Truncamiento de una v.a.

- Partimos de variable aleatoria X que puede ser de cualquier tipo (continua, discreta o mixta).
- Supongamos que sabemos que el evento {X ∈ B} ha ocurrido, para cierto B tal que P(X ∈ B) > 0.
  ¿Podemos mejorar nuestro conocimiento de X sabiendo que los valores de X estan´ restringidos a B? Si! Considerando la distribución condicional de X al evento $\{X \in B\}$

Definición variable truncada:

Truncar una variable X a un conjunto B significa restringirla a tomar valores sobre B. Para a caracterizarla, debemos hallar la distribución condicional de X en B. Es la variable X sabiendo que X esta en un conjunto B.

- Vamos a denotar como X|X∈ B a la v.a. X truncada a los valores de B
- Para hallar la distribución de la variable truncada, sin asumir nada sobre el tipo ´ de VA (si es continua o discreta), buscamos su funcion de distribución

$$F_{X|X\in B}(x) = \mathbb{P}(X \leq x | X \in B) = \frac{\mathbb{P}(\{X \leq x\} \cap \{X \in B\})}{\mathbb{P}(X \in B)}$$

- Considerando el caso X continua o discreta podemos llegar a expresiones mas´ simples.

Variable truncada: caso continuo

- Supongamos que X es continua, con densidad f_X
- Al truncarla obtendremos otra variable continua, es decir, existira una densidad ´ condicional $f_{(X|X \in B)}$
- Para obtenerla, derivamos la funcion de distribución condicional:

$$F_{X|X\in B}(x) = \frac{\mathbb{P}(\{X \leq x\} \cap \{X \in B\})}{\mathbb{P}(X \in B)},$$

  Para obtener:

$$f_{X|X\in B}(x) = \begin{cases} \frac{f_X(x)}{\mathbb{P}(X\in B)} & \text{si } x \in B \\ 0 & \text{si } x \notin B \end{cases}.$$

- Con la notación de indicadora quedaría:

$$f_{X|X\in B}(x) = \frac{f_X(x)}{\mathbb{P}(X \in B)} \mathbf{1}_B(x).$$

Es decir que truncar una VA continua a B equivale a hacer 0 la densidad fuera de B y dividir la densidad por P(X ∈B) (de modo que siga integrando a 1).

Variable truncada: caso discreto

- Supongamos que X es discreta, con funcion de probabilidad p_X

- Al truncarla obtendremos otra variable discreta, es decir, existira una funci ´ on de ´ probabilidad condicional $p(X|X \in B)$
- La funcion de probabilidad condicional sería entonces:

$$p_{X|X\in B}(x) = \frac{\mathbb{P}(\{X = x\} \cap \{X \in B\})}{\mathbb{P}(X \in B)}$$
$$= \begin{cases} \frac{\mathbb{P}(X=x)}{\mathbb{P}(X\in B)} & \text{si } x \in B \\ 0 & \text{si } x \notin B \end{cases}.$$

- Con la notación de indicadora quedaría:

$$p_{X|X\in B}(x) = \frac{\mathbb{P}(X = x)}{\mathbb{P}(X \in B)} \mathbf{1}_B(x).$$

Es decir que truncar una VA discreta a B equivale a hacer 0 a p_X fuera de B y dividir a p_X por P(X ∈ B) (de modo que siga sumando a 1).

EJEMPLO:

Enunciado

El diametro X (en mm) de las arandelas que fabrica cierta máquina tiene función de densidad:

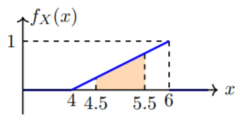
$$f_X(x) = \left(\frac{x}{2} - 2\right) \mathbf{1}_{(4,6)}(x).$$

Las arandelas son aceptables sólo si su diametro está entre 4.5 y 5.5mm. Las arandelas que no son aceptables son descartadas.

1. Halle la probabilidad de que una arandela sea aceptable.
2. Halle la densidad de las arandelas que no son descartadas (aceptables).
3. Halle la densidad de las arandelas que son descartadas.

1) Halle la probabilidad de que una arandela sea aceptable
Debemos hallar:

$$\mathbb{P}(4{,}5 < X < 5{,}5) = \int_{4,5}^{5,5} f_X(x)dx = \frac{1}{2}\left(\frac{1}{4} + \frac{3}{4}\right) = \frac{1}{2}.$$

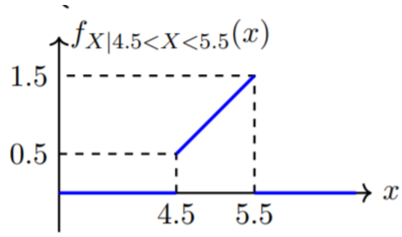


Como es una variable continua no me importa si la X es meno o igual o solo meno a los valores, porque los extremos tienen probabilidad 0

2) Halle la densidad de las arandelas que no son descartadas (aceptables) (aca aparece la idea de truncar)
Aplicamos la fórmula de la densidad truncada:

$$f_{X|X\in(4{,}5,\ 5{,}5)}(x) = \begin{cases} \frac{\left(\frac{x}{2}-2\right)}{\mathbb{P}(4{,}5<X<5{,}5)} = x - 4 & 4{,}5 < x < 5{,}5 \\ 0 & \text{en otro caso} \end{cases}.$$

Tuve que reescalar la densidad, porque si la integraba directamente sin reescalarla, esto no integraba más a 1, osea que no era más una densidad. Al dividir por la probabilidad de que sea aceptable, estoy renormalizando la altura. El grafico tiene la misma forma que antes, pero ahora la amplitud aumento para que el área debajo de la curva siga siendo .

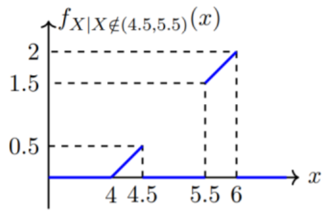


3) Halle la densidad de las arandelas que son descartadas.
Ahora debemos condicionar a que las arandelas no son aceptables:

$$\mathbb{P}(\text{No aceptable}) = \mathbb{P}(X \notin (4{,}5,\ 5{,}5)) = \frac{1}{2}.$$

Aplicamos la fórmula de la densidad truncada:

$$f_{X|X\notin(4{,}5,\ 5{,}5)}(x) = \begin{cases} \frac{\left(\frac{x}{2}-2\right)}{\mathbb{P}(X\notin(4{,}5,\ 5{,}5))} = x-4 & 4 < x < 4{,}5 \text{ o } 5{,}5 < x < 6 \\ 0 & \text{en otro caso} \end{cases}.$$



¡OJO! Dividir la densidad en los dos intervalos (4, 4,5) y (5,5, 6) por la probabilidad de estar su unión $P(X \notin (4.5,5.5) = P(4 < X < 4.5) + P(5.5 < X < 6)$

Mezcla de variables

EJEMPLO INTRODUCTORIO DE MEZCLA

Enunciado

Se dispone de dos cajas con lamparitas:

- Caja 1: las lamparitas tienen una duración (en miles de horas) X_1 con distribución exponencial de media $\mu_1 = 5$.
- Caja 2: las lamparitas tienen una duración (en miles de horas) $X_2 \sim \mathcal{U}(3,6)$.

Se realiza un experimento en dos pasos que consiste en lo siguiente:

- Paso 1: Se elije una de las dos cajas, la caja 1 con probabilidad p, y la caja 2 con probabilidad $1-p$.
- Paso 2: Se extrae una lamparita de la caja elegida y se mide su duración.

Sea Y la duración de la lámpara extraída. ¿Cuál es la distribución de Y?

Veamos el cálculo de la distribución de Y

- Buscamos calcular primero la funcion de distribución de Y:

$$F_Y(y) = \mathbb{P}(Y \leq y)$$

Para poder avanzar, condicionamos al resultado del primer paso del experimento. Para ello definimos:

$$C_1 = \{\text{Se eligio la caja } 1\},$$

y calculamos:

$$F_Y(y) = \mathbb{P}(Y \leq y | C_1)\mathbb{P}(C_1) + \mathbb{P}(Y \leq y | C_1^c)\mathbb{P}(C_1^c).$$

- Luego sabemos que si se elije la caja 1, la duración es X_1, y lo mismo para la 2, es decir:

$$\mathbb{P}(Y \leq y | C_1) = \mathbb{P}(X_1 \leq y) \qquad \mathbb{P}(Y \leq y | C_1^c) = \mathbb{P}(X_2 \leq y).$$

- Reemplazando tenemos:

$$\begin{aligned} F_Y(y) &= \mathbb{P}(X_1 \leq y)\mathbb{P}(C_1) + \mathbb{P}(X_2 \leq y)\mathbb{P}(C_1^c) \\ &= F_{X_1}(y)\mathbb{P}(C_1) + F_{X_2}(y)\mathbb{P}(C_1^c). \end{aligned}$$

  Es decir que la funcion de distribución de Y es el promedio ponderado de las distribuciones de cada caja.
- Finalmente, derivamos para obtener la densidad de Y:

$$f_Y(y) = f_{X_1}P(C_1) + f_{X_2}P(C_1^C)$$

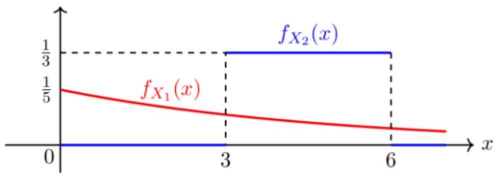
  La función de densidad de Y es el promedio ponderado de las funciones de densidad de las dos cajas.

Observaciones:

- Una vez que uno elige la caja 1, sabe que la duración tendrá distribución´X_1, y si se elije la caja 2, se sabe que será X_2.
- Pero al hablar de Y sin conocer la caja conviven de algún modo ambas posibilidades, de modo que la duración no será ni la de la caja 1 ni la de la caja 2, sino que ponderara ambos universos.

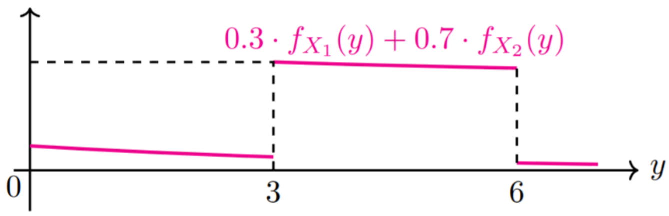
Vamos a ver gráficamente cómo se ve la densidad.

- Supongamos que p = P(C_1) = 0,3.
- En el siguiente grafico vemos la densidad de la duración de cada caja:



- La densidad de Y, la duración de una lámpara elegida con el experimento es el promedio ponderado de ambas:

$$f_Y(y) = 0.3 \cdot f_{X_1}(y) + 0.7 \cdot f_{X_2}(y)$$



Mezcla caso general

- Se tienen n experimentos aleatorios, cada uno asociado a una V.A. $X_1, \dots, X_n$
- Se realiza un experimento aleatorio en dos etapas:
  1) Se elije alguno de los n experimentos anteriores, con

$$p_i = P(elehir\ el\ experimetno\ i)$$

$$\sum_{i=1}^{n} p_i = 1$$

2) Una vez elegido el experimento de entre los n posibles, se hace un ensayo, y se observa un resultado.

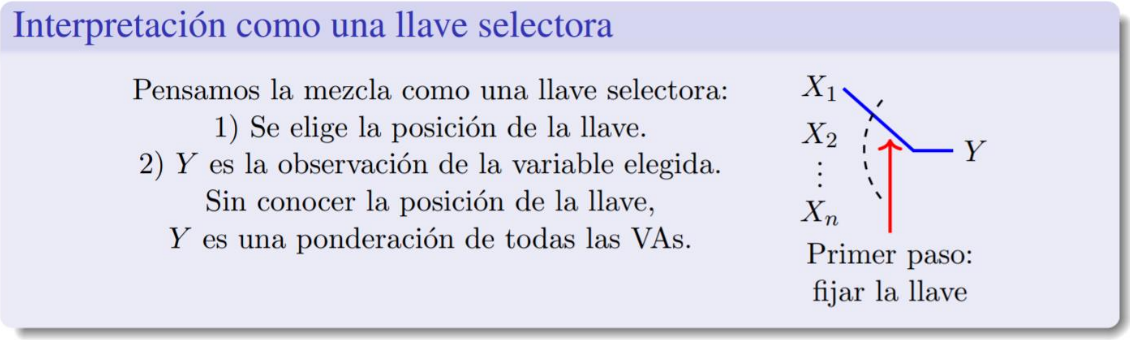
El resultado numérico observado al final del segundo paso le asociamos una variable aleatoria Y. A dicha variable se la llama la ***variable aleatoria mezcla*** de $X_1, \dots, X_n$ con probabilidades $p_1, \dots, p_n$

- Si las variables $X_1, \dots, X_n$ son **continuas,** entonces la mezcla Y es continua y su densidad es:

$$f_Y(y) = p_1 f_{X_1}(y) + \dots + p_n f_{X_n}(y).$$

- Si las variables $X_1, \dots, X_n$ son **discretas,** entonces la mezcla Y es discreta y su función de probabilidad es:

$$p_Y(y) = p_1\, p_{X_1}(y) + \dots + p_n\, p_{X_n}(y).$$



Media de una mezcla

Si Y es una variable mezcla de n variables $X_1, \dots, X_n$ don medias $E[X_1], \dots, E[X_n]$ y probabilades de mezcla $p_1, \dots, p_n$ su media vale

$$E[Y] = \sum_{i=1}^{n} E[X_i] p_i$$

La media es el promedio de las medias, igual que la distribución de Y.

Podemos demostrar fácilmente este resultado. Por ejemplo, para V.A.s continuas:

$$\begin{aligned} E[Y] &= \int_{-\infty}^{\infty} y\, f_Y(y)\, dy = \int_{-\infty}^{\infty} y \sum_{i=1}^{n} p_i\, f_{X_i}(y)\, dy \\ &= \sum_{i=1}^{n} p_i \int_{-\infty}^{\infty} y\, f_{X_i}(y)\, dy \\ &= \sum_{i=1}^{n} E[X_i]\, p_i. \end{aligned}$$

Varianza de una mezcla

Si Y es una variable mezcla de n variables $X_1, \dots, X_n$ con medias $E[X_1], \dots, E[X_n]$, varianzas $V(X_1), \dots, V(X_n)$ y probabilidades de mezcla $p_1, \dots, p_n$, su varianza vale:

$$V(X) = \sum_{i=1}^{n} V(X_i)p_i + \sum_{i=1}^{n} (E[X_i] - E[Y])^2 P_I$$

¡OJO! La varianza NO es el promedio ponderado de las varianzas, como sucede con las medias.

Bayes para la mezcla

- Supongamos que Y es una variable mezcla de n variables $X_1, \dots, X_n$ con medias $E[X_1], \dots, E[X_n]$ y probabilidades de mezcl $p_1, \dots, p_n$
- Supongamos que se observa que la variable mezcla ha tomado un valor Y = y

Queremos saber, en funcion del valor observado ´ y, ¿cual es la probabilidad de que ese valor haya venido de cada VA $X_1, \dots, X_n$

Si definimos:

$$C_1 = \{Fue\ elegida\ la\ variable\ X_i\}$$

queremos hallar:

$$P(C_i|Y = y)$$

- Cuando Y es continua, el evento {Y = y} tiene probabilidad 0.
- Al hacer:

$$\mathbb{P}(C_i|Y = y) = \frac{\mathbb{P}(Y = y, C_i)}{\mathbb{P}(Y = y)} = \frac{0}{0},$$

  nos encontramos con una indeterminación.
- Debemos "agrandar" el evento {Y = y}, considerando la probabilidad:

$$\begin{aligned}\mathbb{P}(C_i|y - \epsilon < Y < y + \epsilon) &= \frac{\mathbb{P}(C_i, y - \epsilon < Y < y + \epsilon)}{\mathbb{P}(y - \epsilon < Y < y + \epsilon)}, \quad (\epsilon > 0)\\ &= \frac{\mathbb{P}(y - \epsilon < Y < y + \epsilon|C_i)\, p_i}{\mathbb{P}(y - \epsilon < Y < y + \epsilon)},\\ &= \frac{p_i \int_{y-\epsilon}^{y+\epsilon} f_{X_i}(x)dx}{\int_{y-\epsilon}^{y+\epsilon} f_Y(x)dx},\end{aligned}$$

- Tomando $\epsilon \to 0$ (con algunas consideraciones) se obtiene la expresión deseada:

$$P(C_i|Y = y) = \frac{p_i f_{X_i}(y)}{f_Y(y)}$$

EJEMPLO DE MEZCLA

Enunciado

Se tienen dos máquinas para producir varillas:

- La máquina I produce varillas con longitud (en metros) uniforme en $(10, 12)$.
- La máquina II produce varillas con longitud (en metros) uniforme en $(8, 12)$.

La máquina I produce el 70 % de las varillas y la otra el resto. La producción de varillas se deposita en un galpón.

1. Se extrae una varilla al azar del galpón y resulta que mide 11 metros. Calcule la probabilidad de que haya sido producida por la máquina II.
2. Halle la media y varianza del largo de una varilla del galpón.

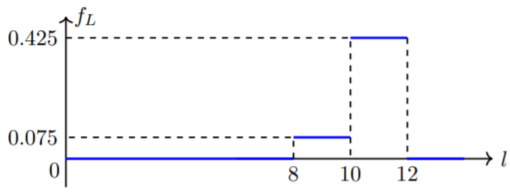
1) Calcule la probabilidad de que haya sido producida por la maquina II.

- Estamos ante un caso de mezcla, de dos uniformes $L_1 \sim \mathcal{U}(10, 12)$ y $L_2 \sim \mathcal{U}(8, 12)$, con probabilidades $p = 0{,}7$, y $1 - p$, respectivamente.
- Si definimos $M_i = \{\text{La pieza viene de la máquina i}\}$, $i = 1, 2$ e L es el largo de la varilla extraída, entonces piden calcular:

$$\mathbb{P}(M_2|L = 11) = \frac{(1 - p) f_{L_2}(11)}{f_L(11)}.$$

Hallamos la densidad de la mezcla primero (aunque podríamos hallar directamente $f_L(11)$:

$$\begin{aligned} f_L(l) &= p\, f_{L_1}(l) + (1 - p)\, f_{L_2}(l) \\ &= \frac{0{,}7}{2}\mathbf{1}_{(10,12)} + \frac{0{,}3}{4}\mathbf{1}_{(8,12)} \\ &= 0{,}35\, \mathbf{1}_{(10,12)} + 0{,}075\, \mathbf{1}_{(8,12)} \end{aligned}$$



De ahi tenemos:

$$\mathbb{P}(M_2|L = 11) = \frac{0{,}3 \cdot \frac{1}{4}}{0{,}425} = \frac{3}{17} \approx 0{,}176.$$

2) Halle la media y varianza del largo de una varilla del galpón

- Dado que hemos calculado la densidad de Y podemos hallar la media y la varianza resolviendo las integrales, pero aqui usamos las fórmulas vistas.
- Es fácil ver que si $X \sim \mathcal{U}(a, b)$:

$$E[X] = \frac{a + b}{2} \qquad V(X) = \frac{(b - a)^2}{12}.$$

entonces:

$$\begin{aligned} E[Y] &= p\, E[L_1] + (1 - p)\, E[L_2] \\ &= 0{,}7 \cdot 11 + 0{,}3 \cdot 10 \\ &= 10{,}7 \quad [\text{m}], \end{aligned}$$

y para la varianza:

$$\begin{aligned} V(X) &= pV(L_1) + (1 - p)V(L_2) + p(E[L_1] - E[Y])^2 + (1 - p)(E[L_2] - E[Y])^2 \\ &= 0{,}7 \cdot \frac{1}{3} + 0{,}3 \cdot \frac{4}{3} + 0{,}7 \cdot (11 - 10{,}7)^2 + 0{,}3 \cdot (10 - 10{,}7)^2 \\ &= \frac{253}{300} \approx 0{,}843. \quad [\text{m}^2] \end{aligned}$$