Probabilidades

Pedro Ovalles
povallesgarcia@usb.ve

Departamento de Cómputo Cientifico y Estadistica
Universidad Simón Bolivar

31 de agosto de 2021

¿Por qué probabilidades?

Elementos de un espacio de probabilidad

Espacio muestral: el conjunto, Ω, de todos los posibles resultados de un experimento aleatorio.

Eventos: todos los subconjuntos de Ω.

Familia $\mathscr{F}$: una σ-álgebra de conjuntos de omega que cumplen

a) $\Omega \in \mathscr{F}$.

b) Si A es un evento, entonces A^c también lo es.

c) Si $A_1, A_2, \ldots$ entonces

$$\bigcup_{i=1}^{\infty} A_i \in \mathscr{F}$$

Probabilidad

Definición

Para un espacio muestral Ω y una familia de eventos $\mathscr{F}$ de Ω, la probabilidad de un evento $A \in \mathscr{F}$ es un número real que cumple con los siguientes axiomas:

1. $P(A) \geq 0$ para todo A.
2. $P(\Omega) = 1$.
3. Si $A_1, A_2, \ldots$ es una sucesión de eventos disjuntos por pares, es decir, tales que $A_i \cap A_j = \emptyset$ para $i \neq j$, entonces $P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$

Propiedades importantes

- $P(\emptyset) = 0$.
- $P(A^c) = 1 - P(A)$.
- $0 \leq P(A) \leq 1$.

Espacios equiprobables

$$P(A) = \frac{\text{Casos favorables}}{\text{Casos totales}}$$

Probabilidad condicional

Definición

La probabilidad condicional de un evento A, suponiendo que ocurrió el evento B es igual a

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

Ley multiplicativa de probabilidades

Teorema

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

$$P(A \cap B) = P(A)P(B|A) = P(B)P(A|B).$$

En general si $A_1, A_2, ..., A_n$ son eventos tales que $P(A_1 \cap A_2 \cap ... \cap A_k) > 0$, entonces

$$P(A_1 \cap A_2 \cap ... \cap A_n)$$

$$= P(A_n|A_1 \cap A_2 \cap ... \cap A_{n-1})P(A_{n-1}|A_1 \cap A_2 \cap ... \cap A_{n-2})...P(A_3|A_1 \cap A_2)$$

$\mathrm{P}(\mathrm{A}_2|A_1)P(A_1)$.

Ley aditiva de probabilidades

Teorema

La probabilidad de la unión de los eventos A y B es

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

En general si A_i es una sucesión finita de eventos, no necesariamente disjuntos, entonces

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i) - \sum_{i \leq j} P(A_i \cap A_j) + \ldots + (-1)^{n-1} P(A_1 \cap A_2 \cap \ldots A_n).$$

Eventos independientes

Se dice que dos eventos A y B son independientes si

$$P(A \cap B) = P(A)P(B)$$

Definición

Definición

Sea un entero positivo k y los conjuntos B_1, B_2, ..., B_k que satisfacen:

1. $\Omega = B_1 \cup B_2 \cup \ldots \cup B_k$
2. $B_i \cap B_j = \emptyset$ para $i \neq j$

De esta manera, se dice que la colección de conjuntos $B_1, B_2, \ldots, B_k$ es una partición de Ω.

Ley de Probabilidad Total

Teorema

Si $B_1, B_2, ..., B_k$ constituye una partición de Ω tal que $P(B_i) > 0$, para $i = 1, 2, ..., k$, entonces para cualquier evento A

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

Teorema de Bayes

Teorema

Si $B_1, B_2, ..., B_k$ constituye una partición de Ω tal que $P(B_i) > 0$, para $i = 1, ..., k$. Entonces

$$P(B_j|A) = \frac{P(A|B_j)P(B_j)}{\displaystyle\sum_{i=1}^{k} P(A|B_i)P(B_i)}.$$

Variable aleatoria

Definición

Una variable aleatoria es una función de valores reales cuyo dominio es un espacio muestral.

Las variables aleatorias pueden ser:

- Discretas
- Continuas
- Mixtas

Funciones de probabilidad

Función de distribución: $F(y) = P(Y \leq y)$.

Función de masa: $p(y) = P(Y = y)$

Función de densidad:

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

Ley de probabilidad

Funciones conjuntas

Función de distribución: $F(x, y) = P(X \leq x, Y \leq y)$.

Función de masa: $p(x, y) = P(X = x, Y = y)$

Función de densidad:

$$F(x, y) = \int \int f(s, t) ds dt$$

Funciones marginales

Caso discreto: $p_X(x) = \sum_y p(x,y)$.

Caso continuo: $f_X(x) = \int f(x,y)dy$.

Funciones condicionales

Caso discreto:

$$p(x|y) = P(X = x|Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)} = \frac{p(x, y)}{p_Y(y)}$$

Caso continuo: $f(x|y) = \dfrac{f(x, y)}{f_Y(y)}$

Variables aleatorias independientes

Caso discreto: $p(x, y) = p_X(x)p_Y(y)$.

Caso continuo: $f(x, y) = f_X(x)f_Y(y)$.

Valor esperado (univariado)

	Discreto	Continuo
$E(X)$	$\sum_x xp(x)$	$\int_{-\infty}^{\infty} xf(x)dx$
$E(g(X))$	$\sum_x g(x)p(x)$	$\int_{-\infty}^{\infty} g(x)f(x)dx$

Varianza

$$V(X) = E[(X - \mu_X)^2] = E[X^2] - \mu_X^2$$

Desviación estándar

La desviación estándar de X es la raiz cuadrada positiva de $V(X)$, y usualmente es denotada como σ, mientras que la varianza es denotada como σ^2.

Propiedades de la media y la varianza

- $E[aX + b] = aE[X] + b$.
- $V(aX + b) = a^2 V(X)$

Covarianza

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

Si X y Y son independientes, entonces $\mathrm{Cov}(X, Y) = 0$.

Correlación

Uno de los problemas de la covarianza es que es dificil de interpretar como medida de dependencia entre las variables, para eso usaremos el coeficiente de correlación que definimos como:

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

donde σ_X y σ_Y son las desviaciones estándar de X y Y respectivamente.

Correlación

Se puede demostrar que $-1 \leq \rho \leq 1$, y se interpreta que si el signo de ρ es positivo entonces ambas variables crecen o decrecen en paralelo, si el signo es negativo se dice que mientras una variable crece la otra decrece (o viceversa). Si $\rho = 0$, se dice que las variables no están correlacionadas.

Modelos de probabilidad discretos

$X \sim$	Rango(X)	$P(X=x)$	$E(X)$	$V(X)$
Ber(p)	$\{0,1\}$	$p^x(1-p)^{1-x}$	p	$p(1-p)$
Bin(n,p)	$\{0,1,\ldots,n-1,n\}$	$\binom{n}{x}p^x(1-p)^{n-x}$	np	$np(1-p)$
Geo(p)	$\{1,2,3,4,\ldots\}$	$(1-p)^{x-1}p$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
BinNeg(k,p)	$\{k,k+1,k+2,\ldots\}$	$\binom{x-1}{k-1}(1-p)^{x-k}p^k$	$\frac{k}{p}$	$\frac{k(1-p)}{p^2}$
Hiper(N,n,k)	$\{0,\ldots,k\}$	$\frac{\binom{n}{x}\binom{N-n}{k-x}}{\binom{N}{k}}$	$\frac{kn}{N}$	$\frac{kn(N-k)(N-n)}{N^2(N-1)}$
Poisson(λ)	$\{0,1,2,3,\ldots\}$	$\frac{\lambda^x}{x!}e^{-\lambda}$	λ	λ

Modelos continuos

$X \sim$	Rango(X)	$P(X=x)$	$E(X)$	$V(X)$
Unif(a,b)	$[a,b]$	$\frac{1}{b-a}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
Nor(μ,σ^2)	$\mathbb{R}$	$\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	μ	σ^2
Gamma(α,β)	$\mathbb{R}^+$	$\frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-\frac{x}{\beta}}$	$\alpha\beta$	$\alpha\beta^2$
Exp(β)	$\mathbb{R}^+$	$\frac{1}{\beta} e^{-\frac{x}{\beta}}$	β	β^2
Beta(α,β)	$(0,1)$	$\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}$	$\frac{\alpha}{\alpha+\beta}$	$\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$

Función generadora de momentos

Definición

Para una variable aleatoria X y algún valor entero positivo i, la esperanza $E[X^i]$ se denomina el momento de orden i de X. Con eso en consideración, para cada valor real t definimos

$$\psi(t) = E[e^{tX}].$$

Función generadora de momentos

Esta función es conocida como la función generadora de momentos de X (f.g.m.). Si X es acotada, entonces el valor esperado anterior existe para cualquier valor de t. Si no es acotada, entonces puede existir para algunos valores y para otros no.

En particular, $\psi(0) = E[e^0] = E[1] = 1$ existe para cualquier variable aleatoria X.

Momento de primer orden

Si en un intervalo alrededor de $t = 0$, se puede probar que $\psi'(t)$ existe en $t = 0$.

$$\psi'(0) = \left[\frac{d}{dt}E[e^{tX}]\right]_{t=0} = E\left[\left(\frac{d}{dt}e^{tX}\right)_{t=0}\right] = E[(Xe^{tX})_{t=0})] = E[X].$$

Momentos sucesivos

Además con esta condición, es claro que como e^t es infinitamente derivable, entonces

$$\psi^{(n)}(0) = \left[\frac{d^n}{dt^n} E[e^{tX}]\right]_{t=0} = E\left[\left(\frac{d^n}{dt^n} e^{tX}\right)_{t=0}\right] = E[(X^n e^{tX})_{t=0})] = E[X^n].$$

Por lo tanto, $\psi'(0) = E[X]$, $\psi''(0) = E[X^2]$, $\psi'''(0) = E[X^3]$, y asi sucesivamente.

f.g.m. de variables independientes

Teorema

Suponga que X_1, ..., X_n son n variables aleatorias independientes; y para $i = 1, ..., n$, sea ψ_i la f.g.m. de X_i. Sea $Y = X_1 + ... + X_n$, y denote ψ como la f.g.m. de Y. Entonces, para cualquier valor de t tal que existe $\psi_i(t)$ para cada i

$$\psi(t) = \prod_{i=1}^{n} \psi_i(t).$$

f.g.m. para distribuciones

Teorema

Si las f.g.m. de dos variables aleatorias X_1 y X_2 son idénticas para todos los valores de t en un intervalo alrededor del punto $t = 0$, entonces las distribuciones de probabilidad X_1 y X_2 deben ser idénticas.

Convergencia en probabilidad

Definición

Suponga que Z_1, Z_2, ... es una sucesión de variables aleatorias. Se dice que la sucesión converge en probabilidad a b si para cualquier $\varepsilon > 0$ dado,

$$\lim_{n\to\infty} P(|Z_n - b| < \varepsilon) = 1.$$

Media muestral

Definición

Dada una muestra aleatoria X_1, X_2, ..., X_n, la media muestral se define como

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Ley de grandes números

Teorema

Suponga que X_1, X_2, ..., X_n constituye una muestra aleatoria con media μ y media muestral $\bar{X}$. Entonces,

$$\lim_{n\to\infty} P(|\bar{X} - \mu| < \varepsilon) = 1$$

para $\varepsilon > 0$. Es decir que la media muestral converge en probabilidad a la media.

Teorema del limite central

Teorema

Sean Y_1, Y_2, ..., Y_n variables aleatorias independientes idénticamente distribuidas con $E(Y_i) = \mu$ y $V(Y_i) = \sigma^2 < \infty$. Definamos

$$U_n = \sqrt{n}\left(\frac{\overline{Y} - \mu}{\sigma}\right)$$

donde $\overline{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$.

Entonces, la función de distribución de U_n converge a la función de distribución normal estándar conforme $n \to \infty$.

Estadistica

En un sentido amplio, la estadistica se define como el arte y la ciencia de reunir datos, analizarlos, presentarlos e interpretarlos.

Se puede clasificar en:

- Estadistica Descriptiva

- Estadistica Inferencial

Estadistica Descriptiva

Estadistica Descriptiva:

Cuando se describe, analiza y representa un grupo de datos utilizando métodos numéricos y gráficos que resumen y presentan la información contenida en ellos.

No pretende ir más allá del conjunto de datos investigados.

Estadistica Inferencial

Estadistica Inferencial:

Cuando apoyándose en el cálculo de probabilidades y a partir de datos muéstrales, efectúa estimaciones, decisiones, predicciones y otras generalizaciones sobre un conjunto mayor de datos.

Definiciones básicas

Población

La población es el conjunto total de individuos u objetos con alguna caracteristica que es de interés estudiar.

Muestra

La muestra es un subconjunto de la población y contiene elementos en los cuales debe estudiarse la caracteristica de interés para la población.

Observación

La observación o dato es cada uno de los valores obtenidos para los elementos incluidos en la muestra.

Estadistico

Es cualquier medida descriptiva de una muestra y se usa como base para estimar el parámetro correspondiente de la población.

Varianza muestral

La varianza muestral se define como

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

Cuartiles

En estadistica, dada una distribución de datos estadisticos, los cuartiles son aquellos valores que la dividen en cuartos. El primer cuartil o cuartil inferior es aquel valor de la variable tal que la cuarta parte (25 %) de las observaciones son inferiores o iguales a él, y el resto (75 %) es superior o igual. El segundo cuartil es la mediana, ya que se trata del valor localizado en la mitad de la distribución. Finalmente, el tercer cuartil o cuartil superior es un valor tal que las tres cuartas partes de las observaciones son inferiores o iguales a él.

Diagramas de cajas

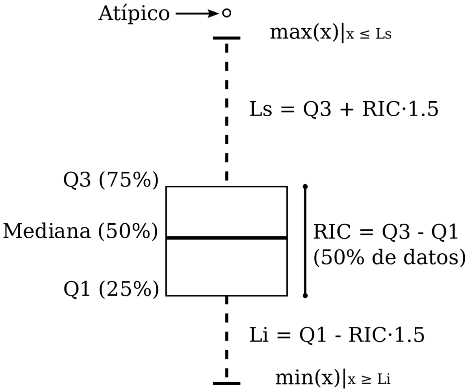

Figura 2: Diagrama de Caja (Boxplot)