Teoría Estadística Elemental I
Teoría (resumida) del 2^{do} Tema

Raúl Jiménez
Universidad Carlos III de Madrid

Noviembre 2011

Consideremos el lanzamiento de un dado, $\Omega = \{1, 2, 3, 4, 5, 6\}$, y supongamos que apostamos al resultado de tal manera que nuestra ganancia es

$$-1 \text{ si el resultado es impar,}$$

$$0 \text{ si el resultado es 2 o 4,}$$

$$2{,}75 \text{ si el resultado es 6.}$$

Se entiende que ganancias negativas son pérdidas positivas. Si el resultado es ω, la ganancia puede expresarse como $X(\omega)$, donde $X : \Omega \to R$ es la función definida por

$$X(1) = X(3) = X(5) = -1$$

$$X(2) = X(4) = 0$$

$$X(6) = 2{,}75$$

X es un ejemplo de una variable aleatoria discreta, las cuales son nuestro actual objeto de estudio.

Definición y ejemplos

Dado un espacio de probabilidad $(\Omega, \mathcal{F}, P)$, una **variable aleatoria discreta** es una función $X : \Omega \to R$ tal que

1. Su conjunto de imágenes $X(\Omega) = \{x \in \mathbb{R} : X(\omega) = x, \text{ para algún } \omega \in \Omega\}$ es un conjunto numerable. Es decir, $X(\Omega) = \{x_i : i \in I\}$, para algún conjunto (finito o infinito) de índices $I \subset \mathbb{N}$.

2. Para todo $x \in \mathbb{R}$ se verifica
$$\{\omega \in \Omega : X(\omega) = x\} \in \mathcal{F}. \tag{1}$$

La primera condición se refiere al hecho de que X toma solamente valores en un conjunto numerable de $\mathbb{R}$. La segunda condición puede parecer oscura al primer vistazo. La idea es que podamos dar probabilidades de que la variable tome cualquiera de sus posibles valores, pero esta probabilidad puede no estar definida si no se satisface (1) para algún x (la probabilidad sólo tiene que estar definida para los eventos pertenecientes a $\mathcal{F}$). Consideremos $\Omega = \mathbb{N}$ y la σ-álgebra $\mathcal{F}$ formada por el vacío, los números pares positivos ($Pares$), los impares positivos ($Impares$) y $\mathbb{N}$.

Sea $P : \mathcal{F} \to [0,1]$ la medida de probabilidad definida por $P(Pares) = P(Impares) = 1/2$ y $X : \Omega \to \mathbb{R}$ la función identidad $X(\omega) = \omega$. Note que

$$\begin{aligned} \{\omega \in \Omega : X(\omega) = x\} &= \{x\} \text{ si } x \in \mathbb{N} \\ &= \emptyset \quad \text{en caso contrario} \end{aligned}$$

Así que no podemos decir con que probabilidad la variable toma el valor 2 o 4, sólo sabemos que es par con probabilidad 1/2 y un número natural con probabilidad 1. Como mencionamos, nos interesa la probabilidad de que la variable tome cualquiera de sus posible valores. A eso apunta la siguiente definición.

Función de masa de probabilidad. La función de masa de probabilidad (fmp) de la variable aleatoria discreta X es la función $p_X : \mathbb{R} \to [0,1]$ definida por

$$p_X(x) = P(X = x) = P(\{\omega \in \Omega : X(\omega) = x\})$$

Ya que $P(X = x)$ es la probabilidad de que X tome el valor x, se tiene que

$$P(X = x) \geq 0 \quad \text{para todo} \quad x \in \mathbb{R}$$

$$P(X = x) = 0 \quad \text{para todo} \quad x \notin X(\Omega).$$

Además, y esta es otra importante propiedad de las funciones de masa de probabilidad,

$$\sum_x P(X = x) = \sum_{x \in X(\Omega)} P(X = x) = P(\Omega) = 1. \tag{2}$$

Esta propiedad caracteriza las funciones de masa de probabilidad de las variables aleatorias discretas en el sentido siguiente:

Si A es un conjunto numerable de $\mathbb{R}$ y $\pi : A \to \mathbb{R}$ satisface

$$\pi \geq 0 \quad \text{y} \quad \sum_{x \in A} \pi(x) = 1, \tag{3}$$

entonces π es la fmp de una variable aleatoria X asociada a un espacio de probabilidades $(\Omega, \mathcal{F}, P)$ tal que $X(\Omega) = A$.

Otro concepto muy importante en teoría de probabilidades es el de **función de distribución de una variable aleatoria**:

La función de distribución de una variable aleatoria X es la función $F_X : \mathbb{R} \to [0,1]$ definida por

$$F_X(x) = P(X \leq x). \tag{4}$$

A partir de la función de distribución de una variable aleatoria discreta podemos calcular su fmp y viceversa. Específicamente,

$$F_X(x) = \sum_{x_i \leq x} P(X = x_i) \quad \text{y} \quad P(X = x) = F(x) - \lim_{\epsilon \to 0^+} F(x - \epsilon)$$

En general, basta determinar una de estas dos funciones para calcular probabilidades de los eventos asociados a una variable aleatoria, que en general son del tipo

$$P(X \in A) = P(\{\omega \in \Omega : X(w) \in A\}) = \sum_{x_i \in A} P(X = x_i)$$

Si F es la función de distribución de una variable aleatoria escribimos $X \sim F$ y si X y Y son variables aleatorias con la misma función de distribución decimos que son igualmente distribuídas y escribimos $X \sim Y$. Veamos algunos ejemplos clásicos:

Distribución Bernoulli. Decimos que X es una variable aleatoria con distribución Bernoulli de parámetro p, y escribimos $X \sim \text{Bernoulli}(p)$, si

$$P(X=1)=p, \quad P(X=0)=1-p, \quad \text{para algún} \;\; p \in [0,1].$$

En el argot, p se entiende como la probabilidad de éxito de un determinado suceso en un experimento y $q = 1-p$ la del fracaso o éxito del complemento.

Distribución Binomial. Decimos que X tiene distribución Binomial con parámetros n y p, $X \sim \text{Bin}(n,p)$, si

$$P(X=k) = \binom{n}{k} p^k q^{n-k}, \quad \text{para} \;\; k=0,1,\ldots,n. \tag{5}$$

Para demostrar que la función definida en (5) satisface (3) es necesario usar la fórmula del binomio de Newton. Así,

$$\sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} = (p+q)^n = 1.$$

$P(X=k)$ en (5) es la probabilidad de observar un total de k éxitos en n experimentos independientes, cada uno con probabilidad p de que sea éxito.

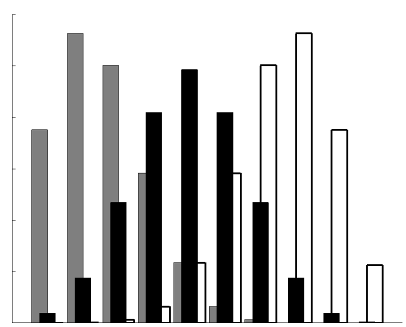

Figura 1: **fmp de una binomial para tres parámetros distintos**

Distribución Geométrica. Decimos que la distribución de X es Geométrica con parámetro p, $X \sim \text{Geo}(p)$, si

$$P(X=n) = q^{k-1}p, \quad \text{para} \;\; n=1,2,3,\ldots. \tag{6}$$

Note que

$$\sum_{k=1}^{\infty} pq^{k-1} = p\sum_{k=0}^{\infty} q^k = p\frac{1}{1-q} = 1$$

La probabilidad (6) es la de requerir exactamente n repeticiones independientes de un mismo experimento hasta observar el primer éxito. Igual que antes, p es la probabilidad de éxito en un experimento y $q = 1-p$.

Distribución Hipergeométrica. X es una variable Hipergeométrica de parámetros N, N_A y n, con $N > \max(N_A, n)$, si

$$P(X = k) = \frac{\binom{N_A}{k}\binom{N-N_A}{n-k}}{\binom{N}{n}}, \quad \text{para} \quad k = 0, 1, \ldots, \min(N_A, n) \tag{7}$$

Para demostrar que esta es una función de masa de probabilidad, es necesario hacer uso de (**??**). La probabilidad (7) es la de extraer k elementos de un conjunto $A \subset \Omega$, cuando se extraen aleatoriamente y sin reposición n elementos de Ω. Aquí $|A| = N_A$ y $|\Omega| = N$.

Distribución de Poisson. X es Poisson de parámetro $\lambda > 0$, $X \sim \text{Poisson}(\lambda)$, si

$$P(X = k) = \frac{1}{k!}\lambda^k e^{-\lambda}, \quad \text{para} \quad k = 0, 1, 2, \ldots. \tag{8}$$

Haciendo uso del desarrollo en serie de Taylor de la función exponencial, es sencillo comprobar que la función definida en (8) satisface (3). Cuando n es grande y p pequeño, haciendo $\lambda = np$, la aproximación

$$\binom{n}{k} p^k q^{n-k} \approx \frac{1}{k!}\lambda^k e^{-\lambda}$$

es buena. De manera que el modelo Poisson puede entenderse como un caso límite del Binomial, cuando el número de experimentos es grande y la probabilidad de éxito de cada experimento es pequeña. Al final del curso formalizamos esta idea.

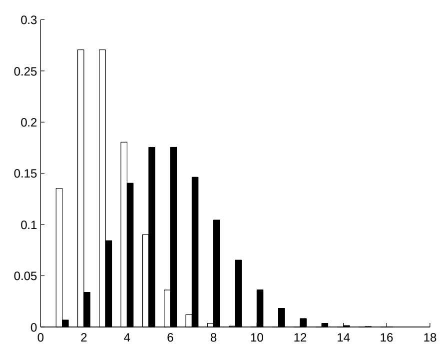

Figura 2: **fmp de una Poisson para tres parámetros distintos**

Esperanza

Consideremos un dado justo. Si este es lanzado un número grande de veces, cada posible resultado aparecerá alrededor de un sexto de las veces y el promedio del número observado será aproximadamente

$$1(1/6) + 2(1/6) + \ldots + 6(1/6) = 3,5$$

El concepto, en su forma más general, lleva a la siguiente definición

Definición. Sea X es una variable aleatoria discreta. La esperanza de X, denotada por $E(X)$ y también llamada valor esperado de X, es el número definido por

$$E[X] = \sum_x \; x \; P(X = x)$$

siempre y cuando la serie converja.

Teorema de transferencia. Si X es una variable discreta y $g : R \to R$ entonces la esperanza de $Y = g(X)$ es

$$E[Y] = E[g(X)] = \sum_x g(x)P(X = x)$$

Prueba

$$\begin{aligned} E[Y] &= \sum_y \; y \; P(Y = y) \\ &= \sum_y \; y \left[\sum_{\{x:g(x)=y\}} P(X = x) \right] \\ &= \sum_y \sum_{\{x:g(x)=y\}} y \; P(X = x) \\ &= \sum_x g(x)P(X = x) \end{aligned}$$

Otra importante valor asociado a una variable X es su varianza $Var(X)$, la cual es una medida de dispersión de la variable en torno a su esperanza. Formalmente, la **varianza de una variable aleatoria** X se define como el valor esperado de la variable $(X - \mu)^2$, siendo μ la esperanza de X. Es decir,

$$\begin{aligned} Var(X) &= E([X - \mu]^2) \\ &= \sum_x (x - \mu)^2 P(X = x) \end{aligned} \tag{9}$$

Proposición 1. $Var(X) = E[X^2] - (E[X])^2$

Prueba

$$\begin{aligned} Var(X) &= E([X - \mu]^2) \\ &= \sum_x (x - \mu)^2 P(X = x) \\ &= \sum_x (x^2 - 2x\mu + \mu^2) P(X = x) \\ &= \sum_x x^2 P(X = x) - 2\mu \sum_x xP(X = x) + \mu^2 \sum_x P(X = x) \\ &= E[X^2] - 2\mu^2 + \mu^2 \\ &= E[X^2] - \mu^2 \\ &= E[X^2] - (E[X])^2 \end{aligned}$$

Dos resultados importantes que podemos demostrar de forma sencilla con la fórmula de tranferencia son:

1. $E(aX+b) = aE(X) + b$.
2. $Var(aX+b) = a^2 Var(X)$.

Esperanza Condicional

Sea X una variable aleatoria discreta y B un evento asociados al mismo espacio de probabilidad. Supongamos que $P(B) > 0$. La **Esperanza Condicional** de X dado el evento B, la cual denotaremos por $E(X|B)$, es el valor esperado asociado a la función de masa de probabilidad condicional

$$P(X = x|B) = \frac{P(\{\omega : X(\omega) = x\} \cap B)}{P(B)}.$$

Esto es,

$$E[X|B] = \sum_x \; x \; P(X = x|B)$$

El siguiente resultado es un versión de la fórmula de probabilidad total (**??**) para valores esperados y de similar utilidad.

Fórmula de particionamiento. Si X es una v.a. discreta y $B_1, B_2, \ldots$ son una partición del espacio muestral, con $P(B_i) > 0$ para cada i, entonces

$$E[X] = \sum_i \; E[X|B_i]P(B_i)$$

Prueba de la fórmula

$$\begin{aligned}
E[X] &= \sum_{i\geq 1} E[X|B_i]P(B_i) \\
&= \sum_{i\geq 1} \left[\sum_x \; x \; P(X = x|B) \right] P(B_i) \\
&= \sum_{i\geq 1} \sum_x \; x \; P(\{X = x\} \cap B_i) \\
&= \sum_x \; x \; P(\{X = x\} \cap (\cup_{i\geq 1} B_i)) \\
&= \sum_x \; x \; P(X = x)
\end{aligned}$$

Ejemplo. Una moneda es lanzada repetidamente. Sea p la probabilidad de obtener cara en cada lanzamiento, con $0 < p = 1 - q < 1$. Vamos a calcular la longitud esperada de la racha inicial (*i.e.* el número de resultados iguales y consecutivos al primero).

Sea H el evento el primer lanzamiento es cara y H^c el evento el primer lanzamiento es sello. El par H, H^c forma una partición del espacio muestral. Si X es la longitud de la racha inicial, es fácil verificar que

$$P(X = k|H) = p^{k-1}q \;\; \text{para} \;\; k = 1, 2, \ldots$$

ya que si H ocurre entonces $X = k$ ocurre sí y sólo sí el primer lanzamiento es seguido por exactamente $k-1$ caras y después un sello. Similarmente,

$$P(X = k|H^c) = q^{k-1}p \;\; \text{para} \;\; k = 1, 2, \ldots$$

Es decir, las distribuciones condicionales son geométricas, así que

$$E[X|H] = \frac{1}{q} \;\text{ y }\; E[X|H^c] = \frac{1}{p}$$

Usando la fórmula de particionamiento obtenemos

$$E[X] = E[X|H]P(H) + E[X|H^c]P(H^c) = \frac{1}{q}p + \frac{1}{p}q = \frac{1}{pq} - 2$$