VARIABLES ALEATORIAS (discretas y continuas)

Se llama **variable aleatoria** a una función que asigna a cada suceso de un espacio muestral un número real.

Pueden ser **discretas**, es decir, que toman un número finito o numerable de valores (a saltos, paso a paso), o **continuas**, es decir, que toman todos los valores intermedios entre dos valores dados

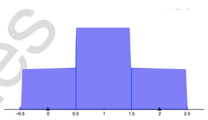
Ejemplo discreta:, X: "*número de caras al lanzar dos monedas*"

El espacio muestral es *E={++,+c,c+,cc)*, siendo +:"cruz" y c:"cara".

La variable aleatoria sería: ++ -> *X=0* +c -> *X=1* c+ -> *X=1* cc -> *X=2*

A esos valores de la variable aleatoria les corresponde una probabilidad:

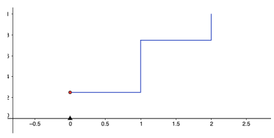
$$P(X=0) = 1/4 = 0{,}25$$
$$P(X=1) = 2/4 = 1/2 = 0{,}5$$
$$P(X=2) = 1/4 = 0{,}25$$



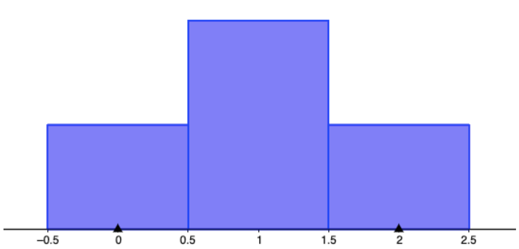
Así se puede definir una **función de probabilidad** $p(x_i) = P(X = x_i)$ que cumple con las propiedades de la probabilidad: $0 \leq p(x_i) \leq 1$ y la suma de todas es 1: $\left(\sum p(x_i) = 1\right)$.

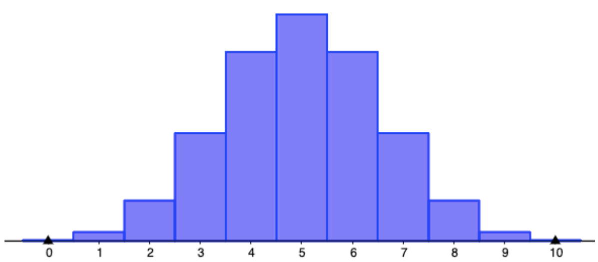
Y también se puede definir una **función de distribución** $F(x) = P(X \leq x)$. En el ejemplo,

$$F(0) = P(X \leq 0) = 0{,}25$$
$$F(1) = P(X \leq 1) = 0{,}25 + 0{,}5 = 0{,}75$$
$$F(2) = P(X \leq 2) = 0{,}25 + 0{,}5 + 0{,}25 = 1$$



De aquí surgen múltiples propiedades y toda una rama de estudio de las matemáticas, pero en este curso estudiaremos una función de distribución concreta: **la distribución binomial** (experimentos dicotómicos éxito/fracaso donde contamos número de éxitos)





LANZAR 2 MONEDAS Y Nº CARAS **LANZAR 10 MONEDAS Y Nº CARAS**

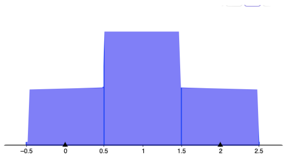
Ejemplo continua:

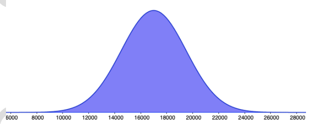
X: "*horas que tarda en fundirse una bombilla seleccionada al azar de un lote*"

En este caso *no tiene sentido hablar de probabilidad en un punto aislado* (probabilidad de que fallen exactamente en 18494 horas, 53 minutos y 12 segundos de uso), sino **por intervalos** (probabilidad de que falle con entre 15000 y 20000 horas de uso).

El concepto equivalente a la función de probabilidad discreta es la **función de densidad**, e igual que la suma de todas las probabilidad discretas tiene que ser uno (suma de todas las barras en el diagrama de barras), el área bajo la curva de la función de densidad también tiene que ser 1 (integral definida de la función de densidad):

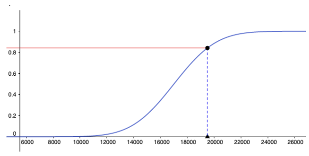
$\sum p(x_i) = 1$ (discreta) $\int_{-\infty}^{\infty} f(x)\,dx = 1$ (continua)





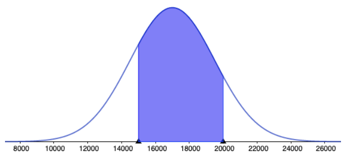
Así, igual que definimos la función de distribución discreta, lo hacemos para la continua. La probabilidad de que X sea menor que un valor dado, es el área bajo la curva hasta ese valor. Y **la herramienta matemática para calcular área bajo la curva es la integral**:

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(x)\,dx$$



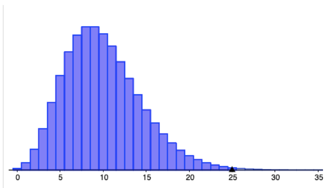
Y se calcularían probabilidades por intervalos, como se dijo, calculando esas integrales, como en el ejemplo probabilidad de que falle con entre 15000 y 20000 horas de uso:

$$P(15000 \leq X \leq 20000) = \int_{15000}^{20000} f(x)\,dx$$

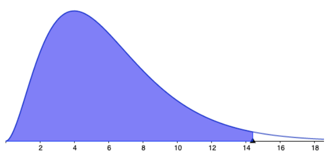


En este curso, como continua, solo estudiaremos la función de **distribución normal**, y no será necesario calcular integrales, sabremos realizar el <u>cálculo a través de una tabla</u> que ya tiene las integrales que necesitamos calculadas.

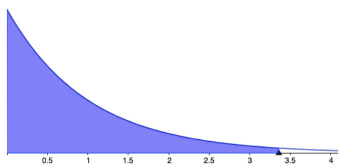
Otros ejemplos de gráficas de distribuciones que no estudiaremos serían:



PASCAL (DISCRETA)



CHI CUADRADO (CONTINUA)

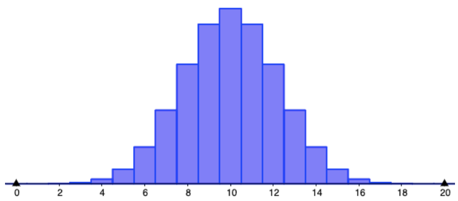


EXPONENCIAL (CONTINUA)

DISTRIBUCIÓN BINOMIAL $B(n, p)$

Es una distribución de probabilidad **discreta** que cuenta el **número de "éxitos"** en un **experimento aleatorio dicotómico** (esto significa que solo puede haber dos resultados posibles, a uno se le llama "éxito" y al otro "fracaso"), repetido un número n de veces.

Número de casos: n
Probabilidad de éxito: p
Probabilidad de fracaso: $q = 1 - p$
Número de éxitos en n casos: k



NOTA: *éxito y fracaso no tienen connotaciones positivas o negativas, se define según lo que convenga para el problema a resolver.*

EJEMPLOS

- Lanzar una moneda 10 veces y contar número de caras es una binomial $B(10, 0{,}5)$
- Lanzar un dado 40 veces y contar número de cincos es una binomial $B(40, 1/6)$
- Sabemos que una de cada mil bombillas que salen de una fábrica sale defectuosa. Contar el número de bombillas defectuosas en un lote de 3000 es una binomial $B(3000, 0{,}001)$

CÁLCULO EN UNA DISTRIBUCIÓN BINOMIAL

La probabilidad de que en una binomial $B(n, p)$ haya k éxitos es: $P(X = k) = \binom{n}{k} p^k q^{n-k}$

Ejemplo: *Si en 10 lanzamientos de dado, queremos saber la probabilidad de que salga 2 veces el número 5, haremos*:

- Es una binomial $B(n, p)$ con $n = 10$, $p = 1/6$, $q = 5/6$ y con $k = 2$ => $B(10, 1/6)$,

- $P(X = 2) = \binom{10}{2} \left(\frac{1}{6}\right)^2 \left(\frac{5}{6}\right)^8 = 0{,}2907$

Ejemplo: *Si en 8 lanzamientos de una moneda, queremos saber la probabilidad de que salgan menos de 2 caras, sería*:

- Es una binomial $B(n, p)$ con $n = 8$, $p = 0{,}5$, $q = 0{,}5$ => $B(8, 0{,}5)$

- Menos de dos caras es que salgan 0 caras o 1 cara:

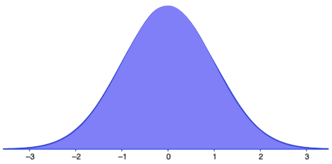
$$P(X < 2) = P(X = 0) + P(X = 1) = \binom{8}{0} 0{,}5^0 \cdot 0{,}5^8 + \binom{8}{1} 0{,}5^1 \cdot 0{,}5^7 =$$

$$= 0{,}0039 + 0{,}0313 = 0{,}0352$$

DISTRIBUCIÓN NORMAL $N(\mu, \sigma)$

Es una distribución de probabilidad **continua**, y la gráfica de su función de densidad es la conocida **campana de Gauss.** Esta gráfica es simétrica respecto a la media ($x = \mu$)

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

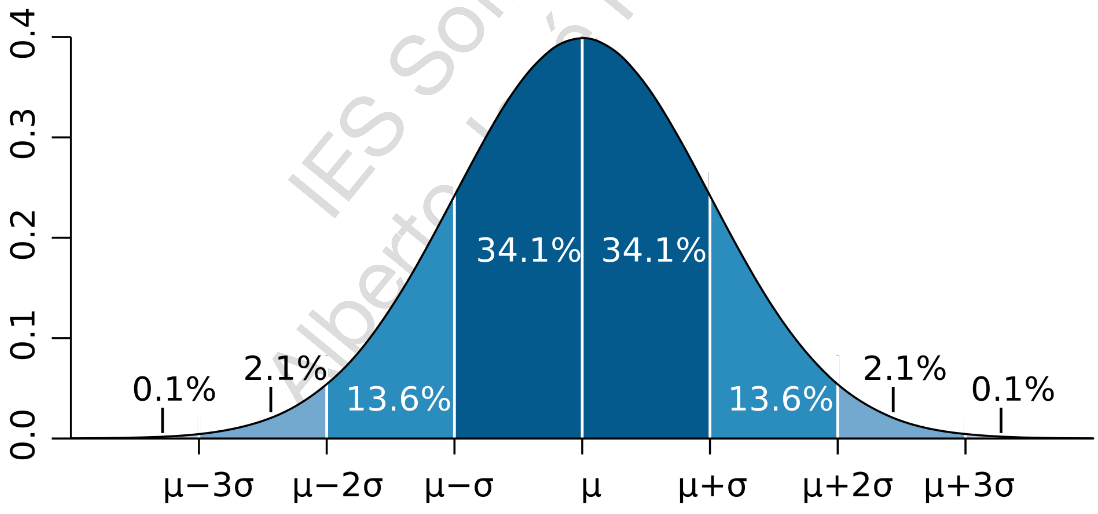


NORMAL (0,1)

Es especialmente conocida e importante, ya que **hay un número enorme de situaciones reales que se pueden estudiar usando como modelo esta distribución**, y también aparece dentro de las matemáticas (nosotros la aplicaremos como aproximación de una binomial cuando n se hace demasiado grande).

Conociendo más la campana de Gauss

Conocer las características gráficas de la campana nos permite interpretar mejor una situación bajo el modelo de la normal sin realizar los cálculos. No es necesario para los ejercicios, pero ayuda a tener una visión más completa. El aspecto de una campana para una distribución $N(\mu, \sigma^2)$ es:



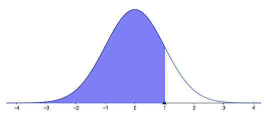
DE AINALI - TRABAJO PROPIO, CC BY-SA 3.0, HTTPS://COMMONS.WIKIMEDIA.ORG/W/INDEX.PHP?CURID=3141713

- Es simétrica respecto de su media μ, que coincide con moda y mediana.
- Los puntos de inflexión de la curva se dan para $\mu \pm \sigma$.
- Entre $[\mu - \sigma, \mu + \sigma]$ está un 68,26% de la población.
- Entre $[\mu - 2\sigma, \mu + 2\sigma]$ está un 95,44% de la población.
- Entre $[\mu - 3\sigma, \mu + 3\sigma]$ está un 99,47% de la población

CÁLCULO EN UNA DISTRIBUCIÓN NORMAL: TABLA $N(0,1)$

El cálculo de probabilidades en una distribución normal $N(\mu, \sigma)$ se basa en el área bajo la curva en un determinado intervalo, y eso se haría con la integral definida:

$$P(X \leq b) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{b} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \, dx$$



Pero en la práctica, se trabaja con una **tabla de valores** con los cálculos para la $N(0,1)$, y un proceso para convertir cualquier normal $N(\mu, \sigma)$ en $N(0,1)$ llamado **tipificación**.

$$\text{Si } X \to N(\mu, \sigma) \implies Z = \frac{X - \mu}{\sigma} \to N(0,1)$$

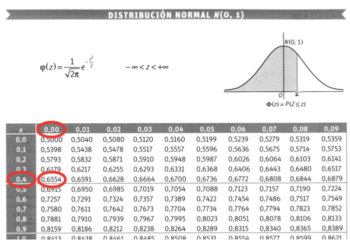
Ejemplo: *ABAU MAT I 2021 EXT. El grosor de las planchas de acero que se producen en una cierta fábrica sigue una distribución normal de media 8 mm y desviación típica 0.5 mm. Calcula la probabilidad de que una plancha elegida al azar tenga un grosor menor de 8,2 mm.*

X: "grosor en mm de una plancha de acero" es $N(8, 0{,}5)$

Tipificamos: $X \to N(8{,}0{,}5) \implies Z = \dfrac{X - 8}{0{,}5} \to N(0,1)$

$$P(X \leq 8{,}2) \overset{tipif.}{=} P\left(Z \leq \frac{0{,}2}{0{,}5}\right) = P(Z \leq 0{,}4) \overset{tabla}{=} 0{,}6554$$

En la tabla se busca así:



DISTRIBUCIÓN NORMAL $N(0, 1)$

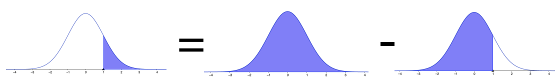
$\varphi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$ $\quad -\infty < z < +\infty$

$\Phi(z) = P(Z \leq z)$

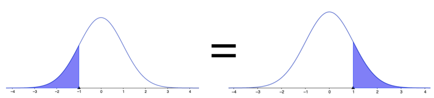
x	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621

Pero **la tabla solo nos ofrece** $P(Z \leq k)$ (áreas a la izquierda), para valores de k positivos.
¿**Cómo podemos actuar en otras situaciones**?

a) Para áreas a la derecha, el complementario $P(Z \geq k) = 1 - P(Z \leq k)$

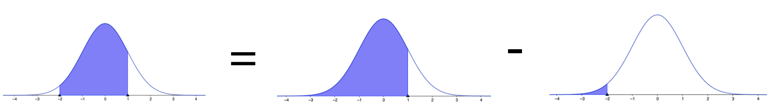


b) Para valores negativos, $P(Z \leq -k) \overset{simetr}{=} P(Z \geq k) = 1 - P(Z \leq k)$



(Caso anterior)

c) Para áreas entre dos valores, separamos: $P(a \leq Z \leq b) = P(Z \leq b) - P(Z \leq a)$



USO GENERAL DE LA NORMAL, CALCULO DE X (tipificado a Z), TABLA DIRECTA

Ejemplo: *Calcula la probabilidad de que una plancha elegida al azar tenga un grosor entre 7,6 mm y 8,2 mm.*

$$P(7{,}6 \le X \le 8{,}2) \overset{tipif.}{=} P(-\frac{0{,}4}{0{,}5} \le Z \le \frac{0{,}2}{0{,}5}) = P(-0{,}8 \le Z \le 0{,}4) =$$

$$= P(Z \le 0{,}4) - P(Z \le -0{,}8) = P(Z \le 0{,}4) - P(Z \ge 0{,}8) =$$

$$= P(Z \le 0{,}4) - (1 - P(Z \le 0{,}8)) = P(Z \le 0{,}4) + P(Z \le 0{,}8) - 1 =$$

$$\overset{tabla}{=} 0{,}6554 + 0{,}7881 - 1 = \mathbf{0,4435}$$

USO DE LA NORMAL EN SENTIDO INVERSO, CÁLCULO DEL k, TABLA INVERSA (ap b))

Ejemplo*: ABAU MAT II ORD 2022 8.*

a) *Calcule el valor de $P(-2 \le X \le 7)$ si X sigue una distribución normal de media 1 y desviación típica 3.*

b) *Calcule el valor de α que hace que $P(\mu - \alpha \le X \le \mu + \alpha) = 0{,}8064$ si X sigue una distribución normal de media μ y desviación típica 4.*

a) Tipificamos: $X \to N(1,3) \implies Z = \dfrac{X-1}{3} \to N(0,1)$

$$P(-2 \le X \le 7) \overset{tipif.}{=} P(-1 \le Z \le 2) = P(Z \le 2) - P(Z \le -1) =$$

$$= P(Z \le 2) - (1 - P(Z \le 1)) = P(Z \le 2) + P(Z \le 1) - 1 \overset{tabla}{=}$$

$$= 0{,}9772 + 0{,}8413 - 1 = \mathbf{0,8185}$$

b) Tipificamos: **Ejemplo***: ABAU MAT I ORD 2022 8.*

a) *Calcule*

$$0{,}8064 = P(\mu - \alpha \le X \le \mu + \alpha) \overset{tipif.}{=} P(-\frac{\alpha}{4} \le Z \le \frac{\alpha}{4}) =$$

$$= P(Z \le \frac{\alpha}{4}) - P(Z \le -\frac{\alpha}{4}) = P(Z \le \frac{\alpha}{4}) - P(Z \ge \frac{\alpha}{4}) =$$

$$= P(Z \le \frac{\alpha}{4}) - (1 - P(Z \le \frac{\alpha}{4})) = 2P(Z \le \frac{\alpha}{4}) - 1$$

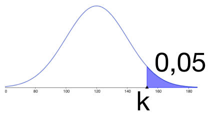
Despejamos $P(Z \le \dfrac{\alpha}{4}) = \dfrac{0{,}8064 + 1}{2} = 1{,}8064 \implies P(Z \le \dfrac{\alpha}{4}) = 0{,}9032$

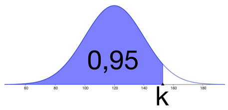
Y, de la tabla, buscando 0,8064 llegamos a fila 1,3 y columa 0,00:

$$\frac{\alpha}{4} = 1{,}3 \implies \alpha = \mathbf{5,2}$$

Ejemplo: *ABAU MAT II ORD 2023 8. (**Caso 5% de los mejores**)*
b) Para conceder becas de estudio, un organismo valora los méritos presentados y asigna a cada candidato una puntuación que indica más méritos cuanto mayor es su valor. Este año, la puntuación sigue una distribución normal de media 100 y desviación típica 20, y se toma la decisión de conocer la beca al 5% mejor del conjunto de solicitantes. ¿Qué puntuación es necesario alcanzar para obtener la beca?

b) El **5% de los mejores** se representa como (la incógnita es el valor de k, el dato el área):





$$P(X \geq k) = 0{,}05 \overset{compl.}{\Longrightarrow} P(X < k) = 0{,}95 \overset{tip.}{\Longrightarrow} P(Z \leq \frac{k-100}{20}) = 0{,}95$$

(tabla inversa)

El valor de la tabla para 0,95 está entre la columna 1,64 y 1,65, por lo que tomamos **1,645**

$$\frac{k-100}{20} = 1{,}645 \implies \mathbf{k = 132,9}$$ -> puntuación necesaria

Recomiendo pensar como sería el caso 5% de los peores, por ejemplo.

Ejemplo: *ABAU MAT II EXT. 2023 8. (**Caso valor superado por el 67% (67% de los mejores)**)*
Para un determinado grupo de pacientes, la tensión arterial sistólica (medida en mmHg) sigue una distribución normal de media 123,6 y desviación típica 17,8. a) Calcule la probabilidad de que un paciente elegido al azar tenga una tensión comprendida entre 100 y 120 mmHg. b) Luego, obtenga el valor de la tensión que es superado por el 67% de los pacientes.

a)

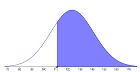
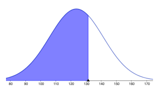
$$P(100 \leq X \leq 120) \overset{tipif.}{=} P(\frac{100-123{,}6}{17{,}8} \leq Z \leq \frac{120-123{,}6}{17{,}8}) = P(-1{,}33 \leq Z \leq -0{,}2) =$$

$$= P(Z \leq -0{,}22) - P(Z \leq -1{,}33) \overset{sim.}{=} P(Z \geq 0{,}22) - P(Z \geq 1{,}33) \overset{comp}{=}$$

$$= (1 - P(Z \leq -0{,}22)) - (1 - P(Z \leq -1{,}33)) = P(Z \leq 1{,}33) - P(Z \leq 0{,}2) =$$

$$= 0{,}9032 - 0{,}5793 = \mathbf{0,3239}$$

b) ***Valor superado por el 67% (67% de los mejores)***

$$P(X \geq k) = 0{,}67 \overset{tip.}{\Longrightarrow} P(Z \leq \frac{k-123{,}6}{17{,}8}) = 0{,}67 \overset{sim.}{=} P(Z \leq -\frac{k-123{,}6}{17{,}8}) = 0{,}67 \xrightarrow{tabla\ inversa} 0{,}44$$

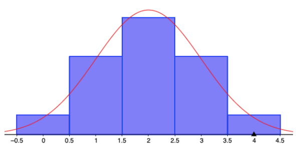
La búsqueda inversa en la tabla nos lleva al valor de fila/columna 0,44:

$$-\frac{k-123{,}6}{17{,}8} = 0{,}44 \implies \mathbf{k = 115,76}$$ valor de tensión buscado

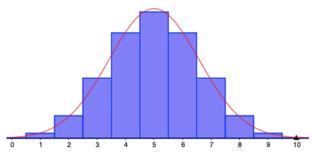
APROXIMACIÓN DE LA BINOMIAL POR LA NORMAL $B(n,p) \to N(np, \sqrt{npq})$

Cuando en una distribución binomial n se hace muy grande, realizar los cálculos puede ser demasiado costoso. En determinados casos, merece la pena considerar la variable como continua y aproximar la binomial a través de la distribución normal.

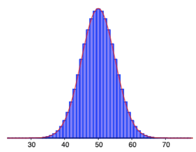
Vemos en el gráfico como el aspecto de la distribución binomial se va acercando a la normal cuando subimos el número de casos, si la probabilidad de de éxito es $p = 0{,}5$:



N=4

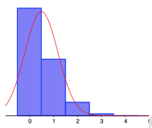


N=10



N=100

Pero si la probabilidad de éxito es muy pequeña o muy grande, o n es pequeño, el ajuste ya no es tan bueno:



N=4, P=0,05

Por lo tanto, el criterio para considerar que podemos realizar una **aproximación de calidad de la binomial usando la normal será**:

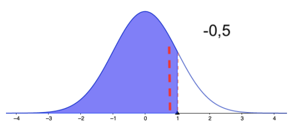
Aproximación **buena si** $\begin{cases} np > 3 \\ nq > 3 \end{cases}$ Aproximación **muy buena si** $\begin{cases} np > 5 \\ nq > 5 \end{cases}$

Entonces convertiremos: $B(n,p) \to N(np, \sqrt{npq})$

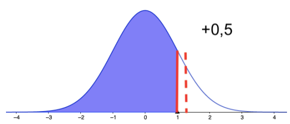
Corrección de continuidad (Yates)

Para poder realizar la aproximación aún mejor, hay que aplicar una corrección para que se ajuste mejor el **paso de discreto a continuo**. Consiste en **sumar o restar medio punto al valor** k**,** dependiendo de si está incluido o no en en intervalo (“*coger o quitar un cachito al intervalo*”).

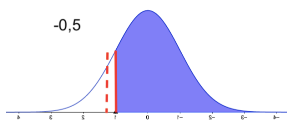
$$P(X < k) = P(X' \le k - 0{,}5)$$



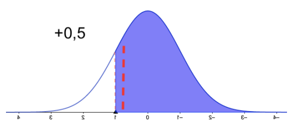
$$P(X \le k) = P(X' \le k + 0{,}5)$$



$$P(X \ge k) = P(X' \ge k - 0{,}5)$$



$$P(X > k) = P(X' \ge k + 0{,}5)$$



Ejemplo: *ABAU MAT I SEPT 2018 B 4. En un bombo tenemos 10 bolas idénticas numeradas del 0 al 9 y cada vez que hacemos una extracción devolvemos la bola al bombo.*

a) Si hacemos 5 extracciones, calcula la probabilidad de que el 7 salga menos de 2 veces.

b) Si hacemos 100 extracciones, calcula la probabilidad de que el 7 salga menos de 9 veces.

a) Es una situación binomial $B(5, 0{,}1)$. Nos piden

a) $P(X < 2) = P(X = 0) + P(X = 1) = \binom{5}{0} 0{,}1^0 \cdot 0{,}9^5 + \binom{5}{1} 0{,}1^1 \cdot 0{,}9^4 =$

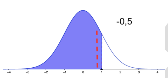
$= 0{,}5905 + 0{,}3281 = \mathbf{0,9185}$

b) Con 100 extracciones y cálculo $P(X < 9)$ es muy costoso, así que comprobamos si procede aproximar por la normal:

$$\begin{cases} np = 100 \cdot 0{,}1 = 10 > 5 \\ nq = 100 \cdot 0{,}9 = 90 > 5 \end{cases} \Longrightarrow \text{La aproximación sería muy buena.}$$

$$X \in B(100, 0{,}1) \xrightarrow{aprox.} X' \in N(100 \cdot 0{,}1, \sqrt{100 \cdot 0{,}1 \cdot 0{,}9}) = N(10, 3)$$

Calculamos:



$$P(X < 9) \overset{corr.}{=} P(X' \leq 8{,}5) \overset{tipif.}{=} P\left(Z \leq \frac{8{,}5 - 10}{3}\right) = P(Z \leq -0{,}5) =$$

$$= P(Z \geq 0{,}5) = 1 - P(Z \leq 0{,}5) \overset{tabla}{=} 1 - 0{,}6915 = \mathbf{0,3085}$$