Introducción a la probabilidad y estadística CM274

César Lara Avila

28 de agosto de 2017

https://github.com/C-Lara

2. Combinatoria para probabilidad

Introducción

Existen muchas técnicas, basadas en la ley de probabilidad total para contar todas las posibilidades. Es usual describir este problema en términos de cómo se pueden elegir `bolas distinguibles` de una caja asumiendo que la elección es `aleatoria`, es decir, cada bola en la caja es igualmente probable que se elija.

Alternativamente, el problema puede ser expresado en términos de cómo se pueden insertar `bolas indistinguibles` en cajas distinguibles.

Esta primera forma se llama **problema de selección** y la última forma se llama **problema de asignación**. Formularemos nuestra notas en términos del problema de selección.

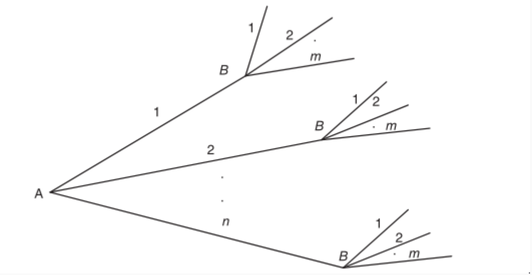
Principio del conteo

Las técnicas de conteo consideradas en un espacio muestral con puntos igualmente probables se basan en dos principios de conteo fundamentales relativos a eventos mutuamente exclusivos A y B:

- `Principio 1:` Si los eventos A y B pueden ocurrir en n y m maneras respectivamente, entonces A y B pueden ocurrir juntos de nm maneras, llamada regla de multiplicación.
- `Principio 2:` Si los eventos A y B pueden ocurrir en n y m maneras respectivamente, entonces A o B pueden ocurrir juntos de $n + m$ maneras, llamada regla de suma.

Principio del conteo(1)

El Principio 1 se establece fácilmente ya que A puede ocurrir de n maneras y luego debe ser seguida por cada manera en que B puede ocurrir. Un diagrama de árbol, mostrado en la siguiente figura ilustra el resultado.



El Principio 2 simplemente usa la palabra o en un sentido exclusivo.

Permutaciones

Una disposición (arreglo) de ítems, también llamada una `secuencia ordenada de ítems`, se dice que es una permutación de los ítems. Una secuencia ordenada de n ítems se llama una n-permutación.

Con n ítems distintos, hay $n!$ permutaciones posibles. Este es el número de diferentes maneras en que los n ítems distintos pueden ser colocados, ordenados

Ejemplo

Considera la palabra PUPPY. Si los cinco caracteres fueran distintos, el número de permutaciones que se obtendría seria 5!. Sin embargo, como el carácter P aparece tres veces en PUPPY, el 5! debe ser dividido por 3!, para obtener $5!/3! = 5 \times 4 = 20$. El número de diferentes maneras en que los caracteres de la palabra PUPPY se pueden colocar o disponer es por lo tanto 20.

Permutaciones con reemplazo

En el caso de permutaciones con reemplazo, nuestro objetivo es contar el número de formas en que se pueden seleccionar k bolas entre n bolas distinguibles. Después de cada bola es elegida y sus características son registradas, se sustituye en la caja y se elige la bola siguiente.

Si $k = 1$(se elige una bola), entonces el número de permutaciones posibles es n, ya que cualquiera de las n bolas puede ser elegida. Si $k = 2$, entonces cualquiera de las n bolas puede ser elegida como la primera bola y luego reemplazada en la caja. Para cada una de estas n elecciones de la primera bola, la siguiente bola elegida puede ser también cualquiera de las n bolas distinguibles y por lo tanto el número de permutaciones cuando $k = 2$ es $n \times n = n^2$.

Este razonamiento puede continuar y demostrar que el número de permutaciones obtenidas para cualquier k es n^k.

Permutaciones con reemplazo(1)

Supongamos ahora que hay n_1 maneras de elegir un primer elemento (ítem) y n_2 formas de elegir un segundo elemento(ítem). Entonces el número de pares ordenados distintos es igual a $n_1 n_2$.

En términos de bolas distinguibles en cajas, esto puede ser visto como el número de maneras en que una bola puede ser elegida de una primera caja que contiene n_1 bolas distinguibles y una segunda bola elegida de una segunda caja que contiene n_2 bolas distinguibles.

Si hay n_i formas de elegir un ítem i, para $i = 1, 2, \ldots, k$, entonces el número de k-tuplas ordenadas distintas es igual a $n_1 n_2 \ldots n_k$.

Ejemplo

El número de códigos de cuatro dígitos que se pueden obtener utilizando el sistema de números decimales (con diez dígitos de 0 a 9 inclusive) es $10^4 = 10{,}000$. Estos son los códigos que van desde 0000 hasta 9999.

Permutaciones sin reemplazo

Pasemos ahora al problema de contar el número de diferentes permutaciones obtenidas al seleccionar k bolas de entre n bolas distinguibles en el caso de que una vez que se elija una bola en particular, esta no se devuelve a la caja.

Denotamos a este número como $P(n, k)$ y asignamos los valores $P(n, 0) = 1, n = 0, 1, \ldots,$ por convención.

Decimos que $P(n, k)$ es el <u>número de permutaciones</u> de n objetos tomados k a la vez.

Si $k = 1$, entonces cualquier bola puede ser elegida y el número total de permutaciones simplemente es n. Si $k = 2$, entonces cualquiera de las n bolas puede ser elegida como la primera, para cada una de estas n elecciones posibles, quedan $n - 1$ bolas en la caja, cualquiera de las cuales puede ser elegida como la segunda. Así, el número total de permutaciones para $k = 2$ es igual a $n(n - 1)$.

Permutaciones sin reemplazo(1)

Con $k = 3$, hay n posibilidades diferentes para la primera bola, $n - 1$ para la segunda y $n - 2$ para la tercero, lo que da $P(n, 3) = n(n-1)(n-2)$.

Podemos ahora generalizar esto a cualquier arbitrario $k \leq n$ para obtener

$$P(n, k) = n(n-1)(n-2)\dots(n-k+1) = \frac{n!}{(n-k)!}.$$

Sea $n = 4, k = 3$ y distinguimos las cuatro bolas por medio de las letras A, B, C y D. Tenemos $P(4, 3) = 4 \times 3 \times 2 = 24$. Esas 24 posibilidades, están dadas por:

```
ABC, ABD, ACB, ACD, ADB, ADC,

BAC, BAD, BCA, BCD, BDA, BDC,

CAB, CAD, CBA, CBD, CDA, CDB,

DAB, DAC, DBA, DBC, DCA, DCB.
```

Debes observar que distinguimos entre ABC y ACB, ya que el orden es importante, pero no incluimos la misma letra más de una vez (por ejemplo, AAB no está presente), esto es por que una vez que las letras son usadas no pueden ser utilizadas de nuevo.

Combinaciones sin reemplazo

Sea al caso en el que las bolas se seleccionan sin reemplazo, pero el orden en que se seleccionan no tiene importancia. Lo único que importa es que se han elegido algunas bolas y la forma en que se seleccionaron no es de interés.

Por ejemplo, podemos haber elegido una bola verde, una bola roja y una bola negra, pero no sabemos o no nos preocupamos cuál de estas tres fue elegida primera, segunda o tercera.

Sea $C(n, k)$ que denota el número de maneras en que se pueden seleccionar k bolas sin reemplazo e independientemente de su orden en una caja que contiene n bolas. Este símbolo se dice que es el número de combinaciones de n elementos tomados k a la vez, no teniendo en cuenta el orden.

Combinaciones sin reemplazo(1)

Dado que cualquier secuencia de k ítems puede ser dispuesta en $k!$ permutaciones, se deduce que debe haber $k!$ veces más permutaciones (sin reemplazo) de k ítems distintos de lo que hay en combinaciones (sin reemplazo), ya que el orden importa en permutaciones, pero no en combinaciones. En otras palabras, debemos tener

$$k!C(n,k) = P(n,k) \quad \text{para} \quad n = 0,1,2,\dots \quad \text{y} \quad k = 0,1,\dots,n.$$

Esto conduce al siguiente resultado conocido:

$$C(n,k) = \frac{P(n,k)}{k!} = \frac{n!}{k!(n-k)!} \quad \text{para} \quad n = 0,1,2,\dots \quad \text{y} \quad k = 0,1,\dots,n.$$

Combinaciones sin reemplazo(2)

Sean A, B, C, D y E cinco elementos distinguibles de los cuales dos deben ser escogidos (lo cual implica, sin reemplazo, y no importa cuál viene primero: es decir, una combinación). El número de formas en que estos dos pueden ser elegidos está dado por $C(5,2) = 5!/(2!3!) = 10$. Estos son los resultados:

```
AB AC AD AE

BC BD BE

CD CE

DE
```

$C(n,k)$ se denomina coeficiente binomial ya que es el coeficiente del término $p^k q^{n-k}$ en la expansión del binomio $(p+q)^n$. Formas alternativas para escribir este coeficiente binomial son:

$$C(n,k) = C_k^n = \binom{n}{k}.$$

Combinaciones sin reemplazo(3)

Para calcular los coeficientes binomiales, notamos que

$$\begin{aligned}\binom{n}{k} &= \frac{n!}{k!(n-k)!} = \frac{n}{k}\left(\frac{(n-1)!}{(k-1)!(n-k)!}\right)\\ &= \frac{n}{k}\left(\frac{(n-1)!}{(k-1)!([n-1]-[k-1])!}\right) = \frac{n}{k}\binom{n-1}{k-1}\end{aligned}$$

Lo que conduce a,

$$\binom{n}{k} = \frac{n}{k} \times \frac{n-1}{k-1} \times \cdots \times \frac{n-(k-1)}{k-(k-1)} = \prod_{j=1}^{k} \frac{n-j+1}{j}.$$

y esta expresión computacionalmente es más eficiente que formar factoriales y tomar razones.

Ejemplo explicativo

Consideremos ahora el problema de colocar k bolas distinguibles en n cajas diferentes de tal manera que el número de bolas en la caja i sea k_i para $i = 1, 2, \ldots, n$.

Asumimos que $\sum_{i=1}^{n} k_i = k$, de modo que cada bola se pone en una de las cajas y de manera que no queda ninguna.

El número de combinaciones obtenidas al seleccionar las primeras k_1 bolas para la caja 1 es $C(k, k_1)$. Esto deja $k - k_1$ bolas, de las cuales k_2 son escogidas y colocadas en la caja 2. El número de combinaciones obtenidas al seleccionar estas k_2 bolas de las $k - k_1$ bolas es $C(k - k_1, k_2)$, de modo que el total obtenido de las dos primeras cajas es $C(k, k_1) \times C(k - k_1, k_2)$.

Ejemplo explicativo(1)

Continuando de esta manera, vemos que el número total de combinaciones está dado por:

$$C(k, k_1) \times C(k - k_1, k_2) \times C(k - k_1 - k_2, k_3) \times \cdots \times C\left(k - \sum_{i=1}^{n-1} k_i, k_n\right),$$

donde $C(k - \sum_{i=1}^{n-1} k_i, k_n) = C(k_n, k_n)$. Sustituyendo en la fórmula para los números de combinación tenemos:

$$\begin{aligned}\frac{k!}{k_1!(k-k_1)!} \times \frac{(k-k_1)!}{k_2!(k-k_1-k_2)!} \times \frac{(k-k_1-k_2)!}{k_2!(k-k_1-k_2-k_3)!} \times \cdots \\ = \frac{k!}{k_1!k_2!\cdots k_n!} \equiv \binom{k}{k_1, k_2, \ldots k_n}.\end{aligned}$$

Estos son los llamados <u>coeficientes multinomiales</u>.

Combinaciones con reemplazo

Cuando se seleccionan k bolas con reemplazo de una caja que contiene n bolas distinguibles. Se puede demostrar que esto es idéntico al problema de contar el número de combinaciones cuando se seleccionan k bolas sin sustitución de una caja que contiene un total de $n + k - 1$ bolas distinguibles, es decir:

$$\frac{(n+k-1)!}{(n-1)!k!}.$$

Representamos las n bolas por estrellas adyacentes y consideramos la posibilidad de insertar $k - 1$ barras entre las estrellas para separar las barras en k en grupos. Por ejemplo, para $n = 12$ y $k = 5$, la siguiente es una representación de un grupo de 12 bolas distinguibles en 5 urnas, donde el tamaño de las urnas 1, 2, 3, 4 y 5 son 2, 4, 0, 3 y 3, respectivamente:

$$\ast\ \ast \mid \ast\ \ast\ \ast\ \ast \mid \mid \ast\ \ast\ \ast \mid \ast\ \ast\ \ast$$

Combinaciones con reemplazo(1)

Se debe tener en cuenta que en la agrupación, pueden haber urnas vacías. Hay un total de $n + k - 1$ posiciones, de las cuales n son estrellas y $k - 1$ son barras. Por lo tanto, el número de formas de colocar n bolas indistinguibles en k urnas etiquetadas es el mismo que el número de formas de escoger n posiciones entre $n - k + 1$ espacios para las estrellas, con todas las posiciones restantes tomadas como barras.

El número de maneras en que esto se puede hacer es: $\binom{n+k-1}{n}$. Pero $\binom{n+k-1}{n} = \binom{n+k-1}{k-1}$ puede interpretarse como el número de formas de elegir las posiciones de las barras y tomar todas las posiciones restantes como estrellas.

Ejemplo explicativo

Veamos el caso que $n = 4$ y diferenciamos las bolas con las letras A, B, C y D. Si elegimos $\mathrm{k} = 3$ entonces la fórmula nos dice que hay $(4+3-1)!/(3! \times 3!) = 20$ combinaciones diferentes. Estas son:

`AAA AAB AAC AAD ABB ABC ABD ACC ACD ADD`

`BBB BBC BBD BCC BCD BDD`

`CCC CCD CDD`

`DDD.`

Debes de observar que, aunque las cuatro bolas se distinguen, la misma bola puede aparecer más de una vez en una combinación dada. Esto se debe a que después de que una bola ha sido elegida, esta se sustituye y se puede elegir una vez más.

Observa también que no incluimos combinaciones como BAA o CBA porque con combinaciones, el orden no es importante y estas son equivalentes a AAB y ABC, respectivamente.