Probabilidad - Clase 5
Estadística. Teorema de Bernoulli

Ernesto Mordecki

Facultad de Ciencias, Universidad de la República. Montevideo, Uruguay

Curso de la Licenciatura en Matemática - 2020

Figura: Jacob Bernoulli (1654-1705) Basilea, Suiza.

Contenidos

Esquema de Bernoulli

- El espacio de probabilidad es

$$\Omega_n = \{\omega = (\omega^{(1)}, \dots, \omega^{(n)}) :$$
$$\text{cada } \omega^{(i)} \text{ es éxito } \omega_1, \text{ o fracaso } \omega_2 \ (i = 1, \dots, n).\}$$

- La σ-álgebra es $\mathcal{P}(\Omega_n)$
- la probabilidad de cada suceso elemental

$$\mathbf{P}(\omega) = p^k q^{n-k}.$$

donde k es la cantidad de éxitos de ω

Distribución binomial

- $\mu(\omega)$ es cantidad de éxitos en $\omega \quad (0 \leq \mu(\omega) \leq m.)$
- Tenemos un teorema que prueba que

$$P_n(k) = \mathbf{P}(\mu = k) = \binom{n}{k} p^m q^{n-k}.$$

para $m = 0, 1, \ldots, n$.

Estadística: estimación de p

Problema: Observamos los resultados de un esquema de Bernoulli con n experimentos. Desconocemos $p \in (0, 1)$. Nos proponemos *estimar* el parámetro p. Mas precisamente:

- Sabemos que observamos un esquema de Bernoulli
- Observamos una realización ω. En particular sabemos $\mu(\omega)$
- Desconocemos $p \in (0, 1)$
- ¿Cómo podemos calcular (o al menos aproximar) p?
- una propuesta intuitiva es tomar:

Estimación de máxima verosimiltud

$$\hat{p} \sim \frac{\mu(\omega)}{n}$$

$$\boxed{\text{¿Podemos dar un fundamento a esta intuición?}}$$

La propuesta es: Calculemos la probabilidad de la observación ω y elijamos el p que haga máxima esa probabilidad. Tenemos

- $\mathbf{P}(\omega, p) = p^k q^{n-k}$
- Ahora conocemos k pero desconocemos p (por eso escribimos la probabilidad como función de p).
- Resolvemos el problema

$$\max \mathbf{P}(\omega, p) \text{ para } p \in (0, 1)$$

Hagamos las cuentas

- El máximo de $\mathbf{P}(\omega, p)$ es el mismo que el de $\log \mathbf{P}(\omega, p)$. Tenemos entonces

$$\ell(p) = \log \mathbf{P}(\omega, p) = \log p^k q^{n-k} = k \log p + (n-k) \log q$$

- Derivamos
$$\frac{\partial \ell(p)}{\partial p} = \frac{k}{p} - \frac{n-k}{1-p} = 0.$$

- Operando: $k(1-p) - p(n-k) = k - kp - pn + pk = 0$

- De aquí:

$$\hat{p} = \frac{k}{n} \text{ es el } \textit{estimador de máxima verosimilitud}.$$

Repasando

- Tenemos una estrategia de estimación, la *máxima verosimilitud* (máxima "creencia")
- Calculamos la probabilidad de obtener la *muestra* que observamos, como función del parámetro (p en este caso) a estimar
- Calculamos el máximo de esa probabilidad como función de p
- Obtenemos el estimador de máxima verosimilitud $\hat{p}$.
- El sombrero $\hat{\theta}$ se usa para indicar que se está *estimando* un parámetro θ desconocido.

Conclusión

- En probabilidades tenemos un espacio

$$(\Omega, \mathcal{A}, \mathbf{P})$$

  con **P** conocida.

- A partir de esto calculamos probabilidades de sucesos que nos interesan. Después ocurre el azar (experimento)

- En estadística los experimentos se realizan primero, surgen *datos*, el ω, pero se desconoce **P**

- Con esos datos se *estima* la probabilidad **P**

Paseo al azar y triangulo de Pascal

- Consideremos una partícula que se desplaza, en cada paso (unidad de tiempo) una unidad de distancia para arriba o para abajo al azar.
- Este azar consiste en subir con probabilidad p y bajar con probabilidad q, números positivos que suman uno.
- Cada paso se realiza en forma independiente.
- Llamamos *paseo al azar* a dicho movimiento.
- El caso $p = q = 1/2$ se denomina *paseo al azar simétrico*.

Un paseo al azar

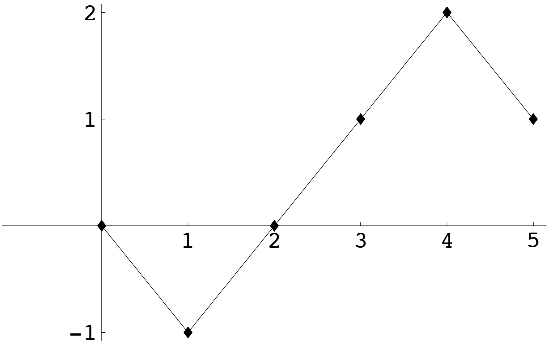

Figura: Una trayectoria del paseo al azar

Propiedades

- En n pasos la particula puede recorrer 2^n caminos diferentes
- Nos interesa saber la cantidad de caminos diferentes que conducen a una altura determinada en una cantidad de pasos dada.
- En general: para el primer paso tenemos dos caminos, uno que sube y otro que baja.

- En el segundo paso, de los cuatro caminos que se obtienen (2^2), tenemos uno que llega a una altura 2, otro que llega a -2, y dos caminos que llegan al nivel cero.
- La cantidad de caminos que llegan a los niveles -3,-1,1 y 3 en tres pasos se obtienen sumando los caminos que llegaban en dos pasos a los niveles ubicados una unidad mas arriba y una mas abajo, porque precisamente de esos niveles llegamos en un paso más hasta el nivel indicado.

Si representamos estas cantidades en una tabla obtenemos el *triángulo de Pascal*:

$$
\begin{array}{ccccccccccccc}
 & & & & & & 1 & & & & & & \\
 & & & & & 1 & & 1 & & & & & \\
 & & & & 1 & & 2 & & 1 & & & & \\
 & & & 1 & & 3 & & 3 & & 1 & & & \\
 & & 1 & & 4 & & 6 & & 4 & & 1 & & \\
 & 1 & & 5 & & 10 & & \mathbf{10} & + & \mathbf{5} & & 1 & \\
1 & & 6 & & 15 & & 20 & & \mathbf{15} & & 6 & & 1
\end{array}
$$

Regla de formación del triángulo de Pascal

- cada línea se obtiene de la superior sumando los dos números que se encuentran arriba a la derecha y a la izquierda (asumiendo un cero cuando no hay número): por ejemplo $15 = 5 + 10$.
- Según vimos, para una fila dada, por ejemplo la última (pongamos $n = 6$), estos números cuentan la cantidad de caminos posibles por los que la partícula llega a las diferentes alturas en 6 pasos.

- Como la cantidad total de caminos que puede seguir la partícula es 2^6, tenemos que

$$1 + 6 + 15 + 20 + 15 + 6 + 1 = 2^6 = 64,$$

como se verifica también sumando.

- El triángulo es simétrico (porque es simétrico el paseo si cambiamos arriba por abajo en su construcción).

- Resulta además, que las combinaciones de n tomadas de a m coinciden con el número que aparece en en m-ésimo lugar de la n-ésima fila del triángulo de Pascal.
- Esto resulta de observar, que, numerando de 1 a n los pasos, para contar cada camino que sube en m pasos, tenemos que elegir los m momentos de subir de esos n números, correspondiendo a cada elección un subconjunto de m elementos del conjunto $\{1, 2, \ldots, n\}$ de pasos.

Propiedades

- Como consecuencia, de la simetría del triángulo de Pascal resulta que
$$C^n_m = C^n_{n-m}$$
la igualdad de las combinaciones complementarias,
- Además
$$C^n_m = C^{n-1}_{m-1} + C^{n-1}_m$$
la formula de Stieffel de las combinaciones, que es la regla de formación del triángulo.

Cálculo de probabilidades

- Como las elecciones en cada paso son independientes, si designamos mediante k a la cantidad de veces que la partícula sube, tenemos para un camino individual que sube k veces (y por lo tanto baja $n-k$ veces) una probabilidad de $p^k q^{n-k}$.
- Si $P_n(k)$ es probabilidad de subir k veces en n pasos (la altura es $2k-n$) tenemos

$$P_n(k) = \binom{n}{k} p^k q^{n-k}.$$

- En conclusión, tenemos la distribución binomial del Esquema de Bernoulli.

Ley de los grandes números de Bernoulli

- Supongamos que repetimos un experimento con dos resultados posibles que denominamos éxito y fracaso
- Designemos mediante μ la cantidad de éxitos.
- Esto incluye el caso particular de tirar una moneda n veces y contar la cantidad de veces que obtenemos cara.

- Queremos estudiar, para valores grandes de n, el comportamiento del cociente μ/n, es decir, de la frecuencia de éxitos.
- Este comportamiento viene dado por el siguiente resultado, obtenido por J. Bernoulli, aparecido en 1713.
- Se fundamenta la intuición, que dice que la frecuencia asintótica de los éxitos es p. En el caso de tirar una moneda, esta intuición dice que, "aproximadamente, la mitad de las veces aparce cara".

Teorema de Bernoulli. Ley de los grandes números

Teorema

Dados $\varepsilon > 0$, $\eta > 0$, *números arbitrariamente pequeños, existe un número natural* n_0 *tal que, para todo* $n \geq n_0$,

$$\mathbf{P}\left(\left|\frac{\mu}{n} - p\right| < \varepsilon\right) \geq 1 - \eta.$$

En palabras,

- dado $\varepsilon > 0$ la probabilidad de que la diferencia entre la frecuencia observada $\frac{\mu}{n}$ y p sea menor que ε es arbitrariamente cercana a 1 cuando n es suficientemente grande.

Para demostrar este teorema, recordando que

$$\sum_{k=0}^{n} P_n(k) = \sum_{k=0}^{n} C_k^n p^k q^{n-k} = 1, \tag{1}$$

obtenemos primero dos fórmulas auxiliares

Lema

Valen las siguientes identidades:

$$\sum_{k=0}^{n} k P_n(k) = np, \tag{2}$$

$$\sum_{k=0}^{n} k^2 P_n(k) = n^2 p^2 + npq. \tag{3}$$

Demostración

Comencemos con la primer fórmula. Tenemos

$$
\begin{aligned}
\sum_{k=0}^{n} kP_n(k) = \sum_{k=1}^{n} kP_n(k) &= \sum_{k=1}^{n} k\frac{n!}{k!(n-k)!}p^k q^{n-k} \\
&= \sum_{k=1}^{n} \frac{n!}{(k-1)!(n-k)!}p^k q^{n-k} \\
&= np\sum_{k=1}^{n} \frac{(n-1)!}{(k-1)!(n-k)!}p^{k-1} q^{n-k} \\
&= np\sum_{k=0}^{n-1} \frac{(n-1)!}{k!(n-k)!}p^{k-1} q^{n-k} \\
&= np\sum_{k=0}^{n-1} C_k^{n-1} p^{k-1} q^{n-k} = np,
\end{aligned}
$$

donde utilizamos (1) para $n-1$ de la cual obtenemos que $\sum_{k=0}^{n-1} C_k^{n-1} = 2^{n-1}$. Transformemos ahora la segunda fórmula:

$$\begin{aligned}\sum_{k=0}^{n} k^2 P_n(k) &= \sum_{k=0}^{n} k(k-1+1)P_n(k)\\ &= \sum_{k=0}^{n} kP_n(k) + \sum_{k=0}^{n} k(k-1)P_n(k).\end{aligned}$$

El primer sumando es el que acabamos de calcular. Con el segundo procedemos de forma análoga, simplificando los factores k y $k-1$. En efecto

$$
\begin{aligned}
\sum_{k=0}^{n} k(k-1)P_n(k) &= \sum_{k=2}^{n} k(k-1)P_n(k) \\
&= \sum_{k=2}^{n} k(k-1)\frac{n!}{k!(n-k)!}p^k q^{n-k} \\
&= \sum_{k=2}^{n} \frac{n!}{(k-2)!(n-k)!}p^k q^{n-k} \\
&= n(n-1)p^2 \sum_{k=2}^{n} \frac{(n-2)!}{(k-2)!(n-k)!}p^{k-2} q^{n-k} \\
&= n(n-1)p^2 \sum_{k=0}^{n-2} \frac{(n-2)!}{k!(n-2-k)!}p^k q^{n-2-k} \\
&= n(n-1)p^2 \sum_{k=0}^{n-2} C_k^{n-2} p^k q^{n-2-k} = n(n-1)p^2,
\end{aligned}
$$

En la fórmula anterior utilizamos que

$$\sum_{k=0}^{n-2} C_k^{n-2} p^k q^{n-2-k} = 1$$

En conclusión

$$\sum_{k=0}^{n} k^2 P_n(k) = np + n^2p^2 - np^2 = n^2p^2 + npq.$$

Esto concluye la demostración del lema.

Demostración del Teorema

Comencemos observando que los dos sucesos $\left|\frac{\mu}{n} - p\right| < \varepsilon$, $\left|\frac{\mu}{n} - p\right| \geq \varepsilon$ son contrarios, por lo que

$$\mathbf{P}\left(\left|\frac{\mu}{n} - p\right| < \varepsilon\right) = 1 - \mathbf{P}\left(\left|\frac{\mu}{n} - p\right| \geq \varepsilon\right)$$

Según la regla de la suma de probabilidades, tenemos

$$\mathbf{P}\left(\left|\frac{\mu}{n} - p\right| \geq \varepsilon\right) = \sum P_n(k)$$

donde la suma abarca todos los k tales que $\left|\frac{k}{n} - p\right| \geq \varepsilon$. Para estos valores de k se verifica también

$$\frac{\left(\frac{k}{n} - p\right)^2}{\varepsilon^2} \geq 1.$$

Basados en esta última desigualdad, obtenemos

$$\mathbf{P}\left(\left|\frac{\mu}{n} - p\right| \geq \varepsilon\right) \leq \sum \frac{\left(\frac{k}{n} - p\right)^2}{\varepsilon^2} P_n(k)$$

donde la suma se extiende a los mismos valores de k que antes. Es claro que al sumar en todos los valores de k desde 0 a n, obtenemos un resultado mayor.

En consecuencia

$$\mathbf{P}\left(\left|\frac{\mu}{n}-p\right|\geq\varepsilon\right)\leq\sum_{k=0}^{n}\frac{\left(\frac{k}{n}-p\right)^2}{\varepsilon^2}P_n(k)=\frac{1}{n^2\varepsilon^2}\sum_{k=0}^{n}(k-np)^2P_n(k)$$

$$=\frac{1}{n^2\varepsilon^2}\Big[\sum_{k=0}^{n}k^2P_n(k)-2np\sum_{k=0}^{n}kP_n(k)+n^2p^2\sum_{k=0}^{n}P_n(k)\Big]$$

$$=\frac{1}{n^2\varepsilon^2}[n^2p^2+npq-2n^2p^2+n^2p^2]=\frac{pq}{n\varepsilon^2}$$

En los últimos cálculos utilizamos las fórmulas (1), (2) y (3). Concluyendo

$$\mathbf{P}\left(\left|\frac{\mu}{n}-p\right|<\varepsilon\right)=1-\mathbf{P}\left(\left|\frac{\mu}{n}-p\right|\geq\varepsilon\right)\geq 1-\frac{pq}{n\varepsilon^2}. \qquad (4)$$

Esta desigualdad demuestra el teorema.

Aplicación

Nos interesa calcular la probabilidad de que entre 400 niños que nacen, la la cantidad de varones se desvie del valor que se espera, que es 200, en no más de 20. **Solución** Aplicamos el ejemplo anterior, suponiendo la probabilidad de nacimiento de un varón igual a $1/2$, y μ cuenta la cantidad de varones que nacen, con un total de $n = 400$ nacimientos. Tenemos que calcular

$$\mathbf{P} = \mathbf{P}\left(\left|\mu - 200\right| < 20\right) = \mathbf{P}\left(\left|\frac{\mu}{n} - \frac{1}{2}\right| < \frac{20}{n}\right).$$

Dado que el cálculo numérico de esta probabilidad es díficil, utilizamos la cota (4) obtenida en el teorema. Tenemos $\varepsilon = 20/400 = 1/20$, y obtenemos

$$\mathbf{P} \geq 1 - \frac{1}{4n^2\varepsilon^2} = 1 - \frac{1}{4 \times 400 \times (1/20)^2} = 3/4.$$

Un cálculo aproximado, pero más preciso, establece que $\mathbf{P} \sim 0,95$.

Consecuencias en estadística del Teorema de Bernoulli

- Consideramos el problema de estimación de p a partir de una *muestra* que observamos: ω.
- El estimador de máxima verosimilitud de p es, según vimos
$$\hat{p} = \frac{k}{n}$$
- El teorema de Bernoulli establece que
$$\mathbf{P}\left(\left|\hat{p} - p\right| > \varepsilon\right) \leq \frac{pq}{n\varepsilon^2} \to_n 0.$$

- También decimos que

$$\hat{p} \text{ converge en probabilidad a } p,$$

cuando ocurre el resultado anterior, y se escribe

$$\hat{p} \xrightarrow{\mathbf{P}} p, \quad \text{cuando } n \to \infty$$

- En estadística se dice que

$$\hat{p} \text{ es es un \textit{estimador insesgado} de } p.$$