Teoría Econométrica I - EAE- 250-A

Introducción al curso, Probabilidades y Variables Aleatorias

Tatiana Rosá

Instituto de Economía - Pontificia Universidad Católica de Chile

Agosto 2021

Detalles administrativos

- Clases 15:30-16:50, Martes y Jueves
- Ayudantía: Miercoles a las 17:00
- Horario de atención: Miercoles 08:00 a 09:30 (agendarse por email)
- Contacto: tatiana.rosa@uc.cl
- Textos sugeridos:
    - Amemiya, T., 1994. Introduction to statistics and econometrics. Harvard University Press.
    - Casella, G. and Berger, R.L., 2021. Statistical inference. Cengage Learning.
    - Greene, W.H., 2000. Econometric analysis 4th edition. International edition, New Jersey: Prentice Hall, pp.201-215.
    - Goldberger, A.S. and Goldberger, A.S.G., 1991. A course in econometrics. Harvard University Press.
    - Wooldridge, J.M., 2010. Econometric analysis of cross section and panel data. MIT press

Ayudantías

- Vamos tenes ayudantías casi todas las semanas
- Algunas serán ayudantías tradicionales y otras en Matlab
- En las ayudantías tradicionales se corregirán los problem sets
- Ayudante: José Alejo Eyzaguirre Ercilla (`jeeyzaguirre@uc.cl`)
- Hay que bajar la licencia de Matlab disponible en canvas

Evaluaciones y aprobación del curso

- Controles para asegurar la continuidad del aprendizaje (5%)
- Problem sets (15%)
- Tareas grupales (20%)
- Prueba (20%)
- Examen final (40%)
- EL curso se aprueba con una calificación promedio de 4 siempre que se obtenga al menos un 3 en el examen final.

Que busco yo con este curso...

- ¡¡Qué aprendan más y mejor econometría !!
- Este curso está orientado a estudiantes de magister o doctorado en economía y corresponde a un primer curso de econometría teórica de un semestre de duración
- Si bien este es un curso teórico intentaré ligarlo con aplicaciones de economía empírica
- Objetivo; que entiendan los fundamentos de la econometría, loas aspectos más algebraicos, más estadísticos, sus propiedades, sus limitaciones
- Trabajaremos mucho con álgebra lineal: deben estar cómodos con eso.
- Tip personal: este ramo se disfruta si no lo corren de atrás....

Probabilidades

- Probabilidad: es un medida de la cualidad de posible de un evento cuyo resultado es incierto
- La axiomatización de la teoría de la probabilidad que conocemos (Kolmogorov) se da en un marco experimental
- Este experimento tiene resultados posibles que son elementos de un ***espacio muestral.***

  $\Rightarrow$ Mismo experimentos muchas veces: diferentes resultados!
- Medida: es una función que asigna un número (que en nuestro caso sería una probabilidad) a los subconjuntos de un conjunto dado

  $\Rightarrow$ Solo podremos asignar una medida a los conjuntos que son medibles, nuestro ***espacio muestral.***

Espacio Muestral, Eventos y Sigma-algebra

DEFINICIONES:

- El ***espacio muestral*** Ω es el conjunto de todos los posibles resultados (realizaciones) de un experimento.

- Elementos individuales $\omega \in \Omega$ se llaman ***resultados elementales*** o resultados
- Un subconjunto $\mathcal{B} \in \Omega$ (una colección de resultados) es llamado un **σ-álgebra** si:

  (i) $\emptyset \in \mathcal{B}$.
  (ii) $B \in \mathcal{B} \Rightarrow B^C \in \mathcal{B}$ ($\mathcal{B}$ es cerrado bajo complementación).
  (iii) $B_1, B_2, ... \in \mathcal{B} \Rightarrow \bigcup_{i=1}^{\infty} B_i \in \mathcal{B}$ ($\mathcal{B}$ es cerrado bajo uniones contables).

  Ejemplo: si $\Omega = \{1, 2, 3, 4\}$, un σ-algebra de Ω es la colección $\mathcal{B} = \{\emptyset, \{1, 2\}, \{3, 4\}, \Omega\}$.

Función de Probabilidad y espacio de probabilidad

Podemos definir ahora una ***función de probabilidad*** de forma axiomatica

- Una función $P : \mathcal{B} \to [0,1]$ definido en una σ-álgebra $\mathcal{B}$ es una ***función de probabilidad*** si
  (i) $P(\Omega) = 1$.
  (ii) $P(A) \geq 0, A \in \mathcal{B}$.
  (iii) $P\left(\bigcup_{i=1}^{\infty} B_i\right) = \sum_{i=1}^{\infty} P(B_i)$ cuando $B_1, B_2, ... \in \mathcal{B}$ sean disjuntos de a pares ($B_i \bigcap B_j = \emptyset, \forall i \neq j$).

  $\rightarrow$ la función de probabilidad mapea desde el σ-álgebra $\mathcal{B}$ al compacto $[0,1]$ y no desde Ω.
- Un ***espacio de probabilidad*** es una tripleta $(\Omega, \mathcal{B}, P)$ donde Ω es un espacio muestral, $\mathcal{B}$ es una σ-álgebra de eventos y P una función de probabilidad.

  $\rightarrow$ Nos van a interesar los espacios donde las realizaciones elementales ω son (vectores de) números reales.

  $\rightarrow$ Estos espacios se pueden cosntruir a partir de variables aleatorias

Variables aleatorias

- Una ***Variable Aleatoria*** es una función medible con valores reales definida en Ω, la denotamos $X : \Omega \to \mathbb{R}$.
- Un ***vector aleatorio*** es un vector de variables aleatorias.
- Cualquier variable aleatoria $X : \Omega \to \mathbb{R}$ induce un espacio de probabilidad $(\mathbb{R}, \mathcal{B}(\mathbb{R}), P_X)$, donde $\mathcal{B}(\mathbb{R})$ es una σ-álgebra definida en $\mathbb{R}$ y $P_X = P \circ X^{-1}$; eso es,

$$P_X(B) = P \circ X^{-1}(B) = P\left(\{\omega : X(\omega) \in B\}\right), \quad B \in \mathcal{B}(\mathbb{R})$$

- Las propiedades de una variable aleatoria X están completamente caracterizadas por la función de probabilidad P_X.
- Una caracterización alternativa es provista por la función de distribución acumulada de X.

Función de distribución acumulada

- Sea X una variable aleatoria definida en el espacio de probabilidad $(\Omega, \mathcal{B}, P)$. La ***función de distribución acumulada (cdf)*** de X es la función $F_X : \mathbb{R} \to [0, 1]$ definida por

$$F_X(x) = P_X((-\infty, x]) = P_X(X \leq x), \quad x \in \mathbb{R}$$

- Si conocemos P_X automáticamente conocemos F_X (Casella y Berger, Teorema 1.5.10)
- Dado que F_X es más fácil de trabajar que P_X (su dominio es $\mathbb{R}$ y de esta forma se puede graficar la función), es mucho más conveniente caracterizar las propiedades de una variable aleatoria X en términos de F_X
- **Definición:** (Casella y Berger, Teorema 1.5.3) *Una función* $F : \mathbb{R} \to [0, 1]$ *es una cdf si y solo si las siguientes tres condiciones se satisfacen:*
  (i) $\lim_{x \to -\infty} F(x) = 0$ *y* $\lim_{x \to +\infty} F(x) = 1$.
  (ii) $F(x)$ *es no decreciente.*
  (iii) $F(x)$ *es continua por la derecha.*

Función continua por la derecha

- Una función $F : \mathbb{R} \to \mathbb{R}$ es ***continua por la derecha*** en $x_0 \in \mathbb{R}$ si para cualquier $\epsilon > 0$, existe un $\delta > 0$ tal que $|F(x) - F(x_0)| < \epsilon$ cuando $x_0 < x < x_0 + \delta$.

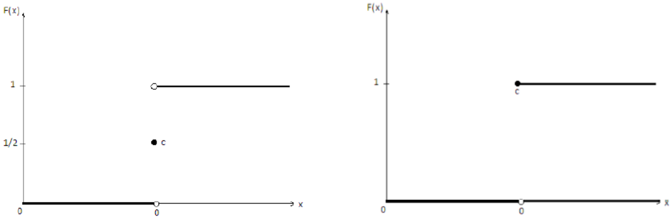

Figure: *No continua por la derecha (izq.) y Continua por al derecha (der.)*

- ***Tip:*** Necesitamos que si nos movemos de izquierda a derecha todos los intervalos sean abiertos por la derecha

Ejemplos de cdf

- Una variable aleatoria X tiene una *distribución Bernoulli* con parámetro $p \in [0,1]$, denotada $X \sim Ber(p)$, si
$$F_X(x) = \begin{cases} 0 & \text{para} \quad x < 0 \\ 1-p & \text{para} \quad 0 \leq x < 1 \\ 1 & \text{para} \quad x \geq 1 \end{cases}$$
- Una variable aleatoria X tiene una *distribución uniforme* en $[0,1]$, denotada $X \sim U[0,1]$, si
$$F_X(x) = \begin{cases} 0 & \text{para} \quad x < 0 \\ x & \text{para} \quad 0 \leq x < 1 \\ 1 & \text{para} \quad x \geq 1 \end{cases}$$
- Una variable aleatoria X tiene una *distribución normal estándar*, denotada $X \sim \mathcal{N}(0,1)$, si
$$F_X(x) = \int_{-\infty}^{x} \phi(t)dt, \quad x \in \mathbb{R},$$
donde
$$\phi(t) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}t^2\right), \quad t \in \mathbb{R}$$

VA continuas y discretas

(i) X es una ***variable aleatoria discreta*** si existe una función $f_X : \mathbb{R} \to [0,1]$ tal que

$$F_X(x) = \sum_{t \leq x} f_X(t) \quad \forall x \in \mathbb{R}$$

La función f_X es la ***función de masa de probabilidad (pmf)*** de X.

(ii) X es una ***variable aleatoria continua*** si existe una función $f_X : \mathbb{R} \to \mathbb{R}_+$ tal que

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt \quad \forall x \in \mathbb{R}$$

Cualquier función de este tipo es una ***función de densidad de probabilidad (pdf)*** de X.

Observación. La cdf de una variable aleatoria discreta es una *step function*.

Ejemplos de pmf y pdf

- Si $X \sim Ber(p)$, X es discreta con pmf

$$f_X(x) = \begin{cases} 1-p & \text{para} \quad x = 0 \\ p & \text{para} \quad x = 1 \\ 0 & \text{otro caso.} \end{cases}$$

- Si $X \sim U[0,1]$, la cdf de X es no diferenciable (en 0 y en 1). A pesar de esto X es continua con pdf

$$f_X(x) = \begin{cases} 1 & \text{para} \quad 0 \leq x \leq 1 \\ 0 & \text{otro caso.} \end{cases}$$

- Si $X \sim \mathcal{N}(0,1)$. X es continua con pdf

$$f_X(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right), \quad x \in \mathbb{R}$$

Definición pmf y pdf (Casella y Berger, Teorema 1.6.5)

- *Una función* $f : \mathbb{R} \to [0,1]$ *es una pmf de una variable aleatoria discreta si y solo si*
  - (i) $f(x) \geq 0$
  - (ii) $\sum_{x \in \mathbb{R}} f(x) = 1$
- *Una función* $f : \mathbb{R} \to \mathbb{R}_+$ *es una pdf de una variable aleatoria continua si y solo si*
  - (i) $f(x) \geq 0$
  - (ii) $\int_{-\infty}^{+\infty} f(x) dx = 1$
- Este teorema nos da condiciones generales que deben satisfacer las pdf y pmf.
- Podemos ser más explícitos:
  - Para la pmf, tenemos que $f(x) = Pr(X = x)$
  - Para la pdf, usando el Teorema Fundamental del Cálculo tenemos que $f(x) = \frac{\partial F_X(x)}{\partial x}$.
- Recuerdo: cuando X es un variable aleatoria continua, tenemos que

$$Pr(a < X < b) = \int_a^b f_X(x) dx$$

Propiedades de las Variables Aleatorias

- Las propiedades distribucionales de una variable aleatoria X están típicamente caracterizadas por:
  - (i) La cdf F_X directamente;
  - (ii) La pdf (pmf) f_X directamente si X es continua (discreta); o
  - (iii) Definiendo $X = g(Z)$, donde $g : \mathbb{R} \to \mathbb{R}$ es una función (medible) y Z es una variable aleatoria con cdf F_Z conocida.
- La alternativa (iii) es muy común en estadística y econometría
- Si X es una variable aleatoria con cdf F_X, ¿cuáles son las de propiedades distribucionales de una variable aleatoria $Y = g(X)$?
- La VA Y va a inducir un espacio de probabilidad de la forma $(\mathbb{R}, \mathcal{B}(\mathbb{R}), P_Y)$.
- Las propiedades de Y dependen de las propiedades de g y de X.
  - Si X es discreta, también lo es Y
  - Si X es continua, las propiedades de Y dependen tb de las propiedades de g

Funciones de VAs

- $X \sim \mathcal{N}(0,1)$.
- Considere las variables aleatorias $g_1(X)$, $g_2(X)$ y $g_3(X)$, donde

$$g_1(x) = x, \quad x \in \mathbb{R}$$

$$g_2(x) = \begin{cases} 0 & \text{para} \quad x \leq 0 \\ 1 & \text{para} \quad x > 0 \end{cases}$$

$$g_3(x) = \begin{cases} 0 & \text{para} \quad x \leq 0 \\ x & \text{para} \quad x > 0 \end{cases}$$

- $g_1(X)$ es continua (en efecto, $g_1(X) \sim \mathcal{N}(0,1)$)
- $g_2(X) \sim Ber(\frac{1}{2})$ es discreta
- $g_3(X)$ no es ni discreta ni continua.
- Las transformaciones g_2 y g_3 son populares en econometría y serán vistas con mayor detalle más adelante

cdf, pmf y pdf defunciones de VAs

- La cdf de Y está dada por:

$$F_Y(y) = P(\{\omega : Y(\omega) \leq y\}) = P(\{\omega : g(X(\omega)) \leq y\}), \quad y \in \mathbb{R}$$

y es relativamente fácil encontrarla cuando:

(i) Y es discreta; o

(ii) g es monótona (ver **Casella y Berger, Teorema 2.1.3**).

Sea X una variable aleatoria con cdf F_X y $Y = aX + b$, donde $a > 0$, $b \in \mathbb{R}$

- La cdf de Y es

$$F_Y(y) = F_X\left(\frac{y-b}{a}\right), \quad y \in \mathbb{R}$$

- Si X es discreta con pmf f_X, Y es discreta con pmf f_Y dada por

$$f_Y(y) = f_X\left(\frac{y-b}{a}\right), \quad y \in \mathbb{R}$$

- Si X es continua con pdf f_X, Y es continua con pdf f_Y dada por

$$f_Y(y) = \frac{1}{a} f_X\left(\frac{y-b}{a}\right), \quad y \in \mathbb{R}$$

Escalas y localizaciones

- Sea X una variable aleatoria con cdf F_X y $Y = aX + b$
- Al mantener $a = 1$ fijo y variando b, una ***familia de localizaciones*** de distribuciones es generada
- Al mantener $b = 0$ fijo y variando a, una ***familia de escalas*** de distribuciones es generada.
- Al variar a y b, una *familia de localización-escala* de distribuciones es generada
- Las familias de localización-escala de distribuciones son muy comúnes en estadística y econometría.
- Ejemplo: $Y = \sigma X + \mu$ y $X \sim \mathcal{N}(0, 1)$
- Con distintos μ y $\sigma > 0$ tenemos $Y \sim \mathcal{N}(\mu, \sigma^2)$ y decimos que Y *se distribuye normal* con media μ y varianza σ^2, donde los conceptos de media y varianza serán definido

Generalización transformaciones no lineales de VAs

- Sea X una variable aleatoria con cdf $F_X(x)$.
- Sea $Y = g(X)$ y $\mathcal{X} = \{x : f_X(x) > 0\}$ e $\mathcal{Y} = \{y : y = g(x),\ x \in \mathcal{X}\}$
  a) Si g es creciente en $\mathcal{X}$, $F_Y(y) = F_X(g^{-1}(y))$ para $y \in \mathcal{Y}$.
  b) Si g es decreciente en $\mathcal{X}$ y $\mathcal{X}$ una variable aleatoria continua, $F_Y(y) = 1 - F_X(g^{-1}(y))$ para $y \in \mathcal{Y}$.
- Si la pdf de Y es continua la podemos obtener derivando la cdf teniendo cuidado con la monotonicidad de g (Casella y Berger 2.1.5)
- Sea X una variable aleatoria con pdf $f_X(x)$. Sea $Y = g(X)$, donde g es una función monótona. Suponga que $f_X(x)$ es continua en $\mathcal{X}$ y que $g^{-1}(y)$ tiene derivada continua en $\mathcal{Y}$. Luego, la *pdf* de Y está dada por

$$f_Y(y) = \begin{cases} f_X(g^{-1}(y))|\frac{d}{dy}g^{-1}(y) & y \in \mathcal{Y} \\ 0 & y \notin \mathcal{Y} \end{cases}$$

- La prueba es directa, sólo debemos derivar y aplicar regla de la cadena.

Momentos de las VAs

(i) Sea X una variable aleatoria discreta con pmf f_X y sea $g : \mathbb{R} \to [0, 1]$ cualquier función. El ***valor esperado*** de $g(X)$, denotado por $\mathbb{E}(g(X))$, es

$$\mathbb{E}(g(X)) = \sum_{x \in \mathbb{R}} g(x) f_X(x),$$

provisto que $\sum_{x \in \mathbb{R}} g(x) f_X(x) < +\infty$. En caso contrario, decimos que el valor esperado no existe.

(ii) Sea X una variable aleatoria continua con pdf f_X y sea $g : \mathbb{R} \to \mathbb{R}$ cualquier función. El ***valor esperado*** de $g(X)$, denotado por $E(g(X))$, es

$$E(g(X)) = \int_{-\infty}^{+\infty} g(x) f_X(x) dx,$$

provisto que $\int_{-\infty}^{+\infty} g(x) f_X(x) < +\infty$. En caso contrario, decimos que el valor esperado no existe.

Media y Varianza

- La ***media*** de una VA X se denota por $\mu = E(X)$
- La ***varianza*** de X se denota por

$$Var(X) = \sigma^2 = \mathbb{E}\left((X-\mu)^2\right) = \mathbb{E}(X^2) - \mu^2$$

- La ***desviación estandar*** de X se denota $\sigma = \sqrt{Var(X)}$
- Ejemplo: Si $X \sim \mathcal{N}(0,1)$, entonces

$$\mathbb{E}(X) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x \exp\left(-\frac{1}{2}x^2\right) dx = -\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right)\Bigg|_{x=-\infty}^{+\infty} = 0$$

$$\begin{aligned} Var(X) = \mathbb{E}(X^2) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x^2 \exp\left(-\frac{1}{2}x^2\right) dx \\ &= -\frac{1}{\sqrt{2\pi}} x \exp\left(-\frac{1}{2}x^2\right)\Bigg|_{x=-\infty}^{+\infty} + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \exp\left(-\frac{1}{2}x^2\right) dx = 1. \end{aligned}$$

Otros momentos de la distribución de una VA

- El *k-ésimo momento* de una variable aleatoria X es

$$\mu_k = E\left(X^k\right), \quad k \in \mathbb{N} = \{1, 2, ...\}.$$

- El *k-ésimo momento central* de X es $\mu_k = E\left((X-\mu)^k\right)$.
- ¿Son media y varianza momentos centrales?
- Otros momentos utilizados frecuentemente:
  - **Coeficiente de asimetría** (*skewness*): $E\left((X-\mu)^3\right)$
  - **kurtosis** : $E\left((X-\mu)^4\right)$.
- La media y varianza de transformaciones afines de X (Casella y Berger, Teorema 2.2.5a y Teorema 2.3.4)
- Si a y b son constantes y X una variable aleatoria, entonces

$$E(aX+b) = aE(X) + b \text{ y}$$

$$Var(aX+b) = a^2 Var(X),$$

Desigualdad de Jensen

- Una función $g : \mathbb{R} \to \mathbb{R}$ es ***convexa*** si $g(\lambda x + (1-\lambda)y) \leq \lambda g(x) + (1-\lambda)g(y)$ para todo x, y y cualquier $\lambda \in (0,1)$.
  $\Rightarrow$ *Tip:* si grafico la función y trazo una recta entre dos puntos cualquiera la recta siempre estará por encima
- Una función $g : \mathbb{R} \to \mathbb{R}$ es ***cóncava*** si $-g$ es convexa.
- Una función g dos veces diferenciable es convexa si y solo si $g'' \geq 0$ para todo x.
- En particular, una función afín g es convexa y cóncava y la siguiente famosa desigualdad generaliza el resultado $E(g(X)) = g(E(X))$
- **Teorema Desigualdad de Jensen** (Casella y Berger, Teorema 4.7.7)
  Si X es una variable aleatoria y g es una función convexa, entonces

$$E(g(X)) \geq g(E(X)).$$

Graficamos la Desigualdad de Jensen

Desigualdad de Chebychev

- **Desigualdad de Chebychev** (Casella y Berger, Teorema 3.6.1)

  Si X es una variable aleatoria y g es una función no negativa entonces

$$P(g(X) \geq r) \leq \frac{E(g(X))}{r}, \quad \forall r > 0.$$

- **EJEMPLO:** Sea X una variable aleatoria con $\mathbb{E}(X) = \mu$. Para cualquier $r > 0$,

$$P(|X| > r) \leq \frac{\mathbb{E}|X|}{r}$$

$$P(|X| > r) = P(X^2 > r^2) \leq \frac{\mathbb{E}(X^2)}{r^2}$$

$$P(|X - \mu| > r) = P\left((X - \mu)^2 > r^2\right) \leq \frac{Var(X)}{r^2}$$

Vector Aleatorio Bivariado

- Un ***vector aleatorio bivariado*** es un vector (X, Y), donde X e Y son variables aleatorias (definidas en el mismo espacio de probabilidad $(\Omega, \mathcal{B}, P)$)
- Un vector aleatorio bivariado $(X, Y) : \Omega \to \mathbb{R}^2$ induce un espacio de probabilidad $(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2), P_{X,Y})$, donde $\mathcal{B}(\mathbb{R}^2)$ es una σ-álgebra de Borel definida en $\mathbb{R}^2$ y

$$P_{X,Y}(B) = P(\{\omega : (X(\omega), Y(\omega)) \in B\}), \quad B \in \mathcal{B}(\mathbb{R}^2)$$

Distribución conjunta

- **Definición.** Una ***cdf conjunta*** de (X, Y) es la función $F_{X,Y} : \mathbb{R}^2 \to [0, 1]$ definida por

$$F_{X,Y}(x, y) = P_{X,Y}((-\infty, x] \times (-\infty, y]) = P(\{\omega : X(\omega) \leq x, Y(\omega) \leq y\}), \quad (x, y$$

$\Rightarrow$ *Tip:* Podemos pensarlo como la probabilidad de que los eventos pasen al mismo tiempo

- Notar que si conocemos $F_{X,Y}$ conocemos $P_{X,Y}$ y viceversa
- $F : \mathbb{R}^2 \to [0, 1]$ es una cdf conjunta si y solo si:
  (i) $\lim_{x \to -\infty} F(x, y) = 0$ para cualquier y, $\lim_{y \to -\infty} F(x, y) = 0$ para cualquier x y en donde $\lim_{x \to +\infty, y \to +\infty} F(x, y) = 1$.
  (ii) F es no decreciente, esto es, $F(x', y') \geq F(x, y)$ cuando $x' \geq x$ y $y' \geq y$.
  (iii) F es continua por la derecha; esto es, para cualquier $\epsilon > 0$ y cualquier $(x_0, y_0) \in \mathbb{R}^2$, existe un $\delta > 0$ tal que $|F(x, y) - F(x_0, y_0)| < \epsilon$ cuando $x_0 \leq x < x_0 + \delta$ y $y_0 < y < y_0 + \delta$.

Gráfico de una distribución normal bivariada

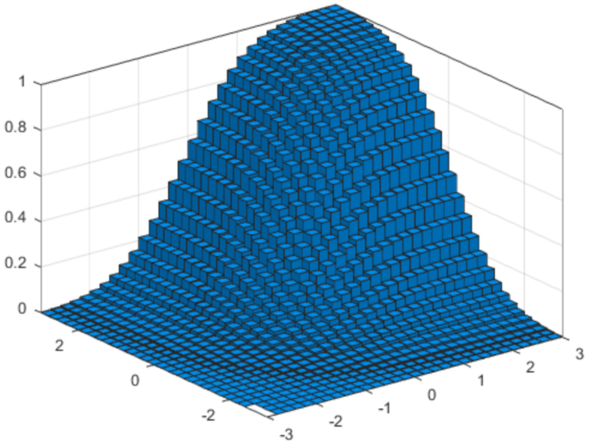

Figure: *Distribución normal bivariada*

Pmf y cdf conjuntas

- **Definición.** Sea (X, Y) un vector aleatorio bivariado con cdf conjunta $F_{X,Y}$.
  (i) (X, Y) es un ***vector aleatorio discreto*** si existe $f_{X,Y}$ no negativa tal que

$$F_{X,Y}(x,y) = \sum_{s \leq x, t \leq y} f_{X,Y}(s,t) \quad \forall (x,y) \in \mathbb{R}^2$$

  (ii) (X, Y) es un ***vector aleatorio continuo*** si existe $f_{X,Y}$ no negativa tal que

$$F_{X,Y}(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(t,s) ds dt \quad \forall (x,y) \in \mathbb{R}^2$$

- **Definición.** (Casella y Berger, p. 142 y 145)
  (i) Una función $f : \mathbb{R}^2 \to [0,1]$ es una pmf conjunta de un vector aleatorio discreto *ssi:*

$$\sum_{(x,y) \in \mathbb{R}^2} f(x,y) = 1.$$

  (ii) Una función $f : \mathbb{R}^2 \to \mathbb{R}_+$ es una pdf conjunta de un vector aleatorio continuo*ssi:*

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x,y) dy dx = 1.$$

Gráfico de una pdf normal bivariada

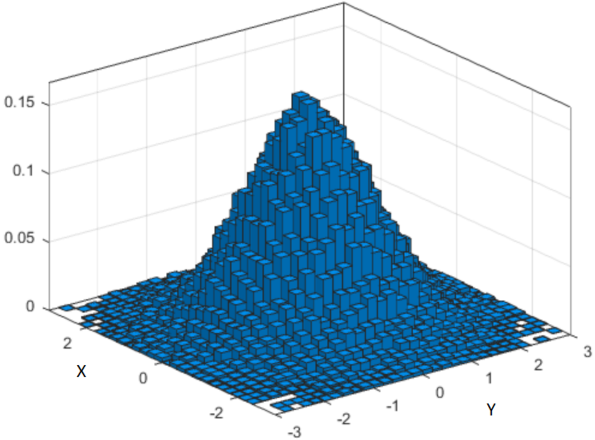

Figure: *Densidad de una normal bivariada*

Distribuciones marginales

- **Definición.** Sea (X, Y) es vector aleatorio bivariado. La cdf de X se llama la ***cdf marginal*** de X
  (i) Si (X, Y) es discreto, X es una variable aleatoria discreta y su pmf se llama la ***pmf marginal*** de X
  (ii) Si (X, Y) es continuo, X es una variable aleatoria continua y una pdf de X se llama la ***pdf marginal*** de X.
- La distribución conjunta de (X, Y) determina las distribuciones marginales de X e Y.
  (i) Si (X, Y) es discreto con pmf $f_{X,Y}$, entonces la pmf marginal f_X de X satisface (Casella y Berger, Teorema 4.1.6)

$$f_X(x) = \sum_{y \in \mathbb{R}} f_{X,Y}(x, y) \quad \forall x \in \mathbb{R}.$$

  (ii) Si (X, Y) es continuo con pdf $f_{X,Y}$, entonces una pdf marginal de X satisface:

$$f_X(x) = \begin{cases} \int_{-\infty}^{+\infty} f_{X,Y}(x, y)dy & \text{si} \quad \int_{-\infty}^{+\infty} f_{X,Y}(x, y)dy < +\infty \\ 0 & \text{si} \quad \int_{-\infty}^{+\infty} f_{X,Y}(x, y)dy = +\infty \end{cases}, \quad x \in \mathbb{R}.$$

Pdf conjunta y marginal de una VA normal bivariada

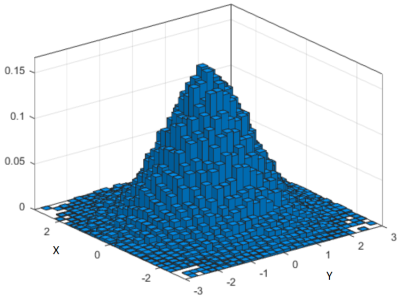

Figure: *Densidad de una normal bivariada*

Distribuciones condicionales

- Introduciremos la distribución condicional de X (dado $Y = y$)
- La distribución condicional de X dado $Y = y$ está bien definida aún si (X, Y) no es ni discreto ni continuo, al igual que las esperanzas

  $\Rightarrow$ A nosotros nos alcanzas los casos continuos/discretos
- **Definición.** (i) Sea (X, Y) un vector aleatorio discreto bivariado con una pmf conjunta $f_{X,Y}$ y una pmf marginal f_Y de Y. Para cualquier y tal que $f_Y(y) > 0$, la ***pmf condicional de X dado $Y = y$*** es la función $f_{X|Y}(\cdot|y) : \mathbb{R} \to [0, 1]$ dada por

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x, y)}{f_Y(y)}, \quad x \in \mathbb{R}$$

  (ii) Sea (X, Y) un vector aleatorio continuo bivariado con una pdf conjunta $f_{X,Y}$ y una pdf marginal f_Y de Y. Para cualquier y tal que $f_Y(y) > 0$, la ***pdf condicional de X dado $Y = y$*** es la función $f_{X|Y}(\cdot|y) : \mathbb{R} \to \mathbb{R}_+$ dada por

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x, y)}{f_Y(y)}, \quad x \in \mathbb{R}$$

Distribución conjunta, marginal y condicional

- Ordenando la ecuación que define $f_{X|Y}(x|y)$, llegamos a la importante relación

$$f_{X,Y}(x,y) = f_{X|Y}(x|y) \cdot f_Y(y)$$

- **Las distribuciones marginales de** X **e** Y **no determinan la distribución conjunta de** (X, Y) a menos que la distribución condicional en $Y = y$ sea igual a la distribución marginal de X para todos los valores de y
- **La distribución conjunta de** (X, Y) **siempre determina las distribuciones marginales de** X **e** Y

  $\Rightarrow$ lo contrario no se mantiene a menos que X e Y sean independientes!
- **Independencia Definición.** Sea (X, Y) un vector aleatorio discreto (continuo) bivariado con pmf (pdf) conjunta $f_{X,Y}$ y pmfs (pdfs) marginales f_X y f_Y. Las variables aleatorias X e Y son ***independientes*** si

$$f_{X,Y}(x,y) = f_X(x) f_Y(y), \quad \forall (x,y) \in \mathbb{R}^2.$$

Ejemplo: Resultados conjuntos

	Male	Female	Total
Football	120	75	195
Rugby	100	25	125
Other	50	130	180
	270	230	500

Figure: *Resultados conjuntos*

Ejemplo: Pmf conjunta, marginal y condicional

	Male	Female	Total
Football	0.24	0.15	0.39
Rugby	0.2	0.05	0.25
Other	0.1	0.26	0.36
	0.54	0.46	1

Figure: *Pmf*

Esperanza condicional

- Para cualquier y fijo con $f_Y(y) > 0$, la pmf (pdf) condicional $f_{X|Y}(\cdot|y)$ es una pmf (pdf) entonces tiene sentido definir las esperanzas condicionales con respecto a la distribución de X condicional en $Y = y$

- **Definición.** (i) Sea (X, Y) un vector aleatorio discreto y sea $g : \mathbb{R} \to \mathbb{R}$ una función. Para cualquier y tal que $f_Y(y) > 0$, el ***valor esperado condicional de $g(X)$ dado $Y = y$*** es denotado por $\mathbb{E}_{X|Y}(g(X)|y)$ y es dado por

$$\mathbb{E}_{X|Y}(g(X)|y) = \sum_{x\in\mathbb{R}} g(x) f_{X|Y}(x|y),$$

provisto que $\sum_{x\in\mathbb{R}} g(x) f_{X|Y}(x|y) < +\infty$.

Esperanza condicional

(ii) Sea (X, Y) un vector aleatorio continuo y sea $g : \mathbb{R} \to \mathbb{R}$ una función. Para cualquier y tal que $f_Y(y) > 0$, el ***valor esperado condicional de*** $g(X)$ ***dado*** $Y = y$ es denotado por $\mathbb{E}_{X|Y}(g(X)|y)$ y es dado por

$$\mathbb{E}_{X|Y}(g(X)|y) = \int_{-\infty}^{+\infty} g(x) f_{X|Y}(x|y) dx,$$

provisto que $\int_{-\infty}^{+\infty} g(x) f_{X|Y}(x|y) dx < +\infty$. En caso contrario, decimos que el valor esperado condicional no existe.

- Para distinguir entre las esperanzas condicionales de la esperanza de $g(X)$ con respecto a la distribución marginal de X ocasionalmente denotaremos lo último como $\mathbb{E}_X(g(X))$

Media y varianza condicional

- La *media condicional de X dado $Y = y$* es $\mathbb{E}_{X|Y}(X|y)$
- La *varianza condicional de X dado $Y = y$* es

$$Var_{X|Y}(X|y) = \mathbb{E}_{X|Y}\left((X - \mathbb{E}_{X|Y}(X|y))^2\right) = \mathbb{E}_{X|Y}(X^2|y) - \mathbb{E}_{X|Y}(X|y)^2$$

- Para cualquier y fijo, tanto la media condicional $\mathbb{E}_{X|Y}(X|y)$ como la varianza condicional $Var_{X|Y}(X|y)$ son solo números fijos.
- Viendo a $\mathbb{E}_{X|Y}(X|\cdot)$ y a $Var_{X|Y}(X|\cdot)$ como funciones de y!!!

  $\Rightarrow$ Podemos definir las variables aleatorias $\mathbb{E}_{X|Y}(X|Y)$ y $Var_{X|Y}(X|Y)$

Teoremas sobre la media y varianza condicional

- Establecen importantes relaciones entre los momentos de estas variables (funciones de Y) y momentos de X

Ley de Esperanzas Iteradas (Casella y Berger, Teorema 4.4.3).

- *Para cualquier vector aleatorio bivariado* (X, Y)*,*

$$\mathbb{E}_X(X) = \mathbb{E}_Y(\mathbb{E}_{X|Y}(X|Y)),$$

- Si algún lado existe también existe el otro y son iguales

Identidad de la Varianza Condicional (Casella y Berger, Teorema 4.4.7)

- *Para cualquier vector aleatorio bivariado* (X, Y)*,*

$$Var_X(X) = \mathbb{E}_Y(Var_{X|Y}(X|Y)) + Var_Y(\mathbb{E}_{X|Y}(X|Y)),$$

- Si algún lado existe también existe el otro y son iguales

Esperanzas respecto a las pdf/pmf conjuntas

El caso discreto

- Sea (X, Y) un vector aleatorio discreto bivariado con pmf conjunta $f_{X,Y}$
- Sea $g : \mathbb{R}^2 \to \mathbb{R}$ una función.
- El ***valor esperado de*** $g(X, Y)$, denotado $\mathbb{E}(g(X, Y))$, es

$$\mathbb{E}(g(X, Y)) = \sum_{(x,y)\in\mathbb{R}^2} g(x, y) f_{X,Y}(x, y),$$

provisto que $\sum_{(x,y)\in\mathbb{R}^2} g(x, y) f_{X,Y}(x, y) < +\infty$.

El caso continuo

- Sea (X, Y) un vector aleatorio continuo bivariado con pdf conjunta $f_{X,Y}$
- Sea $g : \mathbb{R}^2 \to \mathbb{R}$ una función
- El ***valor esperado de*** $g(X, Y)$, denotado $\mathbb{E}(g(X, Y))$, es

$$\mathbb{E}(g(X, Y)) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(x, y) f_{X,Y}(x, y) dydx,$$

provisto que $\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(x, y) f_{X,Y}(x, y) dydx < +\infty$

Esperanzas, covarianzas y coeficiente de correlación

- Si (X, Y) es un vector aleatorio bivariado entonces

$$\mathbb{E}(X+Y) = \mathbb{E}(X) + \mathbb{E}(Y)$$

- Sea (X, Y) un vector aleatorio bivariado. La ***covarianza*** de X e Y es

$$Cov(X, Y) = \mathbb{E}((X-\mathbb{E}(X))(Y-\mathbb{E}(Y))) = \mathbb{E}(XY)-\mathbb{E}(X)\mathbb{E}(Y) = Cov(Y, X).$$

- La ***correlación*** de X e Y es el ***coeficiente de correlación*** ρ_{XY} definido por

$$\rho_{XY} = \frac{Cov(X, Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}}.$$

- Sea (X, Y) un vector aleatorio bivariado. La ***media (vector)*** de (X, Y) es

$$\mathbb{E}\begin{pmatrix} X \\ Y \end{pmatrix} = \begin{pmatrix} \mathbb{E}(X) \\ \mathbb{E}(Y) \end{pmatrix}.$$

Matriz de covarianzas e Independencia

- La ***matriz de covarianza*** de (X, Y) es

$$Var\begin{pmatrix} X \\ Y \end{pmatrix} = \begin{pmatrix} Var(X) & Cov(X, Y) \\ Cov(Y, X) & Var(Y) \end{pmatrix}.$$

$\Rightarrow$ La matriz de covarianza de cualquier vector aleatorio bivariado (X, Y) es simétrica y semidefinida positiva. La matriz de covarianza es singular si y solo si $|\rho_{XY}| = 1$.

Independencia Teorema (Casella y Berger, Teorema 4.5.5)

- Sea (X, Y) un vector aleatorio bivariado
- Si X e Y son independientes, entonces:

$$Cov(X, Y) = \rho_{XY} = 0$$

Varianza y coeficiente de correlación

- La varianza de la variable aleatoria $g(X, Y)$ está definida de la forma natural:

$$Var(g(X,Y)) = E\left((g(X,Y) - E(g(X,Y)))^2\right)$$

- En el caso especial donde $g(x, y) = x + y$ una caracterización útil de $Var(g(X, Y))$ es la siguiente:
- **Teorema (Casella y Berger, Teorema 4.5.6).** *Si (X, Y) es un vector aleatorio bivariado, entonces*

$$Var(X+Y) = Var(X) + Var(Y) + 2Cov(X,Y)$$

$\Rightarrow Var(aX + bY)$?

- **Teorema (Casella y Berger, Teorema 4.5.7).** *Si (X, Y) es un vector aleatorio bivariado entonces:*
  1. $|\rho_{XY}| \leq 1$
  2. $|\rho_{XY}| = 1$ sii existen números $a \neq 0$ y b tal que $P(Y = aX + b) = 1$.

Varianza y coeficiente de correlación

- La desigualdad $|\rho_{XY}| \leq 1$ es un caso especial del siguiente resultado:
  Desigualdad de Cauchy-Schwarz (Casella y Berger, Teorema 4.7.3).
- Si (X, Y) es un vector aleatorio bivariado. entonces:

$$|E(XY)| \leq E|XY| \leq (E(X^2))^{1/2}(E(Y^2))^{1/2}.$$

- La Desigualdad de Cauchy-Schwarz es un caso especial (cuando $p = q = 2$) de la Desigualdad de Hölder (Casella y Berger, Teorema 4.7.2) de acuerdo a la cual:

$$|E(XY)| \leq E|XY| \leq (E(|X|^p))^{1/p}(E(|Y|^q))^{1/q}$$

cuando p y q son números positivos tales que $p^{-1} + q^{-1} = 1$.

Distribuciones Multivariadas

- Vamos a generalizar lo visto en $\mathbb{R}^2$ para $\mathbb{R}^n$
- Un ***vector aleatorio*** *n − dimensional* es un vector $X = (X_1, \dots, X_n)'$, donde $X_1, \dots, X_n$ son variables aleatorias (definidas en el mismo espacio de probabilidad $(\Omega, \mathcal{B}, P)$
- Un vector aleatorio n-dimensional $X : \Omega \to \mathbb{R}^n$ induce un espacio de probabilidad $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), P_X)$, donde $\mathcal{B}(\mathbb{R}^n)$ es una σ-álgebra de Borel definida en $\mathbb{R}^n$ y

$$P_X(B) = P(\{\omega : (X(\omega)) \in B\}), \quad B \in \mathcal{B}(\mathbb{R}^n).$$

- La ***cdf conjunta de un vector aleatorio n-dimensional X*** es la función $F_X : \mathbb{R}^n \to [0, 1]$ definida por

$$\begin{aligned} F_X(x) &= P_X((-\infty, x_1] \times \dots \times (-\infty, x_n]) \\ &= P(\{\omega : X_1(\omega) \le x_1, \dots, X_n(\omega) \le x_n\}), \quad x = (x_1, \dots, x_n)' \in \mathbb{R}^n. \end{aligned}$$

pdf/pmf Multivariadas

- Sea X un vector aleatorio n-dimensional con cdf conjunta F_X.
  (i) X es un ***vector aleatorio discreto*** si existe una función no negativa f_X tal que

$$F_X(x) = \sum_{t \leq x} f_X(t) \quad \forall x \in \mathbb{R}^n$$

  donde $''t \leq x''$ es una abreviación para $t_1 \leq x_1, \ldots, t_n \leq x_n$. La función f_X es la *pmf conjunta* de X.
  (ii) X es un vector aleatorio continuo si existe una función no negativa f_X tal que

$$F_X(x) = \int_{t \leq x} f_X(t) dt \quad \forall x \in \mathbb{R}^n$$

  Cualquier función que cumpla lo anterior es una *pdf conjunta* de X.

Valor esperado

- **Definición.** (i) Sea X un vector aleatorio discreto n-dimensional con pmf conjunta f_X y sea $g : \mathbb{R}^n \to \mathbb{R}$ una función. El ***valor esperado de*** $g(X)$, denotado por $\mathbb{E}(g(X))$, es

$$\mathbb{E}(g(X)) = \sum_{x \in \mathbb{R}^n} g(x) f_X(x),$$

provisto que $\sum_{x \in \mathbb{R}^n} g(x) f_X(x) < +\infty$

(ii) Sea X un vector aleatorio continuo n-dimensional con pdf conjunta f_X y sea $g : \mathbb{R}^n \to \mathbb{R}$ una función. El ***valor esperado de*** $g(X)$, denotado por $\mathbb{E}(g(X))$, es

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

provisto que $\int_{\mathbb{R}^n} g(x) f_X(x) dx < +\infty$.

Notación matricial

- De forma más general, sea

$$g = \begin{pmatrix} g_{11} & \cdots & g_{1m} \\ \vdots & \ddots & \vdots \\ g_{k1} & \cdots & g_{km} \end{pmatrix} : \mathbb{R}^n \to \mathbb{R}^{k\times m}$$

una función con valores matriciales. El ***valor esperado de*** $g(X)$, denotado por $\mathbb{E}(g(X))$, es

$$\mathbb{E}(g(X)) = \begin{pmatrix} \mathbb{E}(g_{11}(X)) & \cdots & \mathbb{E}(g_{1m}(X)) \\ \vdots & \ddots & \vdots \\ \mathbb{E}(g_{k1}(X)) & \cdots & \mathbb{E}(g_{km}(X)) \end{pmatrix}$$

Media y matriz de varianzas y covarianzas

- Sea X un vector aleatorio n-dimensional. La ***media (vector) de X***, denotada $\mathbb{E}(X)$, es

$$\mathbb{E}(X) = \mu = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_n \end{pmatrix},$$

donde $\mu_i = E(X_i)$, $1 \leq i \leq n$

La ***matriz de covarianza*** de X, denotada $Var(X)$, es

$$Var(X) = E\left((X-\mu)(X-\mu)'\right) = \Sigma = \begin{pmatrix} \sigma_{11} & \cdots & \sigma_{1n} \\ \vdots & \ddots & \vdots \\ \sigma_{n1} & \cdots & \sigma_{nn} \end{pmatrix},$$

donde $\sigma_{ij} = Cov(X_i, X_j)$, $1 \leq i, j \leq n$

Algunas propiedades

- Sea $X = (X_1, \ldots, X_n)'$ un vector aleatorio n-dimensional. Si $a_1, \ldots, a_n$ y $b_1, \ldots, b_n$ son constantes, entonces:

$$E\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i E(X_i), \text{ y}$$

$$Cov\left(\sum_{i=1}^{n} a_i X_i, \sum_{j=1}^{n} b_j X_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_i b_j Cov(X_i, X_j),$$

- Como caso especial del último resultado tenemos que:

$$Var\left(\sum_{i=1}^{n} a_i X_i\right) = Cov\left(\sum_{i=1}^{n} a_i X_i, \sum_{j=1}^{n} a_j X_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j Cov(X_i, X_j),$$

- y si las VA no están correlacionadas ($Cov(X_i, X_j) = 0$):

$$Var\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i^2 Var(X_i)$$

De forma matricial

- Una reexpresión de estas identidades puede obtenerse al definir los vectores $a = (a_1, \dots, a_n)'$ y $b = (b_1, \dots, b_n)'$. Específicamente, tenemos que

$$
\begin{aligned}
E(a'X) &= a'E(X) \\
Cov(a'X, b'X) &= a'Var(X)b, \text{ y} \\
Var(a'X) &= a'Var(X)a.
\end{aligned}
$$

- Para cualquier vector aleatorio $X = (X_1, \dots, X_n)'$ la matriz de covarianza $\Sigma = Var(X)$ es simétrica porque $Cov(X_i, X_j) = Cov(X_j, X_i)$ para cualquier $1 \leq i, j \leq n$.
- Σ es semidefinida positiva porque $a'\Sigma a = Var(a'X) \geq 0$ para cualquier vector no nulo $a \in \mathbb{R}^n$
- En efecto, Σ es definida positiva a menos que podamos encontrar un vector no nulo a tal que $Var(a'X) = 0$.

Particiones y cdf, pdf y pmf marginales

- Sea X un vector aleatorio n-dimensional y particione X en la k-ésima fila como

$$X = \begin{pmatrix} Y \\ Z \end{pmatrix},$$

donde $Y = (X_1, \ldots, X_k)'$ y $Z = (X_{k+1}, \ldots, X_n)'$.

- La cdf de Y se llama la ***cdf marginal de*** Y
- Si X es discreto, Y es un vector aleatorio discreto y su pmf se llama la ***pmf marginal de*** Y
- Si X es continuo, Y es un vector aleatorio continuo y su pdf se llama la ***pdf marginal de*** Y

Condicionales

- Si $(Y', Z')'$ es un vector aleatorio discreto con pmf conjunta $f_{Y,Z}$ y pmf marginal f_Z de Z, la ***pmf condicional de Y dado $Z = z$*** es la función $f_{Y|Z}(\cdot|z)$ dada por

$$f_{Y|Z}(y|z) = \frac{f_{Y,Z}(y,z)}{f_Z(z)},$$

para cualquier $y \in \mathbb{R}^k$ y cualquier $z \in \mathbb{R}^{n-k}$ tal que $f_Z(z) > 0$

- Si $(Y', Z')'$ es un vector aleatorio continuo con pdf conjunta $f_{Y,Z}$ y pdf marginal f_Z de Z, la ***pdf condicional de Y dado $Z = z$*** es la función $f_{Y|Z}(\cdot|z)$ dada por

$$f_{Y|Z}(y|z) = \frac{f_{Y,Z}(y,z)}{f_Z(z)},$$

para cualquier $y \in \mathbb{R}^k$ y cualquier $z \in \mathbb{R}^{n-k}$ tal que $f_Z(z) > 0$

Independencia

- Sean $X_1, \ldots, X_n$ vectores aleatorios discretos (continuos) (no necesariamente de la misma dimensión) con pmf (pdf) conjunta $f_{X_1,\ldots,X_n}$ y pmfs (pdfs) marginales $f_{X_1}, \ldots, f_{X_n}$. Los vectores aleatorios $X_1, \ldots, X_n$ son ***mutuamente independientes*** si

$$f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = f_{X_1}(x_1) \cdot \ldots \cdot f_{X_n}(x_n) \qquad \forall x_1, \ldots, x_n.$$

  $\Rightarrow$ La independencia mutua se preserva bajo transformaciones de los vectores aleatorios individuales

- En este curso los vectores aleatorios multivariados serán a menudo:
  (i) normalmente distribuidos (o funciones con valores vectoriales de vectores aleatorios normalmente distribuidos); y/o
  (ii) vectores aleatorios mutuamente independientes con distribuciones marginales idénticas.

Casos especiales: la distribución normal multivariada

- Un vector aleatorio n-dimensional $X = (X_1, \ldots, X_n)'$ está ***normalmente distribuido con media (vector)***

$$\mu = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_n \end{pmatrix}$$

y matriz de covarianza

$$\Sigma = \begin{pmatrix} \sigma_{11} & \cdots & \sigma_{1n} \\ \vdots & \ddots & \vdots \\ \sigma_{n1} & \cdots & \sigma_{nn} \end{pmatrix},$$

denotado $X \sim \mathcal{N}(\mu, \Sigma)$, si X es continuo con pdf conjunta f_X dada por

$$f_X(x) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)\right), \quad x \in \mathbb{R}^n$$

- Como la terminología sugiere, X tiene media $\mu = \mathbb{E}(X)$ y matriz de covarianza $\Sigma = Var(X) = \mathbb{E}\left((X-\mu)(X-\mu)'\right)$

Conjuntas y marginales de la normal multivariada

- Cuando X se distribuye normal también lo está cualquier subvector de X. De forma más general tenemos el siguiente resultado.

Teorema (Ruud, Lema 10.3). *Suponga que $X \sim \mathcal{N}(\mu, \Sigma)$ es un vector aleatorio n-dimensional. Si $A \in \mathbb{R}^{m \times n}$ tiene rango m y $b \in \mathbb{R}^m$, entonces*

$$AX + b \sim \mathcal{N}\left(A\mu + b, A\Sigma A'\right).$$

Aplicación

- Para cualquier $i \in \{1, \ldots, n\}$, sea e_i el i-ésimo vector unitario en $\mathbb{R}^n$ (es decir, $e_i \in \mathbb{R}^n$ tiene un uno en la i-ésima posición y ceros en cualquier otro lugar)
- Fijando $A = e_i$, $b = 0$ y aplicando el teorema tenemos:

$$X_i \sim \mathcal{N}(\mu_i, \sigma_{ii}).$$

- **Importante:** normalidad conjunta implica normalidad marginal, mientras que lo contrario no necesariamente es cierto

Condicionales de la normal multivariada

Teorema (Ruud, Lemas 10.4 y 10.5). *Si*

$$\begin{pmatrix} Y \\ Z \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu_Y \\ \mu_Z \end{pmatrix}, \begin{pmatrix} \Sigma_{YY} & \Sigma_{YZ} \\ \Sigma_{ZY} & \Sigma_{ZZ} \end{pmatrix}\right), \text{ entonces}$$

$$Y|Z = z \sim \mathcal{N}\left(\mu_Y - \Sigma_{YZ}\Sigma_{ZZ}^{-1}(z - \mu_Z), \Sigma_{YY} - \Sigma_{YZ}\Sigma_{ZZ}^{-1}\Sigma_{ZY}\right).$$

- Los vectores aleatorios Y y Z son (mutuamente) independientes si y solo si $\Sigma_{YZ} = 0$.

La distribución chi cuadrado

- Una variable aleatoria X tiene una ***una distribución chi cuadrado con p grados de libertad***, denotada $X \sim \chi^2(p)$, si X es continua con una pdf f_X dada por

$$f_X(x) = \begin{cases} 0 & \text{para} \quad x \leq 0 \\ \frac{1}{\Gamma(p/2)2^{p/2}} x^{(p/2)-1} \exp\left(-\frac{1}{2}x\right) & \text{para} \quad x > 0 \end{cases},$$

  donde Γ es la función gamma.

- Un hecho útil acerca de la distribución normal multivariada es el siguiente:

  Lema (Ruud, Lema 10.2). *Suponga que $X \sim \mathcal{N}(\mu, \Sigma)$ es un vector aleatorio p-dimensional. Entonces*

$$(X - \mu)'\Sigma^{-1}(X - \mu) \sim \chi^2(p).$$

Muestras Aleatorias

- **Definición.** Sea $X = (X_1, \ldots, X_n)'$ un vector aleatorio n-dimensional. Las variables aleatorias $X_1, \ldots, X_n$ se llaman ***muestra aleatoria*** si es que son **mutuamente independientes** y tienen **distribuciones marginales idénticas**
- En este caso decimos que $X_1, \ldots, X_n$ son ***variables aleatorias independientes e idénticamente distribuidas (i.i.d.)***
- Si $X_1, \ldots, X_n$ es una muestra aleatoria de una distribución con cdf F, la cdf conjunta de $(X_1, \ldots, X_n)'$ es

$$F_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = \prod_{i=1}^{n} F_{X_i}(x_i) = \prod_{i=1}^{n} F(x_i),$$

- Análogamente, la pmf (pdf) conjunta de una muestra aleatoria de una distribución discreta (continua) con pmf (pdf) f es

$$f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = \prod_{i=1}^{n} f_{X_i}(x_i) = \prod_{i=1}^{n} f(x_i).$$

Estadístico

- Sea $X_1, \dots, X_n$ una muestra aleatoria y sea $T : \mathbb{R}^n \to \mathbb{R}^k$ una función (medible)
- El vector aleatorio $Y = T(X_1, \dots, X_n)$ se llama ***estadístico*** y su distribución se llama ***distribución muestral de*** Y
- A este nivel de generalidad, cualquier función de $X_1, \dots, X_n$ es un estadístico
- Estudiaremos aquellos estadísticos que son relevantes cuando estamos en presencia de una muestra aleatoria de una distribución normal

Momentos muestrales

- La ***media muestral*** es el estadístico definido por

$$\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i.$$

- La ***varianza muestral*** es

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2 = \frac{1}{n-1} \left(\sum_{i=1}^{n} X_i^2 - n\overline{X}^2 \right),$$

- Mientras que $S = \sqrt{S^2}$ se llama la ***desviación estándar muestral***

$\Rightarrow$ Una justificación parcial para el uso de $(n-1)$ en la definición es provista en la parte (c) del siguiente teorema, el cual caracteriza algunas propiedades (momentos) elementales de $\overline{X}$ y S^2.

Momentos muestrales y los momentos poblacionales

- **Teorema (Casella y Berger, Teorema 5.2.6).** *Sea $X_1, \ldots, X_n$ una muestra aleatoria de una distribución con media μ y varianza σ^2. Entonces*
  (a) $\mathbb{E}(\overline{X}) = \mu$.
  (b) $Var(\overline{X}) = \sigma^2/n$.
  (c) $\mathbb{E}(S^2) = \sigma^2$.
- **Demostración.**
  (a) $\mathbb{E}(\overline{X}) = \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^n X_i\right) = \frac{1}{n}\mathbb{E}\left(\sum_{i=1}^n X_i\right) = \frac{1}{n}\sum_{i=1}^n \mathbb{E}(X_i) = \frac{1}{n}\sum_{i=1}^n \mu = \mu$.
  (b) $Var(\overline{X}) = Var\left(\frac{1}{n}\sum_{i=1}^n X_i\right) = \frac{1}{n^2}Var\left(\sum_{i=1}^n X_i\right) = \frac{1}{n^2}\sum_{i=1}^n Var(X_i) = \frac{\sigma^2}{n}$,
  (c)

$$\mathbb{E}(S^2) = \mathbb{E}\left(\frac{1}{n-1}\left(\sum_{i=1}^n X_i^2 - n\overline{X}^2\right)\right) = \frac{1}{n-1}\mathbb{E}\left(\sum_{i=1}^n X_i^2 - n\overline{X}^2\right)$$

$$= \frac{1}{n-1}\left(\sum_{i=1}^n \mathbb{E}(x_i^2) - n\mathbb{E}\left(\overline{X}^2\right)\right)$$

$$= \frac{1}{n-1}\left(\sum_{i=1}^n (\sigma^2 + \mu^2) - n\left(\frac{\sigma^2}{n} + \mu^2\right)\right) = \sigma^2.$$

Distribución muestral de los estadísticos

- Si la distribución subyacente es conocida la distribución muestral de los estadísticos pueden ser encontrada
- **Teorema (Casella y Berger, Teorema 5.3.1).** *Sea $X_1, \ldots, X_n$ una muestra aleatoria de una distribución $\mathcal{N}(\mu, \sigma^2)$. Entonces*
  (a) $(\overline{X})$ *y* S^2 *son independientes.*
  (b) $(\overline{X}) \sim \mathcal{N}(\mu, \sigma^2/n)$.
  (c) $(n-1)S^2/\sigma^2 \sim \chi^2(n-1)$.
- Demostraciones: Ver Casella y Berger, Teorema 5.3.1