Capítulo 5 Distribuciones de probabilidad multivariantes

5.1 Introducción

La intersección de dos o más eventos es frecuentemente de interés para un experimentador. Suponga que $Y_1, Y_2, \ldots, Y_n$ denota los resultados de n intentos sucesivos de un experimento. Un conjunto específico de resultados o mediciones muestrales puede ser expresado en términos de la intersección de los n eventos $(Y_1 = y_1), (Y_2 = y_2), \ldots, (Y_n = y_n)$, que denotaremos como $(Y_1 = y_1, Y_2 = y_2, \ldots, Y_n = y_n)$, que denotaremos como $(Y_1 = y_1, Y_2 = y_2, \ldots, Y_n = y_n)$ o bien como $(y_1, y_2, \ldots, y_n)$. El cálculo de la probabilidad de esta intersección es esencial para hacer inferencias de la población de la cual se tomo.

5.2 Distribuciones de probabilidad bivariantes y multivariantes

Se pueden definir muchas variables aleatorias sobre el mismo espacio muestral. Por ejemplo al tirar un par de dados el espacio muestral es de 36 puntos. Que corresponden a los valores posibles de ambos dados. Se puede decir que los valores que cada dado puede tomar es una variable aleatoria entonces al tirar ambos se estaría buscando la probabilidad conjunta de dos variables aleatorias.

Definición 5.1

Sean Y_1 y Y_2 variables aleatorias discretas. La *función de probabilidad conjunta* (o bivariante) para Y_1 y Y_2 está dada por

$$p(y_1, y_2) = P(Y_1 = y_1, Y_2 = y_2)$$

cuando $-\infty < y_1 < \infty, -\infty < y_2 < \infty$.

Teorema 5.1

Si Y_1 y Y_2 son variables aleatorias discretas con función de probabilidad conjunta $p(y_1, y_2)$, entonces

1. $p(y_1, y_2) \geq 0$ para toda y_1, y_2.
2. $\sum_{y_1 y_2} p(y_1, y_2) = 1$, donde la suma es para todos los valores (y_1, y_2) a los que se asignan probabilidades difernetes de cero.

Al igual que en el caso univariado, la función de probabilidad conjunto para variables aleatorias discretas a veces se denomina *función de masa de probabilidad conjunta*. De la misma manera la distinción entre variables aleatorias continuas conjuntas y discretas conjuntas puede ser caracterizado en términos de sus funciones de distribución.

Definición 5.2

Para cualesquiera variables aleatorias Y_1 y Y_2, la función de distribución (bivariante) conjunta $F(y_1, y_2)$ es

$$F(y_1, y_2) = P(Y_1 \leq y_1, Y_2 \leq y_2)$$

cuando $-\infty < y_1 < \infty, -\infty < y_2 < \infty$

Para dos variables discretas Y_1 y Y_2, $F(y_1, y_2)$ está dada por

$$F(y_1, y_2) = \sum_{t_1 \leq y_1} \sum_{t_2 \leq y_2} p(t_1, t_2)$$

Se dice que dos variables aleatorias son coninuas conjuntas si su función de distribución conjunta $F(y_1, y_2)$ es continua en ambos segmentos.

Definición 5.3

Sean Y_1 y Y_2 variables aleatorias continuas con función de distribución conjunta $F(y_1.y_2)$. Si existe una función no negativa $f(y_1.y_2)$, tal que

$$F(y_1, y_2) = \int_{-\infty}^{y_1} \int_{-\infty}^{y_2} f(t_1, t_2) dt_1 dt_1$$

Para toda $-\infty < y_1 < \infty, -\infty < y_2 < \infty$, entonces se dice que Y_1 y Y_2 son *variables aleatorias continuas conjuntas*. La función $f(y_1, y_2)$ recibe el nombre de *función de densidad de probabilidad conjunta*.

De la misma manera que en el caso univariado, las funciones de distribución acumulativas satisfacen ciertas propiedades las cuales se muestran en el siguiente teorema.

Teorema 5.2

Si Y_1 y Y_2 son variables aleatorias con función de distribución conjunta $F(y_1, y_2)$, entonces

1. $F(-\infty, \infty) = F(-\infty, y_2) = F(y_1, -\infty) = 0$.
2. $F(\infty, \infty) = 1$.
3. Si $y_1^* \geq y_1$ y $y_2^* \geq y_2$, entonces

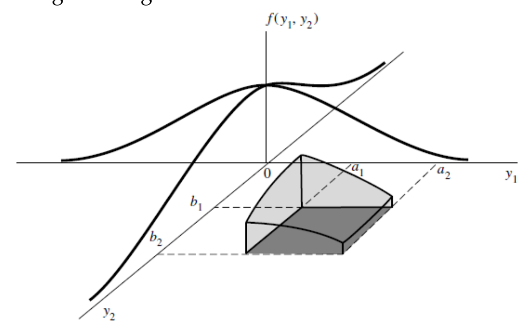
$$F(y_1^*, y_2^*) - F(y_1^*, y_2) - F(y_1, y_2^*) + F(y_1, y_2) \geq 0$$

Teorema 5.3

Si Y_1 y Y_2 son variables aleatorias continuas conjuntas con una función de densidad conjunta dada por $f(y_1, y_2)$, entonces

1. $f(y_1, y_2) \geq 0$ para toda y_1, y_2.
2. $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(y_1, y_2) dy_1 dy_2 = 1$.

Para el caso univariado, las áreas bajo la densidad de probabilidad para un intervalo corresponden a probabilidades. De igual manera la función de probabilidad bivariante $f(y_1, y_2)$ traza una superficie de densidad de probabilidad sobre el plano (y_1, y_2) como se muestra en la siguiente figura.



Los volúmenes bajo esta superficie representan probabilidades. Así, $P(a_1 \leq Y_1 \leq a_2, b_1 \leq Y_2 \leq b_2)$ es el volumen sombreado que se ve en la figura anterior y a continuación.

$$\int_{b_1}^{b_2} \int_{a_1}^{a_2} f(y_1, y_2) dy_1 dy_2$$

Esos métodos vistos se pueden generalizar para una función de probabilidad (o función de densidad de probabilidad) para la intersección de n eventos $(Y_1 = y_1, Y_2 = y_2, \ldots, Y_n = y_n)$ La función de probabilidad conjunta correspondiente al caso discreto está dada por

$$p(y_1, y_2, \ldots, y_n) = P(Y_1 = y_1, Y_2 = y_2, \ldots, Y_n = y_n)$$

En el caso continuo se tiene lo siguiente,

$$\begin{aligned} P(Y_1 \leq y_1, Y_2 \leq y_2, \ldots, Y_n \leq y_n) &= F(y_1, \ldots, y_n) \\ &= \int_{-\infty}^{y_1} \int_{-\infty}^{y_2} \cdots \int_{-\infty}^{y_n} f(t_1, t_2, \ldots, t_n) \\ & dt_n \cdots dt_1 \end{aligned}$$

para todo conjunto de número reales $(y_1, y_2, \ldots, y_n)$. Por último la función de denisdad conjunta de $Y_1, \ldots, Y_n$ está dada por $f(y_1, \ldots, y_n)$.

5.3 Distribuciones de probabilidad marginal y condicional

Recuerde que los valores distintos tomados por una variable aleatoria discreta represntan eventos mutuamente excluyentes. De manera análoga, para todos los distintos pares de valores y_1, y_2, los eventos bivariados $(Y_1 = y_1, Y_2 = y_2)$, representados por (y_1, y_2), son eventos mutuamente excluyentes. De esto se deduce que el evento univariado $(Y_1 = y_1)$ es la unión de eventos bivariados del tipo $(Y_1 = y_1, Y_2 = y_2)$, con la unión tomada para todos los posibles valores de y_2.

Definición 5.4

a Sean Y_1 y Y_2 variables aleatorias discretas conjuntas con función de probabilidad $p(y_1, y_2)$. Entonces las *funiciones de probabilidad marginal* de Y_1 y Y_2, respectivamente, están dadas por

$$p_1(y_1) = \sum_{\text{todos } y_2} p(y_1, y_2)$$
$$p_2(y_2) = \sum_{\text{todos } y_2} p(y_1, y_2)$$

b Sean Y_1 y Y_2 variabels aleatorias continuas conjuntas con función de densidad conjunta $f(y_1, y_2)$. Entonces las *funciones de densidad marginal* de Y_1 y Y_2, respectivamente, están dadas por

$$f_1(y_1) = \int_{-\infty}^{\infty} f(y_1, y_2) dy_2$$
$$f_2(y_2) = \int_{-\infty}^{\infty} f(y_1, y_2) dy_1$$

Llevemos ahora nuestra atención a distribuciones condicionales, viendo primero al caso discreto.

Definición 5.5

Si Y_1 y Y_2 son variables aleatorias discretas conjuntas con función de probabilidad conjunta $p(y_1, y_2)$ y funciones de probabilidad marginal $p_1(y_1)$ y $p_2(y_2)$, respectivamente, entonces la *función de probabilidad discreta condicional* de Y_1 dada Y_2 es

$$p(y_1|y_2) = P(Y_1 = y_1|Y_2 = y_2) = \frac{P(Y_1 = y_1, Y_2 = y_2)}{P(Y_2 = y_2)}$$
$$= \frac{p(y_1, y_2)}{p_2(y_2)}$$

siempre que $p_2(y_2) > 0$.

En el caso continuo no se obtiene la función de probabilidad condicional de forma sencilla. Esto es porque si Y_1 y Y_2 son continuas, $P(Y_1 = y_1|Y_2 = y_2)$ no se puede definir por que $(Y_1 = y_1)$ y $(Y_2 = y_2)$ son eventos con probabilidad cero.

Definición 5.6

Si Y_1 y Y_2 son variables aleatorias continuas conjuntas con función de densidad conjunta $f(y_1, y_2)$, entonces la *función de distribución condicional* de Y_1 dado que $Y_2 = y_2$ es

$$F(y_1|y_2) = P(Y_1 \leq y_1|Y_2 = y_2)$$

Como se esta tratando con una variable aleatoria continua entonces es necesario utilizar integrales. Entonces si se quiere encontrar $F(y_1)$ no se puede sumar sobre todos los valores de Y_2. Pero se puede hacer algo análogo al multiplicarlo por $f_2(y_2)$ para obtener

$$F(y_1) = \int_{-\infty}^{\infty} F(y_1|y_2) f_2(y_2) dy_2$$

De consideraciones previas se tiene

$$F(y_1) = \int_{-\infty}^{y_1} f_1(t_1) dt_1 = \int_{-\infty}^{y_1} \left[\int_{-\infty}^{\infty} f(t_1, y_2) dy_2 \right] dt_1$$
$$= \int_{-\infty}^{\infty} \int_{-\infty}^{y_1} f(t_1, y_2) dt_1 dy_2$$

De estas dos expresiones para $F(y_1)$, debemos tener

$$F(y_1|y_2) f_2(y_2) = \int_{-\infty}^{y_1} f(t_1, y_2) dt_1$$
$$F(y_1|y_2) = \int_{-\infty}^{\infty} \frac{f(t_1, y_2)}{f_2(y_2)} dt_1$$

Definición 5.7

Sean Y_1 y Y_2 variables aleatorias continuas conjuntas con densidad conjunta $f(y_1, y_2)$ y densidades marginales $f_1(y_1)$ y $f_2(y_2)$, respectivamente. Para cualquier y_2 tal que $f_2(y_2) > 0$, la densidad condicional de Y_1 dada $Y_2 = y_2$ está dada por

$$f(y_1|y_2) = \frac{f(y_1, y_2)}{f_2(y_2)}$$

y para cualquier y_1 tal que $f_1(y_1) > 0$, la densidad condicional de Y_2 dada $Y_1 = y_1$ está dada por

$$f(y_2|y_1) = \frac{f(y_1, y_2)}{f_1(y_1)}$$

5.4 Variables aleatorias independientes

Se ha estudiado la independencia de eventos en capítulo anteriores. Esto se puede extender a variables aleatorias.

Definición 5.8

Sea Y_1 que tiene una función de distribución $F_1(y_1)$ y sea Y_2 que tiene una función de distribución $F_2(y_2)$, y $F(y_1, y_2)$ es la función de distribución conjunta de Y_1 y Y_2. Entonces se dice que Y_1 y Y_2 son *independientes* si y sólo si

$$F(y_1, y_2) = F_1(y_1) F_2(y_2)$$

para todo par de números reales (y_1, y_2). Si Y_1 y Y_2 no son independientes, se dice que son *dependientes*.

Teorema 5.4

Si Y_1 y Y_2 son variables discretas con función de probabilidad conjunta $p(y_1, y_2)$ y funciones de probabilidad marginal $p_1(y_1)$ y $p_2(y_2)$, respectivamente, entonces Y_1 y Y_2 son independientes si y sólo si

$$p(y_1, y_2) = p_1(y_1) p_2(y_2)$$

para todos los pares de número reales (y_1, y_2).
Si Y_1 y Y_2 son variables aleatorias continuas con función de densidad conjunta $f_1(y_1, y_2)$ y funciones de densidad marginal $f_1(y_1)$ y $f_2(y_2)$, respectivamente, entonces Y_1 y Y_2 son independientes si y sólo si

$$f(y_1, y_2)) f_1(y_1) f_2(y_2)$$

para todos los pares de número reales (y_1, y_2).

Teorema 5.5

Sean Y_1 y Y_2 que tienen densidad conjunta $f(y_1, y_2)$ que es positiva si y sólo si $a \leq y_1 \leq b$ y $c \leq y_2 \leq d$, para constantes a, b, c y d; y $f(y_1, y_2) = 0$ en otro caso. Entonces Y_1 y Y_2 son variables aleatorias independientes si y sólo si

$$f(y_1, y_2) = g(y_1) h(y_2)$$

donde $g(y_1)$ es una función no negativa de y_1 solamente y $h(y_2)$ es una función no negativa de y_2 solamente.

El beneficio del Teorema 5.5 es que en realidad no necesitamos obtener las densidades marginales. De hecho, las funciones $g(y_1)$ y $h(y_2)$ no necesitan ser funciones de densidad.
Finalmente la definición 5.8 se puede generalizar a n dimensiones. Suponga que se tiene n variables aleatorias $Y_1, \ldots, Y_n$, donde Y_i tiene función de distribución $F_i(y_i)$, para $i = 1, 2, \ldots, n$; y que estas n variables aleatorias tengan función de distribución conjunta $F(y_1, \ldots, y_n)$. Entonces $Y_1, Y_2, \ldots, Y_n$ son independientes si y sólo si

$$F(y_1, y_2, \ldots, y_n) = F_1(y_1) \cdots F_n(y_n)$$

para todos los números reales $y_1, y_2, \ldots, y_n$, son las formas equivalentes obvias para los casos discretos y continuos.

5.5 El valor esperado de una función de variables aleatorias

Definición 5.9

Sea $g(Y_1, Y_2, \ldots, Y_k)$ una función de las variables aleatorias discretas, $Y_1, Y_2, \ldots, Y_k$, que tienen función de probabilidad $p(y_1, y_2, \ldots, y_k)$. Entonces el *valor esperado* de $g(Y_1, Y_2, \ldots, Y_k)$ es

$$E[g(Y_1, \ldots, Y_k)] = \sum_{\text{toda } y_k} \cdots \sum_{\text{toda } y_1} g(y_1, \ldots, y_k) p(y_1, \ldots, y_k)$$

Si $Y_1, Y_2, \ldots, Y_k$ son variables aleatorias continuas con función de denisdad conjunta $f(y_1, y_2, \ldots, y_k)$, entonces

$$E[g(Y_1, \ldots, Y_k)] = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(y_1, \ldots, y_k) \times f(y_1, \ldots, y_k) \\ dy_1 \ldots, dy_k$$

Se puede demostrar que la Definición 5.9 es consistente con la definición 4.5, en la que se define el valor esperado de una variable aleatorias univariada de la siguiente manera. Considere dos variables aleatorias Y_1 y Y_2 con función de densidad $f(y_1, y_2)$. Deseamos hallar el valor esperado de $g(Y_1, Y_2) = Y_1$. De la definición anterior se tiene.

$$\begin{aligned} E(Y_1) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y_1 f(y_1, y_2) dy_2 dy_1 \\ &= \int_{-\infty}^{\infty} y_1 \left[\int_{-\infty}^{\infty} f(y_1, y_2) dy_2 \right] dy_1 \end{aligned}$$

La cantidad dentro del paréntesis, por definición, es la función de densidad marginal para Y_1. Por lo tanto se puede obtener lo siguiente

$$E(Y_1) = \int_{-\infty}^{\infty} y_1 f_1(y_1) dy_1$$

5.6 Teoremas especiales

Teorema 5.6

Sea c una constante. Entonces

$$E(c) = c$$

Teorema 5.7

Sea $g(Y_1, Y_2)$ una función de las variables aleatorias Y_1 y Y_2 y sea c una constante. Entonces

$$E[cg(Y_1, Y_2)] = cE[g(Y_1, Y_2)]$$

Teorema 5.8

Sean Y_1 y Y_2 variables aleatorias y $g_1(Y_1, Y_2)$, $g_2(Y_1, Y_2)$, $\ldots$, $g_k(Y_1, Y_2)$ funciones de Y_1 y Y_2. Entonces

$$\begin{aligned} &E[g_1(Y_1, Y_2) + g_2(Y_1, Y_2) \cdots + g_k(Y_1, Y_2)] \\ &\quad = E[g_1(Y_1, Y_2)] + E[g_2(Y_1, Y_2)] + \cdots + E[g_k(Y_1, Y_2)] \end{aligned}$$

Teorema 5.9

Sean Y_1 y Y_2 variables aleatorias independientes y sean $g(Y_1)$ y $h(Y_2)$ funciones sólo de Y_1 y Y_2, respectivamente. Entonces

$$E[g(Y_1)h(Y_2)] = E[g(Y_1)]E[h(Y_2)]$$

siempre que existan los valores esperados.

Demostración

Daremos la demostración del resultado para el caso continuo. Denotemos con $f(y_1, y_2)$ la densidad conjunta de Y_1 y Y_2. El producto $g(Y_1)h(Y_2)$ es una función de Y_1 y Y_2. Entonces, por la Definición 5.9 y la suposición de que Y_1 y Y_2 son independientes,

$$\begin{aligned} E[g(Y_1)h(Y_2)] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(y_1)h(y_2)f(y_1, y_2) dy_2 dy_1 \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(y_1)h(y_2)f_1(y_1)f_2(y_2) dy_2 dy_1 \\ &= \int_{-\infty}^{\infty} g(y_1)f_1(y_1) \left[\int_{-\infty}^{\infty} h(y_2)f_2(y_2) dy_2 \right] dy_1 \\ &= \int_{-\infty}^{\infty} g(y_1)f_1(y_1)E[h(Y_2)] dy_1 \\ &= E[h(Y_2)] \int_{-\infty}^{\infty} g(y_1)f_1(y_1) dy_1 \\ &= E[g(Y_1)]E[h(Y_2)] \end{aligned}$$

La demostración para el caso discreto sigue un modo análogo.

5.7 Covarianza de dos variables aleatorias

Intuitivamente consideramos la dependencia de dos variables aleatorias Y_1 y Y_2 como un procesos en el que una de las variables, por ejemplo Y_1, aumenta o disminuye cuando Y_2 cambia. Concentraremos nuestra atención en dos medidas de dependencia: la covarianza entre dos variables aleatorias y su coeficiente de correlación.

Definición 5.10

Si Y_1 y Y_2 son variables aleatorias con medias μ_1 y μ_2, respectivamente, la *covarianza* de Y_1 y Y_2 es

$$Cov(Y_1, Y_2) = E[(Y_1 - \mu_1)(Y_2 - \mu_2)]$$

Cuanto mayor sea el valor absoluto de la covarianza de Y_1 y Y_2, mayor será la dependencia lineal entre Y_1 y Y_2. Los valores positivos indican que Y_1 aumenta cuando Y_2 aumenta; los valores negativos indican que Y_1 disminuye cuando Y_2 aumenta. Un valor de cero indica que no hay dependencia lineal entre Y_1 y Y_2.

Desafortunadamente, es difícil utilizar la covarianza como medida absoluta de dependencia por que su valor depende de la escala de medición. Este problema se puede solucionar al estandarizar su valor y ocupar el *coeficiente de correlación*, ρ, definida como

$$\rho = \frac{Cov(Y_1, Y_2)}{\sigma_1 \sigma_2}$$

donde σ_1 y σ_2 son desviaciones estándar de Y_1 y Y_2, respectivamente. Además el coeficiente de correlación ρ satisface la desigualdad $-1 \leq \rho \leq 1$. Entonces, $\rho > 0$ indica que Y_1 aumenta a mediad que Y_2 aumenta y $\rho = +1$ implica correlación perfecta, con todos los puntos cayendo en una recta con pendiente positiva. Un valor de $\rho = 0$ implica cero covarianza y que no hay correlación. Un coeficiente $\rho < 0$ implica una disminución en Y_2 cuando Y_1 aumenta, y $\rho = -1$ implica correlación perfecta cayendo sobre una recta con pendiente negativa.

Teorema 5.10

Si Y_1 y Y_2 son variables aleatorias con medias μ_1 y μ_2, respectivamente, entonces

$$\begin{aligned} Cov(Y_1, Y_2) &= E[(Y_1 - \mu_1)(Y_2 - \mu_2)] \\ &= E(Y_1 Y_2) - E(Y_1)E(Y_2) \end{aligned}$$

Demostración

$$Cov(Y_1,Y_2) = E[(Y_1-\mu_1)(Y_2-\mu_2)]$$
$$= E(Y_1Y_2 - \mu_1Y_2 - \mu_2Y_1 + \mu_1\mu_2)$$

Del Teorema 5.8, el valor esperado de una suma es igual a la suma de los valores esperados; y del Teorema 5.7, el valor esperado de una constante multiplicado por una función de variables aleatorias es la constante por el valor esperado. Entonces,

$$Cov(Y_1,Y_2) = E(Y_1Y_2) - \mu_1E(Y_2) - \mu_2E(Y_1) + \mu_1\mu_2$$

Como $E(Y_1)=\mu_1$ y $E(Y_2)=\mu_2$, se deduce que

$$Cov(Y_1,Y_2) = E(Y_1Y_2) - E(Y_1)E(Y_2)$$
$$= E(Y_1Y_2) - \mu_1\mu_2$$

Teorema 5.11

Si Y_1 y Y_2 son variables aleatorias independientes, entonces

$$Cov(Y_1,Y_2) = 0$$

Así, las variables aleatorias independientes deben ser no correlacionadas.

Demostración

El Teorema 5.10 establece que

$$Cov(Y_1,Y_2) = E(Y_1Y_2) - \mu_1\mu_2$$

Como Y_1 y Y_2 son independientes, el Teorema 5.9 implica que

$$E(Y_1Y_2) = E(Y_1)E(Y_2) = \mu_1\mu_2$$

y el resultado deseado se deduce de inmediato.

5.8 Valor esperado y varianza de funciones lineales de variables aleatorias

Si $a_1, a_2, \dots, a_n$ son constantes, será necesario calcular el valor esperado y varianza de una función lineal de las variables aleatorias $Y_1, Y_2, \dots, Y_n$. Esta se puede definir de la siguiente manera

$$U_1 = a_1Y_1 + a_2Y_2 + a_3Y_3 + \cdots + a_nY_n = \sum_{i=1}^{n} a_iY_i$$

Teorema 5.12

Sean $Y_1, Y_2, \dots, Y_n$ y $X_1, X_2, \dots, X_m$ variables aleatorias con $E(Y_i)=\mu_i$ y $E(X_i)=\xi_i$. Defina

$$U_1 = \sum_{i=1}^{n} a_iY_i \qquad U_2 = \sum_{j=1}^{m} b_jX_j$$

para las constantes $a_1, a_2, \dots, a_n$ y $b_1, b_2, \dots, b_m$. Entonces se cumple lo siguiente:

a $E(U_1) = \sum_{i=1}^{n} a_i\mu_i$.

b

$$V(U_1) = \sum_{i=1}^{n} a_i^2V(Y_i) + 2\sum\sum_{1\le i<j\le n} a_ia_jCov(Y_i,Y_j)$$

donde la doble suma es para todos los pares (i,j) con $i<j$.

c $Cov(U_1,U_2) = \sum_{i=1}^{n}\sum_{j=1}^{m} a_ib_jCov(Y_i,Y_j)$

Demostración

El teorema consta de tres partes, de las cuales (a) viene directamente de los Teoremas 5.7 y 5.8. Para demostrar (b) recurrimos a la definición de varianza y escribimos

$$V(U_1) = E[U_1 - E(U_1)]^2 = E\left[\sum_{i=1}^{n} a_iY_i - \sum_{i=1}^{n} a_i\mu_i\right]^2$$
$$= E\left[\sum_{i=1}^{n} a_i(Y_i-\mu_i)\right]^2$$
$$= E\left[\sum_{i=1}^{n} a_i^2(Y_i-\mu_i)^2 + \sum_{i=1}^{n}\sum_{\substack{i=1\\ i\neq j}}^{n} a_ia_j(Y_i-\mu_i)(Y_j-\mu_j)\right]$$
$$= \sum_{i=1}^{n} a_i^2E(Y_i-\mu_i)^2 + \sum_{i=1}^{n}\sum_{\substack{i=1\\ i\neq j}}^{n} a_ia_jE\left[(Y_i-\mu_i)(Y_j-\mu_j)\right]$$

Por las definiciones de varianza y covarianza tenemos

$$V(U_1) = \sum_{i=1}^{n} a_i^2V(Y_i) + \sum_{i=1}^{n}\sum_{\substack{i=1\\ i\neq j}}^{n} a_ia_jCov(Y_i,Y_j)$$

Como $Cov(Y_i,Y_j) = Cov(Y_j,Y_i)$, podemos escribir

Demostración

$$V(U_1) = \sum_{i=1}^{n} a_i^2V(Y_i) + 2\sum\sum_{1\le i<j\le n} a_ia_jCov(Y_i,Y_j)$$

Se pueden usar pasos similares para obtener (c). Tenemos

$$Cov(U_1,U_2) = E\left\{[U_1 - E(U_1)]\,[U_2 - E(U_2)]\right\}$$
$$= E\left[\left(\sum_{i=1}^{n} a_iY_i - \sum_{i=1}^{n} a_i\mu_i\right)\left(\sum_{j=1}^{m} b_jX_j - \sum_{j=1}^{m} b_j\xi_j\right)\right]$$
$$= E\left\{\left[\sum_{i=1}^{n} a_i(Y_i-\mu_i)\right]\left[\sum_{j=1}^{m} b_j(X_j-\xi_j)\right]\right\}$$
$$= E\left[\sum_{i=1}^{n}\sum_{j=1}^{m} a_ib_j(Y_i-\mu_i)(X_j-\xi_j)\right]$$
$$= \sum_{i=1}^{n}\sum_{j=1}^{m} a_ib_jE[(Y_i-\mu_i)(X_j-\xi_j)]$$
$$= \sum_{i=1}^{n}\sum_{j=1}^{m} a_ib_jCov(Y_i,X_j)$$

Al observar que $Cov(Y_i,Y_i) = V(Y_i)$, podemos ver que (b) es un caso especial de (c).

5.9 Distribución de probabilidad multinomial

Recuerde del Capítulo 3 que una variable aleatoria binomial resulta de un experimento que consiste en n intentos con dos posibles resultados por intento. Un experimento multinomial es una generalización del experimento binomial.

Definición 5.11

Un *experimento multinomial* posee las siguiente propiedades:

1. El experimento consta de n intento idénticos.
2. El resultado de cada intento cae en una de k clases o celdas.
3. La probabilidad de que el resultado de un solo intento caiga en la celda i es p_i. $i = 1, 2, \dots, k$ y sigue siendo el mismo de un intento a otro. Observe que $p_1 + p_2 + p_3 + \cdots + p_k = 1$.
4. Los intento son independientes.
5. Las variables aleatorias de interés son $Y_1, Y_2, \dots, Y_k$, donde Y_i es igual al número de intento para los cuales el resultado cae en la celda i. Observe que $Y_1 + Y_2 + Y_3 + \cdots + Y_k = n$

Definición 5.12

Suponga que $p_1, p_2, \ldots, p_k$ son tales que $\sum_{i=1}^{k} p_i = 1$, y $p_i > 0$ para $i = 1, 2, \ldots, k$. Se dice que las variables aleatorias $Y_1, Y_2, \ldots, Y_k$ tienen una *distribución multinomial* para parámetro n y $p_1, p_2, \ldots, p_k$ si la función de probabilidad conjunta de $Y_1, Y_2, \ldots, Y_k$ está dada por

$$p(y_1, y_2, \ldots, y_k) = \frac{n!}{y_1! y_2! \cdots y_k!} p_1^{y_1} p_2^{y_2} \cdots p_k^{y_k}$$

donde, para cada $i = 0, 1, 2, \ldots, n$ y $\sum_{i=1}^{k} y_i = n$.

Muchos experimentos en los que aparece una clasificación son multinomiales. Observe que el experimento binomial es un caso especial del experimento multinomial (cuando hay $k = 2$ clases).

Teorema 5.13

Si $Y_1, Y_2, \ldots, Y_k$ tienen una distribución binomial con parámetros n y $p_1, p_2, \ldots, p_k$, entonces

1. $E(Y_i) = np_i, V(Y_i = np_i q_i$.
2. $Cov(Y_s, Y_t) = -np_s p_t$, si $s \neq t$.

Demostración

La distribución marginal de Y_i se puede usar para obtener la media y varianza. Recuerde que Y_i puede ser interpretada como el número de intento que caen en la celda i. Imagine todas las celdas, excluyendo la i, combinadas en una sola celda grande. Entonces cada intento resultará en la celda i o en una celda que no sea la i, con probabilidades p_i y $1 - p_i$, respectivamente. Entonces Y_i posee una distribución de probabilidad marginal binomial. EN consecuencia,

$$E(Y_i) = np_i \qquad V(Y_i) = np_i q_i$$

donde $q_i = 1 - p_i$. Los mismos resultados se pueden obtener al establecer los valores esperados y evaluar. Por ejemplo,

$$E(Y_1) = \sum_{y_1} \sum_{y_2} \cdots \sum_{y_k} y_1 \frac{n!}{y_1! y_2! \cdots y_k!} p_1^{y_1} p_2^{y_2} \cdots p_k^{y_k}$$

Como ya hemos deducido el valor esperado y la varianza de Y_i, dejamos la sumatoria de este valor esperado para el lector interesado.

Demostración

La demostración de la parte 2 usa el Teorema 5.12. Considere el experimento multinomial como una sucesión de n intento independientes y defina, para $s \neq t$,

$$U_i = \begin{cases} 1, & \text{si el intento } i \text{ resulta en clase } s, \\ 0, & \text{de otro modo} \end{cases}$$

y

$$W_i = \begin{cases} 1, & \text{si el intento } i \text{ resulta en clase } t, \\ 0, & \text{de otro modo} \end{cases}$$

Entonces

$$Y_s = \sum_{i=1}^{n} U_i \qquad Y_t = \sum_{j=1}^{n} W_j$$

(Como $U_i = 1$ o 0 dependiendo de si el i-ésimo intento resultó en clase s, Y_s es simplemente la suma de una serie de número 0 y 1. Ocurre un 1 en la suma cada vez que observamos un artículo de la clase s y 0 cada vez que observamos cualquier otra clase. Entonces, Y_s es simplemente el número de veces que se observe la clase s. Una interpretación similar se aplica a Y_t.)
Observe que U_i y W_i no pueden ser iguales a 1 (el i-ésimo artículo no puede estar simultáneamente en las clases s y t). Entonces el producto $U_i W_i$ siempre es igual a cero y $E(U_i W_i) = 0$. Los siguientes resultados nos permiten evaluar $Cov(Y_s, Y_t)$:

$$\begin{aligned} E(U_i) &= p_s \\ E(W_j) &= p_t \\ Cov(U_i, W_j) &= 0 \\ &\text{si } i \neq j \text{ porque los intentos son independientes} \\ Cov(U_i, W_i) &= E(U_i W_i) - E(U_i)E(W_i) = 0 - p_s p_t \end{aligned}$$

Del Teorema 5.12 tenemos entonces

$$\begin{aligned} Cov(Y_s, Y_t) &= \sum_{i=1}^{n} \sum_{j=1}^{n} Cov(U_i, W_j) \\ &= \sum_{i=1}^{n} Cov(U_i, W_i) + \sum_{i \neq j}\sum Cov(U_i, W_j) \\ &= \sum_{i=1}^{n} (-p_s p_t) + \sum_{i \neq j}\sum 0 \\ &= -np_s p_t \end{aligned}$$

La covarianza es negativa, lo que ya se esperaba, porque un gran número de resultados en la celda s forzaría al número de la celda t a ser pequeño.

5.10 Distribución normal bivariante (opcional)

La distribución normal multivariante es muy importante en la teoría moderna de estadística. En general, la función de densidad normal multivariante se define para k variabeles aleatorias continuas, $Y_1, Y_2, \ldots, Y_k$. Debido a su complejidad, presentaremos sólo la función de densidad bivariante ($k = 2$):

$$f(y_1, y_2) = \frac{e^{-Q/2}}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \qquad -\infty < y_1 < \infty, -\infty < y_2 < \infty$$

donde

$$Q = \frac{1}{1-\rho^2}\left[\frac{(y_1-\mu_1)^2}{\sigma_1^2} - 2\rho\frac{(y_1-\mu_1)(y_2-\mu_2)}{\sigma_1\sigma_2} + \frac{(y_2-\mu_2)^2}{\sigma_2^2}\right]$$

La función de densidad normal bivariante es una función de cinco parámetros: $\mu_1, \mu_2, \sigma_1^2, \sigma_2^2$ y ρ.

5.11 Valores esperados condicionales

Definición 5.13

Si Y_1 y Y_2 son dos variables aleatorias cualesquiera, el *valor esperado condicional* de $g(Y_1)$, dado que $Y_2 = y_2$, se define que es

$$E(g(Y_1)|Y_2 = y_2) = \int_{-\infty}^{\infty} g(y_1) f(y_1|y_2) dy_1$$

si Y_1 y Y_2 son continuas conjuntamente y

$$E(g(Y_1)|Y_2 = y_2) = \sum_{\text{toda } y_1} g(y_1) p(y_1|y_2)$$

si Y_1 y Y_2 son discretas conjuntamente.

En genera, el valor esperado condicional de Y_1 dada $Y_2 = y_2$ es una función de y_2. Si ahora hacemos variar Y_2 en todos sus posibles valores, podemos considerar el valor esperado condicional $E(Y_1|Y_2)$ como una función de la variable aleatoria Y_2. Como $E(Y_1|Y_2)$ es una función de la variable aleatoria Y_2, también es una variable aleatorias; y como tal, tiene media y varianza.

Teorema 5.14

Si Y_1 y Y_2 son dos variables aleatorias, entonces

$$E(Y_1) = E[E(Y_1|Y_2)]$$

donde en el lado derecho de la ecuación el valor esperado interior es con respecto a la distribución condicional de Y_1 dada Y_2 y el valor esperado exterior es con respecto a la distribución de Y_2.

Demostración

Suponga que Y_1 y Y_2 son continuas conjuntamente con función de densidad conjunta $f(y_1, y_2)$ y densidades marginales $f_1(y_1)$ y $f_2(y_2)$, respectivamente. Entonces

$$\begin{aligned} E(Y_1) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y_1 f(y_1, y_2) dy_1 dy_2 \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y_1 f(y_1|y_2) f_2(y_2) dy_1 dy_2 \\ &= \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} y_1 f(y_1|y_2) dy_1 \right] f_2(y_2) dy_2 \\ &= \int_{-\infty}^{\infty} E(Y_1|Y_2 = y_2) f_2(y_2) dy_2 \\ &= E[E(Y_1|Y_2)] \end{aligned}$$

La demostración es semejante para el caso discreto.

La varianza condicional de Y_1 dada $Y_2 = y_2$ está definida por analogía con una varianza ordinaria, de nuevo utilizando la densidad condicional o función de probabilidad de Y_1 dada $Y_2 = y_2$ en lugar de la densidad ordinaria o función de probabilidad Y_1. Esto es,

$$V(Y_1|Y_2 = y_2) = E(Y_1^2|Y_2 = y_2) - [E(Y_1|Y_2 = y_2)]^2$$

Como en el caso de la media condicional, la varianza condicional es una función de y_2. Si dejamos que Y_2 tome todos sus valores posibles, podemos definir $V(Y_1|Y_2)$ como una variable aleatoria que es una función de Y_2. Específicamente, si $g(y_2) = V(Y_1|Y_2 = y_2)$ es una función particular del valor observado y_2, entonces $g(Y_2) = V(Y_1|Y_2)$ es la *misma función* de la variable aleatorias, Y_2.

Teorema 5.15

Si Y_1 y Y_2 representan variables aleatorias, entonces

$$V(Y_1) = E[V(Y_1|Y_2)] + V[E(Y_1|Y_2)]$$

Demostración

Como se indicó antes, $V(Y_1|Y_2)$ está dada por

$$V(Y_1|Y_2) = E(Y_1^2|Y_2) - [E(Y_1|Y_2)]^2$$

y

$$E[V(Y_1|Y_2)] = E[E(Y_1^2|Y_2)] - E\left\{[E(Y_1|Y_2)]^2\right\}$$

Por definición,

$$V[E(Y_1|Y_2)] = E\left\{[E(Y_1|Y_2)]^2\right\} - \{E[E(Y_1|Y_2)]\}^2$$

La varianza de Y_1 es

$$\begin{aligned} V(Y_1) &= E[Y_1^2] - [E(Y_1)]^2 \\ &= E\{E[Y_1^2|Y_2]\} - \{E[E(Y_1|Y_2)]\}^2 \\ &= E\{E[Y_1^2|Y_2]\} - E\{[E(Y_1|Y_2]^2\} + E\{[E(Y_1|Y_2]^2\} \\ &\quad - \{E[E(Y_1|Y_2)]\}^2 \\ &= E[V(Y_1|Y_2)] + V[E(Y_1|Y_2)] \end{aligned}$$