CAP 1: Distribuciones conjuntas, condicionales e independencia.

DEF. (vector aleatorio) Un vector aleatorio k-dimensional $X = (X_1, .., X_k)^T$ es una función (medible) desde el espacio de probabilidad Ω a $\mathbb{R}^k$, esto es.

$$X : \Omega \longrightarrow \mathbb{R}^k.$$

que asigna un vector $X(\omega)$ a cada resultado ω.

OBS: frecuentemente k variables aleatorias son nuestro interés. Por ejemplo, para un paciente se puede recolectar la información de su edad, presión sanguínea, peso, sexo, nivel de colesterol.

Para introducir algunas ideas consideraremos el caso $k=2$ para posteriormente revisar las definiciones generales $(k>2)$.

Distribuciones bivariadas.

Para un par de variables aleatorias X e Y se tiene la función de densidad conjunta como $f(x,y) = P(X=x, Y=y)$.

Notación: Escribiremos.

$$P(X=x, Y=y) = P(\{X=x\} \cap \{Y=y\}).$$

Ejemplo: Distribución bivariada con X e Y tomando valores 0 o 1.

$X \backslash Y$	0	1	
0	1/9	2/9	1/3
1	3/9	4/9	2/3
	1/3	2/3	1.

En este caso tenemos que

$$f(1,1) = P(X=1, Y=1) = \frac{4}{9}$$

~~DEF~~: Decimos que $f(x,y)$ es la función de densidad del vector aleatorio (X,Y) si

(i) $f(x,y) \geq 0$ para todo (x,y),

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

(iii) Para cualquier conjunto $A \subset \mathbb{R}^2$,

$$P((X,Y) \in A) = \iint_A f(x,y)\,dx\,dy$$

OBS: Para el caso de distribuciones discretas la definición es análoga.

✱ (REVISAR Definición en el Ejemplo anterior).

Ejemplo: Sea (X, Y) con densidad

$$f(x,y) = \begin{cases} x+y, & 0 \le x \le 1,\ 0 \le y \le 1, \\ 0, & \text{en otro caso} \end{cases}$$

Entonces.

$$\int_0^1 \int_0^1 (x+y)\,dx\,dy = \int_0^1 \left[\int_0^1 x\,dx \right] dy$$

$$+ \int_0^1 \left[\int_0^1 y\,dx \right] dy = \int_0^1 \frac{1}{2}\,dy + \int_0^1 y\,dy$$

$$= \frac{1}{2} + \frac{1}{2} = 1.$$

es decir f es una densidad

La distribución conjunta de (X, Y) puede ser completamente caracterizada mediante la función de distribución conjunta (cdf), $F(x,y)$ definida como

$$F(x,y) = P(X \le x, Y \le y).$$

La función de distribución acumulada tiene las siguientes propiedades.

(i) $F(x,y)$ es función monótona creciente y continua a la derecha en (x,y).

(ii) $0 \leq F(x,y) \leq 1$.

(iii) $F(-\infty,y) = F(x,-\infty) = 0$.

(iv) $F(+\infty,+\infty) = 1$.

Aunque la cdf no es muy buena para el caso discreto, pero tiene las siguientes relaciones de interés

$$F(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(s,t)\, dt\, ds.$$

Así, desde el teorema fundamental del cálculo sigue que

$$\frac{\partial^2 F(x,y)}{\partial x\, \partial y} = f(x,y).$$

RESULTADO Sea (X, Y) un vector aleatorio bivariado con densidad conjunta $f_{X,Y}(x,y)$ entonces las densidades marginales de X e Y, $f_X(x) = P(X = x)$ y $f_Y(y) = P(Y = y)$ son respectivamente dadas por

(a) Para el caso discreto:

$$f_X(x) = \sum_y f(x,y) = \sum_y P(X = x, Y = y).$$

$$f_Y(y) = \sum_x f(x,y) = \sum_x P(X = x, Y = y).$$

(b) Para el caso de variables continuas.

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,dy \quad , \quad f_Y(y) = \int_{-\infty}^{\infty} f(x,y)\,dx.$$

DEM: Solo caso (a). Considere $f_X(x)$, entonces para cualquier $x \in \mathbb{R}$, sea

$$S_x = \{(x,y) : -\infty < y < +\infty\}$$

(esto es S_x es la línea en el plano con la coordenada igual a x).

De este modo, para $x \in \mathbb{R}$

$$f_X(x) = P(X = x) = P(X = x, -\infty < Y < \infty)$$

$$\left(\text{pues } P(-\infty < Y < +\infty) = 1\right).$$

$$= P((x, Y) \in S_x) \quad (\text{por la def de } S_x)$$

$$= \sum_{(x,y) \in S_x} f_{X,Y}(x, y)$$

$$= \sum_{y} f_{X,Y}(x, y).$$

Ejemplo: Suponga que

$$f_{X,Y}(x, y) = e^{-(x+y)}, \quad x, y \geq 0.$$

De este modo

$$f_X(x) = e^{-x} \int_0^{\infty} e^{-y} \, dy = e^{-x}$$

Obs: Para variables aleatorias continuas podemos usar la siguiente relación

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

$$= \int_{-\infty}^{x} \int_{-\infty}^{+\infty} f_{X,Y}(t,y)\,dy\,dx$$

$$= F_{X,Y}(x,+\infty)$$

Además

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

y análogamente

$$F_Y(y) = F_{X,Y}(+\infty, y).$$

DISTRIBUCIONES CONDICIONALES E INDEPENDENCIA.

DEF La función de densidad condicional para el caso discreto es definida como

$$f_{X|Y}(x|y) = P(X=x \mid Y=y)$$

$$= \frac{P(X=x, Y=y)}{P(Y=y)}$$

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

si $f_Y(y) > 0$

OBS: A pesar de la notación $f_{X|Y}(x|y)$ se debe destacar que $f_{X|Y}(\cdot|y)$ es función de x.

Para el caso continuo, tenemos.

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}, \quad f_Y(y) > 0$$

De este modo

$$P(X \in A | Y = y) = \int_A f_{X|Y}(x|y)\, dx.$$

<u>Obs</u>: Este cálculo de probabilidades contrasta con el caso discreto donde no es necesario "integrar" (sumar).

Además note que

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

$$= \frac{f_{Y|X}(y|x) f_X(x)}{f_Y(y)}.$$

y

$$f_{X|Y}(x|y) = \frac{f_{Y|X}(y|x)\, f_X(x)}{\int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dx}$$

$$= \frac{f_{Y|X}(y|x)\, f_X(x)}{\int_{-\infty}^{\infty} f_{Y|X}(y|x)\, f_X(x)\, dx}$$

que corresponde al teorema de Bayes para densidades.

OBS: Note que $c = \int_{\mathbb{R}} f_{Y|X}(y|x) f_X(x)\, dx$ corresponde a la constante de normalización de $f_{X|Y}(x|y)$, luego basta calcular

$$f_{X|Y}(x|y) \propto f_{Y|X}(y|x)\, f_X(x)$$

Ejemplo: Considere $Y|X = x \sim N(\mu, \sigma^2/x)$. $X \sim \text{Gama}\left(\frac{\nu}{2}, \frac{\nu}{2}\right)$. Entonces obtenga la distribución de $X|Y$.

previo: Sea $Z \sim \text{Gama}(a, b)$. Entonces.

$$f_Z(z) = \frac{b^a}{\Gamma(a)} z^{a-1} e^{-bz}, \quad a, b > 0, \quad z \geq 0.$$

En esta parametrización de la Gama tenemos que $E(Z) = a/b$.

Note que

$$f_Y(y) = \int_0^\infty f_{X,Y}(x,y)\, dx$$

con

$$f_{X,Y}(x,y) = f_{Y|X}(y|x)\, f_X(x).$$

$$= (2\pi\sigma^2/x)^{-1/2} \exp\left\{-\frac{x}{2}\left(\frac{y-\mu}{\sigma}\right)^2\right\}$$

$$\times \frac{(\nu/2)^{\nu/2}}{\Gamma(\nu/2)} x^{\frac{\nu}{2}-1} \exp\left(-\frac{\nu}{2}x\right).$$

Ahora

$$f_{X|Y}(x|y) \propto x^{1/2} \cdot \exp\left(-\frac{x}{2} y^2\right)$$

$$\cdot\ x^{\nu/2-1} \exp\left(-\frac{x}{2}\nu\right).$$

$$= x^{\frac{\nu+1}{2}-1} \exp\left\{-\frac{1}{2}(y^2+\nu)\,x\right\}.$$

que corresponde al "kernel" de una densidad Gamma $\left(\frac{\nu+1}{2}, \frac{\nu+y^2}{2}\right)\left(\stackrel{d}{=} X|Y\right)$.

OBS: En este caso no fue necesario calcular la "desagradable" constante $f_Y(y)$ en la densidad $f_{X|Y}(x|y)$.

(Además $X|Y$ es del mismo tipo que X esto ocurre cuando las familias son conjugadas)

En efecto, la densidad condicional debe satisfacer las condiciones de una densidad dadas anteriormente.

$$\int_{-\infty}^{\infty} f_{Y|X}(y|x)\,dy = \int_{-\infty}^{\infty} \frac{f_{X,Y}(x,y)}{f_X(x)}\,dy$$

$$= \frac{1}{f_X(x)} \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy = \frac{f_X(x)}{f_X(x)} = 1.$$

Además es obvio que $f_{Y|X}(y|x) \geq 0$.

Los momentos de una v.a. condicional se definen de forma análoga

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

$$E(Y|x) = \int_{-\infty}^{\infty} y\, f_{Y|X}(y|x)\,dy.$$

Análogamente

$$Var(Y|X) = E(Y^2|X) - (E(Y|X))^2$$

Ejemplo: Considere el vector aleatorio (X, Y) con densidad conjunta

$$f(x,y) = e^{-y} \quad , \quad 0 < x < y < \infty$$

Para obtener $Var(Y|X)$ primeramente se debe calcular la densidad condicional $f(y|x) = f(x,y)/f(x)$. Además tenemos que

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\, dy \quad , \quad y > x$$

$$= \int_x^{\infty} e^{-y}\, dy = e^{-x}$$

(Si $x \leq 0$ tenemos que $f(x,y) = 0$ para cualquier y, de ahí que $f_X(x) = 0$)

De este modo, para cualquier $x > 0$ (pues $f_X(x) > 0$) Tenemos

$$f(y|x) = \frac{f(x,y)}{f_X(x)} = \frac{e^{-y}}{e^{-x}} = e^{-(y-x)}, \quad y > x$$

(mientras que $y \leq x \Rightarrow f(y|x) = f(x,y)/f_X(x) = 0/e^{-x} = 0$) De este modo

$$E(Y|X=x) = \int_x^{\infty} y\, e^{-(y-x)}\, dy = 1+x$$

y de ahí que

$$Var(Y|x) = \int_x^{\infty} y^2 e^{-(y-x)}\, dy - \left(\int_x^{\infty} y\, e^{-(y-x)}\, dy\right)^2$$

$$= 1.$$

~~DEF~~: Sea (X,Y) un vector aleatorio con densidad conjunta $f(x,y)$ y densidades marginales $f_X(x)$, $f_Y(y)$, respectivamente. Entonces X e Y se dicen independientes si, para todo $x,y \in \mathbb{R}$.

$$f(x,y) = f_X(x) f_Y(y).$$

OBS: Analogamente, si $F(x,y)$ representa la la función de distribución acumulada del vector (X,Y) y $G(x)$, $H(y)$ corresponden a las cdf de X e Y, respectivamente. Entonces X e Y son independientes si

$$F(x,y) = G(x) H(y).$$

* NOTACIÓN: En ocasiones indicaremos que X e Y son independientes como $X \perp Y$.

OBS: Note que, si X e Y son independientes, entonces la densidad condicional de Y dado $X = x$ es dada por

$$f(y|x) = \frac{f(x,y)}{f_X(x)} = \frac{f_X(x) f_Y(y)}{f_X(x)} = f_Y(y).$$

RESULTADO Sean X, Y variables aleatorias independientes. Entonces.

(a) $P(X \in A, Y \in B) = P(X \in A) P(Y \in B)$
(es decir, los eventos $\{X \in A\}$ y $\{Y \in B\}$ son independientes)

(b) Sea $g(x)$ una función solo de x y $h(y)$ una función solo de y. Entonces

$$E\{g(X) h(Y)\} = E\{g(X)\} E\{h(Y)\}.$$

dem: Solo parte (b). Tenemos que

$$E\{g(x)\,h(y)\} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x)\,h(y)\,f(x,y)\,dx\,dy$$

$$\overset{\text{ind}}{=} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x)\,h(y)\,f_X(x)\,f_Y(y)\,dx\,dy$$

$$= \int_{-\infty}^{\infty} g(x)\,f_X(x)\,dx \int_{-\infty}^{\infty} h(y)\,f_Y(y)\,dy$$

$$= E\{g(x)\}\,E\{h(y)\}.$$

TAREA: probar parte (a). Sugerencia: Usar $g(\cdot)$, $h(\cdot)$ como funciones indicadoras.

EXTENSIONES AL CASO MULTIVARIADO ($k > 2$)

Sea $X = (X_1, .., X_k)^T$ donde $X_1, .., X_k$ son variables aleatorias, diremos que X es vector de parámetros k-dimensional.

Usaremos la notación $f(x)$, $F(x)$ para indicar las funciones de densidad y distribución del vector aleatorio X. En este caso

$$F(x) = P(X \leq x) \left(= P(X_1 \leq x_1, .., X_k \leq x_k) \right)$$

que corresponde a la probabilidad del evento $\bigcap_{i=1}^{k} \{X_i \leq x_i\}$.

Por otro lado, asumiremos que existe f no negativa tal que

$$F(x) = \int_{-\infty}^{x} f(u)\, du \quad , \quad x \in \mathbb{R}^k .$$

Además (por el Teorema fundamental del cálculo) tenemos que la función de densidad f es tal que

$$f(x) = \frac{\partial^k F(x)}{\partial x_1 \cdots \partial x_n}$$

con $\int_{\mathbb{R}^k} f(x)\, dx = 1$.

Considere la siguiente partición $x = (x_1^T, x_2^T)^T$ donde x_1 y x_2 son vectores $k_1 \times 1$ y $k_2 \times 1$ ($k = k_1 + k_2$). Tenemos que $x_i \sim F_i$ para $i = 1, 2$. es decir x_1 y x_2 son marginales del vector x. Además

$$F_1(s) = F(s, +\infty), \quad F_2(t) = F(+\infty, t)$$

para $s \in \mathbb{R}^{k_1}$, $t \in \mathbb{R}^{k_2}$ o alternativamente

$$f_1(s) = \int_{\mathbb{R}^{k_2}} f(s, u)\, du, \quad f_2(t) = \int_{\mathbb{R}^{k_2}} f(u, t)\, du$$

Si $f_1(x) > 0$ entonces la densidad condicional de X_2 dado $X_1 = x_1$ es dada por

$$f_{X_2|X_1=x_1}(u) = \frac{f_X(x_1, u)}{f_1(x_1)}$$

$$= \frac{f_X(x_1, u)}{\int_{\mathbb{R}^{k_2}} f_X(x_1, t)\, dt}.$$

DEF: Si $X_1, .., X_n$ son independientes y cada una tiene la misma distribución marginal con CDF F, decimos que $X_1, .., X_n$ son IID (independientes e identicamente distribuidas) y escribimos

$$X_1, .., X_n \sim F.$$

Si F tiene densidad f podemos escribir $X_1, .., X_n \sim f$.

OBS. Cuando $x_1, .., x_n \overset{iid}{\sim} f$ también decimos que $x_1, .., x_n$ es una muestra aleatoria de tamaño n desde f (o f).

Además, si $x_1, .., x_n$ son independientes, es fácil notar que

$$f(x) = \prod_{i=1}^{n} f_{x_i}(x_i).$$

CASOS IMPORTANTES:

1: **Distribución multinomial.** Sea $p = (p_1, .., p_k)^T$ donde $p_j \geq 0$ y $\sum_{j=1}^{k} p_j = 1$. donde p_j representa la probabilidad de éxito asociada a la k-ésima categoría. Sea $x = (x_1, .. x_k)^T$ donde x_j representa el número de veces que se produce el j-ésimo suceso. y $n = \sum_{j=1}^{k} x_k$

Decimos que $x \sim \text{Multinomial}(n, p)$.
con función de probabilidad

$$f(x) = \binom{n}{x_1 \dots x_k} p_1^{x_1} \dots p_k^{x_k}$$

donde

$$\binom{n}{x_1 \dots x_k} = \frac{n!}{x_1! \dots x_k!}$$

<u>RESULTADO</u>: Sea $x \sim \text{Multinomial}(n, p)$ donde $x = (x_1, \dots, x_k)^T$, $p = (p_1, \dots, p_k)^T$. Entonces la distribución marginal de $x_j \sim \text{Binomial}(n, p_j)$

2. DISTRIBUCIÓN NORMAL MULTIVARIADA

Considere $X = (x_1, \dots, x_k)^T$, tal como el caso univariado la distribución normal multivariada es caracterizada por un vector de medias $\mu = (\mu_1, \dots, \mu_k)^T$ y una matriz de covarianzas $\Sigma = (\sigma_{ij}) > 0$ y escribimos $X \sim N_k(\mu, \Sigma)$.

Para introducir ideas considere $z_1, \dots, z_k$ IID con distr. $N_1(0, 1)$ y suponga $z = (z_1, \dots, z_k)^T$. Entonces

$$f_z(z) = \prod_{i=1}^{k} f_{z_i}(z_i)$$

$$= \prod_{i=1}^{k} (2\pi)^{-1/2} \exp\left\{-\frac{1}{2} z_i^2\right\}$$

$$= (2\pi)^{-k/2} \exp\left\{-\frac{1}{2} \sum_{i=1}^{k} z_i^2\right\}$$

$$= (2\pi)^{-k/2} \exp\left(-\frac{1}{2} z^T z\right)$$

y anotamos $z \sim N_k(0, \mathbb{I}_k)$

más generalmente, $x \sim N_k(\mu, \Sigma)$ tiene densidad

$$f_x(x) = |2\pi \Sigma|^{-1/2} \exp\left\{-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)\right\}$$

Dado que Σ es simétrica y definida positiva tenemos que $\Sigma = BB^T$. esto lleva al resultado siguiente

RESULTADO si $z \sim N_k(0, I)$ y $x = \mu + Bz$ entonces $x \sim N_k(\mu, \Sigma)$. Asimismo si $x \sim N_k(\mu, \Sigma)$ entonces $B^{-1}(x-\mu) \sim N_k(0, I)$.

RESULTADO. Sea $x \sim N_k(\mu, \Sigma)$ y considere la partición

$$x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \quad \mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

Sala Ayudantía.

Entonces

(a) La distribución marginal de $x_i \sim N_{k_i}(\mu_i, \Sigma_{ii})$ para $i=1,2$.

(b) La distribución condicional de $x_2 | x_1 = x_1$ es

$$x_2 | x_1 = x_1 \sim N_{k_2}\left(\mu_2 + \Sigma_{21}\Sigma_{11}^{-1}(x_1 - \mu_1),\ \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\right)$$

(c) $U = (x-\mu)^T \Sigma^{-1} (x-\mu) \sim \chi^2_k$.

Obs: En efecto, recuerde que $z = (z_1, \ldots, z_k)^T \sim N_k(0, I)$ (es decir $z_1, \ldots, z_k$ son IID)

$$U = z^T z = \sum_{i=1}^{k} z_i^2 \sim \chi^2_k.$$