Clase 1: Repaso de Probabilidades

Pablo Barceló
DCC, Universidad de Chile

1. Motivación

La estadística es una noción fundamental para la ciencia de los datos. Por otro lado, las probabilidades constituyen el lenguaje de los fundamentos de la estadística, y por ende es importante conocerlas. Por ejemplo, para entender qué significa que una muestra tenga sesgo, cuáles son las garantías que nos entrega la regresión lineal, o cuál es la confianza con que podemos aseverar la conclusión de un test, es fundamental entender ciertas nociones de probabilidad. En este capítulo presentamos algunas de esas nociones.

Aunque en la vida cotidiana comúnmente las personas hablan acerca de la *probabilidad* de que algo ocurra, no es para nada claro qué es lo que tienen en mente cuando lo dicen. La teoría de probabilidades intenta dar cuenta de esto. Aún así, no hay una sola acepción de la noción de probabilidad, y al menos es posible identificar tres corrientes principales:

- *Interpretación frecuentista:* La probabilidad de un evento se refiere a la frecuencia de una salida en un experimento si pudiéramos repetir tal experimento un gran número de veces bajo las mismas condiciones. Por ejemplo, si lanzamos muchas veces una moneda esperamos que salga cara aproximadamente la mitad de las veces. Pero. ¿cuántas veces debemos repetir el experimento? ¿Qué significa que este deba ser realizado bajo las mismas condiciones? ¿Qué desviación del valor esperado se considera tolerable?

- *Interpretación clásica:* Basada en la noción de salidas *igualmente factibles*; por ejemplo, cara y sello son las dos salidas igualmente factibles del experimento que consiste en lanzar una moneda. Por tanto, la probabilidad de cada evento debe ser 1/2. Pero, ¿qué significa "igualmente factibles"? ¿Cómo asignamos probabilidades a salidas de experimentos que no son "igualmente factibles"?

- *Interpretación subjetiva:* Representa la percepción individual de que una cierta salida de un experimento pueda ocurrir. Esta interpretación es positiva porque es más agnóstica que las anteriores, y además considera la observación y contexto como partes de la definición. Por otro lado, es ciertamente difícil de formalizar.

De todas formas, más allá de la interpretación filosófica que demos a la noción de probabilidad, trabajaremos con una noción matemática que se abstrae de estas interpretaciones y desarrolla una teoría pragmática basada en una noción frecuentista bien definida. Además, ha demostrado ser de enorme utilidad en la práctica por lo que pensamos se ajusta bien a la noción real de probabilidad existente en el mundo.

2. Nociones Básicas

Experimento: Proceso que entrega un conjunto de resultados que podemos determinar antes de tiempo (por ejemplo, lanzamiento de un dado, lanzamiento de tres monedas, testeo de una ampolleta, etc.) No distinguimos entre un experimento y el conjunto de resultados que este entrega.

Evento: Cualquier conjunto de resultados posibles del experimento (por ejemplo, que el dado entregue un resultado par, que al menos dos monedas den sello, que la ampolleta dure más de cinco días, etc.)

Por el momento nos enfocaremos en experimentos que entregan un conjunto finito $S = \{s_1, \ldots, s_n\}$ de resultados. Por tanto, un evento E correponde a un subconjunto de S.

Noción de probabilidad (caso simple): Asumamos inicialmente que todos los resultados del experimento S son igualmente posible. Entonces la *probabilidad del evento* $E \subseteq S$ se define como $|E|/|S|$ y se denota por $Pr(E)$.

Ejercicio: ¿Cuál es la probabilidad de que al tirar un dado obtengamos un número par? Respuesta: 1/2.

Ejercicio: ¿Cuál es la probabilidad de que al lanzar tres monedas obtengamos exactamente dos caras? Respuesta: 3/8.

Métodos de conteo: Existen métodos de conteo que nos permiten determinar la cardinalidad de un evento.

Regla del producto: Suponga que un experimento consiste de dos fases independientes realizadas una tras la otra. La primera fase puede realizarse de m formas y la segunda de n formas. Entonces el experimento puede realizarse de mn formas. Por ejemplo, la cardinalidad del conjunto de eventos en el lanzamiento de dos dados es $6 \times 6 = 36$.

Ejercicio: ¿Cuántas posibles salidas tiene el experimento de lanzar 5 veces una moneda? Respuesta: 2^5 salidas.

Ejercicio: ¿De cuántas formas pueden alinearse a 6 personas? Respuesta: $6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1$. En general, alinear n objetos se puede realizar de $n \cdot (n-1) \cdots 1$ formas. Este número se se conoce como $n!$

Ejercicio: ¿De cuántas formas es posible alinear a tres personas entre un grupo de seis? Respuesta: $6 \cdot 5 \cdot 4$. En general, alinear a k personas entre n personas, para $1 \leq k \leq n$, se puede realizar de $n(n-1)(n-2) \cdot (n-k+1)$ maneras. Note que este número corresponde a $n!/(n-k)!$

Combinaciones: ¿Cuántos grupos de tres personas podemos elegir entre seis personas? Sabemos que podemos alinear a estas personas de $6 \cdot 5 \cdot 4$ formas. Sin embargo, estamos contando cada grupo en 3! alineaciones diferentes. Por tanto, el número de grupos es realmente $(6 \cdot 5 \cdot 4)/3!$. En general, podemos elegir $n!/\big((n-k)! \cdot k!\big)$ grupos diferentes de k elementos en un grupo de n elementos, para $1 \leq k \leq n$. Este número se denota usualmente como $\binom{n}{k}$.

Ejercicio: ¿Cuál es la probabilidad de que una mano de póker de cinco cartas contenga exactamente cuatro cartas del mismo tipo? Respuesta: Hay $\binom{52}{5}$ manos de póker de cinco cartas. De ellas $13 \cdot 48$ contienen cuatro cartas del mismo tipo. La probabilidad es por tanto $(13 \cdot 48)/\binom{52}{5}$.

Noción de probabilidad (caso general): ¿Qué sucede cuando algunos resultados del experimento ocurren más frecuentemente que otros? A cada $s \in S$ le asignamos una probabilidad $0 \leq Pr(s) \leq 1$ de tal forma que $\sum_{s \in S} p(s) = 1$. Entonces la probabilidad del evento $E \subseteq S$ corresponde a $\sum_{s \in E} Pr(s)$.

Ejercicio: Considere un dado en el que el número 1 sale dos veces más frecuentemente que cualquier otro número. ¿Cuál es la probabilidad de que al lanzar este dado obtengamos un número impar? Respuesta: 4/7.

Ejercicio: Considere un dado cargado en el que sale sello con probabilidad $1/3$. ¿Cuál es la probabilidad de que al lanzar el dado dos veces salgan dos sellos? Respuesta: Cada lanzamiento del dado es independiente. Por tanto, la probabilidad de que salgan dos caras es igual al producto de las probabilidades de que en cada lanzamiento salga una cara. Esto es 1/9.

Propiedades de la noción de probabilidad:

1. Para todo evento E se cumple que $0 \leq Pr(E) \leq 1$.
2. $Pr(S) = 1$.
3. Si E y F son eventos, entonces $Pr(E \cup F) = Pr(E) + Pr(F) - Pr(E \cap F)$.
4. Si E es un evento y $\bar{E}$ es el evento $S \setminus E$, entonces $Pr(\bar{E}) = 1 - Pr(E)$.

Ejercicio: ¿Cuál es la probabilidad de que no salga un 1 al lanzar un dado? Respuesta: 1 - Pr(sale un 1) = 1 - 1/6 = 5/6.

Ejercicio: Un paciente tiene probabilidad 0,7 de tener infección viral y 0,4 de tener infección bacterial. ¿Cuál es la probabilidad de que tenga ambas? Respuesta: Sea $A =$ tener infección viral y $B =$ tener infección bacterial. Entonces

$Pr(A \cup B) = 1$, $Pr(A) = 0,7$ y $Pr(B) = 0,4$. Dado que $Pr(A \cup B) = Pr(A) + Pr(B) - Pr(A \cap B)$, concluímos que la probabilidad $Pr(A \cap B)$ de que el paciente tenga ambas infecciones es 0,1.

Ejercicio: ¿Cuál es la probabilidad de que en un grupo de n personas no existan dos con el mismo cumpleaños? Respuesta: El número de salidas posibles es 366^n. El número de salidas que asignan cumpleaños distintos a todas estas personas es $366 \cdot 365 \ldots (366 - n + 1)$. Por tanto, la probabilidad es $\frac{366 \cdot 365 \ldots (366-n+1)}{366^n}$. Considere ahora la probabilidad de que en un grupo con n personas hayan dos con el mismo cumpleaños. Esta corresponde entonces a $1 - \frac{366 \cdot 365 \ldots (366-n+1)}{366^n}$. A medida que n crece esta probabilidad también lo hace. Es posible demostrar que el menor n para el cual esta probabilidad es mayor a $1/2$ es 23. Con 60 personas la probabilidad ya es de 0.994.

Eventos independientes: Dos eventos E y F son independientes si $Pr(E \cap F) = Pr(E) \cdot Pr(F)$.

Ejercicio: Sea A el evento que corresponde a que el lanzamiento de un dado entregue resultado par y B que salga un valor en $\{1, 2, 3, 4\}$. ¿Son estos eventos independientes? La probabilidad de A es $1/2$ y la de B es $2/3$. Por otro lado, $Pr(A \cap B) = 1/3$. Es decir, $Pr(A \cap B) = Pr(A) \cdot Pr(B)$ y los eventos son independientes.

Ejercicio: Asuma que una familia tiene dos hijos. ¿Cuál es la probabilidad de que ambos sean hombres? ¿Cuál es la probabilidad de que al menos uno sea hombre? ¿Cuál es la probabilidad de que ambos eventos ocurran? ¿Qué puede concluir al respecto? Respuesta: La probabilidad de que al menos uno de los hijos sea hombre (evento A) es 3/4. La probabilidad de que ambos sean hombres (evento B) es 1/4. La probabilidad de que ocurran ambos eventos (es decir, $Pr(A \cap B)$) es también 1/4. Esto quiere decir que A y B no son independientes ya que $Pr(A \cap B) \neq Pr(A) \cdot Pr(B)$.

Ejercicio: Dos máquinas son operadas de forma independiente. Los eventos A y B corresponden a que la primera o segunda máquina fallen, respectivamente. Sabemos que $Pr(A) = 1/3$ y $Pr(B) = 1/4$. ¿Cuál es la probabilidad de que al menos una de las máquinas falle? Respuesta: Sabemos que $Pr(A \cup B) = Pr(A) + Pr(B) - Pr(A \cap B)$. Es decir, la probabilidad de que A o B fallen es $1/3 + 1/4 - Pr(A \cap B)$. Pero A y B son independientes, por lo que $Pr(A \cap B) = 1/3 \cdot 1/4 = 1/12$. Concluímos que $Pr(A \cup B) = 1/3 + 1/4 - 1/12 = 1/2$.

Probabilidades condicionales: El saber que un evento ocurre puede modificar la probabilidad de que otro lo haga. A esto lo llamamos probabilidades condicionales. En general, escribimos $Pr(A \mid B)$ para referirnos a la probabilidad del evento A dado que ocurre B. Podemos asumir entonces que $Pr(B) > 0$ y que $Pr(A \mid B) = Pr(A \cap B)/Pr(B)$.

Ejercicio: Sabemos que al lanzar dos dados la suma de los resultado es impar. Llamemos B a este evento. ¿Cuánto afecta esto al evento A que corresponde a que la suma sea además menor a ocho? Respuesta: Tenemos que $Pr(A) = 1/2$ y $Pr(B) = 1/2$. Por otro lado, $Pr(A \cap B) = 1/3$. Por tanto, $Pr(A \mid B) = 2/3$. Es decir, saber que B ocurre aumenta la probabilidad que A lo haga.

Teorema de Bayes: Suponga que queremos determinar si un mensaje es spam utilizando las ocurrencias de ciertas palabras en el mensaje. Si conocemos (a) el porcentaje de mensajes que son spam, (b) el porcentaje de mensajes que son spam y en los que estas palabras ocurren, y (c) el porcentaje de mensajes que no son spam y en los que estas palabras también ocurren, entonces podemos utilizar para esto un importante resultado de la teoría de probabilidades conocido como el teorema de Bayes.

El teorema se enuncia de la siguiente forma. Suponga que E y F son eventos tal que $Pr(E) > 0$ y $Pr(F) > 0$. Entonces:

$$Pr(F \mid E) = \frac{Pr(E \mid F)Pr(F)}{Pr(E \mid F)Pr(F) + Pr(E \mid \bar{F})Pr(\bar{F})}.$$

Ejercicio: Suponga que la palabra "Rolex" aparece en 250 de 2000 mensajes que sabemos que son spam y en tan solo 5 de 1000 mensajes que no lo son. Calcule la probabilidad de que un mensaje sea spam si contiene a la palabra "Rolex", asumiendo que existe la misma probabilidad de que un mensaje sea o no sea spam. Respuesta: Por Teorema de Bayes esto corresponde a $\frac{(250/2000) \cdot 1/2}{\big((250/2000) \cdot 1/2 + (5/1000) \cdot 1/2\big)} = \frac{1/8}{1/8 + 1/200} \approx 0,962$.

3. Variables Aleatorias

Variable aleatoria: Una *variable aleatoria* $X : S \to \mathbb{R}$ asigna un valor real positivo $X(s)$ a cada resultado del experimento S. Por ejemplo, se lanza una moneda 10 veces y X_c cuenta el número de caras en cada resultado.

Dado que S es finito el conjunto de valores de la forma $X(s)$ también es finito. Este tipo de variables aleatorias se denomina *discretas*.

Dado $r \in \mathbb{R}$, denotamos por $Pr(X = r)$ la probabilidad del evento que contiene todos los $s \in S$ tal que $X(s) = r$. Por ejemplo, $Pr(X_c = 5)$.

Función de probabilidad: Definimos $f : \mathbb{R} \to [0,1]$ de tal forma que $f(r) = Pr(X = r)$ para todo $r \in \mathbb{R}$. Note que $\sum_{r\in\mathbb{R}} f(r) = 1$. Por otro lado la *función de probabilidad acumulativa* corresponde a $F : \mathbb{R} \to [0,1]$ de tal forma que $F(r) = \sum_{r'\leq r} f(r')$. Esta función tiene especial sentido en algunos escenarios que veremos más adelante en los que el espacio de muestras es infinito.

Ejercicio: Una fábrica hace ampolletas con probabilidad $1/10$ de ser defectuosas. Sea X_a una variable aleatoria que cuenta el número de ampolletas defectuosas luego de fabricar 3 ampolletas. Calcule el valor de la función de probabilidad $f : \mathbb{R} \to [0,1]$ definida por X_a. Respuesta: Denotamos por d, o una salida defectuosa y no defectuosa, respectivamente. Note que X_a toma valores en $\{0,1,2,3\}$. La única salida s del experimento tal que $X_a(s) = 0$ es ddd, la que ocurre con probabilidad $(1/10)^3$. Esto quiere decir que $f(0) = P(X_a = 0) = (1/10)^3$. Por otro lado, existen tres salidas s tal que $X_a(s) = 1$. Estas son *doo*, *odo*, y *ood*, cada una de las cuales ocurre con probabilidad $0,1 \cdot (0,9)^2$. Por tanto, $f(1) = Pr(X_a = 1) = 3\cdot 0,1\cdot(0,9)^2$. De igual forma concluimos que $f(2) = Pr(X_a = 2) = 3\cdot 0,9\cdot(0,1)^2$ y $f(3) = Pr(X_a = 3) = (0,9)^3$.

Distribución binomial: Un experimento tiene dos posibles resultados s y t, los que ocurren con probabilidad p y $1-p$, respectivamente. Asuma que el experimento se repite n veces y X es una variable aleatoria que cuenta el número de veces en que el resultado fue s. Entonces $P(X = k)$, para $0 \leq k \leq n$, es $\binom{n}{k}p^k(1-p)^{n-k}$. La función de probabilidad asociada con esta variable aleatoria se llama *distribución binomial*.

4. Valor Esperado y Varianza

Valor esperado: Representa el "centro de gravedad" de nuestra variable aleatoria. Formalmente lo definimos como:

$$E(X) := \sum_{r\in\mathbb{R}} r \cdot Pr(X = r) = \sum_{r\in\mathbb{R}} r \cdot f(r).$$

Ejercicio: Asuma que X toma valores 2, 3, 5 y 8 con probabilidades 0,1, 0,3, 0,2 y 0,4, respectivamente. ¿Cuál es el valor esperado de X? Respuesta: $E(X) = 2 \cdot 0,1 + 3 \cdot 0,3 + 5 \cdot 0,2 + 8 \cdot 0,4 = 5,3$.

Ejercicio: ¿Cuál es el valor esperado de X_a en nuestro ejercicio anterior? Respuesta: $E(X_a) = 0 \cdot (0,1)^3 + 1 \cdot 3 \cdot 0,1 \cdot (0,9)^2 + 2 \cdot 3 \cdot 0,9 \cdot (0,1)^2 + 3 \cdot (0,9)^3$.

Propiedades del valor esperado:

1. $E(aX + b) = aE(X) + b$.
2. $E(X + Y) = E(X) + E(Y)$.

Ejercicio: Un recipiente contiene bolas rojas y verdes, de tal forma que el porcentaje de bolas rojas es $0 \leq p \leq 1$. Se sacan n bolas al azar con reemplazo. ¿Cuál es el número esperado de bolas rojas obtenidas? Respuesta: Sea X el número de bolas rojas al elegir n bolas al azar con reemplazo del recipiente. Entonces $X = X_1 + \cdots + X_n$, donde X_i es una variable aleatoria que toma valor 1 si la i-ésima bola de la muestra es roja y toma valor 0 si no ($1 \leq i \leq n$). Entonces $E(X) = E(X_1+\cdots+X_n) = E(X_1)+\cdots+E(X_n)$. Por definición, $E(X_i) = 0\cdot Pr(X_i = 0)+1\cdot Pr(X_i = 1) = Pr(X_i = 1)$. Esto último coincide con la probabilidad de que la i-ésima bola elegida sea roja, lo que equivale a p por definición. Concluímos que $E(X) = np$.

Varianza: Mide cuán distribuído esta $f(x)$ en torno a $E(X)$. Se define como:

$$Var(X) \;=\; E((X-E(X))^2) \;=\; E(X^2)-E(X)^2.$$

Ejercicio: Compare las varianzas de las siguientes funciones de probabilidad:

$$f(x) \;=\; \begin{cases} 0,5, & \text{si } x=0 \\ 0,5, & \text{si } x=1 \end{cases} \qquad\qquad g(x) = \begin{cases} 0,5, & \text{si } x=0 \\ 0,499, & \text{si } x=1 \\ 0,001 & \text{si } x=10000 \end{cases}$$

Respuesta: El valor esperado de la variable que define $f(x)$ es $0,5$ y de la variable que define $g(x)$ es $10,5$. Esto implica que en el primer caso la varianza es $0,25$ y en el segundo es cercano a 10^5.

Propiedad de la varianza: Si $X_1,\ldots,X_n$ son variables aleatorias independientes, entonces $Var(X_1+\cdots+x_n) = Var(X_1)+\cdots+Var(X_n)$.

Ejercicio: Repita el ejercicio de las bolas verdes y rojas, pero ahora calcule la varianza de X. Respuesta: Dado que $X = X_1+\cdots+X_n$ y los X_i's son independientes entre si, tenemos que $Var(X) = Var(X_1)+\cdots+Var(X_n)$. Por otro lado, $Var(X_i) = E(X_i^2)-E(X_i)^2 = E(X_i)-E(X_i)^2 = p-p^2 = p(1-p)$. Concluímos que $Var(X) = np(1-p)$.

Desigualdad de Chebyshev: Sea $X : S \to \mathbb{R}$ una variable aleatoria. Entonces

$$Pr(|X(s)-E(X)| \geq r) \;\leq\; Var(X)/r^2.$$

Ejemplo: Vuelva al ejercicio de las bolas verdes y rojas. La probabilidad de que el número de bolas rojas obtenidas difiera en al menos $\sqrt{n}$ del valor esperado np es a lo más $\frac{np(1-p)}{\sqrt{n}^2}$ por la desigualdad de Chebyshev. Por tanto, es a lo más $p(1-p)$.

Mediana: Definamos $Pr(X \leq r)$ como $\sum_{r' \leq r} Pr(X = r')$. De igual forma definimos $Pr(X \geq r)$. La *mediana* de X se define como un valor r tal que tanto $Pr(X \leq r)$ como $Pr(X \geq r)$ son mayores o iguales a $1/2$. Es posible demostrar que una mediana de X siempre existe.

Ejercicio: Asuma que X toma valores 1, 2, 3, y 4 con probabilidad 0,1, 0,4, 0,3, y 0,2, respectivamente. Encuentre una mediana para X. Respuesta: Es fácil demostrar que tanto 2 como 3 son medianas. Esto demuestra que la mediana no es única.