Capítulo 3

DISTRIBUCIÓN BINOMIAL Y DISTRIBUCIÓN NORMAL

3.1. Introducción

Estudiaremos en este tema dos de las distribuciones de probabilidad más importantes y que son imprescindibles a la hora de adentrarnos en el estudio de la inferencia estadística. La distribución binomial es uno de los primeros ejemplos de las llamadas distribuciones discretas (que sólo pueden tomar un número finito, o infinito numerable, de valores). Fue estudiada por Jakob Bernoulli (Suiza, 1654-1705), quién escribió el primer tratado importante sobre probabilidad, "Ars conjectandi" (El arte de pronosticar). Los Bernoulli formaron una de las sagas de matemáticos más importantes de la historia. La distribución normal es un ejemplo de las distribuciones continuas, y aparece en multitud de fenómenos sociales. Fue estudiada, entre otros, por J.K.F. Gauss (Alemania, 1777-1855), uno de los más famosos matemáticos de la historia. La gráfica de la distribución normal en forma de campana se denomina Campana de Gauss.

3.2. La distribución binomial o de Bernoulli

La distribución binomial está asociada a experimentos del siguiente tipo:

- Realizamos n veces cierto experimento en el que consideramos sólo la posibilidad de éxito o fracaso.

- La obtención de éxito o fracaso en cada ocasión es independiente de la obtención de éxito o fracaso en las demás ocasiones.

- La probabilidad de obtener éxito o fracaso siempre es la misma en cada ocasión.

Veámoslo con un *ejemplo*

Tiramos un dado 7 veces y contamos el número de cincos que obtenemos. ¿Cuál es la probabilidad de obtener tres cincos?.

Este es un típico ejemplo de distribución binomial, pues estamos repitiendo 7 veces el experimento de lanzar un dado. ¿Cuál es nuestro éxito?.

Evidentemente, sacar un 5, que es en lo que nos fijamos.

El fracaso, por tanto, será no sacar 5, sino sacar cualquier otro número.

Por tanto, Éxito = E = "sacar un 5" $\Longrightarrow p(E) = \dfrac{1}{6}$

Fracaso = F = "no sacar un 5" $\Longrightarrow p(F) = \dfrac{5}{6}$

Para calcular la probabilidad que nos piden, fijémonos en que nos dicen que sacamos 3 cincos y por lo tanto tenemos 3 éxitos y 4 fracasos, ¿de cuántas maneras pueden darse estas posibilidades?. Podríamos sacar 3 cincos en las 3 primeras tiradas y luego 4 tiradas sin sacar cinco, es decir: EEEFFFF Pero también podríamos sacar EFEFFFE, es decir que en realidad estamos calculando de cuántas maneras se pueden ordenar 4 fracasos y 3 éxitos. Recordando las técnicas combinatorias, este problema se reduce a calcular las permutaciones con elementos repetidos:

$$P_7^{3,4} = \frac{7!}{3! \cdot 4!} = \frac{7 \cdot 6 \cdot 5}{3 \cdot 2 \cdot 1} = 35\text{formas}$$

Y por tanto, como $p(E) = \dfrac{1}{6}$ y tengo 3 éxitos y $p(F) = \dfrac{5}{6}$ y tengo 4 fracasos:

$$p(\text{tener 3 éxitos y 4 fracasos}) = 35 \cdot \frac{1}{6} \cdot \frac{1}{6} \cdot \frac{1}{6} \cdot \frac{5}{6} \cdot \frac{5}{6} \cdot \frac{5}{6} \cdot \frac{5}{6} = 0'0781$$

Formalizando lo obtenido, en una variable binomial con 7 repeticiones y con probabilidad de éxito $\dfrac{1}{6}$, la probabilidad de obtener 3 éxitos es 0'0781, y lo expresaríamos:

$$\text{Bin}\left(7; \frac{1}{6}\right), \text{entonces } p(X = 3) = 0'0781$$

Como repetir este proceso sería bastante penoso en la mayoría de los casos, lo mejor es recurrir a la siguiente fórmula que expresa la probabilidad de obtener cierto número de éxitos en una distribución binomial:

Definición de distribución binomial:

Si realizamos n veces un experimento en el que podemos obtener éxito, E, con probabilidad p y fracaso, F, con probabilidad q ($q = 1 - p$), diremos que estamos ante una distribución binomial de parámetros n y p, y lo representaremos por Bin(n;p). En este caso la probabilidad de obtener k éxitos viene dada por:

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

Nota:

Observar que las probabilidades de éxito y fracaso son complementarias, es decir, q = 1-p y p = 1-q, por lo que basta saber una de ellas para calcular la otra.

Ejemplo:

Antes teníamos Bin$\left(7; \dfrac{1}{6}\right)$, y queríamos calcular p(X=3) (obtener 3 éxitos). Aplicando la fórmula:

$$p(X = 3) = \binom{7}{3} \cdot \left(\frac{1}{6}\right)^3 \cdot \left(\frac{5}{6}\right)^4 = 0'0781$$

Ejemplo:

Supongamos que la probabilidad de que una pareja tenga un hijo o una hija es igual. Calcular la probabilidad de que una familia con 6 descendientes tenga 2 hijos.

En este caso Éxito = E = "tener hijo" y p(E) = 0'5.

Fracaso = F = "tener hija" y p(F) = 0'5.

Estamos por tanto ante una binomial Bin(6;0'5) y nos piden p(X=2).

Si aplicamos la fórmula es:

$$p(X = 2) = \binom{6}{2} \cdot (0'5)^2 \cdot (0'5)^4 = 0'2344$$

Nota:

La elección de éxito o fracaso es subjetiva y queda a elección de la persona que resuelve el problema, pero teniendo cuidado de plantear correctamente lo que se pide. En el caso concreto del ejemplo anterior, si:

Éxito = "tener hija", como nos piden la probabilidad de que una familia con 6 hijos tenga 2 hijos, si el éxito es tener hija hemos de plantearnos cuál es la probabilidad de tener 4 éxitos (4 hijas), es decir:

$$p(X = 4) = \binom{6}{4} \cdot (0'5)^4 \cdot (0'5)^2 = 0'2344$$

Evidentemente sale lo mismo, pero hay que ser consecuente a la hora de elegir el éxito y el fracaso y la pregunta que nos hagan.

3.2.1. El uso de las tablas de la distribución binomial

La distribución binomial se encuentra tabulada por lo que es fácil calcular probabilidades sin necesidad de hacer demasiadas cuentas. Para usar las tablas de la distribución binomial es necesario conocer:

- El número de veces que se realiza el experimento (n).
- La probabilidad de éxito (p).
- El número de éxitos (k).

La probabilidad p se busca en la primera fila (valores desde 0'01 hasta 0'5).

El número de veces que se realiza el experimento, en la primera columna (valores desde 2 a 10) y el número de éxitos a su lado.

Por ejemplo en el caso anterior, Bin (6;0'5) , p(X=2), la columna p=0'5 es la última, y cuando n=6 y k=2 encontramos 0'2344, el valor que habíamos calculado.

Nota importante: El caso en que $p > 0'5$, no se encuentra tabulado.

La razón es bien sencilla. Si $p > 0'5$, entonces $q < 0'5$ y basta intercambiar los papeles de éxito y fracaso para que podamos utilizar la tabla.

Ejemplo:

La probabilidad de que un alumno de 2º de Bachillerato apruebe las Matemáticas es de 0'7. Si consideramos un grupo de 8 alumnos, ¿cuál es la probabilidad de que cinco de ellos aprueben las Matemáticas?.

Si éxito = "aprobar" y fracaso = "suspender", entonces p = 0'7 y q = 0'3.

Tenemos, por tanto, una Bin(8;0'7).

Nos piden calcular p(X=5), que no se puede calcular mediante las tablas porque p = 0'7 y sólo tenemos hasta p = 0'5. Por tanto si intercambiamos éxito = "suspender" y fracaso ="aprobar" entonces p = 0'3, q = 0'7, es decir la nueva binomial es Bin(8;0'3) y nos piden que aprueben 5 de 8, es decir que suspendan 3 de 8 o lo que es lo mismo, que tengamos 3 éxitos, p(X=3), y buscando en la tabla es p(X=3) = 0'2541.

También, desde luego podríamos haber utilizado la fórmula desde el principio, utilizar la Bin(8;0'7) y olvidarnos de tablas para hacer:

$$p(X = 5) = \binom{8}{5} \cdot (0'7)^5 \cdot (0'3)^3 = 0'254$$

3.2.2. Probabilidades acumuladas

Es posible que nos pidan no sólo la probabilidad de que ocurran un cierto número de éxitos en concreto, sino que ocurran como mucho "k" éxitos o preguntas similares. En el ejemplo anterior, por ejemplo, podrían pedirnos:

a) ¿Cuál es la probabilidad de que aprueben como mucho 2 alumnos?.

Si éxito = aprobar y fracaso = suspender, p= 0'7 y q = 0'3, entonces nos piden $p(X \leq 2)$. En este caso, basta pensar en que para que aprueben 2 alumnos como mucho, puede que aprueben 2, 1 o ninguno, es decir:

$$p(X \leq 2) = p(X = 0) + p(X = 1) + p(X = 2) = 0'0001 + 0'0012 + 0'01 = 0'1013$$

(haz las cuentas)

b) ¿Cuál es la probabilidad de que aprueben entre 3 y 6 alumnos (inclusive)?.

Del mismo modo:

$$p(3 \leq X \leq 6) = p(X = 3) + p(X = 4) + p(X = 5) + p(X = 6) =$$
$$= 0'0467 + 0'1361 + 0'2541 + 0'2965 = 0'7334$$

Hemos de tener en cuenta que para la distribución binomial, en las tablas sólo se admiten valores hasta n=10 (10 repeticiones del experimento). Para valores de $n > 10$, inevitablemente hemos de utilizar la fórmula.

Ejemplo:

Los alumnos de cierta clase se encuentran en una proporción del 67 % que estudian inglés y el resto francés.

Tomamos una muestra de 15 alumnos de la clase, calcular:

a) Probabilidad de que al menos encontremos tres alumnos de inglés.

b) Probabilidad de que los 15 alumnos estudien inglés.

c) Probabilidad de que estudien inglés entre 7 y 10 alumnos.

Si éxito = estudiar inglés, p = 0'67 y fracaso = estudiar francés, q = 1-0'67 = 0'33. Manejamos por tanto una Bin(15;0'67)

a) $p(X \geq 3) = p(X = 3) + p(X = 4) + p(X = 5) + p(X = 6) + \ldots + p(X = 15)$.

Una opción es calcular estas 13 probabilidades y sumarlas. Como hay que aplicar la fórmula para calcular cada una, la tarea se puede hacer bastante larga. Otra opción, más sencilla, es pasar al complementario. El complementario de encontrar al menos 3 alumnos de inglés es encontrar como mucho 2 alumnos de inglés, $p(X \leq 2)$.

Es decir,

$$p(X \geq 3) = 1 - p(X < 3) = 1 - p(X \leq 2) = 1 - (p(X = 0) + p(X = 1) + p(X = 2))$$

y sólo tenemos que calcular 3 probabilidades: $p(X = 0) \approx 0$, p(X=1) = 0'000001, p(X=2) = 0'000026 (¡compruébalo!).

Por lo cual,

$$p(X \geq 3) = 1 - (0 + 0'000001 + 0'000026) = 1 - 0'000027 = 0'999973$$

b) p(X=15) = 0'0025 (aplica la fórmula).

c)

$$p(7 \leq X \leq 10) = p(X = 7) + p(X = 8) + p(X = 9) + p(X = 10) =$$
$$= 0'0549 + 0'1114 + 0'1759 + 0'2142 = 0'5564.$$

3.2.3. Media y desviación típica en una distribución binomial

Aunque no se demostará, en una distribución binomial Bin(n;p), el número esperado de éxitos o media, viene dado por $\bar{x} = n \cdot p$. (Recordemos que la media es una medidad de centralización).

La desviación típica, σ , que es una medida de dispersión y mide lo alejados que están los datos de la media, viene dada por $\sigma = \sqrt{n \cdot p \cdot q}$.

3.3. La distribución Normal

Al estudiar aspectos tan cotidianos como:

- Caracteres morfológicos de individuos (personas, animales, plantas) de una misma raza. como tallas, pesos, envergaduras, etc.

- Caracteres fisiológicos, como el efecto de una misma dosis de un fármaco, o de una misma cantidad de abono.

- Caracteres sociológicos, como el consumo de ciertos productos por individuos de un mismo grupo humano.

- Caracteres psicológicos, como el cociente intelectual, grado de adaptación a un medio.

- Caracteres físicos, como la resistencia a la rotura de ciertas piezas. . .

todos ellos tienen en común que se distribuyen “normalmente”. ¿Qué quiere decir esta expresión?. Pués, por ejemplo, si hacemos una estadística para conocer la altura de 1400 mujeres y representamos los resultados en un diagrama de barras, obtenemos:

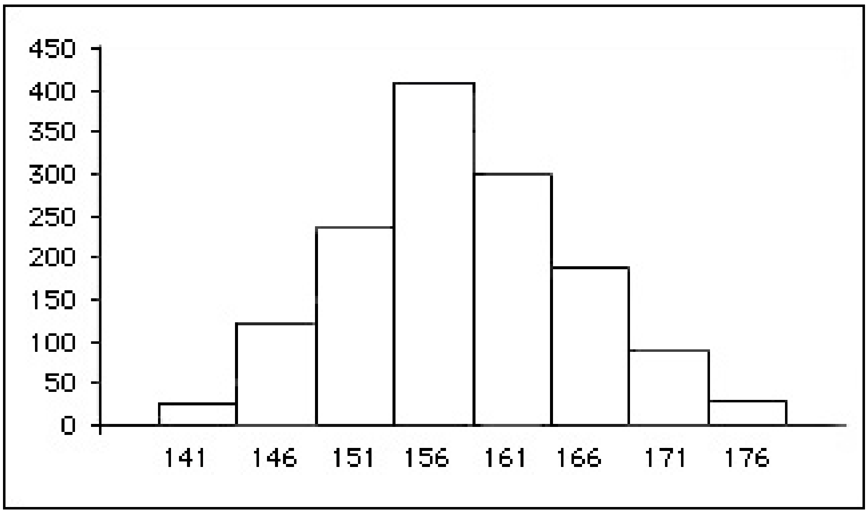

Figura 3.1: Distribución de estaturas de 1400 mujeres

Las gráficas de este tipo son muy corrientes: Hay pocos individuos en los extremos y un aumento paulatino hasta llegar a la parte central del recorrido, donde está la mayoría de ellos.

Definición: Diremos que una distribución de probabilidad sigue una *distribución normal* de media $\overline{x}$ y desviación típica σ, y lo representaremos por $N(\overline{x};\sigma)$ cuando la representación gráfica de su función de densidad es una curva positiva continua, simétrica respecto a la media, de máximo en la media, y que tiene 2 puntos de inflexión , situados a ambos lados de la media ($\overline{x}-\sigma$ y $\overline{x}+\sigma$ respectivamente) y a distancia de σ ella, es decir de la forma:

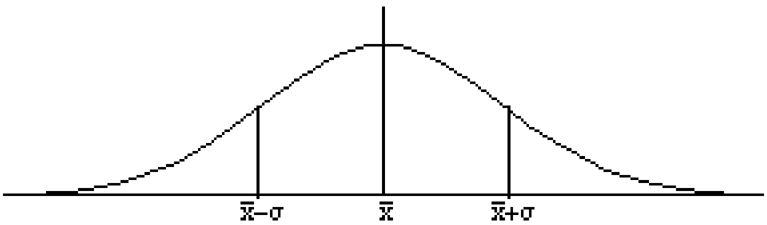

Figura 3.2: Distribución normal $N(\overline{x};\sigma)$. El máximo está en $(\overline{x}, \frac{1}{\sqrt{2\cdot\pi\cdot\sigma^2}})$

Dependiendo de los valores que tomen $\overline{x}$ y σ, la gráfica de esta función puede ser más o menos alargada, achatada, etc..., pero en cualquier caso siempre tiene las mismas condiciones de simetría, continuidad, etc reseñadas anteriormente.

El concepto de función de densidad introducido anteriormente no se estudiará con profundidad. Baste decir que la función de densidad determina la forma de cada distribución de probabilidad. En el caso de la distribución normal de parámetros $\overline{x}$ y σ, dicha función viene dada por:

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

Propiedad:

El área encerrada bajo la curva normal $N(\overline{x}; \sigma)$ siempre es 1.

La demostración de este resultado no es nada sencilla e implica el uso de resultados matemáticos que exceden el nivel de este curso.

De entre todas las curvas normales $N(\overline{x}; \sigma)$, la más sencilla, usada y conocida es aquella que tiene por media 0 y por desviación típica 1, N(0, 1).

Esta normal estándar se suele representar por Z.

La gráfica de esta curva se denomina campana de Gauss y se puede observar en la figura:

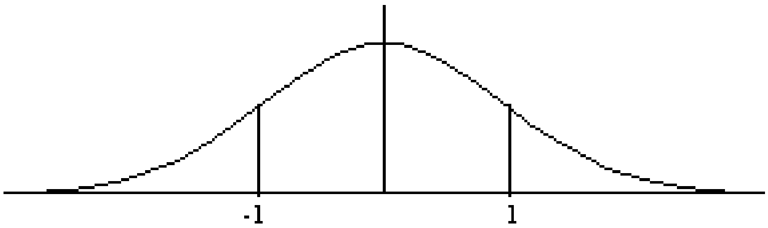

Figura 3.3: Distribución normal $N(0; 1)$. El máximo está en $(0, \frac{1}{\sqrt{2 \cdot \pi}})$

Su función de densidad será:

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

Puesto que el área bajo esta curva normal es 1, podemos definir una probabilidad de la siguiente manera:

Para un valor cualquiera k, definimos la probabilidad de que la distribución Z, N(0;1) , sea menor o igual que k como:

$p(Z \leq k)$= "Área encerrada bajo la curva normal N(0,1) desde $-\infty$ hasta k"

(es decir la parte rayada de la figura siguiente).

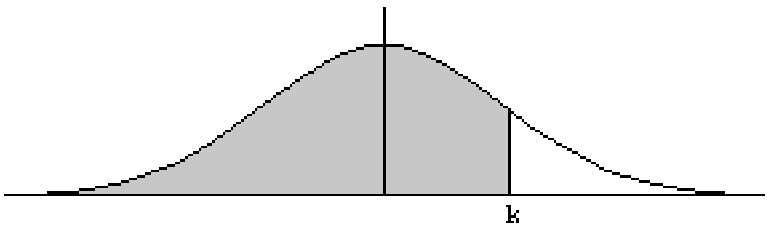

Figura 3.4: Área encerrada por la curva normal desde $-\infty$ hasta k

Ahora bien, ¿cómo calcular dicha área?. Fácil: Dichas áreas o probabilidades se encuentran tabuladas.

3.3.1. Uso de las tablas de la distribución normal N(0;1)

La normal 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 correpondientes a valores tales como 3'62 y mayores ya valen 0'9999 (prácticamente 1). Por eso, para estos valores mayores que 3'99, diremos que la probabilidad es aproximadamente 1. Así:

$$p(Z \leq 5'62) \approx 1$$

aunque no aparezca en la tabla.

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=3'2), basta decir que p(Z=3'2)=0.

Este tipo de distribuciones en las cuales la probabilidad de tomar un valor concreto es 0 se demoniman *distribuciones continuas*, para diferenciarlas de otras en las que esto no ocurre, como por ejemplo la binomial, que es una *distribución discreta*.

Así, al pasar al complementario, si tenemos $Z \geq k$, su complementario será $Z < k$, pero como incluir k no influye en la probabilidad,al calcular probabilidades podemos escribir:

$$p(Z \geq k) = 1 - p(Z < k) = 1 - p(Z \leq k)$$

Sólo se puede hacer esto en distribuciones continuas, en el caso de la binomial esto no se puede hacer y hay que ser cuidadosos con el paso al complementario.

Ejercicio: Buscar en la tabla de la normal estándar N(0;1) las probabilidades:
a) $p(Z \leq 1'15)$ b) $p(Z \leq 0'5)$ c) $p(Z \leq 0'82)$ d) $p(Z \leq 1'05)$ e) $p(Z \leq 4'27)$
f) $p(Z \leq 18'09)$

3.3.2. Cálculo de otras probabilidades

1. Si k es positivo y queremos calcular $p(Z \geq k)$, es decir el área rayada:

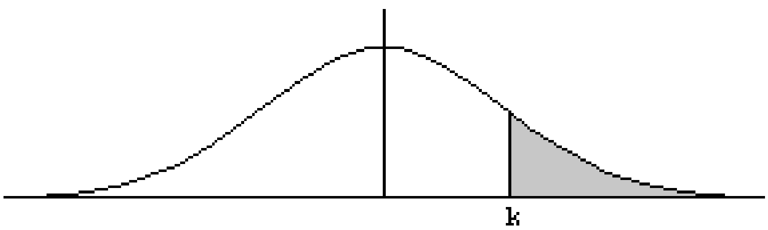

Figura 3.5: $p(Z \geq k)$. Basta pasar al complementario

basta pasar al complementario, es decir: $p(Z \geq k) = 1 - p(Z \leq k)$ y esta última probabilidad ya se encuentra tabulada.

Ejercicio: Calcular $p(Z \geq 0'3)$ y $p(Z \geq 2'07)$.

2. Si k es positivo y queremos calcular $p(Z \leq -k)$, es decir el área: por simetría, $p(Z \leq -k) =$

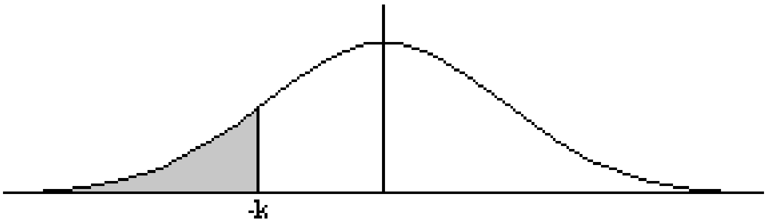

Figura 3.6: $p(Z \leq -k)$.Las probabilidades de valores negativos no están tabuladas

$p(Z \geq k)$ y ésta se calcula como en el caso anterior. Se puede observar la igualdad de áreas en la figura:

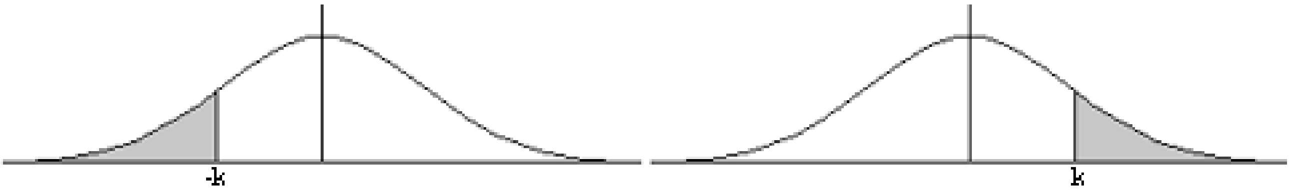

Figura 3.7: $p(Z \leq -k) = p(Z \geq k)$. La simetría permite reducir este caso al anterior

Ejercicio: Calcular $p(Z \leq -0'78)$ y $p(Z \leq -3'2)$.

3. Si k es positivo y queremos calcular $p(Z \geq -k)$, es decir el área rayada:

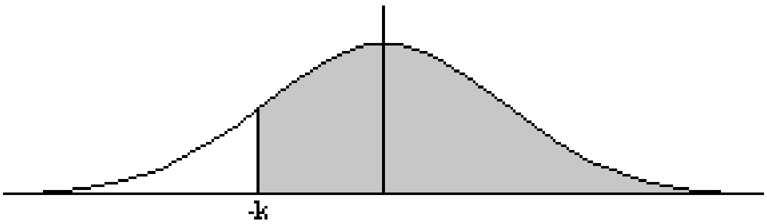

Figura 3.8: $p(Z \geq -k)$

entonces, por simetría $p(Z \geq -k) = p(Z \leq k)$:

Figura 3.9: $p(Z \geq -k) = p(Z \leq k)$.La simetría permite reducir este caso al que ya está tabulado

Ejercicio: Calcular $p(Z \geq -0'96)$ y $p(Z \geq -1'01)$.

4. Probabilidades comprendidas entre dos valores,$p(k_1 \leq Z \leq k_2)$,es decir el área rayada:

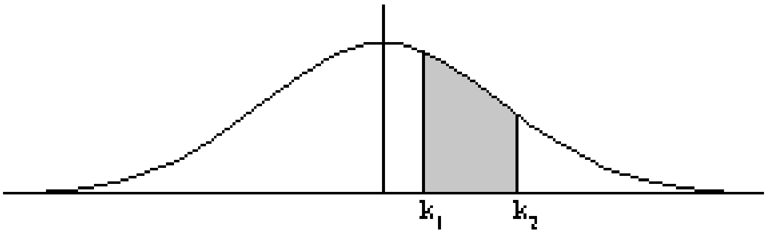

Figura 3.10: $p(k_1 \leq Z \leq k_2)$. Probabilidad comprendida entre dos valores

se calcula restando las áreas:

Figura 3.11: $p(Z \leq k_2)$ en la primera imagen.$p(Z \leq k_1)$ en la segunda. Al restar obtenemos el área pedida.

Se quita la parte correspondiente a $Z \leq k_1$,$p(Z \leq k_2) - p(Z \leq k_1)$.

Ejercicio: Calcular $p(-0'96 \leq Z \leq 1'49)$ y $p(-1'32 \leq Z \leq -0'57)$.

Ejercicio: Calcular p(Z=2), $p(Z \leq 2)$, $p(Z \geq 2)$, $p(Z \leq -2)$, $p(Z \geq -2)$, $p(-2 \leq Z \leq 2)$, $p(0'81 \leq Z \leq 1'33)$.

3.3.3. Cálculo de probabilidades en normales $N(\overline{x};\sigma)$

Si no tenemos una distribución N(0;1), sino una $N(\overline{x};\sigma)$ cualquiera, ¿cómo calcular probabilidades, si no tenemos tabla salvo para N(0;1)?. El siguiente resultado nos da la respuesta.

Propiedad:

Si X sigue una distribución $N(\overline{x};\sigma)$, entonces la variable $Z = \dfrac{X-\overline{x}}{\sigma}$ sigue una distribución N(0,1). (El paso de la variable X $\longrightarrow N(\overline{x};\sigma)$ a la Z $\longrightarrow$ N(0;1) se denomina *tipificación* de la variable X).

Ejemplo:

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

Sea X la distribución de los soldados , X es una N(168,8). Nos piden $p(166 \leq X \leq 170)$.

Utilizando el resultado anterior, primero restamos $\overline{x}$=168 en la desigualdad:

$$p(166 \leq X \leq 170) = p(166-168 \leq X-168 \leq 170-168) = p(-2 \leq X-168 \leq 2)$$

Y ahora dividimos entre $\sigma = 8$, con lo que acabamos de tipificar:

$$p(166 \leq X \leq 170) = p(-2 \leq X-168 \leq 2) = p\left(\frac{-2}{8} \leq \frac{X-168}{8} \leq \frac{2}{8}\right)$$

Llamando a $\dfrac{X-168}{8} = Z$, ésta ya es normal N(0,1) y se encuentra en las tablas:

$$p(166 \leq X \leq 170) = p(-0'25 \leq Z \leq 0'25) = p(Z \leq 0'25) - p(Z \leq -0'25) =$$
$$= (tablas) = 0'5987 - 0'4013 = 0'1974.$$

(pues $p(Z \leq -0'25) = p(Z \geq 0'25) = 1 - p(Z \leq 0'25) = 1 - 0'5987 = 0'4013$).

Ejercicios: 1) En una distribución N(22,5), calcula: $p(X \leq 27)$,$p(X \geq 27)$,$p(X \geq 125)$, $p(15 \leq X \leq 20)$, $p(17 \leq X \leq 30)$.

2) Los pesos de 60 soldados siguen una distribución N(67,5). Calcula la probabilidad de que el peso sea:

a) mayor de 80 kg.
b) 50 kg. o menos
c) menos de 60 kg.
d) 70 kg.
e) Entre 60 y 70 kg inclusive.

3.3.4. Otro uso de las tablas

Hasta ahora nos han dado la distribución normal N(0;1) y nos pedían $p(Z \leq k)$ siendo k un cierto número, y nos pedían calcular dicha probabilidad.

Ahora bien, otra pregunta puede ser: Dado que en una normal N(0;1) sabemos que $p(Z \leq k) = 0'9573$, ¿quién es k?.

La resolución es bien sencilla. Basta buscar 0'9573 *dentro de la tabla* de la distribución normal, y lo encontramos en el cruce de la fila 1'7 con la columna 2, y por lo tanto k debe ser 1'72.

Ejercicio: Calcular k si:
a) $p(Z \leq k) = 0'8078$.
b) $p(Z \geq k) = 0'0028$.

En caso de que el valor a buscar no aparezca directamente dentro de la tabla de la distribución normal, pueden ocurrir dos posibilidades:

a) Si el valor se encuentra entre dos valores de la tabla y a la misma distancia (aproximadamente) de cada uno de ellos, por ejemplo: $p(Z \leq k) = 0'7982$. En este caso el valor buscado será la media entre los valores extremos.

Si buscamos en la tabla este valor no aparece directamente, sino que se encuentra entre los valores 0'7967 (que corresponde a 0'83) y 0'7996 (que corresponde a 0'84). Por tanto el valor de k será:

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

b) Si el valor está entre dos valores, pero muy cercano a uno de ellos, directamente tomamos este valor, por ejemplo: $p(Z \leq k) = 0'7970$. El valor más cercano es 0'9767 (que corresponde a 0'83) y como el valor buscado está muy cerca de él, entonces directamente k=0'83.

Si la distribución no es normal N(0;1), sino $N(\overline{x};\sigma)$, tendremos que tipificar previamente.

Por ejemplo, si X sigue una normal N(6;3) y $p(X \leq k) = 0'9082$, calcula k.

Tipificando:

$$p\left(\frac{X-6}{3} \leq \frac{k-6}{3}\right) = 0'9082 \longrightarrow p\left(Z \leq \frac{k-6}{3}\right) = 0'9082$$

Y buscando en la tabla,

$$\frac{k-6}{3} = 1'33 \Rightarrow k - 6 = 3'99 \Rightarrow k = 9'99$$

Ejercicios:

1. Calcular k si $p(X \leq k) = 0'6141$ y X sigue una N(15,4).
2. De una variable normal $N(\overline{x}; \sigma)$ se sabe que $p(X \leq 7) = 0'9772$ y $p(X \leq 6'5) = 0'8413$. Calcular:
   a) $\overline{x}$ y σ.
   b) $p(5'65 \leq X \leq 6'25)$
   c) El número k tal que $p(X > k) = 0'3$

3.4. Relación entre la distribución binomial y la distribución normal

Es un hecho comprobado que cuando tenemos una distribución Bin(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 $Bin\left(100; \frac{1}{6}\right)$, y nos piden $p(20 \leq X \leq 33)$.

Es inviable aplicar las tablas (pues repetimos el experimento 100 veces) y tampoco la fórmula pues es inviable calcular, por ejemplo,

$$p(X = 32) = \binom{100}{32} \cdot \left(\frac{1}{6}\right)^{32} \cdot \left(\frac{5}{6}\right)^{68}$$

¿Cómo resolver el problema?. Del siguiente modo:

Teorema Central del Límite:

La distribución binomial Bin(n;p) se aproxima a una curva normal de media $\overline{x} = n \cdot p$ y desviación típica $\sigma = \sqrt{n \cdot p \cdot q}$, cuando n tiende a ∞, es decir, cuando n se hace muy grande.

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

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. Algunos 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 cuál, 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 correción por continuidad al aproximar una distribución binomial por una normal.

En el caso anterior,$\overline{x} = n \cdot p = \frac{100}{6} = 16'67$ y $\sigma = \sqrt{n \cdot p \cdot q} = \sqrt{\frac{500}{36}} = 3'73$. De modo que, como $n \geq 30$, $n \cdot p = 16'67 \geq 5$ y $n\dot{q} = 83'33 \geq 5$, se pude aproximar la binomial por la normal, es decir:

$$X \longrightarrow Bin\left(100; \frac{1}{6}\right) \approx Y \longrightarrow N(16'67; 3'73)$$

Entonces:

$$p(20 \leq X \leq 33) \underbrace{\approx}_{(*)} p(20 - 0'5 \leq Y \leq 33 + 0'5) = p\left(\frac{19'5 - 16'67}{3'73} \leq \frac{Y - 16'67}{3'73} \leq \frac{33'5 - 16'67}{3'73}\right) =$$

$$= p(0'89 \leq Z \leq 4'51) = p(Z \leq 4'51) - p(Z \leq 0'89) \approx 1 - 0'8133 = 0'1867$$

Notemos que en el paso señalado por (*) hemos cambiado X(binomial) por Y(normal) y se ha realizado la corrección por continuidad.