UNIDAD 7: DISTRIBUCIÓN BINOMIAL Y NORMAL

CONTENIDO

1. FACTORIAL DE UN NÚMERO. NÚMEROS COMBINATORIOS

Estos conceptos nos harán falta cuando tratemos la distribución binomial.

- Factorial de un número

Se llama factorial de un número natural n mayor que 1, al producto de los números naturales desde 1 hasta n y se representa por $n!$.

Así, $n! = n\cdot(n-1)\cdot(n-2)\cdots 3\cdot 2\cdot 1$

Se define, por convenio que:

$$1! = 1 \qquad 0! = 1$$

Ejemplos:

$3! = 3\cdot 2\cdot 1 = 6 \qquad 4! = 4\cdot 3\cdot 2\cdot 1 = 24 \qquad 5! = 5\cdot 4\cdot 3\cdot 2\cdot 1 = 120$

- Números combinatorios

Se define el número combinatorio del número natural m sobre el número natural n como:

$$\binom{m}{n} = \frac{m!}{n!\cdot(m-n)!} \text{ con } n \le m$$

Se tiene que:

a) $\binom{m}{0} = 1$ b) $\binom{m}{m} = 1$ c) $\binom{0}{0} = 1$ d) $\binom{m}{1} = 1$

Ejemplos:

$\binom{7}{5} = \frac{7!}{5!\cdot 2!} = \frac{7\cdot 6\cdot 5!}{5!\cdot 2} = 21 \qquad \binom{7}{1} = \frac{7!}{1!\cdot 6!} = 7 \qquad \binom{8}{5} = \binom{8}{3} = \frac{8!}{3!\cdot 5!} = \frac{8\cdot 7\cdot 6\cdot 5!}{6\cdot 5!} = 56$

2. VARIABLES ALEATORIAS

Consideremos el experimento de lanzar 3 una moneda 3 veces. Tenemos que su espacio muestral es

$$E = \{CCC, CCX, CXC, XCC, CXX, XCX, XXC, XXX\}$$

Donde *C* representa el suceso salir cara y *X* representa el suceso salir cruz.

Para poder estudiar matemáticamente este experimento, es necesario asignar a cada uno de estos resultados un número, que haremos siguiendo un criterio. Por ejemplo, a cada resultado le corresponde el nº de caras obtenidas. Y tendremos una aplicación de los sucesos del espacio muestral en un subconjunto de los números reales

$$\begin{aligned} CCC &\to 3 \\ CCX &\to 2 \\ CXC &\to 2 \\ XCC &\to 2 \\ CXX &\to 1 \\ XCX &\to 1 \\ XXC &\to 1 \\ XXX &\to 0 \end{aligned}$$

A cada uno de estos criterios o aplicaciones se le llama *variable aleatoria*

Definición: Dado el espacio muestral E asociado a un experimento aleatorio, se llama variable aleatoria a toda aplicación X del espacio muestral E en un subconjunto de los números reales

$$X : E \to S \subseteq \mathbb{R}$$

Al conjunto de los valores de $\mathbb{R}$ asignados a los elementos de E se le llama *recorrido* de la variable aleatoria, y se suele representar por $\{x_1, x_2, ..., x_n\}$

Definición: Una variable aleatoria X se llama *discreta* si toma un nº finito de valores o un infinito numerable.

Ejemplo: Tenemos dos urnas, y cada una de ellas contiene tres bolas numeradas del 1 al 3. Se extrae una bola de cada urna y se anotan la suma de los números obtenidos. En este caso se trata de una variable aleatoria discreta sobre el espacio muestral $E = \{(1,1),(1,2),(1,3),(2,1),(2,2),(2,3),(3,1),(3,2),(3,3)\}$ y cuyo recorrido es $\{x_1 = 2, x_2 = 3, x_3 = 4, x_4 = 5, x_5 = 6\}$

Definición: Una variable aleatoria X se llama *continua* si toma infinito valores en un intervalo de la recta real.

Ejemplo: La variable que representa la altura de todas las mujeres de Sevilla de 25 años es continua pues puede tomar todos los valores que van desde la más baja a la más alta

Definición: Se llama *función o distribución de probabilidad* de una variable aleatoria discreta X a la aplicación que a cada valor x_i de la variable le hace corresponder la probabilidad de que la variable tome dicho valor, y lo expresamos como $P(X = x_i)$

Ejemplo: Considerando el ejemplo inicial de lanzar 3 monedas y anotar el nº de caras como variable aleatoria X podemos construir una tabla con los valores de la variable aleatoria y sus correspondientes probabilidades

X	$x_1 = 0$	$x_2 = 1$	$x_3 = 2$	$x_4 = 3$
$P(X = x_i)$	$P(X = 0) = \frac{1}{8}$	$P(X = 1) = \frac{3}{8}$	$P(X = 2) = \frac{3}{8}$	$P(X = 3) = \frac{1}{8}$

Propiedades (importantes):

- En toda función de probabilidad se verifica que $P(X=x_1)+P(X=x_2)+\ldots+P(X=x_n)=1$, que de forma más resumida se pone como $\sum_{i=1}^{n} P(X=x_i)=1$
- Se tiene que $P(X=x_i)\geq 0 \quad \forall i=1,\ldots,n$

PARÁMETROS DE UNA DISTRIBUCIÓN DE PROBABILIDAD

En las distribuciones de probabilidad podemos proceder como en las distribuciones estadísticas y definir medidas centrales y de dispersión. Se usan habitualmente dos:

- *Media o esperanza matemática*: Se nota por μ y se obtiene de la siguiente expresión

$$\mu = x_1 \cdot P(X=x_1)+x_2 \cdot P(X=x_2)+\ldots+x_n \cdot P(X=x_n)$$

- *Desviación típica*: Se representa por σ y se calcula con la siguiente fórmula:

$$\sigma = \sqrt{x_1^{\,2} \cdot P(X=x_1)+x_2^{\,2} \cdot P(X=x_2)+\ldots+x_n^{\,2} \cdot P(X=x_n)-\mu^2}$$

Ejemplo: Consideremos el experimento de lanzar 3 una moneda 3 veces. Sabemos que su espacio muestral es $E=\{CCC, CCX, CXC, XCC, CXX, XCX, XXC, XXX\}$ donde *C* representa el suceso salir cara y *X* representa el suceso salir cruz. Consideremos la variable aleatoria que a cada suceso elemental le asocia el número de caras obtenidas.

Se trata de la variable $X: E \to \{0,1,2,3\}$ que actúa de la siguiente forma:

$$\begin{aligned} X(ccc)&=3\\ X(ccx)&=2\\ X(cxc)&=2\\ X(xcc)&=2\\ X(cxx)&=1\\ X(xcx)&=1\\ X(xxc)&=1\\ X(xxx)&=0 \end{aligned}$$

Tenemos que:

$$P(X=3)=\frac{1}{8} \qquad P(X=2)=\frac{3}{8} \qquad P(X=1)=\frac{3}{8} \qquad P(X=0)=\frac{1}{8}$$

Ahora podemos calcula la media y la desviación típica.

MEDIA:

$$\mu = x_1 \cdot P(X=x_1)+x_2 \cdot P(X=x_2)+\ldots+x_n \cdot P(X=x_n) = 0 \cdot P(X=0)+1 \cdot P(X=1)+2 \cdot P(X=2)+3 \cdot P(X=3)$$

$$=0 \cdot \frac{1}{8}+1 \cdot \frac{3}{8}+2 \cdot \frac{3}{8}+3 \cdot \frac{1}{8}=\frac{12}{8} \Rightarrow \mu=\frac{3}{2}=1,5$$

DESVIACIÓN TÍPICA:

$$\sigma = \sqrt{x_1^{\,2} \cdot P(X=x_1)+x_2^{\,2} \cdot P(X=x_2)+\ldots+x_n^{\,2} \cdot P(X=x_n)-\mu^2} = \sqrt{0^2 \cdot \frac{1}{8}+1^2 \cdot \frac{3}{8}+2^2 \cdot \frac{3}{8}+3^2 \cdot \frac{1}{8}-\left(\frac{3}{2}\right)^2} \Rightarrow \sigma=\sqrt{\frac{24}{8}-\frac{9}{4}}$$

$$\Rightarrow \sigma=\sqrt{\frac{6}{8}}=\sqrt{\frac{3}{4}}=0,87$$

3. DISTRIBUCIÓN BINOMIAL O DE BERNOULLI

La **distribución binomial o de Bernoulli** es un caso particular de distribución de probabilidad discreta que tiene las siguientes características:

- En cada prueba del experimento aleatorio sólo son posibles dos resultados: el suceso A, llamado *éxito*, y su contrario o complementario, $\overline{A}$, llamado *fracaso*.
- En cada experimento se realizan n pruebas idénticas.
- La probabilidad de A, denotada por p, no varía de una prueba a otra. Y, por tanto, tampoco varía la probabilidad del fracaso $\overline{A}$, que se denota por q, y que obviamente es $q=1-p$.
- Cada prueba es *independiente* de las otras

Notaremos como $B(n,p)$ a la distribución binomial de parámetros n (nº de experiencias o pruebas realizadas) y p (probabilidad de éxito)

Se llama **variable aleatoria binomial**, $X=B(n,p)$, a aquella que expresa el nº de éxitos obtenidos al realizar las experiencias de una distribución binomial. Estas variables son discretas pues pueden tomar los valores $0,1,2,\ldots n-1,n$

Se tiene que: $P(X=a)=\binom{n}{a}\cdot p^a\cdot q^{n-a}$

Ejemplo: Lanzamos un dado 10 veces y observamos si la puntuación obtenida es múltiplo de 3. Se trata de una distribución binomial cuyos parámetros son:

$n=10$ el nº de pruebas realizadas

$p=\frac{2}{6}=\frac{1}{3}$ la probabilidad de éxito (sólo hay dos casos posibles de 6, que salga 3 ó 6) y $q=1-\frac{1}{3}=\frac{2}{3}$ la probabilidad de fracaso

La variable aleatoria binomial X puede tomar los valores $0,1,2,\ldots,9,10$ que es el nº de veces que se ha obtenido un múltiplo de 3. Por tanto, $X=B(10,\frac{1}{3})$

Por ejemplo, si nos piden la probabilidad de que al lanzar 10 monedas nos salgan exactamente 4 veces múltiplos de 3 será: $P(X=4)=\binom{10}{4}\cdot\left(\frac{1}{3}\right)^4\cdot\left(\frac{2}{3}\right)^{10-4}=\frac{10!}{4!\cdot 6!}\cdot\left(\frac{1}{3}\right)^4\cdot\left(\frac{2}{3}\right)^6 \Rightarrow$ Operando y usando la calculadora tenemos:

$P(X=4)=\frac{10\cdot 9\cdot 8\cdot 7\cdot 6!}{4\cdot 3\cdot 2\cdot 6!}\cdot\frac{2^6}{3^{10}}=10\cdot 3\cdot\cdot 7\cdot\frac{2^6}{3^{10}}=0,23$. Como vemos es un cálculo pesado y tedioso.

El problema aún más grave viene cuando nos preguntan, por ejemplo:

$P(X\geq 4)=P(X=4)+P(X=5)+P(X=6)+P(X=7)+P(X=8)+P(X=9)+P(X=10)$, lo cual ya se hace cuando menos insoportable. Veremos más adelante como hacer estos cálculos mediante una aproximación.

Parámetros: Se tiene que en una distribución binomial $X=B(n,p)$

- *Media*: La media es $\mu=n\cdot p$
- *Desviación típica*: La desviación típica es $\sigma=\sqrt{n\cdot p\cdot q}$

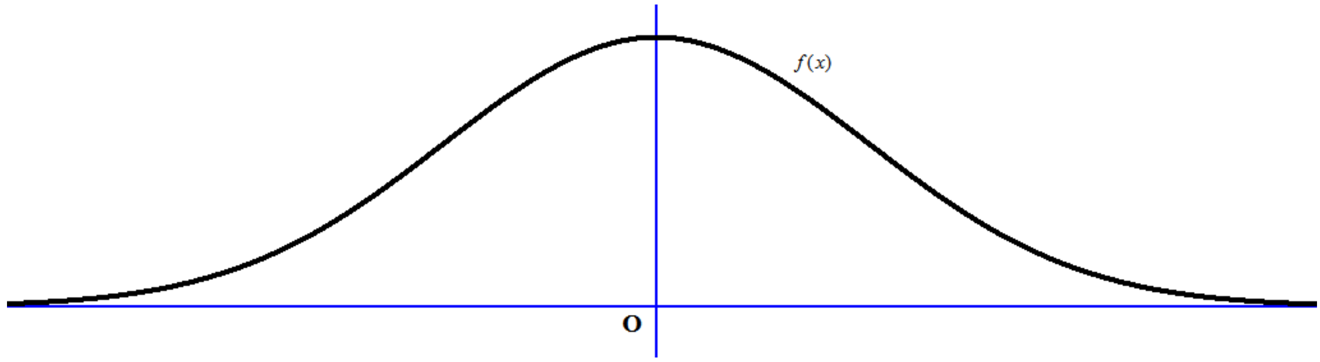
4. DISTRIBUCIÓN NORMAL O DE GAUSS

La distribución normal o de Gauss es una variable aleatoria continua. Y en las variables continuas tenemos que:

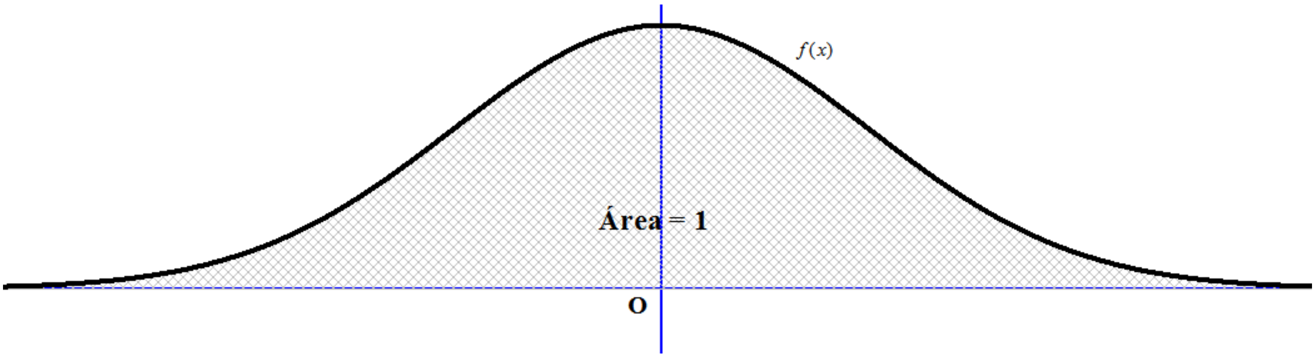
- Puede tomar un nº infinito de valores en la recta real
- En ellas la probabilidad de un valor concreto es cero. Las probabilidades siempre están asociadas a intervalos y las representaremos por $P(a \le X \le b)$ ó $P(X \ge a)$ ó $P(X \le b)$
- Para calcular esas probabilidades se usa la función de densidad o función de distribución, que es el área que limita con el eje OX en el intervalo dado.

La función de densidad se suele notar por $f(x)$ y tiene que cumplir las siguientes propiedades

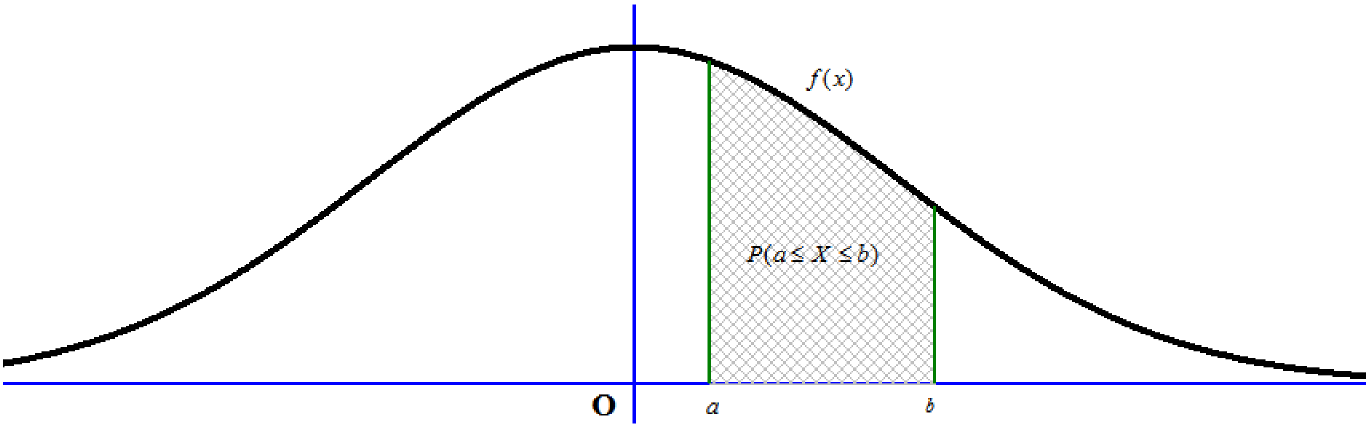
- Ha de ser definida positiva, $f(x) \ge 0$



- El área comprendida entre la función de densidad $f(x)$ y el eje OX ha de ser 1



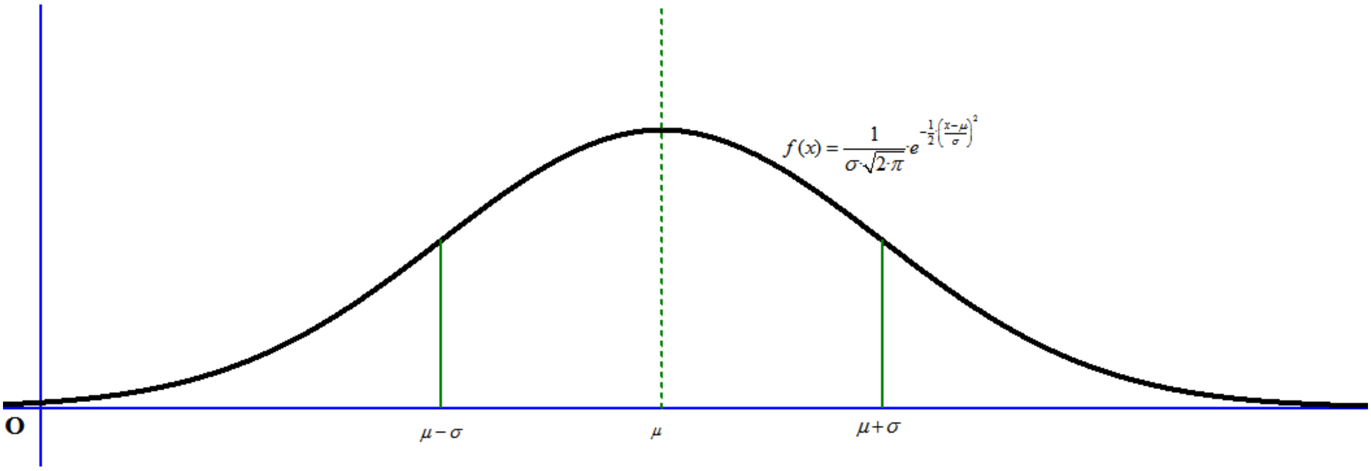
La probabilidad de que la variable aleatoria continua tome un valor en el intervalo $[a,b]$ es al área del recinto limitado por la gráfica, el eje OX y las rectas $x=a$ y $x=b$



Definición: Una distribución normal o de Gauss es una variable aleatoria con media μ y desviación típica σ cuya función de densidad es: $f(x)=\frac{1}{\sigma\cdot\sqrt{2\cdot\pi}}\cdot e^{-\frac{1}{2}\cdot\left(\frac{x-\mu}{\sigma}\right)^2}$

Los valores μ y σ se llaman parámetros de la distribución normal y la distribución se presenta por $N(\mu,\sigma)$

La función de densidad es como sigue:



Obviamente como la función es simétrica y el área encerrada entre la función y el eje OX es 1, podemos concluir que:

$$P(X\leq\mu)=\frac{1}{2}=0,5 \quad \text{y que} \quad P(X\geq\mu)=\frac{1}{2}=0,5$$

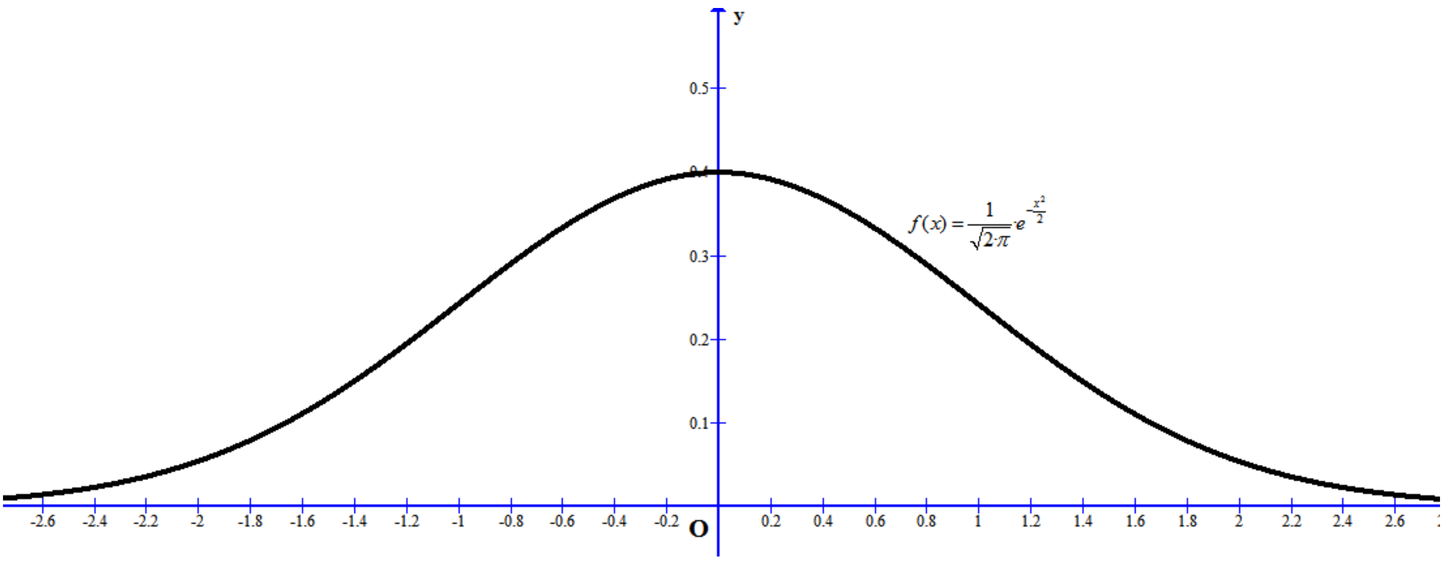
DISTRIBUCIÓN NORMAL ESTÁNDAR

Como veremos cualquier distribución normal con media μ y desviación típica σ puede asociarse a una distribución normal de media 0 y desviación típica1. La $N(0,1)$ es la llamada *distribución normal estándar*

Se suele notar por $Z=N(0,1)$

La función de densidad correspondiente es: $f(x)=\frac{1}{\sqrt{2\cdot\pi}}\cdot e^{-\frac{x^2}{2}}$

Y la gráfica es la siguiente:

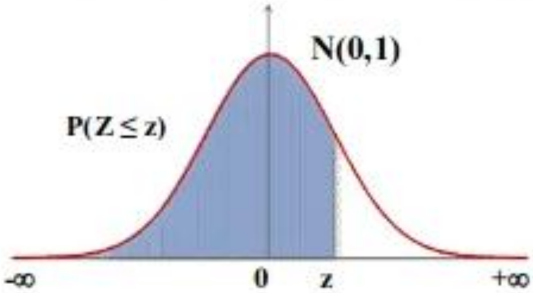


Como sabemos el área encerrada entre esta curva y le eje OX es 1, es decir, $P(-\infty \leq Z \leq +\infty) = 1$

La ventaja de la distribución tipificada frente a las demás distribuciones normales es que las probabilidades $P(Z \leq a)$ se encuentran baremadas (o tabuladas) en la tabla de la distribución normal

Dicha tabla es la siguiente:

FUNCIÓN DE DISTRIBUCIÓN NORMAL N(0,1)



z	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
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
2,6	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964
2,7	0,99653	0,99664	0,99674	0,99683	0,99693	0,99702	0,99711	0,99720	0,99728	0,99736
2,8	0,99744	0,99752	0,99760	0,99767	0,99774	0,99781	0,99788	0,99795	0,99801	0,99807
2,9	0,99813	0,99819	0,99825	0,99831	0,99836	0,99841	0,99846	0,99851	0,99856	0,99861
3,0	0,99865	0,99869	0,99874	0,99878	0,99882	0,99886	0,99889	0,99893	0,99896	0,99900
3,1	0,99903	0,99906	0,99910	0,99913	0,99916	0,99918	0,99921	0,99924	0,99926	0,99929
3,2	0,99931	0,99934	0,99936	0,99938	0,99940	0,99942	0,99944	0,99946	0,99948	0,99950
3,3	0,99952	0,99953	0,99955	0,99957	0,99958	0,99960	0,99961	0,99962	0,99964	0,99965
3,4	0,99966	0,99968	0,99969	0,99970	0,99971	0,99972	0,99973	0,99974	0,99975	0,99976
3,5	0,99977	0,99978	0,99978	0,99979	0,99980	0,99981	0,99981	0,99982	0,99983	0,99983
3,6	0,99984	0,99985	0,99985	0,99986	0,99986	0,99987	0,99987	0,99988	0,99988	0,99989
3,7	0,99989	0,99990	0,99990	0,99990	0,99991	0,99991	0,99992	0,99992	0,99992	0,99992
3,8	0,99993	0,99993	0,99993	0,99994	0,99994	0,99994	0,99994	0,99995	0,99995	0,99995
3,9	0,99995	0,99995	0,99996	0,99996	0,99996	0,99996	0,99996	0,99996	0,99997	0,99997
4,0	0,99997	0,99997	0,99997	0,99997	0,99997	0,99997	0,99998	0,99998	0,99998	0,99998

Nota: En el interior de la tabla se da la probabilidad de que la variable aleatoria Z, con distribución N(0,1), esté por debajo del valor z.

Utilización de la tabla de distribución normal estándar

Veamos ejemplos de cómo calcular probabilidades de una variable normal estándar $Z = N(0,1)$

La normal $Z = N(0,1)$ se encuentra tabulada, para valores a partir de 0 y hasta 3'99. Si por ejemplo queremos calcular $P(Z \leq 2,78)$, hemos de realizar los pasos:

1. Buscar la parte entera y las décimas en la primera columna (en este caso 2,7).
2. Buscar las centésimas en la primera fila (en este caso 8).
3. En el punto común a la fila y la columna que hemos encontrado, tenemos la probabilidad buscada, en este caso 0,9973.

Por tanto $P(Z \leq 2,78)$ = 0.9973.

Si queremos calcular una probabilidad de un valor mayor que 3'99, basta fijarse en que las probabilidades correspondientes a valores tales como 4,1 y mayores ya valen 0,9999 (prácticamente 1). Por eso, para estos valores mayores que 4,1, diremos que la probabilidad es aproximadamente 1. Así: $P(Z \leq 5,83) \simeq 1$ aunque no aparezca en la tabla.

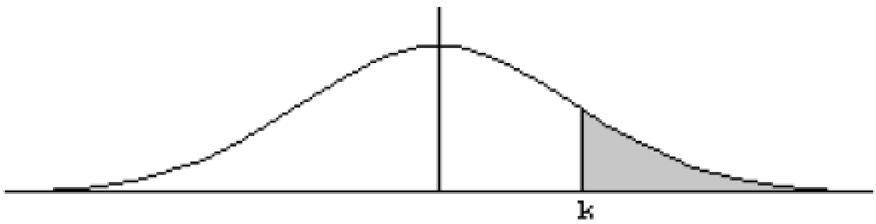
Ejemplo: Comprobar en la tabla que $P(Z \leq 1,31) = 0,9049$

Por otra parte, fijémonos en que en este tipo de distribuciones no tiene sentido plantearse probabilidades del tipo $P(Z = k)$ ya que siempre valen 0, al no encerrar ningún área. Por tanto, si nos pidiesen $P(Z = 2,69)$, basta decir que $P(Z = 2,69) = 0$.

Con esto anterior, podemos decir que $P(Z \leq k)$ = $P(Z < k)$, pues incluir o no k, no influye en el área. Nosotros por convenio siempre usaremos los signos $\leq$ o $\geq$

Veamos ahora el proceso para calcular otras probabilidades:

a) Si k es positivo y queremos calcular $P(Z \geq k)$, es decir el área sombreada siguiente:

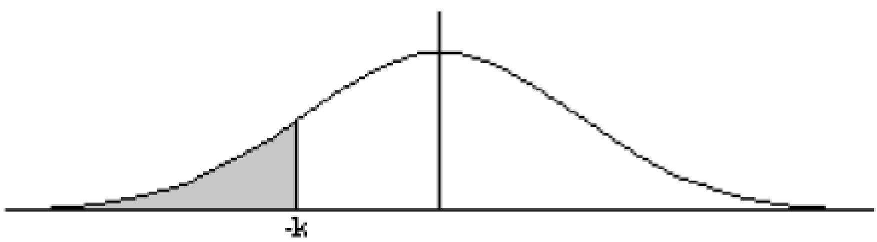


Basta pasar al complementario, es decir, $P(Z \geq k) = 1 - P(Z \leq k)$, que ya podemos buscar en la tabla

Ejemplo: Calcular $P(Z \geq 1,64)$.

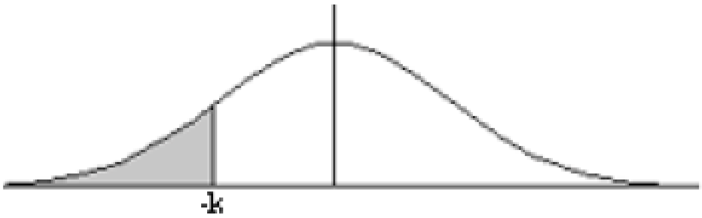
Teniendo en cuenta lo anterior tenemos que $P(Z \geq 1,64) = 1 - P(Z \leq 1,64) \Rightarrow P(Z \geq 1,64) = 1 - 0,9495 = 0,0505$

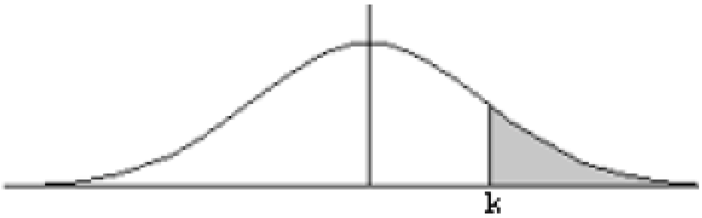
b) Si k es positivo y queremos calcular $P(Z \leq -k)$, es decir el área sombreada siguiente:



Por simetría vemos que es igual al área del caso anterior, $P(Z \le -k)$ = $P(Z \ge k) = 1 - P(Z \le k)$

Lo cual lo observamos en la comparación de las siguientes figuras

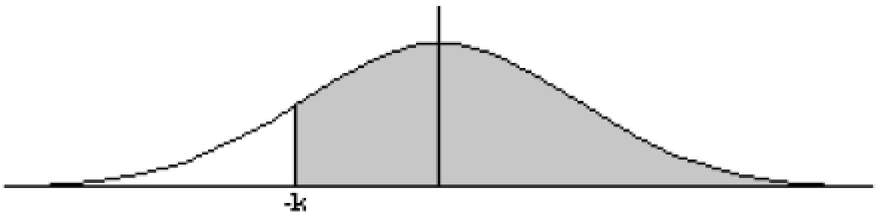




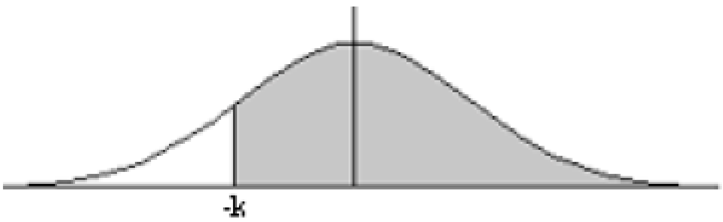
<u>Ejemplo</u>: Calcular $P(Z \le -1,64)$.

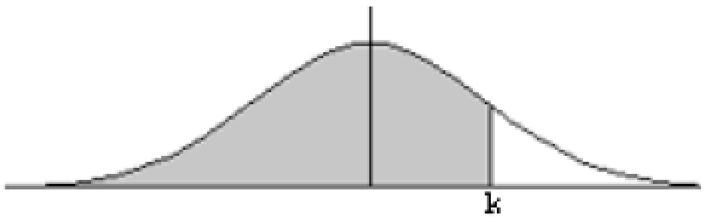
Teniendo en cuenta lo anterior tenemos que $P(Z \le -1,64) = P(Z \ge 1,64) = 1 - P(Z \le 1,64)$
$\Rightarrow P(Z \ge 1,64) = 1 - 0,9495 = 0,0505$

c) Si k es positivo y queremos calcular $P(Z \ge -k)$, es decir el área sombreada siguiente:



Por simetría vemos que $P(Z \ge -k) = P(Z \le k)$ como se observa en la siguiente figura

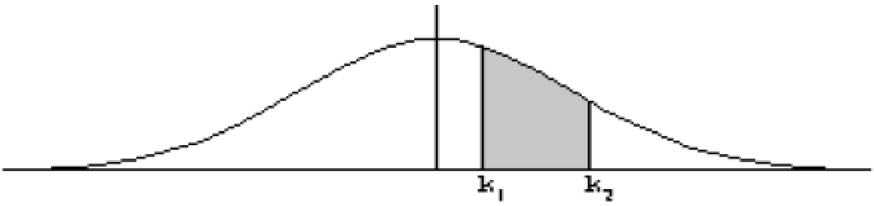




<u>Ejemplo</u>: Calcular $P(Z \ge -0,74)$.

Por lo anterior, tenemos que $P(Z \ge -0,74) = P(Z \le 0,74)$ y usando la tabla, $P(Z \ge -0,74) = 0,7704$

d) Probabilidades comprendidas entre dos valores $P(k_1 \le Z \le k_2)$, es decir el área sombreada siguiente:



En este caso se calcula restando dos áreas $P(k_1 \leq Z \leq k_2) = P(Z \leq k_2) - P(Z \leq k_1)$

Ejemplo: Calcula $P(-1,2 \leq Z \leq 0,82)$

Por lo dicho anteriormente $P(-1,2 \leq Z \leq 0,82) = P(Z \leq 0,82) - P(Z \leq -1,2)$

Calculamos usando las tablas y las simetrías: $\begin{cases} P(Z \leq 0,82) = 0,7939 \\ P(Z \leq -1,2) = P(Z \geq 1,2) = 1 - P(Z \leq 1,2) = 1 - 0,849 = 0,151 \end{cases}$

Por tanto, $P(-1,2 \leq Z \leq 0,82) = 0,7939 - 0,151 = 0,6429$

Cálculo de probabilidades en normales $N(\mu,\sigma)$

Si tenemos una distribución normal cualquiera $X = N(\mu,\sigma)$, entonces para calcular probabilidades se hace un cambio de variable que se llama *tipificación*. El cambio a realizar es $Z = \dfrac{X-u}{\sigma}$ que es una normal estándar $N(0,1)$

Veamos con un ejemplo como operar.

Ejemplo: Las estaturas de 600 soldados se distribuyen de acuerdo a una distribución normal de media 168 cm y desviación típica 8 cm. ¿Cuántos soldados miden más de 170 cm? ¿Cuántos soldados miden entre 166 y 170 cm?

Tenemos una $X = N(168,8)$, y nos pregunta en primer lugar: $P(X \geq 170)$ Si tipificamos tenemos que

$$P(X \geq 170) = P(\frac{X-168}{8} \geq \frac{170-168}{8}) = P(Z \geq 0,25) = 1 - P(Z \leq 0,25) = 1 - 0,5987 \Rightarrow P(X \geq 170) = 0,4013$$

El 40,13% de 600 soldados es 240,78 soldados, es decir, 241 soldados.

En segundo lugar, nos pide calcular $P(166 \leq X \leq 170)$. Pasamos a tipificar la variable para convertirla en una $Z = N(0,1)$

Por tanto $P(166 \leq X \leq 170) = P(\frac{166-168}{8} \leq \frac{X-168}{8} \leq \frac{170-168}{8}) = P(-0,25 \leq Z \leq 0,25)$

Y como $P(-0,25 \le Z \le 0,25) = P(Z \le 0,25) - P(Z \le -0,25)$ Calculamos cada término usando las tablas

$$\begin{cases} P(Z \le 0,25) = 0,5987 \\ P(Z \le -0,25) = P(Z \ge 0,25) = 1 - P(Z \le 0,25) = 1 - 0,5987 = 0,4013 \end{cases}$$

Por tanto $P(166 \le X \le 170) = P(-0,25 \le Z \le 0,25) = 0,5987 - 0,4013 = 0,1974$

Ejemplo: Sabiendo que tenemos una distribución normal estándar, $Z = N(0,1)$, y que se cumple que $P(Z \ge k) = 0,2010$. Vamos a calcular k

Como $P(Z \ge k) = 1 - P(Z \le k) = 0,2010 \Rightarrow P(Z \le k) = 1 - 0,2010 \Rightarrow P(Z \le k) = 0,7990$

Buscamos la cantidad 0,7990 en la tabla, que no se encuentra, pero está comprendido entre los valores 0.7967 (que corresponde a 0,83) y 0.7995 (que corresponde a 0,84). Para k se suele tomar la media aritmética de estos dos valores

$$k = \frac{0,83 + 0,84}{2} = 0,835$$

NOTA: En este caso hubiese sido más favorable o aproximado tomar k = 0,84 puesto que el valor 0,7995 está muy próximo a 0,7990, en comparación con lo cercano que está 0,7967

5. RELACIÓN ENTRE LA DISTRIBUCIÓN BINOMIAL Y LA DISTRIBUCIÓN NORMAL

Es un hecho comprobado que cuando tenemos una distribución binomial $X = B(n, p)$ a medida que n crece, es difícil hacer uso de las fórmulas y/o tablas.

Por ejemplo, tiramos un dado 100 veces, calcular la probabilidad de obtener entre 20 y 33 cincos (inclusive).

Si Éxito = “obtener cinco”, entonces $p = \frac{1}{6}$ y Fracaso = “no obtener cinco” y $q = \frac{5}{6}$

Tenemos una $X = B(100, \frac{1}{6})$, y nos piden $P(20 \le X \le 33)$

Es inviable aplicar la fórmula pues son cálculos muy complicados. ¿Cómo resolver el problema?

Del siguiente modo:

Teorema Central del Límite:

La distribución binomial $X = B(n, p)$ se aproxima a una curva normal de media $\mu = n \cdot p$ y desviación típica $\sigma = \sqrt{n \cdot p \cdot q}$, cuando $n \to +\infty$ n, es decir, cuando *n* se hace muy grande.

La aproximación se puede aplicar (es una buena aproximación) sólo si *n* es grande, en concreto $n \ge 30$ y además $n \cdot p \ge 5$ y $n \cdot q \ge 5$. Si no se cumplen estas condiciones NO podemos aproximar la binomial que tengamos por una distribución normal.

Ejemplo: Dada la binomial anterior $X = B(100, \frac{1}{6})$, como $n = 100 \geq 30$, $n \cdot p = 100 \cdot \frac{1}{6} = 16,67 \geq 5$ y $n \cdot q = 100 \cdot \frac{5}{6} = 83,33 \geq 5$ cumple las condiciones, podemos aproximarla por una

$$X = B(100, \frac{1}{6}) \approx Y = N\left(100 \cdot \frac{1}{6}, \sqrt{100 \cdot \frac{1}{6} \cdot \frac{5}{6}}\right) = N\left(16,67; 3,73\right)$$

APRECIACIONES

En caso de que podamos aproximar, debemos tener en cuenta que estamos pasando de una variable discreta (binomial) a una continua (normal), y por tanto son distribuciones diferentes. El "precio" que hay que pagar por pasar de una a otra se denomina "corrección por continuidad" y consiste en hacer determinados ajustes para que la aproximación realizada sea lo más precisa posible.

Así, si nos piden $P(X = k)$ en una distribución binomial X, y aproximamos X por una distribución normal Y, no podemos calcular directamente $P(Y = k)$ porque, como ya se ha comentado anteriormente, en una distribución continua todas estas probabilidades valen 0. La corrección por continuidad consiste en tomar un pequeño intervalo de longitud 1 alrededor del punto k.

De otro modo, si nos piden $P(X = k)$ con *X* binomial, con la aproximación normal *Y* deberemos calcular $P(k - 0,5 \leq Y \leq k + 0,5)$

Del mismo modo se razona en el caso de probabilidades acumuladas en la binomial.

Ejemplos:

Si nos piden $P(X < k)$ con X binomial, aproximando por Y normal calcularemos $P(Y \leq k - 0.5)$

La explicación de que haya que restar $0,5$ y no sumarlo es que queremos que X sea menor estrictamente que k, con lo cual, si sumase $0,5$, el propio k aparecería en la probabilidad a calcular y NO debe aparecer.

Por contra, si debiésemos calcular $P(X \leq k)$, con X binomial, fijémonos que ahora k SÍ está incluido en la probabilidad y por tanto al aproximar por la normal *Y* deberíamos calcular $P(Y \leq k + 0,5)$.

Comprender estos dos hechos es fundamental para realizar bien la corrección por continuidad al aproximar una distribución binomial por una normal.

En el caso anterior, $\mu = n \cdot p = 100 \cdot \frac{1}{6} = 16,67$ y $\sigma = \sqrt{n \cdot p \cdot q} = \sqrt{100 \cdot \frac{1}{6} \cdot \frac{5}{6}} = 3,73$

De modo que, como $n \geq 30$, $n \cdot p = 16,67 \geq 5$ y $n \cdot q = 83,33 \geq 5$ n se pude aproximar la binomial por la normal, es decir:

$$X = B\left(100; \frac{1}{6}\right) \approx Y = N\left(16,67; 3,73\right)$$

Entonces:

$$P(20 \leq X \leq 33) = P(20 - 0,5 \leq Y \leq 33 + 0,5) \text{ (tipificamos)} = P(\frac{19,5 - 16,67}{3,73} \leq \frac{Y - 16,67}{3,73} \leq \frac{33,5 - 16,67}{3,73})$$

$$= P(0,89 \leq Z \leq 4,51) = P(Z \leq 4,51) - P(Z \leq 0,89) = 1 - 0,8133 = 0,1867$$