

Inferencia Estadística

Dra. Graciela González Farías
Dr. Ulises Márquez



Maestría en Cómputo
Estadístico.

CIMAT Monterrey.



Agradecimientos

En forma de agradecimiento, se enlistan personas que han contribuido de una u otra forma en la construcción de estas notas a través de los años:

- Víctor Muñiz
- Juan Antonio López
- Sigfrido Iglesias González
- Rodrigo Macías Paéz
- Edgar Jiménez
- Todos los estudiantes que han colaborado con sugerencias y comentarios sobre estas notas.

Estas notas son de uso exclusivo para enseñanza y no pretende la sustitución de los textos y artículos involucrados.

Temario

- 1 Variables aleatorias y distribuciones de probabilidad.
 - a) Distribuciones de probabilidad de variables aleatorias discretas.
 - b) Procesos de Poisson.
 - c) Distribuciones de probabilidad de variables aleatorias continuas.
 - d) Métodos gráficos para la identificación de distribuciones.
 - e) Estimación de densidades.
 - f) Distribuciones de probabilidad de vectores aleatorios.
 - g) Esperanzas condicionales y regresión.
 - h) Modelos jerárquicos, compuestos y mezclas de variables aleatorias.
 - i) Transformaciones de variables aleatorias.
 - j) Simulación de variables aleatorias.
 - k) Convergencia de variables aleatorias y el Teorema del Límite Central.

Temario

- ② Distribuciones muestrales y métodos de estimación.
 - a) Propiedades de los estimadores.
 - b) Estimadores no insesgados
 - c) Distribuciones muestrales.
 - d) Principio de máxima verosimilitud.
 - e) Estimación puntual.
 - f) Bootstrap y jackknife.

Temario

- ③ Pruebas de Hipótesis e intervalos de confianza.
 - a) Definición de conceptos.
 - b) Potencia de la prueba.
 - c) Pruebas para dos poblaciones normales independientes.
 - d) Pruebas para medias en muestras pareadas.
 - e) Pruebas básicas de varianzas.
 - f) Pruebas para proporciones.
 - g) Conceptos de estimación bayesiana.
 - h) Temas optativos de modelos para presentaciones finales, por ejemplo:
 - ① Pruebas no-paramétricas clásicas.
 - ② Pruebas de permutaciones.
 - ③ Estimación no paramétrica (suavizadores y splines).
 - ④ Modelos gráficos probabilistas.
 - ⑤ Entre muchos otros.

Evaluación y acreditación

- Dos exámenes parciales, 18 de septiembre y 6 de noviembre: **15 %, cada uno.**
- Evaluación de las tareas (de 2 tipos) y actividades en clase y asistencia: **40 %.**
- Un examen final, consistente en una exposición donde se entrega un reporte y se hace una presentación de 1/2 hora. La presentación debe incluir antecedentes, metodología, un ejemplo práctico y compartir el código. Deberán entregar a los instructores y a sus compañeros el resumen. Adicionalmente, deberán dejar un ejercicio sobre el tema a sus compañeros que calificarán en forma honesta: **30 %.**

Las tareas tienen una frecuencia quincenal e incluyen TODOS los ejercicios dejados en las notas y requerirán en general el uso de recursos computacionales.

Textos

- **Larry Wasserman (2004) . All of Statistics, A concise course in Statistical Inference. Springer.**
- F.M. Dekking, C. Kraaikamp, H.P. Lopuhaa L.E. Meester (2005). A Modern Introduction to Probability and Statistics, Understanding Why and How. Springer text in Statistics.
- John A. Rice (1995). Mathematical Statistics and Data Analysis, Second Edition. Duxbury Press.
- Casella & Berger. (2002). Statistical Inference, Second Edition . Duxbury Press.
- Richard J. Larsen and Morris L. Marx (2011). An Introduction to Mathematical Statistics and its Applications. Fifth Edition. Prentice Hall.

Modelos probabilísticos

Modelos probabilísticos

Recordemos que una variable aleatoria discreta es aquella que sólo toma un número **contable** de valores (finito o infinito). Por ejemplo:

- El número de plantas con daños visibles producidos por una plaga.
- El número de individuos a favor de un partido político.
- El número de televisores con defectos en su selector de canales de un lote de 100 televisores.
- El número de personas en la fila en un centro de servicio al público, entre las 9 y 10 de la mañana.

Notamos que en cada una de esas situaciones uno lleva a cabo algún tipo de **conteo**.

Modelos probabilísticos

En un principio uno debería:

- Examinar cada caso;
- Ver cuáles son las **condiciones específicas** en que se realiza el muestreo;
- Establecer los supuestos de simplicidad que sean factibles; y,
- Determinar el modelo probabilístico que mejor describa el comportamiento de la característica bajo estudio (verificando su validez).

Modelos probabilísticos

Este procedimiento general ha dado lugar a un cierto número de modelos que aparecen frecuentemente en las aplicaciones. Así, lo que haremos aquí, es construir estos **modelos particulares** formando un catálogo básico que nos permita referenciar nuestras situaciones particulares a alguno de estos. En la construcción del catálogo, contemplamos varios puntos:

- 1 **Supuestos** necesarios para identificar el uso del modelo.
- 2 **Construcción** del modelo: función de probabilidad y de probabilidad acumulada.
- 3 **Momentos**: Media, Varianza, Función generatriz de momentos (cuando aplique).

Modelos de variables discretas

Distribución uniforme discreta

Cuando se tiene un número **finito** de resultados de un experimento, x_1, x_2, \dots, x_n , cada uno de ellos **igualmente probable**, se dice que se tiene una variable aleatoria que puede ser modelada por una distribución uniforme discreta. Notación:

$$X \sim U(x_1, x_2, \dots, x_n).$$

Se lee: X se distribuye como una variable aleatoria uniforme.

Este tipo de modelo aparece típicamente en la selección de **muestras al azar**.

El único parámetro de la distribución es n , el número posible de resultados.

Distribución uniforme discreta

Como la suma de todas las probabilidades debe ser uno y todos los valores son equiprobables, entonces la función de masa está dada por

$$f(x) = \begin{cases} 1/n & \text{para } x = x_1, x_2, \dots, x_n \\ 0 & \text{en otro caso} \end{cases}$$

Bajo esta definición, claramente $f(x) \geq 0$ para toda x , y

$$\sum_{i=1}^n f(x_i) = n \left(\frac{1}{n} \right) = 1.$$

El caso más común de esta distribución es cuando la variable toma los valores enteros

$$\{1, 2, \dots, n\}.$$

El resto de los resultados los estableceremos para cuando la variable aleatoria uniforme toma estos valores.

Distribución uniforme discreta

El parámetro (valor que identifica unívocamente al modelo) de la distribución es n , el número total de objetos. Se dice en este caso que se trata de un espacio de **probabilidad equiprobable**.

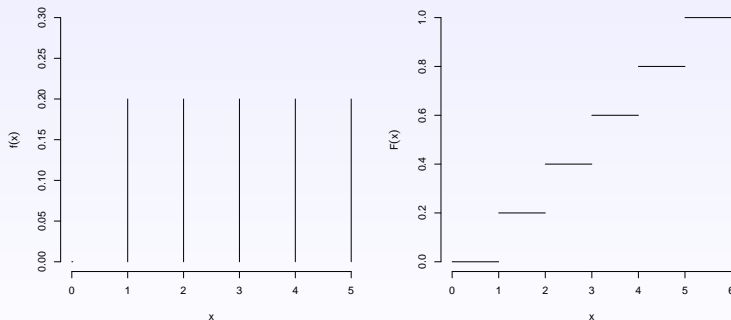


Figura: Distribución uniforme para $n = 5$.

Distribución uniforme discreta

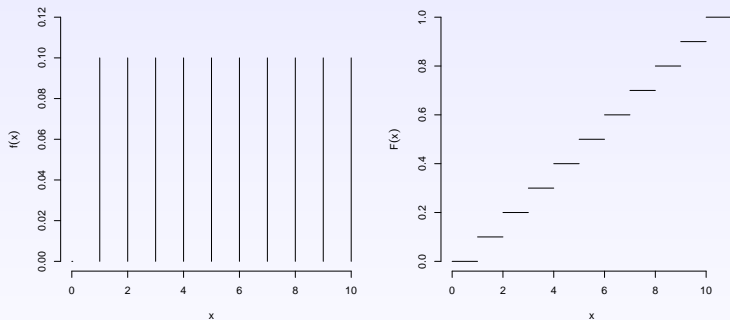


Figura: Distribución uniforme $n = 10$.

Distribución uniforme discreta

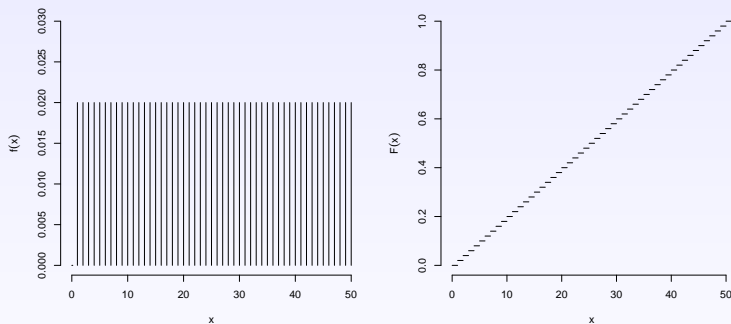


Figura: Distribución uniforme $n = 50$.

Distribución uniforme discreta

Media y Varianza:

$$E(X) = \sum_{x=1}^n x \frac{1}{n} = \frac{1 + 2 + \cdots + n}{n} = \frac{n+1}{2},$$

y

$$\begin{aligned} V(X) &= E(X^2) - [E(X)]^2 = \sum_{x=1}^n x^2 \frac{1}{n} - \left(\frac{n+1}{2} \right)^2 \\ &= \frac{(n+1)(2n+1)}{6} - \left(\frac{n+1}{2} \right)^2 = \frac{2n^2 + 3n + 1}{6} - \frac{n^2 + 2n + 1}{4} \\ &= \frac{2(2n^2 + 3n + 1) - 3(n^2 + 2n + 1)}{12} = \frac{4n^2 + 6n + 2 - 3n^2 - 6n - 3}{12} \\ &= \frac{n^2 - 1}{12}. \end{aligned}$$

Distribución uniforme discreta

Función Generatriz de Momentos:

$$\begin{aligned}
 M_X(t) &= E\left(e^{tX}\right) = \sum_{x=1}^n e^{tx} \cdot f(x) = \sum_{x=1}^n e^{tx} \cdot \frac{1}{n} \\
 &= \frac{1}{n} \sum_{x=1}^n (e^t)^x = \frac{1}{n} \left[\frac{e^t - (e^t)^{n+1}}{1 - e^t} \right] = \frac{1}{n} \cdot \frac{e^t(1 - e^{nt})}{1 - e^t}, \quad \forall t.
 \end{aligned}$$

Nota: $M_X(0) = 0$, aplicando la regla de L'Hopital.

Se puede consultar más información sobre la función característica en:

[en.wikipedia.org/wiki/Characteristic_function_\(probability_theory\)](https://en.wikipedia.org/wiki/Characteristic_function_(probability_theory))

Distribución Bernoulli

Definamos un experimento en el cual hay únicamente **dos posibles resultados**: a uno de ellos le llamamos “**éxito**”, al otro “**fracaso**”. Este tipo de variables aparece frecuentemente en nuestros conteos, por ejemplo, si pensamos en clasificar nuestros productos como: defectuoso, no defectuoso; grande o pequeño; azul o blanco; sí o no, etc.

Generalmente le asignamos un valor de **1** al “**éxito**” y un valor de **0** al “**fracaso**”. Notemos que la asignación de los valores numéricos es arbitraria. Este procedimiento sirve de base para la construcción de otras distribuciones de gran utilidad.

Distribución Bernoulli

Definamos $X = \#$ de “éxitos”, entonces

$$x = \begin{cases} 1 & \text{si } X = \text{“éxito”}, \text{ esto con probabilidad } p \\ 0 & \text{si } X = \text{“fracaso”}, \text{ esto con probabilidad } (1 - p) = q \end{cases}$$

o bien,

$$f(x) = p^x(1 - p)^{1-x}, \quad \text{para } x = 0, 1; \quad (\text{Distribución Bernoulli}).$$

Notación: $X \sim \text{Bernoulli}(p)$. El único parámetro de la distribución de esta variable aleatoria es p , la probabilidad de éxito.

Distribución Bernoulli

Por ejemplo:

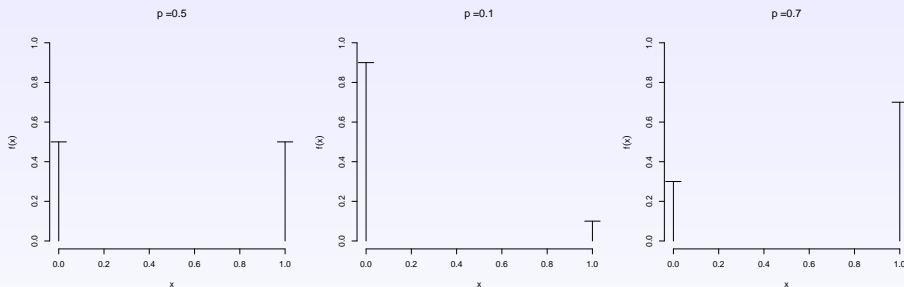


Figura: Función de densidad para $X \sim \text{Bernoulli}(p)$.

El experimento que guió al modelo, dos posibles resultados con probabilidades p y q ($p + q = 1$), se denomina **experimento Bernoulli**.

Distribución Bernoulli

Media y Varianza:

$$E(X) = \sum_{x=0}^1 x \cdot f(x) = 0 \cdot (1 - p) + 1 \cdot p = p$$

$$E(X^2) = \sum_{x=0}^1 x^2 \cdot f(x) = 0^2 \cdot (1 - p) + 1^2 \cdot p = p$$

$$V(X) = E(X^2) + [E(X)]^2 = p - p^2 = p(1 - p) = pq$$

Función Generatriz de Momentos:

$$\begin{aligned} M_X(t) &= E(e^{tX}) = \sum_{x=0}^1 e^{tx} \cdot f(x) = e^{t \cdot 0}(1 - p) + e^{t \cdot 1}p \\ &= (1 - p) + pe^t = q + pe^t, \quad \forall t. \end{aligned}$$

Distribución Binomial e Hipergeométrica

Para las siguientes distribuciones definiremos nuestra variable aleatoria de la siguiente manera:

$X = \#$ de “éxitos” en n observaciones.

X es el número de unos o “éxitos” que se presenten en la muestra de tamaño n . Lo que asigna un patrón distinto de comportamiento, es la **forma como se realizan (condiciones) las n observaciones.**

Distribución Binomial

Supongamos que podemos hacer:

- 1 n observaciones independientes.
- 2 la probabilidad de éxito en cada observación permanece constante, esto es, siempre es p .

En otras palabras, estamos asumiendo que nuestra población es suficientemente grande como para tomarla como “infinita” y que podemos garantizar la independencia entre observaciones, esto es, no obtenemos información adicional para predecir el siguiente resultado sólo porque ya observamos al (los) anterior (es). Esto refleja un comportamiento de procesos que lo podríamos denominar estable. Estas condiciones deben estar presentes al menos durante el período en que se realiza el estudio.

Distribución Binomial

Por ejemplo, pensemos en un proceso industrial, producción de mangueras para gas. Las clasificaremos como defectuosas (éxito) o no defectuosas, de acuerdo a si cumplen o no con el tamaño requerido. Se toma una muestra de tamaño tres y se hacen las mediciones de cada una de las mangueras.

- Asumimos que nuestra población son todas las mangueras que pasan por ese proceso ($\#$ muy grande),
- Se asume que no hay ninguna causa que motive desperfectos sistemáticos,
- También notemos que, bajo estas condiciones, la probabilidad de que una manguera no sea de las medidas requeridas deberá permanecer constante para las 3 observaciones y, más aún, esta probabilidad está dada por otro mecanismo independiente al conteo que nos atañe en este momento.

Distribución Binomial

Por ejemplo, p se podría determinar como una **proporción** observada a través del tiempo, o bien porque conozcamos la **ley** de fallas en cortes, de la maquinaria que se emplea. Para nuestros propósitos, p es un valor dado.

Notemos primero que el número posible de valores que puede tomar X es $\{1, 2, \dots, n\}$. Ahora obtengamos la ley de comportamiento asociada con esta variable:

$$\begin{aligned} P(X = 0) &= P(\text{ninguna manguera defectuosa en la muestra de } n = 3) \\ &= P(X_1 = 0, X_2 = 0, X_3 = 0) \end{aligned}$$

donde X_i denota la i -ésima observación. Cabe notar que cada X_i toma el valor de 0 ó 1 con probabilidades constantes q y p , respectivamente. Esto es, cada X_i **representa una variable Bernoulli(p) y son independientes entre sí.**

Distribución Binomial

Entonces:

$$\begin{aligned} P(X_1 = 0, X_2 = 0, X_3 = 0) &= P(X_1 = 0)P(X_2 = 0)P(X_3 = 0) \\ &= qqq = q^3 = \binom{3}{0} q^3 \end{aligned}$$

A manera de ejercicio, puedes verificar que:

$$\begin{aligned} P(X = 1) &= P(\{X_1 = 1, X_2 = 0, X_3 = 0\} \cup \{X_1 = 0, X_2 = 1, X_3 = 0\} \cup \{X_1 = 0, X_2 = 0, X_3 = 1\}) \\ &= 3pq^2 = 3pq^2 = \binom{3}{1} p^1 q^2, \quad \mathbf{1 \text{ "éxito"}}; \end{aligned}$$

$$\begin{aligned} P(X = 2) &= P(\{X_1 = 1, X_2 = 1, X_3 = 0\} \cup \{X_1 = 1, X_2 = 0, X_3 = 1\} \cup \{X_1 = 0, X_2 = 1, X_3 = 1\}) \\ &= 3p^2q = 3p^2q = \binom{3}{2} p^2 q, \quad \mathbf{2 \text{ "éxitos"}}; \end{aligned}$$

$$\begin{aligned} P(X = 3) &= P(\{X_1 = 1, X_2 = 1, X_3 = 1\}) \\ &= ppp = 3p^3 = \binom{3}{3} p^3, \quad \mathbf{3 \text{ "éxitos"}}. \end{aligned}$$

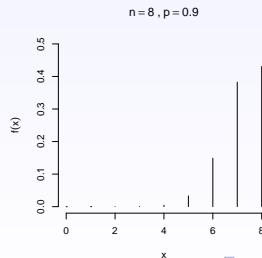
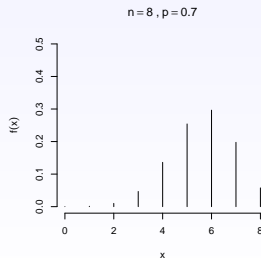
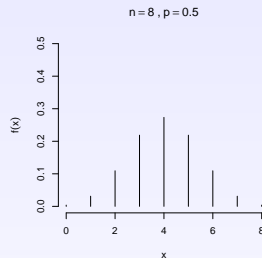
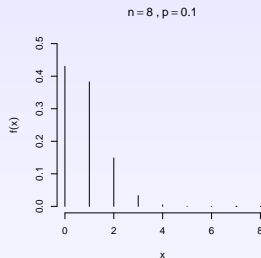
Distribución Binomial

Esto es, tenemos la **Distribución Binomial**

$$f(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & \text{para } x = 0, 1, \dots, n \\ 0 & \text{en otro caso} \end{cases}$$

En este caso se dice que X sigue una **distribución binomial con parámetros (n, p)** . Notación: $X \sim B(n, p)$.

Distribución Binomial



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:

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

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

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

Distribución Binomial

Media y Varianza

Utilizaremos la generatriz de momentos. Sabemos que $E(X) = M'_X(0)$ y $E(X^2) = M''_X(0)$, entonces

$$M'_X(t) = n(q + pe^t)^{n-1}(pe^t),$$

$$M'_X(0) = n(q + pe^0)^{n-1}(pe^0).$$

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

$$M'_X(0) = n(1)^{n-1}p = np,$$

y, por lo tanto,

$$E(X) = np.$$

Distribución Binomial

Ahora,

$$M_X''(t) = n(q + pe^t)^{n-1}pe^t + pe^t [n(n-1)(q + pe^t)^{n-2}pe^t],$$
$$E(X^2) = M_X''(0) = np + n(n-1)p^2.$$

entonces

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

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

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

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

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

y su varianza

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

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

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

Distribución Binomial

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

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

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

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

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. $\text{Binomial}(n, p)$.

¿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

$X = \#$ de “éxitos” en n 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 $X = \# \text{ de botes con leche en mal estado}$.

- 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 muestreo sin reemplazo.

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 .

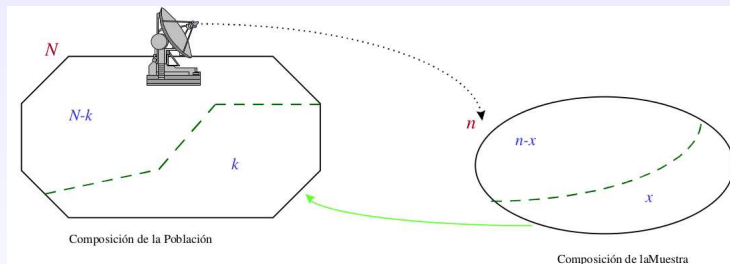
Distribución Hipergeométrica

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	← Inferencia	x elementos con la característica de interés
$N - k$ sin la característica	← 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](#)):

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

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.

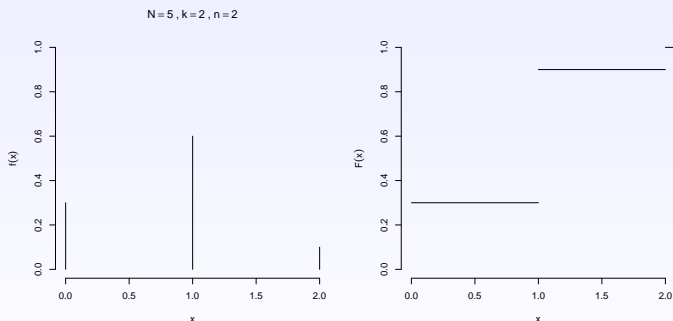


Figura: Funciones de masa y de distribución acumulada para $X \sim \text{Hiper}(5, 2, 2)$.

Distribución Hipergeométrica

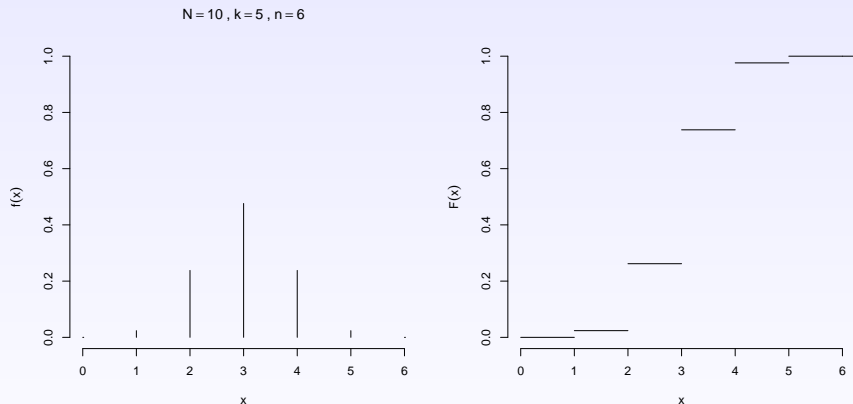


Figura: Funciones de masa y de distribución acumulada para $X \sim \text{Hiper}(10, 5, 6)$.

Distribución Hipergeométrica

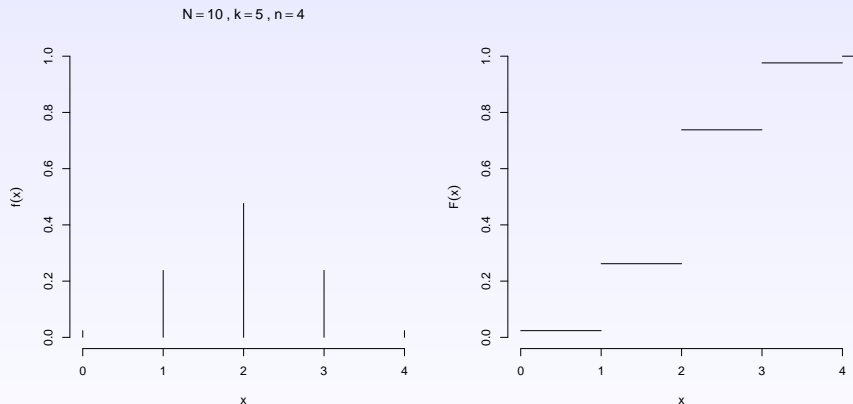


Figura: Funciones de masa y de distribución acumulada para $X \sim \text{Hiper}(10, 5, 4)$.

Distribución Hipergeométrica

Media y Varianza

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

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

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 2).

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

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

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.

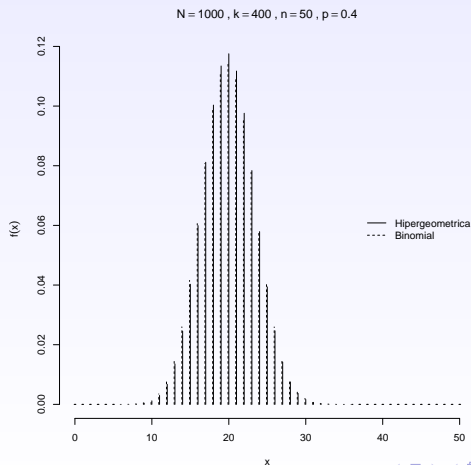
Es posible demostrar formalmente que si $p = k/N$ se mantiene constante entonces

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

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.

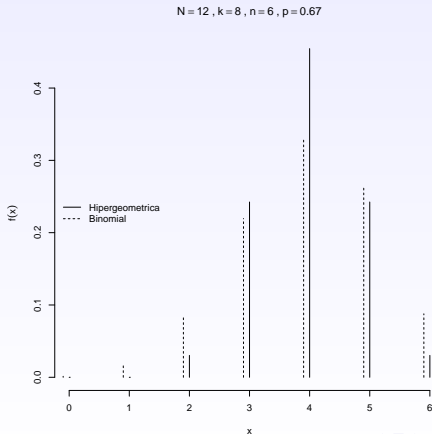
Distribución Hipergeométrica

Por ejemplo, tenemos que la aproximación binomial para $\text{Hiper}(1000, 400, 50)$ 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 $\lim_{N \rightarrow \infty} \frac{N-n}{N-1}$, 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

$X = \#$ de observaciones **hasta obtener un 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 hasta obtener un éxito.

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 $(1 - p = q)$:

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

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

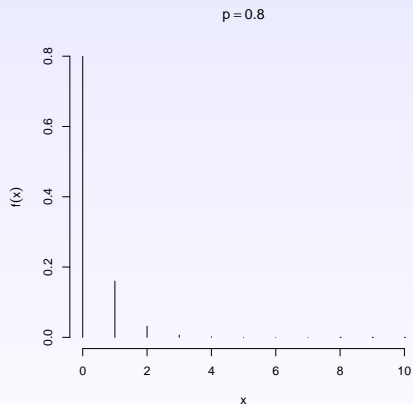
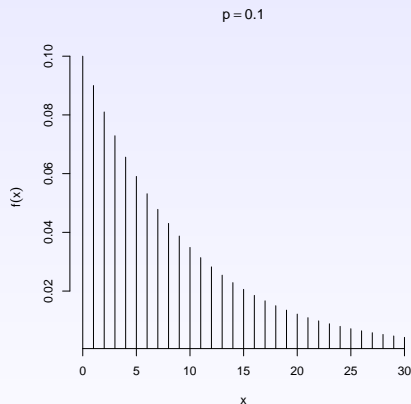


Figura: Funciones de masa para $\text{Geom}(0.1)$ y $\text{Geom}(0.8)$.

Distribución Geométrica

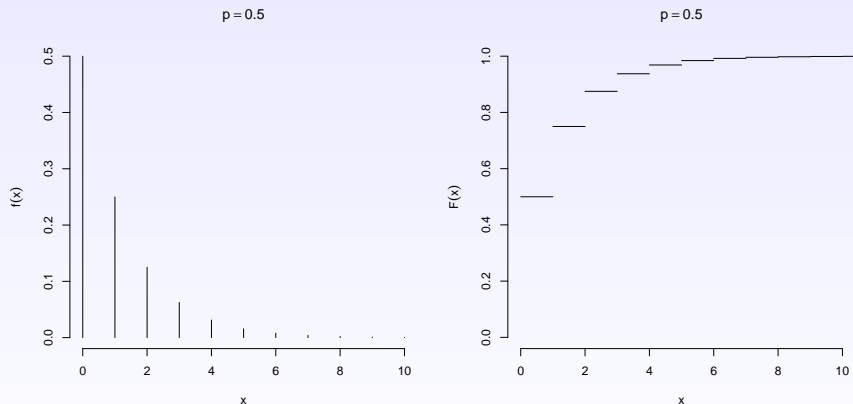


Figura: Funciones de masa y de distribución acumulada para $\text{Geom}(0.5)$.

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,

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

Por otro lado, la función de distribución de una v.a. $\text{Geom}(p)$ 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

$$P(X > 1 + 2 | X > 1) = P(X > 2),$$

$$P(X > 1 + 2 | X > 2) = P(X > 1),$$

$$P(X > 1 + 1 | X > 1) = P(X > 1).$$

Distribución Geométrica

Bosquejo: Otra forma de visualizar esto es pensando que se “recorre el valor más bajo posible de X ”. En el primer ejemplo numérico es como si desechará el 1 como origen y pusiera mi origen en 2 (que ahora sería 1). *Esto equivale a decir que si están haciendo inspecciones cada hora para verificar la vida útil de algún componente, el hecho de saber que no ha fallado en la observación 100, por ejemplo, (esto es, después de 100 horas de uso) no nos dice nada para responder a la pregunta: ¿cuál será la probabilidad de que falle en las siguientes 20 horas? La probabilidad de que falle es la misma que cuando el componente es nuevo y te haces la misma pregunta: ¿cuál será la probabilidad de que falle en las siguientes 20 horas?* Esto sólo tiene sentido si se está trabajando con un proceso que se encuentra en estado de estabilidad.

Distribución Geométrica

Media y Varianza

Ya que

$$\sum_{n=1}^{+\infty} nz^{n-1} = \frac{1}{(1-z)^2} \quad |z| < 1,$$

el valor esperado de $X \sim \text{Geom}(p)$ es

$$\begin{aligned} E(X) &= \sum_{x=1}^{+\infty} x \cdot f(x) = \sum_{x=1}^{+\infty} x \underbrace{(1-p)^{x-1}}_q p = p \sum_{x=1}^{+\infty} x q^{x-1} \\ &= p \left(\frac{1}{(1-q)^2} \right) = p \left(\frac{1}{p^2} \right) = \frac{1}{p}. \end{aligned}$$

Distribución Geométrica

Para el cálculo de la varianza, nos queda

$$\begin{aligned}V(X) &= E(X^2) - [E(X)]^2, \\E(X^2) &= E(X^2 - X + X) = E(X^2 - X) + E(X) \\&= E(X(X - 1)) + E(X),\end{aligned}$$

pero

$$\begin{aligned}E(X(X - 1)) &= \sum_{x=1}^{+\infty} x(x - 1) \cdot f(x) = \sum_{x=1}^{+\infty} x(x - 1)q^{x-1}p \\&= p \sum_{x=1}^{+\infty} x(x - 1)q^{x-1} = pq \sum_{x=1}^{+\infty} x(x - 1)q^{x-2} \\&= pq \left(\frac{2}{(1 - q)^3} \right) = pq \left(\frac{2}{p^3} \right) = \frac{2(1 - 2p)}{p^2}.\end{aligned}$$

Distribución Geométrica

La cantidad $E(X(X - 1))$ recibe el nombre de **Segundo Momento Factorial**, de aquí tenemos que

$$E(X^2) = E(X(X - 1)) + E(X) = \frac{2(1 - p)}{p^2} + \frac{1}{p} = \frac{2 - 2p + p}{p^2} = \frac{2 - p}{p^2},$$

y entonces

$$V(X) = \frac{2 - p}{p^2} - \left(\frac{1}{p}\right)^2 = \frac{2 - p - 1}{p^2} = \frac{1 - p}{p^2}.$$

Distribución Geométrica

Función Generatriz de Momentos:

$$\begin{aligned}
 M_X(t) &= E(e^{tX}) = \sum_{x=1}^{+\infty} e^{tx} \cdot f(x) = \sum_{x=1}^{+\infty} e^{tx} \cdot q^{x-1} p = p \sum_{x=1}^{+\infty} (e^t)^x \cdot q^{x-1} \\
 &= pe^t \sum_{x=1}^{+\infty} (e^t)^{x-1} \cdot q^{x-1} = pe^t \sum_{x=1}^{+\infty} (e^t q)^{x-1} \underset{(j=x-1)}{=} pe^t \sum_{j=0}^{+\infty} (e^t q)^j,
 \end{aligned}$$

usando la identidad $\sum_{n=0}^{+\infty} z^n = \frac{1}{1-z}$, para $|z| < 1$, nos queda

$$M_X(t) = pe^t \left(\frac{1}{1 - qe^t} \right) = \frac{pe^t}{1 - qe^t} \quad \text{si } |qe^t| < 1.$$

Distribución Geométrica

Ejemplo: Una tienda ofrece un premio a las personas que logren juntar las cinco letras de la palabra “valor”; para ello, se les regala una letra al azar de entre las cinco en cada compra que realicen. Lo que interesa aquí es la **cantidad de viajes necesarios (compras)** para juntar las cinco letras distintas.

Se podría pensar que la v.a. de interés, $X = \#$ de viajes necesarios hasta juntar las cinco letras de la palabra “valor”, es geométrica ya que:

- cada resultado es **independiente** del anterior,
- el número posible de viajes, realizados entre cada consecución de letras distintas, se puede considerar como **“infinito”**.

Distribución Geométrica

Sin embargo, la probabilidad de un “éxito” cada vez que compre **irá cambiando**: la primera vez la probabilidad de juntar una letra es 1; la segunda vez que compre ya tendrá en su poder una letra, con lo que la probabilidad de “éxito” cambiaría de 1 a la probabilidad de que reciba una letra distinta de la que se le dio inicialmente y esta sería $4/5$ (4 favorables de 5 posibles); y así seguiría hasta obtener otro “éxito” (letra distinta), con lo que p cambiaría a $3/5$, etc.

Entonces, se puede pensar que en cada viaje la probabilidad es mantenida fija hasta que haya “éxito”, y juntando esto con los dos supuestos anteriores, se tendrían v.a.’s geométricas (exceptuando la primera), con p 's distintas, cada vez que se lograra un “éxito”. Se puede visualizar lo anterior usando una recta numérica:



Distribución Geométrica

Así, nuestra v.a. de interés podría considerarse como la **suma de todos los viajes a la tienda**, esto es, la suma de cada una de las v.a.'s geométricas definidas:

$$X = X_1 + X_2 + X_3 + X_4 + X_5,$$

donde $X_i \sim \text{Geom}(p_i)$ $i = 2, 3, 4, 5$, con $p_2 = 4/5$, $p_3 = 3/5$, $p_4 = 2/5$, $p_5 = 1/5$. Entonces, podemos encontrar el **número esperado de viajes** calculando

$$\begin{aligned} E(X) &= E\left(\sum_{i=1}^5 X_i\right) = E(X_1) + E(X_2) + E(X_3) + E(X_4) + E(X_5) \\ &= 1 + \frac{1}{p_2} + \frac{1}{p_3} + \frac{1}{p_4} + \frac{1}{p_5} = 1 + \frac{1}{4/5} + \frac{1}{3/5} + \frac{1}{2/5} + \frac{1}{1/5} \\ &= 11.42 \text{ viajes,} \end{aligned}$$

esto es, entre 11 ó 12 viajes.

Distribución Geométrica

Ejemplo: Para el Sorteo Tec, si cada vez se expenden 150,000 boletos y 4 automóviles como premio, ¿en cuántos sorteos en promedio hay que participar hasta sacarse un automóvil? ¿Cuánto hay que invertir en promedio si cada boleto cuesta 1,050 pesos?

Como p es fija, hay independencia y “podemos” comprar boletos por siempre, entonces con $X = \#$ de participaciones hasta sacarse un automóvil, $E(X) = \frac{150,000}{4} = 37,500$. Además, habría que invertir $1,050 \times 37,500 = 39,375,000$ pesos.

Con dos rifas por año, uno esperaría ganarse un auto en unos ¡18,750 años! Cuántos años tendrías que participar en promedio si quieres ganarte la casa del primer premio?

Distribución Binomial Negativa

Considera nuevamente el ejemplo del premio al juntar las letras de la palabra “valor”, pero suponte que las probabilidades de éxito p_i se mantienen **constantes** (por ejemplo pensando que el premio se da a la persona que junte 5 letras “V”, con lo que p_i siempre sería $1/5$). Como antes, la variable aleatoria $X =$ **# de viajes necesarios hasta obtener 5 “éxitos”**, entonces

$$X = \sum_{i=1}^5 X_i, \text{ donde } X_i \sim \text{Geom}(p) \text{ son independientes y } p \text{ constante.}$$

Esto es equivalente al segundo criterio mencionado para la detención o ajuste de un proceso (ver diapositiva 56): hasta obtener r (por ejemplo 3) artículos defectuosos (debido tal vez a lo costoso de parar la línea al obtener el primer defectuoso), en donde

Distribución Binomial Negativa

- cada **éxito** tiene la **misma** probabilidad p , ya que se trata del mismo proceso y se asume que éste es estable (no hay fallas sistemáticas),
- cada X_i es **independiente**;

en este caso la variable aleatoria de interés es

$X = \#$ de observaciones **hasta** encontrar r “éxitos”,

la cual se puede redefinir como

$$X = \sum_{i=1}^r X_i, \text{ donde cada } X_i \sim \text{Geom}(p) \text{ es independiente y } p \text{ constante.}$$

La v.a. X definida como la suma anterior se dice que sigue una distribución **Binomial Negativa**: $X \sim \text{BN}(r, p)$. Aquí, inclusive $X_1 \sim \text{Geom}(p)$.

Distribución Binomial Negativa

A continuación se muestra un dibujo similar al hecho para la distribución geométrica (ver diapositiva 76), aquí obs = observaciones y Geo=Geom:



Media y Varianza

El valor esperado de la v.a. $X \sim \text{BN}(r, p)$ es

$$E(X) = E\left(\sum_{i=1}^r X_i\right) = \sum_{i=1}^r E(X_i) = \sum_{i=1}^r \frac{1}{p} = \frac{r}{p};$$

para la varianza usamos el hecho de que las X_i 's son independientes

$$V(X) = V\left(\sum_{i=1}^r X_i\right) = \sum_{i=1}^r V(X_i) = \sum_{i=1}^r \frac{1-p}{p^2} = r \left(\frac{1-p}{p^2}\right).$$

Distribución Binomial Negativa

Función Generatriz de Momentos

Usando la estructura de suma se tiene que la función generatriz de momentos de una distribución $BN(r,p)$ está dada por

$$\begin{aligned}M_X(t) &= E(e^{tX}) = E\left(e^{t\sum_{i=1}^r X_i}\right) = E\left(e^{tX_1+tX_2+\dots+tX_r}\right) \\&= E(e^{tX_1})E(e^{tX_2}) \dots E(e^{tX_r}) \\&= \left(\frac{pe^t}{1-qe^t}\right) \left(\frac{pe^t}{1-qe^t}\right) \dots \left(\frac{pe^t}{1-qe^t}\right) = \left(\frac{pe^t}{1-qe^t}\right)^r,\end{aligned}$$

para $|qe^t| < 1$.

Distribución Binomial Negativa

Recalquemos los supuestos:

- Cada vez que se realice una observación la probabilidad de que ocurra un éxito es fija e igual al valor p :

Nuevamente, el proceso “no sabe” o “no se acuerda” de lo que ha pasado antes de la observación que se esté realizando. Cuando uno detecte artículos defectuosos, sabiendo la probabilidad que el modelo asigna a ese caso, se toma una decisión respecto al proceso. Ver[15].

- Las observaciones son independientes.

Como antes, el que una observación haya sido o no “éxito”, no da información sobre el resultado de la siguiente observación.

- El primer posible valor de la variable aleatoria es r .

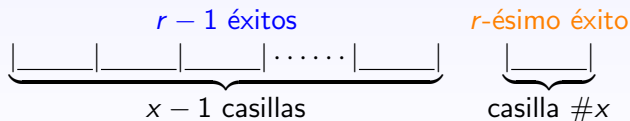
Puesto que la condición es encontrar r defectuosos, se tienen que realizar mínimo r pruebas.

- El número de observaciones puede continuar indefinidamente antes de encontrar el r -ésimo defectuoso.

Distribución Binomial Negativa

Una forma de obtener explícitamente a la función de probabilidad de la v.a. Binomial Negativa se da a continuación.

El evento de que el r -ésimo “éxito” ocurra en $X = x$ equivale a pensar que en las $x - 1$ observaciones anteriores hay $r - 1$ “éxitos”, los cuales pudieron ocurrir en *cualquiera* de las anteriores $x - 1$ observaciones. Si agrupamos en casillas las observaciones, lo anterior puede visualizarse de la siguiente manera



Distribución Binomial Negativa

Debido a la **independencia** sólo se tienen que multiplicar las probabilidades de cada casilla para obtener la probabilidad de r “éxitos” en x pruebas:

- En la última prueba hay “éxito” con probabilidad p .
- Los $r - 1$ “éxitos” anteriores pueden ocurrir en cualquier orden. El número de maneras en que $r - 1$ “éxitos” se pueden acomodar en $x - 1$ casillas está dado por $\binom{x-1}{r-1}$, y cada uno de ellos tiene asignada una probabilidad p .
- Hay $(x - 1) - (r - 1) = x - r$ fracasos cada uno con probabilidad $(1 - p)$.

Distribución Binomial Negativa

Entonces, la probabilidad de obtener r éxitos en x pruebas es

$$\begin{aligned} P(X = x) &= p \left[\binom{x-1}{r-1} p^{r-1} \right] (1-p)^{x-r} \\ &= \binom{x-1}{r-1} (1-p)^{x-r} p^r, \end{aligned}$$

con $x = r, r+1, r+2, \dots$

Esta es la **Distribución Binomial Negativa** cuyos **parámetros** son r y p .
Notemos que n sigue siendo aleatorio.

Distribución Binomial Negativa

Relación entre una v.a. $X \sim B(n, p)$ con una v.a. $Y \sim \text{BN}(r, p)$

Recuerda que

$X = \#$ de éxitos en n pruebas Bernoulli independientes,

$Y = \#$ de observaciones hasta tener r éxitos en pruebas Bernoulli independientes.

Consideremos la probabilidad de que en los n intentos se observen a lo más $r - 1$ “éxitos”, esto es,

$$P(X \leq r - 1).$$

Observemos que en los n intentos hay a lo más $r - 1$ éxitos si y sólo si en los n intentos se observan menos de r éxitos, es decir que la probabilidad anterior es igual a la probabilidad

$$P(Y > n).$$

Distribución Binomial Negativa

Entonces

$$P(X \leq r - 1) = P(Y > n) = 1 - P(Y \leq n).$$

Cabe recalcar que esto **no** es una aproximación, es una identidad.

Ejemplo: Si un proceso nos da **dos** equipos con defectos “severos” se detiene la producción. La probabilidad de ese tipo de defectos es **0.05**.
¿Cuál es la probabilidad de hacer más de 10 equipos sin detener la producción?

La variable aleatoria es $Y = \#$ de equipos revisados hasta obtener 2 con defectos “severos”.

Distribución Binomial Negativa

Supuestos:

- probabilidad **fija** e igual a p (no hay fallas sistemáticas)
- resultados **independientes**

Entonces $Y \sim \text{BN}(r = 2, p = 0.05)$, de donde

$$P(Y > 10) = 1 - P(Y \leq 10) = 1 - \sum_{y=2}^{10} \binom{y-1}{2-1} (0.05)^2 (0.95)^{y-2} = 0.914.$$

Utilizando la relación con la distribución binomial, para

$X \sim B(n = 10, p = 0.05)$ tenemos que

$$P(Y > 10) = P(X \leq 2 - 1) = \sum_{x=0}^1 \binom{10}{x} (0.05)^x (0.95)^{10-x} = 0.914.$$

Distribución Binomial Negativa

Los cálculos de la generatriz de momentos, media y varianza se pueden obtener a partir de la función de probabilidad de la binomial negativa, aunque de hecho ya lo hicimos mediante la representación en sumas de la distribución binomial negativa. Por completez se incluye el siguiente este material.

Deduciremos la generatriz de momentos para obtener la media y la varianza de $X \sim \text{BN}(r, p)$. Para calcular la generatriz de momentos se utilizarán los siguientes resultados:

$$\binom{n}{z} = \binom{n}{n-z}, \quad (1)$$

$$(1-x)^{-r} = \sum_{j=0}^{+\infty} \binom{j+r-1}{j} x^j, \quad \text{para } |x| < 1. \quad (2)$$

Distribución Binomial Negativa

La generatriz de momentos es entonces

$$\begin{aligned}
 M_X(t) &= E(e^{tX}) = \sum_{x=r}^{+\infty} e^{tx} \cdot f(x) = \sum_{x=r}^{+\infty} e^{tx} \binom{x-1}{r-1} q^{x-r} p^r \\
 &= p^r \sum_{x=r}^{+\infty} e^{tx} \binom{x-1}{r-1} q^{x-r} = p^r (e^t)^r \sum_{x=r}^{+\infty} \binom{x-1}{r-1} (qe^t)^{x-r} \\
 &= p^r (e^t)^r \sum_{j=0}^{+\infty} \binom{j+r-1}{r-1} (qe^t)^{x-r} \quad j = x - r \\
 &= (pe^t)^r \sum_{j=0}^{+\infty} \binom{j+r-1}{(j+r-1)-(r-1)} (qe^t)^{x-r} \quad \text{de (1)} \\
 &= (pe^t)^r \sum_{j=0}^{+\infty} \binom{j+r-1}{j} (qe^t)^{x-r} \\
 &\stackrel{\text{de (2)}}{=} (pe^t)^r (1 - qe^t)^{-r} = \frac{(pe^t)^r}{(1 - qe^t)^r} = \left(\frac{pe^t}{1 - qe^t} \right)^r, \quad |qe^t| < 1.
 \end{aligned}$$

Distribución Binomial Negativa

Por lo tanto

$$M_X(t) = \left(\frac{pe^t}{1 - qe^t} \right)^r, \quad \text{para } |qe^t| < 1.$$

Cuando $r = 1$ la binomial negativa es equivalente a la Geométrica; es decir, la binomial negativa se puede considerar como una generalización de la geométrica. Así,

$$\mu = \mu'_1 = E(X) = M'_X(t)|_{t=0},$$

$$\begin{aligned} M'_X(t) &= r \left(\frac{pe^t}{1 - qe^t} \right)^{r-1} \frac{(1 - qe^t)pe^t - pe^t(-qe^t)}{(1 - qe^t)^2} \\ &= r \left(\frac{pe^t}{1 - qe^t} \right)^{r-1} \frac{pe^t - pqe^{2t} + pqe^{2t}}{(1 - qe^t)^2} = r \left(\frac{pe^t}{1 - qe^t} \right)^{r-1} \frac{pe^t}{(1 - qe^t)^2}, \\ &= \frac{r(pe^t)^r}{(1 - qe^t)^{r+1}}. \end{aligned}$$

$$M'_X(0) = \frac{r(pe^0)^r}{(1 - qe^0)^{r+1}} = \frac{rp^r}{(1 - q)^{r+1}} = \frac{rp^r}{p^{r+1}} = \frac{r}{p}.$$

Distribución Binomial Negativa

Por lo tanto,

$$E(X) = r/p.$$

Por otra parte,

$$\mu'_2 = E(X^2) = M''_X(t)|_{t=0},$$

$$M''_X(t) = \frac{(1 - qe^t)^{r+1} r \cdot r(pe^t)^{r-1} pe^t - r(pe^t)^r (r+1)(1 - qe^t)^r (-qe^t)}{[(1 - qe^t)^{r+1}]^2},$$

$$\begin{aligned} M''_X(0) &= \frac{(1 - q)^{r+1} r^2 p^{r-1} p - rp^r (r+1)(1 - q)^r (-q)}{(1 - q)^{2(r+1)}} \\ &= \frac{p^{r+1} r^2 p^r + r(r+1)p^r p^r q}{p^{2r+2}} \\ &= \frac{p^{2r+1} r^2 + r(r+1)p^{2r} q}{p^{2r+2}} = \frac{rp^{2r}(rp + (r+1)q)}{p^{2r+2}} \\ &= \frac{r \left(\overbrace{rp + rq}^r + q \right)}{p^2} = \frac{r(r+q)}{p^2}. \end{aligned}$$

Distribución Binomial Negativa

Así que

$$E(X^2) = \frac{r(r+q)}{p^2}.$$

Entonces,

$$\begin{aligned} V(X) &= E(X^2) - [E(X)]^2 = \frac{r(r+q)}{p^2} - \left(\frac{r}{p}\right)^2 \\ &= \frac{r(r+q) - r^2}{p^2} = \frac{r^2 + rq - r^2}{p^2} = \frac{rq}{p^2} = \frac{r(1-p)}{p^2}, \end{aligned}$$

lo que implica

$$V(X) = \frac{r(1-p)}{p^2}.$$

Distribución Binomial Negativa

A modo de resumen, notemos lo siguiente:

- Notemos que en un contexto teórico el valor de r puede ser **distinto de un número entero**.
- Además, la variable aleatoria BINOMIAL, cuando $n = 1$, nos da una variable aleatoria que tiene distribución Bernoulli; es decir, la Bernoulli es un caso particular de la binomial.
- Similarmente, una variable aleatoria BINOMIAL NEGATIVA, con $r = 1$, nos da una variable aleatoria con distribución GEOMÉTRICA; es decir, la geométrica es un caso particular de la binomial negativa.

Distribución Poisson

Supóngase ahora que estamos interesado en el número de “éxitos”, pero ahora preguntándonos por la posibilidad de que ocurran en un cierto **intervalo de tiempo o espacio**, por ejemplo

- Número de defectos en una plancha de acero de 1 m^2 ,
- Número de bacterias en 1 cm^3 de agua potable,
- Número de entregas de materia prima entre las 8:00 A.M. y la 1:00 P.M.,
- Número de llamadas que llegan a un conmutador en un minuto.

En estos casos interesa la variable aleatoria

$X = \#$ de “éxitos” obtenidos en un cierto **intervalo (temporal o espacial)**.

Distribución Poisson

Nuestros supuestos son:

- El número promedio de “éxitos”, λ (la intensidad), sobre el intervalo dado se mantiene constante.
Se considera que el comportamiento es estable. Incluso, si se estuviera interesado en un múltiplo t (no necesariamente entero) del intervalo original, el valor λ_1 correspondiente será $\lambda_1 = \lambda t$.
- Los éxitos aparecen en forma aleatoria en intervalos de la misma magnitud, los intervalos no se traslapan y son independientes entre ellos.
Se considera que si en un sub-intervalo del intervalo dado ocurre un determinado número de “éxitos”, no implica que en otro sub-intervalo de la misma amplitud tendrá ese mismo número, mayor que él, o menor. Sólo se espera que en promedio se tenga el mismo número.

Distribución Poisson

*Observa que no se hace un número de pruebas hasta obtener “éxitos”, sino que se cuenta el número de éxitos obtenidos en un **continuo** espacial o temporal y por ende, cuando el intervalo tiende a cero la probabilidad de éxito también tiende a cero.*

- El número posible de “éxitos” es **infinito**.

Distribución Poisson

Mecanismo para asignación de probabilidades

Supongamos que tenemos un proceso Poisson y lo observamos en un cierto intervalo, digamos de una unidad, y asumimos que λ es el número promedio de éxitos en esa unidad observada.

Por otra parte, consideremos esa misma unidad y que:

- Se realizan 10 observaciones y se cuenta el número de “éxitos”, cada uno con probabilidad p , obtenidos en 10 subdivisiones. Es natural pensar que hay una probabilidad relativamente pequeña de observar más de un éxito, en cada subdivisión.
- Se realizan 10 observaciones y se cuenta el número de “éxitos”, cada uno con probabilidad p (más pequeña que en el anterior caso), obtenidos en mil subdivisiones del intervalo de tal manera que p será tan pequeña que hay una probabilidad de cero de observar más de un éxito, en cada subdivisión.

Distribución Poisson

El número de “éxitos” Y en n pruebas independientes con probabilidad p se puede modelar ahora como una v.a. Binomial, cuya media es $E(Y) = np$, que debe coincidir con el número promedio de éxitos λ en el intervalo considerado.

Notemos que $\lambda = np$ implica que $p = \lambda/n$ y como λ es fija, p será más y más pequeña a medida que n crece. Esto nos lleva a establecer el siguiente resultado:

La distribución de Y es **Poisson** con $\lambda = np$ constante, cuando el número de observaciones tiende a infinito en el intervalo, con p siendo más pequeña conforme n aumenta.

Distribución Poisson

Demostración:

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

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad np = \lambda \rightarrow p = \frac{\lambda}{n}$$

$$\begin{aligned} f(x) &= \frac{n!}{x!(n-x)!} \left(\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^{n-x} \\ &= \frac{n(n-1)(n-2)\cdots(n-(x-1))(n-x)!}{x!(n-x)!} \frac{\lambda^x}{n^x} \left(1 - \frac{\lambda}{n}\right)^{n-x} \\ &= \frac{1 \left(1 - \frac{1}{n}\right) \left(1 - \frac{2}{n}\right) \cdots \left(1 - \frac{x-1}{n}\right)}{x!} \lambda^x \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-x} \\ &= \frac{1 \left(1 - \frac{1}{n}\right) \left(1 - \frac{2}{n}\right) \cdots \left(1 - \frac{x-1}{n}\right)}{x!} \lambda^x \left[\left(1 - \frac{\lambda}{n}\right)^{-\frac{n}{\lambda}}\right]^{-\lambda} \left(1 - \frac{\lambda}{n}\right)^{-x}. \end{aligned}$$

Distribución Poisson

Por otro lado,

- $1 \left(1 - \frac{1}{n}\right) \left(1 - \frac{2}{n}\right) \cdots \left(1 - \frac{x-1}{n}\right) \rightarrow 1$ cuando $n \rightarrow +\infty$
- $\left(1 - \frac{\lambda}{n}\right)^{-x} \rightarrow 1$ cuando $n \rightarrow +\infty$
- $x!$ y λ^x quedan igual

El único término faltante es $\left(1 - \frac{\lambda}{n}\right)^{-\frac{n}{\lambda}}$. Observe que $\lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n}\right)^{-\frac{n}{\lambda}}$ tiene la forma $1^{-\infty}$ que es una forma indeterminada.

Distribución Poisson

Sea $y = \left(1 - \frac{\lambda}{n}\right)^{-\frac{n}{\lambda}}$, entonces

$$\log(y) = -\frac{n}{\lambda} \log\left(1 - \frac{\lambda}{n}\right) = \frac{-\log\left(1 - \frac{\lambda}{n}\right)}{\frac{\lambda}{n}}.$$

De aquí tenemos,

$$\lim_{n \rightarrow \infty} \log(y) = \lim_{n \rightarrow \infty} \frac{-\log\left(1 - \frac{\lambda}{n}\right)}{\frac{\lambda}{n}}, \quad \text{forma indeterminada del tipo } \frac{0}{0};$$

aplicando la Regla de L'Hopital

$$\lim_{n \rightarrow \infty} -\frac{\frac{1}{(1-\lambda/n)} \cdot \left(\frac{\lambda}{n^2}\right)}{\left(-\frac{\lambda}{n^2}\right)} = \lim_{n \rightarrow \infty} \frac{1}{\left(1 - \frac{\lambda}{n}\right)} = 1 \quad \Rightarrow \quad \lim_{n \rightarrow \infty} \log(y) = 1$$

Distribución Poisson

Ahora, usando el hecho de que si f es una función continua se cumple

$$\lim_{x \rightarrow a} f(y) = f\left(\lim_{x \rightarrow a} y\right);$$

y, por ser la función logaritmo natural una función continua, tenemos

$$\log\left(\lim_{n \rightarrow \infty} y\right) = 1 \quad \Rightarrow \quad \lim_{n \rightarrow \infty} y = e.$$

De aquí que

- $\left(1 - \frac{\lambda}{n}\right)^{-\frac{n}{\lambda}} \rightarrow e$, cuando $n \rightarrow \infty$.

Distribución Poisson

Usando todos los resultados anteriores tenemos que la función de probabilidad de la **Distribución Poisson** está dada por

$$f(x) = \frac{\lambda^x e^{-\lambda}}{x!}, \quad \text{para } x = 0, 1, 2, \dots$$

Notación: $X \sim \text{Poisson}(\lambda)$. El único parámetro del modelo es λ .