Universidad de Chile.		Rodrigo Assar
Facultad de Ciencias Físicas y Matemáticas	$MA34B-3$	Andrés Iturriaga
Departamento de Ingeniería Matemática.		Víctor Riquelme

Distribución Multinomial

Resumen

En el presente artículo se presenta una introducción a la *distribución multinomial*. Se trata la distribución de un vector aleatorio de frecuencias, la relacion entre la dist. multinomial y la dist. binomial, esperanza, varianza y ejemplos.

1. Introducción

Considere una población con artículos pertenecientes a k categorias distintas. Supongase que se extrae un artículo de dicha población, y se quiere ver de que tipo es. Podemos modelar lo anterior por una variable aleatoria X, que indica a que categoría pertenece el artículo. Llamemos $y_1, \ldots, y_k$ a las distintas categorias. Entonces X toma valores en el conjunto $\{y_1, \ldots, y_k\}$, y definimos las probabilidades $p_i = \mathbb{P}(X = y_i)$. Es claro que $\sum_{i=1}^{k} p_i = 1$.

Supongase ahora que se toma una MAS de tamaño n con reposición (o si el tamaño de la población es "grande" da lo mismo). Definamos el vector aleatorio $\vec{N} = (N_1, \ldots, N_k)$ que indica en cada componente $i-$ésima la frecuencia de ocurrencia del tipo y_i en la MAS. Entonces la distribucion de $\vec{N}$ es una multinomial de parametros n y $\vec{p} = (p_1, \ldots, p_k)$:

$$\mathbb{P}(N_1 = n_1, \ldots, N_k = n_k) = \binom{n}{n_1, \ldots, n_k} {p_1}^{n_1} \ldots {p_k}^{n_k} \mathbf{1}_{\{\sum_{i=1}^{k} n_i = n\}}(n_1, \ldots, n_k)$$

$$\text{con } \binom{n}{n_1, \ldots, n_k} = \frac{n!}{n_1! n_2! \ldots n_k!}$$

2. De donde viene?

La deducción tiene dos patas:

- La parte de las probabilidades
- El coeficiente que acompaña, asociado a $(n_1, \ldots, n_k)$, que llamaremos $\alpha_{(n_1,\ldots,n_k)}$.

Primero, la parte de las probabilidades es relativamente facil. Es convencerse que la probabilidad de obtener una configuración de n_1 objetos de tipo y_1, n_2 objetos de tipo y_2,..., n_k objetos de tipo y_k (si no importara el orden en que salen con respecto al total) es ${p_1}^{n_1} \ldots {p_k}^{n_k}$.

La segunda parte tiene que ver con la cantidad de las configuraciones anteriores posibles. Para ello, definamos:

- α_1={numero de formas de elegir n_1 art. de tipo y_1 entre los n disponibles}
$$= \frac{n!}{(n-n_1)!n_1!}$$
- α_2={numero de formas de elegir n_2 art. de tipo y_2 entre los $n-n_1$ restantes disponibles}
$$= \frac{(n-n_1)!}{(n-n_1-n_2)!n_2!}$$
$$\vdots$$
- α_{k-1}={numero de formas de elegir n_{k-1} art. de tipo y_{k-1} entre los $n-n_1-\ldots-n_{k-2}$ restantes disponibles}
$$= \frac{(n-n_1-\ldots-n_{k-2})!}{(n-n_1-\ldots-n_{k-2}-n_{k-1})!n_{k-1}!}$$
- α_k={numero de formas de elegir n_k art. de tipo y_k entre los $n-n_1-\ldots-n_{k-2}-n_{k-1}=n_k$ disponibles}
$$= 1 = \frac{(n-n_1-\ldots-n_{k-1})!}{(n-n_1-\ldots-n_{k-1}-n_k)!n_k!}$$

Es relativamente facil convencerse de que $\alpha_{(n_1,\ldots,n_k)} = \alpha_1 \ldots \alpha_k$, y desarrollando un poco la expresion de la derecha se obtiene el coeficiente multinomial.

3. Relaciones entre Multinomial y Binomial

Para el caso en que $k = 2$, uno se puede convencer que la distribución multinomial coincide con la binomial: interpretando que si no se está en la categoría y_1, se está fuera de la categoría y_1. Como $p_1 + p_2 = 1$, $q \equiv p_2 = 1 - p_1$ y definimos $p = p_1$. De igual forma, $n_2 = n - n_1$. Reemplazando en la distribución multinomial los valores anteriores, se obtiene que $\mathbb{P}(N_1 = n_1, N_2 = n_2) = \mathbb{P}(N_1 = n_1)$, donde N_1 se distribuye como una binomial de parámetros n y $p = p_1$.

Para el caso en que se tienen k categorías nos interesará la distribución marginal de N_i. Primero, si seguimos el razonamiento anterior, el hecho que no se seleccione un elemento de la categoría y_i significa que se selecciona un elemento de el resto de categorías. Esto se hace con probabilidad $1 - p_i$, por lo que la distribución marginal de N_i debiera ser una binomial de parámetros n y p_i.

Hacendo el cálculo:

$$\begin{aligned}
\mathbb{P}(N_i = n_i) &= \sum_{\substack{\{n_1,\ldots,n_k\}\setminus\{n_i\} \\ \sum_{j=1}^{k} n_i = n}} \mathbb{P}(N_1 = n_1, \ldots, N_i = n_i, \ldots, N_k = n_k) \\
&= \sum_{\substack{\{n_1,\ldots,n_k\}\setminus\{n_i\} \\ \sum_{j\neq i} n_j = n - n_i}} \binom{n}{n_1, \ldots, n_k} {p_1}^{n_1} \ldots {p_{i-1}}^{n_{i-1}} {p_i}^{n_i} {p_{i+1}}^{n_{i+1}} \ldots {p_k}^{n_k} \\
&= {p_i}^{n_i} \sum_{\substack{\{n_1,\ldots,n_k\}\setminus\{n_i\} \\ \sum_{j\neq i} n_j = n - n_i}} \frac{n!}{n_1! \ldots n_{i-1}! n_i! n_{i+1}! \ldots n_k!} {p_1}^{n_1} \ldots {p_{i-1}}^{n_{i-1}} {p_{i+1}}^{n_{i+1}} \ldots {p_k}^{n_k} \\
&= \frac{n!}{(n-n_i)! n_i!} {p_i}^{n_i} \sum_{\substack{\{n_1,\ldots,n_k\}\setminus\{n_i\} \\ \sum_{j\neq i} n_j = n - n_i}} \frac{(n-n_i)!}{n_1! \ldots n_{i-1}! n_{i+1}! \ldots n_k!} {p_1}^{n_1} \ldots {p_{i-1}}^{n_{i-1}} {p_{i+1}}^{n_{i+1}} \ldots {p_k}^{n_k} \\
&= \frac{n!}{(n-n_i)! n_i!} {p_i}^{n_i} (p_1 + \ldots + p_{i-1} + p_{i-1} + \ldots + p_k)^{n-n_i} \qquad \text{de la fórmula multinomial} \\
&= \frac{n!}{(n-n_i)! n_i!} {p_i}^{n_i} (1 - p_i)^{n-n_i} \\
&= \binom{n}{n_i} {p_i}^{n_i} (1 - p_i)^{n-n_i}
\end{aligned}$$

Observacion: Como se dijo, la interpretacion es como si hubieran dos clases (categorias): $x_1 = y_i$, y $x_2 = \bigcup_{j\neq i} y_j$. Entonces $(M_1, M_2) = (N_i, \sum_{j\neq i} N_j)$ es un vector de frecuencias para las categorías x_1 y x_2, y la distribuciÓn del vector es Multinomial de parametros n y $q_1 = p_i$,

$q_2 = \sum_{j \neq i} p_j$. La distribución marginal de M_2 es una binomial de parámetros n y q_2.

4. Esperanza, Varianza, Covarianza

Como se dijo anteriormente, la distribucion de N_i es Binom(n, p_i). Por lo tanto, se tiene que

$$\mathbb{E}(N_i) = np_i$$

$$\mathbb{V}ar(N_i) = np_i(1 - p_i)$$

Como ya se vio, $N_i + N_j \sim$Binom$(n, p_i + p_j)$, por lo que

$$\begin{aligned} \mathbb{V}ar(N_i + N_j) &= \mathbb{V}ar(N_i) + \mathbb{V}ar(N_j) + 2\mathbb{C}ov(N_i, N_j) \\ n(p_i + p_j)(1 - p_i - p_j) &= np_i(1 - p_i) + np_j(1 - p_j) + 2\mathbb{C}ov(N_i, N_j) \\ \mathbb{C}ov(N_i, N_j) &= -np_ip_j \end{aligned}$$

5. Ejemplo sencillo

Supongase que el 23 % de las personas que asisten a cierto partido de "baseball" viven a menos de 10 millas del estadio, el 59 % de ellas viven a entre 10 y 50 millas del estadio, y el 18 % vive a mas de 50 millas. Se seleccionan al azar 20 personas entre los asistentes al partido (que son miles). Calcular la probabilidad de que siete de los seleccionados vivan a menos de 10 millas, ocho vivan entre 10 y 50 millas, y cinco vivan a mas de 50 millas del estadio.

Solución Comenzamos por identificar todos los elementos del problema:
$n = 20$ (número de personas seleccionadas), $k = 3$ (cantidad de grupos de clasificación de las personas); $y_1 =$\{Personas que viven a menos de 10 millas del estadio\}, $y_2 =$\{Personas que viven a entre 10 y 50 millas del estadio\}, $y_3 =$\{Personas que viven a más de 50 millas del estadio\}; $p_1 = 0{,}23, p_2 = 0{,}59, p_3 = 0{,}18$

Definiendo (N_1, N_2, N_3) el vector correspondiente a las frecuencias, se pide calcular

$$\begin{aligned} \mathbb{P}(N_1 = 7, N_2 = 8, N_3 = 5) &= \binom{20}{7, 8, 5} 0{,}23^7 0{,}59^8 0{,}18^5 \\ &= \frac{20!}{7!8!5!} 0{,}23^7 0{,}59^8 0{,}18^5 \\ &= 0{,}0094 \end{aligned}$$

6. Problemas

6.1. Problema 1

(a) Suponga que las variables aleatorias $(X_1, \ldots, X_k)$ son independientes y que $X_i \sim$Poisson(λ_i) $\forall i \in \{1, \ldots, k\}$. Demuestre que para todo $n \in \mathbb{N}$ la distribución del vector aleatorio $(X_1, \ldots, X_k)$ condicional a que $\sum_{i=1}^{k} X_i = n$ es una multinomial de parametros n y $\vec{p} = (p_1, \ldots, p_k)$, con

$$p_i = \frac{\lambda_i}{\sum_{j=1}^{k} \lambda_j}$$

(b) A una heladeria llegan clientes de tres tipos distintos: *normal*, *golozo* y *premium*, y el número de clientes de cada tipo (que llega en una hora) son Poisson con tasa $\lambda_n = 100, \lambda_g = 50$ y $\lambda_p = 20$. Si se sabe que el número total de clientes que llegó en una hora es de 500 personas, calcular la probabilidad de que hayan llegado más de 200 clientes premium.

6.2. Problema 2

Se lanzan cinco dados equilibrados. Cual es la probabilidad de que el número 1 y el número 4 aparezcan el mismo número de veces?

6.3. Problema 3

(a) Supongase que el 16 % de los estudiantes de un colegio son alumnos de primero medio, el 14 % de segundo, el 38 % de tercero, y el 32 % de cuarto. Si se seleccionan al azar 15 estudiantes, cual es la probabilidad de que al menos 8 estudiantes sean de primero o segundo?

(b) Sea X_3 el número de estudiantes de tercero y X_4 el número de estudiantes de cuarto. Calcule el número esperado de alumnos de cada curso (en la muestra), y la esperanza y varianza de $X_3 - X_4$

Referencias

[Degroot,1988] Degroot, M.; *"Probabilidad y Estadística, Segunda Edición"*; Addison-Wesley Iberoamericana, S.A.; p.p. 283-286; 1988.