Universidad de Chile
Facultad De Ciencias Físicas y Matemáticas
Departamento de Ingeniería Matemática
MA3403 - Probabilidades y Estadística

# Modelos de Variables Aleatorias

Alberto Vera Azócar, albvera@ing.uchile.cl

## 1. Nociones Generales

Para dar una definición formal de variable aleatoria necesitaríamos primero introducir nociones de medibilidad y de espacios de probabilidad, esto se escapa al alcance del curso, por lo tanto aceptaremos una definición mucho menos rigurosa, pero que será cierta para todos los casos que nos interesen.

**Definición 1.1.** *Sea $\Omega$ un espacio de eventos con una medida de probabilidad $\mathbb{P}$. Se dirá variable aleatoria a cualquier función $X : \Omega \to \mathbb{R}$.*

Lo que nos gustará calcular en el contexto de v.a. será por ejemplo "que $X$ tome un valor menor o igual a cero", lo que denotamos como $\mathbb{P}(X \leq 0)$, o bien "que $X$ tome un valor en el conjunto $A$", lo que denotamos como $\mathbb{P}(X \in A)$. Esto motiva a la siguiente definición:

**Definición 1.2.** *Sea $X : \Omega \to \mathbb{R}$ una v.a., definimos su función de distribución (o de probabilidad acumulada) como*

$$F_X(x) := \mathbb{P}(X \leq x)$$

Supongamos que nos interesa calcular la probabilidad de que $X$ tome un valor en el intervalo $[a, b]$, con $a < b$. A éste evento lo denotaremos por $\{a \leq X \leq b\}$. Utilizaremos la lógica de conjuntos, recordando que si $A \subseteq B$ entonces $\mathbb{P}(B \setminus A) = \mathbb{P}(B) - \mathbb{P}(A)$.

$$\begin{aligned}
\mathbb{P}(\{a \leq X \leq b\}) &= \mathbb{P}(\{X \leq b\} \setminus \{X < a\}) \\
&= \mathbb{P}(\{X \leq b\}) - \mathbb{P}(\{X < a\}) \\
&= \mathbb{P}(\{X \leq b\}) - \mathbb{P}(\{X \leq a\} \setminus \{X = a\}) \\
&= \mathbb{P}(\{X \leq b\}) - [\mathbb{P}(\{X \leq a\}) - \mathbb{P}(\{X = a\})]
\end{aligned}$$

Con esto podemos escribir la probabilidad que buscamos en términos de la función de distribución como:

$$\mathbb{P}(X \in [a, b]) = F_X(b) - F_X(a) + \mathbb{P}(X = a) \tag{1}$$

Si conociéramos $\mathbb{P}(X \in [a, b])$ para todo intervalo $[a, b]$, entonces nos quedaríamos tranquilos y diríamos que conocemos el comportamiento de $X$. La pregunta natural sería: ¿es posible reconstruir la v.a. teniendo solo $F_X$?. La respuesta es afirmativa; la función $F_X$ tiene toda la información de la v.a. (y más). Para ver esto nos fijamos en (1), si tuviéramos $\mathbb{P}(X = a)$, entonces conoceríamos el comportamiento de la v.a.

Para obtener $\mathbb{P}(X = a)$, tomemos un número $t$ grande ($t >> 1$), es claro que $\mathbb{P}(X = a) \approx \mathbb{P}\left(X \in \left(a - \frac{1}{t}, a\right]\right)$, más aún, la aproximación será la igualdad en el límite $t \to \infty$. Formalmente se tiene:

$$
\begin{aligned}
\mathbb{P}(X=a) &= \lim_{t\to\infty} \mathbb{P}\left(\left\{X \in \left(a-\frac{1}{t}, a\right]\right\}\right) \\
&= \lim_{t\to\infty} \mathbb{P}\left(\{X \le a\} \setminus \left\{X \le a-\frac{1}{t}\right\}\right) \\
&= \lim_{t\to\infty} \mathbb{P}(\{X \le a\}) - \mathbb{P}\left(\left\{X \le a-\frac{1}{t}\right\}\right) \\
&= \lim_{t\to\infty} F_X(a) - F_X\left(a-\frac{1}{t}\right) \\
&= F_X(a) - \lim_{t\to\infty} F_X\left(a-\frac{1}{t}\right) \\
&= F_X(a) - \lim_{y\to a^-} F_X(y) \\
\Rightarrow \mathbb{P}(X=a) &= F_X(a) - F_X(a^-) \qquad (2)
\end{aligned}
$$

donde en (2) hemos definido $F_X(a^-) := \lim_{y\to a^-} F_X(y)$. Queda finalmente juntando este resultado con (1)

$$\mathbb{P}(X \in [a,b]) = F_X(b) - F_X(a^-)$$

Se concluye que si $X$ es una v.a. y conocemos su función de distribución $F_X$, entonces para cualquier intervalo $[a,b]$ sabemos calcular $\mathbb{P}(X \in [a,b])$.

Finalmente, interpretamos $F_X(a^-)$ como una probabilidad:

$$F_X(a^-) = \mathbb{P}(X < a)$$

## 1.1. Variables Aleatorias Discretas

**Definición 1.3.** *Decimos que una v.a. $X : \Omega \to \mathbb{R}$ es discreta si existe $A \subseteq \mathbb{R}$ a lo más numerable tal que*

$$\mathbb{P}(X \in A) = 1$$

Es decir, si $X$ es discreta puede tomar a lo más una cantidad numerable de valores. Al conjunto $A$ donde la v.a. $X$ toma valores usualmente lo denotamos como $S(X)$ y lo llamamos el "soporte de $X$".

En éste contexto hablaremos de ley de la v.a. o bien de densidad discreta a la función $p_X : A \to \mathbb{R}_{\ge 0}$ tal que $p_X(k) := \mathbb{P}(X = k)$. Podemos pensar que le estamos asignando a cada elemento de $A$ una probabilidad de ocurrencia, las restricciones son por supuesto:

(i)
$$p_X(k) \ge 0 \quad \forall k \in A$$

(ii)
$$\sum_{k\in A} p_X(k) = 1$$

De la definición, es directo que en el caso discreto se tiene la siguiente igualdad:

$$F_X(x) = \sum_{k\le x} p_X(k) = \sum_{\substack{k\in A \\ k\le x}} p_X(k)$$

Lo anterior se puede ver claro en el ejemplo siguiente:

**Ejemplo 1.1.** *Sea $\Omega = \{w_1, w_2, w_3\}$ el conjunto de los sucesos posibles, sobre este conjunto definimos la medida de probabilidad $\mathbb{P}(\{w_i\}) = q_i$, donde las constantes $q_i$ son todas positivas y obviamente $q_1 + q_2 + q_3 = 1$.*

*Consideremos la v.a. $X$ definida mediante:*

$$X(w) := \begin{cases} -1 & si\ w = w_1 \\ 0 & otro\ caso \end{cases}$$

*Calcularemos la ley $p_X(k)$ y la distribución $F_X(x)$ de esta v.a.*

*Observamos que el conjunto $A$ donde la v.a. toma valores es $A = \{-1, 0\}$, luego para la ley solo debemos buscar $p_X(-1)$ y $p_X(0)$ ya que el resto será cero.*

*Tenemos que $p_X(-1) = \mathbb{P}(X = -1) = \mathbb{P}(\{w_1\}) = q_1$.*

*Asimismo $p_X(0) = \mathbb{P}(X = 0) = \mathbb{P}(\{w_1\}^c) = \mathbb{P}(\{w_2, w_3\}) = q_2 + q_3$. Deducimos que la ley de $X$ viene dada por*

$$p_X(k) = \begin{cases} q_1 & si\ k = -1 \\ q_2 + q_3 & si\ k = 0 \\ 0 & otro\ caso \end{cases}$$

*Ahora calcularemos la función de distribución: para cada $x \in \mathbb{R}$ nos preguntamos por la probabilidad de que nuestra v.a. tome un valor menor o igual a $x$. Obviamente si $x < -1$ la probabilidad es 0 (ya que nuestra v.a. nunca toma un valor menor a -1). En cambio si $x \geq 0$, $\mathbb{P}(X \leq x) = 1$ (ya que $X$ es menor o igual a cero siempre).*

*Finalmente tomemos $x \in [-1, 0)$, tendremos $\mathbb{P}(X \leq x) = \mathbb{P}(X = -1)$, ya que el único valor menor o igual a $x$ sería -1. Se deduce:*

$$F_X(x) = \begin{cases} 0 & si\ x < -1 \\ q_1 & si\ x \in [-1, 0) \\ 1 & si\ x \geq 0 \end{cases}$$

*Propuesto: verificar que se cumple $\mathbb{P}(X = 0) = F_X(0) - F_X(0^-)$*

## 1.2. Variables Aleatorias Continuas

**Definición 1.4.** *Decimos que $X : \Omega \to \mathbb{R}$ es una variable aleatoria continua si su función de distribución $F_X(x) : \mathbb{R} \to [0, 1]$ es una función continua.*

Hay una propiedad que es muy deseable al trabajar en este contexto, y es que la función de distribución se pueda escribir como una integral. Sabemos que $F_X$ es continua, ¿significa eso que podemos escribir $F_X$ como una integral definida? la respuesta es negativa. Para todas las aplicaciones que nos interesan en este curso, estudiaremos las variables aleatorias absolutamente continuas, y como no usaremos otras, las llamaremos sin ambiguedad v.a. continuas.

**Definición 1.5.** *Decimos que $X : \Omega \to \mathbb{R}$ es una variable aleatoria absolutamente continua si existe una función $f_X : \mathbb{R} \to \mathbb{R}_{\geq 0}$, que llamamos densidad de la v.a., tal que se puede escribir*

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

Como ya vimos, nos interesa mucho calcular cosas del estilo $\mathbb{P}(X \in A)$, para esto el resultado siguiente es muy útil, lo aceptaremos sin demostración.

**Teorema 1.1.** *Sea $X : \Omega \to \mathbb{R}$ una v.a. absolutamente continua, con densidad $f_X$, entonces para cualquier conjunto $A \subseteq \mathbb{R}$ se tiene que*

$$\mathbb{P}(X \in A) = \int_A f_X(x)dx$$

Lo anterior nos dice que una v.a. continua tiene probabilidad cero de tomar el valor $\bar{x}$ para cualquier $\bar{x} \in \mathbb{R}$, en efecto:

$$\mathbb{P}(X = \bar{x}) = \mathbb{P}(X \in \{\bar{x}\}) = \int_{\bar{x}}^{\bar{x}} f_X(x)dx = 0$$

## 1.3. Propiedades

A continuación definiremos tres propiedades sobre una v.a., la más importante es la esperanza, este concepto es sin lugar a dudas uno de los más relevantes en el ámbito de las probabilidades.

**Definición 1.6.** *Sea $X$ una v.a., definimos su esperanza (o valor esperado) mediante*

$$\sum_k k \cdot \mathbb{P}(X = k)$$

*si $X$ es discreta, o bien*

$$\int_{\mathbb{R}} x f_X(x)dx$$

*si $X$ es continua.*

Ya vimos en clases su intuición como la de realizar varias veces un experimento y preguntarme por el resultado promedio del experimento, es decir, si tengo una moneda que sale cara con probabilidad 0.6, entonces si la tiro muchas veces yo **esperaría** tener al rededor de un 60 % de caras.

La esperanza no tiene porque ser finita, en caso de que la integral (o la suma) diverja, diremos que $X$ no tiene definida su esperanza o bien que su esperanza es infinita (no tengan miedo a decir que algo vale infinito).

Sea $g : \mathbb{R} \to \mathbb{R}$ una función y $X$ una variable aleatoria, es común no querer el valor esperado de $X$, sino que el valor esperado de $g(X)$, para esto tendríamos que considerar una nueva v.a. $Y = g(X)$, a esta v.a. calcularle su ley y finalmente calcular $\mathbb{E}(Y)$ mediante la definición. Este es un proceso engorroso y no se recomienda. Cuando queramos $g(X)$ usaremos el siguiente teorema (que aceptaremos sin demostración)

**Teorema 1.2.** *Sea $X$ una v.a. y $g : \mathbb{R} \to \mathbb{R}$ una función, entonces el valor esperado de $g(X)$ es*

$$\mathbb{E}(g(X)) = \sum_k g(k)P(X = k)$$

*si $X$ es discreta o*

$$\mathbb{E}(g(X)) = \int_{\mathbb{R}} g(x) f_X(x)dx$$

*si $X$ es continua.*

**Definición 1.7.** *Sea $X$ una v.a., si existe $\mathbb{E}(X)$ y $\mathbb{E}(X^2)$, es decir ambos son finitos, entonces se define la varianza de $X$ mediante*

$$\mathbb{V}ar(X) = \mathbb{E}([X - \mathbb{E}(X)]^2)$$

**Ejemplo 1.2.** *Sea $X$ una v.a. discreta con ley dada por*

$$p_X(k) = \begin{cases} p_0 & si\ k = 0 \\ p_1 & si\ k = 1 \\ p_2 & si\ k = 2 \\ 0 & otro\ caso \end{cases}$$

*Calcularemos $\mathbb{V}ar(X)$ por definición*

$$\begin{aligned} \mathbb{E}(X) &\triangleq \sum_k k \cdot p_k \\ &= 0 \cdot p_0 + 1 \cdot p_1 + 2 \cdot p_2 \end{aligned}$$

*De ahora en adelante pensamos en $\mathbb{E}(X)$ como un número conocido (y que vale $\mu := p_1 + 2p_2$). Consideremos la función $g : \mathbb{R} \to \mathbb{R}$ dada por $g(x) = (x-\mu)^2$, usando el Teorema anterior tenemos que*

$$\begin{aligned} \mathbb{V}ar(X) &\triangleq \mathbb{E}([X - \mathbb{E}(X)]^2) \\ &= \mathbb{E}(g(X)) \\ &= \sum_k g(k) \cdot p_k \\ &= g(0) \cdot p_0 + g(1) \cdot p_1 + g(2) \cdot p_2 \\ &= \mu^2 p_0 + (1-\mu)^2 p_1 + (2-\mu)^2 p_2 \end{aligned}$$

**Definición 1.8.** *Sea $X$ una v.a. aleatoria, sea $B$ el conjunto donde $\mathbb{E}(e^{tX})$ existe, es decir, $B = \{t \in \mathbb{R} | \mathbb{E}(e^{tX}) < \infty\}$, definimos la función generadora de momentos $\varphi_X : B \to \mathbb{R}_+$ mediante*

$$\varphi_X(t) = \mathbb{E}(e^{tX})$$

# 2. Modelos

Ahora nos concentraremos en aprender ejemplos notables de variables aleatorias y enumerar sus propiedades.

## 2.1. Bernoulli

Esta v.a. es nombrada así en honor al matemático Jacob Bernoulli, quien en el siglo XVIII la utilizó para enunciar por primera vez uno de los teoremas más importantes de las matemáticas: la ley de los grandes números.

La v.a. de Bernoulli corresponde al experimento de lanzar una vez una moneda que sale en cara ($X = 1$) con probabilidad $p$ y sello ($X = 0$) con probabilidad $1 - p$.

**Definición 2.1.** *Decimos que $X$ es v.a. Bernoulli de parámetro $0 < p < 1$, que denontamos $X \sim Bernoulli(p)$, si su ley se puede escribir*

$$p_X(k) = p^k (1-p)^{1-k} \qquad k \in \{0, 1\}$$

**Proposición 2.1.** *Sea $X \sim Bernoulli(p)$*

*(i)* $\mathbb{E}(X) = p$

*(ii)* $\mathbb{V}ar(X) = p(1-p)$

*(iii)* $\varphi_X(t) = (pe^t + 1 - p)$

Demostración

(iii) Por definición $\varphi_X(t) = \mathbb{E}(e^{tX})$, como $X$ tiene solo dos valores posibles, lo haremos de forma directa:

$$\begin{aligned} \varphi_X(t) &= \mathbb{E}(e^{tX}) \\ &= e^{t\cdot 1} \cdot \mathbb{P}(X=1) + e^{t\cdot 0} \cdot \mathbb{P}(X=0) \\ &= e^t \cdot p + 1 \cdot (1-p) \end{aligned}$$

□

## 2.2. Binomial

Consideramos el experimento de lanzar $n$ veces una moneda que sale cara con probabilidad $p$ y sello con probabilidad $1-p$. Nos interesa conocer la probabilidad de que de los $n$ lanzamientos resulten $k$ caras, esto corresponde a $\mathbb{P}(X=k)$. Asumimos que las $n$ tiradas son independientes.

**Definición 2.2.** *Decimos que la v.a. $X$ tiene ley binomial de parámetros $0 < p < 1$, $n \in \mathbb{N}$, que denotamos $X \sim Bin(n,p)$, si su ley se puede escribir*

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

**Proposición 2.2.** *Sea $X \sim Bin(n,p)$*

*(i)* $\mathbb{E}(X) = np$

*(ii)* $\mathbb{V}ar(X) = np(1-p)$

*(iii)* $\varphi_X(t) = (pe^t + 1 - p)^n$

*(iv) Si $X_1, \cdots, X_n$ son v.a. Bernoulli independientes de parámetro $p$, entonces*

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

*(v) Sean $X_1, \cdots, X_n$ v.a. binomiales independientes, tales que $X_i \sim Bin(n_i, p)$, entonces*

$$\sum_{i=1}^{n} X_i \sim Bin\left(\sum_{i=1}^{n} n_i, p\right)$$

Demostración

(iv) Definamos $Y = \sum_{i=1}^{n} X_i$, es claro que $Y$ no puede tomar un valor superior a $n$, entonces $\mathbb{P}(Y > n) = 0$, tomemos entonces $k \leq n$. Para que $Y$ tome el valor $k$ es necesario que hayan exactamente $k$ variables que tomen el valor 1 y $n-k$ que tomen el valor 0.

El número de casos en que hay $k$ variables con valor 1 es $\binom{n}{k}$ y por independencia cada uno tiene probabilidad $p^k(1-p)^{n-k}$, se concluye entonces que $\mathbb{P}(Y = k) = \binom{n}{k} p^k (1-p)^{n-k}$

(v) Es directo notando que cada $X_i$ se puede escribir como suma de $n_i$ v.a. independientes de bernoullí, formalmente se tiene $X_i = \sum_{j=1}^{n_i} Y_j^i$, donde los $Y_1^i, Y_2^i, \cdots, Y_{n_i}^i$ son bernoulli independientes de parámetro $p$. Así la v.a. $Z = \sum_{i=1}^{n} X_i$ es suma de $n_1 + n_2 + \cdots n_n$ v.a. bernoullí independientes.

(i)

$$\begin{aligned}
\mathbb{E}(X) &\triangleq \sum_{k=0}^{n} k\mathbb{P}(X=k) \\
&= \sum_{k=0}^{n} k\binom{n}{k} p^k (1-p)^{n-k} \\
&= \sum_{k=0}^{n} k\frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \\
&= \sum_{k=1}^{n} k\frac{n!}{k(k-1)!(n-k)!} p^k (1-p)^{n-k} \\
&= \sum_{k=1}^{n} n\frac{(n-1)!}{(k-1)!(n-k)!} p^k (1-p)^{n-k} \\
&= \sum_{k=1}^{n} n\binom{n-1}{k-1} p^k (1-p)^{n-k} \\
&= n\sum_{k=0}^{n-1} \binom{n-1}{k} p^{k+1} (1-p)^{n-(k+1)} \\
&= np\sum_{k=0}^{n-1} \binom{n-1}{k} p^k (1-p)^{n-(k+1)}
\end{aligned}$$

□

## 2.3. Geométrica

Ahora consideramos el experimento de lanzar una moneda indefinidamente, el experimento termina cuando se observa una cara por primera vez. Nos interesa saber en que momento saldrá la primera cara, es decir, si $X = k$ en el lanzamiento $k$ se observó la primera cara.

Es importante tener claro que mediremos, ya que si lo que nos interesa es determinar el número de sellos observados hasta concluir el experimento, entonces hay que reformular la v.a. (propuesto).

**Definición 2.3.** *Decimos que $X$ es v.a. Geométrica de parámetro $0 < p < 1$, que denotamos $X \sim Geo(p)$, si su ley puede escribirse como*

$$p_X(k) = (1-p)^{k-1} p \qquad k \in \mathbb{N}\backslash\{0\}$$

Una propiedad rara, pero muy deseada para el modelamiento, es la llamada "pérdida de memoria". Pensemos que $X$ representa cuanto tiempo pasa hasta que falle una máquina, uno pensaría que mientras más tiempo pase, es cada vez más probable que la máquina falle. Si $X$ tiene pérdida de memoria, entonces no importa cuanto tiempo pase, la probabilidad de que la máquina falle de aquí a $t$ no cambia. Formalmente se tiene:

**Definición 2.4.** *Sea $X : \Omega \to \mathbb{R}$ una v.a., decimos que $X$ tiene la propiedad de pérdida de memoria si $X$ cumple $\forall t, s > 0$ :*

$$\mathbb{P}(X > t + s | X > s) = \mathbb{P}(X > t)$$

Es fácil ver que se tiene la siguiente caracterización de la v.a. geométrica

**Proposición 2.3.** *Sea $X : \Omega \to \mathbb{R}$ una v.a. se tiene que*

$$X \sim Geo(p) \iff \forall m \in \mathbb{N}, \quad \mathbb{P}(X > m) = (1-p)^m$$

**Proposición 2.4.** *Sean $X \sim Geo(p)$, $Y \sim Geo(q)$ independientes.*

*(i)* $\mathbb{E}(X) = \frac{1}{p}$

*(ii)* $\mathbb{V}ar(X) = \frac{1-p}{p^2}$

*(iii)* $\varphi_X(t) = \frac{pe^t}{1-e^t(1-p)}$ *para* $t < \ln(\frac{1}{1-p})$

*(iv)* $mín\{X, Y\} \sim Geo(p + q - pq)$

*(v) $X$ tiene pérdida de memoria.*

Demostración

(iii) Por definición

$$\begin{aligned} \varphi_X(t) &\triangleq \mathbb{E}(e^{tX}) \\ &= \sum_{k=1}^{\infty} e^{tk} p(1-p)^{k-1} \\ &= \sum_{k=0}^{\infty} e^{t(k+1)} p(1-p)^k \\ &= pe^t \sum_{k=0}^{\infty} e^{tk}(1-p)^k \end{aligned}$$

Para que la serie converja se necesita imponer la condición $e^t(1-p) < 1$, con esto se tiene finalmente

$$\varphi_X(t) = \frac{pe^t}{1 - e^t(1-p)}$$

□

## 2.4. Poisson

La v.a. de Poisson debe su nombre al matemático francés Siméon Denis Poisson y fue formulada en el siglo XVIII, esta v.a. es de gran relevancia en los procesos estocásticos debido a que modela llegadas de entidades a un sistema.

Podemos pensar que tenemos un sistema al que llegan clientes, la cantidad de clientes que llega $(X)$ puede ser cualquiera, nos preguntamos entonces por la probabilidad de que lleguen $k$ clientes, lo que corresponde a $\mathbb{P}(X = k)$.

**Definición 2.5.** *Decimos que $X$ es v.a. de Poisson de parámetro $\lambda > 0$, que denotamos $X \sim Poiss(\lambda)$, si su ley se puede escribir*

$$p_X(k) = e^{-\lambda}\frac{\lambda^k}{k!} \qquad k \in \mathbb{N} \cup \{0\}$$

**Proposición 2.5.** *Sea $X \sim Poiss(\lambda)$.*

*(i)* $\mathbb{E}(X) = \lambda$

*(ii)* $\mathbb{V}ar(X) = \lambda$

*(iii)* $\varphi_X(t) = e^{\lambda(e^t-1)}$

*(iv) Si $X_1, \cdots, X_n$ son v.a. de poisson independientes, tales que $X_i \sim Poiss(\lambda_i)$, entonces:*

$$\sum_{i=1}^{n} X_i \sim Poiss\left(\sum_{i=1}^{n} \lambda_i\right)$$

Demostración

(iv) Lo haremos por inducción en $n$, donde el caso $n = 2$ es realmente el paso inductivo, por lo que quedará propuesto concluir $n \Rightarrow n+1$. Sean entonces $X_1 \sim Poiss(\lambda_1), X_2 \sim Poiss(\lambda_2)$ y definamos $Y = X_1 + X_2$

$$\begin{aligned}
\mathbb{P}(Y = k) &= \mathbb{P}(X_1 + X_2 = k) \\
&= \sum_{j=0}^{k} \mathbb{P}(X_1 = j, X_2 = k - j) \\
&= \sum_{j=0}^{k} \mathbb{P}(X_1 = j) \cdot \mathbb{P}(X_2 = k - j) \\
&= \sum_{j=0}^{k} e^{-\lambda_1}\frac{\lambda_1^j}{j!} \cdot e^{-\lambda_2}\frac{\lambda_2^{k-j}}{(k-j)!} \\
&= e^{-\lambda_1}e^{-\lambda_2}\sum_{j=0}^{k} \frac{\lambda_1^j}{j!} \cdot \frac{\lambda_2^{k-j}}{(k-j)!} \\
&= \frac{e^{-\lambda_1}e^{-\lambda_2}}{k!}\sum_{j=0}^{k} \frac{k!}{j!(k-j)!}\lambda_2^{k-j}\lambda_1^j \\
\Rightarrow \mathbb{P}(Y = k) &= \frac{e^{-(\lambda_1+\lambda_2)}}{k!}(\lambda_1 + \lambda_2)^k
\end{aligned}$$

□

**Nota histórica:** Es interesante que la primera aplicación de ésta v.a. fue propuesta en 1898 por Ladislaus Bortkiewicz, se usó para modelar el número de soldados del ejército de Prusia que morían accidentalmente por una patada de caballo.

## 2.5. Uniforme

La v.a. uniforme corresponde al experimento de elegir un punto al azar en el trazo $[a, b]$, nos interesa contestar la pregunta ¿cuál es la probabilidad de que $X$ esté en cierto intervalo?.

Una sencilla aproximación a modelar esto es bajo los supuestos: cualquier intervalo que no esté contenido en $[a, b]$ tiene probabilidad cero de contener a $X$ y cualquier intervalo contenido en $[a, b]$ tendrá una probabilidad de contener a $X$ proporcional a su largo.

**Definición 2.6.** *Decimos que $X$ es una v.a. uniforme en $[a, b]$, $a < b$, lo que denotamos $X \sim U[a, b]$, si su densidad tiene la forma*

$$f_X(x) = \frac{1}{b-a} \mathbb{1}_{\{a \leq x \leq b\}}$$

**Proposición 2.6.** *Sea $X \sim U[a, b]$*

*(i)* $\mathbb{E}(X) = \frac{a+b}{2}$

*(ii)* $\mathbb{V}ar(X) = \frac{(b-a)^2}{12}$

*(iii)* $\varphi_X(t) = \frac{e^{bt} - e^{at}}{(b-a)t}$

## 2.6. Exponencial

La variable aleatoria más relevante (después de la normal) es la exponencial, esto pues es la única v.a. continua que satisface la propiedad de pérdida de memoria. Esta v.a. se utiliza para modelar tiempos, en particular tiempos entre llegada, por ejemplo el tiempo que pasa desde que se recibe una llamada hasta que se recibe la siguiente llamada en un call center, o bien el tiempo entre llegada de clientes a un sistema.

**Definición 2.7.** *Decimos que $X$ es una v.a. exponencial de parámetro $\lambda > 0$ , lo que denotamos $X \sim exp(\lambda)$, si su densidad tiene la forma*

$$f_X(x) = \lambda e^{-\lambda x} \mathbb{1}_{\{x \geq 0\}}$$

Hay que tener cuidado con la notación, ya que algunos autores definen la exponencial con la densidad de la forma $f(x) = \frac{1}{\lambda} e^{-\frac{x}{\lambda}}$, luego las propiedades no serán las mismas que enunciaremos acá.

**Proposición 2.7.** *Sea $X \sim exp(\lambda)$*

*(i)* $\mathbb{E}(X) = \frac{1}{\lambda}$

*(ii)* $\mathbb{V}ar(X) = \frac{1}{\lambda^2}$

*(iii)* $\varphi_X(t) = (1 - \frac{t}{\lambda})^{-1}$

*(iv) $X$ tiene la propiedad de pérdida de meoria.*

*(v) Si $X_1, \cdots, X_n$ son v.a. de exponenciales independientes, tales que $X_i \sim exp(\lambda_i)$, entonces:*

$$\min_{i=1,\cdots,n} X_i \sim exp\left(\sum_{i=1}^{n} \lambda_i\right)$$

Demostración

(iv) Lo que hay que demostrar es que dados $s, t > 0$, $\mathbb{P}(X > t+s|X > s) = \mathbb{P}(X > t)$, para eso calculamos

$$\mathbb{P}(X > r) = \int_r^\infty f_X(x)dx = \int_r^\infty \lambda e^{-\lambda x}dx = e^{-\lambda r}$$

Luego por definición de condicionalidad tenemos

$$\begin{aligned}
\mathbb{P}(X > t+s|X > s) &\triangleq \frac{\mathbb{P}(X > t+s, X > s)}{\mathbb{P}(X > s)} \\
&= \frac{\mathbb{P}(X > t+s)}{\mathbb{P}(X > s)} \\
&= \frac{e^{-\lambda(t+s)}}{e^{-\lambda s}} \\
&= e^{-\lambda t} \\
&= \mathbb{P}(X > t)
\end{aligned}$$

(v) Esta demostración se puede hacer de forma directa (sin necesidad de inducción) cuando uno se da cuenta que el evento "el mínimo es mayor que $y$" corresponde al evento "todos son mayores que $y$". Sea $Y = \min_{i=1,\cdots,n} X_i$

$$\begin{aligned}
\mathbb{P}(Y > y) &= \mathbb{P}(X_1 > y, X_2 > y, \cdots, X_n > y) \\
&= \prod_{i=1}^{n} \mathbb{P}(X_i > y) \\
&= \prod_{i=1}^{n} \exp(-\lambda_i y) \\
&= \exp\left(-y \sum_{i=1}^{n} \lambda_i\right)
\end{aligned}$$

□

## 2.7. Normal

La v.a. normal es sin duda la más relevante en el estudio de las probabilidades, esto se debe al Teorema del Límite Central que nos permite aproximar probabilidades mediante una normal.

**Definición 2.8.** *Decimos que $X$ es una v.a. normal de parámetros $\mu \in \mathbb{R}$ y $\sigma^2$ ($\sigma > 0$) , lo que denotamos $X \sim \mathcal{N}(\mu, \sigma^2)$, si su densidad tiene la forma*

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Cuando los parámetros de una normal son $\mu = 0$ y $\sigma = 1$ decimos que se trata de una normal estándar.

Es fácil ver que la densidad de una normal no tiene primitiva explícita, así que se utiliza de forma indirecta la función de probabilidad acumulada

**Definición 2.9.** *Para $x \in \mathbb{R}$ definimos la función de probabilidad acumulada de una normal estándar como $\Phi$, es decir:*

$$\Phi(x) := \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$$

**Proposición 2.8.** *Sea* $X \sim \mathcal{N}(\mu, \sigma^2)$

*(i)* $\mathbb{E}(X) = \mu$

*(ii)* $\mathbb{V}ar(X) = \sigma^2$

*(iii)* $\varphi_X(t) = \exp(\mu t + \frac{\sigma^2 t^2}{2})$

*(v) Sean* $a, b \in \mathbb{R}$ *con* $b \neq 0$*, se tiene que* $a + Xb \sim \mathcal{N}(a + b\mu, b^2\sigma^2)$

*(vi) Si* $X_1, \cdots, X_n$ *son v.a. normales independientes, tales que* $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$*, entonces*

$$\sum_{i=1}^{n} X_i \sim \mathcal{N}\left(\sum_{i=1}^{n} \mu_i, \sum_{i=1}^{n} \sigma_i^2\right)$$

## 2.8. Gamma

La v.a. Gamma es una que generaliza a muchas otras variables aleatorias, i.e. hay otras v.a. importantes que son casos particulares de la Gamma.

Esta v.a. es un poco complicada de trabajar porque es difícil de integrar, pero es necesaria muchas veces.

Comenzaremos definiendo la función $\Gamma$, que es en parte una extensión de la función factorial para $x \in \mathbb{R}$.

**Definición 2.10.** *Se define la función Gamma mediante*

$$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt$$

*se puede verificar que si* $n \in \mathbb{N}$*, entonces* $\Gamma(n) = (n-1)!$

**Definición 2.11.** *Decimos que* $X$ *es una v.a. Gamma de parámetros* $s, \lambda > 0$ *, lo que denotamos* $X \sim Gamma(s, \lambda)$*, si su densidad tiene la forma*

$$f_X(x) = \frac{x^{s-1} e^{-\lambda x} \lambda^s}{\Gamma(s)} \mathbb{1}_{\{x \geq 0\}}$$

**Proposición 2.9.** *Sea* $X \sim \mathcal{N}(\mu, \sigma^2)$

*(i)* $\mathbb{E}(X) = \frac{s}{\lambda}$

*(ii)* $\mathbb{V}ar(X) = \frac{s}{\lambda^2}$

*(iii)* $\varphi_X(t) = (1 - \frac{t}{\lambda})^{-s}$

*(iv) Sean* $X_1, \cdots, X_n$ *i.i.d. exponenciales de parámetro* $\lambda$*, entonces*

$$\sum_{i=1}^{n} X_i \sim Gamma(n, \lambda)$$

Demostración

(i)

$$\begin{aligned}
\mathbb{E}(X) &\triangleq \int_{\mathbb{R}} x f_X(x) dx \\
&= \int_0^\infty \frac{x^s e^{-\lambda x} \lambda^s}{\Gamma(s)} dx \\
&= -\frac{x^s}{\lambda \Gamma(s)} e^{-\lambda x} \lambda^s \Big|_0^\infty - \int_0^\infty -\frac{e^{-\lambda x}}{\lambda} \frac{s x^{s-1} \lambda^s}{\Gamma(s)} dx \qquad (3) \\
&= \frac{s}{\lambda} \int_0^\infty \underbrace{\frac{x^{s-1} e^{-\lambda x} \lambda^s}{\Gamma(s)}}_{f_X(x)} dx \\
&= \frac{s}{\lambda}
\end{aligned}$$

Donde en (3) se integró por partes, usando $u = x^s \Rightarrow du = s x^{s-1}$ y $dv = \frac{e^{-\lambda x} \lambda^s}{\Gamma(s)} \Rightarrow v = -\frac{e^{-\lambda x} \lambda^s}{\lambda \Gamma(s)}$

□