Repaso: Probabilidad y Variables Aleatorias

Cecilia Galarza

Procesos Estocásticos
Facultad de Ingeniería, Universidad de Buenos Aires

Modelo probabilístico

- Un modelo matemático es una herramienta de análisis de la realidad que se construye sobre un conjunto de hipótesis.

- Ej: *Sea x_T la posición de un móvil luego de T segundos de haber pasado por x_0 desplazándose con velocidad constante v*

  $$x_T = vT + x_0.$$

  Si x_0 es conocido, entonces el modelo es determinístico, pero si x_0 es desconocido o no puede ser observado, es posible considerar un modelo probabilístico que permita analizar el comportamiento del móvil.

- Algunos ejemplos de fenómenos probabilístico son:
  - Transmisión de señales en redes de comunicación inalámbrica;
  - Procesamiento de señales digitales de voz, video, imagen;
  - Análisis de la información generada en la web.

Espacio de probabilidad

Siguiendo lo aprendido en Proba, caracterizamos un experimento aleatorio mediante un *espacio de probabilidad* $(\Xi, \mathcal{A}, \mathbb{P})$, donde:

- Ξ, *espacio muestral*, es el conjunto de todos los posibles resultados del experimento;
- $\mathcal{A}$, *conjunto de eventos*, contiene subconjuntos de Ξ tal que
  - $\Xi \in \mathcal{A}$;
  - Si $A \in \mathcal{A}$, entonces $A^c \in \mathcal{A}$;
  - Si $A_1, A_2, \ldots \in \mathcal{A}$, entonces $A_1 \cup A_2 \cup \ldots \in \mathcal{A}$.
- $\mathbb{P}$ es la función de probabilidad asociada al experimento que satisface
  - $\forall A \in \mathcal{A}, \mathbb{P}(A) \geq 0$.
  - $\mathbb{P}(\Xi) = 1$.
  - $\forall A_1, A_2$ tales que $A_1 \cap A_2 = \emptyset$, $\mathbb{P}(A_1 \cup A_2) = \mathbb{P}(A_1) + \mathbb{P}(A_2)$.
  - Sea $\{A_i\}_{i=1}^{+\infty}$ una secuencia de subconjuntos tales que $A_i \cap A_j = \emptyset$ con $i \neq j$. Luego, $\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$.

Espacio de probabilidad

- El conjunto de eventos: $\mathcal{A}$
  - Contiene subconjuntos de Ξ llamados *eventos*.
  - Los eventos se eligen de acuerdo a aquello que se observa y se quiere asignar una probabilidad de ocurrencia.

- Propiedades de $\mathbb{P}$:
  - $\mathbb{P}(\emptyset) = 0$.
  - $0 \leq \mathbb{P}(A) \leq 1$ para todo $A \in \mathcal{A}$.
  - $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$ para todo $A \in \mathcal{A}$.
  - Sean $A, B \in \mathcal{A}$ tal que $B \subseteq A$. Entonces, $\mathbb{P}(B) \leq P(A)$.
  - $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B) \leq \mathbb{P}(A) + \mathbb{P}(B)$ para todo $A, B \in \mathcal{A}$.
  - *Desigualdad de Boole o cota de la unión.* De forma más general, para cualesquiera $A_i \in \mathcal{A}$ tenemos que

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) \leq \sum_{i=1}^{\infty} \mathbb{P}(A_i).$$

Probabilidad condicional y regla de Bayes

- Cuando sabemos que un evento A ocurrió y queremos evaluar la probabilidad de otro evento B, utilizamos la probabilidad condicional de B dado A, dada por

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

  Notar que para que la probabilidad condicional esté bien definida, necesitamos que $\mathbb{P}(A) > 0$. O sea, no podemos condicionar a eventos de probabilidad nula.

- Observando que $\mathbb{P}(A \cap B) = \mathbb{P}(B|A)\mathbb{P}(A) = \mathbb{P}(A|B)\mathbb{P}(B)$, obtenemos la famosa *regla de Bayes*:

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

  asumiendo que $\mathbb{P}(B) \neq 0$.

Regla de la cadena de probabilidades

- Generalizando la relación $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B|A)$ a n eventos $A_1, \ldots, A_n$, obtenemos la regla de la cadena para probabilidades:

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

- Probamos la igualdad mediante inducción matemática. El caso base se cumple por definición de probabilidad condicional. Para el caso inductivo, suponemos que

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

La probabilidad del evento $A_1 \cap A_2 \ldots \cap A_n$ resulta:

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

Ley de probabilidad total y teorema de Bayes

- Sea $\{A_1, \ldots, A_n\}$ una partición de Ξ, es decir
  - $A_i \neq \emptyset$ para todo i.
  - $\bigcup_{i=1}^{n} A_i = \Xi$.
  - $A_i \cap A_j = \emptyset$ para todo $i \neq j$.

- Entonces, $\forall B \in \mathcal{A}$ tenemos:

$$\mathbb{P}(B) = \sum_{i=1}^{n} \mathbb{P}(B \cap A_i) = \sum_{i=1}^{n} \mathbb{P}(B|A_i)\mathbb{P}(A_i).$$

  - *Demostración*:

$$B = B \cap \Xi = B \cap (A_1 \cup A_2 \cup \ldots A_n) = (B \cap A_1) \cup \ldots (B \cap A_n).$$

Como A_i son disjuntos dos a dos, entonces $B \cap A_i$ también lo son y el resultado se obtiene a partir de las propiedades de $\mathbb{P}$.

Eventos independientes

- Decimos que dos eventos A y B son independientes si saber que ocurrió uno de ellos no afecta la probabilidad del otro, es decir,

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

- A partir de la relación anterior y la definición de probabilidad condicional obtenemos la definición más habitual de independencia:

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

Eventos independientes

- La independencia de n eventos $A_1, \ldots, A_n$ se puede definir de forma inductiva. Decimos que $A_1, \ldots, A_n$ son independientes si cualesquiera k de ellos son independientes y

$$\mathbb{P}\left(\bigcap_{i=1}^{n} A_i\right) = \prod_{i=1}^{n} \mathbb{P}(A_i).$$

  Notar que el concepto de independencia es más fuerte que el de independencia dos a dos.

- Por ejemplo, A_1, A_2 y A_3 son independientes si los eventos A_i y A_j son independientes para cualesquiera $i \neq j$ y

$$\mathbb{P}(A_1 \cap A_2 \cap A_3) = \mathbb{P}(A_1)\mathbb{P}(A_2)\mathbb{P}(A_3).$$

Independencia condicional

- Dos eventos A y B son condicionalmente independientes dado otro evento C si se cumple que
$$\mathbb{P}(A|B \cap C) = \mathbb{P}(A|C).$$
Notar de nuevo que asumimos que $\mathbb{P}(C) > 0$ y $\mathbb{P}(B \cap C) > 0$ para que las probabilidades condicionales estén bien definidas.

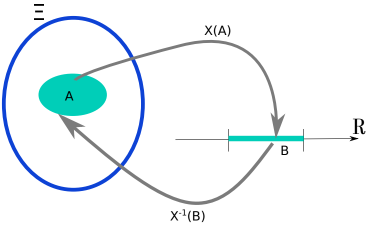
- De forma equivalente, la independencia condicional se puede expresar como
$$\mathbb{P}(A \cap B|C) = \mathbb{P}(A|C)\mathbb{P}(B|C).$$

- Para ver que se cumple esta igualdad podemos operar de la siguiente forma:
$$\mathbb{P}(A \cap B|C) = \frac{\mathbb{P}(A \cap B \cap C)}{\mathbb{P}(C)} = \mathbb{P}(A|B, C)\frac{\mathbb{P}(B \cap C)}{\mathbb{P}(C)} = \mathbb{P}(A|C)\mathbb{P}(B|C).$$

Variable aleatoria

Una variable aleatoria $X : \Xi \to \mathbb{R}$ es una función que mapea el espacio muestral en $\mathbb{R}$

- Es posible asociar una función de probabilidad a una variable aleatoria utilizando la probabilidad del evento asociado a X^{-1}.
- Para definir un evento en $\mathbb{R}$ utilizamos la preimagen dada por la función X^{-1}. Haciendo abuso de notación, en lugar de utilizar $\mathbb{P}(X^{-1}(B))$, vamos a escribir $\mathbb{P}(X \in B)$.



$$\begin{aligned}\mathbb{P}[X(\xi) \in B] &= \mathbb{P}\left[\{\omega \in \Xi : X(\xi) \in B\}\right] \\ &= \mathbb{P}(X^{-1}(B)) = \mathbb{P}[A]\end{aligned}$$

Función de distribución

- Para caracterizar a una variable aleatoria utilizamos la *función de distribución* (CDF, *Cumulative Distribution Function*) :

$$
\begin{aligned}
F_X &: \mathbb{R} \to \mathbb{R} \\
F_X(x) &= \mathbb{P}(X \leq x) = \mathbb{P}(\xi \in \Xi : X(\xi) \leq x).
\end{aligned}
$$

- La función de distribución de una variable aleatoria siempre está bien definida y nos permite calcular la probabilidad de cualquier evento.

Funciones de distribución: Propiedades

- *Monótona no decreciente.* Si $x_1 \leq x_2$, entonces $F_X(x_1) \leq F_X(x_2)$.
- *Continua por derecha.* $\lim_{x \to x_0^+} F_X(x) = F_X(x_0^+) = F_X(x_0)$.
- *Límite por izquierda.* $\lim_{x \to -\infty} F_X(x) = F_X(-\infty) = 0$.
- *Límite por derecha.* $\lim_{x \to +\infty} F_X(x) = F_X(\infty) = 1$.
- *Probabilidad de intervalo semiabierto.*
  $\mathbb{P}(x_1 < X \leq x_2) = F_X(x_2) - F_X(x_1)$ para cualesquiera $x_1 \leq x_2$.
- *Probabilidad de intervalo cerrado.*
  $\mathbb{P}(x_1 \leq X \leq x_2) = F_X(x_2) - F_X(x_1^-)$ para cualesquiera $x_1 \leq x_2$.
- *Probabilidad de un punto.* $\mathbb{P}(X = x_0) = F_X(x_0) - F_X(x_0^-)$.

Variable aleatoria continua y función de densidad de probabilidad

X es una VA continua (VAC) si F_X es derivable. En ese caso, definimos su función de densidad de probababilidad (PDF *Probability Density Function*) como

$$f_X(x) = \frac{dF_X}{dx}(x).$$

Función de densidad de probabilidad: Propiedades

- Si X es VAC, $\mathbb{P}(X = x_0) = F_X(x_0) - F_X(x_o^-) = 0$.
- Como la CDF es monótona no decreciente, entonces $f_X(x) \geq 0$ para todo $x \in \mathbb{R}$.
- Integrando a partir de la definición de la PDF, tenemos que

$$F_X(x) = \int_{-\infty}^{x} f_X(u)du.$$

- A partir de las propiedades de la CDF, obtenemos

$$\int_{-\infty}^{\infty} f_X(x)dx = F_X(+\infty) = 1, \qquad \mathbb{P}(x_1 < X \leq x_2) = \int_{x_1}^{x_2} f_X(u)du.$$

Ejemplos de PDF continuas

- *Uniforme.* La VA uniforme de parámetros a, b con $a < b$ tiene la siguiente PDF:

$$f_X(x) = \frac{1}{b-a}, \qquad a < x < b.$$

- *Exponencial.* La VA exponencial de parámetro $\lambda > 0$ tiene la siguiente PDF:

$$f_X(x) = \lambda\, e^{-\lambda x}, \qquad x \geq 0.$$

- *Gaussiana.* La VA Gaussiana de parámetros μ y $\sigma > 0$ tiene la siguiente PDF:

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \qquad x \in \mathbb{R}.$$

- *Rayleigh.* La VA Rayleigh de parámetro $\sigma > 0$ tiene la siguiente PDF:

$$f_X(x) = \frac{x}{\sigma^2} e^{-\frac{x^2}{2\sigma^2}}, \qquad x \geq 0.$$

Variable aleatoria discreta y función de masa de probabilidad

X es una VA discreta (VAD) si F_X es constante a tramos y tiene una cantidad finita o infinita numerable de discontinuidades. En este caso, sea $\mathcal{X}$ el conjunto de discontinuidades de F_X. Luego, definimos la función de masa de probabilidad (PMF, *Probability Mass Function*) como

$$p_X(x) = F_X(x) - F_X(x^-), \quad x \in \mathcal{X},$$

Propiedades de la PMF

- $\mathbb{P}(X = x_0) = p_X(x_0)$ para todo $x_0 \in \mathcal{X}$ y $\mathbb{P}(X = x_0) = 0$ si $x_0 \notin \mathcal{X}$.
- Como la CDF es monótona no decrecienta, $p_X(x) > 0$ para todo $x \in \mathcal{X}$.
- A partir de la definición de la PMF, tenemos que

$$F_X(x) = \sum_{u \in \mathcal{X} : u \leq x} p_X(u).$$

- A partir de las propiedades de la CDF, obtenemos

$$\sum_{x \in \mathcal{X}} p_X(x) = F_X(+\infty) = 1, \qquad \mathbb{P}(x_1 < X \leq x_2) = \sum_{u \in \mathcal{X} : x_1 < u \leq x_2} p_X(u).$$

- Asociamos a la VAD la PDF generalizada:

$$f_X(x) = \sum_{y \in \mathcal{X}} p_X(y) \delta(x - y)$$

Ejemplos de PMF discretas

- *Bernoulli.* $X \sim \text{Ber}(p)$, $p \in [0,1]$:

$$p_X(x) = \begin{cases} 1-p & \text{si } x = 0, \\ p & \text{si } x = 1. \end{cases}$$

- *Uniforme (discreta).* $X \sim U(N)$:

$$p_X(x) = \frac{1}{N}, \qquad x = 1, \ldots, N.$$

- *Geométrica.* $X \sim Geo(p)$, $p \in (0,1]$:

$$p_X(x) = (1-p)^{k-1} p, \qquad x \in \mathbb{N}_0.$$

- *Binomial.* $X \sim Bin(p)$, $p \in [0,1]$:

$$p_X(x) = \binom{n}{k} p^k (1-p)^{n-k}, \qquad x = 0, 1, \ldots, n.$$

- *Poisson.* $X \sim Poisson(\lambda)$, $\lambda > 0$:

$$p_X(x) = \frac{\lambda^x e^{-\lambda}}{x!}, \qquad x = 1, 2 \ldots$$

Esperanza de una VA

- Sea una función $g : \mathbb{R} \to \mathbb{R}$.
- Luego, aplicando el operador esperanza tenemos:

$$\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x) dF_X(x) = \int_{-\infty}^{\infty} g(x) f_X(x) dx.$$

  donde $f_X(x)$ es la PDF generalizada de X.
- Si X es VAD, tenemos que

$$f_X(x) = \sum_{y \in \mathcal{X}} p_X(y) \delta(x - y),$$

  y la integral resulta

$$\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x) \left[\sum_{y \in \mathcal{X}} p_X(y) \delta(x - y) \right] dx = \sum_{y \in \mathcal{X}} p_X(y) g(y).$$

Propiedades del operador esperanza

- *Existencia.* El resultado del operador esperanza no siempre existe. Incluso si existe puede ser infinito.
  - Ejemplo:VA Cauchy $X \sim \mathcal{C}(0,1)$

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x\, f_X(x)dx = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{x}{1+x^2}dx = \frac{1}{2\pi}\log(1+x^2)\Big|_{-\infty}^{\infty} =?,$$

$$\mathbb{E}[X^2] = \int_{-\infty}^{\infty} x^2 f_X(x)dx = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{x^2}{1+x^2}dx = \frac{x-\arctan(x)}{\pi}\Big|_{-\infty}^{\infty} = \infty.$$

- *Linealidad.* El operador $\mathbb{E}$ es lineal. Para toda función $g(.)$ y $h(.)$,

$$\mathbb{E}[\alpha g(X) + \beta h(X)] = \alpha\mathbb{E}[g(X)] + \beta\mathbb{E}[h(X)] \quad , \quad \forall \alpha, \beta \in \mathbb{R}$$

En particular, si $g(X) = X$ y $h(X) = 1$,

$$\mathbb{E}[\alpha X + \beta] = \alpha\mathbb{E}[X] + \beta.$$

Desigualdad de Jensen

- En general, salvo la excepción para funciones lineales o afines recién vista, tenemos que

$$\mathbb{E}[g(X)] \neq g(\mathbb{E}[X]).$$

- Decimos que $g : D \to \mathbb{R}$ es convexa si D es un conjunto convexo[1] y satisface $g(\alpha x_1 + (1-\alpha)x_2) \leq \alpha g(x_1) + (1-\alpha)g(x_2)$ para todo $\alpha \in [0, 1]$ y $x_1, x_2 \in D$. Si g es convexa decimos que $-g$ es cóncava.
- *Desigualdad de Jensen.* Si g es una función convexa, vale lo siguiente

$$\mathbb{E}[g(X)] \geq g(\mathbb{E}[X]).$$

Si, en cambio, g es cóncava, tenemos $\mathbb{E}[g(X)] \leq g(\mathbb{E}[X])$.

[1] Esto significa que si $x_1, x_2 \in D$, entonces $\alpha x_1 + (1-\alpha)x_2 \in D$ para todo $\alpha \in [0, 1]$.

Desigualdad de Jensen: Ejemplos de aplicación

$$\mathbb{E}[X^2] \geq (\mathbb{E}[X])^2$$

$$\mathbb{E}[e^X] \geq e^{\mathbb{E}[X]},$$

$$\mathbb{E}[\sqrt{X}] \leq \sqrt{\mathbb{E}[X]}$$

$$\mathbb{E}[\log(X)] \leq \log(\mathbb{E}[X]).$$

Momentos de una variable aleatoria

- Un caso particular de interés es considerar funciones de la forma $g(x) = x^n$ o $g(x) = (x-a)^n$ con $n \in \mathbb{N}$ y $a \in \mathbb{R}$. Esto nos lleva al concepto de momentos.
- El *momento* de orden n de una VA X lo denotamos m_n y se define como
$$m_n = \mathbb{E}[X^n] = \int_{-\infty}^{\infty} x^n f_X(x)dx.$$
- El *momento centrado* de orden n de una VA X se define como
$$\eta_n = \mathbb{E}[(X-m_1)^n] = \int_{-\infty}^{\infty} (x-m_1)^n f_X(x)dx.$$

Media de una variable aleatoria

La media es el momento de primer orden de una VA. En general a la media de una VA X la vamos a denotar μ_X en lugar de m_1:

$$\mu_X = \mathbb{E}[X] = \int_{-\infty}^{\infty} x\, f_X(x) dx.$$

Si la VA es discreta, como vimos antes, podemos simplificar esta expresión a

$$\mu_X = \mathbb{E}[X] = \sum_{x \in \mathcal{X}} x\, p_X(x).$$

Varianza de una variable aleatoria

La varianza de una VA es el momento centrado de segundo orden. En general, a la varianza de una VA X la vamos a denotar σ_X^2 o $\mathbb{V}(X)$ en lugar de η_2:

$$\sigma_X^2 = \mathbb{V}(X) = \mathbb{E}[(X - \mu_X)^2] = \int_{-\infty}^{\infty} (x - \mu_X)^2 \, f_X(x) dx.$$

De nuevo, si X es una VAD, podemos simplificar esta expresión y nos queda

$$\sigma_X^2 = \mathbb{V}(X) = \sum_{x \in \mathcal{X}} (x - \mu_X)^2 \, p_X(x).$$

Notar que $\sigma_X^2 \geq 0$.

- El desvío estándar $\sigma_X = \sqrt{\sigma_X^2}$ es una medida de la dispersión promedio que tienen las realizaciones de X alrededor de la media μ_X cuando la cantidad de realizaciones N es suficientemente grande.

Variables aleatorias degeneradas

Definición. X es una VA degenerada si $\mathbb{V}(X) = \sigma_X^2 = 0$.

Teorema

Una X es una VA degenerada si y solo si $\mathbb{P}(X = \mu_X) = 1$.

Demostración:

- ($\Rightarrow$) Según la desigualdad de Chebyshev y dado que $\sigma_X^2 = 0$,

$$\mathbb{P}(|X - \mu_X| \geq a) = 0, \qquad \forall a > 0.$$

  Esto implica que $\mathbb{P}(X = \mu_X) = 1$.

- ($\Leftarrow$) La PMF de X es $p_X(\mu_X) = 1$, $p_X(x) = 0$ para $x \neq \mu_X$. Luego,

$$\mathbb{V}(X) = \mathbb{E}[(X - \mu_X)^2] = p_X(\mu_X)(\mu_X - \mu_X)^2 = 0.$$

Transformación de una variable aleatoria

- Sea una función $g : \mathbb{R} \to \mathbb{R}$.

- Definimos una nueva VA $Y = g(X)$, es decir,

$$Y(\xi) = g(X(\xi)), \qquad \xi \in \Xi.$$

- En general, tenemos que la

$$F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(g(X) \leq y) = \int_{R_y} f(x)dx,$$

donde $R_y = \{x : g(x) \leq y\}$. Del mismo modo, podemos definir la PDF (posiblemente generalizada) de Y como

$$f_Y(y) = \frac{dF_Y}{dy}(y) = \frac{d}{dy}\int_{R_y} f(x)dx.$$

Transformación lineal o afín

- Sea $g(X) = aX + b$, donde $a, b \in \mathbb{R}$ son parámetros constantes.
- La CDF de Y resulta

$$F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(aX + b \leq y) = \mathbb{P}(aX \leq y - b).$$

  - Si $a = 0$, tenemos que $Y = b$ es una VA degenerada. $F_Y(y) = u(y - b)$ y no depende de F_X.
  - Si $a > 0$,

$$F_Y(y) = \mathbb{P}\left(X \leq \frac{y-b}{a}\right) = F_X\left(\frac{y-b}{a}\right).$$

  - Si $a < 0$,

$$F_Y(y) = \mathbb{P}\left(X \geq \frac{y-b}{a}\right) = 1 - F_X\left(\frac{y-b}{a}\right).$$

Tranformación monótona y suave

- Si $g(.)$ es una función monótona y suave, existe g^{-1}.
- Si g es creciente ($g'(y) > 0$)

$$F_Y(y) = \mathbb{P}(g(X) \leq y) = \mathbb{P}(X \leq g^{-1}(y)) = F_X(g^{-1}(y)).$$

$$f_Y(y) = \frac{\partial}{\partial y} F_X(g^{-1}(y)) = f_X(g^{-1}(y)) \frac{d}{dy}(g^{-1})(y) = \frac{f_X(g^{-1}(y))}{g'(y)}.$$

- Si g es decreciente ($g'(y) < 0$)

$$F_Y(y) = \mathbb{P}(g(X) \leq y) = \mathbb{P}(X \geq g^{-1}(y)) = 1 - F_X(g^{-1}(y)).$$

$$f_Y(y) = -\frac{\partial}{\partial y} F_X(g^{-1}(y)) = -\frac{f_X(g^{-1}(y))}{g'(y)}.$$

- En ambos casos

$$f_Y(y) = \frac{f_X(g^{-1}(y))}{|g'(y)|}.$$

Transformación general

- Consideremos una transformación que no es ni lineal ni monótona, por ejemplo, $Y = X^2$.

$$F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(X^2 \leq y).$$

  - Si $y < 0$, $F_Y(y) = \mathbb{P}(X^2 \leq y < 0) = 0$.
  - Si $y = 0$, $F_Y(y) = \mathbb{P}(X^2 \leq 0) = \mathbb{P}(X = 0) = F_X(0) - F_X(0^-)$.
  - Si $y > 0$,
    $F_Y(y) = \mathbb{P}(X^2 \leq y) = \mathbb{P}(-\sqrt{y} \leq X \leq \sqrt{y}) = F_X(\sqrt{y}) - F_X(-\sqrt{y})$.
- Derivando esta expresión,

$$f_Y(y) = \frac{f_X(\sqrt{y}) + f_X(-\sqrt{y})}{2\sqrt{y}} \mathbb{1}\{y > 0\} + \mathbb{P}(X = 0)\, \delta(y).$$

La PDF resulta con dos términos para $y > 0$ y uno para $y = 0$, lo cual está relacionado con la cantidad de soluciones de la ecuación $y = x^2$. En general, si $y = g(x)$ tiene k soluciones, habrá k términos en la PDF.