

Inferencia Estadística

Maestría en Análisis Estadístico y Computación

Capítulo 2: Variables aleatorias y distribuciones de probabilidad multivariadas

Dra. L. Leticia Ramírez Ramírez

Enero–Mayo, 2020



Distribuciones Multivariadas

- Modelos de Probabilidad Conjunta Discretos

- Modelos de Probabilidad Conjunta Continuos

- El modelo Multinomial

- El modelo Normal Bivariado

 - Regresión

- Más Funciones de variables aleatorias

 - Estadísticos de Orden

 - Mezclas de distribuciones

Distribuciones Multivariadas

Introducción

Hasta aquí hemos construido modelos para una sola variable aleatoria como si existieran siempre per se, sin nada que interactúe en su medio.

Sin embargo, en muchos casos, estaremos interesados en estudiar en forma conjunta más de una característica.

Por ejemplo, si medimos el porcentaje de zinc puro en una cierta reacción, no podemos modelar el comportamiento de este porcentaje sin tomar en cuenta las condiciones bajo las cuales fue producido, algunas de ellas serán “constantes”, pero otras pueden variar, esto es, uno esperaría cambios en los valores de los porcentajes de zinc si las revoluciones por minuto a las que se mueve el reactor cambian, o la velocidad del extractor varía, o si la temperatura no se puede mantener fija, etc.

Introducción

En casos como éstos, algunas de las variables determinan el comportamiento de otra (es un modelo del tipo Input-Output). Formas simplificadas de estos modelos son los que usamos frecuentemente en análisis de regresión, diseño de experimentos, etc.

Considera

$$y = \mu_y + e$$

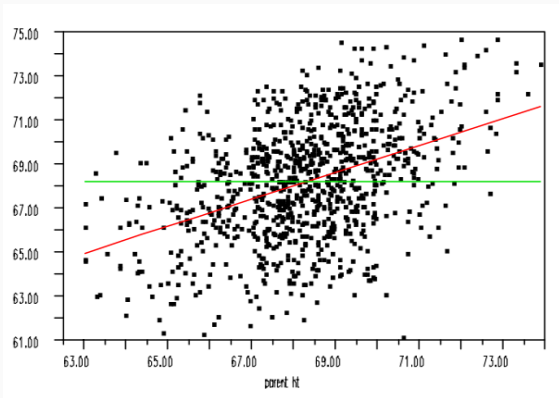
Aquí μ_y representa a la media de la variable aleatoria Y . Sin embargo, esta media cambia por la acción o presencia de otras variables, esto es:

$$\mu_y = f(x_1, x_2, x_3)$$

la media es una función (tal vez lineal, tal vez no) de las otras variables que de alguna manera quedan involucradas en el estudio.

Introducción

Por mencionar un ejemplo clásico en la literatura, supongamos que de la altura de los padres queremos saber cuál será la altura de los hijos, (predecir). Galton, en un estudio realizado sobre 942 padres e hijos, encontró algo como lo que se muestra en la gráfica siguiente:



¿Qué quiere decir esto?

La línea horizontal representa el valor promedio de las alturas de los hijos, ignorando las alturas de los padres, sin embargo la línea con pendiente creciente nos da una mejor idea de lo que está pasando con el comportamiento de las alturas de los hijos.

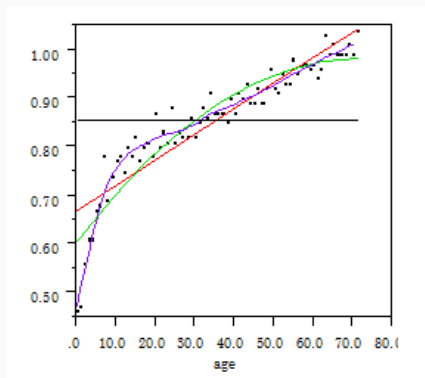
Se esperaba que, entre más altos sean los padres, más lo serán los hijos, relación que sólo se cumple en cierta medida.

Vean el alto nivel de dispersión que tenemos en este caso, lo cual, típicamente oscurece las conclusiones.

Introducción

Otro caso podría ser el crecimiento de los anillos en los troncos de los árboles (Dendrocronología) como un medio para determinar su edad.

Se muestra una gráfica de 72 mediciones en donde se tomaron árboles de la misma variedad con diferente número de años:



Introducción

Aquí, μ_y , el radio promedio del anillo de esa variedad de árboles, es una función no-lineal del tiempo, lo cual es completamente lógico, se espera tienda a estabilizarse después de un cierto número (grande) de años.

Las curvas muestran diferentes opciones de cómo modelar el comportamiento medio de los radios. La línea horizontal de nuevo, representa la media de los radios, sin tomar en cuenta la edad. Claramente *you are missing something!*

¿Es la edad una variable aleatoria? La respuesta en general es sí. El proceso es el siguiente, seleccionamos al azar árboles de esta variedad en una zona donde creemos que podemos encontrar tanto árboles jóvenes como viejos. Mediante algún procedimiento alternativo (por ejemplo pruebas con carbono 14) establecemos la edad del árbol y finalmente hacemos las mediciones de los diámetros de los anillos.

En este caso es inoperante pensar edad como un factor fijo y controlable por el investigador. Ambas mediciones son entonces variables aleatorias. De otra forma implicaría que para completar el estudio tendríamos que esperar 80 años!!!

En la industria por ejemplo, existen muchas situaciones en donde sí se puede tener el control de los factores. Recordemos el caso de producción de zinc puro, aquí, el valor de la temperatura, presión, revoluciones por minuto del revolvedor, del abanico de succión, etc. al menos en *teoría* permanecen fijos o pueden ser manipulados por el operador. Esto resulta en una simplificación muy útil cuando se plantea el modelo y como dijimos antes, son modelos que se trabajan en diseños de experimentos.

Introducción

En otros casos, no existe una relación de *dependencia*, sino que todas afectan a todas en algún nivel y nuestro interés se centra en entender tales *interdependencias*.

Consideremos otro ejemplo clásico en la literatura (**Fisher Iris data**), la construcción de un criterio de clasificación de variedades de plantas *setosa*, *versicolor* y *virginica* en base a *las mediciones del sépalo y el pétalo* (longitud y ancho).

Una vista de los datos en términos de las primeras 3 de estas características se presenta en la siguiente diapositiva.

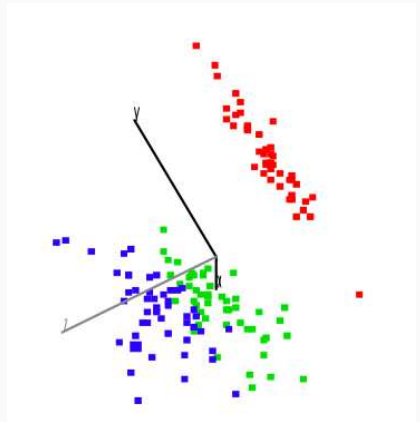
Introducción

Se ve claramente la separación entre variedades y más aún la diferencia de versicolor y virginica con setosa.

En cada planta se miden dos características, por lo que no se espera que esos valores entre sí sean independientes.

Nos interesa saber en qué medida estas 3 mediciones nos permiten decidir si la variedad es setosa, versicolor o virginica, y lo que explotamos es la estructura de dependencias entre las mediciones de sépalos y pétalos. Esto se conoce como análisis discriminante (se estudia en

Análisis Multivariado).



Para poder establecer soluciones a problemas como los mencionados antes, y muchos otros, necesitamos una serie de definiciones y conceptos básicos que son los que trataremos en esta sección y que, cabe mencionar, es lo único que estableceremos en el curso sobre estructuras multivariadas.

Muchos de estos conceptos también nos ayudarán a hacer las conexiones entre los modelos y las inferencias que podamos hacer sobre ellos a partir de información muestral.

Para simplificar la notación trabajaremos casi exclusivamente el caso bivariado dado que la mayoría de los resultados son directamente extendibles al caso de variables aleatorias.

Nuestros objetos de estudio son ahora parejas de variables (X, Y) .

Existen diferentes combinaciones, por ejemplo, ambas variables son discretas (modelos discretos), ambas continuas (modelos continuos) o una variable siendo discreta y la otra continua (modelos de mezclas).

Daremos una explicación extensiva de los primeros dos y sólo trataremos algunos casos especiales de modelos de mezclas.

Modelos de Probabilidad Conjunta Discretos

Modelos de Probabilidad Conjunta Discretos

Sólo para fines de ilustrar los conceptos, comenzaremos con un ejemplo muy simple. Supongamos que estamos muestreando 3 baterías de un lote con las siguientes características:

- 3 baterías son nuevas
- 4 son reconstruidas
- 5 tienen algún tipo de problema

Es un lote que no desearíamos tener en las manos! Por desgracia en la práctica no sabemos lo que hay, sólo lo que vemos en la muestra y de allí tratamos de inferir sobre las cualidades del lote completo; más aún, sobre el proceso de donde provienen!!

Modelos de Probabilidad Conjunta Discretos

Si definimos

X = # de baterías nuevas en la muestra,

Y = # de baterías reconstruidas en la muestra,

no es necesario definir una tercer variable Z para el resto de las baterías dado que con dos es suficiente para describir el comportamiento total de los elementos de la muestra, como veremos a continuación.

Podemos en principio, calcular las probabilidades

$$p(x, y) = P(X = x, Y = y),$$

para $x = 0, 1, 2, 3$ y $y = 0, 1, 2, 3$.

Modelos de Probabilidad Conjunta Discretos

Calculemos algunas de ellas:

$$p(0,0) = \frac{\# \text{ de casos favorables}}{\# \text{ de casos totales}} = \frac{\binom{5}{3}}{\binom{12}{3}} = \frac{10}{220}.$$

Aunque, realmente,

$$\# \text{ de casos favorables} = \binom{5}{3} \binom{3}{0} \binom{4}{0}.$$

Esto define lo que podríamos llamar una distribución Hipergeométrica Multivariada.

Modelos de Probabilidad Conjunta Discretos

Continuando con el proceso anterior

$$\begin{aligned}p(0, 1) &= \frac{\binom{5}{2} \binom{4}{1}}{\binom{12}{3}} = \frac{40}{220}, & p(0, 2) &= \frac{\binom{4}{2} \binom{5}{1}}{\binom{12}{3}} = \frac{30}{220}, \\p(0, 3) &= \frac{\binom{4}{3}}{\binom{12}{3}} = \frac{4}{220}, & p(1, 0) &= \frac{\binom{3}{1} \binom{5}{2}}{\binom{12}{3}} = \frac{30}{220}, \\p(1, 1) &= \frac{\binom{3}{1} \binom{4}{1} \binom{5}{1}}{\binom{12}{3}} = \frac{60}{220}, & p(1, 2) &= \frac{\binom{3}{1} \binom{4}{2} \binom{5}{0}}{\binom{12}{3}} = \frac{18}{220}, \\p(2, 0) &= \frac{\binom{3}{2} \binom{4}{0} \binom{5}{1}}{\binom{12}{3}} = \frac{15}{220}, & p(2, 1) &= \frac{\binom{3}{2} \binom{4}{1}}{\binom{12}{3}} = \frac{12}{220}, \\p(3, 0) &= \frac{\binom{3}{3}}{\binom{12}{3}} = \frac{1}{220}.\end{aligned}$$

Modelos de Probabilidad Conjunta Discretos

Arreglando los resultados anteriores en una tabla, se tiene que

x/y	0	1	2	3
0	10/220	40/220	30/220	4/220
1	30/220	60/220	18/220	0
2	15/220	12/220	0	0
3	1/220	0	0	0

Lo que obtenemos es lo que se conoce como *distribución conjunta* de X y Y . Cada entrada en la tabla no es más que la probabilidad de la intersección de dos eventos $\{X = x, Y = y\}$.

¿Cuándo sabemos que hemos construido un modelo de probabilidad conjunta válido? La respuesta, al igual que antes, radica en el cumplimiento de las reglas básicas de probabilidad.

Modelos de Probabilidad Conjunta Discretos

Definición

Sean X y Y dos variables aleatorias discretas. Se define la función **función de probabilidad conjunta de X y Y** como

$$f(x, y) = P(X = x, Y = y),$$

para cada par de valores (x, y) en el rango de X, Y . La cantidad representa la probabilidad de la intersección de los dos eventos: $X = x$ y $Y = y$.

La función de probabilidad conjunta satisface que:

1. $0 \leq f(x, y) \leq 1 \quad \forall (x, y),$
2. $\sum_x \sum_y f(x, y) = 1.$

Modelos de Probabilidad Conjunta Discretos

Notemos que nuestra tabla cumple con ambas propiedades:

x/y	0	1	2	3	$P\{X = x\}$
0	10/220	40/220	30/220	4/220	84/220
1	30/220	60/220	18/220	0	108/220
2	15/220	12/220	0	0	27/220
3	1/220	0	0	0	1/220
$P\{Y = y\}$	56/220	112/220	48/220	4/220	1

De aquí, se pueden definir otras distribuciones de interés. Por ejemplo, es claro que el último renglón en la tabla representa la distribución de $Y = \text{"\# de baterías reconstruidas en la muestra"}$ y que la última columna es la distribución de $X = \text{"\# de baterías nuevas en la muestra"}$. A éstas distribuciones se les conoce como las distribuciones *marginales* (por estar en los márgenes).

Modelos de Probabilidad Conjunta Discretos

Notemos que para obtener las distribuciones marginales lo que se hizo fue ignorar la presencia de la otra variable y , por ende, sumar sobre cada renglón o cada columna, según sea el caso.

Definición

Las distribuciones marginales de X y Y , denotadas por $p(x) = f_X(x)$ y $p(y) = f_Y(y)$, se definen como

$$f_X(x) = \sum_y p(x, y) \quad \text{para cada } x,$$

y

$$f_Y(y) = \sum_x p(x, y) \quad \text{para cada } y.$$

Bajo estos casos (más de una variable) también podemos contestarnos otras preguntas que involucran el conocimiento previo de la información, por ejemplo:

Si se sabe que X tomó el valor de 2, ¿cuál es el comportamiento de Y ?

Claramente esto restringe sus posibles valores y por tanto sus probabilidades cambian.

Modelos de Probabilidad Conjunta Discretos

A saber, tendríamos que calcular

$$P\{Y = 0 \mid X = 2\} = \frac{P\{X = 2, Y = 0\}}{P\{X = 2\}} = \frac{\frac{15}{220}}{\frac{27}{220}} = \frac{15}{27},$$

y así para cada valor de Y .

x/y	0	1	2	3	$P\{X = x\}$
0					
1					
2	15/220	12/220	0	0	27/220
3					
$P\{Y = y \mid X = 2\}$	15/27	12/27	0	0	1

Modelos de Probabilidad Conjunta Discretos

Es como olvidarnos del resto de la información: ahora, en lugar de las 220 posibles muestras iniciales, solo nos quedan 27, aquellas que tienen el valor de 2 asignado a X .

Este proceso de asignar probabilidades es de particular importancia, pues responde a un tipo de problemas, los más comunes.

Regresa sobre los ejemplos con los que iniciamos este capítulo y verás que nos preguntamos, por ejemplo, sobre el % de zinc cuando la temperatura es xxxx, etc. En otras palabras, estamos frecuentemente interesados en saber cómo se comporta nuestra característica bajo estudio sujeta a la presencia de otras variables, las cuales podemos fijar a diferentes valores y estudiar su efecto en la variable respuesta.

Estas distribuciones reciben el nombre de **distribuciones Condicionales**.

Modelos de Probabilidad Conjunta Discretos: Distrib. Condicional

Definición

Sean x y y dos variables aleatorias con distribución conjunta $f(x, y)$ y marginales definidas por $f_X(x)$ y $f_Y(y)$. Entonces

- i) La distribución condicional de Y dado $X = x$ es

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

para cada valor de y , donde $f_X(x) = \sum_y f(x, y)$ para cada x .

- ii) La distribución condicional de X dado Y es:

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

para cada valor de x , donde $f_Y(y) = \sum_x f(x, y)$ para cada y .

Modelos de Probabilidad Conjunta Discretos: Distrib. Condicional

Así las cosas, se pueden obtener tantas distribuciones condicionales para Y como valores se tengan de X , y viceversa. En nuestro ejemplo, podríamos calcular 3 distribuciones condicionales para Y y otras 3 para X .

Al igual que en el caso de una variable, son las condiciones del muestreo las que nos dirán como construir el modelo conjunto.

En muchas situaciones es a través de la recopilación histórica, de archivo o planeando un experimento para estos fines, que se pueden construir tablas de distribuciones conjuntas, utilizando frecuencias relativas, que se consideran estables (léase, correctas).

Modelos de Probabilidad Conjunta Discretos: Distrib. Condicional

Ejemplo

Se quiere **verificar la programación** del tiempo que debe durar la luz verde en un semáforo que se encuentra en una cierta intersección, con vuelta a la izquierda. Como no se tiene información al respecto, se envía a un estudiante graduado a hacer observaciones sobre el número de carros X y el número de camiones Y que llegan en un ciclo (entre verde y verde); y con esto se construye la tabla de distribuciones conjunta. Se plantean una serie de preguntas y con ello confirmas qué tan eficiente fue el ciclo planeado.

Modelos de Probabilidad Conjunta Discretos: Distrib. Condicional

Los resultados fueron los siguientes

y	x			
	$p(x, y)$	0	1	2
	0	.025	.015	.010
	1	.050	.030	.020
	2	.125	.075	.050
	3	.150	.090	.060
	4	.100	.060	.040
	5	.050	.030	.020

Modelos de Probabilidad Conjunta Discretos: Distrib. Condicional

Ejercicio:

1. Verifica que es una tabla válida de probabilidades conjuntas.
2. Calcula las distribuciones marginales para carros y camiones.
3. ¿Cuál es la probabilidad de que se tengan exactamente un carro y un camión en un ciclo dado?
4. Supongamos que la vuelta a la izquierda tiene una capacidad máxima de 5 carros y un camión es equivalente a 3 carros. ¿Cuál es la probabilidad de que se sature la línea en un ciclo dado?
5. ¿Cuál es la probabilidad de que se tenga exactamente un carro, dado que ya se tiene un camión en un ciclo predeterminado y se sabe que no habrá ningún otro camión más? Contesta esta misma pregunta para $X = 0, 1, 2, 3, 4, 5$.

Modelos de Probabilidad Conjunta Discretos: Independencia

Con el concepto de distribuciones condicionales, podemos refinar nuestra definición, hasta ahora operativa, de **independencia**. Este es posiblemente uno de los supuestos más usados en la construcción de modelos y por lo mismo uno de los que más se abusa.

¿Qué significaría que $P(Y = y \mid X = x) = P(Y = y)$?

Básicamente nos dice que X , al menos cuando se fija en su nivel x , no agrega información adicional sobre el comportamiento de Y . Entonces, si esta condición se cumple sin importar el valor que propongamos de X , se dirá que X y Y son independientes.

Modelos de Probabilidad Conjunta Discretos: Independencia

Más formalmente tenemos la siguiente definición.

Definición

Dos variables aleatorias X y Y se dice que son independientes si, para todo par de valores x y y

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

Nota:

$$P(Y = y | X = x) = \frac{f(x, y)}{f_X(x)} \Rightarrow \begin{aligned} f(x, y) &= P(Y = y | X = x)f_X(x) \\ &= P(Y = y)f_X(x) = f_Y(y)f_X(x) \end{aligned}$$

“Si las variables aleatorias son independientes, la conjunta es el producto de las marginales.”

¿Es el número de carros en un ciclo independiente del número de camiones?

Modelos de Probabilidad Conjunta Continuos

Los conceptos son los mismos, sólo hay que ajustar las definiciones a las características de las variables continuas.

Comencemos con un caso concreto:

Ejemplo

Un grupo de estudiantes del Programa Emprendedor establecieron un restaurante de comida rápida (fast food) e hicieron un estudio de tiempos para evaluar el servicio al cliente. Definamos Y_1 como el tiempo total que tarda un cliente desde que entra al local hasta que deja la ventanilla de servicio y Y_2 el tiempo que espera el cliente en la fila hasta llegar a la ventilla de servicio. Claramente $Y_1 \geq Y_2$.

Ejemplo (Continuación)

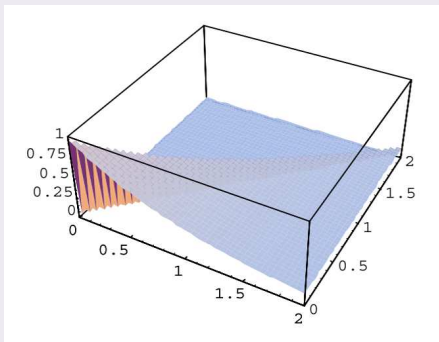
Con un poco más de esfuerzo y aprovechando lo que aprendieron en sus cursos de matemáticas, lograron establecer el modelo del comportamiento conjunto dado por:

$$f(y_1, y_2) = \begin{cases} e^{-y_1} & 0 \leq y_2 \leq y_1 \leq \infty \\ 0 & \text{en otra parte} \end{cases}$$

Modelos de Probabilidad Conjunta Continuos

Ejemplo (Continuación)

Esperaban un comportamiento de caída exponencial, así que tomaron datos y ajustaron el mejor modelo en esta familia, con un software especializado.



Notas:

- Por supuesto que es irreal pensar que tome valores infinitos. No puede sobrepasar lo que dura abierto el restaurante y ningún cliente en un negocio de este tipo, soportaría una espera larga. Sin embargo, y como ya lo hicimos notar desde el inicio del curso, ésta es una hipótesis de simplicidad. Notemos que a los tiempos grandes, si el modelo es adecuado, les corresponde una probabilidad prácticamente de cero por lo que resulta irrelevante si se quitan de antemano. El eliminar esos valores conllevaría por otro lado, a establecer modelos que son bastante más complejos de manejar desde su planteamiento y, a fin de cuentas, generarían respuestas casi idénticas en la región de interés.

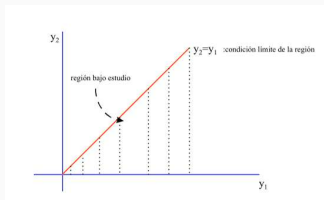
Modelos de Probabilidad Conjunta Continuos

Notas:

- Resulta evidente de la forma de la función y de la gráfica que ahora tenemos definida una superficie y en lugar de valores por intervalo, tenemos regiones en el plano que definen a los valores de las variables. Por ejemplo, aquí, la región en donde la función toma valores distintos de cero es:

$$0 \leq y_2 \leq y_1 \leq \infty$$

esto se representa como:



Modelos de Probabilidad Conjunta Continuos

Ahora bien, necesitamos primero que nada verificar si éste es un modelo de probabilidad. Para ello y viendo las definiciones de la sección previa, extendamos la definición al caso continuo.

Definición

Una función de 2 variables $f(x, y)$ definida sobre el plano xy se llama función de densidad de probabilidad conjunta (densidad conjunta) de X y Y si y sólo si :

$$P((x, y) \in A) = \int_A f(x, y) dx dy \quad \forall \text{ región } A \text{ del plano } xy.$$

Esto nos dice que las probabilidades se obtienen como volúmenes bajo la superficie definida. La superficie debe verse siempre sobre el plano xy .

Modelos de Probabilidad Conjunta Continuos

La función de densidad conjunta satisface que:

1. $f(x, y) \geq 0, \forall (x, y)$. La función no puede ser negativa -cs-.)
2. $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dx dy = 1$. (El volumen total debe ser la unidad.)

a) Mostrar que la función determinada por nuestros emprendedores, es realmente una función de densidad conjunta.

Esto significa que la

$$\begin{aligned}\int_0^{\infty} \int_0^{y_1} e^{-y_1} dy_2 dy_1 &= \int_0^{\infty} e^{-y_1} y_2 \Big|_0^{y_1} dy_1 = \int_0^{\infty} y_1 e^{-y_1} dy_1 \\ &= -y_1 e^{-y_1} \Big|_0^{\infty} + \int_0^{\infty} e^{-y_1} dy_1 \\ &= 0 - e^{-y_1} \Big|_0^{\infty} = 0 + 1 = 1,\end{aligned}$$

por lo que la distribución conjunta está bien definida.

Modelos de Probabilidad Conjunta: Función de distribución acumulada

A continuación estudiaremos el concepto de función acumulada, el cual es muy útil pues nos permite calcular probabilidades. Daremos su definición pero no nos preocuparemos, en general, de calcular dicha función explícitamente; sólo nos enfocaremos en la idea que de ella se desprende que nos facilite los cálculos posteriormente.

Definición

La función de distribución conjunta o función de distribución acumulada conjunta de X y Y se define como:

$$F(x, y) = P(X \leq x, Y \leq y) = \begin{cases} \int_{-\infty}^y \int_{-\infty}^x f(s, t) ds dt & X \text{ y } Y \text{ continuas} \\ \sum_{t \leq y} \sum_{s \leq x} f(s, t) & X \text{ y } Y \text{ discretas} \end{cases}$$

b) Calcular las densidades marginales de Y_1 y Y_2 . Esto es, determinar los comportamientos por separado del tiempo total que tarda un cliente desde que entra al local hasta que deja la ventanilla de servicio, y el tiempo que dura en la fila hasta llegar a la ventanilla de servicio.

Para responder esta pregunta, primero tenemos que definir que entendemos ahora por **marginales**: sólo recordemos que significa ignorar la presencia de la(s) otra(s) variable(s) y esto se logra “sumando” sobre todos los valores de las variables que no son de nuestro interés en ese momento.

Geométricamente se vería como “aplastar” la gráfica sobre el eje que nos interesa.

Definición

Las distribuciones **marginales** de v.a. continuas X y Y , denotadas por $f_X(x)$ y $f_Y(y)$, se definen como

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

y

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

Nota: Para el caso discreto intercambiamos las \sum por \int .

Ahora sí, volviendo sobre el ejemplo, tenemos

$$\begin{aligned}f_{Y_1}(y_1) &= \int_0^{y_1} e^{-y_1} dy_2 = y_2 e^{-y_1} \Big|_0^{y_1} \\&= y_1 e^{-y_1} \quad 0 \leq y_1 \leq \infty \quad \text{una Gamma}(\alpha = 2, \beta = 1)\end{aligned}$$

y

$$\begin{aligned}f_{Y_2}(y_2) &= \int_{y_2}^{\infty} e^{-y_1} dy_1 = -e^{-y_1} \Big|_{y_2}^{\infty} \\&= e^{-y_2} \quad 0 \leq y_2 \leq \infty \quad \text{una exp}(\beta = 1)\end{aligned}$$

Modelos de Probabilidad Conjunta Continuos

No siempre las distribuciones marginales resultan en distribuciones conocidas...

Definición

Dos variables aleatorias continuas se dice que son **independientes** si su distribución conjunta se puede expresar como el producto de sus marginales

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

Más adelante volveremos sobre esta definición.

Ahora plantearemos una serie de preguntas, para algunas de las cuales tendremos que definir algunos conceptos en el contexto varias variables.

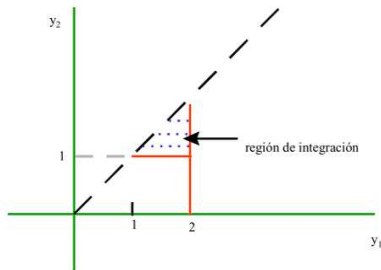
Modelos de Probabilidad Conjunta Continuos

Recordemos que: Y_1 : tiempo total que tarda un cliente desde que entra al local hasta que deja la ventanilla de servicio

Y_2 : tiempo que espera el cliente en la fila hasta llegar a la ventanilla de servicio.

Definimos el evento $A =$ *el tiempo total es menor a 2 minutos y el tiempo en la fila es mayor a uno*. Entonces

$$P(A) = P(Y_1 < 2, Y_2 > 1).$$



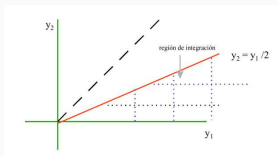
Así tenemos que

$$\begin{aligned}P(Y_1 < 2, Y_2 > 1) &= \int_1^2 \int_{y_2}^2 e^{-y_1} dy_1 dy_2 \\&= \int_1^2 [-e^{-y_1}]_{y_2}^2 dy_2 \\&= \int_1^2 [-e^{-2} + e^{-y_2}] dy_2 \\&= [-e^{-2}y_2 - e^{-y_2}]_1^2 \\&= -2e^{-2} - e^{-2} - [-e^{-2} - e^{-1}] \\&= e^{-1} - 2e^{-2} \\&= 0.0972088.\end{aligned}$$

Modelos de Probabilidad Conjunta Continuos

c) Ahora definimos el evento $A =$ el tiempo total de espera sea mayor al doble del tiempo en la fila. Determinar

$$P(A) = P(Y_1 \geq 2Y_2).$$

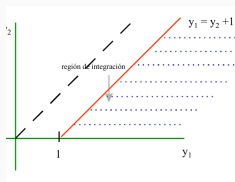


$$\begin{aligned} P(Y_1 \geq 2Y_2) &= \int_0^\infty \int_{2y_2}^\infty e^{-y_1} dy_1 dy_2 = \int_0^\infty [-e^{-y_1}]_{2y_2}^\infty dy_2 \\ &= \int_0^\infty e^{-2y_2} dy_2 = -\frac{e^{-2y_2}}{2} \Big|_0^\infty = \frac{1}{2}. \end{aligned}$$

Modelos de Probabilidad Conjunta Continuos

d) La variable aleatoria $Y_1 - Y_2$ representa el tiempo que se tarda en la ventanilla de servicio. Sea el evento $A =$ el tiempo en la ventanilla sea mayor o igual que un minuto. Calcular:

$$P(A) = P(Y_1 - Y_2 \geq 1)$$



$$\begin{aligned} P(Y_1 - Y_2 \geq 1) &= \int_0^{\infty} \int_{y_2+1}^{\infty} e^{-y_1} dy_1 dy_2 = \int_0^{\infty} [-e^{-y_1}]_{y_2+1}^{\infty} dy_2 \\ &= \int_0^{\infty} e^{-(y_2+1)} dy_2 = e^{-(y_2+1)} \Big|_0^{\infty} = e^{-1} = 0.36788. \end{aligned}$$

e) Si se sabe que el tiempo total de espera y de atención para un cliente será de más de dos minutos, calcular la probabilidad de que el tiempo que dicho cliente espera para que le atiendan sea menor de un minuto.

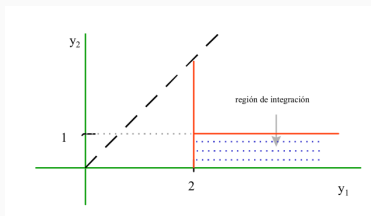
Se sabe que $Y_1 > 2$, queremos la probabilidad de que $Y_2 < 1$. Esto es, queremos calcular una probabilidad condicional del tipo

$$P(Y_2 < 1 | Y_1 > 2) = \frac{P(Y_2 < 1, Y_1 > 2)}{P(Y_1 > 2)}.$$

En este caso sólo nos resta calcular la probabilidad indicada en el numerador y usar el resultado de las marginales, para obtener la probabilidad en el denominador.

Modelos de Probabilidad Conjunta Continuos

La región de integración para el numerador es



$$\begin{aligned} P(Y_2 < 1, Y_1 > 2) &= \int_0^1 \int_2^{\infty} e^{-y_1} dy_1 dy_2 = \int_0^1 [-e^{-y_1}]_2^{\infty} dy_2 \\ &= \int_0^1 e^{-2} dy_2 \\ &= y_2 e^{-2} \Big|_0^1 = e^{-2}. \end{aligned}$$

Modelos de Probabilidad Conjunta Continuos

Para el denominador:

$$\begin{aligned} P(Y_1 > 2) &= \int_2^{\infty} f_{Y_1}(y_1) dy_1 = \underbrace{\int_2^{\infty} y_1 e^{-y_1} dy_1}_{\substack{u = y_1 & dv = e^{-y_1} dy_1 \\ du = dy_1 & v = -e^{-y_1}}} = -y_1 e^{-y_1} \Big|_2^{\infty} + \int_2^{\infty} e^{-y_1} dy_1 \\ &= 2e^{-2} - e^{-y_1} \Big|_2^{\infty} = 2e^{-2} + e^{-2} = 3e^{-2} \\ \therefore P(Y_2 < 1 | Y_1 > 2) &= \frac{e^{-2}}{3e^{-2}} = \frac{1}{3} \end{aligned}$$

Modelos de Probabilidad Conjunta Continuos: Valor Esperado

Nota: Podrás observar que en todo lo anterior (y en lo que sigue), no necesitamos ver el comportamiento de la superficie definida por la densidad conjunta, pero lo que sí es crucial es establecer la región de interés en el plano para definir el volumen que nos de la probabilidad requerida.

f) Calcular el tiempo medio que un cliente estará en la ventanilla.

¿Será muy probable que un cliente pase más de dos minutos en la ventanilla de servicio?

Nos preguntamos por $E(Y_1 - Y_2)$. Nosotros hemos aprendido cómo calcular valores de sumas de esperanzas en principio para cuando sólo teníamos una variable aleatoria. Primero daremos algunas definiciones para el caso de dos (o más variables aleatorias) y algunas de sus propiedades y volveremos más adelante para responder a las preguntas en este apartado.⁵²

Valores Esperados

Demos un pequeño recorrido por lo que sabemos de valores esperados:

$$E(X) = \int_{-\infty}^{+\infty} xf_X(x)dx \quad y \quad E(Y) = \int_{-\infty}^{+\infty} yf_Y(y)dy.$$

Si tenemos dos variables aleatorias, estos mismos valores se pueden obtener directamente (es como calcular la marginal dentro del procedimiento de obtención del esperado).

Modelos de Probabilidad Conjunta Continuos: Valor Esperado

Recordemos que ahora existe una densidad conjunta (o probabilidad conjunta):

$$E(X) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} xf(x, y)dx dy = \int_{-\infty}^{+\infty} x \left\{ \int_{-\infty}^{+\infty} f(x, y)dy \right\} dx = \int_{-\infty}^{+\infty} xf(x)dx.$$

La integral en rojo, no es otra cosa que la marginal $f(x)$, de donde volvemos a la definición inicial.

Análogamente para Y :

$$\begin{aligned} E(X + Y) &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (x + y)f(x, y)dx dy \\ &= \int_{-\infty}^{+\infty} xf(x)dx + \int_{-\infty}^{+\infty} yf(y)dy \\ &\quad \text{(y usando los resultados vistos arriba)} \\ &= E(X) + E(Y). \end{aligned}$$

Modelos de Probabilidad Conjunta Continuos

Con esto podemos ya responder a la primera parte de nuestra pregunta, sobre el tiempo medio que el cliente estará en la ventanilla, *pues el resultado anterior es válido, ya sea si sumamos o restamos*:

$$E(Y_1 - Y_2) = E(Y_1) - E(Y_2)$$

y estos valores se pueden obtener directamente de sus marginales.

Como vimos que la distribución de Y_1 es una $\text{Gamma}(2,1)$ y la de Y_2 una $\text{exponencial}(1)$

$$E(Y_1 - Y_2) = E(Y_1) - E(Y_2) = 2 \cdot 1 - 1 = 1. \quad (1 \text{ minuto en promedio})$$

Modelos de Probabilidad Conjunta Continuos

Por otra parte, como $Y_1 - Y_2$ es el tiempo que pasa en la ventanilla de servicio

$$\begin{aligned}\Rightarrow \text{queremos} \quad P(Y_1 - Y_2 > 2) &= \int_0^{\infty} \int_{y_2+2}^{\infty} e^{-y_1} dy_1 dy_2 \\ &= \int_0^{\infty} [-e^{-y_1}]_{y_2+2}^{\infty} dy_2 \\ &= \int_0^{\infty} e^{-(y_2+2)} dy_2 \\ &= -e^{-y_2+2} \Big|_0^{\infty} = e^{-2} = 0.135335,\end{aligned}$$

o sea que sólo el 13.53% de las veces un cliente tendrá que esperar en la ventanilla más de dos minutos.

Volvamos sobre las definiciones

A pesar de que estamos trabajando con parejas (o vectores), no definimos valores esperados sobre el vector más allá de aquellas operaciones sobre los elementos del mismo que nos den un valor real, esto es, trabajaremos sólo con los casos donde:

$$g(x, y) \rightarrow \mathbb{R} \quad (\text{un valor en los reales}).$$

Ejemplos de tales funciones son sumas, restas, productos, logaritmos, etc. Cualquier operación que no nos regrese un vector como resultado.

Así, podemos definir en forma un poco más general, el cálculo del valor esperado con distribuciones conjuntas.

Definición

Sean X y Y variables aleatorias con función de distribución conjunta $f(x, y)$ (o densidad conjunta) y $g(X, Y)$, una función real, entonces

$$E(g(X, Y)) = \begin{cases} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(x, y) f(x, y) dx dy & \text{si } x, y \text{ son continuas} \\ \sum_{\forall x} \sum_{\forall y} g(x, y) p(x, y) & \text{si } x, y \text{ son discretas} \end{cases}$$

Modelos de Probabilidad Conjunta Continuos

A manera de ejercicio y para establecer un resultado que es muy útil que ya hemos usado repetidas veces, calculemos $E(XY)$

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

sin más información, sólo nos restaría integrar para cada caso particular.

Sin embargo si sabemos que X y Y son independientes, podemos reducir mucho el trabajo de la siguiente manera:

$$\begin{aligned} E(XY) &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} xyf(x, y) dx dy = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} xyf(y)f(x) dx dy \\ &\quad \text{(por la independencia)} \\ &= \int_{-\infty}^{+\infty} xf(x) dx \int_{-\infty}^{+\infty} yf(y) dy = E(X)E(Y). \end{aligned}$$

Modelos de Probabilidad Conjunta Continuos

Lo que el resultado anterior nos dice es que, bajo independencia, el valor esperado del producto es el producto de los valores esperados (esto lo usamos para establecer propiedades de la generatriz de momentos).

El resultado también se sostiene para el caso de variables discretas.

Esta definición general de valor esperado nos da también la oportunidad de extender el concepto a el cálculo de varianzas. Recordemos que conceptualmente estamos haciendo lo mismo, esto es, construyendo una medida de la variabilidad o dispersión de los valores de las variables alrededor de su valor medio; sólo que aquí podemos hablar de la dispersión de X , de la de Y , o de cualquier nueva variable definida por la transformación bajo estudio, por ejemplo, $X + Y$, $\ln(XY)$, etc.

Modelos de Probabilidad Conjunta Continuos

Tenemos que

$$V(X) = \int_{-\infty}^{+\infty} (x - \mu_x)^2 f(x) dx \quad \text{con } \mu_x = E(X).$$

Tomando $g(x, y) = (x - \mu_x)^2$, obtenemos que

$$\begin{aligned} V(X) &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(x, y) f(x, y) dx dy \\ &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (x - \mu_x)^2 \left\{ \int_{-\infty}^{+\infty} f(x, y) dy \right\} dx \\ &= \int_{-\infty}^{+\infty} (x - \mu_x)^2 f(x) dx. \end{aligned}$$

Lo mismo para Y .

Modelos de Probabilidad Conjunta Continuos

Definición

Sean X y Y variables aleatorias con función de distribución conjunta (o densidad conjunta) $f(x, y)$ y $g(x, y)$ una función en los reales,

$$V(g(X, Y)) = \begin{cases} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} [g(x, y) - \mu_{g(x, y)}]^2 f(x, y) dx dy & \text{si } x, y \text{ son continuas} \\ \sum_{\forall x} \sum_{\forall y} [g(x, y) - \mu_{g(x, y)}]^2 p(x, y) & \text{si } x, y \text{ son discretas} \end{cases}$$

donde $\mu_{g(x, y)} = E(g(X, Y))$.

En forma compacta:

$$V(g(X, Y)) = E[g(x, y) - \mu_{g(x, y)}]^2 = E(g(X, Y))^2 - \{\mu_{g(x, y)}\}^2.$$

Modelos de Probabilidad Conjunta Continuos

Calculemos, por ejemplo, $V(X + Y)$

$$\begin{aligned} V(X + Y) &= E[X + Y - \{E(X + Y)\}]^2 = E[X - E(X) + Y - E(Y)]^2 \\ &= E(X - E(X))^2 + E(Y - E(Y))^2 + 2E[(X - E(X))(Y - E(Y))] \\ &= V(X) + V(Y) + 2Cov(X, Y). \end{aligned}$$

Definiendo

$$\begin{aligned} Cov(X, Y) &= E[(X - E(X))(Y - E(Y))] \quad \text{la covarianza entre } X \text{ y } Y \\ &= E(XY) - E(X)E(Y). \end{aligned}$$

Modelos de Probabilidad Conjunta Continuos

¿Qué mide este término generado como el producto cruzado de las diferencias de las desviaciones de los valores con respecto a sus respectivas medias?

La mayoría de nosotros hemos visto este término antes. Veamos que pasa con éste cuando asumimos que X y Y son **independientes**

$$E\{[X - E(X)][Y - E(Y)]\} = E[X - E(X)]E[Y - E(Y)] = 0,$$

esto es, si las variables son independientes entonces la *covarianza* es cero. Esto nos dice que este término nos da una forma de medir dependencias entre X y Y .

Por desgracia, sólo mide un tipo de dependencias: *las dependencias lineales*. **Esto nos dice que si $\text{Cov}(X, Y) = 0$, no necesariamente es cierto que las variables sean independientes, simplemente dice que no hay dependencia lineal, pero puede ser de otro tipo.**

Modelos de Probabilidad Conjunta Continuos

Por ejemplo si consideramos las variables X y $Y = X^2$, *claramente existe una relación cuadrática entre X y Y .*

Resultado: Si X y Y son variables aleatorias independientes

$$\text{Cov}(X, Y) = 0 \quad \text{y} \quad \text{Var}(X \pm Y) = \text{Var}(X) + \text{Var}(Y),$$

de otra forma

$$\text{Var}(X \pm Y) = \text{Var}(X) + \text{Var}(Y) \pm 2\text{Cov}(X, Y).$$

Ya antes habíamos invocado este resultado sin demostración. Estos resultados se extienden directamente a n variables aleatorias discretas o continuas.

Modelos de Probabilidad Conjunta Continuos

Como con medias y varianzas, en donde tenemos sus formas correspondientes en términos de valores muestrales, podemos definir lo que se conoce como la *covarianza muestral*.

Sea $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ una muestra aleatoria, entonces la *covarianza muestral* se define como:

$$d = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{n - 1}$$

Notemos que la muestra ahora, está formada de parejas de valores.

Modelos de Probabilidad Conjunta Continuos

En general es difícil darse una idea adecuada del nivel de dependencia lineal a partir de los valores de covarianzas dado que estos dependen de la escala en que manejamos la información. Por esa razón se ideó un índice ρ conocido como **coeficiente de correlación**:

$$\rho = \frac{\text{Cov}(X, Y)}{\sigma_x \sigma_y} = E \left(\frac{X - E(X)}{\sigma_x} \right) \left(\frac{Y - E(Y)}{\sigma_y} \right).$$

La última igualdad implica que el coeficiente de correlación se puede pensar como la covarianza de las variables estandarizadas.

Esta definición tiene varias ventajas:

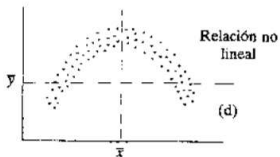
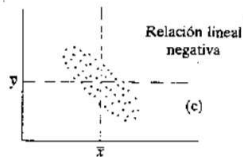
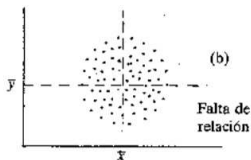
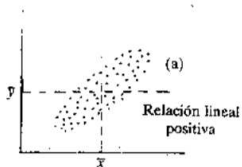
1. ρ es invariantes bajo cambios de escala;
2. $-1 \leq \rho \leq 1$;
3. Es 1 ó -1 sólo cuando la relación lineal es exacta;
4. $\rho = 0$ (esto es $\text{Cov}(X,Y)=0$) indica que no hay dependencia lineal.

El correspondiente coeficiente de correlación muestral se define como

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{s_x s_y}.$$

Modelos de Probabilidad Conjunta Continuos

Algunas ilustraciones del comportamiento de este coeficiente:



Modelos de Probabilidad Conjunta Continuos

$Y_1 - Y_2$ el tiempo que pasa el cliente en la ventanilla. ¿Qué tanta variabilidad hay en estos tiempos?

$$V(Y_1 - Y_2) = V(Y_1) + V(Y_2) - 2\text{Cov}(Y_1, Y_2)$$

Ya conocemos las varianzas, porque corresponden a las de las distribuciones marginales

$$V(Y_1) = 2 \cdot 1^2 = 2 \quad y \quad V(Y_2) = 1$$

Mostraremos todos los cálculos sólo a manera de ejercicio de las definiciones que hemos dado hasta aquí.

Para calcular estas varianzas y covarianzas necesitamos obtener

$$E(Y_1^2), \quad E(Y_2^2) \quad y \quad E(Y_1 Y_2).$$

Modelos de Probabilidad Conjunta Continuos

$$\begin{aligned} E(Y_1^2) &= \int_{-\infty}^{+\infty} y_1^2 f_{Y_1}(y_1) dy_1 = \underbrace{\int_0^{\infty} y_1^3 e^{-y_1} dy_1}_{\substack{u = y_1^3 & dv = e^{-y_1} dy_1 \\ du = 3y_1^2 dy_1 & v = -e^{-y_1}}} \\ &= -y_1^3 e^{-y_1} \Big|_0^{\infty} + \int_0^{\infty} 3y_1^2 e^{-y_1} dy_1 \\ &= \underbrace{\int_0^{\infty} 3y_1^2 e^{-y_1} dy_1}_{\substack{u = 3y_1^2 & dv = e^{-y_1} dy_1 \\ du = 6y_1 dy_1 & v = -e^{-y_1}}} = -3y_1^2 e^{-y_1} \Big|_0^{\infty} + \underbrace{\int_0^{\infty} 6y_1 e^{-y_1} dy_1}_{\substack{u = 6y_1 & dv = e^{-y_1} dy_1 \\ du = 6 dy_1 & v = -e^{-y_1}}} \\ &= -6y_1 e^{-y_1} \Big|_0^{\infty} + \int_0^{\infty} 6e^{-y_1} dy_1 \\ &= -6e^{-y_1} \Big|_0^{\infty} = 6 \end{aligned}$$

Modelos de Probabilidad Conjunta Continuos

$$\begin{aligned} E(Y_2^2) &= \underbrace{\int_0^{\infty} y_2^2 e^{-y_2} dy_2}_{\substack{u = y_2^2 & dv = e^{-y_2} dy_2 \\ du = 2y_2 dy_2 & v = -e^{-y_2}}} \\ &= -y_2^2 e^{-y_2} \Big|_0^{\infty} + \underbrace{\int_0^{\infty} 2y_2 e^{-y_2} dy_2}_{\substack{u = 2y_2 & dv = e^{-y_2} dy_2 \\ du = 2dy_2 & v = -e^{-y_2}}} \\ &= -2y_2 e^{-y_2} \Big|_0^{\infty} + \int_0^{\infty} 2e^{-y_2} dy_2 \\ &= -2e^{-y_2} \Big|_0^{\infty} = 2 \end{aligned}$$

Modelos de Probabilidad Conjunta Continuos

$$E(Y_1 Y_2) = \int_0^{\infty} \underbrace{\int_{y_2}^{\infty} y_1 y_2 e^{-y_1} dy_1}_{u = y_1 y_2 \quad dv = e^{-y_1} dy_1} dy_2$$

$$du = y_2 dy_1 \quad v = -e^{-y_1}$$

$$= \int_0^{\infty} \left[-y_1 y_2 e^{-y_1} \Big|_{y_2}^{\infty} + \int_{y_2}^{\infty} y_2 e^{-y_1} dy_1 \right] dy_2$$

$$= \int_0^{\infty} \{ y_2^2 e^{-y_2} + [-y_2 e^{-y_1}]_{y_2}^{\infty} \} dy_2$$

$$= \int_0^{\infty} (y_2^2 e^{-y_2} + y_2 e^{-y_2}) dy_2$$

$$= 2 + 1 = 3$$

$$\therefore V(Y_1) = 6 - 2^2 = 2 \quad V(Y_2) = 2 - 1^2 = 1$$

$$\therefore \text{Cov}(Y_1, Y_2) = E(Y_1 Y_2) - E(Y_1)E(Y_2) = 3 - 2 * 1 = 1$$

$$\therefore V(Y_1 - Y_2) = 2 + 1 - 2(1) = 1$$

Modelos de Probabilidad Conjunta Continuos

Esto es, se tiene una desviación estándar de minuto. De acuerdo a Chebyshev, al menos el 75% de los tiempos estarán en el intervalo $(2 - 2(1), 2 + 2(1)) = (0, 4)$: *pero por las reglas empíricas*, sabemos que la probabilidad contenida en dos desviaciones estándar alrededor de la media es del orden del 90 y tantos %. La pregunta sería, esa variación se puede reducir sin aumentar los costos, ni para ellos, ni para los clientes?. O, podemos vivir con ella, sin mayores consecuencias?

Nota: El coeficiente de correlación $\rho = \frac{1}{\sqrt{2}\sqrt{1}} = 0.71$

De qué otra forma podemos medir dependencias? Bueno, una forma no siempre reconocida como tal, pero ampliamente usada en lo que a sus consecuencias se refiere, es mediante los valores medios asociados a las distribuciones condicionales:

Modelos de Probabilidad Conjunta Continuos

Definición

Las distribuciones **condicionales** de $X|Y = y$ y $Y|X = x$ denotadas por $f_{X|y}(x)$ y $f_{Y|x}(y)$ se definen de la siguiente forma.

Para cada valor de x y y

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

$f_{X|y}(x)$ y $f_{Y|x}(y)$ son funciones de densidad en X y Y , respectivamente. Sus valores pueden depender (en el sentido de función matemática, de y y x , respectivamente, pero como los valores son dados no representan un factor aleatorio).

$$\int_{-\infty}^{+\infty} f_{X|y}(x) dx = \int_{-\infty}^{+\infty} \frac{f(x, y)}{f(y)} dx = \frac{1}{f(y)} \int_{-\infty}^{+\infty} f(x, y) dx = \frac{1}{f(y)} f(y) = 1$$

Modelos de Probabilidad Conjunta Continuos

Nota: de aquí podemos ver cómo se deriva el concepto de independencia:

Y no modifica su comportamiento por saber como se comporta X :

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

Si X y Y son independientes, entonces

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

Del ejemplo del “fast food”, si se quiere conocer cuál sería el comportamiento de los tiempos que el cliente toma en la fila, dado que el tiempo total debería ser de 2 minutos, tendríamos:

$$f_{Y_2|Y_1}(y_2) = \frac{f(y_1, y_2)}{f(y_1)} = \frac{e^{-y_1}}{y_1 e^{-y_1}} = \frac{1}{y_1} \quad 0 \leq y_2 \leq y_1 \leq \infty$$

Notas:

- 1) si existen condiciones en el rango de la variable, éstas deben aparecer con las condicionales. No así cuando calculamos marginales, pues en ese caso, se *barre* sobre la otra variable, mientras que aquí, *condicionas* con la otra variable.
- 2) como recomendación general, calcula primera la condicional en forma genérica, y después evalúa en los valores particulares.

Modelos de Probabilidad Conjunta Continuos

Como $y_1 = 2$, sólo evaluamos y obtenemos la distribución condicional para estos tiempos :

$$f_{Y_2|Y_1=2}(y_2) = \frac{1}{2} \quad 0 \leq y_2 \leq 2$$

Esto es, la distribución resultante es una $U(0, 2)$. *(No siempre resultan distribuciones conocidas)*

Bajo estas circunstancias es muy sencillo obtener su valor medio,
 $\mu_{Y_2|Y_1=2} = 1$ *(1 minuto en promedio, dado que tarda exactamente 2 minutos, en tiempo total)*

Modelos de Probabilidad Conjunta Continuos

La notación para los valores esperados calculados a partir de las condicionales es:

- $E(Y|X = x) = \int_{-\infty}^{+\infty} yf_{Y|x}(y)dy$, o bien
- $E(X|Y = y) = \int_{-\infty}^{+\infty} xf_{X|y}(x)dx$

A $E(Y|X = x)$ **Se le conoce como la Regresión de Y sobre X.**

Cuando $E(Y|X = x)$ es una función lineal, entonces se le llama la regresión lineal (teórica) de Y sobre X.

Cuando planteamos nuestros modelos, tales como $y = \mu_y + e$, , deberíamos entender

$$\mu_y = E(Y|X = x)$$

En general es difícil obtener las distribuciones condicionales y de allí, sus valores esperados, los cuales además, casi nunca son funciones lineales de las x 's. Entonces lo que planteamos a través de modelos como estos es simplemente una forma de aproximar al comportamiento medio de Y dada la información en X (que puede ser más de una variable).

Modelos de Probabilidad Conjunta Continuos

Un ejemplo más: Sean X y Y dos variables aleatorias con función de densidad conjunta:

$$f(x, y) = \begin{cases} kx(y - x) & 0 < x < 1 \quad -x < y < x \\ 0 & \text{en otra parte} \end{cases}$$

a) Encontrar el valor de la constante k .

$$\begin{aligned} \Rightarrow 1 &= \int_0^1 \int_{-x}^x kx(y - x) dy dx = \int_0^1 kx \left(\frac{y^2}{2} - xy \right) \Big|_{-x}^x dx \\ &= \int_0^1 kx \left[\left(\frac{x^2}{2} - x^2 \right) - \left(\frac{x^2}{2} + x^2 \right) \right] dx \\ &= \int_0^1 kx \left(\frac{-x^2}{2} - \frac{3x^2}{2} \right) dx = \int_0^1 -2kx(x^2) dx = \int_0^1 -2kx^3 dx \\ &= -2k \left[\frac{x^4}{4} \right]_0^1 = -\frac{2k}{4} = -\frac{k}{2} = 1. \quad \text{por lo tanto } k = -2 \end{aligned}$$

Y por lo tanto la función de densidad es:

$$f(x, y) = \begin{cases} -2x(y - x) & 0 < x < 1 \quad -x < y < x \\ 0 & \text{en otra parte} \end{cases}$$

o equivalentemente,

$$f(x, y) = \begin{cases} 2x(x - y) & 0 < x < 1 \quad -x < y < x \\ 0 & \text{en otra parte} \end{cases}$$

Modelos de Probabilidad Conjunta Continuos

Nota sobre marginales:

Para el caso de más de dos variables se define la marginal en forma análoga. Por ejemplo, si tenemos tres variables aleatorias X , Y y Z con función de probabilidad o de densidad $f(x, y, z) \Rightarrow \sum_y \sum_z f(x, y, z)$ ó $\int_y \int_z f(x, y, z) dx dy$ según sea el caso. Aquí es posible definir también la marginal conjunta de dos de las tres variables, por ejemplo X y Y . En ese caso la marginal conjunta se define como:

$$f_{X,Y}(x, y) = \sum_z f(x, y, z) \quad \text{o} \quad \int_z f(x, y, z) dz$$

Estamos listos para ver dos modelos particulares de los cuales encontramos multitud de aplicaciones: El modelo Multinomial y la Normal Bivariada.

El modelo Multinomial

La distribución multinomial es una generalización de la distribución binomial. Este tipo de modelos es adecuado para encuestas de opinión, investigación de mercados, inspección industrial, etc. Además tiene otros muchos usos para generar pruebas estadísticas como las de bondad de ajuste, entre otras.

El modelo Multinomial

Un experimento multinomial consiste en:

1. n pruebas idénticas;
2. El resultado de cada observación puede ser clasificado en una de k -clases o casillas;
3. La probabilidad de ser clasificado en la casilla i es p_i $i = 1, 2, \dots, k$, y permanece constante de observación a observación. Más aún $p_1 + p_2 + \dots + p_k = 1$;
4. Las pruebas son independientes;
5. Las variables aleatorias son X_1, X_2, \dots, X_k donde $X_i = \#$ de observaciones en la casilla i , $i = 1, 2, \dots, k$. Nótese que $X_1 + X_2 + \dots + X_k = n$.

El modelo Multinomial

La función de probabilidad conjunta de X_1, X_2, \dots, X_k está dada por

$$p(x_1, x_2, \dots, x_k) = \frac{n!}{x_1! x_2! \dots x_k!} p_1^{x_1} \cdot p_2^{x_2} \cdots p_k^{x_k}$$

donde $\sum_{i=1}^k p_i = 1$ y $\sum_{i=1}^k x_i = n$.

Nota: Una generalización para multinomios del desarrollo del binomio de Newton está dada por

$$\sum_{x_1} \sum_{x_2} \cdots \sum_{x_k} \frac{n!}{x_1! x_2! \dots x_k!} p_1^{x_1} \cdot p_2^{x_2} \cdots p_k^{x_k} = (p_1 + p_2 + \dots + p_k)^n = 1^n = 1.$$

Esto asegura que la suma de la función de probabilidad sobre todos los valores posibles de las variables es 1.

Resultado

La distribución marginal de X_i , para $i = 1, 2, \dots, k$, es *Binomial*(n, p_i).

Si sólo queremos la distribución de X_i , i.e. el $\#$ de éxitos en la casilla i , podemos pensar que sólo hay dos tipos de casillas: i o no i , con p_i como la probabilidad asociada a la casilla i y $1 - p_i$ como la probabilidad asociada a las casillas complementarias.

Para ilustrar conceptos, la demostración se hará sólo para el caso $k = 3$. El resultado es intuitivamente fácil de entender y manejar.

El modelo Multinomial

Sea (X, Y, Z) un vector aleatorio con probabilidad conjunta:

$$f(x, y, z) = \frac{n!}{x!y!z!} p_X^x p_Y^y p_Z^z,$$

con $x + y + z = n$. Entonces, tenemos que

$$x = 0, 1, \dots, n;$$

$$\Rightarrow y = 0, 1, \dots, n - x;$$

$$\Rightarrow z = 0, 1, \dots, n - x - y.$$

El modelo Multinomial

Ahora

$$\begin{aligned}f_X(X) &= \sum_{y=0}^{n-x} \sum_{z=0}^{n-x-y} \frac{n!}{x!y!z!} p_X^x p_Y^y p_Z^z \\&= \sum_{y=0}^{n-x} \frac{n!}{x!y!(n-x-y)!} p_X^x p_Y^y (1-p_X-p_Y)^{n-x-y} \\&= \frac{n!}{x!(n-x)!} p_X^x (1-p_X)^{n-x} \sum_{y=0}^{n-x} \frac{(n-x)!}{y!(n-x-y)!} p_Y^y \frac{(1-p_X-p_Y)^{n-x-y}}{(1-p_X)^{n-x}} \\&= \frac{n!}{x!(n-x)!} p_X^x (1-p_X)^{n-x} \cdot \\&\quad \underbrace{\sum_{y=0}^{n-x} \frac{(n-x)!}{y!(n-x-y)!} \left(\frac{p_Y}{1-p_X}\right)^y \left(1 - \frac{p_Y}{1-p_X}\right)^{n-x-y}}_{\text{1 por ser la suma de una v.a Binomial } \left(n-x, \frac{p_Y}{1-p_X}\right)} \\&= \frac{n!}{x!(n-x)!} p_X^x (1-p_X)^{n-x} \\&= \binom{n}{x} p_X^x (1-p_X)^{n-x}.\end{aligned}$$

El modelo Multinomial

El mismo argumento se utiliza para X y Z .

De este resultado se sigue entonces que

$$E(X_i) = np_i, \quad V(X_i) = np_i q_i.$$

Más aún, notemos que el número de resultados por casilla no son independientes y

$$Cov(X_i, X_j) = -np_i p_j$$

Para explicar esto, usaremos los siguientes argumentos.

El modelo Multinomial

1. Supongamos que $X = (X_1, \dots, X_K)$ una variable aleatoria multinomial con distribución

$$P(X = x) = P(X_1 = x_1, \dots, X_K = x_K) = \frac{n!}{x_1! \cdots x_K!} p_1^{x_1} \cdots p_K^{x_K}$$

donde $x_1 + \cdots + x_K = n$.

2. Consideremos el caso específico de una variable multinomial como la de arriba, pero con una sola tirada ($n = 1$):

$$P(W = w) = P(W_1 = w_1, \dots, W_K = w_K) = p_1^{w_1} \cdots p_K^{w_K}$$

donde $w_1 + \cdots + w_K = 1$. Como es una sola tirada, es claro que si, por ejemplo, ocurre la primera casilla ($w_1 = 1$) entonces todas las otras casillas son cero. También, cada casilla (i.e. W_i) es una Bernoulli, con probabilidad de éxito igual a p_i , o sea que $E(W_i) = p_i$ y $\text{Var}(W_i) = p_i(1 - p_i)$.

El modelo Multinomial

Además, la variable W además cumple:

- Media de W (usando notación de vectores como vectores columna)

$$E(W) = E \begin{bmatrix} W_1 \\ \vdots \\ W_K \end{bmatrix} = \begin{bmatrix} E(W_1) \\ \vdots \\ E(W_K) \end{bmatrix} = \begin{bmatrix} p_1 \\ \vdots \\ p_K \end{bmatrix}$$

- Varianza de W (i.e. la matriz de varianzas y covarianzas) (o, simplemente, la matriz de covarianza de W)

$$\text{Var}(W) = \begin{bmatrix} \text{Var}(W_1) & \text{Cov}(W_1, W_2) & \cdots & \text{Cov}(W_1, W_K) \\ \text{Cov}(W_2, W_1) & \text{Var}(W_2) & \cdots & \text{Cov}(W_2, W_K) \\ \vdots & \vdots & \ddots & \vdots \\ \text{Cov}(W_K, W_1) & \text{Cov}(W_K, W_2) & \cdots & \text{Var}(W_K) \end{bmatrix}.$$

El modelo Multinomial

Los elementos diagonales de la matriz anterior sabemos que son $\text{Var}(W_i) = p_i(1 - p_i)$. Ahora, los elementos fuera de la diagonal son

$$\text{Cov}(W_i, W_j) = E(W_i W_j) - E(W_i)E(W_j) = 0 - p_i p_j.$$

La razón por la cual $E(W_i W_j) = 0$ es la siguiente: el producto $W_i W_j$ sólo puede tomar el valor 0, pues (W_i, W_j) puede valer $(0,0)$, $(1,0)$, $(0,1)$ y en cada caso $W_i W_j = 0$, pero nunca puede ser $(1,1)$, pues se trata de una sola tirada.

Entonces

$$\text{Var}(W) = \begin{bmatrix} p_1(1 - p_1) & -p_1 p_2 & \cdots & -p_1 p_K \\ -p_2 p_1 & p_2(1 - p_2) & \cdots & -p_2 p_K \\ \vdots & \vdots & \ddots & \vdots \\ -p_K p_1 & -p_K p_2 & \cdots & p_K(1 - p_K) \end{bmatrix}.$$

El modelo Multinomial

3. Finalmente, si X es una variable aleatoria multinomial con n tiradas (independientes), entonces X se puede escribir como la suma de n multinomiales de una sola tirada

$$X = Z_1 + Z_2 + \cdots + Z_n$$

donde cada Z_j es como la W del punto anterior.

Entonces,

- Media de X

$$\begin{aligned} E(X) &= E(Z_1 + \cdots + Z_n) = E(Z_1) + \cdots + E(Z_n) \\ &= \begin{bmatrix} p_1 \\ \vdots \\ p_K \end{bmatrix} + \cdots + \begin{bmatrix} p_1 \\ \vdots \\ p_K \end{bmatrix} = \begin{bmatrix} np_1 \\ \vdots \\ np_K \end{bmatrix}. \end{aligned}$$

- Varianza de X

$$\text{Var}(X) = \text{Var}(Z_1 + \cdots + Z_n) = \text{Var}(Z_1) + \cdots + \text{Var}(Z_n)$$

(varianza de una suma es la suma de las varianzas bajo independencia), entonces

$$\text{Var}(X) = \begin{bmatrix} np_1(1-p_1) & -np_1p_2 & \cdots & -np_1p_K \\ -np_2p_1 & np_2(1-p_2) & \cdots & -np_2p_K \\ \vdots & \vdots & \ddots & \vdots \\ -np_Kp_1 & -np_Kp_2 & \cdots & np_K(1-p_K) \end{bmatrix}.$$

El modelo Multinomial

Para demostrar esto, recordemos que cada una de las observaciones o pruebas sí son independientes y definamos

$$U_s = \begin{cases} 1 & \text{si la prueba } s \text{ cae en la casilla } i \\ 0 & \text{otro caso} \end{cases}$$

$$W_k = \begin{cases} 1 & \text{si la prueba } k \text{ cae en la casilla } j \\ 0 & \text{en otro caso} \end{cases}$$

Es claro que $E(U_s) = p_i$; $E(W_k) = p_j$ y

$$\text{Cov}(U_s, W_k) = \begin{cases} 0 & s \neq k \\ -p_i p_j & s = k \end{cases}$$

El modelo Multinomial

Entonces $X_i = \sum_{s=1}^n U_s$; $X_j = \sum_{k=1}^n W_k$, y

$$\begin{aligned} \text{Cov}(X_i, X_j) &= \text{Cov}\left(\sum_{s=1}^n U_s, \sum_{k=1}^n W_k\right) \\ &= \sum_{s=1}^n \sum_{k=1}^n \text{Cov}(U_s, W_k) \\ &= \sum_{s=1}^n \text{Cov}(U_s, W_s) + \sum_{s \neq k} \sum_{k=1}^n \text{Cov}(U_s, W_k) \\ &= \sum_{s=1}^n (E(U_s W_s) - E(U_s)E(W_s)) \\ &= \sum_{s=1}^n (0 - p_i p_j) = -n p_i p_j \end{aligned}$$

$U_s W_s = 0$ siempre, de su definición.

- Por definición, la función generatriz de momentos del vector $X = (X_1, X_2, \dots, X_K)$ está dada por

$$\begin{aligned}M_{X_1, X_2, \dots, X_K}(t) &= E(e^{t'X}) \\ &= E(e^{t_1X_1 + t_2X_2 + \dots + t_KX_K}),\end{aligned}$$

para $t = (t_1, \dots, t_K)$.

- Notemos que t tiene la misma dimensión que $X = (X_1, \dots, X_K)$.
- También es factible tomar derivadas cruzadas, las cuales al evaluarse en cero nos dan momentos cruzados que sirven para formar las covarianzas, esto es, nos dan $E(X^i X^j)$.

El modelo Multinomial

La generatriz de momentos de la Multinomial, usando el teorema de Newton para multinomio

$$\begin{aligned}M_{X_1, X_2, \dots, X_K}(t_1, t_2, \dots, t_K) &= E(e^{t_1 X_1 + t_2 X_2 + \dots + t_K X_K}) \\&= (p_1 e^{t_1} + p_2 e^{t_2} + \dots + p_k e^{t_k})^n\end{aligned}$$

Así, por ejemplo, si tomamos

$$\begin{aligned}M_{X_1, X_2, \dots, X_K}(t_1, 0, \dots, 0) &= (p_1 e^{t_1} + p_2 e^0 + \dots + p_k e^0)^n \\&= (p_1 e^{t_1} + p_2 + \dots + p_k)^n \\&= (p_1 e^{t_1} + (1 - p_1))^n \\&= (p_1 e^{t_1} + q_1)^n,\end{aligned}$$

que es la función generatriz de momentos de una *Binomial*(n, p_1). Esto certifica el hecho de que las marginales son binomiales.

También se pueden obtener otros resultados , por ejemplo

$$\begin{aligned}M_{X_1, X_2, \dots, X_K}(t, t, \dots, 0) &= (p_1 e^t + p_2 e^t + p_3 \dots + p_k)^n \\ &= ((p_1 + p_2) e^t + (1 - p_1 - p_2))^n\end{aligned}$$

la cual corresponde a la función generatriz de momentos de una v.a. *Binomial*($n, p_1 + p_2$).

Evaluar en la misma t , equivale a juntar las clasificaciones, a sumar las variables. En la Teoría de la Probabilidad esto implica la convolución de las variables aleatorias.

El modelo Multinomial

Ejemplo

Se fabrica una barra de un largo específico. Suponiendo que el largo verdadero X (en pulgadas) sigue una distribución uniforme $[10, 12]$. Si se fabrican 10 de tales barras, calcular la probabilidad de obtener 5 de longitud menor a 10.5 pg. y 2 de longitud mayor que 11.8 pg.

El modelo Multinomial

Ejemplo

Se fabrica una barra de un largo específico. Suponiendo que el largo verdadero X (en pulgadas) sigue una distribución uniforme $[10, 12]$. Si se fabrican 10 de tales barras, calcular la probabilidad de obtener 5 de longitud menor a 10.5 pg. y 2 de longitud mayor que 11.8 pg.

Definamos:

$$A_1 = \{X > 10.5\}, A_2 = \{10.5 \leq X \leq 11.8\}, A_3 = \{X > 11.8\}$$

y las variables

$X_1 = \#$ de barras de la clase A_1

$X_2 = \#$ de barras de la clase A_2

$X_3 = \#$ de barras de la clase A_3

El modelo Multinomial

De aquí $p_1 = 0.25$, $p_2 = 0.65$ y $p_3 = 0.1$. Estas probabilidades se calculan a partir de la distribución Uniforme que modela la longitud de las barras.

Por ejemplo, para p_1 tenemos que

$$p_1 = \int_{10}^{10.5} \frac{1}{2} dx = \left[\frac{1}{2}x \right]_{10}^{10.5} = \frac{1}{2}(10.5 - 10) = 0.5 \frac{1}{2} = 0.25.$$

Ahora obtenemos

$$P(X_1 = 5, X_2 = 3, X_3 = 2) = \frac{10!}{5!3!2!} (0.25)^5 (0.65)^3 (0.1)^2 = 0.0067583.$$

El modelo Multinomial

Una propiedad importante de la distribución Multinomial es que, *cualquier variable aleatoria, continua o discreta* puede ser reducida a este modelo, particionando su rango en k intervalos (sin traslape). Por ejemplo, si X es una v. a. continua:

$$(-\infty, c_1); [c_1, c_2); [c_2, c_3); \dots [c_{k-1}, \infty)$$

con $p_i = \int_{c_{i-1}}^{c_i} f_X(x) dx$ $i = 1, 2, \dots, k$ donde $c_0 = -\infty$ y $c_k = \infty$.

Esto es útil para la pruebas de Bondad de Ajuste.

Ejercicio

La tasa de muertes por accidente de tráfico de un año reciente (fatalidades por 100 millones de millas recorridas) se listan para cada uno de los 50 estados de los Estados Unidos. Construyeron un histograma de estas observaciones usando como clases 2.0-2.9, 3.0-3.9, etc. Suponiendo que la distribución de estas tasas puede ser aproximada por una distribución normal, determinaron las frecuencias esperadas para cada una de las clases. (Con $\mu = 5.3$ y $\sigma = 1.3$).

El modelo Multinomial

Tasa de Muertes	Frecuencia Observada	Frecuencia Esperada
< 2.9	1	1.8
3.0-3.9	4	5.7
4.0-4.9	16	12.2
5.0-5.9	15	14.9(*)
6.0-6.9	7	10.3
7-0-7.9	4	4.1
> 8.0	3	1.0
		$\Sigma = 50$

El modelo Multinomial

- Cada intervalo en la tabla es una de las clasificaciones posibles, las cuales son mutuamente excluyentes.
- Las 50 observaciones se asumen independientes en el sentido de que una no debería dar información sobre donde se clasificarán las otras.
- Las probabilidades se obtienen de la distribución que se asume que siguen los datos. Por ejemplo, para calcular la frecuencia esperada (*) se tiene que:

$$\begin{aligned}P(5 < X < 5.9) &\approx P\left(\frac{4.95 - 5.3}{1.3} \leq Z \leq \frac{5.95 - 5.3}{1.3}\right) \\&= P(-0.27 \leq Z \leq 0.50) \\&\approx P(Z \leq 0.50) - P(Z \leq -0.27) \\&= 0.6915 - 0.3936 = 0.2979,\end{aligned}$$

$$\Rightarrow (0.2979) \cdot 50 = 14.895 = 14.9.$$

El modelo Multinomial

Como tenemos las frecuencias observadas para cada clasificación, lo que se hace ahora es compararlas, por ejemplo se usa el criterio siguiente:

$$\sum_{i=1}^k \frac{(\text{frec. observada} - \text{frec. esperada})^2}{\text{frec. esperada}},$$

esto nos da una medida de la discrepancia que nos permite decir si es suficientemente grande como para creer que los datos no puede provenir de esa distribución.

Estas ideas del cómo tomar decisiones las discutiremos más a fondo en el último capítulo, cuando hablemos de Pruebas de Hipótesis y alguno de ustedes expondrá el tema de Bondad de Ajuste.

El modelo Normal Bivariado

El modelo normal bivariado

Si $X \sim N(\mu_x, \sigma_x^2)$ y $Y \sim N(\mu_y, \sigma_y^2)$ son independientes, entonces

$$\begin{aligned}f_{(X,Y)}(x,y) &= f_X(x)f_Y(y) \\&= \frac{1}{\sqrt{2\pi}\sigma_x} e^{-\frac{1}{2}\left(\frac{x-\mu_x}{\sigma_x}\right)^2} \frac{1}{\sqrt{2\pi}\sigma_y} e^{-\frac{1}{2}\left(\frac{y-\mu_y}{\sigma_y}\right)^2} \\&= \frac{1}{2\pi\sigma_x\sigma_y} e^{-\frac{1}{2}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2 + \left(\frac{y-\mu_y}{\sigma_y}\right)^2\right]},\end{aligned}$$

este es un caso particular de la distribución conjunta Normal Bivariada (caso particular).

El modelo normal bivariado

Recordemos que si X_1, X_2, \dots, X_n son variables aleatorias independientes (una muestra aleatoria, por ejemplo) y definimos $Y = \sum_{i=1}^n X_i$, entonces

$$\begin{aligned} M_Y(t) &= E(e^{tY}) = E(e^{t(X_1+X_2+\dots+X_n)}) \\ &= E(e^{tX_1+tX_2+\dots+tX_n}) \\ &= E(e^{tX_1} \cdot e^{tX_2} \dots e^{tX_n}) \\ &= E(e^{tX_1})E(e^{tX_2}) \dots E(e^{tX_n}) \text{ por independencia} \\ &= M_{X_1}(t)M_{X_2}(t) \dots M_{X_n}(t) \end{aligned}$$

Más aún, si X_1, X_2, \dots, X_n tienen la misma distribución

$$M_Y(t) = [M_X(t)]^n$$

El modelo normal bivariado

Nota: Además es fácil verificar que si $W = aX + bY$ entonces, (una combinación lineal de dos variables normales independientes)

$$\begin{aligned}M_W(t) &= E[e^{(aX+bY)t}] = E(e^{atX})E(e^{btY}) = M_X(at) \cdot M_Y(bt) \\&= [e^{a\mu_x t + \frac{1}{2}(a^2\sigma_x^2)t^2}][e^{b\mu_y t + \frac{1}{2}(b^2\sigma_y^2)t^2}] \\&= e^{(a\mu_x + b\mu_y)t + \frac{1}{2}(a^2\sigma_x^2 + b^2\sigma_y^2)t^2},\end{aligned}$$

i.e. $W \sim \text{Normal}(a\mu_x + b\mu_y, a^2\sigma_x^2 + b^2\sigma_y^2)$.

Esto quiere decir que cualquier combinación lineal de variables aleatorias normales independientes también sigue una distribución normal.

No sólo somos capaces de saber cuál es la media y la varianza de la suma, sino que además nos permite conocer su distribución!!. Este es un resultado particularmente útil para las **distribuciones muestrales**.

El modelo normal bivariado

Ejemplo

Las resistencias de ciertos resistores tienen una distribución normal con media de 100 ohms y desviación estándar de 10 ohms. Dos de tales resistencias se conectan en serie resultando que la resistencia total del circuito es la suma de las resistencias individuales. Determinar la probabilidad de que la resistencia total sea mayor que 220 ohms.

Solución: Tenemos que

$$R_i \sim N(100, 10) \quad i = 1, 2 \text{ son independientes}$$

y, por tanto,

$$M_{R_i(t)} = e^{100t + \frac{1}{2}100t^2}.$$

Entonces $R_T = R_1 + R_2$ y

$$M_{R_T}(t) = (e^{100t + \frac{1}{2}100t^2})^2 = e^{200t + \frac{1}{2}200t^2},$$

lo que nos dice que $R_T \sim N(200, \sigma_{R_T}^2 = 200)$. Esto implica

$$P(R_T > 220) = P(Z > \frac{220 - 200}{\sqrt{200}}) = P(Z > 1.41) = 0.0793.$$

Lectura: Ver Montgomery D.C.(1985) Introduction to Statistical Quality Control, pag. 396-403

Ejercicio

En un área determinada, cierto material es seleccionado al azar y pesado en dos tiempos distintos. Sea W = peso del material (verdadero) y X_1, X_2 las dos mediciones hechas. Uno puede pensar estas mediciones como

$$X_1 = W + e_1$$

$$X_2 = W + e_2$$

donde e_1 y e_2 son los errores de medición. Supongamos que e_1 y e_2 son independientes entre sí e independientes de W , y que $V(e_1) = V(e_2) = \sigma_e^2$. expresar a ρ , el coeficiente de correlación entre X_1 y X_2 , en términos de σ_W^2 (la varianza del peso verdadero) y σ_e^2 , interpreta el resultado.

El modelo normal bivariado

Ahora consideraremos el caso siguiente: X y Y no son variables independientes pero (X, Y) tiene una distribución **Normal Bivariada**.

En este caso la función de densidad conjunta de las variables es

$$f_{X,Y}(x,y) = \frac{\exp\left(-\frac{1}{2(1-\rho^2)}\left\{\left(\frac{x-\mu_x}{\sigma_x}\right)^2 - 2\rho\left(\frac{x-\mu_x}{\sigma_x}\right)\left(\frac{y-\mu_y}{\sigma_y}\right) + \left(\frac{y-\mu_y}{\sigma_y}\right)^2\right\}\right)}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}}.$$

El modelo normal bivariado

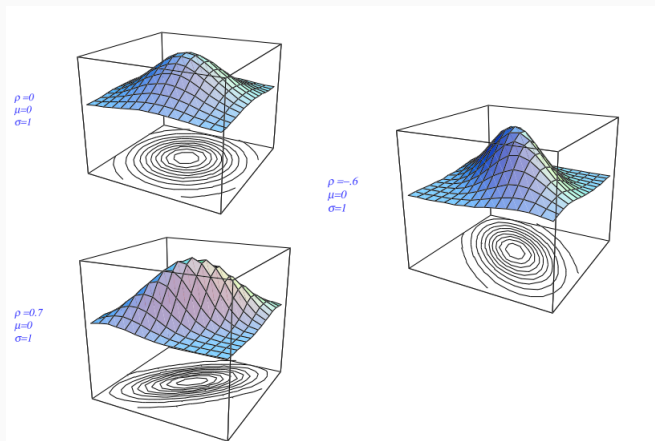


Figure 1: Gráficas de Normales Bivariadas, con dependencia cero, positiva y negativa.

El modelo normal bivariado

- Si cada una sigue una distribución normal, y no son independientes, **no podemos** asegurar que la distribución conjunta sea normal bivariada.
- De la expresión anterior es inmediato verificar que si $\rho = 0$ entonces

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

y por lo tanto X y Y son independientes. Esto es, si X y Y tienen **una distribución normal bivariada y la covarianza (ó correlación) es cero** entonces X y Y deben ser independientes.

- Bajo normalidad bivariada, la **única forma de dependencia permitida es la lineal**.

Estos resultados parecen intrascendentes, pero son la base de la mayoría de los modelos que trabajamos en regresión o diseño de experimentos, y la razón de la popularidad de la distribución normal.

El modelo normal bivariado

La distribución marginal de X es

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{+\infty} f_{X,Y}(x,y) dy \\ &= \int_{-\infty}^{+\infty} \frac{e^{-\frac{1}{2(1-\rho^2)} \left\{ \left(\frac{x-\mu_x}{\sigma_x} \right)^2 + \left(\frac{y-\mu_y}{\sigma_y} \right)^2 - 2\rho \left(\frac{x-\mu_x}{\sigma_x} \right) \left(\frac{y-\mu_y}{\sigma_y} \right) \right\}}}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} dy \end{aligned}$$

Considerando las transformaciones $u = \frac{x-\mu_x}{\sigma_x}$ y $v = \frac{y-\mu_y}{\sigma_y} \Rightarrow dv = \frac{1}{\sigma_y} dy$, la notación se simplifica,

El modelo normal bivariado

$$\begin{aligned}f_X(x) &= \int_{-\infty}^{+\infty} \frac{1}{2\pi\sigma_x(1-\rho^2)} e^{-\frac{1}{2(1-\rho^2)}u^2} e^{-\frac{1}{2(1-\rho^2)}\{v^2-2\rho uv\}} dv \\&= \frac{1}{\sqrt{2\pi}\sigma_x} e^{-\frac{u^2}{2(1-\rho^2)}} \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\{(v-\rho u)^2-\rho^2 u^2\}} dv \\&\quad \text{(se completa el cuadrado en el exponente)} \\&= \frac{1}{\sqrt{2\pi}\sigma_x} e^{-\frac{1}{2}\left(\frac{u^2}{(1-\rho^2)}-\frac{\rho^2 u^2}{(1-\rho^2)}\right)} \underbrace{\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}\sqrt{1-\rho^2}} e^{-\frac{1}{2}\frac{(v-\rho u)^2}{1-\rho^2}} dv}_{N(\rho u, (1-\rho^2))} \\&\quad \text{(se reconoce la distribución para evitar la integración)} \\&= \frac{1}{\sqrt{2\pi}\sigma_x} e^{-\frac{1}{2}u^2} \\&= \frac{1}{\sqrt{2\pi}\sigma_x} e^{-\frac{1}{2}\left(\frac{x-\mu_x}{\sigma_x}\right)^2}\end{aligned}$$

$$\therefore X \sim N(\mu_x, \sigma_x^2)$$

Similarmente, $f_Y(y) = \frac{1}{\sqrt{2\pi}\sigma_y} e^{-\frac{1}{2}\left(\frac{y-\mu_y}{\sigma_y}\right)^2} \therefore Y \sim N(\mu_y, \sigma_y^2)$

En resumen

Normal Bivariada \implies Marginales Normales,

pero Marginales normales **no implican** Normal Bivariada (a menos que sean variables aleatorias normales independientes)

Ahora, consideremos la distribución condicional de Y dado X

$$\begin{aligned}f_{Y|X}(y|x) &= \frac{f_{X,Y}(x,y)}{f_X(x)} = \frac{\frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} e^{-\frac{u^2}{2(1-\rho^2)}} e^{-\frac{1}{2(1-\rho^2)}\{v^2-2\rho uv\}}}{\frac{1}{\sqrt{2\pi}\sigma_x} e^{-\frac{1}{2}u^2}} \\&= \frac{1}{\sqrt{2\pi}\sigma_y\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\{v^2-2\rho uv+\rho^2 u^2\}} \\&= \frac{1}{\sqrt{2\pi}\sigma_y\sqrt{1-\rho^2}} e^{-\frac{1}{2}\left[\frac{v-\rho u}{\sqrt{1-\rho^2}}\right]^2}.\end{aligned}$$

El modelo normal bivariado

Es decir,

$$f_{Y|X}(y|x) = \frac{\exp \left\{ -\frac{1}{2\sigma_Y^2(1-\rho^2)} \left[y - \left\{ \mu_Y + \rho \frac{\sigma_Y}{\sigma_X} (x - \mu_X) \right\} \right]^2 \right\}}{\sqrt{2\pi}\sigma_Y\sqrt{1-\rho^2}},$$

(truco algebraico para reconocer la distribución)

lo cual implica que

$$(Y|X = x) \sim N \left(\mu_{Y|X} = \mu_Y + \rho \frac{\sigma_Y}{\sigma_X} (x - \mu_X), \sigma_{Y|X}^2 = \sigma_Y^2 (1 - \rho^2) \right),$$

y, por tanto,

$$E(Y|X = x) = \rho \frac{\sigma_Y}{\sigma_X} x + \left(\mu_Y - \rho \frac{\sigma_Y}{\sigma_X} \mu_X \right).$$

El modelo normal bivariado

Para $t = (t_1, t_2)$, la función generatriz de momentos conjunta del vector $W = (X, Y)$ está dada por

$$\begin{aligned} M_{X,Y}(t_1, t_2) &= E(e^{t'W}) = E(e^{t_1x+t_2y}) \\ &= \exp \left\{ t_1\mu_x + t_2\mu_y + \frac{1}{2}(\sigma_x^2 t_1^2 + 2\rho\sigma_x\sigma_y t_1 t_2 + \sigma_y^2 t_2^2) \right\} \end{aligned}$$

Ejercicio: Demostrar la igualdad anterior y entregar su desarrollo junto con la Tarea 3.

Ejercicio

1. Determina qué variables aleatorias corresponden a las generatrices $M_{X,Y}(t, 0)$ y $M_{X,Y}(0, t)$
2. Si X y Y tienen una distribución normal bivariada, y $U = X + Y, W = X - Y$.
Determina el coeficiente de correlación de U y W .

El modelo normal bivariado

Ejemplo

La distribución de dos elementos de un tipo de material (en gramos) sigue una distribución normal bivariada, donde

$$\mu_x = 4, \mu_y = 6, \sigma_x^2 = \sigma_y^2 = 1 \text{ y } \text{Cov}(X, Y) = 0.8$$

¿Cuál es el valor esperado del elemento Y si se observan 6 grs. del elemento X ?

Solución:

$$\rho = \frac{\text{Cov}(X, Y)}{\sigma_x \sigma_y} = \frac{0.8}{1 \cdot 1} = 0.8$$

$$\begin{aligned} \Rightarrow E(Y|X) &= E(Y|X = 6) = \mu_y + \rho \frac{\sigma_y}{\sigma_x} (x - \mu_x) = 6 + 0.8\left(\frac{1}{1}\right)(6 - 4) = \\ &= 6 + 0.8(2) = 7.6 \text{ grs.} \end{aligned}$$

Más Funciones de variables aleatorias

Resultado 4: Muestra Aleatoria y Estadísticos de Orden.

Un concepto fundamental es el siguiente: al tomar una muestra aleatoria

$$X_1, X_2, \dots, X_n;$$

de alguna población, lo que se está considerando es **un conjunto de variables aleatorias i.i.d.**

Las características principales de una m.a. son:

- Cada v. a. está definida sobre la misma población, por lo tanto la distribución de cada una será exactamente la misma.
- Los posibles resultados de cada una no afectan los posibles resultados de las otras.

Ejemplo (Estadísticos de Orden)

Existen muchos casos en los que la cantidad importante es el valor menor de una m.a., que por el mismo carácter aleatorio de la m.a. también será una v. a. Equivalentemente, la cantidad importante puede ser el valor mayor de la m. a., y también es una v.a. Las dos variables aleatorias así definidas se llaman respectivamente

- primer estadístico de orden $X_{(1)}$,
- n-ésimo estadístico de orden $X_{(n)}$.

Estos estadísticos de orden resultan ser una transformación de variables aleatorias:

- $X_{(1)} = \min\{X_1, X_2, \dots, X_n\},$
- $X_{(n)} = \max\{X_1, X_2, \dots, X_n\}.$

Aplicaciones:

- Cuando un sistema deja de funcionar al fallar alguna de sus n componentes, la duración total puede ser vista como la duración más corta de cualquiera de sus componentes. Cada componente puede tener asignada una v.a. de duración X_i , entonces, podríamos definir como la v.a. de interés a $X_{(1)}$ = *el menor valor de una m.a. de tamaño n* ; en símbolos, $X_{(1)} = \min\{X_1, X_2, \dots, X_n\}$. (**SISTEMA EN SERIE**)
- Un sistema que deja de funcionar hasta que el último componente falla; aquí interesaría $X_{(n)}$ el valor mayor de una m.a. de tamaño n . En símbolos, $X_{(n)} = \max\{X_1, X_2, \dots, X_n\}$. (**SISTEMA EN PARALELO**)

Aplicaciones (continúa):

- En una prueba de resistencia de material de construcción, la variable crítica puede ser la máxima resistencia posible si se toma una m.a. de tamaño n (n mediciones). Nos interesaría, $X_{(n)}$ = *el mayor valor de la m. a. de tamaño n* ; en símbolos, $X_{(n)} = \max\{X_1, X_2, \dots, X_n\}$.
- También puede interesar el rango de la m.a., esto es $R = X_{(n)} - X_{(1)}$. Esta es una medición alternativa de la dispersión muy usada en la industria de la manufactura, la cual representa otra transformación de variables aleatorias.

Las densidades de $X_{(1)}$ y $X_{(n)}$ se deducen a continuación usando el método de la acumulada.

Densidad de $X_{(n)}$: Llamaremos $Y_n = X_{(n)}$, y a sus valores y $F_{Y_n}(y)$ a su función de distribución acumulada. Entonces,

$$F_{Y_n}(y) = P\{Y_n \leq y\} = P\{X_1 \leq y, X_2 \leq y, \dots, X_n \leq y\};$$

si Y_n es menor que y entonces cada uno de los valores de las X_i tienen que ser menor que y . Puesto que las X_i son independientes, la anterior probabilidad se calcula como el producto de las probabilidades de cada evento:

$$\begin{aligned} F_{Y_n}(y) &= \prod_{i=1}^n P(X_i \leq y) \quad (\text{independencia}) \\ &= [P\{X \leq y\}]^n \quad (\text{distribuciones idénticas}) \\ &= [F_X(y)]^n. \end{aligned}$$

Entonces,

$$\begin{aligned}f_{Y_n}(y) &= \frac{d}{dy} [F_X(y)]^n \\&= n[F_X(y)]^{n-1} \frac{d}{dy} [F_X(y)] \quad (\text{Regla de la Cadena}) \\&= n[F_X(y)]^{n-1} f_X(y). \quad (\text{Teorema Fundamental del Cálculo})\end{aligned}$$

Densidad de $X_{(1)}$: Llamaremos $Y_1 = X_{(1)}$, y a sus valores y $F_{Y_1}(y)$ a su función de distribución acumulada. Entonces,

$$F_{Y_1}(y) = P\{Y_1 \leq y\} = 1 - P\{Y_1 > y\} = 1 - P\{X_1 > y, X_2 > y, \dots, X_n > y\};$$

el valor más pequeño Y_1 es mayor que y cuando cada uno de los valores de las X_i lo sean. Puesto que las variables aleatorias son independientes, la anterior probabilidad se calcula como el producto de las probabilidades de cada evento:

$$\begin{aligned} F_{Y_1}(y) &= 1 - \prod_{i=1}^n P(X_i > y) && \text{(independencia)} \\ &= 1 - [P\{X > y\}]^n && \text{(distribuciones idénticas)} \\ &= 1 - [1 - F_X(y)]^n. && \text{(complemento)} \end{aligned}$$

Después,

$$\begin{aligned}f_{Y_1}(y) &= \frac{d}{dy}[1 - [1 - F_X(y)]^n] \\&= -n[1 - F_X(y)]^{n-1} \frac{d}{dy}[1 - F_X(y)] \quad (\text{Regla de la Cadena}) \\&= -n[1 - F_X(y)]^{n-1}[-f_X(y)] \quad (\text{Teorema Fundamental del Cálculo}) \\&= n[1 - F_X(y)]^{n-1}f_X(y).\end{aligned}$$

Ejemplo

Considera una m.a. X_1, X_2, \dots, X_n de tamaño n , tomada de una población $Unif(0, 1)$. Calcula las densidades de los estadísticos de orden primero y n -ésimo. Recordando que

$$f_X(x) = \begin{cases} 1 & 0 < x < 1 \\ 0 & \text{c.o.p.} \end{cases} \quad F_X(x) = \begin{cases} 0 & x < 0 \\ x & 0 < x < 1 \\ 1 & x > 1 \end{cases}$$

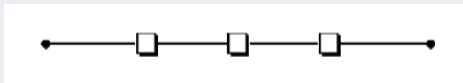
Lo que implica,

$$f_{Y_n}(y) = n[F_X(y)]^{n-1}f_X(y) = ny^{n-1} \cdot 1 \quad \text{para } 0 < y < 1,$$

$$f_{Y_1}(y) = n[1 - F_X(y)]^{n-1}f_X(y) = n(1 - y)^{n-1} \cdot 1 \quad \text{para } 0 < y < 1.$$

Ejemplo (Sistema en Serie)

El siguiente sistema en serie funcionará siempre que cada una de las componentes (idénticas) mostradas fallen, cada componente funciona independientemente de las otras. Supóngase que X = tiempo de vida de cada componente del sistema y que $X \sim \exp(\beta)$. Determinar el comportamiento del tiempo de vida del sistema, es decir, de $Y_1 = \min(X_1, X_2, X_3)$.



Recordemos que

$$f_X(x) = \begin{cases} 0 & x < 0 \\ \frac{1}{\beta} e^{-\frac{x}{\beta}} & x > 0 \end{cases} \quad F_X(x) = \begin{cases} 0 & x < 0 \\ 1 - e^{-\frac{x}{\beta}} & x > 0 \end{cases}$$

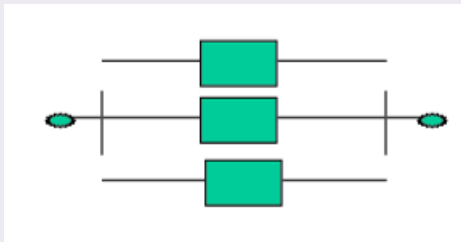
Entonces,

$$\begin{aligned} f_{Y_1}(y) &= n[1 - F_X(y)]^{n-1} f_X(y) = 3[1 - (1 - e^{-\frac{y}{\beta}})]^{3-1} \cdot \frac{1}{\beta} e^{-\frac{y}{\beta}} \\ &= \frac{3}{\beta} (e^{-\frac{y}{\beta}})^2 e^{-\frac{y}{\beta}} = \frac{3}{\beta} e^{-\frac{3y}{\beta}} \quad \text{para } y > 0. \end{aligned}$$

Es decir que $Y_1 \sim \exp(\frac{\beta}{3})$.

Ejemplo (Sistema en Paralelo)

El siguiente sistema en paralelo funcionará hasta que fallen todas las componentes (idénticas) mostradas, cada componente funciona independientemente de las otras. Supóngase que X = tiempo de vida de cada componente del sistema y que $X \sim \exp(\beta)$. Determinar el comportamiento del tiempo de vida del sistema, es decir, de $Y_3 = \max\{X_1, X_2, X_3\}$.



Luego,

$$\begin{aligned}f_{Y_3}(y) &= n[F_X(y)]^{n-1}f_X(y) = 3[1 - e^{-\frac{y}{\beta}}]^{3-1} \cdot \frac{1}{\beta}e^{-\frac{y}{\beta}} \\&= \frac{3}{\beta}[1 - e^{-\frac{y}{\beta}}]^2e^{-\frac{y}{\beta}} \quad \text{para } y > 0.\end{aligned}$$

Para un tratamiento más amplio sobre estadísticos de orden puedes revisar: David, H.A.; Nagaraja, H.N. (2003). *Order Statistics. Wiley Series in Probability and Statistics*. doi:10.1002/0471722162.

Mezclas de distribuciones.

Ejemplo

Denotemos con Y el tiempo (en cientos de horas) de cierto tipo de componentes electrónicos.

Estos componentes frecuentemente fallan cuando son colocados en el sistema. Se ha observado que las probabilidades de falla inmediata son de $1/4$. Cuando el componente no ha fallado, sus tiempos de vida se pueden modelar mediante una distribución *exponencial* ($\beta = 1$).

¿Para este sistema, cuál es la distribución de los tiempos de vida de estos componentes?

Y es una mezcla de dos variables aleatorias. **La función de probabilidad se puede escribir como:**

$$F(y) = c_1 F_1(y) + c_2 F_2(y),$$

donde $c_1 + c_2 = 1$.

Mezclas de distribuciones

Y es una mezcla de dos variables aleatorias. **La función de probabilidad se puede escribir como:**

$$F(y) = c_1 F_1(y) + c_2 F_2(y),$$

donde $c_1 + c_2 = 1$.

Para nuestro ejemplo, X_1 tiene probabilidad 1 en el punto 0 y X_2 tiene una distribución exponencial:

$$F_1(y) = \begin{cases} 0 & y < 0 \\ 1 & y \geq 0 \end{cases},$$

$$F(y) = \frac{1}{4} F_1(y) + \frac{3}{4} F_2(y),$$

$$F_2(y) = \int_0^y e^{-x} dx = 1 - e^{-y}, \quad y > 0.$$