



FACULTAD DE MATEMÁTICAS
PONTIFICIA UNIVERSIDAD
CATÓLICA DE CHILE

EYP1113 - Probabilidad y Estadística

Taller R - 04: Modelos de probabilidad

Ricardo Aravena C. - Cristian Capetillo C. - Ingrid Guevara R.
Ricardo Olea O. - Bladimir Morales T., Daniel Saavedra M.

Facultad de Matemáticas
Departamento de Estadística
Pontificia Universidad Católica de Chile

Segundo Semestre 2024

Contenido I

Motivación

Modelos de Probabilidad

Aplicaciones

Motivación

Cuando nos enfrentamos a un conjunto de datos y calculamos estadísticas tales como la media o varianza, y presentamos gráficos como lo sería un histograma, un boxplot, o un gráfico de dispersión, estamos haciendo lo que se denomina **análisis descriptivo** de los datos.

Más allá de conocer nuestra muestra estaremos interesados más bien en la **población** de la cual proviene. Para esto, dado que no conocemos tal población debemos **suponer** alguna **estructura** en ella, y luego utilizar la información que nos brinda nuestra muestra para acabar de comprenderla.

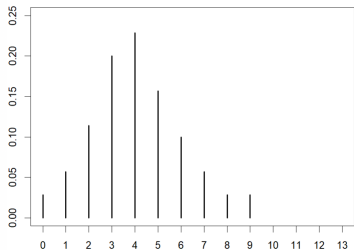
Informalmente, nuestros datos son el resultado de un proceso aleatorio, dígame una maquinita que arroja números, y nosotros queremos reducir nuestra incertidumbre sobre dicha maquinita.



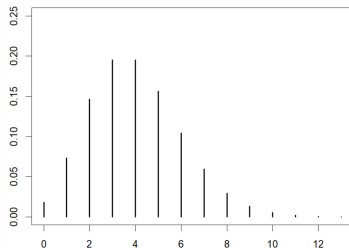
Motivación: Distribución Poisson

La distribución Poisson es una distribución de probabilidad con soporte en los enteros no negativos, esto es, $\{0, 1, 2, \dots\}$. Su función de cuantía depende de un único parámetro $\lambda > 0$ y está dada por

$$f(x) = \frac{e^{-\lambda} \lambda^x}{x!}.$$



(a) Datos de conteo.

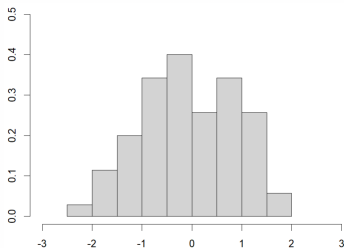


(b) Distribución $Pois(\lambda = 4)$.

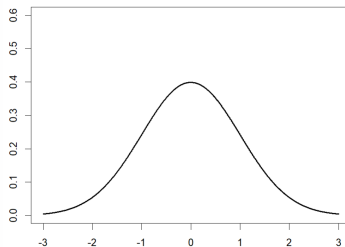
Motivación: Distribución Normal

La distribución Normal es sin duda la distribución más conocida y útil de toda la estadística. Con soporte en \mathbb{R} , su función de densidad depende de dos parámetros $(\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}^+$ y está dada por

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{1}{2\sigma^2}(x - \mu)^2 \right\}.$$



(a) Datos continuos en \mathbb{R} .

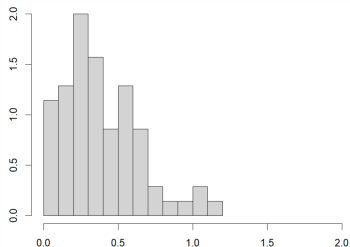


(b) Distribución $\mathcal{N}(\mu = 0, \sigma^2 = 1)$.

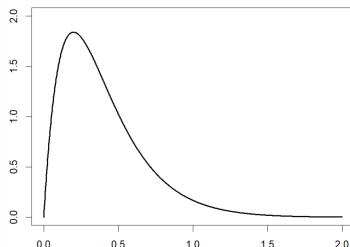
Motivación: Distribución Gamma

La distribución Gamma es una distribución con soporte en \mathbb{R}_+ . Su función de densidad depende de dos parámetros (α, β) , ambos mayores a 0 y está dada por

$$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}.$$



(a) Datos continuos en \mathbb{R}_+ .



(b) Distribución $\mathcal{G}(\alpha = 2, \beta = 5)$.

Modelos de Probabilidad

En R, muchos de los modelos usuales ya vienen incorporados para su uso, por ejemplo:

Modelo	Comando
Binomial	<code>_binom()</code>
Poisson	<code>_pois()</code>
Uniforme	<code>_unif()</code>
Normal	<code>_norm()</code>
Exponencial	<code>_exp()</code>
Gamma	<code>_gamma()</code>
Chi Cuadrado	<code>_chisq()</code>
t-Student	<code>_t()</code>
Fisher	<code>_f()</code>

Modelos de Probabilidad

Cada comando se puede utilizar de cuatro formas diferentes:

- ▶ **ddist**(x, \dots). Entrega la función de cuantía o densidad evaluada en x .
- ▶ **pdist**(q, \dots). Entrega la función de distribución acumulada evaluada en q , i.e., $P(X \leq q)$.
- ▶ **qdist**(p, \dots). Entrega el valor de x tal que $P(X \leq x) = p$, es decir, el cuantil p .
- ▶ **rdist**(n, \dots). Genera una muestra aleatoria proveniente de la distribución.

Ejemplo: Alturas de estudiantes

Existe una base de datos en *R* que contiene medidas de alturas y pesos. Ya que estas variables suelen seguir una distribución normal en poblaciones humanas, podemos suponer tal distribución como la que genera nuestras observaciones. Los datos están disponibles en la librería *gcookbook* y se llaman `heightweight`.



Ejemplo: Número de llamadas por hora

Se tiene la información del número de llamadas recibidas por un centro de emergencias en un hospital, en intervalos de una hora. Por ejemplo, de 09:00 a.m. a 09:59 a.m. se tiene 1 llamada recibida, de 10:00 a 10:59 a.m. se tienen 5 llamadas recibidas, de 11:00 a 11:59 a.m. se tienen 5 llamadas, etc. Esta información se puede ver en el archivo `bd_p.xlsx`.

¿Cómo podríamos modelar esta información?



Ejemplo: Uso de funciones

Si $X \sim \text{Exp}(\nu)$, $x > 0$, entonces:

- ▶ Para obtener la densidad $f_X(x)$ en el punto x se usa el comando `dexp(x,rate= ν)`.
- ▶ Para calcular la probabilidad acumulada $P(X \leq q)$ usamos el comando `pexp(q,rate= ν)`.
- ▶ Para calcular cuantiles de la distribución se usa el comando `qexp(p,rate= ν)`.
- ▶ Para generar n variables aleatorias provenientes de la distribución se utiliza el comando `rexp(n,rate= ν)`.

Distribuciones de Probabilidad

Distribución Exponencial

Suponga que $X \sim \text{Exp}(\nu = 3)$:

- ▶ Calcule $f_X(1)$.
- ▶ Calcule $P(X \leq 1.5) = F_X(1.5)$.
- ▶ Si $P(X \leq k) = 0.5$, obtenga el valor de k .
- ▶ Genere una muestra de tamaño $n = 1000$.