

3

parte

Capítulo 20

VERIFICACIÓN DE LAS HIPÓTESIS ESTADÍSTICAS

20.1. Nociones fundamentales y problemas de la estadística matemática

20.1.1. Espacio muestral. La teoría de las probabilidades propone, para resolver diferentes problemas, métodos que exigen el conocimiento de diversas características probabilísticas. Al resolver problemas prácticos estas características no pueden ser conocidas a priori. La estadística matemática nos enseña cómo se determinan las características probabilísticas a base de datos empíricos.

Uno de los más importantes problemas está relacionado con la determinación de la distribución de la magnitud aleatoria según las observaciones de ésta. A este problema general se reducen muchos particulares, tales como la determinación de la probabilidad de uno o varios sucesos, la distribución conjunta de algunas variables, etc.

Considerando la definición de la función de distribución como un problema de estadística matemática hay que tener en cuenta que este término necesita definición. Según entendemos la definición estadística obtendremos distintos problemas de estadística matemática.

La sucesión de observaciones independientes de la magnitud aleatoria es el material de partida para resolver problemas de estadística matemática. Con otras palabras, suponemos que existe un experimento probabilístico cuando se observa la magnitud aleatoria, y tienen lugar n realizaciones independientes de este experimento. Los valores de la magnitud aleatoria que observamos x_1, x_2, \dots, x_n se llama muestra aleatoria; el número de observaciones, volumen de la muestra. El conjunto de todos los vectores $\{x_1, x_2, \dots, x_n\}$ que pueden observarse durante n realizaciones del experimento, forman el espacio muestral. Desde el punto de vista de la teoría de probabilidades la muestra $\{x_1, x_2, \dots, x_n\}$ es una sucesión de magnitudes aleatorias independientes igualmente distribuidas. El problema de estadística surge si la función general de distribución de las magnitudes x_i no es conocida.

El surgimiento de distintos problemas estadísticos (los principales están formulados más abajo) depende de cuál es la clase de posibles funciones de distribución y qué se necesita conocer acerca de la función de distribución

20.1.2. Verificación de una hipótesis simple. La función desconocida de distribución F pertenece a cierta clase de distribuciones \mathfrak{F} . De las consideraciones apriorísticas se puede deducir que $F = F_0 \in \mathfrak{F}$. A base de las observaciones hechas se necesita confirmar o rechazar esta hipótesis. Por ejemplo, \mathfrak{F} es un conjunto de distribuciones normales, F_0 , una distribución normal con media 0 y varianza 1.

20.1.3. Verificación de una hipótesis compuesta. La función desconocida de distribución F pertenece a \mathfrak{F} , $\mathfrak{F}_0 \subset \mathfrak{F}$. Hay que verificar la hipótesis: $F \in \mathfrak{F}_0$. Por ejemplo, la verificación de la hipótesis de que la magnitud con distribución normal tiene media 0. En este caso \mathfrak{F} coincide con la clase de todas las distribuciones normales y \mathfrak{F}_0 , con la clase de distribuciones normales con media 0.

La verificación de las hipótesis simple y compuesta se reduce a la verificación de la hipótesis estadística, para resolverla se utiliza el criterio de aceptación. Este criterio se define por la fijación del dominio crítico G en el espacio muestral. Si la muestra se halla dentro del dominio crítico, la hipótesis se rechaza. La calidad del criterio se determina por la probabilidad de rechazar la hipótesis verdadera. Cuanto menor es esta probabilidad tanto mejor es este criterio. Por otro lado el criterio se caracteriza por las probabilidades de no rechazar (aceptar) la hipótesis falsa (esta probabilidad depende, naturalmente, de cuál es la distribución real). También es conveniente hacer estas probabilidades lo más pequeñas posible.

20.1.4. Estimación del parámetro de distribución. Se supone que la función desconocida de distribución pertenece a alguna familia de distribuciones $F(\theta, x)$ que depende de cierto parámetro $\theta \in \Theta$ donde Θ es el conjunto en una recta o en un espacio euclídeo de dimensión finita. Esto significa que la distribución depende de uno o varios parámetros reales. Así, por ejemplo, la familia de distribuciones normales en la recta depende de dos parámetros reales: del valor medio y de la varianza. Se necesita estimar el parámetro (o varios parámetros reales) a partir de las observaciones. Para construir estimaciones se utilizan las estadísticas, funciones de los valores muestrales. Ejemplos de estadísticas son:

media muestral

$$\bar{x}_n = \frac{1}{n} \sum_{k=1}^n x_k;$$

varianza muestral

$$s_n^2 = \frac{1}{n} \sum_{k=1}^n (x_k - \bar{x}_n)^2$$

(aquí x_1, \dots, x_n es una muestra de volumen n). A título de estimación del parámetro real θ se utiliza cierta estadística $\hat{\theta}_n(x_1, x_2, \dots, x_n)$ que se considera como valor aproximado del parámetro desconocido. La calidad de la estimación se determina por la distribución de la magnitud

$$\hat{\theta}_n(x_1, x_2, \dots, x_n) - \theta$$

(es evidente que esta distribución depende del valor del parámetro buscado). La estimación será buena si esta distribución está bastante

concentrada cerca del cero. En la práctica suelen limitarse sólo con los dos primeros momentos de estimación. En este caso la estimación se caracteriza por el desplazamiento

$$M_0 \hat{\theta}_n(x_1, x_2, \dots, x_n) - \theta$$

y la varianza

$$M_0 [\hat{\theta}_n(x_1, x_2, \dots, x_n) - M_0 \hat{\theta}_n(x_1, x_2, \dots, x_n)]^2$$

(M_0 es la esperanza matemática suponiendo que la distribución real coincide con $F(\theta, x)$). Las estimaciones para las cuales el desplazamiento es igual a cero se llaman insesgadas. La calidad de las estimaciones insesgadas se determina por el valor de la varianza: cuanto menor es ésta, tanto mejor es la estimación.

20.1.5. Estimación confidencial de los parámetros. El planteamiento del problema es el mismo que en el punto antecedente. Sin embargo, en vez de la estimación del parámetro se construye el dominio confidencial para el parámetro, o sea, un dominio S en Θ tal para el cual la probabilidad de que S contenga el valor real del parámetro no sea menor que $1 - \alpha$ (este número, que se llama nivel de confianza, debe ser bastante próximo a 1). El dominio confidencial se construye respecto de los valores muestrales. Está dado por una función determinada en el espacio muestral cuyos valores son dominios en Θ . Para estimar uno de los parámetros reales se utilizan intervalos confidentiales dados por dos estadísticas que determinan los extremos del intervalo. Sea $S(x_1, x_2, \dots, x_n) \in \Theta$ un dominio confidencial. La calidad del dominio confidencial se caracteriza por el nivel de confianza, así como por la forma y dimensiones del dominio.

20.1.6. Problemas generales de soluciones estadísticas. Como regla la definición de la función de distribución del parámetro desconocido, la aceptación de una u otra hipótesis es parte integrante de cierto problema más general que consiste en tomar alguna solución. Por ejemplo, la aceptación de cierta hipótesis puede ser tal solución. Otro ejemplo más complejo: al seleccionar un nuevo tipo de cultivo es preciso, en cada etapa del experimento, aceptar determinada solución referente a cómo realizar la selección de las semillas y después, tomar la solución definitiva de que el cultivo satisface las exigencias necesarias. La base para aceptar una u otra solución es el material estadístico en observación (en el ejemplo citado, los datos acerca de unas u otras propiedades de las semillas obtenidas en las parcelas experimentales). Cada problema de este tipo tiene un conjunto de posibles soluciones D . La regla para aceptar la solución se da por la función $d(x_1, x_2, \dots, x_n)$ en el espacio muestral, la cual toma valores de D y se llama función resolutive. Se supone que las distribuciones probables de la muestra pertenecen a cierto conjunto de distribuciones \mathcal{P} . Con el fin de estimar la calidad de la regla para tomar la solución se utiliza cierta función de pérdida $W(d, P)$ que determina nuestra pérdida después de tomar la solución d si la distribución real de la muestra es $P \in \mathcal{P}$ (la pérdida puede ser negativa). Es natural buscar tales reglas para las cuales la pérdida media es la mínima.

20.1.7. Análisis sucesivo. Entre las reglas para tomar soluciones juegan un papel muy importante las reglas de sucesión. En tal caso, es conveniente considerar un espacio muestral de dimensión infinita, ya que al resolver se utilizan muestras de volumen tan grande como se

quiere. La regla de sucesión indica para qué n , según la muestra (x_1, x_2, \dots, x_n) , se debe cesar la observación y qué solución se toma en este caso. Así, pues, durante la sucesiva distinción de dos hipótesis las magnitudes x_1, x_2, \dots se observan sucesivamente y para cada $n = 1, 2, \dots$ se adopta una de las soluciones: d_1 significa que se acepta la primera hipótesis; d_2 , la segunda; d_3 , es necesario realizar una observación más (es decir, añadir a la muestra x_1, x_2, \dots, x_n la observación x_{n+1}).

La regla de sucesión para tomar soluciones puede ser descrita por dos sucesiones de funciones. Sea $e_n(x_1, x_2, \dots, x_n) = 1$, si las observaciones cesan en el n -ésimo paso, y $e_n(x_1, x_2, \dots, x_n) = 0$ en el caso contrario; $d_n(x_1, x_2, \dots, x_n)$ es la solución que debe tomarse en el n -ésimo paso; ésta es la función que se determina en el subconjunto del espacio muestral donde $e_n = 1$ y que toma valores de D . Entre las reglas se escoge la que minimiza alguna pérdida, tomando en consideración el número de observaciones necesarias según la regla.

20.1.8. Estadística de sucesos aleatorios. Para los procesos aleatorios se resuelvan los mismos problemas que para las observaciones independientes: verificación de las hipótesis y estimación de los parámetros de las distribuciones. La particularidad de los problemas estadísticos de los procesos aleatorios consiste en que suele observarse una sola trayectoria del proceso aleatorio y a base de esta observación se toman las soluciones estadísticas. De este modo, la estadística de los procesos aleatorios es la de una observación (independiente). Sin embargo, éstas son observaciones no de una magnitud aleatoria sino de un número infinito de ellas (de valores del proceso en distintos momentos de tiempo) enlazadas entre sí. Por eso la estadística de los procesos aleatorios se puede caracterizar como la estadística de observaciones enlazadas. Señalemos algunas propiedades de la estadística de los procesos aleatorios. Primero, los parámetros de los cuales dependen las distribuciones son frecuentemente de dimensión infinita (así, la familia de distribuciones de los procesos gaussianos depende de dos parámetros funcionales: del valor medio y de la función de correlación). Segundo, a pesar de que tenemos sólo una observación se puede de una manera cierta escoger una de las hipótesis o determinar con absoluta exactitud el valor del parámetro. En los problemas clásicos para las distribuciones regulares no hay tal efecto.

Ya que, en general, las distribuciones de los procesos aleatorios no pueden ser dadas de modo efectivo, no siempre es posible resolver problemas estadísticos de una manera constructiva utilizando todas las distribuciones de dimensión finita. En los casos elementales sólo se emplean funciones de momentos hasta cierto orden. Adquirió un desarrollo muy amplio la siguiente tendencia.

20.1.9. Métodos lineales en la estadística matemática. Estos métodos reúnen problemas estadísticos en los cuales se utilizan funciones de momentos de los dos primeros órdenes. Estos problemas, como regla, se refieren a la estimación de parámetros de los cuales depende el valor medio o la función de correlación del proceso. En este caso, se emplean sólo estadísticas lineales y cuadráticas, es decir, funcionales lineales y cuadráticas, según la trayectoria en observación del proceso. Los métodos lineales son, de hecho, la aplicación de la teoría de los espacios de Hilbert en los problemas de la estadística matemática.

20.2. Procedimientos de verificación de las hipótesis

20.2.1. Esquema general de la construcción de criterios estadísticos (no randomizados). Uno de los problemas fundamentales de la estadística matemática es la construcción y estudio de las propiedades de los procedimientos estadísticos de verificación de las hipótesis según los resultados dados de las observaciones.

En la estadística matemática la hipótesis estadística afirma: el parámetro desconocido θ de la distribución iulcal de probabilidades P_θ pertenece al subconjunto dado $H \subset \Theta$ del conjunto de los valores probables del parámetro θ . El subconjunto complementario $K = \Theta \setminus H$ se llama **alternativa de la hipótesis H** .

EJEMPLO. Se observa un suceso aleatorio, la probabilidad p de su comienzo es desconocida. Como hipótesis estadística se considera la suposición de que $p \leq p_0$ ($0 < p_0 < 1$).

La hipótesis H se llama **simple** si el conjunto H consta de un solo valor del parámetro θ ; en el caso contrario, H es una hipótesis **compuesta**.

La aplicación práctica de la estadística matemática consiste en que se verifica la correspondencia real de los resultados reales de los experimentos a la hipótesis supuesta. Con este fin se construye el **procedimiento de verificación de la hipótesis (criterio de aceptación)** que permite, a base de las observaciones, aceptar o rechazar la hipótesis dada.

El espacio muestral X se parte en dos subconjuntos disjuntos: X_0 y X_1 . La regla de verificación de la hipótesis se formula así. Si los resultados de las observaciones son $x \in X_0$ se considera que la hipótesis dada H se confirma por los datos empíricos, o sea, la hipótesis H se admite. Si el valor muestral es $x \in X_1$ se afirma que la hipótesis dada H no concuerda con los resultados de las observaciones, o sea, H se rechaza. El conjunto X_0 se llama **dominio de aceptación de la hipótesis**; el conjunto X_1 se llama **dominio crítico**. Por abreviar, el conjunto X_1 a veces se llama **criterio de la hipótesis H** . Ya que el suceso $x \in X_1$ es aleatorio la hipótesis dada se admite o se rechaza después de observar el suceso aleatorio que tiene cierta probabilidad $P_\theta\{x \in X_1\}$ para cada θ prefijado. La aplicación del procedimiento de verificación de la hipótesis entraña errores de dos géneros: **rechazar la hipótesis cuando es verdadera (error de primer género)**; **aceptar la hipótesis cuando es falsa (error de segundo género)**.

Al construir los procedimientos de verificación de las hipótesis es deseable procurar obtener los valores mínimos de los errores de ambos géneros. En la mayoría de los situaciones prácticamente importantes es imposible construir criterios de aceptación con errores de primero y segundo géneros, cuan se quiera pequeños. Las reglas de verificación de las hipótesis tienen sentido estadístico, es decir, al aplicar múltiplemente una regla determinada el por ciento del número de soluciones falsas se expresa por las probabilidades de errores de primero y segundo géneros.

Si los datos experimentales no concuerdan con la hipótesis dada según el criterio escogido, esto significa que los datos muestrales $x \in X_1$. Entonces, con la probabilidad de error de primer género, se observa un suceso aleatorio que contradice la hipótesis. Si la probabilidad de este suceso aleatorio es pequeña esto significa que se observa

prácticamente un suceso imposible. En este caso, la hipótesis dada debe ser rechazada con certeza práctica.

Cuando los datos experimentales concuerdan con la hipótesis supuesta esto no significa que es imposible concordar estos datos con otra hipótesis. Al aplicar criterios estadísticos a base de las observaciones es imposible demostrar una u otra hipótesis. Sólo se puede afirmar que los resultados de las observaciones no contradicen la hipótesis aceptada.

De este modo, las deducciones adoptadas a base de los datos estadísticos se formulan así: *los datos experimentales concuerdan con la hipótesis dada (o le contradicen).*

20.2.2. Función de la potencia de un criterio. La probabilidad de error de primer género $\beta(\theta) = P_{\theta}(X_1)$ considerada como función del parámetro $\theta \in \Theta$ se llama **función de la potencia de un criterio**:

El valor máximo admisible α del error de primer género para el criterio dado se denomina **nivel de significación del criterio**.

$$\sup_{\theta \in H} P_{\theta}(X_1) \leq \alpha.$$

Si para el criterio se cumple la condición $P_{\theta}(X_1) = \alpha$ cuando $\theta \in H$, el dominio crítico X_1 se llama **semejante al espacio muestral**.

Generalmente el nivel de significación α se elige a base de razonamientos prácticos según se trata la hipótesis. Si la hipótesis supuesta es muy verosímil, el nivel de significación se elige bastante pequeño. En este caso, la hipótesis se rechazará con menor probabilidad. Al elegir el nivel de significación es conveniente tomar en consideración el comportamiento de la función de potencia del criterio $\beta(\theta) = P_{\theta}(X_1)$ para los valores alternativos del parámetro $\theta \in K$. La propiedad deseable del criterio es su carácter insesgado que se determina por las condiciones

$$\begin{aligned} P_{\theta}(X_1) &\leq \alpha && \text{para } \theta \in H; \\ P_{\theta}(X_1) &> \alpha && \text{para } \theta \in K. \end{aligned}$$

Puede resultar que para el nivel dado de significación la potencia del criterio $\beta(\theta)$ es muy pequeña si $\theta \in K$, o sea, es grande el error de segundo género del criterio de la hipótesis falsa.

Es natural considerar óptimo el criterio que logra para el nivel dado de significación el valor máximo de la función de potencia del criterio (problema de Neumann — Pearson).

Como regla, el criterio, que es óptimo para el valor dado del parámetro $\theta \in K$, depende de éste.

El criterio X_1^* se llama **uniformemente más potente (U.M.P.)** si para cualquier otro criterio X_1 se cumplen las condiciones:

$$\begin{aligned} P_{\theta}(X_1^*) &\leq P_{\theta}(X_1) && \text{para } \theta \in H; \\ P_{\theta}(X_1^*) &\geq P_{\theta}(X_1) && \text{para } \theta \in K. \end{aligned}$$

El **criterio randomizado** se determina por la función crítica $\varphi(x)$ en el espacio muestral X cuyo valor es la probabilidad de rechazo de la hipótesis H para el valor dado de x de dos resultados de las observaciones. La potencia del criterio randomizado con la fun-

ción crítica $\varphi(x)$ se determina por la correlación

$$\beta(\theta) = M_{\theta} \varphi(x) = \int_{\bar{X}} \varphi(x) P_{\theta}(dx).$$

En particular, cuando $\varphi(x) = 1$ para $x \in X_1$ y $\varphi(x) = 0$ para $x \in X \setminus X_1$, es decir, cuando la función crítica es función característica del conjunto X_1 , el criterio determinado por la función φ resulta ser no randomizado con el dominio crítico X_1 .

Cuando la muestra es simple y aleatoria los datos estadísticos son los resultados de las observaciones de los valores de la magnitud aleatoria ξ en la sucesión de experimentos independientes. En este caso, el espacio muestral X es un espacio euclídeo n -dimensional: $x = (x_1, x_2, \dots, x_n)$, donde x_k ($k = \overline{1, n}$) son magnitudes aleatorias independientes igualmente distribuidas. El número n de elementos de la sucesión muestral se llama volumen de la muestra.

El criterio de aceptación con el dominio crítico X_1 se llama conciliable, si tenemos la correlación

$$\lim_{n \rightarrow \infty} P_{\theta}(X_1) = 1 \quad \text{para } \theta \in K.$$

La conciliabilidad del criterio significa que el error de segundo género —la probabilidad de aceptación de la hipótesis falsa— tiende a cero para $n \rightarrow \infty$.

Para comparar distintos criterios entre sí se utilizan las medidas de eficiencia asintótica de los criterios, que se basan en el análisis de la velocidad de convergencia de la función de potencia en el entorno del parámetro $\theta \in H$.

20.3. Criterios de verificación de las hipótesis estadísticas

20.3.1. Criterio de Neumann—Pearson. Para la hipótesis simple H_0 a la cual le corresponde la distribución $P_0(x)$, en comparación con la alternativa simple K a la cual corresponde la distribución $P_1(x)$, la función crítica $\varphi(x)$ del criterio óptimo de Neumann—Pearson para el nivel dado de significación α se determina por las condiciones:

$$\int_X \varphi(x) p_0(x) dx = \alpha;$$

$$\varphi(x) = \begin{cases} 1, & p_1(x) \geq C_{\alpha} p_0(x); \\ 0, & p_1(x) < C_{\alpha} p_0(x), \end{cases}$$

donde $p_0(x)$ y $p_1(x)$ son las densidades de las distribuciones P_0 y P_1 , a partir de la medida inicial ν . El criterio de Neumann—Pearson es el criterio uniformemente más potente del nivel α .

La aplicación práctica del criterio de Neumann—Pearson consiste en comprobar la desigualdad $p_1(x) \geq C_{\alpha} p_0(x)$ para los resultados dados de las observaciones x . Si se cumple esta desigualdad la hipótesis H_0 se rechaza; en el caso contrario la hipótesis H_0 se acepta.

La constante C_{α} en el criterio de Neumann—Pearson se determina por la cuantila de distribución de la magnitud aleatoria

$T(\xi) = \frac{P_1(\xi)}{P_0(\xi)}$ para la hipótesis H (ξ tiene la distribución P_0):

$$P_0(T(\xi) \geq C_\alpha) = \alpha.$$

Si las probabilidades $P_0(T(\xi) \geq C_\alpha)$ no dependen de los valores alternativos del parámetro θ , el criterio de Neumann—Pearson de verificación de la hipótesis simple H_0 es uniformemente más potente con respecto a todas las alternativas.

Los criterios basados en la utilización de la distribución de estadísticas $T(\xi) = \frac{P_1(\xi)}{P_0(\xi)}$ se llaman **criterios de la relación de verosimilitud**. Tienen muchas propiedades útiles (véase [37]).

EJEMPLO 1. Sean $P(x; a, \sigma)$ las densidades normales de distribución con valores medios a y varianzas σ^2 . Consideremos la hipótesis simple $H_0: a = a_0$ para el valor conocido del parámetro σ en comparación con la clase de alternativas $k: a > a_0$ con el mismo valor del parámetro σ . En la elección aleatoria simple $x = (x_1, x_2, \dots, x_n)$ es un vector n -dimensional de magnitudes aleatorias independientes igualmente distribuidas x_1, x_2, \dots, x_n . En este caso, el criterio de Neumann—Pearson se determina por la desigualdad

$$\exp \left\{ - \sum_{k=1}^n \frac{(x_k - a)^2}{2\sigma^2} \right\} \geq I_\alpha \exp \left\{ - \sum_{k=1}^n \frac{(x_k - a_0)^2}{2\sigma^2} \right\}.$$

Después de la determinación por logaritmos esta desigualdad se reduce a la forma siguiente:

$$(a - a_0) \sum_{k=1}^n x_k \geq I_\alpha.$$

Para las clases de alternativas $a > a_0$ el dominio crítico se da por la desigualdad

$$\sum_{k=1}^n x_k \geq C_\alpha,$$

donde la constante C_α se determina de la condición

$$P_0 \left(\sum_{k=1}^n x_k \geq C_\alpha \right) = \alpha.$$

El criterio $\sum_{k=1}^n x_k \geq C_\alpha$ de la hipótesis simple $H: a = a_0$ es uniformemente más potente para la clase de alternativas $a > a_0$. Análogamente, para la clase de alternativas $a < a_0$ el criterio uniformemente más potente de la hipótesis simple $H_0: a = a_0$ se determina por la desigualdad $\sum_{k=1}^n x_k \leq C_\alpha$.

Si el conjunto de los valores alternativos del parámetro a contiene puntos tanto a la izquierda como a la derecha de a_0 , el criterio uniformemente más potente no existe.

EJEMPLO 2. Para la hipótesis simple $H_0 : a = a_0$, cuando el valor del parámetro σ es desconocido, en comparación con la alternativa $K : a > a_0$ el criterio de Neumann—Pearson se determina por el dominio crítico

$$\frac{\bar{x} - a_0}{s} \geq C_\alpha; \quad \bar{x} = \frac{1}{n} \sum_{k=1}^n x_k; \quad s^2 = \frac{1}{n-1} \sum_{k=1}^n (x_k - \bar{x})^2.$$

EJEMPLO 3. Para la hipótesis simple $H_0 : \sigma = \sigma_0$ en comparación con la alternativa $K : \sigma < \sigma_0$ el criterio uniformemente más potente se determina por la correlación $s^2 \geq C_\alpha$.

20.3.2. Los criterios de la relación de verosimilitud se construyen empleando las propiedades de la función de la relación de verosimilitud $l(x) = \frac{P_\theta(x)}{P_{\theta_0}(x)}$.

Resulta natural la elección del dominio crítico de tal manera que con $x \in X_1$ la relación de verosimilitud adquiera los valores máximos posibles para todos los valores alternativos del parámetro $\theta \in K$.

El criterio de la relación de verosimilitud para la hipótesis simple $H_\theta : \theta = \theta_0$ con respecto a la alternativa compuesta $\theta \in K$ se determina por la estadística

$$l(x) = \frac{\sup_{\theta \in K} P_\theta(x)}{P_{\theta_0}(x)}.$$

El dominio crítico tiene el aspecto de $X_1 = \{x : l(x) \geq C_\alpha\}$, donde la constante C_α se determina por la condición: $P_{\theta_0}(X_1) = \alpha$. Por ejemplo, para la muestra $x = (x_1, x_2, \dots, x_n)$ de la población normal $P(x; a, \sigma)$ con parámetros desconocidos a y σ el criterio de la relación de verosimilitud para la hipótesis $H_0 : a = a_0$ se determina por la estadística

$$l(x) = \left(1 + \frac{t^2(x)}{n-1}\right)^{-n/2},$$

donde $t(x) = \sqrt{n}(\bar{x} - a_0)/s$ tiene distribución de Student con $n-1$

grados de libertad. (Aquí $\bar{x} = \frac{1}{n} \sum_{k=1}^n x_k$, $s^2 = \frac{1}{n-1} \sum_{k=1}^n (x_k - \bar{x})^2$).

El criterio de la relación de verosimilitud de la hipótesis compuesta $\theta \in H$ con respecto a la alternativa compuesta $\theta \in K$ se determina por la estadística

$$l(x) = \frac{\sup_{\theta \in K} P_\theta(x)}{\sup_{\theta_0 \in H} P_{\theta_0}(x)}.$$

20.3.3. Criterio χ^2 . Para verificar la hipótesis simple H_θ , a la cual corresponde la distribución discreta (p_1, p_2, \dots, p_m) , $\sum_{k=1}^m p_k =$

= 1 se emplea la estadística

$$\chi^2 = \sum_{h=1}^m \frac{(v_h - np_h)^2}{np_h}$$

donde v_1, v_2, \dots, v_m son las frecuencias de los resultados de las observaciones en la muestra de volumen $n = \sum_{h=1}^m v_h$. Para $n \rightarrow \infty$ la distribución de la estadística χ^2 tiende a la distribución χ^2 con m grados de libertad y densidad de probabilidad

$$K_{m-1}(x) = \frac{1}{2^{\frac{m-1}{2}} \Gamma\left(\frac{m-1}{2}\right)} x^{\frac{m-2}{2}} e^{-\frac{x}{2}}, \quad x > 0.$$

La distribución límite no depende del tipo de la distribución discreta inicial. El criterio χ^2 se define del modo siguiente. Sea χ^2_α la cuantila de la distribución límite $\chi^2: \alpha = P(\chi^2 > \chi^2_\alpha)$. Entonces el dominio crítico del criterio χ^2 se determina por la desigualdad para la estadística $\chi^2: \chi^2 > \chi^2_\alpha$.

El criterio χ^2 se emplea también para verificar una hipótesis simple referente a la distribución inicial cuando se agrupan los valores de la magnitud aleatoria ξ que se observa. En este caso, $p_h = P\{\xi \in X_h\}$ donde X_h son grupos en los cuales está partido el conjunto de los valores probables de la magnitud aleatoria ξ . En la práctica el criterio χ^2 es bastante efectivo cuando todas las frecuencias esperadas $np_h \geq 10$.

EJEMPLO 4. En la sucesión de los experimentos independientes se observa un suceso aleatorio cuya probabilidad p ($0 < p < 1$) es desconocida. Sea v_n la frecuencia de las observaciones del suceso en la muestra de volumen n . Entonces, para n grandes la estadística $\chi^2 = \frac{(v - np)^2}{np(1-p)}$ está distribuida aproximadamente con una densidad

$$K_1(x) = \frac{1}{\sqrt{2\pi x}} e^{-\frac{x}{2}}, \quad x > 0.$$

Las tablas existentes de cuantilas de la distribución χ^2 (véase [37]) dan la posibilidad de verificar la hipótesis acerca de que en la serie dada de observaciones la probabilidad del suceso aleatorio es igual al número dado p .

El criterio χ^2 se emplea también cuando tenemos varias series independientes de observaciones.

EJEMPLO 5. Sean v_1, v_2, \dots, v_m las frecuencias de las observaciones en las muestras de volumen n_1, n_2, \dots, n_m del suceso aleatorio cuya probabilidad se supone igual a p .

Para verificar la hipótesis supuesta se puede utilizar la estadística $\chi^2 = \sum_{h=1}^m \frac{(v_h - d_h p)^2}{n_h p(1-p)}$ cuya distribución para grandes

$n = \sum_{h=1}^m n_h$ es próxima a la distribución χ^2 con m grados de libertad.

Si la distribución inicial depende de los parámetros desconocidos $\theta_1, \theta_2, \dots, \theta_r$, entonces, a título de criterio de aceptación, se emplea la estadística

$$\hat{\chi}^2 = \sum_{h=1}^m \frac{[v_h - n p_h(\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_r)]^2}{n p_h(\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_r)},$$

en la cual $\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_r$ son estimaciones de los parámetros desconocidos construidos según los resultados de las observaciones. A condiciones determinadas la estadística $\hat{\chi}^2$ en límite para $n \rightarrow \infty$ tiene la distribución χ^2 con $m - r - 1$ grados de libertad.

EJEMPLO 6. Valiéndonos del ejemplo 5, utilicemos la estimación de la probabilidad desconocida $\hat{p} = \frac{1}{n} \sum_{h=1}^m v_h$.

La estadística $\chi^2 = \sum_{h=1}^m \frac{(v_h - n_h \hat{p})^2}{n_h \hat{p} (1 - \hat{p})}$ con $m - 1$ grados de libertad

se utiliza para verificar la hipótesis de la homogeneidad de las muestras: la suposición de que en todas las m series de las observaciones independientes la probabilidad del suceso en observación queda invariable.

20.3.4. Criterios no paramétricos. Uno de los principales problemas estadísticos no paramétricos es el problema de verificación de la concordancia de los datos muestrales con la hipótesis de que la función inicial de distribución $F(x)$ está prefijada y no contiene parámetros desconocidos. Si la función inicial de distribución $F(x)$ es continua, entonces, a título de criterio no paramétrico, se emplea la estadística de Kolmogórov

$$\sqrt{n} D_n = \sqrt{n} \sup_{-\infty < x < +\infty} |F_n(x) - F(x)|,$$

cuya distribución no depende del aspecto de $F(x)$ y para la cual es conocida la distribución límite (véase el punto 20.4).

El dominio crítico del criterio para el nivel de significación dado se determina por la desigualdad

$$\sqrt{n} D_n > d_\alpha,$$

donde d_α es la cuantila de la distribución límite de Kolmogórov: $1 - K(d_\alpha) = \alpha$.

El problema de verificación de la homogeneidad de los datos estadísticos se plantea del modo siguiente. A partir de dos series de observaciones independientes x_1, x_2, \dots, x_n e y_1, y_2, \dots, y_n hay que verificar la hipótesis referente a que los resultados de las observaciones en ambas series fueron obtenidos como resultado de los experimentos con magnitudes aleatorias con la misma función de distribución $F(x)$. Si la función inicial de distribución $F(x)$ es continua, entonces,

a título del criterio de homogeneidad de los datos muestrales, se emplea la estadística de Smirnov cuya distribución no depende del aspecto de la función $F(x)$:

$$\sqrt{\frac{nm}{n+m}} D_{nm} = \sqrt{\frac{nm}{n+m}} \max_{-\infty < x < +\infty} |F_n(x) - F_m(x)|,$$

donde $F_n(x)$ y $F_m(x)$ son funciones empíricas de distribución de la primera y la segunda series de observaciones, respectivamente. El dominio crítico del criterio de homogeneidad se determina por la desigualdad

$$\sqrt{\frac{nm}{n+m}} D_{nm} \geq d_\alpha,$$

donde d_α es la cuantila de distribución de la estadística de Smirnov para n y m pequeñas, en tanto que si n y m son grandes, d_α es la cuantila de la distribución límite de la estadística de Smirnov: $1 - K(d_\alpha) = \alpha$. En caso particular, cuando los volúmenes de las muestras son iguales ($n = m$) la distribución exacta de la estadística de Smirnov tiene una expresión analítica bastante simple.

Los criterios no paramétricos de orden se construyen a base de las estadísticas de la serie variacional, las cuales no dependen de los valores concretos de los términos de la serie variacional.

El criterio de las series de Wald-Wolfowitz está basado en la estadística U_{nm} , es decir, en el número de series de los valores observados de la primera y la segunda muestras en la serie variacional general $x_1 \leq x_2 \leq \dots \leq x_{n+m}$, donde cada x_h es x_{ih} o y_{jh} .

La distribución de la estadística U_{nm} no depende del aspecto de $F(x)$ de la distribución inicial

$$U_{nm} = \sum_{h=1}^{n+m} v_h,$$

donde las magnitudes aleatorias v_h son independientes y toman valores 1 ó 0 con probabilidades iguales a 1/2. Si n y m son grandes la distribución de la estadística U_{nm} es asintóticamente normal con media $\frac{nm}{2}$ y variancia $\frac{nm}{2}(n+m+1)$. Para m

fijado y $n \rightarrow \infty$ la estadística $\frac{1}{n} U_{nm}$ en límite está distribuida como

$U_m = \sum_{t=1}^m W_t$, donde W_t son independientes, uniformemente distribuidas en el intervalo (0, 1).

Hay otros criterios no paramétricos (véase [37]).

20.3.5. Criterio sucesivo de la relación de verosimilitud. En la práctica los experimentos se realizan sucesivamente. En cada etapa hay posibilidad de decidir si es necesario continuar o cesar los experimentos. El volumen de la muestra durante el análisis estadístico sucesivo no se fija de antemano y es una magnitud aleatoria. Sean H_0 y H_1 dos hipótesis alternativas sobre el aspecto de la densidad de distribución $p_0(x)$ o $p_1(x)$ de la magnitud aleatoria en observación. El

criterio sucesivo de la relación de verosimilitud (de Wald) se construya según los resultados de las observaciones independientes $x_1, x_2, \dots, x_n, \dots$ de modo siguiente. Se fijan dos constantes A y B que determinan la partición del espacio muestral según la relación de verosimilitud

$$\prod_{h=1}^n \frac{p_1(x_h)}{p_0(x_h)}$$

para cada n en tres dominios:

$$X_1: \prod_{h=1}^n \frac{p_1(x_h)}{p_0(x_h)} \geq B \text{ es el dominio de aceptación de la hipótesis } H_1;$$

$$X_0: \prod_{h=1}^n \frac{p_1(x_h)}{p_0(x_h)} \leq A \text{ es el dominio de aceptación de la hipótesis } H_0;$$

$$X_{01}: A < \prod_{h=1}^n \frac{p_1(x_h)}{p_0(x_h)} < B \text{ es el dominio de continuación de los experimentos.}$$

La calidad de criterio sucesivo de la relación de verosimilitud se determina por los errores de primer género $\alpha = P_0(X_1)$ y de segundo género $\beta = P_1(X_0)$, así como por el número medio de observaciones M_{0v} y M_{1v} ; el número aleatorio de observaciones v se determina por las condiciones:

$$\prod_{h=1}^{v-1} \frac{p_1(x_h)}{p_0(x_h)} \in (A, B), \quad \prod_{h=1}^v \frac{p_1(x_h)}{p_0(x_h)} \notin (A, B).$$

El criterio sucesivo de la relación de verosimilitud exige, por término medio, menor número de observaciones que el criterio con volumen fijo de la muestra con los mismos errores de primero y segundo géneros.

El criterio sucesivo se termina con probabilidad 1 tanto para la hipótesis H_0 , como para la hipótesis H_1 , es decir, $P_0(v < \infty) = P_1(v < \infty) = 1$.

La determinación exacta de las fronteras A y B y del volumen medio de la muestra M_{0v} en el criterio sucesivo está vinculada con grandes dificultades. Sin embargo, hay desigualdades útiles para las aplicaciones:

$$A \geq \frac{\beta}{1-\alpha}, \quad B \leq \frac{1-\beta}{\alpha};$$

$$M_{0v} \leq \frac{(1-\alpha) \log B + \alpha \log A}{M_{0z}};$$

$$M_{0v} \geq \frac{(1-\alpha) \log \frac{B}{1-\alpha} + \alpha \log \frac{1-\beta}{\alpha}}{M_{0z}}.$$

$$\text{Aquí } z = \log \frac{p_1(\xi)}{p_0(\xi)}.$$

20.4. Distribución de la muestra

20.4.1. **Serie variacional.** La muestra de volumen finito n es el material inicial para el análisis estadístico, obtenido como resultado de la elección aleatoria simple de una población madre, que se define por la magnitud aleatoria ξ con la función de distribución $P(x)$:

$$x_1, x_2, \dots, x_n. \quad (4.1)$$

es decir, una sucesión de magnitudes aleatorias independientes igualmente distribuidas x_k , $1 \leq k \leq n$, con la función general de distribución $P(x)$.

La sucesión de valores muestrales ordenada según las magnitudes

$$x_1^{(n)} \leq x_2^{(n)} \leq \dots \leq x_n^{(n)} \quad (4.2)$$

se llama **serie variacional**. Los elementos de la muestra iguales entre sí se numeran en orden arbitrario.

Los términos de la serie variacional $x_m^{(n)}$ ($m = 1, 2, \dots, n$) se llaman **estadísticas de orden (de rango)**. El número $\lambda_m = \frac{m}{n}$ se llama **rango** del término $x_m^{(n)}$.

La estadística $v_n(x)$ igual al número de valores de la muestra menores que x

$$(v_n(x) = m) = \{x_m^{(n)} < x \leq x_{m-1}^{(n)}\}, \quad m = 0, 1, \dots, n \quad (4.3)$$

se llama **frecuencia empírica**. La magnitud aleatoria $v_n(x)$ es igual al número de apariciones del suceso $\{\xi < x\}$ en n experimentos independientes, así que la frecuencia empírica $v_n(x)$ tiene distribución binomial con el parámetro $p = P\{\xi < x\} = P(x)$:

$$P(v_n(x) = m) = C_n^m P^m (1 - P(x))^{n-m}, \quad m = 0, 1, \dots, n. \quad (4.4)$$

La distribución de los términos de la serie variacional (estadísticas de orden) se determina simplemente según la distribución de la frecuencia empírica:

$$P(x_m^{(n)} < x) = P(v_n(x) \geq m) = \sum_{k=m}^n C_n^k P^k (1 - P(x))^{n-k}. \quad (4.5)$$

En particular, un aspecto muy simple lo tienen las distribuciones de los términos extremos de la serie variacional $x_1^{(n)}$ y $x_n^{(n)}$:

$$\left. \begin{aligned} P(x_1^{(n)} < x) &= 1 - (1 - P(x))^n; \\ P(x_n^{(n)} < x) &= P^n(x). \end{aligned} \right\} \quad (4.6)$$

Se puede representar la distribución de los términos de la serie variacional de otro modo más cómodo para el análisis:

$$P(x_m^{(n)} < x) = \frac{n!}{(m-1)!(n-m)!} \int_0^{P(x)} y^{m-1} (1-y)^{n-m} dy. \quad (4.7)$$

La distribución (4.7) pertenece al tipo de distribuciones β .

Si la distribución inicial de la población madre $P(x)$ tiene la densidad $p(x) = \frac{dP}{dx}$, entonces la distribución de las estadísticas de orden tiene la densidad en forma

$$\frac{d}{dx} P(x_m^{(n)} < x) = \frac{n!}{(m-1)!(n-m)!} P^{m-1}(x) (1-P(x))^{n-m} p(x). \quad (4.8)$$

En los problemas del control estadístico de la calidad de la producción se emplea a menudo la estadística $R_n = x_n^{(n)} - x_1^{(n)}$ que se llama recorrido de la muestra. La distribución del recorrido tiene el aspecto

$$P(x_n^{(n)} - x_1^{(n)} < t) = n \int_{-\infty}^{\infty} [P(x+t) - P(x)]^{n-1} dP(x). \quad (4.9)$$

Es difícil emplear las distribuciones exactas de las estadísticas de orden en el análisis estadístico, ya que dependen considerablemente de la distribución inicial. Es natural esperar que al crecer indefinidamente el volumen de la muestra n esta dependencia debe disminuir. La representación aproximada de las estadísticas de orden para n grandes se llama representación asintótica. La transformación de los términos de la serie variacional según la fórmula

$$z_m^{(n)} = nP(x_m^{(n)}) \quad (4.10)$$

da la densidad de la distribución $g_m^{(n)}(z)$ de las magnitudes aleatorias $z_m^{(n)}$ en la forma

$$g_m^{(n)}(z) = C_{n-1}^{m-1} \left(\frac{z}{n}\right)^{m-1} \left(1 - \frac{z}{n}\right)^{n-m}. \quad (4.11)$$

Para z fijadas esta distribución de Bernoulli (según m), si $n \rightarrow \infty$, se aproxima por la distribución de Poisson

$$g_m^{(n)}(z) \approx \frac{z^{m-1} e^{-z}}{(m-1)!}, \quad z > 0 \quad (4.12)$$

que según z es la densidad de la distribución gamma con parámetro $m-1$.

Análogamente la transformación de los términos extremos de la serie variacional según la fórmula $z_{n-m}^{(n)} = n(1 - P(x_{n-m}^{(n)}))$ nos ofrece la densidad de la distribución $g_{n-m}^{(n)}(z)$ de las magnitudes aleatorias $z_{n-m}^{(n)}$ en la forma

$$g_{n-m}^{(n)}(z) = C_{n-1}^m \left(\frac{z}{n}\right)^m \left(1 - \frac{z}{n}\right)^{n-m-1}. \quad (4.13)$$

Para $n \rightarrow \infty$ y m y z fijados $g_{n-m}^{(n)}(z)$ se aproxima por la densidad de distribución gamma con parámetro m :

$$g_{n-m}^{(n)}(z) \approx \frac{z^m e^{-z}}{m!}, \quad z > 0. \quad (4.14)$$

Las distribuciones $x_i^{(n)}$ y $x_{n-m}^{(n)}$ se pueden emplear para hallar la representación asintótica de los términos extremos de la serie variacional para m fijados.

EJEMPLO 1. La distribución inicial $P(x)$ es uniforme en el segmento $[-a, a]$. Entonces, para el término mínimo $x_i^{(n)}$ y el término máximo $x_n^{(n)}$ de la serie variacional existen las representaciones asintóticas

$$\left. \begin{aligned} x_i^{(n)} &\simeq -a + \frac{2a}{n} z; \\ x_n^{(n)} &\simeq a - \frac{2a}{n} z, \end{aligned} \right\} \quad (4.15)$$

donde z es una magnitud aleatoria con distribución gamma y parámetro $m = 0$, es decir, con densidad $g(z) = e^{-z}$, $z > 0$.

EJEMPLO 2. La distribución $P(x)$ tiene densidad de la forma $p(x) = \frac{1}{2}e^{-|x|}$. Entonces, los términos mínimo y máximo de la serie variacional tienen las representaciones asintóticas en la forma

$$\left. \begin{aligned} x_i^{(n)} &\simeq v - \ln \frac{n}{2}; \\ x_n^{(n)} &\simeq -v + \ln \frac{n}{2}; \end{aligned} \right\} \quad (4.16)$$

donde v es una magnitud aleatoria con densidad de distribución $g(v) = e^v - e^{-v}$.

En el caso del rango finito, cuando $m = [nq]$, es decir, para m enteros que satisfacen las desigualdades $0 < \frac{m}{n} \leq q < \frac{m+1}{n} \leq 1$, la transformación de los términos de la serie variacional con rango finito $x_q^{(n)} = P(x_{[nq]}^{(n)})$ nos da la densidad $g_q(z)$ de la distribución $x_q^{(n)}$ en la forma

$$g_q(z) = C_{n-1}^{m-1} z^{m-1} (1-z)^{n-m},$$

que en el entorno del punto $z = q$ se aproxima por la densidad de distribución normal con parámetros $(q, \sqrt{\frac{q(1-q)}{n}})$, donde q es el valor medio, $\sqrt{\frac{q(1-q)}{n}}$ es la desviación típica, lo que permite representar asintóticamente la magnitud aleatoria z para $n \rightarrow \infty$ en la forma

$$x_q^{(n)} \simeq q + u \sqrt{\frac{q(1-q)}{n}},$$

donde u tiene distribución normal con parámetros $(0, 1)$. En este caso, los términos medios de la serie variacional $x_{[nq]}^{(n)}$ pueden ser representados asintóticamente en forma de

$$x_{[nq]}^{(n)} \simeq x_q + \frac{u}{p(x_q)} \sqrt{\frac{q(1-q)}{n}}$$

donde x_q es cuantila de orden q de la distribución $P(x)$, es decir, $P(x_q) = q$, $p(x)$ es la densidad de la distribución inicial.

20.4.2. Función empírica de distribución. Anteriormente fue deducida la frecuencia empírica $v_n(x)$ igual al número de valores muestrales menores que x y que tiene una distribución binomial con parámetro $P(x)$.

Se llama función empírica de distribución la función $\bar{P}_n(x)$ determinada por la relación

$$\bar{P}_n(x) = \frac{v_n(x)}{n}. \quad (4.17)$$

De otro modo, la función empírica de distribución se determina por la siguiente relación:

$$\bar{P}_n(x) = \begin{cases} 0, & x \leq x_1^{(n)}, \\ \frac{m}{n}, & x_m^{(n)} < x \leq x_{m+1}^{(n)}, \\ 1, & x > x_n^{(n)}. \end{cases} \quad (4.18)$$

El gráfico de la función empírica de distribución es una línea escalonada con saltos múltiples a la magnitud $1/n$ en los puntos determinados por los términos de la serie variacional $x_1^{(n)} \leq x_2^{(n)} \leq \dots \leq x_n^{(n)}$. La función empírica de distribución tiene todas las propiedades de la distribución de probabilidades. La función empírica de distribución se llama también distribución de la muestra.

Para x fijado la esperanza matemática $\bar{P}_n(x)$ es igual a $P(x)$: $M\bar{P}_n(x) = P(x)$. Por consiguiente, según la ley de los grandes números cuando $n \rightarrow \infty$ para cada x la función empírica de distribución converge, con respecto a la probabilidad, a la distribución teórica inicial $P(x)$. Además, tiene lugar el

Teorema de Glivenko. La función empírica de distribución $P_n(x)$ converge uniformemente según x con probabilidad 1, para $n \rightarrow \infty$, a la distribución teórica $P(x)$:

$$P \left\{ \lim_{n \rightarrow \infty} \sup_{-\infty < x < +\infty} |F_n(x) - F(x)| = 0 \right\} = 1.$$

A título de la medida probable de la desviación de la función empírica de distribución $\bar{P}_n(x)$ con respecto a la teórica $P(x)$ para x fijado se puede hacer uso de que la diferencia $\bar{P}_n(x) - P(x)$ es asintóticamente normal con media 0 y varianza $\frac{P(x)(1-P(x))}{n}$ (x). Pero esta medida de desviación no es uniforme según x . Un papel importante en la estadística matemática lo desempeñó el análisis de la estadística introducida por A. N. Kolmogórov:

$$D_n = \sup_{-\infty < x < +\infty} |P_n(x) - P(x)|.$$

Teorema de Kolmogórov. Si la función de distribución $P(x)$ es continua, entonces

$$\lim_{n \rightarrow \infty} P\left\{\frac{1}{n} \sup_{-\infty < x < +\infty} |\bar{P}_n(x) - P(x)| < z\right\} = K(z) = \sum_{k=-\infty}^{+\infty} (-1)^k e^{-2k^2 z^2}, \quad z > 0. \quad (4.19)$$

La función $K(z)$ está tabulada [8]. La estadística D_n se emplea en el criterio no paramétrico de aceptación (véase el punto 20.3.4.) de los resultados empíricos con hipótesis sobre la distribución teórica inicial. Señalemos que la distribución de la estadística D_n no depende de la forma de la distribución continua inicial $P(x)$.

El cálculo práctico de la estadística D_n no cuesta mucho trabajo, ya que la desviación máxima de la función empírica de distribución con respecto a la función teórica continua de distribución se logra en los puntos de saltos $\bar{P}_n(x)$, así que

$$D_n = \max_{1 \leq m \leq n} \left\{ \left| \frac{m-1}{n} - P(x_m^{(n)}) \right|, \left| \frac{m}{n} - P(x_n^{(n)}) \right| \right\} \quad (4.20)$$

o de otro modo

$$D_n = \max_{1 \leq m \leq n} \left\{ \left[\frac{2m-1}{2n} - P(x_m^{(n)}) \right] + \frac{1}{2n} \right\}. \quad (4.21)$$

En los criterios unilaterales de aceptación se puede emplear la distribución de las estadísticas de Smirnov:

$$D_n^+ = \sup_{-\infty < x < +\infty} [\bar{P}_n(x) - P(x)] \quad (4.22)$$

o

$$D_n^- = - \inf_{-\infty < x < +\infty} [\bar{P}_n(x) - P(x)]. \quad (4.22')$$

Estas estadísticas tienen distribuciones iguales:

$$P\{D_n^+ \geq x\} = P\{D_n^- \geq x\} = \sum_{k=0}^{[n(1-x)]} C_n^k x \left(x + \frac{k}{n}\right)^{k-1} \left(1 - x - \frac{k}{n}\right)^{n-k}, \quad 0 < x < 1. \quad (4.23)$$

Teorema de Smirnov. Si la distribución $P(x)$ es continua, entonces

$$\lim_{n \rightarrow \infty} P\{\sqrt{n} D_n^+ < z\} = 1 - e^{-2z^2}, \quad z > 0. \quad (4.24)$$

Es conocida también la distribución límite conjunta de las estadísticas D_n^+ y D_n^- .

Teorema. Si la distribución $P(x)$ es continua, entonces

$$\lim_{n \rightarrow \infty} P \left\{ \sqrt{\frac{n}{2}} D_n^* < z, \sqrt{\frac{n}{2}} D_n^- < v \right\} = \\ = 1 + 2 \sum_{h=1}^{\infty} e^{-2h^2(z+v)^2} - \sum_{h=1}^{\infty} [e^{-2(hv+(h-1)z)^2} + e^{-2((h-1)v+hz)^2}].$$

Hay también desarrollos asintóticos para las distribuciones de las estadísticas D_n , D_n^* y D_n^- .

20.5. Distribución de las características muestrales

Se llaman muestrales (o empíricas) las características de la función empírica de distribución.

Las características muestrales son magnitudes aleatorias, funciones a partir de los valores muestrales, es decir, estadísticas representables en forma de

$$\bar{t}(x_1, x_2, \dots, x_n) = \int_{-\infty}^{\infty} t(x) d\bar{P}(x). \quad (5.1)$$

Ya que la función empírica de distribución $\bar{P}(x)$ sirve de estimación de la distribución inicial $P(x)$ (véase 20.5.2) se debe esperar que las características muestrales también pueden servir de estimaciones de las correspondientes características de la distribución inicial, lo que explica la importancia del estudio de las distribuciones de las estadísticas muestrales y de sus características numéricas.

Convengamos que en adelante las características numéricas de las estadísticas muestrales (es decir, de la función empírica $\bar{P}(x)$) se designen con la misma letra que las características numéricas correspondientes de la población madre (es decir, de la distribución inicial $P(x)$), solamente con una raya encima.

20.5.1. Momentos muestrales (estadísticos). El valor medio de la muestra se determina por la relación

$$\bar{x} = \frac{1}{n} \sum_{h=1}^n x_h = \int_{-\infty}^{\infty} x d\bar{P}(x). \quad (5.2)$$

Las características numéricas del valor medio se calculan fácilmente tomando en consideración que \bar{x} es la suma de las magnitudes aleatorias independientes igualmente distribuidas. Por ejemplo,

$$\left. \begin{aligned} M\bar{x} &= m; \\ D\bar{x} &= \frac{\sigma^2}{n}, \end{aligned} \right\} \quad (5.3)$$

donde m es el valor medio; σ^2 , la varianza de la población madre

$$m = \int_{-\infty}^{\infty} x dP(x), \quad \sigma^2 = \int_{-\infty}^{\infty} (x-m)^2 dP(x).$$

De las fórmulas (5.3) se deduce de inmediato que la media muestral \bar{x} converge según la probabilidad al valor medio de m de la población madre para $n \rightarrow \infty$. Además, la magnitud de desviación de la media muestral con respecto a su esperanza matemática $\sqrt{n}(\bar{x}-m)$ es asintóticamente normal con parámetros $(0, \sigma^2)$.

La varianza muestral (estadística) se determina, corrientemente, por la relación

$$\bar{s}^2 = \frac{1}{n} \sum_{h=1}^n (x_h - \bar{x})^2. \quad (5.4)$$

Las principales características numéricas de la varianza muestral tienen el aspecto

$$\left. \begin{aligned} M\bar{s}^2 &= \left(1 - \frac{1}{n}\right) \sigma^2; \\ D\bar{s}^2 &= \frac{m_4 - m_2^2}{n} - \frac{2(m_4 - 2m_2^2)}{n^2} + \frac{m_4 - 3m_2^2}{n^3}. \end{aligned} \right\} \quad (5.5)$$

Aquí a m_2 y m_4 les corresponden el segundo y el cuarto momentos centrales de la población madre, es decir,

$$m_2 = \sigma^2, \quad m_4 = \int_{-\infty}^{\infty} (x-m)^4 dP(x).$$

De la primera fórmula (5.5) se deduce que la estadística

$$\frac{n}{n-1} \bar{s}^2 = \frac{1}{n-1} \sum_{h=1}^n (x_h - \bar{x})^2$$

es la estimación insesgada de la varianza. Al mismo tiempo la magnitud de la desviación de la varianza muestral con respecto a la varianza de la población madre $\sqrt{n}(\bar{s} - \sigma^2)$ es asintótica normal con parámetros $(0, m_4 - m_2^2)$.

Los momentos muestrales superiores (centrales) se determinan por la relación

$$\bar{m}_r = \frac{1}{n} \sum_{h=1}^n (x_h - \bar{x})^r, \quad r \geq 2. \quad (5.6)$$

Es fácil calcular las características numéricas de los momentos muestrales superiores. Se debe emplear la evidente correlación $Mx_h^r = a_r =$

$= \int_{-\infty}^{\infty} x^r dP(x)$ y la independencia de las magnitudes aleatorias x_k .

Las expresiones exactas de las características numéricas de los momentos muestrales \bar{m}_r para $r \geq 3$ son voluminosas, pero, sin embargo, si n son grandes, las expresiones asintóticas se simplifican considerablemente:

$$M\bar{m}_r = m_r + O\left(\frac{1}{n}\right); \quad (5.7)$$

$$D\bar{m}_r = \frac{1}{n} [m_{2r} - 2im_{r-1}m_{r+1} - m_r^2 + r^2 m_2 m_{r-1}^2] + O\left(\frac{1}{n^2}\right). \quad (5.8)$$

Lo mismo que en el caso de los dos primeros momentos muestrales, los momentos muestrales superiores \bar{m}_r son asintóticamente normales (para $n \rightarrow \infty$) con media y varianza determinadas por los términos principales de las fórmulas (5.7) y (5.8).

20.5.2. Funciones de los momentos muestrales. Cuando se determinan las características numéricas de la función de los momentos muestrales es útil el siguiente

Teorema. Sea dada la función $H(\bar{m}_r, \bar{m}_s)$ de los momentos muestrales \bar{m}_r y \bar{m}_s que no depende explícitamente de n y satisface las condiciones:

1) la función $H(u, v)$ es diferenciable continuamente dos veces en el entorno del punto m_r, m_s ;

2) para todos los valores de $x_k, k = \overline{1, n}$, la función $H(\bar{m}_r, \bar{m}_s) = H(x_1, x_2, \dots, x_n, n)$ satisface la estimación: $|H| < Cn^p$, donde C y p son constantes no negativas.

Entonces, el valor medio y la varianza de la variable aleatoria $H(\bar{m}_r, \bar{m}_s)$ pueden ser representados por las fórmulas asintóticas:

$$MH = H(m_r, m_s) + O\left(\frac{1}{n}\right); \quad (5.9)$$

$$DH = D\bar{m}_r \frac{\partial H}{\partial m_r}(m_r, m_s) + 2M[(\bar{m}_r - m_r)(\bar{m}_s - m_s)] \times \\ \times \frac{\partial H}{\partial m_r}(m_r, m_s) \frac{\partial H}{\partial m_s}(m_r, m_s) + D\bar{m}_s \frac{\partial H}{\partial m_s}(m_r, m_s) + O\left(\frac{1}{n^{3/2}}\right). \quad (5.10)$$

El teorema citado es válido también para la función de cualquier número de argumentos en el caso de muestras multidimensionales.

Cuando se cumple solamente la primera condición del teorema la estadística $H(\bar{m}_r, \bar{m}_s)$ es asintóticamente normal (para $n \rightarrow \infty$) con media y varianza determinadas por los términos principales de la fórmula (5.9) y (5.10). Notemos que el término principal de la fórmula (5.10) puede resultar igual a cero. En este caso, $\sqrt{n}(H - MH)$ es asintóticamente normal con varianza nula, es decir, converge según la probabilidad a cero. No se excluye la posibilidad de que para cierto $p > \frac{1}{2}$ $n^p(H - MH)$ tiene distribución límite no trivial que no es obligatoriamente normal.

20.5.3. Las distribuciones exactas de las características muestrales son visibles sólo en casos excepcionales. Con más plenitud está estudiado el caso de la población madre normal.

Si la función inicial de distribución $P(x)$ de valores muestrales x_k , $k = \overline{1, n}$, es normal con parámetros (m, σ^2) , entonces la media

muestral $\bar{x} = \frac{1}{n} \sum_{k=1}^n x_k$ y la varianza muestral \bar{s}^2 son independientes,

además \bar{x} está normalmente distribuido con parámetros $(m, \frac{\sigma^2}{n})$,

mientras que $\frac{n}{\sigma^2} \bar{s}^2$ tiene distribución χ_{n-1}^2 con $n-1$ grados de

libertad, o de otro modo, $\frac{n}{n-1} \bar{s}^2$ tiene la misma distribución que la media aritmética de $n-1$ cuadrados de las magnitudes normales independientes con parámetros $(0, \sigma^2)$. Además, la relación

$$t = \sqrt{n-1} \frac{\bar{x} - m}{\bar{s}} \quad (5.11)$$

tiene la distribución de Student (véase el cap. 6) con $(n-1)$ grados de libertad.

TEORÍA DE ESTIMACIÓN DE LOS PARÁMETROS

21.1. Problemas de estimación
y propiedades de las estimaciones

21.1.1. Planteamiento del problema. Sea ξ un elemento aleatorio que toma valores en un espacio medible (X, \mathfrak{B}) . Se considera conocido que la distribución del elemento ξ pertenece a cierta clase de distribuciones $\{P_\theta, \theta \in \Theta\}$ en (X, \mathfrak{B}) la cual depende del parámetro θ . Esto significa que en el conjunto Θ , llamado conjunto de valores admisibles del parámetro, existe tal θ_0 que la distribución del elemento ξ coincide con P_{θ_0} , es decir, $P\{\xi \in \Gamma\} = P_{\theta_0}(\Gamma)$, $\Gamma \in \mathfrak{B}$. El valor θ_0 se llama valor real del parámetro. Se supone que el valor real del parámetro es desconocido y el problema consiste en estimar a base del experimento con el elemento ξ el valor real del parámetro θ . En la práctica, el experimento de esta índole consiste ordinariamente en realizar la muestra de volumen n , es decir, de n observaciones (mediciones) independientes sobre el elemento aleatorio ξ . El resultado de la i -ésima observación se designará con x_i así que a consecuencia de n observaciones obtendremos un punto en el espacio X^n que se llama espacio muestral. A partir de las observaciones x_1, x_2, \dots, x_n se construye la estimación del valor real del parámetro. De este modo, la estimación $\theta^* = \theta^*(x_1, x_2, \dots, x_n)$ es una función dada en el espacio muestral X^n que toma el valor en el conjunto Θ . Sustituyendo en esta función los resultados de las observaciones en vez de los argumentos obtendremos el valor del parámetro θ^* que se toma en calidad de la estimación del valor real del parámetro. Se puede construir un número infinito de tales funciones y surge la pregunta acerca de cuáles de ellas son preferibles. La respuesta no es unívoca, pues se puede introducir diversos criterios de la calidad de las estimaciones.

21.1.2. Principios de construcción de las estimaciones. Es natural considerar que la calidad de la estimación θ^* depende de la proximidad de θ^* al valor real del parámetro. Sin embargo, se necesita precisar el término «proximidad».

Primero, se puede introducir de distintos modos la noción de proximidad en el conjunto Θ . Por ejemplo, si Θ es un espacio métrico es posible considerar como medida de proximidad entre los elementos $\theta_1, \theta_2 \in \Theta$ la distancia entre ellos. De una manera más general, en el conjunto Θ se puede introducir la función de pérdidas $r(\theta_1, \theta_2)$, $\theta_1, \theta_2 \in \Theta$, es decir, una función no negativa interpretada como las pérdidas que sufrimos si aceptamos a título de estimación del valor real del parámetro el valor θ_2 , mientras que el valor real es igual a θ_1 . En este caso la estimación θ^* es tanto más próxima a θ cuanto menores son las pérdidas $r(\theta, \theta^*)$.

Segundo, ya que los resultados de las observaciones x_1, x_2, \dots, x_n son aleatorios, la estimación $\theta^*(x_1, x_2, \dots, x_n)$ es también un elemento aleatorio, y por eso, la proximidad de θ^* a θ debe entenderse en cierto sentido promedio. Por ejemplo, se puede considerar que θ^* es próximo a θ si son pequeñas las pérdidas medias

$$\begin{aligned} M_{\theta^*}(\theta, \theta^*(x_1, \dots, x_n)) &= \\ &= \int_X \dots \int_X r(\theta, \theta^*(x_1, \dots, x_n)) P_{\theta}(dx_1) \dots P_{\theta}(dx_n). \end{aligned}$$

Sin embargo, para la estimación dada $\theta^*(x_1, \dots, x_n)$ estas pérdidas pueden ser pequeñas con unos θ y bastante grandes con otros θ . Claro está, si existiese tal estimación $\theta^*(x_1, x_2, \dots, x_n)$ que para otra estimación cualquiera $\theta^*(x_1, \dots, x_n)$ con todos los $\theta \in \Theta$ se cumpliera la desigualdad

$$M_{\theta^*}(\theta, \theta^*(x_1, x_2, \dots, x_n)) \leq M_{\theta^*}(\theta, \theta^*(x_1, x_2, \dots, x_n))$$

entonces, desde el punto de vista de la medida elegida de proximidad se convendría preferirla a otra estimación cualquiera. No obstante, hablando en general, tal estimación no existe. Por eso, eligiendo la estimación del parámetro desconocido se tienen en cuenta algunas consideraciones complementarias.

Se puede, por ejemplo, elegir la estimación $\theta^*(x_1, x_2, \dots, x_n)$ de tal modo que el valor de la función

$$\sup_{\theta \in \Theta} M_{\theta^*}(\theta, \theta^*(x_1, x_2, \dots, x_n))$$

sea el mínimo. Este principio de elección de las estimaciones lleva el nombre de principio de minimax y las estimaciones correspondientes, si existen, se llaman minimáximas. Procediendo así tratamos de minimizar la pérdida máxima relacionada con la elección de una u otra estimación.

Otro procedimiento para elegir la estimación es el llamado bayesiano: se considera que hay algunos razonamientos apriorísticos sobre la preferencia de unos u otros valores del parámetro θ . Con otras palabras, en el espacio Θ se considera dada una distribución apriorística $\mu(d\theta)$. La estimación $\theta^*(x_1, \dots, x_n)$ se elige de tal modo que el valor de la integral

$$\int_{\Theta} M_{\theta^*}(\theta, \theta^*(x_1, x_2, \dots, x_n)) \mu(d\theta)$$

sea el mínimo. Son posibles otros principios de construcción de las estimaciones.

2f.1.3. Desigualdad de Cramer—Rao. En adelante, con respecto al conjunto Θ , supondremos que es un intervalo en R^1 o un dominio en el espacio euclídeo d -dimensional R^d . El conjunto X , como regla, coincidirá con R^1 .

Llamemos insesgada la estimación $\theta^*(x_1, x_2, \dots, x_n)$ si para todos los $\theta \in \Theta$

$$M_{\theta^*}(\theta, \theta^*(x_1, x_2, \dots, x_n)) = \theta,$$

En la clase de estimaciones insesgadas es natural considerar la mejor aquella estimación cuya distribución para todos los θ se concentra «más densamente» alrededor del valor medio, es decir, alrededor de θ . Cuando θ es un parámetro unidimensional, de medida de tal concentración de la distribución puede servir la varianza de distribución. De esto modo, llegamos al problema de determinar en la clase de todas las estimaciones insesgadas una estimación $\theta^*(x_1, x_2, \dots, x_n)$ tal para la cual, con cada $\theta \in \Theta$, el valor de la función

$$M_{\theta} (\theta^* (x_1, x_2, \dots, x_n) - \theta)^2$$

sea el mínimo. Resulta que esta expresión está acotada por abajo por cierta función de θ , así que si para una estimación $\theta^*(x_1, x_2, \dots, x_n)$ esta cota inferior para la varianza se obtiene con todos los θ , ésta será la estimación buscada.

Supongamos que la distribución $P_{\theta}(dx)$, $\theta \in \Theta$ (aquí θ es un parámetro unidimensional), tiene la densidad $p(\theta, x)$ con respecto a alguna medida σ -finita $\nu(dx)$ en (X, \mathfrak{B}) . En particular, $P_{\theta}(dx)$ puede ser una distribución discreta concentrada en los puntos $x_1, x_2, \dots \in X$. Además, los puntos x_1, x_2, \dots no dependen de θ y $p(\theta, x_h)$ es la masa (probabilidad) correspondiente al punto x_h , así que $\sum_h p(\theta, x_h) = 1$.

Supongamos, luego, que las densidades $p(\theta, x)$ son diferenciables según θ , con la particularidad de que para cualquier conjunto medible Γ de X

$$\frac{\partial}{\partial \theta} \int_{\Gamma} p(\theta, x) \nu(dx) = \int_{\Gamma} \frac{\partial p(\theta, x)}{\partial \theta} \nu(dx).$$

Hagamos

$$I(\theta) = \int_X \left[\frac{\partial \log p(\theta, x)}{\partial \theta} \right]^2 p(\theta, x) \nu(dx), \quad \theta \in \Theta.$$

La magnitud $I(\theta)$ se llama cantidad de información sobre el parámetro θ , contenida en una observación. En el caso discreto $I(\theta)$ se escribe en forma de

$$I(\theta) = \sum_{h=1}^{\infty} \left[\frac{\partial \log p(\theta, x_h)}{\partial \theta} \right]^2 p(\theta, x_h).$$

La cantidad de información sobre θ contenida en n observaciones independientes x_1, x_2, \dots, x_n es igual a $nI(\theta)$.

Sea $\theta^*(x_1, x_2, \dots, x_n)$ cualquier estimación insesgada del parámetro θ . Para ciertas condiciones de regularidad tiene lugar la desigualdad

$$\sigma_{\theta}^2(\theta^*) = M_{\theta} (\theta^*(x_1, x_2, \dots, x_n) - \theta)^2 \geq \frac{1}{nI(\theta)}, \quad (1.1)$$

además, la igualdad se logra cuando, y sólo cuando,

$$\sum_{h=1}^n \frac{\partial \log p(\theta, x_h)}{\partial \theta} = \lambda [\theta^*(x_1, \dots, x_n) - \theta]$$

casi para todos los $x \in R^n$ con respecto a la medida $p(\theta, x_1) \dots p(\theta, x_n) \nu(dx_1) \dots \nu(dx_n)$. Aquí λ no depende de x_1, x_2, \dots, x_n , no obstante, puede depender de θ .

Las mencionadas condiciones de regularidad consisten en el cumplimiento de la igualdad

$$M_0 \frac{\partial \log L(\theta, x_1, \dots, x_n)}{\partial \theta} = \int_X \dots \int_X \frac{\partial \log L}{\partial \theta} \times \\ \times L \nu(dx_1) \dots \nu(dx_n) = 0$$

donde $L(\theta, x_1, \dots, x_n) = p(\theta, x_1) p(\theta, x_2) \dots p(\theta, x_n)$, así como en la posibilidad de diferenciar según θ la igualdad

$$\int_X \dots \int_X \theta^* (x_1, \dots, x_n) L(\theta, x_1, \dots, x_n) \nu(dx_1) \dots \nu(dx_n) = \theta.$$

La desigualdad (1.1) se llama desigualdad de Cramer—Roa y da la cota inferior para la varianza de la estimación insesgada. Si para la estimación $\theta^*(x_1, x_2, \dots, x_n)$ en la desigualdad (1.1) se obtiene la igualdad, entonces tal estimación se llama eficiente. Así, pues, entre las estimaciones regulares insesgadas, las estimaciones eficientes tienen varianza mínima. Llamemos eficiencia de la estimación $\theta^*(x_1, x_2, \dots, x_n)$ la relación entre la cota inferior para la varianza de la estimación y la varianza real de la estimación:

$$\text{eff}(\theta^*) = \frac{1}{nI(\theta)} \frac{1}{\sigma_{\theta}^2(\theta^*)}.$$

Es evidente que $\theta \leq \text{eff}(\theta^*) \leq 1$. Dos estimaciones eficientes del mismo parámetro coinciden casi con seguridad para cada θ .

21.1.4. Estimaciones suficientes. Designemos con $Q_{\theta}(dx_1, dx_2, \dots, dx_n)$ la medida en (X^n, \mathfrak{A}^n) determinada por la fórmula

$$Q_{\theta}(dx_1, dx_2, \dots, dx_n) = P_{\theta}(dx_1) P_{\theta}(dx_2) \dots P_{\theta}(dx_n).$$

La estimación $\theta^*(x_1, x_2, \dots, x_n)$ se llama suficiente para el parámetro θ si la distribución condicional

$$Q_{\theta}(A/\theta^*(x_1, x_2, \dots, x_n) = t), \quad A \in \mathfrak{A}^n,$$

no depende de θ . La estimación $\theta^*(x_1, x_2, \dots, x_n)$ es suficiente cuando y sólo cuando la distribución condicional de otra estimación cualquiera, a condición de $\theta^*(x_1, x_2, \dots, x_n) = t$, no depende de θ . La siguiente afirmación da el criterio de suficiencia de la estimación suponiendo que existo la densidad de distribución.

Teorema 1. Supongamos que las medidas $P_{\theta}(dx)$, $\theta \in \Theta$, son absolutamente continuas con respecto a cierta medida σ -finita $\nu(dx)$, dada en (X, \mathfrak{A}) , y sea $p(\theta, x) = \frac{dP_{\theta}}{d\nu}$. La estimación $\theta^*(x_1, x_2, \dots, x_n)$ es suficiente para el parámetro θ cuando, y sólo cuando, tiene lugar la representación

$$p(\theta, x_1) p(\theta, x_2) \dots p(\theta, x_n) = g_{\theta}(\theta^*(x_1, x_2, \dots, x_n)) \times \\ \times h(x_1, x_2, \dots, x_n)$$

donde $g_\theta(\theta^*)$ y $h(x_1, x_2, \dots, x_n)$ son funciones medibles no negativas, con la particularidad de que g_θ depende de x_1, x_2, \dots, x_n solamente mediante la estimación $\theta^*(x_1, x_2, \dots, x_n)$ y $h(x_1, x_2, \dots, x_n)$ no depende de θ .

La siguiente afirmación subraya la importancia del concepto de estimación suficiente en la teoría de estimación.

Teorema 2. Supongamos que $\theta^*(x_1, x_2, \dots, x_n)$ es la estimación suficiente para el parámetro θ y $\theta_1^*(x_1, x_2, \dots, x_n)$ es la estimación insesgada arbitraria del parámetro θ . Entonces, para todos los $\theta \in \Theta$

$$M_\theta [f(\theta^*(x_1, x_2, \dots, x_n)) - \theta]^2 \leq$$

$$\leq M_\theta [\theta_1^*(x_1, x_2, \dots, x_n) - \theta]^2$$

donde $f(t) = M_\theta \{\theta_1^*(x_1, x_2, \dots, x_n) / \theta^*(x_1, x_2, \dots, x_n) = t\}$ (como sigue de lo dicho anteriormente, la función $f(t)$ no depende de θ). En este caso, la función $f(\theta^*(x_1, x_2, \dots, x_n))$ es también una estimación insesgada del parámetro θ .

De este modo, teniendo para el parámetro θ la estimación insesgada arbitraria $\theta_1^*(x_1, x_2, \dots, x_n)$ y la estimación suficiente $\theta^*(x_1, x_2, \dots, x_n)$ podemos construir una nueva estimación insesgada del parámetro θ la cual tendrá para todos los θ varianza menor que la estimación inicial.

La estimación suficiente $\theta^*(x_1, x_2, \dots, x_n)$ se llama completa si de lo que para cierta función $\varphi(\theta^*(x_1, x_2, \dots, x_n))$ se cumple la relación $M_\theta \varphi(\theta^*(x_1, x_2, \dots, x_n)) = 0$, se deduce que $\varphi(\theta^*(x_1, x_2, \dots, x_n)) = 0$ es casi cierta con respecto a la medida Q_θ para cualquier $\theta \in \Theta$. Si existe la estimación insesgada completa $\theta^*(x_1, x_2, \dots, x_n)$ del parámetro θ , entonces para otra estimación cualquiera $\theta_1^*(x_1, x_2, \dots, x_n)$ de parámetro θ se cumple la desigualdad

$$M_\theta [\theta_1^*(x_1, x_2, \dots, x_n) - \theta]^2 \geq$$

$$\geq M_\theta [\theta^*(x_1, x_2, \dots, x_n) - \theta]^2,$$

es decir, en este caso la estimación $\theta^*(x_1, x_2, \dots, x_n)$ es eficiente.

21.1.5. Estimaciones de los parámetros multidimensionales. Supongamos que el conjunto Θ de los valores admisibles del parámetro es un dominio (abierto) en el espacio euclideo d -dimensional R^d . En este caso, la estimación $\theta^*(x_1, x_2, \dots, x_n)$ representa también un vector d -dimensional. Como en el caso unidimensional planteemos el problema sobre la búsqueda entre todas las estimaciones insesgadas tales cuya distribución tuviese máximo grado de concentración alrededor del valor medio. En el caso multidimensional la medida cómoda de esta concentración es una elipsoide de dispersión.

Sea ξ un vector aleatorio d -dimensional con media a y matriz de segundos momentos $V = \|v_{ij}\| : a = M\xi, v_{ij} = M\xi^i \xi^j, i, j = 1, 2, \dots, d$, donde ξ^i es el i -ésimo componente del vector ξ . Sea S un elipsoide en R^d con el centro en un punto a tal que si se concentra dentro de él la distribución uniforme (es decir, una distribución con densidad constante), entonces esta distribución tendrá a título de media el vector a , y a título de matriz de los segundos momentos, la matriz V . El elipsoide de este tipo se llama elipsoide de dispersión. En el caso unidimensional se convierte en el intervalo $(a - \sigma\sqrt{3}, a + \sigma\sqrt{3})$ donde a y σ^2 son la esperanza matemática y la varianza de la magnitud ξ , respectivamente. En el caso general la ecuación del

elipsoide de dispersión tiene el aspecto de

$$(V^{-1}(x-a), x-a) = d+2,$$

donde V^{-1} es la matriz inversa a la matriz V , x es un punto móvil del elipsoide, (y, x) es el producto escalar en R^d .

Supongamos ahora que $\{P_\theta, \theta \in \Theta\}$ es una familia de distribuciones en (X, \mathfrak{B}) las cuales tienen la densidad $p(\theta, x)$ respecto de alguna medida σ -finita $\nu(dx)$ dada en \mathfrak{B} . Sea θ un parámetro d -dimensional. Pongamos

$$i_{kr}(\theta) = \int \frac{\partial \log p(\theta, x)}{\partial \theta^k} \frac{\partial \log p(\theta, x)}{\partial \theta^r} p(\theta, x) \nu(dx)$$

donde $k, r = 1, 2, \dots, d$ y designemos por $I(\theta)$ la matriz con elementos $i_{kr}(\theta)$. La matriz $I(\theta)$ se llama matriz de información. Además, sea $\theta^*(x_1, x_2, \dots, x_n)$ cierta estimación insesgada del parámetro θ . Anotemos con $S_{n*}(\theta)$ el elipsoide de dispersión para la distribución del vector $\theta^*(x_1, x_2, \dots, x_n)$ según la medida $Q_\theta(dx_1, dx_2, \dots, dx_n) = p(\theta, x_1) p(\theta, x_2) \dots p(\theta, x_n) \nu(dx_1) \dots \nu(dx_n)$.

El análogo multidimensional de la desigualdad de Cramer—Roach dice: para cualquier estimación insesgada $\theta^*(x_1, x_2, \dots, x_n)$ del parámetro θ , a ciertas condiciones de regularidad, el elipsoide $S_{n*}(\theta)$ contiene un elipsoide cuya ecuación tiene por expresión

$$n(I(\theta)(u-\theta), u-\theta) = d+2, \quad (1.2)$$

donde n es el número de observaciones, u es el punto corriente del elipsoide ($u \in R^d$). En el caso límite ambos elipsoides coinciden. En este caso la estimación $\theta^*(x_1, x_2, \dots, x_n)$ se llama conjuntamente eficiente. La matriz de los segundos momentos para la estimación conjuntamente eficiente coincide con $n^{-1}I^{-1}(\theta)$. Se llama eficiencia de la estimación $\theta^*(x_1, x_2, \dots, x_n)$ la relación entre el volumen del elipsoide (1.2) y el volumen del elipsoide $S_{n*}(\theta)$. De un modo evidente, para el caso multidimensional, se aplican la noción de estimación suficiente, así como las propiedades de estimaciones suficientes expuestas en los teoremas 1, 2.

21.1.6. Propiedades asintóticas de las estimaciones. Supongamos que el volumen de la muestra n crece, es decir, $n \rightarrow \infty$, y que nos interese por las propiedades asintóticas de las estimaciones.

La estimación $\theta^*(x_1, x_2, \dots, x_n)$ del parámetro θ se llama conciliable si para $n \rightarrow \infty$, $\theta^*(x_1, x_2, \dots, x_n) \rightarrow \theta$ en probabilidad. A veces, tales estimaciones se llaman débilmente conciliables a diferencia de las estimaciones fuertemente conciliables para las cuales la convergencia correspondiente se realiza con la probabilidad 1.

Luego, notemos que las estimaciones eficientes no existen siempre ni mucho menos. Sin embargo, en muchos casos existen las estimaciones asintóticamente eficientes.

Llamemos eficiencia asintótica de la estimación $\theta^*(x_1, x_2, \dots, x_n)$ al límite

$$e_0(\theta) = \lim_{n \rightarrow \infty} \text{eff}(\theta^*) = \lim_{n \rightarrow \infty} \frac{1}{nI(\theta)\sigma_{\theta}^2(\theta^*)},$$

si éste existe. En muchos casos $\sigma_{\theta}^2(\theta^*) \sim \frac{c}{n}$ para $n \rightarrow \infty$ y por eso, en estos casos, $e_n(\theta^*) = [cI(\theta)]^{-1}$ existe. Es evidente que $0 \leq e_n(\theta^*) \leq 1$. Si $e_n(\theta^*) = 1$ la estimación $\theta^*(x_1, x_2, \dots, x_n)$ se llama asintóticamente eficiente.

21.2. Métodos de construcción de las estimaciones

21.2.1. Método de los momentos. Tenga la magnitud aleatoria la distribución perteneciente a la familia de las distribuciones $\{P_\theta, \theta \in \Theta\}$, donde Θ es un dominio en R^d . Supongamos que existen los primeros d momentos de distribución P_θ y hagamos

$$m_r(\theta) = \int_X x^r P_\theta(dx), \quad r = 1, 2, \dots, d.$$

Aquí X coincide con R^1 , o X es un conjunto numerable. Teniendo n observaciones independientes x_1, x_2, \dots, x_n sobre la magnitud aleatoria ξ construyamos los momentos muestrales

$$\bar{m}_r = \frac{1}{n} \sum_{k=1}^n x_k^r, \quad r = 1, 2, \dots, d.$$

El método de los momentos consiste en igualar los momentos muestrales a los teóricos. Obtenemos el sistema de ecuaciones

$$m_r(\theta) = \bar{m}_r, \quad r = 1, 2, \dots, d,$$

con respecto a d incógnitas $\theta^1, \theta^2, \dots, \theta^d$. Si existe la solución única $\theta_r^*(x_1, x_2, \dots, x_n) = f_r(\bar{m}_1, \bar{m}_2, \dots, \bar{m}_d)$, $r = 1, \dots, d$, de este sistema y las funciones f_r son continuas, entonces la estimación que se obtiene $\{\theta_r^*, r = 1, 2, \dots, d\}$ es una estimación conciliable del parámetro θ . No obstante, hablando en general, las estimaciones obtenidas mediante el método de momentos no son eficientes.

21.2.2. Método del máximo de verosimilitud. Supongamos que la familia de distribuciones $\{P_\theta, \theta \in \Theta\}$ en el espacio medible (X, \mathfrak{B}) tiene la densidad de distribución $p(\theta, x)$ respecto de cierta medida σ -finita $\nu(dx)$ dada en \mathfrak{A} . Si X es discreto, $p(\theta, x)$ es la probabilidad de que $\xi = x$ si el valor real del parámetro es igual a θ . Sean x_1, \dots, x_n los resultados de n observaciones independientes de la magnitud aleatoria ξ . Según el método de la verosimilitud máxima a título de la estimación $\theta^*(x_1, x_2, \dots, x_n)$ del parámetro θ se elige tal función de las observaciones que da el máximo de la función

$$p(\theta, x_1) p(\theta, x_2) \dots p(\theta, x_n) = L(\theta, x_1, x_2, \dots, x_n)$$

la cual se llama función de verosimilitud. Si para $\theta = \theta^*(x_1, x_2, \dots, x_n)$ la función de verosimilitud alcanza el valor máximo, entonces para este mismo θ el valor máximo lo alcanza también la función $\log L(\theta, x_1, x_2, \dots, x_n)$. A menudo, es más cómodo utilizar la función $\log L$. Las estimaciones obtenidas por el método de verosimilitud máxima se llaman estimaciones de verosimilitud máxima.

Para encontrar las estimaciones de verosimilitud máxima es necesario resolver la ecuación ($k = 1, 2, \dots, d$)

$$\frac{\partial \log L(\theta, x_1, x_2, \dots, x_n)}{\partial \theta^k} = 0$$

si $\theta = (\theta^1, \theta^2, \dots, \theta^d)$. Las ecuaciones de este tipo se llaman ecuaciones de verosimilitud. Al resolver las ecuaciones de verosimilitud es conveniente rechazar las soluciones de tipo $\theta = \text{const}$ y estudiar sólo las que dependen de x_1, x_2, \dots, x_n y caen en el dominio Θ de valores admisibles del parámetro. Se debe también tener en cuenta que la función de verosimilitud puede tomar el valor máximo en la frontera del dominio Θ .

Las estimaciones de verosimilitud máxima poseen las dos importantes propiedades siguientes:

A) Si existe la estimación suficiente $\theta^*(x_1, x_2, \dots, x_n)$ para el parámetro θ , cada solución de la ecuación de verosimilitud es función de $\theta^*(x_1, x_2, \dots, x_n)$.

B) Si para el parámetro θ existe la estimación eficiente (en el caso multidimensional, conjuntamente eficiente) $\theta^*(x_1, x_2, \dots, x_n)$, la ecuación de verosimilitud tiene la única solución $\theta^*(x_1, x_2, \dots, x_n)$.

21.2.3. Comportamiento asintótico de las estimaciones de verosimilitud máxima. Sea Θ un intervalo en R^1 , $X = R^1$ y $\nu(dx) = dx$, donde dx es la medida de Lebesgue en R^1 . Formulemos un teorema que muestra que las estimaciones de verosimilitud máxima tienen una serie de buenas propiedades cuando $n \rightarrow \infty$.

Teorema 1. Supongamos que la densidad $p(\theta, x)$ satisface las condiciones:

1) para cada $\theta \in \Theta$ y para casi todo x existen las derivadas

$$\frac{\partial^k \log p(\theta, x)}{\partial \theta^k}, \quad k=1, 2, 3;$$

2) para cada $\theta \in \Theta$ están cumplidas las desigualdades

$$\left| \frac{\partial p(\theta, x)}{\partial \theta} \right| \leq G_1(x), \quad \left| \frac{\partial^2 p(\theta, x)}{\partial \theta^2} \right| \leq G_2(x),$$

$$\left| \frac{\partial^2 p(\theta, x)}{\partial \theta^3} \right| \leq G_2(x),$$

donde las funciones $G_1(x)$ y $G_2(x)$ son integrables en R^1 según la medida de Lebesgue y

$$\sup_{\theta \in \Theta} \int_{R^1} G_2(x) p(\theta, x) dx < \infty;$$

3) para cada $\theta \in \Theta$ la integral

$$I(\theta) = \int_{R^1} \left[\frac{\partial \log p(\theta, x)}{\partial \theta} \right]^2 p(\theta, x) dx$$

es finita y positiva.

Entonces la ecuación de verosimilitud tiene la solución $\theta^*(x_1, x_2, \dots, x_n)$ que es una estimación conciliable asintóticamente eficiente y

asintóticamente normal del parámetro θ , lo que significa que la magnitud

$$I(\theta) \sqrt{n} (\theta^*(x_1, x_2, \dots, x_n) - \theta)$$

es asintóticamente normal con parámetros $(0, 1)$ si el valor real del parámetro es igual a θ .

Este teorema se generaliza para el caso de una magnitud aleatoria discreta, así como para el del parámetro multidimensional θ .

21.2.4. Método del mínimo de χ^2 . Sean x_1, x_2, \dots, x_n observaciones independientes de la magnitud aleatoria ξ con valores en (X, \mathfrak{B}) , cuya distribución pertenece a la clase de distribuciones $\{P_\theta, \theta \in \Theta\}$. Supongamos que el espacio X está partido en r conjuntos medibles disjuntos X_1, X_2, \dots, X_r . Designemos por n_i el número de observaciones en la muestra x_1, x_2, \dots, x_n caídas en el conjunto X_i . Si el conjunto X es finito, es decir, la magnitud aleatoria toma sólo un número finito de valores, se puede considerar que X_i es el conjunto de un punto. De este modo, están agrupados los resultados de las observaciones.

Compongamos la magnitud

$$\chi^2 = \sum_{i=1}^r \frac{(n_i - np_i(\theta))^2}{np_i(\theta)},$$

donde $p_i(\theta) = P_\theta(X_i)$, $i = 1, 2, \dots, r$, $\theta \in \Theta$. La estimación $\theta^*(x_1, x_2, \dots, x_n)$ se llama estimación según el método del mínimo de χ^2 si se obtiene minimizando la magnitud χ^2 según θ . Si θ es un parámetro d -dimensional, entonces para encontrar la estimación según el método del mínimo de χ^2 obtenemos el sistema de ecuaciones

$$\sum_{i=1}^r \left[\frac{n_i - np_i(\theta)}{r_i(\theta)} + \frac{(n_i - np_i(\theta))^2}{2np_i^2(\theta)} \right] \frac{\partial p_i(\theta)}{\partial \theta^k} = 0, \quad k = 1, 2, \dots, d.$$

Por sus propiedades asintóticas las estimaciones, obtenidas mediante el método del mínimo de χ^2 , son muy próximas a las estimaciones de verosimilitud máxima. Por ejemplo, a ciertas condiciones, con la probabilidad 1, se tiene sólo una raíz conciliable de las ecuaciones correspondientes, que da a la magnitud χ^2 el mínimo absoluto.

21.3. Dominios confidenciales

21.3.1. Noción del dominio confidencial. En algunos casos es importante no sólo dar la estimación para un parámetro desconocido de distribución, sino que también indicar el dominio donde supuestamente debe estar el valor real del parámetro. Este dominio está construido a partir de los resultados de las observaciones y por eso puede variar de una muestra a otra y, por lo tanto, es un dominio aleatorio. Por consiguiente, se puede hablar de la probabilidad de que este dominio recubre el valor real del parámetro. Eligiendo cierto número bastante pequeño $\varepsilon > 0$ podemos proponernos construir la regla que nos permita asignar a los resultados de las observaciones tal dominio en el conjunto paramétrico que, con la probabilidad $1 - \varepsilon$, el valor

real del parámetro se contenga en este dominio. Esto significa que en una serie larga de muestras nos equivocamos sólo el 100 ε % de casos.

Los dominios en cuestión se llaman dominios confidenciales y el número $1 - \varepsilon$, coeficiente de confianza.

21.3.2. Construcción de los dominios confidenciales. Sean x_1, x_2, \dots, x_n las observaciones independientes de la magnitud aleatoria ξ cuya distribución contiene un parámetro desconocido que varía en el espacio R^1 y sea θ^* (x_1, x_2, \dots, x_n) cualquier estimación del parámetro θ . Designemos con $q_\theta(dt)$ la distribución de la estimación θ^* suponiendo que el valor real del parámetro coincide con θ , es decir,

$$q_\theta(dt) = Q_\theta\{\theta^*(x_1, x_2, \dots, x_n) \in dt\},$$

donde $Q_\theta(dx_1, dx_2, \dots, dx_n)$ es una distribución en el espacio muestral R^n que se determina por la fórmula

$$Q_\theta(dx_1, dx_2, \dots, dx_n) = P_\theta(dx_1) P_\theta(dx_2) \dots P_\theta(dx_n).$$

Si la distribución $q_\theta(dt)$ no tiene átomos, entonces según $\varepsilon > 0$ fijado se puede siempre elegir los números $a_1(\theta, \varepsilon)$ y $a_2(\theta, \varepsilon)$ de tal modo que $a_1(\theta, \varepsilon) < a_2(\theta, \varepsilon)$ y

$$\int_{\{t < a_1(\theta, \varepsilon)\}} q_\theta(dt) + \int_{\{t > a_2(\theta, \varepsilon)\}} q_\theta(dt) = \varepsilon.$$

Naturalmente, esta elección no es unívoca. Supongamos que se puede elegir estas funciones de modo que sean continuas según θ y que cada una de las ecuaciones $a_i(\theta, \varepsilon) = t$, $i = 1, 2$, $t \in \Theta$, tenga la única solución $c_i(t, \varepsilon)$, $i = 1, 2$. Entonces las correlaciones

$$Q_\theta\{a_1(\theta, \varepsilon) < \theta^* < a_2(\theta, \varepsilon)\} = 1 - \varepsilon$$

y

$$Q_\theta\{c_1(\theta^*, \varepsilon) < \theta < c_2(\theta^*, \varepsilon)\} = 1 - \varepsilon$$

son equivalentes.

Así pues, conociendo la distribución de la estimación $\theta^*(x_1, x_2, \dots, x_n)$ según $\varepsilon > 0$ fijado podemos construir el intervalo de confianza $(c_1(\theta^*, \varepsilon), c_2(\theta^*, \varepsilon))$ con coeficiente de confianza $1 - \varepsilon$.

Notemos que si la distribución $q_\theta(dt)$ tiene átomos, los números $a_1(\theta, \varepsilon)$ y $a_2(\theta, \varepsilon)$ se deben elegir partiendo de la desigualdad

$$\int_{\{t < a_1(\theta, \varepsilon)\}} q_\theta(dt) + \int_{\{t < a_2(\theta, \varepsilon)\}} q_\theta(dt) \leq \varepsilon,$$

ya que en este caso pueden no existir tales a_1 y a_2 para los que se cumpla la igualdad correspondiente. Si continuamos construyendo del modo mencionado arriba, obtendremos un intervalo confidencial con coeficiente de confianza no menor que $1 - \varepsilon$.

Evidentemente, partiendo de distintas estimaciones θ^* del parámetro θ obtendremos diversos intervalos confidenciales. Es deseable que la longitud del intervalo confidencial sea lo más pequeña posible. Por eso, al construir intervalos confidenciales es natural basarse en las estimaciones eficientes o asintóticamente eficientes que se obtienen, por ejemplo, mediante el método de verosimilitud máxima.

21.3.3. Un método de construcción de los intervalos confidenciales. Supongamos que la distribución $P_\theta(dx)$ no tiene átomos, y exista la función $g(\theta, x_1, x_2, \dots, x_n)$, $\theta \in \Theta$, $x_1, x_2, \dots, x_n \in R^1$, que posee las propiedades:

- 1) $g(\theta, x_1, x_2, \dots, x_n)$ es continua y monótona según θ ;
- 2) la función $Q_\theta\{g(\theta, x_1, x_2, \dots, x_n) < \alpha\}$ no depende de θ para cada $\alpha \in R^1$ (véase la definición de la medida Q_θ en el p. 21.3.2).

A base de $\varepsilon > 0$ fijado elegimos los números $a_1(\varepsilon)$ y $a_2(\varepsilon)$ de tal modo que

$$Q_\theta\{a_1(\varepsilon) < g(\theta, x_1, x_2, \dots, x_n) < a_2(\varepsilon)\} = 1 - \varepsilon.$$

En virtud de la condición 2) los números $a_1(\varepsilon)$ y $a_2(\varepsilon)$ no dependen de θ .

Designemos con c_i , $i = 1, 2$, los números que satisfacen las correlaciones $g(c_i, x_1, x_2, \dots, x_n) = a_i(\varepsilon)$, $i = 1, 2$. La magnitud c_i depende sólo de x_1, x_2, \dots, x_n y ε . Es fácil ver que

$$Q_\theta\{c_1 < \theta < c_2\} = 1 - \varepsilon$$

y, por eso, $[c_1, c_2]$ es el intervalo confidencial con coeficiente de confianza $1 - \varepsilon$.

Pongamos

$$F_\theta(x) = \int_{-\infty}^x P_\theta(dy), \quad x \in R^1$$

y supongamos que F_θ es continua y monótona según θ . Como es fácil comprobar, la función

$$F_\theta(x_1) F_\theta(x_2) \dots F_\theta(x_n)$$

satisface las condiciones 1), 2) y entonces puede ser utilizada para construir intervalos confidenciales. Además, ya que $F_\theta(x)$ es continua según x

$$Q_\theta\{E_\theta(x_h) < \alpha\} = \begin{cases} 0 & \text{para } \alpha \leq 0; \\ \alpha & \text{para } 0 \leq \alpha \leq 1; \\ 1 & \text{para } \alpha > 1. \end{cases}$$

la suma $\sum_{h=1}^n \log E_\theta(x_h)$ tiene la distribución gamma

$$\begin{aligned} Q_\theta\left\{-\log a_2 < -\sum_{k=1}^n \log F_\theta(x_k) < -\log a_1\right\} &= \\ &= \int_{\log a_2}^{-\log a_1} \frac{1}{\Gamma(n)} t^{n-1} e^{-t} dt. \end{aligned}$$

A partir de $\varepsilon > 0$ fijado se puede elegir los números a_1 y a_2 , $a_1 < a_2$ de tal modo que la integral a la derecha sea igual a $1 - \varepsilon$.

De aquí obtenemos

$$Q_{\theta} \left\{ a_1 < \prod_{h=1}^n F_{\theta}(x_h) < a_2 \right\} = 1 - \varepsilon.$$

Ya que la función $\prod_{h=1}^n F_{\theta}(x_h)$ es continua y monótona según θ , existen tales c_1 y c_2 dependientes sólo de ε y x_1, x_2, \dots, x_n que (c_1, c_2) es el intervalo confidencial para θ con coeficiente de confianza $1 - \varepsilon$.

Si la función $g(\theta, x_1, x_2, \dots, x_n)$ satisface la condición 2) es continua según θ , pero no es obligatoriamente monótona, entonces en vez del intervalo confidencial se obtiene cierto dominio confidencial. Este mismo principio puede ser utilizado también para la construcción de los dominios confidenciales cuando θ es un parámetro multidimensional.

21.3.4. Método de Bayes. Aplicando el método de construcción de los intervalos confidenciales basado en la fórmula de Bayes, partimos de la suposición de que el mismo parámetro θ es aleatorio. Se supone también que conocemos la distribución a priori de θ . Designemos con $\varphi(\theta)$ la densidad de esta distribución respecto de la medida de Lebesgue (recordemos que Θ es un intervalo R^1).

Luego, sea θ^* (x_1, x_2, \dots, x_n) cierta estimación del parámetro θ , construida a base de las observaciones independientes x_1, x_2, \dots, x_n de la variable aleatoria ξ . Supongamos que la distribución $q_{\theta}(dt)$ de la estimación θ^* (véase el p. 21.3.2) es también absolutamente continua respecto de la medida de Lebesgue con densidad $g(\theta, t)$. Conforme a la fórmula de Bayes la densidad de distribución de la magnitud θ para θ^* fijado es igual a

$$\psi(\theta/\theta^*) = \frac{g(\theta, \theta^*) \varphi(\theta)}{\int_{\Theta} g(\theta, \theta^*) \varphi(\theta) d\theta}.$$

Por eso la probabilidad condicional de que el parámetro θ está entre los límites c_1 y c_2 a condición de que θ^* fijado se expresa por la fórmula

$$P\{c_1 < \theta < c_2/\theta^*\} = \int_{c_1}^{c_2} \psi(t/\theta^*) dt.$$

Ahora, por medio de $\varepsilon > 0$ fijado podemos determinar los números $c_1(\theta^*, \varepsilon)$ y $c_2(\theta^*, \varepsilon)$ de tal modo que

$$P\{c_1(\theta^*, \varepsilon) < \theta < c_2(\theta^*, \varepsilon)\} = 1 - \varepsilon.$$

Así, pues, para θ^* está obtenido el intervalo confidencial con coeficiente de confianza $1 - \varepsilon$.

Este método no siempre es cómodo, ya que en algunos casos no hay razón de considerar aleatorio el parámetro θ , si es aleatorio, no siempre es conocida su distribución a priori.

ESTIMACIONES DE LOS PARÁMETROS DE ALGUNAS DISTRIBUCIONES

22.1. Estimaciones de los parámetros de distribución normal

22.1.1. Estimación de la media con varianza conocida. Sean x_1, x_2, \dots, x_n observaciones independientes de la magnitud aleatoria ξ con densidad de distribución

$$f(\theta, x) = \frac{1}{\sqrt{2\pi} \sigma} e^{-\frac{(x-\theta)^2}{2\sigma^2}}, \quad x \in R^1,$$

donde θ es un parámetro desconocido y σ , conocido. Hagamos

$$\theta^*(x_1, x_2, \dots, x_n) = \bar{x} = \frac{1}{n} \sum_{h=1}^n x_h.$$

Es evidente que θ^* es una magnitud aleatoria normalmente distribuida con parámetros $M_{\theta^*} = \theta$ y $M_{\theta^*}[\theta^* - \theta]^2 = \frac{\sigma^2}{n}$. De este modo, la estimación θ^* es insesgada y conciliable.

Luego, ya que $I(\theta) = \frac{1}{\sigma^2}$, el segundo miembro de la desigualdad de Cramer-Roa en el caso considerado es igual a $\frac{\sigma^2}{n}$, lo que significa que la estimación θ^* es eficiente.

La magnitud aleatoria $\frac{\sqrt{n}(\theta^* - \theta)}{\sigma}$ tiene distribución normal con los parámetros (0, 1). Hallando con ayuda de $\varepsilon > 0$ fijado (por ejemplo, de las tablas) tal número c_ε , que

$$\frac{1}{\sqrt{2\pi}} \int_{c_\varepsilon}^{c_\varepsilon} e^{-\frac{x^2}{2}} dx = 1 - \varepsilon,$$

obtenemos

$$Q_0 \left\{ -c_\varepsilon < \frac{\sqrt{n}(\theta^* - \theta)}{\sigma} < c_\varepsilon \right\} = 1 - \varepsilon.$$

de donde

$$Q_0 \left\{ \theta^* - \frac{c_\varepsilon}{\sqrt{n}} \sigma < \theta < \theta^* + \frac{c_\varepsilon}{\sqrt{n}} \sigma \right\} = 1 - \varepsilon.$$

Aquí $Q_0(dx_1, \dots, dx_n) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{h=1}^n (x_h - \theta)^2 \right\} \times dx_1, \dots, dx_n$. De este modo $\left(\theta^* - \frac{c_\varepsilon}{\sqrt{n}} \sigma, \theta^* + \frac{c_\varepsilon}{\sqrt{n}} \sigma \right)$ es el intervalo confidencial con coeficiente de confianza $1 - \varepsilon$.

22.1.2. Estimación de la varianza para la media conocida. En este caso

$$p(\theta, x) = \frac{1}{\sqrt{2\pi\theta}} \exp \left\{ -\frac{(x-a)^2}{2\theta} \right\}, \quad x \in R^1,$$

donde $\theta \in (0, \infty)$ es un parámetro desconocido y a , conocida. La estimación insesgada eficiente para θ será

$$\theta^*(x_1, x_2, \dots, x_n) = \frac{1}{n} \sum_{h=1}^n (x_h - a)^2,$$

donde x_1, x_2, \dots, x_n son los resultados de las observaciones independientes. La varianza de la estimación θ^* es igual a

$$M_0(\theta^* - \theta)^2 = \frac{2\theta^2}{n}.$$

La magnitud $\frac{n\theta^*}{\theta}$ tiene distribución χ^2 con n grados de libertad. Para construir el intervalo confidencial con coeficiente de confianza $1 - \varepsilon$ elegimos los números a_1 y a_2 (por las tablas) de modo que

$$Q_0 \left\{ a_1 \leq \frac{n\theta^*}{\theta} < a_2 \right\} = 1 - \varepsilon,$$

donde

$$Q_0(dx_1, dx_2, \dots, dx_n) = (2\pi\theta)^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2\theta} \sum_{h=1}^n (x_h - a)^2 \right\} dx_1 dx_2 \dots dx_n.$$

Entonces

$$Q_0 \left\{ \frac{n\theta^*}{a_2} < \theta < \frac{n\theta^*}{a_1} \right\} = 1 - \varepsilon,$$

así que $\left(\frac{n\theta^*}{a_2}, \frac{n\theta^*}{a_1} \right)$ es el intervalo buscado

22.1.3. Estimación de la varianza para la media desconocida. El problema es el mismo que en el punto 22.1.2, sin embargo, el parámetro a se considera desconocido. La estimación insesgada conciliable

para el parámetro θ será

$$\theta^*(x_1, x_2, \dots, x_n) = \frac{1}{n-1} \sum_{h=1}^n (x_h - \bar{x})^2, \quad n > 1,$$

donde $\bar{x} = \frac{1}{n} \sum_{h=1}^n x_h$. Para ésta

$$M_{\theta}(\theta^* - \theta)^2 = \frac{2\theta^2}{n-1}.$$

La magnitud $\frac{(n-1)\theta^*}{\theta}$ tiene distribución χ^2 con $(n-1)$ -ésimo grado de libertad y el intervalo confidencial se construye del mismo modo que en el punto 22.1.2.

22.1.4. Estimación de la media para la varianza desconocida. El problema es el mismo que en el punto 22.1.1, no obstante el parámetro σ es desconocido. Como antes, la estimación $\theta^* = \bar{x}$ no es sesgada y es conciliable. Para construir el intervalo confidencial empleemos el hecho de que la magnitud $\frac{\bar{x} - \theta}{s}$, donde $s^2 = \frac{1}{n} \sum_{h=1}^n (x_h - \bar{x})^2$ tiene

la distribución de Student con $(n-1)$ -ésimo grado de libertad. La densidad de esta distribución tiene el aspecto de

$$S_{n-1}(x) = \frac{1}{\sqrt{\pi}} \frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)} (1+x^2)^{-\frac{n}{2}}.$$

Al determinar el número c_v de la relación

$$2 \int_0^{c_v} S_{n-1}(x) dx = 1 - \varepsilon$$

obtendremos

$$Q_{\theta} \left\{ -c_v < \frac{\bar{x} - \theta}{s} < c_v \right\} = 1 - \varepsilon,$$

donde $Q_{\theta}(dx_1, dx_2, \dots, dx_n) = (2\pi\sigma^2)^{-n/2} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{h=1}^n (x_h - \theta)^2 \right\} \times$
 $\times dx_1 dx_2 \dots dx_n$.

Así, pues, $(\bar{x} - c_v s, \bar{x} + c_v s)$ es el intervalo confidencial para el parámetro θ con coeficiente de confianza $1 - \varepsilon$.

22.1.5. Estimación conjunta de los parámetros de la media y de la varianza. En la distribución normal sean desconocidos los parámetros

a y σ . Resolviendo la ecuación de verosimilitud obtendremos:

$$a^* = \frac{1}{n} \sum_{h=1}^n x_h = \bar{x}, \quad (\sigma^*)^2 = \frac{1}{n} \sum_{h=1}^n (x_h - \bar{x})^2 = s^2.$$

No obstante, la estimación $(\sigma^*)^2$ es sesgada. Pongamos $b^* = \frac{n}{n-1} s^2$. Entonces (a^*, b^*) será la estimación insesgada para los parámetros (a, σ^2) .

La elipse óptima (véase el p. 214.5) tiene la ecuación

$$\frac{(u-a)^2}{\sigma^2} + \frac{(v-\sigma)^2}{2\sigma^4} = \frac{4}{n}.$$

La elipse de dispersión para la estimación (a^*, b^*) se da por la ecuación

$$\frac{(u-a)^2}{\sigma^2} + \frac{n-1}{n} \frac{(v-\sigma^2)^2}{2\sigma^4} = \frac{4}{n}.$$

La estimación eficiente conjunta (a^*, b^*) es igual a $\frac{n-1}{n}$, así que la estimación (a^*, b^*) es asintóticamente eficiente conjunta. Las magnitudes a^* y b^* son independientes.

22.2. Estimaciones de los parámetros de las distribuciones binomial y de Poisson

22.2.1. **Distribución binomial.** Sea que la magnitud ξ toma los valores $0, 1, 2, \dots$ con probabilidad $p_h(\theta) = C_N^h \theta^h (1-\theta)^{N-h}$, $k = 0, 1, 2, \dots, N$, respectivamente, con la particularidad de que el parámetro θ , $0 < \theta < 1$, es desconocido. Sean k_1, k_2, \dots, k_n los resultados de n observaciones independientes de la magnitud ξ . Hagamos

$$\theta^* (k_1, k_2, \dots, k_n) = \frac{1}{nN} \sum_{i=1}^n k_i.$$

Entonces θ^* es la estimación insesgada del parámetro θ para la cual

$$M_\theta (\theta^* - \theta)^2 = \frac{\theta(1-\theta)}{nN}.$$

Do otro lado, no es difícil calcular que

$$I(\theta) = \frac{N}{\theta(1-\theta)}.$$

De aquí sigue que θ^* es la estimación eficiente del parámetro θ .

Para construir el intervalo confidencial empleemos el hecho de que la magnitud $\frac{\sqrt{nN}(\theta^* - \theta)}{\sqrt{\theta(1-\theta)}}$ es asintóticamente normal con los pará-

metros (0, 1). Admitiendo aproximadamente que

$$Q_0 \left\{ -a_\varepsilon < \frac{1 - \sqrt{Nn}(\theta^* - \theta)}{\sqrt{\theta(1-\theta)}} < a_\varepsilon \right\} = \frac{2}{\sqrt{2\pi}} \int_0^{a_\varepsilon} e^{-\frac{t^2}{2}} dt$$

y hallando a_ε de la condición

$$\frac{2}{\sqrt{2\pi}} \int_0^{a_\varepsilon} e^{-\frac{t^2}{2}} dt = 1 - \varepsilon,$$

obtendremos aproximadamente

$$Q_0 \left\{ -a_\varepsilon < \frac{1 - \sqrt{Nn}(\theta^* - \theta)}{\sqrt{\theta(1-\theta)}} < a_\varepsilon \right\} = 1 - \varepsilon,$$

donde Q_0 se determina por la igualdad

$$Q_0(k_1, k_2, \dots, k_N) = p_{k_1}(\theta) p_{k_2}(\theta) \dots p_{k_N}(\theta), \quad k_i = 0, 1, 2, \dots, N.$$

Entonces, los extremos del intervalo confidencial con coeficiente de confianza $1 - \varepsilon$ son las raíces de la ecuación cuadrática

$$(Nn - a_\varepsilon^2) x^2 - (2Nn\theta^* + a_\varepsilon^2) x + Nn\theta^{*2} = 0.$$

En particular, si en el esquema estudiado ponemos $N = 1$, obtendremos que el resultado de la i -ésima observación de la magnitud ξ es la aparición o la no aparición del suceso $A = \{\xi = 1\}$. La probabilidad de este suceso es igual a θ . El papel de su estimación lo juega $\theta^* = \frac{v}{n}$ donde v es el número de aquellas observaciones durante las cuales se realizó el suceso A .

22.2.2. Distribución de Poisson. Tenga ξ una distribución de Poisson, lo que significa que ξ puede tomar los valores $0, 1, 2, \dots$ con las probabilidades $p_k(\theta) = \frac{\theta^k}{k!} e^{-\theta}$, $0 < \theta < \infty$, siendo θ un parámetro desconocido.

Es una estimación insesgada del parámetro θ la estimación

$$\theta^* = \frac{1}{n} \sum_{i=1}^n k_i$$

donde k_1, k_2, \dots, k_n son los resultados de las observaciones independientes de la magnitud ξ . En este caso,

$$M_0(\theta^* - \theta)^2 = \frac{\theta}{n}.$$

Ya que $I(\theta) = \theta^{-1}$, θ^* es una estimación eficiente. Después, la magnitud

$$\sqrt{\frac{n}{\theta}}(\theta^* - \theta)$$

es asintóticamente normal con los parámetros $(0, 1)$. Por eso, para n grandes se verifica la igualdad aproximada

$$Q_{\theta} \left\{ -a_{\varepsilon} < \sqrt{\frac{n}{\theta}} (\theta^* - \theta) < a_{\varepsilon} \right\} = 1 - \varepsilon,$$

donde a_{ε} se elige del mismo modo que en el p. 22.1 y la medida Q_{θ} se determina por la fórmula

$$Q_{\theta}(k_1, k_2, \dots, k_n) = \frac{\theta^{k_1+k_2+\dots+k_n}}{k_1! k_2! \dots k_n!} e^{-n\theta}, \quad k_i = 0, 1, 2, \dots$$

De aquí obtenemos que los extremos del intervalo confidencial con coeficiente de confianza $1 - \varepsilon$ son raíces de la ecuación cuadrática

$$x^2 - \left(2\theta^* + \frac{a_{\varepsilon}^2}{n} \right) x + \theta^{*2} = 0$$

Subrayemos una vez más que aquí, del mismo modo que en el p. 22.2.1, el intervalo confidencial está construido aproximadamente, sin embargo al crecer n el error tiende a cero.

22.3. Estimaciones de los parámetros de la distribución uniforme y de la distribución Γ

22.3.1. Distribución uniforme en el segmento con extremo fijo.

Sean x_1, x_2, \dots, x_n observaciones independientes de la magnitud aleatoria ξ , distribuida uniformemente en el segmento $[0, \theta]$, con el parámetro θ desconocido. Pongamos $x_n^* = \max_{1 \leq h \leq n} x_h$. Es fácil ver que la estimación x_n^* es suficiente para el parámetro θ y la función de verosimilitud

$$L(\theta, x_1, x_2, \dots, x_n) = \begin{cases} \frac{1}{\theta^n} & \text{si } \max_{1 \leq i \leq n} x_i \leq \theta, \quad x_i \geq 0; \\ 0 & \text{en los demás casos} \end{cases}$$

adquiere el valor máximo para $\theta = x_n^*$. No obstante, la estimación x_n^* es sesgada. La estimación insesgada del parámetro θ será

$$\theta^* = \frac{n+1}{n} x_n^*.$$

En este caso θ^* tendrá la varianza menor entre todas las estimaciones insesgadas, igual a

$$M_{\theta}(\theta^* - \theta)^2 = \frac{\theta^2}{n(n+2)}.$$

Designemos con $Q_{\theta}(dx_1, dx_2, \dots, dx_n)$ la medida en R^n , cuya densidad respecto de la medida de Lebesgue es igual a $L(\theta, x_1, x_2, \dots, x_n)$. Señalemos que la distribución de la magnitud $\frac{\theta^*}{\theta}$ no depen-

de de θ :

$$Q_0 \left\{ \frac{\theta^*}{\theta} < \alpha \right\} = \begin{cases} 0, & \text{si } \alpha \leq 0; \\ \left(\frac{\alpha n}{n+1} \right)^n, & \text{si } \alpha \in \left[0, 1 + \frac{1}{n} \right]; \\ 1, & \text{si } \alpha > 1 + \frac{1}{n}. \end{cases}$$

Según $\varepsilon > 0$ dado escogemos a_ε de tal modo que

$$1 - \varepsilon = Q_0 \left\{ \frac{\theta^*}{\theta} > a_\varepsilon \right\} = 1 - \left(\frac{a_\varepsilon n}{n+1} \right)^n,$$

de donde obtendremos $a_\varepsilon = \frac{n+1}{n} \sqrt[n]{1-\varepsilon}$. Así pues,

$$Q_0 \left\{ x_n^* < \theta < \frac{x_n^*}{\sqrt[n]{1-\varepsilon}} \right\} = 1 - \varepsilon,$$

es decir, el intervalo $\left(x_n^*, \frac{x_n^*}{\sqrt[n]{1-\varepsilon}} \right)$ es confidencial con coeficiente de confianza $1 - \varepsilon$.

22.3.2. Distribución uniforme en un segmento con extremos desconocidos. Si la magnitud ξ tiene una distribución uniforme en el segmento $[\theta_1, \theta_2]$ con parámetros desconocidos θ_1, θ_2 ($0 < \theta_1 < \theta_2$), entonces las estimaciones insesgadas con la varianza mínima para los parámetros θ_1 y θ_2 serán, respectivamente, las estimaciones

$$\hat{\theta}_1^* = \frac{nx_1^*}{n-1} - \frac{x_n^*}{n-1};$$

$$\hat{\theta}_2^* = \frac{nx_n^*}{n-1} - \frac{x_1^*}{n-1};$$

donde $x_1^* = \min_{1 \leq k \leq n} x_k$, $x_n^* = \max_{1 \leq k \leq n} x_k$. Las estimaciones insesgadas con varianza mínima para el punto medio $\frac{\theta_1 + \theta_2}{2}$ y el recorrido $\theta_2 - \theta_1$ serán, respectivamente, las estimaciones

$$\hat{\theta}_1^* = \frac{x_n^* + x_1^*}{2};$$

$$\hat{\theta}_2^* = \frac{n-1}{n-1} (x_n^* - x_1^*).$$

La distribución conjunta de las magnitudes x_1^* y x_n^* se determina por la densidad

$$f(x_1, x_2) = n(n-1)(x_2 - x_1)^{n-2}(\theta_2 - \theta_1)^{-n},$$

donde $\theta_1 \leq x_1 \leq x_2 \leq \theta_2$.

22.3.3. Estimación del parámetro de escala en la distribución Γ . Supongamos que la magnitud aleatoria ξ tiene una distribución Γ

$$P_{\theta}(dx) = \begin{cases} \frac{x^{\lambda-1} e^{-\frac{x}{\theta}}}{\Gamma(\lambda) \theta^{\lambda}} dx, & \text{si } x > 0, \\ 0, & \text{si } x \leq 0 \end{cases}$$

con el parámetro desconocido $\theta \in (0, \infty)$ y el conocido $\lambda \in (0, \infty)$. Señalemos que para $\lambda = 1$ esta distribución se convierte en la distribución exponencial

$$P_{\theta}^0(dx) = \begin{cases} 0^{-1} e^{-\frac{x}{\theta}} dx, & \text{si } x > 0, \\ 0, & \text{si } x \leq 0, \end{cases}$$

Sean x_1, x_2, \dots, x_n las observaciones independientes de la magnitud aleatoria ξ . La estimación eficiente del parámetro θ es

$$\theta^* = \frac{1}{n\lambda} \sum_{h=1}^n x_h,$$

además

$$M_0(\theta^* - \theta)^2 = \frac{\theta^2}{n\lambda}.$$

Indiquemos, luego, que la distribución de la magnitud $\frac{n\lambda\theta^*}{\theta}$ es también distribución Γ independiente de θ :

$$Q_0 \left\{ \frac{n\lambda\theta^*}{\theta} < x \right\} = \begin{cases} \int_0^x \frac{y^{n\lambda-1} e^{-y}}{\Gamma(n\lambda)} dy, & \text{si } x \geq 0, \\ 0, & \text{si } x < 0, \end{cases}$$

donde la medida $Q_0(dx_1, dx_2, \dots, dx_n)$ en R^n se determina por la fórmula

$$Q_0(dx_1, dx_2, \dots, dx_n) = \begin{cases} \prod_{h=1}^n \frac{x_h^{\lambda-1} e^{-\frac{x_h}{\theta}}}{\theta^{\lambda} \Gamma(\lambda)} dx_h, & \text{si } x_h \geq 0, h = 1, 2, \dots, n; \\ 0 & \text{en el caso contrario.} \end{cases}$$

De este modo, al obtener los números a_1 y a_2 de la correlación

$$Q_0 \left\{ a_1 + n\lambda < \frac{n\lambda\theta^*}{\theta} < a_2 + n\lambda \right\} = 1 - \varepsilon,$$

determinamos en intervalo confidencial $\left(\frac{n\lambda\theta^*}{n\lambda + a_2}, \frac{n\lambda\theta^*}{n\lambda + a_1} \right)$ con coeficiente de confianza $1 - \varepsilon$.

22.3.4. Estimación del valor medio de la distribución Γ . Supongamos otra vez que la magnitud aleatoria ξ tiene a título de su distribución la distribución Γ^θ

$$P_\theta(dx) = \begin{cases} \frac{x^{\theta-1} e^{-x}}{\Gamma(\theta)} dx, & \text{si } x > 0, \\ 0, & \text{si } x \leq 0 \end{cases}$$

con el parámetro desconocido θ , $0 < \theta < \infty$. No es difícil calcular que en este caso

$$I(\theta) = \frac{d^2 \log \Gamma(\theta)}{d\theta^2}.$$

El método de momentos lleva a la estimación $\theta^* = \frac{1}{n} \sum_{k=1}^n x_k$ donde x_1, x_2, \dots, x_n son observaciones independientes de la magnitud aleatoria ξ . Para ella $M_0 \theta^* = \theta$ y $M_0 (\theta^* - \theta)^2 = \frac{\theta}{n}$. La eficiencia de la estimación θ_1^* se determina por la fórmula

$$\text{eff}(\theta_1^*) = \frac{1}{\theta \frac{d^2 \log \Gamma(\theta)}{d\theta^2}},$$

así que ésta no depende de n y siempre es menor que uno. La ecuación de verosimilitud tiene el aspecto de

$$\frac{1}{n} \sum_{k=1}^n \log x_k - \frac{d \log \Gamma(\theta)}{d\theta} = 0.$$

La única raíz positiva de esta ecuación define la estimación de verosimilitud máxima θ^* . Es asintóticamente eficiente y es asintóticamente normal con parámetros $(\theta, [n \frac{d^2 \log \Gamma(\theta)}{d\theta^2}]^{-1/2})$.

MÉTODO DE LOS CUADRADOS MÍNIMOS

23.1. Estimaciones del método de los cuadrados mínimos

23.1.1. Principio de los cuadrados mínimos. Sean representados los resultados de las observaciones de algún parámetro desconocido $\theta = (\theta_1, \theta_2, \dots, \theta_m)'$ del cual conocemos que puede tomar los valores del dominio Θ del espacio m -dimensional R^m , en forma de

$$\left. \begin{aligned} y_1 &= f_1(\theta, \varepsilon_1); \\ y_2 &= f_2(\theta, \varepsilon_2); \\ &\vdots \\ y_n &= f_n(\theta, \varepsilon_n), \end{aligned} \right\} \quad (1.1)$$

donde $f_k(\theta, \varepsilon_k)$ son funciones conocidas salvo el parámetro θ ; ε_k es el error (ruido) de la k -ésima observación. Se supone que la esperanza matemática del vector $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)'$ de los errores de la observación es igual al vector nulo: $M\varepsilon = \bar{0}$, lo que corresponde a la suposición sobre la ausencia de errores sistemáticos en la observación.

En la estadística matemática se llama método de los cuadrados mínimos el método de estimación estadística puntual (por puntos) fundado en el siguiente principio de los cuadrados mínimos: a título de estimación del parámetro desconocido θ en (1.1) se elige tal valor $\hat{\theta} \in \Theta$ para el cual se logra el mínimo de la suma de cuadrados:

$$s(\theta) = \sum_{k=1}^n [y_k - f_k(\theta, \varepsilon_k)]^2. \quad (1.2)$$

es decir,

$$\sum_{k=1}^n [y_k - f_k(\hat{\theta}, \varepsilon_k)]^2 = \min_{\theta \in \Theta} \sum_{k=1}^n [y_k - f_k(\theta, \varepsilon_k)]^2. \quad (1.3)$$

La estimación $\hat{\theta}$ que minimiza la suma de cuadrados $s(\theta)$ se llama estimación del método de los cuadrados mínimos o estimación MCM del parámetro θ .

Como método de estimación estadística puntual, el método de los cuadrados mínimos se utiliza ampliamente al manejar los resultados de mediciones, en el análisis de regresión, la planificación de experimentos, la econometría, etc.

A diferencia, por ejemplo, del método de verosimilitud máxima el empleo del método de los cuadrados mínimos no supone la infor-

nación apriorística sobre el tipo de la distribución de magnitudes aleatorias ε_k , $k = \overline{1, n}$. Sin embargo, en un modelo tan general como (1.1) las estimaciones MCM del parámetro no poseen propiedades óptimas, hasta cierto grado buenas, excepto el caso cuando ε_k están distribuidas normalmente.

23.1.2. Modelo lineal general. Existe una situación prácticamente importante cuando las estimaciones MCM, poseen, incluso para pequeños volúmenes de muestras, la propiedad de ser óptimas, lo que consiste en que éstas son insesgadas y tienen varianza mínima. Este es el caso del modelo lineal cuando $\Theta \in R^m$, $m < n$ o

$$\left. \begin{aligned} y_1 &= x_{11}\theta_1 + x_{12}\theta_2 + \dots + x_{1m}\theta_m + \varepsilon_1; \\ y_2 &= x_{21}\theta_1 + x_{22}\theta_2 + \dots + x_{2m}\theta_m + \varepsilon_2; \\ &\dots \\ y_n &= x_{n1}\theta_1 + x_{n2}\theta_2 + \dots + x_{nm}\theta_m + \varepsilon_n \end{aligned} \right\} \quad (1.4)$$

o en forma matricial

$$y = X\theta + \varepsilon, \quad (1.5)$$

donde $y = (y_1, y_2, \dots, y_n)'$ es el vector columna de las observaciones, $X = \{x_{ij}, i = \overline{1, n}, j = \overline{1, m}, m < n\}$, la matriz de coeficientes conocidos, $\theta = (\theta_1, \theta_2, \dots, \theta_m)'$, el vector columna de los parámetros desconocidos y a estimar, $\varepsilon = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)'$, es el vector columna de los errores aleatorios.

De ordinario se supone que $M\varepsilon = 0$ y $C(\varepsilon) = M[\varepsilon\varepsilon'] = \sigma^2 I$, donde σ^2 es la varianza (desconocida) de las magnitudes aleatorias incorrelacionadas igualmente distribuidas ε_k , $k = \overline{1, n}$, ya que si $M[\varepsilon\varepsilon'] = \delta^2 V$, donde V es una matriz conocida, la sustitución de

$$y_0 = V^{-\frac{1}{2}} y, \quad x_{0j} = V^{-\frac{1}{2}} x_j, \quad \varepsilon_0 = V^{-\frac{1}{2}} \varepsilon$$

lleva a $M[\varepsilon_0\varepsilon_0'] = \sigma^2 I$. La estimación del parámetro θ en (1.5) según el método de los cuadrados mínimos consiste en determinar el vector $\hat{\theta}$ que minimiza la suma de los cuadrados

$$s(\hat{\theta}) = \sum_{i=1}^n \left\{ y_i - \sum_{j=1}^m x_{ij}\hat{\theta}_j \right\}^2 = (y - X\hat{\theta})'(y - X\hat{\theta}) \quad (1.6)$$

23.1.3. Estimaciones MCM en el modelo lineal regular. Sea en el modelo lineal (1.5) del $X'X \neq 0$. Este modelo se llama regular.

Es estimación MCM $\hat{\theta}$ del parámetro θ la solución del sistema de ecuaciones normales

$$X'X\hat{\theta} = X'y, \quad (1.7)$$

es decir,

$$\hat{\theta} = (X', X)^{-1} X'y, \quad (1.8)$$

además, $M\hat{\theta} = \theta$, $M(\hat{\theta}\hat{\theta}') - (\hat{\theta} - \theta)(\hat{\theta} - \theta)' = \sigma^2 (X'X)^{-1}$.

La estimación insesgada para σ^2 es

$$s^2 = \frac{1}{n-m} (y - X\hat{\theta})'(y - X\hat{\theta}) = \frac{1}{n-m} \sum_{i=1}^n \left[y_i - \sum_{j=1}^m x_{ij}\hat{\theta}_j \right]^2, \quad (1.9)$$

23.1.4. Mediciones directas. Los resultados de las mediciones independientes reiteradas se representan en forma de

$$y_i = \theta + \varepsilon_i, \quad i = \overline{1, n},$$

donde θ es la magnitud que se mide, ε_i son los errores de las mediciones. La matriz X en el modelo lineal (1.5) para el caso dado representa un vector columna de unidades. Si $D\varepsilon_h = \sigma^2$ hablan de las mediciones equiexactas directas.

La estimación del MCM $\hat{\theta}$ para mediciones equiexactas directas tiene el aspecto de

$$\hat{\theta} = \frac{1}{n} \sum_{i=1}^n y_i, \quad D\hat{\theta} = \frac{\sigma^2}{n}. \quad (1.10)$$

La estimación insesgada para la varianza σ^2 tiene forma de

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \hat{\theta})^2.$$

Si $D\varepsilon_h = \frac{\sigma^2}{p_h}$, donde p_h son conocidos y, hablando en general, son distintos, se habla de mediciones no equiexactas directas.

La estimación MCM $\hat{\theta}$ para las mediciones no equiexactas directas tiene el aspecto

$$\hat{\theta} = \frac{\sum_{i=1}^n p_i y_i}{\sum_{i=1}^n p_i}, \quad (1.11)$$

además,

$$D\hat{\theta} = \frac{\sigma^2}{\sum_{h=1}^n p_h}.$$

a estimación insesgada para σ^2 es

$$s^2 = \frac{1}{(n-1) \sum_{h=1}^n p_h} \sum_{h=1}^n p_h (y_h - \hat{\theta})^2. \quad (1.12)$$

23.1.5. Estimaciones MCM en el modelo lineal degenerado. Si el modelo lineal (1.5) es degenerado, es decir, $\det X'X = 0$, es cómodo definir las estimaciones MCM en términos de las matrices inversas generalizadas, también llamadas pseudoinversas.

Sea Π una matriz $m \times m$ de proyección que se define unívocamente por las desigualdades

$$\Pi X'X - X'X\Pi = 0, \quad \Pi^2 = \Pi.$$

Se llama **seudoinversa (inversa generalizada)** de la matriz $X'X$ la matriz $(X'X)^{(-)}$ definida por la igualdad

$$(X'X)^{(-)} = (X'X + \Pi)^{-1} - \Pi. \quad (1.13)$$

Si $\det X'X \neq 0$, $\Pi = 0$ y, por consiguiente, $(X'X)^{(-)} = (X'X)^{-1}$, es decir, en este caso la matriz pseudoinversa coincide con la inversa.

Si el modelo lineal es degenerado, el método de los cuadrados mínimos define el subespacio de vectores en el que se consigue el mínimo de la suma de cuadrados $s(\theta)$ en (1.6) y el cual tiene el aspecto

$$(X'X)^{(-)}X'y + \Pi h, \quad (1.14)$$

donde h es un vector arbitrario de R^m . Por esta causa, en el modelo lineal degenerado se estima no el parámetro θ , sino una función lineal $B\hat{\theta}$ de este parámetro, donde $B = \{b_{ij}, i=1, k, j=1, m\}$ es la matriz $k \times m$ definida, k es un número arbitrario, pero fijado.

La función lineal $B\hat{\theta}$ del parámetro θ se llama **estimadora** si $B\Pi = 0$. La estimación MCM $B\hat{\theta}$ de la función estimadora $B\theta$ tiene por expresión

$$B\hat{\theta} = B(X'X)^{(-)}X'y, \quad (1.15)$$

además $MB\hat{\theta} = B\theta$. $M \{B\hat{\theta} - B\theta\} [B\hat{\theta} - B\theta]' = \sigma^2 B(X'X)^{(-)}B'$.

La estimación insesgada para σ^2 es

$$s^2 = \frac{1}{n-r} (y - X\hat{\theta})'(y - X\hat{\theta}) = \frac{1}{n-r} \sum_{i=1}^n \left[y_i - \sum_{j=1}^m x_{ij}\hat{\theta}_j \right]^2, \quad (1.16)$$

donde $\hat{\theta} = (X'X)^{(-)}X'y$, r es el rango de la matriz $X'X$.

Si $B_1\hat{\theta}$ y $B_2\hat{\theta}$ son dos funciones estimadoras,

$$\text{Cov}(B_1\hat{\theta}, B_2\hat{\theta}) = \sigma^2 B_1(X'X)^{(-)}B_2'. \quad (1.17)$$

23.1.6. Modelos lineales con parámetros aleatorios. Analicemos el modelo lineal de la forma

$$y = X\theta + \varepsilon, \quad (1.18)$$

donde $y = (y_1, y_2, \dots, y_n)'$ es un vector de los valores a observar, $X = \{x_{ij}, j=1, n, i=1, m\}$, una matriz de coeficientes conocidos, $\theta = (\theta_1, \theta_2, \dots, \theta_m)'$, un vector aleatorio desconocido con esperanza matemática $M\theta = a$ y una matriz de covariación $M(\theta - a)(\theta - a)' = C$, $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)'$ es un vector de errores independiente de θ , $M\varepsilon = 0$, $M\varepsilon\varepsilon' = \sigma^2 I$.

El problema de la estimación del vector desconocido θ por el método de los cuadrados mínimos consiste en determinar la función lineal $\hat{\theta}$ del vector y de los valores a observar, en la cual la magnitud

aleatoria

$$s(\hat{\theta}) = (y - X\hat{\theta})' (y - X\hat{\theta}) = \sum_{j=1}^n [y_j - \sum_{i=1}^m x_{ij}\hat{\theta}_i]^2 \quad (1.19)$$

tiene la esperanza matemática mínima. Si el vector a de las esperanzas matemáticas del parámetro θ es conocido, entonces

$$\hat{\theta} = (I - L)X a + L y, \quad (1.20)$$

donde $L = CX'(\sigma^2 I + X'CX)^{-1}$, además, $M\hat{\theta} = a$, $M(\hat{\theta} - \theta)(\hat{\theta} - \theta)' = C - CX'(\sigma^2 I + X'CX)^{-1}XC$.

Si el vector a es desconocido, entonces

$$\hat{\theta} = (X'X)^{-1}X'y \quad (1.21)$$

con la particularidad de que $M\hat{\theta} - \theta' = \sigma^2(X'X)^{-1}$ y la estimación MCM del parámetro aleatorio θ es la misma que en el caso del parámetro determinado.

Esto es cierto, si incluso $\det X'X = 0$. Naturalmente, aquí, como en el caso determinado, hay que considerar las funciones lineales a estimar del parámetro θ y en vez de $(X'X)^{-1}$ usar la matriz pseudoinversa $(X'X)^{(-1)}$.

23.1.7. Propiedades de las estimaciones MCM. Las estimaciones MCM en el modelo lineal general

$$y = X\theta + \varepsilon, \quad M\varepsilon = 0, \quad \text{Cov } \varepsilon = \sigma^2 I, \quad \det X'X \neq 0 \quad (1.22)$$

poseen las siguientes propiedades.

- La estimación MCM $\hat{\theta}$ es insesgada.
- Si $\lim_{n \rightarrow \infty} \frac{1}{n}(X'X)$ es una matriz no degenerada, la estimación

MCM $\hat{\theta}$ es conciliable y el vector $\sqrt{n}(\hat{\theta} - \theta)$ es asintóticamente normal.

c) **Teorema de Gauss-Márkov.** La estimación MCM $\hat{\theta}$ tiene varianza mínima en la clase de todas las estimaciones lineales insesgadas del parámetro θ , con la particularidad de que $\hat{\theta}$ e $y - X\hat{\theta}$ son incorrelacionados.

En la clase de las estimaciones insesgadas la estimación MCM $\hat{\theta}$ en el modelo lineal no posee, hablando en general, varianza mínima.

Si el vector ε en (1.22) está distribuido normalmente, la estimación MCM $\hat{\theta}$ del parámetro θ , además de las propiedades a), b), c), posee las siguientes.

- $\hat{\theta}$ es la estimación del método de verosimilitud máxima.
- $\hat{\theta}$ tiene varianza mínima en la clase de todas las estimaciones insesgadas del parámetro θ .
- $\hat{\theta}$ tiene distribución normal con esperanza matemática θ y matriz de covariación $\sigma^2(X'X)^{-1}$.

g) La magnitud aleatoria

$$\frac{(y - X\hat{\theta})'(y - X\hat{\theta})}{\sigma^2} = \frac{1}{\sigma^2} \sum_{i=1}^n \left[y_i - \sum_{j=1}^m x_{ij}\hat{\theta}_j \right]^2$$

tiene distribución χ^2 con $n - m$ grados de libertad.

h) Las estimaciones $\hat{\theta}$ y $s^2 = \frac{1}{n-m} (y - X\hat{\theta})'(y - X\hat{\theta})$ son suficientes para la estimación de θ y σ^2 .

i) $\hat{\theta}$ e $y - X\hat{\theta}$ son independientes.

23.2. Modelos lineales de regresión

23.2.1. Nivelación de los datos. En las aplicaciones de la estadística a menudo se exige estimar el carácter de la dependencia entre las variables estadísticas observadas, por ejemplo, entre los parámetros de los procesos tecnológicos y la producción acabada, la luminosidad de las estrellas y sus dimensiones, la cantidad de precipitaciones atmosféricas en sector dado y el rendimiento de la cosecha en éstos, etc.

En este caso, el problema fundamental consiste en la nivelación (alisamiento) de los datos experimentales con ayuda de curvas especialmente elegidas llamadas líneas o superficies de regresión que con mayor o menor seguridad caracterizan la dependencia de correlación entre las variables en observación.

Sean η y ξ dos magnitudes aleatorias (hablando en general, dependientes).

Definición. Las funciones

$$f_{\eta/\xi}(x) = M[\eta/\xi = x], \quad g_{\xi/\eta}(y) = M[\xi/\eta = y]$$

se llaman regresiones (teóricas) de η en ξ y de ξ en η , respectivamente.

En el caso de regresión de η en ξ , por ejemplo, la magnitud aleatoria ξ se considera independiente y se llama variable de regresión.

La regresión tiene la siguiente propiedad de ser óptima: las magnitudes $M[\eta - \varphi(\xi)]^2$ y $M[\xi - \psi(\eta)]^2$ alcanzan el mínimo con $\varphi(x) = f_{\eta/\xi}(x)$ y $\psi(y) = g_{\xi/\eta}(y)$, respectivamente.

En calidad de estimaciones estadísticas de las regresiones teóricas se eligen, como regla, las funciones linealmente dependientes de los parámetros desconocidos a estimar. Con este fin a menudo se usan diferentes escalas funcionales.

Las funciones más usadas en el análisis de regresión estadística son

$$y = \theta_0 + \theta_1 x;$$

$$y = \theta_0 + \theta_1 x + \dots + \theta_k x^k;$$

$$y = \theta_0 + \theta_1 \ln x;$$

$$y = \theta_0 + \theta_1 x + \theta_2 \frac{1}{x};$$

$$y = \frac{1}{\theta_0 + \theta_1 x};$$

$$y = \frac{i}{\theta_0 + \theta_1 x + \dots + \theta_n x^n}$$

$$\ln y = \theta_0 + \theta_1 x;$$

$$\ln y = \theta_0 + \theta_1 \ln x.$$

23.2.2. Estimación de los parámetros de regresión lineal. Sean x_i el valor de la variable de regresión e y_i el valor de la variable dependiente que se obtiene en la i -ésima observación, $i = \overline{1, n}$. El modelo lineal de regresión puede ser representado en la forma

$$p(y_i) = r(\theta, x_i) + \varepsilon_i, \quad i = \overline{1, n}, \quad (2.1)$$

donde $p(y)$ y $r(\theta, x)$ son funciones que se consideran preferibles para describir la relación estadística entre la variable de regresión y la variable dependiente, con la particularidad de que $r(\theta, x)$ es lineal según $\theta = (\theta_1, \dots, \theta_m)'$, ε_i son magnitudes aleatorias incorrelacionadas con esperanzas matemáticas nulas.

La estimación del vector $\theta = (\theta_1, \dots, \theta_m)'$ en (2.1) puede ser obtenida por el método de los cuadrados mínimos como el vector que minimiza la suma de cuadrados

$$s(\theta) = \sum_{i=1}^n [p(y_i) - r(\theta, x_i)]^2. \quad (2.2)$$

El modelo de regresión (2.1) es un caso particular del modelo lineal inesgado $y = X\theta + \varepsilon$, det $X'X \neq 0$. Por eso, si la matriz de covarianza $\text{Mse}' = \sigma^2 I$, entonces la estimación MCM $\hat{\theta}$ del parámetro θ en $r(\theta, x)$ es la solución del sistema de las ecuaciones normales $X_r' X_r \hat{\theta} = X_r' y_p$, y, por consiguiente,

$$\hat{\theta} = (X_r' X_r)^{-1} X_r' y_p, \quad (2.3)$$

donde $y_p = (p(y_1), p(y_2), \dots, p(y_n))$. La función $y = p^{-1}(r(\theta, x))$ describe la línea de regresión, estadística o muestral.

EJEMPLO 1. Regresión lineal simple. Sean $p(y) = y$, $r(\theta, x) = \theta_1 + \theta_2 x$. Entonces

$$X_r = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \dots & \dots \\ 1 & x_n \end{pmatrix}, \quad \theta = \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix};$$

$$\hat{\theta} = \begin{pmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \end{pmatrix} = \begin{pmatrix} \frac{\sum x_i^2 \sum y_i - \sum x_i \sum x_i y_i}{n \sum x_i^2 - (\sum x_i)^2} \\ \frac{n \sum x_i y_i - \sum x_i \sum y_i}{n \sum x_i^2 - (\sum x_i)^2} \end{pmatrix}.$$

La línea de regresión muestral es

$$y = \hat{\theta}_1 + \hat{\theta}_2 x.$$

EJEMPLO 2. Regresión polinomial. Sean $p(y) = y$, $r(\theta, x) = \theta_1 + \theta_2 x + \theta_3 x^2 + \dots + \theta_m x^{m-1}$. Entonces

$$X_r = \begin{pmatrix} 1 & x_1 x_1^2 & \dots & x_1^{m-1} \\ 1 & x_2 x_2^2 & \dots & x_2^{m-1} \\ \dots & \dots & \dots & \dots \\ 1 & x_n x_n^2 & \dots & x_n^{m-1} \end{pmatrix}, \quad \theta = \begin{pmatrix} \theta_1 \\ \theta_2 \\ \dots \\ \theta_m \end{pmatrix}.$$

La estimación MCM $\hat{\theta}$ es la solución del sistema de las ecuaciones normales

$$\begin{pmatrix} n & \sum x_i & \sum x_i^2 & \dots & \sum x_i^{m-1} \\ \sum x_i & \sum x_i^2 & \sum x_i^3 & \dots & \sum x_i^m \\ \dots & \dots & \dots & \dots & \dots \\ \sum x_i^{m-1} & \sum x_i^m & \sum x_i^{m+1} & \dots & \sum x_i^{2(m-1)} \end{pmatrix} \begin{pmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \\ \vdots \\ \hat{\theta}_m \end{pmatrix} = \begin{pmatrix} \sum y_i \\ \sum x_i y_i \\ \vdots \\ \sum x_i^{m-1} y_i \end{pmatrix}.$$

La línea de regresión muestral es

$$y = \hat{\theta}_1 + \hat{\theta}_2 x + \hat{\theta}_3 x^2 + \dots + \hat{\theta}_m x^{m-1}.$$

EJEMPLO 3. Regresión exponencial. Sean $p(y) = \ln y$, $r(\theta, x) = \theta_1 + \theta_2 x$. Entonces

$$y_p = \begin{pmatrix} \ln y_1 \\ \ln y_2 \\ \dots \\ \ln y_n \end{pmatrix}, \quad X_r = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \dots & \dots \\ 1 & x_n \end{pmatrix}, \quad \theta = \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix},$$

$$\hat{\theta} = \begin{pmatrix} \frac{n \sum x_i \ln y_i - \sum x_i \sum \ln y_k}{n \sum x_i^2 - (\sum x_i)^2} \\ \frac{\sum x_i^2 \sum \ln y_k - \sum x_i \sum x_k \ln y_k}{n \sum x_i^2 - (\sum x_i)^2} \end{pmatrix}.$$

La línea de regresión muestral es

$$y = e^{\hat{\theta}_1 + \hat{\theta}_2 x}.$$

23.2.3. Regresión polinomial. La aplicación práctica de la regresión polinomial tiene un defecto esencial: el aumento del grado de polinomio de alisamiento exige realizar de nuevo todos los cálculos.

En este caso es útil emplear los polinomios ortogonales de Chebichev. En vez de estimar el parámetro $\theta = (\theta_1, \dots, \theta_m)'$ en el

modelo

$$y_i = \theta_1 + \theta_2 x_i + \theta_3 x_i^2 + \dots + \theta_m x_i^{m-1} + v_i$$

se estima el parámetro $\alpha = (\alpha_0, \alpha_1, \dots, \alpha_{m-1})'$ en el modelo

$$y_i = \alpha_0 \varphi_0(x_i) + \alpha_1 \varphi_1(x_i) + \dots + \alpha_{m-1} \varphi_{m-1}(x_i) + v_i, \quad (2.4)$$

donde $\varphi_k(x)$ son polinomios de grado k , que poseen la siguiente propiedad de ortogonalidad:

$$\sum_{i=1}^n \varphi_k(x_i) \varphi_l(x_i) = 0, \quad k \neq l.$$

El modelo (2.4) en forma matricial tiene el aspecto

$$y = \Phi \alpha + v, \quad (2.5)$$

donde

$$\Phi = \{\varphi_{ij}, \quad i = \overline{1, n}, \quad j = \overline{0, m-1}\}, \quad \varphi_{ij} = \varphi_j(x_i)$$

Para el parámetro α la estimación MCM tiene la forma

$$\hat{\alpha} = (\Phi' \Phi)^{-1} \Phi' y, \quad (2.6)$$

es decir,

$$\hat{\alpha}_h = \frac{\sum_{i=1}^n \varphi_h(x_i) y_i}{\sum_{i=1}^n \varphi_h^2(x_i)}.$$

Los polinomios ortogonales de Chébishev $\varphi_h(x)$ se calculan por la fórmula

$$\varphi_0(x) \equiv 1, \quad \varphi_{k+1}(x) = \frac{\det M_k(x)}{\det M_k}, \quad k = \overline{0, m-2}, \quad (2.7)$$

donde

$$M_k = \begin{pmatrix} n \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 & \dots & \sum_{i=1}^n x_i^k \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 & \sum_{i=1}^n x_i^3 & \dots & \sum_{i=1}^n x_i^{k+1} \\ \dots & \dots & \dots & \dots & \dots \\ \sum_{i=1}^n x_i^k & \sum_{i=1}^n x_i^{k+1} & \sum_{i=1}^n x_i^{k+2} & \dots & \sum_{i=1}^n x_i^{2k} \end{pmatrix};$$

$$M_h(x) = \begin{pmatrix} n & \sum_{i=1}^n x_i & \dots & \sum_{i=1}^n x_i^{h+1} \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 & \dots & \sum_{i=1}^n x_i^{h+2} \\ \dots & \dots & \dots & \dots \\ \sum_{i=1}^n x_i^h & \sum_{i=1}^n x_i^{h+1} & \dots & \sum_{i=1}^n x_i^{2h+1} \\ 1 & x & \dots & x^{h+1} \end{pmatrix}$$

o bien de las relaciones recurrentes $\varphi_0(x) \equiv 1$,

$$\varphi_h(x) = x^h - \frac{\sum_{i=1}^n x_i^h \varphi_{h-1}(x_i)}{\sum_{i=1}^n \varphi_{h-1}^2(x_i)} \varphi_{h-1}(x \dots) - \dots - \frac{\sum_{i=1}^n x_i^h \varphi_0(x_i)}{\sum_{i=1}^n \varphi_0^2(x_i)} \varphi_0(x). \quad (2.8)$$

EJEMPLO 4.

$$\varphi_1(x) = \frac{\det \begin{vmatrix} n & \sum_{i=1}^n x_i \\ 1 & x \end{vmatrix}}{\det |n|} = x - \bar{x},$$

$$\varphi_2(x) = \frac{\det \begin{vmatrix} n & \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 & \sum_{i=1}^n x_i^3 \\ 1 & x & x^2 \end{vmatrix}}{\det \begin{vmatrix} n & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 \end{vmatrix}} =$$

$$= x^2 - \frac{\sum_{i=1}^n (x_i - \bar{x}) x_i^2}{\sum_{i=1}^n (x_i - \bar{x})^2} (x - \bar{x}) - \frac{\sum_{i=1}^n x_i^2}{n},$$

donde

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i.$$

ESTADÍSTICA DE LOS PROCESOS ALEATORIOS

24.1. Distinción de las hipótesis

Al distinguir las hipótesis para los procesos aleatorios se observa la trayectoria del proceso aleatorio $x(t)$, $t \in [0, T]$; respecto de las distribuciones de dimensión finita hay algunas hipótesis una de las cuales debe ser elegida a base de la trayectoria en observación. Actualmente está elaborada la solución del problema más simple, o sea, de dos hipótesis elegir una. Una aplicación importante de este problema es la detección de la señal en el fondo de un ruido. Por ejemplo, al receptor llega una información. Es necesario determinar si existe o no en la información llegada una señal útil en el fondo de un ruido aleatorio.

24.1.1. Formulación general del problema de la distinción de dos hipótesis referentes a la distribución del proceso aleatorio. Se observa la trayectoria del proceso aleatorio $x(t)$, $t \in [0, T]$ sobre la cual es conocido que debe sin falta pertenecer a cierto espacio de las funciones $F_{[0, T]}$ dadas en $[0, T]$ (por ejemplo, al espacio de funciones continuas o al espacio de funciones sin discontinuidades de primera especie, etc.) Respecto del proceso aleatorio hay dos hipótesis H_i ($i = 1, 2$). Según la hipótesis H_i la medida correspondiente al proceso en el espacio $F_{[0, T]}$ es μ_i (ésta está dada en el σ -álgebra generada por conjuntos cilíndricos; véase el cap. 9) Valiéndose de la observación hay que resolver qué hipótesis es preferible.

Sea R cierta regla según la cual se acepta una u otra hipótesis. La forma más general de esta regla es la siguiente: para cada trayectoria posible $x(\cdot)$ debe estar prefijada la probabilidad $p(x(\cdot))$ (p es la funcional medible de la trayectoria) de aceptar la hipótesis H_1 , si se observa la trayectoria $x(\cdot)$; $1 - p(x(\cdot))$ es la probabilidad de aceptar la hipótesis H_2 . Esta regla se caracteriza por las probabilidades de errores $\alpha_{12} = P(H_1/H_2)$, es decir, aceptar la hipótesis H_1 si H_2 es cierta; $\alpha_{21} = P(H_2/H_1)$, o sea, aceptar la hipótesis H_2 si H_1 es cierta. Las expresiones para α_{12} y α_{21} están dadas por las fórmulas

$$\alpha_{12} = \int p(x(\cdot)) \mu_2(dx);$$

$$\alpha_{21} = \int (1 - p(x(\cdot))) \mu_1(dx).$$

Es natural buscar tales reglas que hagan mínimas las probabilidades de errores. Como veremos más adelante, en muchos casos podemos

limitarnos a las reglas no randomizadas para las cuales $p(x(\cdot))$ toma sólo los valores 0 ó 1. Entonces, $F_{[0, T]}$ se parte en dos subconjuntos: G_1 y $G_2 = F_{[0, T]} - G_1$; si $x(\cdot) \in G_1$, se toma la hipótesis H_1 ; si $x(\cdot) \in G_2$, se adopta H_2 . Las probabilidades de errores están dadas por la expresión

$$\alpha_{ij} = \int \mu_j(G_i) \quad (i, j = 1, 2).$$

24.1.2. Continuidad absoluta de las medidas en los espacios funcionales. Al resolver los problemas de estadística de los procesos aleatorios juega un papel importante la continuidad absoluta o la singularidad de las medidas correspondientes a los procesos en observación. Sean $F_{[0, T]}$ un espacio fijado; $\mathfrak{F}[0, T]$, una σ -álgebra de los subconjuntos de este espacio, generada por conjuntos cilíndricos; μ_1 y μ_2 , dos medidas en $\mathfrak{F}[0, T]$ que corresponden a los procesos aleatorios (o al mismo proceso para distintas hipótesis). Las medidas μ_1 y μ_2 son singulares entre sí (ortogonales) si existen tales conjuntos disjuntos $G_i \in \mathfrak{F}[0, T]$ que $\mu_1(G_2) = \mu_2(G_1) = 0$. La medida μ_2 es absolutamente continua respecto de μ_1 , si $\mu_2(C) = 0$ para todas las $C \in \mathfrak{F}[0, T]$ para las cuales $\mu_1(C) = 0$. En este caso existe una función $\mathfrak{F}[0, T]$ -medible $\rho(x(\cdot))$ tal que para cada $C \in \mathfrak{F}[0, T]$

$$\mu_2(C) = \int_C \rho(x) \mu_1(dx). \quad (1.1)$$

Esta función $\rho(x(\cdot))$ se llama densidad de la medida μ_2 respecto de μ_1 . Si μ_1 es también absolutamente continua respecto de μ_2 , las medidas μ_1 y μ_2 se llaman equivalentes, cuando y sólo cuando la función $\rho(x(\cdot))$ de (1.1) sea positiva casi en todos los puntos de la medida μ_1 . En este caso

$$\mu_1(C) = \int_C \frac{1}{\rho(x)} \mu_2(dx). \quad (1.2)$$

Para cualesquiera dos medidas (finitas) μ_1 y μ_2 se pueden indicar tales conjuntos disjuntos dos a dos Δ_1, Δ_2 y Δ que $\mu_1(\Delta_2) = \mu_2(\Delta_1) = 0$, y que en Δ estas medidas son equivalentes, es decir, existe tal función medible $\rho(x(\cdot))$, definida en Δ , que para cada C

$$\left. \begin{aligned} \mu_2(C \cap \Delta) &= \int_{C \cap \Delta} \rho(x) \mu_1(dx); \\ \mu_1(C \cap \Delta) &= \int_{C \cap \Delta} \rho^{-1}(x) \mu_2(dx). \end{aligned} \right\} \quad (1.3)$$

24.1.3. El criterio de Neumann — Pearson da la regla que distingue las hipótesis y según la cual para $\alpha_{12} \leq \varepsilon$ ($0 < \varepsilon < 1$) α_{21} es la mínima. La regla mencionada es óptima en cierto sentido. Efectivamente, sea que existe una regla con probabilidades de errores $\bar{\alpha}_{12}$ y $\bar{\alpha}_{21}$.

Con ayuda del criterio de Neumann—Pearson se puede construir una regla según la cual $\alpha_{12} \leq \bar{\alpha}_{12}$ y α_{21} es la mínima posible (es decir, $\alpha_{21} \leq \bar{\alpha}_{21}$).

Describamos el criterio en función de las propiedades de continuidad absoluta mutua de las medidas μ_1 y μ_2 , correspondientes al proceso en observación según las hipótesis H_1 y H_2 .

a. Sean μ_1 y μ_2 ortogonales. Entonces se puede indicar tal conjunto G_1 que $\mu_1(G_1) = 1$, $\mu_2(G_1) = 0$. Si $x(\cdot) \in G_1$, aceptamos la hipótesis H_1 ; si $x(\cdot) \in G_2$ aceptamos la hipótesis H_2 . Las probabilidades de errores son

$$\alpha_{12} = \mu_2(G_1) = 0,$$

$$\alpha_{21} = \mu_1(F_{[0, T]} - G_1) = 1 - \mu_1(G_1) = 0.$$

De este modo, en el caso considerado puede existir una regla infalible de distinción de las hipótesis, es decir, tal regla con la que $\alpha_{12} = \alpha_{21} = 0$.

b. Sean μ_1 y μ_2 equivalentes y sea $\rho(x(\cdot))$ la densidad de μ_2 respecto de μ_1 , $\rho(x(\cdot)) > 0$ para casi todas las $x(\cdot)$ según la medida μ_1 . Indiquemos

$$R_\lambda = \{x(\cdot) : \rho(x(\cdot)) < \lambda\}; \quad \Gamma_\lambda = \{x(\cdot) : \rho(x(\cdot)) = \lambda\}.$$

Satisfaga $\bar{\lambda}$ la relación

$$\mu_2(R_{\bar{\lambda}}) \leq \varepsilon, \quad \mu_2(R_{\bar{\lambda}} \cup \Gamma_{\bar{\lambda}}) \geq \varepsilon.$$

Para $\varepsilon \in (0, 1)$ esta $\bar{\lambda}$ existe, ya que para λ creciente desde 0 hasta ∞ , $\mu_1(R_\lambda)$ crece de 0 hasta 1. Analicemos tres subcasos:

1) $\mu_2(R_{\bar{\lambda}}) = \varepsilon$. Pongamos $G_1 = R_{\bar{\lambda}}$, $G_2 = F_{[0, T]} - G_1$, en este caso $\alpha_{12} = \mu_2(R_{\bar{\lambda}}) = \varepsilon$,

$$\alpha_{21} = \mu_1(F_{[0, T]} - G_1) = 1 - \mu_1(R_{\bar{\lambda}});$$

2) $\mu_2(R_{\bar{\lambda}}) < \varepsilon$, $\mu_2(R_{\bar{\lambda}} \cup \Gamma_{\bar{\lambda}}) = \varepsilon$, en este caso $G_1 = R_{\bar{\lambda}} \cup \Gamma_{\bar{\lambda}}$, $G_2 = F_{[0, T]} - G_1$, $\alpha_{12} = \varepsilon$, $\alpha_{21} = 1 - \mu_1(R_{\bar{\lambda}}) - \mu_1(\Gamma_{\bar{\lambda}})$;

3) $\mu_2(R_{\bar{\lambda}}) < \varepsilon$, $\mu_2(R_{\bar{\lambda}} \cup \Gamma_{\bar{\lambda}}) > \varepsilon$, entonces se puede construir el siguiente criterio randomizado, dado por la funcional de probabilidad $p(x(\cdot))$: $p(x(\cdot)) = 1$, si $x(\cdot) \in R_{\bar{\lambda}}$; $p(x(\cdot)) = 0$, si $x(\cdot) \in$

$\in F_{[0, T]} - R_{\bar{\lambda}} - \Gamma_{\bar{\lambda}}$, $p(x(\cdot)) = \frac{\varepsilon - \mu_2(R_{\bar{\lambda}})}{\mu_2(\Gamma_{\bar{\lambda}})}$, si $x(\cdot) \in \Gamma_{\bar{\lambda}}$. Cabe señalar

que cuando la medida μ_2 (y también μ_1) es continua, es decir, no existen trayectorias con probabilidad positiva, entonces en el caso 3) también se puede construir el criterio no randomizado: existe tal $D \subset \Gamma_{\bar{\lambda}}$ que $\mu_2(D) = \varepsilon - \mu_2(R_{\bar{\lambda}})$. Poniendo $G_1 = R_{\bar{\lambda}} \cup D$, $G_2 = F_{[0, T]} - R_{\bar{\lambda}} - D$ obtenemos la regla con $\alpha_{12} = \varepsilon$ y α_{21} mínimo,

c. En el caso general se pueden indicar tales $\Delta_1, \Delta_2, \Delta$ que se verifiquen las igualdades (1.3) y $\mu_2(\Delta_1) = \mu_1(\Delta_2) = 0$. Sea

$$R_\lambda = \{x(\cdot) \in \Delta : \rho(x(\cdot)) < \lambda\} \cup \Delta_1; \Gamma_\lambda = \{x(\cdot) \in \Delta : \rho(x(\cdot)) = \lambda\}.$$

Si $\varepsilon \geq 1 - \mu_2(\Delta_2)$, entonces, eligiendo $G_1 = \Delta_1 \cup \Delta$, $G_2 = \Delta_2$, tendremos

$$\begin{aligned} \alpha_{12} &= \mu_2(\Delta_1 \cup \Delta) = 1 - \mu_2(\Delta_2) \leq \varepsilon, \\ \alpha_{21} &= \mu_1(\Delta_2) = 0. \end{aligned}$$

Si $\varepsilon < 1 - \mu_2(\Delta_2)$, elegimos $\bar{\lambda}$ de tal modo que $\mu_2(R_{\bar{\lambda}}) \leq \varepsilon$, $\mu_2(R_{\bar{\lambda}} \cup \Gamma_{\bar{\lambda}}) \geq \varepsilon$ y procedemos del mismo modo que en el caso b.

Así, pues, para construir el mejor criterio según Neumann-Pearson es necesario saber construir los conjuntos Δ_1 en los cuales están concentradas las medidas singulares entre sí y, en el caso de las medidas absolutamente continuas, la densidad de una medida respecto de la otra, así como hay que conocer la distribución de $\rho(x(\cdot))$ para una u otra hipótesis.

24.2. Distinción de las hipótesis para los procesos con incrementos independientes

24.2.1. Designaciones principales. Sea $x(t)$ la trayectoria que se observa del proceso y que está determinada en $[0, T]$, $x(t) \in R^1$, $x(\cdot) \in D_{[0, T]}$ donde $D_{[0, T]}$ es el espacio de las funciones sin discontinuidades de segunda especie. Según la hipótesis, H_k ($k = 1, 2$) $x(t)$ es un proceso continuo estocástico con incrementos independientes cuya función característica se da por la fórmula

$$\begin{aligned} M \exp \{i z x(t)\} &= \exp \left\{ i z \gamma_k(t) - \frac{1}{2} b_k(t) z^2 + \right. \\ &\left. + \int \left(e^{i z x} - 1 - \frac{i z x}{1+x^2} \right) \Pi_k(t, dx) \right\}, \end{aligned} \quad (2.1)$$

donde las funciones $\Pi_k(t, A)$ para todas las A borelianas que están a una distancia positiva del punto 0, son continuas y no decrecen, $\gamma_k(t)$ son continuas, $\gamma_k(0) = b_k(0) = \Pi_k(0, A) = 0$ (es decir, se supone que $x(0) = 0$; si no es así, se puede considerar la función $x(t) - x(0)$). Introduzcamos las medidas en los subconjuntos borelianos $[0, T] \times R^1$:

$$\pi_k(B) = \int_{(t, x) \in B} d_t \Pi_k(t, dx).$$

Luego, sean Λ_1, Λ_2 y Λ tales conjuntos disjuntos en $[0, T] \times R^1$ que $\Lambda_1 \cup \Lambda_2 \cup \Lambda = [0, T] \times R^1$, $\pi_1(\Lambda_2) = \pi_2(\Lambda_1) = 0$, y en Λ las medidas π_1 y π_2 son equivalentes: e iste tal función medible posi-

tiva $f(t, x)$ que

$$\pi_2(\Lambda \cap B) = \int_{\Lambda \cap B} f(t, x) \pi_1(dt \times dx);$$

$$\pi_1(\Lambda \cap B) = \int_{\Lambda \cap B} f^{-1}(t, x) \pi_2(dt \times dx).$$

Determinemos la medida empírica de saltos:

$$v_{x(\cdot)}(B) = \sum_t \chi_B(t, x(t+0) - x(t-0)) \quad (2.2)$$

(ya que el proceso no tiene discontinuidades de segunda especie, esta magnitud es finita para todos los B para los cuales $B \cap \{[0, T] \times (-\varepsilon, \varepsilon)\} = \emptyset$). Designemos con μ_h la medida correspondiente al proceso $x(t)$ en $D_{[0, T]}$ para la hipótesis H_h .

24.2.2. Condiciones de ortogonalidad. Distinción infalible de las hipótesis. Las medidas μ_1 y μ_2 son ortogonales si se cumple siquiera una de las condiciones:

- 1) $\pi_1(\Lambda_1) + \pi_2(\Lambda_2) = +\infty$;
- 2) para cierto $c \in (0, 1)$

$$\pi_2\{(s, x) : |1 - f(s-x)| > c\} = \infty;$$

- 3) $\int \frac{(1-f(s, x))^2}{1+f^2(s, x)} \pi_1(ds \times dx) = +\infty$;

- 4) $b_1(s) = b_2(t)$ (para cierto $t \in [0, T]$);

- 5) sean $\int \frac{(1-f(s, x))^2}{1+f^2(s, x)} \pi_1(ds \times dx) < \infty$

y $b_1(t) = b_2(t) = b(t)$ para todos los $t \in [0, T]$, entonces está definida la función

$$a_2(t) = \int \frac{y}{1+y^2} [II_1(t, dy) - II_2(t, dy)], \quad (2.3)$$

las medidas μ_h serán ortogonales si la función $\gamma(t) = \gamma_2(t) - \gamma_1(t) + a_2(t)$ no es absolutamente continua respecto de $b(t)$, o bien

$$\int_0^T \left(\frac{d\gamma(t)}{db(t)} \right)^2 db(t) = \infty. \quad (2.4)$$

Indiquemos las reglas infalibles de elección de la hipótesis en cada uno de estos casos.

1a) Sea $\pi_1(\Lambda_1) = +\infty$. Designemos con G_1 el conjunto de $x(\cdot) \in D_{[0, T]}$ para los cuales $v_{x(\cdot)}(\Lambda_1) > 0$. Si $x(\cdot) \in G_1$ aceptamos la hipótesis H_1 ; si $x(\cdot) \notin G_1$ admitimos la hipótesis H_2 .

1b) Sea $\pi_2(\Lambda_2) = +\infty$. Designemos con G_2 el conjunto de $x(\cdot) \in D_{[0, T]}$ para los cuales $v_{x(\cdot)}(\Lambda_2) > 0$. Si $x(\cdot) \in G_2$ aceptamos la hipótesis H_2 ; si $x(\cdot) \notin G_2$ admitimos la hipótesis H_1 .

2a) Sea $\pi_1 \{(s, x) : f(s, x) > 1 + c\} = +\infty$. Elegimos la sucesión creciente de los conjuntos medibles B_n de modo que $\pi_1(B_n) > n$, $f(s, x) > 1 + c$ para $(s, x) \in B_n$. Designemos con G_1 el conjunto de tales $x(\cdot)$ para los cuales

$$\lim_{n \rightarrow \infty} \frac{1}{\pi_1(B_n)} v_{x(\cdot)}(B_n) = 1.$$

Si $x(\cdot) \in G_1$ admitimos la hipótesis H_1 ; si $x(\cdot) \notin G_1$ aceptamos la hipótesis H_2 .

2b) Sea $\pi_1 \{(s, x) : f(s, x) < 1 - c\} = +\infty$. Elegimos la sucesión creciente de conjuntos medibles B_n de modo que $\pi_1(B_n) > n$, $f(s, x) < 1 - c$ para $(s, x) \in B_n$. El conjunto G_1 se construye igualmente que en 2a).

3) Elegimos tal sucesión de particiones $\Lambda = \bigcup_{k=1}^{m_n} V_{nk}$ que para la magnitud

$$\theta_n = \sum_{k=1}^{m_n} \pi_1(V_{nk}) \frac{(\pi_2(V_{nk}) - \pi_1(V_{nk}))^2}{\pi_1^2(V_{nk}) + \pi_2^2(V_{nk})}$$

se cumpla la relación $\sum_{n=1}^{\infty} \theta_n^{-1} < \infty$. Entonces la hipótesis H_1 se acepta si

$$\lim_{n \rightarrow \infty} \frac{1}{\theta_n} \sum_{k=1}^{m_n} [v_{x(\cdot)}(V_{nk}) - \pi_1(V_{nk})] \times \\ \times \frac{\pi_2(V_{nk}) - \pi_1(V_{nk})}{\sqrt{\pi_1^2(V_{nk}) + \pi_2^2(V_{nk})}} = 0,$$

si esta igualdad no se verifica, se acepta la hipótesis H_2 .

4) Supongamos que

$$\int_{\Lambda} \frac{(1 - f(s, x))^2}{1 + f^2(s, x)} \pi_1(ds \times dx) < \infty,$$

$x^0(t)$ está determinado por la relación (2.3). Introduzcamos el proceso

$$x^0(t) = x(t) - \int \gamma \left[v_{x(\cdot)}(ds \times dy) - \frac{1}{1 + \gamma^2} \pi_1(ds \times dy) \right]$$

(la integral se comprende como el límite de la integral en el dominio $[0, T] \times \{y : |y| \geq \varepsilon\}$ para $\varepsilon \downarrow 0$). Para cada hipótesis H_k $x^0(t)$ es un proceso continuo de incrementos independientes con media $\gamma_1(t)$ para la hipótesis H_1 , $\gamma_2(t) + a_2(t)$ para la hipótesis H_2 y la variancia $b_k(t)$ para la hipótesis H_k . Sea t_0 tal que $b_1(t_0) \neq b_2(t_0)$. Consideraremos que las funciones $b_k(t)$ crecen estrictamente. Analicemos dos casos.

4a) Sea con cierto $\delta > 0$ $b_1(t_0) < (1-\delta)b_2(t_0)$. Entonces, para cada n se puede elegir tales puntos $t'_{n1} < t''_{n1} \leq t'_{n2} < t''_{n2} < \dots$, que $(1-\delta)[b_2(t''_{nh}) - b_2(t'_{nh})] \geq b_1(t''_{nh}) - b_1(t'_{nh}) > 0$. Tomemos G_1 igual al conjunto de $x(\cdot)$ para las cuales

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{h=1}^n \frac{1}{b_1(t''_{nh}) - b_1(t'_{nh})} [x^0(t_{nh}) - x^0(t'_{nh}) - \gamma_1(t''_{nh}) + \gamma_1(t'_{nh})]^2 = 1.$$

4b) Sea con cierto $\delta > 0$ $b_2(t_0) < (1-\delta)b_1(t_0)$. Elegimos los puntos $t'_{n1} < t''_{n1} \leq t'_{n2} < t''_{n2} \leq \dots$ de tal modo que $(1-\delta) \times [b_1(t''_{nh}) - b_1(t'_{nh})] \geq b_2(t''_{nh}) - b_2(t'_{nh}) > 0$. Designemos con G_2 el conjunto de $x(\cdot)$ para las cuales

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{h=1}^n \frac{1}{b_2(t''_{nh}) - b_2(t'_{nh})} [x^0(t''_{nh}) - x^0(t'_{nh}) - \gamma_2(t''_{nh}) + \gamma_2(t'_{nh}) - a_2(t''_{nh}) + a_2(t'_{nh})]^2 = 1.$$

En cada uno de estos casos la hipótesis H_h se admite si $x(\cdot) \in G_h$; si $x(\cdot) \notin G_h$ se acepta otra hipótesis.

5) Si $\gamma(t)$ no es absolutamente continua respecto de $b(t)$, entonces se puede elegir la sucesión de particiones del segmento $[0, T] : 0 = t_{n0} < t_{n1} < \dots < t_{nm_n} = T$ de modo que las magnitudes

$$\theta_n = \sum_{h=1}^{m_n} \frac{(\gamma(t_{nh}) - \gamma(t_{nh-1}))^2}{b(t_{nh}) - b(t_{nh-1})}$$

para un $\alpha > 0$ satisfagan la desigualdad $\theta_n > n^\alpha$. Designemos con G_1 el conjunto de $x(\cdot)$ para los cuales

$$\lim_{n \rightarrow \infty} \frac{1}{\theta_n} \sum_{h=1}^{m_n} \left[\frac{x^0(t_{nh}) - x^0(t_{nh-1}) - \gamma_1(t_{nh}) + \gamma_1(t_{nh-1})}{b(t_{nh}) - b(t_{nh-1})} \right] \times \\ \times [b(t_{nh}) - b(t_{nh-1})] = 0.$$

Aceptemos la hipótesis H_1 para $x(\cdot) \in G_1$ y la hipótesis H_2 , si $x(\cdot) \notin G_1$.

2A.2.3. Condiciones de equivalencia de las medidas. Sean cumplidas las condiciones

I. $\Lambda_1 \cup \Lambda_2 = \emptyset$.

II. $\int_{\Lambda} \frac{(1-f(s, x))^2}{1+f^2(s, x)} \pi_1(ds \times dx) < \infty$.

III. $b_1(t) = b_2(t) = b(t)$.

IV. La función $\gamma(t)$ es absolutamente continua respecto de $b(t)$ y

$$\int_0^T \left(\frac{d\gamma(t)}{db(t)} \right)^2 db(t) < \infty.$$

Entonces las medidas μ_1 y μ_2 son equivalentes y la densidad de μ_2 respecto de μ_1 se da por la fórmula

$$\begin{aligned} \rho(x(\cdot)) = \exp \left\{ \int_0^T \frac{d\gamma(t)}{db(t)} d[x^0(t) - \gamma_1(t)] - \right. \\ \left. - \frac{1}{2} \int_0^T \left(\frac{d\gamma(t)}{db(t)} \right)^2 db(t) + \right. \\ \left. + \int_{|1-f(s,y)| \leq c} \ln f(s,y) [v_{x(\cdot)}(ds \times dy) - \pi_1(ds \times dy)] + \right. \\ \left. + \int_{|1-f(s,y)| > c} \ln f(s,y) v_{x(\cdot)}(ds \times dy) + \right. \\ \left. + \int_{|1-f(s,y)| \leq c} [\ln f(s,y) - f(s,y) + 1] \pi_1(ds \times dy) + \right. \\ \left. + \int_{|1-f(s,y)| > c} (1-f(s,y)) \pi_1(ds \times dy) \right\} \quad (2.5) \end{aligned}$$

(las integrales en la fórmula (2.5) deben entenderse como estocásticas).

Demos la distribución $\rho(x(\cdot))$ mediante la función característica de la magnitud $\ln \rho(x(\cdot))$. Para la hipótesis H_1

$$\begin{aligned} M \exp \{ iz \ln \rho(x(\cdot)) \} = \exp \left\{ -\frac{1}{2} z^2 \int_0^T \left(\frac{d\gamma(t)}{db(t)} \right)^2 db(t) + \right. \\ \left. + \int [e^{iz \ln f(s,y)} - 1 - iz (f(s,y) - 1)] \pi_1(ds \times dy) + iz \alpha_1 \right\}, \end{aligned}$$

donde $\alpha_1 = -\frac{1}{2} \int_0^T \left(\frac{d\gamma(t)}{db(t)} \right)^2 db(t)$. Para la hipótesis H_2 ,

$$\begin{aligned} M \exp \{ iz \ln \rho(x(\cdot)) \} = \exp \left\{ -\frac{1}{2} z^2 \int_0^T \left(\frac{d\gamma(t)}{db(t)} \right)^2 db(t) + \right. \\ \left. + \int \left[e^{iz \ln f(s,y)} - 1 - iz \frac{f(s,y) - 1}{2 - 2f(s,y) + f^2(s,y)} \right] \times \right. \\ \left. \times \pi_2(ds \times dy) + iz \alpha_2 \right\}, \end{aligned}$$

donde

$$\alpha_2 = + \frac{1}{2} \int_0^T \left(\frac{d\gamma(t)}{db(t)} \right)^2 db(t) + \int \left(\frac{(1-f(s,y))^2}{2-2f(s,y)+f^2(s,y)} \right) \pi_1(ds \times dy).$$

24.2.4. Caso general. Es evidente que si $v_{x(\cdot)}(\Lambda_1 \cup \Lambda_2) > 0$ se puede con certeza elegir H_i para tal i con el cual $v_{x(\cdot)}(\Lambda_i) > 0$ ($v_{x(\cdot)}(\Lambda_1)$ y $v_{x(\cdot)}(\Lambda_2)$ no pueden ser positivos simultáneamente). Si $v_{x(\cdot)}(\Lambda_1 \cup \Lambda_2) = 0$, la distribución de $x(\cdot)$ coincide con la distribución del proceso con incrementos independientes $x^*(t)$ cuya función característica para la hipótesis H_i tiene el aspecto

$$\exp \left\{ i\gamma_h^*(t)z - \frac{1}{2} b_h(t)z^2 + \int_{\{s < t\} \cap \Lambda} \int \left(e^{izy} - 1 - \frac{izy}{1+y^2} \right) \pi_h(ds \times dy) \right\},$$

donde

$$\gamma_h^*(t) = \gamma_h(t) + \int_{\{s < t\} \cap \Lambda_h} \int \frac{y}{1+y^2} \pi_h(ds \times dy).$$

Las medidas μ_k^* correspondientes al proceso $x^*(\cdot)$ en $D_{[0, T]}$ para la hipótesis H_k serán equivalentes, además la densidad $\rho(x(\cdot))$ de la medida μ_2^* respecto de μ_1^* se determina por la fórmula (2.5) si en ella se sustituye $\pi_1(B)$ por $\pi_1^*(B) = \pi_1(B \cap \Lambda)$. De este modo, si a base de la observación es imposible distinguir con certeza las hipótesis, entonces para construir la mejor regla se puede usar los resultados del punto anterior.

24.2.5. Definición del parámetro del proceso homogéneo de Poisson. Sea $x(t)$ una función escalonada, todos los saltos de la cual son iguales a 1, $x(0) = 0$. La hipótesis H_k consiste en que $x(t)$ es un proceso homogéneo de Poisson con parámetros λ_k ($k = 1, 2$). De este modo, para la hipótesis H_k

$$P(x(t) = n) = \frac{\lambda_k^n}{n!} e^{-\lambda_k}.$$

En este caso, la medida $\pi_k(ds \times dx)$ tiene para $B \in R^1$ por expresión

$$\pi_k(ds \times B) = \lambda_k ds \chi_B(1).$$

Entonces las medidas μ_k son equivalentes y

$$\rho(x(\cdot)) = \left(\frac{\lambda_2}{\lambda_1} \right)^{x(T)} e^{\lambda_1 - \lambda_2}.$$

Es evidente que el dominio $\rho(x(\cdot)) < \lambda$ coincide con el dominio $x(T) < \lambda_1 - \lambda_2 + \frac{\ln \lambda}{\ln \frac{\lambda_2}{\lambda_1}}$. Por eso, la regla, para la cual $\alpha_{12} = \pi$

y α_{21} es mínima, se construye del modo siguiente. Designemos con n_e tal número que

$$\sum_{h < n_e} \frac{(T\lambda_2)^h}{h!} e^{-T\lambda_2} \leq \varepsilon \leq \sum_{h \leq n_e} \frac{(T\lambda_2)^h}{h!} e^{-T\lambda_2}$$

($\sum_{h < 0} = 0$). Entonces, para $x(T) < n_e$ se acepta la hipótesis H_1 , para $x(T) > n_e$ se admite la hipótesis H_2 . Si $x(T) = n_e$, con la probabilidad

$$\left[\varepsilon - \sum_{h < n_e} \frac{(T\lambda_2)^h}{h!} e^{-T\lambda_2} \right] \left(\frac{(T\lambda_2)^{n_e}}{n_e!} e^{-T\lambda_2} \right)^{-1}$$

se acepta la hipótesis H_2 , y con la probabilidad

$$\left[\sum_{h \leq n_e} \frac{(T\lambda_2)^h}{h!} e^{-T\lambda_2} - \varepsilon \right] \left(\frac{(T\lambda_2)^{n_e}}{n_e!} e^{-T\lambda_2} \right)^{-1}$$

se admite la hipótesis H_1 .

24.2.6. Determinación del proceso homogéneo medio de Wiener.

Sea $x(t)$ un proceso continuo; para la hipótesis H_1 éste es el proceso de Wiener con media 0 y varianza t ; para la hipótesis H_2 , con media γt y varianza t . Las medidas μ_1 y μ_2 son equivalentes y

$$\rho(x(\cdot)) = \exp \left\{ \gamma x(T) - \frac{\gamma^2}{2} T \right\}.$$

Es evidente que el dominio $\{\rho(x(\cdot)) < \lambda\}$ tiene el aspecto

$$x(T) < \frac{\ln \lambda + \frac{\gamma^2}{2} T}{\gamma}$$

(consideramos $\gamma > 0$). Para la hipótesis H_1 $x(T)$ tiene distribución normal con medio 0 y varianza T ; para la hipótesis H_2 , distribución normal con media γT y varianza T . Satisfaga $c(\varepsilon)$ la relación

$$\varepsilon = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{c(\varepsilon)} e^{-u^2/2} du$$

y sea $\lambda = c(\varepsilon) \sqrt{T} + \gamma T$. Aceptemos la hipótesis H_1 si $x(T) < \lambda$ y la hipótesis H_2 si $x(T) \geq \lambda$. Además, $\alpha_{12} = \varepsilon$,

$$\alpha_{21} = \frac{1}{\sqrt{2\pi}} \int_{c(\varepsilon) + \gamma \sqrt{T}}^{\infty} e^{-u^2/2} du.$$

24.3. Distinción de las hipótesis para los procesos difusivos

24.3.1. Definición de operador de difusión. Sea $x(t)$, $t \in [0, T]$ la trayectoria de un proceso difusivo en R^m ; para la hipótesis H_h el vector de traslación del proceso difusivo será $a_h(t, x)$ y el operador de difusión, $B_h(t, x)$. Supongamos que las funciones $a_h(t, x)$, $B_h(t, x)$ están determinadas y son continuas para $t \in [0, T]$, $x \in R^m$. Entonces, conociendo la trayectoria del proceso $x(t)$ se puede determinar $B_h \times \times(t, x(t))$, $t \in [0, T]$ si es cierta la hipótesis H_h . Esto se puede hacer del modo siguiente. Para $z \in R^m$ ponemos

$$\lambda(t, z) = \lim_{n \rightarrow \infty} \sum_{k=0}^{2^n - 1} \left(x \left(\frac{k+1}{2^n} t \right) - x \left(\frac{k}{2^n} t \right), z \right)^2 \quad (3.1)$$

((\cdot, \cdot) es un producto escalar en R^m), el límite de (3.1) existe con la probabilidad 1, cualquiera que sea la hipótesis válida. Entonces

$$\lambda(t, z) = \int_0^t (B_h(s, x(s)) z, z) ds, \quad (3.2)$$

si es cierta la hipótesis H_h .

Si en la trayectoria en observación, para algunas $t \in [0, T]$ y $z \in R^m$

$$\int_0^t (B_1(s, x(s)) z, z) ds \neq \int_0^t (B_2(s, x(s)) z, z) ds$$

(en virtud de la continuidad de las funciones subintegrales esto se realiza cuando para ciertas $t \in [0, T]$ y $z \in R^m$ $(B_1(t, x(t)) z, z) \neq (B_2(t, x(t)) z, z)$), entonces la igualdad (3.2) puede cumplirse sólo para un k . En este caso eligiendo la hipótesis H_k , si (3.2) se verifica para el k dado, obtendremos la regla infalible de distinción de las hipótesis.

Ahora, sea a lo largo de la trayectoria en observación

$$(B_1(t, x(t)) z, z) = (B_2(t, x(t)) z, z)$$

para todas las $z \in R^m$. Entonces $B_1(t, x(t)) = B_2(t, x(t))$. Por eso podemos considerar que $B_1(t, x) = B_2(t, x)$ para todas las $t \in [0, T]$ y $x \in R^m$.

24.3.2. Condición de equivalencia de las medidas. Sea μ_h una medida correspondiente al proceso en observación para la hipótesis H_h (la distribución $x(0)$ se considera dada o independiente de la elección de la hipótesis). Designemos $B(t, x) = B_1(t, x) = B_2(t, x)$, $a(t, x) = a_2(t, x) - a_1(t, x)$. Para que las medidas μ_1 y μ_2 sean equivalentes es suficiente que para todo t, x exista el vector $b(t, x) \in R^m$ tal que

$$a(t, x) = B(t, x) b(t, x)$$

y, con la probabilidad 1,

$$\int_0^T (a(t, x(t)), b(t, x(t))) dt < \infty.$$

Cumpléndose estas condiciones la densidad de la medida μ_x respecto de μ_t tiene el aspecto

$$\rho(x(\cdot)) = \exp \left[\int_0^T (b(t, x(t)), dx(t)) - \frac{1}{2} \int_0^T (b(t, x(t)), a_1(t, x(t)) + a_2(t, x(t))) dt \right]. \quad (3.3)$$

Sea $c(t, x) = (b(t, x), a_1(t, x) + a_2(t, x))$. Si queremos encontrar la distribución $\rho(x(\cdot))$ para la hipótesis H_h introduzcamos la magnitud

$$\xi_t = \int_t^T (b(s, x(s)), dx(s)) - \frac{1}{2} \int_t^T c(s, x(s)) ds.$$

Designemos con M_t^x esperanza matemática condicional a condición de que $x(t) = x$. Hagamos

$$u_\lambda(t, x) = M_t^x e^{i\lambda \xi_t},$$

entonces, la función $u_\lambda(t, x)$ satisfaco la siguiente ecuación para $t \in [0, T]$:

$$\frac{\partial u_\lambda(t, x)}{\partial t} + \frac{1}{2} [(B(t, x) \nabla, \nabla) + (a_1(t, x) + i\lambda b(t, x) \nabla)] u_\lambda - \left[i\lambda c(t, x) + \frac{\lambda^2}{2} (a(t, x), b(t, x)) \right] u_\lambda = 0 \quad (3.4)$$

(aquí ∇ es un vector con las componentes $\frac{\partial}{\partial x^1}, \dots, \frac{\partial}{\partial x^m}$, donde x^1, \dots, x^m son las coordenadas del punto x). Para $t=T$ la función $u_\lambda(t, x)$ satisfaco la condición de frontera $u_\lambda(T, x) = 1$.

24.3.3. Procesos homogéneos según el espacio. Sea $a_h(t, x) = a_h(t)$, $B_h(t, x) = B_h(t)$, es decir, los coeficientes del proceso difusivo no dependen de la coordenada espacial. En este caso, el proceso $x(t)$ será un proceso con incrementos independientes. De (3.2) se deduce que

$$\lambda(t, z) = \int_0^t (B_h(s) z, z) ds$$

si es cierta la hipótesis H_h . Por eso, las hipótesis se distinguen con certeza si para cierto $t B_1(t) \neq B_2(t)$.

Sea $B_1(t) = B_2(t) = B(t)$, $a_2(t) - a_1(t) = a(t)$. Designemos con E el conjunto de tales $t \in [0, T]$ para las cuales $a(t)$ no pertenece al dominio de los valores del operador $B(t)$. Sea $P(t)$ el operador de proyección sobre el dominio de los valores del operador $B(t)$. Si la medida de Lebesgue del conjunto E es positiva las hipótesis se distinguen infaliblemente:

si es cierta la hipótesis H_1 , entonces

$$\int_0^T |P(t)(x(t) - a_1(t))|^2 dt = 0,$$

si es cierta la hipótesis H_2 , entonces

$$\int_0^T |P(t)(x(t) - a_1(t))|^2 dt > 0.$$

Supongamos que casi para todas las t el vector $a(t)$ pertenece al dominio de los valores del operador $B(t)$, es decir, que existe tal vector $b(t)$ que

$$a(t) = B(t)b(t)$$

(para que el vector $b(t)$ se determine unívocamente lo elegimos en el dominio de los valores de $B(t)$; esto es posible ya que $B(t)$ es simétrico). Cabe señalar que $(a(t), b(t)) > 0$. La condición

$$\int_0^T (a(t), b(t)) dt < \infty \quad (3.5)$$

es suficiente y necesaria para la continuidad absoluta de las medidas. Si (3.5) se cumple, las medidas μ_1 y μ_2 son equivalentes y la densidad de la medida μ_2 respecto de μ_1 tiene la forma

$$\rho(x(\cdot)) = \exp \left\{ \int_0^T (b(t), dx(t)) - \frac{1}{2} \int_0^T (b(t), a_1(t) + a_2(t)) dt \right\}.$$

Para cualquiera de las hipótesis la $\rho(x(\cdot))$ tiene distribución normal con la varianza

$$\int_0^T (a(t), b(t)) dt$$

y la media

$$\left\{ \begin{array}{l} -\frac{1}{2} \int_0^T (a(t), b(t)) dt \text{ para la hipótesis } H_1; \\ \frac{1}{2} \int_0^T (a(t), b(t)) dt \text{ para la hipótesis } H_2. \end{array} \right.$$

Señalemos, además, la regla de distinción infalible de las hipótesis a condición de que

$$\int_0^T (a(t), b(t)) dt = +\infty. \quad (3.6)$$

Elegimos la sucesión de las funciones $b_n(t)$ tal que

$$\int_0^T (B(t) b_n(t), b_n(t)) dt = \int_0^T (a(t), b_n(t)) dt \geq n$$

(esto es posible en virtud de (3.6)). Entonces elegimos la hipótesis H_1 si

$$\lim_{n \rightarrow \infty} \left(\int_0^T (B(t) b_n(t), b_n(t)) dt \right)^{-1} \int_0^T b_n(t) d[x(t) - a_1(t)] = 0$$

y la hipótesis H_2 si esta condición no se cumple.

24.4. Distinción de las hipótesis del valor medio del proceso gaussiano

Sea $x(t)$, $t \in [0, T]$, la trayectoria de un proceso unidimensional gaussiano con la función de correlación, continua prefijada $R(t, s)$. El valor medio del proceso para la hipótesis H_1 es igual a 0, y para la hipótesis H_2 , a la función continua dada $a(t)$ (si el valor medio para la hipótesis H_1 fuese $a_1(t)$, se podría considerar $x(t) - a_1(t)$ en vez de $x(t)$).

24.4.1. Condiciones de ortogonalidad. Distinción infalible de las hipótesis. Designemos con $L_2[0, T]$ el espacio de las funciones $g(t)$ en $[0, T]$ de cuadrado integrable, éste es un espacio de Hilbert con

producto escalar $(g_1, g_2) = \int_0^T g_1(t) g_2(t) dt$.

Sea $Rg(t) = \int_0^T R(t, s) g(s) ds$ un operador lineal en $L_2[0, T]$.

Este operador lineal es totalmente continuo y tiene una sucesión de valores propios y de funciones propias $\{\lambda_k, \varphi_k\}$:

$$\int_0^T R(t, s) \varphi_k(s) ds = \lambda_k \varphi_k(t),$$

las funciones φ_k son ortogonales dos a dos y completas en el dominio de los valores del operador R , $\lambda_k > 0$ y $\sum \lambda_k < \infty$.

1) Si $a(t)$ no se desarrolla en una serie según las funciones $\varphi_k(t)$, entonces las medidas μ_1 y μ_2 , correspondientes al proceso para las hipótesis H_1 y H_2 , son ortogonales. La regla que distingue las hipó-

tesis consiste en lo siguiente: sea

$$\hat{a}(t) = a(t) - \sum_{h=1}^{\infty} \varphi_h(t) \int_0^T \varphi_h(s) a(s) ds \quad (\langle \varphi_h, \varphi_h \rangle = 1).$$

Si $\int_0^T x(t) \hat{a}(t) dt = 0$, aceptamos la hipótesis H_1 ; si $\int_0^T x(t) \hat{a}(t) dt \neq 0$, admitimos la hipótesis H_2 .

2) Supongamos que $a(t)$ se desarrolla según las funciones propias φ_h :

$$a(t) = \sum_{h=1}^{\infty} \alpha_h \varphi_h(t), \quad \alpha_h = \int_0^T a(t) \varphi_h(t) dt.$$

Si

$$\sum_{h=1}^{\infty} \frac{\alpha_h^2}{\lambda_h} = \infty \quad (4.1)$$

las medidas μ_1 y μ_2 son ortogonales. La regla de distinción se construye del modo siguiente: elegimos la subsucesión m_n de modo que

$$\sum_{h=1}^{m_n} \frac{\alpha_h^2}{\lambda_h} \geq n.$$

Si

$$\lim_{n \rightarrow \infty} \left(\sum_{h=1}^{m_n} \frac{\alpha_h^2}{\lambda_h} \right)^{-1} \sum_{h=1}^{m_n} \frac{\alpha_h}{\lambda_h} \int_0^T x(t) \varphi_h(t) dt = 0,$$

entonces se acepta la hipótesis H_1 , si esta condición no se cumple se admite la hipótesis H_2 .

La regla arriba citada de distinción de las hipótesis usa los valores propios y las funciones propias del operador integral R . Se puede proponer otras reglas que no utilizan estos datos.

Sea $\{\psi_k(t), k = 1, 2, \dots\}$ un sistema de funciones ortonormalizado aleatorio en $L_2[0, T]$. Designemos

$$x_k = \int_0^T x(t) \psi_k(t) dt;$$

$$a_k = \int_0^T a(t) \psi_k(t) dt;$$

$$r_{jh} = M \int_0^T x(t) \psi_h(t) dt \int_0^T x(t) \psi_j(t) dt$$

(la esperanza matemática se toma para la hipótesis H_1). Sea $b_h^{(n)}$ la solución de un sistema de ecuaciones lineales

$$a_h = \sum_{j=1}^n r_{jk} b_j^{(n)}.$$

Si se cumple (4.1), entonces

$$\lim_{n \rightarrow \infty} \sum_{h=1}^n a_h b_h^{(n)} = +\infty.$$

Elijamos m_n de modo que

$$\sum_{h=1}^{m_n} a_h b_h^{(m_n)} \geq n.$$

Entonces, la regla de elección de la hipótesis consiste en lo siguiente: si

$$\lim_{n \rightarrow \infty} \left(\sum_{h=1}^{m_n} a_h b_h^{(m_n)} \right)^{-1} \sum_{h=1}^{m_n} x_h b_h^{(m_n)} = 0$$

elegimos la hipótesis H_1 ; si la última relación no se cumple, elegimos la hipótesis H_2 .

24.4.2. Condiciones de equivalencia de las medidas. En las designaciones del punto anterior

$$\sum_{h=1}^{\infty} \frac{\alpha_h^2}{\lambda_h} < \infty \quad (4.2)$$

es la condición suficiente y necesaria de equivalencia de las medidas μ_1 y μ_2 . Si (4.2) se cumple la densidad de la medida μ_2 respecto de la medida μ_1 tiene la forma

$$\rho(x(\cdot)) = \exp \left\{ \sum_{h=1}^{\infty} \frac{x_h \alpha_h}{\lambda_h} - \frac{1}{2} \sum_{h=1}^{\infty} \frac{\alpha_h^2}{\lambda_h} \right\}. \quad (4.3)$$

Si existe la solución $b(t)$ de la ecuación de Fredholm de primer género

$$a(t) = \int_0^T R(T(t, s) b(s)) ds \quad (4.4)$$

la densidad $\rho(x(\cdot))$ puede ser escrita de una manera más evidente

$$\rho(x(\cdot)) = \exp \left\{ \int_0^T x(s) b(s) ds - \frac{1}{2} \int_0^T a(s) b(s) ds \right\} \quad (4.5)$$

(cabe señalar que en muchos casos la solución de la ecuación (4.4) existe como una función generalizada y las integrales de (4.5) se pueden transformar en ordinarias integrando por partes). En $\rho(x(\cdot))$ tiene

distribución normal con la varianza

$$d = \sum_{h=1}^{\infty} \frac{\alpha_h^2}{\lambda_h}$$

para ambas hipótesis; el valor medio es igual a $-\frac{d}{2}$ para la hipótesis H_1 y a $\frac{d}{2}$ para la hipótesis H_2 . Si es difícil resolver la ecuación (4.4) se puede usar el siguiente criterio aproximado de distinción de las hipótesis.

Sean $a_h^{(n)}$ los mismos que en el punto anterior. Aceptemos la hipótesis H_1 si $\sum_{h=1}^n b_h^{(n)} x_h < \lambda$ y la hipótesis H_2 si $\sum_{h=1}^n b_h^{(n)} x_h \geq \lambda$. Además,

$$\lambda = c(\varepsilon) \sqrt{d_n} \cdot \frac{1}{2} d_n,$$

donde

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{c(\varepsilon)} e^{-u^2/2} du = \varepsilon; \quad d_n = \sum_{h=1}^n a_h b_h^{(n)}$$

En este caso $\alpha_{12} \rightarrow \varepsilon$ y

$$\alpha_{21} = \frac{1}{\sqrt{2\pi}} \int_{c(\varepsilon) + \sqrt{d_n}}^{\infty} e^{-u^2/2} du.$$

Ya que $d_n \rightarrow d = \sum_{h=1}^{\infty} \frac{\alpha_h^2}{\lambda_h}$ y para la regla óptima de distinción con $\alpha_{12} = \varepsilon$

$$\alpha_{21} = \frac{1}{\sqrt{2\pi}} \int_{c(\varepsilon) + \sqrt{d}}^{\infty} e^{-u^2/2} du.$$

entonces $\alpha_{21}^{opt} \rightarrow \alpha_{21}$.

24.4.3. Caso de procesos estacionarios. Sea $R(t, s) = r(t - s)$, es decir, $x(t)$ es un proceso estacionario para la hipótesis H_1 . Designemos con $P(\lambda)$ la función espectral del proceso. Sea W_T el espacio de las funciones $g(\lambda)$ que pueden ser representadas en la forma

$$g(\lambda) = \int_{-T}^T e^{i\lambda t} q(t) dt.$$

donde φ son funciones de valor complejo en $[-T, T]$ para las cuales $\int_{-T}^T |\varphi(t)|^2 dt < \infty$. Designemos, luego, con $W_T(F)$ la clausura W_T en la métrica determinada por la ecuación

$$\|g\|_F^2 = \int |g(\lambda)|^2 dF(\lambda).$$

Para la equivalencia de las medidas μ_1 y μ_2 es necesario y suficiente que la función $a(t)$ para $t \in [-T, T]$ tenga la representación

$$a(t) = \int e^{-i\lambda t} b(\lambda) dF(\lambda),$$

donde $b(\lambda) \in W_T(F)$. Si esta condición está cumplida, la densidad $\rho(x(\cdot))$ tiene el aspecto

$$\rho(x(\cdot)) = \exp \left\{ \int b(\lambda) dy(\lambda) - \frac{1}{2} \int |b(\lambda)|^2 dF(\lambda) \right\} \quad (4.6)$$

donde $y(\lambda)$ es la medida espectral del proceso $x(t)$:

$$x(t) = \int e^{i\lambda t} dy(\lambda).$$

Señalemos que la integral estocástica en (4.6) puede ser calculada como

$$\lim_{n \rightarrow \infty} \int_{-T}^T \varphi_n(t) x(t) dt,$$

donde φ_n es tal sucesión de funciones que $\|b - b_n\|_F \rightarrow 0$, si sólo

$$b_n(\lambda) = \int_{-T}^T e^{i\lambda t} \varphi_n(t) dt.$$

24.5. Distinción de las hipótesis sobre la función de correlación del proceso gaussiano

Se observa la trayectoria del proceso unidimensional gaussiano $x(t)$, $t \in [0, T]$, cuyo valor medio es 0; respecto de la función de correlación hay dos hipótesis: según la hipótesis H_1 la trayectoria es igual a $R_1(t, s)$ y según la hipótesis H_2 , a $R_2(t, s)$; se supone que $R_k(t, s)$ son funciones continuas.

24.5.1. Condiciones de ortogonalidad. Como antes designemos con μ_k la medida correspondiente al proceso para la hipótesis H_k , $k = 1, 2$. Siempre se puede considerar que estas medidas están concentradas en el espacio $L_2[0, T]$.

1) Si existe tal sucesión de funciones $g_n(t) \in L_2[0, T]$ que

$$\lim_{n \rightarrow \infty} \int_0^T \int_0^T R_2(t, s) g_n(t) g(s) dt ds \times \\ \times \left(\int_0^T \int_0^T R_1(t, s) g_n(t) g_n(s) dt ds \right)^{-1} = +\infty, \quad (5.1)$$

las medidas μ_1 y μ_2 son ortogonales. Obtendremos la regla infalible de distinción de la hipótesis eligiendo tal sucesión de las funciones $\psi_n(t)$ que

$$\int_0^T \int_0^T R_1(t, s) \psi_n(t) \psi_n(s) dt ds = 1; \\ \int_0^T \int_0^T R_2(t, s) \psi_n(t) \psi_n(s) dt ds \geq n$$

Aceptamos la hipótesis H_1 si

$$\left(\int_0^T \int_0^T R_2(t, s) \psi_n(t) \psi_n(s) dt ds \right)^{-1/2} \int_0^T x(t) \psi_n(t) dt \rightarrow 0,$$

y la hipótesis H_2 si esta condición no se cumple. De modo análogo se construye la regla de elección si la relación (5.1) se cumple permutand los índices 1 y 2.

2) Supongamos que existe tal $\delta > 0$ que

$$\delta \int_0^T \int_0^T R_1(t, s) g(t) g(s) dt ds \leq \\ \leq \int_0^T \int_0^T R_2(t, s) g(t) g(s) dt ds \leq \frac{1}{\delta} \int_0^T \int_0^T R_1(t, s) g(t) g(s) dt ds.$$

Construyamos cierta sucesión de tales funciones $\psi_k(t)$ que

$$\int_0^T \int_0^T R_2(t, s) \psi_k(t) \psi_l(s) dt ds = 0 \quad \text{para } k \neq l.$$

Si

$$\sum_{k=1}^{\infty} \left[\int_0^T \int_0^T R_2(t, s) \psi_k(t) \psi_k(s) dt ds \times \right. \\ \left. \times \left(\int_0^T \int_0^T R_1(t, s) \psi_k(t) \psi_k(s) dt ds \right)^{-1} - 1 \right]^2 = +\infty, \quad (5.2)$$

entonces las medidas μ_1 y μ_2 son ortogonales. Sean

$$\int_0^T \int_0^T R_1(t, s) \psi_h(t) \psi_h(s) dt ds = 1;$$

$$\int_0^T \int_0^T R_2(t, s) \psi_h(t) \psi_h(s) dt ds = 1 + \delta_h.$$

Obtendremos la regla infalible de distinción de las hipótesis eligiendo m_n de tal modo que

$$\sum_{h=1}^{m_n} \delta_h^2 > n$$

y aceptando la hipótesis H_1 si

$$\lim_{n \rightarrow \infty} \left(\sum_{h=1}^{m_n} \delta_h^2 \right)^{-1} \sum_{h=1}^{m_n} \left[\left(\int_0^T \psi_h(t) x(t) dt \right)^2 - 1 \right] \delta_h = 0$$

y la hipótesis H_2 , si esta condición no se cumple.

24.5.2. Condiciones de equivalencia. Introduzcamos junto con la función $R_h(t, s)$ la función $R_h^{1/2}(t, s)$ que es una función simétrica de cuadrado integrable en $[0, T] \times [0, T]$ la cual satisface la relación

$$R_h(t, s) = \int_0^T R_h^{1/2}(t, u) R_h^{1/2}(u, s) du.$$

Si $\varphi_n^{(k)}(t)$ y $\lambda_n^{(k)}$, $k = 1, 2, n = 1, 2, \dots$, son, respectivamente, funciones propias y valores propios del operador integral

$$R_h \varphi(t) = \int_0^T R_h(t, s) \varphi(s) ds.$$

entonces la función $R_h^{1/2}(t, s)$ puede ser construida del modo siguiente:

$$R_h^{1/2}(t, s) = \sum_{h=1}^{\infty} \sqrt{\lambda_n^{(h)}} \varphi_n^{(h)}(t) \varphi_n^{(h)}(s) \quad (5.3)$$

(la serie a la derecha converge a la media cuadrática en $[0, T] \times [0, T]$). Para la equivalencia de las medidas μ_1 y μ_2 es necesario y suficiente que exista tal función de cuadrado integrable $D(t, s)$ que

$$R_1(t, s) - R_2(t, s) = \int_0^T \int_0^T R_1^{1/2}(t, u) D(u, v) R_2^{1/2}(v, s) du dv. \quad (5.4)$$

Sea $D(t, s)$ esta función. Designemos con $\theta_h(t)$ las funciones propias y con δ_h los valores propios de la integral simétrica con núcleo $D(t, s)$:

$$\delta_h \theta_h(t) = \int_0^T D(t, s) \theta_h(s) ds.$$

(De la relación (5.4) se deduce que $\delta_h > -1$). Entonces, la densidad de la medida μ_2 respecto de μ_1 tiene la forma

$$\rho(x(\cdot)) = \exp \left\{ -\frac{1}{2} \sum_{h=1}^{\infty} \left[\xi_h^2 \frac{\delta_h}{1+\delta_h} - \ln(1+\delta_h) \right] \right\}, \quad (5.5)$$

donde ξ_h se determina de la función $x(t)$ mediante la relación

$$\xi_h = \sum_{n=1}^{\infty} \frac{1}{V \lambda_n^2} \int_0^T \varphi_n^{(2)}(s) x(s) ds \int_0^T \theta_h(t) \varphi_n^{(2)}(t) dt. \quad (5.6)$$

Para encontrar la distribución de la magnitud $\rho(x(\cdot))$ se debe tener en cuenta que las magnitudes ξ_h , para cada una de las hipótesis, son una sucesión de magnitudes independientes gaussianas con media 0 y varianza 1 para la hipótesis H_2 y varianza $1 + \delta_h$ para la hipótesis H_1 . Por eso

$$M e^{i s \ln \rho(x(\cdot))} = \prod_{h=1}^{\infty} \frac{(1+\delta_h)^{s^2/2}}{(1+is\delta_h)^{1/2}}$$

para la hipótesis H_1 y

$$M e^{i s \ln \rho(x(\cdot))} = \prod_{h=1}^{\infty} \frac{(1+\delta_h)^{s^2/2}}{\left(1 + \frac{is\delta_h}{1+\delta_h}\right)^{1/2}}$$

para la hipótesis H_2 .

Como vemos la determinación de la densidad de una medida respecto de otra y de la distribución de esta densidad en el caso de distintas varianzas lleva a la necesidad de definir las funciones propias y los valores propios de los operadores integrales. A veces, al calcular la densidad y su distribución, se puede no recurrir al cálculo del operador D y la magnitud ξ_h . Analicemos la ecuación

$$\lambda \int_0^T R_2(t, s) \psi(s) ds = \int_0^T R_1(t, s) \psi(s) ds. \quad (5.7)$$

Esta ecuación, en el caso de la equivalencia de las medidas, tiene soluciones (posiblemente, generalizadas) para el conjunto numerable λ y no más. Designemos estos valores de λ con λ_h y las soluciones correspondientes con ψ_h . Entonces $\lambda_h = 1 + \delta_h$ y la densidad $\rho(x(\cdot))$

tiene el aspecto

$$\rho(x(\cdot)) = \exp \left\{ -\frac{1}{2} \left[\sum_{h=1}^{\infty} \int_0^T \left(x(t) \psi_h(t) dt \frac{\lambda_h - 1}{\lambda_h} - \ln \lambda_h \right) \right] \right\} \quad (5.8)$$

(se supone que ψ_h están normalizadas de modo que

$$\int_0^T \int_0^T R_2(t, s) \psi_h(t) \psi_h(s) dt ds = 1). \text{ La solución}$$

generalizada de la ecuación

$$\lambda_h \int_0^T R_2(t, s) \psi_h(s) ds = \int_0^T R_1(t, s) \psi_h(s) ds,$$

se determina por la sucesión de las funciones $\psi_h^{(n)}(t)$ para las cuales

$$\int_0^T \int_0^T R_2(t, s) \psi_h^{(n)}(t) \psi_h^{(n)}(s) dt ds = 1;$$

$$\lim_{n \rightarrow \infty} \left[\int_0^T \left(\lambda_h \int_0^T R_2(t, s) \psi_h^{(n)}(s) ds - \int_0^T R_1(t, s) \psi_h^{(n)}(s) ds \right)^2 dt \right] = 0.$$

24.5.3. Distinción de las hipótesis para los procesos estacionarios.

Sea $x(t)$ un proceso estacionario con la media 0 y la función de correlación $R_h(t)$ para la hipótesis H_h , $k = 1, 2$. Sea $F_h(\lambda)$ la función espectral del proceso para la hipótesis H_h :

$$R_h(t) = \int e^{i\lambda t} dF_h(\lambda).$$

Designemos con W_T^2 el conjunto de funciones del aspecto

$$\psi(\alpha, \beta) = \int_{-T}^T \int_{-T}^T e^{i\alpha s + i\beta t} g(s, t) ds dt,$$

donde $g(s, t)$ es acotada y medible en $[-T, T] \times [-T, T]$. Luego, sea $W_T^2(F_1)$ la clausura W_T^2 en la norma

$$\|\psi\|_{F_1} = \int \int |\psi(\alpha, \beta)|^2 dF_1(\alpha) dF_1(\beta).$$

Para la equivalencia de las medidas μ_1 y μ_2 es necesario y suficiente que exista tal función $b(\alpha, \beta)$ de $W_T^2(F_1)$ que sea válida la igualdad

$$R_2(t-s) - R_1(t-s) = \int \int e^{-i\alpha t + i\beta s} b(\alpha, \beta) dF_1(\alpha) dF_2(\beta),$$

en este caso la densidad tiene por expresión

$$\rho(x(\cdot)) = \exp \left\{ \frac{1}{2} \int \int \Phi(\alpha, \beta) dy(\alpha) \overline{dy(\beta)} + c \right\}, \quad (5.9)$$

donde $y(\alpha)$ es la medida espectral correspondiente al proceso $x(t)$:

$$x(t) = \int e^{i\lambda t} dy(\lambda),$$

la función $\Phi(\alpha, \beta)$ está ligada con $b(\alpha, \beta)$ por la correlación $b(\alpha, \beta) = \Phi(\alpha, \beta) + \int \Phi(\alpha, \gamma) \overline{b(\gamma, \beta)} dF_1(\gamma)$ y la integral estocástica múltiple en (5.9) se determina como la integral según la medida ν en $(-\infty, \infty) \times (-\infty, \infty)$ para la cual

$$\begin{aligned} \nu(\{\alpha, \beta\} \times \{\gamma, \delta\}) &= \\ &= [y(\delta) - y(\gamma)] \times [y(\beta) - y(\alpha)] - F_1(\{\alpha, \beta\} \cap \{\gamma, \delta\}) \\ &\left(F_1(\Delta) = \int_{\Delta} dF_1(\lambda) \right). \end{aligned}$$

La constante c en la fórmula (5.9) se determina de la igualdad

$$c = -\frac{1}{2} \sum_{h=1}^{\infty} \ln(1 + \lambda_h),$$

donde λ_h son los valores propios del operador

$$V_{\mathcal{L}}(\beta) = \int b(\alpha, \beta) g(\alpha) dF_1(\alpha)$$

en $W_T(F_1)$. Señalemos también que la integral en (5.9) se puede escribir en la forma

$$\sum_{h, j} c_{hj} \left[\int g_h(\alpha) dy(\alpha) \overline{\int f_j(\beta) dy(\beta)} - \int g_h(\alpha) \overline{f_j(\alpha)} dF_1(\alpha) \right],$$

si

$$b(\alpha, \beta) = \sum_{h, j} c_{hj} g_h(\alpha) \overline{f_j(\beta)}.$$

24.5.4. Distinción de las hipótesis sobre la densidad espectral. Supongamos que las funciones espectrales $F_h(\lambda)$ tienen densidades espectrales $f_h(\lambda)$. Citemos algunas condiciones suficientes de la continuidad absoluta y de la ortogonalidad de las medidas en términos de las densidades espectrales.

1. Supongamos que están cumplidas las condiciones:

1) para ciertos c_1 y c_2

$$c_1 | \varphi_0(\lambda) |^2 \leq f_1(\lambda) \leq c_2 | \varphi_0(\lambda) |^2.$$

donde $\varphi_0(\lambda)$ para cierto $s > 0$ tiene el aspecto

$$\varphi_0(\lambda) = \int_{-s}^s e^{i\lambda t} g(t) dt \quad (5.10)$$

con una función de cuadrado integrable $g(t)$;

$$2) \int \left[\frac{f_2(\lambda) - f_1(\lambda)}{f_1(\lambda)} \right]^2 d\lambda < \infty.$$

Entonces, cualquiera que sea $T > 0$, las medidas μ_1 y μ_2 correspondientes a $x(t)$, $t \in [0, T]$, para las hipótesis H_1 y H_2 serán equivalentes.

II. Supongamos que para $|\lambda|$ bastante grandes y ciertos c_1 y c_2 positivos se cumple la relación

$$c_1 \ll |\lambda|^\alpha f_1(\lambda) \leq c_2,$$

donde $\alpha > 1$. Si

$$\int \frac{|f_2(\lambda) - f_1(\lambda)|^2}{1 + \lambda^{2\alpha}} d\lambda < \infty,$$

entonces las medidas μ_1 y μ_2 , correspondientes a $x(t)$, $t \in [0, T]$ para las hipótesis H_1 , H_2 , son equivalentes para todos los $T > 0$.

III. Exista la función analítica entera $\varphi_0(\lambda)$ de tipo exponencial no superior a $s < T$ para la cual con $0 < c_1 < c_2$ está cumplida la desigualdad

$$c_1 \ll |\varphi_0(\lambda)|^2 f_1(\lambda) \leq c_2.$$

Entonces, si

$$\int \int \frac{\sin^2(T-s)(\alpha-\beta)}{(\alpha-\beta)^2} \frac{f_2(\alpha) - f_1(\alpha)}{f_1(\alpha)} \frac{f_2(\beta) - f_1(\beta)}{f_1(\beta)} d\alpha d\beta < +\infty \quad (5.11)$$

as medidas μ_1 y μ_2 son ortogonales.

IV. Si $f_1(\lambda)$ y $f_2(\lambda)$ son densidades racionales (fraccionales), la condición necesaria y suficiente de equivalencia es la condición

$$\lim_{\lambda \rightarrow \infty} \frac{f_2(\lambda)}{f_1(\lambda)} = 1.$$

Si $\lim_{\lambda \rightarrow \infty} \frac{f_2(\lambda)}{f_1(\lambda)} \neq 1$ la regla, que distingue infaliblemente las hipótesis, consiste en lo siguiente. Sea m tal que existe un límite finito, distinto de cero,

$$\lim_{\lambda \rightarrow \infty} \lambda^{2m} [f_1(\lambda) + f_2(\lambda)] = a.$$

Entonces existen también los límites

$$\lim_{\lambda \rightarrow \infty} \lambda^{2m} f_1(\lambda) = a_1, \quad \lim_{\lambda \rightarrow \infty} \lambda^{2m} f_2(\lambda) = a_2, \quad (5.12)$$

con la particularidad de que $a_1 + a_2 = a \neq 0$, $a_1 \neq a_2$. De (5.12) se deduce la existencia de la $m - 1$ -ésima derivada de la trayectoria en observación $\bar{x}(t) = x^{(m-1)}(t)$. Con la probabilidad 1 existe el lí-

mite

$$\lim_{n \rightarrow \infty} \sum_{k=1}^{2^n} \left[\bar{x} \left(\frac{k}{2^n} T \right) - \bar{x} \left(\frac{k-1}{2^n} T \right) \right]^2 = a_1, \quad (5.13)$$

si es cierta la hipótesis H_1 . Calculando el primer miembro de (5.13) y comparándolo con los valores a_1 se puede elegir infaliblemente la hipótesis verdadera.

24.6. Estimaciones de los parámetros de las distribuciones para procesos aleatorios

24.6.1. Planteamiento del problema. Supongamos que acerca de la trayectoria del proceso en observación $x(t)$, $t \in [0, T]$ se conoce que es la trayectoria de un proceso al cual corresponde la medida μ_θ en el espacio de la función $F_{[0, T]}$. El parámetro θ , que debe estimarse durante la observación, varía en cierto conjunto paramétrico Θ . La particularidad de los problemas de la estadística de los procesos aleatorios consiste en que este parámetro, como regla, varía en un espacio de dimensión infinita (por ejemplo, el papel de parámetro puede jugarlo un valor medio desconocido que en principio puede ser cualquier función de $F_{[0, T]}$). Supongamos que existe la medida ν en $F_{[0, T]}$ respecto de la cual todas las medidas μ_θ son absolutamente continuas y

$$\frac{d\mu_\theta}{d\nu}(x) = \rho(\theta, x). \quad (6.1)$$

En este caso la familia de medidas $\{\mu_\theta, \theta \in \Theta\}$ se llama regular. Se supone que Θ es una variedad lineal, $\rho(\theta, x)$ es una función suficientemente regular. Entonces, la estimación del parámetro se puede determinar, por ejemplo, mediante el método de verosimilitud máxima. Señalamos que para los procesos con incrementos independientes, los procesos difusivos, así como los gaussianos las condiciones de regularidad y el aspecto de la densidad se pueden deducir de los resultados de los pp. 24.2—24.5.

Es más interesante (y más específico, precisamente, para la estadística de los procesos aleatorios) el caso de la familia singular dos a dos de medidas $\{\mu_\theta, \theta \in \Theta\}$. Entonces, si $\theta_1 \neq \theta_2$ las medidas μ_{θ_1} y μ_{θ_2} son ortogonales. Por eso, es natural esperar que el parámetro θ se puede determinar infaliblemente de una sola observación. Como estimación del parámetro θ se entenderá la función $\hat{\theta}(x)$ definida en $F_{[0, T]}$ con valores en Θ . Sea Θ un espacio métrico separable completo (o un conjunto boreliano en el espacio de este tipo). Con respecto a $\hat{\theta}(x)$ se supondrá que es medible respecto de las σ -álgebras $\tilde{\mathfrak{F}}_{[0, T]}$ generadas en $F_{[0, T]}$ por los conjuntos cilíndricos y \mathfrak{B} son las σ -álgebras de los conjuntos borelianos en Θ , es decir, que la preimagen de todo conjunto boreliano en Θ es medible en $F_{[0, T]}$. La estimación del parámetro θ se llama conciliable si

$$\mu_\theta \{x : \hat{\theta}(x) = \theta\} = 1, \quad \forall \theta \in \Theta. \quad (6.2)$$

Si es posible determinar infaliblemente el parámetro esto significa que existe la estimación conciliable del parámetro θ . Los ejemplos muestran que existen tales familias ortogonales dos a dos para las cuales no hay estimación conciliable. Por eso son de interés el problema referente a la existencia de estimaciones conciliables, así como los métodos de su construcción en caso de su existencia. Más abajo se dan algunos métodos de construcción de las estimaciones conciliables.

24.8.2. Métodos de proyección. Supongamos que las medidas μ_θ son tales que existen el valor medio del proceso $a_\theta(t)$ y la función de correlación $R_\theta(t, s)$. Luego, existan dos variedades lineales L_1 y L_2 en $F_{[0, T]}$ con intersección nula tales que $a_\theta(\cdot) \in L_1$ para todos los θ , y si θ es el valor real del parámetro, $x(t) - a_\theta(t)$, con la probabilidad 1, pertenece a L_2 . Lo último significa que si $\varphi_k(\theta, t)$ son las funciones propias del operador integral con núcleo $R_\theta(t, s)$ y $\lambda_k(\theta)$ son los valores propios correspondientes, es decir,

$$\lambda_k(\theta) \varphi_k(\theta, t) = \int_0^T R_\theta(t, s) \varphi_k(\theta, s) ds.$$

entonces

$$\sum_{k=1}^{\infty} \left[\int_0^T (x(s) - a_\theta(s)) \varphi_k(\theta, s) ds \right] \varphi_k(\theta, t) \in L_2.$$

Después, sea $a_\theta(\cdot)$ la aplicación biunívoca Θ en L_1 . Designemos $L = L_1 + L_2$ y con P el operador de «proyección» de L sobre L_1 ; si $z(t) = y_1(t) + y_2(t)$, donde $y_i(t) \in L_i$ (esta representación es única para todos los $z(t) \in L$), $Pz(t) = y_1(t)$. Entonces para nuestras suposiciones

$$\mu_\theta(\{z(\cdot) : Pz(t) = a_\theta(t)\}) = 1.$$

Conociendo $a_\theta(t)$ y, en virtud de que la aplicación $a_\theta(t)$ es biunívoca, podemos determinar según $a_\theta(t)$ el parámetro θ . De este modo, usando el método de proyección, el problema principal consiste en construir el operador de proyección P .

EJEMPLO. La familia de medidas μ_θ es familia de medidas gaussianas con valor medio

$$a_\theta(t) = \int_0^T B(t, s) \theta(s) ds, \quad (6.3)$$

$\theta(s) \in L_2[0, T]$, $\int_0^T \int_0^T B^2(t, s) dt ds < \infty$, el operador de correlación

$R_\theta(t, s) = R(t, t)$ no depende de θ .

Sean $\varphi_k(t)$ y λ_k las funciones propias y los valores del operador integral con núcleo $R(t, s)$, respectivamente. Designemos con $R^{1/2}$ la variedad lineal de funciones $z(t)$ de $L_2[0, T]$ de tipo $z(t) =$

$= \sum_{k=1}^{\infty} \alpha_k \sqrt{\lambda_k} \varphi_k(t)$, donde $\sum \alpha_k^2 < \infty$ y con B , la variedad lineal de las

funciones que pueden ser representadas en forma del segundo miembro de (6.3)

Supongamos que $B \cap R^{1/2} = (0)$, lo que significa que para todos los $\theta(s)$, para los cuales $\int_0^T \theta^2(s) ds > 0$,

$$\sum_{k=1}^{\infty} \frac{1}{\lambda_k} \left(\int_0^T \int_0^T B(t, s) \theta(s) ds \varphi_k(t) dt \right)^2 = +\infty.$$

Introduzcamos en $B + R^{1/2}$ el producto escalar: si

$$y(t) = \sum_{k=1}^{\infty} \alpha_k \sqrt{\lambda_k} \varphi_k(t) + \int_0^T B(t, s) \theta(s) ds,$$

entonces

$$(y(t), y(t))_1 = \sum_{k=1}^{\infty} \alpha_k^2 + \int_0^T \theta^2(s) ds.$$

Supongamos que $B(t, s)$ es una función simétrica y $\{\varphi_k(t), \mu_k\}$ es una sucesión de funciones propias y de valores propios, correspondiente al operador integral con núcleo $B(t, s)$. Sea

$$y(t) = \sum_{k=1}^{\infty} \alpha_k \sqrt{\lambda_k} \varphi_k(t) + \sum_{k=1}^{\infty} \beta_k \mu_k \psi_k(t), \quad (6.4)$$

donde $\sum \alpha_k^2 + \sum \beta_k^2 < \infty$. Mostremos cómo se puede construir el operador P que pone la expresión

$$\sum_{k=1}^{\infty} \beta_k \mu_k \psi_k(t)$$

en correspondencia con la función $y(t)$ del tipo (6.4).

Designemos con $\beta_k^{(n)}$ tales números que la expresión

$$\sum_{k=1}^n \frac{1}{\lambda_k} \left\{ \int_0^T \left[y(t) - \sum_{m=1}^{\infty} \gamma_m \mu_m \psi_m(t) \right] \varphi_k(t) dt \right\}^2$$

tome el valor mínimo para $\gamma_k = \beta_k^{(n)}$. Entonces $\beta_k = \lim_{n \rightarrow \infty} \beta_k^{(n)}$ y

$$P y(t) = \sum_{k=1}^{\infty} \lim_{n \rightarrow \infty} \beta_k^{(n)} \psi_k(t).$$

De este modo

$$a_{\theta}(t) = P x(t),$$

donde $x(t)$ es la trayectoria observada.

24.6.3. Método de minimización de la funcional cuadrática. Sea la familia $\{\mu_\theta, \theta \in \Theta\}$ la misma que el punto anterior. Señalemos que en el ejemplo considerado en él la estimación de la media se construía minimizando cierta sucesión de funcionales cuadráticas. En el caso general se puede analizar alguna sucesión de funcionales cuadráticas

$$K_n(x(\cdot)) = \int_0^T \int_0^T K_n(t, s) x(t) x(s) dt ds$$

y buscar la estimación de la media de $a_\theta(t)$ en forma del límite de la sucesión de las funciones $a_0^{(n)}(t)$ que dan el mínimo a la expresión

$$\int_0^T \int_0^T K_n(t, s) [x(t) - z(t)] [x(s) - z(s)] dt ds$$

para $z(\cdot)$ que varía en el conjunto M de las medias probables $\{a_\theta(\cdot), \theta \in \Theta\}$. Para que esta estimación sea conciliable (en particular, para que exista el límite $a_0^{(n)}(t)$) es suficiente que se cumplan las condiciones siguientes:

1) para $\theta_1 \neq \theta_2$

$$\lim_{n \rightarrow \infty} \int_0^T \int_0^T K_n(t, s) [a_{\theta_1}(t) - a_{\theta_2}(t)] [a_{\theta_1}(s) - a_{\theta_2}(s)] dt ds = +\infty,$$

2) existe tal métrica ρ en Θ que si $\xi(t) = x(t) - a_\theta(t)$ (θ es el valor real del parámetro), entonces, con la probabilidad 1,

$$\lim_{n \rightarrow \infty} \frac{\int_0^T \int_0^T K_n(t, s) \xi(t) [a_{\theta_1}(s) - a_0(s)] dt ds}{\int_0^T \int_0^T K_n(t, s) [a_{\theta_1}(t) - a_0(t)] [a_{\theta_1}(s) - a_0(s)] dt ds} = 0$$

uniformemente según tales θ_1 para los cuales $\rho(\theta, \theta_1) \geq \varepsilon$, cualquiera que sea $\varepsilon > 0$. Si se cumplen estas condiciones y θ_n se determina valiéndose de la igualdad $a_0^{(n)}(t) = a_{\theta_n}(t)$, entonces θ_n converge en probabilidad (en la métrica ρ) hacia 0.

24.6.4. Método de verosimilitud máxima. Consideremos dos variantes de este método.

A. Exista, para cualesquiera t_1, \dots, t_n de $[0, T]$, la densidad conjunta siempre positiva de magnitudes $x(t_1), \dots, x(t_n)$

$$p_\theta(t_1, \dots, t_n, x_1, \dots, x_n)$$

si θ es el valor real del parámetro. Eligiendo un valor $\theta_0 \in \Theta$ introduzcamos las funciones

$$f_n(\theta, x(\cdot)) = \frac{p_\theta(t_1^{(n)}, \dots, t_n^{(n)}, x(t_1^{(n)}), \dots, x(t_n^{(n)}))}{p_{\theta_0}(t_1^{(n)}, \dots, t_n^{(n)}, x(t_1^{(n)}), \dots, x(t_n^{(n)}))} \quad (6.5)$$

$\hat{\theta}_n$ es una funcional de la trayectoria en observación) donde

$$t) = t_0^{(n)} < t_1^{(n)} < \dots < t_n^{(n)} = T \quad \text{y} \quad \max_h (t_h^{(n)} - t_{h-1}^{(n)}) \rightarrow 0$$

para $n \rightarrow \infty$. Si el proceso $x(t)$ es estocásticamente continuo cualquiera que sea el valor real del parámetro, entonces, si θ_0 es el valor real del parámetro, $\lim_{n \rightarrow \infty} f_n(\theta, x(\cdot)) = 0$, con probabilidad 1, para todos los $\theta \neq \theta_0$ y para $\theta = \theta_0$ este límite es igual a 1.

Sea $\hat{\theta}_n$ un valor para el cual $f_n(\theta, x(\cdot))$ alcanza el máximo (se supone que $f_n(\theta, x(\cdot))$ es continua según θ y el mismo θ varía en un compacto). Es natural elegir a título de la estimación la magnitud $\hat{\theta}_n$. Hay que analizar la conciliabilidad de esta estimación en cada caso concreto.

B. Supongamos que $x(t) \in L_2[0, T]$. Elegimos un sistema ortonormalizado de funciones en $L_2[0, T]$ $\{\varphi_k(t), k = 1, 2, \dots\}$ y hagamos

$$x_k = \int_0^T \varphi_k(t) x(t) dt.$$

Sea $p_0^{(n)}(y_1, \dots, y_n)$ la densidad conjunta de las magnitudes x_1, \dots, x_n ; si θ es el valor real del parámetro

$$f_n(\theta, x(\cdot)) = \frac{p_0^{(n)}(x_1, \dots, x_n)}{p_{\theta_0}^{(n)}(x_1, \dots, x_n)}. \quad (6.6)$$

La estimación $\hat{\theta}_n$ se determina como el punto donde la última función alcanza el máximo.

24.6.5. Método de Bayes. Estén cumplidas las condiciones del punto anterior. Designemos con $f_n(\theta, x(\cdot))$ la función determinada por la igualdad (6.5) o (6.6). Supongamos que Θ es un conjunto abierto convexo en el espacio normalizado lineal, demos en Θ tal medida boreliana ν que la medida de cualquier conjunto abierto sea positiva. En calidad de la estimación de Bayes del parámetro θ se toma la sucesión de estimaciones

$$\hat{\theta}_n = \frac{\int \theta f_n(\theta, x(\cdot)) \nu(d\theta)}{\int f_n(\theta, x(\cdot)) \nu(d\theta)}.$$

Hay que investigar la conciliabilidad de esta estimación en cada caso concreto.

ESTADÍSTICA DE LOS PROCESOS ALEATORIOS ESTACIONARIOS EN AMPLIO SENTIDO

25.1. Propiedades de las estimaciones estadísticas para las características de procesos estacionarios

25.1.1. Problemas de la estadística de procesos estacionarios. Supongamos que se observa un proceso aleatorio $\{\xi(t), t \in T\}$ y los razonamientos apriorísticos o el test estadístico realizado de antemano permiten considerar que el proceso $\{\xi(t), t \in T\}$ tiene el aspecto de:

- 1) $\xi_0(t) = \xi_0(t), t \in T$, o bien
- 2) $\xi_0(t) = m + \xi_0(t), t \in T$, o bien

$$3) \xi(t) = \sum_{h=1}^r \theta_h a_h(t) + \xi_0(t), \text{ donde en cada caso } \xi_0(t) \text{ es un}$$

proceso aleatorio estacionario en amplio sentido con esperanza matemática nula.

Sea $x(t), t \in T_0$, la trayectoria del proceso $\xi(t)$ observada durante el tiempo T_0 , donde T_0 puede ser un intervalo en el eje de tiempo $T_0 = [a, b]$ en el caso del tiempo continuo o una sucesión de momentos de observación: $T_0 = \{t_k, k = \overline{1, n}\}$.

En el caso 1) se requiere estimar a base de la observación $x(t), t \in T_0$, la función espectral o la densidad espectral del proceso $\xi(t)$. En el caso 2) la función espectral del proceso $\xi_0(t)$ se supone conocida y hay que estimar a base de las observaciones $x(t), t \in T_0$, la media desconocida m . En el caso 3) la función espectral del proceso $\xi_0(t)$ también se supone conocida y hay que estimar a base de las observaciones $x(t), t \in T_0$, los parámetros desconocidos $\theta_1, \theta_2, \dots, \theta_r$ de la

regresión $A(t) = \sum_{h=1}^r \theta_h a_h(t)$, donde las funciones $a_h(t), k = \overline{1, r}$, se suponen conocidas.

Estos son los problemas fundamentales de la estadística de procesos estacionarios. Pueden existir variantes, por ejemplo, la estimación previa de la función espectral para la estimación posterior de la media o los parámetros de la regresión.

25.1.2. Propiedades de las estimaciones. Sea $\hat{\mu}$ una estadística destinada a resolver uno de los problemas mencionados, la cual representa una funcional de la trayectoria en observación $x(t), t \in T_0$: $\hat{\mu} = g(x(t), t \in T_0)$.

En el conjunto de estadísticas $\hat{\mu}$ es natural elegir las que tienen «buenas» propiedades. Las propiedades más deseables de las estadísticas son las siguientes.

1. Linealidad (la funcional $g(\cdot)$ debe ser lineal).
2. Carácter insesgado. Si h es la característica a estimar del proceso

$\{\xi(t), t \in T\}$ y μ , la estadística destinada a estimar h , se exige que $M\hat{\mu} = h$.

3. Conciliabilidad. La estadística $\hat{\mu}$ debe converger, en probabilidad, a h cuando aumenta el intervalo de observación.

4. Eficiencia. La estadística $\hat{\mu}$ debe tener la varianza mínima entre las estadísticas de la clase dada.

A veces hay que limitarse a exigencias más débiles que las exigencias de carácter insesgado y eficiencia.

2'. Carácter insesgado asintótico. $M\hat{\mu} \rightarrow h$ cuando el intervalo de observación crece ilimitadamente.

4'. Eficiencia asintótica. La estadística $\hat{\mu}$ debe tener la varianza asintóticamente mínima entre las estadísticas de la clase dada cuando el intervalo de observación crece ilimitadamente.

Para verificar la conciliabilidad de la estimación es suficiente cerciorarse de que su varianza tiende a cero.

25.2. Estimaciones de la media desconocida

25.2.1. Media de tiempo (estimación media aritmética). Sea $x(t)$, $t \in T_0$, la trayectoria del proceso $\xi(t) = m + \xi_0(t)$, $t \in T_0$ donde $\xi_0(t)$ es un proceso estacionario en amplio sentido con media nula y función covariacional $B(t)$. Se supone que el tiempo t es continuo y $T_0 = [a, b]$.

Entre las estimaciones insesgadas lineales del proceso estacionario medio la forma más simple la tiene la estadística \bar{m} , que se llama media de tiempo o estimación media aritmética:

$$\bar{m} = \frac{1}{b-a} \int_a^b x(t) dt, \quad (2.1)$$

Si el proceso estacionario en observación es ergódico la estimación media aritmética \bar{m} es conciliable. La estimación de la media m según el método de los cuadrados mínimos es la estimación media aritmética \bar{m} .

Determinemos la clase M_g de estimaciones insesgadas lineales de la media desconocida

$$M_g = \left\{ \hat{\mu} : \hat{\mu} = \int_a^b g(t) x(t) dt \right\},$$

donde las funciones $g(t)$ pertenecen a la clase de tales funciones equicontinuas y uniformemente limitadas en $[a, b]$ que $\int_a^b g(t) dt = 1$, y $x(t)$, $t \in [a, b]$, es la trayectoria del proceso $\{\xi(t), t \in T\}$ que tiene densidad espectral continua en cero.

Teorema 1. La estimación media aritmética \bar{m} tiene varianza asintóticamente mínima en la clase M_g .

Do este modo, entre las estimaciones $\hat{\mu} \in M_g$ para $b - a \rightarrow \infty$ no existen estimaciones más eficientes que la estimación media aritmética.

Se puede obtener una clase de estimaciones insesgadas lineales considerablemente más amplia que $M_{\mathcal{K}}$ investigando las estimaciones "en suspensión" del tipo

$$\hat{\mu} = \sum_{k=1}^n c_k^{(n)} x(t_k), \quad a = t_1 < t_2 < \dots < t_n = b, \quad (2.2)$$

donde $\sum_{k=1}^n c_k^{(n)} = 1$ para cualquier $n \geq 1$.

Sea $M_{\hat{\mu}}$ la clausura de la clase de estimaciones del tipo (2.2) en media cuadrática.

Teorema 2. En $M_{\hat{\mu}}$ existe, salvo la equivalencia, la única estimación \hat{m} de la media desconocida que tiene varianza mínima, con la particularidad de que

$$M\hat{m}x(t) \equiv C, \quad t \in [a, b], \quad (2.3)$$

donde $C = \inf_{\hat{\mu} \in M_{\hat{\mu}}} D\hat{\mu}$ es la varianza de la estimación \hat{m} .

25.2.2. Cálculo de las estimaciones de la media a base del pronóstico. Uno de los métodos de construcción de las estimaciones eficientes de la media para los procesos estacionarios ergódicos se basa en el análisis del pronóstico construido a partir de las observaciones $x(t)$, $t \in T_0 = [a, b]$.

Sea $\hat{x}(t)$ para $t \in [a, b]$ el mejor pronóstico insesgado lineal de las estimaciones $x(t)$, $t \in [a, b]$, es decir, $M\hat{x}(t) = m$, $M|\hat{x}(t) - \xi(t)|^2 = \min$ para todos los $t \in [a, b]$.

Introduzcamos la estadística $\hat{\mu}_A$ determinada como media del tiempo, construida mediante el pronóstico $\hat{x}(t)$: para $[a, b] \subset [-A, A]$

$$\hat{\mu}_A = \frac{1}{2A} \left[\int_{-A}^a \hat{x}(t) dt + \int_a^b x(t) dt + \int_b^A \hat{x}(t) dt \right]. \quad (2.4)$$

Para la estimación insesgada lineal \hat{m} con varianza mínima se verifica el

Teorema 3.

$$\hat{m} = \lim_{A \rightarrow \infty} \hat{\mu}_A.$$

Si el proceso $\xi_0(t)$ es tal que $\int_{-\infty}^{\infty} B(t) dt < \infty$, entonces el resultado del teorema 3 puede ser escrito de una manera más cómoda para el cálculo.

Sea $\hat{x}_0(t)$ el mejor pronóstico lineal de los valores $x(t)$, $t \in [a, b]$, hecho en supuesto de que $m = 0$ (si $t \in [a, b]$, $\hat{x}_0(t) = x(t)$).

Teorema 4. Si $\int_{-\infty}^{\infty} B(t) dt < \infty$ existe una constante d tal que

$$\hat{m} = d \int_{-\infty}^{\infty} \hat{x}_0(t) dt$$

y d se determina del único modo de la condición del carácter insesgado:
 $M\hat{m} = m$.

EJEMPLO 1. Sea igual a $B(t) = e^{-\alpha|t|}$ la función covariacional del proceso $\xi(t) = m + \hat{\xi}_0(t)$ y m desconocida.

Se observa la trayectoria $x(t)$, $t \in [a, b]$. Suponiendo que $m = 0$ encontramos el mejor pronóstico lineal

$$\begin{cases} \hat{x}_0(b+\tau) = e^{-\alpha\tau}x(b), & \tau > 0; \\ \hat{x}_0(t) = x(t), & t \in [a, b]; \\ \hat{x}_0(a-\tau) = e^{-\alpha\tau}x(a), & \tau > 0. \end{cases}$$

$$\begin{aligned} \text{Por consiguiente, } \hat{m} &= d \int_{-\infty}^{\infty} \hat{x}_0(t) dt = d \left[\int_{-\infty}^a \hat{x}_0(t) dt + \int_a^b x(t) dt + \right. \\ &\left. + \int_b^{\infty} \hat{x}_0(t) dt \right] = d \left[\frac{x(a)}{\alpha} + \int_a^b x(t) dt + \frac{x(b)}{\alpha} \right]. \end{aligned}$$

La condición del carácter insesgado da

$$\begin{aligned} m = M\hat{m} &= dM \left[\frac{x(a)}{\alpha} + \int_a^b x(t) dt + \frac{x(b)}{\alpha} \right] = \\ &= d \left[\frac{m}{\alpha} + m(b-a) + \frac{m}{\alpha} \right], \end{aligned}$$

de donde

$$d = \frac{\alpha}{2 + \alpha(b-a)}.$$

De este modo

$$\begin{aligned} \hat{m} &= \frac{x(a) + \int_a^b x(t) dt + x(b)}{2 + \alpha(b-a)}, \\ D\hat{m} &= \frac{2}{2 + \alpha(b-a)}. \end{aligned} \quad (2.5)$$

Para comparar: la varianza de la estimación media aritmética \bar{m}

$$D\bar{m} = \frac{2[e^{-\alpha(b-a)} - 1 + \alpha(b-a)]}{\alpha^2(b-a)^2}, \quad (2.6)$$

además, $D\bar{m} > D\hat{m}$, $\frac{D\hat{m}}{D\bar{m}} \rightarrow 1$, $b-a \rightarrow \infty$.

25.2.3. Ecuaciones del tipo de Wiener—Hopf. En una serie de casos se puede obtener la fórmula explícita para la estimación lineal

insesgada \hat{m} de la media desconocida m , partiendo de la representación formal

$$\hat{m} = \int_a^b x(t) dG(t). \quad (2.7)$$

La función $G(t)$ debe satisfacer la condición sobre el carácter insesgado de $\int_a^b dG(t) = 1$ y es la solución de la ecuación integral de tipo de Wiener-Hopf

$$\int_a^b B(t-s) dG(s) = C, \quad t \in [a, b] \quad (2.8)$$

que se obtiene de (2.3) para las estimaciones de este tipo.

Para los procesos de densidad espectral racional fraccional, la ecuación (2.8) siempre tiene una solución que contiene combinaciones lineales de funciones delta de Dirac y sus derivadas. Tal solución puede ser hallada explícitamente (véase el p. 25.3.3).

EJEMPLO 2. En el caso de un proceso estacionario de Márkov, la ecuación

$$\int_a^b e^{-\alpha(t-s)} dG(s) = C, \quad t \in [a, b],$$

tiene la solución

$$G(t) = \frac{C}{2} [u(t-a) - \alpha t + u(b-t)],$$

donde

$$u(t) = \begin{cases} 0, & t < 0; \\ 1, & t \geq 0. \end{cases}$$

La estimación $\hat{m} = \int_a^b x(t) dG(t)$ coincide de modo natural con la mencionada en el ejemplo 1.

EJEMPLO 3. Tenga el proceso $\xi_q(t)$ densidad espectral racional fraccional del tipo

$$f(\lambda) = \frac{1}{|Q(i\lambda)|^2},$$

donde $Q(z) = \sum_{h=0}^q q_h z^h$, q_h son números reales (proceso de autorregresión de orden q).

En este caso la solución de la ecuación (2.8) tiene el aspecto

$$\frac{dG(t)}{dt} = \frac{Cq_0}{2\pi} \{q_0 + q_1 [\delta(b-t) + \delta(t-a)] + q_2 [\delta'(b-t) - \delta'(t-a)] + \dots + q_n [\delta^{(n-1)}(b-t) + (-1)^{n-1} \delta^{(n-1)}(t-a)]\},$$

donde $\delta(t)$ es la función delta de Dirac, $\delta^{(k)}(t)$ es su k -ésima derivada. Por consiguiente

$$\hat{m} = \frac{Cq_0}{2\pi} \left\{ q_0 \int_a^b x(t) dt + q_1 [x(b) + x(a)] + \dots + q_n [x^{(n-1)}(b) + (-1)^{n-1} x^{(n-1)}(a)], \right.$$

donde

$$C = D\hat{m} = \frac{2\pi}{q_0 [2q_1 + q_0(b-a)]}.$$

25.2.4. Método de Yaglom. Tenga el proceso $\xi_a(t)$ densidad espectral racional fraccional

$$f(\lambda) = \left| \frac{P(i\lambda)}{Q(i\lambda)} \right|^2,$$

donde $P(z)$ es un polinomio de grado p , $Q(z)$, un polinomio de grado q ,

$p < q$ y sea $x(t) = \int_{-\infty}^{\infty} e^{i\lambda t} d\zeta(\lambda)$ la representación espectral de la trayectoria $x(t)$, $t \in [a, b]$. El método de Yaglom consiste en representar la mejor estimación insesgada lineal \hat{m} en la forma

$$\hat{m} = \int_{-\infty}^{\infty} \varphi_{(a, b)}(\lambda) d\zeta(\lambda) \quad (2.9)$$

y señalar las condiciones que determinan unívocamente la característica espectral $\varphi_{(a, b)}(\lambda)$ de la estimación \hat{m} y que permiten calcularla de manera eficiente.

Teorema de Yaglom. Para los procesos con densidades espectrales racionales fraccionales la característica espectral $\varphi_{(a, b)}(\lambda)$ en (2.9) se determina unívocamente por las condiciones:

a) $\varphi_{(a, b)}(\lambda)$ es una función completa tal que

$$\int_{-\infty}^{\infty} |\varphi_{(a, b)}(\lambda)|^2 f(\lambda) d\lambda < \infty;$$

b) $\varphi_{(a, b)}(\lambda)$ puede ser representada en la forma

$$\varphi_{(a, b)}(\lambda) = e^{i\lambda a} \frac{w_a(\lambda) \overline{Q(i\lambda)}}{\lambda |P(i\lambda)|^2} + e^{i\lambda b} \frac{w_b(\lambda) Q(i\lambda)}{\lambda |P(i\lambda)|^2}, \quad (2.10)$$

donde $\varphi_a(\lambda) = \frac{w_a(\lambda) \overline{Q(i\lambda)}}{\lambda |P(i\lambda)|^2}$ es una función analítica en el semiplano superior, $w_a(0) \neq 0$; $\varphi_b(\lambda) = \frac{w_b(\lambda) Q(i\lambda)}{\lambda |P(i\lambda)|^2}$ es una función analítica en el semiplano inferior, $w_b(0) = 0$;

c) $\lim_{\lambda \rightarrow 0} \varphi_{(a, b)}(\lambda) = 1$ (condición del carácter insesgado).

Las funciones $w_a(\lambda)$ y $w_b(\lambda)$, en virtud de la segunda parte de la condición a), pueden ser sólo polinomios de grado no superior a p . Para la varianza de la estimación \hat{m} , el método de Yaglom da la fórmula

$$D\hat{m} = \int_{-\infty}^{\infty} |\varphi_{(a, b)}(\lambda)|^2 f(\lambda) d\lambda = 2\pi \left| \frac{w_a(0)}{q_0} \right|^2 = 2\pi \left| \frac{w_b(0)}{q_0} \right|^2. \quad (2.11)$$

EjemPlo 4. Tenga el proceso $\xi_0(t)$ la densidad espectral $f(\lambda) = B \frac{\lambda^2 + \alpha^2}{\lambda^2 + \alpha^4}$ (modelo mixto de autorregresión y sumación móvil).

A partir de la observación $x(t)$, $t \in [a, b]$, hay que dar la mejor estimación insesgada de la media m del proceso $\xi(t) = m + \xi_0(t)$.

En este caso la condición b) del teorema de Yaglom da

$$\varphi_{a, b}(\lambda) = e^{i\lambda a} \frac{w_a(\lambda) [\lambda^2 + i\sqrt{2}\alpha\lambda - \alpha^2]}{a\lambda(\lambda^2 + \alpha^2)} + e^{i\lambda b} \frac{w_b(\lambda) [\lambda^2 - i\sqrt{2}\alpha\lambda - \alpha^2]}{\lambda(\lambda^2 + \alpha^2)}, \quad (2.12)$$

donde $w_a(\lambda)$ y $w_b(\lambda)$ son algunos polinomios de orden no superior a 1, cuyos coeficientes deben ser elegidos de tal modo que satisfagan las condiciones a), b) y c).

El segundo miembro de (2.12) es más cómodo representarlo en la forma

$$\varphi_{a, b}(\lambda) = e^{i\lambda a} \left[c_a^0 + \frac{c_a^1}{\lambda} + \frac{c_a^2}{\lambda - i\alpha} + \frac{c_a^3}{\lambda + i\alpha} \right] + e^{i\lambda b} \left[c_b^0 + \frac{c_b^1}{\lambda} + \frac{c_b^2}{\lambda - i\alpha} + \frac{c_b^3}{\lambda + i\alpha} \right], \quad (2.13)$$

donde los coeficientes c_a^k , c_b^k , $k=0, 3$ deben ser determinados.

De la condición a) se deduce que

$$\left. \begin{aligned} c_a^1 + c_b^1 &= 0; \\ e^{-a\alpha} c_a^2 + e^{-b\alpha} c_b^2 &= 0; \\ e^{a\alpha} c_a^3 + e^{b\alpha} c_b^3 &= 0. \end{aligned} \right\} \quad (2.14)$$

La condición del carácter insesgado c) da:

$$c_a^0 + c_b^0 + i(ac_a^1 + bc_b^1) + i \frac{c_a^2 + c_b^2}{\alpha} - i \frac{c_a^3 + c_b^3}{\alpha} = 1. \quad (2.15)$$

De la condición b) se deduce que el coeficiente de $e^{i\alpha\lambda}$ en el segundo miembro de (2.13) se anula para $\lambda = \frac{\pm 1 - i}{\sqrt{2}} \alpha$ y el coeficiente de $e^{i\beta\lambda}$ se anula para $\lambda = \frac{\pm 1 + i}{\sqrt{2}} \alpha$, lo que da las cuatro ecuaciones para definir $c_a^k, c_b^k, k=0,3$, las cuales faltan en (2.14) y (2.15). La resolución de las ecuaciones correspondientes nos ofrece

$$\hat{m} = d \left\{ \left(\sqrt{2} \operatorname{sh} \frac{\alpha T_0}{2} + \operatorname{ch} \frac{\alpha T_0}{2} \right) [x(a) + x(b)] + \right. \\ \left. + \alpha \left(\operatorname{sh} \frac{\alpha T_0}{2} + \sqrt{2} \operatorname{ch} \frac{\alpha T_0}{2} \right) \int_a^b x(t) dt - \right. \\ \left. - \sqrt{2} \alpha \int_a^b \operatorname{ch} \left(\frac{a+b}{2} - t \right) x(t) dt \right\},$$

donde $T_0 = b - a$,

$$d = \left[(2 + \sqrt{2} \alpha T_0) \operatorname{ch} \frac{\alpha T_0}{2} + \alpha T_0 \operatorname{sh} \frac{\alpha T_0}{2} \right]^{-1}, \\ D\hat{m} = \frac{2\pi B \left(\operatorname{sh} \frac{\alpha T_0}{2} + \sqrt{2} \operatorname{ch} \frac{\alpha T_0}{2} \right)}{\alpha (2 + \sqrt{2} \alpha T_0) \operatorname{ch} \frac{\alpha T_0}{2} + \alpha^2 T_0 \operatorname{sh} \frac{\alpha T_0}{2}}.$$

25.3. Estimaciones de los parámetros de la regresión

25.3.1. Estimaciones de los parámetros de la regresión mediante el método de los cuadrados mínimos. Supongamos que se observa el proceso del aspecto

$$\xi(t) = \sum_{h=1}^r \theta_h a_h(t) + \xi_0(t), \quad (3.1)$$

donde $\xi_0(t)$ es un proceso estacionario con media nula, $a_h(t), k = \overline{1, r}$, son funciones no aleatorias conocidas que se suponen ser linealmente independientes, $\theta_h, k = \overline{1, r}$, son parámetros desconocidos. El problema referente a la definición de las estimaciones de los parámetros θ_h de la realización $x(t), t \in [a, b]$, del proceso $\xi(t)$ se llama problema de la estimación de los parámetros de la regresión

$$A(t, \theta_1, \theta_2, \dots, \theta_r) = \sum_{h=1}^r \theta_h a_h(t).$$

En las aplicaciones radiotécnicas de los procesos estacionarios $A(t) = A(t, \theta_1, \dots, \theta_r)$ se llama señal (útil), $\xi_0(t)$ ruido estacionario.

rio. En las aplicaciones económicas, biológicas, sociológicas $A(t)$ se llama *trend* (tendencia).

El aspecto más simple lo tienen las estimaciones insesgadas lineales de los parámetros de la regresión, calculadas por el método de los cuadrados mínimos, es decir, las estimaciones $\tilde{\theta}_k$ que minimizan la funcional cuadrática:

$$\int_a^b \left| x(t) - \sum_{h=1}^r \theta_h a_h(t) \right|^2 dt.$$

Si $a_k(t) \in L_2[a, b]$, $k = \overline{1, r}$, entonces

$$\tilde{\theta}_k = \sum_{j=1}^r c_{kj}^{-1} \int_a^b \overline{a_j(t)} x(t) dt, \quad (3.2)$$

donde c_{kj}^{-1} es el (k, j) -ésimo elemento de la matriz inversa a la matriz

$$c_{kj} = \int_a^b \overline{a_k(t)} a_j(t) dt.$$

Señalemos que cuando se calculan los parámetros de la regresión por el método de los cuadrados mínimos no se supone el conocimiento de las propiedades de correlación y espectrales del proceso $\xi(t)$.

Si suponemos que la función espectral $F_0(\lambda)$ del proceso $\xi_0(t)$ es absolutamente continua y la densidad espectral $f_0(\lambda)$ es acotada y casi siempre positiva, podemos afirmar mucho más:

Teorema 1. *Para que las estimaciones $\tilde{\theta}_k$ de los parámetros θ_k por el método de los cuadrados mínimos sean conciliables, es necesario y suficiente que para todos los $\rho_1, \rho_2, \dots, \rho_r$ la función $a(t) = \sum_{k=1}^r \rho_k a_k(t)$ satisfaga la condición*

$$\int_{-\infty}^{\infty} |a(t)|^2 dt = \infty.$$

Suponiendo que el proceso $\xi_0(t)$ tiene densidad espectral racional fraccional y las funciones básicas $a_k(t)$ tienen el aspecto

$$a_k(t) = e^{i\omega_k t} t^{m_k}, \quad (3.3)$$

donde m_k son números enteros no negativos, ω_k , números reales (en este caso $A(t)$ se llama *regresión trigonométrica polinomial*), las estimaciones $\tilde{\theta}_k$ de los parámetros θ_k de la regresión son asintóticamente eficientes en el siguiente sentido.

Sean $G(a, b)$ una matriz covariacional de las mejores estimaciones insesgadas lineales de los parámetros de la regresión (su aspecto explícito se da en el punto siguiente), $\tilde{G}(a, b)$, una matriz covariacional

de las estimaciones $\tilde{\theta}_h$ de parámetros θ_h , determinados por el método de los cuadrados mínimos. Entonces $\tilde{G}(a, b) \geq G(a, b)$ (es decir, $\tilde{G}(a, b) - G(a, b)$ es una matriz no negativamente determinada y existe una función no decreciente y no negativa $g(t)$ tal que

$$\lim_{b-a \rightarrow \infty} g(b-a) \tilde{G}(a, b) = \lim_{b-a \rightarrow \infty} g(b-a) G(a, b) \neq 0.$$

25.3.2. Mejores estimaciones insesgadas lineales de los parámetros de la regresión. Si se conocen la función espectral $F(\lambda)$ y, por consiguiente, la función de correlación $B(t)$ del proceso $\xi_0(t)$, entonces se supone corrientemente que las funciones $\alpha_h(t)$, que forman la base de la regresión $A(t)$, son tales que el proceso $\xi(t)$ permite la representación

espectral $\xi(t) = \int_{-\infty}^{\infty} e^{it\lambda} d\zeta(\lambda)$, donde $d\zeta(\lambda) = \sum_{h=1}^r \theta_h \overline{\alpha_h(\lambda)} dF(\lambda) + d\zeta_0(\lambda)$, $\zeta_0(\lambda)$ es un proceso espectral que corresponde al proceso $\xi_0(t)$: $\xi_0(t) = \int_{-\infty}^{\infty} e^{it\lambda} d\zeta_0(\lambda)$, y $\alpha_h(\lambda)$ son funciones de cuadrado integrable según la medida espectral $F(\cdot)$:

$$\int_{-\infty}^{\infty} |\alpha_h(\lambda)|^2 dF(\lambda) < \infty$$

(o, abreviadamente, $\alpha_h(\lambda) \in L_2(F)$) las cuales son soluciones de la ecuación integral del tipo de Wiener-Hopf:

$$\int_{-\infty}^{\infty} e^{it\lambda} \alpha_h(\lambda) dF(\lambda) = a_h(t). \quad (3.4)$$

EJEMPLO 1. Sea $\xi(t) = \theta a(t) + \xi_0(t)$ y sean continuas casi todas las trayectorias del proceso $\xi_0(t)$. Si la función $a(t)$ tiene discontinuidad en el punto $t_0 \in (a, b)$, entonces valiéndose de la única realización de $x(t)$, $t \in [a, b]$, se puede definir exactamente los valores del parámetro θ , a saber, la estimación

$$\theta_h = \frac{1}{a(t_0+) - a(t_0-)} [x(t_0+h) - x(t_0-h)],$$

con la probabilidad 1, para $h \rightarrow 0$ converge al valor exacto θ . Para la función $a(t)$ del ejemplo dado la ecuación (3.4) no tiene soluciones en $L_2(F)$.

Teorema 2. Para que la ecuación (3.4) tenga solución en $L_2(F)$ es necesario y suficiente que

$$\inf \int_{-\infty}^{\infty} |\psi(\lambda)|^2 dF(\lambda) \neq 0$$

ó lo que es equivalente,

$$\inf_{j, l} \sum c_j c_l b(t_j - t_l) > 0,$$

donde \inf se toma según $\Psi(\lambda)$ que son sumas finitas de la forma

$$\sum_j c_j e^{i\lambda t_j}, t_j \in T,$$

y tales, que $\sum_j c_j a_k(t_j) = 1, k = \overline{1, r}$. Si la solución de la ecuación (3.4) existe, es única en $L_2(P)$.

Si las soluciones de las ecuaciones (3.4) están obtenidas el problema de la estimación de los parámetros de la regresión se reduce a resolver el sistema lineal de ecuaciones algebraicas.

Teorema 3. Si la función espectral $F(\lambda)$ y las funciones básicas $a_k(t), k = \overline{1, r}$, de la regresión $A(t) = \sum \theta_k a_k(t)$ satisfacen las condiciones del teorema 1 y $\alpha_k(\lambda)$ son soluciones de las ecuaciones (3.4), entonces las estimaciones insesgadas lineales $\hat{\theta}_k$ de los parámetros θ_k que tienen varianza mínima se determinan por las fórmulas

$$\hat{\theta}_k = \sum_{j=1}^r c_{kj} \int_{-\infty}^{\infty} \overline{\alpha_j(\lambda)} d\zeta(\lambda), \quad (3.5)$$

donde $\zeta(\lambda)$ es la representación espectral de la trayectoria $x(t), t \in [a, b]$, $c_{kj} = \text{cov}(\hat{\theta}_k, \hat{\theta}_j)$ es el (kj) -ésimo elemento de la matriz C inversa a la matriz $D = (d_{kj}, k, j = \overline{1, r})$, donde

$$d_{kj} = \int_{-\infty}^{\infty} \overline{\alpha_k(\lambda)} f(\lambda) \alpha_j(\lambda) d\lambda.$$

Si C_1 es la matriz covariacional de las estimaciones insesgadas lineales distintas de $\hat{\theta}_k, k = \overline{1, r}$, determinadas por la igualdad (3.5), entonces $C \leq C_1$ en el sentido de que la matriz $C_1 - C$ está determinada no negativamente.

25.3.3. Solución de las ecuaciones del tipo de Wiener-Hopf para las densidades espectrales racionales fraccionales. En este caso, prácticamente más importante, cuando la densidad espectral $f(\lambda)$ del proceso $\xi_0(t)$ es racional fraccional, es decir,

$$f(\lambda) = \left| \frac{P(i\lambda)}{Q(i\lambda)} \right|^2,$$

donde $P(z) = \sum_{k=1}^q p_k z^k, Q(z) = \sum_{k=1}^q q_k z^k, p < q$, la solución de las ecuaciones (3.4), así como la solución del problema de la estimación de los parámetros de la regresión, pueden ser halladas en forma explícita.

Teorema 4. Si los polinomios $P(z)$ y $Q(z)$ tienen ceros sólo en el semiplano izquierdo, ninguna de las raíces del polinomio $P(z)$ no es pura-

mente imaginaria y las funciones $a_k(t)$, $k = \overline{1, r}$, satisfacen las condiciones del teorema 1, entonces las soluciones $\alpha_k(\lambda)$ de la ecuación

$$\int_{-\infty}^{\infty} e^{i b \lambda} \alpha_k(\lambda) \left| \frac{P(i\lambda)}{Q(i\lambda)} \right|^2 d\lambda = a_k(t)$$

tiene el aspecto

$$\alpha_k(\lambda) = e^{i a \lambda} \sum_{j=0}^{n-m-1} c_{jh}^{(a)}(i\lambda)^j + e^{i b \lambda} \sum_{j=0}^{n-m-1} c_{jh}^{(b)}(i\lambda)^j + \int_a^b e^{i \lambda t} c_h(t) dt,$$

donde $c_h(t) = \frac{1}{2\pi} Q\left(\frac{d}{dt}\right) Q\left(-\frac{d}{dt}\right) v_h(t)$, $v_h(t)$ es la solución de la ecuación diferencial

$$P\left(\frac{d}{dt}\right) P\left(-\frac{d}{dt}\right) v_h(t) = a_k(t), \quad t \in (a, b)$$

con las condiciones de frontera

$$\lim_{t \uparrow a} \frac{d^l}{dt^l} Q\left(-\frac{d}{dt}\right) v_h(t) = \lim_{t \downarrow b} \frac{d^l}{dt^l} Q\left(\frac{d}{dt}\right) v_h(t) = 0,$$

$$l = \overline{0, m-1};$$

$$c_{jh}^{(a)} = \lim_{t \downarrow 0} \frac{1}{2\pi} \sum_{l=j+m+1}^n q_l Q\left(-\frac{d}{dt}\right) v_h(t);$$

$$c_{jh}^{(b)} = \lim_{t \uparrow b} \frac{1}{2\pi} \sum_{l=j+m+1}^n (-1)^l q_l Q\left(\frac{d}{dt}\right) v_h(t).$$

25.4. Estimaciones de la densidad espectral y de la función espectral de las sucesiones estacionarias

25.4.1. Periodograma. Sea $\{\xi_k(t), t \in T\}$ una sucesión a catoría estacionaria en amplio sentido (serie de tiempo) y sea $x(t), t \in T_a$, la trayectoria del proceso $\xi(t)$, donde $T_a = \{t_k, k = \overline{1, n}\}$, $t_k \geq T$.

La base de la mayoría de las estadísticas destinadas a estimar la función espectral y la densidad espectral del proceso $\{\xi_k(t), t \in T\}$ es una estadística que se llama periodograma y que se define por la igualdad

$$I_n(\lambda) = \frac{1}{2\pi n} \left| \sum_{t \in T_a} x(t) e^{i t \lambda} \right|^2.$$

Observación. Para los procesos con tiempo continuo el periodograma se determina como

$$I_{(a, b)}(\lambda) = \frac{1}{2\pi(b-a)} \left| \int_a^b x(t) e^{i t \lambda} dt \right|^2, \quad (4.2)$$

donde $x(t)$, $t \in [a, b]$, es la trayectoria del proceso estacionario en estudio. Los resultados formulados más abajo tienen análogos continuos correspondientes.

Si $f(\lambda)$ es la densidad espectral del proceso $\xi(t)$, $t \in T$, entonces

$$\lim_{n \rightarrow \infty} M I_n(\lambda) = f(\lambda), \quad (4.3)$$

es decir, el periodograma es una estimación asintóticamente insesgada de la densidad espectral. Sin embargo,

$$\lim_{n \rightarrow \infty} \text{cov} \{I_n(\lambda_1), I_n(\lambda_2)\} = \begin{cases} 2f^2(0), & \lambda_1 = \lambda_2 = 0; \\ f^2(\lambda), & \lambda_1 = \lambda_2 = \lambda; \\ 0, & \lambda_1 \neq \lambda_2, \end{cases} \quad (4.4)$$

o sea, el periodograma no es una estimación conciliable para la densidad espectral.

El periodograma $I_n(\lambda)$ considerada como un proceso aleatorio según λ tiene para n grandes trayectorias fluctuantes en virtud de (4.4).

25.4.2. Estimaciones de la función espectral. Sea

$$\hat{F}_n(\lambda) = \int_{-\pi}^{\lambda} I_n(\mu) d\mu \quad (4.5)$$

y sea $F(\lambda)$ una función espectral del proceso $\xi(t)$, $t \in T$. La estadística $\hat{F}_n(\lambda)$ es una estimación asintóticamente insesgada de la función espectral $F(\lambda)$:

$$\lim_{n \rightarrow \infty} M \hat{F}_n(\lambda) = F(\lambda). \quad (4.6)$$

Si $\xi(t)$, $t \in T$, es un proceso ergódico, la estadística $\hat{F}_n(\lambda)$ es una estimación conciliable de la función espectral, lo que justifica la definición de $\hat{F}_n(\lambda)$ como una función espectral empírica.

Si $F(\lambda)$ es absolutamente continua,

$$\lim_{n \rightarrow \infty} \sup_{\lambda} |\hat{F}_n(\lambda) - F(\lambda)| = 0. \quad (4.7)$$

Sea $\xi(t)$, $t \in T$, una sucesión estacionaria regular y $\xi(t) = \sum_{k \in T} c_k \zeta(t-k)$ su representación en forma de una sumación

móvil, $M \zeta(t) = 0$, $M \zeta^2(t) = 1$. Pongamos $\hat{F}_n^{(0)}(\lambda) = \int_0^{\lambda} I_n(\mu) d\mu$,

$F^{(0)}(\lambda) = \int_0^{\lambda} f(\mu) d\mu$, donde $f(\lambda)$ es la densidad espectral del proceso $\xi(t)$, $t \in T$.

Teorema 1. Si se cumplen las condiciones: a) $\xi(t)$ tiene el cuarto momento finito μ_4 ; b) $f(\lambda)$ es absolutamente continua; c) $c_k = O(k^\beta)$,

$\beta < -\frac{3}{2}$, entonces

$$\lim P \left\{ \max_{0 \leq \lambda \leq \pi} \sqrt{n} |\hat{F}_n^{(0)}(\lambda) - F^{(0)}(\lambda)| \leq z \right\} = P \left\{ \max_{0 \leq \lambda \leq \pi} |\eta(\lambda)| \leq z \right\}, \quad (4.8)$$

donde $\eta(\lambda)$ es un proceso gaussiano con $M\eta(\lambda) = 0$, $M\eta(\lambda)\eta(\mu) = (\mu_n - 3)F^{(0)}(\lambda)F^{(0)}(\mu) + 2\pi \int_0^{\min(\lambda, \mu)} f^2(s) ds$.

La igualdad (4.8) recuerda el resultado correspondiente para la estadística de Kolmogórov, no obstante a diferencia de la última, en el caso dado la distribución límite depende de la función en estimación.

25.4.3. Estimaciones de la densidad espectral. A título de estimación puntual de la densidad espectral $f(\lambda)$ de la sucesión estacionaria $\{\xi(t), t \in T\}$, según la observación $x(t)$, $t \in T_n = \{t_k, k = \overline{1, n}\}$, se eligen las estadísticas $\hat{f}_n(\lambda)$ de la forma

$$\hat{f}_n(\lambda) = \int_{-\pi}^{\pi} W_n(\lambda - \mu) J_n(\mu) d\mu, \quad (4.9)$$

donde $J_n(\lambda)$ es el periodograma, mientras que las funciones de espesor $W_n(\lambda)$, que se llaman *ventanas espectrales*, se eligen de modo que

1) $W_n(\lambda)$ tenga máximo fuertemente expresado en cero;

2) $\int_{-\pi}^{\pi} W_n(\lambda) d\lambda = 1$; σ

3) $D\hat{f}_n(\lambda) \rightarrow 0$ para $n \rightarrow \infty$.

La condición 1) «corta» la estimación de la frecuencia exigida, que en virtud de 1) resulta asintóticamente insesgada, y en virtud de 3), conciliable.

Si se cumplen las condiciones:

1) $\{\xi(t), t \in T\}$ es un proceso regular y $\xi(t) = \sum_{k=0}^{\infty} c_k \xi^*(t-k)$, su

representación en forma de una sumación móvil;

2) $M\xi^2(t) = 1$, $M\xi^4(t) < \infty$;

3) $c_k = O(|k|^{-(2+\delta)})$ para algún $\delta > 0$;

4) $\left| \frac{W_n^*(\mu)}{W_n^*(0)} - 1 \right| \rightarrow 0$ con $n \rightarrow \infty$ para $|\mu| \leq \frac{c}{n}$,

donde $W_n^*(\mu) = W_n(\mu) * W_n(\mu)$ es la convolución $W_n(\mu)$ con sí misma, c es cierta constante positiva, entonces la estadística $\hat{f}_n(\lambda)$ es asintóticamente normal con la media

$$M\hat{f}_n(\lambda) = \int_{-\pi}^{\pi} W_n(\lambda - \mu) f(\lambda) d\mu + O\left(\frac{\ln n}{n}\right) \quad (4.10)$$

y la varianza

$$D\hat{f}_n(\lambda) \sim \frac{2\pi}{n} \int_{-\pi}^{\pi} W_n^2(\lambda - \mu) f^2(\mu) d\mu \sim \frac{2\pi f^2(\lambda)}{n} \int_{-\pi}^{\pi} W_n^2(\mu) d\mu. \quad (4.11)$$

La estadística $\hat{f}_n(\lambda)$ es una estimación conciliable para $f(\lambda)$ si $\lim_{n \rightarrow \infty} \frac{1}{n} \int_{-\pi}^{\pi} W_n^2(\lambda) d\lambda = 0$.

Entre el conjunto de las ventanas espectrales hay una clase que se usa con más frecuencia en la práctica estadística, o sea, la clase de ventanas espectrales que permiten la representación de la forma

$$W_n(\lambda) = 2 \sum_{-n+1}^{n-1} k_n(l) e^{-i\lambda l}, \quad (4.12)$$

donde $k_n(l) = k\left(\frac{l}{m_n}\right)$, $\{m_n\}$ es cierta sucesión ilimitadamente creciente de números positivos enteros, tal que $\frac{m_n}{n} \rightarrow 0$, $n \rightarrow \infty$, $k(x)$ una función par acotada que satisface las condiciones: $k(0) = 1$, $|k(x)| < 1$ para $x \neq 0$, $\int_{-\infty}^{\infty} k^2(x) dx < \infty$.

Supongamos que la sucesión estacionaria $\xi(t)$ es regular, $\xi(t) = \sum_{k=0}^{\infty} c_k \zeta(t-k)$ es su representación en forma de una sumación móvil, $\sum_0^{\infty} |c_k| < \infty$, las magnitudes aleatorias independientes $\zeta(t)$ tienen medias nulas y el cuarto momento finito. Sea también $B(t)$ una función de correlación $\xi(t)$.

La representación referente al carácter asintótico del desplazamiento $f(\lambda) - Mf_n(\lambda)$ nos la ofrece el

Teorema 2. Si

1) para cierto $q > 1$ existe y es finito el límite

$$k_q = \lim_{x \rightarrow \infty} \frac{1 - k(x)}{|x|^q};$$

2) $\frac{n}{m_n^q} \rightarrow \infty$, $n \rightarrow \infty$;

3) $\sum_{t \in T} |t^q B(t)| < \infty$,

entonces

$$\lim_{n \rightarrow \infty} m_n^q [f(\lambda) - M\hat{f}_n(\lambda)] = \frac{k_q}{2\pi} \sum_{t \in T} |t|^q B(t) e^{it\lambda} < \infty. \quad (4.13)$$

Las propiedades covariacionales asintóticas de las estadísticas $\hat{f}_n(\lambda)$ las describe el Teorema 3.

$$\lim_{n \rightarrow \infty} \frac{n}{m_n} \text{cov} [\hat{f}_n(\lambda_1), \hat{f}_n(\lambda_2)] = \begin{cases} \lambda_1 \neq \pm \lambda_2; \\ f^2(\lambda) \int_{-\infty}^{\infty} k^2(t) dt, \lambda_1 = \pm \lambda_2 = \lambda \neq 0; \\ 2f^2(\lambda) \int_{-\infty}^{\infty} k^2(t) dt, \lambda_1 = \lambda_2 = 0, \pm \pi. \end{cases}$$

En las condiciones del teorema 2 se consigue la más rápida convergencia de la varianza $D\hat{f}_n(x)$ hacia cero si

$$m_n = O\left(n^{\frac{1}{1+2q}}\right).$$

25.4.4. Ejemplos de ventanas espectrales. Designemos con $B_n(t) = \frac{1}{n-t} \sum_{j=1}^{n-t} x_{j+t} \bar{x}_j$ la estimación de la función de correlación $B(t)$

(función de correlación empírica).

1. Transformación de Fourier finita (estimación de Daniels)

$$W_n(\lambda) = \begin{cases} \frac{m_n}{2}, & |\lambda| \leq \frac{\pi}{m_n}; \\ 0, & |\lambda| > \frac{\pi}{m_n}, \end{cases}$$

$$\hat{f}_n(\lambda) = \frac{m_n}{2\pi} \sum_{t=-n+1}^{n-1} \left(1 - \frac{|t|}{n}\right) B_n(t) \frac{\text{sen} \frac{\pi t}{n}}{\pi t} e^{it\lambda};$$

$$k(x) = \frac{\text{sen} \pi x}{\pi x}.$$

2. Estimación cortada

$$W_n(\lambda) = 2 \frac{\text{sen} \frac{2m_n+1}{2} \lambda}{\text{sen} \frac{\lambda}{2}};$$

$$\hat{f}_n(\lambda) = \frac{1}{2\pi} \sum_{t=-m_n}^{m_n} \left(1 - \frac{|t|}{n}\right) B_n(t) e^{it\lambda};$$

$$k(x) = \begin{cases} 1, & |x| \leq 1; \\ 0, & |x| > 1. \end{cases}$$

3. Estimación de Bartlett

$$W_n(\lambda) = \frac{\text{sen}^2 \frac{m_n \lambda}{2}}{m_n \text{sen}^2 \frac{\lambda}{2}};$$

$$\hat{f}_n(\lambda) = \frac{1}{2\pi} \sum_{t=-m_n}^{m_n} \left(1 - \frac{|t|}{n}\right) \left(1 - \frac{|t|}{m_n}\right) B_n(t) e^{it\lambda},$$

$$k(x) = \begin{cases} 1 - |x|, & |x| \leq 1, \\ 0, & |x| > 0. \end{cases}$$

4. Estimación Tukey-Hanning

$$\hat{f}_n(\lambda) = \frac{1}{2} \hat{f}'_n(\lambda) + \frac{1}{2} \hat{f}_n\left(\lambda - \frac{\pi}{m_n}\right) + \frac{1}{n} \hat{f}_n\left(\lambda + \frac{\pi}{m_n}\right),$$

donde $\hat{f}'_n(\lambda)$ es la estimación cortada del ejemplo 2;

$$k(x) = \begin{cases} \frac{1}{2} (1 + \cos \pi x), & |x| \leq 1; \\ 0, & |x| > 1. \end{cases}$$

25.5. Estimaciones de los parámetros de la densidad espectral

Supongamos que se observa la trayectoria $x(t)$, $t \in T_0$, del proceso estacionario $\{\xi(t), t \in T\}$. $M\xi(t) = 0$, $D\xi(t) = \sigma^2$, cuya densidad espectral está determinada, salvo uno o varios parámetros pertenecientes a cierto conjunto paramétrico Θ :

$$f(\lambda) = f(\lambda, \theta), \quad \theta \in \Theta.$$

Hay que estimar σ^2 y $\theta \in \Theta$.

En supuesto de que el proceso $\xi(t)$, $t \in T$, es regular para cualesquiera $\theta \in \Theta$ y $\xi(t) = \sum_{k=0}^{\infty} c_k(\theta) \zeta(t-k)$ es su representación en forma

de la sumación móvil, a título de las estimaciones $\hat{\sigma}_n^2$ y $\hat{\theta}_n$ de los desconocidos σ^2 y θ se puede elegir tales valores de $\hat{\sigma}_n^2$ y $\hat{\theta}_n$ para los cuales se alcanza el mínimo de la expresión

$$n \ln \sigma^2 + \frac{W_n(\hat{\theta}_n, x(t), t \in T_0)}{\sigma^2}, \quad (4.14)$$

donde $W_n(\theta, x(t), t \in T_0) = n \sum_{-n+1}^{n-1} W(t, \theta) \left(1 - \frac{|t|}{n}\right) B_n(t)$; $B_n(t) =$

$= \frac{1}{n-t} \sum_{j=t}^{n-t} x(t_{j+t}) x(t_j)$ es función de correlación empírica, $W(t, \theta)$

es coeficiente de z^t en el desarrollo de Laurent de la función

$$v(z, \theta) = \frac{1}{u(z, \theta) u\left(\frac{1}{z}, \theta\right)}, \quad u(z, \theta) = \sum_{k=0}^{\infty} c_k(\theta) z^k \quad (\text{se supone que}$$

este desarrollo es posible en el anillo que contiene circunferencia unitaria).

Teorema 1. Si se cumplen las condiciones:

1) las magnitudes aleatorias $\xi(t)$ en la representación del proceso $\xi(t)$ en forma de la sumación móvil son independientes, están igualmente distribuidas y tienen los cuartos momentos finitos;

2) el conjunto paramétrico Θ es compacto en el espacio euclídeo de dimensión finita R^k , $\theta = (\theta_1, \theta_2, \dots, \theta_k)$;

3) $|v(z, \theta^1)|^2 \neq |v(z, \theta^2)|^2$ para casi todos los $|z| = 1$, $\theta^1, \theta^2 \in \Theta$, $\theta^1 \neq \theta^2$;

4) $f(\lambda, \theta)$ y $f^{-1}(\lambda, \theta)$ son continuas en $[-\pi, \pi]$, entonces las estimaciones $\hat{\sigma}_n^2$ y $\hat{\theta}_n$ son estimaciones conciliables σ^2 y θ .

Si, además, se cumple la condición:

5) la función $\frac{1}{|v(z, \theta)|^2}$ tiene derivadas continuas según θ_j hasta el tercer orden incluso en el entorno del valor real θ_0 del parámetro θ y

$$\sum_{k=1}^{\infty} k |c_k(\theta_0)| < \infty,$$

entonces la distribución del vector $\frac{1}{\sqrt{n}}(\hat{\theta}_n - \theta_0)$ converge para $n \rightarrow \infty$ al vector normal con media cero y matriz covariacional G_0^{-1} donde G_0 tiene elementos

$$r_{lm}^0 = \lim_{\theta \rightarrow \theta_0} \frac{1}{4\pi} \int_{-\pi}^{\pi} \frac{\partial \ln |u(c^{i\lambda}, \theta)|^2}{\partial \theta_l} \frac{\partial \ln |u(c^{i\lambda}, \theta)|^2}{\partial \theta_m} d\lambda$$

($l, m = \overline{1, k}$) (en supuesto de que G_0 es no degenerada).

Alfabeto griego

Imprenta	Escritura	Nombre	Imprenta	Escritura	Nombre
A α	<i>Α α</i>	alfa	Ν ν	<i>Ν ν</i>	nu
B β	<i>Β β</i>	beta	Ξ ξ	<i>Ξ ξ</i>	xi
Γ γ	<i>Γ γ</i>	gamma	Ο ο	<i>Ο ο</i>	ómicron
Δ δ	<i>Δ δ</i>	delta	Π π	<i>Π π</i>	pi
Ε ε	<i>Ε ε</i>	epsilon	Ρ ρ	<i>Ρ ρ</i>	rho
Z ζ	<i>Ζ ζ</i>	zeta	Σ σ	<i>Σ σ</i>	sigma
H η	<i>Η η</i>	eta	ς	<i>ς</i>	
Θ θ	<i>Θ θ</i>	theta	Τ τ	<i>Τ τ</i>	tau
I ι	<i>Ι ι</i>	iota	Υ υ	<i>Υ υ</i>	upsilon
K κ	<i>Κ κ</i>	kappa	Φ φ	<i>Φ φ</i>	phi
Λ λ	<i>Λ λ</i>	lambda	Χ χ	<i>Χ χ</i>	ji o chi
M μ	<i>Μ μ</i>	mu	Ψ ψ	<i>Ψ ψ</i>	psi
			Ω ω	<i>Ω ω</i>	omega

Alfabeto gótico

Imprenta	Escritura	Valor	Imprenta	Escritura	Valor
Na	<i>Ūa</i>	a	Ūn	<i>Ūn</i>	n
Bb	<i>Ūb</i>	b	Do	<i>Ūo</i>	o
Cc	<i>Ūc</i>	c	Ɔp	<i>Ɔp</i>	p
Dd	<i>Ūd</i>	d	Dq	<i>Ūq</i>	q
Ee	<i>Ūe</i>	e	Rr	<i>Ūr</i>	r
Ff	<i>Ūf</i>	f	Ūß	<i>Ūß</i>	s
Gg	<i>Ūg</i>	g	Ɔt	<i>Ūt</i>	t
Hh	<i>Ūh</i>	h	Uu	<i>Ūu</i>	u
Ii	<i>Ūi</i>	i	Ɔo	<i>Ūo</i>	v
Jj	<i>Ūj</i>	j	Ɔw	<i>Ūw</i>	w
Kk	<i>Ūk</i>	k	Xx	<i>Ūx</i>	x
Ll	<i>Ūl</i>	l	Ɔy	<i>Ūy</i>	y
Mm	<i>Ūm</i>	m	Ɔz	<i>Ūz</i>	z

BIBLIOGRAFÍA

1. *Anderson T. W.* The statistical analysis of time series. John Wiley, New York — London — Sydney — Toronto, 1971.
2. *Bartlett M. S.* An Introduction to Stochastic Processes, Cambridge University Press, 1955.
3. *Бикялис А.* О центральной предельной теореме в R^k . I. — Литов. мат. сб., 1971, вып. 1, с. 27—58; II — Литов. мат. сб., 1972, вып. 2, с. 73—84; III — Литов. мат. сб., 1972, вып. 3, с. 19—35. (*Bikialis A.* Sobre el teorema del limite central en R^k .)
4. *Billingsley P.* Ergodic Theory and Information, John Wiley, New York, 1965.
5. *Blackwell D. and Girshick M. A.* Theory of Games and Statistical Decisions, John Wiley, New York, 1954.
6. *Большев Л. П., Смирнов Н. В.* Таблицы математической статистики. М., «Наука», 1965, 464 с. (*Bolshev L. N., Smirnov N. V.* Tablas de la estadística matemática.)
7. *Боровков А. А.* Вероятностные процессы в теории массового обслуживания. М., «Наука», 1972, 367 с. (*Borovkov A. A.* Procesos probabilísticos en la teoría de las colas.)
8. *Боровков А. А.* Курс теории вероятностей. М., «Наука», 1972, 287 с. (*Borovkov A. A.* Curso de la teoría de probabilidades.)
9. *Bochner S.* Lectures on Fourier Integrals, Princeton University Press, Princeton, 1959.
10. *Wald A.* Sequential Analysis, John Wiley, New York, 1947.
11. *Van der Waerden B. L.* Mathematische Statistik, Springer Verlag, Berlin — Göttingen — Heidelberg, 1957.
12. *Вентцель А. Д.* Курс теории случайных процессов. М., «Наука», 1975, 319 с. (*Ventsel A. D.* Curso de la teoría de procesos aleatorios.)
13. *Гузман И. И., Скороход А. В.* Введение в теорию случайных процессов. М., «Наука», 1965, 654 с. (*Guzman I. I., Skorohod A. V.* Introducción a la teoría de procesos aleatorios.)
14. *Гузман И. И., Скороход А. В.* Теория случайных процессов. В 3-х т. Т. 1—3. М., «Наука», 1971—1975. (*Guzman I. I., Skorohod A. V.* Teoría de procesos aleatorios.)
15. *Гнеденко Б. В.* Курс теории вероятностей. Изд. 4-е. М., «Наука», 1965, 400 с. (*Gnedenko B. V.* Curso de la teoría de probabilidades.)
16. *Гнеденко Б. В., Колмогоров А. Н.* Предельные распределения для сумм независимых случайных величин. Л.—М., Гостехиздат, 1949, 264 с. (*Gnedenko B. V., Kolmogorov A. N.* Distribuciones límites para las sumas de magnitudes aleatorias independientes.)

17. Grenander U. Stochastic Processes and Statistical Inference, Almqvist and Wilsells Boltryckeri A. B., 1950.
18. Doetsch G. Anleitung zum practischen Gebrauch der Laplace — Transformation und der z — Transformation, Dritte Aufgabe, R. Oldenburg, München, Wien, 1967.
19. Диткин В. А., Прудников А. П. Операционное исчисление. Изд. 2-е. М., «Высш. школа», 1975. 407 с. (Ditkin V. A., Prudnikov A. P. Cálculo operacional.)
20. Doob J. L. Stochastic Processes, John Wiley, New York, 1953.
21. Дынкин Е. В. Основания теории марковских процессов. М., Физматгиз, 1959. 227 с. (Dyunkin E. V. Fundamentos de la teoría de procesos de Márkov.)
22. Дынкин Е. В. Марковские процессы. М., Физматгиз, 1963. 859 с. (Dyunkin E. V., Procesos de Márkov.)
23. Дынкин Е. В., Юшкевич А. А. Теоремы и задачи о процессах Маркова. М., «Наука», 1967. 231 с. (Dyunkin E. V., Yushkevich A. A. Teoremas y problemas de los procesos de Márkov.)
24. Zaks Sh. The theory of Statistical Inference, John Wiley, New York, 1974.
25. Ибрагимов И. А., Линник Ю. В. Независимые и стационарно связанные величины. М., «Наука», 1965. 524 с. (Ibragimov I. A., Linnik Yu. V. Magnitudes independientes y magnitudes estacionarias relacionadas.)
26. Ибрагимов И. А., Розанов Ю. А. Гауссовские случайные процессы. М., «Наука», 1970. 384 с. (Ibragimov I. A., Rozanov Yu. A. Procesos aleatorios de Gauss.)
27. Ито К. Вероятностные процессы. Б-ка сб. «Математика». Вып. 1—2. Пер. с японского. М., изд-во иностр. лит., 1960—1963. (Ito K. Procesos probabilísticos, original en japonés.)
28. Ито К., McKean H. P. Diffusion processes and their sample paths, Springer Verlag, Berlin — Heidelberg — New York, 1965.
29. Karlin S. A First Course in Stochastic Processes. Academic Press, New York — London, 1968.
30. Kemeny J. G. and Snell L. J. Finite Markov Chains, The University Series in Undergraduate Mathematics, 1959.
31. Kendall M. and Stuart D. Distribution Theory (2nd ed.), Charles Griffin and Company Limited, London, 1962.
32. Kendall M. and Stuart D. Inference and Relationship (2nd ed.), Charles Griffin and Company Limited, London, 1967.
33. Колмогоров А. Н. Основные понятия теории вероятностей. Изд. 2-е. М., «Наука», 1974. 119 с. (Kolmogorov A. M. Conceptos fundamentales de la teoría de probabilidades.)
34. Cramer H. Mathematical Methods of Statistics, Princeton University Press, 1946.
35. Cramer H., Leadbetter M. R. Stationary and Related Stochastic Processes, John Wiley, New York, 1967.
36. Lehman E. L. Testing Statistical Hypotheses, John Wiley, New York, 1959.
37. Линник Ю. В. Метод наименьших квадратов и основы математико-статистической теории обработки наблюдений. Изд. 2-е. М., Физматгиз, 1962. 349 с. (Linnik Yu. V. Método de cuadrados mínimos y fundamentos de la teoría matemática estadística de elaboración de las observaciones.)
38. Липцер Р. М., Ширков А. Н. Статистика случайных процессов.

- М., «Наука», 1974. 696 с. (*Lévy R. M., Shiriáev A. N.* Estadística de procesos aleatorios.)
39. *Loève M.* Probability Theory (2nd ed.), Van Nostrand, Princeton, 1960.
 40. *Meyer P. A.* Probability and Potentials, Blaisdel Publishing Society, Calcutta 1967.
 41. *Монин А. С., Яглом А. М.* Статистическая гидромеханика. Ч. 1—2. М., «Наука», 1965—1967. (*Monin A. S., Yaglom A. M.* Hidromecánica estadística.)
 42. *Петров В. В.* Суммы независимых случайных величин. М., «Наука», 1972. 414 с. (*Petrov V. V.* Sumas de magnitudes aleatorias independientes.)
 43. *Писаренко В. Ф., Розанов Ю. А.* О некоторых задачах для стационарных процессов, приводящих к интегральным уравнениям, родственному уравнению Винера — Хопфа. Проблемы передачи информации, 1963, вып. 14. (*Pisarenko V. F., Rozánov Yu. A.* Acerca de los problemas para procesos estacionarios que llevan a las ecuaciones integrales afines a la ecuación de Wiener — Hoph. Problemas de la transmisión de información.)
 44. *Прохоров Ю. В.* Сходимость случайных процессов и предельные теоремы теории вероятностей.— Теория вероятностей и ее применения. 1956, вып. 1, с. 177—238. (*Próhorov Yu. V.* Convergencia de los procesos aleatorios y teoremas del límite de la teoría de probabilidades.)
 45. *Прохоров Ю. В., Розанов Ю. А.* Теория вероятностей. (Основные понятия. Предельные теоремы. Случайные процессы.) М., «Наука», 1973. 494 с. (*Próhorov Yu. V., Rozánov Yu. A.* Teoría de probabilidades. Conceptos fundamentales. Teoremas del límite. Procesos aleatorios.— Teoría de probabilidades y sus aplicaciones.)
 46. *Ramachandran B.* Advanced theory of characteristic functions, Statistical Publishing Society, Calcutta, 1967.
 47. *Rao C. R.* Linear statistical inference and its applications, John Wiley, New York, 1965.
 48. *Розанов Ю. А.* Стационарные случайные процессы. М., Физматгиз, 1963. 284 с. (*Rozánov Yu. A.* Procesos aleatorios estacionarios.)
 49. *Романовский В. И.* Дискретные цепи Маркова. М.—М., Гостехиздат, 1949. 436 с. (*Romanovski V. I.* Cadenas discretas de Márkov.)
 50. *Романовский В. И.* Математическая статистика. М., ГОНТИ, 1938. 527 с. (*Romanovski V. I.* Estadística matemática.)
 51. *Рытов С. М.* Введение в статистическую радиофизику. М., «Наука», 1966. 404 с. (*Rýtov S. M.* Introducción a la radiofísica estadística.)
 52. *Сазонов В. В.* О скорости сходимости в многомерной центральной предельной теореме.— Теория вероятностей и ее применение. 1968, вып. 1, с. 191—194. (*Sazónov V. V.* Acerca de la velocidad de convergencia en el teorema multidimensional del límite. Teoría de probabilidades y su aplicación.)
 53. *Сарымсаков Т. А.* Основы теории процессов Маркова. М., Гостехиздат, 1954. 208 с. (*Sarymsákov T. A.* Fundamentos de la teoría de procesos de Márkov.)
 54. *Свейников А. А.* Прикладные методы в теории случайных функ-

- кий. Изд. 2-е. М., «Наука», 1968. 463 с. (*Svéshnikov A. A. Métodos aplicados en la teoría de las funciones aleatorias.*)
55. *Севастьянов В. А. Ветвящиеся процессы.* М., «Наука», 1971. 436 с. (*Sevastianov V. A. Procesos ramificados.*)
 56. *Сираджинов С. Х. Предельные теоремы для однородных цепей Маркова.* Ташкент, Изд-во АН УзССР, 1955. 81 с. (*Siradzhi-nov S. J. Teoremas del límite para las cadenas homogéneas de Márkov.*)
 57. *Скоруход А. В. Исследования по теории случайных процессов.* К., Изд-во Киевского ун-та, 1961. 216 с. (*Skorojod A. V. Investi-gaciones en la teoría de los procesos aleatorios.*)
 58. *Скоруход А. В. Случайные процессы с независимыми прираще-ниями.* М., «Наука», 1964. 278 с. (*Skorojod A. V. Procesos aleato-rios con incrementos independientes.*)
 59. *Скоруход А. В. Элементы теории вероятностей и случайных процессов.* К., «Высшая школа», 1975. 218 с. (*Skorojod A. V. Elementos de la teoría de probabilidades y de los procesos alea-torios.*)
 60. *Скоруход А. В., Слободенюк Ш. Ш. Предельные теоремы для случайных блужданий.* К., «Наук. думка», 1970. 302 с. (*Skoro-jod A. V. Slobodenuk N. P. Teoremas del límite para las fluctua-ciones aleatorias.*)
 61. *Spitzer F. Principles of Random Walk,* Princeton University Press, 1964.
 62. *Wilks S. S. Mathematical statistics,* John Wiley, New York, 1962.
 63. *Feller W. An Introduction to Probability Theory and its Applica-tions,* volume 1 (2nd ed.), John Wiley, New York, 1957.
 64. *Feller W. An Introduction to Probability Theory and its Applica-tions,* volume 11, John Wiley, New York, 1966.
 65. *Hunt G. A. Markoff Processes and Potentials,* Illinois Journal of Mathematics, vol. I (1957), vol. II, № 3 (1957) vol. II, № 2 (1958).
 66. *Harris T. E. The theory of Branching Processes,* Springer Ver-lag, Berlin — Göttingen — Heidelberg, 1963.
 67. *Hannan E. J. Time series Analysis,* Methuen, London, 1960.
 68. *Hannan E. J. Multiple Time Series,* John Wiley, New York, 1970.
 69. *Хинчин А. Я. Теория корреляции стационарных случайных процессов.* — УМН, 1938, вып. 5, с. 42—51. (*Ginchin A. Ya. Teoria de correlación de los procesos aleatorios estacionarios.*)
 70. *Ченцов Н. Н. Статистические решающие правила и оптимальные выводы.* М., «Наука», 1972. 520 с. (*Chentsov N. N. Reglas estadís-ticas de resolución y deducciones óptimas.*)
 71. *Chung Kai Lat. Markov Chains with Stationary Transition Pro-babilities,* Springer — Verlag, Berlin — Göttingen — Heidel-berg, 1960.
 72. *Ширяев А. Н. Статистический последовательный анализ.* М., «Наука», 1969. 229 с. (*Shiriáev A. N. Análisis sucesivo estadístico.*)
 73. *Яглом А. М. Введение в теорию стационарных случайных функ-ций.* УМН, VII, 1952, вып. 5, с. 3—168. (*Yaglom A. M. Intro-ducción a la teoría de las funciones aleatorias estacionarias.*)
 74. *Яглом А. М. Некоторые классы случайных полей в n-мерном пространстве, родственные стационарным случайным процес-сам. — Теория вероятностей и ее применения, 2, 1957, № 3, с. 293—333. (Yaglom A. M. Algunas clases de campos aleatorios*

en el espacio n -dimensional afines a los procesos aleatorios estacionarios. Teoría de probabilidades y sus aplicaciones.)

75. Яглом А. М. Спектральное представление для различных классов случайных функций.— Тр. IV Всесоюз. мат. съезда. Т. 1, 1963, с. 250—273. (Yaglom A. M. Representación espectral para diferentes clases de funciones aleatorias.)
76. Яглом А. М. Экстраполирование, интерполирование и фильтрация стационарных процессов с рациональной спектральной плотностью.— Тр. Моск. мат. о-ва, 4, 1955, с. 333—374. (Yaglom A. M. Extrapolación, interpolación y filtración de los procesos estacionarios con la densidad espectral racional.)
77. Blumenthal R. M., Gettoor R. K. Markov Processes and Potential Theory. Academic Press, 1968. 313 p.
78. Doetsch G. Handbuch der Laplace — Transformation, Bd. 1—3. Basel. Birkhauser Verl, 1950—1956.
79. Elderton W. P. Frequency Curves and Correlation. Cambridge, University Press, 1938. 199 p.
80. Grenander U. On Empirical spectral analysis of stochastic processes.— Arkiv för Matematik. Bd. 1, H. 6, 1952, p. 503—531.
81. Grenander U., Rosenblatt M. Statistical analysis of stationary time series. N.—Y. 1957.
82. Hirschman I., Widder D. The Convolution transform. Princeton, 1955. 268 p.
83. Kemeny J. G., Snell J. L., Knapp A. W. Denumerable Markov chains. N.—Y.—L., Van Nostrand, 1966. 433 p.
84. Moysal I. E. Multiplicative population chains.— Pros. Roy. Soc., A266, 1962, p. 1327.
85. Sazonov V. V. On multidimensional central limit theorem.— Sankhya. Ser. A, 39, part 2, 1968, p. 181—209.
86. Wold H. A Study in the Analysis of Stationary Time Series. Ed. 2. Stockholm, Almqvist — Wiksell, 1938. 236 p.

INDICE ALFABETICO DE MATERIAS

- Acotación en probabilidad 56
- Algebra de conjuntos 15
- Alternativa:
 - compuesta 469
 - simple 469
- Análisis sucesivo 467
- Anticipo 149
- Ausencia del efecto posterior 113, 118
- Axioma:
 - de adición de probabilidades 19
 - de adición ampliado 19
 - de continuidad 19
- Axiomas de probabilidad 16
- σ -álgebra 18
 - de Borel 20
 - numerablemente engendrada 38

- Cadena de Markov 164
 - aperiódica 191
 - ergódica 199
 - homogénea 173
 - irreducible 190
 - irreversible 191
 - periódica 181
 - positivamente reversible 196
 - reversible nula 196
- Campo aleatorio 280
 - con incrementos independientes 283
 - funciones de momento 281
 - funciones de momento centrales 281
 - funciones muestrales 280

- Campo aleatorio
 - gaussiano 283
 - homogéneo 287
 - en la esfera 298
 - isótropo 296, 298
 - medible 284
 - medida espectral 288, 292
 - densidad 288
 - separable 285
 - valor medio 281
- Cantidad de información 489
- Característica:
 - de la funcional W 57
 - de una martingala integrable de modo cuadrático 310
- Características muestrales 483
- Carga invariante 180
- Clase:
 - aperiódica 186, 191
 - ergódica 186
 - periódica 191
- Clases de estados comunicantes 190
- Clases ergódicas 185
- Coefficiente:
 - de amplificación del filtro 250
 - de confianza 496
 - de correlación 32
 - de correlación recíproca 237
 - de difusión 362, 431
 - pe interrupción 362
 - de mezclado fuerte 278
 - de traslado 362, 431
- Componente de factorización 158
 - negativo 158
 - positivo 158
- Composición 59

- Condición:**
 de equivalencia de las medidas 525, 538
 de estimación insesgada 233
 de compacidad débil 420
 de Cramer 92
 de Doblin 184
 de Kolmogórov—Chentsov 286
 de Liapunov 78
 de Lindeberg 77, 82
 de mezclado fuerte 278
 de ortogonalidad (singularidad) de medidas 523, 532, 536
 de pequeñez uniforme 75
 de ramificación 394
 de realización física del filtro 251
- Condiciones de convergencia:**
 hacia una distribución aritmética 103
 hacia una distribución degenerada 102, 105
 hacia una distribución general de Poisson 103
 hacia una distribución normal 103, 107
- Conjunto:**
 boreliano 20
 cilíndrico 204
- Continuidad:**
 con la probabilidad 1 214, 321
 en media cuadrática 224
 estocástica 206
 —uniforme 338
- Convergencia:**
 casi por cierto 53
 con la probabilidad 1 53
 en media cuadrática 49, 223
- Convergencia**
 en media del orden r 49
 en probabilidad 46
- Convolución** 59, 63, 94
- Correlación:**
 de Mills 121
 de verosimilitud 472
- Covariación** 32, 224
- Criterio:**
 conciliable 471
 de carácter totalmente monótono 63
 de la hipótesis 469
 de la relación de verosimilitud 472, 473
- de Kolmogórov—Smirnov 476
 de Neumann—Pearson 471
 de Pearson 473
 de reversibilidad 151
 de series de Wald—Wolfowitz 476
 óptimo 470
 para distinguir la cadena de Márkov 168
 randomizado 470
 sucesivo 476
 uniformemente más potente 470
- Criterios:**
 de aceptación 469
 no paramétricos 475
 no paramétricos de orden 476
- Cumulante** 376
- Densidad:**
 de distribución 22
 —condicional 36, 39
 espectral 241, 288, 290
 —racional fraccionaria 254
- Descomposición:**
 de Doob 308
 de Doob—Meyer 309
 de Lebesgue 241
 de Levi 371
 de Hiesz 308
 de Wold 263
- Desigualdad:**
 de Berry—Esseen 85
 de Cramer—Rao 488
 de Chébishev 49
 de Doob 302
 de Esseen 85
 de Kolmogórov 52
- Desviación:**
 estándar 32
 media 120
- Distinción infalible de las hipótesis** 532
- Distribución:**
 absolutamente continua 22
 aritmética (véase distribución reticular)
 beta 122
 binomial 23, 44, 59, 110
 binomial generalizada 115
 binomial negativa 112
 conjunta 24

continua 22
 de arco seno 124
 —generalizada 124
 de Bernoulli 110
 de Borel—Tanner 117
 de Cauchy 124
 de dimensiones finitas 203
 de Dirichlet 141
 de Erlang 122
 de Kapteyn 130
 de Laplace (véase distribución
 exponencial doble)
 de magnitud del anticipo 383
 de máximo 160
 —del proceso de Wiener 391
 de Maxwell 127
 de Pareto 130
 de Pascal 112
 de Poisson 23, 59, 66, 114
 —generalizada 99
 —compleja 60
 de Polya 114
 de producto de dos magnitudes
 aleatorias independientes 29
 de razón de dos magnitudes
 aleatorias independientes 29
Distribución:
 de razón de la varianza de
 Fisher 132
 de Rayleigh 127
 de Simpson 118
 de Snedekor 128
 de Student 127
 —multidimensional 141
 de suma de dos magnitudes
 aleatorias independientes 28,
 59
 de Sherman 131
 de Weibull—Gnedenko 132
 de Wishart 142
 divisible infinitamente 96
 degenerada 110
 discreta 25
 estable 100, 142
 exponencial 64, 118
 exponencial doble 124
 exponencial potencial 122
 gamma 64, 66, 121
 geométrica 23, 113
 hiperexponencial 119
 hipergeométrica 45, 113
 inicial 314
 logarítmica normal 129

logarítmica 116
 logística 130
 marginada 37
 marginal 25
 multinomial (véase distribu-
 ción polinomial)
 no central 126
 normal 66, 120
 —bidimensional 71
Distribución:
 —estándar 121
 —multidimensional 71, 136
 — — degenerada 136
 — — no degenerada 136, 137
 reticular 23
 polinomial 25, 45, 136
 potencial 131
 semicontinua inferiormente 157
 simétrica 52
 uniforme 66
 χ 126
 χ^2 124
 χ no central 126
 F (véase distribución de Sne-
 dekor)
 t (véase distribución de Stu-
 dent)
 z (véase distribución de la
 razón de varianza de Fisher)
Distribuciones:
 de Burr 133 $\frac{1}{2}$
 de Pearson 134
Dominio:
 confidencial 495
 crítico 469
 de aceptación de la hipótesis
 469
Ecuación:
 de Fokker—Planck 432
 de Kolmogórov—Chapman 169,
 313
 de Kolmogórov inversa 431
 —normal 432
 de regeneración 147
 de verosimilitud 494
 de Wiener—Hopf 551
 diferencial estocástica
 Elección aleatoria 478
 Elipses de iguales probabilidades
 139

- Endomorfismo 274
- Equivalencia estocástica 204, 210
- Error:
- de filtración 258
 - de interpolación 258
 - de primer género 469
 - de pronosticación 258
 - de segundo género 469
- Espacio:
- de magnitudes aleatorias de Hilbert 223
 - de valores de un proceso estacionario en amplio sentido 238
 - físico 203
 - medible 19
 - muestral 465
 - probabilístico 19
- Esperanza matemática 29, 30
- condicional 36, 39
- Esquema de series 80
- Estadística 466
- de Kolmogórov 476
 - de orden 478
 - de rango 478
 - de Smirnov 482
 - de Smirnov—Pearson 125
 - suficiente 490
- Estado:
- absorbente 343
 - de paso 343
 - de retención 343
 - irreversible 191
 - no real 190
 - real 190
 - reversible 191
 - nulo 195
 - positivo 195
- Estados comunicantes 190
- Estimación:
- cortada 563
 - de Bartlett 564
 - de Daniels 563
 - de densidad espectral 561
 - de función espectral 560
 - de parámetros de distribución de Bernoulli 502
 - de distribución exponencial 506
 - de distribución de Poisson 503
 - de los parámetros de distribución gamma 504
 - de distribución normal 99
 - de regresión 555
 - distribución uniforme 504
 - de los procesos estacionarios 548
- Estimación
- de Tykey—Hanning 564
 - insesgada 467
- Estimaciones:
- conciliables 543
 - del método de cuadrados mínimos 508
 - de la varianza 500
 - de la verosimilitud máxima 493
 - eficientes conjuntas 502
 - insesgadas 467
 - insesgadas asintóticas 549
 - eficientes 490, 549
 - suficientes 490
- Experimento determinista 13
- Experimento aleatorio 13
- Exponente característico 143
- Extrapolación 265
- Extrapolación de un proceso aleatorio 233
- Filtro:
- característica de frecuencia 249
 - coeficiente de amplificación 250
 - de alta frecuencia 251
 - de baja frecuencia 251
 - de banda 250
 - fase 250
 - físicamente realizable 251
 - de frecuencia media 251
 - función impulsora de transición 250
- Filtración 257
- lineal 270
- Fluctuación aleatoria 164
- con absorción 167
 - con las fronteras 165
 - con reflexión 165
 - en el esquema de Bernoulli 154
 - exponencial 161
 - irreversible 151
 - oscilante 152
 - que se aleja a $\pm \infty$ 152
 - reversible 151
 - semicontinua 157
- Flujo de σ -álgebras 300

- Fórmula:**
 de Bayes 35
 de esperanza matemática total 36
 de esperanzas matemáticas reiteradas 39
 de inversión 67
 —con suavización 67
 de Ito 438, 444
 —generalizada 460
- Fórmula:**
 de Kotélnikov—Shannon 270
 de Levi—Ginčin 373
 de multiplicación de las probabilidades 35
 de probabilidad total 35
- Frecuencia** 15
 empírica 478
- Frontera:**
 accesible 367
 atractiva 368
 cautivadora 367
 de escape 367
 de reflexión 367
 inaccesible 367
 natural 367
 regular 367
 repelente 368
- Función:**
 armónica 180
 característica 65
 —del proceso homogéneo con incrementos independientes 375
 centradora 371
 coespectral 240
 —coespectral matricial 240
 continua en media cuadrática 224
 de autocorrelación 237
 de correlación 205, 235, 281
 —recíproca 236
 de covariación 235
 de distribución 22
 —absolutamente continua 22
 —condicional 39
- Función**
 —conjunta 39
 —marginal 25
 —multidimensional 20
 —unidimensional 20
 de la potencia de un criterio 470
- de momento 205, 281
 de regeneración 147, 151
 de variación correcta 133
 de variación lenta 60, 108
 de variación regular
 empírica 481
 espectral 240
 —empírica 500
 —matricial 240
 — — cuadrática 240
 — — — matricial 240
 estimadora 511
 estructural 227
 excesiva 180, 355
 generadora 57
 —conjunta 57
 —del proceso de regeneración 146
 —del proceso ramificado 396, 400, 406, 411
 marginal 25
 muestral 204
 positivamente definida 224
 resolutive 467
 subarmónica 180
 superarmónica 180
 totalmente monótona
- Función**
 de verosimilitud 494
- Funciones aleatorias de Márkov** 312
- Funcional:**
 aditiva 352
 característica 205
 multiplicativa 326
 superior 154
- Hipótesis:**
 alternativa 469
 estadística 466
 compuesta 466
 simple 468
- Identidad:**
 de factorización 157
 de Pollaczek—Spitzer 160
 de Wald 154, 162, 305
- Igualdad de Parseval**
 Independencia 40

en conjunto 41
de los grupos de magnitudes
aleatorias 41
Integral estocástica 227, 231,
434, 441, 456
Interpolación 257
Intervalo confidencial 497
Intervalo de regularidad 366

Juego probabilístico 305

Ley:

de arco seno 302
de arco seno. — local 155
de cero y de unidad 43, 306
de entrada 315
de logaritmo reiterado 380, 380
de los grandes números 46,
50-52
para un campo aleatorio 289,
291
reforzada 53, 54, 307

Ley local:

de arco seno 155
de logaritmo reiterado (véase
ley de l. reiterado)

Límite:

en media cadrática 49
inferior 43
superior 43

Magnitud aleatoria 21

de Hilbert 223
de valores enteros 23

Magnitud aleatoria marginada 37,
73

Magnitudes aleatorias independi-
entes 41

Magnitudes en escalera 158

Magnitud independiente del fu-
turo 324

Martingala:

integrable de modo cuadrático
310
local 309

Matriz:

de correlación 33, 236

de covariación 33, 236
de difusión 363, 431
de precisión 140
de información 492
estocástica 189, 344
inversa generalizada 511
semiestocástica 344

Medio:

de tiempo 549
probabilística 30

Medias temporales 247

Mediciones directas:

equiexactas 510
no equiexactas 510

Medida:

aleatoria (véase medida estocás-
tica)
espectral 240, 243, 288, 290
estacionaria 180
estocástica 227
estocástica ortogonal 227
de saltos de un proceso con
incrementos independientes
374
invariante 180
normada 19
vectorial 236

Medidas:

absolutamente continuas
equivalentes 218
estocásticas 307
— elementales 227
ortogonales 520
singulares 218

Método

de Bayes 547
de cuadrados mínimos 508
de ecuaciones diferenciales 389
de los momentos 493
de minimización de la funcional
cuadrática 546
del mínimo 495
de proyección 544
de verosimilitud máxima 546
de Wiener 260

Momento:

absoluto 31
central 31
de arruinamiento 156
de interrupción 316, 325
de la primera llegada 178
de Márkov 177, 304, 321, 339
de regeneración 146

- de salto 343
- espectral 241
- factorial 58
- mixto 32
- muestral 484
- Momento riguroso en escalera 158
- Movimiento browniano (véase Proceso de Wiener)
- Muestra:
 - aleatoria
 - distribución 484
 - homogeneidad 476
 - recorrido 479
 - volumen 465, 471
- Nivel de significación del criterio 470
- Núcleo:
 - del potencial
 - de L resolvente 336
- Operador:
 - característico 342
 - de Blaschke—Privalov 361
 - de desplazamiento del tiempo 333
 - de difusión 529
 - infinitesimal 335
- Parámetro de estabilidad 143
- Período de la clase de estados comunicantes 191
- Periodograma 559
- Polinomios:
 - de Chébishev 516
 - de Chébishev—Hermite 87
- Potencial 157, 181
- Principio:
 - de minimax 488
 - de cuadrados mínimos 508
- Probabilidad:
 - condicional 37
 - de degeneración de un proceso ramificado 397
 - definición clásica 17
 - geométrica 18
 - de paso 170, 313, 332
 - conservativa 338
 - con prohibición 196
 - de Feller 337
 - estocástica continua 336
 - normal 325
- Problema
 - de arruinamiento 156
 - de Dirichlet 182
- Proceso:
 - aleatorio 203
 - con incrementos independientes 371
 - con incrementos ortogonales 230 230
 - de autorregresión 256
 - de difusión 361, 430
 - de espera 156
 - de Márkov
 - casi continuo a la izquierda 323
 - de Feller 337
 - en amplio sentido 241
 - estándar 323
 - homogéneo 331
 - interrumpido 315, 324
 - irregular 343
 - normal 325
 - de pérdida 396
 - de Poisson 208
 - de regeneración 146, 153
 - con retardo 148
 - general 149
 - reglado 149
 - de reproducción y pérdida 351
 - de ruido blanco 252
 - de sumación deslizante 239, 256
 - de Wiener 209
 - espectral 242
 - estacionario:
 - con densidad espectral racional fraccionaria 254
 - determinista 263
 - en amplio sentido 236
 - en estrecho sentido 272
 - ergódico 276
 - regular 263
 - singular 263
 - totalmente indeterminista 263
 - estacionario gaussiano 237

- Proceso:
 estándar con incrementos ortogonales 239
 homogéneo según el espacio 530
 medible 210
 progresivo medible 322
 ramificado 399
 — aperiódico 408
 — con el número finito de tipos de partículas 410
 — co un mismo tipo de partículas 399
 — crítico 403
 — degenerativo 397
 — general de Márkov 414
 — indecomponible 413
 — periódico 408
 — regular 413
 separable 213
 singular 263
 — subcrítico 403
 — supercrítico 403
 sin discontinuidad de segunda especie 215
 Propiedad de Márkov 312
 rigurosa de Márkov 177, 321, 339
 Puntos on escalera superiores rigurosos 153
- Rango del proceso estacionario 242
 máximo 242
 Regla de tres sigmas 121
 Regresión:
 exponencial 515
 línea 514
 polinomial 515
 superficie 513
 trigonométrica polinomial 556
 Relación de verosimilitud 472
 Representación espectral 291
 Resolvente:
 de fluctuación aleatoria 157
 del semigrupo de operadores 336
 Resultados equiposibles 17
 Retraso 149
- Semigrupo:
 contrayente 335
 de Márkov 334
 de operadores 335
 Semiinvariante 69
 Semimartingala 300
 Serie de Cramer 92
 Serie variacional 478
 Señal útil 233, 258
- Sistema:
 de ecuaciones de Kolmogórov 346
 — primero 345
 — segundo 346
 de ecuaciones normales 509
 — sin efecto residual 164
 — sin memoria 164
 Solución mínima 347
 de Yaglom 261
 Subclase cíclica 186
 Submartingala 300
 Subproceso de proceso de Márkov 329
- Sucesión:
 estacionaria 180
 estándar de magnitudes aleatorias incorrelacionadas 239
 de magnitudes aleatorias independientes 44, 48
 de magnitudes aleatorias uniformemente integrables 50
 de sucesos independientes 43
 de σ -álgebras independientes 44
- Suceso:
 cierto 14
 elemental 14
 imposible 14
- Sucesos:
 álgebra 14
 diferencia 14
 grupo completo 14
 intersección 14
 producto 14
 suma 14
 union 14
- Sucesos independientes 43
 Supermartingala 300
 Superposición 59
 Suma normada 73
 Sustitución aleatoria del tiempo 357
- Tabú-probabilidad 196
- Teorema:
 acerca de las tres series 55
 de Birkhoff—Ginčin 275
 de Bochner—Ginčin 239

- de Borel—Cantelli 43
- de Chentsov 286
- de continuidad 60, 63, 68
- de Gauss—Márkov 512
- de Glivenko 481
- de Gnedenko 85
- de Kolmogórov, 204, 482
- de Kolmogórov—Rózanov 279
- de Levi—Lindeberg 74, 81
- de Liapunov 76
- de Lindeberg—Feller 77, 81
- de Moivre—Laplace 46, 85
- de Poisson 47
- de Radon—Nikodym 219
- de regeneración 147
 - nodal 148
- de Smirnov 482
- de Tauber 60, 64
- de Yaglom 262, 553
- del límite central 73, 120
 - multidimensional 80, 81, 82
 - para procesos estacionarios en amplio sentido 248
 - para procesos estacionarios en estrecho sentido 279
- del límite local 83
 - de Gnedenko 85
 - de Moivre—Laplace 85
- ergódico 199
 - para las cadenas de Márkov 198
 - para procesos estacionarios en amplio sentido 247
 - para procesos estacionarios en estrecho sentido 275
- Teorema integral de Moivre—Laplace 46, 74, 80
- Tiempo de vida 316
- Tiempo local 376
- Transformaciones:
 - de Laplace 61
 - del espacio fásico 354
 - que conserva la medida 265
- Valor real del parámetro en estimación 487
- Varianza 31
- Varianza muestral 486
- Ventana espectral 561
- Zona:
 - estrecha 92
 - de convergencia normal 91
 - monomial 92

A nuestros lectores:

«Mir» edita libros soviéticos traducidos al español, inglés, francés, árabe y otros idiomas extranjeros. Entre ellos figuran las mejores obras de las distintas ramas de la ciencia y la técnica: manuales para los centros de enseñanza superior y escuelas tecnológicas; literatura sobre ciencias naturales y médicas. También se incluyen monografías, libros de divulgación científica y ciencia ficción. Dirijan sus opiniones a la Editorial «Mir», 1 Rizhski per., 2, 129820, Moscú, 1—410, GSP, URSS.

TÍTULOS DE NUESTRO SELLO EDITORIAL:

Irodov I.

LEYES FUNDAMENTALES DE LA MECÁNICA

La finalidad de este libro es la de concentrar la atención sobre las principales leyes de la mecánica (de movimiento y de conservación de la energía, del impulso y del momento de impulso), así como mostrar el modo de aplicar estas leyes en la resolución de distintos problemas concretos. El libro contiene dos partes: 1) mecánica clásica y 2) mecánica relativa. En la primera parte, las leyes de la mecánica se analizan con aproximación newtoniana, o sea, con velocidades de movimiento considerablemente menores que la de la luz; en la segunda, con velocidades comparables a la de la luz.

El libro está dedicado a los estudiantes de los primeros cursos de institutos superiores y universidades con programa ampliado de física. Asimismo, puede ser útil a los estudiantes de cursos superiores y docentes.

Krasnov M., Kiseliov A., Makárenko G.

ANÁLISIS VECTORIAL

La buena preparación matemática del ingeniero moderno, indudablemente contribuye a nuevos logros de la técnica en sus distintas especialidades. Una de las disciplinas matemáticas de gran significado en la formación de un ingeniero, es el análisis vectorial, incluido hoy en los programas de los cursos de matemática superior de institutos y universidades.

La colección de problemas de análisis vectorial propuesta, contiene el mínimo necesario de problemas y ejercicios del curso de análisis vectorial correspondiente al programa de los institutos técnicos de enseñanza superior.

El libro puede ser considerado como un curso breve de análisis vectorial, en el que se comunican sin demostración los hechos básicos, ilustrándolos en ejemplos concretos. Por eso, esta obra puede ser utilizada, por una parte, para repetir los fundamentos del análisis vectorial, y, por otra, como libro de texto para quienes, sin entrar en la demostración de algunas proposiciones y teoremas, desean dominar la técnica de operación del análisis vectorial.

Efimov N.

GEOMETRÍA SUPERIOR

En este libro se examina un gran número de problemas. Se da la argumentación matemática de: la geometría euclídea, las geometrías no euclídeas de Lobachevski y Riemann, la geometría proyectiva, la geometría de Minkovski y las cuestiones geométricas de la teoría especial de la relatividad y una noción general de las formas topológicas de la geometría de curvatura constante. La obra se divide en las tres partes. El material principal se expone en las primeras dos partes. El material de la tercera parte—nociones principales de la geometría de curvatura constante—puede ser aprovechado en el trabajo de los círculos matemáticos.

El libro se caracteriza por la claridad de su exposición y es comprensible para amplios círculos de lectores, aunque las cuestiones que trata, por así decirlo, no siempre son sencillas. Esta monografía ha sido reeditada varias veces en la Unión Soviética y en otros países. Está destinada a los estudiantes de centros docentes superiores así como a todas aquellas personas que se interesan por las matemáticas.