

Métodos estadísticos

Rodrigo Barrera

Variable aleatoria

Una variable aleatoria es una aplicación que asocia cada suceso elemental un número real. Conviene definir este concepto con precisión, puesto que es la idea fundamental que permite dar un tratamiento riguroso los fenómenos aleatorios.

Variable aleatoria: Sea $(\Omega, \mathcal{A}, \mathcal{P})$ un espacio probabilístico asociado un experimento aleatorio. Una variable aleatoria es una aplicación definida sobre que toma valores en el conjunto de los números reales $X : \Omega \rightarrow \mathbb{R} | a \rightarrow X(a) \in \mathbb{R}$ si se verifica que $\forall x \in \mathbb{R}$ el conjunto $\{a \in \Omega | X(a) \leq x\} \in \mathcal{A}$

Variable aleatoria

Discreta

Una variable aleatoria puede tomar un conjunto de valores finito, infinito numerable una infinidad no numerable de valores reales. Será discreta continua, por tanto, según sea su contradominio imagen.

Variable aleatoria discreta: Una variable aleatoria es discreta es aquella que toma un conjunto finito o infinito numerable de valores reales¹.

La variable aleatoria definida por el número de caras que salen cuando se lanzan dos monedas es una variable discreta; su contradominio está formado por los puntos del conjunto: $\{0, 1, 2\}$

¹Se dice que un conjunto es infinito contable (o infinito numerable o que la cardinalidad del conjunto es infinita contable), si existe una correspondencia uno a uno entre los elementos del conjunto y los elementos de \mathbb{N} .

Variable aleatoria

Continuas

Hay variables aleatorias que tienen por imagen toda la recta real algún intervalo de la misma (a, b) , $[a, b)$, $(a, b]$, $(-\infty, a)$, $[a, \infty)$, ... este tipo de variable aleatoria, que no toma valores aislados, le llamamos variable aleatoria continua, esto es:

Variable aleatoria continua: Una variable aleatoria es continua es aquella que toma valores en una escala continua.

Distribuciones discretas

Si tenemos una variable aleatoria discreta, como puede ser el número de caras que aparecen al lanzar dos monedas, es conveniente representar las probabilidades con que toma la variable aleatoria cada uno de los valores numéricos x , por una fórmula.

Esta fórmula debe ser una función de x , $f(x)$, esto es;

$$f(x) = P(X = x) = P(\{a \in \Omega | X(a) = x\})$$

Al conjunto de los pares ordenados $(x, f(x))$ le llamaremos distribución de probabilidad —o función masa de probabilidad— de la variable aleatoria X .

Distribución de probabilidad

El conjunto de pares ordenados $(x, f(x))$ es una distribución de probabilidad de la variable aleatoria X si;

$$1. - f(x) \geq 0$$

$$2. - \sum f(x) = 1$$

$$3. - P(X = x) = f(x)$$

Ahora estamos en condiciones de asociar la variable aleatoria una función real —de una variable real—, que permite construir los modelos matemáticos adecuados para tratar los problemas originados por fenómenos que se rigen por la ley del azar.

Función de distribución

Sea una variable aleatoria con distribución de probabilidad $f(x)$. Se llama función de distribución de la variable aleatoria la función F ;

$$F : \mathbb{R} \rightarrow \mathbb{R}, \text{ tal que}$$
$$F(x) = P(X \leq x) = \sum_{r \leq x}, \forall r \in \mathbb{R}$$

Variables aleatorias continuas

La función real de una variable real² $f(x)$ es una **función de densidad** de la variable aleatoria X si verifica;

$$f(x) \geq 0, \forall x \in R$$

$$\int_{-\infty}^{\infty} f(x) dx = 1$$

$$P(a < X < b) = \int_a^b f(x) dx$$

²Función real de variable real es toda correspondencia f que asocia a cada elemento de un determinado subconjunto de números reales, llamado dominio, otro número real.

Variables aleatorias continuas

Función de distribución de una variable aleatoria continua: Sea una variable aleatoria continua con función de densidad $f(x)$. Se llama función de distribución de la variable aleatoria la función real de una variable real $F(x)$ tal que;

$$F(x) = P(X \leq x) = \int_{-\infty}^x f(t)dt, \forall x \in R$$

De lo anterior se desprende que;

$$P(a < X < b) = \int_a^b f(x)dx = F(b) - F(a)$$
$$f(x) = \frac{dF(X)}{dx}$$

Esperanza matemática

Sea X una variable aleatoria con distribución de probabilidad $f(x)$. La esperanza matemática o media de X ; denotada por μ_X es;

$$\mu_X = E(X) = \sum_x xf(x)$$

Si X es discreta; y si X es continua;

$$\mu_X = E(X) = \int_{-\infty}^{\infty} xf(x)dx$$

Esperanza matemática

Sea una variable aleatoria con distribución de probabilidad $f(x)$ sea $h(X)$ una variable aleatoria. Entonces la esperanza matemática de $h(X)$ es;

$$\mu_{h(X)} = E[h(X)] = \sum_x h(x)f(x)$$

Si X es discreta; y si X es continua, es

$$\mu_{h(X)} = E[h(X)] = \int_{-\infty}^{\infty} h(x)f(x)dx$$

Propiedades de la esperanza matemática

Sea una variable aleatoria. Si las funciones de X , $g(X)$ y $h(X)$, son dos variables aleatorias tales que existen $E[g(X)]$ y $E[h(X)]$, entonces existe también $E[g(X) \pm h(X)]$ y es;

$$E[g(X) \pm h(X)] = E(g(X)) \pm E(h(X)) \quad (1)$$

Si $g(X)$ es una función de la variable aleatoria tal que existe $E[g(X)]$ y es a es un número real cualquiera, existe $E[ag(X)]$ y es;

$$E[ag(X)] = aE[g(X)] \quad (2)$$

Propiedades de la esperanza matemática

Si son dos números reales y X una variable aleatoria, se verifica:

$$E[aX \pm b] = aE[X] \pm b$$

$$E[aX] = aE[X]$$

$$E[b] = b$$

Esperanza matemática

Algunas demostraciones

Sea $p(x)$ la f.m.p. de X .

$$\begin{aligned} E(cX) &= \sum_x cxp(x) \\ &= c \sum_x xp(x) \\ &= cE(X) \end{aligned}$$

Esperanza matemática

Algunas demostraciones

Sea $p(x)$ la f.m.p. de X .

$$\begin{aligned}E(X + c) &= \sum_x (x + c)p(x) \\&= \sum_x (xp(x) + cp(x)) \\&= \sum_x (xp(x)) + \sum_x (cp(x)) \\&= E(X) + c \sum_x (p(x)) \\&= E(X) + c(1) \\&= E(X) + c\end{aligned}$$

Varianza y desviación típica

La varianza de la variable aleatoria es una buena medida de la dispersión. No obstante, esta medida no está dada en las mismas unidades que X , por lo que se utiliza su raíz cuadrada positiva, que conocemos como desviación típica. Para variables aleatorias, se define:

Sea una variable aleatoria con distribución de probabilidad $f(x)$. La varianza de X se denota por σ^2 es la esperanza de la variable aleatoria $(X - \mu)^2$.

Varianza y desviación típica

Si X es una v.a. discreta, es;

$$\sigma^2 = E[(X - \mu)^2] = \sum_x (x - \mu)^2 f(x)$$

y, si X es una v.a. continua, es

$$\sigma^2 = E[(X - \mu)^2] = \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx$$

Desviación típica: La desviación típica de la variable aleatoria X , que se representa por σ o σ_X , es la raíz cuadrada positiva de la varianza de X .

Propiedades de la varianza

- Si X es una v.a., $\sigma^2 = E(X^2) - \mu^2$
- Sea una variable aleatoria con distribución de probabilidad $f(x)$, sea $g(X)$ una función de la variable aleatoria X . Entonces la varianza de la variable aleatoria $g(X)$ es
$$\sigma_{g(X)}^2 = E\{[g(X) - \mu_{g(X)}]^2\}$$
 - $\sum_x [g(x) - \mu_{g(X)}]^2$
 - $\int_{-\infty}^{\infty} [g(x) - \mu_{g(X)}]^2$
- Si son dos números reales, se verifica $\sigma_{aX+b} = a\sigma_X$

Distribuciones

Uniforme

Sea la variable aleatoria X que puede asumir valores x_1, x_2, \dots, x_k con idéntica probabilidad. Entonces la distribución uniforme discreta viene dada por:

$$f(x) = P(X = x) = \frac{1}{k}$$

$$\mu = E(X) = \sum_{\forall x_i} x_i f(x) = \frac{\sum_{i=1}^k x_i}{k}$$

$$\sigma^2 = E[(x - \mu)^2] = \sum_{\forall x_i} (x_i - \mu)^2 f(x) = \frac{\sum_{i=1}^k (x_i - \mu)^2}{k}$$

Distribución

Binomial

Las características de un experimento aleatorio Bernoulli son:

El experimento tiene solamente dos posibles resultados mutuamente excluyentes denominados éxito (E) y fracaso (F). De esta manera el espacio muestral está dado por $\Omega = (E, F)$. Sea X una v.a. Bernoulli, se define como sigue;

$$X : \Omega \rightarrow \mathbb{R}$$

Donde el rango de la v.a. es $\{1, 0\}$ La función de probabilidad está dada por;

$$p(x) = p^x(1 - p)^{1-x}, x = 0, 1$$

Distribución

Binomial

Consideremos un proceso de Bernoulli, en el que la probabilidad de éxito en una prueba particular es p , siendo $q = 1 - p$ la probabilidad de fracaso. Entonces la distribución de probabilidad de la variable aleatoria X ="número de éxitos en un suceso compuesto de pruebas particulares" es llamada distribución binomial, siendo;

$$B(x; n, p) = P(X = x) = \binom{n}{x} p^x q^{n-x}, x = 0, 1, 2, \dots, n$$

Distribución

Binomial

El nombre de binomial se debe que las probabilidades $B(x; n, p)$ corresponden los $n + 1$ términos del desarrollo del binomio $(p + q)^n$:

$$\begin{aligned}(p + q)^n &= \binom{n}{0} q^n + \binom{n}{1} p q^{n-1} + \dots + \binom{n}{n} p^n \\&= B(0; n, p) + B(1; n, p) + \dots + B(n; n, p) \\&= \sum_{x=0}^n \binom{n}{x} p^x q^{n-x}\end{aligned}$$

Distribución

Binomial

Si X es unva v.a. binomial;

$$E(X) = np$$

$$V(X) = np(1 - p)$$

Distribución

Poisson

Considere un experimento tal que;

- Se observa la realización de hechos de cierto tipo durante un cierto periodo de tiempo o a lo largo de un espacio de observación
- Los hechos a observar tienen naturaleza aleatoria ; pueden producirse o no de una manera no determinística.
- La probabilidad de que se produzcan un número x de éxitos en un intervalo de amplitud t no depende del origen del intervalo (Aunque, sí de su amplitud)
- Si en estas circunstancias aleatorizamos de forma que la variable aleatoria X signifique o designe el “número de hechos que se producen en un intervalo de tiempo o de espacio”, la variable X se distribuye con una distribución de parámetro λ .

Distribución

Poisson

La variable aleatoria que representa el número de sucesos que tienen lugar en un intervalo de tiempo en una región específica t , se llama variable aleatoria de Poisson, cuya distribución de probabilidad representaremos por;

$$P(x; \lambda t) = P(X = x) = e^{-\lambda t} \frac{(\lambda t)^x}{x!}, x = 0, 1, 2, \dots$$

Distribución

Poisson

Demuestre que;

$$\sum_{x=0}^{\infty} P(x, \mu) = 1$$

$$\sum_{x=0}^{\infty} P(x, \mu) = \sum_{x=0}^{\infty} e^{-\mu} \frac{\mu^x}{x!} = e^{-\mu} \sum_{x=0}^{\infty} \frac{\mu^x}{x!} = 1$$

Pues $\sum_{x=0}^{\infty} \frac{\mu^x}{x!} = e^{\mu}$.

Distribución

Poisson

La media la varianza de la distribución de Poisson $P(x; \mu)$ vienen dadas por;

$$E(X) = \mu = \lambda t$$

$$Var(X) = \mu = \lambda t$$

Distribución

Geométrica

La distribución geométrica es un modelo adecuado para aquellos procesos en los que se repiten pruebas hasta la obtención de un éxito.

- El proceso consta de un número no definido de pruebas o experimentos separados o separables. El proceso concluirá cuando se obtenga por primera vez el resultado deseado (éxito).
- Cada prueba puede dar dos resultados mutuamente excluyentes : A y no A
- La probabilidad de obtener un resultado A en cada prueba es p y la de obtener un resultado no A es q siendo $p + q = 1$.

Distribución

Geométrica

La función de cuantía $P(x)$ hará corresponder a cada valor de X la probabilidad de obtener el primer éxito precisamente en el k -ésimo ensayo. Esto es, $P(X)$ será la probabilidad del suceso obtener $k - 1$ resultados “no A ” y un éxito o resultado A en la prueba número k ; es decir;

$$P(\underbrace{A^c \cap A^c \cap \dots \cap A^c}_{(k-1)\text{-veces}} \cap A) = q^{k-1}p$$

Distribución

Geométrica

$$\mu = \frac{1}{p}$$

$$\sigma^2 = \frac{q}{p^2}$$

Distribución

Binomial Negativa

Esta distribución o modelo se deriva de un proceso experimental puro o de Bernouilli en el que se presenten las siguientes condiciones;

- El proceso consta de un número no definido de pruebas separadas o separables . El proceso concluirá cuando se obtenga un determinado número de resultados favorables r .
- Cada prueba puede dar dos resultados posibles mutuamente excluyentes A y A^c .
- La probabilidad de obtener un resultado A en cada una de las pruebas es p .
- Los ensayos son independientes.

Distribución

Binomial Negativa

En una sucesión de ensayos Bernoulli independientes. sea X la v.a. que denota aquel ensayo en el cual se produce el r -ésimo éxito, entonces La f.m.p. es;

$$P(X = x; r, p) = \binom{x-1}{r-1} (1-p)^{x-r} p^r \quad (3)$$

$$E(X) = \frac{r}{p}. \quad \text{Var}(X) = \frac{r(1-p)}{p^2}$$

Distribución

Hipergeométrica

La distribución hipergeométrica puede derivarse de un proceso experimental puro o de Bernouilli con las siguientes características;

- El proceso consta de n ensayos, separados o separables de entre un conjunto de N ensayos posibles.
- Cada una de los ensayos puede dar únicamente dos resultados mutuamente excluyentes: A y A^c .
- En el primer ensayo las probabilidades son : $P(A) = p$.

Distribución

Hipergeométrica

Suponga la extracción aleatoria de n elementos de un conjunto formado por N elementos totales, de los cuales N_p son del tipo A y N_q son del tipo A^c . Si realizamos las extracciones sin devolver los elementos extraídos, y sea X el número de elementos del tipo A que extraemos en n extracciones; X seguirá una distribución hipergeométrica de parámetros N, n, p .

La función de cuantía de una distribución Hipergeométrica hará corresponder a cada valor de la variable X ($x = 0, 1, 2, \dots, n$) la probabilidad del suceso obtener x resultados del tipo A y $(n - x)$ resultados del tipo A^c en las n pruebas realizadas de entre las N posibles.

Distribución

Hipergeométrica

Función de cuantía;

$$P(X = x) = \frac{\binom{N_p}{x} \binom{N_q}{n-x}}{\binom{N}{n}}$$

$$E(X) = np$$

$$Var(X) = np(1 - p) \left(\frac{N - n}{N - 1} \right)$$

Si el tamaño n de la muestra sin reemplazo es pequeña con relación a N , la probabilidad de cada extracción varía muy levemente. En la práctica cuando $\frac{n}{N} < 0,1$ se aproxima la distribución hipergeométrica a la binomial con $p = \frac{N_p}{N}$ y n .

Convergencia Binomial-Poisson

Se puede probar que la distribución binomial tiende a converger a la distribución de Poisson cuando el parámetro n tiende a infinito y el parámetro p tiende a ser cero, de manera que el producto de n por p sea una cantidad constante. De ocurrir esto la distribución binomial tiende a un modelo de Poisson de parámetro $\lambda = np$.

$$\begin{aligned} P(X = x) &= \binom{n}{x} p^x (1 - p)^{n-x}; \lambda = np \\ &= \frac{n(n-1)(n-2)\dots(n-x+1)}{x!} \frac{\lambda^x}{n^x} \left(1 - \frac{\lambda}{n}\right)^{n-x} \\ &= \frac{n}{n} \left(\frac{n-1}{n}\right) \dots \left(\frac{n-x+1}{n}\right) \frac{\lambda^x}{x!} \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-x} \end{aligned}$$

Convergencia Binomial-Poisson

Ahora debería ser relativamente fácil ver que si tomamos el límite de n cuando se acerca al infinito, manteniendo x y λ fijos, las primeras x fracciones en esta expresión tenderían hacia 1, al igual que el último factor en la expresión. El penúltimo factor, tiende hacia $e^{-\lambda}$, y el factor restante permanece sin cambios ya que no depende de n .

Recuerde que una típica definición de e^x está dada por;

$$e^x = \lim_{n \rightarrow \infty} \left(1 + \frac{x}{n}\right)^n$$

Otra caracterización típica de e^x es;

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}$$

Distribución uniforme

Sea X una variable aleatoria continua, que toma todos los valores de un intervalo o $[a, b]$, donde a y b son números reales y $a < b$. Si la función de densidad de probabilidad de X está dada por;

$$f(x) = \begin{cases} \frac{1}{b-a}, & a \leq x \leq b \\ 0, & \text{en otro caso.} \end{cases}$$

Se dice que X se distribuye uniformemente en el intervalo $[a, b]$.

Distribución uniforme

Una variable aleatoria continua uniformemente distribuida, representa la analogía a los resultados igualmente posibles en el sentido siguiente: Para cualquier sub-intervalo $[c, d]$, donde $a \leq c \leq d \leq b$ es la misma para todo los subintervalos que tienen la misma longitud.

$$P[c \leq X \leq d] = \int_c^d \frac{dx}{b-a} = \frac{d-c}{b-a}$$

Solo depende de la longitud del intervalo y no de la ubicación del intervalo. La función de distribución acumulativa está dado por;

Distribución uniforme

$$F(x) = P(X \leq x) = \int_{-\infty}^x \frac{1}{b-a} dx = \int_a^x \frac{1}{b-a} dx = \frac{x-a}{b-a}$$

Más formalmente;

$$f(a, b) = \begin{cases} 0 & , x \leq a \\ \frac{x-a}{b-a} & , a \leq x < b \\ 1 & , x \geq b \end{cases}$$

Distribución uniforme

Media y varianza

$$E(X) = \int_a^b \frac{x}{b-a} dx = \frac{a+b}{2}$$

$$Var(X) = \int_a^b \frac{x^2}{b-a} dx - \left(\frac{b+a}{2}\right)^2 = \frac{(b-a)^2}{12}$$

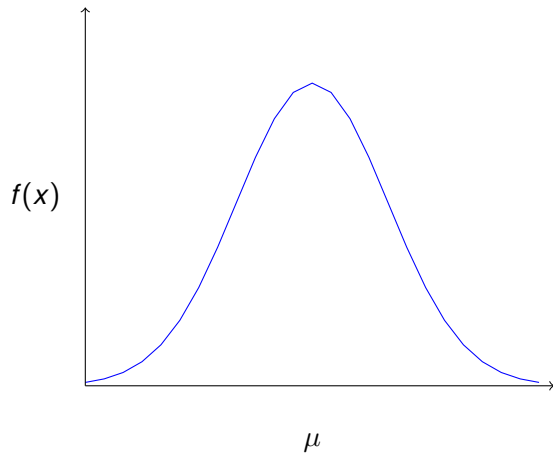
Distribución normal

La distribución de probabilidad continua más importante en todo el campo de la estadística, es con toda seguridad, la distribución normal debido a que en la práctica muchos fenómenos, industriales, científicos, o de la vida diaria pueden describirse por esta distribución. También mediante esta distribución se obtienen aproximaciones a otras leyes de probabilidad.

Una v.a. continua X , se dice que se distribuye normalmente, con media $\mu \in \mathbb{R}$ y $\sigma^2 \in \mathbb{R}^+$; Su función de densidad está determinada por;

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp \left\{ -(1/2) \left[\frac{x - \mu}{\sigma} \right]^2 \right\}$$

Distribución normal



Distribución normal

Si Z es una v.a. que tiene distribución normal con media $\mu = 0$ y varianza $\sigma^2 = 1$, entonces $Z = \frac{X - \mu}{\sigma}$ es una distribución normal estándar. Su función de densidad es

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}, -\infty < z < \infty$$

La función de distribución de Z se denota por $\phi(z)$ y está dado por;

$$\phi(z) = P[Z \leq z] = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$$

Distribución normal

La distribución normal, también llamada distribución gaussiana, la función de distribución más común para variables independientes generadas aleatoriamente. Su curva familiar en forma de campana es omnipresente en los informes estadísticos, desde el análisis de encuestas y el control de calidad hasta la asignación de recursos. El gráfico de la distribución normal se caracteriza por dos parámetros: la media, o promedio, que es el máximo del gráfico y sobre el cual el gráfico es siempre simétrico; y la desviación estándar, que determina la cantidad de dispersión lejos de la media. Una pequeña desviación estándar (en comparación con la media) produce un gráfico empinado, mientras que una desviación estándar grande (nuevamente en comparación con la media) produce un gráfico plano.

Distribución normal

La probabilidad de que una variable aleatoria caiga dentro de cualquier rango dado de valores es igual a la proporción del área encerrada bajo el gráfico de la función entre los valores dados y por encima del eje x . Debido a que el denominador de la función de distribución, conocido como coeficiente de normalización, hace que el área total encerrada por el gráfico sea exactamente igual a la unidad, las probabilidades se pueden obtener directamente del área correspondiente, es decir, un área de 0.5 corresponde a una probabilidad de 0.5. Aunque estas áreas se pueden determinar con cálculo, las tablas se generaron en el siglo 19 para el caso especial de $\mu = 0$ y $\sigma = 1$, conocido como la distribución normal estándar, y estas tablas se pueden usar para cualquier distribución normal después de que las variables se reescalen adecuadamente.

Distribución normal

El término *distribución gaussiana* se refiere al matemático alemán Carl Friedrich Gauss, quien desarrolló por primera vez una función exponencial de dos parámetros en 1809 en relación con los estudios de errores de observación astronómica. Este estudio llevó a Gauss a formular su ley del error observacional y a avanzar en la teoría del método de aproximación de mínimos cuadrados. Otra famosa aplicación temprana de la distribución normal fue por el físico británico James Clerk Maxwell, quien en 1859 formuló su ley de distribución de velocidades moleculares, más tarde generalizada como la ley de distribución de Maxwell-Boltzmann.

Distribución normal

El matemático francés Abraham de Moivre, en su *Doctrine of Chances* (1718), señaló por primera vez que las probabilidades asociadas con variables aleatorias generadas discretamente (como las que se obtienen lanzando una moneda o tirando un dado) pueden ser aproximadas por el área bajo el gráfico de una función exponencial. Este resultado fue extendido y generalizado por el científico francés Pierre-Simon Laplace, en su *Théorie analytique des probabilités* (1812), en el primer teorema del límite central, que demostró que las probabilidades para casi todas las variables aleatorias independientes e idénticamente distribuidas convergen rápidamente al área bajo una función exponencial, es decir, a una distribución normal. El teorema del límite central permitía que los problemas hasta entonces intratables, particularmente aquellos que involucraban variables discretas, se manejaran con cálculo.

Normal estandar

Sea X una variable aleatoria $N(5, 4)$. (a) ¿Cuál es la probabilidad de que X tome valores entre 4 y 7? (b) ¿Cuál es la probabilidad que tome valores mayores que 10?

Respuestas;

- $(a) = 0,5324$
- $(b) = 0,0062$

Si X es una variable aleatoria $N(650, 625)$. Hallar la constante $c > 0$ tal que,
 $P(|X - 650| \leq c) = 0,9544$

Normal estandar

Los registros de pérdida de peso por evaporación de cierto producto empacado muestran una pérdida media de 6.45 gramos con una desviación estándar de 1.30. Asumiendo una distribución normal. ¿Cuál es la probabilidad de que si se extraen dos paquetes al azar de un lote ambas muestran una pérdida de más de 8.00 gramos?

¿Cuál es la probabilidad al seleccionar cinco paquetes que, por lo menos, uno de ellos muestre pérdida de más de 8.00 gramos?

Teorema de De Moivre-Laplace

Sean X_1, X_2, \dots, X_n v.a. independientes con distribución de probabilidad $Ber(1, p)$. Entonces

$$\frac{S_n - np}{\sqrt{np(1-p)}} \rightarrow Z \sim N(0, 1) (n \uparrow \infty)^3 \quad (4)$$

donde $S_n = X_1 + \dots + X_n \sim Bin(n, p)$

³¿Sabía usted que existen distintos tamaños del infinito?

Teorema central del límite

Una variable que es el resultado de la suma de muchos efectos independientes entre sí y sin que ninguno domine al total es aproximadamente normal.

Sean X_1, X_2, \dots, X_n v.a. i.i.d. con igual media e igual varianza. Entonces;

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \rightarrow Z \sim N(0, 1) (n \uparrow \infty) \quad (5)$$

donde $S_n = X_1 + \dots + X_n$.

Teorema central del límite

En una avión hay 100 personas. Sea X_i el peso (en libras) de la i -ésima persona en el avión. Suponga que los X_i son i.i.d., y $E(X_i) = 170$ y $\sigma(X_i) = 30$. Encuentre la probabilidad que el peso total de las personas en el avión exceda las 18.000 libras.

Recuerde que $Var(X + Y) = Var(X) + Var(Y) + 2Cov(X, Y)$. Y si X e Y son independiente, entonces $Cov(X, Y) = 0$. Demostrar.

Teorema central del límite

Sean X_1, X_2, \dots, X_{25} v.a. i.i.d. con función de masa de probabilidad;

$$P_X(k) = \begin{cases} 0.6 & k=1 \\ 0.4 & k=-1 \\ 0 & \text{e.o.c.} \end{cases}$$

Utilizando el T.C.L estime $P(4 \leq Y \leq 6)$

Teorema central del límite

¿Cuál ? Un dado es lanzado 420 veces. ¿Cuál es la probabilidad que la suma de dichos datos?

Aproximación normal a la Poisson

La aproximación de la distribución de Poisson a la normal, se hace teniendo en cuenta que si, X_1, X_2, \dots, X_n son variables aleatorias independientes de Poisson, cada una con parámetro λ , entonces $X = \sum X_i$ es una v.a. Poisson con parámetro $n\lambda$.

Por el teorema central del límite, la variable aleatoria;

$$Z = \frac{X - n\lambda}{\sqrt{n\lambda}}$$

tiene aproximadamente una distribución $N(0,1)$, para n suficientemente grande. La aproximación de la distribución de Poisson a la normal se mejora conforme aumenta el valor del parámetro $n\lambda$, de la suma.

Aproximación normal a la Hipergeométrica

Sea N el tamaño de la población finita constituida por objetos de dos clases A y B. Suponga que hay M objetos de clase A y $N - M$ de clase B. Se extrae una muestra de tamaño n sin reemplazo de la población, y se define la variable aleatoria como sigue;

X : número de objetos de la clase A en la muestra de tamaño n . $R_X : \{0, 1, 2, \dots, \min(M, n)\}$.

Recuerde además que $E(X) = \frac{nM}{N}$ y $Var(X) = \frac{nM}{N}(1 - \frac{M}{N})(\frac{N-n}{N-1})$

Así se define la v.a. X_i : número de objetos de la clase A obtenida en la i ésima extracción.

$R_{X_i} = \{0, 1\}$.

Aproximación normal a la Hipergeométrica

Entonces, la v.a. $X = \sum X_i$, suponiendo que los X_i son aproximadamente independientes, para n suficientemente grande, por el TCL se tiene que;

$$Z = \frac{X - \frac{nM}{N}}{\sqrt{\frac{nM}{N} \left(1 - \frac{M}{N}\right) \frac{N-n}{N-1}}}$$

Aproximación de la hipergeométrica a la binomial

Si el tamaño n de la muestra sin reemplazo es pequeña con relación a N , la probabilidad de cada extracción varía muy levemente. En la práctica cuando n es menor que el 10% de N se aproxima la distribución hipergeométrica a la distribución binomial con $p = \frac{N_p}{N}$ y n .

Aproximación de la hipergeométrica a la binomial

Un auditor del Departamento de impuesto sobre la renta está seleccionando una muestra de seis declaraciones de impuestos de personas de una profesión particular, para una posible auditoría. Si dos o más de ellas indican deducciones "no autorizadas", se auditará todo el grupo (población) de 100 declaraciones. Si el 25% de las declaraciones es incorrecta, determinar: (a) la verdadera distribución de probabilidad del número de declaraciones incorrectas en la muestra. ¿Cuales son los parámetros?. Halle la probabilidad de una auditoría más detallada. (b) Utilice una aproximación a la verdadera distribución de probabilidad para hallar la probabilidad de una auditoría más detallada.

Distribución exponencial

Dada una variable aleatoria X que tome valores reales no negativos diremos que tiene una distribución exponencial de parámetro $\alpha \geq 0$, si y sólo si su función de densidad tiene la expresión:

$$f(x) = \alpha e^{-\alpha x}$$

$$\text{Función de distribución} = \int_0^x \alpha e^{-\alpha t} dt = 1 - e^{-\alpha x}$$

$$E(X) = \frac{1}{\alpha}$$

$$\text{Var}(X) = \frac{1}{\alpha^2}$$

Distribución exponencial

Un monitor emite una señal de advertencia cuando se necesita una acción como parte de un proceso de producción. El intervalo, X horas, entre señales sucesivas sigue una distribución exponencial con parámetro 0.08.

- Halle la probabilidad de que el intervalo entre las siguientes dos señales sea:
 - Entre 10 y 20 horas.
 - Menos de dos horas.
 - Más de 50 horas.
- Indique la media y la desviación estándar de los intervalos entre sucesivas señales.
- Después de una señal de advertencia, ¿cuál es el tiempo más largo que podría durar el proceso de producción sin supervisión mientras se asegura que la probabilidad de perder la siguiente señal sea menos de 0.01?

Relación entra la Poisson y Exponencial

La distribución exponencial tiene una relación especial con la distribución de Poisson. Hemos visto que la distribución de Poisson describe el número de ocurrencias de eventos (éxitos) por unidad de medida (intervalo de tiempo, una área determinada etc.). Por ejemplo, numero de llamadas telefónicas por minuto. Otro ejemplo es: número de carros que llegan a la caseta de peaje cada hora. En cambio la distribución exponencial describe el valor de la medida por ocurrencia de eventos (éxito).

Relación entra la Poisson y Exponencial

Así pues, las dos distribuciones se puede utilizar para describir el mismo - fenómeno, la distribución de Poisson describe el número de ocurrencias de eventos por unidad de media, y la exponencial, describe valor de la medida - entre ocurrencias sucesivas de eventos.

Consideremos un proceso de Poisson como sigue: $X(\omega)$ = número de ocurrencias de un evento en un período t ; definimos el parámetro X del proceso de Poisson como el número esperado de ocurrencia por unidad de tiempo. Entonces, el número esperado de ocurrencias en el intervalo t será λt , luego la distribución de probabilidad de X es;

$$P[X = x] = \frac{(\lambda t)^x e^{-\lambda t}}{x!}, x = 0, 1, 2, \dots$$

Supongamos ahora que estamos interesados en la distribución de probabilidad del intervalo de tiempo entre ocurrencias de los eventos. Es decir, la variable aleatoria está definida ahora por: $T(t)$ = intervalo de tiempo entre ocurrencias sucesivas de eventos. Luego, T es continua obviamente y su rango es;

$$R_T = t \in \mathbb{R}$$

Relación entra la Poisson y Exponencial

Un vendedor vende periódicos en una esquina. Los periódicos que - vende son eventos de un proceso de Poisson con parámetro $\lambda = 50$ por hora. Si alguien acaba de comprarle un periódico. ¿Cuál es la probabilidad que - transcurran al menos 2 minutos antes que venda otro? ¿De que no pasen más de 5 minutos?. Si ya han transcurridos 5 minutos desde la última venta. ¿Cuál es la probabilidad que transcurran al menos 2 minutos más para su siguiente venta?

Distribución χ^2

Sean Y_1, Y_2, \dots, Y_n v.a. independientes distribuidas normalmente, cada una con medio 0 y varianza 1; con $X = Y_1^2 + Y_2^2 + \dots + Y_n^2$, se dirá que X sigue una distribución χ^2 con n grados de libertad.

$$\text{Función de densidad } f_X(x) = \frac{1}{2^{\frac{n}{2}} \Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} e^{-x/2}$$

Donde Γ representa la función gamma. La v.a. χ^2 con n grados de libertad tiene $E(X) = n$ y $Var(X) = 2n$.

Distribución χ^2

Observe que las distribuciones χ^2 son una familia de distribuciones continuas positivamente asimétricas; sin embargo cuando n ,aumenta la χ^2 se aproxima a una distribución normal. Por esta razón, es que en la práctica, cuando a es grande ($n > 30$), la probabilidad de la χ^2 puede calcularse empleando aproximación normal

Variables aleatorias bidimensionales

Dado un espacio muestral Ω asociado a un experimento aleatorio, una variable aleatoria bidimensional es una aplicación que a un mismo elemento del espacio muestral le hace corresponder 2 números reales, es decir, una aplicación;

$$(X, Y); s \in \Omega \rightarrow (X(s), Y(s)) \in \mathbb{R}^2 \quad (6)$$

Variables aleatorias bidimensionales

Dicho de otro modo, una variable aleatoria bidimensional resulta de cuantificar dos aspectos, X e Y , de un experimento aleatorio. Podríamos decir que $(X, Y); s \in \Omega \rightarrow (X(s), Y(s)) \in \mathbb{R}^2$ es una variable aleatoria bidimensional si y sólo X e Y son ambas variables aleatorias (simples).

Variables aleatorias bidimensionales discretas

Una variable aleatoria bidimensional (X, Y) se dice discreta cuando sus dos componentes X e Y son ambas discretas. Una variable aleatoria bidimensional discreta queda completamente caracterizada una vez que conocemos el conjunto de todos los posibles pares de valores que puede tomar y la probabilidad de que este valor sea tomado.

La función de probabilidad conjunta de una variable aleatoria bidimensional (X, Y) es $P(X = x_i, Y = y_j) = p_{ij}$ entendiendo que si la variable no puede tomar un par concreto, su probabilidad es 0.

Variables aleatorias bidimensionales discretas

Propiedades

Sea (X, Y) una variable aleatoria bidimensional discreta.

- $0 \leq p_{ij} \leq 1; \forall i, j$
- $\sum_{i,j} p_{ij} = 1$

Se denomina función de distribución conjunta a la función $F : \mathbb{R}^2 \rightarrow [0, 1]$ dada por;

$$F(x, y) = P(X \leq x, Y \leq y) = \sum_{x_i \leq x} \sum_{y_j \leq y} p_{ij}$$

Variables aleatorias bidimensionales discretas

Sea (X, Y) una variable aleatoria bidimensional discreta con función de probabilidad conjunta $P(X = x_i, Y = y_j)$ La función de probabilidad marginal de X viene dada por;

$$p_X(x_i) = P(X = x_i) = \sum_j P(X = x_i, Y = y_j) = \sum_j p_{ij} \quad (7)$$

La función de probabilidad marginal de Y viene dada por;

$$p_Y(y_j) = P(Y = y_j) = \sum_i P(X = x_i, Y = y_j) = \sum_i p_{ij} \quad (8)$$

Variables aleatorias bidimensionales discretas

Función de densidad marginal

Sea (X, Y) una variable aleatoria bidimensional discreta con función de probabilidad conjunta $P(X = x_i, Y = y_j)$

La función de densidad marginal de X viene dada por;

$$F_X(x) = P(X \leq x) = \sum_{x_i \leq x} \sum_j p_{ij} \quad (9)$$

La función de densidad marginal de Y viene dada por;

$$F_Y(y) = P(Y \leq y) = \sum_{y_j \leq y} \sum_i p_{ij} \quad (10)$$

V.a. bidimensionales discretas

Función de probabilidad condicionada

Sea (X, Y) una variable aleatoria bidimensional discreta con función de probabilidad conjunta $P(X = x_i, Y = y_j)$. Si $P(Y = y_{j_0}) > 0$ se define la función de probabilidad de X condicionada a $Y = y_{j_0}$ como;

$$p_{X|Y}(x_i|y_{j_0}) = \frac{P(X = x_i|Y = y_{j_0})}{P(Y = y_{j_0})} = \frac{p_{ij_0}}{\sum_i p_{ij_0}} \quad (11)$$

V.a. bidimensionales continuas

Una variable aleatoria bidimensional (X, Y) se dice continua cuando sus dos componentes X e Y son ambas continuas.

La variable aleatoria bidimensional (X, Y) tiene una distribución de probabilidad conjunta continua si existe una función integrable $f : \mathbb{R} \rightarrow [0, +\infty)$, tal que para cualquier subconjunto A del plano se tiene que;

$$P([X, Y] \in A) = \int \int_A f(x, y) dx dy$$

La función f se denomina función de densidad conjunta.

V.a. bidimensionales continuas

La probabilidad de que el vector (X, Y) pertenezca a una región se calcula por integración. Como consecuencia de ello, cualquier punto del plano o cualquier curva en el plano tiene probabilidad cero. Además, dado que el total debe tener probabilidad 1, si $f(x, y)$ es una función de densidad conjunta, entonces;

$$\int \int_{\mathbb{R}^2} f(x, y) dx dy = 1$$

V.a. bidimensionales continuas

Sea (X, Y) una variable aleatoria bidimensional continua con función de densidad conjunta $f(x, y)$. Se llama función de distribución conjunta a la función $F : \mathbb{R}^2 \rightarrow [0, 1]$ dada por;

$$F(x, y) = P(X \leq x, Y \leq y) = \int_{-\infty}^x \int_{-\infty}^y f(s, t) ds dt$$

Naturalmente, se tiene que;

$$f(x, y) = \frac{\partial^2 F(x, y)}{\partial x \partial y}$$

Esperanza matemática

Sea (X, Y) una v.a. bidimensional. Se define el valor esperado o esperanza matemática de una función $g(X, Y)$ como:

Caso discreto;

$$E[g(X, Y)] = \sum_i \sum_j g(x_i, y_i) p_{ij}$$

Caso continuo;

$$E[g(X, Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) f(x, y) dy dx$$

Esperanza matemática

Propiedades

$$E[aX + bY + c] = aE[X] + bE[Y] + c$$

Si X e Y son independientes, entonces $E[g(X)g(Y)] = E[g(X)]E[g(Y)]$

Covarianza $\text{Cov}(X, Y) = E(XY) - E(X)E(Y)$.

Esperanza condicional

Sea (X,Y) una v.a. bidimensional y sea y un número real, tal que $P[Y = y] > 0$, entonces;

$$E(X|Y = y) = \sum_{x \in R_X} xP[X = x|Y = y]$$

En particular, si $P[X = x_0] > 0$, se tiene;

$$\begin{aligned} E(X|X = x_0) &= \sum_{x \in R_X} xP[X = x|X = x_0] \\ &= \sum_{x \in R_X} x \frac{P[X = x, X = x_0]}{P[X = x_0]} \\ &= x_0 \frac{P[X = x_0]}{P[X = x_0]} = x_0 \end{aligned}$$

Varianza condicional

$$\begin{aligned}\sigma_{X|y}^2 &= E[(X - E(X|y))^2|y] \\ &= \sum_{x \in R_X} (x - E(X|y))^2 p(x|y)\end{aligned}$$

Se tienen también que $\sigma_{X|y}^2 = E(X^2|y) - (E(X|y))^2$

Propiedades

$$E(aX + b|Y = y) = aE(X|Y = y) + b; \forall y/p_Y(y) > 0$$

Sea $H(X, Y)$ una función de dos variables;

$$E[H(X, Y)|Y = y] = \sum_{x \in R_X} H(x, y)P[X = x|Y = y]$$

Teorema: Sean X e Y dos v.a. independientes, $\Rightarrow E(X|Y = y) = E(X); \forall y/p_Y(y) > 0$

Independencia de variables aleatorias

Con anterioridad vimos que si dos sucesos A y B son independiente, entonces $P(A \cap B) = P(A)P(B)$. Análogamente, se dirá que dos variables aleatorias X e Y son independiente si;

$$P(X = x_i, Y = y_j) = p_{ij} = p_i \cdot p_j.$$

$$f(x, y) = F_X(x)f_Y(y)$$

$$F(x, y) = F_X(x)F_Y(y)$$

Covarianza

Sean X e Y v.a. con medias μ_X y μ_Y respectivamente, entonces;

$$\begin{aligned} \text{Cov}(X, Y) &= \sigma_{XY} = E[(X - \mu_X)(Y - \mu_Y)] \\ &= \sum_{(x,y) \in R_{X \times Y}} (x - \mu_X)(y - \mu_Y)p(x, y) \end{aligned}$$

Si los valores más pequeños de X son asociados con valores grandes Y , y viceversa, entonces la covarianza es negativa. En cambio la covarianza será positiva cuando los valores grandes de X , se asocian con valores grandes de Y , y valores pequeños de X son asociados con valores pequeños de Y . Si la covarianza es cero, se dice que X e Y no están correlacionadas, aunque pueden ser dependientes o independientes.

Covarianza

Sean X e Y v.a. con medias μ_X y μ_Y respectivamente, entonces;

$$\text{Cov}(X, Y) = E(XY) - \mu_X \mu_Y$$

Si X e Y son v.a. independientes, entonces

$$\text{Cov}(X, Y) = 0$$

Si X e Y son variables aleatorias con función de probabilidad conjunta y varianzas finitas, entonces,

$$\text{Var}(aX \pm bY) = a^2 \text{Var}(X) + b^2 \text{Var}(Y) \pm 2ab \text{Cov}(X, Y)$$

Correlación

Sean X e Y v.a. con desviación estandar σ_X y σ_Y respectivamente. El coeficiente de correlación de X e Y se define por;

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}$$

Propiedades de la correlación

$$-1 \geq \rho \leq 1$$

$$\rho(X, Y) = \rho(Y, X)$$

Si X e Y son independientes $\Rightarrow \rho = 0$

Si $\rho \pm 1 \Rightarrow \exists$ dependencia funcional lineal entre X e Y .