

Master en Estadística Aplicada y Estadística para el Sector Público

INFERENCIA ESTADÍSTICA

José María Montero
Lorenzo



2008
2009

TEMA 1: FUNDAMENTOS DE LA INFERENCIA ESTADÍSTICA

TEMA 1. EPÍGRAFE 1.1.

1.1. Introducción. Concepto de muestra aleatoria. Distribución de la muestra. Estadísticos y su distribución en el muestreo. Función de distribución empírica y sus características. Teorema de Glivenko-Cantelli.

ESQUEMA DE TRABAJO:

1. Introducción.
2. Concepto de muestra aleatoria.
3. Distribución de la muestra.
4. Concepto de estadístico y su distribución en el muestreo.
5. Función de distribución empírica y sus características: La media y la varianza muestral aleatoria.
6. Teorema de Glivenko-Cantelli.

Adicional:

Ejercicios.

Demostración del teorema de Glivenko-Cantelli (con dos anexos).

Ejercicio de arrastre: Ejercicio 1.3. de Martín Pliego, F.J.; Montero Lorenzo, J.M.; Ruiz-Maya Pérez, L. (2000): “Problemas de Inferencia Estadística” 2ª ed. AC.

1. Introducción.

“El progreso científico va unido a la experimentación: el investigador realiza un experimento y obtiene varios datos; sobre la base de estos datos se sacan ciertas conclusiones, y estas operaciones suelen ir más allá de los materiales y operaciones del experimento particular realizado. En otras palabras, puede ocurrir que el científico generalice ciertas conclusiones del experimento particular a todas las clases de experimentos semejantes. Tal tipo de extensión de lo particular a lo general se denomina inferencia inductiva y es un procedimiento para hallar nuevo conocimiento científico.

Es bien sabido que la inferencia inductiva constituye un proceso arriesgado. En efecto, es un teorema de lógica que toda inferencia inductiva exacta es imposible. Una generalización perfectamente válida no puede hacerse; sin embargo sí cabe hacer inferencias inseguras, y el grado de incertidumbre es susceptible de medición si el experimento se ha realizado de acuerdo con determinados principios. Una de las misiones de la estadística consiste en conseguir técnicas para efectuar inferencias inductivas y para medir el grado de incertidumbre de tales inferencias. La medida de la incertidumbre viene expresada en probabilidad y esa es la razón por la cual hemos dedicado tanta extensión a la teoría de la probabilidad.....

Ejemplo:

Imaginemos un almacén que contiene, por ejemplo, 10 millones de semillas, de las cuales sabemos que producen flores blancas o rojas. La información que deseamos es: ¿Cuántas de estos 10 millones de semillas (o qué porcentaje) producirán flores blancas?. La única manera de estar seguros de dar una respuesta correcta a esa pregunta es plantar todas las semillas y observar el número de las que producen flores blancas. Sin embargo, esto no es posible, pues deseamos vender las semillas; aunque no quisiéramos vender las semillas, preferiríamos obtener una respuesta sin invertir tanto esfuerzo. Naturalmente, sin plantar todas las semillas y observar el color de la flor producida por cada una no podremos conocer con certeza el número de semillas que producen flores blancas.

Otra idea que se nos ocurre es: ¿Podemos plantar unas pocas semillas y, basándonos en los colores de estas pocas flores, hacer una afirmación sobre el número de semillas de los 10 millones que producirán flores blancas?. La respuesta es que no es posible hacer una predicción exacta acerca de cuántas flores blancas producirán las semillas, pero sí cabe hacer una afirmación probabilística si seleccionamos esas pocas semillas de cierta manera.

Esto es inferencia inductiva: Seleccionamos unas pocas de los 10 millones de semillas, las plantamos y observamos el número de las que producen flores blancas y, basándonos en estas pocas semillas, hacemos una predicción sobre cuántas de los 10 millones producirán flores blancas; a partir del conocimiento del color de unas pocas generalizamos al total de los 10 millones. No podemos estar seguros de nuestra respuesta, pero sí cabe tener confianza en ella en el sentido de la relación entre frecuencia y probabilidad”.

Mood, A.M; Graybill, F.A. (1976): “Introducción a la Teoría de la Estadística” (4ª ed.) Aguilar, pp.160-162.

En otros términos: Una determinada característica (el color de la flor) de una población (10 millones de semillas) sigue una distribución de probabilidad con un parámetro desconocido (binomial con parámetro desconocido p : probabilidad de que la flor sea blanca) y tomamos una muestra del valor de dicha característica en un subconjunto de elementos poblacionales para hacer inferencias sobre p .

Dicha inferencia puede ser de dos tipos:

- Estimación: No sabemos nada acerca del parámetro objeto de interés y formulamos una propuesta en base a la información que proporciona la muestra. Dicha propuesta puede ser puntual o en forma de intervalo.
- Contrastación: Se formula una conjetura acerca del parámetro de interés y la información muestral nos llevará a rechazarla o no con el apoyo de una regla de decisión o contraste. Evidentemente, la conjetura no tiene por qué referirse únicamente al valor de algún parámetro.

2. Concepto de muestra aleatoria.

Para realizar inferencias es necesario tomar información de la población a investigar, pero ¿Cómo debe ser el proceso de selección de la muestra de la población?

1. ¿Cuánta información hay que tomar?

Cuanta más mejor. Lo ideal es observar exhaustivamente la característica de interés en la **población** (llevar a cabo un censo), pero eso no es generalmente viable por muchos motivos (sobre todo económicos). Por ello, se observará únicamente en un subconjunto de un determinado tamaño de los elementos de la población a investigar. Este subconjunto concreto de observaciones de la población se denomina **MUESTRA (de observaciones) de tamaño n** aunque también se puede definir muestra como el resultado de repetir n veces un experimento aleatorio (cuando la población no consista en un número finito de entes y requiera experimentación). El conjunto de todas las muestras de tamaño n que se pueden formar a partir de las observaciones de los elementos o unidades de la población se denomina **espacio de las muestras o espacio muestral**.

2. ¿Con qué criterios debe tomarse dicha información?

Criterio ideal: Que sea una representación a escala de la población a investigar.

Ejemplo: Si en una población hay 1.000 bolas del mismo tamaño, 750 blancas y 250 negras, una muestra perfectamente representativa de la población deberá contener las tres cuartas partes de las bolas de color blanco y la cuarta parte restante de color negro.

3. ¿Se consigue la representación a escala en la realidad?

En la realidad las muestras no son representaciones perfectas o idílicas de la población. En el ejemplo anterior, un posible resultado si el tamaño de la muestra es 100 podría ser: 78

bolas blancas y 22 negras. Es más, como en la realidad no se conoce la distribución poblacional ni siquiera se puede saber de antemano si una muestra es o no representativa.

4. ¿Cómo se garantiza la representatividad de la muestra y, por tanto, la validez del proceso inferencial?

Las desviaciones respecto de la “perfecta representatividad” que se atribuyan al proceso de selección de la muestra no invalidan en absoluto los resultados del proceso inferencial, siempre y cuando tengan origen aleatorio, es decir, sean debidas al azar. El azar se encargará de proporcionar una muestra con la representatividad deseada.

Por tanto, para que la muestra sea representativa, la elección de los elementos de la población de los que se tomará información sobre la característica de interés debe de hacerse en condiciones de azar. Si se procede de esta manera la **muestra** se denominará **probabilística o aleatoria**¹. El valor de las observaciones de la característica de interés no se conoce a priori por lo que la muestra aleatoria (de observaciones) consistirá en n variables aleatorias muestrales X_1, X_2, \dots, X_n . Una vez tomada la información, de lo que se dispone es de un conjunto de valores numéricos x_1, x_2, \dots, x_n denominados realización muestral o, simplemente, muestra (de observaciones). Además, esta forma de proceder nos permitirá conocer, en términos de probabilidad, el error que se comete al utilizar la muestra como reflejo de la población.

En el muestreo probabilístico o aleatorio se distinguen dos modalidades, dependiendo del procedimiento aleatorio de extracción utilizado:

1. Muestreo con reemplazamiento:

¹ Un ejemplo de selección no probabilística son las muestras opináticas.

1.a) Cuando la población de interés no consiste en objetos tangibles a partir de los cuales se selecciona un cierto número n para formar la muestra (caso de muchos experimentos que involucren fenómenos aleatorios en la ingeniería o las ciencias físicas: tiempo de espera en un servicio, nivel de concentración de un contaminante, lanzamiento de un dado, etc.)

En estos casos se diseña un experimento y se lleva a cabo para proporcionar la observación X_1 . Se repite el experimento bajo las mismas condiciones y se obtiene X_2 y el proceso continúa hasta tener n observaciones de la característica de interés.

1.b) Cuando la población de interés consiste en un conjunto finito de objetos tangibles y se seleccionan un cierto número n para formar la muestra procediéndose como sigue: Se extrae un elemento de la población, se hacen sobre él las observaciones oportunas y se devuelve a la población. Esta técnica de muestreo es un caso especial de la anterior, dado que la población no se afecta tras cada extracción.

2. Muestreo sin reemplazamiento:

Cuando la población de interés consiste en un conjunto finito de objetos tangibles y se seleccionan un cierto número n para formar la muestra procediéndose como sigue: Tras la extracción del elemento y la observación del valor de la característica de interés, dicho elemento no se devuelve a la población.

Tanto en el muestreo con reemplazamiento como sin él las distribuciones de las variables muestrales son iguales entre sí e iguales a la distribución de probabilidad de la población de la cual proceden. Sin embargo en el muestreo sin reemplazamiento las variables muestrales no se distribuyen independientemente, cosa que sí sucede cuando existe reemplazamiento.

Ejemplo:

Sea una urna con 100 bolas, de las cuales 20 están marcadas con el número 1, 30 con el 2 y 50 con el 3. Se extraen dos bolas al azar. Determine la distribución de las variables muestrales X_1 y X_2 cuando la muestra se extrae con y sin reemplazamiento.

Solución:

La variable poblacional ξ ="puntuación de la bola extraída" presenta la siguiente distribución de probabilidad:

$\xi=x_i$	$P(\xi=x_i)$
1	0,20
2	0,30
3	0,50
Total	1

Cuando el muestreo es con reemplazamiento:

La probabilidad de que X_1 tome el valor 1 es

$$P[(X_1=1 \cap X_2=1) \cup (X_1=1 \cap X_2=2) \cup (X_1=1 \cap X_2=3)] = 0,2 \cdot 0,2 + 0,2 \cdot 0,3 + 0,2 \cdot 0,5 = 0,2$$

La probabilidad de que X_1 tome el valor 2 es

$$P[(X_1=2 \cap X_2=1) \cup (X_1=2 \cap X_2=2) \cup (X_1=2 \cap X_2=3)] = 0,3 \cdot 0,2 + 0,3 \cdot 0,3 + 0,3 \cdot 0,5 = 0,3$$

La probabilidad de que X_1 tome el valor 3 es

$$P[(X_1 = 3 \cap X_2 = 1) \cup (X_1 = 3 \cap X_2 = 2) \cup (X_1 = 3 \cap X_2 = 3)] = 0,5 \cdot 0,2 + 0,5 \cdot 0,3 + 0,5 \cdot 0,5 = 0,5$$

con lo que X_1 se distribuye igual que la población.

La probabilidad de que X_2 tome el valor 1 es

$$P[(X_1 = 1 \cap X_2 = 1) \cup (X_1 = 2 \cap X_2 = 1) \cup (X_1 = 3 \cap X_2 = 1)] = 0,2 \cdot 0,2 + 0,3 \cdot 0,2 + 0,5 \cdot 0,2 = 0,2$$

La probabilidad de que X_2 tome el valor 2 es

$$P[(X_1 = 1 \cap X_2 = 2) \cup (X_1 = 2 \cap X_2 = 2) \cup (X_1 = 3 \cap X_2 = 2)] = 0,2 \cdot 0,3 + 0,3 \cdot 0,3 + 0,5 \cdot 0,3 = 0,3$$

La probabilidad de que X_2 tome el valor 3 es

$$P[(X_1 = 1 \cap X_2 = 3) \cup (X_1 = 2 \cap X_2 = 3) \cup (X_1 = 3 \cap X_2 = 3)] = 0,2 \cdot 0,5 + 0,3 \cdot 0,5 + 0,5 \cdot 0,5 = 0,5$$

con lo que X_2 se distribuye igual que la población.

Además, puede comprobarse que las variables muestrales son independientes puesto que

$$P(X_1 = x_1 \cap X_2 = x_2) = P(X_1 = x_1) \cdot P(X_2 = x_2)$$

sean cuales sean los valores que tomen las dos variables muestrales.

En caso de no reemplazamiento:

La probabilidad de que X_1 tome el valor 1 es

$$P[(X_1 = 1 \cap X_2 = 1) \cup (X_1 = 1 \cap X_2 = 2) \cup (X_1 = 1 \cap X_2 = 3)] = \frac{20}{100} - \frac{19}{99} + \frac{20}{100} - \frac{30}{99} + \frac{20}{100} - \frac{50}{99} = 0,2$$

La probabilidad de que X_1 tome el valor 2 es

$$P[(X_1 = 2 \cap X_2 = 1) \cup (X_1 = 2 \cap X_2 = 2) \cup (X_1 = 2 \cap X_2 = 3)] = \frac{30}{100} - \frac{20}{99} + \frac{30}{100} - \frac{29}{99} + \frac{30}{100} - \frac{50}{99} = 0,3$$

La probabilidad de que X_1 tome el valor 3 es

$$P[(X_1 = 3 \cap X_2 = 1) \cup (X_1 = 3 \cap X_2 = 2) \cup (X_1 = 3 \cap X_2 = 3)] = \frac{50}{100} - \frac{20}{99} + \frac{50}{100} - \frac{30}{99} + \frac{50}{100} - \frac{49}{99} = 0,5$$

con lo que X_1 se distribuye igual que la población.

La probabilidad de que X_2 tome el valor 1 es

$$P[(X_1=1 \cap X_2=1) \cup (X_1=2 \cap X_2=1) \cup (X_1=3 \cap X_2=1)] = \frac{20}{100} - \frac{19}{99} + \frac{30}{100} - \frac{20}{99} + \frac{50}{100} - \frac{20}{99} = 0,2$$

La probabilidad de que X_2 tome el valor 2 es

$$P[(X_1=1 \cap X_2=2) \cup (X_1=2 \cap X_2=2) \cup (X_1=3 \cap X_2=2)] = \frac{20}{100} - \frac{30}{99} + \frac{30}{100} - \frac{29}{99} + \frac{50}{100} - \frac{30}{99} = 0,3$$

La probabilidad de que X_2 tome el valor 3 es

$$P[(X_1=1 \cap X_2=3) \cup (X_1=2 \cap X_2=3) \cup (X_1=3 \cap X_2=3)] = \frac{20}{100} - \frac{50}{99} + \frac{30}{100} - \frac{50}{99} + \frac{50}{100} - \frac{49}{99} = 0,5$$

con lo que X_2 se distribuye igual que la población.

Sin embargo la probabilidad conjunta no coincide con el producto de las probabilidades marginales. A modo de ejemplo:

$$P(X_1=1 \cap X_2=1) \neq P(X_1=1) \cdot P(X_2=1)$$

$$\frac{20}{100} - \frac{19}{99} \neq \frac{20}{100} - \frac{20}{99}$$

por lo que las variables muestrales no son independientes.

Pues bien, en lo que sigue, y salvo que se diga expresamente lo contrario, las muestras serán aleatorias. Y más aún, serán **muestras aleatorias simples**, entendiendo como tales las aleatorias con reemplazamiento. Como en ellas todos los elementos de la población tendrán siempre la misma probabilidad de ser seleccionados para la muestra, se tiene que:

- a) La distribución de cada variable muestral es idéntica a la de la variable poblacional.
- b) Las variables muestrales son independientes entre sí.

Definición: Una muestra aleatoria simple de tamaño n está formada por n variables muestrales

X_1, X_2, \dots, X_n independientes e idénticamente distribuidas, con la misma distribución de probabilidad que la característica poblacional, ξ , a investigar.

Por tanto, cuando una muestra se tome de forma aleatoria y con reemplazamiento² la denominaremos **muestra aleatoria simple**. Estas muestras son las que subyacen en todo el programa de inferencia estadística puesto que, como se dijo en la introducción, se podían hacer afirmaciones probabilísticas acerca de la población si la muestra se seleccionaba de determinada manera. Pues bien, el caso de muestreo aleatorio simple resultará, a estos efectos, de particular importancia³.

NOTA: Sea cual sea el procedimiento de muestreo, los principios de la Inferencia estadística son comunes para todos ellos. Sin embargo, para que un problema de Inferencia Estadística esté bien formulado tiene que incluir el procedimiento de muestreo con el que se obtienen las observaciones. Como verán más adelante, las propiedades de los estimadores dependen de su distribución de probabilidad y ésta depende del procedimiento de muestreo. Por consiguiente, la formación de estimadores no es una operación independiente del procedimiento de muestreo que se adopte.

En la práctica, para obtener una m.a.s. se suele recurrir a una tabla de números aleatorios mediante el siguiente procedimiento⁴:

- 1) Se enumeran los miembros de la población de 1 a N.
- 2) Se elige de forma arbitraria un lugar en la tabla de números aleatorios. Por ejemplo: fila 3, columna 5.

² O en condiciones de independencia de las variables muestrales.

³ No obstante, en caso de que la población sea infinita o muy grande no haremos distinción sobre si el muestreo es con o sin reemplazamiento, pues ello será irrelevante para nuestro propósito.

⁴ Si la población es infinita, simplemente se realizan n experimentaciones de forma independiente.

3) Avanzando por filas o por columnas seleccionamos los n números distintos contados a partir del seleccionado.

Una tabla de cuatro dígitos y otra de cinco pueden verse en Casas Sánchez, J.M. (1996): “Inferencia Estadística para Economía y Administración de Empresas”. CEURA.

3. Distribución de la muestra

La distribución de probabilidad (conjunta) de la muestra vendrá dada:

- a) En el caso discreto por las muestras que conforman el espacio muestral y sus respectivas probabilidades.
- b) En el caso continuo por la función de densidad conjunta de las variables muestrales.

Sea una muestra genérica $\{x_1, x_2, \dots, x_n\}$. En el caso discreto su probabilidad de obtención es

$$P(X_1 = x_1; X_2 = x_2; \dots; X_n = x_n) = \\ = P(X_1 = x_1) \cdot P(X_2 = x_2 / X_1 = x_1) \cdot \dots \cdot P(X_n = x_n / X_1 = x_1; X_2 = x_2; \dots; X_{n-1} = x_{n-1})$$

y si m.a.s.

$$= P(\xi = x_1) \cdot P(\xi = x_2) \cdot \dots \cdot P(\xi = x_n)$$

En el caso continuo se tiene que

$$P(x_1 < X_1 \leq x_1 + d x_1) = f(x_1) d x_1$$

$$P(x_2 < X_2 \leq x_2 + d x_2 / x_1 < X_1 \leq x_1 + d x_1) = f(x_2 / x_1) d x_2$$

.....

.....

$$P(x_n < X_n \leq x_n + d x_n / x_1 < X_1 \leq x_1 + d x_1; \dots; x_{n-1} < X_{n-1} \leq x_{n-1} + d x_{n-1}) = f(x_n / x_1; \dots; x_{n-1}) d x_n$$

con lo que la probabilidad elemental de obtención de un resultado muestral $\{x_1, x_2, \dots, x_n\}$ es

$$P(x_1 < X_1 \leq x_1 + d_{x_1}; x_2 < X_2 \leq x_2 + d_{x_2}; \dots; x_n < X_n \leq x_n + d_{x_n})$$

$$= f(x_1) \cdot f(x_2/x_1) \cdot \dots \cdot f(x_n/x_1, x_2, \dots, x_{n-1}) d_{x_1} d_{x_2} \dots d_{x_n}$$

y si m.a.s.

$$= f(x_1) \cdot f(x_2) \cdot \dots \cdot f(x_n) d_{x_1} d_{x_2} \dots d_{x_n} \quad \text{donde } f_{x_i}(x_i) = f_{\xi}(x) \quad \forall i$$

y como

$$P(x_1 < X_1 \leq x_1 + d_{x_1}; x_2 < X_2 \leq x_2 + d_{x_2}; \dots; x_n < X_n \leq x_n + d_{x_n}) = f(x_1; x_2; \dots; x_n) d_{x_1} d_{x_2} \dots d_{x_n}$$

entonces

$$f(x_1; x_2; \dots; x_n) = f(x_1) f(x_2) \dots f(x_n) \quad \text{con } f_{x_i}(x_i) = f_{\xi}(x) \quad \forall i$$

Ejemplo: Sea una población binomial $(1;p)$ de la que se extrae una m.a.s. de tamaño 3. Determine la distribución de probabilidad que gobierna la muestra.

Dado que una variable aleatoria binomial $(1,p)$ únicamente puede tomar los valores 0 (fracaso) y 1 (éxito), con probabilidades p y q respectivamente, y que la muestra es aleatoria simple (las variables muestrales son independientes y se distribuyen $B(1;p)$), se tiene la siguiente distribución de probabilidad de la muestra:

$X_1=x_1; X_2=x_2; X_3=x_3$	$P(X_1=x_1; X_2=x_2; X_3=x_3)$
1 ; 1; 1	p^3
1 ; 1; 0	$p^2 q$
1; 0 ; 1	$p^2 q$
0 ; 1 ; 1	$p^2 q$
1 ; 0 ; 0	pq^2

0 ; 1 ; 0	pq^2
0 ; 0 ; 1	pq^2
0 ; 0 ; 0	q^3

o también

$$P(X_1 = x_1; X_2 = x_2; X_3 = x_3) = P(X_1 = x_1) \cdot P(X_2 = x_2) \cdot P(X_3 = x_3)$$

$$= p^{x_1} q^{1-x_1} \cdot p^{x_2} q^{1-x_2} \cdot p^{x_3} q^{1-x_3} = p^{x_1+x_2+x_3} q^{3-x_1-x_2-x_3} \quad x_1, x_2, x_3 = 0, 1$$

Ejemplo: Sea una población $N(\mu; \sigma)$, de la cual se extrae una muestra aleatoria simple de tamaño n . Determine la función de densidad conjunta de la muestra.

$$f(x_1; x_2; \dots; x_n) = f(x_1) \cdot f(x_2) \cdot \dots \cdot f(x_n)$$

$$= \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{1}{2} \frac{(x_1 - \mu)^2}{\sigma^2}} \cdot \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{1}{2} \frac{(x_2 - \mu)^2}{\sigma^2}} \cdot \dots \cdot \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{1}{2} \frac{(x_n - \mu)^2}{\sigma^2}}$$

$$= \left(\frac{1}{\sigma \sqrt{2\pi}} \right)^n e^{-\frac{1}{2} \sum_{i=1}^n \frac{(x_i - \mu)^2}{\sigma^2}} \quad -\infty < x_i < \infty$$

4. Estadísticos y su distribución en el muestreo.

Uno de los problemas fundamentales de la Estadística se presenta cuando queremos estudiar una característica ξ de una población, de la cual conocemos su modelo de su distribución de probabilidad pero no algún parámetro θ del mismo. En este caso se toma una muestra aleatoria simple de tamaño n de la población y se calcula el valor de alguna función de las observaciones que se supone es la que mejor estima el valor del parámetro desconocido.

DEFINICIÓN: Se denomina estadístico a cualquier función real de las variables muestrales X_1, X_2, \dots, X_n .

Ejemplos de estadísticos:

$$f(X_1, X_2, \dots, X_n) = X_1 + X_2 + \dots + X_n$$

$$f(X_1, X_2, \dots, X_n) = \frac{X_1 + X_2 + \dots + X_n}{n}$$

$$f(X_1, X_2, \dots, X_n) = \frac{X_1^3 + X_2^3 + \dots + X_n^3}{n}$$

$$f(X_1, X_2, \dots, X_n) = \min(X_1; X_2; \dots; X_n)$$

En general, denotaremos un estadístico por $T(\mathbf{X}) = f(X_1, X_2, \dots, X_n)$ y, evidentemente, es una variable aleatoria puesto que es función de las variables muestrales que son aleatorias. Por tanto, cada estadístico tendrá su propia distribución de probabilidad en el muestreo (función de cuantía o función de densidad).

La distribución de probabilidad en el muestreo de un estadístico es un concepto fundamental en la inferencia estadística, por cuanto nos permitirá analizar la bondad de dicho estadístico, en relación a otros, a la hora de su utilización para la realización de inferencias. Así mismo, permitirá la evaluación probabilística de los resultados que proporcione.

En la gran mayoría de las ocasiones los parámetros poblacionales a estimar son, como es lógico, los más importantes: la media y la varianza. Por ello, parece de sentido común que la media muestral $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ y la varianza muestral $S_X^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$ sean estadísticos relevantes cuya distribución de probabilidad en el muestreo necesita ser conocida por su amplia utilización. Estos estadísticos a los que se les confiere la característica de “ser útiles para estimar” se denominan **estimadores**.

Ejemplo: Sea una población con función de densidad de probabilidad⁵

$$f(x) = \frac{2}{\theta} x e^{-\frac{x^2}{\theta}} \quad x > 0 \quad \theta > 0. \quad \text{Determine la distribución en el muestreo del}$$

$$\text{estadístico } T(X) = \sum_{i=1}^n X_i^2.$$

Solución: La función de distribución de $Y=X^2$ es:

$$F_Y(y) = P(Y \leq y) = P(X^2 \leq y) = P(-\sqrt{y} \leq X \leq \sqrt{y}) = F_{\text{sub}X}(\sqrt{y}) - F_X(-\sqrt{y}) = F_X(\sqrt{y}) \quad y > 0$$

siendo X la variable poblacional. La función de densidad de $Y=X^2$ será por tanto:

$$f_Y(y) = f_X(\sqrt{y}) \frac{1}{2\sqrt{y}} \quad y > 0$$

$$f_Y(y) = \frac{2\sqrt{y}}{\theta} e^{-\frac{y}{\theta}} \frac{1}{2\sqrt{y}} = \frac{1}{\theta} e^{-\frac{y}{\theta}} \quad y > 0$$

que no es sino la función de densidad de una Gamma (1;1/θ) o exponencial de parámetro 1/θ. Y como las variables Y_i son independientes y se distribuyen idénticamente, se tiene que el

⁵ Correspondiente a una distribución de Rayleigh.

estadístico $T(\mathbf{X})$ seguirá un modelo Gamma $(n; 1/\theta)$ dada la reproductividad o aditividad del modelo gamma $(r; \mu)$ respecto a μ .

$$f_T(t) = \frac{\frac{1}{\theta^n}}{\Gamma(n)} t^{n-1} e^{-\frac{t}{\theta}} \quad t > 0$$

Ejemplo: Sea una población con función de densidad de probabilidad

$$f(x) = \frac{\mu}{(1+x)^{1+\mu}} \quad x > 0 \quad \mu > 0. \quad \text{Determine la distribución en el muestreo del}$$

$$\text{estadístico } T(\mathbf{X}) = \sum_{i=1}^n \ln(1+X_i).$$

Solución: Considérese la variable aleatoria $Y = \ln(1+X)$. Entonces:

$F_Y(y) = P(Y \leq y) = P(\ln(1+X) \leq y) = P(1+X \leq e^y) = P(X \leq e^y - 1) = F_X(e^y - 1) \quad y > 0$
siendo X la variable poblacional. La función de densidad de $Y = \ln(1+X)$ será por tanto:

$$f_Y(y) = f_X(e^y - 1) e^y \quad y > 0$$

$$f_Y(y) = \frac{\mu}{(1+e^y-1)^{1+\mu}} e^y = \mu (e^y)^{-(1+\mu)} e^y = \mu e^{-\mu y} \quad y > 0$$

que no es sino la función de densidad de una Gamma $(1; \mu)$. Y como las variables muestrales se distribuyen independientemente e igual que la población entonces, dada la reproductividad del modelo exponencial respecto de μ , $T(\mathbf{X}) = \sum_{i=1}^n \ln(1+X_i)$ seguirá una distribución gamma $(n; \mu)$

$$f_T(t) = \frac{\mu^n}{\Gamma(n)} t^{n-1} e^{-\mu t} \quad t > 0$$

5. Función de distribución empírica y sus características⁶.

Como ha podido apreciarse en el punto 3, la distribución de la muestra suele identificarse con su función de cuantía o de densidad conjunta (según que la distribución poblacional sea discreta o continua), que proporcionan la probabilidad o densidad de probabilidad con la cual puede presentarse cada muestra concreta en el proceso de muestreo.

En un sentido totalmente distinto, **a cada muestra concreta** se le puede asociar una función de distribución denominada función de distribución empírica $F_n^*(x)$ que se define como

$$F_n^*(x) = \frac{\text{Número de elementos de la muestra} \leq x}{n}$$

En otros términos, representa la frecuencia con la que los elementos seleccionados en la muestra no superan el valor x de la característica bajo estudio. Por tanto es, siempre, una función discreta (escalonada) y no tiene relación directa con la distribución de la muestra ni con la de la

⁶ En este epígrafe y el siguiente se ha tomado como base el texto VELEZ IBARROLA, R. ; GARCÍA PÉREZ, A. (1994): "Principios de Inferencia Estadística". UNED. Madrid. Pp. 12-13, 34 y 36-39.

población, en el sentido de que una muestra concreta tendría la misma función de distribución fuese cual fuese la población de origen. Sí existe una relación indirecta e importante por cuanto cada valor de la muestra se obtiene aleatoriamente de la distribución de la población de interés y no de otra. Por ello, es razonable esperar que la función de distribución empírica proporcione una imagen aproximada de la distribución de la población de la cual se extrajo la muestra.

Ejemplo:

Sea una muestra de tamaño 5 de una población en la cual se investiga una variable aleatoria ξ . Tal muestra es $\{2, -2, 0, 1, -2\}$. Determine la función de distribución empírica de la muestra.

Solución:

La función de distribución empírica de la muestra es:

$$\left\{ \begin{array}{ll} 0 & x < -2 \\ \frac{2}{5} & -2 \leq x < 0 \\ \frac{3}{5} & 0 \leq x < 1 \\ \frac{4}{5} & 1 \leq x < 2 \\ 1 & 2 \leq x \end{array} \right.$$

Características:

Como hemos señalado anteriormente, la función de distribución empírica es una función de distribución discreta que asigna probabilidad $1/n$ a cada una de las observaciones muestrales, de forma que todos sus momentos, denominados momentos muestrales, existen y valen:

Momentos de orden r respecto del origen: $a_r = \frac{\sum_{i=1}^n x_i^r}{n}$.

Momentos de orden r respecto de la media: $m_r = \frac{\sum_{i=1}^n (x_i - \bar{x})^r}{n}$.

Es evidente que, entre los momentos respecto del origen cobra una especial importancia la media aritmética muestral y entre los momentos respecto de la media resulta especialmente

interesante la varianza muestral. Por ello, y dado que ambos son variables aleatorias, a continuación pasamos a estudiar su esperanza y su varianza, características que no dependen del modelo de distribución de probabilidad de la población.

Esperanza y Varianza de la media muestral aleatoria

ESPERANZA DE LA MEDIA MUESTRAL ALEATORIA.

Como

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n} = \frac{X_1 + X_2 + \dots + X_n}{n}$$

entonces

$$E(\bar{X}) = \frac{1}{n} E\left[\sum_{i=1}^n X_i\right] = \frac{1}{n} \sum_{i=1}^n E(X_i) = \frac{1}{n} n E(\xi) = E(\xi)$$

ya que, al tratarse de una m.a.s., las variables muestrales se distribuyen igual que la población de la cual proceden y, por tanto, tendrán su misma esperanza.

VARIANZA DE LA MEDIA MUESTRAL ALEATORIA.

$$V(\bar{X}) = \frac{1}{n^2} V\left[\sum_{i=1}^n X_i\right]$$

y como las variables muestrales son independientes la varianza de la suma de todas ellas es igual a la suma de sus varianzas, por lo que

$$V(\bar{X}) = \frac{1}{n^2} V\left[\sum_{i=1}^n X_i\right] = \frac{1}{n^2} \sum_{i=1}^n V(X_i) = \frac{1}{n^2} n V(\xi) = \frac{\sigma^2}{n}$$

Este hecho es de suma importancia en la estadística aplicada por cuanto que significa que, sea cual sea la distribución de probabilidad de la población, siempre que ésta tenga varianza finita la distribución de la media muestral se concentrará tanto más alrededor de la media poblacional cuanto más aumente el tamaño de la muestra. En consecuencia, cuanto mayor sea el tamaño de la muestra más confianza tenemos en que la media muestral sea una buena estimación de la media poblacional.

Otro hecho de relevancia es que, independientemente del modelo de distribución de la variable poblacional, para tamaños muestrales elevados la media muestral aleatoria tiende a distribuirse como una normal con la esperanza y varianza anteriormente expuestas (Teorema Central del Límite).

Esperanza y Varianza de la varianza muestral aleatoria.

ESPERANZA DE LA VARIANZA MUESTRAL ALEATORIA.

Como

$$S_x^2 = \frac{\sum_{i=1}^n (X_i - \mu)^2}{n} - (\bar{X} - \mu)^2$$

entonces

$$E(S_x^2) = E\left[\frac{\sum_{i=1}^n (X_i - \mu)^2}{n} - (\bar{X} - \mu)^2\right] = E\left[\frac{\sum_{i=1}^n (X_i - \mu)^2}{n}\right] - E(\bar{X} - \mu)^2$$

y como la esperanza de una suma de variables aleatorias coincide con la suma de las esperanzas de éstas, sean independientes o no

$$E(S_x^2) = \frac{\sum_{i=1}^n E(X_i - \mu)^2}{n} - E(\bar{X} - \mu)^2 = \frac{\sum_{i=1}^n V(\xi)}{n} - V(\bar{X}) = V(\xi) - \frac{V(\xi)}{n} = \frac{n-1}{n} V(\xi) = \frac{n-1}{n} \sigma^2$$

VARIANZA DE LA VARIANZA MUESTRAL ALEATORIA.

$$V[S_x^2] = \frac{\mu_4 - \sigma^4}{n} - 2 \frac{\mu_4 - 2\sigma^4}{n^2} + \frac{\mu_4 - 3\sigma^4}{n^3}$$

de laboriosa demostración (Puede verse en Rohatgi, V.K. (1976): "An Introduction to Probability Theory and Mathematical Statistics". John Wiley and Sons).

Esperanza y Varianza de la cuasivarianza muestral aleatoria.

Como ha podido apreciarse el valor esperado de la varianza muestral aleatoria no es la varianza poblacional, por lo que se define la cuasivarianza poblacional, estadístico cuya esperanza si coincidirá con la varianza poblacional.

ESPERANZA DE LA CUASIVARIANZA MUESTRAL ALEATORIA.

$$E[S_x^{*2}] = E\left[\frac{n}{n-1} S_x^2\right] = \frac{n}{(n-1)} E(S_x^2) = \frac{n}{(n-1)} \frac{(n-1)}{n} V(\xi) = V(\xi) = \sigma^2$$

VARIANZA DE LA CUASIVARIANZA MUESTRAL ALEATORIA.

$$\begin{aligned} V[S_x^{*2}] &= V\left[\frac{n}{n-1} S_x^2\right] = \frac{n^2}{(n-1)^2} V(S_x^2) \\ &= \frac{n^2}{(n-1)^2} \left[\frac{\mu_4 - \sigma^4}{n} - 2 \frac{\mu_4 - 2\sigma^4}{n^2} + \frac{\mu_4 - 3\sigma^4}{n^3} \right] \\ &= \frac{\mu_4}{n} + \frac{(3-n)\sigma^4}{n(n-1)} \end{aligned}$$

En caso en que la distribución sea normal, como

$$\frac{\mu_4}{\sigma^4} = 3 \quad \text{entonces} \quad \mu_4 = 3\sigma^4$$

con lo que

$$V[S_x^{*2}] = \frac{\mu_4}{n} + \frac{(3-n)\sigma^4}{n(n-1)} = \frac{3\sigma^4}{n} + \frac{(3-n)\sigma^4}{n(n-1)} = \frac{2\sigma^4}{n-1}$$

6. Teorema de Glivenko-Cantelli o Teorema Fundamental de la Estadística.

a) Cuestiones previas.

Sea una m.a.s. Para cada $x \in \mathcal{R}$ fijo, el valor de la función de distribución es una variable aleatoria⁷ que se puede expresar como

$$F_n^*(x) = \frac{1}{n} \sum_{i=1}^n I_{(-\infty, x]}(X_i)$$

donde la función indicadora $I_{(-\infty, x]}(X_i)$ toma el valor 1 si X_i toma un valor en el intervalo $(-\infty, x]$ y 0 en caso contrario.

Si el procedimiento de muestreo es aleatorio simple, entonces las variables $I_{(-\infty, x]}(X_i)$ son independientes y tienen la misma distribución de probabilidad: binomial $[1; F(x)]$ puesto que

$$\begin{cases} P[I_{(-\infty, x]}(X_i) = 1] = P[X_i \leq x] = F_\xi(x) \\ P[I_{(-\infty, x]}(X_i) = 0] = P[X_i > x] = 1 - F_\xi(x) \end{cases}$$

Por tanto, $\sum_{i=1}^n I_{(-\infty, x]}(X_i) = n F_n^*(x)$, que representa el número de elementos de la muestra de tamaño n que son iguales o inferiores a x , sigue una distribución binomial $[n; F(x)]$ y,

⁷ Evidentemente, como las X_i son variables aleatorias, para cada $x \in \mathcal{R}$ fijo el valor de la función de distribución empírica en ese punto es también una variable aleatoria.

en consecuencia, para $k = 0, 1, 2, \dots, n$, la probabilidad de que en una muestra de tamaño n el número de valores iguales o inferiores a x sea k viene dada por

$$P\left[\sum_{i=1}^n I_{(-\infty, x]}(X_i) = k\right] = \binom{n}{k} [F_{\xi}(x)]^k [1 - F_{\xi}(x)]^{n-k}$$

En otros términos: la probabilidad de que en una muestra de tamaño n el porcentaje de valores menores o iguales que x sea k/n es

$$P\left[F_n^*(x) = \frac{k}{n}\right] = \binom{n}{k} [F_{\xi}(x)]^k [1 - F_{\xi}(x)]^{n-k}$$

Además, como las variables aleatorias indicadoras $I_{(-\infty, x]}(X_i)$ son independientes y están idénticamente distribuidas, el Teorema Central del Límite nos lleva a que cuando el tamaño de la muestra es elevado ($n \rightarrow \infty$) entonces

$$\sum_{i=1}^n I_{(-\infty, x]}(X_i) = n F_n^*(x) \rightarrow^d N\left(n F(x); \sqrt{n F(x)(1 - F(x))}\right)$$

por lo que, para grandes tamaños muestrales

$$F_n^*(x) \rightarrow_{\text{aprox}} N\left(F_{\xi}(x); \sqrt{\frac{F_{\xi}(x)(1 - F_{\xi}(x))}{n}}\right)$$

La lección que se desprende de este hecho es que a medida que aumenta el tamaño muestral la distribución en el muestreo de la función de distribución empírica (para un x determinado) está cada vez más concentrada alrededor del valor de la función de distribución poblacional en ese punto⁸, por lo que para un tamaño de muestra suficientemente grande es muy probable que la muestra que se obtenga proporcione un valor de $F_n^*(x)$ muy próximo al valor desconocido de $F(x)$.

⁸ Lógico, puesto que disminuye la variabilidad (en torno a $F(x)$) de los valores de la función de distribución empírica en x .

Para precisar más⁹, supóngase que el proceso de muestreo se prolonga indefinidamente, observándose sucesivamente los valores de una sucesión de variables aleatorias independientes con distribución de probabilidad común $\{X_i\}_{i=1}^{\infty}$.

Como $\{I_{(-\infty, x]}(X_i)\}_{i=1}^{\infty}$ es una sucesión de variables aleatorias independientes idénticamente distribuidas y con esperanza $F_{\xi}(x)$, **la ley fuerte de los grandes números** permite concluir que, cuando el tamaño muestral tiende a infinito

$$\frac{1}{n} \sum_{i=1}^n I_{(-\infty, x]}(X_i) = F_n^*(x) \xrightarrow{c.s.} F_{\xi}(x)$$

lo que significa que $\forall \varepsilon \exists m \in \mathbb{N}$ tal que

$P\left[\left\{\left|F_m^*(x) - F(x)\right| < \varepsilon \cap \left\{\left|F_{m+1}^*(x) - F(x)\right| < \varepsilon \cap \left\{\left|F_{m+2}^*(x) - F(x)\right| < \varepsilon \cap \dots\right\}\right\} > 1 - \varepsilon\right]$
es decir, a partir de un cierto tamaño muestral m , las desviaciones aleatorias $\left|F_{m+i}^*(x) - F(x)\right|$ son simultáneamente muy débiles.

⁹ En el sentido de pasar de convergencia en distribución a convergencia casi segura, que es una condición más fuerte.

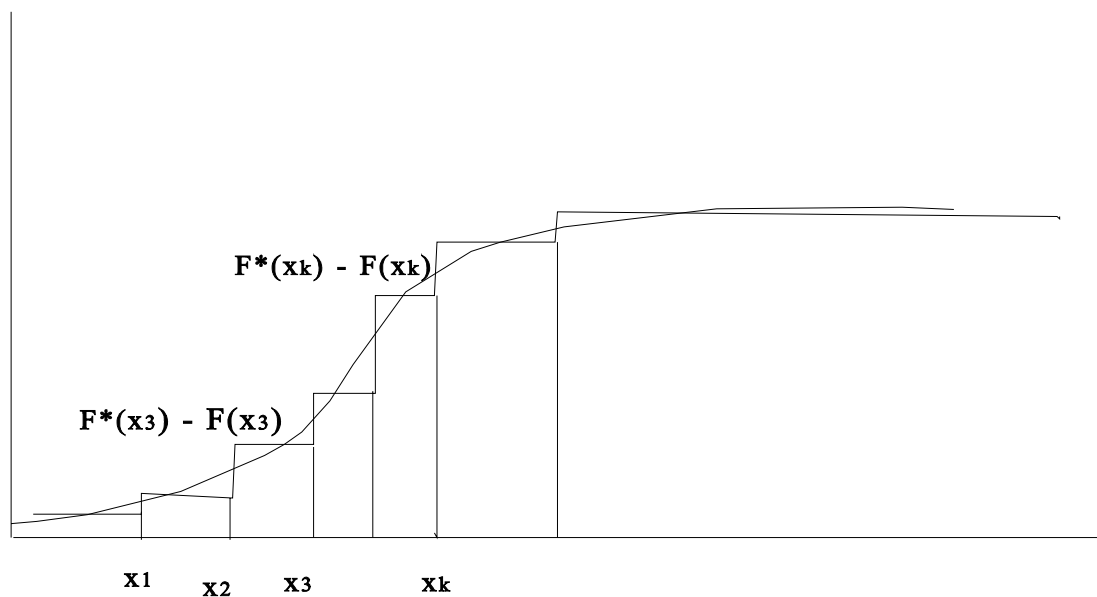
b) Teorema de Glivenko-Cantelli.

Aunque la consecuencia anterior de la ley fuerte de los grandes números es de gran interés ya que garantiza conocer, con la precisión que se desee, el valor de $F_{\xi}(x)$ en cada $x \in \mathcal{R}$ fijo, lo realmente interesante a nuestros efectos es estudiar dicho tipo de convergencias de la función de distribución empírica cuando se consideran todos los valores x a la vez. El principal resultado a estos efectos lo proporciona el Teorema Central de la Estadística o Teorema de Glivenko-Cantelli que demostrará que $F_n^*(x)$ converge a $F_{\xi}(x)$, uniformemente en x , con probabilidad 1.

Teorema de Glivenko-Cantelli:

Sea $\{X_i\}_{i=1}^{\infty}$ una sucesión de variables aleatorias independientes idénticamente distribuidas y con la misma distribución que la población de la cual proceden. Si:

- a) $F_n^*(x)$ es la función de distribución muestral asociada a la m.a.s. (X_1, X_2, \dots, X_n) .
- b) y $\Delta_n = \sup_{x \in \mathcal{R}} |F_n^*(x) - F_{\xi}(x)|$, habiendo uno para cada tamaño muestral,
- c) Entonces $P[\lim_{n \rightarrow \infty} (\Delta_n \geq \varepsilon)] = 0$ es decir $\Delta_n \xrightarrow{c.s.} 0$.



La explicación del Teorema es la siguiente:

- 1) Para cada tamaño muestral n se tienen un conjunto de diferencias $|F_n^*(x_i) - F(x_i)|$ que, a su vez, tienen multitud de cotas superiores de las cuales la más pequeña es Δ_n .
- 2) A medida que aumenta n disminuye Δ_n y existe un N tal que a partir de él ($\forall n > N$) todos los Δ_n tienden a cero. Es decir $\lim_{n \rightarrow \infty} \Delta_n = 0$ con probabilidad 1.

Es decir, imaginando una banda de amplitud ϵ , arbitrariamente estrecha, alrededor de la distribución teórica $F(x)$, el Teorema de Glivenko-Cantelli garantiza que hay probabilidad 1 (convergencia casi segura) de que la distribución muestral $F_n^*(x)$ llegue a estar contenida dentro de esa banda si se hace crecer suficientemente el tamaño muestral.

Ejercicio (Adaptado de Vélez Ibarrola y García Pérez)

Sea una m.a.s. de tamaño 40 de una distribución exponencial de media 3.

- a) Calcule la probabilidad de que los valores de la función de distribución muestral y teórica difieran en menos de 0,01 en el punto $x=1$.**
- b) Determine el valor del tamaño muestral para que dicha probabilidad sea aproximadamente 0,98.**

Solución:

- a) Sabemos que

$$f(x) = \frac{1}{3} e^{-\frac{x}{3}} \quad x \geq 0$$

y que

$$F(x) = \int_0^x \frac{1}{3} e^{-\frac{u}{3}} du = - \left[e^{-\frac{u}{3}} \right]_0^x = 1 - e^{-\frac{x}{3}} \quad x \geq 0$$

con lo que

$$F(1) = 1 - e^{-\frac{1}{3}} = 0,28341$$

En consecuencia, la solución a la pregunta propuesta es

$$P\left[\left|F_n^*(I) - F(I)\right| < 0,01\right] =$$

$$P\left[\left|F_n^*(I) - 0,28341\right| < 0,01\right] =$$

$$P\left[0,27341 < F_n^*(I) < 0,29341\right] =$$

$$P\left[40 - 0,27341 < 40 - F_n^*(I) < 40 - 0,29341\right] =$$

$$P\left[10,9364 < 40 - F_n^*(I) < 11,7364\right] \text{ y como } 40 - F_n^*(I) = B(40; 0,28347)$$

$$P\left[40 - F_n^*(I) = 11\right] = \binom{40}{11} 0,28347^{11} (1 - 0,28347)^{29} = 0,139$$

$$P\left[\left|F_n^*(I) - F(I)\right| < 0,01\right] = 0,98$$

$$P\left[\left|F_n^*(I) - 0,28341\right| < 0,01\right] = 0,98$$

$$P\left[0,27341 < F_n^*(I) < 0,29341\right] = 0,98$$

b)

$$F_n^*(x) \rightarrow_{\text{aprox}} N\left(F_\xi(x); \sqrt{\frac{F_\xi(x)(1 - F_\xi(x))}{n}}\right)$$

y como, para grandes tamaños muestrales
en nuestro caso

$$P\left[0,27341 < N\left(0,28341; \sqrt{\frac{0,28341 - 0,71653}{n}}\right) < 0,29341\right] = 0,98$$

con lo que

$$0,01 = P \left[N(0;1) < \frac{0,27341 - 0,28341}{\sqrt{\frac{0,28341 - 0,71653}{n}}} \right] = P \left[N(0;1) < -\frac{0,01\sqrt{n}}{0,4506} \right]$$

$$0,01 = P \left[N(0;1) > \frac{0,29341 - 0,28341}{\sqrt{\frac{0,28341 - 0,71653}{n}}} \right] = P \left[N(0;1) > \frac{0,01\sqrt{n}}{0,4506} \right]$$

de tal forma que

$$-2,33 = -\frac{0,01\sqrt{n}}{0,4506} ; \quad 2,33 = \frac{0,01\sqrt{n}}{0,4506}$$

de donde se tiene, finalmente, que $n = 10.934$.

Ejercicio (Adaptado de Vélez Ibarrola y García Pérez)

Sea una m.a.s. de tamaño 50 de una distribución de Poisson con $\lambda=3$.

- a) Calcule la probabilidad de que, en el punto $x=2$, $F_n^*(x)$ y $F(x)$ difieran en menos de 0,03.**
b) ¿Qué tamaño muestral hay que tomar para que dicha probabilidad sea aproximadamente 0,99?.

Solución:

a) Sabemos que $P(\xi = x_i) = e^{-\lambda} \frac{\lambda^{x_i}}{x_i!} \quad x_i = 0, 1, 2, 3, \dots,$

con lo que $F(2) = e^{-2} \left[\frac{2^0}{0!} + \frac{2^1}{1!} + \frac{2^2}{2!} \right] = 0,67667$

En consecuencia, la solución a la pregunta propuesta es

$$P \left[\left| F_n^*(2) - F(2) \right| < 0,03 \right] =$$

$$P \left[\left| F_{subn}^*(2) - 0,67667 \right| < 0,03 \right] =$$

$$P \left[0,64667 < F_n^*(2) < 0,70667 \right] =$$

$$P \left[50 - 0,64667 < 50 - F_n^*(2) < 50 - 0,70667 \right] =$$

$$P \left[32,3335 < 50 - F_n^*(2) < 35,3335 \right] \text{ y como } 50 - F_n^*(2) = B(50; 0,67667)$$

$$= \binom{50}{33} 0,67667^{33} (1 - 0,67667)^{17} + \binom{50}{34} 0,67667^{34} (1 - 0,67667)^{16} + \binom{50}{35} 0,67667^{35} (1 - 0,67667)^{15} = 0,3492$$

b)

$$P \left[\left| F_n^*(2) - F(2) \right| < 0,03 \right] = 0,99$$

$$P \left[0,64667 < F_n^*(2) < 0,70337 \right] = 0,99$$

$$F_n^*(x) \rightarrow_{\text{aprox}} N \left(F_\xi(x); \sqrt{\frac{F_\xi(x)(1 - F_\xi(x))}{n}} \right)$$

y como, para grandes tamaños muestrales
en nuestro caso

$$P \left[0,64667 < N \left(0,67667; \sqrt{\frac{0,67667 - 0,32333}{n}} \right) < 0,70667 \right] = 0,99$$

con lo que

$$0,005 = P \left[N(0;1) < \frac{0,64667 - 0,67667}{\sqrt{\frac{0,67667 - 0,32333}{n}}} \right] = P \left[N(0;1) < -\frac{0,03\sqrt{n}}{0,21879} \right]$$

$$0,005 = P \left[N(0;1) > \frac{0,70667 - 0,67667}{\sqrt{\frac{0,67667 - 0,32333}{n}}} \right] = P \left[N(0;1) > \frac{0,03\sqrt{n}}{0,21879} \right]$$

de tal forma que

$$-2,58 = -\frac{0,03\sqrt{n}}{0,4677} ; \quad 2,58 = \frac{0,03\sqrt{n}}{0,4677}$$

con lo que $n = 1.618$.

Demostración del Teorema de Glivenko-Cantelli. (Con dos Anexos)

- 1) Hemos visto que $F_n^*(x) \xrightarrow{c.s.} F_\xi(x)$ para cada $x \in \mathfrak{R}$ fijo.
- 2) De la misma manera, considerando las funciones indicadoras $I_{(-\infty, x)}(X_i)$ hubiésemos llegado a la conclusión de que $F_n^*(x^-) \xrightarrow{c.s.} F_\xi(x^-)$ para cada $x \in \mathfrak{R}$ fijo.
- 3) Para cada número natural k , y $j=1,2, \dots, k$, considérense los puntos

$$x_{j,k} = \min \left\{ x \in R \mid F_\xi(x^-) \leq \frac{j}{k} \leq F_\xi(x) \right\} = \min \left\{ x \in R \mid P(\xi < x) \leq \frac{j}{k} \leq P(\xi \leq x) \right\}$$

es decir, el mínimo cuantil de orden r/k de la población (ANEXO 1).

- 4) Considérense los siguientes sucesos:

$$A_{j,k} = \{F_n^*(x_{j,k}) \rightarrow F_\xi(x_{j,k})\}$$

$$B_{j,k} = \{F_n^*(\tilde{x}_{j,k}) \rightarrow F_\xi(\tilde{x}_{j,k})\}$$

$$D_k = \bigcap_{j=1}^k (A_{j,k} \cap B_{j,k})$$

$$D = \bigcap_{k=1}^{\infty} D_k$$

donde:

-El suceso D_k hace referencia a que, para un k fijo, todas las diferencias $|F_n^*(x_{j,k}) - F_\xi(x_{j,k})|$ y $|F_n^*(\tilde{x}_{j,k}) - F_\xi(\tilde{x}_{j,k})|$ converjan a cero (convergencia casi segura), al tender n a infinito.

- D es el suceso relativo a que ello ocurra simultáneamente para todos los valores de k .

Según la Ley fuerte de los grandes números, $P(A_{j,k}) = P(B_{j,k}) = 1$ para cualquier j y cualquier k , luego $P(D_k) = 1$ (ANEXO 2), para cualquier k y, por la misma razón, $P(D) = 1$.

Obsérvese ahora que si $x \in [x_{j,k}; x_{j+1,k})$ entonces, por ser la función de distribución monótona no decreciente, se tiene que

$$\begin{cases} F_\xi(x_{j,k}) \leq F_\xi(x) \leq F_\xi(\tilde{x}_{j+1,k}) \\ F_n^*(x_{j,k}) \leq F_n^*(x) \leq F_n^*(\tilde{x}_{j+1,k}) \end{cases}$$

Entonces:

$$F_n^*(x_{j,k}) - F(\tilde{x}_{j+1,k}) \leq F_n^*(x) - F(x) \leq F_n^*(\tilde{x}_{j+1,k}) - F(x_{j,k})$$

de donde se desprende que

$$\begin{cases} F_n^*(x) - F(x) \leq F_n^*(\tilde{x}_{j+l,k}) - F(x_{j,k}) \\ F_n^*(x) - F(x) \geq F_n^*(x_{j,k}) - F(\tilde{x}_{j+l,k}) \end{cases}$$

y como $F(\tilde{x}_{j+l,k}) - F(x_{j,k}) \leq \frac{l}{k}$ ¹⁰ entonces

$$\begin{cases} F_n^*(x) - F(x) \leq F_n^*(\tilde{x}_{j+l,k}) - F(x_{j,k}) \leq F_n^*(\tilde{x}_{j+l,k}) - F(\tilde{x}_{j+l,k}) + \frac{l}{k} \\ F_n^*(x) - F(x) \geq F_n^*(x_{j,k}) - F(\tilde{x}_{j+l,k}) \geq F_n^*(x_{j,k}) - F(x_{j,k}) - \frac{l}{k} \end{cases}$$

¹⁰ De la definición de $x_{j,k}$ sabemos que

$$\begin{cases} F(\tilde{x}_{j,k}) \leq \frac{j}{k} \leq F(x_{j,k}) \\ F(\tilde{x}_{j+l,k}) \leq \frac{j+l}{k} \leq F(x_{j+l,k}) \end{cases}$$

por lo cual $F(\tilde{x}_{j+l,k}) - F(x_{j,k}) \leq \frac{j+l}{k} - \frac{j}{k} \leq F(x_{j+l,k}) - F(x_{j,k})$, de donde se desprende

que $0 \leq F(\tilde{x}_{j+l,k}) - F(x_{j,k}) \leq \frac{l}{k}$, por lo que $F(x_{j,k}) \geq F(\tilde{x}_{j+l,k}) - \frac{l}{k}$

con lo cual, si para k y n fijos $\delta_n^{(k)}$ es la mayor entre todas las diferencias $|F_n^*(x_{j,k}) - F(x_{j,k})|$ y $|F_n^*(x_{j+l,k}) - F(x_{j+l,k})|$, se tiene

$$\begin{cases} F_n^*(x) - F(x) \leq \delta_n^{(k)} + \frac{1}{k} \\ F_n^*(x) - F(x) \geq -\delta_n^{(k)} - \frac{1}{k} \end{cases}$$

por lo que, en valor absoluto

$$|F_n^*(x) - F(x)| \leq \delta_n^{(k)} + \frac{1}{k}$$

Y como $\delta_n^{(k)} + \frac{1}{k}$ son cotas superiores, el supremo será menor o igual que todas ellas: $\Delta_n \leq \delta_n^{(k)} + \frac{1}{k}$.

Si se verifica la ocurrencia del suceso D , para cualquier $k \in \mathbb{N}$ y cualquier $\varepsilon > 0$, entonces

$\delta_n^{(k)} < \varepsilon$ a partir de un n en adelante, PORQUE:

- Si se verifica A_{jk} la diferencia $|F_n^*(x_{j,k}) - F(x_{j,k})| < \varepsilon_1$.
- Si se verifica B_{jk} la diferencia $|F_n^*(x_{j+l,k}) - F(x_{j+l,k})| < \varepsilon_2$.
- Si se verifica D_k es que se verifican A_{jk} y $B_{jk} \forall j$.
- Si se verifica D es que se verifican todos los A_{jk} y $B_{jk} \forall j$ y $\forall k$.

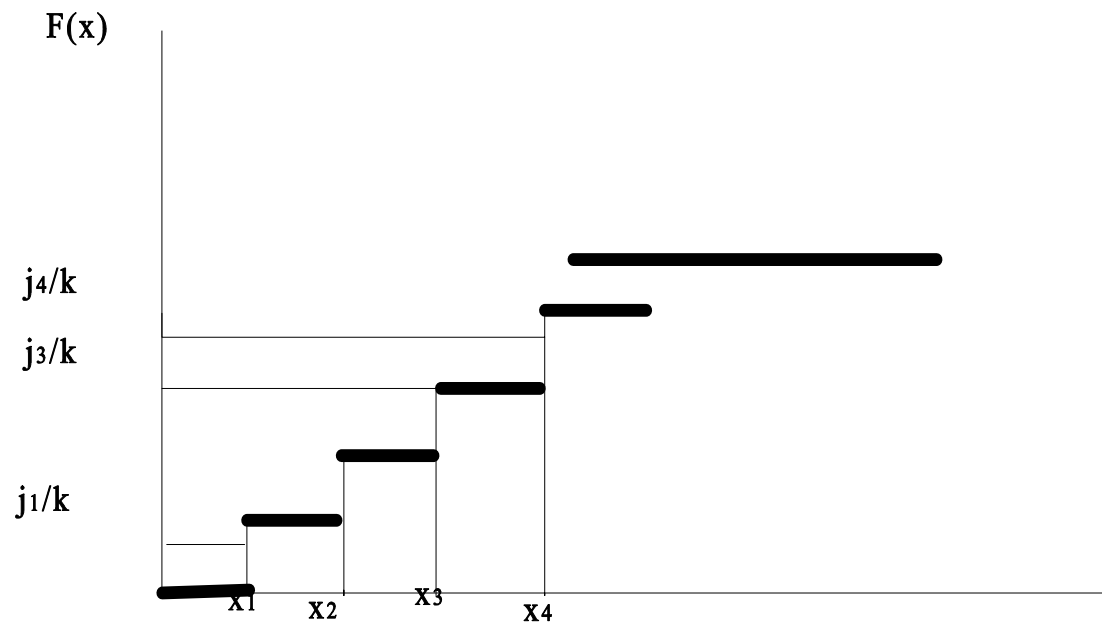
y como $\delta_n^{(k)}$ es el máximo de A_{jk} y $B_{jk} \forall j$ y $\forall k$ y A_{jk} y B_{jk} son inferiores a ε , entonces $\delta_n^{(k)} < \varepsilon$.

Entonces, a partir de un n en adelante, $\Delta_n \leq \delta_n^{(k)} + \frac{1}{k} \leq \varepsilon + \frac{1}{k}$. Y como k es arbitrario, entonces siempre que se verifique el suceso D se tiene que $\lim_{n \rightarrow \infty} \Delta_n = 0$. Y como hay probabilidad 1 de ocurrencia del suceso D también la habrá de que $\lim_{n \rightarrow \infty} \Delta_n = 0$.

ANEXO 1: Determinación de x_{jk} .

Distribuciones discretas:

- Si j/k coincide entre dos peldaños x_{jk} es el primero del peldaño superior.
- Si j/k coincide con un peldaño, todos los puntos del peldaño, más el inmediatamente siguiente, verifican la condición. El mínimo es el primero del peldaño.



Distribuciones continuas:

- Para cada j/k hay un sólo valor x_{jk} .

Puntos que verifican $P(\xi < x) \leq \frac{j_l}{k} \leq P(\xi \leq x)$: sólo x_1 .

Puntos que verifican $P(\xi < x) \leq \frac{j_3}{k} \leq P(\xi \leq x)$: $[x_3 - x_4]$. El mínimo es x_3 .

Puntos que verifican $P(\xi < x) \leq \frac{j_4}{k} \leq P(\xi \leq x)$: sólo x_4 .

ANEXO 2

Teorema Auxiliar: “Si cada uno de los sucesos de una secuencia finita o infinito numerable $E_1, E_2, \dots, E_n, \dots$ tiene probabilidad 1, entonces la probabilidad de que ocurran a la vez también es 1”

La demostración puede verse, por ejemplo, en Gnedenko, B.V. (1968): “The Theory of Probability”, Chelsea Publishing Company, New York, pp.446-448.

Demostración Anexo 2:

Consideremos primeramente dos sucesos E_1, E_2 tales que $P(E_1) = P(E_2) = 1$. Entonces:

$$1^{\circ} P(E_1 \cup E_2) = 1 \quad \text{ya que} \quad \begin{cases} E_1 \subset E_1 \cup E_2 \\ E_2 \subset E_1 \cup E_2 \end{cases} \quad \text{y} \quad P(E_1) = P(E_2) = 1$$

$$2^{\circ} P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2)$$

$$\text{lo que implica que } P(E_1 \cap E_2) = P(E_1) + P(E_2) - P(E_1 \cup E_2) = 1 + 1 - 1 = 1$$

Por inducción matemática se concluye que para cualesquiera n sucesos para los cuales $P(E_1) = P(E_2) = \dots = P(E_n) = 1$ se satisface que $P(E_1 \cap E_2 \cap \dots \cap E_n) = 1$.

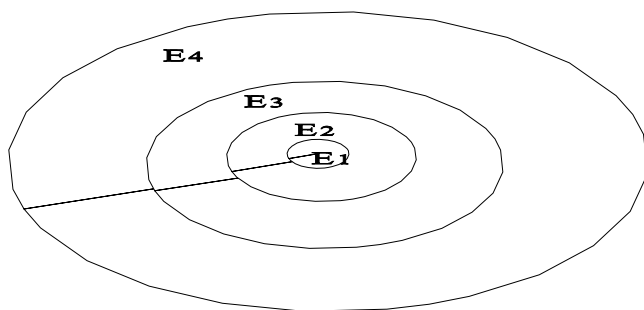
Sea ahora una sucesión infinita de sucesos $E_1, E_2, \dots, E_n, \dots$ tal que $P(E_1) = P(E_2) = \dots = P(E_n) = \dots = 1$.

Como

$$E_1 \cap E_2 \cap E_3 \cap \dots = E_1 \cap (E_1 \cap E_2) \cap (E_1 \cap E_2 \cap E_3) \dots$$

y cada suceso del lado derecho de la ecuación implica el anterior (en el sentido de que está incluido en él??), se tiene, por el Lema 1, que:

$$P(E_1 \cap E_2 \cap E_3 \cap \dots) = \lim_{n \rightarrow \infty} P(E_1 \cap E_2 \cap E_3 \cap \dots \cap E_n)$$



Lema 1:

Si un suceso E es equivalente a la ocurrencia simultánea de una colección infinita de sucesos E_1, E_2, \dots , es decir, $E = E_1 \cap E_2 \cap E_3 \cap \dots$, y si cada suceso E_{n+1} implica el

$$P(E) = \lim_{n \rightarrow \infty} P(E_n)$$

precedente E_n , ENTONCES:

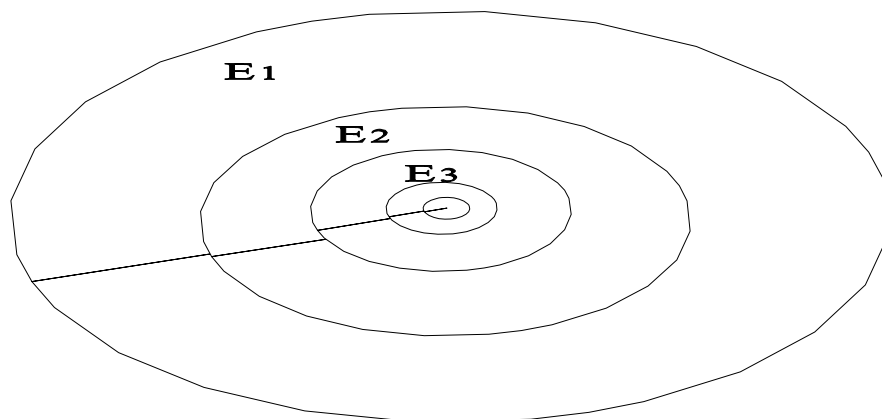
Demostración:

El suceso E_1 puede ser expresado, como unión de sucesos mutuamente excluyentes, de dos formas distintas:

$$1^a) E_1 = (E_1 \cap \bar{E}_2) \cup (E_2 \cap \bar{E}_3) \cup \dots \cup (E_1 \cap \bar{E}_2) \cup (E_{n-1} \cap \bar{E}_n) \cup E_n$$

$$2^a) E_1 = (E_1 \cap \bar{E}_2) \cup (E_2 \cap \bar{E}_3) \cup \dots \cup (E_1 \cap \bar{E}_2) \cup (E_{n-1} \cap \bar{E}_n) \cup \dots \cup E$$

Nota: como cada suceso implica la ocurrencia del anterior, imaginemos que E_1 incluye a E_2 y E_2 incluye a E_3 y así sucesivamente.



Entonces, en 1ª, cada suceso de la derecha es uno de los donuts y el último es el centro o agujero.

En 2ª ocurre lo mismo pero los donuts son infinitos por lo que E es el centro o agujero.

Y como

$$P(E_1) = P(E_1 \cap \overline{E_2}) + P(E_2 \cap \overline{E_3}) + \dots + P(E_1 \cap \overline{E_2}) + P(E_{n-1} \cap \overline{E_n}) + P(E_n)$$

y

$$P(E_1) = P(E_1 \cap \overline{E_2}) + P(E_2 \cap \overline{E_3}) + \dots + P(E_1 \cap \overline{E_2}) + P(E_{n-1} \cap \overline{E_n}) + \dots + P(E)$$

comparando ambas expresiones se tiene que

$$P(E) = P(E_n) - \sum_{k=n}^{\infty} P(E_k \cap \overline{E_{k+1}})$$

y como el sumatorio es el resto de una serie convergente, ENTONCES

$$P(E) = \lim_{n \rightarrow \infty} P(E_n)$$

Corolario:

Si $E_1 \subset E_2 \subset \dots$ y $\overline{E} = E_1 + E_2 + \dots$ ENTONCES

$$P(E) = \lim_{n \rightarrow \infty} P(E_n)$$

Demostración: Como

$$\overline{E_1} \supset \overline{E_2} \supset \dots \text{ y } \overline{E} = \overline{E_1} \cap \overline{E_2} \cap \dots$$

se sigue del lema 1 que

$$1 - P(E) = P(\overline{E}) = \lim_{n \rightarrow \infty} P(\overline{E_n}) = \lim_{n \rightarrow \infty} (1 - P(E_{subn})) = 1 - \lim_{n \rightarrow \infty} P(E_n)$$

En consecuencia,

$$P(E) = \lim_{n \rightarrow \infty} P(E_n)$$

TEMA 1. EPÍGRAFE 1.2.

Distribuciones en el muestreo asociadas a poblaciones normales. Distribuciones de la media, varianza y diferencia de medias.

Esquema de trabajo:

Cuestiones previas:

1. Importancia de la distribución normal o por qué un epígrafe aparte para las distribuciones en el muestreo asociadas a poblaciones normales.
2. La reproducción del modelo normal en las combinaciones lineales de variables normales o propiedad aditiva de la distribución normal.

Desarrollo del epígrafe:

Caso de una población.

1. Distribución de la media muestral aleatoria con varianza poblacional conocida.
2. Lema de Fisher-Cochran: Independencia de la media y la varianza muestrales aleatorias.
3. Distribución de la varianza muestral aleatoria.
4. Distribución de la media muestral aleatoria con varianza desconocida.

Caso de dos poblaciones:

5. Distribución de la diferencia de medias muestrales aleatorias (con varianzas poblacionales conocidas).
6. Generalización del lema de Fisher-Cochran.
7. Distribución de la diferencia de medias muestrales aleatorias (con varianzas poblacionales desconocidas).
8. Distribución del cociente de varianzas muestrales aleatorias.

Ejercicios complementarios.

Importancia de la distribución normal o por qué un epígrafe aparte para las distribuciones en el muestreo asociadas a poblaciones normales.

¿Por qué “Distribuciones en el muestreo asociadas a poblaciones normales”? ¿Por qué merecen un capítulo aparte la distribución de la media, la varianza, la diferencia de medias y, en su caso, el cociente de varianzas cuando la población de la que se extrae la muestra sigue una ley normal?

Para dar respuesta a esta pregunta, reproduciremos un par de párrafos del texto de Canavos, G.C. (1990): “Probabilidad y Estadística”, McGraw Hill, pp. 131 y 132:

“La distribución normal o Gaussiana es indudablemente la más importante y la de mayor uso de las distribuciones de probabilidad. Es la piedra angular de la inferencia estadística en el análisis de datos, puesto que las distribuciones de muchas estadísticas muestrales tienden hacia la distribución normal conforme crece el tamaño de la muestra..... Un gran número de estudios muestran que la distribución normal proporciona una adecuada representación, por lo menos en una primera aproximación, de las distribuciones de una gran cantidad de variables físicas. Algunos ejemplos específicos incluyen datos meteorológicos como la temperatura y la precipitación pluvial, mediciones efectuadas en organismos vivos, calificaciones en pruebas de actitud, mediciones físicas de partes manufacturadas, errores de instrumentación y otras desviaciones de las normas establecidas, etc”.

La reproducción del modelo normal en combinaciones lineales de variables normales ó propiedad aditiva de la distribución normal.

1) Sabemos que la función característica de una suma de variables aleatorias independientes coincide con el producto de las funciones características de dichas variables aleatorias.

2) Sabemos que $\varphi_{(c_{-}\xi)}(t) = \varphi_{\xi}(c_{-}t)$

3) En consecuencia para n variables muestrales independientes X_1, X_2, \dots, X_n

$$\varphi_{(a_1 X_1 + a_2 X_2 + \dots + a_n X_n)}(t) = \varphi_{X_1}(a_1 t) \varphi_{X_2}(a_2 t) \dots \varphi_{X_n}(a_n t)$$

4) En el caso en que la muestra (m.a.s.) proceda de una población $N(\mu; \sigma)$

$$\begin{aligned} \varphi_{(a_1 X_1 + a_2 X_2 + \dots + a_n X_n)}(t) &= \\ &= e^{i a_1 t \mu - \frac{1}{2} \sigma^2 a_1^2 t^2} \dots e^{i a_n t \mu - \frac{1}{2} \sigma^2 a_n^2 t^2} \\ &= e^{it(a_1 + a_2 + \dots + a_n)\mu - \frac{1}{2}(a_1^2 + a_2^2 + \dots + a_n^2)\sigma^2 t^2} \end{aligned}$$

ya que todas las variables muestrales, además de ser independientes, se distribuyen igual que la población de la cual proceden y, por tanto, todas ellas tienen media μ y desviación típica σ .

Como puede observarse, la función característica de una combinación lineal de variables muestrales (m.a.s.) procedentes de una población normal obedece a la función característica de una normal con media la media poblacional ponderada por la suma de los coeficientes a_i y con varianza la varianza poblacional ponderada por la suma de los cuadrados de dichos coeficientes. **Por tanto, si la muestra se toma de una población normal, los estadísticos que se formen como de combinaciones lineales de las variables muestrales tendrán**

1. Distribución Normal.

2. Con esperanza la esperanza poblacional multiplicada por la suma de los coeficientes de la combinación lineal.

3. Con varianza la varianza poblacional multiplicada por la suma de los cuadrados de los coeficientes de la combinación lineal.

A) Una población $[N(\mu; \sigma)]$.

1. Distribución de la media muestral aleatoria (con varianza poblacional conocida).

Sabemos que:

$$E(\bar{X}) = \mu \quad V(\bar{X}) = \frac{\sigma^2}{n}$$

sea cual sea la distribución de probabilidad de la población.

En nuestro caso la población es normal y el estadístico media muestral es una combinación lineal de variables normales por lo que

$$\bar{X} \sim N\left(\mu; \frac{\sigma}{\sqrt{n}}\right)$$

llegándose a la siguiente expresión pivotal

$$\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0; 1) = \xi^* \quad ; \quad \bar{X} = \mu + \frac{\sigma}{\sqrt{n}} \xi^*$$

expresión que, al relacionar las medias muestral y poblacional mediante una distribución de probabilidad conocida, nos permitirá llevar a cabo inferencias sobre un parámetro tan importante como la media poblacional en base a la media muestral si la varianza de la población es conocida.

No menos importante que la media poblacional es la varianza poblacional¹, por lo que se hace necesario el conocimiento de la distribución de probabilidad de la varianza muestral para

¹ El orden esperado en el desarrollo de este epígrafe, cuando de una población se trata, sería el siguiente:

1. Distribución de la media muestral aleatoria con varianza poblacional conocida.
2. Distribución de la media muestral aleatoria con varianza poblacional desconocida.
3. Distribución de la varianza muestral aleatoria (con media poblacional desconocida, caso general, o conocida, caso inusual).

Sin embargo,

- 1) La determinación de la distribución de la varianza muestral aleatoria (con media poblacional desconocida) exige la utilización del lema de Fisher-Cochran.
- 2) La determinación de la distribución de la media muestral aleatoria con varianza poblacional desconocida exige tanto la utilización del lema de Fisher-Cochran como el conocimiento de la distribución de la varianza muestral aleatoria (con media poblacional desconocida, lógicamente).

De lo expuesto se deduce el orden adoptado en el desarrollo de estas cuestiones en el caso de una población.

formular inferencias sobre ella. La media de la población puede ser conocida o desconocida; sin embargo, como es sumamente raro el primero de los casos adoptaremos el supuesto de desconocimiento de la misma. Bajo esta suposición, el conocimiento de la distribución en el muestreo de la varianza muestral aleatoria exige previamente el conocimiento del lema de Fisher-Cochran.

Lema de Fisher-Cochran o independencia de la media y la varianza muestral en poblaciones normales².

Teorema: Para una m.a.s de tamaño n procedente de una $N(\mu; \sigma)$ el estadístico \bar{X} y el vector $((X_1 - \bar{X}) \ (X_2 - \bar{X}) \ \dots \ (X_n - \bar{X}))$ se distribuyen independientemente.

Corolario: Si se extrae una m.a.s. de una población $N(\mu; \sigma)$, los estadísticos \bar{X} y S_x^2 se distribuyen independientemente.

Demostración del Teorema:

Sea $\phi_{\bar{X}(X_1 - \bar{X})(X_2 - \bar{X}) \dots (X_n - \bar{X})}(t_{S_1} t_{S_2} \dots t_{S_n})$ la función característica conjunta de $\bar{X} (X_1 - \bar{X}) (X_2 - \bar{X}) \dots (X_n - \bar{X})$.

Entonces,

² Otra demostración puede verse en Arnaiz, G. (1986): "Introducción a la Estadística Teórica", (4ª ed.) Lex Nova, págs 465 a 469.

$$\begin{aligned}
\phi_{\bar{X}(X_1-\bar{X})(X_2-\bar{X})\dots(X_n-\bar{X})}(t s_1 s_2 \dots s_n) &= E \left[e^{i t \bar{X} + i s_1 (X_1 - \bar{X}) + \dots + i s_n (X_n - \bar{X})} \right] \\
&= E \left[e^{i t \bar{X} + \sum_{j=1}^n i s_j (X_j - \bar{X})} \right] = E \left[e^{\sum_{j=1}^n \left[\frac{i t}{n} + i (s_j - \bar{s}) \right] X_j} \right] \\
&= E \left[e^{i \sum_{j=1}^n a_j X_j} \right]
\end{aligned}$$

donde $a_j = \frac{t}{n} + (s_j - \bar{s})$ son los coeficientes de una combinación lineal de variables muestrales normales tal que

$$\sum_{j=1}^n a_j = \sum_{j=1}^n \left(\frac{t}{n} + (s_j - \bar{s}) \right) = \frac{n t}{n} + \sum_{j=1}^n (s_j - \bar{s}) = t$$

$$\sum_{j=1}^n a_j^2 = \frac{n t^2}{n^2} + \frac{2 t}{n} \sum_{j=1}^n (s_j - \bar{s}) + \sum_{j=1}^n (s_j - \bar{s})^2 = \frac{t^2}{n} + \sum_{j=1}^n (s_j - \bar{s})^2$$

En consecuencia

donde el segundo factor, no es sino la función característica conjunta de $(X_1 - \bar{X}) (X_2 - \bar{X}) \dots (X_n - \bar{X})$ ya que si la función característica conjunta de dos variables se factoriza en el producto de una función de t y otra de s, entonces ambas variables son independientes. Además, si uno de estos factores es una función característica el otro también lo es (Lindgren B.W. (1993): “Statistical Theory”, 4ª ed., Chapman & Hall, p. 131).

En virtud de este teorema la media muestral aleatoria y el vector de diferencias se distribuyen independientemente y, dado que $e^{i t \mu - \frac{1}{2} \frac{\sigma^2}{n} t^2}$ es la función característica de la media muestral aleatoria cuando la m.a.s. se toma de una población normal, $e^{-\frac{1}{2} \sigma^2 \sum_{i=1}^n (s_i - \bar{s})^2}$ es la función característica n-dimensional del vector de diferencias. En consecuencia la media muestral aleatoria y la varianza muestral aleatoria se distribuyen independientemente.

Inciso: Función característica conjunta de $(X_1 - \bar{X}) \ (X_2 - \bar{X}) \ \dots\dots\dots (X_n - \bar{X})$.

$$\phi_{(X_1 - \bar{X})(X_2 - \bar{X}) \dots\dots (X_n - \bar{X})}(s_1 \text{ sub2 } \dots\dots s_n) = E \left[e^{i \sum_{j=1}^n s_j (X_j - \bar{X})} \right]$$

y como

$$\sum_{j=1}^n s_j (X_j - \bar{X}) = \sum_{j=1}^n (s_j - \bar{s})(X_j - \bar{X}) = \sum_{j=1}^n (s_j - \bar{s}) X_j$$

entonces

$$\phi_{(X_1 - \bar{X})(X_2 - \bar{X}) \dots\dots (X_n - \bar{X})}(s_1 \text{ sub2 } \dots\dots s_n) = E \left[e^{i \sum_{j=1}^n (s_j - \bar{s}) X_j} \right]$$

$$= E \left[e^{i(s_1 - \bar{s})X_1} _ e^{i(s_2 - \bar{s})X_2} _ \dots\dots _ e^{i(s_n - \bar{s})X_n} \right]$$

y como las variables muestrales son independientes

$$\phi_{(X_1 - \bar{X})(X_2 - \bar{X}) \dots\dots (X_n - \bar{X})}(s_1 \text{ sub2 } \dots\dots s_n) =$$

$$= E \left[e^{i(s_1 - \bar{s})X_1} \right] _ E \left[e^{i(s_2 - \bar{s})X_2} \right] _ \dots\dots _ E \left[e^{i(s_n - \bar{s})X_n} \right]$$

$$= \varphi_{X_1}(s_1 - \bar{s}) _ \varphi_{X_2}(s_2 - \bar{s}) _ \dots\dots _ \varphi_{X_n}(s_n - \bar{s})$$

$$= e^{i(s_1 - \bar{s})\mu - \frac{1}{2}\sigma^2(s_1 - \bar{s})^2} _ e^{i(s_2 - \bar{s})\mu - \frac{1}{2}\sigma^2(s_2 - \bar{s})^2} _ \dots\dots _ e^{i(s_n - \bar{s})\mu - \frac{1}{2}\sigma^2(s_n - \bar{s})^2}$$

$$= e^{-\frac{1}{2}\sigma^2 \sum_{i=1}^n (s_i - \bar{s})^2}$$

3. Distribución de la varianza muestral aleatoria (no se conoce la media poblacional).

Sabemos que:

$$S_x^2 = \frac{\sum_{i=1}^n (X_i - \mu)^2}{n} - (\bar{X} - \mu)^2 \rightarrow n S_x^2 = \sum_{i=1}^n (X_i - \mu)^2 - n(\bar{X} - \mu)^2 \rightarrow$$

1.

$$\frac{n S_x^2}{\sigma^2} = \sum_{i=1}^n \left(\frac{X_i - \mu}{\sigma} \right)^2 - \left(\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \right)^2$$

2. Si la población de la que se extrae la m.a.s. es $N(\mu; \sigma)$ entonces $X_i \rightarrow N(\mu; \sigma)$,

$\bar{X} \sim N\left(\mu; \frac{\sigma}{\sqrt{n}}\right)$ y $\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0; 1)$. En consecuencia,

$$\left(\frac{X_i - \mu}{\sigma} \right)^2 \sim \chi_1^2 \quad \sum_{i=1}^n \left(\frac{X_i - \mu}{\sigma} \right)^2 \sim \chi_n^2 \quad \left(\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \right)^2 \sim \chi_1^2$$

3. Por el Lema de Fisher-Cochran, \bar{X} y S_x^2 se distribuyen independientemente.

De 1.) se deduce que $\sum_{i=1}^n \left(\frac{X_i - \mu}{\sigma} \right)^2 = \left(\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \right)^2 + \frac{n S_x^2}{\sigma^2}$, y como \bar{X} y S_x^2 son

independientes:

$$\varphi_{\sum_{i=1}^n \left(\frac{X_i - \mu}{\sigma} \right)^2}(t) = \varphi_{\left(\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \right)^2}(t) \cdot \varphi_{\frac{n S_x^2}{\sigma^2}}(t)$$

y por tanto

$$\left(\frac{\frac{1}{2}}{\frac{1}{2} - it} \right)^{\frac{n}{2}} = \left(\frac{\frac{1}{2}}{\frac{1}{2} - it} \right)^{\frac{1}{2}} \cdot \varphi_{\frac{n S_x^2}{\sigma^2}}(t)$$

con lo que

$$\varphi_{\frac{n S_x^2}{\sigma^2}}(t) = \left(\frac{\frac{1}{2}}{\frac{1}{2} - it} \right)^{\frac{n-1}{2}}$$

que no es sino la función característica de una ji-cuadrado con n-1 grados de libertad, por lo que,

dada la unicidad de las funciones características se puede concluir que

$$\frac{n S_x^2}{\sigma^2} = \chi_{n-1}^2 \rightarrow S_x^2 = \frac{\sigma^2}{n} \chi_{n-1}^2$$

Ya disponemos por tanto de una expresión (expresión pivotal) que liga la varianza poblacional con la varianza muestral a través de una distribución conocida y tabulada. Esta expresión será de indudable importancia a la hora de realizar inferencias acerca de la varianza de una población normal con media desconocida sobre la base de la varianza de una m.a.s.³

Corolario:

Como la esperanza de una chi cuadrado son sus grados de libertad y la varianza el doble de sus grados de libertad, entonces la esperanza y la varianza de la varianza muestral aleatoria son, para m.a.s. procedentes de una población normal

$$E[S_x^2] = E\left[\frac{\sigma^2}{n} \chi_{n-1}^2\right] = \frac{\sigma^2}{n} (n-1) = \frac{n-1}{n} \sigma^2$$

$$V[S_x^2] = V\left[\frac{\sigma^2}{n} \chi_{n-1}^2\right] = \frac{\sigma^4}{n^2} 2(n-1) = \frac{n-1}{n^2} 2 \sigma^4$$

Por otra parte, sabíamos que, fuese cual fuese la distribución de probabilidad de la

³ Si μ fuese conocida podríamos realizar inferencias sobre σ^2 en base a la expresión $\sum_{i=1}^n \frac{(X_i - \mu)^2}{n}$. Y como $\sum_{i=1}^n \frac{(X_i - \mu)^2}{\sigma^2} = \chi_n^2$, entonces $\sum_{i=1}^n \frac{(X_i - \mu)^2}{n} = \frac{\sigma^2}{n} \chi_n^2$.

población, $E(S_x^2) = \frac{n-1}{n} \sigma^2$ y $V[S_x^2] = \frac{\mu_4 - \sigma^4}{n} - 2 \frac{\mu_4 - 2\sigma^4}{n^2} + \frac{\mu_4 - 3\sigma^4}{n^3}$. Pero en el caso

normal, como $\mu_4 = 3\sigma^4$, se tiene que $V[S_x^2] = \frac{n-1}{n^2} 2\sigma^4$.

4. Distribución de la media muestral aleatoria (varianza poblacional desconocida).

Pasamos a continuación a desarrollar la distribución de la media muestral cuando la m.a.s. procede de una población normal con varianza desconocida. Dicha distribución será de utilidad para realizar inferencias sobre la media poblacional (lógicamente también desconocida), en base a la media muestral, en una tesitura en la que se desconoce la varianza de la población.

Sabemos (de 1) que $\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} = N(0; 1) = \xi^*$.

Sin embargo, esta expresión pivotal no resulta de utilidad para realizar inferencias sobre μ en caso de que la varianza poblacional sea desconocida (caso, por otra parte, muy frecuente). En consecuencia, tendremos que arbitrar algún procedimiento que la elimine, de tal forma que tras dicha eliminación se conozca la distribución de probabilidad de la expresión resultante.

La eliminación de σ se lleva a cabo dividiendo la expresión anterior por

$\sqrt{\frac{1}{n-1} \frac{n S_x^2}{\sigma^2}}$ donde, como es sabido, $\frac{n S_x^2}{\sigma^2} \sim \chi_{n-1}^2$.

Entonces se tiene que, dado que la media y la varianza muestrales se distribuyen independientemente (lema de Fisher-Cochran),

$$\frac{\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}}{\sqrt{\frac{1}{n-1} \frac{n S_x^2}{\sigma^2}}} = \frac{\bar{X} - \mu}{\frac{S_x}{\sqrt{n-1}}} = \frac{(\bar{X} - \mu) \sqrt{n-1}}{S_x} = t_{n-1} \quad ; \quad \bar{X} = \mu + t_{n-1} \frac{S_x}{\sqrt{n-1}}$$

expresión pivotal que relaciona la media muestral y la media poblacional sin necesidad de conocer la varianza de la población y que permitirá inferencias sobre μ en base a \bar{X} sin conocer σ^2 .

B) Dos Poblaciones [$N(\mu_1; \sigma_1)$ y $N(\mu_2; \sigma_2)$]

5. Distribución de la diferencia de medias muestrales aleatorias (con varianzas poblacionales conocidas).

Si se tiene interés en la diferencia de dos medias poblacionales un enfoque viable es formular la inferencia en base a la diferencia entre las medias procedentes de dos m.a.s. (una de cada población).

Sean dos poblaciones en las cuales nos interesamos por una variable aleatoria, denominada ξ_1 en la primera población y ξ_2 en la segunda, tal que

$$\xi_1 \sim N(\mu_1; \sigma_1)$$

$$\xi_2 \sim N(\mu_2; \sigma_2)$$

De la primera se extrae una m.a.s. de tamaño n ($X_1; X_2; \dots; X_n$) y de la segunda otra de tamaño m ($Y_1; Y_2; \dots; Y_m$), muestras independientes.

Entonces se tiene que

$$\bar{X} \sim N\left(\mu_1; \frac{\sigma_1}{\sqrt{n}}\right) \quad \bar{Y} \sim N\left(\mu_2; \frac{\sigma_2}{\sqrt{m}}\right)$$

y, como las combinaciones lineales de variables normales presentan distribución normal,

$$(\bar{X} - \bar{Y}) \sim N\left((\mu_1 - \mu_2); \sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}\right)$$

teniéndose la siguiente expresión pivotal:

$$\frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}} = \xi^* \quad (\bar{X} - \bar{Y}) = (\mu_1 - \mu_2) + \xi^* \sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}$$

de utilidad para establecer inferencias sobre la diferencia entre las medias de dos poblaciones normales en base a la diferencia entre las medias de las muestras tomadas de ellas, siempre y cuando se conozcan las varianzas poblacionales.

En el caso particular de que las dos poblaciones tengan la misma varianza, la expresión anterior se particulariza en:

$$\frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sigma \sqrt{\frac{1}{n} + \frac{1}{m}}} = \xi^* \quad (\bar{X} - \bar{Y}) = (\mu_1 - \mu_2) + \xi^* \sigma \sqrt{\frac{1}{n} + \frac{1}{m}}$$

6. Generalización del lema de Fisher-Cochran.

Sabemos que $\frac{n S_x^2}{\sigma_1^2} \sim \chi_{n-1}^2$ y $\frac{m S_y^2}{\sigma_2^2} \sim \chi_{m-1}^2$ y como las muestras se toman de forma independiente, las varianzas muestrales se distribuyen independientemente y, por tanto, $\frac{n S_x^2}{\sigma_1^2} + \frac{m S_y^2}{\sigma_2^2} \sim \chi_{n+m-2}^2$ puesto que el modelo chi-cuadrado es reproductivo respecto de los grados de libertad.

Además, $\frac{n S_x^2}{\sigma_1^2} + \frac{m S_y^2}{\sigma_2^2}$ es independiente de \bar{X} y de \bar{Y} y, por consiguiente, de la diferencia de ambas $(\bar{X} - \bar{Y})$.

7. Distribución de la diferencia de medias muestrales aleatorias (con varianzas poblacionales desconocidas pero iguales).

La tesitura en la que se conoce el valor de las varianzas de las dos poblaciones es ciertamente rara, siendo lo normal que éstas sean desconocidas.

En el caso en que las varianzas poblacionales sean desconocidas la expresión pivotal

$$\frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}} = \xi^*$$

no resulta de utilidad para la realización de inferencias acerca de la diferencia entre las medias poblacionales, siendo necesaria una expresión con distribución de probabilidad conocida que no dependa de las varianzas poblacionales.

¿Como eliminar las varianzas poblacionales?. Ello sólo es posible si ambas son iguales.

En ese caso se tiene que:

$$a) \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sigma \sqrt{\frac{1}{n} + \frac{1}{m}}} = \xi^*$$

$$b) \frac{n S_x^2 + m S_y^2}{\sigma^2} \sim \chi_{n+m-2}^2$$

c) Por la generalización del teorema de Fisher-Cochran $n S_x^2 + m S_y^2$ se distribuye independientemente de $(\bar{X} - \bar{Y})$.

d) En consecuencia,

$$\frac{\frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sigma \sqrt{\frac{1}{n} + \frac{1}{m}}}}{\sqrt{\frac{1}{n+m-2} \frac{n S_x^2 + m S_y^2}{\sigma^2}}} = \frac{\xi^*}{\sqrt{\frac{1}{n+m-2} \chi_{n+m-2}^2}} = t_{n+m-2}$$

Simplificando:

$$\frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{1}{n} + \frac{1}{m}}} = \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{1}{n+m-2} (n S_x^2 + m S_y^2)}} = \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{n S_x^2 + m S_y^2}} \frac{\sqrt{n+m-2}}{\sqrt{\frac{1}{n} + \frac{1}{m}}} = t_{n+m-2}$$

expresión pivotal que relaciona la diferencia de medias muestrales con la diferencia de medias poblacionales sin necesitar del conocimiento de la varianza poblacional (recuérdese que es la misma en ambas poblaciones).

La expresión anterior también se suele escribir como

$$\frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{n S_x^2 + m S_y^2}{n+m-2}}} \frac{1}{\sqrt{\frac{1}{n} + \frac{1}{m}}} = \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{S_p \sqrt{\frac{1}{n} + \frac{1}{m}}} = t_{n+m-2}$$

donde S_p^2 recibe el nombre de estimador combinado (pooled) de la varianza común σ^2 . Nótese que el estimador combinado es el promedio ponderado de las dos cuasivarianzas muestrales, siendo los ponderadores los grados de libertad.

$$S_p^2 = \frac{n S_x^2 + m S_y^2}{n+m-2} = \frac{(n-1) S_x^{*2} + (m-1) S_y^{*2}}{n+m-2}$$

Llegados a este punto la pregunta natural es la siguiente: ¿Cuál es la distribución de la diferencia de medias muestrales si las varianzas poblacionales son desconocidas y distintas?.

La situación descrita se conoce como el problema de **Behrens-Fisher** que sobrepasa nuestro ámbito. No obstante, se han propuesto algunas aproximaciones.

Hoel, P.G. (1976): Introducción a la Estadística Matemática, Ariel., p. 280, propone estimar las varianzas poblacionales a través de las cuasivarianzas muestrales y entonces:

A) Si los tamaños de cada muestra son grandes (digamos que mayores que 30) entonces las cuasivarianzas muestrales son muy buenos estimadores de las varianzas poblacionales, por lo que

$$\frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_x^{*2}}{n} + \frac{S_y^{*2}}{m}}} \underset{\text{—aprox}}{\xi^*}$$

B) Si las muestras son pequeñas, la expresión anterior se aproximará por una t de Student con v grados de libertad, tomando por valor de v el entero más próximo. (Solución de Welch, la más popular)

8. Distribución del cociente de varianzas muestrales.

Sabemos que

$$\frac{n S_x^2}{\sigma_1^2} = \chi_{n-1}^2 \quad \frac{m S_y^2}{\sigma_2^2} = \chi_{m-1}^2$$

$$S_x^2 = \frac{\sigma_1^2 \chi_{n-1}^2}{n} \quad S_y^2 = \frac{\sigma_2^2 \chi_{m-1}^2}{m}$$

por lo cual

ambos independientes. Por tanto,

$$\frac{S_x^2}{S_y^2} = \frac{\frac{1}{n} \sigma_1^2 \chi_{n-1}^2}{\frac{1}{m} \sigma_2^2 \chi_{m-1}^2} = \frac{\frac{n-1}{n} \sigma_1^2 \frac{1}{n-1} \chi_{n-1}^2}{\frac{m-1}{m} \sigma_2^2 \frac{1}{m-1} \chi_{m-1}^2} = \frac{\frac{n-1}{n} \sigma_1^2}{\frac{m-1}{m} \sigma_2^2} F_{n-1; m-1}$$

o bien

$$\frac{S_x^{*2}}{S_y^{*2}} = \frac{\sigma_1^2}{\sigma_2^2} F_{n-1; m-1}$$

En caso de conocerse las medias poblacionales μ_1 y μ_2 podíamos haber

utilizado $\sum_{i=1}^n \frac{(X_i - \mu_1)^2}{n}$ en vez de $\sum_{i=1}^n \frac{(X_i - \bar{X})^2}{n}$ y $\sum_{i=1}^m \frac{(Y_i - \mu_2)^2}{m}$ en vez de $\sum_{i=1}^m \frac{(Y_i - \bar{Y})^2}{m}$.

y como

$$\sum_{i=1}^n \frac{(X_i - \mu_1)^2}{\sigma_1^2} = \chi_n^2$$

$$\sum_{i=1}^m \frac{(Y_i - \mu_2)^2}{\sigma_2^2} = \chi_m^2$$

y, además, se distribuyen independientemente, entonces

$$\frac{\frac{1}{n} \sum_{i=1}^n \frac{(X_i - \mu_1)^2}{\sigma_1^2}}{\frac{1}{m} \sum_{i=1}^m \frac{(Y_i - \mu_2)^2}{\sigma_2^2}} = F_{n;m} \quad ; \quad \frac{\sum_{i=1}^n \frac{(X_i - \mu_1)^2}{n}}{\sum_{i=1}^m \frac{(Y_i - \mu_2)^2}{m}} = \frac{\sigma_1^2}{\sigma_2^2} F_{n;m}$$

Ejercicios complementarios del epígrafe 1.2.

Ejercicio: Sea una muestra aleatoria simple de tamaño 10 de una población $N(\mu; 2)$. Determine:

- a) Probabilidad de que la media muestral y la poblacional difieran en más de 0,5.
 b) El tamaño muestral necesario para que, con una probabilidad de 0,9, las medias muestral y poblacional difieran en menos de 0,1.

Solución:

a)

$$\begin{aligned}
 P\left[|\bar{X} - \mu| > 0,5\right] &= P\left[\frac{|\bar{X} - \mu|}{\frac{\sigma}{\sqrt{n}}} > \frac{0,5}{\frac{\sigma}{\sqrt{n}}}\right] = P\left[|\xi^*| > \frac{0,5}{\frac{\sigma}{\sqrt{n}}}\right] \\
 &= P\left[|\xi^*| > \frac{0,5}{\frac{2}{\sqrt{10}}}\right] = P\left[|\xi^*| > 0,79\right] = 2 P\left[\xi^* > 0,79\right] = 0,4296
 \end{aligned}$$

b)

$$\begin{aligned}
 P\left[|\bar{X} - \mu| < 0,1\right] &= P\left[\frac{|\bar{X} - \mu|}{\frac{\sigma}{\sqrt{n}}} < \frac{0,1}{\frac{\sigma}{\sqrt{n}}}\right] = P\left[|\xi^*| < \frac{0,1}{\frac{\sigma}{\sqrt{n}}}\right] \\
 &= P\left[|\xi^*| < \frac{0,1}{\frac{2}{\sqrt{n}}}\right] = P\left[|\xi^*| < 0,05 \sqrt{n}\right] = 1 - 2 P\left[\xi^* > 0,05 \sqrt{n}\right]
 \end{aligned}$$

y como dicha probabilidad tiene que ser 0,9, se tiene que

$$P\left[\left|\bar{X} - \mu\right| < 0,1\right] = 1 - 2 P\left[\xi^* > 0,05 \sqrt{n}\right] = 0,9$$

$$P\left[\xi^* > 0,05 \sqrt{n}\right] = 0,05$$

$$0,05 \sqrt{n} = 1,645$$

$$n = 1.082$$

Ejercicio: Sea una muestra aleatoria simple tomada de una $N(\mu; \sigma)$ con μ conocida y σ desconocida. Compare las distribuciones en el muestreo, esperanza y varianza de los estadísticos

$$T_1(X) = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1} \quad T_2(X) = \frac{\sum_{i=1}^n (X_i - \mu)^2}{n}$$

Solución:

Se sabe que

$$\frac{n S_x^2}{\sigma^2} = \chi_{n-1}^2$$

$$\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\sigma^2} = \chi_{n-1}^2$$

$$\sum_{i=1}^n (X_i - \bar{X})^2 = \sigma^2 \chi_{n-1}^2$$

$$T_1(X) = S_x^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1} = \frac{\sigma^2 \chi_{n-1}^2}{n-1}$$

Por otro lado

$$\frac{\sum_{i=1}^n (X_i - \mu)^2}{\sigma^2} = \chi_n^2$$

$$\sum_{i=1}^n (X_i - \mu)^2 = \sigma^2 \chi_n^2$$

$$T_2(X) = \frac{\sum_{i=1}^n (X_i - \mu)^2}{n} = \frac{\sigma^2 \chi_n^2}{n}$$

En consecuencia,

$$E[T_1(X)] = E\left[\frac{\sigma^2 \chi_{n-1}^2}{n-1}\right] = \frac{\sigma^2}{n-1} E[\chi_{n-1}^2] = \frac{\sigma^2}{n-1} (n-1) = \sigma^2$$

$$E[T_2(X)] = E\left[\frac{\sigma^2 \chi_n^2}{n}\right] = \frac{\sigma^2}{n} E[\chi_n^2] = \frac{\sigma^2}{n} n = \sigma^2$$

y

$$V[T_1(X)] = V\left[\frac{\sigma^2 \chi_{n-1}^2}{n-1}\right] = \frac{\sigma^4}{(n-1)^2} V[\chi_{n-1}^2] = \frac{\sigma^4}{(n-1)^2} 2(n-1) = \frac{2\sigma^4}{n-1}$$

$$V[T_2(X)] = V\left[\frac{\sigma^2 \chi_n^2}{n}\right] = \frac{\sigma^4}{n^2} V[\chi_n^2] = \frac{\sigma^4}{n^2} 2n = \frac{2\sigma^4}{n}$$

En consecuencia, el valor esperado de ambos estimadores es el mismo, pero la variabilidad del segundo en torno a la varianza poblacional es menor que la del primero (sobre todo para muestras de escaso tamaño).

Ejercicio: Sea X una variable aleatoria con distribución $N(\mu_1; \sigma_1)$ siendo μ_1 conocida y σ_1 desconocida. Sea Y otra variable aleatoria, independiente de X , con distribución $N(\mu_2; \sigma_2)$ siendo desconocidos sus dos parámetros. Determine un estadístico razonable para obtener información acerca del cociente de varianzas poblacionales en base a dos muestras de tamaños

n_1 y n_2 tomadas de X e Y , respectivamente, así como su distribución en el muestreo.

Solución:

Sabemos que

$$\frac{\sum_{i=1}^{n_1} (X_i - \mu)^2}{\sigma_1^2} = \chi_{n_1}^2 \quad y \quad que \quad \frac{n_2 S_y^2}{\sigma_2^2} = \frac{\sum_{i=1}^{n_2} (Y_i - \bar{Y})^2}{\sigma_2^2} = \chi_{n_2-1}^2$$

Por tanto, como ambos estadísticos se distribuyen independientemente,

$$\frac{\frac{1}{n_1} \sum_{i=1}^{n_1} (X_i - \mu)^2}{\sigma_1^2} = \frac{\sigma_2^2}{\sigma_1^2} \frac{\sum_{i=1}^{n_1} (X_i - \mu)^2}{\sum_{i=1}^{n_2} (Y_i - \bar{Y})^2} \frac{n_2 - 1}{n_1} = F_{n_1; n_2-1}$$

es decir,

$$\frac{\frac{\sum_{i=1}^{n_1} (X_i - \mu)^2}{n_1}}{\frac{\sum_{i=1}^{n_2} (Y_i - \overline{Y})^2}{n_2 - 1}} = \frac{\sigma_1^2}{\sigma_2^2} F_{n_1; n_2-1}$$

con lo que se tiene el estadístico y su distribución de probabilidad en el muestreo.

Nota: Tengase en cuenta que la esperanza y varianza de una F de Snedecor con v_1 grados de libertad en el numerador y v_2 en el denominador es:

$$E(F_{v_1; v_2}) = \frac{v_2}{v_2 - 2} \quad v_2 > 2$$

$$V(F_{v_1; v_2}) = \frac{v_2^2 (2v_2 + 2v_1 - 4)}{v_1 (v_2 - 2)^2 (v_2 - 4)} \quad v_2 > 4$$

con lo que si en la población X hubiésemos utilizado la media muestral en vez de la poblacional,

aunque la esperanza del estimador hubiese sido la misma, la varianza hubiese sido mayor.

TEMA 1. EPÍGRAFE 1.3.

Estadísticos ordenados. Distribución del menor y mayor valor. Distribución del recorrido.

Esquema de trabajo:

A) INTRODUCCIÓN:

1. Variables muestrales y Estadísticos ordenados.

B) CASO DISCRETO:

1. Distribución del i -ésimo valor de la muestra.
2. Particularizaciones al menor y mayor valor de la muestra.
3. Distribución conjunta del i -ésimo y j -ésimo valor de la muestra.
4. Particularización a la distribución conjunta del menor y mayor valor de la muestra.
5. Distribución del recorrido.

C) CASO CONTINUO:

1. Distribución del i -ésimo valor de la muestra.
2. Particularizaciones al menor y mayor valor de la muestra.
3. Distribución conjunta del i -ésimo y j -ésimo valor de la muestra.
4. Particularización a la distribución conjunta del menor y mayor valor de la muestra.
5. Distribución del recorrido.

D) Ejercicios prácticos.

E) Apéndice: Función de distribución de los estadísticos ordenados.

A) INTRODUCCIÓN:

1. Variables Muestrales y Estadísticos Ordenados.

Sea una población en la que se investiga la variable aleatoria ξ . Si de esta población se toma una muestra aleatoria se tienen las variables muestrales X_1, X_2, \dots, X_n . Si las variables muestrales se ordenan por orden de magnitud, se obtiene una muestra aleatoria ordenada U_1, U_2, \dots, U_n , tal que $U_1 \leq U_2 \leq \dots \leq U_n$.

U_i es una variable aleatoria que representa el i -ésimo valor de la muestra. Dicho i -ésimo valor de la muestra puede provenir de la primera variable muestral, la segunda, ..., o la n -ésima. Evidentemente, U_1, U_2, \dots, U_n son variables aleatorias no independientes, y aunque X_1, X_2, \dots, X_n estén igualmente distribuidas (caso, entre otros, de m.a.s.) U_1, U_2, \dots, U_n no lo están.

Ejemplo:

	X_1	X_2	X_3	X_4	X_5
muestra 1	10	6	11	3	25
muestra 2	4	8	2	22	10
muestra 3	30	8	15	11	9
muestra 4	21	7	11	4	8
	U_1	U_2	U_3	U_4	U_5
muestra 1	3	6	10	11	25
muestra 2	2	4	8	10	22
muestra 3	8	9	11	15	30
muestra 4	4	7	8	11	21

Así, en la muestra 1

$$X_1 = 10, X_2 = 6, X_3 = 11, X_4 = 3, X_5 = 25.$$

Sin embargo,

$$U_1 = 3, U_2 = 6, U_3 = 10, U_4 = 11, U_5 = 25$$

La distribución de probabilidad de estas nuevas variables, los estadísticos ordenados, es, en ocasiones, de considerable interés: el máximo o mínimo de la cotización de un determinado activo bursátil, el valor máximo o mínimo de la temperatura de un lugar, etc.

B) CASO DISCRETO:

1. Distribución del i -ésimo valor de la muestra.

Sea una población discreta de la cual se extraen muestras¹ de tamaño n , cada una de ellas con una determinada probabilidad (distribución de probabilidad de la muestra). Pues bien, la función de cuantía del i -ésimo valor de la muestra, U_i , estará formada por los valores de la población acompañados de la probabilidad de obtención de una muestra en la que U_i sea dicho valor.

2. Particularizaciones: Menor y mayor valor de la muestra.

Evidentemente, en el caso en que $i=1$ se estará obteniendo la distribución de probabilidad del menor valor de la muestra. En caso de que $i=n$, la distribución de probabilidad así obtenida será la del mayor valor de la muestra.

¹ Se opta por este encabezamiento por ser más amplio que el de m.a.s.

Ejemplo:

Sea una población formada por los elementos $\{1, 2, 3, 4\}$ de la cual se extraen muestras de tamaño 2 con reemplazamiento. Determinése la distribución de probabilidad del primer valor de la muestra.

La distribución de probabilidad de la muestra es la siguiente:

$\{X_1=x_1; X_2=x_2\}$	$P \{X_1=x_1; X_2=x_2\}$	$\{X_1=x_1; X_2=x_2\}$	$P \{X_1=x_1; X_2=x_2\}$
$\{1; 1\}$	1/16	$\{3; 1\}$	1/16
$\{1; 2\}$	1/16	$\{3; 2\}$	1/16
$\{1; 3\}$	1/16	$\{3; 3\}$	1/16
$\{1; 4\}$	1/16	$\{3; 4\}$	1/16
$\{2; 1\}$	1/16	$\{4; 1\}$	1/16
$\{2; 2\}$	1/16	$\{4; 2\}$	1/16
$\{2; 3\}$	1/16	$\{4; 3\}$	1/16
$\{2; 4\}$	1/16	$\{4; 4\}$	1/16

siendo la función de cuantía de U_1

$U_1 = u_1$	$P (U_1 = u_1)$
1	$P[\{1;1\} \cup \{1;2\} \cup \{1;3\} \cup \{1;4\} \cup \{2;1\} \cup \{3;1\} \cup \{4;1\}] = 7/16$
2	$P [\{2;2\} \cup \{2;3\} \cup \{2;4\} \cup \{3;2\} \cup \{4;2\}] = 5/16$
3	$P [\{3;3\} \cup \{3;4\} \cup \{4;3\}] = 3/16$
4	$P [\{4;4\}] = 1/16$

3. Distribución conjunta del i-ésimo y j-ésimo valor de la muestra.

La función de cuantía conjunta del i-ésimo y j-ésimo valor de la muestra estará formada por los pares de valores que tomen ambos estadísticos y sus probabilidades, que no son otras que las probabilidades de extracción de una muestra en la que ambos valores constituyan el i-ésimo y el j-ésimo valor de la muestra.

En el ejemplo anterior la distribución conjunta de los estadísticos U_1 y U_2 vendrá dada por

U_2					
4	$\frac{2}{16}$ $P[\{1;4\} \cup \{4;1\}]$	$\frac{2}{16}$ $P[\{2;4\} \cup \{4;2\}]$	$\frac{2}{16}$ $P[\{3;4\} \cup \{4;3\}]$	$\frac{1}{16}$ $P[\{4;4\}]$	U_1
3	$\frac{2}{16}$ $P[\{1;3\} \cup \{3;1\}]$	$\frac{2}{16}$ $P[\{2;3\} \cup \{3;2\}]$	$\frac{1}{16}$ $P[\{3;3\}]$	0	
2	$\frac{2}{16}$ $P[\{1;2\} \cup \{2;1\}]$	$\frac{1}{16}$ $P[\{2;2\}]$	0	0	
1	$\frac{1}{16}$ $P[\{1;1\}]$	0	0	0	
	1	2	3	4	

4. Particularización al menor y mayor valor de la muestra.

Si $i=1$ y $j=n$ entonces se tiene la distribución conjunta del menor y mayor valor de la muestra.

5. Distribución del Recorrido Muestral.

Se denomina recorrido muestral R a la diferencia entre el mayor y el menor valor de la muestra. En consecuencia $R = U_n - U_1$. La distribución de probabilidad de la variable R estará formada por todas las posibles diferencias no negativas entre U_n y U_1 acompañadas de sus probabilidades de ocurrencia (suma de las probabilidades de las muestras que dan lugar a dichas diferencias). En el ejemplo que venimos arrastrando:

$R=r$	$P [R=r]$
0	$P [\{1;1\} \cup \{2;2\} \cup \{3;3\} \cup \{4;4\}] = 4/16$
1	$P [\{1;2\} \cup \{2;1\} \cup \{2;3\} \cup \{3;2\} \cup \{3;4\} \cup \{4;3\}] = 6/16$
2	$P [\{1;3\} \cup \{3;1\} \cup \{2;4\} \cup \{4;2\}] = 4/16$
3	$P [\{1;4\} \cup \{4;1\}] = 2/16$

C) CASO CONTINUO:

1. Distribución del i-ésimo valor de la muestra.

Si u_i es el i-ésimo valor de la muestra, tiene que ocurrir que²:

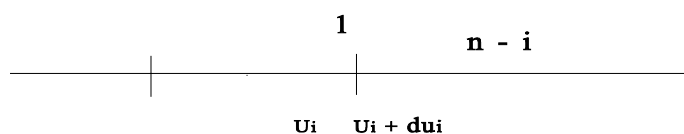
- 1) alguna variable muestral debe tomar dicho valor. En realidad, por tratarse del caso continuo, lo que tiene que ocurrir es que alguna variable muestral tome un valor en el intervalo $(u_i ; u_i + \delta u_i]$. Evidentemente, se realiza el supuesto de que la probabilidad de que dos o más variables muestrales tomen un valor perteneciente al intervalo es despreciable.
- 2) $(i-1)$ variables muestrales deben tomar valores inferiores a u_i .
- 3) $(n-i)$ variables muestrales deben tomar valores superiores a $u_i + \delta u_i$.

² Imaginemos que se trata de que el 7 sea el quinto valor de la muestra. La probabilidad de este suceso no es sino la probabilidad de las muestras que tienen:

- Cuatro valores inferiores o iguales al 7, uno, dos, tres,entre 7 y $7 + \delta u_5$, y el resto mayores que $7 + \delta u_5$.
- Tres valores inferiores o iguales a 7, dos, tres, cuatro, entre 7 y $7 + \delta u_5$, y el resto mayores que $7 + \delta u_5$.
- Dos valores inferiores o iguales a 7, tres, cuatro, cinco, entre 7 y $7 + \delta u_5$, y el resto mayores que $7 + \delta u_5$.
- Un valor inferior o igual a 7, cuatro, cinco, seis, entre 7 y $7 + \delta u_5$, y el resto mayores que $7 + \delta u_5$.
- Ningún valor inferior o igual a 7, cinco, seis, siete, entre 7 y $7 + \delta u_5$, y el resto mayores que $7 + \delta u_5$.

Sin embargo, al suponer que la probabilidad de que dos o más variables muestrales tomen un valor perteneciente al intervalo 7 y $7 + \delta u_5$ es despreciable, el único caso posible es el expuesto.

$i - 1$



de tal manera que

1) La probabilidad de que una variable muestral X_i tome un valor en el intervalo $(u_i ; u_i + \delta u_i]$ es:

$$P(u_i < X_i \leq u_i + \delta u_i) = P(u_i < \xi \leq u_i + \delta u_i) = F_\xi(u_i + \delta u_i) - F_\xi(u_i)$$

2) La probabilidad de que una variable muestral X_i tome un valor igual o inferior a u_i es:

$$P(X_i \leq u_i) = P(\xi \leq u_i) = F_\xi(u_i)$$

3) La probabilidad de que una variable muestral X_i tome un valor superior a $u_i + \delta u_i$ es

$$P(X_i > u_i + \delta u_i) = P(\xi > u_i + \delta u_i) = 1 - F_\xi(u_i + \delta u_i)$$

Dado que se pueden llevar a cabo $\frac{n!}{(i-1)! 1! (n-i)!}$ ordenaciones de las n variables aleatorias

muestrales que verifiquen las condiciones anteriores (una toma un valor en el intervalo $(u_i ; u_i + \delta u_i]$, $(i-1)$ toman valores no superiores a u_i y $(n-i)$ valores superiores a $u_i + \delta u_i$) y en el caso de muestreo aleatorio simple todas ellas tienen la misma probabilidad, $[F_\xi(u_i)]^{i-1} [F_\xi(u_i + \delta u_i) - F_\xi(u_i)] [1 - F_\xi(u_i + \delta u_i)]^{n-i}$, se tiene que, para m.a.s., la probabilidad de que u_i sea el i -ésimo valor de la muestra es:

$$P(u_i < U_i \leq u_i + \delta u_i) =$$

$$= \frac{n!}{(i-1)! 1! (n-i)!} [F_\xi(u_i)]^{i-1} [F_\xi(u_i + \delta u_i) - F_\xi(u_i)] [1 - F_\xi(u_i + \delta u_i)]^{n-i}$$

con lo que, dividiendo primeramente por δu_i y pasando posteriormente al límite cuando $\delta u_i \rightarrow 0$, se tiene

$$\lim_{\delta u_i \rightarrow 0} \frac{P(u_i < U_i \leq u_i + \delta u_i)}{\delta u_i} =$$

$$= \lim_{\delta u_i \rightarrow 0} \frac{n!}{(i-1)! 1! (n-i)!} [F_\xi(u_i)]^{i-1} \left[\frac{F_\xi(u_i + \delta u_i) - F_\xi(u_i)}{\delta u_i} \right] [1 - F_\xi(u_i + \delta u_i)]^{n-i}$$

o bien:

$$\lim_{\delta u_i \rightarrow 0} \frac{G_{U_i}(u_i + \delta u_i) - G_{U_i}(u_i)}{\delta u_i} =$$

$$= \frac{n!}{(i-1)! 1! (n-i)!} \lim_{\delta u_i \rightarrow 0} [F_{\xi}(u_i)]^{i-1} \lim_{\delta u_i \rightarrow 0} \left[\frac{F_{\xi}(u_i + \delta u_i) - F_{\xi}(u_i)}{\delta u_i} \right] \lim_{\delta u_i \rightarrow 0} [1 - F_{\xi}(u_i + \delta u_i)]^{n-i}$$

con lo que, finalmente, la función de densidad del i-ésimo valor de la muestra es:

$$g_{U_i}(u_i) = \frac{n!}{(i-1)! 1! (n-i)!} [F_{\xi}(u_i)]^{i-1} f_{\xi}(u_i) [1 - F_{\xi}(u_i)]^{n-i}$$

2) Particularizaciones:

$$g_{U_1}(u_1) = n f_{\xi}(u_1) [1 - F_{\xi}(u_1)]^{n-1}$$

$$g_{U_n}(u_n) = n [F_{\xi}(u_n)]^{n-1} f_{\xi}(u_n)$$

3. Distribución conjunta del i-ésimo y j-ésimo valor de la muestra (i < j).

Para obtener la distribución conjunta en el muestreo (aleatorio simple) de U_i y U_j (función de densidad) partimos del cálculo de la probabilidad del siguiente suceso: $\{u_i < U_i \leq u_i + \delta u_i; u_j < U_j \leq u_j + \delta u_j\}$, para lo cual:

a) Una variable muestral debe tomar un valor en el intervalo $(u_i; u_i + \delta u_i]$. La probabilidad de que X_i lo haga es

$$P(u_i < X_i \leq u_i + \delta u_i) = P(u_i < \xi \leq u_i + \delta u_i) = F_{\xi}(u_i + \delta u_i) - F_{\xi}(u_i)$$

Evidentemente, se realiza el supuesto de que la probabilidad de que dos o más variables muestrales tomen un valor perteneciente al intervalo es despreciable.

b) Una variable muestral debe tomar un valor en el intervalo $(u_j; u_j + \delta u_j]$. La probabilidad de que X_j lo haga es

$$P(u_j < X_j \leq u_j + \delta u_j) = P(u_j < \xi \leq u_j + \delta u_j) = F_{\xi}(u_j + \delta u_j) - F_{\xi}(u_j)$$

De nuevo se realiza el supuesto de que la probabilidad de que dos o más variables muestrales tomen un valor perteneciente al intervalo es despreciable.

c) (i-1) variables muestrales deben tomar valores iguales o inferiores a u_i . La probabilidad de que una variable muestral concreta, X_i , tome un valor igual o inferior a u_i es

$$P(X_i \leq u_i) = P(\xi \leq u_i) = F_\xi(u_i)$$

d) (n-j) variables muestrales deben tomar valores superiores a $u_j + \delta u_j$. La probabilidad de que X_i (una variable muestral cualesquiera) tome un valor superior a $u_i + \delta u_i$ es

$$P(X_i > u_j + \delta u_j) = P(\xi > u_j + \delta u_j) = 1 - F_\xi(u_j + \delta u_j)$$

e) Las (j-i-1) restantes variables aleatorias deben tomar valores en el intervalo $(u_i + \delta u_i; u_j]$. La probabilidad de que X_i lo haga es

$$P(u_i + \delta u_i < X_i \leq u_j) = P(u_i + \delta u_i < \xi \leq u_j) = F_\xi(u_j) - F_\xi(u_i + \delta u_i)$$

Dado que se pueden llevar a cabo $\frac{n!}{(i-1)!1!(j-i-1)!1!(n-j)!}$ ordenaciones de las n

variables aleatorias muestrales que verifiquen las condiciones anteriores y que por tratarse de muestreo aleatorio simple todas ellas tienen la misma probabilidad,

$$[F_\xi(u_i)]^{i-1} [F_\xi(u_i + \delta u_i) - F_\xi(u_i)] [F_\xi(u_j) - F_\xi(u_i + \delta u_i)]^{j-i-1} [F_\xi(u_j + \delta u_j) - F_\xi(u_j)] [1 - F_\xi(u_j + \delta u_j)]^{n-j}$$

se tiene que

$$P(u_i < U_i \leq u_i + \delta u_i; u_j < U_j \leq u_j + \delta u_j) =$$

$$= \frac{n!}{(i-1)!1!(j-i-1)!1!(n-j)!} \cdot$$

$$[F_\xi(u_i)]^{i-1} [F_\xi(u_i + \delta u_i) - F_\xi(u_i)] [F_\xi(u_j) - F_\xi(u_i + \delta u_i)]^{j-i-1}$$

$$[F_\xi(u_j + \delta u_j) - F_\xi(u_j)] [1 - F_\xi(u_j + \delta u_j)]^{n-j}$$

con lo que, dividiendo en ambos lados de la igualdad por $\delta u_i \delta u_j$ y haciendo que la amplitud de los intervalos tienda a cero ($\delta u_i \rightarrow 0$ y $\delta u_j \rightarrow 0$)

$$\begin{aligned}
& \lim_{\delta u_i \rightarrow 0 \delta u_j \rightarrow 0} \frac{P(u_i < U_i \leq u_i + \delta u_i ; usubj < U_j \leq u_j + \delta u_j)}{\delta u_i \delta u_j} \\
&= \lim_{\delta u_i \rightarrow 0 \delta u_j \rightarrow 0} \left[\frac{G(u_i + \delta u_i ; u_j + \delta u_j) - G(u_i ; u_j)}{\delta u_i \delta u_j} \right] \\
&= g(u_i ; u_j) \\
&= \lim_{\delta u_i \rightarrow 0 \delta u_j \rightarrow 0} \frac{n!}{(i-1)! l! (j-i-1)! l! (n-j)!} - \\
& \quad - [F_\xi(u_i)]^{i-1} \left[\frac{F_\xi(u_i + \delta u_i) - F_\xi(u_i)}{\delta u_i} \right] [F_\xi(u_j) - F_\xi(u_i + \delta u_i)]^{j-i-1} \\
& \quad - \left[\frac{F_\xi(u_j + \delta u_j) - F_\xi(u_j)}{\delta u_j} \right] [l - F_\xi(u_j + \delta u_j)]^{n-j}
\end{aligned}$$

de tal forma que

$$\begin{aligned}
& g(u_i ; u_j) = \\
&= \frac{n!}{(i-1)! l! (j-i-1)! l! (n-j)!} - \\
& \quad - \lim_{\delta u_i \rightarrow 0} [F_\xi(u_i)]^{i-1} \lim_{\delta u_i \rightarrow 0} \frac{[F_\xi(u_i + \delta u_i) - F_\xi(u_i)]}{\delta u_i} \lim_{\delta u_i \rightarrow 0} [F_\xi(u_j) - F_\xi(u_i + \delta u_i)]^{j-i-1} \\
& \quad - \lim_{\delta u_j \rightarrow 0} \frac{[F_\xi(u_j + \delta u_j) - F_\xi(u_j)]}{\delta u_j} \lim_{\delta u_j \rightarrow 0} [l - F_\xi(u_j + \delta u_j)]^{n-j} = \\
&= \frac{n!}{(i-1)! l! (j-i-1)! l! (n-j)!} - \\
& \quad - [F_\xi(u_i)]^{i-1} f_\xi(u_i) [F_\xi(u_j) - F_\xi(u_i)]^{j-i-1} f_\xi(u_j) [l - F_\xi(u_j)]^{n-j}
\end{aligned}$$

4. Particularizaciones: Distribución conjunta del menor y mayor valor de la muestra.

Particularizando la expresión anterior en $i=1$ y $j=n$ se tiene que

$$\begin{aligned} g(u_1; u_n) &= \frac{n!}{0! 1! (n-2)! 1! 0!} f_{\xi}(u_1) [F_{\xi}(u_n) - F_{\xi}(u_1)]^{n-2} f_{\xi}(u_n) \\ &= n(n-1) f_{\xi}(u_1) [F_{\xi}(u_n) - F_{\xi}(u_1)]^{n-2} f_{\xi}(u_n) \end{aligned}$$

5. Distribución del recorrido.

Se define recorrido (R) de una muestra como la diferencia entre el mayor y el menor de sus valores. En otros términos, $R = U_n - U_1$.

Para obtener la distribución del recorrido muestral acudimos al teorema fundamental del cambio de variable bidimensional, cuya notación adecuamos a nuestro caso.

- Sea $g(u_1; u_n)$ la función de densidad conjunta de la variable aleatoria bidimensional $(U_1; U_n)$.

- Sean $r=z(u_1; u_n)$ y $s=w(u_1; u_n)$ dos transformaciones que constituyen una aplicación inyectiva de $R^2 (U_1; U_n)$ en $R^2 (R, S)$ tal que existe en su conjunto imagen la transformación inversa

$$u_1=z^{-1}(r, s), u_n=w^{-1}(r, s).$$

- **Existen, y son continuas**, las derivadas parciales $\frac{\partial u_1}{\partial r}, \frac{\partial u_1}{\partial s}, \frac{\partial u_n}{\partial r}, \frac{\partial u_n}{\partial s}$.

ENTONCES: la función de densidad conjunta de (R;S) viene dada por

$$h_{(R;S)}(r; s) = g_{(U_1; U_n)}[z^{-1}(r; s); w^{-1}(r; s)] |J|$$

siendo

$$J = \frac{\partial(u_1, u_n)}{\partial(r, s)} = \begin{vmatrix} \frac{\partial u_1}{\partial r} & \frac{\partial u_1}{\partial s} \\ \frac{\partial u_n}{\partial r} & \frac{\partial u_n}{\partial s} \end{vmatrix}$$

En nuestro caso se tiene que:

$$1. \text{ Transformación directa: } \begin{cases} R = U_n - U_1 \\ S = U_1 \end{cases}$$

$$2. \text{ Transformación inversa } \begin{cases} U_n = R + S \\ U_1 = S \end{cases}$$

3. Jaccobiano de la transformación inversa

$$J = \frac{\partial(u_1, u_n)}{\partial(r, s)} = \begin{vmatrix} \frac{\partial u_1}{\partial r} & \frac{\partial u_1}{\partial s} \\ \frac{\partial u_n}{\partial r} & \frac{\partial u_n}{\partial s} \end{vmatrix} = \begin{vmatrix} 0 & 1 \\ 1 & 1 \end{vmatrix} = -1$$

siendo su valor absoluto unitario.

En consecuencia, la función de densidad conjunta de las nuevas variables R y S es

$$\begin{aligned} h_{(R,S)}(r; s) &= g_{(u_1, u_n)}[u_1(r; s); u_n(r; s)] |J| \\ &= n(n-1) f_{\xi}(s) [F_{\xi}(r+s) - F_{\xi}(s)]^{n-2} f_{\xi}(r+s) \end{aligned}$$

$$\text{cuyo campo de variación es } \begin{cases} -\infty < s < \infty \\ r \geq 0 \end{cases}$$

Por tanto, a partir de la función de densidad conjunta de R y S, la distribución marginal del recorrido (función de densidad) se obtiene como

$$h_R(r) = \int_{-\infty}^{\infty} h(r; s) ds = \int_{-\infty}^{\infty} n(n-1) f_{\xi}(s) [F_{\xi}(r+s) - F_{\xi}(s)]^{n-2} f_{\xi}(r+s) ds$$

con campo de variación $r \geq 0$.

D) EJERCICIOS PRÁCTICOS.

Caso discreto: 1.11., 1.12, 1.13, 1.14 de Martín Pliego, J., Montero Lorenzo, J.M. y Ruiz-Maya, L. (2000): “Problemas de Inferencia”. 2ª ed. AC, Madrid.

Caso continuo: 1.15, 1.16, 1.17 de Martín Pliego, J., Montero Lorenzo, J.M. y Ruiz-Maya, L. (2000): “Problemas de Inferencia”. 2ª ed. AC, Madrid.

ADICIONALES:

Ejercicio: Determine la función de densidad del subrecorrido $R_{ij} = U_j - U_i$.

El teorema fundamental del cambio de variable bidimensional establece que

$$h_{(R_{ij}; S)}(r_{ij}; s) = g_{(u_i; u_j)}[u_i(r_{ij}; s); u_j(r_{subij}; s)] |J|$$

donde U_i y U_j son las antiguas variables, R_{ij} y S las nuevas y $|J|$ el valor absoluto del determinante jacobiano de la transformación inversa.

Por tanto, se tiene que:

$$1. \text{ Transformación directa: } \begin{cases} R_{ij} = U_j - U_i \\ S = U_i \end{cases}$$

$$2. \text{ Transformación inversa } \begin{cases} U_j = R_{ij} + S \\ U_i = S \end{cases}$$

3. Determinante jacobiano de la transformación inversa

$$J = \frac{\partial(u_i, u_j)}{\partial(r, s)} = \begin{vmatrix} \frac{\partial u_i}{\partial r} & \frac{\partial u_i}{\partial s} \\ \frac{\partial u_j}{\partial r} & \frac{\partial u_j}{\partial s} \end{vmatrix} = \begin{vmatrix} 0 & 1 \\ 1 & 1 \end{vmatrix} = -1$$

siendo su valor absoluto unitario.

Y como la función de densidad conjunta de U_i y U_j es

$$g(u_i, u_j) = \frac{n!}{(i-1)! 1! (j-i-1)! 1! (n-j)!} -$$

$$- [F_\xi(u_i)]^{i-1} f_\xi(u_i) [F_\xi(u_j) - F_\xi(u_i)]^{j-i-1} f_\xi(u_j) [1 - F_\xi(u_j)]^{n-j}$$

entonces la función de densidad conjunta de las nuevas variables R_{ij} y S es

$$h_{(R_{ij}; S)}(r_{ij}; s) = g_{(u_i; u_j)}[u_i(r_{ij}; s); u_j(rs_{ubij}; s)] |J|$$

$$= \frac{n!}{(i-1)! (j-i-1)! (n-j)!} [F(s)]^{i-1} f(s) [F(s+r_{ij}) - F(s)]^{j-i-1} f(s+r_{ij}) [1 - F(s+r_{ij})]^{n-j}$$

con

$$\begin{cases} -\infty < u_i < \infty \\ -\infty < u_i < u_j < \infty \end{cases} \rightarrow \begin{cases} -\infty < s < \infty \\ -\infty < s < r_{ij} + s < \infty \end{cases} \rightarrow \begin{cases} -\infty < s < \infty \\ 0 < r_{ij} < \infty \end{cases}$$

con lo que la función de densidad marginal de R_{ij} es

$$h_{R_{ij}}(r_{ij}) = \frac{n!}{(i-1)! (j-i-1)! (n-j)!} \int_{-\infty}^{\infty} [F(s)]^{i-1} f(s) [F(s+r_{ij}) - F(s)]^{j-i-1} f(s+r_{ij}) [1 - F(s+r_{ij})]^{n-j} ds$$

con $0 < r_{ij} < \infty$.

Ejercicio: Demuestre que para una distribución uniforme $(0;1)$ el subrecorrido $(i;j)$ se

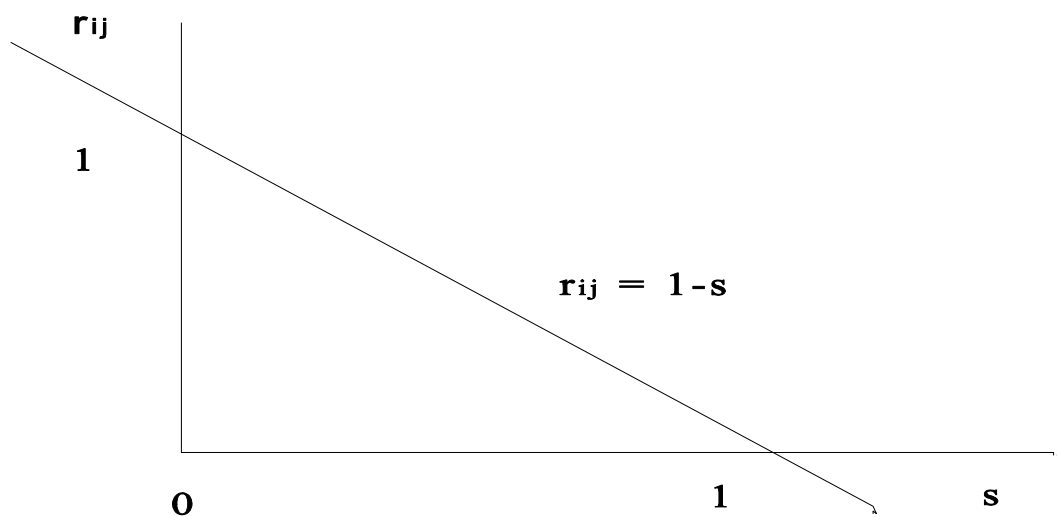
distribuye exactamente como el estadístico de orden $j-i$.

En este caso se tiene que, de acuerdo con el resultado del ejercicio anterior, la función de densidad conjunta de las nuevas variables R_{ij} y S es

$$\begin{aligned} h_{(R_{ij}; S)}(r_{ij}; s) &= g_{(u_i; u_j)}[u_i(r_{ij}; s); u_j(r_{ij}; s)] |J| \\ &= \frac{n!}{(i-1)!(j-i-1)!(n-j)!} s^{i-1} r_{ij}^{j-i-1} (1-r_{ij}-s)^{n-j} \end{aligned}$$

El campo de variación de la variable bidimensional $(R_{ij}; S)$ se obtiene como sigue

$$\begin{cases} 0 < u_i < 1 \\ u_i < u_j < 1 \end{cases} \rightarrow \begin{cases} 0 < s < 1 \\ s < r_{ij} + s < 1 \end{cases} \rightarrow \begin{cases} 0 < s < 1 \\ 0 < r_{ij} < 1-s \end{cases}$$



Por tanto, finalmente, la función de densidad conjunta del subrecorrido es

$$\begin{aligned}
h_{R_{ij}}(r_{ij}) &= \int_0^{1-r_{ij}} h_{(R_{ij};S)}(r_{ij};s) ds = \int_0^{1-r_{ij}} \frac{n!}{(i-1)!(j-i-1)!(n-j)!} s^{i-1} r_{ij}^{j-i-1} (1-r_{ij}-s)^{n-j} ds \\
&= \frac{n!}{(i-1)!(j-i-1)!(n-j)!} r_{ij}^{j-i-1} \int_0^{1-r_{ij}} s^{i-1} (1-r_{ij}-s)^{n-j} ds
\end{aligned}$$

donde $\int_0^{1-r_{ij}} s^{i-1} (1-r_{ij}-s)^{n-j} ds$ se resuelve integrando sucesivamente por partes hasta que el

exponente de la “s” se anule (desde la iteración h=1 hasta la h=i). Tomando

$$\left\{ \begin{array}{ll} u = s^{i-h} & du = (i-h) s^{i-h-1} ds \\ dv = (1-r_{ij}-s)^{n-j+h-1} ds & v = -\frac{(1-r_{ij}-s)^{n-j+h}}{n-j+h} \end{array} \right.$$

se obtiene $\int_0^{1-r_{ij}} s^{i-1} (1-r_{ij}-s)^{n-j} ds = \frac{(i-1)!}{(n-j+i)!} (1-r_{ij})^{n-j+i}$ con lo que, finalmente,

$$h_{R_{ij}}(r_{ij}) = \frac{n!}{(j-i-1)!(n-j+i)!} r_{ij}^{j-i-1} (1-r_{ij})^{n-j+i} \quad 0 < r_{ij} < 1$$

que no es sino la función de densidad del (j-i)-ésimo valor de una muestra de tamaño n tomada de una población U(0;1).

Adicional: En lo que se refiere a la resolución de la integral $\int_0^{1-r_{ij}} s^{i-1} (1-r_{ij}-s)^{n-j} ds$, a modo de ejemplo, supongamos que $i=4$. Entonces la resolución de la misma se lleva cabo en las siguientes iteraciones:

Iteración $h=1$: Haciendo

$$\left\{ \begin{array}{ll} u = s^{i-1} & du = (i-1) s^{i-2} ds \\ dv = (1-r_{ij}-s)^{n-j} ds & v = -\frac{(1-r_{ij}-s)^{n-j+1}}{n-j+1} \end{array} \right.$$

con lo que

$$\begin{aligned} \int_0^{1-r_{ij}} s^{i-1} (1-r_{ij}-s)^{n-j} ds &= \left[-\frac{s^{i-1} (1-r_{ij}-s)^{n-j+1}}{n-j+1} \right]_0^{1-r_{ij}} + \frac{i-1}{n-j+1} \int_0^{1-r_{ij}} s^{i-2} (1-r_{ij}-s)^{n-j+1} ds \\ &= \frac{i-1}{n-j+1} \int_0^{1-r_{ij}} s^{i-2} (1-r_{ij}-s)^{n-j+1} ds \end{aligned}$$

Iteración $h=2$: Haciendo

$$\left\{ \begin{array}{ll} u = s^{i-2} & du = (i-2) s^{i-3} ds \\ dv = (1-r_{ij}-s)^{n-j+1} ds & v = -\frac{(1-r_{ij}-s)^{n-j+2}}{n-j+2} \end{array} \right.$$

con lo que

$$\int_0^{l-r_{ij}} s^{i-1} (1 - r_{ij} - s)^{n-j} ds =$$

$$\frac{i-1}{n-j+1} \left[\left[-\frac{s^{i-2} (1 - r_{ij} - s)^{n-j+2}}{n-j+2} \right]_0^{l-r_{ij}} + \frac{i-2}{n-j+2} \int_0^{l-r_{ij}} s^{i-3} (1 - r_{ij} - s)^{n-j+2} ds \right] =$$

$$\frac{i-1}{n-j+1} \frac{i-2}{n-j+2} \int_0^{l-r_{ij}} s^{i-3} (1 - r_{ij} - s)^{n-j+2} ds$$

Iteración $h=3$: Haciendo

$$\left\{ \begin{array}{ll} u = s^{i-3} & du = (i-3) s^{i-4} ds \\ dv = (1 - r_{ij} - s)^{n-j+2} ds & v = -\frac{(1 - r_{ij} - s)^{n-j+3}}{n-j+3} \end{array} \right.$$

con lo que

$$\int_0^{l-r_{ij}} s^{i-1} (1 - r_{ij} - s)^{n-j} ds =$$

$$\frac{i-1}{n-j+1} \frac{i-2}{n-j+2} \left[\left[-\frac{s^{i-3} (1 - r_{ij} - s)^{n-j+3}}{n-j+3} \right]_0^{l-r_{ij}} + \frac{i-3}{n-j+3} \int_0^{l-r_{ij}} s^{i-4} (1 - r_{ij} - s)^{n-j+3} ds \right] =$$

$$\frac{i-1}{n-j+1} \frac{i-2}{n-j+2} \frac{i-3}{n-j+3} \int_0^{l-r_{ij}} s^{i-4} (1 - r_{ij} - s)^{n-j+3} ds$$

Iteración $h=4$: No hay más que resolver la integral inmediata

$$\int_0^{l-r_{ij}} s^{i-4} (1 - r_{ij} - s)^{n-j+3} ds = \int_0^{l-r_{ij}} (1 - r_{ij} - s)^{n-j+3} ds$$

con lo que

$$\begin{aligned}
\int_0^{1-r_{ij}} s^{i-1} (1-r_{ij}-s)^{n-j} ds &= \frac{i-1}{n-j+1} \frac{i-2}{n-j+2} \frac{i-3}{n-j+3} \int_0^{1-r_{ij}} (1-r_{ij}-s)^{n-j+3} ds = \\
&\frac{i-1}{n-j+1} \frac{i-2}{n-j+2} \frac{i-3}{n-j+3} \left[-\frac{(1-r_{ij}-s)^{n-j+4}}{n-j+4} \right]_0^{1-r_{ij}} \\
&\frac{i-1}{n-j+1} \frac{i-2}{n-j+2} \frac{i-3}{n-j+3} \frac{(1-r_{ij})^{n-j+4}}{n-j+4} = \frac{(i-1)!}{(n-j+i)!} (1-r_{ij})^{n-j+i} \\
&\frac{(n-j)!}{(n-j)!}
\end{aligned}$$

Ejercicio: Demuestre que para la distribución exponencial standard ($\mu=1$) el subrecorrido ($i;j$) se distribuye exactamente como el estadístico de orden $j-i$ en una muestra de tamaño $(n-i)$.

En primer lugar, recuérdese que para la distribución exponencial standard

$$f_{\xi}(x) = e^{-x} \quad x > 0$$

$$F_{\xi}(x) = 1 - e^{-x} \quad x > 0$$

En este caso se tiene que, de acuerdo con el resultado del ejercicio anterior, la función de densidad conjunta de las nuevas variables R_{ij} y S es

$$\begin{aligned}
h_{(R_{ij};S)}(r_{ij};s) &= g_{(u_i;u_j)}[u_i(r_{ij};s); u_j(r_{subij};s)] | J | \\
&= \frac{n!}{(i-1)!(j-i-1)!(n-j)!} [1-e^{-s}]^{i-1} e^{-s} [e^{-s} - e^{-s-r_{ij}}]^{j-i-1} e^{-s-r_{ij}} [e^{-s-r_{ij}}]^{n-j}
\end{aligned}$$

con ambas variables variando entre 0 e infinito.

En consecuencia, la función de densidad marginal de R_{ij} se obtiene como

$$h_{R_{ij}}(r_{ij}) = \frac{n!}{(i-1)!(j-i-1)!(n-j)!} [e^{-r_{ij}}]^{n-j+1} [1-e^{-r_{ij}}]^{j-i-1} \int_0^{\infty} [1-e^{-s}]^{i-1} [e^{-s}]^{n-i+1} ds \quad 0 < r_{ij} < \infty$$

e integrando reiteradamente por partes se tiene que

$$h_{R_{ij}}(r_{ij}) = \frac{(n-i)!}{(j-i-1)!(n-j)!} [e^{-r_{ij}}]^{n-j+1} [1-e^{-r_{ij}}]^{j-i-1} \quad 0 < r_{ij} < \infty$$

Adicional: Supongamos $U_i = U_3$. Haciendo

$$\begin{cases} u = (1 - e^{-s})^{i-1} & du = (i-1)(1 - e^{-s})^{i-2} e^{-s} ds \\ dv = (e^{-s})^{n-i+1} ds & v = -\frac{1}{n-i+1} (e^{-s})^{n-i+1} \end{cases}$$

se tiene que

$$\begin{aligned} & \int_0^\infty (1 - e^{-s})^{i-1} [e^{-s}]^{n-i+1} ds = \\ & = \left[- (1 - e^{-s})^{i-1} \frac{1}{n-i+1} (e^{-s})^{n-i+1} \right]_0^\infty + \frac{i-1}{n-i+1} \int_0^\infty (1 - e^{-s})^{i-2} (e^{-s})^{n-i+2} ds \\ & = \frac{i-1}{n-i+1} \int_0^\infty (1 - e^{-s})^{i-2} (e^{-s})^{n-i+2} ds \end{aligned}$$

Haciendo

$$\begin{cases} u = (1 - e^{-s})^{i-2} & du = (i-2)(1 - e^{-s})^{i-3} e^{-s} ds \\ dv = (e^{-s})^{n-i+2} ds & v = -\frac{1}{n-i+2} (e^{-s})^{n-i+2} \end{cases}$$

se tiene que

$$\begin{aligned} & \frac{i-1}{n-i+1} \int_0^\infty (1 - e^{-s})^{i-2} [e^{-s}]^{n-i+2} ds = \\ & = \frac{i-1}{n-i+1} \left[- (1 - e^{-s})^{i-2} \frac{1}{n-i+2} (e^{-s})^{n-i+2} \right]_0^\infty + \frac{i-2}{n-i+2} \int_0^\infty (1 - e^{-s})^{i-3} (e^{-s})^{n-i+3} ds \\ & = \frac{i-1}{n-i+1} \frac{i-2}{n-i+2} \int_0^\infty (1 - e^{-s})^{i-3} (e^{-s})^{n-i+3} ds \\ & = \frac{i-1}{n-i+1} \frac{i-2}{n-i+2} \int_0^\infty (e^{-s})^n ds = \frac{i-1}{n-i+1} \frac{i-2}{n-i+2} \frac{1}{n} \end{aligned}$$

Nótese que si, por ejemplo, $n=10$ entonces $(n-i+1)=8$, $(n-i+2)=9$ y $n=10$, con lo que

$$\frac{n!}{(n-i+1)(n-i+2)n} = (n-i)!$$

y que además $(i-1) \cdot (i-2) = 2! = (i-1)!$, con lo que

$$h_{R_{ij}}(r_{ij}) = \frac{(n-i)!}{(j-i-1)!(n-j)!} [e^{-r_{ij}}]^{n-j+1} [1 - e^{-r_{ij}}]^{j-i-1} \quad 0 < r_{ij} < \infty$$

En la expresión resultante se observa que para la distribución exponencial standard, el subrecoredo $(i;j)$ se distribuye exactamente como el $(j-i)$ -ésimo estadístico ordenado de una muestra de tamaño $(n-i)$ de una población exponencial standard.

Ejercicio: Compruebe que si las m.a.s. se extraen de una población $U(0,1)$ entonces la distribución del i -ésimo estadístico ordenado sigue un modelo Beta ($i; n-i+1$). Si la muestra es de tamaño 10, calcule la esperanza y la varianza del mayor valor de la muestra.

Primeramente recordemos que, para la distribución Beta ($\mu; r$) se tiene que

$$f(x) = \frac{(\mu + r - 1)!}{(\mu - 1)!(r - 1)!} x^{\mu-1} (1-x)^{r-1} \quad 0 \leq x \leq 1$$

$$E(\xi) = \frac{\mu}{\mu + r} \quad V(\xi) = \frac{\mu - r}{(\mu + r)^2 (\mu + r + 1)}$$

Hecho este recordatorio, la función de densidad del i -ésimo valor de la muestra es

$$g_{U_i}(u_i) = \frac{n!}{(i-1)!(n-i)!} [F_\xi(u_i)]^{i-1} f_\xi(u_i) [1 - F_\xi(u_i)]^{n-i} \quad 0 \leq u_i \leq 1$$

que en el caso de que la muestra proceda de una población $U(0,1)$ se transforma en

$$g_{U_i}(u_i) = \frac{n!}{(i-1)!(n-i)!} [u_i]^{i-1} [1-u_i]^{n-i} \quad 0 \leq u_i \leq 1$$

que no es sino la función de densidad de una Beta ($i; n-i+1$).

Para m.a.s. de tamaño 10, U_{10} se distribuye como una Beta (10; 1), con lo que

$$E(U_{10}) = \frac{10}{10+1} = \frac{10}{11}$$

$$V(U_{10}) = \frac{10-1}{(10+1)^2 (10+1+1)} = 0,0377$$

Nótese que el cálculo de la esperanza y la varianza a través de la función de densidad del estadístico U_{10} hubiese sido sencillo por tratarse del mayor valor de la muestra. No lo es tanto en otros casos.

E. Apéndice: Función de distribución de los estadísticos ordenados.

Una forma sencilla de obtener la función de distribución de cualquier estadístico ordenado parte de la función de distribución empírica de la muestra.

La función de distribución del i-ésimo valor de la muestra es

$$\begin{aligned}
 G_{U_i}(u_i) &= P(U_i \leq u_i) \\
 &= P(\text{en el muestreo al menos } i \text{ valores muestrales no excedan a } u_i) \\
 &= P(n F_n^*(u_i) \geq i) \\
 &= \sum_{h=i}^n \binom{n}{h} [F_{\xi}(u_i)]^h [1 - F_{\xi}(u_i)]^{n-h}
 \end{aligned}$$

Nota aclaratoria: Para que el 5º valor de la muestra sea menor o igual que 10 al menos tiene que haber 5 valores en la muestra menores o iguales que 10:

- Si son 5 los valores menores o iguales que 10 el quinto también lo será.
- Si son 6 los valores menores o iguales que 10 el quinto también lo será.
- Si son 7 los valores menores o iguales que 10 el quinto también lo será. Y así, sucesivamente.

Mediante derivación de la expresión anterior se obtienen las funciones de densidad de los estadísticos ordenados. Efectivamente:

$$\begin{aligned}
 g_{U_i}(u_i) &= \\
 &\left[\sum_{h=i}^n \binom{n}{h} h [F_{\xi}(u_i)]^{h-1} f_{\xi}(u_i) [1 - F_{\xi}(u_i)]^{n-h} \right] - \\
 &- \left[\sum_{h=i}^n \binom{n}{h} [F_{\xi}(u_i)]^h (n-h) [1 - F_{\xi}(u_i)]^{n-h-1} f_{\xi}(u_i) \right]
 \end{aligned}$$

pero el segundo sumando se anula para $i=n$ con lo que la expresión anterior se puede escribir como:

$$\begin{aligned}
 g_{U_i}(u_i) &= \\
 &\left[\sum_{h=i}^n \binom{n}{h} h [F_{\xi}(u_i)]^{h-1} f_{\xi}(u_i) [1 - F_{\xi}(u_i)]^{n-h} \right] - \\
 &- \left[\sum_{h=i}^{n-1} \binom{n}{h} [F_{\xi}(u_i)]^h (n-h) [1 - F_{\xi}(u_i)]^{n-h-1} f_{\xi}(u_i) \right]
 \end{aligned}$$

$$g_{U_i}(u_i) =$$

$$\left[\sum_{h=i}^n \frac{n!}{(h-1)!(n-h)!} [F_{\xi}(u_i)]^{h-1} f_{\xi}(u_i) [1 - F_{\xi}(u_i)]^{n-h} \right] -$$

$$- \left[\sum_{h=i}^{n-1} \frac{n!}{(h)!(n-h-1)!} [F_{\xi}(u_i)]^h [1 - F_{\xi}(u_i)]^{n-h-1} f_{\xi}(u_i) \right]$$

operando:

donde, sumando para h-1 en vez de para h en el segundo sumando, se tiene

$$g_{U_i}(u_i) =$$

$$\left[\sum_{h=i}^n \frac{n!}{(h-1)!(n-h)!} [F_{\xi}(u_i)]^{h-1} f_{\xi}(u_i) [1 - F_{\xi}(u_i)]^{n-h} \right] -$$

$$- \left[\sum_{h=i+1}^n \frac{n!}{(h-1)!(n-h)!} [F_{\xi}(u_i)]^{h-1} [1 - F_{\xi}(u_i)]^{n-h} f_{\xi}(u_i) \right]$$

y cancelando términos se tiene finalmente que

$$g_{U_i}(u_i) = \frac{n!}{(i-1)!(n-i)!} [F_{\xi}(u_i)]^{i-1} f_{\xi}(u_i) [1 - F_{\xi}(u_i)]^{n-i}$$

Ejemplo 1: Obtenga la función de densidad del mayor valor de la muestra a partir de su función de distribución.

La función de distribución de U_n es $G_{U_n}(u_n) = [F_{\xi}(u_n)]^n$ con lo que $g_{U_n}(u_n) = n [F_{\xi}(u_n)]^{n-1} f_{\xi}(u_n)$.

Ejemplo 2: Sea una población de la que se extraen m.a.s. de tamaño 4. Obténgase la función de densidad del tercer valor de la muestra a partir de su función de distribución.

Particularizando la expresión genérica de la derivada de la función de distribución en $h=3$ y $h=4$ y sumando ambos resultados se tiene que:

$$\begin{aligned}
 g_{U_3}(u_3) &= \binom{4}{3} \left[3 [F_{\xi}(u_3)]^2 f_{\xi}(u_3) [1 - F_{\xi}(u_3)] - [F_{\xi}(u_3)]^3 [1 - F_{\xi}(u_3)]^0 f_{\xi}(u_3) \right] + \\
 &\quad + \binom{4}{4} \left[4 [F_{\xi}(u_3)]^3 f_{\xi}(u_3) [1 - F_{\xi}(u_3)]^0 - [F_{\xi}(u_3)]^4 [1 - F_{\xi}(u_3)]^{-1} f_{\xi}(u_3) \right] \\
 &= 4 \left[3 [F_{\xi}(u_3)]^2 f_{\xi}(u_3) [1 - F_{\xi}(u_3)] - [F_{\xi}(u_3)]^3 f_{\xi}(u_3) \right] + 4 [F_{\xi}(u_3)]^3 f_{\xi}(u_3) \\
 &= 4 \left[3 [F_{\xi}(u_3)]^2 f_{\xi}(u_3) [1 - F_{\xi}(u_3)] - 4 [F_{\xi}(u_3)]^3 f_{\xi}(u_3) + 4 [F_{\xi}(u_3)]^3 f_{\xi}(u_3) \right] \\
 &= 4 \left[3 [F_{\xi}(u_3)]^2 f_{\xi}(u_3) [1 - F_{\xi}(u_3)] \right]
 \end{aligned}$$

8. Recubrimientos.

Dados dos estadísticos ordenados U_i y U_j se denomina recubrimiento, V_{ij} , a la probabilidad de que la variable aleatoria poblacional pertenezca al intervalo $(U_i ; U_j]$.

$$V_{ij} = P[u_i < \xi \leq u_j] = F_{\xi}(u_j) - F_{\xi}(u_i)$$

siendo V_{ij} una variable aleatoria.

Obtengamos en nuestro ejemplo la distribución de probabilidad del recubrimiento v_{1n} .

Los valores que toma V_{1n} son, además del cero

$$P[1 < \xi \leq 2] = 1/4$$

$$P[1 < \xi \leq 3] = 2/4$$

$$P[1 < \xi \leq 4] = 3/4$$

$$P[2 < \xi \leq 3] = 1/4$$

$$P[2 < \xi \leq 4] = 2/4$$

$$P[3 < \xi \leq 4] = 1/4$$

con lo que la función de cuantía de V_{1n} es

$V_{1n}=v_{1n}$	$P[V_{1n}=v_{1n}]$
0	$P[\{1;1\} \cup \{2;2\} \cup \{3;3\} \cup \{4;4\} \cup \{4;3\} \cup \{3;4\}] = 6/16$
1/4	$P[\{1;2\} \cup \{2;1\} \cup \{2;3\} \cup \{3;2\} \cup \{3;4\} \cup \{4;3\}] = 6/16$
2/4	$P[\{1;3\} \cup \{3;1\} \cup \{2;4\} \cup \{4;2\}] = 4/16$
3/4	$P[\{1;4\} \cup \{4;1\}] = 2/16$