

VARIABLES ALEATORIAS. BINOMIAL E NORMAL

Índice

1. Variables aleatorias	1
2. Variables aleatorias discretas: a distribución binomial	3
2.1. Características dunha variable aleatoria discreta	3
2.2. Distribución binomial	4
3. Variables aleatorias continuas: a distribución normal.....	6
3.1. Función de distribución dunha variable aleatoria continua	6
3.2. Función de densidade dunha variable aleatoria continua.....	7
3.3. Distribución normal	8
4. Aproximación da distribución binomial pola normal.....	10
5. Táboa da función de probabilidade da distribución binomial	12
6. Táboa da distribución normal estándar $N(0, 1)$	14

1. Variables aleatorias

Tíranse dúas moedas e cóntase o número de caras, este número pode ser 0, 1 ou 2. Repítese a experiencia 5000 veces (en realidade, simúlase a experiencia cun ordenador) e na táboa seguinte anotouse a cantidade de veces que saíu cada un dos resultados, é dicir, a frecuencia absoluta de cada suceso:

número de caras	0	1	2
frecuencia absoluta	1222	2509	1269

Como a experiencia repetiuse 5000 veces, dividindo as frecuencias absolutas por este número, obtéñense as frecuencias relativas, e a nova táboa é

número de caras	0	1	2
frecuencia relativa	0'2444	0'5018	0'2538

Este é un exemplo de **distribución de frecuencias** (relativas neste caso.) Obsérvase que a suma das frecuencias relativas é 1, que é a frecuencia relativa do suceso seguro: $f_r(0) + f_r(1) + f_r(2) = 0'2444 + 0'5018 + 0'2538 = 1$

Segundo se viu, se se aumenta indefinidamente o número de veces que se repite o experimento, as frecuencias relativas acabarán tendendo á probabilidade de cada un dos sucesos. Dado que o espazo mostral que se obtén ao tirar dúas moedas é $E = \{cc, c+, +c, ++\}$, as probabilidades dos diferentes números de caras son: $P(0) = 0'25$, $P(1) = 0'5$ e $P(2) = 0'25$.

Tense agora, unha **distribución de probabilidade**, que é o límite da distribución de frecuencias relativas cando o número de veces que se repite a experiencia tende a infinito.

número de caras	0	1	2
probabilidade	0'25	0'5	0'25

Tamén a suma de todas as probabilidades é 1.

Tanto a distribución de frecuencias, como a distribución de probabilidade están asociadas a unha variable aleatoria, que neste caso é $X = \text{"número de caras"}$. En xeral, unha **variable aleatoria** é unha lei ou función que toma valores asociados a un certo experimento aleatorio.

O exemplo anterior, a variable $X = \text{"número de caras"}$, é un exemplo de **variable aleatoria discreta**. Chámase así ás variables aleatorias que só poden tomar un certo número de valores, ben finito, $X = 0, 1, 2$, como o exemplo anterior, ou se o número é infinito, de maneira que entre cada dous non haxa posibilidade de atopar un terceiro, por exemplo, se os valores da variable son $X = 0, 1, 2, 3, \dots$. Con todo, os exemplos de variables aleatorias discretas que se estudarán serán sempre do primeiro tipo, é dicir, cun número de valores posibles finito.

Cando a variable pode tomar todos os valores dun certo intervalo de números reais, entón dise que é unha **variable aleatoria continua**. Por exemplo, se se elixe unha xudía verde ao azar e se mide a súa lonxitude en centímetros, a variable $X = \text{"lonxitude en cm"}$, é unha variable aleatoria continua.

Exemplo:

Das seguintes variables aleatorias, indicar cal é discreta e cal é continua:

- O número de chamadas telefónicas que recibe unha determinada central entre as 9 e as 10 horas dun día elixido ao azar.
Pode que o número de chamadas durante esa hora sexa 0, 1, 2, 3, ... Polo tanto, é unha variable aleatoria discreta.
- Tíranse dous dados e súmanse as súas puntuacións.
As posibles sumas son 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12. Polo tanto, é unha variable aleatoria discreta.
- A hora á que chega o autobús a unha parada, da que se sabe que está comprendida entre as 5 e as 5 e media da tarde.
Neste caso, os posibles valores da variable son todos os números reais comprendidos no intervalo $[5, 5'5]$ (entendendo por 5'5 as cinco e media). Entón, trátase dunha variable aleatoria continua.
- O peso dunha persoa elixida ao azar.
Os valores desta variable poden ser dende uns poucos gramos (o peso dun acabado de nacer) ata o peso máximo dunha persoa, que nalgúns casos extremos como se sabe, pode chegar a ser de varios centenares de quilogramos. Polo tanto, é unha variable continua.
- A contía do primeiro premio dun sorteo de lotería primitiva en euros.
Como se sabe, a contía do primeiro premio dun sorteo de lotería primitiva depende da recadación do sorteo e do número de gañadores do primeiro premio. Entón, os valores que pode tomar inclúen un amplo intervalo de posibilidades e, aínda que realmente a fracción máis pequena é un céntimo, pódese considerar a efectos prácticos unha variable aleatoria continua.

- O número de coches dunha familia escollida ao azar.
O número de coches dunha familia calquera sempre será un número enteiro, maior ou igual que 0. Polo tanto, trátase dunha variable aleatoria discreta.

2. Variables aleatorias discretas: a distribución binomial

2.1. Características dunha variable aleatoria discreta

Vólvese ao exemplo do lanzamento de dúas moedas. Se se considera a variable aleatoria X = "número de caras", esta variable aleatoria discreta podía tomar os valores $X = 0, 1, 2$. A función que asigna a cada valor da variable a súa probabilidade correspondente chámase **función de probabilidade**. Neste caso, como se viu antes, $P(X = 0) = 0'25$, $P(X = 1) = 0'5$ e $P(X = 2) = 0'25$.

Se se denota por x_i os valores que pode tomar a variable, e p_i as probabilidades respectivas de cada un dos valores, é dicir, $P(X = x_i) = p_i$; a función de probabilidade pódese representar na táboa seguinte:

x_i	p_i
0	0'25
1	0'5
2	0'25

Para que unha función de probabilidade dunha variable aleatoria discreta estea ben definida é preciso que a suma de todas as probabilidades sexa 1, é dicir, $\sum p_i = 1$.

Do mesmo xeito que as variables estatísticas, as variables aleatorias tamén teñen asociados certos parámetros que permiten describilas. Vaise estudar a continuación.

Chámase **media, esperanza matemática ou valor esperado** dunha variable aleatoria discreta X ao número $\mu = \sum x_i \cdot p_i$.

É dicir, a suma dos produtos dos valores da variable x_i polas súas probabilidades correspondentes p_i . No exemplo, $\mu = \sum x_i \cdot p_i = 0 \cdot 0'25 + 1 \cdot 0'5 + 2 \cdot 0'25 = 1$.

Este valor é o análogo á media $\bar{x}$ dunha variable estatística unidimensional. É o valor que se espera obter como media se o experimento aleatorio se repite un número de veces moi alto.

Chámase **varianza** dunha variable aleatoria discreta X ao número

$$\sigma^2 = \sum (x_i - \mu)^2 \cdot p_i.$$

A varianza dunha variable aleatoria é análoga á varianza dunha variable estatística.

Tamén se pode simplificar o seu cálculo mediante a fórmula $\sigma^2 = \sum x_i^2 \cdot p_i - \mu^2$.

No exemplo, $\sigma^2 = \sum x_i^2 \cdot p_i - \mu^2 = 0^2 \cdot 0'25 + 1^2 \cdot 0'5 + 2^2 \cdot 0'25 - 1^2 = 0'5$.

Por último, chámase **desviación típica** dunha variable aleatoria discreta X á raíz cadrada da súa varianza, é dicir, $\sigma = \sqrt{\sum x_i^2 \cdot p_i - \mu^2}$.

Que no exemplo é $\sigma = \sqrt{0'5} = 0'707$.

2.2. Distribución binomial

Tense un experimento aleatorio que só produce dúas posibilidades que vanse chamar **éxito** e **fracaso**. A probabilidade de éxito é p e a de fracaso é $1 - p = q$. Repítese a experiencia n veces, cada vez que se repite o resultado é independente do anterior. Se se chama $X = \text{"número de éxitos dos } n\text{"}$, entón dise que X é unha variable aleatoria que segue unha **distribución binomial**, o que se indicará da forma $X \sim B(n, p)$, onde n é o número de veces que se repite a experiencia e p é a probabilidade de obter éxito cada vez que se repite. É evidente que a binomial é unha variable aleatoria discreta, xa que os valores que pode tomar son $X = 0, 1, 2, 3, \dots, n$.

Quizais a situación máis sinxela na que aparece a binomial sexa no lanzamento dunha moeda. Por exemplo, lánzase unha moeda $n = 4$ veces. Cada vez que se lanza pode saír cara con probabilidade $p = \frac{1}{2}$, ou cruz con probabilidade $q = 1 - p = \frac{1}{2}$. Se se considera éxito ao feito de saír cara, a variable aleatoria $X = \text{"número de caras segue"} una binomial $B\left(4, \frac{1}{2}\right)$.$

Vaise calcular a súa función de probabilidade:

$P(X = 0) = P(4 \text{ cruces}) = P(+ + + +) = \left(\frac{1}{2}\right)^4$; xa que as catro probas son independentes.

$P(X = 1) = P(1 \text{ cara e 4 cruces}) = P(c + + +) + P(+ c + +) + P(+ + c +) + P(+ + + c) = 4 \cdot \left(\frac{1}{2}\right)^4$; onde $c + + +, + c + +, + + c +, + + + c$ son as diferentes ordes en que poden aparecer a cara e as cruces.

$P(X = 2) = P(2 \text{ caras e 2 cruces}) =$

$P(c c + +) + P(+ c + c) + P(c + c +) + P(+ + c c) + P(c + + c) + P(+ c c +) = 6 \cdot \left(\frac{1}{2}\right)^4$;

tamén aquí consideráronse as diferentes formas de ordenar as dúas cara e as dúas cruces, que son seis en total.

$P(X = 3) = P(3 \text{ caras e 1 cruz}) = P(1 \text{ cara e 3 cruces}) = 4 \cdot \left(\frac{1}{2}\right)^4$.

$P(X = 4) = P(4 \text{ caras}) = P(4 \text{ cruces}) = \left(\frac{1}{2}\right)^4$.

Por unha banda, pódese comprobar que a suma destas probabilidade é 1, co que a función de probabilidade está ben definida; e por outra, os factores polos que aparece multiplicado $\left(\frac{1}{2}\right)^4$ en cada caso son 1, 4, 6, 4, 1, que é a cuarta fila do triángulo de Tartaglia.

Lémbrese que o **triángulo de Tartaglia** é o seguinte, no que cada número obtense mediante a suma dos dous que están inmediatamente sobre el.

			1		1			
		1		2		1		
	1		3		3		1	
	1	4		6		4		1
1		5	10		10	5		1
				...				

Por outra banda, os elementos de calquera fila n ésima deste triángulo pódense calcular mediante os números combinatorios $\binom{n}{0}, \binom{n}{1}, \binom{n}{2}, \dots, \binom{n}{n}$ onde cada número combinatorio calcúlase $\binom{n}{r} = \frac{n!}{r!(n-r)!}$

En xeral, pódese comprobar que a función de probabilidade da binomial $B(n, p)$ vén dada por

$$P(X = k) = \binom{n}{k} \cdot p^k \cdot q^{n-k}$$

que é a probabilidade de que se produzan k éxitos, con $k = 0, 1, 2, \dots, n$.

A dificultade á hora de resolver un problema asociado á binomial non está na utilización da fórmula, senón na identificación da binomial adecuada.

Exemplos:

- Nunha biblioteca o 10 % dos libros son de matemáticas. Elíxense 5 libros ao azar. Cal é a probabilidade de que se collesen máis de 3 libros de matemáticas?

Cada vez que se escolle un libro tense unha probabilidade $p = 0'10$ de elixir un de matemáticas, e unha probabilidade $q = 1 - p = 0'90$ de elixir un libro que non sexa de matemáticas. Se se considera a variable aleatoria $X =$ “número de libros de matemáticas elixidos”, a variable X segue unha binomial $B(5, 0'1)$.

Quérese calcular a probabilidade de que haxa máis de 3 de matemáticas, é dicir, $P(X > 3) = P(X = 4) + P(X = 5)$. Polo tanto, hai que utilizar a fórmula da función de probabilidade da binomial dúas veces: $P(X > 3) = P(X = 4) + P(X = 5) =$

$$= \binom{5}{4} \cdot 0'1^4 \cdot 0'9^1 + \binom{5}{5} \cdot 0'1^5 \cdot 0'9^0 = 0'00045 + 0'00001 = 0'00046$$

Como cabía esperar, unha probabilidade moi baixa. Pois xa era moi difícil, con só un 10 %, elixir un só libro de matemáticas.

- Nunha cidade, o 40 % dos estudantes que promocionan a bacharelato ten suspensa algunha materia de 4.º de ESO. Elíxense 6 estudantes de 1.º de Bacharelato ao azar, cal é a probabilidade de que a metade deles teña algunha materia suspensa de ESO?

Verifícase que en cada proba hai dous únicos resultados: ter algunha suspensa ou non. Ademais o resultado de cada proba é independente do anterior. Se se considera a variable aleatoria $X =$ “número de estudantes que ten algunha materia suspensa”, a probabilidade de atopar un estudante con algún suspenso é constante $p = 0'4$. É, polo tanto, unha distribución binomial de parámetros $n = 6$ e $p = 0'4$: $B(6, 0'4)$.

A metade de 6 é 3, a probabilidade buscada é $P(X = 3) = \binom{6}{3} \cdot 0'4^3 \cdot 0'6^3 = 0'2765$.

Táboa da distribución binomial

Os valores da función de probabilidade dunha distribución binomial con $n = 10$ e p múltiplo de 0'05 están tabulados. Os resultados do exemplo anterior pódense atopar na táboa da distribución binomial que se inclúe ao final do documento.

n	k	p		
		...	0,4	...
...	...	...	↓	...
6	0	...	...	...
	1	...	...	...
	2	...	...	...
	3	→	0,2765	...

Parámetros:

- **Media:** $\mu = n \cdot p$
- **Varianza:** $\sigma^2 = n \cdot p \cdot q$
- **Desviación típica:** $\sigma = \sqrt{n \cdot p \cdot q}$

3. Variables aleatorias continuas: a distribución normal

3.1. Función de distribución dunha variable aleatoria continua

Supóñase que se ten un grupo moi numeroso de alumnos e alumnas que coñeceron a nota final dunha determinada materia. Supóñase tamén que esta nota determinouse tendo en consideración moitos factores, e non unicamente un exame, por exemplo, a entrega de traballos, a participación na clase, a realización de varios controis etc. Todo iso fará que, se a cualificación está comprendida entre 0 e 10 como adoita ser habitual, a nota final dun alumno concreto ou dunha alumna concreta poida ser case calquera número real do intervalo $[0, 10]$. Se se elixe un alumno ou unha alumna ao azar e se considera a súa nota, a variable aleatoria $X = \text{"nota final"}$ é un exemplo de variable aleatoria continua. Os posibles valores que pode tomar a variable X son todos os números reais do intervalo $[0, 10]$ (polo menos, teoricamente).

Agora ben, como son as probabilidades asociadas a esta variable? En primeiro lugar, dadas as posibilidades, non parece que teña moito sentido preguntarse pola probabilidade de que un alumno ou unha alumna elixida ao azar obtivese, por exemplo, un 4'589 de nota final. Terá sentido máis ben, preguntarse por un determinado rango de posibilidades, é dicir, terá sentido preguntarse pola probabilidade de que un alumno ou unha alumna obtivese unha nota superior a 5, ou entre 2 e 8, ou menor que 8 etc. Isto é unha característica das variables aleatorias continuas, as probabilidades puntuais son nulas, é dicir, $P(X = 4'589) = 0$, e o mesmo para calquera outro valor.

Por todo o anterior, nas variables aleatorias continuas, estarase máis interesado en como está distribuída a probabilidade ao longo do intervalo no que toma valores, que nas probabilidades de valores concretos, que son todas nulas. Esta distribución da probabilidade vén determinada polo que se denomina precisamente **función de distribución**, que se define:

$$F(x) = P(X \leq x)$$

Noutras palabras, $F(x)$ é a probabilidade de que a variable X tome un valor inferior ou igual a x .

Para entender mellor o significado da función de distribución, vaise tentar imaxinar como é a función de distribución da variable aleatoria do exemplo da nota final do exame.

En primeiro lugar, hai uns valores evidentes, $F(10) = P(X \leq 10) = 1$, xa que, con toda seguridade a nota de calquera alumno ou calquera alumna será menor ou igual que 10. E o mesmo ocorrerá para valores maiores que 10, é dicir, tamén a probabilidade de que alguén saque unha nota inferior a 11 será 1.

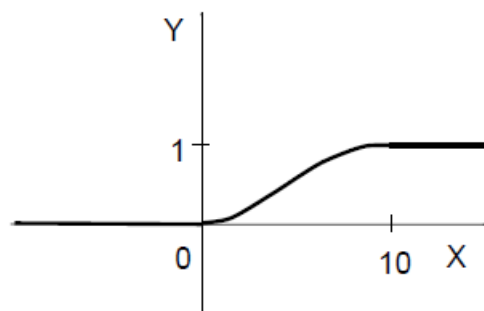
$F(0) = P(X \leq 0) = 0$ xa que, ninguén sacará unha nota inferior a 0, e teoricamente ninguén obterá 0, polo que se comentou antes das probabilidades puntuais (aínda que lamentablemente sábese que algunhas veces ocorre.) O que si ocorre, con toda seguridade, é que $F(-1) = P(X \leq -1) = 0$, e o mesmo para calquera outro valor negativo.

Xa se sabe entón que a función $F(x)$ vale 0 en $x = 0$ e vale 1 en $x = 10$. Só falta ver que ocorre entre 0 e 10.

Aínda que non se teñan datos da probabilidade de obter unha cualificación menor ou igual que 3, por exemplo, seguro que é máis probable obter unha nota menor ou igual que 4, e isto á súa vez menos probable que obter unha nota menor ou igual que 5 etc.

Noutras palabras, a medida que se aumenten as posibilidades de nota, aumentará a probabilidade. Isto fai que a función de distribución verifique $F(1) \leq F(2) \leq F(3) \leq F(4) \leq \dots$ e o mesmo para calquera outro valor. E, como ben se sabe, isto significa que a función de distribución é crecente. Concluindo, a función de distribución é crecente e crece dende 0 ao 10. Entón, unha posible gráfica sería a do debuxo.

O feito de que a función sexa continua dedúcese precisamente do feito de que a variable aleatoria tamén sexa continua. Se non o fose, a variable podería non tomar todos os valores do intervalo $[0, 10]$. (As variables aleatorias discretas tamén teñen unha función de distribución, que se define exactamente da mesma forma, pero naquel caso a función de distribución non é continua).



Salvo o do intervalo $[0, 10]$, que pode ser

calquera outro, todo o anterior pódese aplicar a calquera función de distribución de calquera variable aleatoria continua:

A función de distribución $F(x)$ asociada a unha variable aleatoria continua X é unha función continua, crecente, non nula (xa que é unha probabilidade) e a súa gráfica crece dende 0 a 1.

3.2. Función de densidade dunha variable aleatoria continua

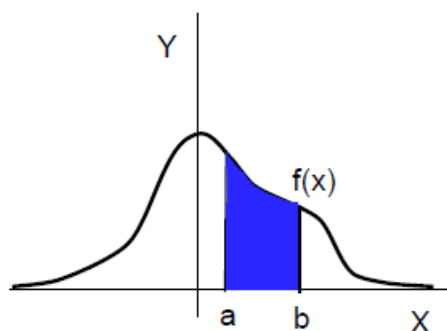
Nas variables aleatorias discretas utilizouse a función de probabilidade para calcular as probabilidades asociadas á variable. Para as variables aleatorias continuas dispónse da función de distribución, aínda que se utilizará máis a miúdo o equivalente á función de probabilidade das discretas, que aquí se chama **función de densidade** de probabilidade.

A función de densidade dunha variable aleatoria continua X non é máis que a derivada da súa función de distribución, denotarase por $f(x)$. Entón,

$$f(x) = F'(x)$$

Véxanse as características e propiedades desta nova función:

- Dado que a función de distribución $F(x)$ é crecente, $f(x) \geq 0$, xa que a derivada dunha función crecente sempre é maior ou igual que 0. Isto fai que a súa gráfica sempre se atope por enriba do eixe de abscisas.
- Pódese demostrar que as probabilidades asociadas á variable X son as áreas por debaixo da gráfica de $f(x)$, de feito verifícase que a probabilidade $P(a < X \leq b)$ é a área que hai entre o eixe OX e a gráfica de $f(x)$ no intervalo $[a, b]$. Expresarase da forma seguinte: $P(a < X \leq b) = \text{Área}(f(x), [a, b])$



- Admitindo a propiedade anterior, pódese deducir outra característica da función de densidade. Se a probabilidade é a área que hai por debaixo da curva, necesariamente a área total por debaixo da función de densidade sempre é 1.

Por outra banda, como as probabilidades puntuais son nulas, é dicir, $P(X = b) = 0$, á hora de calcular a probabilidade $P(a < X \leq b)$ é indiferente que as desigualdades sexan estritas ou non, é dicir,

$$P(a < X < b) = P(a < X \leq b) = P(a \leq X \leq b) = P(a \leq X < b)$$

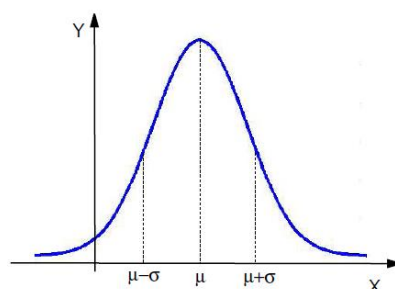
O único que importa é a área que hai por debaixo da gráfica, non se os extremos están ou non incluídos.

3.3. Distribución normal

A **distribución normal** é unha variable aleatoria continua que aparece asociada a multitude de fenómenos. Nun tempo pensouse que practicamente todos os fenómenos estatísticos podíanse axustar a esta distribución, de aí o nome que se lle puxo, "normal", de feito, se un fenómeno non a verificaba, dicíase que era un fenómeno "anormal". Hoxe en día sábese que, sobre todo aquelas variables nas que interveñen multitude de factores independentes, seguen unha variable, se non normal, moi parecida. Entre os exemplos que se poden citar atópanse: case todas as características naturais dos seres vivos, como a súa lonxitude ou o seu peso; multitude de fenómenos sociais e económicos como a taxa de natalidade, a distribución da renda etc. Todos son casos nos que interveñen multitude de factores independentes entre si. Por exemplo, na estatura dunha persoa inflúen factores xenéticos, hábitos alimenticios etc.

Dise que X é unha variable aleatoria continua que segue unha distribución normal de media μ e desviación típica σ , o que se denotará $X \sim N(\mu, \sigma)$, se a súa función de densidade é:

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



Propiedades da función de densidade:

- A función cumpre $f(x) \geq 0$, é dicir, a súa gráfica sempre se atopa por enriba do eixe X .
- Ten forma de campá, de feito adóitaselle chamar **campá de Gauss**.
- Ten un máximo absoluto no punto $x = \mu$.
- Ten dous puntos de inflexión que se atopan en $x = \mu - \sigma$ e $x = \mu + \sigma$.
- A gráfica é simétrica con respecto da recta vertical que pasa polo seu máximo.
- O eixe X é asíntota horizontal de $f(x)$.
- A área por debaixo da súa gráfica é 1.

Para calcular probabilidades asociadas a unha normal, sería preciso calcular áreas sobre os intervalos correspondentes por debaixo da gráfica anterior. Non obstante, isto non é sinxelo, entre outras cousas porque a integral da función de densidade da normal, que é a ferramenta coa que habería que calcular a área, non se pode calcular de maneira exacta. O único que se pode facer é aproximala numericamente. Con todo, pódese comprobar que se $X \sim N(\mu, \sigma)$, entón a variable

$$Z = \frac{X - \mu}{\sigma}$$

é una normal $N(0, 1)$, é dicir, unha normal de media $\mu = 0$ e desviación típica $\sigma = 1$. Á normal $N(0, 1)$ chámasele **normal estándar** e para este caso particular os valores das áreas, é dicir, os valores da súa función de distribución están tabulados, polo que non é preciso calculalos cada vez. O proceso mediante o cal se pasa dunha normal $X \sim N(\mu, \sigma)$ a unha normal estándar $Z \sim N(0, 1)$ chámase **tipificación da variable**.

Resumindo, para calcular probabilidades dunha variable $X \sim N(\mu, \sigma)$, primeiro hai que tipificar a variable, e despois hai que utilizar a táboa da normal estándar para determinar os valores da probabilidade.

Tipificación da variable

Sexa $X \sim N(3, 2)$ e supóñase que se quere calcular $P(X \leq 4)$, entón, na expresión $X \leq 4$ réstase a media e divídese pola desviación típica en ambos os lados da desigualdade:

$$P(X \leq 4) = P\left(\frac{X - 3}{2} \leq \frac{4 - 3}{2}\right) = P(Z \leq 0.5)$$

onde $Z \sim N(0, 1)$ que ao ser a normal estándar poderase calcular mediante os valores da táboa como se verá despois.

Se a probabilidade a calcular fose da forma $P(-2 < X < 3)$, por exemplo, teríase que restar a media e dividir pola desviación típica nos tres membros da desigualdade, da forma seguinte:

$$P(-2 < X < 3) = P\left(\frac{-2 - 3}{2} < \frac{X - 3}{2} < \frac{3 - 3}{2}\right) = P(-2.5 < Z < 0)$$

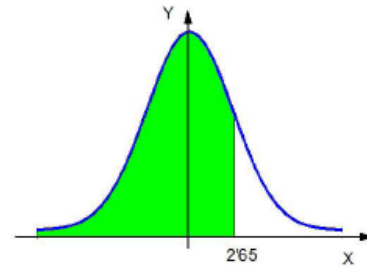
onde, novamente, $Z \sim N(0, 1)$.

Uso da táboa da normal estándar $N(0, 1)$

Supóñase entón que a variable xa está tipificada e tense que calcular unha probabilidade que está en termos de Z , con $Z \sim N(0, 1)$. A táboa da normal estándar, que se incluíu no apartado seguinte ten tabulados con catro decimais os valores das probabilidades (ou áreas, segundo a gráfica que se achega coa táboa) correspondentes a $P(Z \leq k)$ para valores de k comprendidos entre 0 e 3.89. Para valores de k maiores que 3.89, a probabilidade é practicamente 1, polo que non é preciso que estén na táboa. Vaise ir vendo con exemplos as diferentes posibilidades:

- $P(Z \leq 2'65)$

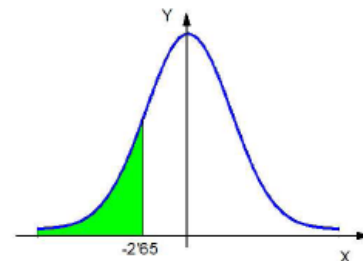
A área que se quiere calcular é a que se coloreou no debuxo. Esta área pódese atopar directamente na táboa, para iso hai que buscar o número no que coinciden a fila 2'6 e a columna 0'05, que é 0'9960. Polo tanto, $P(Z \leq 2'65) = 0'9960$.



- $P(Z \leq -2'65)$

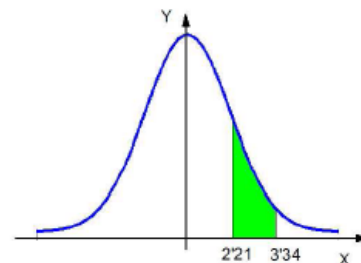
A área que se quiere calcular é a que se coloreou no debuxo. A táboa só ofrece valores da área á derecha de $Z = 0$. Non obstante, como a gráfica da función de densidade é simétrica, a área dende $-\infty$ ata $-2'65$ é igual que a área dende $2'65$ ata $+\infty$, e esta última pódese calcular mediante o contrario:

$$P(Z \leq -2'65) = P(Z > 2'65) = 1 - P(Z \leq 2'65) = 1 - 0'9960 = 0'0040.$$



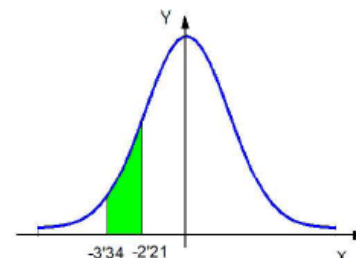
- $P(2'21 < Z \leq 3'34)$

A área que se quiere calcular agora é a que se coloreou no debuxo, comprendida entre 2'21 e 3'34. Para calcular esta área basta con restar á área ata 3'34, a área ata 2'21. É dicir,
 $P(2'21 < Z \leq 3'34) = P(Z \leq 3'34) - P(Z \leq 2'21) = 0'9996 - 0'9864 = 0'0132$.



- $P(-3'34 < Z \leq -2'21)$

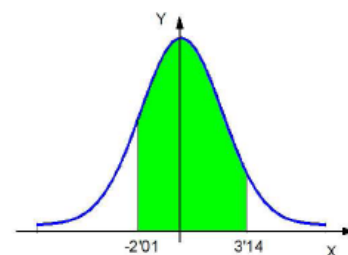
A área que se quiere calcular agora é a que se coloreou no debuxo, e debido á simetría da función de densidade, esta área é exactamente igual á do debuxo anterior, que xa se calculou antes. Polo tanto, $P(-3'34 < Z \leq -2'21) = P(2'21 < Z \leq 3'34) = 0'0132$.



- $P(-2'01 < Z \leq 3'14)$

A área que se quiere calcular agora é a que se coloreou no debuxo. Procédese como nos casos anteriores:

$$\begin{aligned} P(-2'01 < Z \leq 3'14) &= P(Z \leq 3'14) - P(Z \leq -2'01) = \\ &= P(Z \leq 3'14) - P(Z > 2'01) = \\ &= P(Z \leq 3'14) - (1 - P(Z \leq 2'01)) = 0'9970. \end{aligned}$$



4. Aproximación da distribución binomial pola normal

Se n é grande, o cálculo de probabilidades nunha distribución binomial $B(n, p)$ pode ser demasiado laborioso. Cando isto acontece a distribución binomial pódese aproximar por unha normal, pero cales son os parámetros desta distribución normal?

Pódese demostrar que cando n é suficientemente grande a distribución binomial $B(n, p)$ pódese aproximar pola normal $N(n \cdot p, \sqrt{n \cdot p \cdot (1-p)})$, sempre que $n \cdot p \geq 5$ e $n \cdot (1-p) \geq 5$.

É evidente que estas condicións ocorren canto maior sexa n e canto máis preto estea p de 0'5.

Cando X é unha $B(n, p)$ pódese aproximar por outra variable Y que sexa

$N(n \cdot p, \sqrt{n \cdot p \cdot (1-p)})$ pero, como unha é discreta e a outra continua, é preciso facer unha **corrección de continuidade** que se establece da forma seguinte:

$$P(X = x) = P(x - 0.5 \leq Y \leq x + 0.5)$$

$$P(X \leq x) = P(Y \leq x + 0.5)$$

$$P(X < x) = P(Y \leq x - 0.5)$$

$$P(X \geq x) = P(Y \geq x - 0.5)$$

$$P(X > x) = P(Y > x + 0.5)$$

Exemplo:

Sábase que un determinado fármaco produce efectos secundarios no 20 % dos pacientes que se tratan con el. Tómake unha mostra de 50 pacientes os que se lles administra o fármaco. Cal é a probabilidade de que haxa como máximo 8 pacientes que teñan efectos secundarios?

Tense que $X = \text{"número de pacientes que teñen efectos secundarios"} \sim B(50, 0.2)$.

A probabilidade buscada é:

$$P(X \leq 8) = \binom{50}{0} \cdot 0.8^{50} + \binom{50}{1} \cdot 0.2 \cdot 0.8^{49} + \dots + \binom{50}{8} \cdot 0.2^8 \cdot 0.8^{42}$$

É obvio que calcular os oito sumandos e logo sumar é laborioso.

Como $n \cdot p = 50 \cdot 0.2 = 10 > 5$ e $n \cdot (1 - p) = 50 \cdot 0.8 = 40 > 5$, pódese aproximar a

$B(50, 0.2)$ pola variable aleatoria Y que segue unha distribución

$$B(50 \cdot 0.2, \sqrt{50 \cdot 0.2 \cdot 0.8}) = N(10, 2.82).$$

Como a variable X é discreta e Y continua, faise unha corrección de 0.5, polo que

$$\text{resulta de modo aproximado que } P(X \leq 8) = P(Y \leq 8 + 0.5) = P\left(\frac{Y - 10}{2.82} \leq \frac{8.5 - 10}{2.82}\right) =$$

$$= P(Z \leq -0.53) = P(Z \geq 0.53) = 1 - P(Z \leq 0.53) = 1 - 0.7019 = 0.2981$$

5. Táboa da función de probabilidade da distribución binomial

$$P(X = k) = \binom{n}{k} \cdot p^k \cdot q^{n-k}$$

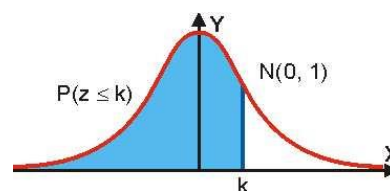
Exemplo: Se X é unha binomial de parámetros $n = 4$ e $p = 0'3$, a probabilidade de que X tome o valor 0 é **0'2401**, a de que tome o valor 1 é **0'4116**, o valor 2 é **0'2646**, o valor 3 é **0'0756** e o valor 4 é **0'0081**.

		valores de p									
n	k	0'05	0'1	0'15	0'2	0'25	0'3	0'35	0'4	0'45	0'5
1	0	0'9500	0'9000	0'8500	0'8000	0'7500	0'7000	0'6500	0'6000	0'5500	0'5000
	1	0'0500	0'1000	0'1500	0'2000	0'2500	0'3000	0'3500	0'4000	0'4500	0'5000
2	0	0'9025	0'8100	0'7225	0'6400	0'5625	0'4900	0'4225	0'3600	0'3025	0'2500
	1	0'0950	0'1800	0'2550	0'3200	0'3750	0'4200	0'4550	0'4800	0'4950	0'5000
	2	0'0025	0'0100	0'0225	0'0400	0'0625	0'0900	0'1225	0'1600	0'2025	0'2500
3	0	0'8574	0'7290	0'6141	0'5120	0'4219	0'3430	0'2746	0'2160	0'1664	0'1250
	1	0'1354	0'2430	0'3251	0'3840	0'4219	0'4410	0'4436	0'4320	0'4084	0'3750
	2	0'0071	0'0270	0'0574	0'0960	0'1406	0'1890	0'2389	0'2880	0'3341	0'3750
	3	0'0001	0'0010	0'0034	0'0080	0'0156	0'0270	0'0429	0'0640	0'0911	0'1250
4	0	0'8145	0'6561	0'5220	0'4096	0'3164	0'2401	0'1785	0'1296	0'0915	0'0625
	1	0'1715	0'2916	0'3685	0'4096	0'4219	0'4116	0'3845	0'3456	0'2995	0'2500
	2	0'0135	0'0486	0'0975	0'1536	0'2109	0'2646	0'3105	0'3456	0'3675	0'3750
	3	0'0005	0'0036	0'0115	0'0256	0'0469	0'0756	0'1115	0'1536	0'2005	0'2500
	4	0'0000	0'0001	0'0005	0'0016	0'0039	0'0081	0'0150	0'0256	0'0410	0'0625
5	0	0'7738	0'5905	0'4437	0'3277	0'2373	0'1681	0'1160	0'0778	0'0503	0'0313
	1	0'2036	0'3281	0'3915	0'4096	0'3955	0'3602	0'3124	0'2592	0'2059	0'1563
	2	0'0214	0'0729	0'1382	0'2048	0'2637	0'3087	0'3364	0'3456	0'3369	0'3125
	3	0'0011	0'0081	0'0244	0'0512	0'0879	0'1323	0'1811	0'2304	0'2757	0'3125
	4	0'0000	0'0005	0'0022	0'0064	0'0146	0'0284	0'0488	0'0768	0'1128	0'1563
	5	0'0000	0'0000	0'0001	0'0003	0'0010	0'0024	0'0053	0'0102	0'0185	0'0313
6	0	0'7351	0'5314	0'3771	0'2621	0'1780	0'1176	0'0754	0'0467	0'0277	0'0156
	1	0'2321	0'3543	0'3993	0'3932	0'3560	0'3025	0'2437	0'1866	0'1359	0'0938
	2	0'0305	0'0984	0'1762	0'2458	0'2966	0'3241	0'3280	0'3110	0'2780	0'2344
	3	0'0021	0'0146	0'0415	0'0819	0'1318	0'1852	0'2355	0'2765	0'3032	0'3125
	4	0'0001	0'0012	0'0055	0'0154	0'0330	0'0595	0'0951	0'1382	0'1861	0'2344
	5	0'0000	0'0001	0'0004	0'0015	0'0044	0'0102	0'0205	0'0369	0'0609	0'0938
	6	0'0000	0'0000	0'0000	0'0001	0'0002	0'0007	0'0018	0'0041	0'0083	0'0156
7	0	0'6983	0'4783	0'3206	0'2097	0'1335	0'0824	0'0490	0'0280	0'0152	0'0078
	1	0'2573	0'3720	0'3960	0'3670	0'3115	0'2471	0'1848	0'1306	0'0872	0'0547
	2	0'0406	0'1240	0'2097	0'2753	0'3115	0'3177	0'2985	0'2613	0'2140	0'1641
	3	0'0036	0'0230	0'0617	0'1147	0'1730	0'2269	0'2679	0'2903	0'2918	0'2734
	4	0'0002	0'0026	0'0109	0'0287	0'0577	0'0972	0'1442	0'1935	0'2388	0'2734
	5	0'0000	0'0002	0'0012	0'0043	0'0115	0'0250	0'0466	0'0774	0'1172	0'1641
	6	0'0000	0'0000	0'0001	0'0004	0'0013	0'0036	0'0084	0'0172	0'0320	0'0547
	7	0'0000	0'0000	0'0000	0'0000	0'0001	0'0002	0'0006	0'0016	0'0037	0'0078

		valores de p									
n	k	0'05	0'1	0'15	0'2	0'25	0'3	0'35	0'4	0'45	0'5
8	0	0'6634	0'4305	0'2725	0'1678	0'1001	0'0576	0'0319	0'0168	0'0084	0'0039
	1	0'2793	0'3826	0'3847	0'3355	0'2670	0'1977	0'1373	0'0896	0'0548	0'0313
	2	0'0515	0'1488	0'2376	0'2936	0'3115	0'2965	0'2587	0'2090	0'1569	0'1094
	3	0'0054	0'0331	0'0839	0'1468	0'2076	0'2541	0'2786	0'2787	0'2568	0'2188
	4	0'0004	0'0046	0'0185	0'0459	0'0865	0'1361	0'1875	0'2322	0'2627	0'2734
	5	0'0000	0'0004	0'0026	0'0092	0'0231	0'0467	0'0808	0'1239	0'1719	0'2188
	6	0'0000	0'0000	0'0002	0'0011	0'0038	0'0100	0'0217	0'0413	0'0703	0'1094
	7	0'0000	0'0000	0'0000	0'0001	0'0004	0'0012	0'0033	0'0079	0'0164	0'0313
	8	0'0000	0'0000	0'0000	0'0000	0'0000	0'0001	0'0002	0'0007	0'0017	0'0039
9	0	0'6302	0'3874	0'2316	0'1342	0'0751	0'0404	0'0207	0'0101	0'0046	0'0020
	1	0'2985	0'3874	0'3679	0'3020	0'2253	0'1556	0'1004	0'0605	0'0339	0'0176
	2	0'0629	0'1722	0'2597	0'3020	0'3003	0'2668	0'2162	0'1612	0'1110	0'0703
	3	0'0077	0'0446	0'1069	0'1762	0'2336	0'2668	0'2716	0'2508	0'2119	0'1641
	4	0'0006	0'0074	0'0283	0'0661	0'1168	0'1715	0'2194	0'2508	0'2600	0'2461
	5	0'0000	0'0008	0'0050	0'0165	0'0389	0'0735	0'1181	0'1672	0'2128	0'2461
	6	0'0000	0'0001	0'0006	0'0028	0'0087	0'0210	0'0424	0'0743	0'1160	0'1641
	7	0'0000	0'0000	0'0000	0'0003	0'0012	0'0039	0'0098	0'0212	0'0407	0'0703
	8	0'0000	0'0000	0'0000	0'0000	0'0001	0'0004	0'0013	0'0035	0'0083	0'0176
	9	0'0000	0'0000	0'0000	0'0000	0'0000	0'0000	0'0001	0'0003	0'0008	0'0020
10	0	0'5987	0'3487	0'1969	0'1074	0'0563	0'0282	0'0135	0'0060	0'0025	0'0010
	1	0'3151	0'3874	0'3474	0'2684	0'1877	0'1211	0'0725	0'0403	0'0207	0'0098
	2	0'0746	0'1937	0'2759	0'3020	0'2816	0'2335	0'1757	0'1209	0'0763	0'0439
	3	0'0105	0'0574	0'1298	0'2013	0'2503	0'2668	0'2522	0'2150	0'1665	0'1172
	4	0'0010	0'0112	0'0401	0'0881	0'1460	0'2001	0'2377	0'2508	0'2384	0'2051
	5	0'0001	0'0015	0'0085	0'0264	0'0584	0'1029	0'1536	0'2007	0'2340	0'2461
	6	0'0000	0'0001	0'0012	0'0055	0'0162	0'0368	0'0689	0'1115	0'1596	0'2051
	7	0'0000	0'0000	0'0001	0'0008	0'0031	0'0090	0'0212	0'0425	0'0746	0'1172
	8	0'0000	0'0000	0'0000	0'0001	0'0004	0'0014	0'0043	0'0106	0'0229	0'0439
	9	0'0000	0'0000	0'0000	0'0000	0'0000	0'0001	0'0005	0'0016	0'0042	0'0098
	10	0'0000	0'0000	0'0000	0'0000	0'0000	0'0000	0'0000	0'0001	0'0003	0'0010

6. Táboa da distribución normal estándar $N(0, 1)$

Os valores que aparecen na táboa corresponden ás áreas debaixo da curva normal ata o valor de k que aparece nesa fila (coa segunda cifra decimal correspondente á columna). É dicir, son as probabilidades de que a variable aleatoria sexa menor ou igual a ese valor de k .



Exemplo: A probabilidade de que unha variable aleatoria normal de media 0 e desviación típica 1, $N(0, 1)$, sexa menor ou igual a 0'6 é **0'7257** e de que sexa menor ou igual a 0'68 é de **0'7517**.

k	0	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	0'9965	0'9966	0'9967	0'9968	0'9969	0'9970	0'9971	0'9972	0'9973	0'9974
2'8	0'9974	0'9975	0'9976	0'9977	0'9977	0'9978	0'9979	0'9979	0'9980	0'9981
2'9	0'9981	0'9982	0'9982	0'9983	0'9984	0'9984	0'9985	0'9985	0'9986	0'9986
3'0	0'9987	0'9987	0'9987	0'9988	0'9988	0'9989	0'9989	0'9989	0'9990	0'9990
3'1	0'9990	0'9991	0'9991	0'9991	0'9992	0'9992	0'9992	0'9992	0'9993	0'9993
3'2	0'9993	0'9993	0'9994	0'9994	0'9994	0'9994	0'9994	0'9995	0'9995	0'9995
3'3	0'9995	0'9995	0'9995	0'9996	0'9996	0'9996	0'9996	0'9996	0'9996	0'9997
3'4	0'9997	0'9997	0'9997	0'9997	0'9997	0'9997	0'9997	0'9997	0'9997	0'9998
3'5	0'9998	0'9998	0'9998	0'9998	0'9998	0'9998	0'9998	0'9998	0'9998	0'9998
3'6	0'9998	0'9998	0'9999	0'9999	0'9999	0'9999	0'9999	0'9999	0'9999	0'9999
3'7	0'9999	0'9999	0'9999	0'9999	0'9999	0'9999	0'9999	0'9999	0'9999	0'9999
3'8	0'9999	0'9999	0'9999	0'9999	0'9999	0'9999	0'9999	0'9999	0'9999	0'9999