Apuntes Unidad 3

Aproximación normal de la distribución binomial

FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS UNIVERSIDAD DE CHILE

CMM Centro de Modelamiento Matemático

El cálculo de probabilidades acumuladas en una distribución binomial, en general, puede tornarse largo y engorroso, especialmente si el número de éxitos y el número de intentos son grandes. Dado esto, a continuación aprenderás a utilizar Microsoft Excel y Google Sheets para facilitar este tipo de cálculos.

CÁLCULO DE PROBABILIDADES ACUMULADAS UTILIZANDO HERRAMIENTAS TECNOLÓGICAS

Para calcular las probabilidades acumuladas en una distribución binomial en Microsoft Excel debemos introducir en una celda el siguiente comando:

=DISTR.BINOM.N()

Dentro del paréntesis se ingresan en orden y separadas por un punto y coma (;), el número de éxitos (k), ensayos (n), probabilidad de éxito (p) y finalmente se ingresa "VERDADERO" o "FALSO", dependiendo de lo que busquemos calcular. Si lo que buscamos calcular es una probabilidad acumulada (que la variable aleatoria tome como máximo cierto valor) debemos ingresar "VERDADERO", mientras que si queremos calcular la probabilidad de que la variable aleatoria tome exactamente un valor debemos ingresar "FALSO".

De la misma manera, para calcular probabilidades de una variable aleatoria que sigue una distribución binomial en Google Sheets, el procedimiento es análogo al anterior, solo que el comando a utilizar es:

=BINOM.DIST()

Observación: Luego de probar varios cálculos utilizando las fórmulas anteriores, es posible observar que mientras más "grande" sea el valor de n, mejor será la aproximación de una distribución binomial por medio de una curva normal. Lo que se entiende por un n "grande" es arbitrario, pero en general un $n \geq 30$ es considerado lo suficientemente "grande" para una buena aproximación por medio de la curva normal.

APROXIMACIÓN NORMAL DE LA DISTRIBUCIÓN BINOMIAL

Una distribución binomial permite modelar experimentos en donde se está interesado en observar un número particular de "éxitos" en un número definido de "intentos" independientes entre sí.

Para una variable aleatoria X que sigue una distribución binomial con parámetros n y p, podemos calcular la probabilidad acumulada hasta un valor k mediante la siguiente expresión:

$$\begin{aligned} F_X(k) =& \mathrm{P}(X \leq k) \\ =& \mathrm{P}(X = 0) + \ldots + \mathrm{P}(X = k) \end{aligned}$$

El cálculo de probabilidades acumuladas con la distribución binomial puede resultar muy largo de calcular, especialmente si el número de éxitos y el número de intentos son grandes. Por eso mismo, resulta interesante realizar una aproximación normal de la distribución binomial.

Ahora, es posible demostrar que si X sigue una distribución binomial con parámetros n y p, en la medida que el valor de n sea grande (mayor que 30), la probabilidad acumulada de la variable X se puede estimar a partir de la probabilidad acumulada de una variable aleatoria Y que sigue una distribución normal.

$$F_X(k) = \mathrm{P}(X \leq k) \approx \mathrm{P}(Y \leq k)$$

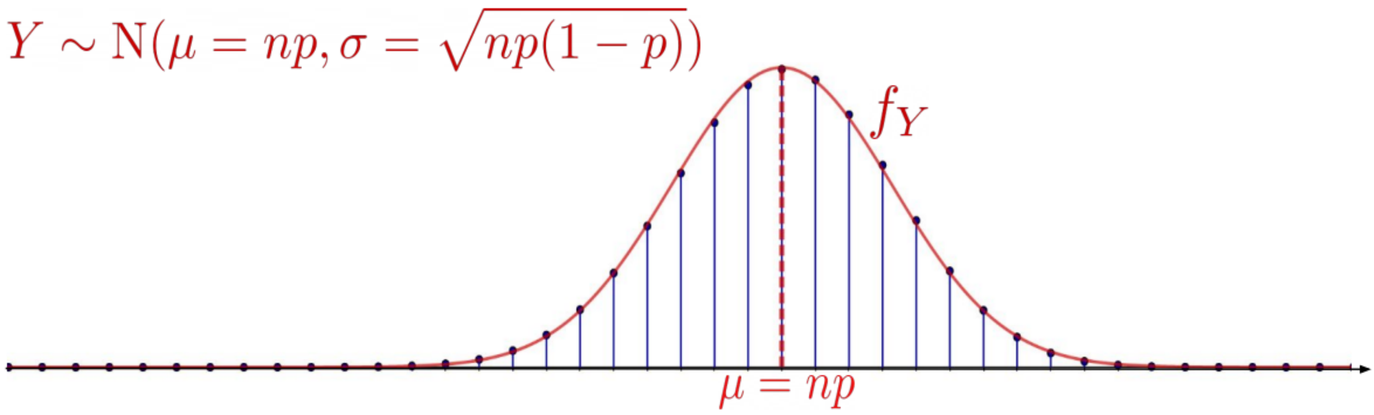
Gráficamente, esto se puede interpretar de la siguiente manera:

Mientras más grande sea n mejor será la aproximación. Recordemos que el valor esperado y la desviación estándar de la variable X son:

$$\mathrm{E}(X) = np \qquad\qquad \mathrm{DE}(X) = \sqrt{np(1-p)}$$

Luego, los parámetros de la distribución normal que aproximará las probabilidades acumuladas de la variable aleatoria X están definidos por el valor esperado y la desviación estándar de la distribución binomial de origen:

$$Y \sim \mathrm{N}(\mu, \sigma) \qquad \mu = np \qquad \sigma = \sqrt{np(1-p)}$$



Ahora, es natural preguntarse, ¿Cómo es posible aproximar normal una distribución binomial si una variable aleatoria es continua y la otra es discreta? A continuación, veremos cómo responder a esta interrogante.

CORRECCIÓN POR CONTINUIDAD

Consideremos una variable aleatoria discreta, X, que sigue una distribución binomial con parámetros n y p. Recordemos que, si n es grande, en vez de utilizar una suma de probabilidades binomiales para calcular la probabilidad acumulada $F_X(k)$, podemos aproximar este valor por $P(Y \leq k)$, donde Y es una variable aleatoria que sigue una distribución normal:

$$\mathrm{N}\left(\mu = np, \sigma = \sqrt{np(1-p)}\right)$$

Tal y como mencionamos anteriormente, notemos que la distribución binomial está definida para una variable aleatoria discreta, a diferencia de la distribución normal que está definida para un intervalo en los números reales. De lecciones anteriores sabemos que la probabilidad de que una variable aleatoria continua tome un valor constante es cero, a diferencia de una variable aleatoria discreta. Por este motivo, para aproximar una variable aleatoria que sigue una distribución binomial por una normal, usualmente se realiza la siguiente corrección:

$$\underbrace{P(X \leq k)}_{\text{X sigue una distribución binomial}} \approx \underbrace{P(Y \leq k + \overbrace{0{,}5}^{\text{Corrección}})}_{\text{Y sigue una distribución normal}}$$

A la acción de sumar $0,5$ se denomina corrección por continuidad. Para ver el efecto de esta acción, consideremos una variable aleatoria X que sigue una distribución binomial con $n = 50$ y $p = 0,6$, la que podemos aproximar mediante una variable aleatoria Y que sigue una distribución normal con parámetros:

$$\mu = np = 50 \cdot 0{,}6 = 30$$

$$\sigma = \sqrt{np(1-p)} = \sqrt{50 \cdot 0{,}6 \cdot 0{,}4} = \sqrt{12}$$

Es decir, $Y \sim N(\mu = 30,\ \sigma = \sqrt{12})$. La siguiente tabla muestra el valor de la probabilidad $P(X \leq 32)$ y de las aproximaciones usando la variable aleatoria Y, sin y con usar la corrección por continuidad:

Probabilidad exacta	Probabilidad sin usar corrección por continuidad.	Probabilidad usando corrección por continuidad.
$Binomial(n, p)$	$Y \sim N(\mu, \sigma)$	$Y \sim N(\mu, \sigma)$
$P(X \leq 32) = 0{,}763$	$P(X \leq 32) \approx P(Y \leq 32) = 0{,}718$	$P(X \leq 32) \approx P(Y \leq 32 + 0{,}5) = P(Y \leq 32{,}5) = 0{,}765$

Como se puede observar, la aproximación usando la corrección por continuidad entrega un valor mucho más cercano a la probabilidad $P(X \leq 32)$.

Curso: Probabilidad y estadística descriptiva
Unidad 3 : Situaciones o fenómenos que se modelan por medio de las distribuciones binomial y normal.
Tema: Teorema Central del Límite.
Contenido: Aproximación normal de la distribución binomial

SÍNTESIS

- En esta lección interpretamos y analizamos la información obtenida por medio de una encuesta de respuesta dicotómica a fin de aplicar la distribución binomial a un problema real.

- También, aprendimos a calcular las probabilidad puntual de que una variable aleatoria binomial X tome exactamente un valor k, y la probabilidad acumulada de que X tome valores menores o iguales a k utilizando una hoja de cálculo.

    - En Microsoft Excel utilizamos el comando =DISTR.BINOM.N(), mientras que en Google Sheets utilizamos el comando =BINOM.DIST() Ambos comandos necesitan los mismo 4 argumentos en el siguiente orden: (número de éxitos; ensayos; probabilidad de éxito; VERDADERO o FALSO)

    - El último argumento es falso si estamos interesados en una probabilidad puntual de que $X = k$, y es verdadero si estamos interesados en calcular la probabilidad acumulada de que $X \leq k$.

- Finalmente, vimos de forma intuitiva que, al aumentar el valor de n, una distribución binomial se puede aproximar por una normal, escogiendo adecuadamente sus parámetros.

Considere la variable aleatoria discreta X que sigue una distribución binomial con parámetros n y p.

- Si el valor de n es grande podemos aproximar el valor de la probabilidad acumulada por el área bajo la curva de una variable aleatoria Y que sigue una distribución normal.

- Los parámetros de la distribución normal corresponden al valor esperado y a la desviación estándar de la distribución binomial de origen.

$$Y \sim \mathrm{N}(\mu = np, \sigma = \sqrt{np(1-p)})$$

- Una mejor aproximación de la probabilidad $P(X \leq k)$ se obtiene sumando $0,5$ al valor de k.

$$F_X(k) = \mathrm{P}(X \leq k) \approx \mathrm{P}(Y \leq k + 0{,}5)$$

- A la acción de sumar $0,5$ se denomina corrección por continuidad.