1. VARIABLE ALEATORIA DISCRETA - DISTRIBUCIÓN BINOMIAL

Una variable aleatoria X sigue una **distribución binomial** si cumple:

1) Estudiamos un experimento compuesto por **n experimentos independientes** entre sí.

2) El resultado de cada experimento elemental ha de admitir sólo dos categorías, a las que se denomina **éxito y fracaso**.
El valor de ambas posibilidades ha de ser constante en todos los experimentos, y se denotan como ***p* y *q*** respectivamente, o *p* y 1-*p* de forma alternativa.

3) Se designa por ***X* a la variable que mide el número de éxitos** que se han producido en los *n* experimentos.

Cuando se dan estas circunstancias, se dice que la variable *X* sigue una distribución de probabilidad binomial, y se denota **X ~ *B(n,p)***.

La función de probabilidad de la binomial es: $P(X=k)=\binom{n}{k}p^k q^{n-k}$ donde q = 1-p

Siendo $\binom{n}{k}$ un número combinatorio: $\binom{n}{k}=\frac{n!}{k!(n-k)!}$

Media: $\mu = \sum_{k=0}^{n} k\cdot P(X=k)=\sum_{k=0}^{n} k\cdot\binom{n}{k}p^k q^{n-k}=n\cdot p$

Varianza: $\sigma^2=\sum_{k=0}^{n} k^2\cdot P(X=k)$ - μ^2 = $n\cdot p\cdot q$ **Desviación típica:** $\sigma=\sqrt{\sigma^2}=\sqrt{n\cdot p\cdot q}$

2.1. DISTRIBUCIÓN BINOMIAL - EJERCICIO MODELO:

Ejercicio: La probabilidad de ganar un sorteo es 1/3. Suponiendo que se van a jugar cuatro sorteos:
a) Describe la variable aleatoria que cuenta el número de sorteos ganados. Justifica si se trata o no de una variable binomial.
b) Haz la tabla de la función de probabilidad.
c) Calcula la media y la varianza de esta variable.
d) Probabilidad de que gane la mitad de los sorteos.
e) Probabilidad de que gane más de la mitad de los sorteos.
f) Probabilidad de que gane algún sorteo.
g) Probabilidad de que pierda a lo sumo 1 sorteo

SOLUCIÓN:

a) Describe la variable aleatoria que cuenta el número de sorteos ganados. Justifica si se trata o no de una variable binomial:

X = n.º de sorteos ganados X sigue una distribución binomial con n = 4 y p = ⅓ porque son 4 experimentos aleatorios independientes e iguales con dos únicos resultados posibles "ganar" o "perder": X ~ B(4, 1/3)

b) Haz la tabla de la función de probabilidad

k (n.º de sorteos ganados	P(X=k)
0	$P(X=0)=\binom{4}{0}\left(\frac{1}{3}\right)^0\left(\frac{2}{3}\right)^{4-0}=\frac{16}{81}$
1	$P(X=1)=\binom{4}{1}\left(\frac{1}{3}\right)^1\left(\frac{2}{3}\right)^{4-1}=\frac{32}{81}$
2	$P(X=2)=\binom{4}{2}\left(\frac{1}{3}\right)^2\left(\frac{2}{3}\right)^{4-2}=\frac{24}{81}$
3	$P(X=3)=\binom{4}{3}\left(\frac{1}{3}\right)^3\left(\frac{2}{3}\right)^{4-3}=\frac{8}{81}$
4	$P(X=4)=\binom{4}{4}\left(\frac{1}{3}\right)^4\left(\frac{2}{3}\right)^{4-4}=\frac{1}{81}$

La suma de todas las probabilidades es 1.

c) Calcula a partir de la tabla la media y la varianza de esta variable:

Media: $\mu= \sum_{k=0}^{4} k\cdot P(X=k)=\frac{4}{3}$

Varianza: $\sigma^2=\sum_{k=0}^{4} k^2\cdot P(X=k)-\quad \mu^2 = \frac{8}{3}-\left(\frac{4}{3}\right)^2=\frac{24}{9}-\frac{16}{9}=\frac{8}{9}$

Desviación típica: $\sigma=\sqrt{\sigma^2}=\sqrt{\frac{8}{9}}=\frac{2\sqrt{2}}{3}$

d) Probabilidad de que gane la mitad de los sorteos:

$$P(X=2)=\binom{4}{2}\left(\frac{1}{3}\right)^2\left(\frac{2}{3}\right)^{4-2}=\frac{24}{81}=0{,}2963$$

e) Probabilidad de que gane más de la mitad de los sorteos:

$$P(X>2)=P(X=3)+P(X=4)=\binom{4}{3}\left(\frac{1}{3}\right)^3\left(\frac{2}{3}\right)^{4-3}+\binom{4}{4}\left(\frac{1}{3}\right)^4\left(\frac{2}{3}\right)^{4-4}=\frac{9}{81}=\frac{1}{9}=0{,}1111$$

f) Probabilidad de que gane algún sorteo:

$$P(X>0)=1-P(X=0)=1-\binom{4}{0}\left(\frac{1}{3}\right)^0\left(\frac{2}{3}\right)^{4-0}=1-\frac{16}{81}=\frac{65}{81}=0{,}8025$$

g) Probabilidad de que pierda a lo sumo 1 sorteo

Podemos definir la variable aleatoria Y = n.º de sorteos perdidos   Y ~ B(4, ⅔)

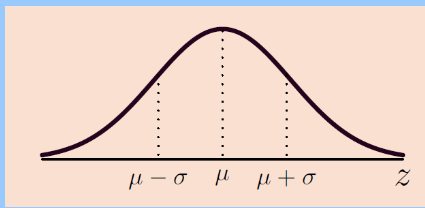
$$P(Y\leq 1)=P(Y=0)+P(X=1)=\binom{4}{0}\left(\frac{2}{3}\right)^0\left(\frac{1}{3}\right)^{4-0}+\binom{4}{1}\left(\frac{2}{3}\right)^1\left(\frac{1}{3}\right)^{4-1}=\frac{1}{81}+\frac{8}{81}=\frac{1}{9}=0{,}1111$$

2. DISTRIBUCIÓN NORMAL

La función de distribución normal juega un papel central en la estadística ya que, además de sus interesantes propiedades de reproductividad y de aproximación de otras distribuciones, sirve para modelizar una gran cantidad de situaciones prácticas.

Definición - Distribución Normal: Una variable aleatoria X se dice que tiene una distribución de probabilidades normal con media μ y varianza σ^2 y se denomina **N(μ, σ^2)** si su función de densidad es de la forma:

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-u)^2}{2\sigma^2}} \quad , \quad -\infty < x < +\infty$$



La media y varianza de una variable normal coinciden con los parámetros μ y σ^2 de la misma.
Su gráfica se llama **campana de Gauss**.

La probabilidad se determina con $P(a<x<b)=\int_a^b f(x)dx$ que corresponde al área de la región limitada por f(x), el eje x y las rectas x = a y x = b.

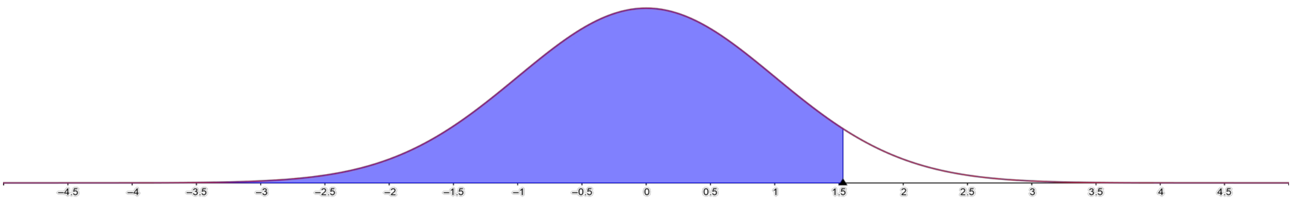
Se cumple: 1º) $P(-\infty<X<+\infty)=\int_{-\infty}^{+\infty} f(x)dx=1$ (El área bajo la curva es uno)

2º) P(X=a)= $\int_a^a f(x)dx=0$

La normal N(0; 1)
Primero estudiaremos la normal típica, que tiene media μ = 0 y desviación típica = 1, N(0; 1) y se designa con la letra Z. Para calcular las probabilidades P(Z ≤ z_0) se usa una tabla.

2.1. USO DE LA TABLA DE LA NORMAL N(0, 1)

Caso I: P(Z ≤ z_0) y z_0 es positivo.
Ejemplo: P(Z ≤ 1,53) buscamos en la tabla la fila del 1,5 y la columna del 0,03 obteniendo P(Z ≤1,53)=0,9370.



Fíjate que la campana de Gauss es simétrica respecto al eje Y con área total igual a 1, por tanto, P(Z ≤ 0) = 0,5.

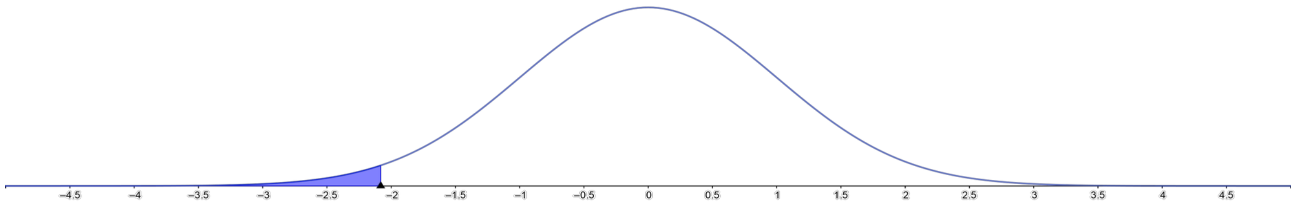
Caso II: $P(Z \le -z_0)$ y z_0 es positivo. $\boxed{P(Z \le -z_0) = 1 - P(Z \le z_0)}$

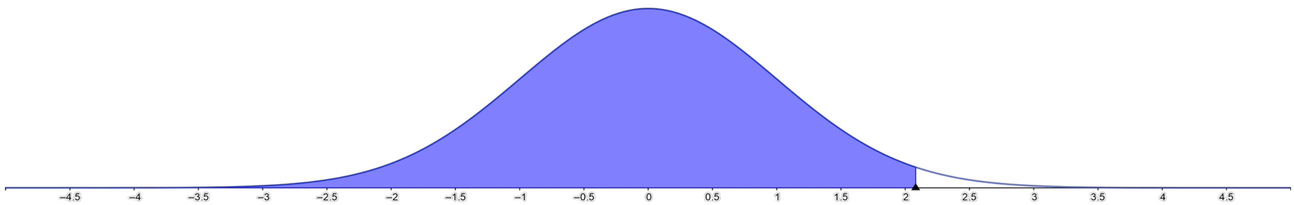
Ejemplo: $P(Z \le -2{,}08) = ?$

Utilizamos la simetría de la curva: $P(Z \le -2{,}08) = P(Z \ge 2{,}08) = 1 - P(Z < 2{,}08) = 1 - P(Z \le 2{,}08)$

Ten en cuenta que $\mathbf{P(Z \le z_0) = P(Z < z_0) + P(Z = z_0) = P(Z < z_0)}$ ya que por definición

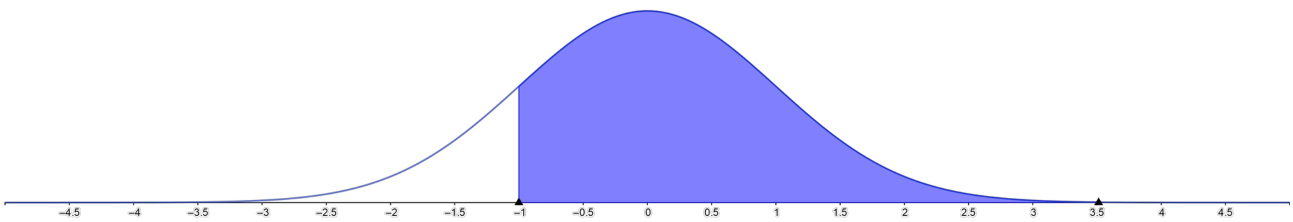
$$P(Z=z_0)=\int_{z_0}^{z_0} f(x)\,dx=0 \quad .$$





Buscamos en la tabla en la fila del 2 y la columna del 0,08 obteniendo $P(Z\le -2{,}08) = 1 – P(Z\le 2{,}08) = 1- 0{,}9812 = 0{,}0188$

Caso III: $\boxed{P(z_0 \le Z \le z_1) = P(Z \le z_1) - P(Z \le z_0)}$



Ejemplo: $P(1 \le Z \le 3{,}51) = P(Z \le 3{,}51) – P(Z \le 1) = 0{,}9998 – 0{,}8413 = 0{,}1585$

2.2. USO INVERSO DE LA TABLA N(0,1)

Nos dan la probabilidad y calculamos el valor de la variable z_0 que acumula dicha probabilidad.

Caso I: $P(Z \le z_0)=p \ge 0{,}5$ Hallar z_0 de la normal $Z \sim N(0, 1)$ tal que $P(Z \le z_0) = p$

Ejemplo: $P(Z \le z_0) = 0{,}7019$ buscamos en la tabla 0,7019 y nos fijamos que corresponde a la fila 0,5 columna 0,03. Por tanto, $z_0 = 0{,}53$.

Caso II: $P(Z \le z_0)=p < 0{,}5$ Para hallar z_0 debemos tener en cuenta que debe ser negativo, por tanto, $P(Z \le z_0) = 1- P(Z \le -z_0) = p$.

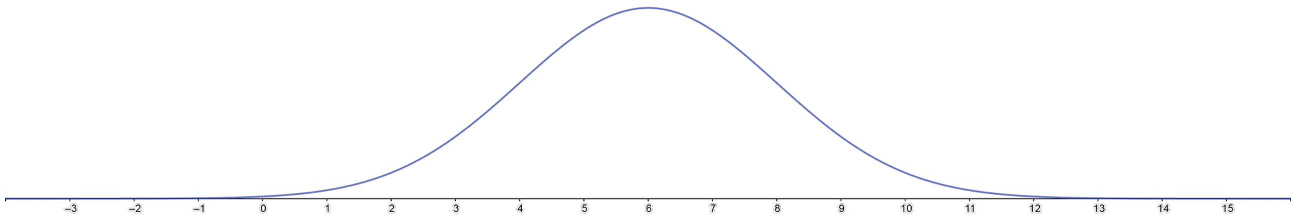
Ejemplo: $P(Z \le z_0) = 0{,}2877 \rightarrow P(Z \le -z_0) = 1 – 0{,}2877 = 0{,}7123 \rightarrow$ Buscamos en la tabla 0,7123 que se corresponde con la fila 0,5 y columna 0,06. Por tanto, $-z_0= 0{,}56$ y $z_0= -0{,}56$.

2.3. TIPIFICAR UNA VARIABLE NORMAL N (μ, σ) - EJERCICIO MODELO

Sea X una variable aleatoria normal $X \sim N(\mu, \sigma)$. Entonces, la variable Z = $(X - \mu) / \sigma$ cumple que $Z \sim N(0,1)$.

Ejemplo: Supongamos que una variable X, que describe la nota media de 2º de Bachillerato del alumnado de un instituto sigue una distribución normal de media 6 y desviación típica 2:

$$X \sim N(6, 2) \rightarrow \quad \frac{X-6}{2} \sim N(0,1)$$



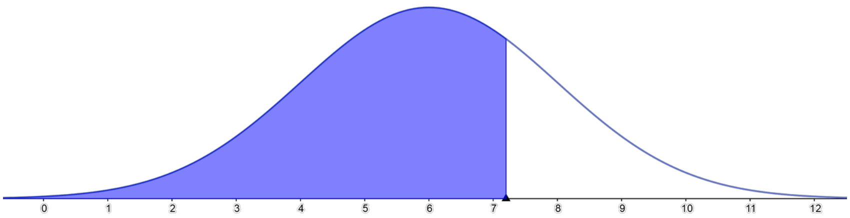
Elegido un alumno al azar, calculemos las probabilidades siguientes:

a) Tenga una nota media de un 7: Evidentemente, P (X = 7) = 0 al ser X una variable continua.

b) Tenga una nota media menor que 7,2:

$$P(X \le 7{,}2) \underbrace{=}_{tipificamos} P\left(\frac{X-6}{2} \le \frac{7{,}2-6}{2}\right)$$

$$P(Z \le 0{,}6) \underbrace{=}_{tabla\ N(0,1)} 0{,}7257$$



c) Tenga una nota media superior a 8:

$$P(X>8) = 1 - P(X \le 8) \underbrace{=}_{tipificamos} 1 - P\left(\frac{X-6}{2} \le \frac{8-6}{2}\right) = 1 - P(Z \le 1) \underbrace{=}_{tabla\ N(0,1)} 1 - 0{,}8413 = 0{,}1587$$

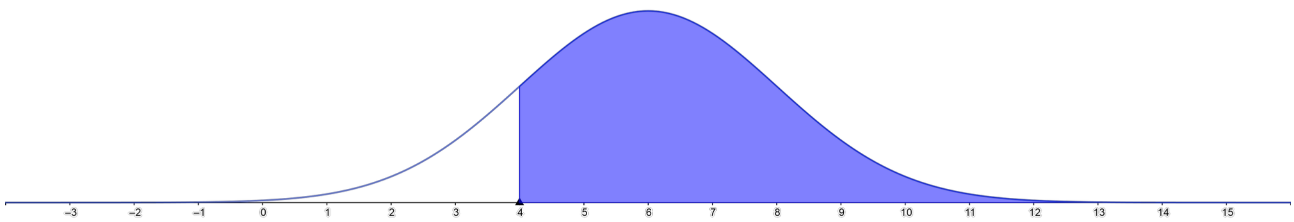
d) Sabiendo que hay 80 alumnos en 2ª de bachillerato, ¿cuántos habrán suspendido?

$$P(X<5) \underbrace{=}_{tipificamos} P\left(\frac{X-6}{2} < \frac{5-6}{2}\right) = P(Z<-0{,}5) \underbrace{=}_{simetría} 1 - P(Z \le 0{,}5) \underbrace{=}_{tabla\ N(0,1)} 1 - 0{,}6915 = 0{,}3085$$

Por lo tanto, habrá $0{,}3085 \cdot 80 = 24{,}68$ aproximadamente 25 alumnos han suspendido.

e) Tenga más de un 4 de nota media:

$$P(X>4) \underbrace{=}_{tipificamos} P\left(\frac{X-6}{2} > \frac{4-6}{2}\right) = P(Z > -1) \underbrace{=}_{simetría} P(Z \le 1) \underbrace{=}_{tabla\ N(0,1)} 0{,}8413$$



f) Tenga entre un 4 y un 6:

$$P(4\le X\le 6)\underbrace{=}_{tipificamos} P\left(\frac{4-6}{2}\le\frac{X-6}{2}\le\frac{6-6}{2}\right)=P(-1\le\ Z\ <0)=P(Z\le 0)-P(Z\le\ -1)=$$

$$\underbrace{=}_{simetría}\ 0,5-(1-P(Z\le 1))\underbrace{=}_{tabla\,N(0,1)}\ 0,5-1+0,8413=0,3413$$

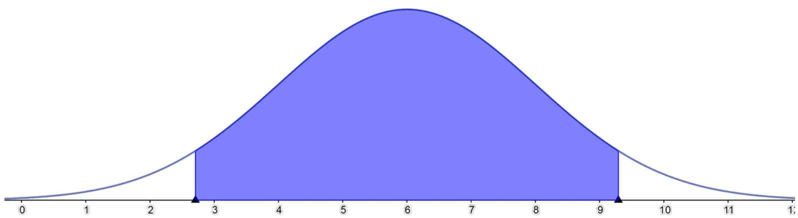
g) Se sabe que sólo un 5% del alumnado puede obtener una matrícula de honor. Halla la nota media que debe alcanzar un alumno/a para que se le conceda cumpla:

$$P(X>z_0)=0,05\Rightarrow P(X\le z_0)=0,95\underbrace{\Rightarrow}_{tipificamos} P\left(\frac{X-6}{2}\le\frac{z_0-6}{2}\right)=P\left(Z\ \le\ \frac{z_0-6}{2}\right)=0,95$$

$$\underbrace{\Rightarrow}_{tabla\,N(0,1)}\ \frac{z_0-6}{2}=1,645\ \Rightarrow\ z_0=\ 1,645\cdot 2+6\ =\ 9,29$$

(0,95 se encuentra entre P(Z≤1,64) = 0,9495 y P(Z≤1,65) = 0,9505)

h) Encuentra un intervalo centrado en la media que encierre el 90% de las notas de los alumnos: Buscamos un intervalo (z_1, z_0) que deje un 5% de las notas a la izquierda y un 5% a la derecha. El valor z_0 ya lo hemos calculado en el apartado anterior. Hallamos z_1 tal que $P(X\le z_1)=0,05\ \Rightarrow$



$$P\left(Z\le\frac{z_1-6}{2}\right)=0,05\ \Rightarrow\ 1-P\left(Z\le-\frac{z_1-6}{2}\right)=0,05\ \Rightarrow\ P\left(Z\le-\frac{z_1-6}{2}\right)=0,95$$

$$\Rightarrow\ -\frac{z_1-6}{2}=1,645\ \Rightarrow\ z_1=6-2\cdot 1,645=2,71$$

En el intervalo (2,71; 9,29) encontraremos el 90% de las notas de 2º de bachillerato.

3. APROXIMACIÓN DE LA BINOMIAL POR LA NORMAL

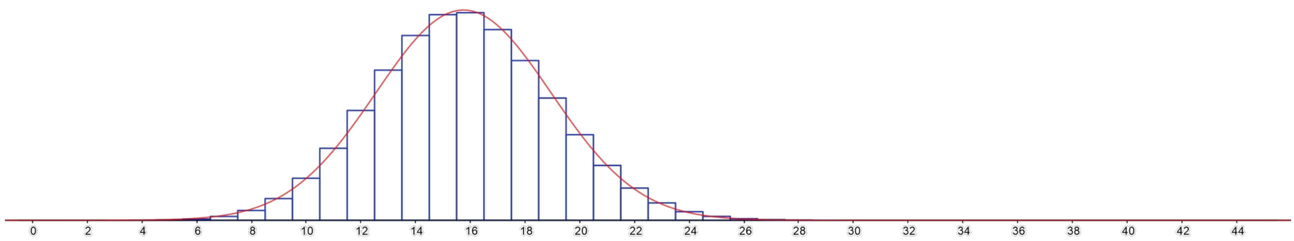
Supongamos que X en una variable aleatoria con distribución de probabilidad binomial B(n, p) y que el parámetro n tiende a infinito mientras que el parámetro p permanece constante. Entonces, puede demostrarse que la función de distribución de la variable aleatoria X tiende a una normal:

Proposición: (Aproximación de la binomial por la normal)

Sea X ~ B(n, p) tal que, n ≥ 30 ; np ≥ 5 y nq ≥ 5. Entonces la variable X se puede aproximar por una distribución normal $\boxed{X' \rightarrow N(np, \sqrt{npq})}$.

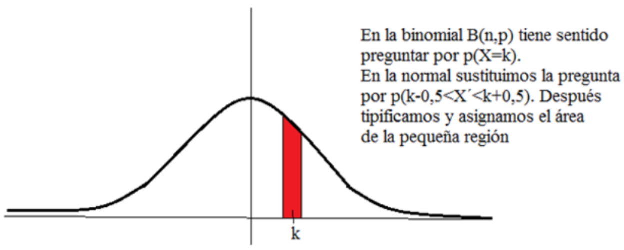
Esta proposición es consecuencia directa de un teorema más general llamado **Teorema Central del Límite.** Esto nos permitirá determinar probabilidades de la distribución binomial utilizando la distribución normal estándar.

Ejemplo: Sea X ~B(45, 0,35), por tanto, cumple las condiciones: n = 45 ≥ 30 ; np = 15,75 ≥ 5 y nq = 29,25 ≥ 5. Podemos aproximarla por una normal X' ~ N(15,75; 3,2).

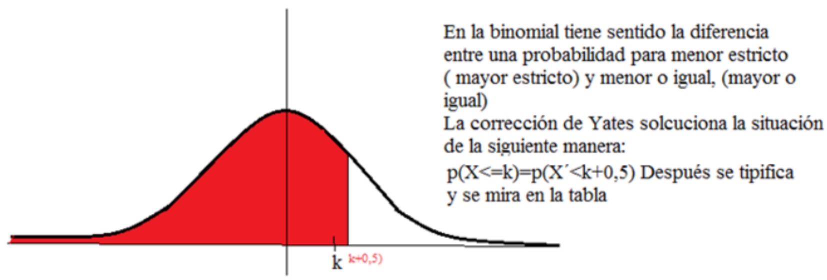


Nota: Como es lógico, el hecho de aproximar una variable discreta como la binomial, por una continua como la normal, da lugar a un problema. El problema es que mientras sí existe la probabilidad de que una binomial tome un valor concreto, en el caso de la normal, esta probabilid-ad es cero. Para resolver este problema, se introducen las llamadas **correcciones de Yates**, que consiste en considerar los valores de la variable discreta X como marcas de clase de intervalos de la forma siguiente:

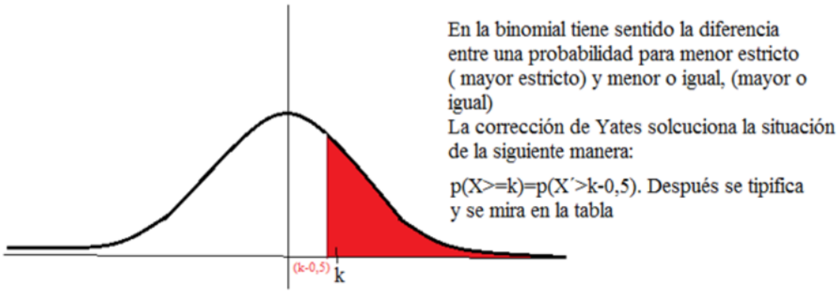
1. P (X = k) ≃ P (k - 0.5 ≤ X' ≤ k + 0.5).



2. P (X ≤ k) ≃ P (X' ≤ k + 0,5)



3. P (X < k) =P (X ≤ k - 1) ≃ P (X' ≤ k - 1 + 0,5) = P (X' ≤ k – 0,5).



4. P (X ≥ k) ≃ P (X' ≥ k – 0,5)

5. P (X > k) = P (X ≥ k + 1) ≃ P (X' ≥ k + 1 - 0,5) = ≃ P (X' ≥> k + 0,5)

Ejemplo: En el ejemplo anterior calculamos P(X = 17):

Sin aproximación por la normal: $P(X=17)=\binom{45}{17}\cdot 0{,}35^{17}\cdot 0{,}65^{28}=0{,}1131$

Aproximando por la normal: $P(X=17)\underbrace{\simeq}_{\substack{aprox.\,normal\\ con\ \ corrección\ \ de\ \ Yates}} P(17-0{,}5\le X'\le 17+0{,}5)=$

$$\underbrace{=}_{tipificación} P\left(\frac{16{,}5-15{,}75}{3{,}2}\le Z\le\frac{17{,}5-15{,}75}{3{,}2}\right)=P(0{,}23\le Z\le 0{,}55)=0{,}7088-0{,}5910=0{,}1178$$

Ejemplo: Durante cierta epidemia de gripe, enferma el 30% de la población. En un aula con 200 estudiantes de Medicina, calcula la probabilidad de que:

a) Ninguno padezca la enfermedad

b) Haya 60 estudiantes con gripe

c) Al menos 40 padezcan la enfermedad

Sea X = "nº de estudiantes enfermos de gripe" es una binomial con n = 200 y p = 0,3

a) $P(X=0)=\binom{200}{0}0{,}3^{0}0{,}7^{200}=1{,}05\cdot 10^{-31}$

b) $P(X=60)=\binom{200}{60}0{,}3^{60}0{,}7^{140}=7{,}04\cdot 10^{51}\cdot 4{,}3\cdot 10^{-32}\cdot 2{,}06\cdot 10^{-22}=0{,}0624$

(algunas calculadoras no pueden calcular este número combinatorio)

Podemos aproximar el resultado por la normal si se cumplen las condiciones: n = 200 ≥30, np = 60≥5 y nq=140≥5. Por tanto, X ~ B(200; 0,3) → X' ~N(60; 6,5)

$$P(X=60)\underbrace{\simeq}_{\substack{aprox.\,normal\\ con\ \ corrección\ \ de\ \ Yates}} P(60-0{,}5\le X'\le 60+0{,}5)\underbrace{=}_{tipificación} P\left(\frac{59{,}5-60}{6{,}5}\le Z\le\frac{60{,}5-60}{6{,}5}\right)=$$

$$=P(-0{,}08\le Z\le 0{,}08)=2\cdot P(Z\le 0{,}08)-1=2\cdot 0{,}5319-1=0{,}0638$$

c) $P(X\ge 40)\underbrace{\simeq}_{\substack{aprox.\,normal\\ con\ \ corrección\ \ de\ \ Yates}} P(X'\ge 40-0{,}5)\underbrace{=}_{tipificación} P\left(Z\ge\frac{39{,}5-60}{6{,}5}\right)=P(Z\ge -3{,}15)=$

$P(Z\le 3{,}15)=0{,}9992$