

Análisis de Datos 2016 - Primera parte

1 Probabilidades

1.1 Definición y propiedades básicas

Consideremos el experimento consistente en realizar un tiro de ruleta. El conjunto de todos los resultados posibles del mismo, que es Ω (“ómega”) $=\{0, 1, 2, \dots, 36\}$ es llamado “espacio de probabilidad” o “espacio muestral”.

Se llaman *eventos* a los subconjuntos de Ω . El conjunto $A = \{2, 4, 6, \dots, 36\}$ es “el evento que salga número par”; el $B = \{1, 2, \dots, 12\}$ es “el evento que salga primera docena”.

Las operaciones habituales con conjuntos tienen una traducción intuitiva en términos probabilísticos: la intersección $A \cap B$ es el evento “ A y B ocurren simultáneamente”; la unión $A \cup B$ es “ocurre al menos uno de los dos”; el complemento A' es el evento “no ocurre A ”; la diferencia $A - B = A \cap B'$ es “ocurre A pero no B ”. Si A y C son disjuntos —o sea, $A \cap C = \emptyset$ —, entonces “ A y C no pueden ocurrir simultáneamente”; si $A \subseteq C$, “siempre que ocurre A , ocurre C ”.

A cada evento A queremos asignarle un número $P(A)$, su “probabilidad”. Daremos una idea intuitiva de la misma. En el caso de la ruleta, si A es el evento de obtener número par, la manera (ideal) de determinar su probabilidad es realizar un gran número N de tiros, registrar la cantidad f_N que sale par, y calcular la proporción de veces que salió par: la “frecuencia relativa” $f_N(A)/N$. Entonces podemos pensar la $P(A)$ como el límite de estas frecuencias relativas cuando $N \rightarrow \infty$.

Se puede verificar fácilmente que la frecuencia relativa tiene las siguientes propiedades:

1. $0 \leq f_N(A)/N \leq 1$ para todo evento A
2. $f_N(\Omega)/N=1$ (donde Ω es el espacio muestral)

3. [*Ley aditiva*] Si los eventos A, B son disjuntos:

$$f_N(A \cup B)/N = f_N(A)/N + f_N(B)/N.$$

Entonces para que coincida con esta idea intuitiva, definimos probabilidad como una función P que a cada evento A le hace corresponder un número $P(A)$ que verifica esas mismas propiedades:

A1. $0 \leq P(A) \leq 1$ para todo evento A

A2. $P(\Omega)=1$ (donde Ω es el espacio muestral)

A3. [*Ley aditiva*] Si los eventos A, B son disjuntos:

$$P(A \cup B) = P(A) + P(B).$$

Algunas propiedades básicas

Dibujando los conjuntos, es fácil verificar que para el complemento A^c vale:

$$P(A^c) = 1 - P(A). \quad (1)$$

En particular, como $\Omega^c = \emptyset$, resulta

$$P(\emptyset) = 0.$$

Es fácil probar que si

$$A \subset B \text{ entonces } P(A) \leq P(B) \quad (2)$$

También a partir de las propiedades anteriores se puede probar a partir que si A y B son cualesquiera:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B). \quad (3)$$

La ley aditiva se puede extender a varios conjuntos. Si A, B, C son disjuntos (o sea, $A \cap B = A \cap C = B \cap C = \emptyset$), entonces se puede verificar que

$$P(A \cup B \cup C) = P(A) + P(B) + P(C). \quad (4)$$

1.2 Espacios equiprobables

En el caso de la ruleta, si suponemos que es equilibrada, podemos traducir esta creencia en la suposición de que los 37 resultados tienen la misma probabilidad (son *equiprobables*). En tal caso, todos tendrán probabilidad $1/37$ (por P2 y (4)). Pero esto no es obligatorio: dicha suposición es sólo un modelo de entre muchos posibles; y si la ruleta estuviera cargada, sería necesario asignar otras probabilidades.

En el caso de la ruleta equilibrada “el evento de que salga número par” $A = \{2, 4, 6, \dots, 36\}$ tiene probabilidad $P(A) = 18/37$; y “el evento de que salga primera docena” $B = \{1, 2, \dots, 12\}$ tiene probabilidad $P(B) = 12/37$. En ese caso $A \cap B = \{2, 4, 6, 8, 10, 12\}$ es “el evento sale un número par de la primera docena”, y tiene probabilidad $P(A \cap B) = 6/37$. El “evento sale un número par o de la primera docena” es $A \cup B$ y se puede calcular su probabilidad como: $P(A \cup B) = 18/37 + 12/37 - 6/37 = 24/37$.

En general en un espacio de probabilidades donde todos los resultados son *equiprobables*, la probabilidad de un evento se calcula como el número de resultados que forman ese evento dividido por el número de resultados de todo el espacio muestral.

También se aplica esta idea al calcular la probabilidad de que un individuo elegido en una población tenga determinada característica. Supongamos que se conoce que el 46% de los individuos de una población tienen sangre del grupo O, el 43% del grupo A, el 8% del grupo B y el 3% del grupo AB. Se elige una persona al azar en dicha población, la probabilidad de que tenga sangre grupo A es 0,43; la probabilidad de que tenga sangre grupo A o grupo B es 0,51 (dado que tener sangre grupo A o tener sangre grupo B son eventos incompatibles o disjuntos).

1.3 Eventos independientes

Consideremos ahora dos mesas de ruleta, atendidas por croupiers distintos, y sean respectivamente A y B los eventos de que salga número par en la primera mesa y “tercera docena” en la segunda. Intuitivamente, uno supondría que el resultado de una mesa no debiera influir en la otra. Esto se formaliza con la siguiente definición: los eventos A, B son *independientes* si

$$P(A \cap B) = P(A)P(B).$$

Pensando nuevamente en los grupos sanguíneos, si se eligen aleatoriamente tres personas (no familiares) en esa población, el grupo sanguíneo de cada una es independiente del de las otras. Entonces, la probabilidad de que las tres tengan grupo O es $0,46 \times 0,46 \times 0,46 = 0,097$.

Por supuesto si dos eventos no son independientes, la probabilidad de que ocurran simultáneamente **no** es el producto. Por ejemplo, si la probabilidad de que un hombre tenga talla superior a 1,80 es 0,2, la probabilidad de que un padre y un hijo tengan talla superior a 1,80 **no** es $0,2 \times 0,2$, ya que las tallas de los hijos tienden a estar relacionadas con las tallas de los padres.

2 Probabilidad condicional

Ahora estudiaremos el concepto de probabilidad condicional. Consideremos el siguiente ejemplo, se selecciona al azar un recién nacido y se realiza un análisis para diagnosticar hipotiroidismo congénito (HC), sea A el evento el recién nacido padece HC, la $P(A)$ es igual a la proporción de recién nacidos con HC en la población; ahora bien, si observamos que el recién nacido es una niña (sea B el evento el recién nacido es de sexo femenino), queremos conocer la probabilidad de que padezca HC sabiendo que es una niña, esto es la proporción de recién nacidos con HC en esa subpoblación, la de recién nacidos de sexo femenino. Utilizaremos la notación $P(A | B)$ para designar la *probabilidad condicional de A dado B*, (en el ejemplo la probabilidad de que el recién nacido padezca HC, sabiendo que es de sexo femenino) que se define cuando $P(B) > 0$ como:

$$P(A | B) = \frac{P(A \cap B)}{P(B)} \quad (5)$$

En este ejemplo $P(A | B) > P(A)$, pues es sabido que el HC es más frecuente en las niñas.

Es evidente, por la definición dada anteriormente, que si dos eventos A y B son independientes y con probabilidades no nulas, entonces $P(A | B) = P(A)$ y $P(B | A) = P(B)$

Ejemplo 2.1 *Supongamos que la prevalencia de una enfermedad E_1 es del 10% y la de otra enfermedad E_2 es del 15%, además sabemos que la probabilidad de que un individuo tenga ambas enfermedades es 0,05. Si se sabe*

que un individuo tiene la primera enfermedad, cuál es la probabilidad de que padezca la segunda?

Si llamamos A = tiene la enfermedad E_1 , y B = tiene la enfermedad E_2 , el enunciado dice que:

$P(A) = 0,10$, $P(B) = 0,15$ y $P(A \cap B) = 0,05$, entonces usando (5):

$$P(B | A) = \frac{0,05}{0,10} = 0,50$$

Esto significa que si sabemos que el individuo padece la primera enfermedad, esto aumenta la probabilidad de que padezca la segunda.

A partir de (5) se deduce el siguiente resultado, llamado *regla de la multiplicación*:

$$P(A \cap B) = P(A | B) \times P(B) \quad (6)$$

Esta regla es importante porque en muchas situaciones se desea conocer la $P(A \cap B)$, cuando $P(A | B)$ y $P(B)$ son conocidas.

Ejemplo 2.2 *Cuatro individuos han respondido a la solicitud de un banco de sangre. Supongamos que se necesita sangre tipo A+ y sólo uno de ellos tiene ese tipo, pero no se sabe cuál. Si los donantes potenciales se seleccionan al azar para determinar su tipo sanguíneo.Cuál es la probabilidad de que haya que determinar el tipo sanguíneo en al menos tres individuos para obtener el tipo deseado?*

Llamemos B = primer donante no es A+ y A = segundo donante no es A+, sabemos que $P(B) = \frac{3}{4}$ y $P(A | B) = \frac{2}{3}$. El evento $A \cap B$ es “ni el primero ni el segundo son tipo A+” = “se determina el tipo en al menos tres individuos”. Usando la regla de la multiplicación:

$$P(A \cap B) = P(A | B) \times P(B) = \frac{2}{3} \times \frac{3}{4} = \frac{1}{2}$$

3 Fórmula de la probabilidad total. Teorema de Bayes.

Ejemplo 3.1 *En cierta comunidad, el 8% de los adultos de más de 50 años de edad padece diabetes. Se conoce que la prueba para diagnosticar esa enfermedad tiene una sensibilidad del 95% (esto significa que la probabilidad de un resultado positivo dado que el individuo está enfermo es 0,95) y la especificidad es del 98% (la probabilidad de obtener un resultado negativo dado que el individuo es sano es 0,98). Si se realiza esa prueba en un individuo de más de 50 años elegido al azar en esa comunidad, cuál es la probabilidad de que el resultado sea positivo?*

Llamemos R^+ al evento el resultado de la prueba es positivo, R^- el resultado es negativo, D al evento el individuo tiene diabetes y ND el individuo no tiene diabetes.

Conocemos lo siguiente:

Prevalencia = $P(D) = 0,08$, y entonces $P(ND) = 0,92$,

Sensibilidad = $P(R^+ | D) = 0,95$, entonces $P(R^- | D) = 0,05$

Especificidad = $P(R^- | ND) = 0,98$, entonces $P(R^+ | ND) = 0,02$

y queremos calcular $P(R^+)$.

En este caso, D y ND son eventos disjuntos y también $D \cup ND = \Omega$, entonces podemos escribir:

$$R^+ = R^+ \cap (D \cup ND) = (R^+ \cap D) \cup (R^+ \cap ND)$$

y aplicando la ley aditiva y la ley de la multiplicación:

$$\begin{aligned} P(R^+) &= P(R^+ \cap D) + P(R^+ \cap ND) = \\ &= P(R^+ | D) \times P(D) + P(R^+ | ND) \times P(ND) \end{aligned}$$

y reemplazando por los valores, tenemos:

$$P(R^+) = 0,95 \times 0,08 + 0,02 \times 0,92 = 0,0944$$

En este ejemplo hemos utilizado la *fórmula de la probabilidad total* que se generaliza como sigue:

Si A_1, A_2, \dots, A_n son n eventos mutuamente incompatibles con $P(A_i) \neq 0$ y que cumplen: $\Omega = A_1 \cup A_2 \cup \dots \cup A_n$, entonces para cualquier evento B , se cumple:

$$P(B) = \sum_{i=1}^n P(B | A_i) \times P(A_i) \quad (7)$$

Para demostrar esta igualdad se siguen los mismos pasos que en el ejemplo:

Para cualquier $B \subset \Omega$, se cumple $B = B \cap \Omega$, por hipótesis $\Omega = \bigcup_{i=1}^n A_i$, entonces $B = B \cap \bigcup_{i=1}^n A_i = \bigcup_{i=1}^n (B \cap A_i)$, y como por hipótesis son eventos incompatibles, usando la ley aditiva y luego la ley de la multiplicación, se cumple:

$$P(B) = \sum_{i=1}^n P(B \cap A_i) = \sum_{i=1}^n P(B | A_i) \times P(A_i)$$

Volviendo al ejemplo anterior, supongamos que al individuo elegido al azar se le realizó la prueba diagnóstica, y esta dio un resultado positivo, ¿cuál es entonces la probabilidad de que dicho individuo tenga realmente diabetes?

Ahora lo que se desea es calcular $P(D | R^+)$, si aplicamos la definición de probabilidad condicional:

$$P(D | R^+) = \frac{P(D \cap R^+)}{P(R^+)}$$

calculamos $P(D \cap R^+)$ por la regla de la multiplicación y reemplazamos $P(R^+)$, tenemos:

$$P(D | R^+) = \frac{P(R^+ | D) \times P(D)}{P(R^+ | D) \times P(D) + P(R^+ | ND) \times P(ND)}$$

Esto se suele llamar valor predictivo positivo (*VPP*) de una prueba diagnóstica, es la *probabilidad de que el individuo este enfermo dado que la prueba dio un resultado positivo*.

En nuestro caso:

$$P(D | R^+) = \frac{0,95 \times 0,08}{0,0944} = 0,8051$$

De la misma manera se define el valor predictivo negativo (*VPN*) de una prueba diagnóstica, que es la *probabilidad de que el individuo esté sano dado que el resultado de la prueba fue negativo*:

$$P(ND | R^-) = \frac{P(R^- | ND) \times P(ND)}{P(R^- | D) \times P(D) + P(R^- | ND) \times P(ND)}$$

En estos ejemplos hemos usado el *teorema de Bayes*, que dice lo siguiente:

Si A_1, A_2, \dots, A_n son n eventos mutuamente incompatibles con $P(A_i) \neq 0$ y que cumplen: $\Omega = A_1 \cup A_2 \cup \dots \cup A_n$, entonces para cualquier evento B tal que $P(B) \neq 0$, se cumple:

$$P(A_k | B) = \frac{P(B | A_k) \times P(A_k)}{\sum_{i=1}^n P(B | A_i) \times P(A_i)} \quad (8)$$

La demostración es evidente a partir de la definición de probabilidad condicional, la regla de la multiplicación y la fórmula de la probabilidad total.

4 Variables aleatorias discretas

Al realizar un experimento aleatorio, muchas veces no estamos interesados en el resultado, sino en una función del mismo. Por ejemplo si tiramos dos veces un dado, podemos estar interesados en saber cuál es la suma, cuántas veces salió un valor en particular, cuál es el máximo de los dos valores observados, etc.

Una *variable aleatoria* es una función que a cada elemento del espacio muestral Ω , le hace corresponder un número.

Ejemplos:

1) Se tira un dado dos veces y se observa $X =$ "el número de veces que sale as".

2) Se tira un dado dos veces y se observa $Y =$ "el máximo de los dos valores".

3) Se tira una moneda hasta que sale cara y se define $Z =$ "el número de tiradas necesarias"

4) Se enciende una lámpara y se observa $T =$ "el tiempo hasta que se quema"

Todas estas son variables aleatorias, si consideramos el conjunto de valores que puede tomar cada una de estas variables vemos que:

$$v_X = \{0, 1, 2\}$$

$$v_Y = \{1, 2, 3, 4, 5, 6\}$$

$$v_Z = \{1, 2, 3, \dots\}$$

$$v_T = (0, \infty)$$

En los dos primeros ejemplos los conjuntos de valores: v_X y v_Y son finitos, en el tercero v_Z es infinito numerable y en el cuarto v_T es no numerable.

Cuando el conjunto de valores que toma una variable aleatoria es *finito* o *numerable*, la variable se denomina *discreta*. Las variables X , Y y Z definidas anteriormente son *v. a. discretas*. Una variable aleatoria *continua* puede tomar cualquier valor en un intervalo de números reales, la T es una *v.a. continua*.

Por el momento nos dedicaremos a estudiar *v. a. discretas*.

Utilizaremos la siguiente notación: $(X = a)$ es un evento de Ω , que está formado por todos los resultados para los que la variable X toma el valor a , y $(X \leq a)$ es el evento de Ω formado por todos los resultados para los que la

v.a. X toma valores menores o iguales que a . Esto se puede escribir:

$$\begin{aligned}(X = a) &= \{\omega \in \Omega; X(\omega) = a\} \\ (X \leq a) &= \{\omega \in \Omega; X(\omega) \leq a\}\end{aligned}$$

Para el primer y segundo ejemplo el espacio muestral es el mismo: $\Omega = \{(1, 1); (1, 2); \dots; (1, 6); (2, 1); (2, 2); \dots; (2, 6); \dots; (6, 1); (6, 2); \dots; (6, 6)\}$

En el primer ejemplo, podemos definir los eventos:

$$\begin{aligned}(X = 0) &= \left\{ \begin{array}{l} (2, 2); (2, 3); \dots; (2, 6); \\ (3, 2); (3, 3); \dots; (3, 6); \\ \dots \\ (6, 2); (6, 3); \dots; (6, 6) \end{array} \right\} \\ (X = 1) &= \left\{ \begin{array}{l} (1, 2); (1, 3); (1, 4); (1, 5); (1, 6); \\ (2, 1); (3, 1); (4, 1); (5, 1); (6, 1) \end{array} \right\} \\ (X = 2) &= \{(1, 1)\}\end{aligned}$$

Si el espacio es equiprobable, es fácil ver que:

$$\begin{aligned}P(X = 0) &= 25/36 \\ P(X = 1) &= 10/36 \\ P(X = 2) &= 1/36\end{aligned}$$

En el segundo ejemplo:

$$\begin{aligned}(Y = 1) &= \{(1, 1)\} \\ (Y = 2) &= \{(1, 2); (2, 1); (2, 2)\} \\ (Y = 3) &= \{(1, 3); (3, 1); (2, 3); (3, 2); (3, 3)\} \\ (Y = 4) &= \{(1, 4); (4, 1); (2, 4); (4, 2); (3, 4); (4, 3); (4, 4)\} \\ (Y = 5) &= \{(1, 5); (5, 1); (2, 5); (5, 2); (3, 5); (5, 3); (4, 5); (5, 4); (5, 5)\} \\ (Y = 6) &= \{(1, 6); (6, 1); (2, 6); (6, 1); (3, 6); (6, 3); (4, 6); (6, 4); (5, 6); (6, 5); (6, 6)\}\end{aligned}$$

Del mismo modo se pueden calcular las probabilidades $P(Y = y)$, para valores de $y = 1, 2, 3, 4, 5, 6$

$$P(Y = 1) = 1/36$$

$$P(Y = 2) = 3/36$$

$$P(Y = 3) = 5/36$$

$$P(Y = 4) = 7/36$$

$$P(Y = 5) = 9/36$$

$$P(Y = 6) = 11/36$$

Función de distribución. La *función de distribución* de una variable aleatoria X (*discreta* o *continua*) se define como

$$F(x) = P(X \leq x) \text{ para todo } x.$$

- es una función no decreciente, ya que si $a < b \implies (X \leq a) \subset (X \leq b)$ y en consecuencia $F(a) \leq F(b)$
- toma valores entre 0 y 1, porque es una probabilidad
- para todo $a < b$ se cumple

$$P(a < X \leq b) = F(b) - F(a) \tag{9}$$

Para verificar esta última propiedad, vemos que si $a < b$ podemos escribir $(X \leq b) = (X \leq a) \cup (a < X \leq b)$ y estos dos eventos son incompatibles, entonces por la ley aditiva de la definición de probabilidad:

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

y de allí se llega a (9)

Función de frecuencia. La *función de frecuencia* de una variable aleatoria discreta X se define para todos los $x \in v_X$ como:

$$f(x) = P(X = x).$$

Permite calcular probabilidades referidas a la X :

$$P(a \leq X \leq b) = \sum_{a \leq x \leq b} f(x) \quad \forall a, b,$$

donde la suma es sobre los valores x que toma X .

Se puede demostrar fácilmente que para cualquier v.a. discreta la función de frecuencia cumple

$$a) \quad f(x) \geq 0, \quad b) \quad \sum_x f(x) = 1, \quad (10)$$

Y se relaciona con la función de distribución a través de

$$F(x) = P(X \leq x) = \sum_{k \leq x} f(k).$$

Es evidente, entonces, que la función de distribución de una variable aleatoria discreta es *escalonada*, con saltos en los valores que toma la variable y constante en el resto.

La función de frecuencia de la v.a. X del ejemplo está dada por:

x	0	1	2
$f(x)$	25/36	10/36	1/36

Se puede ver que verifica (10). La función de distribución para esta variable es:

$$F(x) = \begin{array}{ll} 0 & \text{si } x < 0 \\ 25/36 & \text{si } 0 \leq x < 1 \\ 35/36 & \text{si } 1 \leq x < 2 \\ 1 & \text{si } x \geq 2 \end{array}$$

La función de frecuencia de la Y del ejemplo se muestra en la siguiente tabla:

y	1	2	3	4	5	6
$f(y)$	1/36	3/36	5/36	7/36	9/36	11/36

Que también verifica (10). La función de distribución para esta variable es:

$$F(x) = \begin{array}{ll} 0 & \text{si } x < 1 \\ 1/36 & \text{si } 1 \leq x < 2 \\ 4/36 & \text{si } 2 \leq x < 3 \\ 9/36 & \text{si } 3 \leq x < 4 \\ 16/36 & \text{si } 4 \leq x < 5 \\ 25/36 & \text{si } 5 \leq x < 6 \\ 1 & \text{si } x \geq 6 \end{array}$$

La probabilidad de cualquier evento que se relacione con el máximo de las dos tiradas puede calcularse usando la función de distribución. por ejemplo: sean A el evento “el máximo de las dos tiradas es a lo sumo 3”, B “el máximo de las dos tiradas es 4”, C “el máximo es mayor que 2 y menor que 5”

$$\begin{aligned} P(A) &= P(X \leq 3) = F(3) = 9/36 \\ P(B) &= P(X = 4) = P(X \leq 4) - P(X \leq 3) = F(4) - F(3) \\ P(C) &= P(2 < X < 5) = P(2 < X \leq 4) = F(4) - F(2) \end{aligned}$$

Variables aleatorias independientes. El ejemplo de las dos mesas de ruleta dado anteriormnte motiva la siguiente definición: Las variables X, Y (discretas o continuas) son *independientes* si para todo a, b , los eventos $(X \leq a)$ y $(Y \leq b)$ son independientes. En particular, para variables aleatorias discretas, se puede decir que X e Y son independientes si para todo a, b , los eventos $(X = a)$ y $(Y = b)$ son independientes.

Esta noción será útil para representar los resultados de experimentos que no se influyen mutuamente.

4.1 Distribución binomial

Supongamos que en un hospital hay 3 pacientes internados con determinada enfermedad, a los cuales se les aplica el mismo tratamiento (estos individuos no son parientes). Supongamos que la probabilidad de que un individuo sane en una semana de tratamiento es p . Sea Y el número de individuos que sanan en una semana de tratamiento. Y puede tomar los valores 0 (ninguno sana en una semana), 1, 2 o 3 (todos sanan en una semana). Los posibles resultados y sus respectivas probabilidades se resumen en la siguiente tabla.

Pac 1	Pac 2	Pac 3	Valor de Y	Probabilidad de este resultado
N	N	N	0	$(1-p)(1-p)(1-p) = (1-p)^3$
S	N	N	1	$p(1-p)(1-p) = p(1-p)^2$
N	S	N	1	$(1-p)p(1-p) = p(1-p)^2$
N	N	S	1	$(1-p)(1-p)p = p(1-p)^2$
S	S	N	2	$pp(1-p) = p^2(1-p)$
S	N	S	2	$p(1-p)p = p^2(1-p)$
N	S	S	2	$(1-p)pp = p^2(1-p)$
S	S	S	3	$ppp = p^3$

Si nos interesa unicamente saber cuantos pacientes sanan en la primera semana de tratamiento (el valor de Y), y las respectivas probabilidades, se puede resumir aun más.

Valor de Y	Probabilidad
0	$(1-p)^3$
1	$3p(1-p)^2$
2	$3p^2(1-p)$
3	p^3

esta es una manera de expresar la función de frecuencia de la variable aleatoria Y.

Este es un caso particular de la *distribución binomial*, en general si tenemos n repeticiones de un ensayo que puede tener solo dos resultados, que se suelen llamar “*éxito*” y “*fracaso*”, la probabilidad de “*éxito*” p , es la misma en cada repetición y los resultados de cada repetición son independientes, la variable aleatoria que cuenta el número de “*éxitos*” en los n ensayos tiene *distribución binomial* con parámetros n y p ($X \sim B(n, p)$) Se puede demostrar que su función de frecuencia es

$$f(k) = P(X = k) = \binom{n}{k} p^k (1-p)^{n-k} \quad (k = 0, 1, \dots, n).$$

En el ejemplo dado anteriormente, cada paciente se considera un ensayo: para cada paciente los posibles resultados son S (sana en una semana de tratamiento) y N (no sana en una semana de tratamiento); la probabilidad de sanar es p , la misma para cada individuo; y los resultados de los diferentes individuos son independientes. $X \sim B(3, p)$, donde p es la probabilidad de que un individuo sane en una semana de tratamiento.

La distribución binomial es también útil cuando se toma una muestra de una población finita, siempre que el tamaño de la muestra sea relativamente pequeño respecto del tamaño de la población. Supongamos por ejemplo una población de $N = 10000$ personas, de las cuales $M = 4000$ son fumadores. Se toma una muestra aleatoria de $n = 50$ personas; sea X la cantidad de fumadores en la muestra. Entonces la distribución de la variable aleatoria X es aproximadamente binomial con $n = 50$ y $p = M/N = 0.4$. Esto se puede aplicar en general cuando el tamaño N de la población es “grande”.

4.2 Distribución de Poisson

Una variable aleatoria que toma los valores $0, 1, 2, \dots$ se dice que tiene distribución de Poisson con parámetro $\lambda > 0$, ($X \sim P(\lambda)$) si su función de frecuencia está dada por:

$$f(k) = P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!} \quad (k = 0, 1, 2, \dots)$$

La distribución de Poisson sirve para modelizar el número Y de eventos “raros” que ocurren en el tiempo o en el espacio. Por ejemplo el número de partículas emitidas por una sustancia radiactiva en un intervalo de tiempo, el número de accidentes industriales ocurridos en un período, el número de bacterias en un volumen de agua, el número de árboles de determinada especie distribuidos aleatoriamente en un área, etc.

Algunos de esos ejemplos son procesos temporales, interesa conocer cuántas veces ocurre un evento en un intervalo de tiempo; y otros son procesos espaciales, interesa conocer cuántos “puntos” hay en un volumen o un área. Se denomina proceso temporal de Poisson cuando cumple las siguientes características:

Invariancia: Las condiciones no cambian en el tiempo.

Falta de memoria: Lo que sucede en el intervalo de tiempo $[0, t)$ no influye en lo que suceda en el intervalo $[t, t')$ para $t' > t$.

Sucesos aislados: La probabilidad de que en un intervalo de tiempo muy corto ocurra más de una vez el evento, es despreciable comparada con la probabilidad de que ocurra una o ninguna vez.

Para un proceso de este tipo, si X_t es la variable aleatoria que mide el número de veces que ocurre el evento en un intervalo de tiempo de longi-

tud t , puede verse que X_t es una variable aleatoria discreta cuya función de frecuencia está dada por:

$$f(k) = e^{-ct} \frac{(ct)^k}{k!}$$

entonces X_t tiene distribución de Poisson con parámetro $\lambda = ct$, donde c es una constante positiva que está relacionada con el tipo de proceso.

Las características de un proceso espacial de Poisson se pueden expresar como:

Homogeneidad espacial: La probabilidad de que un punto esté en una región dada, sólo depende del tamaño de esa región (área o volumen) y no de su forma o posición.

No interacción: lo que ocurre en una región es independiente de lo que ocurre en otra, si no se superponen.

La variable aleatoria X_a que mide el número de “puntos” en una región de área o volumen a , tiene distribución de Poisson con parámetro $\lambda = ca$

Ejemplo 4.1 *Suponga que estamos midiendo la emisión de partículas radiactivas, y que este es un proceso de Poisson con tasa 6 por minuto. ¿Cuál es la probabilidad de que en un período de 1/2 minuto ocurran al menos dos emisiones?*

Si X es la variable aleatoria que mide el número de emisiones en 1/2 minuto, esta tiene distribución de Poisson con $\lambda = 6 \times 0,5 = 3$.

$$\begin{aligned} P(X \geq 2) &= 1 - P(X < 2) = 1 - (P(X = 0) + P(X = 1)) = \\ &= 1 - e^{-3} \left(\frac{3^0}{0!} + \frac{3^1}{1!} \right) = 1 - 0,05(1 + 3) = 1 - 0,2 = 0,8 \end{aligned}$$

Ejemplo 4.2 *La distribución de plantas de cierta especie en una zona sigue un proceso de Poisson con una tasa de 6 plantas por metro cuadrado. ¿De qué medida debe ser tomado el radio r de una región circular de muestreo para que la probabilidad de hallar al menos una planta de esa especie en la región de muestreo sea $\geq 0,99$?*

Si la región de muestreo es circular de radio r , el área de esa región es $a = \pi \times r^2$, y la variable aleatoria X_a que mide el número de plantas en esa región tendrá distribución de Poisson con $\lambda = 6 \times \pi \times r^2$, entonces

$$P(X > 0) = 1 - P(X = 0) = 1 - \exp(-6\pi r^2) \geq 0,99$$

y despejando: $r^2 \geq -\ln 0,01/6\pi \implies r \geq 0,4943$

Aproximación de Poisson a la binomial En muchas aplicaciones tratamos con una variable aleatoria con distribución binomial donde n es grande y p es pequeña, en esos casos el cálculo de los números combinatorios es sumamente engorroso, y es útil tener alguna aproximación al valor de la probabilidad. Puede demostrarse que en ese caso:

$$f(k) = \binom{n}{k} p^k (1-p)^{n-k} \simeq e^{-\lambda} \frac{\lambda^k}{k!}, \text{ donde } \lambda = np$$

esta aproximación es aceptable si $p < 0.10$

Ejemplo 4.3 *Se sabe que el peso al nacimiento muy bajo (<1500gr), es una de las causas de morbilidad infantil. Se conoce que en determinada población el porcentaje de niños con muy bajo peso al nacimiento es de 1,2%. Si consideramos los 200 nacimientos en un hospital de esa población, ¿cuál es la probabilidad de que el número de recién nacidos con muy bajo peso en ese grupo sea mayor de 3?*

La variable aleatoria X que cuenta el número de niños con muy bajo peso entre los 200 nacimientos, tiene distribución binomial donde $n = 200$ y $p = 0,012$, y se puede aproximar con una Poisson con $\lambda = 2,4$

$$P(X > 3) = 1 - P(X \leq 3) \simeq 1 - e^{-2.4} \left(1 + 2,4 + \frac{2,4^2}{2!} + \frac{2,4^3}{3!} \right) = 1 - 0,7787 = 0,2213$$

4.3 Algunas medidas resumen de una variable aleatoria discreta

El *valor medio* o *esperado* de una variable aleatoria X se indica con $E(X)$ y es un “promedio ponderado” de los valores que toma, cada valor con un peso dado por su probabilidad. Más formalmente, para una variable discreta con función de frecuencia f , se define:

$$E(X) = \sum_{x \in v_X} x f(x),$$

donde x recorre todos los valores que toma X .

El significado intuitivo del valor medio es el siguiente: imaginemos que el experimento se repite un gran número N de veces, y se toma el promedio de los valores de X observados; entonces EX es el límite de esos promedios cuando $N \rightarrow \infty$.

Si X es una variable aleatoria y h es una función cualquiera, $h(X)$ es una variable aleatoria cuya media se calcula como:

para una variable aleatoria X discreta con función de frecuencia f

$$E(h(X)) = \sum_{x \in v_X} h(x)f(x) \quad (11)$$

Una consecuencia inmediata es que el valor medio tiene la propiedad de *linealidad*: si a, b son constantes

$$E(aX + b) = aE(X) + b. \quad (12)$$

La demostración de esta propiedad es simple usando (11) y considerando $h(x) = ax + b$:

$$E(aX + b) = \sum_{x \in v_X} (ax + b)f(x) = a \sum_{x \in v_X} xf(x) + b \sum_{x \in v_X} f(x) = aE(X) + b$$

La *varianza* de una variable aleatoria da un idea de la dispersión de la distribución de la variable alrededor de su *valor medio*. Para cualquier variable aleatoria X , se define la *varianza* de X como

$$\text{var}(X) = E(X - E(X))^2$$

Para el cálculo de la *varianza* se puede usar una fórmula abreviada

$$\text{var}(X) = E(X^2) - (EX)^2 \quad (13)$$

Demostración: llamamos $E(X) = \mu_X$ y usando (11), donde $h(x) = (x - \mu_X)^2$ tenemos:

$$\begin{aligned} \text{var}(X) &= E(X - \mu_X)^2 = \sum_{x \in v_X} (x - \mu_X)^2 f(x) = \sum_{x \in v_X} (x^2 - 2x\mu_X + \mu_X^2) f(x) = \\ &= \sum_{x \in v_X} x^2 f(x) - 2\mu_X \sum_{x \in v_X} x f(x) + \mu_X^2 \sum_{x \in v_X} f(x) = \\ &= E(X^2) - 2\mu_X^2 + \mu_X^2 = E(X^2) - \mu_X^2 \end{aligned}$$

A partir de lo anterior se ve facilmente que para todo a, b constantes:

$$\text{var}(aX + b) = a^2 \text{var}(X) \quad (14)$$

La *desviación típica* de X es también una medida de dispersión, definida como

$$\text{dt}(X) = \sqrt{\text{var}(X)}.$$

Cumple para todo a, b constantes:

$$\text{dt}(aX + b) = |a| \text{dt}(X). \quad (15)$$

Si X es constante: $X \equiv c$, entonces

$$\text{E}(X) = c \text{ y } \text{dt}(X) = 0.$$

Si X tiene media μ y varianza σ^2 , se deduce de (12) y (15) que la “variable normalizada”

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

tiene media 0 y varianza 1. En efecto

$$\begin{aligned} \text{E}(Z) &= \frac{1}{\sigma} \text{E}(X - \mu) = \frac{1}{\sigma} (\text{E}X - \text{E}\mu) = \frac{1}{\sigma} (\mu - \mu) = 0, \\ \text{dt}(Z) &= \frac{1}{\sigma} \text{dt}(X - \mu) = \frac{1}{\sigma} \text{dt}(X) = \frac{\sigma}{\sigma} = 1. \end{aligned}$$

Se puede probar que la media y varianza de una variable X con distribución binomial son

$$\text{E}(X) = np, \text{ var}(X) = np(1 - p). \quad (16)$$

y la media y varianza de una variable X con distribución de Poisson son

$$\text{E}(X) = \lambda, \text{ var}(X) = \lambda \quad (17)$$

El *cuantil- α* de la variable X con función de distribución F es el número $x(\alpha)$ tal que

$$F(x(\alpha)) = \alpha.$$

Los cuantiles para $\alpha = 0.50, 0.25$ y 0.75 se llaman *mediana* y *cuartiles* de X .

El cuantil de —digamos— $\alpha = 0.45$ se suele llamar también el *percentil* del 45%.

5 Variables aleatorias continuas

Una variable aleatoria *continua* puede tomar cualquier valor en un intervalo de números reales..

Función de densidad. La *función de densidad* de una variable aleatoria continua es una función f que cumple

$$a) f(x) \geq 0, \quad b) \int_{-\infty}^{\infty} f(x)dx = 1. \quad (18)$$

(comparar con (10)) y para todo $a < b$

$$F(b) - F(a) = P(a < X \leq b) = \int_a^b f(x)dx. \quad (19)$$

Entonces es evidente que para una variable aleatoria continua:

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

Es importante resaltar que si X es una v.a. *continua* entonces

$$P(X = a) = 0, \quad \forall a$$

La relación entre la función de *densidad* y la función de *distribución* está dada por:

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

de donde se deduce que la función de distribución de una variable aleatoria continua, es una función continua en todas partes; y que la función de densidad es la derivada de la función de distribución, en todos los puntos en los que esta última sea derivable:

$$f(x) = \frac{dF(x)}{dx}.$$

5.1 Esperanza de una variable aleatoria continua

La *esperanza* o *media* de una variable continua con función de densidad f , es

$$E(X) = \int_{-\infty}^{\infty} xf(x)dx.$$

cuando esta integral está definida.

Si X es una variable aleatoria continua con densidad f y h es una función cualquiera, $h(X)$ es una variable aleatoria cuya *esperanza* se calcula como:

$$E(h(X)) = \int_{-\infty}^{\infty} h(x)f(x)dx$$

cuando esta integral está definida.

La propiedad de *linealidad* del valor esperado también vale para variables aleatorias continuas, así como la definición y propiedades de la *varianza*, y de la *desviación típica*.

Para variables aleatorias continuas se definen los *cuantiles* de la siguiente forma, para cualquier $0 < \alpha < 1$, el *cuantil- α* , es el valor $x(\alpha)$, tal que

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

En particular, el *cuantil-0.5* se llama *mediana* y es el valor $\tilde{\mu}$, tal que:

$$F(\tilde{\mu}) = \int_{-\infty}^{\tilde{\mu}} f(x)dx = 1/2$$

Ejemplo 5.1 Sea X una v. a. con densidad dada por:

$$f(x) = \begin{cases} 0,2 & \text{si } -1 \leq x \leq 0 \\ 0,2 + cx & \text{si } 0 < x \leq 1 \\ 0 & \text{en caso contrario} \end{cases}$$

Obtener $F(x)$

Graficar $f(x)$ y $F(x)$

Calcular $F(-1)$, $F(0)$ y $F(1)$

Calcular $P(0 \leq X \leq 0,5)$

Ejemplo 5.2 Sea X una v.a. con la siguiente función de distribución:

$$F(x) = \begin{array}{ll} 0 & \text{si } x < 0 \\ x/8 & \text{si } 0 \leq x < 2 \\ x^2/16 & \text{si } 2 \leq x \leq 4 \\ 1 & \text{si } x > 4 \end{array}$$

Obtener la función de densidad para X

Calcular $P(1 \leq X \leq 3)$

Calcular $P(X \geq 1, 5)$

5.2 Distribución uniforme

Veamos un ejemplo muy simple: supongamos que una persona toma un colectivo para ir al trabajo, que pasa exactamente cada 5 minutos. Si sale de su casa sin tener en cuenta la hora, el tiempo X , que tiene que esperar en la parada es una variable aleatoria que puede tomar cualquier valor en el intervalo $[0, 5]$, la *función de densidad* para esta variable es:

$$f(x) = \begin{array}{l} 1/5 \text{ si } x \in [0, 5] \\ 0 \text{ en otro caso} \end{array}$$

Es evidente que $f(x) \geq 0$, para todo x ; y que el área total bajo $f(x)$ es igual a 1.

La probabilidad de que tenga que esperar entre 1 y 3 minutos es:

$$P(1 \leq X \leq 3) = \int_1^3 \frac{1}{5} dx = \frac{2}{5}$$

Se dice que esta variable aleatoria tiene distribución uniforme en el intervalo $[0, 5]$

En general se dice que una variable aleatoria tiene *distribución uniforme en el intervalo $[a, b]$* ($X \sim U[a, b]$), si su función de densidad está dada por:

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{si } x \in [a, b] \\ 0 & \text{en otro caso} \end{cases}$$

esta función cumple: $f(x) \geq 0$ para todo x ; y $\int_{-\infty}^{\infty} f(x)dx = 1$.

Calculemos la media y varianza de una variable aleatoria con distribución uniforme en $[a, b]$.

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx = \int_a^b \frac{x}{b-a} dx = \frac{1}{b-a} \frac{b^2 - a^2}{2} = \frac{b+a}{2}$$

entonces la media de una distribución uniforme es el punto medio del intervalo. Para calcular la varianza, calculamos primero

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f(x) dx = \int_a^b \frac{x^2}{b-a} dx = \frac{1}{b-a} \frac{b^3 - a^3}{3} = \frac{a^2 + ab + b^2}{3}$$

entonces

$$\text{var}(X) = EX^2 - (EX)^2 = \frac{a^2 + ab + b^2}{3} - \frac{(b+a)^2}{4} = \frac{(b-a)^2}{12}$$

La *función de distribución* está dada por:

$$F(x) = \begin{cases} 0 & \text{si } x < a \\ \frac{x-a}{b-a} & \text{si } x \in [a, b] \\ 1 & \text{si } x > b \end{cases}$$

Para calcular la mediana $\tilde{\mu}$, podemos hacer:

$$F(\tilde{\mu}) = \frac{\tilde{\mu} - a}{b - a} = 1/2$$

y despejando:

$$\text{med}(X) = \tilde{\mu} = \frac{b+a}{2}$$

en este caso la mediana y la media coinciden. Esto ocurre siempre que la distribución es *simétrica*.

5.3 Distribución exponencial

Una va X se dice que tiene *distribución exponencial* con parámetro λ (donde $\lambda > 0$), y se escribe $X \sim \epsilon(\lambda)$, si su función de densidad está dada por:

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{si } x > 0 \\ 0 & \text{si } x \leq 0 \end{cases}$$

Se ve facilmente que su función de distribución está dada por:

$$F(t) = \int_{-\infty}^t f(x)dx = \begin{cases} 1 - e^{-\lambda t} & \text{si } t > 0 \\ 0 & \text{si } t \leq 0 \end{cases}$$

Relación con un proceso de Poisson: Pensemos en un proceso temporal de Poisson de tasa λ , y consideramos el tiempo T , transcurrido desde el inicio del proceso hasta que se presenta el primer "evento" del proceso de Poisson. Entonces, para cualquier $t > 0$, se ve que $(T > t)$ ocurre si y solo si en el intervalo de tiempo $[0, t]$ no ocurre ninguno de los "eventos" del proceso, si definimos X_t el número de "eventos" que ocurren en el intervalo $[0, t]$, sabemos que $X_t \sim P(\lambda t)$ y podemos decir:

$$P(T > t) = P(X_t = 0) = e^{-\lambda t}$$

Entonces podemos determinar la función de distribución de esta va T :

$$P(T \leq t) = 1 - P(T > t) = 1 - e^{-\lambda t} \quad \text{si } t > 0$$

entonces hemos demostrado que $T \sim \epsilon(\lambda)$.

También puede demostrarse que el tiempo transcurrido **entre dos "eventos" sucesivos** de un proceso de Poisson de tasa λ , tiene distribución exponencial con parámetro λ .

Ejemplo 5.3 *Supongase que se reciben llamadas en una línea telefónica de emergencias las 24 hs del día, según un proceso de Poisson con una tasa 0,5 llamadas por hora. ¿Cuál es la probabilidad de que transcurran más de dos horas entre dos llamadas sucesivas?*

La variable aleatoria T , el tiempo (medido en horas) entre dos llamados sucesivos, tiene distribución exponencial con $\lambda = 0,5$. Entonces:

$$P(T > 2) = 1 - P(T \leq 2) = 1 - (1 - e^{-0.5 \times 2}) = 0,3679$$

La media y varianza de una variable exponencial λ están dadas por:

$$ET = \frac{1}{\lambda}; \quad \text{var}(T) = \frac{1}{\lambda^2}$$

Para calcular la mediana $\tilde{\mu}$

$$F(\tilde{\mu}) = 1 - e^{-\lambda\tilde{\mu}} = 1/2$$

luego

$$\text{med}(T) = \tilde{\mu} = \frac{\ln 2}{\lambda}$$

en este caso puede verse que la mediana es menor que la media.

Propiedad de ausencia de memoria: Dada una va $T \sim \epsilon(\lambda)$, si $t > 0$, y $s > 0$ se cumple:

$$P(T > t + s | T > s) = P(T > t)$$

Demostración: aplicando definicion de probabilidad condicional

$$P(T > t + s | T > s) = \frac{P[(T > t + s) \cap (T > s)]}{P(T > s)} = (a)$$

teniendo en cuenta que $(T > t + s) \subset (T > s)$

$$(a) = \frac{P(T > t + s)}{P(T > s)} = \frac{e^{-\lambda(t+s)}}{e^{-\lambda s}} = \frac{e^{-\lambda t} \times e^{-\lambda s}}{e^{-\lambda s}} = e^{-\lambda t} = P(T > t)$$

5.4 La distribución normal

Una va tiene distribución *distribución normal típica* (o normal $N(0, 1)$), y se escribe: $Z \sim N(0, 1)$, si su densidad está dada por:

$$\varphi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$$

se puede ver que esta función es simétrica respecto de 0

Si Z tiene esta distribución, entonces se prueba que

$$E(Z) = 0, \text{ var}(Z) = 1$$

Llamamos $\Phi(z)$ a su función de distribución; entonces para todo $a < b$.

$$P(a \leq Z \leq b) = \Phi(b) - \Phi(a).$$

Por ser φ simétrica, la Φ cumple

$$\Phi(-z) = 1 - \Phi(z). \quad (20)$$

Las tablas de Φ suelen incluir sólo las $z \geq 0$, y los valores para $z < 0$ se obtienen usando (20).

Por ejemplo, se obtiene de la tabla que $\Phi(1,02) = 0,8461$. Para calcular $\Phi(-1,02)$ se hace $1 - 0,8461 = 0,1539$.

Se deduce de (20) que los cuantiles de la $N(0, 1)$ cumplen

$$z(1 - \alpha) = -z(\alpha), \quad (21)$$

y en particular la mediana

$$\text{med}(Z) = \tilde{\mu} = z(0,5) = 0.$$

La variable X tiene distribución normal con media μ y varianza σ^2 (que se indica $X \sim N(\mu, \sigma^2)$), si su función de densidad está dada por:

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

puede verse que la gráfica de esta función es simétrica respecto de μ .

También puede demostrarse que si $X \sim N(\mu, \sigma^2)$, entonces:

$$E(X) = \mu, \text{ var}(X) = \sigma^2$$

Importante: La familia de distribuciones normales tiene la siguiente propiedad: si $X \sim N(\mu, \sigma^2)$, entonces para cualquier $a \neq 0$ y cualquier b , se verifica que $Y = aX + b \sim N(a\mu + b, a^2\sigma^2)$

Entonces, variable normalizada $Z = (X - \mu)/\sigma$ tiene distribución $N(0, 1)$. O sea que

$$X \sim N(\mu, \sigma^2) \text{ si } X = \mu + \sigma Z \text{ donde } Z \sim N(0, 1). \quad (22)$$

Ejemplo 5.4 Sea $X \sim N(30, 4)$, se desea calcular $P(28 < X < 31)$.

$$\begin{aligned} P(28 < X < 31) &= P((28 - 30)/2 < (X - 30)/2 < (31 - 30)/2) = \\ &= P(-1 < Z < 1/2) = \Phi(0,5) - \Phi(-1) = \Phi(0,5) - (1 - \Phi(1)) = \\ &= 0,6915 - (1 - 0,8413) = 0,5328 \end{aligned}$$

En general, para calcular probabilidades correspondientes a una X con distribución $N(\mu, \sigma^2)$, se la lleva al caso $N(0, 1)$, trabajando con $Z = (X - \mu)/\sigma$. Su función de distribución es entonces

$$F(x) = P(X \leq x) = P\left(Z \leq \frac{x - \mu}{\sigma}\right) = \Phi\left(\frac{x - \mu}{\sigma}\right),$$

y por lo tanto, si $a < b$:

$$P(a \leq X \leq b) = F(b) - F(a) = \Phi\left(\frac{b - \mu}{\sigma}\right) - \Phi\left(\frac{a - \mu}{\sigma}\right).$$

De aquí se puede ver que para cualquier variable aleatoria X con distribución normal, la probabilidad de que X esté dentro de 1 desvío estándar de su media es:

$$\begin{aligned} P(\mu - \sigma < X < \mu + \sigma) &= \Phi(1) - \Phi(-1) = 0,6826 \\ P(\mu - 2\sigma < X < \mu + 2\sigma) &= \Phi(2) - \Phi(-2) = 0,9544 \\ \text{y } P(\mu - 3\sigma < X < \mu + 3\sigma) &= \Phi(3) - \Phi(-3) = 0,9974 \end{aligned}$$

Usando (22), se prueba que los cuantiles de una variable X con distribución $N(\mu, \sigma^2)$ cumplen

$$x(\alpha) = \mu + \sigma z(\alpha), \quad (23)$$

donde $z(\alpha)$ son los cuantiles de $N(0, 1)$.

En particular, la mediana de una $X \sim N(\mu, \sigma^2)$ es

$$\text{med}(X) = x(0.5) = \mu$$

Ejemplo 5.5 Calcularemos los cuantiles de 0.80 y de 0.20 de una variable normal X con media 5 y desviación 2. De la tabla: $\Phi(0,84) = 0,7995$ y $\Phi(0,85) = 0,8023$. Interpolando resulta aproximadamente $z(0,8) = 0,843$, y por lo tanto $x(0,8) = 5 + 2 \times 0,843 = 6,686$. Para el otro cuantil, usamos $z(0,2) = -z(0,8)$ y por lo tanto $x(0,2) = 5 - 2 \times 0,843 = 3,314$.

6 "Propagación de errores": Funciones de una variable

Cuando se realiza cualquier medición, siempre se cometen de errores. Los errores pueden ser sistemáticos, como en el caso de una balanza mal calibrada que siempre da un valor superior al verdadero, o pueden ser aleatorios. En muchas situaciones los errores sistemáticos pueden y deben ser eliminados, por ejemplo calibrando correctamente la balanza, en otras cuando se conoce la magnitud del error, puede corregirse el resultado de la medición sumando (o restando según sea el caso) el valor del error sistemático. Por otra parte los **errores aleatorios** nunca pueden ser completamente eliminados.

Los errores que podemos tratar en estadística son los **errores aleatorios**. Debido a los errores aleatorios cuando repetimos una medición, en idénticas condiciones, los resultados de esas mediciones no son constantes, y fluctúan alrededor de un valor medio.

Entonces, consideremos el resultado de una medición, como una variable aleatoria $X = a + \varepsilon$, donde a es el verdadero valor de la magnitud que se está midiendo y ε es el error aleatorio, que en general, podemos suponer que tiene distribución normal con media 0 (esto indica que es igualmente probable que el error sea positivo o negativo) y varianza σ^2 , el valor de la varianza depende de la precisión del método de medición. Resumiendo, consideramos el resultado de una medición como una variable aleatoria:

$$X = a + \varepsilon, \text{ donde } a \text{ es el verdadero valor y } \varepsilon \sim N(0, \sigma^2) \quad (24)$$

o lo que es equivalente:

$$X \sim N(a, \sigma^2), \text{ donde } a \text{ es el verdadero valor} \quad (25)$$

En muchos casos interesa conocer el valor de una medición indirecta, es decir se está midiendo x , pero se desea conocer $f(x)$, y también interesa conocer cual es el error en la medición indirecta $f(x)$, esto es lo que denominamos "propagación de errores". Por ejemplo, sabemos que la absorbancia a de una solución es el negativo del logaritmo (decimal) de su transmitancia t :

$$a = -\log t.$$

Se tiene una medición de t con error, representada por una variable aleatoria X con media t y desviación σ , entonces la absorvancia también es una variable aleatoria Y :

$$Y = -\log X. \quad (26)$$

Con estos elementos se desea calcular la desviación típica de Y , sabiendo que

$$t = 0.501, \sigma = 0.001.$$

Para resolver este problema, lo plantearemos de modo más general, sea una variable aleatoria Y que es función de otra variable X :

$$Y = h(X),$$

donde X tiene media μ y desviación σ , buscaremos una forma de aproximar su media y su varianza.

La aproximación de la serie de Taylor a $h(X)$ en un entorno de μ es:

$$h(X) \simeq h(\mu) + h'(\mu)(X - \mu) \quad (27)$$

El lado derecho de esta ecuación es una función lineal de X . Si la distribución de X está concentrada en un intervalo sobre el que la función h sea aproximadamente lineal, entonces (27) es una buena aproximación de Y , y puede usarse para aproximar los valores de $E(Y)$ y $dt(Y)$, utilizando (27) y (11):

$$E(h(X)) \simeq h(\mu) + h'(\mu)E(X - \mu) = h(\mu)$$

del mismo modo, usando (27) y (15):

$$dt(h(X)) \simeq |h'(\mu)| dt(X)$$

Por ejemplo, si $Y = X^2$, es $h(x) = x^2$, por lo tanto $h'(x) = 2x$, y en consecuencia

$$dt(Y) \approx 2|\mu|\sigma.$$

En el caso (26) es $\mu = t$, y $h(x) = -\log x$, por lo tanto

$$h'(x) = -\frac{\log e}{x} = -\frac{0,434}{x}$$

(con $e = 2.71828\dots$), lo que da

$$EY \approx -\log 0,501 = 0,30, \quad dt(Y) \approx \frac{0,434 \times 0,001}{0,501} = 0,0008.$$

7 Modelo de mediciones repetidas

Parámetros de una suma de variables aleatorias

Dadas dos variables aleatorias X e Y , la esperanza y varianza de la suma son:

$$E(aX + bY) = aE(X) + bE(Y) \quad (28)$$

$$\text{var}(aX + bY) = a^2\text{var}(X) + b^2\text{var}(Y) + 2abcov(X, Y)$$

donde $\text{cov}(X, Y)$ es un parámetro que depende de la distribución conjunta de las variables aleatorias. Cuando las variables aleatorias son **independientes**, $\text{cov}(X, Y) = 0$, entonces:

$$\begin{aligned} \text{var}(aX + bY) &= a^2\text{var}(X) + b^2\text{var}(Y) \\ \text{var}(X + Y) &= \text{var}(X - Y) = \text{var}(X) + \text{var}(Y) \end{aligned} \quad (29)$$

Modelo de mediciones repetidas

Cuando se realizan mediciones repetidas (con error aleatorio) de una magnitud, los resultados de cada medición pueden ser representados por las variables aleatorias X_1, X_2, \dots, X_n , que se suponen independientes, con la misma distribución, y por consiguiente con la misma media μ y la misma desviación típica σ . En ese caso la media μ es la verdadera magnitud que deseamos medir y la desviación σ , depende de la precisión del método de medición.

Cuando se eligen al azar n individuos de una población, y se mide determinada característica de esos individuos -por ejemplo hemoglobina-, los resultados de esas mediciones, también son variables aleatorias X_1, X_2, \dots, X_n , independientes, con la misma distribución. En este caso μ y σ son la media y la desviación típica de la distribución de esa característica en la población.

En ambos casos ese conjunto de variables aleatorias independientes y con la misma distribución se llama *muestra aleatoria* de esa distribución.

Dada una muestra aleatoria, se define la *media muestral* como:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i.$$

Entonces, la *media muestral* es una variable aleatoria, que cumple

$$E\bar{X} = \mu, \quad \text{var}(\bar{X}) = \frac{\sigma^2}{n}, \quad \text{dt}(\bar{X}) = \frac{\sigma}{\sqrt{n}}. \quad (30)$$

En efecto, aplicando las propiedades de la esperanza:

$$E\bar{X} = \frac{1}{n}E\left(\sum_{i=1}^n X_i\right) = \frac{1}{n}\sum_{i=1}^n EX_i = \frac{1}{n}\sum_{i=1}^n \mu = \mu$$

también aplicando las propiedades de la varianza:

$$\text{var}(\bar{X}) = \frac{1}{n^2}\text{var}\left(\sum_{i=1}^n X_i\right)$$

y por ser variables independientes

$$\text{var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \text{var}(X_i) = \sum_{i=1}^n \sigma^2 = n\sigma^2.$$

A partir de (30) también se deduce que promediar varias mediciones aumenta la precisión. Si se tienen —por ejemplo— $n = 10$ mediciones y se desea duplicar la precisión —o sea, reducir $\text{dt}(\bar{X})$ a la mitad— hace falta tomar $n = 40$ (y no $n = 20$).

Ejemplo 7.1 *En un experimento con $\sigma = 0,5$, se desea determinar cuántas mediciones hacen falta para que $\text{dt}(\bar{X}) \leq 0,2$. Partiendo de*

$$\text{dt}(\bar{X}) = \frac{0,5}{\sqrt{n}} \leq 0,2,$$

se deduce que $\sqrt{n} \geq 0,5/0,2 = 2,5$, lo que implica $n \geq 2,5^2 = 6,25$, y por lo tanto se necesitan por lo menos 7 mediciones.

7.1 Distribución de \bar{X}

Como ya se mencionó, \bar{X} es una variable aleatoria y su distribución depende de la distribución de las X_1, X_2, \dots, X_n . En muchos casos conocer la forma de esta distribución no es simple, lo único que sabemos es que la media de \bar{X} es la misma que la de las X_i y la desviación típica de \bar{X} es la misma de las X_i dividida por \sqrt{n} .

Hay una situación particular en que la distribución de \bar{X} es sencilla, cuando la muestra aleatoria X_1, X_2, \dots, X_n tiene distribución normal.

Se puede demostrar que, si X e Y son independientes y tienen distribución normal, entonces también $Z = aX + bY$ tiene distribución normal. Más precisamente, si las distribuciones de X e Y son respectivamente $N(\mu_1, \sigma_1^2)$ y $N(\mu_2, \sigma_2^2)$, la de Z es $N(a\mu_1 + b\mu_2, a^2\sigma_1^2 + b^2\sigma_2^2)$.

Por lo tanto, dada una muestra aleatoria X_1, X_2, \dots, X_n , donde cada X_i es $N(\mu, \sigma^2)$, entonces $\sum_{i=1}^n X_i$ es $N(n\mu, n\sigma^2)$ y \bar{X} es $N(\mu, \sigma^2/n)$.

Este resultado, sumamente importante, de la familia de distribuciones normales, no vale en general para cualquier distribución.

Ahora bien, si la distribución de las X_1, X_2, \dots, X_n no es normal, en general es difícil determinar la distribución de \bar{X} . Pero el siguiente resultado es de gran ayuda cuando podemos contar con una muestra suficientemente grande.

Teorema del límite central: Sean X_1, X_2, \dots, X_n variables aleatorias independientes y con la misma distribución, con media μ y desviación típica σ (es decir que X_1, X_2, \dots, X_n es una muestra aleatoria de esa distribución). Entonces si n es grande, \bar{X} tiene aproximadamente una distribución normal con media μ y desviación típica σ/\sqrt{n} , del mismo modo $\sum_{i=1}^n X_i$ también tiene distribución aproximadamente normal con media $n\mu$ y desviación típica $\sqrt{n}\sigma$.

Formalmente

$$P(\bar{X} \leq a) = P\left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq \frac{a - \mu}{\sigma/\sqrt{n}}\right) \cong \Phi\left(\frac{a - \mu}{\sigma/\sqrt{n}}\right)$$

y

$$P\left(\sum_{i=1}^n X_i \leq a\right) = P\left(\frac{\sum_{i=1}^n X_i - n\mu}{\sqrt{n}\sigma} \leq \frac{a - n\mu}{\sqrt{n}\sigma}\right) \cong \Phi\left(\frac{a - n\mu}{\sqrt{n}\sigma}\right)$$

Este teorema nos dice que *cualquiera* sea la distribución de las X_1, X_2, \dots, X_n , podemos calcular (al menos en forma aproximada) probabilidades de eventos relacionados con $\sum_{i=1}^n X_i$, o con \bar{X} , siempre que n sea suficientemente grande. Como regla práctica podemos decir que con $n > 30$, se consigue una aproximación aceptable.

Ejemplo 7.2 *Suponga que el consumo diario de calorías de una persona es una variable aleatoria con media $\mu = 3000$ y desviación típica $\sigma = 350$. ¿Cuál es la probabilidad de que el consumo promedio de calorías diarias en el próximo año esté entre 2950 y 3050?*

Los consumos diarios de cada uno de los días del próximo año, los podemos representar con las variables aleatorias X_1, X_2, \dots, X_{365} , que podemos suponer independientes y todas tienen la misma distribución desconocida con media 3000, y desviación típica 350. Nos interesa calcular la probabilidad de que el promedio de esas 365 v. a. esté entre 2950 y 3050, esto es:

$$\begin{aligned} P(2950 \leq \bar{X} \leq 3050) &= P\left(\frac{2950 - 3000}{350/\sqrt{365}} \leq \frac{\bar{X} - 3000}{350/\sqrt{365}} \leq \frac{3050 - 3000}{350/\sqrt{365}}\right) \simeq \\ &\simeq \Phi\left(\frac{3050 - 3000}{350/\sqrt{365}}\right) - \Phi\left(\frac{2950 - 3000}{350/\sqrt{365}}\right) = \Phi(2,73) - \Phi(-2,73) = \\ &= 2\Phi(2,73) - 1 = 2 \times 0,9968 - 1 = 0,9936 \end{aligned}$$

Aplicaciones particulares del Teorema del límite central

- La distribución binomial, cuando n grande, se puede aproximar con una normal con media np y varianza $np(1-p)$. Entonces si $X \sim B(n, p)$ se cumple: $P(X \leq x) \simeq \Phi((x - np)/\sqrt{np(1-p)})$

Esta aproximación es aceptable cuando $np(1-p) > 5$

- La distribución de Poisson, cuando λ es grande, se puede aproximar a una normal con media λ y varianza λ . Entonces si $X \sim P(\lambda)$ se cumple: $P(X \leq x) \simeq \Phi((x - \lambda)/\sqrt{\lambda})$

Esta aproximación es aceptable cuando $\lambda > 30$

8 Propagación de errores: Funciones de dos variables

Se tiene ahora una variable Z que es función de dos variables aleatorias *independientes* X, Y :

$$Z = h(X, Y),$$

y se desea calcular aproximadamente su media μ_Z y desviación σ_Z , conociendo las medias μ_X, μ_Y y las desviaciones σ_X, σ_Y de X e Y .

El caso más fácil es una combinación lineal: $h(x, y) = a + bx + cy$. Aquí se cumple por (28) y (29):

$$\mu_Z = a + b\mu_X + c\mu_Y, \text{ y } \sigma_Z = \sqrt{b^2\sigma_X^2 + c^2\sigma_Y^2}.$$

Por ejemplo, para

$$Z = 2 - 3X + 4Y,$$

con

$$\mu_X = 5, \mu_Y = 6, \sigma_X = 0.1, \sigma_Y = 0.2$$

es

$$\begin{aligned} \mu_Z &= 2 - 3 \times 5 + 4 \times 6 = 11, \\ \sigma_Z &= \sqrt{9 \times 0.01 + 16 \times 0.04} = 0.854. \end{aligned}$$

Veamos el caso para cualquier función $h(x, y)$, sean las derivadas parciales

$$h_x(x, y) = \frac{\partial h(x, y)}{\partial x}, \quad h_y(x, y) = \frac{\partial h(x, y)}{\partial y}.$$

Entonces, si σ_X y σ_Y son “pequeñas”, X, Y son *independientes* y $Z = h(X, Y)$, se puede demostrar que:

$$E(Z) = \mu_Z \approx h(\mu_X, \mu_Y), \quad dt(Z) = \sigma_Z \approx \sqrt{[h_x(\mu_X, \mu_Y)\sigma_X]^2 + [h_y(\mu_X, \mu_Y)\sigma_Y]^2}.$$

Por ejemplo, supongamos $Z = X^2Y^3$, con $EX = 4$, $EY = 2$, $dt(X) = 0.01$, $dt(Y) = 0.02$. Entonces

$$h(x, y) = x^2y^3, \quad h_x(x, y) = 2xy^3, \quad h_y(x, y) = 3x^2y^2,$$

de modo que

$$\mu_Z \approx 16 \times 8, \quad \sigma_Z \approx \sqrt{(64 \times 0.01)^2 + (192 \times 0.02)^2}.$$