Probabilidades y Estadística (C)

Clase Práctica 11

Matías Data

Departamento de Matemática, FCEyN, Universidad de Buenos Aires

06/10/2020

Introducción

- Hoy veremos (algo de) los siguientes conceptos: vectores aleatorios continuos, conjunta y marginal, función de distribución de un vector aleatorio, distribución condicional: caso discreto y continuo, independencia (de nuevo).
- Recordemos brevemente las definiciones y resultados de la teórica que vamos a utilizar.

Definición

Un vector aleatorio (X, Y) se dice continuo si existe una función $f : \mathbb{R}^2 \to \mathbb{R}_{\geq 0}$ (llamada *densidad conjunta*) tal que:

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

En particular:

$$P(a \leq X \leq b,\, c \leq Y \leq d) = \int_c^d \int_a^b f(x,y)\,dx\,dy$$

Introducción

- Si f es la densidad conjunta de un vector aleatorio (X, Y) entonces:

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

- Si queremos calcular la *distribución marginal* de X (resp. Y) entonces debemos integrar en la variable y (resp. x) en toda la recta, esto es:

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

Introducción

Definición

Dado (X, Y) un vector aleatorio discreto, entonces se define la *función de probabilidad condicional* de Y dado $X = x$ (que la notamos $Y\,|_{X=x}$) por:

$$p_{Y\,|\,X=x}(y) = \frac{p_{XY}(x, y)}{p_X(x)} = P(Y = y\,|\,X = x)$$

si $p_X(x) > 0$.

- Esta función define la probabilidad puntual de una variable aleatoria, toma valores no negativos y suma 1.
- Vale la regla de Bayes:

$$p_{Y\,|\,X=x}(y)\,p_X(x) = p_{XY}(x, y) = p_{X\,|\,Y=y}(x)\,p_Y(y)$$

Introducción

Definición

Dado (X, Y) un vector aleatorio continuo, entonces se define la *función de probabilidad condicional* de Y dado $X = x$ (que la notamos $Y\,|_{X=x}$) por:

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

si $f_X(x) > 0$.

- Esta función define una densidad, toma valores no negativos y integra 1.
- Vale la regla de Bayes:

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

Introducción

Definición

Dado (X,Y) un vector aleatorio, decimos que X e Y son independientes si

$$P(a \leq X \leq b,\, c \leq Y \leq d) = P(a \leq X \leq b)\, P(c \leq Y \leq d)$$

para todos a, b, c, d.

- (*caso discreto*) Si (X,Y) es discreto, entonces esto ocurre si y solo si

$$p_{XY}(x,y) = p_X(x)\, p_Y(y)$$

- (*caso continuo*) Si (X,Y) es continuo, si

$$f_{XY}(x,y) = f_X(x) f_Y(y)$$

entonces X e Y son independientes. Si

$$f_{XY}(x,y) \neq f_X(x) f_Y(y)$$

en un punto de continuidad de f_{XY}, f_X y f_Y entonces X e Y no son independientes.

Ejercicio 1

Ejercicio

Sea (X, Y) un vector aleatorio cuya densidad esta dada por:

$$f(x,y) = \begin{cases} \frac{1}{x^3} & \text{si } \; 0 \leq y \leq x, \;\; 1 \leq x \\ 0 & \text{c.c} \end{cases}$$

A) Hallar f_X.
B) Probar que $Y\,|_{X=x} \sim U[0,x]$.
C) Hallar f_Y.
D) Son X e Y independientes?

A) Podemos escribir la función de densidad conjunta utilizando indicadoras como

$$f(x,y) = \frac{1}{x^3}\, \mathbb{1}_{[0,x]}(y)\, \mathbb{1}_{[1,+\infty)}(x)$$

Ejercicio 1

- Queremos calcular la densidad marginal f_X, para esto debemos integrar con respecto a y a la densidad conjunta.
- Pero por como escribimos la indicadora esta integral es fácil, todo lo que es función solo de x sale para afuera de la integral, y la indicadora nos dice donde integrar:

$$
\begin{aligned}
f_X(x) = \int_{-\infty}^{+\infty} f(x,y)\,dy &= \int_{-\infty}^{+\infty} \frac{1}{x^3}\,\mathbb{1}_{[0,x]}(y)\,\mathbb{1}_{[1,+\infty)}(x)\,dy \\
&= \frac{1}{x^3}\mathbb{1}_{[1,+\infty)}(x)\left(\int_{-\infty}^{+\infty} \mathbb{1}_{[0,x]}(y)\,dy\right) \\
&= \frac{1}{x^3}\mathbb{1}_{[1,+\infty)}(x)\left(\int_{0}^{x} 1\,dy\right) = \frac{1}{x^2}\mathbb{1}_{[1,+\infty)}(x)
\end{aligned}
$$

- Luego $f_X(x) = \frac{1}{x^2}\mathbb{1}_{[1,+\infty)}(x)$.

Ejercicio 1

b) Nos piden probar que la distribución condicional de Y dado $X = x$ es uniforme en el intervalo $[0, x]$.

- Para esto calculamos:

$$f_{Y\,|\,X=x}(y) = \frac{f(x,y)}{f_X(x)} = \frac{\frac{1}{x^3}\,\mathbb{1}_{[0,x]}(y)\,\mathbb{1}_{[1,+\infty)}(x)}{\frac{1}{x^2}\,\mathbb{1}_{[1,+\infty)}(x)}$$

$$= \frac{1}{x}\mathbb{1}_{[0,x]}(y)$$

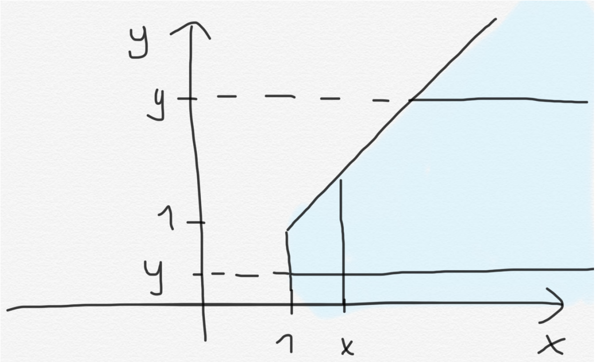
- Luego $f_{Y\,|\,X=x}(y) = \frac{1}{x}\mathbb{1}_{[0,x]}(y)$, es decir, $Y\,|_{X=x} \sim U[0,x]$.
- Aplicación: si queremos simular el vector aleatorio (X, Y) podemos primero simular la variable aleatoria X con una uniforme y la inversa generalizada, y luego una vez que obtenemos la simulación de $X = x$, simular una uniforme en $[0, x]$.

Ejercicio 1

c) Queremos calcular la densidad marginal f_Y.

- Para esto tenemos que integrar respecto a x, pero esto no es fácil con las indicadoras que usamos.
- Vamos a dibujar la región para entender donde debemos integrar. La región es el conjunto

$$\{(x, y) \mid 0 \leq y \leq x,\ x \geq 1\}$$



Ejercicio 1

- Como vemos en el dibujo hay dos casos, si $0 \leq y \leq 1$ entonces tenemos que integrar x entre 1 y $+\infty$. Mientras que si $y \geq 1$ entonces tenemos que integrar x entre y y $+\infty$.
- Calculamos en el caso $y \geq 1$:

$$f_Y(y) = \int_{-\infty}^{+\infty} f(x,y)\,dx = \int_y^{+\infty} \frac{1}{x^3}\,dx$$
$$= -\frac{x^{-2}}{2}\bigg|_y^{+\infty} = \frac{1}{2y^2}$$

- En caso de que $0 \leq y \leq 1$ hay que integrar desde 1 a $+\infty$ y luego $f_Y(y) = 1/2$.
- Por lo tanto tenemos:

$$f_Y(y) = \frac{1}{2}\mathbb{1}_{[0,1]}(y) + \frac{1}{2y^2}\mathbb{1}_{[1,+\infty)}(y)$$

Ejercicio 1

d) Para ver que (X, Y) no son independientes basta ver que

$$f_{XY}(x, y) \neq f_X(x) f_Y(y)$$

en un punto (x, y) de continuidad de f_{XY}, f_X y f_Y.

- Tomemos $x = 2$ e $y = 1.5$.

$$f_{XY}(2, 1.5) = \frac{1}{8} \neq \frac{1}{4} \frac{1}{2(3/2)^2} = f_X(2) f_Y(1.5)$$

y dado que es un punto de continuidad de las funciones, esto muestra que no son independientes.

Ejercicio 2

Ejercicio

El número de personas que concurren en un día a una oficina de correo, este es una v.a. Poisson de parámetro λ. Además se sabe que una persona que concurre al correo será hombre con probabilidad p y mujer con probabilidad $1 - p$ de forma independiente al resto. Si se definen:

$$X = \text{'número hombres que ingresan'}$$

$$Y = \text{'número mujeres que ingresan'}$$

a) Observar que $X\,|_{X+Y=n} \sim \mathrm{Bi}(n, p)$.

b) Cuál es la distribución conjunta de X e Y?

c) Probar que X e Y son Poisson independientes de parámetros λp y $\lambda(1-p)$ respectivamente.

Ejercicio 2

a) Queremos calcular la distribución de X condicional a que $X + Y = n$, es decir, condicional a que entraron n personas al correo.

- Pero si la cantidad de personas que entraron al correo está fija (es conocida), entonces cada una de ellas es hombre con probabilidad p y mujer con probabilidad $(1 - p)$, y esto de forma independiente para cada persona.
- Luego:

$$P(X = k \,|\, X + Y = n) = \binom{n}{k} p^k (1-p)^{n-k} \sim \mathrm{Bi}(n, p)$$

Ejercicio 2

b) Ahora queremos calcular la distribución conjunta de X e Y. Para esto vamos a condicionar en $X+Y$.

- Calculamos

$$
\begin{aligned}
p_{XY}(k,l) &= P(X=k,\ Y=l) \\
&= \sum_{n=0}^{+\infty} P(X=k,\ Y=l \,|\, X+Y=n)\, P(X+Y=n) \\
&= P(X=k,\ Y=l \,|\, X+Y=k+l)\, P(X+Y=k+l) \\
&= \underbrace{P(X=k \,|\, X+Y=k+l)}_{\sim \mathrm{Bi}(k+l,p)} \underbrace{P(X+Y=k+l)}_{\sim P(\lambda)} \\
&= \binom{k+l}{k} p^k (1-p)^l e^{-\lambda} \frac{\lambda^{k+l}}{(k+l)!}
\end{aligned}
$$

- Listo, tenemos la función de probabilidad puntual conjunta.

Ejercicio 2

c) Queremos ver que X e Y son Poisson independientes de parámetros λp y $\lambda(1-p)$ respectivamente.

- Dado que queremos ver que son independientes, la función de probabilidad conjunta se tiene que factorizar como el producto de las marginales.
- Calculamos:

$$p_{XY}(k,l) = \binom{k+l}{k} p^k (1-p)^l e^{-\lambda} \frac{\lambda^{k+l}}{(k+l)!}$$
$$= \frac{(k+l)!}{k!\,l!} p^k (1-p)^l e^{-\lambda(p+(1-p))} \frac{\lambda^{k+l}}{(k+l)!}$$
$$= \left(e^{-\lambda p} \frac{(\lambda p)^k}{k!} \right) \left(e^{-\lambda(1-p)} \frac{(\lambda(1-p))^l}{l!} \right) = p_X(k)\, p_Y(l)$$

- En la última igualdad usamos que estas son las fórmulas para la probabilidad puntual de v.a. Poisson, por lo tanto suman uno y luego son las marginales.
- Luego tenemos que $X \sim P(\lambda p)$ e $Y \sim P(\lambda(1-p))$ independientes.

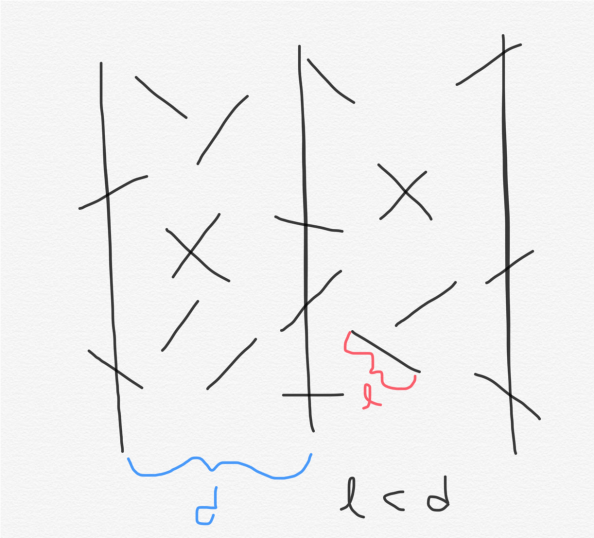
Ejercicio 3

Ejercicio

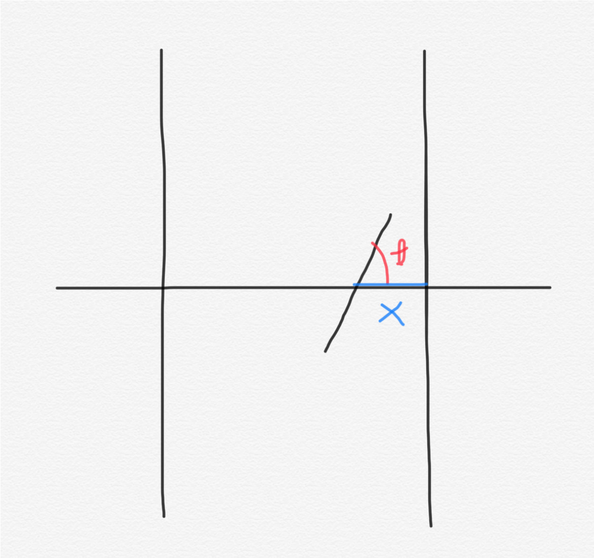
(La aguja de Buffon) Dada una aguja de longitud l, se la arroja al azar en el plano, donde hay dibujadas rectas paralelas a distancia d una de otra, con $l < d$. Cuál es la probabilidad de que la aguja interseque a una recta?

- Para resolver el problema de la aguja de Buffon vamos a definir un vector aleatorio apropiado.
- Podemos pensar que las rectas en el plano son verticales y luego una variable que nos va a importar es la posición en el eje x en el que cae el centro de la aguja.
- No nos importa a que altura del eje y cae, ya que haciendo una traslación vertical no cambian las intersecciones con las rectas.

Ejercicio 3



Ejercicio 3



Ejercicio 3

- Tampoco nos importa entre qué rectas cae, ya que es igual si cae entre la primera y la segunda, o entre la cuarta y la quinta recta.
- Con lo cual podemos pensar que nuestra primera variable de interés es $X \sim U[0, d/2]$, y nos dice la distancia en el eje x en la que cae el centro de la aguja respecto a la recta vertical más cercana.
- La otra variable de interés va a ser el ángulo respecto del eje x con el que cae la aguja.
- Podemos pensar que el ángulo es uniforme entre $-\pi/2$ y $\pi/2$, es independiente de la distancia X, y lo notamos $\theta \sim U[-\pi/2, \pi/2]$.
- Ahora ya podemos formular la pregunta en términos de las variables.

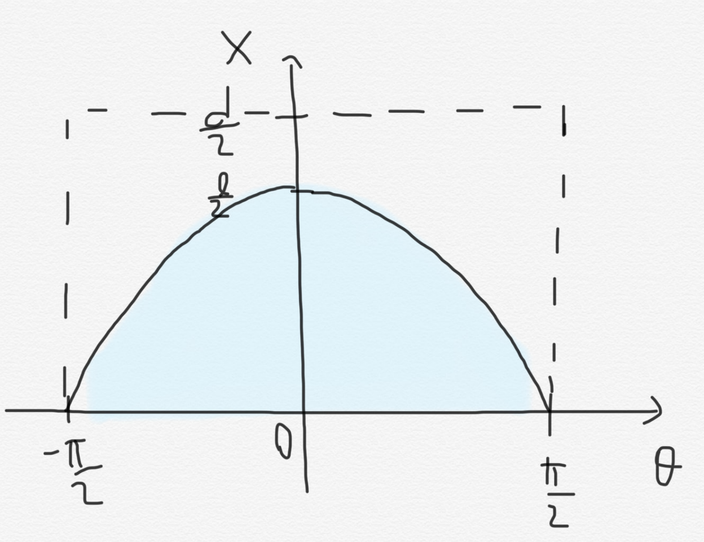
Ejercicio 3

- Para que la aguja interseque a una recta es necesario que sea la más cercana del centro (ya que $l < d$).
- La intersección ocurre si y solo si la distancia a la recta más cercana X es menor a la proyección de la aguja sobre la recta x, esto es $\frac{l}{2}\cos(\theta)$.
- Luego debemos calcular $P(X \leq \frac{l}{2}\cos(\theta))$.
- Para esto debemos integrar a la densidad conjunta en la región:

$$R = \{(\theta, x) \,|\, x \leq \frac{l}{2}\cos(\theta)\}$$

- Graficamos esta región para entenderla y poder calcular la integral.

Ejercicio 3



Ejercicio 3

- Recordamos que $X \sim U[0, d/2]$ y $\theta \sim U[-\pi/2, \pi/2]$ independientes, luego su densidad es el producto, con lo cual la integral es:

$$\iint\limits_R f(x,\theta)dx\,d\theta = \iint\limits_R \frac{1}{d/2}\mathbb{1}_{[0,d/2]}(x)\,\frac{1}{\pi}\mathbb{1}_{[-\pi/2,\pi/2]}(\theta)\,dx\,d\theta$$

$$= \int_{-\pi/2}^{\pi/2}\int_0^{\frac{l}{2}\cos(\theta)} \frac{2}{d\pi}\,dx\,d\theta = \int_{-\pi/2}^{\pi/2}\frac{l}{d\pi}\cos(\theta)\,d\theta$$

$$= \frac{l}{d\pi}\sin(\theta)\Big|_{-\pi/2}^{\pi/2} = \frac{2l}{d\pi}$$

- En la tercera igualdad estamos usando que como $l < d$, si $x < \frac{l}{2}\cos(\theta)$ entonces $x < d/2$.
- Luego la probabilidad de que la aguja interseque a una recta es $\frac{2l}{d\pi}$.
- Podemos usar este resultado para aproximar π, haciendo simulaciones de la aguja en R.