

# MODELOS DE PROBABILIDAD II

DISTRIBUCIÓN GEOMÉTRICA  
DISTRIBUCIÓN BINOMIAL NEGATIVA  
DISTRIBUCIÓN HIPERGEOMÉTRICA  
DISTRIBUCIÓN DE PÓLYA  
INTRODUCCIÓN A LA ESTIMACIÓN BAYESIANA DE PROPORCIONES

---

## DISTRIBUCIÓN GEOMÉTRICA O DE PASCAL

La distribución geométrica es un modelo adecuado para aquellos procesos en los que se repiten pruebas hasta la consecución del éxito a resultado deseado y tiene interesantes aplicaciones en los muestreos realizados de esta manera. También implica la existencia de una dicotomía de posibles resultados y la independencia de las pruebas entre sí.

Proceso experimental del que se puede hacer derivar

Esta distribución se puede hacer derivar de un proceso experimental puro o de Bernouilli en el que tengamos las siguientes características

- **El proceso consta de un número no definido de pruebas o experimentos separados o separables. El proceso concluirá cuando se obtenga por primera vez el resultado deseado (éxito).**
- **Cada prueba puede dar dos resultados mutuamente excluyentes: A y no A**
- **La probabilidad de obtener un resultado A en cada prueba es p y la de obtener un resultado no A es q siendo ( $p + q = 1$ ).**

Las probabilidades p y q son constantes en todas las pruebas, por tanto, las pruebas, son independientes (si se trata de un proceso de "extracción" éste se llevará a cabo con devolución del individuo extraído).

- **(Derivación de la distribución).** Si en estas circunstancias aleatorizamos de forma que tomemos como variable aleatoria  $X =$  el número de pruebas necesarias para obtener por primera vez un éxito o resultado A, esta variable se distribuirá con una distribución geométrica de parámetro p.

$$X \Rightarrow G(p)$$

## Obtención de la función de cuantía

De lo dicho anteriormente, tendremos que la variable X es el número de pruebas necesarias para la consecución del primer éxito. De esta forma la variables aleatoria toma valores enteros a partir del uno;  $\{ 1,2,\dots\dots\dots \}$

La función de cuantía P(x) hará corresponder a cada valor de X la probabilidad de obtener el primer éxito precisamente en la X-sima prueba. Esto es, P(X) será la probabilidad del suceso obtener X-1 resultados "no A" y un éxito o resultado A en la prueba número X teniendo en cuenta que todas las pruebas son independientes y que conocemos sus probabilidades tendremos:

$$\text{suceso} \equiv \underbrace{\bar{A} \text{ y } \bar{A} \text{ y } \dots \text{ y } \bar{A}}_{x-1 \text{ veces}} \text{ y } \underbrace{A}_{\text{una vez}} = \underbrace{\bar{A} \cap \bar{A} \cap \dots \cap \bar{A}}_{x-1 \text{ veces}} \cap \underbrace{A}_{\text{una vez}}$$

dado que se trata de sucesos independientes y conocemos las probabilidades

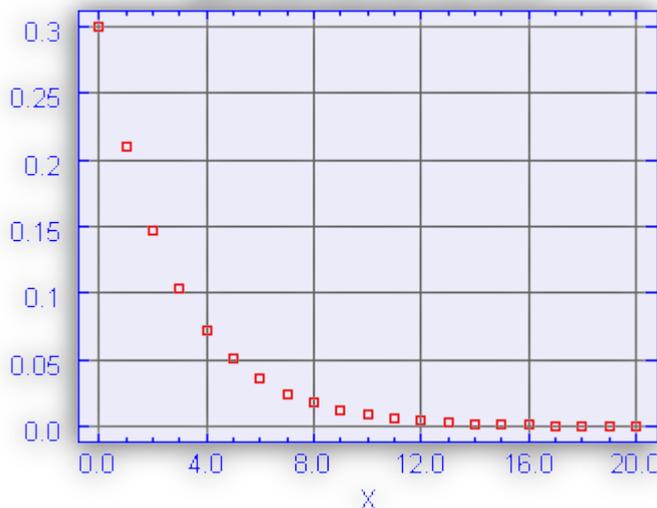
$$P(x) = \underbrace{P(\bar{A}) \cdot P(\bar{A}) \cdot \dots \cdot P(\bar{A})}_{x-1 \text{ veces}} \cdot P(A) = q^{x-1} \cdot p \text{ dado que}$$

$$P(\bar{A}) = q \quad \text{y} \quad P(A) = p$$

luego la función de cuantía quedaría

$$P(x) = q^{x-1} \cdot p$$

geométrica con p=0,3



Algunos autores consideran la aleatorización como "número de pruebas anteriores al primer éxito". De esta manera el conseguir el éxito a la primera sería X=0. En la siguiente representación gráfica de la función de cuantía de la geométrica puede apreciarse este tipo de aleatorización, sin embargo nosotros preferimos, por razones prácticas, utilizar la aleatorización antes comentada

Función de distribución

En base a la función de cuantía se puede expresar la función de distribución de la siguiente manera.

$$F(X) = \sum_{x=1}^X q^{x-1} \cdot p$$

desarrollando la expresión tendríamos

$$F(X) = p(1 + q + q^2 + \dots + q^{X-1}) = p \cdot \frac{1 - q^X}{1 - q}$$

de donde

$$F(X) = 1 - q^X$$

La Función Generatriz de Momentos (F.G.M.) quedaría:

$$\varphi(t) = E[e^{xt}] = \sum_{x=1}^{\infty} e^{xt} \cdot p \cdot q^{x-1} = \frac{p}{q} \sum_{x=1}^{\infty} (e^t q)^x =$$

$$\lim_{x \rightarrow \infty} \frac{p}{q} \left[ qe^t + (qe^t)^2 + (qe^t)^3 + \dots + (qe^t)^x \right] =$$

$$\lim_{x \rightarrow \infty} \frac{p}{q} \frac{(qe^t)^{x+1} - qe^t}{qe^t - 1} = \frac{p}{q} \frac{qe^t}{1 - qe^t} = p \frac{e^t}{1 - qe^t} = \frac{p}{e^{-t} - q} = p(e^{-t} - q)^{-1}$$

por lo que queda establecida que la F.G.M. tiene la expresión

$$\varphi(t) = p(e^{-t} - q)^{-1}$$

En base a la FGM podemos obtener la media y varianza:

Así

$$\alpha_1 = \mu = \varphi'(t)$$

$$\varphi'(t) = (-1)p(e^{-t} - q)^{-2}(-e^{-t}) = p(e^{-t} - q)^{-2}(e^{-t})$$

Haciendo  $t=0$  tendríamos que

$$\mu = \frac{1}{p}$$

La **varianza** sería

$$\sigma^2 = \alpha_2 - \mu^2$$

$$\varphi''(t) = (-2)p(e^{-t} - q)^{-3}(-e^{-t})(e^{-t}) + (-e^{-t})p(e^{-t} - q)^{-2} =$$

Haciendo  $t=0$  tendríamos que

$$\alpha_2 = -2p \cdot p^{-3}(-1) + (-1)p \cdot p^{-2} = \frac{2}{p^2} - \frac{1}{p}$$

De esta manera

$$\sigma^2 = \alpha_2 - \mu^2 = \frac{2}{p^2} - \frac{1}{p} - \frac{1}{p^2} = \frac{1}{p^2} - \frac{1}{p} = \frac{1-p}{p^2} = \frac{q}{p^2}$$

Luego

$$\sigma^2 = \frac{q}{p^2}$$

La **moda** es el valor de la variable que tiene asociada mayor probabilidad el valor de su función de cuantía es el mayor. Es fácil comprobar (véase simplemente la representación gráfica anterior) que

$$P(x_i) \leq P(x=1) \quad \forall x_i$$

Por lo tanto la media de la distribución geométrica es siempre 1.

En cuanto a la **mediana**  $M_e$  será aquel valor de la variable en el cual la función de distribución toma el valor 0,5. Así

$$F(M_e) = 1/2$$

$$F(M_e) = 1 - q^{M_e} = 1/2 \Rightarrow q^{M_e} = 1/2$$

por lo que

$$M_e \ln q = \ln 1 - \ln 2 = -\ln 2 \rightarrow M_e = \frac{-\ln 2}{\ln q}$$

## DISTRIBUCIÓN BINOMIAL NEGATIVA.

Esta distribución puede considerarse como una extensión o ampliación de la distribución geométrica. La distribución binomial negativa es un modelo adecuado para tratar aquellos procesos en los que se repite un determinado ensayo o prueba hasta conseguir un número determinado de resultados favorables (por vez primera). Es por tanto de gran utilidad para aquellos muestreos que procedan de esta manera. Si el número de resultados favorables buscados fuera 1 estaríamos en el caso de la distribución geométrica. Está implicada también la existencia de una dicotomía de resultados posibles en cada prueba y la independencia de cada prueba o ensayo, o la reposición de los individuos muestreados.

Proceso experimental del que puede hacerse derivar

Esta distribución o modelo puede hacerse derivar de un proceso experimental puro o de Bernoulli en el que se presenten las siguientes condiciones

- **El proceso consta de un número no definido de pruebas separadas o separables. El proceso concluirá cuando se obtenga un determinado número de resultados favorables K**
- **Cada prueba puede dar dos resultados posibles mutuamente excluyentes A y no A**
- **La probabilidad de obtener un resultado A en cada una de las pruebas es p siendo la probabilidad de no A, q. Lo que nos lleva a que p+q=1**
- **Las probabilidades p y q son constantes en todas las pruebas. Todas las pruebas son independientes.** Si se trata de un experimento de extracción éste se llevará cabo con devolución del individuo extraído, a no ser que se trate de una población en la que el número de individuos tenga de carácter infinito.
- (Derivación de la distribución) **Si, en estas circunstancias aleatorizamos de forma que la variable aleatoria x sea "el número de pruebas necesarias para conseguir K éxitos o resultados A " ; entonces la variable aleatoria x seguirá una distribución binomial negativa con parámetros p y k**

será entonces

$$x \Rightarrow BN(p, k)$$

La variable aleatoria x podrá tomar sólo valores superiores a k

$$x \in \{k, k+1, k+2, \dots\}$$

El suceso del que se trata podría verse como:

$$\underbrace{\bar{A}y \bar{A}y \bar{A}y \dots y \bar{A}y \overbrace{Ay Ay A \dots y A}^{K \text{ veces}}}_{x \text{ veces}}$$

o lo que es lo mismo

$$\underbrace{\bar{A} \cap \bar{A} \cap \bar{A} \cap \dots \cap \bar{A} \cap \overbrace{A \cap A \cap A \dots \cap A}^{K \text{ veces}}}_{x \text{ veces}}$$

dado que las pruebas son independientes y conocemos que P(A)= p y P(no A)= q

$$\underbrace{q \cdot q \cdot q \cdots q}_{x \text{ veces}} \cdot \underbrace{p \cdot p \cdot p \cdots p}_{K \text{ veces}} = q^{x-k} \cdot p^k$$

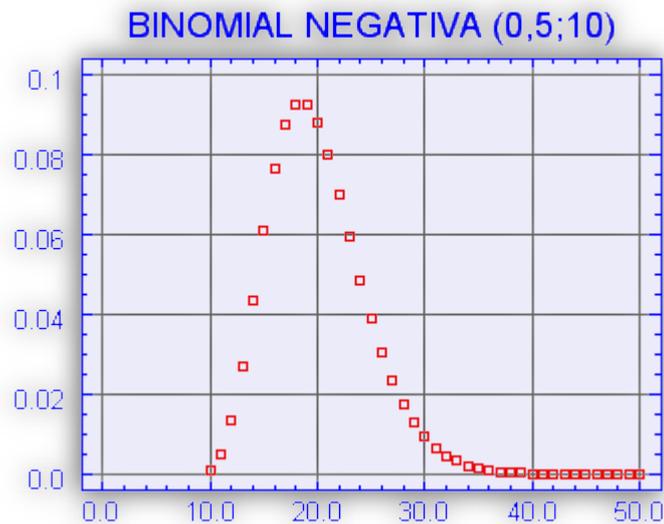
que sería la probabilidad de x si el suceso fuera precisamente con los resultados en ese orden. Dado que pueden darse otros órdenes, en concreto

$$\binom{x-1}{x-k}$$

formas u órdenes distintos . La función de cuantía de la distribución binomial negativa quedará como :

$$P(x) = \binom{x-1}{x-k} q^{x-k} \cdot p^k$$

Como ejemplo la representación gráfica de una variable  $X \Rightarrow BN(0,5;10)$  sería la siguiente .



Como en el caso de la geométrica, algunos autores aleatorizan de distinta manera el mismo proceso. Así X sería el número de fracasos (k) necesarios antes de conseguir el r-ésimo éxito. En este caso el número de pruebas sería k + r (lo que nosotros hemos llamado x) y r lo que nosotros hemos denominado k.

Para este tipo de aleatorización la función de cuantía sería:

$$P(x = k) = \binom{k+r-1}{k} q^k p^r$$

que como se observa es la misma si se realizan los antes nombrados cambios

La función generatriz de momentos será (según nuestra aleatorización) para una  $BN(k,p)$ .

$$\begin{aligned}\varphi(t) &= E[e^{tx}] = \sum_{x_i} e^{tx_i} \binom{x_i-1}{x_i-k} p^k q^{x_i-k} = p^k \sum_{x_i} e^{tx_i} \binom{x_i-1}{x_i-k} q^{x_i-k} = \\ &= p^k (e^{-t} - q)^{-k} = \varphi(t)\end{aligned}$$

Aplicando el teorema de los momentos hallamos media y varianza que resultan ser:

$$\mu = \frac{k}{p} \quad \sigma^2 = \frac{kq}{p^2}$$

No parece necesario recordar que si nos encontramos con una distribución  $BN(k=1,p)$  realmente se trata de una distribución geométrica.

## DISTRIBUCIÓN HIPERGEOMÉTRICA

Hasta ahora hemos analizado distribuciones que modelizaban situaciones en las que se realizaban pruebas que entrañaban una dicotomía (proceso de Bernouilli) de manera que en cada experiencia la probabilidad de obtener cada uno de los dos posibles resultados se mantenía constante. Si el proceso consistía en una serie de extracciones o selecciones ello implicaba la reposición de cada extracción o selección, o bien la consideración de una población muy grande. Sin embargo si la población es pequeña y las extracciones no se reemplazan las probabilidades no se mantendrán constantes. En ese caso las distribuciones anteriores no nos servirán para la modelizar la situación. La distribución hipergeométrica viene a cubrir esta necesidad de modelizar procesos de Bernouilli con probabilidades no constantes (sin reemplazamiento).

La distribución hipergeométrica es especialmente útil en todos aquellos casos en los que se extraigan muestras o se realizan experiencias repetidas sin devolución del elemento extraído o sin retornar a la situación experimental inicial.

Modeliza, de hecho, situaciones en las que se repite un número determinado de veces una prueba dicotómica de manera que con cada sucesivo resultado se ve alterada la probabilidad de obtener en la siguiente prueba uno u otro resultado. Es una distribución fundamental en el estudio de muestras pequeñas de poblaciones pequeñas y en el cálculo de probabilidades de, juegos de azar y tiene grandes aplicaciones en el control de calidad en otros procesos experimentales en los que no es posible retornar a la situación de partida.

La distribución hipergeométrica puede derivarse de un proceso experimental puro o de Bernouilli con las siguientes características:

- El proceso consta de  $n$  pruebas, separadas o separables de entre un conjunto de  $N$  pruebas posibles.
- Cada una de las pruebas puede dar únicamente dos resultados mutuamente excluyentes:  $A$  y no  $A$ .
- En la primera prueba las probabilidades son:  $P(A) = p$  y  $P(\bar{A}) = q$ ; con  $p+q=1$ .

Las probabilidades de obtener un resultado  $A$  y de obtener un resultado no  $A$  varían en las sucesivas pruebas, dependiendo de los resultados anteriores.

- (Derivación de la distribución). Si estas circunstancias a leatorizamos de forma que la variable aleatoria  $X$  sea el número de resultados  $A$  obtenidos en  $n$  pruebas la distribución de  $X$  será una Hipergeométrica de parámetros  $N, n, p$  así

$$X \Rightarrow H(N, n, p)$$

Un típico caso de aplicación de este modelo es el siguiente:

Supongamos la extracción aleatoria de  $n$  elementos de un conjunto formado por  $N$  elementos totales, de los cuales  $Np$  son del tipo  $A$  y  $Nq$  son del tipo  $\bar{A}$  siendo  $(p+q=1)$ . Si realizamos las extracciones sin devolver los elementos extraídos, y llamamos  $X$  al número de elementos del tipo  $A$  que extraemos en  $n$  extracciones  $X$  seguirá una distribución hipergeométrica de parámetros  $N, n, p$

Función de cuantía.

La función de cuantía de una distribución Hipergeométrica hará corresponder a cada valor de la variable  $X$  ( $x = 0, 1, 2, \dots, n$ ) la probabilidad del suceso "obtener  $x$  resultados del tipo  $A$ ", y  $(n-x)$  resultados del tipo no  $A$  en las  $n$  pruebas realizadas de entre las  $N$  posibles.

Veamos:

Hay un total de

$$\binom{Np}{x} \cdot \binom{Nq}{n-x}$$

formas distintas de obtener  $x$  resultados del tipo  $A$  y  $n-x$  del tipo  $\bar{A}$  si partimos de una población formada por  $Np$  elementos del tipo  $A$  y  $Nq$  elementos del tipo  $\bar{A}$

Por otro lado si realizamos  $n$  pruebas o extracciones hay un total de

$$\binom{N}{n}$$

posibles muestras ( grupos de n elementos)

aplicando la regla de Laplace tendríamos

$$P(x) = \frac{\text{casos favorables}}{\text{casos posibles}} \text{ es decir}$$

$$P(x) = \frac{\binom{Np}{x} \cdot \binom{Nq}{n-x}}{\binom{N}{n}}$$

que para valores de X comprendidos entre el conjunto de enteros 0,1,... .n será la expresión de la función de cuantía de una distribución , Hipergeométrica de parámetros N,n,p .

Media y varianza.

Considerando que una variable hipergeométrica de parámetros N, n, p puede considerarse generada por la reiteración de un proceso dicotómico n veces en el que las n dicotomías NO son independientes; podemos considerar que una variable hipergeométrica es la suma de n variables dicotómicas NO independientes.

Es bien sabido que la media de la suma de variables aleatorias (sean éstas independientes o no) es la suma de las medias y por tanto la media de una distribución hipergeométrica será, como en el caso de la binomial :

$$\mu = np$$

En cambio si las variables sumando no son independientes la varianza de la variable suma no será la suma de las varianzas.

Si se evalúa el valor de la varianza para nuestro caso se obtiene que la varianza de una distribución hipergeométrica de parámetros N,n,p es :

si

$$X \Rightarrow H(N, n, p)$$

$$\sigma^2 = npq \left( \frac{N-n}{N-1} \right)$$

para demostración de esta expresión véase Wilks S. ,Mathematical Statistics,1962

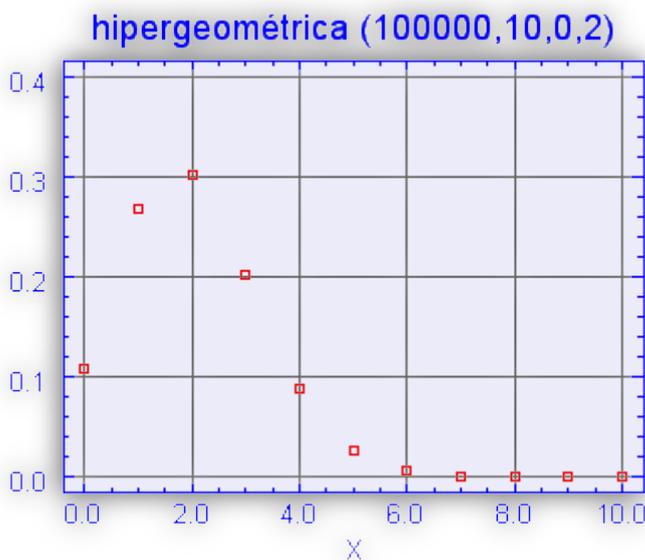
Esta forma resulta ser la expresión de la varianza de una binomial (n, p) afectada por un coeficiente corrector  $[N-n/N-1]$ , llamado coeficiente de exhaustividad o Factor Corrector de Poblaciones Finitas (F.C.P.F.) y que da cuenta del efecto que produce la no reposición de los elementos extraídos en el muestreo.

Este coeficiente es tanto más pequeño cuanto mayor es el tamaño muestral (número de pruebas de n ) y puede comprobarse como tiende a aproximarse a 1 cuando el tamaño de la población N es muy grande . Este último hecho nos confirma lo ya comentado sobre la irrelevancia de la reposición o no cuando se realizan extracciones sucesivas sobre una población muy grande. Con una población muy grande se cual fuere el tamaño de n , el factor corrector sería uno lo que convertiría , en cierto modo a la hipergeométrica en una binomial . Así

Límite de la distribución hipergeométrica cuando N tiende a infinito.

Hemos visto como la media de la distribución hipergeométrica  $[H\{N,n,p\}]$ , tomaba siempre el mismo valor que la media de una distribución binomial  $[B\{n,p\}]$  también hemos comentado que si el valor del parámetro N crecía hasta aproximarse a infinito el coeficiente de exhaustividad tendía a ser 1, y, por lo tanto, la varianza de la hipergeométrica se aproximaba a la de la binomial : puede probarse asimismo , cómo la función de cuantía de una distribución hipergeométrica tiende a aproximarse a la

función de cuantía de una distribución binomial cuando  $N \rightarrow \infty$



Puede comprobarse en la representación gráfica de una hipergeométrica con  $N = 100000$  como ésta, es idéntica a la de una binomial con los mismos parámetros restantes n y p, que utilizamos al hablar de la binomial

Moda de la distribución hipergeométrica

De manera análoga a como se obtenía la moda en la distribución binomial es fácil obtener la expresión de ésta para la distribución hipergeométrica. De manera que su expresión  $X_0$  sería la del valor o valores enteros que verificasen.

$$\frac{N(pn - q) + n - 1}{N + 2} \leq X_0 \leq \frac{N(pn + p) + n + 1}{N + 2}$$

## DISTRIBUCIÓN DE PÓLYA

Se trata de un modelo discreto de gran utilidad en la ciencia actuarial dado que también es conocida como distribución del contagio. Su utilidad radica precisamente en que es capaz de medir la aleatoriedad del "contagio" no sólo de enfermedades, como cabría suponer, sino de la extensión-propagación de la información sobre cualidades de los productos o servicios.

Se basa en un proceso, en cierto sentido, parecido al que sirve para dar lugar a la distribución hipergeométrica.

Sus características son las siguientes:

La distribución de Pólya puede derivarse de un proceso experimental puro o de Bernouilli con las siguientes características:

- **El proceso consta de n pruebas, separadas o separables de entre un conjunto de N pruebas posibles.**
- **Cada una de las pruebas puede dar únicamente dos resultados mutuamente excluyentes: A y no A.**
- **En la primera prueba las probabilidades son:  $P(A) = p$  y  $P(\bar{A}) = q$ ; con  $p + q = 1$ .**

**Las probabilidades de obtener un resultado A y de obtener un resultado no A varían en las sucesivas pruebas, dependiendo de los resultados anteriores.** De manera que aumentan las probabilidades del suceso que haya ocurrido en la prueba anterior, dado que si el resultado es A se aumenta la probabilidad de A para la siguiente prueba en un factor que denominamos de contagio (c). En el caso de que el resultado fuera no A se procedería de la misma manera (aumentando las probabilidades de no A)

- (Derivación de la distribución). **Si estas circunstancias aleatorizamos de forma que la variable aleatoria X sea el número de resultados A obtenidos en n pruebas la distribución de X será una Pólya de parámetros N,n,p,c:**

$$x \Rightarrow Pólya(N, n, p, c)$$

Planteemos un proceso como ejemplo:

En una urna existen  $N$  elementos  $N_1$  de carácter A y  $N_2$  de carácter no A

Evidentemente  $N = N_1 + N_2$

El experimento (prueba) consiste en la visualización del carácter de un elemento extraído

En la primera ocasión la

$$P(A) = \frac{N_1}{N} = p \quad \text{siendo} \quad P(\bar{A}) = \frac{N_2}{N} = q$$

Si el resultado es A se introducen en la urna  $C$  elementos de carácter A

Si el resultado es no A se introducen en la urna  $C$  elementos de carácter no A

Las probabilidades habrán variado .Así

La probabilidad de obtener  $x$  resultados A en  $x$  pruebas sería:

$$\frac{N_1}{N} \cdot \frac{N_1 + c}{N + c} \cdot \dots \cdot \frac{N_1 + (x-1)c}{N + (x-1)c}$$

dado que se realizan  $n$  pruebas , debemos considerar también las  $n-x$  restantes y cuyo resultado sería "no A" .

Así

$$\frac{N_2}{N + xc} \cdot \frac{N_2}{N + (x+1)c} \cdot \dots \cdot \frac{N_2 + (n-x-1)c}{N + (n-1)c}$$

evidentemente realizaremos todas las  $n$  pruebas, luego tendríamos:

$$\frac{N_1}{N} \cdot \frac{N_1 + c}{N + c} \cdot \dots \cdot \frac{N_1 + (x-1)c}{N + (x-1)c} \cdot \frac{N_2}{N + xc} \cdot \frac{N_2}{N + (x+1)c} \cdot \dots \cdot \frac{N_2 + (n-x-1)c}{N + (n-1)c}$$

Dado que no tendrían porqué darse los resultados en ese determinado orden. La función de cuantía quedaría:

$$P(x) = \binom{n}{x} \frac{N_1}{N} \cdot \frac{N_1 + c}{N + c} \cdot \dots \cdot \frac{N_1 + (x-1)c}{N + (x-1)c} \cdot \frac{N_2}{N + xc} \cdot \frac{N_2}{N + (x+1)c} \cdot \dots \cdot \frac{N_2 + (n-x-1)c}{N + (n-1)c}$$

La esperanza matemática de esta distribución quedará como:

$$E[x] = \mu = np$$

Siendo la varianza

$$\sigma^2 = npq \frac{1 + n \frac{c}{N}}{1 + \frac{c}{N}}$$

al cociente  $c/N$  se le conoce como fracción inicial de contagio . Proporción de individuos que se reponen o alcanzan el valor del resultado de la prueba anterior.

Puede comprobarse que si  $C$  ( número de individuos que se introducen en la urna con la característica del resultado anterior) es  $(-1)$ . Estaríamos "no reponiendo" la extracción , estaríamos , por tanto, ante una distribución hipergeométrica . Lo que podemos comprobar con su varianza.

Así : la varianza de la hipergeométrica es

$$\sigma^2 = npq \left( \frac{N-n}{N-1} \right)$$

mientras que varianza de la Pólya hemos visto que era

$$\sigma^2 = npq \frac{1 + n \frac{c}{N}}{1 + \frac{c}{N}}$$

si hacemos  $c = -1$  tendremos

$$\sigma^2 = npq \frac{1 + n \frac{-1}{N}}{1 + \frac{-1}{N}} = npq \frac{1 - \frac{n}{N}}{1 - \frac{1}{N}} = npq \frac{N-n}{N-1}$$

que es la varianza de la hipergeométrica.

## INTRODUCCIÓN A LA ESTIMACIÓN BAYESIANA DE PROPORCIONES

En un gran número de fenómenos que pueden incluirse dentro de los llamados procesos de Bernoulli el parámetro  $p$  (Probabilidad de obtener un éxito en la prueba) puede entenderse como la proporción de individuos que poseen cierta característica .Pensemos en las situaciones en las que las pruebas implicadas en el proceso consistan en la extracción aleatoria (muestreo aleatorio) de individuos de una cierta población y la contemplación de si poseen o no una cierta característica

En muchas situaciones prácticas la proporción con la que se da una característica en una población nos es desconocida y sin embargo puede resultar necesario determinarla o "estimarla". Para ello podemos considerar la realización de una serie de ensayos consistentes en la extracción de individuos y la determinación de cuántos de ellos poseen, la característica. Si no muestreamos a la totalidad de los individuos de la población la proporción no podrá determinarse con absoluta certeza.

Sin embargo, las técnicas estadísticas nos pueden ayudar a estimar la proporción desconocida con un cierto grado de probabilidad. Básicamente éste es un caso particular de un tipo de técnica de inferencia estadística que se conoce con el nombre de estimación.

Podemos utilizar para la estimación únicamente la información suministrada por la muestra (Estimación clásica) o bien podemos utilizar además de la información muestral otros tipos de informaciones no muestrales, que podrían incluir experiencias anteriores, apreciaciones de expertos, ideas de tipo subjetivo, la opinión del investigador o la del "jefe" o el "cliente". En este segundo caso es necesario utilizar métodos Bayesianos de Estimación.

No pretendemos ahora tratar con profundidad las técnicas de la estadística Bayesiana, ni siquiera las técnicas de estimación, en general, sino sólo hacer una pequeña introducción a su utilidad.

Los métodos Bayesianos de inferencia reciben este nombre por que son capaces de sintetizar la información muestral y la llamada "información a priori" (no muestral) utilizando el Teorema de Bayes.

El primer supuesto de la estimación Bayesiana es que la información inicial que se dispone sobre el parámetro que se quiere estimar (en nuestro caso el parámetro  $p$  -la proporción de una característica- puede, expresarse a través de una cierta distribución de probabilidad que se llama distribución a priori o distribución inicial.

Puede considerarse que el parámetro puede tomar un conjunto numerable de valores posibles o bien que puede tomar valores comprendido en un cierto intervalo o en toda la recta real. En el primer caso la distribución a priori será discreta y en el segundo continua.

Uno de los mayores problemas de la estadística bayesiana es, habitualmente, el hecho de poder construir la distribución a priori a partir de la información inicial, pero aquí no nos plantearemos el problema.

Considerando la información disponible antes de realizar ninguna experiencia, la estimación del parámetro  $p$  deberá realizarse a partir de la distribución a priori. La manera de obtener un valor concreto para la estimación del parámetro es algo que debe plantearse, en general, considerado una cierta función de pérdida asociada al error de la estimación. Aquí consideraremos la función de pérdida más habitual: la pérdida cuadrática. La estimación que minimiza la pérdida cuadrática es la media de la distribución

Así pues antes de hacer ninguna experiencia consideraremos la distribución a priori de  $p$  como toda la información disponible; y el mejor resumen de esta información, y por tanto la estimación inicial: la media de la distribución a priori:

$$\hat{p} = E[x]$$

Esta primera estimación de  $p$  puede mejorarse utilizando la información muestral: utilizando los resultados obtenidos, en cierta experiencia consisten en la extracción aleatoria de algunos individuos de la población.

Supongamos que realizamos una muestra de tamaño  $n$  (extraemos  $n$  individuos) y obtenemos  $x$  resultados del tipo que nos interesa. Dependiendo de lo que valga el parámetro  $p$  el resultado obtenido será más o menos verosímil. Esa verosimilitud de la muestra nos vendrá dada por la probabilidad de obtener  $x$  resultados éxito en  $n$  pruebas en función de los posibles valores del parámetro  $p$  desconocido: Así

$$L(x) = P(x/p).$$

Para cada posible valor del parámetro  $p$  esta probabilidad podrá calcularse. No será otra cosa que la función de cuantía de  $x$  en una distribución dicotómica, Binomial, geométrica, Binomial Negativa, o Hipergeométrica (incluso Poisson, aunque ésta sea un proceso de observación) con parámetro  $p$  el valor de cada una de las alternativas. Se tratará de una distribución dicotómica, binomial, etc según las condiciones en las que se realice el muestreo (una o varias pruebas, con reposición o no, un número fijo de extracciones, o extracciones hasta que se produzcan 1 u otro número fijo de éxitos)

Para sintetizar la información a priori y la información muestral nos planteamos determinar las probabilidades de cada posible valor de  $p$  sabiendo que tras muestrear  $n$  individuos se han obtenido  $x$  éxitos:

$$P\left(\frac{p}{x}\right)$$

que aplicando el Teorema de Bayes será:

$$P\left(\frac{p_i}{x}\right) = \frac{P(p_i)P\left(\frac{x}{p_i}\right)}{\sum_{v_i} P(p_i)P\left(\frac{x}{p_i}\right)}$$

A la distribución que asigna a cada posible valor del parámetro la probabilidad de ese el valor condicionada a que la experiencia nos ha dado  $x$  éxitos en  $n$  pruebas se la conoce como distribución a posteriori o distribución final.

Esta distribución nos da toda la información disponible acerca del parámetro desconocido  $p$ , tanto la inicial como la empírica. Y a partir de ella podremos realizar una segunda estimación mejorada del valor del parámetro que será la media de la distribución a posteriori:

$$\hat{p} = E\left[\frac{p}{x}\right]$$

Ejemplo: Supongamos que la proporción de personas que no tienen teléfono en su casa es desconocida. Pero que basándonos en los datos de otras ciudades similares podemos suponer que se encuentra entre 0,05 y 0,01 con las siguientes probabilidades asociadas.

$p_i$	$P(p_i)$
0,01	0,20
0,02	0,20
0,03	0,30
0,04	0,15
0,05	0,15

en principio la estimación inicial sería

$$\hat{p} = E[p] = 0,0285 = 2,85\%$$

Para mejorar esta información realizamos una encuesta al azar preguntando a las personas si tienen teléfono. Resultando que de 20 preguntadas sólo una no tenía teléfono.

La probabilidad de que de 20 personas 1 no tenga teléfono (como la población es muy grande no importa que no haya reemplazamiento) nos vendrá dada por la función de cuantía (para  $X=1$ ) en una  $B\{20,p\}$ , (siendo  $p$  la proporción de personas .que no tienen teléfono):

$$P(x = 1/p) = \binom{20}{1} p^1 (1-p)^{19} = 20p(1-p)^{19}$$

De manera que la verosimilitud de este resultado para cada posible valor de  $p$  será:

$p_i$	$P\left(x = 1/p_i\right)$
0,01	$20 \cdot (0,01) \cdot (0,99)^{19} = 0,1652338$
0,02	$20 \cdot (0,02) \cdot (0,98)^{19} = 0,2724931$
0,03	$20 \cdot (0,03) \cdot (0,97)^{19} = 0,3363678$
0,04	$20 \cdot (0,04) \cdot (0,96)^{19} = 0,3683353$
0,05	$20 \cdot (0,05) \cdot (0,95)^{19} = 0,3773535$

y a partir de las probabilidades a priori y de las verosimilitudes podremos calcular las probabilidades a posteriori, aplicando el teorema de Bayes de la siguiente manera

$$P\left(\frac{p_1}{x=1}\right) = P\left(\frac{0,01}{x=1}\right) = \frac{P(p_1)P\left(\frac{x}{p_1}\right)}{\sum_{vi} P(p_i)P\left(\frac{x}{p_i}\right)} = \frac{0,20 \cdot 0,033 \cdot 0,1652338}{(0,2 \cdot 0,1652338) + (0,2 \cdot 0,2724931) + (0,3 \cdot 0,3363678) + (0,15 \cdot 0,3683353) + (0,15 \cdot 0,3773535)} = \frac{0,03304}{0,30029} = 0,1100245$$

para el segundo valor tendríamos

$$P\left(\frac{p_2}{x=1}\right) = P\left(\frac{0,01}{x=1}\right) = \frac{P(p_2)P\left(\frac{x}{p_2}\right)}{\sum_{vi} P(p_i)P\left(\frac{x}{p_i}\right)} = \frac{0,054498}{0,30029} = 0,1814752$$

siendo , los anteriores y el resto que forman la distribución a posteriori de la probabilidad de tener teléfono en casa , los siguientes valores:

$p_i$	$P\left(\frac{p_i}{x=1}\right)$
0,01	0,1100425
0,02	0,1814752
0,03	0,3360217
0,04	0,1839782
0,05	0,1884826

así la estimación mejorada, tras la realización del proceso bayesiano será la media de la distribución a posterior

$$\hat{p} = E\left[\frac{p}{x=1}\right] = 0,03159383 = 3,15\%$$

Una vez realizado un proceso Bayesiano la distribución a posteriori obtenida puede considerarse la información disponible en ese momento y plantearse realizar otro nuevo ensayo para mejorar de nuevo la estimación no habría más que considerar la distribución final obtenida como distribución a priori del nuevo proceso y repetir el planteamiento con una nueva información muestral.