

Clase 15: Normal bi-variada y regresión lineal

Matías Carrasco

28 de agosto de 2019

Resumen Veremos cómo construir vectores normales bi-variados correlacionados. Usando este tipo de modelos, estudiaremos la recta de regresión y daremos algunas aplicaciones.

Índice

Densidad condicional	1
Regla del producto	4
Esperanza condicional	6
Construcción de variables normales correlacionadas	7
Regresión lineal	10

Densidad condicional

Esta sección trata las probabilidades condicionales dado el valor de una variable aleatoria X con distribución continua. En el caso discreto, la probabilidad condicional de un evento A , dado que X tiene un valor x , se define por

$$P(A|X = x) = \frac{P(A, X = x)}{P(X = x)},$$

siempre que $P(X = x) > 0$. En el caso continuo $P(X = x) = 0$ para todo x , por lo que la fórmula anterior da la expresión indefinida $0/0$. Esto debe ser reemplazado, como en la definición usual del cálculo de una derivada dy/dx , por lo siguiente:

$$P(A|X = x) = \frac{P(A|X \in dx)}{P(X \in dx)}$$

Intuitivamente, $P(A|X = x)$ debe entenderse como $P(A|X \in dx)$, la probabilidad de A dado que X cae en un intervalo muy pequeño cerca de x . Aquí se supone que en el límite de pequeños intervalos, esta posibilidad no depende de qué intervalo se elija cerca de x . Entonces, como una derivada dy/dx , $P(A|X \in dx)$ es una función de x , de ahí la notación $P(A|X = x)$. En términos de límites,

$$P(A|X = x) = \lim_{\Delta x \rightarrow 0} P(A|X \in \Delta x) = \lim_{\Delta x \rightarrow 0} \frac{P(A, X \in \Delta x)}{P(X \in \Delta x)}$$

en donde Δx denota un intervalo de longitud Δx que contiene a x .

Siempre supondremos que el límite existe, excepto quizás para un número finito de puntos excepcionales x , como los puntos finales de

un intervalo que define el rango de X , o lugares donde la densidad de X tiene una discontinuidad.

A menudo, el evento A de interés está determinado por alguna variable aleatoria Y , por ejemplo $A = \{Y > 3\}$. Si (X, Y) tiene densidad conjunta $p(x, y)$, entonces $P(A|X = x)$ se puede calcular mediante integración de la densidad condicional de Y dado $X = x$, definida de la siguiente manera:

Densidad condicional de Y dado $X = x$

Si X e Y tienen densidad conjunta $p(x, y)$, entonces para cada x tal que la densidad marginal $p_X(x) > 0$, la densidad condicional de Y dada $X = x$ se define como la función de densidad de probabilidad

$$p_Y(y|X = x) = \frac{p(x, y)}{p_X(x)}.$$

Intuitivamente, la fórmula para $p_Y(y|X = x)$ se justifica por el siguiente cálculo de la probabilidad de que $Y \in dy$ dado que $X = x$:

$$\begin{aligned} P(Y \in dy|X = x) &= P(Y \in dy|X \in dx) = \frac{P(X \in dx, Y \in dy)}{P(X \in dx)} \\ &= \frac{p(x, y)dx dy}{p_X(x)dx} = p_Y(y|X = x)dy. \end{aligned}$$

La fórmula $\int p(x, y)dy = p_X(x)$, la densidad marginal de X , implica que

$$\int p_Y(y|X = x)dy = 1.$$

Entonces, para cada x fijo con $p_X(x) > 0$, la fórmula para $p_Y(y|X = x)$ da una densidad de probabilidad en y . Esta densidad condicional dada x define una distribución de probabilidad parametrizada por x , llamada distribución condicional de Y dado $X = x$. Ver la Figura 1.

En los ejemplos, la densidad condicional a menudo será una distribución familiar, por ejemplo, una distribución uniforme o normal, con parámetros que dependen de x .

La densidad condicional de Y dado $X = x$ puede entenderse geoméricamente tomando una tajada vertical a través de la superficie de densidad conjunta en x , y renormalizando la función resultante de y por su integral total, que es $p_X(x)$. Las probabilidades condicionales dado $X = x$ de eventos determinados por X e Y pueden calcularse integrando con respecto a esta densidad condicional. Por ejemplo:

$$\begin{aligned} P(Y > b|X = x) &= \int_b^\infty p_Y(y|X = x)dy; \\ P(Y > 2X|X = x) &= \int_{2x}^\infty p_Y(y|X = x)dy. \end{aligned}$$

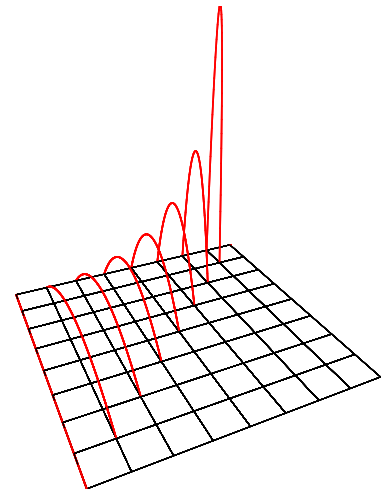
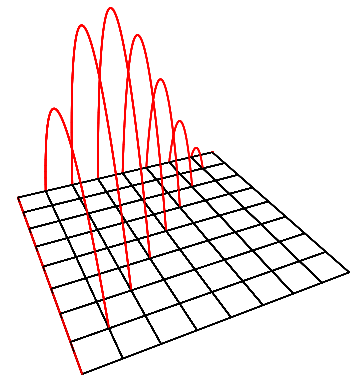
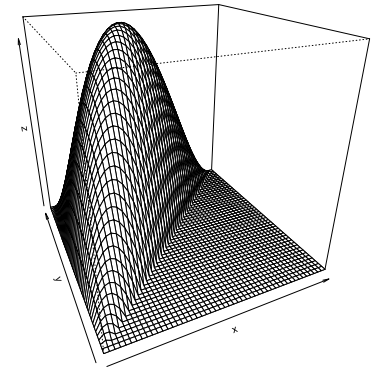


Figura 1: Aquí se muestra las densidades condicionales para la densidad $p(x, y) = 5!x(y - x)(1 - y)$, $0 < x < y < 1$ que se muestra arriba, para varios valores de x . La figura del centro muestra las tajadas con x constante, y en la de abajo aparecen las densidades normalizadas por la marginal.

Dichas expresiones se obtienen formalmente de sus análogos discretos reemplazando una suma por una integral, y reemplazando la probabilidad de un punto individual por el valor de una densidad multiplicada por una longitud infinitesimal.

■ *Ejemplo 1 — Uniforme en un triángulo.* Supongamos que (X, Y) es un punto al azar en el triángulo

$$T = \{(x, y) : x \geq 0, y \geq 0, x + y \leq 2\}.$$

Calculemos la probabilidad de que $Y > 1$ dado que $X = x$. Para ilustrar los conceptos básicos, daremos tres soluciones ligeramente diferentes.

Enfoque intuitivo

Intuitivamente, parece obvio que dado $X = x$, el punto aleatorio (X, Y) es un punto uniforme en el segmento vertical $\{(x, y) : y \geq 0, x + y \leq 2\}$ que tiene longitud $2 - x$. Esta es la distribución condicional de (X, Y) dado que $X = x$. Si x está entre 0 y 1, la proporción de este segmento que está por arriba de la línea horizontal $y = 1$ tiene longitud $(2 - x) - 1 = 1 - x$. De otro modo, la proporción es nula. Luego, la respuesta es

$$P(Y > 1 | X = x) = \begin{cases} (1 - x)/(2 - x) & \text{si } 0 \leq x < 1; \\ 0 & \text{si no.} \end{cases}$$

A partir de la definición de probabilidad condicional

Para ver que la solución anterior coincide con la definición formal

$$P(Y > 1 | X = x) = \lim_{\Delta x \rightarrow 0} P(Y > 1 | X \in \Delta x)$$

observar el siguiente diagrama que muestra los eventos $Y > 1$ y $X \in \Delta x$.

Como el triángulo tiene área 2, la probabilidad de un evento es la mitad de su área. Luego, para $0 \leq x < 1$, $x + \Delta x \leq 1$, tenemos que

$$P(X \in \Delta x) = \frac{1}{2} \Delta x \left(2 - x - \frac{1}{2} \Delta x \right)$$

y por lo tanto

$$P(Y > 1, X \in \Delta x) = \frac{1}{2} \Delta x \left(1 - x - \frac{1}{2} \Delta x \right).$$

Entonces, para $0 \leq x < 1$, deducimos que

$$\begin{aligned} P(Y > 1 | X \in \Delta x) &= \frac{P(Y > 1, X \in \Delta x)}{P(X \in \Delta x)} \\ &= \frac{1 - x - \frac{1}{2} \Delta x}{2 - x - \frac{1}{2} \Delta x} \rightarrow \frac{1 - x}{2 - x} \text{ cuando } \Delta x \rightarrow 0. \end{aligned}$$

Esto confirma la solución anterior cuando $0 \leq x < 1$. La fórmula para $x \geq 1$ es obvia pues el evento $\{Y > 1, X \in \Delta x\}$ es vacío en ese caso.

Cálculo con densidades

Calculemos ahora $P(Y > 1|X = x)$ usando la densidad condicional $p_Y(y|X = x)$. La distribución uniforme en el triángulo T tiene densidad

$$p(x, y) = \begin{cases} 1/2 & \text{si } (x, y) \in T \\ 0 & \text{en otro caso.} \end{cases}$$

Entonces, para $0 \leq x \leq 2$ tenemos

$$p_X(x) = \int_0^\infty p(x, y) dy = \int_0^{2-x} \frac{1}{2} dy = \frac{1}{2}(2-x).$$

Luego

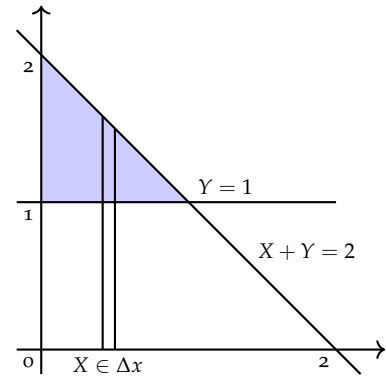
$$p_Y(y|X = x) = \begin{cases} \frac{p(x, y)}{p_X(x)} = \frac{1}{2-x} & \text{si } 0 \leq y \leq 2-x; \\ 0 & \text{si no.} \end{cases}$$

Esto es, dado $X = x$ con $0 \leq x \leq 2$, Y tiene distribución uniforme en $(0, 2-x)$, como es de esperarse intuitivamente. Entonces

$$P(Y > 1|X = x) = \begin{cases} \int_1^{2-x} \frac{dy}{2-x} = \frac{1-x}{2-x} & \text{si } 0 \leq x \leq 1 \\ 0 & \text{si no.} \end{cases}$$

La misma respuesta que antes.

El objetivo de la primera solución es que las distribuciones condicionales a menudo son intuitivamente obvias, y una vez identificadas, pueden usarse para encontrar probabilidades condicionales muy rápidamente. La segunda solución muestra que este tipo de cálculo está justificado por la definición formal. Este método no es recomendado para cálculos de rutina. La tercera solución es esencialmente una versión más detallada de la primera. Aunque bastante pedante en el presente problema, este tipo de cálculo es esencial en problemas más difíciles en los que no se puede adivinar la respuesta mediante un razonamiento intuitivo. ■



Regla del producto

Cuando la densidad de X es conocida, y la densidad condicional de Y dado $X = x$ es conocida para todo x en el recorrido de X , la densidad conjunta del par (X, Y) se calcula mediante la siguiente *regla del producto* para densidades:

$$p(x, y) = p_X(x)p_Y(y|X = x)$$

■ **Ejemplo 2** Supongamos por ejemplo que la densidad de X está dada por

$$p_X(x) = \begin{cases} \lambda^2 x e^{-\lambda x} & \text{si } x \geq 0; \\ 0 & \text{si no;} \end{cases}$$

y que Y tiene distribución condicional uniforme en $(0, x)$ dado que $X = x$.

Calculemos la densidad conjunta del par (X, Y) . La densidad condicional de Y dado que $X = x$ es

$$p_Y(y|X = x) = \begin{cases} 1/x & \text{si } 0 < y < x; \\ 0 & \text{si no.} \end{cases}$$

Por la regla del producto tenemos

$$p(x, y) = p_X(x)p_Y(y|X = x) = \begin{cases} \lambda^2 e^{-\lambda x} & \text{si } 0 < y < x; \\ 0 & \text{si no.} \end{cases}$$

Una vez que tenemos la densidad conjunta, podemos calcular la densidad marginal de Y :

$$p_Y(y) = \int_0^\infty p(x, y) dx = \int_y^\infty \lambda^2 e^{-\lambda x} dx = \lambda e^{-\lambda y} \text{ si } y \geq 0.$$

La densidad es por supuesto nula si $y \leq 0$. ■

■ **Ejemplo 3 — Fórmulas de la probabilidad total y Bayes.** Para una variable aleatoria X con densidad $p_X(x)$ la fórmula de la probabilidad total deviene

$$P(A) = \int P(A|X = x) p_X(x) dx$$

La integral descompone la probabilidad de A de acuerdo a los distintos valores que puede tomar X :

$$P(A|X = x) p_X(x) = P(A|X \in dx) P(X \in dx) = P(A, X \in dx).$$

Al igual que en el caso discreto, a menudo $P(A|X = x)$ se especifica de antemano por la formulación de un problema. Luego $P(A)$ puede calcularse mediante la fórmula integral de la probabilidad total, suponiendo también que se conoce la distribución de X . La regla de Bayes da la densidad condicional de X dado que A ha ocurrido:

$$P(X \in dx|A) = \frac{P(A|X \in dx) P(X \in dx)}{P(A)} = \frac{p_X(x) P(A|X = x) dx}{\int p_X(x) P(A|X = x) dx}$$

■ **Ejemplo 4** Supongamos que Π tiene distribución uniforme en $[0, 1]$. Dado que $\Pi = p$, sea S_n el número de éxitos en n ensayos de Bernoulli con probabilidad de éxito p . Calculemos la distribución de S_n .

Como la densidad de Π es $p_{\Pi}(p) = 1$ si $p \in [0, 1]$ y cero en otro caso, por la fórmula de la probabilidad total tenemos que

$$\begin{aligned} P(S_n = k) &= \int P(S_n = k | \Pi = p) p_{\Pi}(p) dp \\ &= \int_0^1 \binom{n}{k} p^k (1-p)^{n-k} dp = \frac{1}{n+1}. \end{aligned}$$

Es decir, S_n tiene distribución uniforme en $\{0, \dots, n\}$.

Calculemos ahora la densidad condicional de Π dado que $S_n = k$. Usando la regla de Bayes, para $0 < p < 1$ tenemos

$$\begin{aligned} P(\Pi \in dp | S_n = k) &= \frac{P(\Pi \in dp) P(S_n = k | \Pi = p)}{P(S_n = k)} \\ &= (n+1) \binom{n}{k} p^k (1-p)^{n-k} dp. \end{aligned}$$

Calculemos ahora la probabilidad de que el próximo ensayo sea un éxito dado que $S_n = k$.

$$\begin{aligned} &P(\text{un éxito más} | S_n = k) \\ &= \int_0^1 P(\text{un éxito más} | S_n = k, \Pi = p) p_{\Pi}(p | S_n = k) dp \\ &= \int_0^1 p p_{\Pi}(p | S_n = k) dp = \frac{k+1}{n+2} \end{aligned}$$

En particular, para $k = n$, dados n éxitos seguidos, la probabilidad de un éxito más es $(n+1)/(n+2)$. Esta fórmula, para la probabilidad de un éxito más dada una serie de n éxitos en ensayos independientes con probabilidad de éxito desconocida, supuestamente distribuida uniformemente en $(0, 1)$, se conoce como la ley de Laplace. Laplace ilustró su fórmula calculando la probabilidad de que el sol salga mañana, dado que ha salido diariamente durante 5000 años, o $n = 1826213$ días. Pero este tipo de aplicación es dudosa. Tanto la suposición de ensayos independientes con p desconocida, como la distribución a priori uniforme de p , tienen poco sentido en este contexto. ■

Esperanza condicional

La esperanza condicional de Y dada $X = x$, denotada $E(Y|X = x)$, se define como la esperanza de Y con respecto a la distribución condicional de Y dada $X = x$. De manera más general, para una función $g(y)$, suponiendo que Y tiene densidad condicional $p_Y(y|X = x)$ tenemos que

$$E(g(Y)|X = x) = \int g(y) p_Y(y|X = x) dy.$$

Tomando $g(y) = y$ da $EY|X = x$. La integración de la esperanza condicional con respecto a la distribución de X da la esperanza (incondi-

cional)

$$E(g(Y)) = \int E(g(Y)|X = x) p_X(x) dx.$$

Estas fórmulas son extensiones a funciones generales $g(y)$ de la fórmula de probabilidad total anterior, que son el caso especial cuando g es un indicador de un conjunto. Como regla general, valen las mismas propiedades para la esperanza condicional que valen para la esperanza (incondicional).

■ **Ejemplo 5** Punto al azar en un triángulo Supongamos que (X, Y) es un punto al azar en el triángulo $T = \{(x, y) : x \geq 0, y \geq 0, x + y \leq 2\}$. Calculemos $E(Y|X = x)$.

Dado $X = x$, con $0 < x < 2$, Y tiene distribución uniforme en $(0, 2 - x)$. Como la esperanza de esta distribución condicional es $(2 - x)/2$, concluimos que

$$E(Y|X = x) = \frac{2 - x}{2}.$$

Una notación muy usada es la siguiente: $E(Y|X) = (2 - X)/2$. Cuando sabemos que $X = x$, basta reemplazar x por X en la expresión anterior.

■

Construcción de variables normales correlacionadas

En esta clase estudiaremos variables normales correlacionadas (no independientes) mediante cambios lineales de variables que reducen el estudio a lo que vimos en clases anteriores sobre normales independientes.

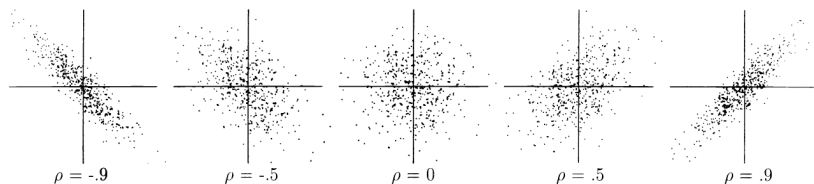


Figura 2: Simulación por computadora de normales bi-variadas para varias correlaciones.

Nubes de puntos como la de arriba son muy comunes. Fueron estudiadas por primera vez por Francis Galton (1822-1911), quién estudió la relación entre variables genéticas, como la altura de los padres y los hijos.

El ingrediente principal es el coeficiente de correlación

$$\rho = \rho_{XY} = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)}\sqrt{\text{Var}(Y)}} \in [-1, 1].$$

Para obtener un par de variables normales correlacionadas X e Y , podemos comenzar con variables independientes X y Z normales están-

dar. Luego tomamos la proyección del par (X, Z) sobre la recta que forma un ángulo θ con el eje X .

De la misma forma que vimos antes, la proyección Y se escribe como

$$Y = X \cos \theta + Z \sin \theta,$$

y sabemos que Y tiene distribución normal estándar.

Por otro lado, la correlación entre X e Y es entonces

$$\begin{aligned} \rho_{XY} &= E(XY) = E(X(X \cos \theta + Z \sin \theta)) \\ &= E(X^2 \cos \theta + XZ \sin \theta) \\ &= \underbrace{E(X^2)}_1 \cos \theta + \underbrace{E(XZ)}_0 \sin \theta = \cos \theta. \end{aligned}$$

Para resumir, $\rho = \cos \theta$. Por ejemplo:

- $\theta = 0 \Rightarrow \rho = 1 \Rightarrow Y = X$
- $\theta = \pi/2 \Rightarrow \rho = 0 \Rightarrow X$ e Y son independientes
- $\theta = \pi \Rightarrow \rho = -1 \Rightarrow Y = -X$

Para cada $\rho \in [-1, 1]$, existe un ángulo θ tal que $\rho = \cos \theta$. Luego, para cada $\rho \in [-1, 1]$ existen X e Y normales estándar con correlación ρ . Como $\sin \theta = \sqrt{1 - \rho^2}$, podemos escribir

$$Y = \rho X + \sqrt{1 - \rho^2} Z$$

en donde X e Y son normales estándar independientes.

Definición de normal bi-variada estándar

Decimos que el par (X, Y) tiene distribución normal bi-variada estándar con correlación ρ si

$$Y = \rho X + \sqrt{1 - \rho^2} Z$$

en donde X y Z son normales estándar independientes.

Veamos algunas propiedades que se deducen de la definición:

1. *Marginales*: como ya lo hemos visto $X \sim N(0, 1)$ e $Y \sim N(0, 1)$.
2. *Condicionales*:
 - a) Dado que $X = x$ entonces $Y \sim N(\rho x, 1 - \rho^2)$.
 - b) Dado que $Y = y$ entonces $X \sim N(\rho y, 1 - \rho^2)$.

3. *Densidad conjunta*: el par (X, Y) tiene densidad conjunta

$$p(x, y) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}(x^2 - 2\rho xy + y^2)\right)$$

Esto se prueba fácilmente usando la regla del producto $p(x, y) = p_X(x) \cdot p_Y(y|X = x)$.

4. *Independencia*: X e Y son independientes si, y solo si, $\rho = 0$. Esto es porque si $\rho = 0$ entonces $Y = Z$.

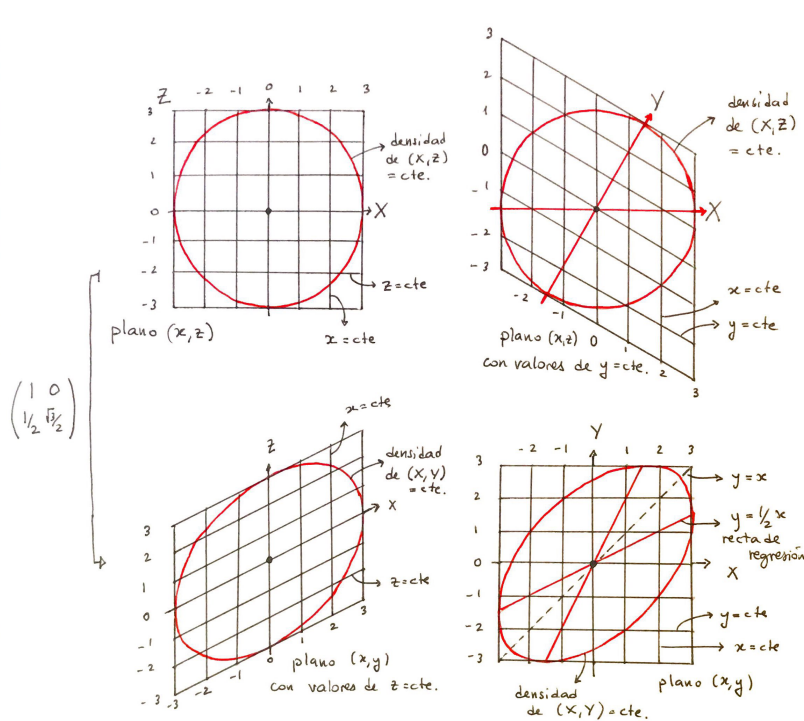
A veces un dibujo vale más que mil palabras. La geometría detrás de las fórmulas que hemos escrito hasta ahora se muestra en el siguiente dibujo en un caso particular. En el dibujo $\rho = \cos(60^\circ) = 1/2$, y por lo tanto

$$\sqrt{1-\rho^2} = \sin(60^\circ) = \sqrt{3}/2, \quad Y = \frac{1}{2}X + \frac{\sqrt{3}}{2}Z.$$

Notar que se trata de un cambio lineal de variables

$$(X, Z) \rightarrow \left(X, \frac{1}{2}X + \frac{\sqrt{3}}{2}Z\right) = \begin{pmatrix} 1 & 0 \\ 1/2 & \sqrt{3} \end{pmatrix} \begin{pmatrix} X \\ Z \end{pmatrix}.$$

En la parte superior del dibujo se muestra el plano de coordenadas (x, z) y en la parte inferior el de coordenadas (x, y) . Las imágenes por el cambio de coordenadas de las rectas paralelas a $z = 0$ son las rectas paralelas a $y = \rho x$. Por su importancia para lo que haremos más adelante le daremos un nombre a esta recta: la recta $y = \rho x$ en el plano (x, y) es la *recta de regresión*.



La definición general de normal bi-variada es la siguiente.

Definición general de normal bi-variada

Decimos que el par (X, Y) tiene distribución normal bi-variada de parámetros

$$\left(\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \rho \right)$$

si el par

$$\left(\frac{X - \mu_X}{\sigma_X}, \frac{Y - \mu_Y}{\sigma_Y} \right)$$

tiene distribución normal bi-variada estándar con correlación ρ .

Desglosando la definición de normal bi-variada estándar dada más arriba, vemos que esto quiere decir que existe Z normal estándar independiente de X tal que

$$\frac{Y - \mu_Y}{\sigma_Y} = \rho \frac{X - \mu_X}{\sigma_X} + \sqrt{1 - \rho^2} Z.$$

Reordenando esta ecuación obtenemos

$$Y = \mu_Y + \rho \frac{\sigma_Y}{\sigma_X} (X - \mu_X) + \sigma_Y \sqrt{1 - \rho^2} Z.$$

Regresión lineal

Cuando queremos estudiar la relación entre un par de variables aleatorias X e Y , es común tratar de entender lo que se conoce como *función de regresión*:

$$R(x) = E(Y|X = x).$$

Esta función se suele usar como método de predicción de la variable Y si sabemos que X vale x . La regresión se dice lineal cuando $R(x) = mx + n$.

Un caso muy importante es cuando sabemos que el par (X, Y) tiene (al menos de forma aproximada) distribución normal bi-variada. En este caso la función de regresión es

$$\begin{aligned} R(x) &= E(Y|X = x) = E\left(\mu_Y + \rho \frac{\sigma_Y}{\sigma_X} (X - \mu_X) + \sigma_Y \sqrt{1 - \rho^2} Z \mid X = x\right) \\ &= \mu_Y + \rho \frac{\sigma_Y}{\sigma_X} (x - \mu_X) + \sigma_Y \sqrt{1 - \rho^2} E(Z|X = x) \\ &= \mu_Y + \rho \frac{\sigma_Y}{\sigma_X} (x - \mu_X), \end{aligned}$$

ya que Z es independiente de X y tiene esperanza nula.

Los coeficientes de la función de regresión se suelen llamar

$$\beta = \rho \frac{\sigma_Y}{\sigma_X}, \quad \alpha = \mu_Y - \rho \frac{\sigma_Y}{\sigma_X} \mu_X,$$

por lo que $R(x) = \beta x + \alpha$.

Veamos algunos ejemplos de aplicación.

■ **Ejemplo 6** Un estudiante de Galton, Karl Pearson, llevó a cabo estudios genéticos sobre la relación entre la altura de padres e hijos. Digamos que como idea general se trata de predecir la altura de un hijo en base a la altura de su padre.

Para esto Pearson colectó muestras de 1078 pares de padres e hijos

$$(x_1, y_1), \dots, (x_{1078}, y_{1078})$$

en donde x_i es la altura del padre e y_i es la altura del hijo.

Los datos de Pearson se resumen en la tabla siguiente.

Padres:	$\mu_X = 1.75$ (m)	$\sigma_X = 0.05$ (m)
Hijos:	$\mu_Y = 1.78$ (m)	$\sigma_Y = 0.05$ (m)
Correlación:	$\rho = 0.5$	

Asumimos que el par (X, Y) tiene distribución normal bi-variada. Queremos predecir la altura del hijo de un padre que mide 1.88 m. Para esto usaremos la recta de regresión. Evaluando los coeficientes

$$\beta = 0.5, \quad \alpha = 0.905$$

Luego,

$$R(1.88) = 0.5 \cdot 1.88 + 0.905 = 1.845 \approx 1.85.$$

Notar que la altura predicha es menor que la altura del padre. Galton llamaba a este fenómeno de regresión a la media. De aquí proviene el nombre “regresión”.

¿Cuál es la probabilidad de que nuestra predicción sea errónea por más de 2.5 cm? Definimos las variables

$$U = \frac{X - \mu_X}{\sigma_X}, \quad V = \frac{Y - \mu_Y}{\sigma_Y}.$$

Por definición estamos asumiendo que (U, V) es normal bi-variado estándar de correlación $\rho = 0.5$. Sea Z normal estándar independiente de U , con

$$V = \rho U + \sqrt{1 - \rho^2} Z.$$

Estamos en la probabilidad

$$\begin{aligned} & P\left(|Y - R(X)| > 0.025 \mid X = 1.88\right) \\ &= P\left(\left|\frac{Y - \mu_Y}{\sigma_Y} - \rho \frac{(X - \mu_X)}{\sigma_X}\right| > \frac{0.025}{\sigma_Y} \mid X = 1.88\right) \\ &= P\left(|V - \rho U| > 0.5 \mid \sigma_X U + \mu_X = 1.88\right) \\ &= P\left(|V - \rho U| > 0.5 \mid U = 2.6\right) \end{aligned}$$

Pero como $V - \rho U = \sqrt{1 - \rho^2}Z$ es independiente de U , esta probabilidad es igual a

$$\begin{aligned} P(|V - \rho U| > 0.5 | U = 2.6) &= P\left(\sqrt{1 - \rho^2}|Z| > 0.5\right) \\ &= P\left(|Z| > \frac{0.5}{\sqrt{1 - \rho^2}}\right) \\ &= P(|Z| > 0.58) = 0.562. \end{aligned}$$

Es decir, hay un 56 % de chances de que nuestra predicción sea errónea por más de 2.5 cm.

Estimemos la altura ahora del padre de un hijo que mide 1.85 m. Una respuesta rápida sería que como la predicción de la altura de un hijo cuyo padre mide 1.88 m es 1.85 m, entonces la predicción de la altura de un padre cuyo hijo mide 1.85 m es 1.88 m. Sin embargo, esto no es así.

Hay dos rectas de regresión. Intercambiando X con Y vemos que

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

Así que la predicción es $R(1.85) = 1.78$. Nuevamente vemos el efecto de regresión a la media.

Calculemos ahora la probabilidad de que tanto el padre como el hijo tengan alturas por arriba de la media. Si pasamos a las unidades típicas (U, V) , estamos interesados en el evento $\{U \geq 0, V \geq 0\}$. Como

$$V = \rho U + \sqrt{1 - \rho^2}Z$$

vemos que

$$P(U \geq 0, V \geq 0) = P\left(U \geq 0, Z \geq -\frac{\rho}{\sqrt{1 - \rho^2}}U\right).$$

La pendiente de la recta es

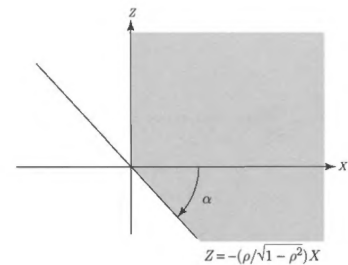
$$-\frac{\rho}{\sqrt{1 - \rho^2}} = \tan \alpha = -\frac{1}{\sqrt{3}},$$

de donde $\alpha = 30^\circ$.

Luego la región que nos interesa forma un ángulo de 120° . Usando la simetría rotacional de la normal bi-variada estándar (U, Z) , vemos que

$$P\left(U \geq 0, Z \geq -\frac{\rho}{\sqrt{1 - \rho^2}}U\right) = \frac{1}{3}.$$

En definitiva $P(X \geq \mu_X, Y \geq \mu_Y) = 1/3$. ■



■ **Ejemplo 7** Supongamos que el par (X, Y) tiene distribución normal bi-variada, y que la probabilidad de que $X \geq \mu_X$ y que $Y \geq \mu_Y$ es $3/8$. ¿Cuánto vale ρ ?

Si transformamos el par (X, Y) a unidades típicas (U, V) , debemos estimar ρ a partir del hecho de que

$$P\left(U \geq 0, Z \geq -\frac{\rho}{\sqrt{1-\rho^2}}U\right) = \frac{3}{8}.$$

Como $3/8 = 135/360$, vemos que la región hace un ángulo de $135^\circ = 90^\circ + 45^\circ$. Luego

$$\tan \alpha = -\frac{\rho}{\sqrt{1-\rho^2}} = \tan(-45) = -1,$$

de donde $\rho = 1/\sqrt{2}$. ■

Sea (X, Y) con distribución normal estándar de correlación ρ . Calculemos $E(Y|a < X < b)$.

Si $X = x$, sabemos que $E(Y|X = x) = \rho x$. Luego

$$\begin{aligned} E(Y|a < X < b) &= \int_a^b E(Y|X = x) p_X(x|a < X < b) dx \\ &= \int_a^b \rho x p_X(x|a < X < b) dx. \end{aligned}$$

Por otro lado, si $x \in (a, b)$ tenemos

$$\begin{aligned} p_X(x|a < X < b) &= P(X \in dx|a < X < b) \\ &= \frac{P(X \in dx, a < X < b)}{P(a < X < b)} \\ &= \frac{\varphi(x)}{\Phi(b) - \Phi(a)} \end{aligned}$$

Integrando resulta

$$E(Y|a < X < b) = \frac{\frac{\rho}{\sqrt{2\pi}} (e^{-a^2/2} - e^{-b^2/2})}{\Phi(b) - \Phi(a)}.$$

■ **Ejemplo 8** La siguiente tabla representa el resumen de los datos (X, Y) con distribución normal bi-variada, en donde X es la nota de un estudiante en un parcial, e Y es la nota en el examen final.

Parcial:	$\mu_X = 65$ (puntos)	$\sigma_X = 18$ (puntos)
Examen:	$\mu_Y = 60$ (puntos)	$\sigma_Y = 20$ (puntos)
Correlación:	$\rho = 0.75$	

Estimemos el valor esperado de la nota de examen de un estudiante que ha obtenido una nota superior a la media en el parcial.

Sea (U, V) las notas en el parcial y examen en unidades típicas. El evento conocido es $\{U > 0\}$. Tomando $a = 0$ y $b = +\infty$ en la fórmula general que derivamos más arriba, vemos que

$$E(V|U > 0) = \frac{\rho}{\sqrt{2\pi}} \left[\frac{1-0}{0.5} \right] = \frac{0.75 \times 2}{\sqrt{2\pi}} \approx 0.6$$

Esta es la respuesta en unidades típicas. Volviendo a la escala de puntos original, la respuesta es $60 + 20 \times 0.6 = 72$. ■