UNIVERSIDAD NACIONAL
DE MAR DEL PLATA

FACULTAD DE CIENCIAS
ECONOMICAS Y SOCIALES

APROXIMACIONES DE DISTRIBUCIONES DE PROBABILIDAD: ENFOQUE EMPÍRICO

Carrera: Licenciatura en Economía
Cátedra: Estadística Metodológica

Autor: Mariano Morettini
Profesor Adjunto

Fecha: septiembre de 2013

Mariano Morettini
Profesor Adjunto – Estadística Metodológica
Septiembre de 2013

APROXIMACIONES DE DISTRIBUCIONES DE PROBABILIDAD:

ENFOQUE EMPÍRICO

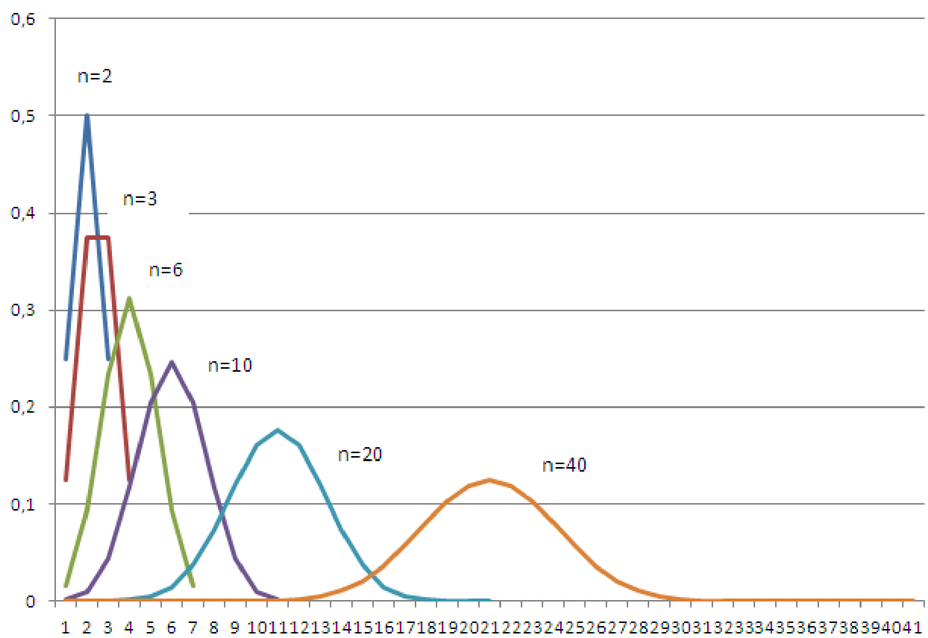
1) Aproximación de la Distribución Binomial a la Distribución Normal.

Cuando la cantidad de experimentos tiende a infinito (generalmente se considera que esto sucede cuando n≥30), la distribución binomial tiende a la distribución normal, siempre y cuando p y q no sean ninguna menor a 0,1.

Veamos qué sucede si calculamos, por distribución binomial, la probabilidad de s con *n* de 2, 3, 6, 10, 20 y 40 y con p=0,5; p=0,2; p=0,1 y p=0,05.

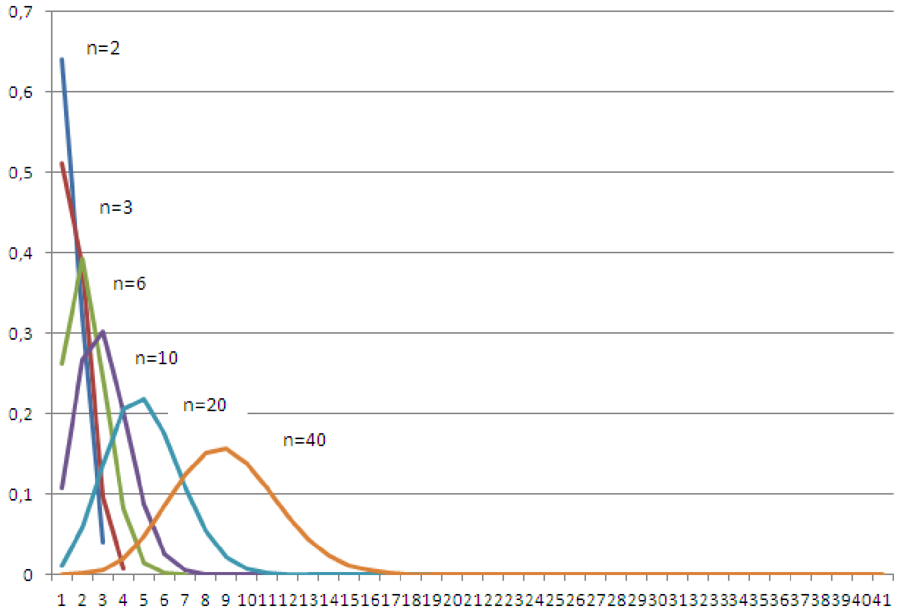
Es necesario destacar que, al ser la binomial una distribución discreta, la función de probabilidad o frecuencia debe graficarse con “bastones”. Sin embargo, como el propósito del presente trabajo es compararla con la distribución normal, que es continua, graficaremos la función de probabilidad binomial como si ésta fuera continua, pero, reiteramos, a los solos efectos comparativos.

Caso 1: Distribución Binomial con p=0,5 y n= 2, 3, 6, 10, 20 y 40



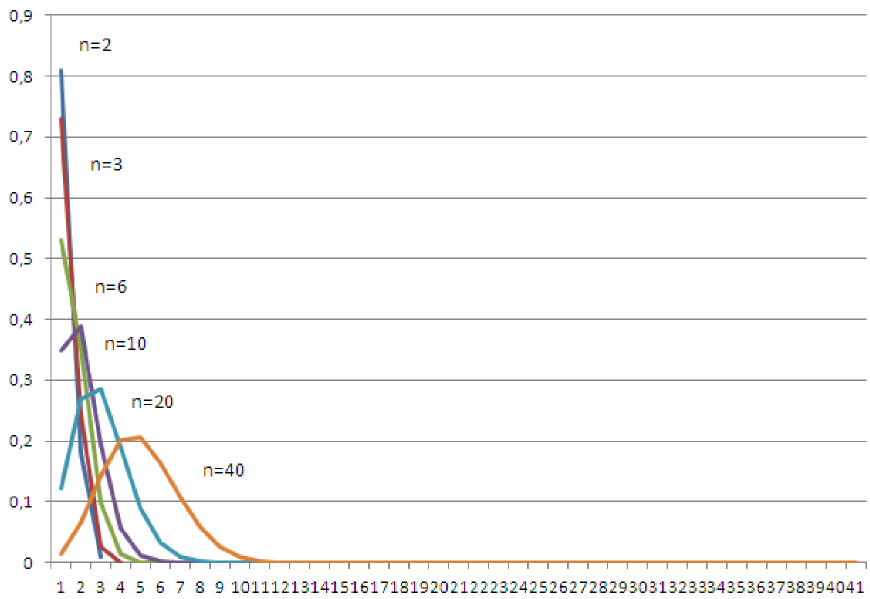
Se puede observar que cuando p=05, la distribución binomial presenta una forma similar a la normal para una cantidad de experimentos baja. Ya a partir de n=10 puede verse una similitud entre ambas distribuciones, y para n=40 ya es prácticamente igual la binomial a la normal.

Caso 2: Distribución Binomial con p=0,2 y n= 2, 3, 6, 10, 20 y 40



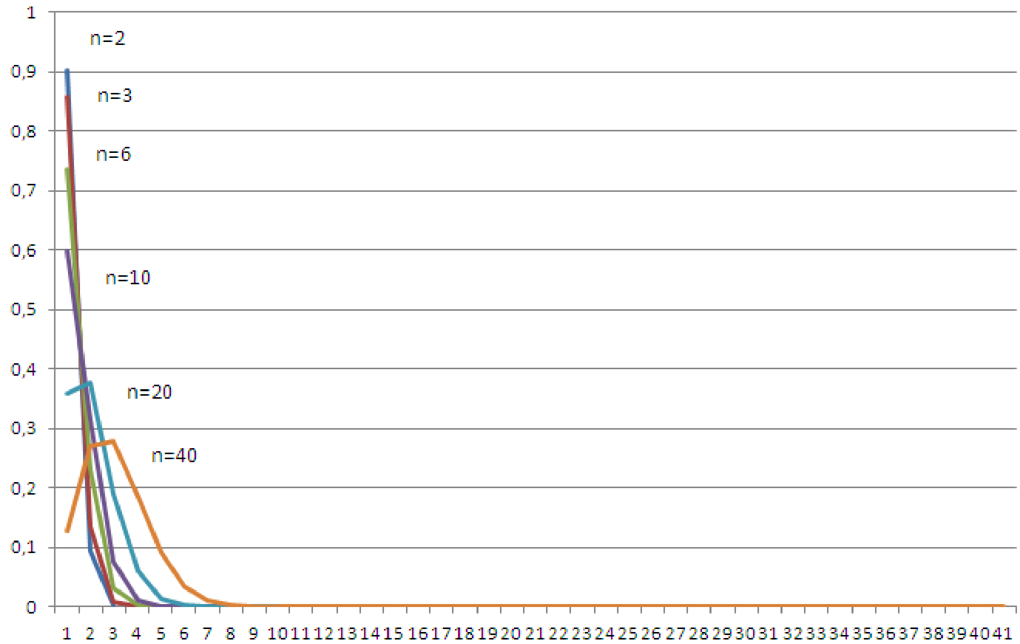
En este caso, p es sensiblemente menor al caso anterior, y puede observarse que para n=10 vemos que ya no es tan similar la distribución binomial a la normal. Sin embargo, para n=40 sí se asemejan ambas, aunque esta similitud sea menor a la del caso anterior para igual cantidad de experimentos.

Caso 3: Distribución Binomial con p=0,1 y n= 2, 3, 6, 10, 20 y 40



Este caso es el límite que, un tanto arbitrariamente, establecimos para aproximar la distribución binomial con la normal, es decir, el caso en que p=0,1. Nótese que para n=10, donde si p=0,5 ya podemos vislumbrar una similitud entre binomial y normal, ahora, que p=0,1, la forma de la función de frecuencia binomial no se parece en nada a la normal. Y para n=40 ya podemos suponer una aproximación, pero no tan clara como en los casos anteriores.

Caso 4: Distribución Binomial con p=0,05 y n= 2, 3, 6, 10, 20 y 40



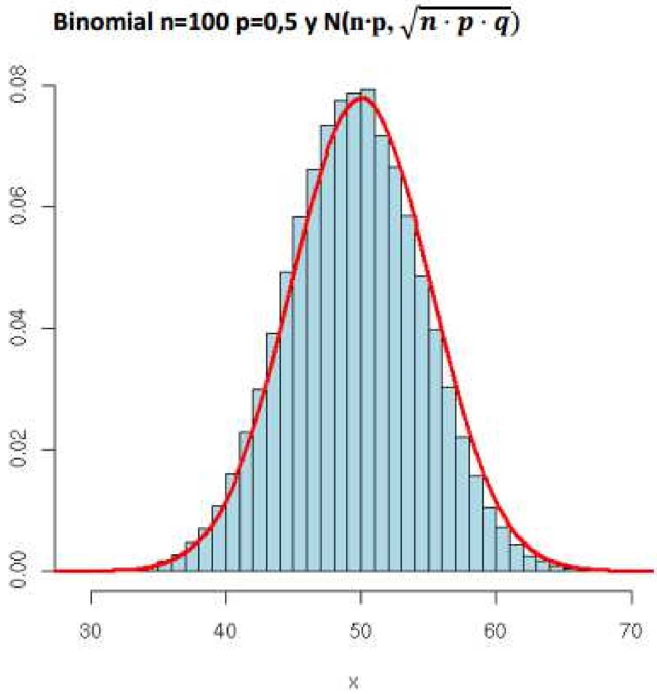
En este último caso, donde p=0,05 es menor al límite impuesto de 0,1, puede verse que incluso para n≥30 la distribución binomial no se asemeja lo suficiente a la distribución normal, con lo cual la aproximación no sería adecuada.

Como conclusión de éstos diferentes casos, podemos decir que, desde un punto de vista empírico, se puede aproximar la distribución binomial a la distribución normal, es decir, bajo determinados supuestos, en lugar de realizar los cálculos de probabilidades (sobre todo los referentes a probabilidades acumuladas) con la distribución binomial, a fin de simplificar el trabajo de cálculo, se puede utilizar la distribución normal. Y esta aproximación sería lo suficientemente buena si n≥30 y si 0,1≤p≤0,9. En términos matemáticos:

$$\lim_{n \to \infty} B(n,p) = N\left(np, \sqrt{npq}\right)$$

Sin embargo, debe quedar claro que si *n* es infinito, no hay coincidencia entre la distribución binomial y la normal, ya que no solo depende de la cantidad de experimentos la buena o mala aproximación de distribuciones, sino que también depende de p.

Para poner un ejemplo de aproximación, compararemos las funciones de probabilidad de las distribuciones Binomial y Normal siguientes: $B(100, 0{,}5)$ y $N(100 * 0{,}5, \sqrt{100 * 0{,}5 * ,05})$.



Puede fácilmente comprobarse que la aproximación de la binomial por la normal, en este caso, es excelente.

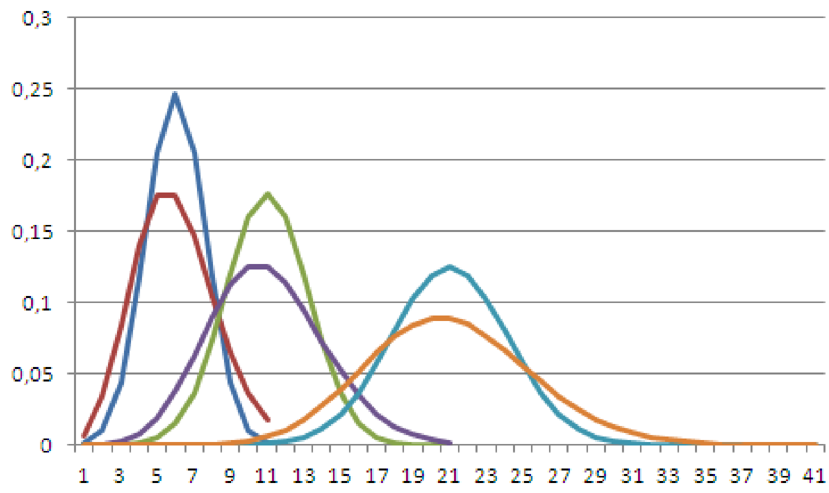
2) Aproximación de la Distribución Binomial a la Distribución de Poisson.

Veremos ahora la comparación entre las probabilidades obtenidas con la distribución binomial y la distribución de Poisson. Ya hemos visto que la distribución de Poisson es un límite de la distribución binomial cuando *n* tiende a infinito y p tiende a 0 (o a 1). Veremos el comportamiento de ambas distribuciones para p=0,5; p=0,25 y p=0,1 y, para cada caso, n=10; n=20 y n=40.

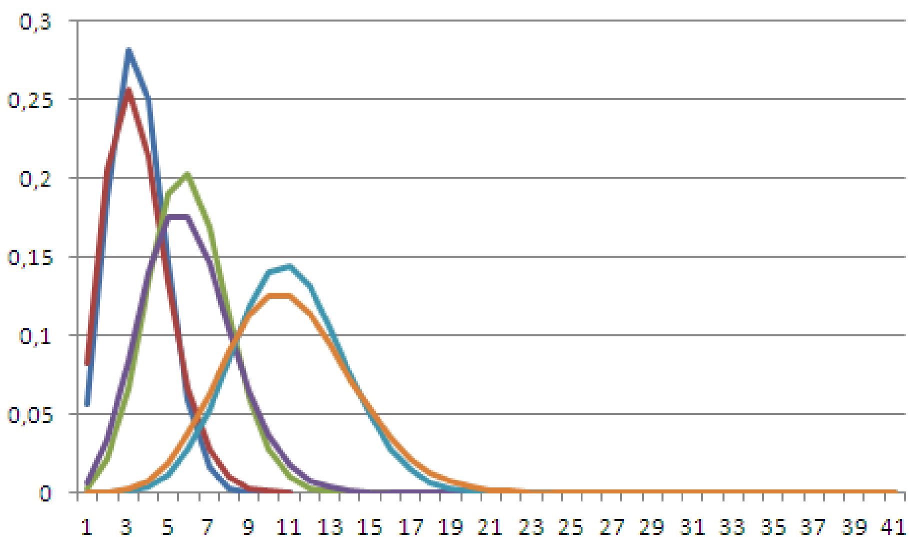
Caso 1: Distribución Binomial y de Poisson con p=0,5 y n= 10, 20 y 40

El primer par de curvas es el correspondiente a n=10; el segundo par de curvas es para n=20 y el tercero para n=40. En todos los casos, la distribución que presenta mayor kurtosis es la correspondiente a la binomial.

Puede observarse en el gráfico que, como p es sensiblemente diferente de 0,1 (y de 0,9), la diferencia entre las distribuciones binomial y de Poisson se mantiene, aún cuando aumenta n, por lo que nunca es una buena aproximación una de la otra.



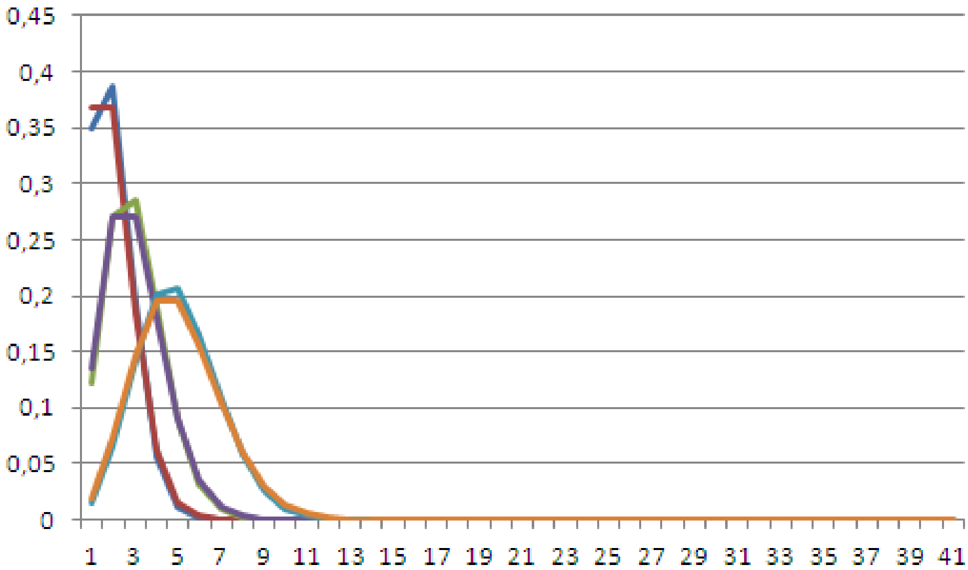
Caso 2: Distribución Binomial y de Poisson con p=0,25 y n= 10, 20 y 40



Cuando p se reduce, en este caso a 0,25, vemos que la distribución binomial se aproxima más a la de Poisson que en el caso anterior, pero siguen subsistiendo diferencias aun cuando *n* es mayor a 30.

Caso 3: Distribución Binomial y de Poisson con p=0,1 y n= 10, 20 y 40

Vemos en este caso, donde p=0,1, que la distribución binomial arroja valores de probabilidad muy similares a los de la distribución de Poisson, y que la similitud es mayor cuando aumenta n. Para el caso de n=40 vemos que las distribuciones son casi equivalentes.



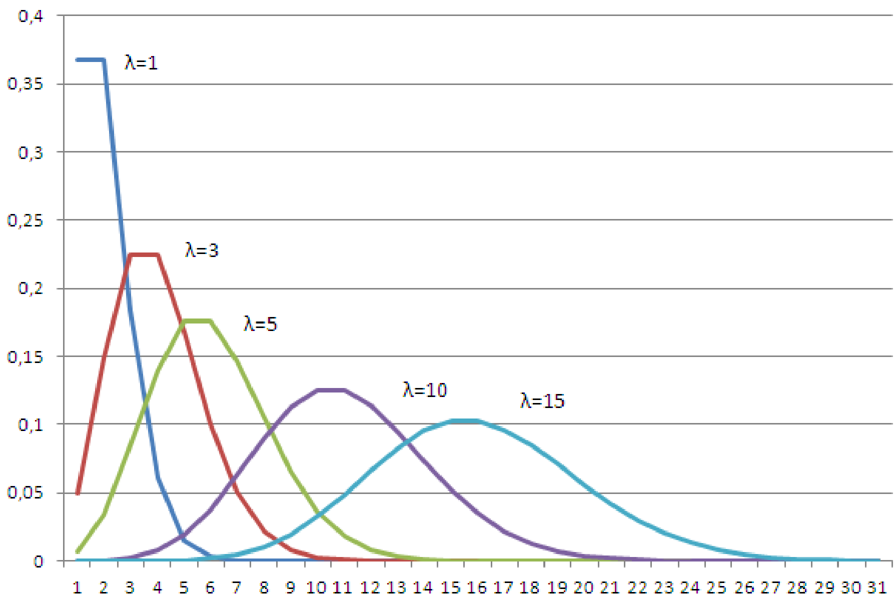
Podemos concluir, de éstas 3 gráficas, que empíricamente la aproximación de la Binomial a Poisson es buena para p≤0,1 y n≥30. Para estos casos, consideramos λ=np. Alternativamente, puede aproximarse la Binomial a Poisson cuando p≥0,9 y n≥30 considerando λ=n(1-p).

3) Aproximación de la Distribución de Poisson a la Distribución Normal.

Para el caso de la Distribución de Poisson, cuando λ≥10, la forma de esta distribución se asemeja lo suficiente a la Distribución Normal como para que puede utilizarse ésta última como aproximación.

Veremos a continuación la Distribución de Poisson para λ=1; λ=3; λ=5; λ=10 y λ=15. Puede verse en ese gráfico que para λ=10 ya la forma de la Distribución de Poisson se asemeja bastante a la Normal. Es de destacar, sin embargo, que, como sucede con la Distribución Binomial, la de Poisson no coincide con la Normal cuando *n* es infinito.

Para realizar la aproximación, debe considerarse que la Distribución Normal tendrá $E(x) = \lambda$ y $\sigma(x) = \sqrt{\lambda}$.



4) Aproximación de la Distribución Hipergeométrica a la Distribución Binomial.

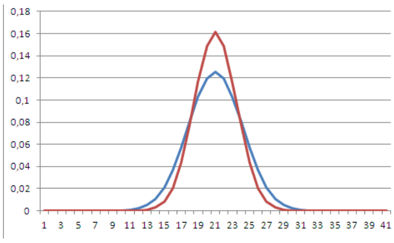
La diferencia entre las distribuciones hipergeométrica y binomial radica en que la probabilidad de éxito es variable en la primera y constante en la segunda. Una forma de convertir un experimento de probabilidad variable a probabilidad constante es realizarlo sin reposición, para el primer caso, y con reposición en el segundo caso.

La esperanza matemática de la cantidad de éxitos a obtener repitiendo el experimento *n* veces en la distribución binomial es: E(x)=np, mientras que en la hipergeométrica es $E(X) = n\frac{N_p}{N}$. Como Np representa la cantidad de elementos en la población que cumple con los requisitos de éxito y *N* es la cantidad total de individuos en la población, surge que el cociente entre ambos es la probabilidad de éxito al realizar el primer experimento, que es igual a p si los experimentos se realizan con reposición. Por eso, la esperanza matemática de ambos modelos coinciden.

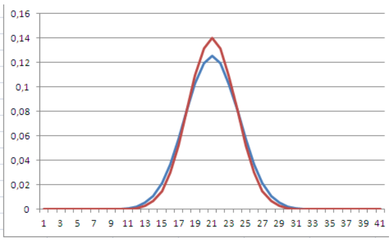
Sin embargo, el desvío standard será de $\sigma(X) = \sqrt{npq}$en la distribución binomial, mientras que en la hipergeométrica será $\sigma(X) = \sqrt{n\frac{N_p}{N}\frac{N_q}{N}\frac{N-n}{N-1}}$.

Realizando análogo razonamiento al del caso de la esperanza matemática, podemos concluir que la diferencia de las expresiones radica en el cociente $\sqrt{\frac{N-n}{N-1}}$ que se añade en la variable hipergeométrica, y que surge por el hecho de trabajar sin reposición.

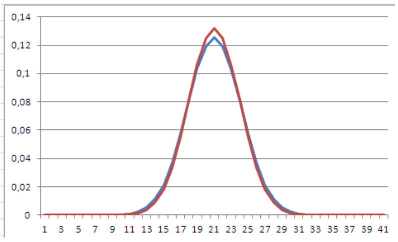
Esto quiere decir que ambas distribuciones poseen la misma esperanza matemática, pero la dispersión será mayor en el caso de la binomial. Sin embargo, puede fácilmente observarse que cuanto menor sea *n* y mayor sea *N* o, en términos más precisos, canto menor sea *n* en relación a N, más cercano a 1 será el cociente $\sqrt{\frac{N-n}{N-1}}$, por lo que el desvío standard de la distribución hipergeométrica tenderá a igualar al de la binomial cuanto menor sea *n* en relación a N. En términos prácticos puede considerarse que esto sucede cuando $\frac{n}{N} < 0{,}05$, es decir, cuando el tamaño de la muestra (o la cantidad de experimentos) es inferior al 5% de la población. Veamos algunas gráficas de funciones de probabilidad:



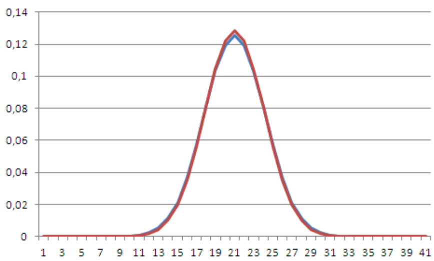
$\frac{n}{N} = 0{,}4$



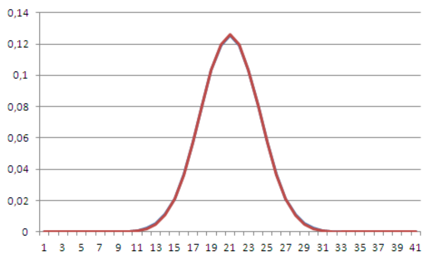
$\frac{n}{N} = 0{,}2$



$\frac{n}{N} = 0{,}1$



$\frac{n}{N} = 0{,}05$

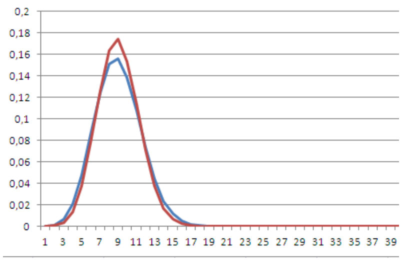


$\frac{n}{N} = 0{,}01$

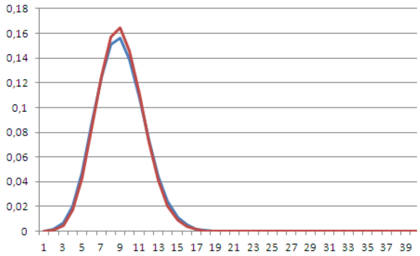
Distribuciones binomial e hipergeométrica para p=0,5

Se evidencia en estos gráficos que ambas distribuciones poseen la misma esperanza matemática y que, al tener la hipergeométrica menor dispersión, es la curva de mayor kurtosis. Al mismo tiempo, puede observarse cómo la diferencia entre ambas distribuciones se va reduciendo a medida que n pasa a ser un porcentaje cada vez más pequeño de N. Para el caso del 5%, límite práctico usualmente establecido, la similitud entre ambas distribuciones es notable, y para el caso del 1% ambas curvas prácticamente coinciden.

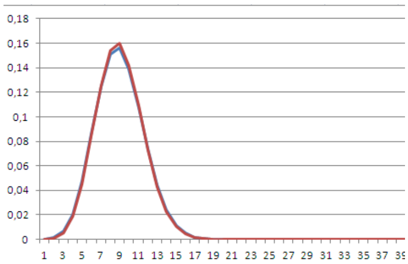
En los gráficos siguientes mostraremos que la aproximación entre ambas distribuciones es igualmente buena, cuando decrece n en relación a N, para el caso en que la probabilidad de éxito difiera significativamente de 0,5.



$\frac{n}{N} = 0{,}2$



$\frac{n}{N} = 0{,}1$



$\frac{n}{N} = 0{,}05$

Distribuciones binomial e hipergeométrica para p=0,2

Vemos aquí, que el hecho de considerar una probabilidad de éxito significativamente diferente de 0,5, lo que consiguió fue, principalemente, desplazar las funciones de probabilidad hacia la izquierda (porque se redujo p y, en consecuencia, se reduce la esperanza matemática), pero la dispersión sigue siendo menor para el caso hipergeométrico, y se aproxima cada vez más a la del modelo binomial a medida que n se reduce en relación a N, como habíamos visto en el caso anterior.