

Repaso de Teoría Estadística

Francisco A. Ramírez

Semestre 2021S2

Repaso de Estadística

Este apéndice revisa un poco de la teoría estadística y de distribución que es utilizada en estos apuntes. Más detalles pueden ser encontrados en...

Variables aleatorias discretas

Una **variable aleatoria** es una variable que puede tomar diferentes resultados dependiendo del “estado de la naturaleza”. Por ejemplo, el resultado de lanzar una vez un dado es aleatorio, con posibles resultados 1,2,3,4,5 y 6. Sea denotada una variable aleatoria arbitraria Y . Si y denota el resultado del experimento del dado (y el dado es justo y lanzado aleatoriamente), la **probabilidad** de cada resultado es $1/6$. Esto se puede denotar como

$$P\{Y = y\} = 1/6$$

para $y = 1, 2, \dots, 6$

La función que vincula los posibles resultados (en este caso $y = 1, 2, \dots, 6$) a las probabilidades correspondientes es **la función de masa de probabilidad**, o más generalmente, la función de distribución de probabilidad. Esta puede ser denotada como:

$$f(y) = P\{Y = y\}$$

Note que $f(y)$ no es una función de la variable aleatoria Y , sino de todos sus posibles resultados. La función $f(y)$ tiene la propiedad que, si sumamos sobre los posibles resultados, el resultado es uno. Esto es,

$$\sum_j f(y_j) = 1$$

El **valor esperado** de una variable aleatoria discreta es el promedio ponderado de todos los posibles resultados, donde los pesos corresponden a la probabilidad de un evento particular. Se denota como:

$$E\{Y\} = \sum_j y_j f(y_j)$$

Note que $E\{Y\}$ no necesariamente corresponde a uno de los posibles resultados. En el experimento del dado, por ejemplo, el valor esperado es 3.5.

Una distribución es **degenerada** si está concentrada en solo un punto, esto es, si $P\{Y = y\} = 1$ para un valor particular de y y cero para el resto de los otros valores.

Variables aleatorias continuas

Una **variable aleatoria continua** puede tomar un número infinito de diferentes resultados. Por ejemplo, cualquier valor en el intervalo $[0, 1]$. En este caso, cada resultado individual tiene una probabilidad de cero. En lugar de una función de masa de probabilidad, se define la **función de densidad de probabilidad** $f(y) \geq 0$ como

$$P\{a \leq Y \leq b\} = \int_a^b f(y)dy$$

En un gráfico, $P\{a \leq Y \leq b\}$ es el área bajo la función $f(y)$ entre los puntos a y b . Tomando la integral de $f(y)$ sobre todos los posibles resultados da:

$$\int_{-\infty}^{\infty} f(y)dy = 1$$

Si Y toma valores solamente en un determinado rango, se asume implícitamente que $f(y) = 0$ en cualquier lugar fuera de este rango.

Se puede definir la **función de densidad acumulada** (CDF) como

$$F(y) = P\{Y \leq y\} = \int_{-\infty}^y f(t)dt$$

Es fácil demostrar que $P\{a \leq Y \leq b\} = F(b) - F(a)$.

El **valor esperado** o **media** de una variable aleatoria continua, denotado usualmente como μ , es definido como

$$\mu = E\{Y\} = \int_{-\infty}^{\infty} yf(y)dy$$

Otra medida de ubicación es la **mediana**, que es el valor m para el que se tiene

$$P\{Y \leq m\} \geq 1/2$$

y

$$P\{Y \geq m\} \leq 1/2$$

De tal manera 50% de las observaciones están debajo de la mediana y 50% por encima. La **moda** es simplemente el valor para el cual $f(y)$ es máxima.

Una distribución es **simétrica** alrededor de la media si $f(\mu - y) = f(\mu + y)$. En este caso la media y la mediana de la distribución son idénticas.

Expectativas y momentos

Si Y y X son variables aleatorias y a y b son constantes, se tiene que

$$E\{aY + bX\} = aE\{Y\} + bE\{X\}$$

lo que muestra que el valor esperado es un operador lineal. Este resultado no se sostiene necesariamente si se consideran transformaciones no lineales de una variable aleatoria. Para una función no lineal g , no se tiene en general que $E\{g(Y)\} = g(E\{Y\})$. Si g es cóncava ($g''(Y) < 0$), la **desigualdad de Jensen**.

Por ejemplo, $E\{\log(Y)\} \leq \log(E\{Y\})$. La implicación de este es que no se puede determinar el valor esperado de una función de Y del valor esperado de Y solamente. Por definición se cumple.

$$E\{g(Y)\} = \int_{-\infty}^{\infty} g(y)f(y)dy$$

La **varianza** de una variable aleatoria, denotada por σ^2 , es una medida de la dispersión de la distribución. Es definida como:

$$\sigma^2 = V\{Y\} = E\{(Y - \mu)^2\}$$

igual al valor esperado de las desviación respecto de la media al cuadrado. Es algunas veces llamada **segundo momento central**. Un resultado útil es:

$$E\{(Y - \mu)^2\} = E\{Y^2\} - 2E\{Y\}\mu + \mu^2 = E\{Y^2\} - \mu^2$$

donde $E\{Y^2\}$ es el segundo momento. Si Y tiene una distribución discreta, su varianza es determinada por:

$$V\{Y\} = \sum_j (y_j - \mu)^2 f(y_j)$$

donde j indexa los diferentes resultados posibles. Para una distribución continua se tiene:

$$V\{Y\} = \int_{-\infty}^{\infty} (y - \mu)^2 f(y)dy$$

Usando estas definiciones, es fácil verificar que:

$$V\{aY + b\} = a^2 V\{Y\}$$

donde a y b son constantes arbitrarias. Con frecuencia se utiliza la **desviación estándar** de una variable aleatoria, denotada por σ , y definida como la raíz cuadrada de la varianza. La desviación estándar es expresada en las mismas unidades que Y .

En la mayoría de los casos la distribución de una variable aleatoria no es completamente descrita por su media y su varianza, y se puede definir el momento k central como:

$$E\{(Y - \mu)^k\}, \quad k = 1, 2, 3, \dots$$

En particular, el tercer momento central es una medida de asimetría de la distribución alrededor de su media, mientras que el cuarto momento que mide el apuntamiento de la distribución. Típicamente, la **asimetría o skewness** es definido como $S \equiv E\{(Y - \mu)^3\}/\sigma^3$, mientras que la **kurtosis** es definida como $K \equiv E\{(Y - \mu)^4\}/\sigma^4$. La kurtosis de una distribución normal es 3, tal que $K - 3$ es referida como **exceso de kurtosis**. Una distribución con exceso de kurtosis positiva es llamada leptocúrtica.

Distribuciones Multivariadas

La **función de distribución conjunta** de dos variables aleatorias Y y X , denotada por $f(y, x)$ es definido como:

$$P\{a_1 < Y < b_1, a_2 < X < b_2\} = \int_{a_1}^{b_1} \int_{a_2}^{b_2} f(y, x)dydx.$$

Si Y y X son **independientes**, se sostiene que $f(y, x) = f(y)f(x)$, tal que:

$$P\{a_1 < Y < b_1, a_2 < X < b_2\} = P\{a_1 < Y < b_1\}P\{a_2 < X < b_2\}$$

En general, la **distribución marginal** de Y es caracterizado por la función de densidad

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

Esto implica que el valor esperado de Y es dado por:

$$E\{Y\} = \int_{-\infty}^{\infty} y f(y) dy = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y f(y, x) dx dy.$$

La **covarianza** entre Y y X es una medida de la dependencia lineal entre dos variables. Es definido como:

$$\sigma_{xy} = cov\{Y, X\} = E\{(Y - \mu_y)(X - \mu_x)\}.$$

donde $\mu_y = E\{Y\}$ y $\mu_x = E\{X\}$. El **coeficiente de correlación** es dada por la covarianza estandarizada por dos desviaciones estándar, esto es,

$$\rho_{yx} = \frac{cov\{Y, X\}}{\sqrt{V\{Y\}V\{X\}}} = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$$

El coeficiente de correlación está siempre entre -1 y 1 y no es afectado por el escalamiento de las variables. El cuadrado del coeficiente de correlación entre 0 y 1 y describe la proporción de varianza en común entre Y y X . Puede ser multiplicado por 100 y expresado como porcentaje. Si $cov\{Y, X\} = 0$, Y y X se dice que **no están correlacionada**.

Cuando a, b, c, d son constantes, se sostiene que:

$$cov\{aY + b, cX + d\} = ac \, cov\{Y, X\}$$

Más aun,

$$cov\{aY + bX, X\} = a \, cov\{Y, X\} + b \, cov\{X, X\} = a \, cov\{Y, X\} + bV\{X\}$$

También se tiene que dos variables Y y X son perfectamente correlacionado ($\rho_{xy} = 1$) si $Y = aX$ para valores diferente de cero de a . Si Y y X estan correlacinados, la varianza de una función linela de Y y X depende de su covarianza. En particular,

$$V\{aY + bX\} = a^2V\{Y\} + b^2V\{X\} + 2ab \, cov\{Y, X\}$$

Si se considera un vector K -dimensional de vector de variables aleatorias, $\vec{Y} = (Y_1, \dots, Y_K)'$, se puede definir su vector de expectativas:

$$E\{\vec{Y}\} = \begin{pmatrix} E\{Y_1\} \\ \vdots \\ E\{Y_K\} \end{pmatrix}$$

y su matriz varianza covarianza (o simplemente **matriz de covarianza**) como:

$$V\{\vec{Y}\} = \begin{pmatrix} V\{Y_1\} & \cdots & cov\{Y_1, Y_K\} \\ \vdots & \ddots & \vdots \\ cov\{Y_K, Y_1\} & \cdots & V\{Y_K\} \end{pmatrix}$$

Note que esta matriz es simétrica. Si consideramos una o más combinaciones lineales de los elementos en \vec{Y} , es decir $R\vec{Y}$, donde R es de dimensión $J \times K$ se tiene que:

$$V\{R\vec{Y}\} = RV\{\vec{Y}\}R'$$

Distribuciones Condicionales

Una distribución condicional describe la distribución de una variable, diga Y , dada la realización de otra variable X . Por ejemplo, si se lanza un dado dos veces, X puede denotar la realización del primer dado y Y puede denotar el total de dos dados. Entonces se puede estar interesado en la distribución de Y condicional a la realización del primer dado. Por ejemplo, cuál es la probabilidad de lanzar 7 en total si la realización del primer dado es 3? O una realización de 3 o menos? La distribución condicional está implícita por la distribución conjunta de dos variables. Se define,

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

Si Y y X son independientes, inmediatamente sigue que $f(y|x) = f(y)$. De la definición anterior sigue que:

$$f(y,x) = f(y|x)f(x),$$

que dice que la distribución conjunta de dos variables puede ser descompuesta en el producto de una distribución condicional y una distribución marginal. De manera similar, se puede escribir:

$$f(y,x) = f(x|y)f(y).$$

La **esperanza condicional** de Y dado $X = x$ es el valor esperado de Y de la distribución condicional. Esto es,

$$E\{Y|X=x\} = E\{Y|x\} = \int yf(y|x)dy$$

La expectativa condicional es una función de x , a menos que Y y X son independientes. De manera similar, se puede definir la varianza condicional como:

$$V\{Y|x\} = \int (y - E\{Y|x\})^2 f(y|x)dy$$

que puede ser escrita como:

$$V\{Y|x\} = E\{Y^2|x\} - (E\{Y|x\})^2.$$

Se tiene que:

$$V\{Y\} = E_x\{V\{Y|X\}\} + V_x\{E\{Y|X\}\},$$

donde E_x y V_x denotan el valor esperado y varianza, respectivamente, obtenida de la distribución marginal de X . Los términos $V\{Y|X\}$ y $E\{Y|X\}$ son funciones de la variable aleatoria X y en consecuencia son variables aleatorias en sí mismas.

Considerese la relación entre dos variables aleatorias Y y X , donde $E\{Y\} = 0$. Entonces, se tiene que Y y X están **no correlacionada** si:

$$E\{YX\} = cov\{Y, X\} = 0$$

Si Y es **independiente en media condicional** de X , significa que:

$$E\{Y|X\} = E\{Y\} = 0.$$

Este resultado es más fuerte que correlación cero porque $E\{Y|X\} = 0$ implica que $E\{Yg(X)\} = 0$ para cualquier función g . Si Y y X son **independientes**, este es también fuerte e implica que:

$$E\{g_1(Y)g_2(Y)\} = E\{g_1(Y)|E\{g_2(Y)\}\},$$

para funciones arbitrarias g_1 y g_2 . Es fácilmente verificable que esto implica independencia en media condicional y correlación cero. Note que $E\{Y|X\} = 0$ no necesariamente implica que $E\{X|Y\} = 0$.

La distribución normal

En econometría, la **distribución normal** juega en un rol central. La función de densidad para una distribución normal con media μ y varianza σ^2 es dada por:

$$f(y) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2} \frac{(y - \mu)^2}{\sigma^2}\right\}$$

donde se puede escribir como $Y \sim N(\mu, \sigma^2)$. Es fácil verificar que la distribución normal es simétrica.

Una distribución normal estándar es obtenida con $\mu = 0$ y $\sigma = 1$. Note que la variable estandarizada $(Y - \mu)/\sigma$ es $N(0, 1)$ si $Y \sim N(\mu, \sigma^2)$. La densidad de una distribución normal, típicamente denotada por ϕ , está dada por:

$$\phi(y) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}y^2\right\}$$

Una propiedad útil de la distribución normal es que una función lineal de una variable normal es también normal. Esto es, si $Y \sim N(\mu, \sigma^2)$, entonces:

$$aY + n \sim N(a\mu + b, a^2\sigma^2)$$

La función de densidad acumulada de la distribución normal no tienen forma cerrada. Se tienen que:

$$P\{Y \leq y\} = P\left\{\frac{Y - \mu}{\sigma} \leq \frac{y - \mu}{\sigma}\right\} = \Phi\left(\frac{y - \mu}{\sigma}\right) = \int_{-\infty}^{(y - \mu)/\sigma} \phi(t) dt$$

donde Φ denota la CDF de una distribución normal estándar. Note que $\Phi(y) = 1 - \Phi(-y)$ debido a la simetría.

La simetría también implica que el tercer momento central de una distribución normal es cero. Puede mostrarse que el cuarto momento de una distribución normal está dada por:

$$E\{(Y - \mu)^4\} = 3\sigma^2.$$

Típicamente estas propiedades del tercer y cuarto momento son explotadas en los contrastes de normalidad.

Si (Y, X) tiene una **distribución normal bivariada** con vector de medias $\mu = (\mu_y, \mu_x)$ y matriz de covarianza:

$$\Sigma = \begin{pmatrix} \sigma_y^2 & \sigma_{yx} \\ \sigma_{yx} & \sigma_x^2 \end{pmatrix}$$

denotado por $(Y, X)' \sim N(\mu, \Sigma)$, la distribución conjunta está dada por:

$$f(y, x) = f(y|x)f(x)$$

donde ambas la **densidad condicional** de Y dado X y la **densidad marginal** de X son normales. La densidad condicional está dada por:

$$f(y|x) = \frac{1}{\sqrt{2\pi\sigma_{y|x}^2}} \exp\left\{-\frac{1}{2} \frac{(y - \mu_{y|x})^2}{\sigma_{y|x}^2}\right\}$$

donde $\mu_{y|x}$ es la **expectativa condicional** de Y dado X , dada por:

$$\mu_{y|x} = \mu_y + (\sigma_{yx}/\sigma_x^2)(x - \mu_x),$$

y $\sigma_{y|x}^2$ es la varianza condicional de Y dado X ,

$$\sigma_{y|x}^2 = \sigma_y^2 - \frac{\sigma_{yx}^2}{\sigma_x^2} = \sigma_y^2(1 - \rho_{yx}^2)$$

con ρ_{yx} denotando el coeficiente de correlación entre Y y X . Estos resultados tienen implicaciones importantes. Primero, si dos (o más) variables tienen una distribución conjunta normal, todas las distribuciones marginales y condicionales son también normales. Segundo, la expectativa condicional de una variable dada las demás es una función lineal (con un término intercepto). Tercero, si $\rho_{yx} = 0$, entonces $f(y|x) = f(y)$, tal que

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

y Y y X son independientes. Entonces si Y y X tienen una distribución conjunta con correlación cero, entonces ellas son automáticamente independientes. Recuerde que en general independencia es más fuerte que no correlación.

Otro resultado importante es que una función lineal de variables normales es también normal, que es, si $(Y, X)' \sim N(\mu, \Sigma)$, entonces

$$aY + bX \sim N(a\mu_y + b\mu_x, a^2\sigma_y^2 + b^2\sigma_x^2 + 2ab\sigma_{yx})$$

Estos resultados pueden ser generalizados a una distribución normal de K variables. Si el vector K dimensional \vec{Y} tiene distribución normal con media el vector μ y matriz de covarianza Σ , esto es,

$$\vec{Y} \sim N(\mu, \Sigma)$$

se tiene que la distribución de $R\vec{Y}$, donde R es una matriz $J \times K$, es una distribución normal J-variante, dada por:

$$R\vec{Y} \sim N(R\mu, R\Sigma R')$$

En modelos con variables dependientes limitadas se encuentran con frecuencia formas de **truncamiento**. Si Y tienen una densidad $f(y)$, la distribución truncada por debajo en un punto c ($Y \geq c$) es dada por:

$$f(y|Y \geq c) = \frac{f(y)}{P\{Y \geq c\}} \quad \text{si } y \geq c, \quad 0 \text{ lo contrario}$$

Si Y es una variable normal estándar, la distribución truncada de $Y \geq c$ tiene media

$$E\{Y|Y \geq c\} = \lambda_1(c)$$

donde:

$$\lambda_1(c) = \frac{\phi(c)}{1 - \Phi(c)},$$

y varianza:

$$V\{Y|Y \geq c\} = 1 - \lambda_1(c)[\lambda_1(c) - c].$$

Si la distribución es truncada por arriba ($Y \leq c$), se tiene que:

$$E\{Y|Y \leq c\} = \lambda_2(c),$$

con

$$\lambda_2(c) = \frac{-\phi(c)}{\Phi(c)}$$

Si Y tiene una densidad normal con media μ y varianza σ^2 , la distribución truncada $Y \geq c$ tiene media:

$$E\{Y|Y \geq c\} = \mu + \sigma\lambda_1(c^*) \geq \mu$$

donde $c^* = (c - \mu)/\sigma$, y, de manera similar,

$$\begin{aligned} E\{Y|X \geq c\} &= \mu_y + \left(\frac{\sigma_{yx}}{\sigma_x^2} \right) [E\{X|X \geq c\} - \mu_x] \\ &= \mu_y + \left(\frac{\sigma_{yx}}{\sigma_x} \right) \lambda_1(c^*) \end{aligned}$$

Otras distribuciones relacionadas

Más allá de la distribución normal, existen otras distribuciones importantes. Primero, se define la **distribución Chi cuadrada** como sigue. Si Y_1, \dots, Y_J es un conjunto de variables independientes con distribución normal estándar, se tiene que:

$$\xi = \sum_{j=1}^J Y_j^2$$

tiene una distribución Chi cuadrada con J grados de libertad. Se denota como $\xi \sim \chi_J^2$. De manera general, si un conjunto de variables normales independientes con media μ y varianza σ^2 , se sigue que:

$$\xi = \sum_{j=1}^J \frac{(Y_j - \mu)^2}{\sigma^2}$$

es Chi cuadrada con J grados de libertad. Aun de forma mas general, si $\vec{Y} = (Y_1, \dots, Y_J)'$ es un vector de variables aleatorias que tiene una distribución normal conjunta con vector de medias μ y matriz de covarianzas (no singular), sigue que:

$$\xi = (\vec{Y} - \mu)' \Sigma^{-1} (\vec{Y} - \mu) \sim \chi_J^2$$

Si x_i tiene una distribución Chi cuadrada con J grados de libertad, se puede mostrar que $E\{\xi\} = J$ y $V\{\xi\} = 2J$.

A continuación, considere la **distribución t** (o distribución de Student). Si X tiene una distribución normal estándar, $X \sim N(0, 1)$, y $\xi \sim \chi_J^2$, y si X y ξ son independientes, la razón sigue una distribución t con J grados de libertad. Como la distribución normal estándar, la distribución t es simétrica y centrada en cero, pero tiene colas más “gruesas”, particularmente para J pequeño. Si J se aproxima a infinito, la distribución t se aproxima a la distribución normal.

Si $\xi_1 \sim \chi_{J_1}^2$ y $\xi_2 \sim \chi_{J_2}^2$, y si ξ_1 y ξ_2 son independientes, se tiene que la razón:

$$f = \frac{\xi_1/J_1}{\xi_2/J_2}$$

tiene una **distribución F** con J_1 y J_2 grados de libertad en el numerador y el denominador respectivamente. La razón inversa:

$$\frac{\xi_2/J_2}{\xi_1/J_1}$$

también tiene una distribución F , pero con J_2 y J_1 grados de libertad respectivamente. La distribución F es por tanto la distribución de la razón de dos variables Chi cuadradas independientes, divididos por sus respectivos grados de libertad. Cuando $J_1 = 1$, ξ_1 es una variable normal cuadrada, por ejemplo $\xi_1 = X^2$, entonces:

$$t^2 = \left(\frac{X}{\sqrt{\xi/J_2}} \right)^2 = \frac{\xi_1}{\xi_2/J_2} = f \sim F_{J_2}^1$$

Entonces con un grado de libertad en el numerador, la distribución F es justamente el cuadrado de la distribución t . Si J_2 es grande, la distribución de

$$J_1 f = \frac{\xi_1}{\xi_2/J_2}$$

es bien aproximada por una distribución Chi cuadrada con J_1 grados de libertad. Para un J_2 grande el denominador es insignificante.

Finalmente, considere la **distribución log-normal**. Si $\log Y$ tiene una distribución normal con media μ y varianza σ^2 , entonces $Y > 0$ tiene una distribución llamada log-normal. La densidad log-normal es usualmente utilizada para describir la distribución poblacional del ingreso (laboral) o la distribución de retornos de activos. Mientras que $E\{\log Y\} = \mu$, se cumple que:

$$E\{Y\} = \exp\left\{\mu + \frac{1}{2}\sigma^2\right\}$$

.