Algunos ejemplos de distribuciones

Alan Reyes-Figueroa

Introducción a la Ciencia de Datos (Aula 07) 30.enero.2021

Variables aleatorias discretas

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

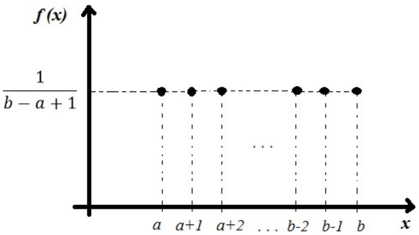
- distribución Rademacher $Rad(p)$,
- distribución Binomial Negativa $NB(r, p)$,
- distribución Hipergeométrica $Hypergeometric(N, K, n)$.

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*.

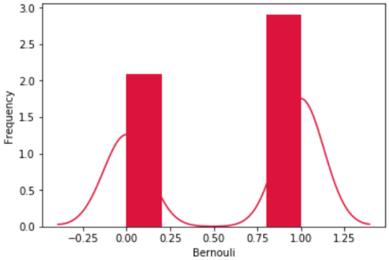


2. Distribución Bernoulli

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

Obs.

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



Variables aleatorias discretas

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

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

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

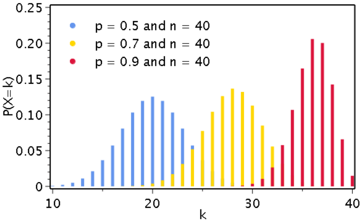
$$X \sim Binom(n,p) \;\Leftrightarrow\; \mathbb{P}(X=k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad \text{para } k = 0, 1, \ldots, n.$$

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, $p = 1/2$.
- $\mathbb{E}(X) = np$, $Var(X) = np(1-p)$.



UVG UNIVERSIDAD DEL VALLE DE GUATEMALA

4. Distribución Geométrica

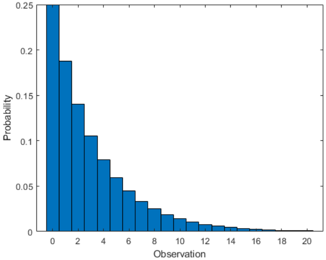
$$X \sim Geom(n,p) \;\Leftrightarrow\; \mathbb{P}(X=k) = p(1-p)^{k-1}, \quad \text{para } k = 1,2,3,\ldots$$

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.
- $\mathbb{E}(X) = \frac{1}{p}$.



5. Distribución Poisson

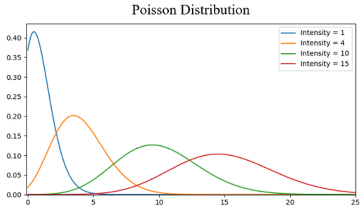
$$X \sim Poisson(\lambda) \;\Leftrightarrow\; \mathbb{P}(X=k) = \frac{e^{-\lambda}\lambda^k}{k!}, \quad \text{para } k = 0, 1, 2, \ldots$$

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.

- 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.



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 $\theta = \frac{1}{\lambda}$:

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

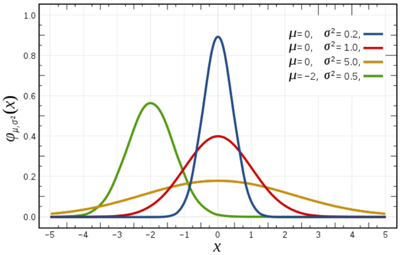
3. Distribución Normal Dados dos parámetros $\mu \in \mathbb{R}$, $\sigma > 0$, la distribución normal tiene densidad

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

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

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



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

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

Teoremas importantes

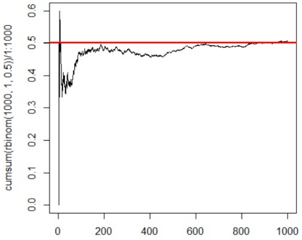
Teorema (Ley débil de los números grandes)

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 A ocurrió en n repiticiones se aproxima a $\mathbb{E}(X)$.



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

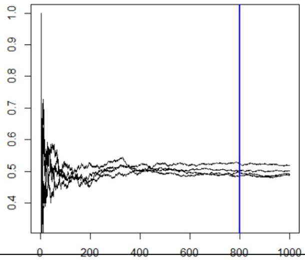
$$\lim_{n\to\infty} \mathbb{P}\Big(\frac{S_n/n - \mu}{\sigma^2/\sqrt{n}} \leq x\Big) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, dt.$$

Consecuencias:

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

o equivalentemente

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



Distribuciones multivariadas

Sea $X = (X_1, X_2, \ldots, X_d) \in \mathbb{R}^d$ un vector aleatorio (esto es, cada componente X_i es una variable aleatoria $X_i \Omega \to \mathbb{R}$).

Definición

*Definimos el **valor esperado** de X como el vector $\mu \in \mathbb{R}^d$ dado por*

$$\mathbb{E}(X) = \mu = (\mu_1, \mu_2, \ldots, \mu_d)^T \in \mathbb{R}^d,$$

donde $\mu_i = \mathbb{E}(X_i)$, para $i = 1, 2, \ldots, d$.

Definición

*Definimos la **varianza** de X como la matriz $\Sigma \in \mathbb{R}^{d \times d}$ dada por*

$$Var(X) = \Sigma = \big(Cov(X_i, X_j)\big)_{i,j}.$$

*La entrada (i, j) de esta matriz corresponde a la covarianza de las variables X_i y X_j. A Σ también se le conoce como la **matriz de covarianza** de X.*

Distribuciones multivariadas

Propiedades

Para cualquier vector aleatorio $X \in \mathbb{R}^d$, la matriz de covarianzas $\Sigma = Var(X)$ satisface

1. *Σ es simétrica (como consecuencia, tiene autovalores reales).*

2. *Σ es semi-definida positiva (todos sus autovalores son no-negativos). En particular, para todo vector $\mathbf{x} \in \mathbb{R}^d$, se cumple que $\mathbf{x}^T\Sigma\mathbf{x} \geq 0$.*

3. *La diagonal de Σ contiene a las varianzas $\sigma_i^2 = Var(X_i)$, para $i = 1, 2, \ldots, d$.*

Sea $X_1, X_2, \ldots, X_n \in \mathbb{R}^d$ es una muestra aleatoria de vectores i.i.d (independientes e identicamente distribuidos), todos con distribución X. Podemos codificar esta muestra dentro de una matriz, $\mathbb{X} \in \ltimes \times$, llamada la **matriz de datos** (cada dato de la muestra es un renglón de $\mathbb{X}$).

$$\mathbb{X} = \begin{pmatrix} x_{11} & x_{12} & \ldots & x_{1d} \\ x_{21} & x_{22} & \ldots & x_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \ldots & x_{nd} \end{pmatrix},$$

donde

$$X_i = (x_{i1}, x_{i2}, \ldots, x_{id}) \in \mathbb{R}^d, \quad i = 1, 2, \ldots, n.$$

Distribuciones multivariadas

Observe que la $i-$ésima columna de $\mathbb{X}$ corresponde a una muestra (de tamaño n) de la variable aleatoria X_i. Podemos entonces restar a cada columna su correspondiente media $\mu_i = \mathbb{E}(X_i)$. Así, obtenemos una versión centrada de la matriz de datos:

$$\mathbb{X}_c = \mathbb{X} - \mu = \begin{pmatrix} x_{11} - \mu_1 & x_{12} - \mu_2 & \ldots & x_{1d} - \mu_d \\ x_{21} - \mu_1 & x_{22} - \mu_2 & \ldots & x_{2d} - \mu_d \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} - \mu_1 & x_{n2} - \mu_2 & \ldots & x_{nd} - \mu_d \end{pmatrix}.$$

Es posible mostrar (con las propiedades de la página siguiente) que la matriz de covarianzas empírica (muestral) se puede escribir como

$$\Sigma = Var(X) = \mathbb{X}_c^T \mathbb{X}_c.$$

Propiedades

Sea $X, Y \in \mathbb{R}^d$ *vectores aleatorios,* $a, b \in \mathbb{R}$*,* $c \in \mathbb{R}^d$ *constantes,* $A \in \mathbb{R}^{p\times d}$ *una matriz constante. Entonces*

1. $\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y)$,
2. $\mathbb{E}(c) = c$,
3. $\mathbb{E}(\mathbb{E}(X)) = \mathbb{E}(X)$,
4. $Var(aX) = a^2 Var(X)$,
5. $\mathbf{Var}(\mathbf{AX}) = \mathbf{A}^T\mathbf{Var}(\mathbf{X})\mathbf{A}$,
6. $Var(aX + bY) = a^2 Var(X) + b^2 Var(Y) + 2ab Cov(X, Y)$,
7. *Si* $X \perp Y$*, entonces* $Cov(X, Y) = 0$,
8. *Si* $X \perp Y$*, entonces* $Var(aX + bY) = a^2 Var(X) + b^2 Var(Y)$.

Distribución normal multivariada

Sea $X = (X_1, X_2, \ldots, X_d) \in \mathbb{R}^d$ un vector aleatorio. Decimos que X sigue una **distribución normal multivariada** $\mathcal{N}_d(\mu, \Sigma)$ si su densidad está dada por

$$f_X(\mathbf{x}) = \int_{\mathbb{R}^d} \frac{1}{(2\pi)^{d/2}\sqrt{|\Sigma|}} \exp\Big(-\frac{1}{2}(\mathbf{x}-\mu)^T\Sigma^{-1}(\mathbf{x}-\mu)\Big)\, d\mathbf{x}.$$

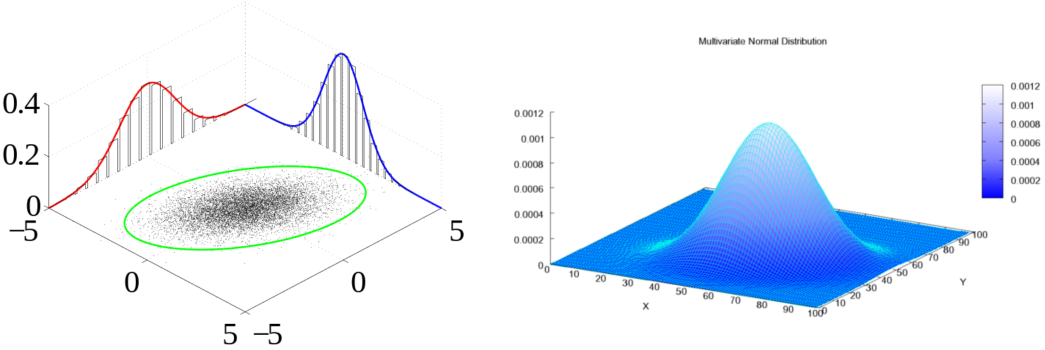
Aquí,

$$\mathbb{E}(X) = \mu = (\mu_1, \mu_2, \ldots, \mu_d)^T,$$

y

$$Var(X) = \Sigma = \begin{pmatrix} Cov(X_1,X_1) & Cov(X_1,X_2) & \ldots & Cov(X_1,X_d) \\ Cov(X_2,X_1) & Cov(X_2,X_2) & \ldots & Cov(X_2,X_d) \\ \vdots & \vdots & \ddots & \vdots \\ Cov(X_d,X_1) & Cov(X_d,X_2) & \ldots & Cov(X_d,X_d) \end{pmatrix}.$$

Distribución normal multivariada



Densidad de una normal bivariada: (a) como nube de puntos, (b) como función.

Distribución normal multivariada

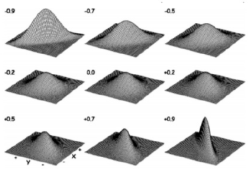
Típicamente la matriz Σ proporciona información sobre la relación entre las variables componentes.

$Cov(X) = [Cov(X_i, X_j)]$ es una matriz simétrica y pos. definida.

Caso $d = 2$:

$$\begin{pmatrix} Var(X_1) & Cov(X_1, X_2) \\ Cov(X_1, X_2) & Var(X_2) \end{pmatrix} = \begin{pmatrix} \sigma_{X_1}^2 & Cor(X_1, X_2)\sigma_{X_1}\sigma_{X_2} \\ Cor(X_1, X_2)\sigma_{X_1}\sigma_{X_2} & \sigma_{X_2}^2 \end{pmatrix}$$

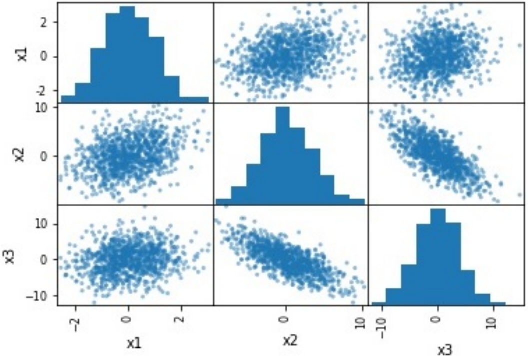
Cambiar $\rho = Cor(X_1, X_2)$:



`http://personal.kenyon.edu/hartlaub/MellonProject/images/Bivariate52.gif`.

Distribución normal multivariada

Una forma práctica de ver esta información de covarianza o correlación entre las componentes es a través de *pair-plots*.



Pairplot de una muestra para una normal 3-variada.

Distribución normal multivariada

Problema: ¿Cómo generar una muestra de una distribución normal d-variada con μ y Σ específicas?

Algoritmo (o receta):

1. Generar d muestras (de tamaño n), independientes, de distribuciones normales estándar $Z_1, Z_2, \ldots, Z_d \in \mathbb{R}^n$, y construir una matriz de datos $\mathbb{Z}$ con las muestras Z_i como columnas.
Como son independientes y estándar el vector $Z = (Z_1, \ldots, Z_d)$ sigue una distribución normal estándar $\mathcal{N}_d(\mathbf{0}, I_d)$.

2. Asegurarse que la matriz Σ es simétrica y positiva definida. Luego, construir descomposición de Cholesky $L^T L = \Sigma$, (el algoritmo que vieron en análisis numérico).

3. Construir la variable aleatoria $X = LZ + \mu$, la cual tiene una matriz de datos dada por $\mathbb{X} = L\mathbb{Z} + \mu$ (la muestra que queremos). De las propiedades anteriores, tenemos que $\mathbb{E}(Z) = \mu$ y $Var(X) = L^T I_d L = L^T L = \Sigma$.