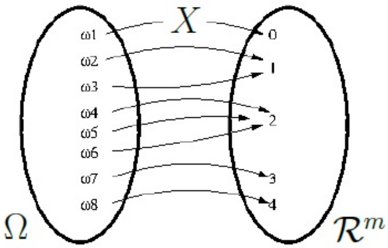
UFM UNIVERSIDAD FRANCISCO MARROQUÍN
VERITAS • LIBERTAS • JVSTITIA
FACULTAD de CIENCIAS ECONÓMICAS

Variables aleatorias

Alan Reyes-Figueroa

Elements of Machine Learning (Aula 03) 15.enero.2025

Variables aleatorias



Definición

Sea $(\Omega, \mathcal{F}, \mathbb{P})$ un espacio de probabilidad. Una ***variable aleatoria*** *(v.a.) es una función mesurable $X : \Omega \to \mathbb{R}$.*

Aquí mesurable significa que si $X : (\Omega, \mathcal{F}, \mathbb{P}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}), \mu)$, entonces la preimagen de cualquier elemento en $\mathcal{B}(\mathbb{R})$ es un elemento de $\mathcal{F}$. Esto es, X^{-1} lleva conjuntos mesurables de $\mathbb{R}$ (bajo la medida de Lebesgue μ), a conjuntos mesurables en $\mathcal{F}$ (bajo la probabilidad $\mathbb{P}$).

A los elementos de $\mathcal{B}(\mathbb{R})$ se les llama los *borelianos* de $\mathbb{R}$.

UFM UNIVERSIDAD FRANCISCO MARROQUÍN

Ejemplo
Elegimos al azar una persona de un grupo. De cada persona tenemos un registro de su edad, altura, peso, $\ldots$

Mapeamos cada persona ω a $X(\omega) = (X_1(\omega), \ldots, X_d(\omega))$, donde por ejemplo $X_1(\omega)$ representa su edad, $X_2(\omega)$ su altura, etc.

Si el grupo de personas corresponde a una base de datos, entonces X regresa los campos de interés de cada registro. Las variables $X_1, \ldots, X_d$ son variables aleatorias.
En este ejemplo llamaremos a X como una variable aleatoria (en realidad X es un vector aleatorio).

Observaciones:

- una variable aleatoria determina una relación determinística.
- una variable aleatoria induce una función de probabilidad.

Definimos $\mathbb{P}_X(\cdot)$ como

$$\mathbb{P}_X(A) = P(\{\omega \in \Omega : X(\omega) \in A\}).$$

Escribimos $\mathbb{P}_X(\cdot)$ como $\mathbb{P}(\cdot)$.
Por ejemplo, $\mathbb{P}(X = x)$ denota $\mathbb{P}_X(X = x) = \mathbb{P}(\{\omega : X(\omega) = x\})$.

$\mathbb{P}(X < a)$ denota $\mathbb{P}_X(X < a) = \mathbb{P}(\omega : X(\omega) < a\})$.

Caso discreto:

Definición

Diremos que X es una variable aleatoria ***discreta*** *si su contradominio $I = X(\Omega)$ es enumerable y $\mathbb{P}_X(i) = \mathbb{P}(X = i)$ existe para cada $i \in I$. (Comunmente se identifica el contradominio I con los naturales).*

Definición

Al conjunto de probabilidades $\{\mathbb{P}_X(i)\}_{i\in I}$ le llamamos la ***distribución*** *de X. (En general, a $\mathbb{P}_X$ se le llama la* ***función de masa de probabilidad****).*

Definición

Si $X \in \mathbb{R}$, llamamos a $F_X(x) = \mathbb{P}(X \leq x)$ la ***función de distribución*** *(acumulativa) de X.*

Caso continuo:

Definición

Considere la función $F : \mathbb{R} \to \mathbb{R}$, dada por

$$F_X(t) = \mathbb{P}(X \leq t).$$

*Diremos que X es una variable aleatoria **continua** si existe una función no-negativa $f_X : \mathbb{R} \to \mathbb{R}$, tal que*

$$F_X(t) = \mathbb{P}(X \leq t) = \int_{-\infty}^{t} f_X(x)\, dx.$$

Definición

*En ese caso, a la función f_X le llamamos la **densidad de probabilidad** de X.*

Propiedades

Obs! La función de densidad f_X no tiene por qué ser continua.

Ya sea en el caso discreto o continuo,

Definición

*Si $x \in \mathbb{R}$, llamamos a $F_X(x) = \mathbb{P}(X \leq x)$ la **función de distribución** (acumulativa) de X.*

En general, definimos la función de distribución para un vector aleatorio $X = (X_1, \ldots, X_d)$ como

$$F_X(x_1, \ldots, x_d) = \mathbb{P}(X_1 \leq x_1, \ldots, X_d \leq x_d), \quad \forall (x_1, \ldots, x_d) \in \mathbb{R}^d.$$

En este caso, llamamos a F_X la **función de distribución conjunta** de $X_1, \ldots, X_d$.

Propiedades

Propiedades de $\mathbb{P}_X$ y f_X:

Propiedad	X discreta	X continua
no-negativa	$\mathbb{P}_X(A) \geq 0$	$f_X(x) \geq 0$
suma total	$\sum_x \mathbb{P}_X(x) = 1$	$\int_{\mathbb{R}} f_X(x)\,dx = 1$
relación entre f_X y F_X	$\mathbb{P}(X = x) = F_X(x) - F_X(x^-)$	$f_X(x) = \dfrac{d}{dx} F_X(x)$
relación entre f_X y F_X	$F_X(x) = \sum_{t \leq x} \mathbb{P}(X = t)$	$F_X(x) = \int_{-\infty}^{x} f_X(t)\,dt$

Propiedades

Propiedades de F_X:

Propiedad	X discreta	X continua
limitada	$0 \leq F_X(x) \leq 1$	$0 \leq F_X(x) \leq 1$
monotonía	F_X no-decreciente	F_X no-decreciente
límite inferior	$F_X(t) = 0, \forall t < \min_{x \in I(\Omega)}$	$\lim_{x \to -\infty} F_X(x) = 0$
límite superior	$F_X(t) = 1, \forall t \geq \max_{x \in I(\Omega)}$	$\lim_{x \to +\infty} F_X(x) = 1$

Además, F_X tiene la propiedad de semi-continuidad inferior: F_X es continua por la derecha, con límites por la izquierda.

Variables aleatorias discretas

- distribución Uniforme $U[a..b]$,
- distribución Bernoulli $Ber(p)$,
- distribución Binomial $Binom(n, p)$,
- distribución Geométrica $Geom(p)$,
- distribución Poisson $Poisson(\lambda)$,

- distribución Rademacher $Rad(p)$,
- distribución Binomial Negativa $NB(r, p)$,
- distribución Hipergeométrica $Hypergeometric(N, K, n)$.

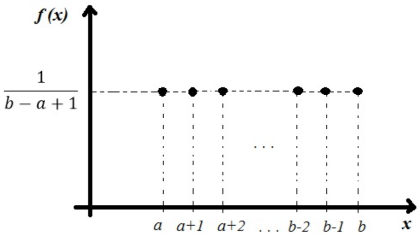
Variables aleatorias discretas

1. Distribución Uniforme

$$X \sim U[a..b] \;\Leftrightarrow\; \mathbb{P}(X = k) = \tfrac{1}{b-a+1}, \quad \text{para } k = a, a+1, \ldots, b.$$

Obs.

- Esta distribución depende de dos parámetros (de localización): a y b.
- El caso $a = b$, con $\mathbb{P}(X = a = b) = 1$ se llama una v.a. *degenerada*.



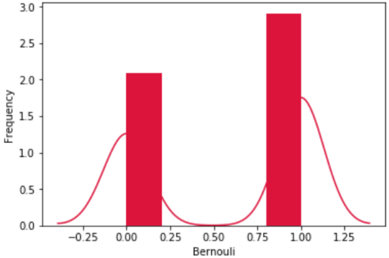
UFM UNIVERSIDAD FRANCISCO MARROQUÍN

2. Distribución Bernoulli

$$X \sim Ber(p) \;\Leftrightarrow\; \mathbb{P}(X=1)=p,\; \mathbb{P}(X=0)=1-p, \quad \text{para } 0 \leq p \leq 1.$$

Obs.

- La distribución es simétrica si, y sólo si, $p = 1/2$.
- $\mathbb{E}(X) = p$, $Var(X) = p(1-p)$.



UFM UNIVERSIDAD FRANCISCO MARROQUÍN

Variables aleatorias discretas

La distribución Bernoulli tiene una hermana gemela: la distribución de Rademacher.

$$X \sim Rad(p) \;\Leftrightarrow\; \mathbb{P}(X=1)=p,\; \mathbb{P}(X=-1)=1-p, \quad \text{para } 0 \leq p \leq 1.$$

Preguntas:

- ¿Cuál es la media y varianza de la distribución $Rad(p)$.
- Sean X, Y v.a., con $X \sim Ber(p)$ y $Y \sim Rad(p)$. Escribir X en términos de Y, y Y en términos de X.

La distribución de Bernoulli es importante para escribir situaciones donde se cuenta la ocurrencia de eventos. La variable $X \sim Ber(p)$ cuenta o indica la ocurrencia del evento de interés.

3. Distribución Binomial

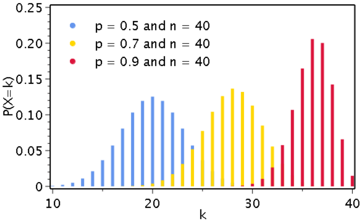
$$X \sim Binom(n,p) \;\Leftrightarrow\; \mathbb{P}(X=k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad \text{para } k = 0, 1, \ldots, n.$$

Interpretación: Si $\{X_i\}_{i=1}$ son v.a. *i.i.d.* con $X_i \sim Ber(p)$, entonces

$$X = \sum_{i=1}^{n} X_i \sim Binom(n,p).$$

Obs.

- La distribución es simétrica si, y sólo si, $p = 1/2$.
- $\mathbb{E}(X) = np$, $Var(X) = np(1-p)$.



Variables aleatorias discretas

4. Distribución Geométrica

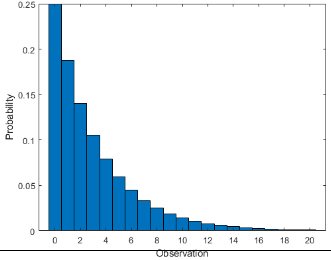
$$X \sim Geom(n,p) \;\Leftrightarrow\; \mathbb{P}(X=k) = p(1-p)^{k-1}, \quad \text{para } k = 1,2,3,\ldots$$

Interpretación: Si $\{X_i\}_{i=1}$ son v.a. *i.i.d.* con $X_i \sim Ber(p)$, entonces

$$X = \text{el momento del primer éxito en } \{X_i\} \sim Geom(p).$$

Obs.

- La probabilidad va decayendo en forma geométrica con k.
- $\mathbb{E}(X) = \frac{1}{p}$.



Variables aleatorias discretas

5. Distribución Poisson

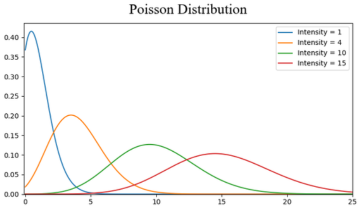
$$X \sim Poisson(\lambda) \;\Leftrightarrow\; \mathbb{P}(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}, \quad \text{para } k = 0, 1, 2, \ldots$$

En este caso,

$\mathbb{P}(X = k) =$ probabilidad de que el evento de interés ocurra k veces.

Obs.

- Cuenta el número de llegadas de un proceso con tiempos exponenciales $Exp(\lambda)$.
- $\mathbb{E}(X) = \lambda$. Representa el número esperado de veces que ocurra el fenómeno durante un intervalo dado.



UFM UNIVERSIDAD FRANCISCO MARROQUÍN

Variables aleatorias continuas

- distribución Uniforme $U[a..b]$,
- distribución Normal $\mathcal{N}(\mu, \sigma^2)$,
- distribución Lognormal $\mathcal{LN}(\mu, \sigma^2)$,

- distribución Exponencial $Exp(\lambda)$,
- distribución Erlang $Erlang(n, \lambda)$,
- distribución Gamma $\Gamma(\alpha, \beta)$,
- distribución Beta $Beta(\alpha, \beta)$,

- distribución Weibull,
- distribución Pareto,
- distribuciones de valores extremos.

1. Distribución Uniforme

$$X \sim U[a..b] \;\Leftrightarrow\; f_X(t) = \tfrac{1}{b-a}\,\mathbf{1}_{[a,b]}(t).$$

Obs. Aquí la función $\mathbf{1}_{[a,b]}$ es una *función indicadora* o *función característica*, que indica cuál es el soporte de la distribución.

Recordemos que para cualquier subconjunto $A \subseteq \mathbb{R}$

$$\mathbf{1}_A(t) = \begin{cases} 1, & t \in A; \\ 0, & t \notin A. \end{cases}$$

Preguntas: ¿Cuál es la función de distribución F_X? ¿$\mathbb{E}(X)$, $Var(X)$?

2. Distribución Exponencial Dado un parámetro $\lambda > 0$, la distribución exponencial tiene densidad

$$f_X(t) = \lambda e^{-\lambda t}\,\mathbf{1}_{[0,\infty)}(t), \quad F_X(t) = (1 - e^{-\lambda t})\,\mathbf{1}_{[0,\infty)}(t).$$

Obs.

- $\mathbb{E}(X) = \frac{1}{\lambda}$ y $Var(X) = \frac{1}{\lambda^2}$.
- En ocasiones, se parametriza en términos de su valor esperado $\theta = \frac{1}{\lambda}$:

$$f_X(t) = \frac{1}{\theta} e^{-t/\theta}\,\mathbf{1}_{[0,\infty)}(t).$$

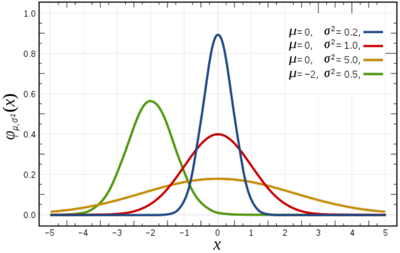
3. Distribución Normal Dados dos parámetros $\mu \in \mathbb{R}$, $\sigma > 0$, la distribución normal tiene densidad

$$f_X(t) = \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \mathbf{1}_{\mathbb{R}}(t).$$

Obs.

- $\mathbb{E}(X) = \mu$ y $Var(X) = \sigma^2$.
- La distribución es simétrica alrededor de μ.
- X no tiene una función de distribución elemental

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\, dt.$$



UFM UNIVERSIDAD FRANCISCO MARROQUÍN

Variables aleatorias continuas

Propiedades

1. *Si* $X \sim \mathcal{N}(\mu, \sigma^2)$*, se tiene que* $Z = \frac{X-\mu}{\sigma} \sim \mathcal{N}(,1)$.
2. *Si* $X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$ *y* $X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$*, y* X_1, X_2 *son independientes, entonces*
$$X_1 + X_2 \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$$
3. *Si* $X \sim \mathcal{N}(\mu, \sigma^2)$*, se tiene que* $-X \sim \mathcal{N}(-\mu, \sigma^2)$.
4. *En general, si* $X \sim \mathcal{N}(\mu, \sigma^2)$*, y* $a, b \in \mathbb{R}$*, entonces* $Y = aX + b$ *es normal, con*
$$Y \sim \mathcal{N}(a\mu + b, a^2\sigma^2).$$

Teorema (Desigualdad de Markov)

Si X es una v.a. no-negativa, $a > 0$, entonces

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}(X)}{a}.$$

Teorema (Desigualdad de Tchebyshev)

Si X es una v.a. con $\mathbb{E}(X)$ y $Var(X)$ finitas, $a > 0$, entonces

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq a) \leq \frac{Var(X)}{a}.$$

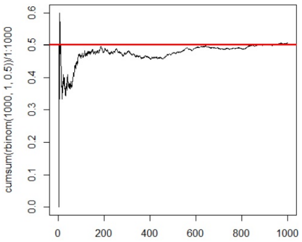
Teorema (Ley débil de los números grandes)

Sean $X_1, X_2, \ldots, X_n$ variables aleatorias i.i.d., con $\mathbb{E}(X_i) < \infty$. Entonces, para todo $\epsilon > 0$

$$\lim_{n\to\infty} \mathbb{P}\left(\left|\frac{X_1 + X_2 + \ldots + X_n}{n} - \mathbb{E}(X)\right| > \epsilon\right) = 0.$$

Interpretación:

Se repite el experimento n veces, con resultados X_i.
$\mathbb{E}(X) = \mathbb{P}(A)$, entonces $\frac{X_1+X_2+\ldots+X_n}{n}$ es el porcentaje de veces que ocurrió A.
La ley débil dice que el porcentaje de veces que A ocurrió en n repiticiones se aproxima a $\mathbb{E}(X)$.



UFM UNIVERSIDAD FRANCISCO MARROQUÍN

Teoremas importantes

Teorema (Teorema Central de Límite)

Sean $X_1, X_2, \ldots, X_n$ variables aleatorias i.i.d., con $\mathbb{E}(X_i) = \mu$, $Var(X_i) = \sigma^2$ finitas. Entonces

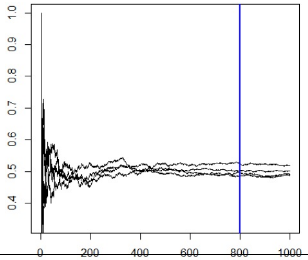
$$\lim_{n\to\infty} \mathbb{P}\Big(\frac{S_n/n - \mu}{\sigma/\sqrt{n}} \leq x\Big) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, dt.$$

<u>Consecuencias</u>:

$$\tfrac{S_n}{n} \sim \mathcal{N}(\mu, \tfrac{\sigma^2}{n}),$$

o equivalentemente

$$X_1 + \ldots + X_n = S_n \sim \mathcal{N}(n\mu, n\sigma^2).$$



UFM UNIVERSIDAD FRANCISCO MARROQUÍN