Distribución Binomial

Se puede verificar que la suma, sobre todos los valores de X, de $f(x)$ es uno haciendo uso del **Teorema del Binomio de Newton**:

$$(a+b)^n = \sum_{x=0}^{n} \binom{n}{x} a^{n-x} b^x,$$

en dónde si se reemplaza a por q y b por p se obtiene

$$(q+p)^n = \sum_{x=0}^{n} \binom{n}{x} q^{n-x} p^x = \sum_{x=0}^{n} f(x),$$

y como $q+p=(1-p)+p=1$, se concluye que

$$\sum_{x=0}^{n} f(x) = (1)^n = 1.$$

Distribución Binomial

Su **Función Generatriz de Momentos** es:

$$M_X(t) = \mathrm{E}\left(e^{tX}\right) = \sum_{x=0}^{n} e^{tx} \cdot f(x) = \sum_{x=0}^{n} e^{tx} \cdot \binom{n}{x} p^x (1-p)^{n-x}$$

$$= \sum_{x=0}^{n} \binom{n}{x} (e^t)^x p^x (1-p)^{n-x} = \sum_{x=0}^{n} \binom{n}{x} (e^t p)^x (1-p)^{n-x}.$$

De acuerdo al binomio de Newton con $a = q, \; b = e^t p$:

$$M_X(t) = \mathrm{E}\left(e^{tX}\right) = (q + pe^t)^n.$$

Distribución Binomial

Media y Varianza

Utilizaremos la generatriz de momentos. Sabemos que $\mathrm{E}(X) = M_X'(0)$ y $\mathrm{E}(X^2) = M_X''(0)$, entonces

$$\begin{aligned} M_X'(t) = &\ n(q + pe^t)^{n-1}(pe^t), \\ M_X'(0) = &\ n(q + pe^0)^{n-1}(pe^0). \end{aligned}$$

Puesto que $e^0 = 1$ y $p + q = 1$, tenemos:

$$M_X'(0) = \ n(1)^{n-1}p = np,$$

y, por lo tanto,

$$\mathrm{E}(X) = np.$$

Distribución Binomial

Ahora,

$$
\begin{aligned}
M_X''(t) &= n(q+pe^t)^{n-1}pe^t + pe^t\left[n(n-1)(q+pe^t)^{n-2}pe^t\right],\\
\mathrm{E}(X^2) &= M_X''(0) = np + n(n-1)p^2.
\end{aligned}
$$

entonces

$$
\begin{aligned}
\mathrm{V}(X) &= \mathrm{E}(X^2) - [\mathrm{E}(X)]^2 = np + n(n-1)p^2 - (np)^2\\
&= np + n^2p^2 - np^2 - n^2p^2 = np - np^2 = np(1-p)\\
&= npq.
\end{aligned}
$$

Distribución Binomial

La variable aleatoria Binomial como una suma de variables aleatorias Bernoulli

Hemos mencionado que cada X_i es una v.a. Bernoulli y que estas son independientes. Considera la variable aleatoria Y definida como la suma de n variables aleatorias Bernoulli(p), esto es, como la suma de ceros y unos producidos por los posibles resultados de las variables Bernoulli

$$Y = \sum_{i=1}^{n} X_i, \qquad \text{con } X_i \sim \text{Bernoulli}(p) \text{ e independientes entre sí.}$$

Podemos imaginarnos lo anterior como realizando n repeticiones de un experimento Bernoulli en el que cada resultado será independiente de los otros, un **muestreo aleatorio**.

Distribución Binomial

Recordando las propiedades del valor esperado y el valor que toma la esperanza de una v.a. *Bernoulli*(p) , tenemos que el valor esperado de Y es

$$\mathrm{E}(Y) = \mathrm{E}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{E}(X_i) = \sum_{i=1}^{n} p = np$$

y su varianza

$$\mathrm{V}(Y) = \mathrm{V}\left(\sum_{i=1}^{n} X_i\right).$$

Veremos posteriormente que la varianza de una suma de variables aleatorias **independientes** es la suma de las varianzas de cada v.a. individual (multiplicada por el cuadrado de su coeficiente respectivo), entonces

$$\mathrm{V}(Y) = \sum_{i=1}^{n} \mathrm{V}(X_i) = \sum_{i=1}^{n} pq = npq.$$

Distribución Binomial

Vemos que estos resultados coinciden con los de la v.a. Binomial(n, p) pero **no** es suficiente para afirmar que $Y \sim$ Binomial(n, p). Si su generatriz es la misma que la de una v.a. binomial podremos afirmar que Y es binomial:

$$
\begin{aligned}
M_Y(t) &= \mathrm{E}(e^{tY}) = \mathrm{E}\left(e^{t\sum_{i=1}^{n} X_i}\right) = \mathrm{E}\left(e^{t(X_1+X_2+\cdots+X_n)}\right) \\
&= \mathrm{E}\left(e^{tX_1} e^{tX_2} \ldots e^{tX_n}\right),
\end{aligned}
$$

de nuevo, como las variables aleatorias X_i son independientes, el último término se puede escribir como

$$
\begin{aligned}
M_Y(t) &= \mathrm{E}\left(e^{tX_1}\right) \mathrm{E}\left(e^{tX_2}\right) \cdots \mathrm{E}\left(e^{tX_n}\right) \\
&= (q + pe^t)(q + pe^t) \cdots (q + pe^t) = (q + pe^t)^n,
\end{aligned}
$$

que es idéntica a la generatriz de momentos de una binomial.

Distribución Binomial

Por lo tanto nuestra v.a. Y, producto de un muestreo determinado, es una v.a. Binomial(n, p).

¿Qué se obtiene si en la sumatoria que define Y usamos $n = 1$?

Es muy importante recalcar el hecho de que cada posible resultado del experimento es una v.a. con las mismas propiedades: misma distribución, mismo parámetro, y que el resultado de cada una de ellas será independiente del resultado de las otras.

Estas condiciones se resumen diciendo que se realizó un **Muestreo Aleatorio**.

Distribución Hipergeométrica

Si cambiamos la forma en que se realiza el muestreo y mantenemos como nuestra v.a. a

$$X = \# \text{ de "éxitos" en } n \text{ observaciones,}$$

nuestro modelo tiene que ser adaptado a las nuevas condiciones.

Supongamos que podemos hacer:

- n observaciones de un conjunto total de N posibles,
- la probabilidad de éxito en cada observación cambia "paso a paso".

Ahora estamos asumiendo que nuestra población es finita (N) y no hay independencia entre observaciones, esto es, el resultado de cada observación será afectado por los resultados anteriores.

Distribución Hipergeométrica

Como **ejemplo**, considera un lote de 25 "botes" de leche comercial de 1 litro, cada uno susceptible de ser clasificado como en mal estado (éxito = agria) o en buen estado. Se toma una muestra de tamaño n, se considera la posibilidad de que se tengan k botes con leche agria y se hace la revisión de cada uno de los botes. Aquí $X = \#$ de botes con leche en mal estado.

- Nuestra población inicial son los 25 "botes",
- La probabilidad de que un bote seleccionado contenga leche en mal estado cambiará conforme vamos haciendo el muestreo debido a que en el lote habrá inicialmente una cierta proporción de botes en mal estado (botes de leche en mal estado / N total de botes en el lote) y al ir sacando los botes esta proporción se verá afectada por los resultados obtenidos con anterioridad. Esto es, se realiza un <u>muestreo sin reemplazo</u>.

Distribución Hipergeométrica

Este tipo de experimento es muy usado cuando en la revisión se tiene que destruir o alterar el objeto, en nuestro caso hay que abrir el bote de leche para revisarlo.

Considera lo siguiente: si $k = 4$, la proporción inicial es de $\frac{4}{25}$, pero si sacamos un bote la proporción cambia a $\frac{3}{24}$ ó $\frac{4}{24}$, dependiendo de si el que se sacó contenía leche en mal estado o no, respectivamente.

Distribución Hipergeométrica

Para poder establecer el conjunto de posibles valores de X, necesitamos conocer, en principio, la proporción exacta de botes en mal estado, esto es, cuántos botes hay en mal estado en el lote de 25, y cuántas observaciones se hacen (n):

- Si hay $k = 3$ botes con leche agria y se toman $n = 5$ botes para revisarlos, X podrá tomar los valores $\{0, 1, 2, 3\}$, i.e. $0 \leq x \leq k$; pero si se revisan $n = 2$, X podrá tomar los valores $\{0, 1, 2\}$, i.e. $0 \leq x \leq n$.
- Algo similar ocurriría si el lote fuera de $N = 10$ botes, si hubiesen $k = 4$ botes con leche agria y si se revisaran $n = 7$: puesto que se están revisando 7 y sólo hay 6 en buen estado, X podría ser $\{1, 2, 3, 4\}$, i.e. $n - (N - k) \leq x \leq k$. ¿Qué valores toma X si $k = 5$ y $n = 7$?

Observa que x **llega hasta el menor de los números** n **y** k**, y que el menor valor de** x **es 0 ó** $n - (N - k)$. Estos casos muestran que los valores de la v.a. dependerán de los valores de k, n y N.

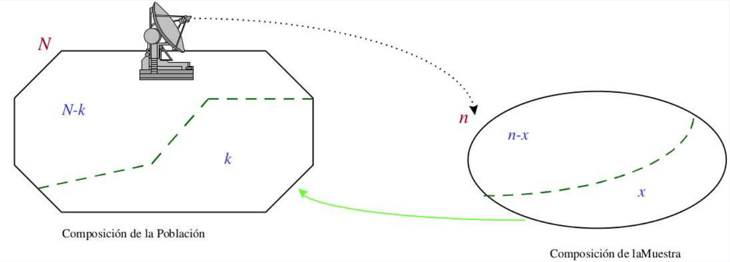
Distribución Hipergeométrica

Las condiciones de muestreo mencionadas suelen aplicarse en la práctica en problema de muestreo de **aceptación en control de calidad**, así como en la **estimación del tamaño de una población finita** N en **procesos de captura-recaptura**. (Ver [4] ejemplo 3.2.10 pág. 86, ejemplo 3.2.9 págs. 84-85. También te recomendamos (deberías leer) el capítulo 4, pág. 42 de [5], para muestreo de aceptación.)

No es de sorprender este tipo de aplicaciones ya que, como dijimos, al tomar n observaciones en las cuales habrá x "éxitos", conoceremos la proporción $\frac{x}{n}$, la cual nos dará información de la verdadera proporción $\frac{k}{n}$.

Distribución Hipergeométrica

Esquemáticamente podemos imaginarnos los parámetros de la distribución hipergeométrica como en la siguiente figura



Población con N objetos		Muestra de n objetos tomados al azar
k elementos con cierta característica de interés	$\leftarrow$ Inferencia	x elementos con la característica de interés
$N-k$ sin la característica	$\leftarrow$ Inferencia	$n-x$ sin la característica de interés

Distribución Hipergeométrica

La ley de comportamiento asociada con nuestra v.a. se obtiene utilizando el enfoque de frecuencia relativa en el que contamos el número de casos favorables y dividimos entre el número de casos totales:

- Nos interesan x "éxitos" de un total de k posibles, y esto lo podemos hacer de $\binom{k}{x}$ formas; pero por cada uno de estos resultados, como se seleccionaron n, obtendríamos $n - x$ "fracasos" de un total de $N - k$ "fracasos" posibles, y esto lo podemos hacer de $\binom{N-k}{n-x}$ formas. Entonces, usando la regla multiplicativa, el número total de **casos favorables** es $\binom{k}{x}\binom{N-k}{n-x}$.
- Al seleccionar n objetos de un total de N, podemos hacerlo de $\binom{N}{n}$ formas.

Distribución Hipergeométrica

Por lo tanto, la probabilidad de observar x "éxitos" en n pruebas está dada por la función de probabilidad (Distribución Hipergeométrica):

$$f(x) = \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}, \qquad \max\{0, n-(N-k)\} \leq x \leq \min\{n, k\}.$$

Sus parámetros son n, k y N. Notación $X \sim \text{Hiper}(N, k, n)$.

Distribución Hipergeométrica

En las siguientes gráficas se muestra la funciones de masa y de distribución acumulada de la distribución hipergeométrica para varios valores de los parámetros.

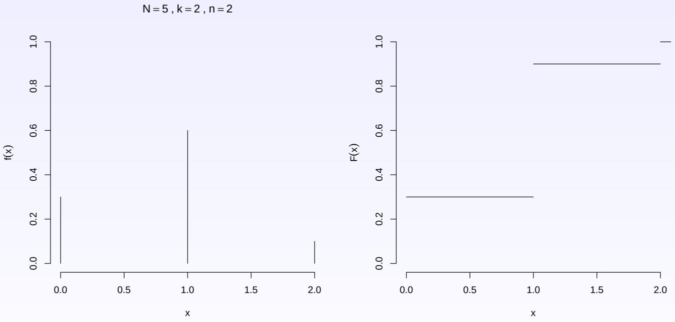


Figura: Funciones de masa y de distribución acumulada para $X \sim$Hiper$(5, 2, 2)$.

Distribución Hipergeométrica

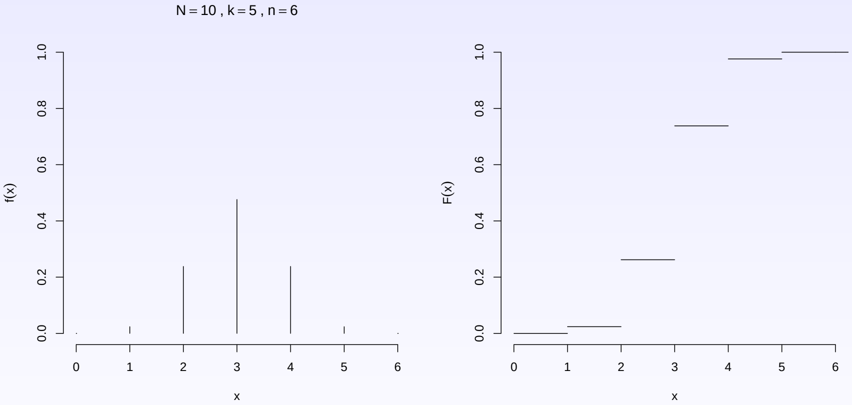


Figura: Funciones de masa y de distribución acumulada para $X \sim$Hiper$(10, 5, 6)$.

Distribución Hipergeométrica

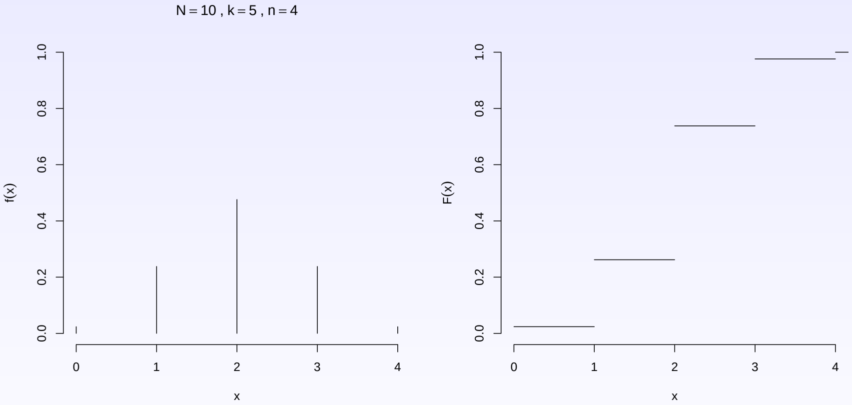


Figura: Funciones de masa y de distribución acumulada para $X \sim$Hiper$(10, 5, 4)$.

Distribución Hipergeométrica

Media y Varianza

La media y la varianza de una v.a. hipergeométrica están dados por:

$$\mathrm{E}(X) = \frac{nk}{N}, \qquad \mathrm{V}(X) = \frac{nk}{N}\left(1 - \frac{k}{N}\right)\left(\frac{N-n}{N-1}\right).$$

Debido a la especial complejidad y a su poco uso práctico, no se muestra la función generatriz de momentos de la hipergeométrica. (Ver referencias adicionales en Clase 4).

Distribución Hipergeométrica

Aproximación de la distribución binomial a la distribución hipergeométrica

Existe una relación entre la distribución binomial y la hipergeométrica: bajo ciertas condiciones, una puede aproximar los valores de probabilidad de la otra.

Observa que, a partir del enfoque de probabilidad como frecuencia relativa, la razón $\frac{k}{N}$ representa la proporción de éxitos p, mientras que $1 - \frac{k}{N}$ la proporción de fracasos q, de lo cual las anteriores fórmulas pueden ser escritas como

$$\mathrm{E}(X) = np, \qquad \mathrm{V}(X) = npq\left(\frac{N-n}{N-1}\right).$$

Distribución Hipergeométrica

Se puede apreciar una cierta similitud con la media y la varianza de la distribución binomial. Si N fuese muy grande, y el tamaño n de la muestra pequeño, tendríamos que

- el factor $\frac{N-n}{N-1}$ tendería a 1;
- p equivaldría a la probabilidad de "éxito" como en la distribución binomial, pudiéndose considerar como **"cuasi-fija"**, ya que no cambiaría mucho al ir realizando las observaciones;
- q equivaldría a la probabilidad de "fracaso", con las mismas consideraciones que para p; e,
- igualmente, podríamos considerar una **"cuasi-independencia"**, suponiendo que no tendríamos "rachas sistemáticas de resultados", esto es, el resultado de una observación no nos daría información del resultado de la siguiente observación.

Distribución Hipergeométrica

Con las condiciones anteriores nuestro modelo sería aproximadamente binomial, y podemos usar esta distribución como una aproximación al modelo hipergeométrico.

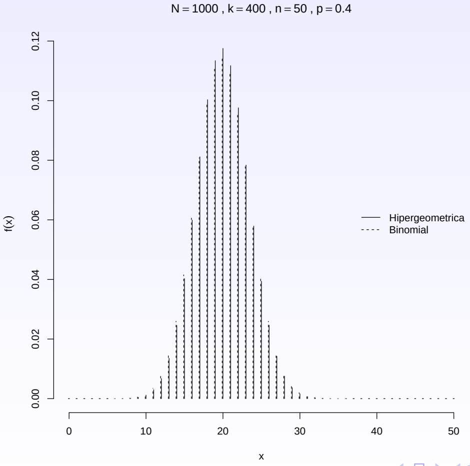
Es posible demostrar formalmente que si $p = k/N$ se mantiene constante entonces

$$\lim_{N\to\infty} \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}} = \binom{n}{x} p^x (1-p)^{n-x}.$$

La pregunta es: ¿qué tan grande deberá ser N comparada con n? Se ha encontrado de la experiencia que cuando la proporción $\frac{n}{N}$ es del orden del 10 %, se tiene una buena aproximación, mejorando cuando $\frac{n}{N}$ disminuye.

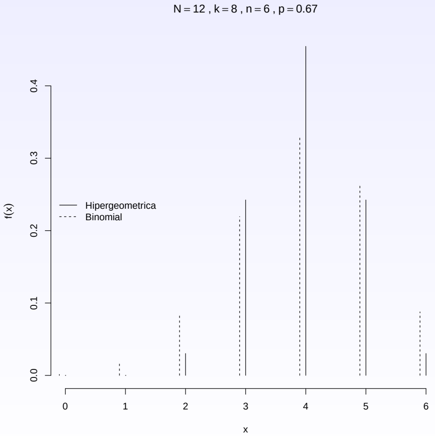
Distribución Hipergeométrica

Por ejemplo, tenemos que la aproximación binomial para Hiper$(1000, 400, 50)$ es excelente.



Distribución Hipergeométrica

En este segundo ejemplo la relación entre N y n es del 50 %, por lo que no se espera una buena aproximación entre las probabilidades asociadas bajo cada distribución.



Distribución Hipergeométrica

El factor $\frac{N-n}{N-1}$, cantidad que nos recuerda la finitud de N, es llamado el **factor de corrección para población finita**. ¿Cuál es $\lim_{N\to\infty} \frac{N-n}{N-1}$, con n fija?

Nota: Entre los materiales del curso se encuentra un texto sobre métodos de captura y recaptura asociados a la distribución hipergeométrica y sobre como calcular los momentos de tal distribución. Es importante leerlos.

Distribución Geométrica y Binomial Negativa

En la mayoría de los procesos industriales existen procedimientos de "monitoreo" a través de gráficos de control. Para ello pueden emplearse diversos criterios, por ejemplo

- detener la producción (o ajustar el proceso) hasta el momento en que se detecta el primer artículo defectuoso (al obtener una "falsa alarma"),
- o bien, hasta que se detectan un cierto número r con $(r > 1)$, de artículos defectuosos.

Analizaremos el modelo correspondiente a un "éxito", y posteriormente lo generalizaremos a r "éxitos", para cualquier valor de r.

Distribución Geométrica

Aunque usamos el ejemplo de la línea de producción para introducir las v.a.'s de esta sección, también las definiremos en términos de la cantidad de experimentos Bernoulli realizados hasta obtener un determinado número de "éxitos".

Uno de los criterios arriba mencionados para mantener bajo control las líneas de producción es detener el proceso (o ajustarlo) hasta obtener un artículo defectuoso. Este criterio puede parecer un tanto estricto, pero si se sabe que el obtener un artículo defectuoso es un evento muy raro, esto es, hay una probabilidad pequeña de que ocurra; o bien, que la obtención de un artículo defectuoso es una situación en extremo crítica o costosa; es factible poner en marcha dicha política.

Distribución Geométrica

La variable aleatoria de interés en este caso es

$$X = \# \text{ de observaciones hasta obtener } \underline{\text{un}} \text{ éxito}.$$

Observa que en ningún momento se está fijando el número de observaciones, como en la binomial y la hipergeométrica, sino que más bien el número de observaciones es la variable aleatoria de interés y lo que es fijo en este caso es el número de éxitos.

La v.a. geométrica se define como el número de observaciones de variables aleatorias Bernoulli(p) independientes
<u>hasta obtener un éxito</u>.

Distribución Geométrica

Nuestros supuestos son:

1. Cada vez que se realice una observación la probabilidad de que ocurra un "éxito" es fija e igual a p.
2. Las observaciones son independientes.
3. El número de observaciones puede continuar indefinidamente antes de encontrar el primer defectuoso.

Distribución Geométrica

El supuesto 1 equivale a decir que estamos pensando en hacer observaciones sobre el mismo proceso y, por lo tanto, todos sus elementos deben tener en principio las mismas características.

Por ejemplo, en la línea de producción, el proceso "no sabe" o "no se acuerda" de lo que ha pasado antes de la observación que se esté realizando, ya que los productos defectuosos ocurren por mero azar (una variación inherente e inevitable), no por alguna falla sistemática. Cuando uno detecta un artículo defectuoso, sabiendo que el modelo asigna una cierta probabilidad a ese evento, se toma una decisión respecto al proceso. Ver [15].

El supuesto 2 equivale a decir que el que una observación resulte o no en "éxito", **no** da información sobre el resultado de la siguiente observación.

Distribución Geométrica

Bajo las condiciones y supuestos anteriores podemos deducir inmediatamente el mecanismo para asignar probabilidades, donde llamamos p a la probabilidad de "éxito" y por ende $(1 - p = q)$:

$$f(x) = q^{x-1}p, \quad x = 1, 2, 3, \ldots . \qquad \text{(Distribución Geométrica)}$$

Notación: $X \sim \text{Geom}(p)$. El único parámetro es p, la probabilidad de éxito.

A continuación se presentan algunas gráficas de funciones de distribución acumulada y de densidad para $\text{Geom}(p)$ y distintos valores de p.

Distribución Geométrica

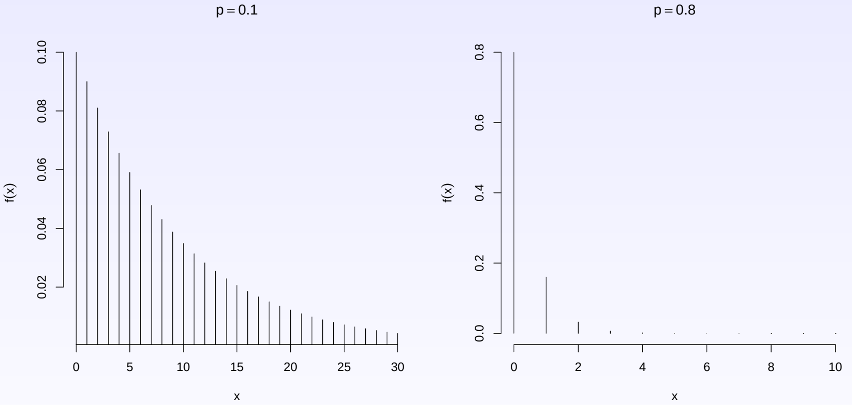


Figura: Funciones de masa para Geom(0.1) y Geom(0.8).

Distribución Geométrica

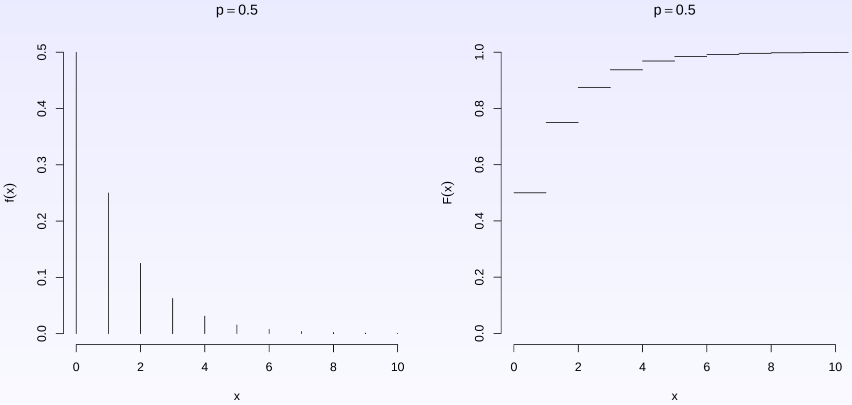


Figura: Funciones de masa y de distribución acumulada para Geom(0.5).

Distribución Geométrica

Esta distribución regularmente se utiliza en la modelación de:

1. Fenómenos con eventos raros, i.e. aquéllos en los que p es **pequeña**.
2. tiempos de espera discretos.

Ejemplo del uso 1: cuando se quiere detectar una enfermedad rara, en lugar de muestrear mediante una campaña masiva toda la población, se puede ir muestreando aleatoriamente de uno en uno, y si conforme se realiza el muestreo no se detecta a nadie enfermo, se puede pensar que no hay problema; pero si se encuentra a alguien con la enfermedad, es motivo para hacer un muestreo más exhaustivo. Algo equivalente se considera en los errores de *transmisión en comunicaciones*, o cualquier sistema considerado como muy *confiable*.

Distribución Geométrica

Ejemplo del uso 2: cuando se tiene que inspeccionar un sistema para ver si está en buenas condiciones para seguir operando, las inspecciones se realizan programadamente cada cierto intervalo de tiempo fijo, y aún cuando una falla haya ocurrido en el lapso entre inspecciones se considera, que esta ocurre al momento de realizar la inspección ("tiempo discreto"). Entonces, la v.a. sería *el número de intervalos de tiempo que se tuvo que esperar hasta encontrar un "éxito"*. Nuevamente se espera que el defecto ocurra por la aleatoriedad inherente y no por desgaste.

Distribución Geométrica

Se ha dicho que el proceso "no sabe" o "no se acuerda" de lo que ha pasado antes de la observación que se esté realizando. Cuando un proceso tiene esta característica, se dice que "NO TIENE MEMORIA". Lo mismo se aplica para la v.a. que describe al proceso.

Matemáticamente: si $X \sim \text{Geom}(p)$, calculemos la probabilidad de que un "éxito" ocurra para una X mayor que un cierto número de pruebas, denotado como la suma de dos números $a + b$, dado que se sabe que el éxito ocurre después de las a pruebas (a y b son enteros positivos). Si el resultado es independiente de que el éxito ocurra después de las primeras a pruebas significará que al proceso "no le importa" lo que haya ocurrido antes del valor que se esté observando.

Distribución Geométrica

En símbolos,

$$
\begin{aligned}
P(X > a+b | X > a) &= \frac{P(X > a+b, X > a)}{P(X > a)} \quad \text{pues } P(A|B) = \frac{P(A \cap B)}{P(B)} \\
&= \frac{P(X > a+b)}{P(X > a)} \quad \text{ya que } X > a+b \text{ es la intersección entre los dos intervalos del numerador} \\
&= \frac{1 - \mathrm{P}(X \leq a+b)}{1 - \mathrm{P}(X \leq a)} \\
&= \frac{1 - F(a+b)}{1 - F(a)}.
\end{aligned}
$$

Por otro lado, la función de distribución de una v.a. Geom(p) es

$$F(x) = \sum_{n=1}^{x} q^{n-1} p = p \sum_{m=0}^{x-1} q^m = p \left(\frac{1 - q^x}{1 - q} \right) = 1 - q^x.$$

Distribución Geométrica

De lo anterior,

$$P(X > a+b|X > a) = \frac{1 - F(a+b)}{1 - F(a)} = \frac{q^{a+b}}{q^a} = q^b = 1-F(b) = P(X > b);$$

esto es, X "**no tiene memoria**".

Como ejemplos numéricos

$$\begin{aligned} P(X > 1+2|X > 1) &= P(X > 2), \\ P(X > 1+2|X > 2) &= P(X > 1), \\ P(X > 1+1|X > 1) &= P(X > 1). \end{aligned}$$