

Repaso de Probabilidad

Clase 1

Sergio Béjar

Departamento de Estudios Políticos

Objetivos para hoy

- Discutir probabilidad y mate básica. Son cosas que debemos saber de cualquier forma.

Probabilidad

Probabilidad se refiere a la posibilidad de que algún evento ocurra.

- Es una característica inevitable del mundo en que vivimos y hay que entenderla.
- Encuentra sus orígenes en las apuestas en los siglos XVII y XVIII.

Consciente o inconscientemente, siempre pensamos en términos probabilísticos.

- e.g., Si voy a 150 en una zona de 95km/hr, puedo recibir una infracción.

La teoría de la probabilidad es la forma matemática de modelar la realidad incierta.
Probability theory is a precursor to statistics and applied mathematics.

Reglas de Probabilidad

Estas son algunas (no todas) reglas de probabilidad importantes.

1. El conjunto de todos los eventos posibles ($E_1 \dots E_n$) es el **espacio muestra**.
 - S es el **conjunto** para un volado $S = \{ \text{Soles, Águilas} \}$.
2. Las probabilidades deben satisfacer la desigualdad $0 \leq p \leq 1$.
3. La suma de las probabilidades en el espacio muestra debe ser igual a 1.
 - Formalmente: $\sum_{E_i \in S} p(E_i) = 1$
4. Si el evento A y evento B son *independientes*, la **probabilidad conjunta** de que ocurran los dos es $p(A, B) = p(A) * p(B)$.
5. Si la probabilidad del evento A depende de que el evento B haya ocurrido, la **probabilidad condicional** de A “dado” B es un poco diferente.

$$p(A | B) = \frac{p(A, B)}{p(B)}$$

Reglas de Probabilidad

La probabilidad condicional implica que los eventos no son completamente independientes.

- es decir, se “superponen” o “intersectan”.

Por lo tanto, tenemos que conocer otras dos reglas de probabilidad.

Probabilidad de Uniones: $p(A \cup B) = p(A) + p(B) - p(A \cap B)$

Probabilidad de Intersecciones: $p(A \cap B) = p(A) + p(B) - p(A \cup B)$

Algunas Aplicaciones Simples

Empecemos con un diagrama de Venn. Asumimos:

- Probabilidad de ser hombre (i.e. $p(A)$) = .5
- Probabilidad de ser obeso (i.e. $p(B)$) = .3

Queremos saber:

- ¿Cuál es la probabilidad de que alguien sea hombre *ú* obeso?
- ¿Cuál es la probabilidad de que alguien sea obeso, dado que es hombre?

Algunas Aplicaciones Simples

¿Cuál es la probabilidad de que alguien sea hombre *ú* obeso?

- Esta pregunta se refiere a una probabilidad de uniones.
- $p(A \cup B) = p(A) + p(B) - p(A \cap B) = .5 + .3 - .1 = .7$
- Restamos la intersección porque algunos hombres son obesos.
- Probabilidad de ser una mujer no obesa: $1 - .7 = .3$

¿Cuál es a probabilidad de que alguien sea obeso, dado que es hombre?

- Esta es una probabilidad condicional.
- $p(A | B) = \frac{p(A, B)}{p(B)} = \frac{.1}{.5} = .2$

Siempre podemos derivar reglas más importantes y complejas de las reglas de probabilidad básicas.

Probabilidad Total y Teorema de Bayes

Recordar: $p(A | B) = \frac{p(A, B)}{p(B)}$. Entomces: $p(A, B) = p(A | B) * p(B)$.

- Además: $p(B, A) = p(B | A) * p(A)$.
- Entonces, obviamente: $p(A, B) = p(B, A)$.
- Por lo tanto: $p(B | A) * p(A) = p(A | B) * p(B)$.

Si lo queremos isolar $p(B | A)$, simplemente lo dividimos por $p(A)$.

$$p(B | A) = \frac{p(A | B) * p(B)}{p(A)}$$

Teorema de Probabilidad Total y Bayes

Es un teorema interesante por sí solo. Es el **Teorema de Probabilidad Total**.

- También lo llamamos **Teorema de Bayes**.

Con solo dos resultados posibles (B and $\sim B$).

$$p(B | A) = \frac{p(A | B)p(B)}{p(A | B)p(B) + p(A | \sim B)p(\sim B)}$$

Una aplicación: El Error del Investigador

Assumimos el siguiente escenario: un investigador muy cuidadoso está recolectando evidencias contra un acusado.

- Las huellas digitales coinciden, lo que sucede con una probabilidad aleatoria de una vez en un millón.
- Puesto de otra forma: $p(\text{fingerprint} \mid \text{innocent}) = \frac{1}{1000000}$.

¿Qué crees que hará el investigador?

Una aplicación: El Error del Investigador

Argumentar que la probabilidad de inocencia es una en un millón.

- En resumen: los investigadores rutinariamente se olvidan de que $p(B \mid A) \neq p(A \mid B)$!

El caso de Sally Clark

Algunos detalles sobre el caso:

- Sally Clark era una abogada británica.
- 13 Dic 1996: su primer hijo muere en la cuna a las 11 semanas.
- 26 Ene 1998: su segundo hijo muere a las 8 semanas.
- 23 Feb 1998: Clark es arrestada acusada de doble asesinato.
- Ella sufría síntomas de depresión post-parto.
- Los dos bebés tenían evidencias de trauma (ligados a intentos de resucitación).

El caso de Sally Clark

El el juicio, los investigadores presentaron a un especialista pediátrico que estimo que la probabilidad de muerte en una cuna de dos bebés saludables era de 1 en 73 millones.

- Poniéndolo de otra forma: $p(\text{two crib deaths} \mid \text{innocent}) = \frac{1}{73000000}$.
- Los investigadores alegaron: $p(\text{innocent} \mid \text{two crib deaths}) = \frac{1}{73000000}$.

9 Nov 1999: Sally Clark fue sentenciada a prisión en vida.

El error en este caso

Vamos a llenar los blancos para ilustrar el error. Va el teorema reelaborado:

$$p(H | D) = \frac{p(D | H)p(H)}{p(D | H)p(H) + p(D | A)p(A)}$$

Aumimos:

- H = ambos bebés fallecen de muerte en cuna.
 - $p(H) = \frac{1}{100000}$. Sí, el experto pediátrico confundió la probabilidad conjunta con la condicional.
- D = ambos bebés fallecen.
 - Trivialmente, $p(D | H) = 1$.
- A = ambos bebés mueren de otras causas (i.e. homicidio).
 - $p(A) = 1 - p(H)$.
- $p(D | A) = \frac{30}{650000}$ in este caso.

Hay *muchas* piezas moviéndose en este caso, pero nos sirve para ilustrar el problema.

El error en este caso

$$\begin{aligned} p(H | D) &= \frac{p(D | H)p(H)}{p(D | H)p(H) + p(D | A)p(A)} \\ &= \frac{.00001}{.00001 + .0000046 * (1 - .00001)} \\ &= \frac{.00001}{.0000145} \\ &= .689 \end{aligned}$$

En otras palabras, la probabilidad de que Sally Clark fuera inocente era mucho mas alta del testimonio dado por los investigadores.

Sally Clark

La Sociedad Británica de Estadística eventualmente encontró el error y lo condenó.

- Ante la falta de contexto apropiado (i.e. la probabilidad de que una madre asesinará a sus dos hijos consecutivamente), la condena a Sally Clark fue errónea.

Clark fue exonerada en el 2003 y murió en 2007.

Conteo, Permutaciones y Combinaciones

¿Qué significa esto?

- Si tenemos un número específico de pasos individuales. . .
- . . . cada uno tienen un conjunto de alternativas. . .
- . . . entonces el número total de alternativas es el producto de ellas en cada paso.

Entonces, para 1, 2, . . . k características diferentes, multiplicamos el correspondiente n_1, n_2, \dots, n_k número de características.

Cuatro Métodos de Conteo

Una forma de conteo sigue reglas de **ordenamiento** y **reemplazo**.

1. Ordenado, con reemplazo
2. Ordenado, sin reemplazo
3. No ordenado, sin reemplazo

Hay un cuarto método (no ordenado, con reemplazo), aunque intuitivo no se usa mucho. No lo vamos a ver.

Ordenado, son Reemplazo

Es el primero y más fácil de los métodos.

- Digamos que tenemos n objetos (e.g. una baraja de 52 cartas).
- Queremos escoger $k < n$ objetos (por ejemplo: 5 cartas).
- Con reemplazo: ponemos la carta de regreso en la baraja y volvemos a escoger.

Ordenado, con Reemplazo

Por el **Teorema Fundamental de Conteo**, siempre hay n selecciones para cada una de las 5 etapas de decisión.

- Puesto de otra forma: es posible escoger un Rey de Corazones 5 veces.
- El número total de combinaciones = $n^k = 52^5 = 380,204,032$.

Ordenado, sin Reemplazo

- En este caso, una vez que tescogemos el Rey de Corazones, se va de la baraja.
- Entonces, el número de cartas se reduce en cada etapa. Formally:

$$n * (n - 1) * (n - 2) * (n - 3) * \dots * (k + 1) * k = \frac{n!}{(n - k)!}$$

Nota: ! = factorial. $5! = 5 * 4 * 3 * 2 * 1$.

- En este ejemplo, $\frac{52!}{(52-5)!} = 311,875,200$.

No ordenado, sin Reemplazo

Un poco más complicado, pero todavía común.

- Informalmente: como ordenado sin reemplazo, pero no podemos ver el orden.

Supongamos que estamos escogiendo pelotas de colores de una urna $S = \{ \text{Blanca, Roja} \}$.

- Aquí, escoger RBR, RRB, and BRR son equivalentes.

Esto genera una pequeña modificación en la fórmula.

$$\frac{n!}{(n-k)!k!} = \binom{n}{k}$$

Una Aplicación: Muestra Aleatoria de una Población

Supongamos que tenemos una población de 150. Queremos entrevistar a 15.
¿Cuántas combinaciones diferentes tenemos?

Ordenado, con Reemplazo: $n^k = 150^{15} = 4.378939 * 10^{32}$

Ordenado, sin Reemplazo: $\frac{n!}{(n-k)!} = 2.123561 * 10^{32}$

No ordenado, sin Reemplazo: $\binom{n}{k} = \binom{150}{15} = 1.623922 * 10^{20}$

Otra : Formación de Gobierno de Coalición

Va una aplicación de política europea.

- Supongamos que tenemos tres partidos (Liberal, Demócrata Cristiano, Verdes).
- Liberals tienen 6 miembros senior. DCs tienen 5 miembros senior. Verdes tienen 4 miembros senior.

Cuántas formas distintas hay de escoger un gabinete con 3 liberales, 2 demócratas cristianos y 3 verdes?

Formación de Gobierno de Coalición

Lo podemos resolver con la regla de conteo **No ordenado, sin reemplazo**.

- No importa el orden.
- Sólo hay que seleccionar el número de miembros por partido que se requiere.

$$\binom{6}{3} \binom{5}{2} \binom{4}{3} = \frac{720}{6(6)} * \frac{120}{2(6)} * \frac{24}{6(1)} = 20 * 10 * 4 = 800$$

Table of Contents

Objetivos para hoy

Probabilidad

Probabilidad Básica