

RESUMEN MA3403 - PROBABILIDADES Y ESTADÍSTICA

Escrito por: Cristian Aguayo Quintana¹

1. Probabilidad

1.1. Modelo probabilístico para experimentos (caso discreto)

1.1.1. Experimento

Un *experimento* es el proceso por medio del cual se hace una observación. Pueden ser *contables* o *incontables*.

1.1.2. Punto muestral

Un *punto muestral* es un resultado individual de un experimento.

1.1.3. Espacio muestral

El *espacio muestral* asociado con un experimento es el conjunto formado por todos los posibles puntos muestrales. Se denota por S .

1.1.4. Evento

Un *evento* es un resultado de un experimento. Puede ser:

- *Simple*: corresponde a un y solo un punto muestral.
- *Compuesto*: corresponde a un número finito o numerable de puntos muestrales distintos.

1.1.5. Probabilidad

Dado S espacio muestral asociado a un experimento. A todo evento $A \subseteq S$ se le asigna un número $\mathbb{P}(A)$ llamado *probabilidad de A* , de modo que se cumplen los siguientes axiomas:

1. $\mathbb{P}(A) \geq 0$
2. $\mathbb{P}(S) = 1$
3. Si A_1, A_2, A_3, \dots forman una secuencia de eventos disjuntos ($A_i \cap A_j = \emptyset$ si $i \neq j$), entonces:

$$\mathbb{P}(A_1 \cup A_2 \cup A_3 \cup \dots) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$$

1.1.6. Método del punto muestral

Los pasos siguientes se usan para hallar la probabilidad de un evento:

1. Definir el experimento y determinar con claridad cómo describir un evento simple.
2. Indicar los eventos simples asociados con el experimento, asegurándose que no se puedan descomponer. Esto define el espacio muestral S .
3. Asignar probabilidades razonables a los puntos muestrales en S , asegurándose que $\mathbb{P}(E_i) \geq 0$ y $\sum_i \mathbb{P}(E_i) = 1$.

¹Si encuentra algún error o tiene alguna sugerencia, escriba a: cristian.aguayo@ing.uchile.cl

4. Definir el evento de interés A , como un conjunto específico de puntos muestrales.
5. Encontrar $\mathbb{P}(A)$ al sumar las probabilidades de los puntos muestrales en A .

1.2. Herramientas para contar espacios muestrales

1.2.1. Espacios equiprobables

Dado un espacio muestral S compuesto por N puntos equiprobables, y un evento A compuesto por n_A puntos muestrales, entonces:

$$\mathbb{P}(A) = \frac{|A|}{|S|} = \frac{n_A}{N}$$

donde $|\cdot|$ es el *cardinal* de un conjunto.

1.2.2. Regla $m \times n$

Con m elementos a_1, a_2, \dots, a_m y n elementos b_1, b_2, \dots, b_n es posible formar $m \times n$ pares que contengan un elemento de cada grupo.

1.2.3. Permutaciones

Un arreglo ordenado de r objetos distintos se denomina *permutación*. El número de formas de ordenar n objetos distintos tomados r a la vez ($r \leq n$) estará dado por

$$P_r^n = n(n-1)(n-2)\dots(n-r+1) = \frac{n!}{(n-r)!}$$

1.2.4. Permutaciones con elementos indistinguibles

El número de maneras de ordenar n objetos donde $\{n_i\}_{i=1}^k$ son indistinguibles y $\sum_{i=1}^k n_i = n$ es:

$$N = \binom{n}{n_1 n_2 \dots n_k} = \frac{n!}{n_1! n_2! \dots n_k!}$$

1.2.5. Combinaciones

El número de *combinaciones* de n objetos tomados r a la vez, es el número de subconjuntos de tamaño r que se pueden formar a partir de los n objetos. Este número estará dado por:

$$C_r^n = \frac{P_r^n}{r!} = \binom{n}{r} = \frac{n!}{r!(n-r)!}$$

1.3. Probabilidad condicional e independencia de eventos

1.3.1. Probabilidad condicional

La *probabilidad condicional* de un evento A dado que un evento B ha ocurrido, es igual a:

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$$

1.3.2. Independencia de eventos

Se dice que dos eventos A y B son *independientes* si se cumple cualquiera de los siguientes casos:

- $\mathbb{P}(A|B) = \mathbb{P}(A)$
- $\mathbb{P}(B|A) = \mathbb{P}(B)$
- $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$

1.4. Leyes de Probabilidad

1.4.1. Ley multiplicativa de probabilidad

La probabilidad de la intersección de dos eventos A y B es:

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B|A) = \mathbb{P}(B)\mathbb{P}(A|B)$$

Si A y B son independientes, entonces:

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$$

1.4.2. Ley aditiva de probabilidad

La probabilidad de la unión de dos eventos A y B es:

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$$

Si A y B son independientes, entonces

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$$

1.4.3. Probabilidad del complemento

Si A es un evento, entonces:

$$\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$$

donde A^c es el *complemento de A* .

1.5. Cálculo de la probabilidad de un evento: el método de composición de evento

1.5.1. Método de composición de evento

1. Definir el experimento
2. Visualizar la naturaleza de los puntos muestrales. Identificar unos pocos para aclarar el modo de pensar del experto en estadística.
3. Escribir una ecuación que exprese el evento de interés A por ejemplo, como una composición de dos o más eventos usando operaciones básicas de conjuntos.
4. Aplicar las leyes de probabilidad a las composiciones obtenidas para hallar $\mathbb{P}(A)$.

1.6. Ley de Probabilidad total y Regla de Bayes

1.6.1. Ley de probabilidad total

Dada una partición disjunta ($B_i \cap B_j = \emptyset$ si $i \neq j$) $\{B_i\}_{i=1}^k$ de S tal que $\mathbb{P}(B_i) > 0$ para $i = 1, 2, \dots, k$, se tiene que para cualquier evento A :

$$\mathbb{P}(A) = \sum_{i=1}^k \mathbb{P}(A|B_i)\mathbb{P}(B_i)$$

1.6.2. Regla de Bayes

Dada $\{B_i\}_{i=1}^k$ de S tal que $\mathbb{P}(B_i) > 0$ para $i = 1, 2, \dots, k$, se define la *regla de Bayes* como:

$$\mathbb{P}(B_j|A) = \frac{\mathbb{P}(A|B_j)\mathbb{P}(B_j)}{\mathbb{P}(A)}$$

2. Variables aleatorias discretas y sus distribuciones de probabilidad

2.1. Definiciones básicas

2.1.1. Variable aleatoria

Una *variable aleatoria* es una función de valor real para la cual el dominio es un espacio muestral. Es decir:

$$Y : S \rightarrow R_y \subseteq \mathbb{R}$$

2.1.2. Variable aleatoria discreta

Una variable aleatoria Y se dice *discreta* si puede tomar sólo un número finito o infinito numerable de valores distintos.

2.2. Distribución de probabilidad para una variable aleatoria discreta

2.2.1. Probabilidad de valores de una variable aleatoria discreta

La probabilidad de que Y tome el valor y , $\mathbb{P}(Y = y)$, se define como la suma de las probabilidades de todos los puntos muestrales en S a los que se les asigna el valor y . A veces, se denotará $\mathbb{P}(Y = y)$ por $p(y)$. Como $p(y)$ es una función que asigna probabilidades a cada valor y de la variable aleatoria Y , suele recibir el nombre de *función de probabilidad para Y* .

2.2.2. Distribución de probabilidad

La *distribución de probabilidad* para una variable aleatoria discreta Y puede ser representada por una fórmula, una tabla o una gráfica que produzca $p(y) = \mathbb{P}(Y = y)$ para toda y .

Para cualquier distribución de probabilidad discreta, lo siguiente debe ser verdadero:

- $0 \leq p(y) \leq 1$
- $\sum_y p(y) = 1$ donde la suma es para todos los y con probabilidad distinta de cero.

2.3. Valor esperado y varianza (caso discreto)

2.3.1. Valor esperado

Sea Y una variable aleatoria discreta con función de probabilidad $p(y)$. Entonces, el *valor esperado* de Y , $\mathbb{E}(Y)$, se define como:

$$\mathbb{E}(Y) = \sum_y yp(y)$$

Dada una variable aleatoria discreta Y con función de probabilidad $p(y)$ y dada $g(Y)$ una función de valor real de Y , entonces, el valor esperado de $g(Y)$ está dado por:

$$\mathbb{E}[g(Y)] = \sum_y g(y)p(y)$$

A partir de esto, es directo que $\mathbb{E}(c) = c$ con $c \in \mathbb{R}$ y que $\mathbb{E}[c \cdot g(Y)] = c \cdot \mathbb{E}[g(Y)]$. Además, se cumple que:

$$\mathbb{E} \left[\sum_{i=1}^k g_i(Y) \right] = \sum_{i=1}^k \mathbb{E}[g_i(Y)]$$

2.3.2. Varianza

Si Y es una variable aleatoria con media $\mathbb{E}(Y) = \mu$, la *varianza* de una variable aleatoria Y se define como el valor esperado de $(Y - \mu)^2$. Esto es:

$$V(Y) = \mathbb{E}[(Y - \mu)^2] = \mathbb{E}(Y^2) - \mu^2$$

La *desviación estándar* de Y es la raíz cuadrada positiva de $V(Y)$.

2.4. Distribuciones de probabilidad para variables aleatorias discretas

2.4.1. Distribución Binomial

Un *experimento binomial* presenta las siguientes propiedades:

1. Consiste en un número fijo n de pruebas idénticas.
2. Cada prueba resulta en uno de dos resultados: éxito s o fracaso f .
3. La probabilidad de éxito en una sola prueba es igual a algún valor p y es el mismo de una prueba a otra. La probabilidad de fracaso es igual a $q = (1 - p)$.
4. Las pruebas son independientes.
5. La variable de interés es Y , el número de éxitos observado durante las n pruebas.

Habiendo descrito esto, se dice que una variable aleatoria Y tiene una *distribución binomial* basada en n pruebas con probabilidad de éxito $p \in [0, 1]$ si y solo si:

$$p(y) = \binom{n}{y} p^y (1 - p)^{n-y} \quad y = 0, 1, 2, \dots, n$$

Para esta distribución, $\mu = \mathbb{E}(Y) = np$ y $V(Y) = np(1 - p)$.

2.4.2. Distribución geométrica

Se dice que una variable aleatoria Y tiene una *distribución geométrica* si y solo si:

$$p(y) = (1 - p)^{y-1}p \quad y = 1, 2, \dots$$

En este caso, $\mu = \mathbb{E}(Y) = \frac{1}{p}$ y $V(Y) = \frac{1-p}{p^2}$.

Para esta distribución, la probabilidad de éxito es igual a p y es constante de una prueba a otra. A diferencia de la distribución binomial, la variable aleatoria geométrica Y es el número de prueba en que ocurre el primer éxito. Es decir, el experimento consiste en una serie de pruebas que concluye con el primer éxito.

2.4.3. Distribución binomial negativa

Se dice que una variable aleatoria Y tiene una *distribución binomial negativa* si y solo si:

$$p(y) = \binom{y-1}{r-1} p^r (1-p)^{y-r} \quad y = r, r+1, r+2, \dots$$

En este caso, $\mu = \mathbb{E}(Y) = \frac{r}{p}$ y $V(Y) = \frac{r(1-p)}{p^2}$.

La distribución geométrica maneja el caso donde estamos interesados en el número de intento en que ocurre el primer éxito. ¿Qué pasa si estamos interesados en el número de intento en que ocurre el segundo, tercer o cuarto éxito?. La distribución binomial negativa es la distribución que se aplica a la variable aleatoria Y igual al número de intento en que ocurre el r -ésimo éxito.

2.4.4. Distribución de Poisson

Se dice que una variable aleatoria Y tiene una *distribución de Poisson* si y solo si

$$p(y) = \frac{\lambda^y}{y!} e^{-\lambda} \quad y = 0, 1, 2, \dots \quad \lambda > 0$$

En este caso, $\mu = \mathbb{E}(Y) = \lambda$ y $V(Y) = \lambda$.

2.5. Momentos y funciones generadoras de momentos

2.5.1. Momento k -ésimo

El k -ésimo momento de una variable aleatoria Y tomada alrededor del origen se define como $\mathbb{E}(Y^k)$ y se denota como μ'_k .

El k -ésimo momento de una variable aleatoria Y tomada alrededor de su media o k -ésimo momento central de Y se define como $\mathbb{E}[(Y - \mu)^k]$ y se denota como μ_k .

2.5.2. Función generadora de momento

La *función generadora de momento* $m(t)$ para una variable aleatoria Y se define como :

$$m(t) = \mathbb{E}(e^{tY})$$

Se dice que una función generadora de momento para Y existe si existe una constante $b > 0$ tal que $m(t)$ es finita para $|t| < b$. Además, si $m(t)$ existe, entonces para cualquier entero positivo k :

$$\left. \frac{d^k m(t)}{dt^k} \right|_{t=0} = m^{[k]}(0) = \mu'_k$$

3. Variables continuas y sus distribuciones de probabilidad

3.1. Distribución de probabilidad para una variable continua

3.1.1. Variable aleatoria continua

Una variable aleatoria Y se dice *continua* si puede tomar cualquier valor en un intervalo.

3.1.2. Función de distribución para variables aleatorias continuas

Sea Y una variable aleatoria. La *función de distribución* de Y denotada por $F(y)$, es tal que

$$F(y) = \mathbb{P}(Y \leq y) \quad -\infty < y < \infty$$

Las funciones de distribución cumplen las siguientes propiedades:

- $F(-\infty) = \lim_{y \rightarrow -\infty} F(y) = 0$
- $F(\infty) = \lim_{y \rightarrow \infty} F(y) = 1$
- $F(y)$ es una función *no decreciente* de y
- $F(y)$ se dice *continua* si es continua $\forall y \in \mathbb{R}$
- Si y es una variable aleatoria continua, entonces $\forall y \in \mathbb{R}$:

$$\mathbb{P}(Y = y) = 0$$

3.1.3. Función densidad de probabilidad

Sea $F(y)$ una función de distribución para una variable aleatoria continua Y . Se denomina *función de densidad de probabilidad* a la función:

$$f(y) = \frac{dF(y)}{dy} = F'(y)$$

siempre que exista esta derivada. Se tiene entonces que:

$$F(y) = \int_{-\infty}^y f(t)dt$$

Las funciones de densidad satisfacen:

- $f(y) \geq 0 \quad \forall y \in \mathbb{R}$
- $\int_{-\infty}^{\infty} f(y)dy = 1$

3.1.4. Funciones de densidad y probabilidades

Si la variable aleatoria continua Y tiene función de densidad $f(y)$ y $a < b$, entonces:

$$\mathbb{P}(a \leq y \leq b) = \int_a^b f(y)dy$$

3.2. Valor esperado y varianza (caso continuo)

3.2.1. Valor esperado

El valor esperado de una variable aleatoria continua Y es:

$$\mathbb{E}(Y) = \int_{-\infty}^{\infty} yf(y)dy$$

siempre que la integral exista. El valor esperado satisface además:

- $\mathbb{E}(c) = c$ con $c \in \mathbb{R}$
- $\mathbb{E}[g(Y)] = \int_{-\infty}^{\infty} g(y)f(y)dy$
- $\mathbb{E}[c \cdot g(Y)] = c \cdot \mathbb{E}[g(Y)]$
- $\mathbb{E}\left[\sum_{i=1}^k g_i(Y)\right] = \sum_{i=1}^k \mathbb{E}[g_i(Y)]$

3.2.2. Varianza

Si Y es una variable aleatoria continua con media $\mathbb{E}(Y) = \mu$, la *varianza* de una variable aleatoria Y se define como el valor esperado de $(Y - \mu)^2$. Esto es:

$$V(Y) = \mathbb{E}[(Y - \mu)^2] = \mathbb{E}(Y^2) - \mu^2$$

La *desviación estándar* de Y es la raíz cuadrada positiva de $V(Y)$.

3.3. Algunas distribuciones de probabilidad para variables continuas

3.3.1. Distribución uniforme

Si $\theta_1 < \theta_2$, se dice que una variable aleatoria Y tiene *distribución de probabilidad uniforme* en el intervalo (θ_1, θ_2) si y solo si la función de densidad de Y es:

$$f(y) = \begin{cases} \frac{1}{\theta_2 - \theta_1} & \theta_1 \leq y \leq \theta_2 \\ 0 & \text{si no} \end{cases}$$

Las constantes que determinan la forma específica de una función de densidad se llaman *parámetros* de la función de densidad.

Para una variable aleatoria uniformemente distribuida en el intervalo (θ_1, θ_2) , se tiene que

$$\mu = \mathbb{E}(Y) = \frac{\theta_1 + \theta_2}{2} \quad \text{y} \quad \sigma^2 = V(Y) = \frac{(\theta_2 - \theta_1)^2}{12}$$

3.3.2. Distribución de probabilidad normal

Se dice que una variable aleatoria Y tiene una *distribución normal de probabilidad* si y sólo si, para $\sigma > 0$ y $-\infty < \mu < \infty$ la función de densidad de Y es:

$$f(y) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(y-\mu)^2/(2\sigma^2)} \quad -\infty < y < \infty$$

Para una variable aleatoria Y normalmente distribuida de parámetros μ y σ , se tiene que:

$$\mu = \mathbb{E}(Y) = \mu \quad \text{y} \quad V(Y) = \sigma^2$$

3.3.3. Distribución de probabilidad gamma

Se dice que una variable aleatoria Y tiene una *distribución gamma* con parámetros $\alpha, \beta > 0$ si y sólo si la función de densidad de Y es:

$$f(y) = \begin{cases} \frac{y^{\alpha-1} e^{-y/\beta}}{\beta^\alpha \Gamma(\alpha)} & 0 \leq y < \infty \\ 0 & \text{si no} \end{cases}$$

donde $\Gamma(\alpha) = \int_0^\infty y^{\alpha-1} e^{-y} dy$. Se tiene además que:

$$\mu = \mathbb{E}(Y) = \alpha\beta \quad y \quad \sigma^2 = V(Y) = \alpha\beta^2$$

Dado v un entero positivo, se dice que una variable aleatoria tiene *distribución χ^2 (ji cuadrado)* con v grados de libertad si y sólo si Y es una variable aleatoria con distribución gamma de parámetros $\alpha = \frac{v}{2}$ y $\beta = 2$.

3.3.4. Distribución de probabilidad exponencial

Se dice que una variable aleatoria Y tiene una *distribución exponencial* con parámetro $\beta > 0$ si y sólo si la función de densidad de Y es:

$$f(y) = \begin{cases} \frac{1}{\beta} e^{-y/\beta} & 0 \leq y < \infty \\ 0 & \text{si no} \end{cases}$$

Notar que esta distribución es un caso particular de la distribución gamma tomando $\alpha = 1$. Es claro entonces que:

$$\mu = \mathbb{E}(Y) = \beta \quad y \quad \sigma^2 = V(Y) = \beta^2$$

3.3.5. Distribución de probabilidad beta

Se dice que una variable aleatoria Y tiene una *distribución beta* con parámetros $\alpha, \beta > 0$ si y sólo si la función de densidad de Y es:

$$f(y) = \begin{cases} \frac{y^{\alpha-1} (1-y)^{\beta-1}}{B(\alpha, \beta)} & 0 \leq y < 1 \\ 0 & \text{si no} \end{cases}$$

donde $B(\alpha, \beta) = \int_0^1 y^{\alpha-1} (1-y)^{\beta-1} dy = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$. Se tiene además que:

$$\mu = \mathbb{E}(Y) = \frac{\alpha}{\alpha+\beta} \quad y \quad \sigma^2 = V(Y) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

3.4. Momentos, funciones generadoras de momentos y teorema de Tchebysheff

Observación: Las definiciones 3.4.1 y 3.4.2 son las mismas de la sección 2.5.

3.4.1. Momento k -ésimo

El k -ésimo momento de una variable aleatoria Y tomada alrededor del origen se define como $\mathbb{E}(Y^k)$ y se denota como μ'_k .

El k -ésimo momento de una variable aleatoria Y tomada alrededor de su media o k -ésimo momento central de Y se define como $\mathbb{E}[(Y - \mu)^k]$ y se denota como μ_k .

3.4.2. Función generadora de momento

La función generadora de momento $m(t)$ para una variable aleatoria Y se define como :

$$m(t) = \mathbb{E}(e^{tY})$$

Se dice que una función generadora de momento para Y existe si existe una constante $b > 0$ tal que $m(t)$ es finita para $|t| < b$. Además, si $m(t)$ existe, entonces para cualquier entero positivo k :

$$\left. \frac{d^k m(t)}{dt^k} \right|_{t=0} = m^{[k]}(0) = \mu'_k$$

Dada Y una variable aleatoria con función de densidad $f(y)$ y $g(Y)$ una función de Y . Entonces, la función generadora de momento para $g(Y)$ es:

$$\mathbb{E} [e^{tg(Y)}] = \int_{-\infty}^{\infty} e^{tg(Y)} f(y) dy$$

3.4.3. Teorema de Tchebysheff

Sea Y una variable aleatoria con media finita μ y varianza σ^2 , Entonces, para cualquier $k > 0$:

$$\mathbb{P}(|Y - \mu| < k\sigma) \geq 1 - \frac{1}{k^2} \quad \text{o} \quad \mathbb{P}(|Y - \mu| \geq k\sigma) \leq \frac{1}{k^2}$$

4. Distribuciones de probabilidad multivariantes

4.1. Distribuciones de probabilidad bivariantes y multivariantes

4.1.1. Función de probabilidad conjunta

Sean Y_1, Y_2 variables aleatorias discretas. La función de probabilidad conjunta para Y_1, Y_2 está dada por

$$p(y_1, y_2) = \mathbb{P}(Y_1 = y_1, Y_2 = y_2) \quad y_1, y_2 \in \mathbb{R}$$

Si Y_1, Y_2 son variables aleatorias discretas con función de probabilidad conjunta $p(y_1, y_2)$, entonces

- $p(y_1, y_2) \geq 0$ para toda y_1, y_2 .
- $\sum_{y_1, y_2} p(y_1, y_2) = 1$ para todos los valores (y_1, y_2) con probabilidad distinta de cero.

4.1.2. Función de distribución conjunta

Para cualesquiera variables aleatorias Y_1, Y_2 , la *función de distribución conjunta* (o *bivariante*) es:

$$F(y_1, y_2) = \mathbb{P}(Y_1 \leq y_1, Y_2 \leq y_2) \quad y_1, y_2 \in \mathbb{R}$$

4.1.3. Variables aleatorias conjuntas y densidad conjunta

Sean Y_1, Y_2 variables aleatorias conjuntas con distribución de distribución conjunta $F(y_1, y_2)$. Si existe una función no negativa $f(y_1, y_2)$ tal que:

$$F(y_1, y_2) = \int_{-\infty}^{y_1} \int_{-\infty}^{y_2} f(t_1, t_2) dt_1 dt_2$$

para $y_1, y_2 \in \mathbb{R}$, entonces, se dice que Y_1 e Y_2 son *variables aleatorias continuas conjuntas*. La función $f(y_1, y_2)$ recibe el nombre de *función de densidad de probabilidad conjunta*.

Si Y_1, Y_2 son variables aleatorias con función de distribución conjunta $F(y_1, y_2)$, entonces:

- $F(-\infty, -\infty) = F(-\infty, y_2) = F(y_1, -\infty) = 0$
- $F(\infty, \infty) = 1$
- Si $y_1^* \geq y_1$ e $y_2^* \geq y_2 \implies F(y_1^*, y_2^*) + F(y_1, y_2) \geq F(y_1^*, y_2) + F(y_1, y_2^*)$

Si Y_1, Y_2 son variables aleatorias con función de densidad conjunta $f(y_1, y_2)$, entonces:

- $f(y_1, y_2) \geq 0$
- $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(y_1, y_2) dy_1 dy_2 = 1$

4.2. Funciones de probabilidad y densidad marginal

4.2.1. Funciones de probabilidad marginal

Sean Y_1, Y_2 variables aleatorias discretas conjuntas con función de probabilidad $p(y_1, y_2)$. Entonces, las *funciones de probabilidad marginal* de Y_1 e Y_2 respectivamente, están dadas por:

$$p_1(y_1) = \sum_{y_2} p(y_1, y_2) \quad p_2(y_2) = \sum_{y_1} p(y_1, y_2)$$

4.2.2. Funciones de densidad marginal

Sean Y_1, Y_2 variables aleatorias continuas conjuntas con función de densidad conjunta $f(y_1, y_2)$. Entonces, las *funciones de densidad marginal* de Y_1 e Y_2 respectivamente, están dadas por:

$$f_1(y_1) = \int_{-\infty}^{\infty} f(y_1, y_2) dy_2 \quad f_2(y_2) = \int_{-\infty}^{\infty} f(y_1, y_2) dy_1$$

4.3. Funciones de probabilidad discreta, distribución y densidad condicional

4.3.1. Función de probabilidad discreta condicional

Si Y_1, Y_2 son variables aleatorias discretas conjuntas con función de probabilidad conjunta $p(y_1, y_2)$ y funciones de probabilidad marginal $p_1(y_1)$ y $p_2(y_2)$ respectivamente, entonces la *función de probabilidad discreta condicional* de Y_1 dada Y_2 es:

$$p(y_1|y_2) = \mathbb{P}(Y_1 = y_1|Y_2 = y_2) = \frac{\mathbb{P}(Y_1 = y_1, Y_2 = y_2)}{\mathbb{P}(Y_2 = y_2)} = \frac{p(y_1, y_2)}{p_2(y_2)}$$

4.3.2. Función de distribución condicional

Si Y_1, Y_2 son variables aleatorias continuas con función de densidad conjunta $f(y_1, y_2)$, entonces, la función de distribución condicional de Y_1 dado $Y_2 = y_2$ es:

$$F(y_1|y_2) = \mathbb{P}(Y_1 \leq y_1|Y_2 = y_2)$$

4.3.3. Densidad condicional

Sean Y_1, Y_2 variables aleatorias continuas conjuntas con densidad conjunta $f(y_1, y_2)$ y densidades marginales $f_1(y_1, y_2)$ y $f_2(y_1, y_2)$ respectivamente. Para cualquier y_2 tal que $f_2(y_2) > 0$, la *densidad condicional* de Y_1 dado $Y_2 = y_2$ es:

$$f(y_1|y_2) = \frac{f(y_1, y_2)}{f_2(y_2)}$$

y para cualquier y_1 tal que $f_1(y_1) > 0$, la *densidad condicional* de Y_2 dado $Y_1 = y_1$ es:

$$f(y_2|y_1) = \frac{f(y_1, y_2)}{f_1(y_1)}$$

4.4. Variables aleatorias independientes y valores esperados

4.4.1. Independencia de variables

A continuación, se enuncian una serie de caracterizaciones que sirven para demostrar independencia de variables aleatorias.

- Sean Y_1, Y_2 variables aleatorias con funciones de distribución $F_1(y_1), F_2(y_2)$ y función de distribución conjunta $F(y_1, y_2)$. Entonces, se dice que Y_1 e Y_2 son *independientes* si y sólo si $\forall y_1, y_2$:

$$F(y_1, y_2) = F_1(y_1)F_2(y_2)$$

- Si Y_1, Y_2 son variables aleatorias discretas con función de probabilidad conjunta $p(y_1, y_2)$ y funciones de probabilidad marginal $p_1(y_1)$ y $p_2(y_2)$, entonces, Y_1 e Y_2 son *independientes* si y sólo si $\forall y_1, y_2$:

$$p(y_1, y_2) = p_1(y_1)p_2(y_2)$$

- Si Y_1, Y_2 son variables aleatorias continuas con función de densidad conjunta $f(y_1, y_2)$ y funciones de densidad marginal $f_1(y_1)$ y $f_2(y_2)$, entonces, Y_1 e Y_2 son *independientes* si y sólo si $\forall y_1, y_2$:

$$f(y_1, y_2) = f_1(y_1)f_2(y_2)$$

- Sean Y_1, Y_2 con densidad conjunta $f(y_1, y_2)$ tal que es positiva si $(y_1, y_2) \in [a, b] \times [c, d]$ con $a, b, c, d \in \mathbb{R}$; y $f(y_1, y_2) = 0$ en otro caso. Entonces, Y_1 e Y_2 son variables aleatorias *independientes* si y sólo si:

$$f(y_1, y_2) = g(y_1)h(y_2)$$

donde $g(y_1), h(y_2)$ son funciones no negativas de y_1 e y_2 respectivamente.

4.4.2. Valores esperados multivariantes

Sea $g(y_1, y_2, \dots, y_k)$ función de variables aleatorias Y_1, Y_2, \dots, Y_k .

- Si Y_1, \dots, Y_k son variables aleatorias discretas con función de probabilidad $p(y_1, \dots, y_k)$. Entonces, el *valor esperado* de $g(Y_1, \dots, Y_k)$ es:

$$\mathbb{E}[g(Y_1, \dots, Y_k)] = \sum_{y_k} \cdots \sum_{y_1} g(y_1, \dots, y_k) p(y_1, \dots, y_k)$$

- Si Y_1, \dots, Y_k son variables aleatorias continuas con función de densidad $f(y_1, \dots, y_k)$. Entonces, el *valor esperado* de $g(Y_1, \dots, Y_k)$ es:

$$\mathbb{E}[g(Y_1, \dots, Y_k)] = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(y_1, \dots, y_k) \times f(y_1, \dots, y_k) dy_1 \cdots dy_k$$

- Si Y_1, Y_2 son variables aleatorias independientes, y $g(Y_1), h(Y_2)$ funciones solo de Y_1 e Y_2 respectivamente, entonces

$$\mathbb{E}[g(Y_1)h(Y_2)] = \mathbb{E}[g(Y_1)] \mathbb{E}[h(Y_2)]$$

4.4.3. Covarianza

Si Y_1, Y_2 son variables aleatorias con medias μ_1 y μ_2 respectivamente. La *covarianza* de Y_1 e Y_2 es:

$$\text{Cov}(Y_1, Y_2) = \mathbb{E}[(Y_1 - \mu_1)(Y_2 - \mu_2)] = \mathbb{E}(Y_1 Y_2) - \mathbb{E}(Y_1)\mathbb{E}(Y_2)$$

Si Y_1, Y_2 son independientes, entonces $\text{Cov}(Y_1, Y_2) = 0$.

Si $\text{Cov}(Y_1, Y_2) = 0$, entonces las variables son *no correlacionadas* y no hay dependencia lineal entre Y_1 e Y_2 .

4.4.4. Coeficiente de correlación

Se define el *coeficiente de correlación* como:

$$\rho = \frac{\text{Cov}(Y_1, Y_2)}{\sigma_1 \sigma_2}$$

donde σ_1, σ_2 son las desviaciones estándar de Y_1 e Y_2 respectivamente.

4.4.5. Valor esperado y varianza de funciones lineales de variables aleatorias

Sean Y_1, \dots, Y_n y X_1, \dots, X_m variables aleatorias con $\mathbb{E}(Y_i) = \mu_i$ y $\mathbb{E}(X_j) = \xi_j$. Definiendo:

$$U_1 = \sum_{i=1}^n a_i Y_i \quad \text{y} \quad U_2 = \sum_{j=1}^m b_j X_j$$

para las constantes a_1, \dots, a_n y b_1, \dots, b_m . Entonces, se cumple lo siguiente:

- $\mathbb{E}(U_1) = \sum_{i=1}^n a_i \mu_i$
- $V(U_1) = \sum_{i=1}^n a_i^2 V(Y_i) + 2 \sum_{1 \leq i < j \leq n} a_i a_j \text{Cov}(Y_i, Y_j)$
- $\text{Cov}(U_1, U_2) = \sum_{i=1}^n \sum_{j=1}^m a_i b_j \text{Cov}(Y_i, X_j)$

4.4.6. Valores esperados condicionales

Si Y_1, Y_2 son variables aleatorias, el valor esperado condicional de $g(Y_1)$ dado $Y_2 = y_2$ está dado como:

$$\mathbb{E}[(g(Y_1)|Y_2 = y_2)] = \sum_{y_1} g(y_1)p(y_1|y_2) \quad \text{y} \quad \mathbb{E}[(g(Y_1)|Y_2 = y_2)] = \int_{-\infty}^{\infty} g(y_1)f(y_1|y_2)dy_1$$

para el caso discreto conjunto y continuo conjunto respectivamente.

Si Y_1, Y_2 son dos variables aleatorias, entonces:

$$\mathbb{E}(Y_1) = \mathbb{E}[\mathbb{E}(Y_1|Y_2)]$$

donde en el lado derecho de la ecuación el valor esperado interior es con respecto a la distribución condicional de Y_1 dada Y_2 y el valor esperado exterior es con respecto a la distribución de Y_2 .

5. Funciones de variables aleatorias

5.1. Métodos para determinar funciones de probabilidad

Considere Y_1, \dots, Y_n variables aleatorias y una función $U(Y_1, \dots, Y_n)$. Para determinar la función de probabilidad de U , se pueden utilizar los siguientes métodos:

5.1.1. Método de las funciones de distribución

1. Localizar la región $U = u$ del espacio (y_1, \dots, y_n) .
2. Localizar la región del espacio $U \leq u$.
3. Determinar $F_U(u) = \mathbb{P}(U \leq u)$ al integrar $f(y_1, \dots, y_n)$ sobre la región $U \leq u$.
4. Determinar $f_U(u)$ al derivar $F_U(u)$.

5.1.2. Método de las funciones generadoras de momento

Antes de describir este método, se presentan algunos teoremas que pueden ser útiles a la hora de emplear dicho método. Sean $\mu_i, \sigma_i^2, m_{Y_i}(t)$ el valor esperado, la varianza y la función generadora de momento de Y_i para cada $i = 1, \dots, n$ respectivamente. Entonces:

- Si $m_{Y_1}(t) = m_{Y_2}(t)$ para todo t , entonces Y_1 e Y_2 tienen la misma distribución de probabilidad.
- Si $U = \sum_{i=1}^n Y_i$, entonces:

$$m_U(t) = m_{Y_1}(t) \cdot \dots \cdot m_{Y_n}(t)$$

- Si Y_1, \dots, Y_n son variables aleatorias independientes normalmente distribuidas y definiendo $Z_i = \frac{Y_i - \mu_i}{\sigma_i}$ para $i = 1, \dots, n$, entonces $\sum_{i=1}^n Z_i^2$ tiene una distribución χ^2 con n grados de libertad.

El método de las funciones generadores de momento consiste en:

1. Encontrar la función generadora de momento para U , $m_U(t)$.
2. Comparar $m_U(t)$ con otras funciones de momento conocidas.

6. Distribuciones muestrales y teorema central del límite

6.1. Distribuciones muestrales relacionadas con la distribución normal

6.1.1. Estadístico

Un *estadístico* es una función de las variables aleatorias observables en una muestra y de constantes conocidas.

6.1.2. Distribución de la media muestral

Sea Y_1, \dots, Y_n una muestra aleatoria de tamaño n de una distribución normal, con media μ y varianza σ^2 . Entonces

$$\bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i$$

está distribuida normalmente con media $\mu_{\bar{Y}}$ y varianza $\sigma_{\bar{Y}}^2 = \frac{\sigma^2}{n}$.

6.1.3. Distribución de variables aleatorias normales estándar e independientes

Sea Y_1, \dots, Y_n una muestra aleatoria de tamaño n de una distribución normal, con media μ y varianza σ^2 . Entonces, $Z_i = \frac{(Y_i - \mu_i)}{\sigma}$ son variables aleatorias normales estándar e independientes, y

$$\sum_{i=1}^n Z_i^2 = \frac{1}{\sigma^2} \sum_{i=1}^n (Y_i - \mu_i)^2$$

tienen una distribución χ^2 con n grados de libertad.

6.1.4. Distribución de la varianza muestral

Sea Y_1, \dots, Y_n una muestra aleatoria de una distribución normal. Se define la *varianza central* como:

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (Y_i - \bar{Y})^2$$

Si Y_1, \dots, Y_n tiene media μ y varianza σ^2 , entonces

$$\frac{(n-1)S^2}{\sigma^2} = \frac{1}{\sigma^2} \sum_{i=1}^n (Y_i - \bar{Y})^2$$

tiene una distribución χ^2 con $(n-1)$ grados de libertad. También, \bar{Y} y S^2 son variables aleatorias independientes.

6.1.5. Distribución t de Student

Sea Z una variable aleatoria normal estándar y sea W una variable con distribución χ^2 con ν grados de libertad. Entonces, si Z, W son independientes:

$$T = \frac{Z}{\sqrt{W/\nu}}$$

tiene *distribución t de Student* con ν grados de libertad.

6.1.6. Distribución F de Fisher-Snedecor

Sean W_1, W_2 variables aleatorias independientes con distribución χ^2 y con ν_1, ν_2 grados de libertad respectivamente. Entonces, se dice que

$$F = \frac{W_1/\nu_1}{W_2/\nu_2}$$

tiene *distribución F de Fisher-Snedecor* con ν_1 grados de libertad en el numerador y ν_2 grados de libertad en el denominador.

6.2. Teorema del Límite central

Sean Y_1, \dots, Y_n variables aleatorias independientes y distribuidas idénticamente con $\mathbb{E}(Y_i) = \mu$ y $V(Y_i) = \sigma^2 < \infty$. Se define

$$U_n = \frac{\sum_{i=1}^n Y_i - n\mu}{\sigma\sqrt{n}} = \frac{\bar{Y} - \mu}{\sigma/\sqrt{n}}$$

Entonces, la función de distribución U_n converge hacia la función de distribución normal estándar cuando $n \rightarrow \infty$. Esto es:

$$\lim_{n \rightarrow \infty} \mathbb{P}(U_n \leq u) = \int_{-\infty}^u \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$$

7. Estimación

7.1. Estimadores, sesgo y error cuadrático medio

7.1.1. Estimador

Un *estimador* es una regla que indica cómo calcular el valor de una estimación con base en las mediciones contenidas en una muestra.

7.1.2. Sesgo de un estimador puntual

Dado un estimador puntual $\hat{\theta}$ de un parámetro θ , se define el *sesgo* de $\hat{\theta}$ como:

$$B(\hat{\theta}) = \mathbb{E}(\hat{\theta}) - \theta$$

7.1.3. Error cuadrático medio de un estimador puntual

El error cuadrático medio de un estimador puntual $\hat{\theta}$ es:

$$\text{MSE}(\hat{\theta}) = \mathbb{E}[(\hat{\theta} - \theta)^2] = V(\hat{\theta}) + [B(\hat{\theta})]^2$$

7.1.4. Error estándar de un estimador

La desviación estándar del estimador $\hat{\theta}$, $\sigma_{\hat{\theta}} = \sqrt{\sigma_{\hat{\theta}}^2}$ suele recibir el nombre de *error cuadrático estándar*.

7.2. Estimadores puntuales comunes y consistencia de estimadores

7.2.1. Estimadores puntuales comunes

Los estimadores puntuales más utilizados son :

Parámetro objetivo θ	Tamaño(s) muestral(es)	Estimador puntual $\hat{\theta}$	$\mathbb{E}(\hat{\theta})$	Error estándar $\sigma_{\hat{\theta}}$
μ	n	\bar{Y}	μ	$\frac{\sigma}{\sqrt{n}}$
p	n	$\hat{p} = \frac{Y}{n}$	p	$\sqrt{\frac{p(1-p)}{n}}$
$\mu_1 - \mu_2$	n_1, n_2	$\bar{Y}_1 - \bar{Y}_2$	$\mu_1 - \mu_2$	$\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$
$p_1 - p_2$	n_1, n_2	$\hat{p}_1 - \hat{p}_2$	$p_1 - p_2$	$\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$

donde σ_i^2 es la varianza de la población i , y las muestras son independientes.

7.2.2. Consistencia de estimadores

Se dice que el estimador $\hat{\theta}_n$ es un *estimador consistente* de θ si, para cualquier $\varepsilon > 0$

$$\lim_{n \rightarrow \infty} \mathbb{P}(|\hat{\theta}_n - \theta| \leq \varepsilon) = 1$$

o bien, de forma equivalente

$$\lim_{n \rightarrow \infty} \mathbb{P}(|\hat{\theta}_n - \theta| > \varepsilon) = 0$$

Un estimador insesgado $\hat{\theta}_n$ para θ es un estimador consistente de θ si

$$\lim_{n \rightarrow \infty} V(\hat{\theta}_n) = 0$$

7.2.3. Ley de los grandes números

Sean Y_1, \dots, Y_n variables aleatorias independientes con $\mathbb{E}(Y_i) = \mu$ y $V(Y_i) = \sigma^2 < \infty$. Entonces, $\hat{\mu}_n = \bar{Y}_n = \frac{1}{n} \sum_{i=1}^n Y_i$ es un estimador consistente de μ . Es decir $\forall \varepsilon > 0$:

$$\lim_{n \rightarrow \infty} \mathbb{P}\left(\left|\frac{1}{n} \sum_{i=1}^n Y_i - \mu\right| \leq \varepsilon\right) = 1$$

7.3. Método de máxima verosimilitud

7.3.1. Verosimilitud de la muestra

Sean y_1, \dots, y_n observaciones muestrales tomadas de variables aleatorias correspondientes Y_1, \dots, Y_n cuya distribución depende de un parámetro θ . Entonces, si Y_1, \dots, Y_n son variables aleatorias discretas, la *verosimilitud de la muestra*, $L(y_1, \dots, y_n|\theta)$ se define como la probabilidad conjunta de y_1, \dots, y_n . Si Y_1, \dots, Y_n son variables aleatorias continuas, entonces $L(y_1, \dots, y_n|\theta)$ se define como la densidad conjunta evaluada en y_1, \dots, y_n .

7.3.2. Método de máxima verosimilitud

Suponga que la función de verosimilitud depende de k parámetros $\theta_1, \dots, \theta_k$. Escoja como estimadores los valores de los parámetros que maximicen la verosimilitud $L(y_1, \dots, y_n|\theta_1, \dots, \theta_k)$.

7.4. Intervalos de confianza

7.4.1. Intervalo de confianza y límites de confianza

Un *estimador de confianza* o *intervalo de confianza* es una regla que especifica el método para usar las mediciones muestrales en el cálculo de dos números que forman los puntos extremos del intervalo. Estos puntos extremos se denominan *límites de confianza superior e inferior*, respectivamente.

7.4.2. Coeficiente de confianza

La probabilidad de que un intervalo de confianza contenga a θ (una cantidad fija) se llama *coeficiente de confianza*.

Dados $\hat{\theta}_L, \hat{\theta}_U$ límites de confianza (aleatorios) inferior y superior respectivamente, para un parámetro θ . Entonces, si

$$\mathbb{P}(\hat{\theta}_L \leq \theta \leq \hat{\theta}_U) = 1 - \alpha$$

la probabilidad $(1 - \alpha)$ es el coeficiente de confianza.

7.4.3. Método del Pivote

Un método muy útil para encontrar intervalos de confianza se llama *método del pivote*. Éste consiste en determinar una cantidad que actúe como pivote y que posea las siguientes características:

1. Que sea una función de las medias muestrales y el parámetro desconocido θ , donde θ sea la única cantidad desconocida.
2. Que su distribución de probabilidad no dependa de θ .

La probabilidad del evento $\mathbb{P}(a \leq Y \leq b)$ no se ve afectada por un cambio de escala o traslación de Y . Por lo tanto, si se conoce la distribución de probabilidad de una cantidad pivote, se pueden utilizar operaciones como las señaladas anteriormente para formar la estimación por intervalos buscada.

7.4.4. Intervalos de confianza en una muestra grande

Para muestras grandes

$$Z = \frac{\hat{\theta} - \theta}{\sigma_{\hat{\theta}}}$$

posee aproximadamente una distribución normal estándar. En consecuencia, Z forma (al menos aproximadamente) una cantidad pivote y el método del pivote se puede emplear para desarrollar intervalos de confianza para el parámetro objetivo θ .

8. Test de hipótesis

8.1. Elementos de una prueba estadística

8.1.1. Hipótesis de investigación

Una *hipótesis de investigación* es una teoría acerca del o los parámetros que se desean apoyar.

8.1.2. Hipótesis nula e hipótesis alternativa

Se denomina *hipótesis nula* H_0 a la hipótesis que se desea contrastar. Su nombre sugiere que H_0 debe identificarse con la hipótesis de no cambio. La *hipótesis alternativa* H_a , es la que, por lo general, se quiere comprobar con base en la información contenida en la muestra.

8.1.3. Estadístico de prueba y región de rechazo

El *estadístico de prueba* es una función de las mediciones muestrales en la que la decisión estadística estará basada.

La *región de rechazo* RR, especifica los valores del estadístico de prueba para el cual la hipótesis nula ha de ser rechazada a favor de la hipótesis alternativa. Si, para una muestra particular, el valor calculado del estadístico de prueba cae en la región de rechazo RR, se rechaza la hipótesis nula H_0 y se acepta la hipótesis alternativa H_a . Si el valor del estadístico de prueba no cae en la región de rechazo, se acepta H_0 .

8.1.4. Errores tipo I y II

Se comete un *error del tipo I* si H_0 es rechazada cuando H_0 es verdadera. La *probabilidad de un error tipo I* está denotada por α y se denomina *nivel de la prueba*.

Se comete un *error del tipo II* si H_0 es aceptada cuando H_a es verdadera. La *probabilidad de un error tipo II* está denotada por β .

Dadas regiones de rechazo RR_1, RR_2 tales que $RR_1 \subseteq RR_2$ (es decir, la región 2 es más grande que la 1) con sus respectivos α_i, β_i . Entonces:

$$\alpha_1 \leq \alpha_2 \quad \text{y} \quad \beta_1 \geq \beta_2$$

8.2. Pruebas comunes con muestras grandes

8.2.1. Pruebas de muestras grandes

Dado un estimador $\hat{\theta}$, si θ_0 es un valor específico de θ , se puede probar $H_0 : \theta = \theta_0$ contra $H_a : \theta > \theta_0$. Si $\hat{\theta}$ es cercana a θ_0 , parece razonable aceptar H_0 . Pero, si en realidad $\theta > \theta_0$, es más probable que $\hat{\theta}$ sea más grande. En consecuencia, valores grandes de $\hat{\theta}$ (mayores a θ_0 en una cantidad apropiada) favorecen el rechazo de $H_0 : \theta = \theta_0$ y una aceptación de $H_a : \theta > \theta_0$.

$$\begin{aligned} H_0 &: \theta = \theta_0 \\ H_a &: \theta > \theta_0 \\ \text{Estadístico de prueba} &: \hat{\theta} \\ \text{Región de rechazo} &: RR = \{\theta > k\} \text{ para alguna selección de } k \end{aligned}$$

El valor real de k en la región de rechazo RR se determina al fijar la probabilidad α de error tipo I (el nivel de la prueba) y escoger k de conformidad. Si H_0 es verdadera, $\hat{\theta}$ tiene distribución aproximadamente normal con media θ_0 y error estándar $\sigma_{\hat{\theta}}$. Por lo tanto, si deseamos una prueba de nivel α :

$$k = \theta_0 + z_{\alpha} \sigma_{\hat{\theta}}$$

es la selección apropiada para k (si Z tiene una distribución normal estándar, entonces z_α es tal que $\mathbb{P}(Z > z_\alpha) = \alpha$). Como

$$\text{RR} = \left\{ \hat{\theta} : \hat{\theta} > \theta_0 + z_\alpha \sigma_{\hat{\theta}} \right\} = \left\{ \hat{\theta} : \frac{\hat{\theta} - \theta_0}{\sigma_{\hat{\theta}}} > z_\alpha \right\}$$

si $Z = \frac{\hat{\theta} - \theta_0}{\sigma_{\hat{\theta}}}$ se usa como estadístico de prueba, la región de rechazo se puede escribir como $\text{RR} = \{z > z_\alpha\}$.

8.2.2. Pruebas de hipótesis de nivel α para muestras grandes

La fórmula precedente para Z es sencillamente:

$$Z = \frac{\text{estimador para el parámetro} - \text{valor del parámetro dado por } H_0}{\text{error estándar del estimador}}$$

Dado el estadístico de prueba Z , se pueden realizar las siguientes *pruebas de hipótesis de nivel α para muestras grandes*:

$$H_0 : \theta = \theta_0$$

$$H_a : \begin{cases} \theta > \theta_0 & (\text{alternativa de cola superior}) \\ \theta < \theta_0 & (\text{alternativa de cola inferior}) \\ \theta \neq \theta_0 & (\text{alternativa de dos colas}) \end{cases}$$

$$\text{Estadístico de prueba} : Z = \frac{\hat{\theta} - \theta_0}{\sigma_{\hat{\theta}}}$$

$$\text{Región de rechazo} : \begin{cases} \{z > z_\alpha\} & (\text{RR de cola superior}) \\ \{z < -z_\alpha\} & (\text{RR de cola inferior}) \\ \{|z| > z_{\alpha/2}\} & (\text{RR de dos colas}) \end{cases}$$

En cualquier prueba particular, sólo una de las alternativas citadas H_a es apropiada. Cualquiera sea la hipótesis alternativa escogida, se debe estar seguro de usar la región de rechazo correspondiente.

8.2.3. Tamaño muestral para una prueba de nivel α

Dados α, β deseados, para consultar el tamaño de la muestra para una cola se tiene la siguiente expresión:

$$n = \frac{(z_\alpha + z_\beta)^2 \sigma^2}{(\mu_a - \mu_0)^2}$$

En caso de no conocer σ^2 , se puede estimar.

8.3. Niveles de significancia o valores p

Si W es un estadístico de prueba, el *valor p o nivel de significancia alcanzado*, es el nivel más pequeño de significancia α para el cual la información observada indica que la hipótesis nula debe ser rechazada. Cuanto más pequeño es el valor de p , más fuerte es la evidencia de que la hipótesis nula debe ser rechazada.

Si fuésemos a rechazar H_0 en favor de H_a para valores pequeños de un estadístico de prueba W , por ejemplo RR: $\{w \leq k\}$, el valor p relacionado con un valor observado w_0 de W está dado por:

$$\text{valor de } p = \mathbb{P}(W \leq w_0, \text{ cuando } H_0 \text{ es verdadera})$$

Análogamente, si fuésemos a rechazar H_0 en favor de H_a para valores grandes de W , por ejemplo RR: $\{w \geq k\}$, el valor p relacionado con un valor observado w_0 de W está dado por:

$$\text{valor de } p = \mathbb{P}(W \geq w_0, \text{ cuando } H_0 \text{ es verdadera})$$

8.4. Pruebas de hipótesis comunes

8.4.1. Pruebas de hipótesis referentes a una varianza poblacional

Dada una muestra aleatoria Y_1, \dots, Y_n de una distribución normal con media μ y varianza σ^2 desconocidas. Se pueden aplicar las siguientes pruebas de hipótesis referentes a varianza:

$$H_0 : \sigma^2 = \sigma_0^2$$

$$H_a : \begin{cases} \sigma^2 > \sigma_0^2 & (\text{alternativa de cola superior}) \\ \sigma^2 < \sigma_0^2 & (\text{alternativa de cola inferior}) \\ \sigma \neq \sigma_0^2 & (\text{alternativa de dos colas}) \end{cases}$$

$$\text{Estadístico de prueba} : \chi^2 = \frac{(n-1)S^2}{\sigma_0^2}$$

$$\text{Región de rechazo} : \begin{cases} \{\chi^2 > \chi_{\alpha}^2\} & (\text{RR de cola superior}) \\ \{\chi < -\chi_{\alpha}^2\} & (\text{RR de cola inferior}) \\ \{\chi > \chi_{\alpha/2}^2 \text{ o } \chi < \chi_{1-\alpha/2}^2\} & (\text{RR de dos colas}) \end{cases}$$

8.4.2. Pruebas de hipótesis sobre una muestra pequeña para μ

Sean Y_1, \dots, Y_n una muestra aleatoria de una distribución normal con $\mathbb{E}(Y_i) = \mu$.

$$H_0 : \mu = \mu_0$$

$$H_a : \begin{cases} \mu > \mu_0 & (\text{alternativa de cola superior}) \\ \mu < \mu_0 & (\text{alternativa de cola inferior}) \\ \mu \neq \mu_0 & (\text{alternativa de dos colas}) \end{cases}$$

$$\text{Estadístico de prueba} : T = \frac{\bar{Y} - \mu_0}{S/\sqrt{n}}$$

$$\text{Región de rechazo} : \begin{cases} \{t > t_{\alpha}\} & (\text{RR de cola superior}) \\ \{t < -t_{\alpha}\} & (\text{RR de cola inferior}) \\ \{|t| > t_{\alpha/2}\} & (\text{RR de dos colas}) \end{cases}$$

8.4.3. Pruebas con muestras pequeñas para comparar dos medias poblacionales

Sean dos muestras independientes de distribuciones normales con $\sigma_1^2 = \sigma_2^2$.

$$H_0 : \mu_1 - \mu_2 = D_0$$

$$H_a : \begin{cases} \mu_1 - \mu_2 > D_0 & (\text{alternativa de cola superior}) \\ \mu_1 - \mu_2 < D_0 & (\text{alternativa de cola inferior}) \\ \mu_1 - \mu_2 \neq D_0 & (\text{alternativa de dos colas}) \end{cases}$$

$$\text{Estadístico de prueba} : T = \frac{\bar{Y}_1 - \bar{Y}_2 - D_0}{S_p \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}, \quad \text{donde } S_p = \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}}$$

$$\text{Región de rechazo} : \begin{cases} \{t > t_\alpha\} & (\text{RR de cola superior}) \\ \{t < -t_\alpha\} & (\text{RR de cola inferior}) \\ \{|t| > t_{\alpha/2}\} & (\text{RR de dos colas}) \end{cases}$$

Aquí, $\mathbb{P}(T > t_\alpha) = \alpha$ y grados de libertad $\nu = n_1 + n_2 - 2$.

8.5. Potencia de las pruebas y lema de Neyman-Person

8.5.1. Potencia de la prueba

Sea W estadístico de prueba y RR es la región de rechazo para una prueba de una hipótesis que involucra el valor de un parámetro θ . Entonces, la potencia de una prueba, denotada por $\text{potencia}(\theta)$, es la probabilidad de que la prueba lleve al rechazo de H_0 cuando el valor real del parámetro es θ . Esto es:

$$\text{potencia}(\theta) = \mathbb{P}(W \text{ en RR cuando el valor del parámetro es } \theta)$$

Si θ_a es un valor de θ en la hipótesis alternativa H_a , entonces:

$$\text{potencia}(\theta_a) = 1 - \beta(\theta_a)$$

8.5.2. Hipótesis simple e hipótesis compuesta

Si se toma una muestra aleatoria de una distribución con parámetro θ , se dice que una hipótesis es *simple* si especifica de manera única la distribución de la población de la cual se toma la muestra. Cualquier hipótesis que no sea simple se denomina hipótesis *compuesta*.

8.5.3. Lema de Neyman-Person

Suponga que se desea probar la hipótesis nula simple $H_0 : \theta = \theta_0$ contra la hipótesis alternativa simple $H_a : \theta = \theta_a$, con base en una muestra aleatoria Y_1, \dots, Y_n de una distribución de parámetro θ . Sea $L(\theta)$ la verosimilitud de la muestra cuando el valor del parámetro es θ . Entonces, para un α dado, la prueba que maximiza la potencia en θ_a tiene una región de rechazo RR, determinada por

$$\frac{L(\theta_0)}{L(\theta_a)} < k$$

donde el valor de k se escoge de modo que la prueba tenga el valor deseado para α . Esta prueba es la más potente en el nivel α para H_0 contra H_a .

9. Análisis de datos categóricos

9.1. Descripción del experimento

Las características que definen a un *experimento multinomial* son:

1. Consiste en n ensayos idénticos.
2. El resultado de cada ensayo cae en exactamente una de k categorías o celdas distintas.
3. La probabilidad de que el resultado de un solo ensayo caiga en una celda particular i es p_i , donde $i = 1, \dots, k$ y continúa igual de un ensayo a otro. Observe que:

$$\sum_{i=1}^k p_i = 1$$

4. Los ensayos son independientes.
5. Estamos interesados en n_1, n_2, \dots, n_k donde n_i es el número de ensayos para los cuales el resultado cae en la celda i . Observe que:

$$\sum_{i=1}^k n_i = n$$

9.2. Test χ^2 y algunos casos particulares

9.2.1. Test χ^2

Suponga que se proponen probabilidades p_1, \dots, p_k y se calcula el valor esperado para cada celda. Entonces, se espera que conteos n_i por celda sean cercanos a np_i . A partir de esto, se define el estadístico de prueba:

$$X^2 = \sum_{i=1}^k \frac{[n_i - \mathbb{E}(n_i)]^2}{\mathbb{E}(n_i)}$$

Si se cumple la hipótesis, $\mathbb{E}(n_i) = np_i$. Luego, para n grande:

$$X^2 = \sum_{i=1}^k \frac{[n_i - np_i]^2}{np_i}$$

tiene una distribución χ^2 . El número de grados de libertad es igual al número de celdas k menos el número de restricciones linealmente independiente y menos el número de parámetros estimados del modelo.

9.2.2. Test de bondad del ajuste

Se desea testear las probabilidades de cada celda o categoría, es decir

$$H_0 : p_1 = p_{1,0}, p_2 = p_{2,0}, \dots, p_k = p_{k,0}$$

donde $p_{i,0}$ es el valor especificado para p_i . La hipótesis alternativa expresa, en general, que al menos una de las igualdades no se cumple. Como la única restricción en las probabilidades por celda es que $\sum_{i=1}^k p_i = 1$, el estadístico de prueba X^2 tiene aproximadamente una distribución χ^2 con $n - 1$ grados de libertad.

9.2.3. Tablas de contingencia

Se utilizan para evaluar la independencia de dos variables aleatorias categóricas. El número de grados de libertad en una tabla de contingencia de f filas y c columnas es $(f - 1) \times (c - 1)$.