Variables Aleatorias

Introducción a la Estadística

Fiona Franco Churruarín
fionafch96@gmail.com

Universidad Torcuato Di Tella

Febrero 2022

Variable aleatoria

Variable que toma valores numéricos determinados por el resultado del espacio muestral de un experimento aleatorio.

Variable Aleatoria

Sea Ω un espacio muestral sobre el que se encuentra definida una función de probabilidad. Sea X una función real definida sobre Ω, de manera que transforme los resultados de Ω en puntos sobre la recta real $\mathbb{R}$. Se dice entonces que X es una variable aleatoria.

Ejemplo 1: "Experimento: Tiro un dado una vez y registro su valor."

Resultado	Valor de la v.a.	Número de ocurrencias	Probabilidad
⚀	1	1	$1/6$
⚁	2	1	$1/6$
⚂	3	1	$1/6$
⚃	4	1	$1/6$
⚄	5	1	$1/6$
⚅	6	1	$1/6$

Variable aleatoria

Ejemplo 2: "Experimento: Tiro un dado dos veces y sumo el valor de cada uno."

Resultado	Valor de la v.a.	Número de ocurrencias	Probabilidad
(1.1)	2	1	$1/36$
(1,2),(2,1)	3	2	$2/36$
(1,3),(2,2),(3,1)	4	3	$3/36$
(1,4),(2,3),(3,2),(4,1)	5	4	$4/36$
(1,5),(2,4),(3,3),(4,2),(5,1)	6	5	$5/36$
(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)	7	6	$6/36$
(2,6),(3,5),(4,4) .(5,3),(6,2)	8	5	$5/36$
(3,6),(4,5),(5,4),(6,3)	9	4	$4/36$
(4,6),(5,5),(6,4)	10	3	$3/36$
(5,6),(6,5)	11	2	$2/36$
(6,6)	12	1	$1/36$

Variable aleatoria

Es posible que conozcamos el **rango** de posibles valores que puede tomar. A este conjunto de valores se lo denomina **soporte** de una variable aleatoria y lo llamaremos S_X

Por ejemplo, si consideramos la variable aleatoria $Y =$'Valor de un dado de 6 caras en una tirada', los posibles valores son $\{1{,}2{,}3{,}4{,}5{,}6\}$. El número 7 no pertenece al soporte de X.

Notación: en general describimos la variable aleatoria con mayúscula (X) y el resultado (realización) con minúscula (x). Por ejemplo el tipo de cambio nominal mañana es una variable aleatoria X, y una posible realización x es 82.

Tipos de variables aleatorias

$$\text{Variables aleatorias} \begin{cases} \text{Cuantitativas (numéricas)} \begin{cases} \text{Discretas} \\ \text{Continuas} \end{cases} \\ \\ \text{Cualitativas (categóricas)} \begin{cases} \text{Ordinales} \\ \text{Nominales} \end{cases} \end{cases}$$

Tipos de variables aleatorias

Variables aleatorias **cualitativas**:

- **nominales:** presentan modalidades no numéricas que no admiten un criterio de orden (por ej. sexo, estado civil).
- **ordinales:** presentan modalidades no numéricas en las que existe un orden (por ej. nivel educativo)

Variables aleatorias **cuantitativas**:

- **discretas:** toman un número contable de valores.
  El soporte es un conjunto finito o infinito numerable
- **continuas:** toman un número incontable de valores
  El soporte es conjunto infinito no numerable

Ejemplos

¿Son las siguientes variables aleatorias **cualitativas** o **cuantitativas**?
¿En cada caso, qué tipo de variables son?

- El precio de los alquileres en Buenos Aires
- La cantidad de semanas hasta completar un proyecto
- El resultado de tirar un dado
- La altura de los niños entre 5 y 8 años
- La elección de un individuo de viajar o no en transporte público
- El grado de satisfacción con algún servicio

Volviendo a las variables aleatorias

Vamos a centrarnos en el estudio de variables aleatorias **cuantitativas**:

- Discretas
- Continuas

Variable aleatoria discreta

Función de probabilidad

Sea X una v.a. discreta que toma valores x_1, x_2,. . . , con sus respectivas probabilidades $P(X = x_1)$, $P(X = x_2)$,. . . (Esta lista de valores numéricos y sus probabilidades puede ser finita o infinita, pero **numerable**).

La **función de probabilidad** (pmf), $p(x)$, asociada a la variable aleatoria X se define como:

$$p(x) = \begin{cases} P(X = x) & \text{si } x \in \{x_1, x_2, ...\} \\ 0 & \text{otro caso} \end{cases}$$

Observe que se cumplen las siguientes propiedades:

(a) $p(x) \geq 0$

(b) $\sum_x p(x) = 1$

Variable aleatoria continua

Función de densidad

Sea X una v.a. continua. La función integrable y no negativa $f(x)$ es la **función de densidad** (pdf) de X si para cualquier intervalo (a, b) se cumple:

$$P(X \in (a, b)) = P(a < X < b) = \int_b^a f(x)dx.$$

Se cumplen las siguientes propiedades:

(a) $f(x) \geq 0$

(b) $\int_{-\infty}^{\infty} f(x)dx = 1$

(c) $P(X = x) = 0$

Es decir, la probabilidad de que la variable tome un valor dentro del intervalo (a, b) se puede calcular o expresar como el área bajo la función de densidad en el intervalo (a, b).

En general (v.a. discretas y continuas)

Función de distribución acumulada

Sea X una variable aleatoria discreta o continua, la **función de distribución acumulada** (cdf) de X es la probabilidad de que la variable aleatoria tome un valor menor o igual a x:

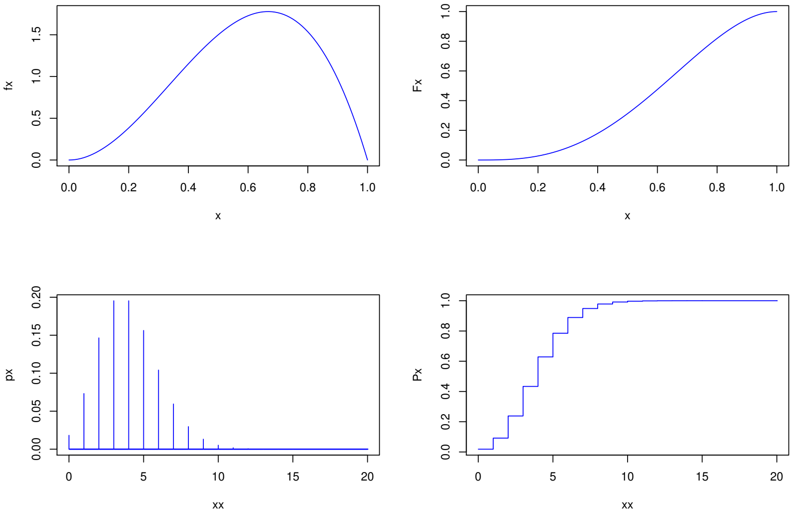
$$F(x) = P(X \leq x).$$

Para v.a. discreta: $F(x) = \sum_{X \leq x} p(x)$
Para v.a. continua: $F(x) = \int_{-\infty}^{x} f(x)dx$

Toda función de distribución acumulada, $F(x)$, satisface las siguientes propiedades:

(a) $0 \leq F(x) \leq 1$
(b) $\lim_{x \to +\infty} F(x) = 1$
(c) $\lim_{x \to -\infty} F(x) = 0$
(d) Si $x_1 \leq x_2$, entonces $F(x_1) \leq F(x_2)$
(e) Si $x_1 \leq x_2$, entonces $P(x_1 < X \leq x_2) = F(x_2) - F(x_1)$

Ejemplo: Funciones de probabilidad, de densidad y de distribución



Distribuciones

En resumen... podemos decir en general que una **distribución de probabilidad** es una función que describe la probabilidad con la que una variable aleatoria toma distintos valores (discreta, función de probabilidad de masa) o cae en distintos intervalos (contínua, función de densidad).

Algunas llevan nombres porque representan situaciones inciertas comunes. Más formalmente, son funciones que sirven para modelizar algunos fenómenos con características comunes. Vamos a ver tres tipos de distribuciones (las más usuales):

- Distribución Binomial (discreta)
- Distribución de Poisson (discreta)
- Distribución Normal (contínua)

A continuación, vemos algunas medidas que se usan para describir características de una variable aleatoria.

Valor esperado (Esperanza Matemática)

El **valor esperado** (o esperanza matemática) es una medida de lo que ocurre en promedio a largo plazo, si se repite el experimento muchísimas veces.

Esperanza de una v.a.

Si X es una variable aleatoria **discreta**, su valor esperado, $E(X)$, es:

$$\mu = E(X) = \sum_{x \in S_X} x P(X = x)$$

Si X es una variable aleatoria **continua**, su valor esperado, $E(X)$, es:

$$\mu = E(X) = \int_{-\infty}^{\infty} x f(x) dx = \int_{S_X} x f(x) dx$$

Es decir, el valor esperado es un **promedio ponderado** de todos los posibles valores que la variable X puede tomar, donde los ponderadores son las probabilidades asociadas a cada x.

Valor esperado (Esperanza Matemática)

Recordar la definición frecuentista de probabilidad, puede demostrarse que la esperanza es un promedio de largo plazo, cuando $n \to \infty$. A la

esperanza se la conoce también con los nombres de:

- media
- valor esperado
- valor promedio
- promedio poblacional

Ejercicio: calcular la media para los dados de las primeras slides.

Propiedades de la esperanza

Si $a \in \mathbb{R}$ es una constante (un número) y X una variable aleatoria, entonces:

- $E(a) = a$
- $E(aX) = aE(X)$
- Si $X \geq 0$, entonces $E(X) \geq 0$

Si tenemos dos variables aleatorias (X e Y). Entonces,

- $E(X + Y) = E(X) + E(Y)$
- $E(X - Y) = E(X) - E(Y)$

No hay reglas genéricas para la esperanza del producto o división de dos variables aleatorias, depende del caso.

Varianza

Varianza de una v.a.

Sea X una variable aleatoria. Se define la varianza de X como

$$\sigma^2 = Var(X) = E\left[(X-\mu)^2\right],$$

donde $\mu = E(X)$. Para cada tipo de variable aleatoria:

$$Var(X) = \begin{cases} \sum_x (x-\mu)^2 p(x) & \text{si } X \text{ es discreta} \\ \int_{-\infty}^{+\infty} (x-\mu)^2 f(x)dx & \text{si } X \text{ es continua} \end{cases}$$

Intuición:

- La varianza es un promedio ponderado de la distancia cuadrática entre la media (μ) y cada x (ponderado por su probabilidad).
- Es mayor cuanto más lejos estén los x de la media, y cuanto mayor sea el peso (probabilidad de ocurrencia) de esos valores.

Varianza y desvío estándar

Si miramos la fórmula, la varianza queda expresada en términos de unidades al cuadrado. Es decir, si la variable X estaba en dólares, la varianza serán dólares al cuadrado.

Una expresión igual de importante es el **desvío estándar** de X, que es simplemente:

Desvío Estándar

Sea X una v.a. con varianza σ^2. El **desvío estándar de X** se define como

$$\sigma = SD(X) = \sqrt{Var(X)},$$

es decir, la raiz cuadrada (positiva) de la varianza.

Esta medida de dispersión de una distribución queda expresada en las unidades originales de la variable aleatoria.

Propiedades de la varianza

Si a y b son dos escalares y X una variable aleatoria:

- $Var(X) \geq 0$
- $Var(a) = 0$
- $Var(aX) = a^2 Var(X)$
- $Var(a + bX) = b^2 Var(X)$
- $Var(X) = E(X^2) - [E(X)]^2$

Si X e Y son dos variables aleatorias, a y b números:

- $V(aX + bY) = a^2 V(X) + b^2 V(Y) + 2abCov(X, Y)$
- **si** X e Y son variables aleatorias **independientes**, $V(aX + bY) = a^2 V(X) + b^2 V(Y)$ (no tienen relación, lo cual implica $Cov(X, Y) = 0$, ver más adelante)

Esperanza y varianza - Ejemplo

Su compañía tiene 1000 empleados en la Argentina. Se sabe que el salario mensual promedio es de \$10.500 con un desvío estándar de \$1.500 (en Argentina se pagan trece sueldos al año, 12 meses más uno de aguinaldo). Siguiendo las recomendaciones de política salarial de la presidente, su compañía ha decidido este año dar un bono salarial navideño equivalente a \$3.600. ¿Cuál es el salario mensual promedio? ¿Cuál es el desvío de la remuneración anual de los empleados bajo esta nueva política?

Esperanza y varianza: ejemplo/aplicación

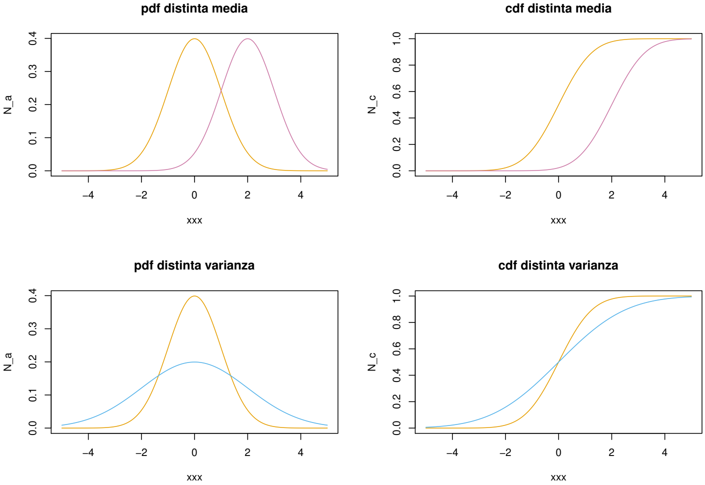
Supongamos que hay dos posibles inversiones:

- Portafolio 1: con probabilidad 0.5 obtenemos ganancias de un 3%, y con probabilidad 0.5 obtenemos pérdidas de 1%. La esperanza del retorno de este portafolio es 1%.
- Portafolio 2: con probabilidad 0.4 obtenemos ganancias de 10%, y con probabilidad 0.6 obtenemos pérdidas de 5%. La esperanza del retorno de este portafolio es 1%.

Podemos estudiar los retornos de estos portafolios como variables aleatorias. Ambas tendrían la misma esperanza, pero su comportamiento es distinto. La esperanza no está capturando la idea de que el portafolio 2 es de alguna forma más riesgoso.

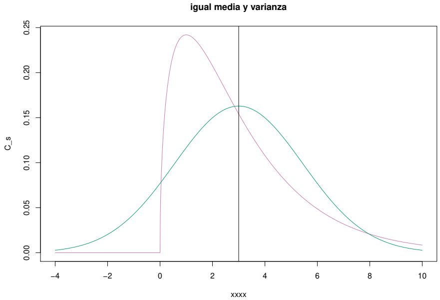
Esperanza y varianza

En resumen: La esperanza es una noción del centro de una distribución, la varianza una medida de la dispersión.



Medidas de forma

Consideremos estas dos distribuciones con $E(X) = 3$ y $Var(X) = 6$. ¿En qué difieren?



Ahora estudiaremos brevemente dos medidas de forma:

- Asimetría
- Kurtosis

Momentos absolutos de una distribución

El **momento absoluto** de orden k de una variable aleatoria X es:

$$m_k = E\left[X^k\right]$$

- Orden 0: $m_0 = E\left[X^0\right] = 1$
- Orden 1: $m_1 = E\left[X^1\right] = \mu$
- Orden 2: $m_2 = E\left[X^2\right]$
- Orden 3: $m_3 = E\left[X^3\right]$
- Orden 4: $m_4 = E\left[X^4\right]$
- ...

Momentos centrados de una distribución

El **momento central** o **centrado** de orden k de una variable aleatoria X es:

$$\mu_k = E\left[(X - E(X))^k\right]$$

- Orden 0: $\mu_0 = E\left[(X - E(X))^0\right] = 1$
- Orden 1: $\mu_1 = E\left[(X - E(X))^1\right] = 0$
- Orden 2: $\mu_2 = E\left[(X - E(X))^2\right] = \sigma^2$
- Orden 3: $\mu_3 = E\left[(X - E(X))^3\right]$
- Orden 4: $\mu_4 = E\left[(X - E(X))^4\right]$
- ...

Asimetría y curtosis

La asimetría y kurtosis se calculan a partir de los momentos centrados de orden 3 y 4, respectivamente.

Asimetría

Sea X una v.a. con varianza $sigma^2$ y momentos centrados μ_k. Se define el **coeficiente de asimetría de Fisher** como

$$\gamma_1 = \frac{\mu_3}{\sigma^3}$$

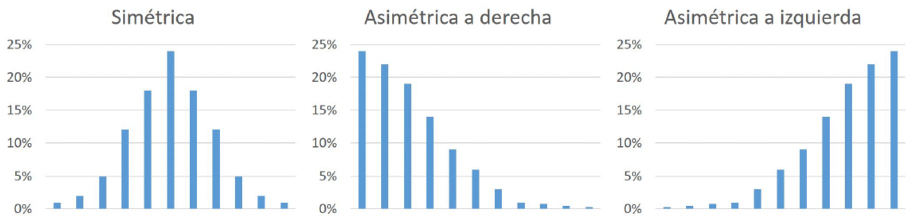
Curtosis

Sea X una v.a. con varianza σ^2 y momentos centrados μ_k. Se definen, respectivamente, el **coeficiente de curtosis** y **coeficiente de exceso de curtosis** como

$$\beta_2 = \frac{\mu_4}{\sigma^4} \quad \text{o} \quad g_2 = \frac{\mu_4}{\sigma^4} - 3$$

Asimetría

La distribución de los datos puede ser **simétrica** o **asimétrica** .

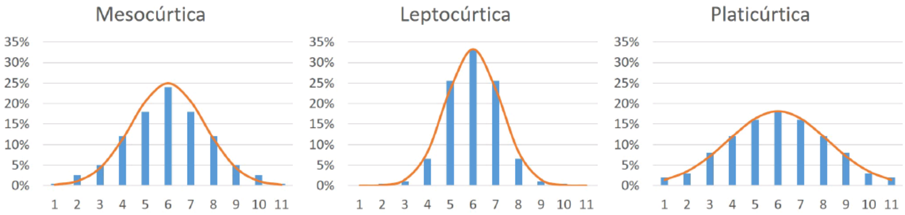


Cuando la distribución es:

- **simétrica**: Media=Mediana=Moda, $\gamma_1 = 0$
- **asimétrica a derecha (positiva)**: Media>Mediana>Moda, $\gamma_1 > 0$
- **asimétrica a izquierda (negativa)**: Media<Mediana<Moda, $\gamma_1 < 0$

Curtosis

La **curtosis** mide el grado de concentración que presentan los valores en la región central de la distribución y el "peso" de las colas de la distribución.



Usando como referencia la distribución normal, la distribución puede ser:

- Platicúrtica (más chata)
- Leptocúrtica (más en punta y colas más pesadas)
- Mesocúrtica (la distribución normal)

Covarianza

La **covarianza** entre dos variables aleatorias X e Y es una medida de la asociación **lineal** que existe entre ambas.

Covarianza

Sean X e Y dos variables aleatorias. Se define la covarianza entre X e Y como

$$\sigma_{XY} = Cov(X, Y) = E[(X - \mu_X)(Y - \mu_Y)],$$

siendo $\mu_X = E(X)$ y $\mu_Y = E(Y)$. Además, se puede calcular como

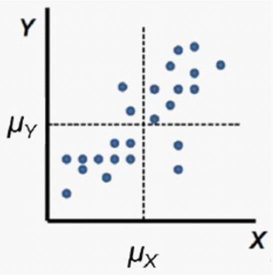
$$Cov(X, Y) = E(XY) - E(X)E(Y)$$

$$Cov(X, Y) = \begin{cases} \sum_x \sum_y (x - \mu_X)(y - \mu_Y) p(x, y) & \text{(caso discreto)} \\ \int_x \int_y (x - \mu_X)(y - \mu_Y) f(x, y)\, dy\, dx & \text{(caso continuo)} \end{cases}$$

¡Para calcularla necesitamos conocer la probabilidad conjunta!

Covarianza

Una covarianza positiva (negativa) implica que existe una asociación lineal positiva (negativa) entre el par de variables aleatorias.



Intuición:

- $Cov(X, Y) > 0$: si $\uparrow X$, $\uparrow Y$ (o si $\uparrow Y$, $\uparrow X$)
- $Cov(X, Y) < 0$: si $\uparrow X$, $\downarrow Y$ (o si $\uparrow Y$, $\downarrow X$)
- $Cov(X, Y) = 0$: **no** hay asociación **lineal** entre X e Y

Covarianza e Independencia

Diremos que dos variables aleatorias X e Y son **independientes** si $p(x, y) = p(x)p(y)$ para todos los valores en el soporte de X e Y.

Bajo independencia, también se verifica que $E(XY) = E(X)E(Y)$

Nota: Si un par de variables aleatorias son independientes, la covarianza entre ellas es 0. Sin embargo, la inversa no es necesariamente verdadera (puede haber otro tipo de relación distinta de la lineal).

Correlación

Como la covarianza depende de las unidades de medición de $X \cdot Y$, muchas veces utilizamos el **coeficiente de correlación** (ρ_{XY}) que lo escribimos de la siguiente manera:

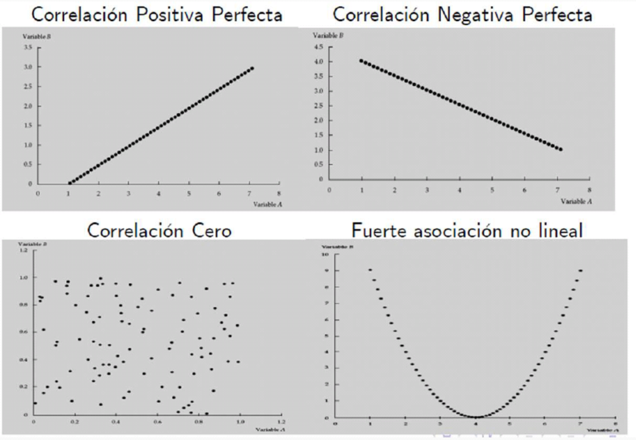
$$\rho_{X,Y} = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$$

ρ_{XY} mantiene el signo de la covarianza y es 0 cuando la covarianza es 0. Además, al dividir por el desvío estándar de cada variable, normalizamos cualquier modificación en las unidades. De esta forma,

$$-1 \leq \rho_{X,Y} \leq 1$$

El coeficiente de correlación es una medida MUY utilizada en la práctica y que retomaremos cuando veamos regresión.

Correlación



Nota: no siempre podemos caracterizar completamente dos variables y la relación que tienen entre sí con media, varianza y covarianza.

Ver: `https://blog.revolutionanalytics.com/2017/05/the-datasaurus-dozen.html`

Probabilidades Binomiales

- Mide el número de éxitos en una secuencia de n ensayos independientes entre sí, con una probabilidad fija p de ocurrencia del éxito entre los ensayos.
- **Ejemplo**: se lanzará una moneda tres veces, ¿cuál es la probabilidad de sacar exactamente una cara?

a. 0.125
b. 0.250
c. 0.333
d. 0.375
e. 0.500

Probabilidades Binomiales

- 8 resultados posibles: HHH, HHT, HTH, THH, HTT, THT, TTH, TTT
- 3 con exactamente una cara: HTT, THT y TTH
- P(sólo 1 cara)=3/8=0.375

- OK. Pero si te preguntan ¿cuál es la probabilidad de sacar una cara cuando se lanza 100 veces una moneda?

Probabilidades Binomiales

Es fácil ver que contar los casos posibles y totales se vuelve engorroso rápidamente. Vamos a darle estructura al problema.

Ensayo de Bernoulli

Se llama ensayo de Bernoulli a un experimento con dos posibles resultados, "éxito" o "fracaso". Si X es una variable aleatoria, entonces

$$X = \begin{cases} 1 & \text{“éxito”} \quad \text{con probabilidad} P(\text{éxito}) \\ 0 & \text{“fracaso”} \quad \text{con probabilidad} P(\text{fracaso}) \end{cases}$$

El ejemplo típico es el del experimento "lanzar una moneda" con "éxito=H" y "fracaso=T", pero es mucho más general que eso. En ciencias sociales y en economía, este tipo de variable aleatoria sirve para codificar información cualitativa. Por ejemplo, cualquier pregunta de encuestas con respuestas tipo "sí/no" puede representarse como un ensayo de Bernoulli.

Probabilidades Binomiales

Podemos resumir la probabilidad de cada resultado en una sola función.

Distribución de Bernoulli

Una variable aleatoria X que es un ensayo de Bernoulli tiene asociada la siguiente función de probabilidad:

$$p(x) = p^x(1-p)^{1-x},$$

donde $p = P(\text{éxito}) = P(X = 1)$

Este tipo de funciones que resumen las probabilidades de variables aleatorias con comportamiento particular se llaman modelos de probabildiad, o más comunmente, distribuciones de probabilidad.

Probabilidades Binomiales

- En nuestro ejemplo, teníamos un experimento distinto: tirábamos una moneda 3 veces, no una.
- Esto es, básicamente, tener 3 ensayos de Bernoulli (dado que cada tirada es independiente), o sea, 3 variables aleatorias X_1, X_2, X_3.
- Usemos lo que sabemos para calcular probabilidades para calcular la probabilidad de sacar sólo una cara

$$
\begin{aligned}
P(\text{solo una cara}) &= P(X_1 = 1, X_2 = 0, X_3 = 0) \\
&= p^{x_1}(1-p)^{1-x_1} \cdot p^{x_2}(1-p)^{1-x_2} \cdot p^{x_3}(1-p)^{1-x_3} \\
&= \text{reemplazando resultados 1 y 0} \\
&= p \cdot (1-p)^2
\end{aligned}
$$

Probabilidades Binomiales

- Una forma más directa es usar la función de probabilidad **binomial**.
- Llamemos n a los intentos, x a los éxitos que consideramos y p a la probabilidad de éxito.

$$P(\text{obtener x exitos en n intentos}) = \binom{n}{k} p^x (1-p)^{n-x}$$

donde $\binom{n}{x} = \frac{n!}{x!(n-x)!}$, se conoce como "combinatoria de n en x"

y $n! = 1 \times 2 \times 3 \times ... \times (n-1) \times n$, se conoce como "factorial de n".

- En Excel: `BINOM.DIST()` o `DISTR.BINOM()`
- La combinatoria aparece porque hay casos de igual probabilidad: la probabildad de sacar HHT es la misma que la de sacar HTH o THH.
- Calcular $P(X=0), ..., P(X=3)$ para un p genérico.

Distribución Binomial

Supongamos que tenemos un experimento aleatorio que puede darnos dos resultados posibles: éxito o fracaso, donde p es la probabilidad de éxito. Si se llevan a cabo n repeticiones independientes, la distribución del número de éxitos X resultantes se conoce como **distribución binomial**.

Distribución Binomial

Sea X con posibles resultados $0, 1, 2, ..., n$. Si X representa el número de éxitos en n intentos (ensayos de Bernoulli), entonces su función de probabilidad es:

$$p(X = x) = \frac{n!}{x!(n-x)!} p^x (1-p)^{(n-x)} \text{ para } x = 0, 1, 2, ..., n$$

Notación: $X \sim Bi(p, n)$

La **esperanza** y la **varianza** de X están dadas por:

- $E(X) = n.p$
- $V(X) = n.p.(1-p)$

Ejemplo

(cualquier parecido con la realidad es pura coincidencia)

Un examen final consta de 10 preguntas multiple choice (con 4 opciones cada una) donde sólo una respuesta es correcta. Un alumno no estudió para el final y no sabe la respuesta de ninguna de las diez preguntas por lo que decide contestar al azar (al no penalizar por respuesta incorrecta, es una buena estrategia tirarle a pegar). ¿Cuál es la probabilidad de aprobar el examen (acertar 5 o más preguntas)?
Probar calcular en el Excel y a mano.

1. 1%
2. 2.5%
3. 5%
4. 8%
5. 19%
6. 25%

Distribución de Poisson

Modela la probabilidad de una cantidad de ocurrencias del mismo evento durante un período de tiempo con cierta frecuencia de ocurrencia media ($\lambda > 0$).

Distribución de Poisson

Sea $X \in \Omega = \{0, 1, 2, \dots\}$, $X \sim Po(\lambda)$ si:

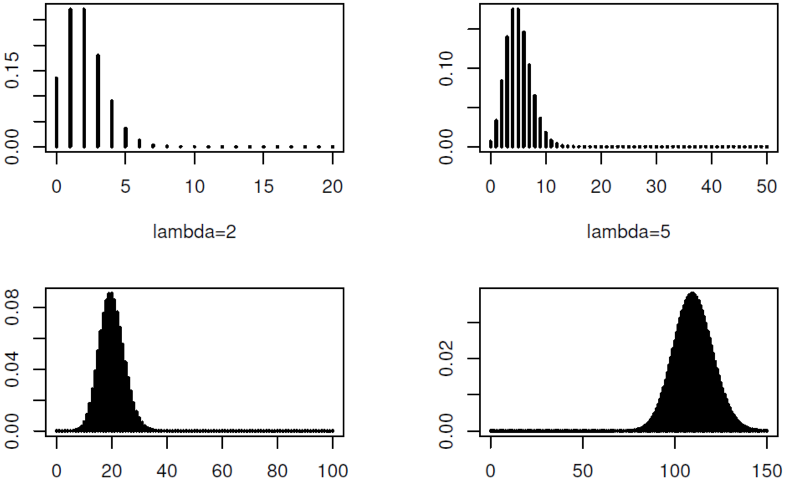
$$p(x) = \begin{cases} \dfrac{e^{-\lambda}\lambda^x}{x!}, & x = 0, 1, 2, \dots \\ 0, & \text{si no} \end{cases}$$

donde: $E(X) = \lambda$, $Var(X) = \lambda$.

Algunas aplicaciones:

- Número de accidentes en un tramo de ruta en un día.
- Número de llegadas en un sistema de colas en una hora.

Distribución de Poisson



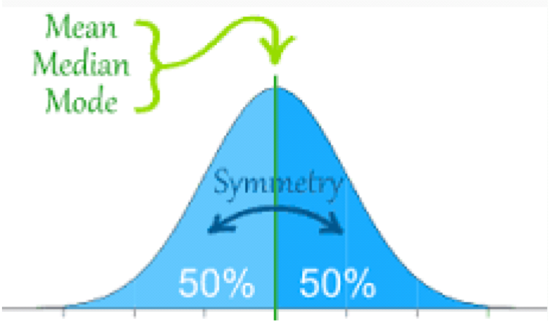
Distribución de Poisson - Ejemplo

El número de personas que ingresan a determinado sitio de internet por día, tiene una distribución del tipo Poisson con media igual a 16.
El manager de un negocio sabe que entran, en promedio, 16 clientes por día.

(a) ¿Cuál es la probabilidad de que en un día ingresen 10 clientes?
(b) ¿Cuál es la probabilidad de que en una hora ingresen 2 o más clientes?
(c) ¿Cree que es factible que en un lapso de 2 días ingresen más de 60 clientes?

Distribución Normal

- Es la distribución más común (de allí su nombre)
- Tiene forma de campana
- Depende sólo de la media y la varianza (2 parámetros)
- Su notación es: $X \sim N(\mu, \sigma^2)$ y se lee como "X sigue una distribución Normal con media μ y varianza σ^2". Donde μ y σ^2 tendrán ciertos valores que definirán la forma de la distribución.



Distribución Normal

This shape of distribution is actually fairly common and can be applied to lots of situations. It's called the **normal distribution**.

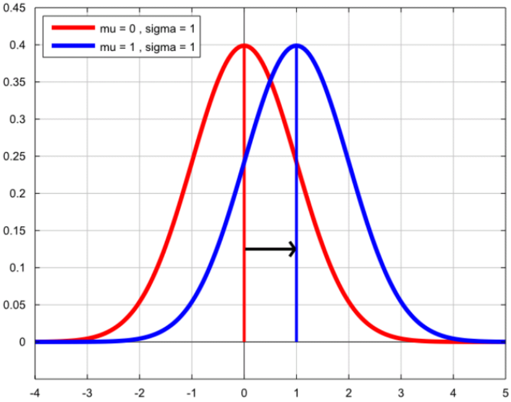
Normalidad en la Realidad

Bajar de **Yahoo Finance** (finance.yahoo.com) una serie histórica del precio de cierre ajustado de alguna acción con frecuencia diaria entre el 4 de enero de 2016 y el 22 de julio de 2016. Computar el retorno y obtener el histograma.

- Ingresar a: finance.yahoo.com
- Escribir el símbolo o nombre de la empresa
- Hacer click en "Historical Data"
- Ingresar el rango de fechas y frecuencia
- Descargar a Excel
- Computar los retornos.
- Ver el histograma (usar el add-in: *Analysis ToolPak*).
- ¿Parece seguir una distribución normal?

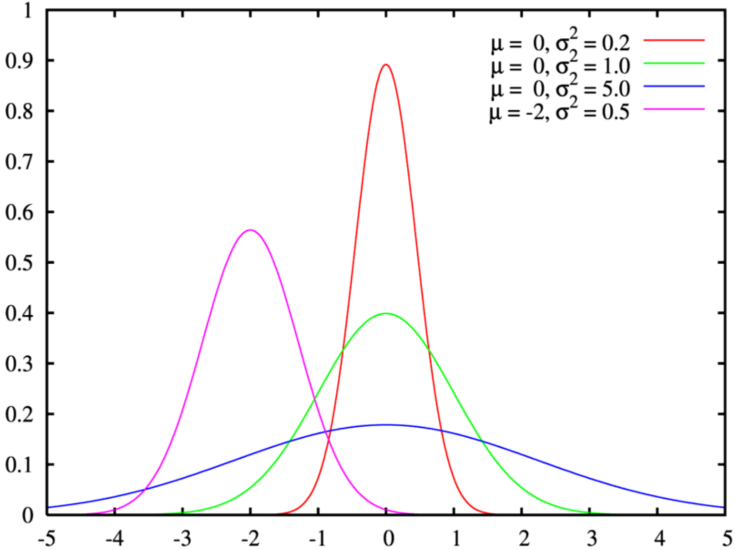
Distribución Normal

Como la distribución es **simétrica**, la media se sitúa en el centro de la curva (coincide con la mediana). Si se cambia μ sin cambiar σ se produce un desplazamiento de la curva de densidad a lo largo del eje horizontal sin que cambie su dispersión.



Distribución Normal

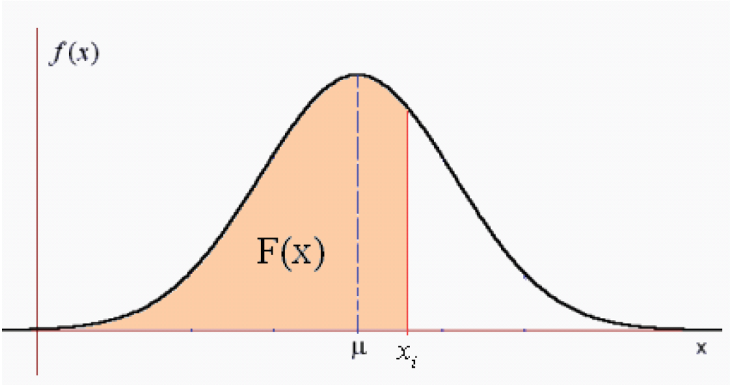
En cambio, a mayor varianza, la curva presenta mayor dispersión:



Función de Distribución Acumulada

Si X es una variable aleatoria normal con media μ y varianza σ^2, esto es, $X \sim N(\mu, \sigma^2)$. Entonces, la **función de distribución acumulada**, $F(x)$, es:

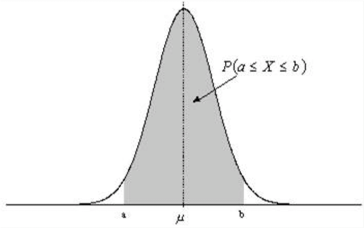
$$F(x) = P(X \leq x)$$



Función de Distribución Acumulada

Entonces, se puede usar la función de distribución acumulada para calcular rangos de probabilidad. Es decir,

$$P(a \leq X \leq b) = F(b) - F(a)$$



Calcular dichas probabilidades puede ser muy tedioso (para cada μ y σ^2 posibles). Afortunadamente, los cálculos se facilitan reexpresando la variable normal como una $N(0,1)$, este cálculo que se denomina **estandarización**.

Distribución Normal Estándar

Si Z es una variable aleatoria normal con media 0 y varianza 1; esto es, $Z \sim N(0,1)$. Entonces Z se dice que sigue una **distribución normal estándar**. Ahora mostraremos como las probabilidades de cualquier variable aleatoria normal pueden ser expresadas en términos de una variable aleatoria normal estándar mediante la **estandarización**.

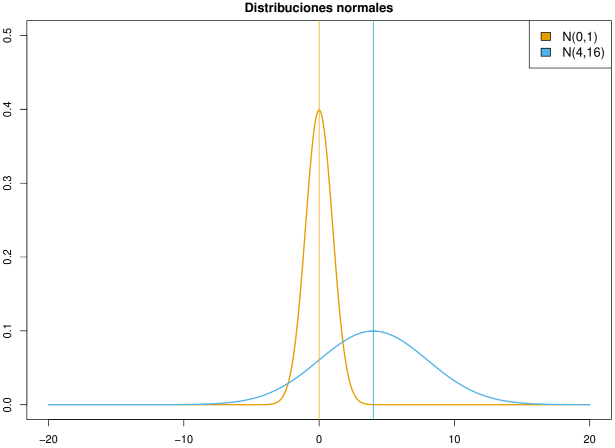
Distribución Normal estándar

Sea $X \sim N(\mu, \sigma^2)$, sustrayendo la media y dividiendo por el desvío estándar obtenemos una variable aleatoria Z que tiene media 0 y varianza 1. Es decir,

$$Z = \frac{X - \mu}{\sigma}, \text{ donde } Z \sim N(0,1)$$

Distribución Normal Estándar

Para facilitar la comparación entre normales se las “estandariza”:



Distribución Normal Estándar

Una regla simple para pensar en la distribución normal es la siguiente:

- A ± 1 desvío estándar, se concentra el 68% de los datos
- A ± 2 desvíos estándar, se concentra el 95% de los datos
- A ± 3 desvíos estándar, se concentra el 99.7% de los datos

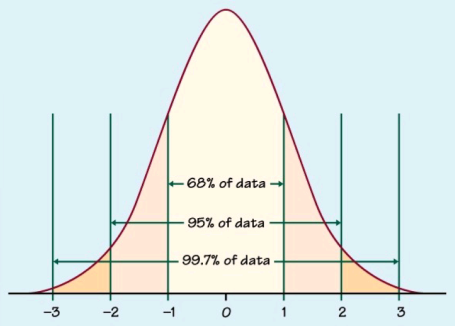
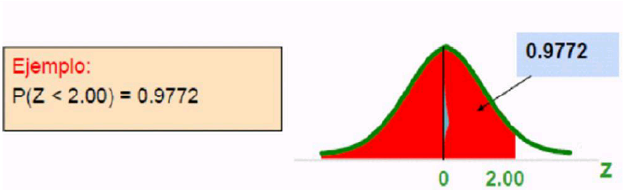


Tabla de la Normal Estándar

La tabla de la Distribución Normal Estándar Acumulada nos da las probabilidades de que Z sea **menor o igual** a un valor dado.

La ventaja de usar la tabla de la normal estándar es que alguien ya calculó las probabilidades!!!



El análogo moderno de "las tablas" es cualquier programa que calcule probabildades, ya sea Excel o algún software estadístico (Stata), o lenguaje de programación (R, Python). También existe una app de iPhone y Android: "Probability Distributions".

Tabla de la Normal Estándar

DISTRIBUCION NORMAL ESTANDARIZADA

z	F(z)	z	F(z)	z	F(z)	z	F(z)
-3.50	0.000	-2.91	0.002	-2.32	0.010	-1.73	0.042
-3.49	0.000	-2.90	0.002	-2.31	0.010	-1.72	0.043
-3.48	0.000	-2.89	0.002	-2.30	0.011	-1.71	0.044
-3.47	0.000	-2.88	0.002	-2.29	0.011	-1.70	0.045
-3.46	0.000	-2.87	0.002	-2.28	0.011	-1.69	0.046
-3.45	0.000	-2.86	0.002	-2.27	0.012	-1.68	0.046
-3.44	0.000	-2.85	0.002	-2.26	0.012	-1.67	0.047
-3.43	0.000	-2.84	0.002	-2.25	0.012	-1.66	0.048
-3.42	0.000	-2.83	0.002	-2.24	0.013	-1.65	0.049
-3.41	0.000	-2.82	0.002	-2.23	0.013	**-1.64**	**0.051**
-3.40	0.000	-2.81	0.002	-2.22	0.013	-1.63	0.052
-3.39	0.000	-2.80	0.003	-2.21	0.014	-1.62	0.053
-3.38	0.000	-2.79	0.003	-2.20	0.014	-1.61	0.054
-3.37	0.000	-2.78	0.003	-2.19	0.014	-1.60	0.055
-3.36	0.000	-2.77	0.003	-2.18	0.015	-1.59	0.056
-3.35	0.000	-2.76	0.003	-2.17	0.015	-1.58	0.057
-3.34	0.000	-2.75	0.003	-2.16	0.015	-1.57	0.058
-3.33	0.000	-2.74	0.003	-2.15	0.016	-1.56	0.059
-3.32	0.000	-2.73	0.003	-2.14	0.016	-1.55	0.061
-3.31	0.000	-2.72	0.003	-2.13	0.017	-1.54	0.062
-3.30	0.000	-2.71	0.003	-2.12	0.017	-1.53	0.063
-3.29	0.001	-2.70	0.003	-2.11	0.017	-1.52	0.064
-3.28	0.001	-2.69	0.004	-2.10	0.018	-1.51	0.066
-3.27	0.001	-2.68	0.004	-2.09	0.018	-1.50	0.067
-3.26	0.001	-2.67	0.004	-2.08	0.019	-1.49	0.068
-3.25	0.001	-2.66	0.004	-2.07	0.019	-1.48	0.069
-3.24	0.001	-2.65	0.004	-2.06	0.020	-1.47	0.071
-3.23	0.001	-2.64	0.004	-2.05	0.020	-1.46	0.072
-3.22	0.001	-2.63	0.004	-2.04	0.021	-1.45	0.074
-3.21	0.001	-2.62	0.004	-2.03	0.021	-1.44	0.075

DISTRIBUCION NORMAL ESTANDARIZADA

z	F(z)	z	F(z)	z	F(z)	z	F(z)
-1.14	0.127	-0.55	0.291	0.04	0.516	0.63	0.736
-1.13	0.129	-0.54	0.295	0.05	0.520	0.64	0.739
-1.12	0.131	-0.53	0.298	0.06	0.524	0.65	0.742
-1.11	0.133	-0.52	0.302	0.07	0.528	0.66	0.745
-1.10	0.136	-0.51	0.305	0.08	0.532	0.67	0.749
-1.09	0.138	-0.50	0.309	0.09	0.536	0.68	0.752
-1.08	0.140	-0.49	0.312	0.10	0.540	0.69	0.755
-1.07	0.142	-0.48	0.316	0.11	0.544	0.70	0.758
-1.06	0.145	-0.47	0.319	0.12	0.548	0.71	0.761
-1.05	0.147	-0.46	0.323	0.13	0.552	0.72	0.764
-1.04	0.149	-0.45	0.326	0.14	0.556	0.73	0.767
-1.03	0.152	-0.44	0.330	0.15	0.560	0.74	0.770
-1.02	0.154	-0.43	0.334	0.16	0.564	0.75	0.773
-1.01	0.156	-0.42	0.337	0.17	0.567	0.76	0.776
-1.00	0.159	-0.41	0.341	0.18	0.571	0.77	0.779
-0.99	0.161	-0.40	0.345	0.19	0.575	0.78	0.782
-0.98	0.164	-0.39	0.348	0.20	0.579	0.79	0.785
-0.97	0.166	-0.38	0.352	0.21	0.583	0.80	0.788
-0.96	0.169	-0.37	0.356	0.22	0.587	0.81	0.791
-0.95	0.171	-0.36	0.359	0.23	0.591	0.82	0.794
-0.94	0.174	-0.35	0.363	0.24	0.595	0.83	0.797
-0.93	0.176	-0.34	0.367	0.25	0.599	0.84	0.800
-0.92	0.179	-0.33	0.371	0.26	0.603	0.85	0.802
-0.91	0.181	-0.32	0.374	0.27	0.606	0.86	0.805
-0.90	0.184	-0.31	0.378	0.28	0.610	0.87	0.808
-0.89	0.187	-0.30	0.382	0.29	0.614	0.88	0.811
-0.88	0.189	-0.29	0.386	0.30	0.618	0.89	0.813
-0.87	0.192	-0.28	0.390	0.31	0.622	0.90	0.816
-0.86	0.195	-0.27	0.394	0.32	0.626	0.91	0.819
-0.85	0.198	-0.26	0.397	0.33	0.629	0.92	0.821

Ejemplo

La cantidad de personas que entran a la sala de emergencias de un hospital público por día puede aproximarse razonablemente mediante una distribución normal con una media de 80 personas y un desvío estándar de 40. ¿Cuál es la probabilidad de que un día lleguen más de 120 pacientes a la sala de emergencias? (Aproxime)

(A) 16%

(B) 50%

(C) 84%

(D) Ninguna de las anteriores

Dos tipos de problemas

Tipo 1. Encontrar una probabilidad para un valor x:

$$P(X < x) = ?$$

- Función DISTR.NORM.ESTAND()
- Función DISTR.NORM()

Tipo 2. Dada una probabilidad, encontrar para qué valor de la distribución Normal se cumple esa probabilidad:

$$P(X < ?) = p$$

- Función DISTR.NORM.ESTAND.INV()
- Función DISTR.NORM.INV()

Dos tipos de problemas

X es una variable aleatoria que representa "la venta semanal de revistas en un kiosko de diarios y revistas". En promedio, un kiosko vende 100 revistas semanales con un desvío de 25. Asuma que las ventas siguen una distribución normal.

Tipo 1. ¿Cuál es la probabilidad de vender menos de 75 revistas?

Tipo 2. ¿Y si me dicen que quieren saber cuál es lo máximo que puede llegar a vender un kiosko con 95% de probabilidad (es decir, el 95% de las veces)?

Intente responder ambas preguntas "a mano" y en el Excel.

Ejercicio 1

La compañía de seguros Segurola S.A. afirma y perjura que el costo promedio de sus siniestros es de \$10.000 con un desvío estándar de \$1.500. Si asumimos que el costo de un siniestro es una variable aleatoria que sigue una distribución normal, con los parámetros que Segurola dice. ¿Cuál es la probabilidad de que un siniestro dado tenga un costo mayor a \$13.489,5?

Ejercicio 2

Usted está evaluando la posibilidad de lanzar un nuevo negocio, el de *service express* de mantenimiento de autos. Las pruebas realizadas en su taller con los mecánicos que se dedicarían a realizar los services muestran que el tiempo requerido por un mecánico para cada service puede ser aproximado por una variable aleatoria con una distribución Normal con media de 22 minutos y una desviación estándar de 5 minutos. ¿Cada cuánto tendría que dar los turnos si desea que la probabilidad de que un cliente espere su turno porque el mecánico no terminó con el turno anterior sea de 5%?

Bibliografía sugerida

- Bertsekas et al. (2000) - Cap. 2 y 3
- Newbold et al. (2013) - Cap 4 y 5
- Wooldridge. Introducción a la Econometría (múltiples ediciones) - Apéndice B