La distribución de Bernoulli y sus ramificaciones

Juan Piccini

Contents

1 Introducción

Este material pretende servir como complemento a algunos temas que por falta de tiempo suelen quedar en el debe.

Su intención es presentar dichos temas de un modo que busca equilibrar lo formal con lo coloquial, para que nos sea ni puro formalismo ("solo personal autorizado") ni puras intuiciones sin soporte ("mucha pluma y poca carne").

2 La distribución de Bernoulli

Es quizás la más básica de las distribuciones, dado que está asociada al experimento aleatorio más simple.

En efecto, consideremos un experimento aleatorio con sólo dos resultados posibles (que llamaremos éxito y fracaso), tal como lanzar una moneda.

Estas etiquetas son arbitarias, podrían ser "cara"y "número"o "1"y "0".

Lo medular es que cada vez que repetimos dicho experimento siempre observaremos uno de dos resultados posibles (siempre los mismos) y que la probabilidad de cada uno siempre es la misma: p es la probabilidad de que al ejecutar el experimento resulte en éxito y $q = 1 - p$ es la probabilidad de obtener fracaso.

Esto es, si lanzamos una moneda entonces siempre es la misma moneda y las condiciones experimentales son siempre iguales (la lanzamos siempre de la misma manera, sobre la misma superficie, etc.).

Asimismo se asume que los resultados de cada repetición del experimento son independientes (cada vez que lanzo la moneda no tengo forma de saber cuál va a ser el resultado, lo que haya sucedido en otros lanzamientos no afecta al resultado de éste).

A cada repetición se le suele llamar **Experimento de Bernoulli**, y a la secuencia de ellos **Proceso de Bernoulli** o Proceso Bernoulli.

Si llamamos X a la variable aleatoria (VA de ahora en más) que indica el resultado obtenido en una realización del experimento (que denotaremos con uno y cero), tendremos que

$$X(w) = \begin{cases} 1 & \text{Si el resultado es éxito} \\ 0 & \text{Si el resultado es fracaso} \end{cases}$$

De lo dicho surge que $P(X = 1) = p$ y $P(X = 0) = 1 - p$.

3 Primera rama: la distribución Binomial

Supongamos ahora que fijamos una cantidad n que serán las veces que repetiremos un experimento Bernoulli.

Al finalizar contamos la cantidad de éxitos obtenidos en esos n ensayos de Bernoulli (que es otra forma de referirse a ello).

Llamemos X a la VA que cuenta la cantidad de éxitos observados en n ensayos de Bernoulli, es claro que X solamente puede arrojar valores $0, 1, 2, \ldots, n$, esto es, $Im(X) = \{0, 1, \ldots, n\}$.

Nos interesa hallar la función de probabilidad o cuantía de X, esto es, nos interesa hallar $P_x : \mathbb{R} \to \mathbb{R}$ definida por $P_X(x) = P(X = x)$, donde $X = x$ es una forma abreviada de escribir el evento $w \in \Omega : X(w) = x$ (recordemos que P es una probabilidad, su dominio son eventos, sucesos, en este caso cada w representa n lanzamientos de moneda).

Para ilustrar supongamos $n = 5,\ x = 2,$ ¿ qué probabilidad hay de obtener 2 caras (éxitos) en 5 lanzamientos de una moneda?

Resulta natural representar el resultado de cada experimento (lanzar 5 veces una moneda) por una 5-upla de ceros y unos, donde en el lugar i pondremos un 1 si en el i-simo lanzamiento salió cara y un cero si fue número.

Por ejemplo una forma en la que puede darse el evento X=2 (o A="2 éxitos en 5 intentos") es (0,1,0,1,0), es una de las realizaciones posibles.

Si llamamos A_i al evento "éxito en el i-simo intento", tenemos que $A = A_1^c \cap A_2 \cap A_3^c \cap A_4 \cap A_5^c$.

Como las repeticiones son independientes, tenemos que
$P(X = 2) = P(A) = P(A_1^c)P(A_2)P(A_3^c)P(A_4)P(A_5^c) = (1-p)p(1-p)p(1-p) = p^2(1-p)^3$.

Notemos que sin importar cuándo ocurren los éxitos y los fracasos (o sea en qué lugares están los ceros y los unos), siempre son 2 éxitos y 3 fracasos, por lo que cualquiera sea la forma en la que tuvieron lugar los 2 éxitos y 3 fracasos, todas ellas tienen la misma probabilidad: $p^2(1-p)^3$.

Por tanto si llamamos $|A|$ a la cantidad de formas en las que puede darse el evento A, tenemos que $P(A) = \sum_{|A|} p^2(1-p)^3 = |A|p^2(1-p)^3$.

Tenemos que $|A|$ es la cantidad de 5-uplas de ceros y unos que podemos formar con 2 unos y 3 ceros.

Notemos que una vez que ubicamos los 2 unos, quedan también determinados los lugares en los que van los 3 ceros (y viceversa).

La cantidad de maneras en las que podemos elegir 2 lugares de entre 5 disponibles es C_2^5, por lo que $P(X=2) = C_2^5 p^2(1-p)^3$.

En fácil ver que en el caso general tendremos $P(X=k) = C_k^n p^k(1-p)^{n-k}$ dado que ahora tendremos C_k^n n-uplas con k unos y n-k ceros, donde cada una de dichas n-uplas tendrá la misma probabilidad de ocurrir, que es $p^k(1-p)^{n-k}$.

Decimos entonces que la VA X tiene una distribución Binomial de parámetros n y p, ya que si fijamos la cantidad de repeticiones n y la probabilidad de éxito p en cada una tenemos todo lo necesario para hallar la función de cuantía de X.

En resumen, X tiene distribución Binomial de parámetros n y p y se escribe $X \sim Bin(n,p)$ cuando $Im(X) = \{0,1,\ldots,n\}$ y $P_X(k) = C_k^N p^k(1-p)^{n-k}\ \forall k = 0,1,2,\ldots,n$.

Puede probarse que en efecto P_X es una función de cuantía, que $\sum_{k=0}^{n} P_X(k) = 1$ (el que $P_X(k) \geq 0$ es obvio).

3.1 Caso límite: la distribución de Poisson

Ahora el experimento aleatorio consiste en observar la aparición de un determinado suceso en un conjunto dado, donde ahora el conjunto es un continuo.

Por ejemplo, cantidad de poros en un metro cuadrado de una plancha de hormigón, cantidad de clientes que ingresan en una hora a un local, cantidad de accesos por minuto a un servidor, cantidad de colonias de bacterias por centímetro cuadrado en una placa de Petri, etc.

La variable de interés sigue siendo contar la cantidad de éxitos (que se dé el evento de interés), pero ahora lo que se observa no ocurre en un conjunto discreto de n elementos, sino que se da en un conjunto continuo con infinitos elementos.

Hay dos hipótesis que son fundamentales:

- Supondremos que lo que está generando los éxitos y los fracasos es estable.

  Esto es, la probabilidad de éxito no cambia en el tiempo, (si lo que observamos se da en el tiempo como por ejemplo la cantidad de pedidos de acceso a un servidor), o no cambia en el espacio (si lo que observamos se da en el espacio, como los poros en una superficie).

- Otra suposición es que los eventos aparecen en forma independiente, por ejemplo el número de sucesos en un intervalo no nos ayuda a predecir el número de sucesos en el intervalo siguiente, la cantidad de colonias de bacterias en un centímetro cuadrado no nos ayuda a predecir la cantidad en otro centímetro cuadrado.

Podemos pensar a este tipo de experimentos como una generalización del Proceso Bernoulli a un soporte continuo y lo llamaremos **Proceso de Poisson**.

Lo que llamaremos VA con distribución Poisson es la VA que cuenta la cantidad de éxitos en una unidad de observación de tamaño fijo.

Por ejemplo cantidad de sucesos en un intervalo de tiempo de longitud fija, cantidad de eventos por centímetro cuadrado, etc.

Si llamamos X a esta variable, es claro que $Im(X) = \{0, 1, 2, \ldots, n, \ldots\}$.

Nos interesa encontrar la función de cuantía $P_X(\lambda) = P(X = \lambda)$.

Por ejemplo, si estamos contando la cantidad de accesos por hora a una página web dada nos interesará calcular la probabilidad de tener λ accesos en una hora.

Veremos que podemos llegar a P_X vía la distribución Binomial. Tomaremos como ejemplo de trabajo el de los accesos por hora a una determinada página web.

El objetivo es hallar $P_X(\lambda) = P(X = \lambda)$, para ello:

1. Primero dividimos el intervalo (esa hora) en n subintervalos de igual longitud. Es claro que si n es suficientemente grande, lograremos subintervalos en los que la cantidad de éxitos (observar accesos en dicho subintervalo) sea 0 o 1, ya que a fuerza de agrandar n tendremos subintervalos tan cortos que harán imposible el que tengan lugar 2 o más accesos en un mismo subintervalo.

   Formalizando un poco más, si representamos al intervalo como $[a, b]$ y marcamos cada evento que ocurre en $[a, b]$, si por ejemplo $\lambda = 137$ entre cada una de las 137 marcas siempre tendremos una distancia positiva, y si elegimos n suficientemente grande tendremos subintervalos de longitud $\frac{b-a}{n}$ menor a esas distancias, por lo que cada subintervalo tendrá a lo sumo una marca.

   Como estamos hablando de λ accesos (éxitos) en esa hora (o sea en los n subintervalos), la VA X_i que cuenta la cantidad de éxitos en el subintervalo i será una Bernoulli de parámetro $p = \frac{\lambda}{n}$, $X_i \sim Ber(p)$, ya que cada subintervalo tiene la misma probabilidad de registrar uno de los λ éxitos.

   Por tanto la VA X que cuenta la cantidad de éxitos en esa hora es una VA Binomial de parámetros n y p, donde $np = \lambda$, np es constante.

   Si dividimos esa hora en cada vez más subintervalos donde cada uno tiene la misma probabilidad $p = \frac{\lambda}{n}$ de que en él ocurra un éxito, cuando n tiende a infinito (y por tanto p tiende a cero ya que $np = \lambda$ siempre) tendremos la probabilidad de observar λ eventos en esa hora, o sea $P(X = \lambda)$.

   Por tanto cuando n es suficientemente grande, $X \approx X_n$, donde
   $X_n \sim Bin(n, \frac{\lambda}{n})$.

2. Ahora debemos calcular el límite de una $Bin(n, \frac{\lambda}{n})$ cuando $n \to +\infty$.

   Como $P(X_n = j) = C_j^n \left(\frac{\lambda}{n}\right)^j \left(1 - \frac{\lambda}{n}\right)^{n-j}$, se trata de hallar para cada $j = 0, 1, \ldots, n$ el $\lim\limits_{n\to+\infty} P(X_n = j)$.

   Esto es, hallar $\lim\limits_{n\to+\infty} C_j^n \left(\frac{\lambda}{n}\right)^j \left(1 - \frac{\lambda}{n}\right)^{n-j}$, donde λ está fijo.

   $$P(X_n = j) = \underbrace{\left(\frac{n!}{(n-j)!j!}\right)}_{C_j^n}\underbrace{\left(\frac{\lambda^j}{n^j}\right)}_{(\frac{\lambda}{n})^j}\underbrace{\left(\frac{(1-\frac{\lambda}{n})^n}{(1-\frac{\lambda}{n})^j}\right)}_{\left(1-\frac{\lambda}{n}\right)^{n-j}}.$$

   Operando y reagrupando, obtenemos $\dfrac{n.(n-1)\ldots.(n-(j-1))\ \lambda^j(1-\frac{\lambda}{n})^n}{\underbrace{(n.n.\ldots.n)}_{n^j} j!(1-\frac{\lambda}{n})^j}$.

   Cuando $n \to +\infty$, el cociente $\frac{n.(n-1)\ldots.(n-(j-1))}{n.n.\ldots.n} = (\frac{n}{n})(\frac{n-1}{n})(\frac{n-2}{n})\ldots\frac{n-(j-1)}{n}$ tiende a 1, dado que para cada factor $\lim\limits_{n\to+\infty} \frac{n-i}{n} = 1$ para $i = 0, 1, \ldots, j-1$.

   El término $\frac{\lambda^j}{j!}$ es una constante que no se ve afectada por n, mientras que $\lim\limits_{n\to+\infty}(1 - \frac{\lambda}{n})^j = 1$ y $\lim\limits_{n\to+\infty}(1 - \frac{\lambda}{n})^n = e^{-\lambda}$.

   Sustituyendo, obtenemos que $\lim\limits_{n\to+\infty} P(X_n = j) = P(X = j) = \frac{\lambda^j}{j!}e^{-\lambda}$.

Decimos entonces que X tiene distribución Poisson de parámetro λ (lo denotamos por $X \sim \mathcal{P}(\lambda)$) cuando $Im(X) = \{0, 1, 2, \ldots\}$ y $P(X = j) = \frac{\lambda^j}{j!}e^{-\lambda}$.

Observaciones

- El parámetro λ representa la cantidad media de eventos por unidad de observación.

  Si los eventos se dan en el tiempo, es la cantidad media de eventos que se producen en una unidad de medida temporal.

  Si por ejemplo tenemos $\lambda = 5$ eventos por hora, en un intervalo de 2 horas es de esperar una media de $2\lambda = 10$ eventos, mientras que un intervalo de 12 minutos es de esperar una media de $\frac{1}{5}\lambda = 1$ evento.

  Esto es, si en media suceden λ eventos por unidad de tiempo, en un intervalo de longitud t_0 en media se darán λt_0 eventos.

  Entonces $P(X = k) = \frac{(\lambda t_0)^k}{k!}e^{-\lambda t_0}$ en un intervalo de longitud t_0.

- En muchas ocasiones en las que hay que trabajar con una $Bin(n, p)$, si n es grande (varias decenas en adelante) y p chico (p.ej. $p \leq 0.1$), en vez de operar con la función de cuantía $C_j^n p^j (1-p)^{n-j}$ (lo que siempre es engorroso), se puede aproximar dicha Binomial por una Poisson de parámetro $\lambda = np$ cuya operatoria es más sencilla.

3.2 La distribución Exponencial

Sigamos con el Proceso de Poisson de parámetro λ, pero ahora la variable de interés es el tiempo entre eventos consecutivos.

Es claro que $Im(X) = (0, +\infty)$.

Entonces dado $t_0 > 0$, $P(X > t_0) = P(\text{cero eventos en } (0, t_0]) = \frac{(\lambda t_0)^0}{0!} e^{-\lambda t_0} = e^{-\lambda t_0}$.

Por tanto $P(X \leq t_0) = F_X(t_0) = 1 - e^{-\lambda t_0}$. Diremos que la VA X tiene distribución exponencial de parámetro λ, $X \sim \mathcal{E}(\lambda)$ cuando su función de distribución es $F_X(t) = \begin{cases} 1 - e^{-\lambda t} & \text{si } t > 0 \\ 0 & \text{si } t \leq 0 \end{cases}$

Si derivamos a F_X obtenemos la función de densidad $f_X(t) = \begin{cases} \lambda e^{-\lambda t} & \text{si } t > 0 \\ 0 & \text{si } t \leq 0 \end{cases}$

4 Segunda rama: la distribución Geométrica

Volvamos a Bernoulli, pero ahora en lugar de repetir una cantidad fija n de veces y contar la cantidad de éxitos ocurridos en esas n veces, fijamos la cantidad de éxitos en uno ($k = 1$) y contamos la cantidad de repeticiones hasta que aparezca el primer éxito inclusive.

Ahora el experimento es repetir hasta que ocurra el primer éxito.

Si pensamos en monedas, lanzamos nuestra moneda una y otra vez hasta que salga la primera cara.

Ahora nuestra VA X cumple que $Im(X) = \{1, 2, 3, \ldots, n, \ldots\}$ y nos interesa hallar $P_X(k) = P(X = k)$.

Si mantenemos la notación $A_i = \{$"éxito en la i-sima vez"$\}$ entonces el evento A="la primera cara se obtuvo en el k-simo lanzamiento"es
$A = A_1^c \cap A_2^c \cap \cdots \cap A_{k-1}^c \cap A_k$, ya que fracasamos las primeras n-1 veces y nuestro primer y único éxito se dió en la última vez, la n-sima.

Recordemos que los resultados en cada vez son independientes, de donde $P(X = k) = P(A_1^c)P(A_2^c) \ldots P(A_{k-1}^c) = P(A_k) = \underbrace{(1-p)(1-p)\ldots(1-p)}_{\text{k-1 veces}}p$.

Por tanto $P(X = k) = (1-p)^{k-1}p$.

Cuando la VA X cumpla que $Im(X) = \{1, 2, \ldots, n, \ldots\}$ y que $P_X(k) = p(1-p)^{k-1}$ diremos que X tiene distribución Geométrica de parámetro p, y lo denotaremos como $X \sim Geo(p)$.

Puede probarse que P_X es una función de cuantía ya que
$$\sum_{n=1}^{+\infty} P_X(n) = \sum_{n=1}^{+\infty} p(1-p)^{n-1} = p\sum_{n=1}^{+\infty}(1-p)^{n-1} = p\sum_{n=0}^{+\infty}(1-p)^n = p.\ \frac{1}{p} = 1.$$

Ya que la serie es una serie geométrica (de ahí el nombre de la distribución) cuya suma es $\frac{1}{1-(1-p)} = \frac{1}{p}$, lo que con el factor p que multiplica nos da un total de uno.

4.1 Generalización: la distribución Binomial Negativa

Ahora en lugar de fijar $k = 1$ tomamos $k \geq 1$ y repetimos el experimento Bernoulli hasta obtener k éxitos.

Sea X la VA que cuenta la cantidad de repeticiones hasta observar el k-simo éxito inclusive.

De nuevo k está fijo, la probabilidad de éxito en cada intento p está fija y n es lo que varía.

Ahora $Im(X) = \{k, k+1, \ldots, n, \ldots\}$ ya que si quiero acumular k éxitos necesito al menos k intentos.

Para ilustrar, fijemos k=3, y supongamos que el tercer (y último) éxito se obtuvo al séptimo intento.

Supongamos que el resultado obtenido se representa por la 7-upla $(0,0,1,0,1,0,1)$, esto es, $A = A_1^c \cap A_2^c \cap A_3 \cap A_4^c \cap A_5 \cap A_6^c \cap A_7$, donde A es el suceso "el tercer éxito se obtuvo en la séptima repetición".

Entonces $P(X=7) = P(A) = P(A_1^c)P(A_2^c)P(A_3)P(A_4^c)P(A_5)P(A_6^c)P(A_7) = (1-p)(1-p)p(1-p)p(1-p)p = p^3(1-p)^4$.

Cualquier otra forma en la que pueda haberse dado el tercer éxito en el séptimo intento será una 7-upla donde en el último lugar hay un uno y en los restantes 6 lugares hay 2 unos y 4 ceros, su probabilidad siempre será un producto de 7 factores de los que 3 valen p y los 4 restantes valen $(1-p)$.

Entonces todas estas 7-uplas son equiprobables, con probabilidad $p^3(1-p)^4$.

Si llamamos $|A|$ a la cantidad de 7-uplas posibles, tendremos $P(A) = \underbrace{p^3(1-p)^4 + p^3(1-p)^4 + \cdots + p^3(1-p)^4}_{|A| \text{ veces}}$ (recordemos que los ensayos son independientes).

Para contar cuántas 7-uplas hay, recordemos que el último lugar siempre tiene un uno, y que si elegimos dónde colocar los 2 unos restantes en los 6 lugares restantes (o sea los $k-1$ unos restantes) también estaremos fijando los 4 ceros restantes (los $n-k$ ceros restantes) y viceversa.

Elegir 2 lugares de entre 6 disponibles se puede hacer de C_2^6 maneras (C_{k-1}^{n-1} en el caso general), de donde $|A| = C_2^6$ ($|A| = C_{k-1}^{n-1}$ en el caso general).

Por tanto $P(X=7) = C_2^6 p^3(1-p)^4$, en el caso general $P(X=n) = C_{k-1}^{n-1} p^k (1-p)^{n-k}$.

Nuevamente puede probarse que tenemos una función de cuantía, que $\sum_{n \geq k}^{+\infty} P(X=n) = \sum_{n \geq k}^{+\infty} C_{k-1}^{n-1} p^k (1-p)^{n-k} = 1 \; \forall k \geq 1$.

Diremos que la VA X tiene distribución Binomial Negativa de parámetros k, p, $X \sim BN(k,p)$.

p nos dice la probabilidad de éxito en cada ensayo, k la cantidad de éxitos en la que nos detendremos y lo que es aleatorio ahora es n, la cantidad de ensayos hasta observar k éxitos.

5 Semejanzas entre las distribuciones Geométrica y Exponencial: La pérdida de memoria

Ambas distribuciones miran lo mismo, cada una en su contexto.

La distribución Geométrica es la de una VA que cuenta la cantidad de intentos hasta observar algo por primera vez, donde ese algo se da en el marco de un proceso Bernoulli.

Si pensamos que cada intento insume la misma cantidad de tiempo, entonces podemos interpretarla como que cuenta el tiempo de espera hasta observar un evento.

Esto es lo mismo que hace una VA con distribución Exponencial, contar el tiempo hasta que se observa un evento que se da en el marco de un proceso Poisson.

Por tanto no es demasiado sorprendente que ambas compartan una propiedad llamada pérdida de memoria.

5.1 Pérdida de Memoria de la Distribución Exponencial

Esta propiedad puede enunciarse en términos de probabilidad condicional:

$$P(X > t + h | X > t) = P(X > h) \tag{1}$$

En palabras: supongamos que X mide el tiempo de vida de una lámpara LED y que $t = 7000$ horas y $h = 1500$ horas. Entonces la probabilidad de que la lámpara siga "viva"(esto es, funcionando) después de 8500 horas sabiendo que hace más de 7000 horas que viene funcionando, es la misma probabilidad de que una lámpara nueva supere las 1500 horas sin romperse.

La lámpara "no recuerda"que ya llevaba 7000 horas encendida.

Esta propiedad puede demostrarse p.ej. a partir de la función de distribución.

$$\begin{aligned}
P(X > t + h | X > t) &= \frac{P(X > t + h, X > t)}{P(X > t)} = && (2)\\
&\frac{P(X > t = h)}{P(X > t)} = && (3)\\
&\frac{1 - (1 - e^{-\lambda(t+h)})}{1 - (1 - e^{\lambda t})} = && (4)\\
&\frac{e^{-\lambda t} e^{\lambda h}}{e^{-\lambda t}} = && (5)\\
&e^{-\lambda h} = P(X > h) && (6)
\end{aligned}$$

Donde se ha utilizado la definición de probabilidad condicional, que $\{X > t + h\} \cap \{X > t\} = \{X > t + h\}$ y que $e^{-\lambda(t+h)} = e^{-\lambda t} e^{-\lambda h}$.

5.2 Pérdida de Memoria de la Distribución Geométrica

Ahora tanto t como h son enteros positivos, llamémosles n y m respectivamente, al igual que antes tenemos:

$$P(X > n + m | X > n) = \frac{P(X > n + m)}{P(X > n)} \tag{7}$$

$\{X > n + m\}$ implica que en los n+m intentos previos no se ha observado el evento de interés, $\{X > n + m\} = \{A_1^c \cap \cdots \cap A_{n+m}^c\}$, de donde

$$P(X > n + m | X > n) = \frac{P(X > n + m)}{P(X > n)} \tag{8}$$

$$= \frac{P(A_1^c)P(A_2^c)\dots P(A_n^c)P(A_{n+1}^c)\dots P(A_{n+m}^c)}{P(A_1^c)P(A_2)^c\dots P(A_n)^c} \tag{9}$$

$$= P(A_{n+1}^c)\dots P(A_{n+m}^c) \tag{10}$$

$$= \underbrace{(1-p)\dots(1-p)}_{\text{m veces}} \tag{11}$$

$$= P(A_1^c)\dots P(A_m^c) \tag{12}$$

$$= P(X > m) \tag{13}$$

Por tanto $P(X > n + m|X > n) = P(X > m)$. Hemos usado que los resultados de cada ensayo son independientes y que la probabilidad de éxito en cada uno de ellos siempre es p y por tanto la de fracaso es $1 - p$.

Por ejemplo, si n=120 y m=200, la probabilidad de que el primer éxito se observe después del 320-avo intento dado que hasta el 120-avo intento aún no había sucedido, es la misma que observarlo después del 200-avo intento empezando de cero.