Distribuciones de Probabilidad

Distribuciones Discretas

Edimer David Jaramillo - Bioestadística 1

Marzo de 2019

Funciones importantes con R

- `r_`: prefijo `r` (*random*) para generación de números aleatorios de alguna distribución específica _ (ejemplo: `rbinom()`, `rnorm()`, `rexp()`).
- `d_`: prefijo `d` (*density*) permite obtener la función de densidad.
- `p_`: prefijo `p` (*probability*) permite obtener probabilidades acumuladas.
- `q_`: prefijo `q` (*quantile*) permite obtener el valor asociado a determinada probabilidad.

Distribución Bernoulli (1/3)

En determinados experimentos suele ocurrir que existen sólo dos resultados posibles:

- Presencia o ausencia
- Sí o no
- Defectuoso o no defectuoso
- Ganó o perdió
- Hembra o macho

Distribución Bernoulli (2/3)

- Probabilidad de éxito (p)
- Probabilidd del fracaso ($q = 1 - p$)
- $p + q = 1$

Una variable aleatoria X tiene distribución Bernoulli si su función de densidad (f.m.p.) es:

$$f(x) = p^x q^{1-x}, x = 0, 1$$

Note que:

- $P(x = 0) = f(0) = p^0 q^{1-0} = q$
- $P(x = 1) = f(1) = p^1 q^{1-1} = p$

Notación:

$$X \sim Bernoulli(p)$$

Distribución Bernoulli (3/3)

- **Esperanza matemática:**

$$E[X] = 1 \; \times p \; + 0 \; \times (1-p) = p$$
$$E[X^2] = 1^2 \; \times p \; + 0^2 \; \times (1-p) = p$$

- **Varianza:**

$$Var[X] = E[X^2] - [E(x)]^2 = p \; - p^2 = p \times (1-p) = p \times q$$

Distribución Binomial (1/7)

Ocurre cuando se cumple lo siguiente:

- Se realizan o repiten n ensayos Bernoulli
- El parámetro p se mantiene constante entre ensayos
- Los ensayos son independientes entre sí

Objetivo: el problema de interés radica en el número de "éxitos" en n casos estudiados (el éxito no necesariamente implica que sea bueno, sólo determina la ocurrencia del evento).

Una variable aleatoria X tiene distibución Binomial si su función de densidad (f.m.p.) es:

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x}; con\ x = 0, 1, 2, ..., n$$

Distribución Binomial (2/7)

Notación:

$$X \sim B(n, p)$$

Donde:

- n: número de ensayos
- p: probabilidad de éxito

Esperanza matemática y varianza:

- $E[X] = n \times p$
- $Var[X] = n \times p \times q = n \times p \times (1 - p)$

Distribución Binomial (3/7)

- **Ejemplo:** Suponga que se toman 10 semillas de un tipo de pasto y se registran los eventos "germinó" o "no germinó" después de cinco días de la siembra. Las semillas están lo suficientemente aisladas como para garantizar respuestas independientes. Si la probabilidad de la germinación es 0.25 (cada semilla), obtener:

  1. Probabilidad de que germinen 7 de las 10 semillas
  2. Probabilidad de que germinen al menos 3 de las 10 semillas
  3. Probabilidad de que germinen máximo 5 semillas
  4. La esperanza de esta variable aleatoria
  5. La varianza

Distribución Binomial (4/7)

- **Solución manual ejemplo 1:**

$$X \sim B(10; 0.25)$$

1. $P(X = 7) = \binom{10}{7} 0.25^7 (1 - 0.25)^{10-7} = 0.0031$
2. $P(X \geq 3) = P(X = 3) + P(X = 4) + P(X = 5) + ..., P(X = 10) = 1 - P(X = 0) + P(X = 1) + P(X = 2) = 0.4744$
3. $P(X \leq 5) = P(X = 0) + P(X = 1) + ... + P(X = 5) = 0.0563 + 0.1877 + 0.2816 + 0.2503 + 0.1460 + 0.0584 = 0.9803$
4. $E[X] = n \times p = 10 \times 0.25 = 2.5$
5. $Var[X] = n \times p \times 1 = 10 \times 0.25 \times (1 - 0.25) = 1.875$

Distribución Binomial (5/7)

- **Solución con R ejemplo 1:**

Funciones importantes:

- `rbinom()`: generar números aleatorios bajo la distribución Binomial.
- `dbinom()`: función masa de probabilidad.
- `pbinom()`: función de distribución acumulada.

Distribución Binomial (6/7)

- **Solución con R ejemplo 1:**

1. $P(X = 7)$:

```
dbinom(x = 7, size = 10, prob = 0.25)
## [1] 0.003089905
```

2. $P(X \geq 3)$:

```
1 - sum(dbinom(x = 0:2, size = 10, prob = 0.25))
## [1] 0.4744072
```

```
1 - pbinom(q = 2, size = 10, prob = 0.25)
## [1] 0.4744072
```

Distribución Binomial (7/7)

- **Solución con R ejemplo 1:**

3. $P(X \leq 5)$:

```
pbinom(5, 10, 0.25)
## [1] 0.9802723
```

4. $E[X]$:

```
(media <- 10 * 0.25)
## [1] 2.5
```

5. $Var[X]$:

```
(Varianza <- 10 * 0.25 * (1 - 0.25))
## [1] 1.875
```

Distribución Poisson (1/7)

La distribución Poisson da un modelo para el número de eventos (conteos) por intervalo.

- Número de huevos de un insecto en una oviposición
- Número de bacterias en una muestra de agua
- Número de insectos por m^2
- Número de clientes por hora
- Llamadas por hora

Una variable aleatoria X tiene distribución Poisson si su función de densidad es:

$$f(x) = \frac{e^{-\lambda}\lambda^{x}}{x!}; x = 0, 1, 2, ...$$

Distribución Poisson (2/7)

Notación:

$$X \sim Poisson(\lambda)$$

- **Esperanza ($E[X]$) y varianza ($Var[X]$)**:
- $E[X] = \lambda$
- $Var[X] = \lambda$

Distribución Poisson (3/7)

- **Ejemplo 1:** Si el número promedio de picaduras de una plaga cualquiera por semilla es 0.2 (es decir, que en promedio, de cada 100 semillas, 20 tienen picadura). Obtener:

    1. ¿Cuántas de 100 semillas no tendrán picadura?
    2. ¿Cuántas tendrán una sola picadura?
    3. ¿Cuántas tendrán más de 1 picadura?

Distribución Poisson (4/7)

- **Solución manual:**
- $P(X=0)=\frac{0.2^0 e^{-0.2}}{0!}=0.819$
- $P(X=1)=\frac{0.2^1 e^{-0.2}}{1!}=0.164$
- $p(X>1)=1-P(X=0)+P(X=1)=1-0.982=0.018$

Distribución Poisson (5/7)

- **Solución con R:**

Funciones importantes:

- `rpois()`: generar números bajo la distribución Poisson
- `dpois()`: función masa de probabilidad
- `ppois()`: función de distribución acumulada

Distribución Poisson (6/7)

- **Solución con R:**

$P(x = 0)$:

```
dpois(x = 0, lambda = 0.2)
```

```
## [1] 0.8187308
```

$P(x = 1)$:

```
dpois(x = 1, lambda = 0.2)
```

```
## [1] 0.1637462
```

Distribución Poisson (7/7)

- **Solución con R:**

$P(x > 1)$:

```
1 - sum(dpois(x = 0:1, lambda = 0.2))
## [1] 0.0175231
```

```
1 - ppois(q = 1, lambda = 0.2)
## [1] 0.0175231
```

```
ppois(q = 1, lambda = 0.2, lower.tail = FALSE)
## [1] 0.0175231
```

Distribución Geométrica (1/4)

De gran utilidad cuando el problema de interés es modelar la distribución del número de ensayos necesarios para encontrar el **primer éxito**.

Una variable aleatoria X tiene distribución Geométrica si su función de densidad (f.m.p) es:

$$f(x) = p(1-p)^{x-1};\ x = 1, 2, 3, ...$$

Notación:

$$X \sim Geo(p)$$

Distribución Geométrica (2/4)

Diferencias entre distribuciones Binomial y Geométrica:

- **Binomial (1):** número fijo de ensayos n
- **Geométrica (1):** número aleatorio de ensayos n
- **Binomial (2):** número de aleatorio de éxitos X
- **Geométrica (2):** número fijo de éxitos X (en este caso, igual a 1)

Esperanza ($E[X]$) y varianza ($Var[X]$):

- $E[x] = \frac{1}{p}$
- $Var[x] = \frac{1-p}{p^2}$

Distribución Geométrica (3/4)

- **Ejemplo:** Suponga que de registros históricos en una fábrica dedicada a producir bombillos, se conoce que la probabilidad de que un bombillo esté dañado es 0.01. Se asume que los bombillos evaluados son independientes. Obtener:
  - La probabilidad de que se necesite analizar exactamente 125 bombillos para encontrar uno defectuoso.
  - La probabilidad de que se necesite analizar máximo 50 bombillos para encontrar uno defectuoso.
  - La esperanza matemática.

Solución manual punto 1:

$$f(x) = p(1-p)^{x-1} = f(125) = 0.01(1-0.01)^{125-1} = 0.0028$$

Distribución Geométrica (4/4)

- `rgeom()`: generar números bajo la distribución Geométrica
- `dgeom()`: función masa de probabilidad
- `pgeom()`: función de distribución acumulada

$P(X = 125)$:

```r
dgeom(x = 125, prob = 0.01)
## [1] 0.002847078
```

$P(X \leq 50)$:

```r
pgeom(q = 50, prob = 0.01)
## [1] 0.401044
```

$E[x]$: 100

Distribución Binomial Negativa (1/4)

Ciertos fenómenos de interés centran su atención en el número de ensayos hasta que ocurran k éxitos (por primera vez). Se puede considerar como una extensión de la distribución geométrica; si el número de éxitos fuese 1, estaríamos en el caso de dicha distribución.

Una variable aleatoria X tiene distribución Binomial Negativa si su función de densidad es:

$$f(x) = \binom{x-1}{k-1} p^k (1-p)^{x-k};\ x = k,\ k+1,\ k+2, ...$$

Notación:

$$X \sim BN(k, p)$$

Distribución Binomial Negativa (2/4)

Diferencias entre distribuciones Binomial y Binomial Negativa:

- **Binomial (1):** número fijo de ensayos n
- **B. Negativa (1):** número aleatorio de ensayos
- **Binomial (2):** número aleatorio de éxitos
- **B. Negativa (2):** número fijo de éxitos k

Esperanza ($E[X]$) y varianza ($Var[X]$):

- $E[X] = \frac{k}{p}$
- $Var[X] = \frac{k(1-p)}{p^2}$

Distribución Binomial Negativa (3/4)

- Una pareja desea intentar tener hijos hasta que tengan dos niñas. Si la probabilidad de nacer mujer es igual a 0.5, obtener:
  - La probabilidad de que la pareja lo logre en exactamente 2 nacimientos
  - La probabilidad de que la pareja lo logre en cuatro o más nacimientos
  - Promedio de nacimientos para lograr tener 2 hijas

Solución manual:

$$X \sim BN(2, 0.5)$$

- $P(X = 2) = f(2) = \binom{2-1}{2-1} 0.5^2 (1 - 0.5)^{4-2} = 0.25$
- $P(X \geq 4) = P(X = 4) + P(X = 5) + ... = 1 - P(X < 4) = 0.5$
- $E[X] = \frac{2}{0.5} = 4$

Distribución Binomial Negativa (4/4)

- `rnbinom()`: generar números bajo la distribución Binomial Negativa
- `dnbinom()`: función masa de probabilidad
- `pnbinom()`: función de distribución acumulada

$P(X = 2)$:

```
dnbinom(x = 0, size = 2, prob = 0.5)
## [1] 0.25
```

$P(X \geq 4)$:

```
1 - sum(dnbinom(x = 0:1, size = 2, prob = 0.5))
## [1] 0.5
```

$E[X]$:

```
(media_bn <- 2/0.5)
## [1] 4
```

Distribución Hipergeométrica (1/6)

- Esta distribución está sujeta a condiciones de **muestreo sin reemplazo** en las que cada elemento es diferente y será elegido hasta completar la muestra, sin restituir los elementos extraídos.
- El tamaño de la población del cual proviene la muestra es conocido.
- Permite responder a la probabilidad de obtener k éxitos en una muestra con n elementos, proveniente de una población de tamaño N.

Distribución Hipergeométrica (2/6)

Una variable aleatoria X tiene distribución Hipergeométrica si su función de densidad es:

$$f(x) = \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}};\ x = 0, 1, ..., k$$

donde:

- N: tamaño de la población
- n: tamaño de la muestra extraída
- k: número de elementos en la población original que pertenecen al evento deseado
- x: número de elementos en la muestra que pertenecen al evento deseado

Notación:

$$X \sim h(n, k, N)$$

Distribución Hipergeométrica (3/6)

Esperanza ($E[X]$) y varianza ($Var[X]$):

- $E[X] = np$
- $Var[X] = np(1-p)\frac{N-n}{N-1}$

Donde:

- $p = \frac{k}{N}$
- $\frac{N-n}{N-1}$: es el factor de corrección para poblaciones finitas

Distribución Hipergeométrica (4/6)

Ejemplo: Un lote de bovinos contiene 100 animales alimentados con la dieta A y 200 animales alimentados con la dieta B. Si se escogen 4 animales al azar, sin reemplazo. Obtener:

1. La probabilidad de que todos los bovinos sean alimentados con la dieta A.
2. La probabilidad de que dos o más bovinos sean alimentados con la dieta A.

Distribución Hipergeométrica (5/6)

Solución manual:

- N: $300 = (200 + 100)$
- n: 4
- k: 100 y 200, para la dieta A y B, respectivamente.

1. $P(X = 4) = \frac{\binom{100}{4}\binom{300-100}{0}}{\binom{300}{4}} = 0.0119$
2. $P(X \geq 2) = \frac{\binom{100}{2}\binom{300-100}{2}}{\binom{300}{4}} + \frac{\binom{100}{3}\binom{300-100}{3}}{\binom{300}{4}} + \frac{\binom{100}{4}\binom{300-100}{0}}{\binom{300}{4}} = 0.408$

Distribución Hipergeométrica (6/6)

- `rhyper()`: generar números bajo la distribución Hipergeométrica
- `dhyper()`: función masa de probabilidad
- `phyper()`: función de distribución acumulada

$P(X = 4)$:

```
dhyper(x = 4, m = 100, n = 200, k = 4)
## [1] 0.01185408
```

$P(X \geq 2)$:

```
sum(dhyper(x = 2:4, m = 100, n = 200, k = 4))
## [1] 0.4074057
```

```
1 - phyper(q = 1, m = 100, n = 200, k = 4)
## [1] 0.4074057
```

Ejercicios

Ejercicio 1

- Suponga que la probabilidad de tener una unidad defectuosa en una línea de ensamble es de 0.05. Obtener:
  1. La probabilidad de que en 20 unidades seleccionadas al azar, dos sean defectuosas.
  2. La probabilidad de que máximo dos de las 20 unidades estén defectuosos.
  3. La probabilidad de que mínimo dos unidades de las 20, estén defectuosas.

1. $P(X = 2)$: 0.1886768
2. $P(X \leq 2)$: 0.9245163
3. $P(X \geq 2)$: 0.2641605

Ejercicio 2

Un examen de opción múltiple contiene 10 preguntas. Cada pregunta tiene cuatro opciones, de las cuales sólo una es la correcta. El examen se aprueba si se responden correctamente mínimo seis preguntas. Si el estudiante "adivina" las preguntas, obtener:

a. La probabilidad de aprobar el examen (0.0197277)
b. La probabilidad de responder correctamente a 6 preguntas (0.016222)
b. Halle la esperanza matemática (2.5)
c. Halle la varianza (1.875)

Ejercicio 3

Suponga que el número de imperfecciones en un alambre delgado de cobre sigue una distribución Poisson con una media de 2.3 (λ) imperfecciones por milímetro. Obtener:

a. La probabilidad de 2 imperfecciones en un milímetro de alambre. (0.2651846)

b. la probabilidad de 10 imperfecciones en 5 milímetros de alambre. (0.1129351)

c. La probabilidad de al menos una imperfección en 2mm de alambre. (0.9899482)