
UNIVERSIDAD AUTONOMA DEL ESTADO DE MEXICO

Curso de probabilidad y estadística con una introducción a R

FACULTAD DE CIENCIAS

JOSÉ ALFREDO SEPÚLVEDA GARÍA
DR. JUAN CARLOS CORONA ORAN
26 DE MARZO DE 2021

Índice general

1. Probabilidad	5
1.1. Teoría de Conjuntos	6
1.1.1. El concepto de conjunto	6
1.1.2. Subconjuntos	6
1.1.3. Conjunto universal y conjunto vacío	7
1.1.4. Operaciones entre conjuntos	8
1.2. Elementos de Probabilidad	9
1.2.1. Espacio muestral de un experimento	9
1.2.2. Muestreo de urnas	10
1.2.3. Experimentos Aleatorios	12
1.2.4. Eventos	12
1.3. Propiedades de Probabilidad	13
1.3.1. Funciones de Probabilidad	13
1.3.2. Algunas Propiedades	13
1.3.3. Asignación de Probabilidades	14
1.4. Probabilidad Condicional	14
1.4.1. Propiedades y Reglas	16
1.4.2. La Regla de la Multiplicación	16
1.5. Teorema de Bayes	17
1.6. Técnicas de Conteo	19
1.6.1. Muestreo Ordenado	20
1.6.2. Muestras Desordenadas	20
2. Distribuciones discretas	23
2.1. Variables Aleatorias Discretas	23
2.1.1. Función de Masa de Probabilidad	23
2.1.2. Media, Varianza, y Desviación Estándar	24
2.2. La Distribución Uniforme Discreta	25
2.3. La Distribución Binomial	26
2.3.1. El Modelo Binomial	26
2.4. Funciones Genradoras de Momentos	28
2.4.1. El Operador de Expectación	28
2.4.2. Funciones Generadoras de Momentos	29
2.5. La Distribución Empírica	31
2.6. Otras Distribuciones Discretas	32
2.6.1. Ensayos de Bernoulli Dependientes	33
2.6.2. Distribución de la variable Tiempo de Espera	34
2.6.3. Procesos de Llegada	36
2.7. Funciones de Variables Aleatorias Discretas	37
3. Distribuciones Continuas	39
3.1. Variables Aleatorias Continuas	39
3.1.1. Funciones de Densidad de Probabilidad	39
3.1.2. Expectación de Variables Aleatorias Continuas	40
3.2. La Distribución Uniforme Continua	43
3.3. La Distribución Normal	44
3.3.1. Cuantiles Normales y la Función Cuantil	45
3.4. Funciones de Variables Aleatorias Continuas	46
3.4.1. El Método de Función de Distribución de Probabilidad	46

3.4.2.	El método de Función de Distribución Acumulada	47
3.5.	Otras Distribuciones Continuas	49
3.5.1.	Distribuciones de Espera de Tiempo	49
3.5.2.	Las Distribuciones Chi Cuadrada, t Student y F Snedecor	50
3.5.3.	Otras Distribuciones Populares	52
4.	Distribuciones de Muestras	55
4.1.	Muestras Aleatorias Simples	55
4.2.	Muestreo de una Distribución Normal	56
4.2.1.	La Distribución de la Media Muestral	56
4.2.2.	La distribución de la Varianza Muestral	56
4.2.3.	Estadística de la Distribución de t Student	57
4.3.	El Teorema del Límite Central	58
4.4.	Distribuciones Muestrales de Estadísticas de Dos Muestras	59
4.4.1.	Diferencia de Medias Muestrales Independientes	59
4.4.2.	Diferencia de Proporciones Muestrales Independientes	59
4.4.3.	Tasa de Varianzas Muestrales Independientes	60
4.5.	Distribuciones Muestrales Simuladas	60
4.5.1.	El Rango Intercuartil	60
4.5.2.	La Desviación Absoluta Mediana	61
5.	Pruebas de Hipótesis de Una Muestra	63
5.1.	Hypótesis, Pruebas y Errores	63
5.2.	Pruebas de Hipótesis y Distribuciones Muestrales	63
5.3.	Trabajando de nuevo con Z	64
5.4.	Prueba de hipótesis con t	66
5.5.	Prueba de una Varianza	67
6.	Pruebas de Hipótesis de Dos Muestras	69
6.1.	Hipótesis para 2	69
6.2.	Pruebas con Z	70
6.3.	Prueba t para dos Muestras	70
6.3.1.	Varianzas Iguales	71
6.3.2.	Varianzas desiguales	72

Capítulo 1

Probabilidad

Introducción

Una de las definiciones técnicas de probabilidad dicta que ésta es la posibilidad de que suceda un fenómeno o un hecho, bajo determinadas circunstancias, es decir, el nivel de certeza que se tiene sobre la circunstancia de cierto evento; el estudio de esta ciencia proporciona métodos para cuantificar las oportunidades y probabilidades asociadas con varios sucesos.

Aunque la definición anterior puede darle al término de probabilidad un aire muy teórico, esta palabra se puede utilizar en el lenguaje coloquial, conservando su definición. Por ejemplo, se habla de la probabilidad de lluvia en cierto día, o de qué tan probable es que un equipo de fútbol le gane a otro, o, en días de pandemia, de la probabilidad de una cuarta ola de COVID-19.

Actualmente, la probabilidad tiene una gran número de aplicaciones, tanto en las ciencias sociales, como en las ciencias naturales. Por ejemplo, en el área de las ciencias sociales es una herramienta fundamental en la planificación estratégica de los movimientos sociales, económicos y laborales de toda la comunidad. La importancia esencial de la aplicación de los métodos de cálculo de la probabilidad reside en su capacidad para estimar o predecir eventos.

En este capítulo se definirán algunos conceptos de la probabilidad, se mostrarán ejercicios resueltos, se enseñará como aplicar dichos conceptos en el lenguaje de programación R, y al final se propondrá una serie de ejercicios a resolver.

1.1. Teoría de Conjuntos

1.1.1. El concepto de conjunto

El concepto de **conjunto** es base para el estudio de la probabilidad (y de la estadística y matemáticas en general). Un conjunto es una colección de elementos. Generalmente, un conjunto se define mediante una letra mayúscula: A, B, C; y un elemento de algún conjunto cualquiera, generalmente por una letra minúscula: a, b, c.

Se usa el símbolo \in para denotar que un elemento a , pertenece a un conjunto A : $a \in A$. Por otra parte, si un elemento no se encuentra en un conjunto, entonces se emplea el símbolo \notin : $b \notin A$. Cuando dos o más elementos pertenecen a un conjunto, esto se expresa separando con una coma los elementos de dicho conjunto: $a, c \in A$.

Es posible definir a un conjunto listando todos sus elementos, pero si la cantidad de estos es muy grande, o si no se conocen todos ellos, se pueden determinar al conjunto mediante alguna propiedad que cumplan todos sus elementos. Al primer método se le conoce como método de expansión, mientras que al segundo como método de compresión.

Definición. Un conjunto es una colección de objetos considerada como un objeto en sí. Los objetos de la colección pueden ser cualquier cosas: personas, números, colores, letras, figuras, etc. Cada uno de los objetos en la colección es un elemento o miembro del conjunto.

Ejemplo 1.1. El conjunto A, que contenga a todas las vocales puede ser representado por $A = \{a, e, i, o, u\}$; pero también puede representarse mediante la propiedad que cumplen todos los elementos, es decir, $A = \{x | x \text{ es una vocal}\}$.

Ejemplo 1.2. Un ejemplo más común es: los números que pueden resultar en la cara superior de un dado, después de lanzarlo: $\{1, 2, 3, 4, 5, 6\}$.

Ejemplo 1.3. Si se considera un conjunto cuyos elementos son infinitos, entonces no es posible describir al conjunto por medio del método de expansión: el conjunto de todos los triángulos en un plano es $\{x | x \text{ es un triángulo en un plano}\}$.

En R

Existen varias maneras de crear conjuntos en R. La primera que se muestra consiste en asignar los elementos del conjunto a un objeto, por ejemplo "A". A representa el conjunto de las vocales del abecedario, como en el **Ejemplo 1.1**:

```
> A<-c('a','e','i','o','u')
> A

[1] "a" "e" "i" "o" "u"
```

Otra forma de crear conjuntos es la siguiente:

```
> U<-c(1:5)
> U

[1] 1 2 3 4 5
```

que corresponde al **Ejemplo 1.2**.

También se puede usar la función `seq()`. Por ejemplo, se busca crear el conjunto de los números pares entre 2 y 10:

```
> P<-seq(from=2, #Desde dónde inicia el conteo
+        to=10,  #Hasta dónde llega
+        by=2)   #Número de pasos entre cada elemento
> P

[1] 2 4 6 8 10
```

1.1.2. Subconjuntos

Si todos los elementos de un conjunto A se encuentran en un conjunto B, entonces se dice que A es un subconjunto de B. Se escribe como $A \subset B$ o $B \supset A$ y se lee "A está contenido en B" o "B contiene a A". Por lo tanto, para todo conjunto A, $A \subset A$.

Propiedades

Si $A \subset B$ y $B \subset A$ entonces A y B son iguales y se escribe $A = B$. En este caso A y B tienen los mismos elementos.

Si A no tiene los mismos elementos que B entonces $A \neq B$.

Si $A \subset B$ pero $A \neq B$ entonces se llama a A un subconjunto propio de B.

Conjunto potencia. La totalidad de los subconjuntos de un conjunto dado A constituye el llamado conjunto potencia o conjunto de partes de A : $P(A) = \{B | B \subseteq A\}$.

Cuando el conjunto A tiene un número finito de elementos, por ejemplo $|A| = n$, el conjunto potencia también es finito y tiene 2^n elementos.

Ejemplo 1.4. $\{a, b, c\}$ es un subconjunto propio del conjunto de las letras del abecedario.

Ejemplo 1.5. $\{i, a, e, u, o\}$ no es un subconjunto propio de $\{a, e, i, o, u\}$ pues ambos conjuntos son iguales, sin importar el orden de los elementos.

Ejemplo 1.6. Al lanzar un dado los resultados posibles cuando el resultado es par son los elementos del conjunto $\{2, 4, 6\}$, que es un subconjunto de los resultados posibles $\{1, 2, 3, 4, 5, 6\}$.

Ejemplo 1.7. Dado el conjunto $A = \{a, b\}$, su conjunto potencia es: $P(A) = \{\emptyset, \{a\}, \{b\}, \{a, b\}\}$, donde \emptyset representa al conjunto vacío.

Teorema. Si $A \subset B$ y $B \subset C$, entonces $A \subset C$

En R

Usualmente al trabajar con bases de datos muy grandes, se busca extraer ciertos subconjuntos de datos con el fin de analizarlos por separado. Es posible realizar esta tarea en R de distintas maneras. El método usado para la extracción de los datos dependerá de la estructura que la base de datos tenga.

La manera más simple es con el uso de corchetes:

```
> calificaciones<-c(9,7,8,10,7,8)
> calificaciones[c(3,5)]
```

```
[1] 8 7
```

Un ejercicio más complejo consiste en extraer subconjuntos de bases de datos más grandes. Para el siguiente ejemplo se importará una base de datos de la educación en los 50 estados de Estados Unidos:

```
> educacion <- read.csv("https://vincentarelbundock.github.io/Rdatasets/csv/robustbase/education.csv",
> colnames(educacion) <- c("X", "ESTADO", "REGION", "POBLACION.URBANA", "INGRESO.PER.CAPITA",
+ "POBLACION.MENOR", "GASTO.EN.EDUCACION")
> View(educacion)
```

Ahora, suponiendo que se busca analizar la división de escuelas de la región 2, además de calcular cuánto dinero se gastó por alumno por cada estado en esta región. Se necesitan 3 variables: ESTADO, POBLACIÓN.MENOR Y GASTOS.POR.EDUCACIÓN. Aquí se muestra una manera de obtener dichos datos:

```
> ed_exp1 <- educacion[c(10:21),
+ c(2,6:7)]
```

Sin embargo, estas maneras básicas de extraer subconjuntos de bases de datos en R pueden resultar tediosas cuando se trabaja con bases de datos más voluminosas. Si, por ejemplo, la base de datos constara de 70 columnas y 5000 filas, ¿cómo se podría encontrar el intervalo de columnas y filas que se necesita? A continuación se muestra otra manera de extraer un subconjunto, con la función "subset":

```
> ed_exp2<-subset (educacion,
+ REGION==2,
+ select=c("ESTADO", "POBLACION.MENOR", "GASTO.EN.EDUCACION"))
```

Aunque estas 3 maneras no son las únicas, sí son las más sencillas y por ahora son suficientes para explicar la extracción y creación de subconjuntos.

1.1.3. Conjunto universal y conjunto vacío

El universo de discurso, conjunto universal o referencial, que normalmente se denota por las letras U , V o E , es un conjunto cuyo objeto de estudio son los subconjuntos del mismo. Anteriormente se consideraba al conjunto universal como el conjunto de todas las cosas, sin embargo está demostrado que este conjunto no existe.

Actualmente se debe dejar en claro sobre cuál conjunto se está tratando. Por ejemplo, si estamos tratando conjuntos cuyos elementos son letras, el conjunto referencial sería el conjunto formado por todas las letras del alfabeto.

Por otro lado, el conjunto vacío es el que no posee elemento alguno. Puesto que lo único que define a un conjunto es la propiedad que satisfacen sus elementos, el conjunto vacío es único. Se representa con el símbolo \emptyset . Otra notación común para el conjunto vacío es la notación extensiva, especificando sus elementos (ninguno) entre llaves: $\{\}$.

Definición. Un conjunto universal es un conjunto formado por todos los objetos de estudio en un contexto dado.

Definición. El conjunto vacío \emptyset es el conjunto de todos los elementos x tal que $x \neq x$.

Ejemplo 1.8. Un conjunto importante es el conjunto R de los números reales como $3, \pi, -\sqrt{2}$, que pueden representarse por puntos en una línea real como el eje x . Si a y b son números reales tales que $a < b$, los subconjuntos $\{x|a \leq x \leq b\}$ y $\{x|a < x < b\}$ de R se denominan intervalos cerrado y abierto respectivamente. Los subconjuntos tales como $\{x|a \leq x < b\}$ o $\{x|a < x \leq b\}$ se denominan intervalos semi-abiertos o semi-cerrados.

Ejemplo 1.9. El conjunto de todos los números reales x tales que $x^2 = -1$, escrito $\{x|x^2 = -1\}$ es el conjunto nulo o vacío ya que no hay números reales cuyos cuadrados sean iguales a -1 . Sin embargo, si se considera a los números complejos el conjunto no es vacío.

Ejemplo 1.10. Al lanzar un dado, el conjunto de todos los resultados posibles es el universo $\{1, 2, 3, 4, 5, 6\}$. El conjunto de los resultados que consisten de las caras 7 u 11 sobre un solo dado es el conjunto nulo.

1.1.4. Operaciones entre conjuntos

Al considerar dos conjuntos A y B , son diversas las operaciones que se pueden definir sobre ellos dos. Sin embargo, todas se basan en las operaciones de unión, intersección y el complemento.

Definición.

1. El complemento de un conjunto, denotado por A^c son todos aquellos elementos que no están A .
2. La intersección de dos conjuntos A y B , donotada como $A \cap B$ es el conjunto de todos los elementos que se encuentran tanto en A como en B .
3. La unión de dos conjuntos A y B , denotada como $A \cup B$ es el conjunto que contiene a todos los elementos de A y B .

Ejemplo 1.11. Un porcentaje bastante grande de programas C++ escritos en una empresa particular, se compilan en la primera ejecución, pero algunos otros no. Supongamos que un experimentos consiste en seleccionar y compilar programas C++ uno por uno hasta encontrar uno que se compile en la primera ejecución. Se denota mediante E (éxito) un programa que se compila en la primera ejecución y uno que no lo hace se denota mediante F (falla). Ahora definiendo

$$A = \{S, FS, FFS\}, B = \{S, FFS, FFFFS\}, C = \{FS, FFFS, FFFFFS\}$$

entonces

$$A^c = \{FFFS, FS, FFS\}, C^c = \{S, FFS, FFFFS\}, A \cup B = \{S, FS, FFS, FFFFS\}, A \cap B = \{S, FFS\}$$

En R

R cuenta con las funciones *union* para unir conjuntos, *intersect* para hallar su intersección y *setdiff* para mostrar el complemento de un conjunto dado otro conjunto:

```
> A<-c(2,3,4,5)
> B<-c(4,7,9,12)
> AuB<-union(A,B)
> AuB

[1] 2 3 4 5 7 9 12

> AiB<-intersect(A,B)
> AiB

[1] 4

> Acomp<-setdiff(B,A)
> Acomp

[1] 7 9 12

> Bcomp<-setdiff(A,B)
> Bcomp

[1] 2 3 5
```

Para visualizar las operaciones de conjuntos existe una herramienta muy útil llamada diagrama de Venn. Un diagrama de Venn es una representación gráfica, normalmente círculos, que muestramn las relaciones existentes entre los conjuntos. Cada círculo es un conjunto diferente. La forma en que estos círculos se sobreponen entre sí muestra todas las posibles relaciones lógicas entre los conjuntos que representan.

Para crear un diagrama de Venn en R dados los conjuntos es necesario descargar y mandar a llamar la paquetería *ggVennDiagram*:

AQUI VAN LOS DIAGRAMAS DE VENN

La función que genera los diagramas de Venn muestra en dentro de los círculos el número de elementos que comparten, o no, los conjuntos A, B y C.

1.2. Elementos de Probabilidad

1.2.1. Espacio muestral de un experimento

Un conjunto que consiste en todos los resultados de un experimento aleatorio se llama espacio muestral y cada uno de los resultados se denomina punto muestral. Con frecuencia habrá más de un espacio muestral que describe los resultados de un experimento pero hay comúnmente sólo uno que suministra la mayoría de la información. Es importante observar que el espacio muestral corresponde al conjunto universal.

Definición. El espacio muestral de un experimento, denotado por Ω , es el conjunto de todos los posibles resultados de dicho experimento.

Ejemplo 1.12. Si se lanza un dado, un espacio muestral de todos los resultados posibles se da por $\{1, 2, 3, 4, 5, 6\}$, mientras que otro es $\{par, impar\}$

Ejemplo 1.13. Si se examinan tres soldaduras en secuencia y se anota el resultado de cada examen, entonces un resultado del experimento es cualquier secuencia de letras N y D de longitud 3, por tanto $\Omega = \{NNN, NND, NDN, NDD, DNN, DND, DDN, DDD\}$

En R

Además de poder simular experimentos aleatorios sencillos con la función *sample*, R cuenta con una librería que permite simular experimentos aleatorios más complejos para mostrar sus espacios muestrales. A continuación se muestra el espacio muestral del experimento que corresponde a lanzar 2 monedas

```
> install.packages("prob", repos = "http://cran.us.r-project.org")
> library(prob)
> tosscoin(2, makespace=TRUE)
```

	toss1	toss2	probs
1	H	H	0.25
2	T	H	0.25
3	H	T	0.25
4	T	T	0.25

Esta paquetería también cuenta con la función *rolldie* que muestra el espacio muestral correspondiente al experimento de tirar n cantidad de dados:

```
> rolldie(2, makespace = TRUE)
```

	X1	X2	probs
1	1	1	0.02777778
2	2	1	0.02777778
3	3	1	0.02777778
4	4	1	0.02777778
5	5	1	0.02777778
6	6	1	0.02777778
7	1	2	0.02777778
8	2	2	0.02777778
9	3	2	0.02777778
10	4	2	0.02777778
11	5	2	0.02777778
12	6	2	0.02777778
13	1	3	0.02777778
14	2	3	0.02777778
15	3	3	0.02777778
16	4	3	0.02777778
17	5	3	0.02777778
18	6	3	0.02777778
19	1	4	0.02777778
20	2	4	0.02777778

```

21 3 4 0.02777778
22 4 4 0.02777778
23 5 4 0.02777778
24 6 4 0.02777778
25 1 5 0.02777778
26 2 5 0.02777778
27 3 5 0.02777778
28 4 5 0.02777778
29 5 5 0.02777778
30 6 5 0.02777778
31 1 6 0.02777778
32 2 6 0.02777778
33 3 6 0.02777778
34 4 6 0.02777778
35 5 6 0.02777778
36 6 6 0.02777778

```

1.2.2. Muestreo de urnas

Este experimento aleatorio es tal vez el más fundamental. Se tiene una urna que contiene un conjunto de objetos distinguibles (en este caso bolas). Se saquea la urna, se toma una bola, y se observa el resultado. En eso consiste el experimento.

Pero existen todo tipo de variaciones en esta temática. Se podría tomar más de una bola, por ejemplo, dos. ¿Cuáles son los posibles resultados del experimento ahora? Depende de cómo se muestree. Se podría elegir una bola, mirarla, regresarla a la urna y muestrear de nuevo. Otra manera sería elegir una bola, mirarla pero sin ponerla dentro de nuevo, y muestrear de nuevo. Ciertamente hay más posibles resultados del experimento en el segundo que en el último. En el segundo caso se dice que el muestreo está hecho sin reemplazo.

Hay más. Supóngase que no se observó qué bola salió primero. Lo único que se observa son dos bolas, y no se conoce el orden en que fueron extraídas. A esto se le conoce como muestreo desordenado porque el orden de la selección no es de importancia con respecto a lo que se observa. Se podrían elegir las bolas y regresarlas a la bolsa antes de mirarlas.

Es importante notar que esta clase general de experimentos aleatorios contiene como caso especial a todos los experimentos aleatorios comunes experimentales. Lanzar una moneda dos veces es equivalente a elegir dos bolas etiquetadas como Sol y Aguila de la urna, con reemplazo. El experimento de lanzar un dado es equivalente a elegir una bola de la urna con 6 elementos, etiquetados del 1 al 6.

En R

El paquete *prob* es capaz de muestrear de urnas con la función *urnsamples*, la cuál tiene los argumentos *x*, *size*, *replace* y *ordered*. El argumento *x* representa la urna de la cuál el muestreo se hace. El argumento *size* indica el tamaño de la muestra. Los argumentos *ordered* y *replace* son de tipo lógico (*TRUE* o *FALSE*) y especifican si el muestreo se realizará ordenado o desordenado y con reemplazo y sin reemplazo. A continuación se discutirá cada caso.

Ejemplo 1.14 Considérese una urna que contiene 3 bolas, etiquetados con los números 1, 2 y 3, respectivamente. Se tomará una muestra de tamaño 2 de la urna

Ordenado, con Reemplazo

Si el muestreo es con reemplazo, entonces se puede obtener cualquier resultado 1, 2 o 3 en cualquier sorteo. Luego, "ordenado" se refiere a que se hará un seguimiento del orden de los sorteos que se realicen. En R se hace de la siguiente manera

```
> urnsamples(1:3, size=2, replace=TRUE, ordered=TRUE)
```

```

X1 X2
1  1  1
2  2  1
3  3  1
4  1  2
5  2  2
6  3  2
7  1  3
8  2  3
9  3  3

```

Hay que notar que las columnas 2 y 4 son idénticas, excepto por el orden en el cuál los números se muestran. Además, hay que notar que cada posible par de los números del 1 al 3 están listados. Este experimento es equivalente a lanzar un dado de 3 caras dos veces.

Ordenado, sin Reemplazo

Aquí el muestreo es sin reemplazo, por lo que no se observará el mismo número en ambas columnas. Sin embargo, el orden es importante aún, por lo tanto se pueden observar los resultados 1 y 2 y 2 y 1.

```
> urnsamples(1:3, size = 2, replace = FALSE, ordered = TRUE)
```

```
  X1 X2
1   1  2
2   2  1
3   1  3
4   3  1
5   2  3
6   3  2
```

Esta lista es justo la que se esperaba. Hay menos columnas en este resultado que en el anterior debido a la restricción del procedimiento de muestreo. Si los números 1, 2 y 3, representaran a "Juan", "María" "Daniel", respectivamente, entonces el experimento podría ser equivalente a elegir dos de las tres personas para servir como presidente y vicepresidente de una compañía, respectivamente, y el espacio muestra mostrado arriba enlista todas las posibles maneras en las que este experimento podría ser realizado.

Desordenado, sin Reemplazo

De nuevo, no se podrá observar el mismo resultado dos veces, pero en este caso, sólo se considerarán aquellos resultados que cuando sean invertidos no duplican a los anteriores.

```
> urnsamples(1:3, size=2, replace=FALSE, ordered=FALSE)
```

```
  X1 X2
1   1  2
2   1  3
3   2  3
```

Este experimento es equivalente a elegir un par de bolas de la urna, y mirar qué etiquetas tienen. Esta es la configuración estándar de la función *urnsamples*, por lo tanto se obtendría el mismo resultado escribiendo únicamente *urnsamples(1 : 3, 2)*.

Desordenado, con Reemplazo

La última posibilidad es tal vez la más interesante. Se reemplazan las bolas después de cada sorteo, pero sin recordar el orden en el que los sorteos se hicieron.

```
> urnsamples(1:3, size=2, replace=TRUE, ordered=FALSE)
```

```
  X1 X2
1   1  1
2   1  2
3   1  3
4   2  2
5   2  3
6   3  3
```

Existen distintas maneras de interpretar este experimento. Una de ellas consiste en el experimento de poner 2 dados de 3 caras en una taza, agitar la taza y mirar los resultados. Cada renglón del espacio muestra es un par potencial que se podría observar. Otra manera es ver cada resultado como un método por separado de distribuir dos pelotas de golf en 3 cajas etiquetadas con 1, 2 y 3. Independientemente de la interpretación, *urnsamples* enlista cada posible manera en que el experimento puede concluir.

Es importante notar que la urna no necesita contener números; podría asignársele el vector $x=c(\text{rojo}, \text{.azul}, \text{verde})$. Pero hay un punto importante que mencionar antes de proceder. En el ejemplo de las bolas, estas eran distinguibles en el sentido de que cada una tenía una etiqueta para distinguirla de las otras bolas en la urna. Una pregunta natural sería, ¿Qué pasaría si la urna contiene elementos indistinguibles, por ejemplo, $x=c(\text{rojo}, \text{rojo}, \text{.azul})$? La respuesta es que la función *urnsamples* se comporta considerando a cada bola en la urna como distinguible, independientemente de su etiquetado. Por lo tanto, aunque haya dos bolas rojas en la urna, cada bola es distinguible una de la otra.

De este modo, cuando el argumento *x* de *urnsample* tiene elementos repetidos, el espacio muestral puede aparentar estar ordenado (*ordered = TRUE*) incluso si se estableció *ordered = FALSE*. Lo mismo sucede con el argumento *replace*.

1.2.3. Experimentos Aleatorios

En la ciencia los experimentos conforman la base de los conocimientos adquiridos, de ahí su importancia. Un principio fundamental es que si se realizan estos experimentos de manera repetida bajo condiciones similares los resultados serán esencialmente idénticos.

Sin embargo, en ocasiones sucede que, aún bajo condiciones similares, los resultados varían ampliamente. Estos experimentos se denominan experimentos aleatorios.

Definición. Un experimento aleatorio es la reproducción controlada de un fenómeno, existiendo incertidumbre sobre el resultado que se obtendrá.

Por tanto, como ejemplos se pueden considerar el lanzamiento de un dado o una moneda, el salto de un atleta, la predicción del clima de cierto día, seleccionar una carta o más de un mazo, medir el tiempo de recorrido entre un lugar u otro, etc.

En R

Un ejemplo sencillo se puede realizar usando la función *sample*:

Supóngase que de un grupo de 8 alumnos, se necesita elegir a 3 de ellos para llevar a cabo una actividad escolar. Los nombres de los alumnos son: Daniel, Alejandro, Antonio, Daniela, Luis, Judith, Gabriel y Juan.

```
> alumnos<-c("Daniel", "Alejandro", "Antonio", "Daniela", "Luis", "Judith", "Gabriel", "Juan")
> sample(alumnos, size=3, replace=FALSE)
```

```
[1] "Luis"      "Alejandro" "Antonio"
```

Al correr el código de nuevo se obtendrán, generalmente, distintos nombres porque la función *sample* ejecuta un experimento aleatorio.

Para simular el experimento de lanzar un dado usando la función *sample* el código queda así:

```
> caras_de_dado<-c(1,2,3,4,5,6)
> sample(caras_de_dado,1)
```

```
[1] 2
```

1.2.4. Eventos

Otro elemento de probabilidad importante en el estudio de los experimentos aleatorios es el concepto de evento. Su importancia se debe a que los eventos son el conjunto de resultados posibles de un experimento, y por tanto los eventos son subconjuntos del espacio muestral. Existen distintos tipos de eventos y cada uno tiene propiedades individuales.

Definición: Un evento es cualquier recopilación de resultados contenidos en el espacio muestral Ω .

Existen distintos tipos de eventos. Algunos de los más importantes se mencionan a continuación:

Eventos independientes y dependientes

Los eventos independientes son aquellos cuyo resultado no depende de un resultado anterior. No importa cuántas veces se realice un experimento la probabilidad de que sucedan los eventos independientes será la misma. Por ejemplo, lanzar una moneda es un evento independiente.

Los eventos dependientes son eventos cuyo resultado depende de un resultado anterior. Esto implica que la probabilidad de ocurrencia de un evento dependiente se verá afectada por un resultado previo. Por ejemplo, sacar 2 canicas una después de otra de una bolsa sin remplazo.

Eventos imposibles y seguros

Un evento que no puede suceder nunca se conoce como evento imposible. Dado que los eventos imposibles nunca sucederán, su probabilidad de ocurrencia es 0. Por ejemplo, el sol nunca girará al rededor de la tierra.

El evento seguro es aquel cuya probabilidad de ocurrencia es siempre 1. Por ejemplo, la tierra girando al rededor del sol.

Eventos simples y compuestos

Si un evento consiste de un único resultado del espacio muestral, se le llama evento simple. El evento de obtener un número menor que 2 después de lanzar un dado, $E = \{1\}$, es un evento simple.

Si un evento consiste de más de un único resultado del espacio muestral, se denomina evento compuesto. Un ejemplo de evento compuesta es lanzar un dado y obtener un número impar: $E = \{1, 3, 5\}$

Eventos complementarios

Cuando hay dos eventos tales que uno de ellos puede ocurrir si y solo si el otro no sucede, entonces tales eventos se conocen como eventos complementarios. La suma de la probabilidad de la ocurrencia de los eventos complementarios debe ser 1 (la razón de esta suma se abordará más a fondo en la siguiente sección). Por ejemplo, sea A el resultado de lanzar una moneda y obtener águila, y sea S el resultado de lanzar la misma moneda y obtener sol. Entonces los eventos A y S son eventos complementarios. Estos eventos son mutuamente excluyentes y exhaustivos.

Eventos mutuamente excluyentes

Los eventos que no pueden suceder al mismo tiempo se conocen como eventos mutuamente excluyentes. Por lo tanto, estos eventos no tienen resultados en común. Podría parecer que este tipo de eventos y los eventos complementarios son los mismos, sin embargo, es importante recordar que la suma de las probabilidades de los eventos complementarios siempre es igual a 1, mientras que los eventos mutuamente excluyentes no necesariamente cumplen con esta propiedad. Por lo tanto se puede afirmar que los eventos complementarios son un subconjunto de los eventos mutuamente excluyentes.

Eventos colectivamente exhaustivos

Se conoce como eventos colectivamente exhaustivos al conjunto de eventos que incluye todos los resultados posibles de un experimento. Por ejemplo, cuando lanzamos una moneda al aire se espera que salga cara o cruz y, se dice que son colectivamente exhaustivos ya que representan todos los resultados posibles del lanzamiento de una moneda.

Eventos igualmente probables

Son aquellos cuya probabilidad de resultar es la misma. Por ejemplo, la probabilidad de obtener águila después de lanzar una moneda es $1/2$ al igual que la probabilidad de obtener sol.

Funciones para encontrar subconjuntos

Puede llegar el momento en el que el subconjunto evento de interés sea complicado de especificar. Sin embargo,

1.3. Propiedades de Probabilidad

1.3.1. Funciones de Probabilidad

Una **función de probabilidad** es una regla que asocia a cada evento A del espacio muestral un único número $P(A) = p$, llamado la probabilidad de A . Para asegurarnos de que las asignaciones de estas probabilidades en cualquier experimento sean congruentes y homogéneas es necesario establecer 3 axiomas:

Axioma 1. Para cualquier evento A , $P(A) \geq 0$

Axioma 2. $P(\Omega) = 1$

Axioma 3. Sea A_1, A_2, A_3, \dots un conjunto de eventos disjuntos, entonces

$$P(A_1 \cup A_2 \cup A_3 \cup \dots) = \sum_{i=1}^{\infty} P(A_i) \quad (1.1)$$

El primer axioma establece y asegura que no pueden haber posibilidades negativas. Dado que el espacio muestral contiene a todos los eventos posibles, entonces la probabilidad de que suceda el espacio muestral siempre es 1, que es la máxima probabilidad posible: el 100 %. El tercer axioma significa que para una secuencia de eventos disjuntos, su probabilidad total debe ser igual a la suma de sus partes. Por ejemplo, la probabilidad de obtener un 1 o un 2 después de lanzar un dado es igual a la probabilidad de obtener un 1 más la probabilidad de obtener un 2.

1.3.2. Algunas Propiedades

Para cualesquiera eventos A y B ,

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

Demostración. Ya que $A \cup A^c = \Omega$ y $A \cap A^c = \emptyset$, entonces

$$1 = P(\Omega) = P(A \cup A^c) = P(A) + P(A^c)$$

2. $P(\emptyset) = 0$

Demostración. Se puede ver que $\Omega^c = \emptyset$, y después se usa la Propiedad 1.

3. Si $A \subset B$, entonces $P(A) \leq P(B)$.

Demostración. Reescribiendo $B = A \cup (B \cap A^c)$, y se puede notar que $A \cap (B \cap A^c) = \emptyset$; por lo tanto

$$P(B) = P(A \cup (B \cap A^c)) = P(A) + P(B \cap A^c) \geq P(A)$$

ya que $P(B \cap A^c) \geq 0$

4. $0 \leq P(A) \leq 1$

Demostración. La desigualdad de la izquierda es inmediata del Axioma 1. La segunda desigualdad es resultado de la Propiedad 3 ya que $A \subset S$.

5. **La Regla de Adición General.**

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

Demostración. Hay que notar que $A \cup B$ se puede separar en 2 eventos que son excluyentes, a saber, A y $B \cap A^c$. Adicional a eso, B es la unión de $A \cap B$ con $A^c \cap B$, entonces $P(B) = P(A \cap B) + P(A^c \cap B)$. Por lo tanto

$$P(A \cup B) = P(A) + P(B \cap A^c) = P(A) + P(B) - P(A \cap B)$$

6. El Teorema de Probabilidad Total

Sean B_1, B_2, \dots, B_n eventos mutuamente exclusivos y exhaustivos. Entonces

$$P(A) = P(A \cap B_1) + P(A \cap B_2) + \dots + P(A \cap B_n)$$

1.3.3. Asignación de Probabilidades

Un modelos de particular interés es el **modelo igualmente probable**. La idea consiste en dividir al espacio muestra Ω en una colección finita de eventos elementales a_1, a_2, \dots, a_N que sean igualmente probables en el sentido de que cada a_i tenga probabilidades iguales de suceder. La función de probabilidad asociada con este modelo debe satisfacer $P(\Omega) = 1$, por el Axioma 2. Por otro lado, debe satisfacer también

$$P(\Omega) = P(a_1, a_2, \dots, a_N) = P(a_1 \cup a_2 \cup \dots \cup a_N) = \sum_{i=1}^N P(a_i)$$

por el Axioma 3. Ya que $P(a_i)$ es la misma para todo i , cada evento debe tener necesariamente una probabilidad de $1/N$.

Para un evento $A \subset \Omega$, se escribe como una colección de resultados elementales: si $A = a_{i_1}, a_{i_2}, \dots, a_{i_k}$ entonces A tiene K elementos y

$$P(A) = P(a_{i_1}) + P(a_{i_2}) + \dots + P(a_{i_k}) = \frac{1}{N} + \frac{1}{N} + \dots + \frac{1}{N} = \frac{k}{N} = \frac{\#A}{\#\Omega}$$

En otras palabras, bajo el modelo igualmente probable, la probabilidad de un evento A está determinada por el número de eventos elementales que contiene A .

Ejemplo 1.15. Considerando el experimento E de lanzar una moneda. Entonces el espacio muestra es $\Omega = \{Sol, Aguila\}$, y bajo el modelo igualmente probable, estos dos resultados tienen probabilidades $P(Sol) = P(Aguila) = 1/2$. Asumiendo que la moneda no está cargada.

Ejemplo 1.16. Considerando el experimento E consiste en lanzar una moneda dos veces. El espacio muestral puede ser representado por $\Omega = \{SS, SA, AA, AS\}$. Dado que la moneda no está cargada y que la moneda se lanza de manera idéntica e independiente ambas veces, es razonable aplicar el modelo igualmente probable.

¿Cuál es la probabilidad $P(\text{al menos 1 Sol})$? Mirando el espacio muestral se ven los elementos SS, SA, AS que tienen al menos 1 Sol; por tanto, $P(\text{al menos 1 Sol}) = 3/4$.

¿Cuál es la probabilidad $P(\text{ningun Sol})$? Se puede notar que el evento $\{\text{ningun Sol}\} = \{\text{al menos un Sol}\}^c$, lo cuál, por la Propiedad 1, es igual a decir que $P(\text{ningun Sol}) = 1 - P(\text{al menos 1 Sol}) = 1 - 3/4 = 1/4$. Este resultado parece obvio en este sencillo ejemplo, sin embargo, el uso de esta propiedad en circunstancias más complicadas resulta sumamente útil.

Ejemplo 1.17. Considérese un mazo de 52 cartas. Estas están usualmente etiquetadas con 4 palos: Espadas, Diamantes, Corazones y Tréboles, y las 13 ranks: 2, 3, 4, ..., 10, J, Q, K y A. Dependiendo del juego, A puede estar 2 lugares debajo o arriba de K.

Sea E el experimento consistente en elegir una carta del mazo, y sea el resultado la cara de la carta. Sean los eventos $A = \{\text{obtener una A}\}$ y $B = \{\text{obtener un trébol}\}$. Es importante recordar que solo se está tomando una carta.

Inmediatamente se tiene que $P(A) = 4/52$ ya que solo hay 4 As en el mazo; similarmente, hay 13 Tréboles en el mazo, lo que implica que $P(B) = 13/52$.

¿Cuál es la probabilidad $P(A \cap B)$? Se puede ver que sólo hay una carta de las 52 que es una A y un Trébol al mismo tiempo, por lo tanto $P(A \cap B) = 1/52$

¿Cuál es la probabilidad $P(A \cup B)$? Por la Propiedad 5

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = 4/52 + 13/52 - 1/52 = 16/52$$

1.4. Probabilidad Condicional

Considera un mazo completo de 52 cartas estándar. Ahora, selecciona 2 cartas del mazo, sucesivamente. Sea $A = \{\text{la primera carta obtenida es un As}\}$ y $B = \{\text{la segunda carta es un As}\}$. Ya que hay 4 Aces en el mazo, es natural asignar $P(A) = 4/52$. Suponiendo que se ve la primera carta. ¿Cuál es la probabilidad de B ahora? Claro que la respuesta depende del valor de la primera carta. Si la primera carta es un As, entonces la probabilidad de que la segunda carta también sea un As es $3/51$, pero si la primera carta no es un As, entonces la probabilidad de que la segunda carta sea un As es $4/51$. La notación para estas probabilidades es

$$P(B|A) = 3/51, P(B|A^c) = 4/51$$

y se lee como la probabilidad de que suceda B dado que sucedió A .

Entonces, por medio de este experimento, es notorio que existen situaciones donde las probabilidades que se asignen a cada evento del espacio muestral, dependerán de resultados conocidos previamente

Definición. La probabilidad condicional de B dada A , denotada por $P(B|A)$, está definida por

$$P(B|A) = \frac{P(A \cap B)}{P(A)}, \text{ si } P(A) > 0$$

Ejemplo 1.18. Se lanza una moneda 2 veces. El espacio muestral es $\Omega = \{SS, SA, AA, AS\}$. Sea el evento $A = \{\text{obtener un Sol}\}$ y $B = \{\text{Obtener un Sol y un Aguila}\}$. Es claro que $P(A) = 3/4$, $P(B) = 2/4$ y $P(A \cap B) = 2/4$. ¿Cuáles son las probabilidades $P(A|B)$ y $P(B|A)$?

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

en otras palabras, una vez que se conoce que ocurrió un Sol y un Aguila, se puede estar seguros de que ocurrió un Sol.

Después

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

lo que significa que dada la información de que ha ocurrido un Sol, ya no es necesario tomar en cuenta el resultado AA , y los 3 resultados restantes son igualmente probables con exactamente 2 resultados que se encuentran en el conjunto B .

Ejemplo 1.19. Se lanza un dado de 6 caras 2 veces. El espacio muestral consiste en todos los pares ordenados (i, j) de los números 1, 2, ..., 6, es decir, $\Omega = \{(1, 1), (1, 2), \dots, (6, 6)\}$. El número de eventos en este espacio muestral es de 36. Sea $A = \{\text{igual número}\}$ y $B = \{\text{suma de resultados mayor a o igual a 8}\}$. El espacio muestral puede ser representado mediante una matriz:

Los resultados que están en el evento A se marcan con el símbolo \times , los resultados que se encuentran en B se marcan con el símbolo \circ , y los que se encuentran tanto en A como en B se marcan con \otimes . Ahora es claro que $P(A) = 6/36$, $P(B) = 15/36$ y $P(A \cap B) = 3/36$. Finalmente,

$$P(A|B) = \frac{3/36}{15/36} = \frac{1}{5}, P(B|A) = \frac{3/36}{6/36} = \frac{1}{2}$$

De nuevo, se puede ver que dado el conocimiento de que B ocurrió (los 15 resultados en el triángulo inferior derecho), hay 3 de los 15 que pertenecen al set A , por lo tanto la probabilidad es $3/15$. Similarmente, dado que A ocurrió (los elementos de la diagonal), hay 3 de los 6 resultados que también pertenecen a B , por tanto, la probabilidad de B dado A es $1/2$.

en R

Continuando con el Ejemplo 1.18, el primer paso es establecer el espacio de probabilidad con la función *rolldie*.

```
> library(prob)
> S<-rolldie(2,makespace=T)
> head(S)
```

```
  X1 X2      probs
1  1  1 0.02777778
2  2  1 0.02777778
3  3  1 0.02777778
4  4  1 0.02777778
5  5  1 0.02777778
6  6  1 0.02777778
```

Luego se definen los eventos

```
> A<-subset(S, X1==X2)
> B<-subset(S, X1+X2>=8)
```

Finalmente se calculan las probabilidades. Para hacerlo, se usa el argumento *given* de la función *prob*:

```
> Prob(A, given = B)
[1] 0.2
> Prob(B, given = A)
[1] 0.5
```

1.4.1. Propiedades y Reglas

El siguiente teorema establece que las probabilidades condicionales se comportan justo como las probabilidades regulares cuando el evento condicionante está fijo.

Teorema. Para cualquier evento fijo A con $P(A) > 0$

1. $P(B|A) \geq 0$, para todos los eventos $B \subset \Omega$,
2. $P(\Omega|A) = 1$,
3. Si B_1, B_2, B_3, \dots son eventos disjuntos, entonces

$$P(\bigcup_{k=1}^{\infty} B_k|A) = \sum_{k=1}^{\infty} P(B_k|A)$$

En otras palabras, $P(\cdot|A)$ es una función de probabilidad. Con este hecho en mente, las siguientes probabilidades son inmediatas:

Proposición. Para cualesquiera A, B y C con $P(A) > 0$,

1. $P(B^c|A) = 1 - P(B|A)$.
2. Si $B \subset C$ entonces $P(B|A) \leq P(C|A)$.
3. $P[(B \cup C)|A] = P(B|A) + P(C|A) - P[(B \cap C)|A]$.
- 4.

1.4.2. La Regla de la Multiplicación.

Para cualesquiera 2 eventos A y B ,

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

Y más generalmente, para los eventos $A_1, A_2, A_3, \dots, A_n$

$$P(A_1 \cap A_2 \cap \dots \cap A_n) = P(A_1)P(A_2|A_1) \dots P(A_n|A_1 \cap A_2 \cap \dots \cap A_{n-1}).$$

La regla de la multiplicación resulta muy importante porque permite encontrar probabilidades en experimentos aleatorios que tienen una estructura secuencial como muestra el siguiente ejemplo.

Ejemplo 1.20. Retomando el experimento de las 52 cartas, la pregunta es ¿cuál es la probabilidad $P(\text{ambos son Ases})$?

$$P(\text{ambos son Ases}) = P(A \cap B) = P(A)P(B|A) = \frac{4}{52} \frac{3}{51} \approx 0,00452.$$

En R

Para el Ejemplo 1.20, se establece el espacio de probabilidad siguiendo 3 pasos. Primero se emplea la función `cards` para obtener un data frame `L` con 2 columnas: `rank` y `suit`. Ambas columnas son almacenadas internamente como factores con 13 y 4 niveles, respectivamente.

Después se muestrean 2 cartas aleatoriamente de `L` por medio de la función `urnsamples`. Esto da como resultado una lista `M` que contiene todos los posibles pares de renglones de `L`. El espacio muestral para este experimento es entonces la lista `M`.

Finalmente se asocia un modelo de probabilidad al espacio muestral. Esto se logra con la función `probspace`. Se asume el modelo igualmente probable por default. Se le asigna a este resultado la variable `N`.

Este objeto `N` es una lista con 2 elementos: los resultados y sus probabilidades. El elemento de resultados es otra lista y el elemento de probabilidades es un vector. De esta manera

```
> library(prob)
> L<-cards()
> M<-urnsamples(L,size=2)
> N<-probspace(M)
```

Ahora que se tiene el espacio de probabilidad `N`, ya es posible calcular algunas probabilidades con la función `prob`. Solo para recordar, el evento que se busca es $P(\text{ambos son Ases})$

```
> Prob(N, all(rank=="A"))
```

```
[1] 0.004524887
```

Ejemplo 1.21. Considérese una urna con 10 pelotas adentro, de las cuales 7 son rojas y 3 de las cuales son verdes. Selecciónense 3 pelotas sucesivamente de la urna. Sea $A = \{\text{La primer pelota es roja}\}$, $B = \{\text{La segunda pelota es roja}\}$, $C = \{\text{La tercera pelota es roja}\}$. Entonces

$$P(\text{las 3 pelotas son rojas}) = P(A \cap B \cap C) = \frac{7}{10} \frac{6}{9} \frac{5}{8} \approx 0,2917.$$

Ejemplo 1.22. Considérense 2 urnas, la primera con 5 bolas rojas y 3 bolas verdes, y la segunda con 2 bolas rojas y 6 bolas verdes. Una persona elige aleatoriamente una bola de la primera urna y la transfiere a la segunda, sin mostrar el color de la bola. Se elige una bola de la segunda urna. ¿Cuál es la probabilidad de que la bola seleccionada sea roja? Sea $A = \{\text{la bola transferida es roja}\}$ y $B = \{\text{la bola seleccionada es roja}\}$. Escribiendo

$$B = \Omega \cap B = (A \cup A^c) \cap B = (A \cap B) \cup (A^c \cap B)$$

Se puede notar que $A \cap B$ y $A^c \cap B$. Por lo tanto

$$P(B) = P(A \cap B) + P(A^c \cap B) = P(A)P(B|A) + P(A^c)P(B|A^c) = \frac{5}{8} \frac{3}{9} + \frac{3}{8} \frac{2}{9} = \frac{21}{72}$$

1.5. Teorema de Bayes

En esta sección se introducirá una regla que permite actualizar las probabilidades a partir de información nueva disponible.

Teorema. Sean B_1, B_2, \dots, B_n eventos mutuamente excluyentes y exhaustivos y sea A un evento con $P(A) > 0$. Entonces

$$P(B_k|A) = \frac{P(B_k)P(A|B_k)}{\sum_{i=1}^n P(B_i)P(A|B_i)}, \quad k = 1, 2, \dots, n.$$

Demostración. La demostración se sigue de observar a $P(B_k \cap A)$ en dos maneras distintas. Para simplificar, supóngase que $P(B_k) > 0$ para toda k . Entonces

$$P(A)P(B_k|A) = P(B_k \cap A) = P(B_k)P(A|B_k)$$

Ya que $P(A) > 0$ puede pasar al otro lado dividiendo

$$P(B_k|A) = \frac{P(B_k)P(A|B_k)}{P(A)}$$

Ahora, recordando que $\{B_k\}$ es una partición, el Teorema de Probabilidad Total dice que el denominador de la expresión anterior es

$$P(A) = \sum_{k=1}^n P(B_k \cap A) = \sum_{k=1}^n P(B_k)P(A|B_k)$$

¿Qué significa? Usualmente en las aplicaciones son dadas las probabilidades a priori $P(B_k)$. Se recolecta información o datos, los cuales se representan con el evento A . Se quiere conocer cómo actualizar la probabilidad $P(B_k)$ a $P(B_k|A)$. La respuesta se encuentra en la regla de Bayes.

Ejemplo 1.33 Archivo de asistentes incorrecto. En este problema, hay 3 asistentes trabajando en una compañía: Luis, José y Rosa. Su principal trabajo consiste en llenar papaleo en el cubículo de llenado cuando los papeles se encuentran disponibles. Los tres asistentes tienen diferentes horarios laborales:

	Luis	José	Rosa
Carga de trabajo	60 %	30 %	10 %

Es decir que Luis trabaja el 60% del tiempo, José el 30% y Rosa el 10% restante, y llenan documentos a la misma velocidad aproximadamente. Suponiendo ahora que una persona elige aleatoriamente uno de los documentos del cubículo, entonces hay 3 opciones, que se pueden representar con los eventos $L = \{\text{Documento llenado por Luis}\}$, $J = \{\text{Documento llenado por Jose}\}$ y $R = \{\text{Documento llenado por Rosa}\}$. ¿Cuáles son las probabilidades de cada evento? Sin más información por el momento, estas probabilidades serían

	Luis	José	Rosa
Probabilidad a priori	0.60	0.30	0.10

Un día la jefa llega, abre el cubículo, y elige un archivo aleatoriamente. La jefa descubre que el documento ha sido llenado de manera incorrecta. La jefa está tan enojada que amenaza con despedir a la persona que cometió el error. La pregunta es: ¿Quién se equivocó?

La jefa decide usar la probabilidad para decidir, y mira directamente al horario de los 3 trabajadores. Ella razona que, ya que los 3 empleados trabajan a la misma velocidad, la probabilidad de que un documento

aleatorio haya sido llenado por cada uno sería proporcional a su carga de trabajo. La jefa notifica a Luis que tiene hasta el final del día para vaciar su escritorio.

Pero Luis argumenta en su defensa que la jefa ha ignorado información adicional. La probabilidad de que Luis haya cometido un error llenando un documento es menor que la de Jo'se y la de Rosa, ya que es un trabajador diligente que pone atención a su trabajo. Luis admite que trabaja más que sus otros dos compañeros, pero que no comete tantos errores como ellos. Por tanto, Luis recomienda que antes de tomar una decisión, la jefa debería actualizar la probabilidad (basada inicialmente en la carga de trabajo únicamente) para incorporar la probabilidad de haber observado un documento mal llenado.

Y resulta que la jefa tiene información acerca de la precisión de llenado de los 2 trabajadores (debido a evaluaciones de desempeño). La información del desempeño se puede representar mediante la siguiente tabla

	Luis	José	Rosa
Índice de error	0.003	0.007	0.010

En otras palabras, en promedio, Luis llena incorrectamente el 0.03 % de sus documentos. De esta manera, Luis es el trabajador más acertado, posteriormente José y finalmente Rosa. Si la jefa tuviera que tomar una decisión basada en la precisión de cada trabajador, entonces Rosa sería despedida. Pero Rosa escucha y argumenta que ella sólo trabaja por un periodo corto en el día, y consecuentemente comete pocos errores; hay una muy pequeña oportunidad de que ella haya llenado incorrectamente el documento.

A la jefa le gustaría utilizar esta información para actualizar las probabilidades de los 3 trabajadores, es decir, quiere utilizar la probabilidad adicional de que el documento fue llenado para actualizar sus creencias acerca del probable culpable. Sea A el evento de que un documento sea mal llenado. A la jefa le gustaría conocer las probabilidades

$$P(L|A), P(J|A) \text{ y } P(R|A)$$

Sólo se mostrará el cálculo para $P(L|A)$ ya que los otros dos cálculos serán similares. Se debe usar la Regla de Bayes en la forma

$$P(L|A) = \frac{P(M \cap A)}{P(A)}$$

Para encontrar $P(L|A)$ hay que recordar que esto es igual a $P(L) \cdot P(A|L)$ por la Regla de la Multiplicación. Ya se conoce que $P(L) = 0,6$ y $P(A|L)$ es el índice de error de Luis, siendo igual a $P(A|L) = 0,003$. Por lo tanto

$$P(L \cap A) = (0,6)(0,003) = 0,0018$$

Usando el mismo procedimiento para los demás trabajadores

$$P(J \cap A) = 0,0021 \text{ y } P(R \cap A) = 0,0010$$

Ahora resta encontrar $P(A)$. La clave aquí es la noción de que si el documento está mal llenado, entonces o Luis o José o Rosa debieron haberlo llenado; no hay nadie más que haga ese trabajo y por tanto cometa ese error. Más aún, las probabilidades son mutuamente excluyentes. Se puede usar el Teorema de Probabilidad Total, de manera que

$$P(A) = P(A \cap L) + P(A \cap J) + P(A \cap R)$$

Usando los resultados calculados arriba

$$P(A) = 0,0018 + 0,0021 + 0,0010 = 0,0049$$

Por lo tanto, la Regla de Bayes indica que

$$P(L|A) = \frac{0,0018}{0,0049} \approx 0,37$$

Esta última cantidad es llamada la probabilidad posterior de que Luis haya llenado incorrectamente el documento, ya que incorpora los datos observados de que un documento elegido aleatoriamente haya sido mal llenado. De este modo se puede obtener la siguiente tabla

	Luis	José	Rosa
Probabilidad Posterior	0.37	0.43	0.20

La conclusión es que José será despedido. Lo que sucede es que hay una interacción entre el tiempo de trabajo y el índice de error de cada trabajador. No es obvio quién será el perdedor, y es necesario consultar la Regla de Bayes para determinar la acción más conveniente.

Ejemplo 1.34. La jefa se toca el corazón y no despedirá a nadie. Pero al día siguiente, la jefa elige un archivo aleatorio de nuevo y vuelve a encontrar que está mal llenado. Para decidir a quien despedir ahora, la jefa usará el mismo procedimiento, con un pequeño cambio. No usará las probabilidades 60 %, 30 % y 10 %; ahora es información vieja. En vez de eso, reemplazará las probabilidades a priori con las probabilidades posteriores recién calculadas. Después del cálculo la jefa tendrá nuevas probabilidades posteriores, más actualizadas que las del día anterior.

De esta manera, las probabilidades encontradas con la Regla de Bayes siempre están actualizándose con base en la información disponible en el momento.

En R

No hay funciones especiales para la Regla de Bayes en el paquete *prob*, pero problemas como el anterior son fáciles de calcular a mano.

Ejemplo 1.35. Continuación del ejemplo de los documentos mal llenados. Se almacenan las probabilidades a priori y los índices de error en vectores

```
> apriori<-c(0.6,0.3,0.1)
> error<-c(0.003,0.007,0.01)
> posterior<-apriori*error
> posterior/sum(posterior)

[1] 0.3673469 0.4285714 0.2040816
```

Estos resultados son similares a los que se calcularon a mano. Más adelante se reemplazará *prior* por *post* para un cálculo actualizado de probabilidad. También se podría elevar el error a alguna potencia para observar cómo se ve afectada la probabilidad posterior.

Ejemplo 1.36. La jefa está incorporando la probabilidad posterior del ejemplo anterior y sabe que los trabajadores han llenado incorrectamente 7 documentos más. ¿Cuáles serían las nuevas probabilidades?

```
> nuevapriori<-posterior
> posterior<-nuevapriori*error^7
> posterior/sum(posterior)

[1] 0.0003355044 0.1473949328 0.8522695627
```

Ahora, la persona con la mayor probabilidad de ser despedida ya no es José, sino Rosa.

Hay dos puntos importantes. Primero, no se dividió *posterior* por la suma de sus entradas hasta el último paso; no es necesario calcularlo, y ahorrará tiempo de cálculo a la computadora. Se puede posponer hasta el momento en que sea necesario usarlo, a saber, hasta que los resultados se quieran conocer en forma de probabilidades.

Segundo, el lector podría preguntarse qué resultado obtendría la jefa si decidiera saltarse el paso intermedio de calcular la probabilidad posterior de un sólo archivo mal llenado. ¿Qué sucedería si empezara a partir de la probabilidad apriori original, después hubiera observado 8 archivos mal llenados, y finalmente hubiera calculado la probabilidad posterior?. Por supuesto que debería ser la misma respuesta.

```
> posterior_rapida<-apriori*error^8
> posterior_rapida/sum(posterior_rapida)

[1] 0.0003355044 0.1473949328 0.8522695627
```

Tal como se esperaba que sucediera.

1.6. Técnicas de Conteo

El modelo igualmente probable es una manera popular y conveniente de analizar experimentos aleatorios. Y cuando el modelo igualmente probable se aplica, encontrar la probabilidad de un evento A tiene como propósito contar el número de resultados que contiene A (junto con el número de eventos en Ω). Por tanto, es de suma importancia ser capaz de contar el número de resultados en eventos de todos los tipos.

Proposición. El principio de multiplicación. Supóngase que un experimento se compone de dos pasos sucesivos. Además, supóngase que el primer paso puede ser realizado de n_1 maneras distintas mientras que el segundo paso puede ser realizado de n_2 maneras distintas. Entonces el experimento puede ser realizado de $n_1 \cdot n_2$ distintas maneras.

Más generalmente, si el experimento se compone de k pasos sucesivos que pueden ser realizados de n_1, \dots, n_k distintas formas, entonces el experimento puede ser realizado de $n_1 n_2 \dots n_k$ distintas maneras.

Ejemplo 1.37. Se quiere ordenar una pizza. Esta deberá tener queso (y salsa marinara), pero se puede elegir uno o más de los siguientes 5 ingredientes: peperoni, salsa, anchoas, aceitunas y chile. ¿Cuántas distintas pizzas son posibles de ordenar?

Hay muchas maneras de analizar el problema, pero la más sencilla es utilizando el Principio de Multiplicación directamente. Es conveniente separar la acción de pedir la pizza en una serie de etapas. En la primera etapa se decidirá si se incluirá peperoni en la pizza (2 posibilidades). En la segunda etapa se decidirá si se le agregará o no salsa a la pizza (de nuevo 2 posibilidades). Se continúa de esta manera hasta decidir si la pizza llevará o no chile.

En cada etapa se presentan 2 opciones de elegir la forma en que se preparará la pizza. El Principio de Multiplicación dice que se deben multiplicar los 2's para encontrar el número total de pizzas posibles: $2 \cdot 2 \cdot 2 \cdot 2 \cdot 2 = 2^5 = 32$.

Ejemplo 1.38. Para comprar un computadora existe la posibilidad de elegir entre 2 procesadores, 3 distintos sistemas operativos, 4 capacidades de memoria RAM, 4 capacidades de memoria ROM, y 10 distintos monitores. ¿Cuántas posibles computadoras se pueden formar? De acuerdo con el principio de multiplicación: $2 \cdot 3 \cdot 4 \cdot 4 \cdot 10 = 960$

1.6.1. Muestreo Ordenado

Imagina una bolsa con n pelotas distinguibles dentro. Ahora sacude la bolsa y elige k pelotas aleatoriamente. ¿Cuántas posibles secuencias se pueden observar?

Proposición. El número de maneras en las que se puede elegir una muestra ordenada de k elementos de una población que tiene miembros distinguibles es

- n^k si el muestreo se hace con reemplazo,
- $n(n-1)(n-2)\dots(n-k+1)$ si el muestreo se hace sin reemplazo.

Recordando la notación factorial:

$$n! = n(n-1)(n-2)\dots 3 \cdot 2 \cdot 1$$

Entonces, el número de permutaciones de n elementos es $n!$

Ejemplo 1.39. Al lanzar una moneda 7 veces ¿cuántas secuencias de Sol y Aguila son posibles?: $2^7 = 128$

Ejemplo 1.40. En un salón con 20 estudiantes, se elige aleatoriamente un presidente de clase, un vicepresidente y un tesorero. ¿Cuántas maneras existen de elegir dichos alumnos?: $20 \cdot 19 \cdot 18 = 6840$.

Ejemplo 1.41. Un grupo de amigos rentó 5 películas para ver en el lapso de 2 noches. Desean ver 3 películas en la primer noche. ¿Cuántas secuencias distintas de 3 películas pueden mirar en una noche?: $5 \cdot 4 \cdot 3$.

1.6.2. Muestras Desordenadas

El número de maneras en las que se puede seleccionar una muestra desordenada de k elementos de una población que tiene n miembros distinguibles es

- $(n-1+k)!/[(n-1)!k!]$ si el muestreo se hace con reemplazo,
- $n!/ [k!(n-k)!]$ si el muestreo es hecho sin reemplazo

La cantidad $n!/ [k!(n-k)!]$ se llama **coeficiente binomial** y juega un rol especial en las matemáticas; se denota como

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

y se lee como las combinaciones de n en k .

Ejemplo 1.42. Rentas 5 películas para ver en el lapso de 2 noches, pero sólo quieres ver 3 películas en la primer noche. Tu amigo quiere que le prestes algunas películas para ver en su casa en la primer noche. Le debes a tu amigo un favor, y dejas que elija 2 películas de las 5 que rentaste. ¿Cuántas opciones tiene tu amigo para elegir?: $\binom{5}{2} = 10$

Ejemplo 1.43. Un experimento consiste en colocar 3 dados de 6 caras dentro de una taza. Después, se agita la taza y se avientan los dados hacia afuera. ¿Cuántas posibilidades en los dados se pueden obtener?: $(6-1+3)!/[(6-1)!3!] = \binom{8}{3} = 56$.

En R

El factorial $n!$ se calcula con el comando *factorial*(n) y el coeficiente binomial $\binom{n}{k}$ con el comando *choose*(n, k).

Los espacios muestrales que hasta ahora se han calculado en este yexto han sido relativamente pequeños, y son fáciles de estudiar visualmente. Sin embargo, es muy sencillo generar espacios muestrales que son

considerablemente grandes. Y aunque R es una herramienta muy poderosa, también tiene límites que deben ser tomados en cuenta.

Pero generalmente no es necesario generar el espacio muestral; basta con contar el número de resultados. La función *nsamp* calcula el número de columnas en un espacio muestra generado por la función *urnsamples* sin tener que dedicar los recursos de memoria necesarios para generar el espacio. Los argumentos son *n*, el número de objetos distinguibles en la urna, *k* el tamaño de la muestra, y *replace*, *ordered*, que indican que el conteo es ordenado y con reemplazo, como se explicó anteriormente.

Ejemplo 1.44. Queremos calcular el número de resultados para cada caso de la función *urnsamples* visto en el Ejemplo 1.14. Se tomó una muestra de tamaño 2 de la urna con 3 elementos distinguibles

```
> nsamp(n=3,k=2,replace=TRUE, ordered=TRUE)
[1] 9

> nsamp(n=3,k=2,replace=FALSE, ordered=TRUE)
[1] 6

> nsamp(n=3,k=2,replace=FALSE, ordered=FALSE)
[1] 3

> nsamp(n=3,k=2,replace=TRUE, ordered=FALSE)
[1] 6
```

Se pueden comparar estas respuestas con el tamaño de las listas generadas en el Ejemplo 1.14, notando que son coincidentes.

El Principio de Multiplicación

Un beneficio de *nsamp* es que está vectorizado por lo que ingresar vectores en vez de números para los argumentos *n*, *k*, *replace* y *ordered* resulta en un vector. Esto se vuelve particularmente conveniente para problemas de combinatoria.

Ejemplo 1.45. Hay 11 artistas y cada uno tiene un portafolio con 7 pinturas para competir en una exhibición de arte. Desafortunadamente, el director de la galería sólo puede elegir 12 pinturas ganadoras para acomodarlas en una fila separada igualmente en 3 muros consecutivos. El director decide darles al primer, segundo y tercer lugar un muro para presentar el trabajo de su elección. Las paredes cuentan con 31 opciones de iluminación separadas cada una. ¿Cuántas formas posibles hay de presentar? Solución: Los jueces elegirán 3 ganadores (clasificados) de los 11 participantes, es decir, *replace* = *Flase*, *ordered* = *TRUE*. Cada artista elegirá 4 de sus 7 pinturas en una fila (*replace* = *FALSE*, *oredered* = *TRUE*), y finalmente, cada una de las 3 paredes tiene 31 posibilidades de iluminación (*replace* = *TRUE*, *ordered* = *TRUE*). Estos 3 números se puede calcular rápidamente mediante

```
> n=c(11,7,31)
> k=c(3,4,3)
> r=c(FALSE,FALSE,TRUE)
> x=nsamp(n,k,rep=r, ord=TRUE)
```

Nótese que *ordered* siempre será *TRUE*; *nsamp* reusará *ordered* y *replace* dada la longitud del vector. Por el Principio de Multiplicación, el número de maneras para completar el experimento es el producto de las entradas de *x*:

```
> prod(x)
[1] 24774195600
```

O alternativamente

```
> (11*10*9)*(7*6*5*4)*313
[1] 260290800
```

o incluso

```
> prod(factorial(c(11,7))/factorial(c(8,3)))*313
```

[1] 260290800

Como un puede suponer, en muchos de los problemas estándar de conteo no hay maneras sustanciales de ahorrar escritura; es equivalente usar *nsamp* a *factorial* y *choose*. Pero la virtud de *nsamp* se encuentra en la recopilación de las fórmulas de conteo relevantes en una sola función.

Ejemplo 1.46. El problema del cumpleaños. Supón que hay n personas juntas en un cuarto. Cada persona anuncia la fecha de sus cumpleaños. La preguntas es: ¿Cuál es la probabilidad de al menos una coincidencia? Si A es $\{\text{al menos una coincidencia}\}$, entonces se busca $P(A)$, pero como se verá más adelante, es mas conveniente calcular $P(A^c)$.

Para simplificar este problema, se ignorarán los años bisiestos y se asumirá que sólo hay 365 días en un año. Después, se asumirá que los cumpleaños están igualmente distribuidos en el curso de un año (lo cuál no es cierto debido a todo tipo de complicaciones).

Entonces, pensando en el espacio muestral, hay 365 posibilidades para el cumpleaños de la primera persona, 365 psoibilidades para el segundo, etc. El número total de secuencias de posibles cumpleaños es $\#(\Omega) = 365^n$.

Ahora, a manera de truco, se puede ver que la única situación en la cuál A no sucede es si no hay coincidencias en los cumpleaños de las personas en la habitación, es decir, sólo cuando el cumpleaños de todos es en días distintos, entonces

$$P(A) = 1 - P(A^c) = 1 - \frac{\#(A^c)}{\#(S)},$$

ya que los resultados son igualmente probables. Supóngase ahora que no hay coincidencias. La primer persona tiene 1 de 365 posibles cumpleaños. La segunda perosna no debe coincidir con la primera, por tanto sólo tiene 1 de 364 cumpleaños disponibles, y así sucesivamente hasta la enésima persona, que tiene $365 - n + 1$ posibilidades restantes para cumplir años. Por el Principio de Multiplicación, se tiene que $\#(A^c) = 365 \cdot 364 \dots (365 - n + 1)$, y

$$P(A) = 1 - \frac{365 \cdot 364 \dots (365 - n + 1)}{365^n} = 1 - \frac{364}{365} \cdot \frac{363}{365} \dots \frac{(365 - n + 1)}{365}$$

Como consecuencia, considérese esto: ¿cuántas personas son necesarias en la habitación para que la probabilidad de al menos una coincidencia sea de 0.50?. Claramente, si sólo hay $n = 1$ personas en la habitación entonces la probabilidad de coincidencia es cero, y si hay $n = 366$ personas en la habitación la probabilidad es del 100 %. Así que ¿cuántas personas se necesitan para que haya una probabilidad igual de coincidencia y no coincidencia?

La intuición podría indicar que $n = 180$. Este razonamiento se debe a que para obtener una probabilidad del 50 % de coincidencia, debería haber una ocupación del 50 % de los días del año.

Sorprendentemente, usando la ecuación anterior, se necesita solamente $n = 23$ estudiantes para tener una probabilidad mayor al 50 % de al menos una coincidencia.

En R Es posible graficar el número de personas en la habitación vs la probabilidad de al menos una coincidencia, de la siguiente manera

poner ejemplo del cumpleaños

Hay un accesorio *Birthday problem* en el menu *Probability* de *RcmdrPlugin.IPSUR*

En la versión base de R, se puede calcular aproximadamente las probabilidades para el caso mpas general de probabilidades distintas a $1/2$, para un número distinto de días en el año, e incluso para más de dos coincidencias.

Capítulo 2

Distribuciones discretas

En este capítulo se introducirán las variables aleatorias, que toman valores en un conjunto finito o infinito contable. Se discutirán las funciones de masa de probabilidad y algunos conceptos importantes, a saber, la media, la varianza y la desviación estándar. Algunas de las distribuciones discretas más importantes serán exploradas a detalle, y el concepto más general de expectación será definido, el cuál deja el camino listo para las funciones generadoras de momento.

Se prestará atención especialmente a la distribución empírica ya que juega un rol fundamental con respecto a la repetición de un muestreo; además, se introducirá un catálogo de variables aleatorias discretas que pueden ser usadas para modelar experimentos.

Hay algunos comentarios acerca de las simulaciones, y se mencionan las transformaciones de variables aleatorias en el caso discreto.

2.1. Variables Aleatorias Discretas

2.1.1. Función de Masa de Probabilidad

Las variables discretas aleatorias están caracterizadas por sus conjuntos de soporte, cuya forma es

$$S_X = \{u_1, u_2, \dots, u_k\} \text{ o } S_X = \{u_1, u_2, u_3, \dots\}.$$

Toda variable discreta aleatoria X tiene asociada consigo una función de masa de probabilidad (FMP) $f_X : S_X \rightarrow [0, 1]$ definida por

$$f_X(x) = P(X = x), \quad x \in S_X.$$

Ya que los valores de la FMP representan probabilidades, entonces cumplen algunas propiedades. En particular, todas las FMP satisfacen

1. $f_X(x) > 0$ para $x \in S$
2. $\sum_{x \in S} f_X(x) = 1$, y
3. $P(X \in A) = \sum_{x \in A} f_X(x)$, para cualquier evento $A \subset S$

Ejemplo 2.1. Al lanzar una moneda tres veces el espacio muestral es

$$S = \{AAA, ASA, SAA, SSA, AAS, ASS, SAS, SSS\}$$

Ahora, sea X el número de Soles observado. Entonces X tiene un conjunto de soporte $S_X = \{0, 1, 2, 3\}$. Asumiendo que la moneda no está cargada y que fue lanzada de la misma manera en cada ocasión, no es poco razonable suponer que los resultados en el espacio muestral son igualmente probables. ¿Cuál es la función de masa de probabilidad de X ? Nota que X vale cero cuando el resultado es AAA , y este evento tiene probabilidad $1/8$. Por lo tanto, $f_X(0) = 1/8$, y usando el mismo razonamiento se puede mostrar que $f_X(3) = 1/8$. Exactamente 3 resultados tienen como valor $X = 1$, por lo tanto, $f_X(1) = 3/8$ y los $3/8$ restantes de probabilidad se encuentran en $f_X(2)$, pues la suma es 1. Se puede representar la función de masa de probabilidad con la siguiente tabla:

$x \in S_X$	0	1	2	3	Total
$f_X(x) = P(X = x)$	1/8	3/8	3/8	1/8	1

2.1.2. Media, Varianza, y Desviación Estándar

Estos conceptos son números asociados con las funciones de masa de probabilidad. Un ejemplo importante es la media μ , también conocida como EX

$$\mu = EX = \sum_{x \in S} x f_X(x),$$

siempre que la serie $\sum |x| f_X(x)$ sea convergente. Otro número importante es la varianza:

$$\sigma^2 = E(X - \mu)^2 = \sum_{x \in S} (x - \mu)^2 f_X(x),$$

la cual puede ser calculada con la ecuación $\sigma^2 = EX - (EX)^2$. Directamente definida por la varianza surge la desviación estándar $\sigma = \sqrt{\sigma^2}$.

Ejemplo 3.2. Calcular la media de X del Ejemplo 3.1.

$$\mu = \sum_{x=0}^3 x f_X(x) = 0 \cdot \frac{1}{8} + 1 \cdot \frac{3}{8} + 2 \cdot \frac{3}{8} + 3 \cdot \frac{1}{8} = 3,5.$$

La interpretación de $\mu = 3,5$ surge del razonamiento de que si se repitiera el experimento aleatorio muchas veces, independientemente cada vez, se observarían muchos resultados correspondientes de la variable aleatoria X , y se tomaría la media muestral de las observaciones, entonces el valor calculado sería cercano a 3.5. La aproximación se acercaría más a medida que se observarían más y más valores de X (otra forma de la Ley de los Números Grandes). Otra manera de verlo es que X es 3.5 en promedio.

Es importante aclarar que X es 3.5 **en promedio**, pues, de hecho, X nunca puede ser 3.5.

Otra función importante relacionada con la función de masa de probabilidad $f_X(x) = P(X = x)$, es la función llamada función de distribución acumulada (FDA), F_X . Se define por la ecuación

$$F_X(t) = P(X \leq t), \quad -\infty < t < \infty$$

Se sabe que todas las funciones de masa de probabilidad satisfacen ciertas propiedades, y de igual manera lo hacen las funciones de distribución acumulada. En particular, cualquier función de distribución acumulada F_X cumple

- F_X es no decreciente ($t_1 \leq t_2$ implica $F_X(t_1) \leq F_X(t_2)$).
- F_X es continua por la derecha ($\lim_{t \rightarrow a+} F_X(t) = F_X(a)$ para toda $a \in \mathbb{R}$)
- $\lim_{t \rightarrow -\infty} F_X(t) = 0$ y $\lim_{t \rightarrow \infty} F_X(t) = 1$

Se puede observar que X tiene la distribución F_X y se escribe $X \sim F_X$. Abusando de la notación también se puede escribir $X \sim f_X$ y para las distribuciones conocidas, que se mostrarán a más adelante, la función de masa de probabilidad o la función de distribución acumulada serán identificadas por el nombre familiar en vez de las ecuaciones que las definen respectivamente.

En R

La media y la varianza de una variable aleatoria discreta es fácil de calcular en la consola. Regresando con el Ejemplo 3.2, se empieza definiendo un vector x que contiene el conjunto soporte de X , y un vector f que contiene los valores de f_X respectivos de los valores de x :

```
> x<-c(0,1,2,3)
> f<-c(1/8,3/8,3/8,1/8)
```

Para calcular la media μ , se necesita multiplicar los correspondientes valores de x y f y se suman. Esto se logra fácilmente en R ya que las operaciones sobre vectores son realizados elemento a elemento:

```
> mu<-sum(x*f)
> mu
```

```
[1] 1.5
```

Para calcular la varianza σ^2 , se resta el valor de μ de cada entrada de x , se eleva al cuadrado, se multiplica por f , y se suman las entradas. La desviación estándar σ es simplemente la raíz cuadrada de σ^2 .

```
> sigma2<-sum((x-mu)^2*f)
> sigma2
```



```
[1] 0.75
> sigma<-sqrt(sigma2)
> sigma
```

```
[1] 0.8660254
```

Finalmente, se pueden encontrar los valores de la función de distribución acumulativa F_X en el conjunto soporte acumulando las probabilidades en f_X con la función *cumsum*

```
> F=cumsum(f)
> F
```

```
[1] 0.125 0.500 0.875 1.000
```

Y es aún más sencillo de calcular con la paquetería *distrEx*. Se define una variable aleatoria X como un objeto, entonces se calculan los números importantes como la media, la varianza y la desviación estándar con las funciones *E*, *var* y *sd*

```
> install.packages("distrEx", repos="http://cran.us.r-project.org")
```

```
package 'distrEx' successfully unpacked and MD5 sums checked
```

```
The downloaded binary packages are in
```

```
C:\Users\meyam\AppData\Local\Temp\RtmpykbYNg\downloaded_packages
```

```
> library(distrEx)
> X<-DiscreteDistribution(supp=0:3,prob=c(1,3,3,1)/8)
> E(X); var(X); sd(X)
```

```
[1] 1.5
```

```
[1] 0.75
```

```
[1] 0.8660254
```

2.2. La Distribución Uniforme Discreta

Se han visto los bloques básicos de las distribuciones discretas y ahora es necesario estudiar modelos particulares que a menudo se encuentran en el campo de la estadística. Es posible que la más fundamental de todas sea la distribución uniforme discreta.

Una variable X con la distribución uniforme discreta en los enteros $1, 2, \dots, m$ tiene función de masa de probabilidad

$$f_X(x) = \frac{1}{m}, \quad x = 1, 2, \dots, m.$$

Se escribe $X \sim \text{disunif}(m)$. Un experimento aleatorio donde esta distribución tiene lugar es la elección de un entero entre 1 y 100. Sea X el número elegido. Entonces $X \sim \text{disunif}(m = 100)$ y

$$P(X = x) = \frac{1}{100}, \quad x = 1, \dots, 100$$

Se encuentra una ecuación directa para la media de $X \sim \text{disunif}(m)$:

$$\mu = \sum_{x=1}^m x f_X(x) = \sum_{x=1}^m x \cdot \frac{1}{m} = \frac{1}{m} (1 + 2 + \dots + m) = \frac{m+1}{2},$$

donde se ha usado la famosa identidad $1 + 2 + \dots + m = m(m+1)/2$. Es decir, si se eligen repetidamente enteros aleatoriamente entre el 1 y m entonces, en promedio, se espera obtener $(m+1)/2$. Para obtener la varianza primero se calcula

$$EX^2 = \frac{1}{m} \sum_{x=1}^m x^2 = \frac{1}{m} \frac{m(m+1)(2m+1)}{6} = \frac{(m+1)(2m+1)}{6}$$

y finalmente,

$$\sigma^2 = EX^2 - (EX)^2 = \frac{(m+1)(2m+1)}{6} - \left(\frac{m+1}{2}\right)^2 = \dots = \frac{m^2-1}{12}.$$

Ejemplo 3.4. Lanza un dado y sea X la cara que queda arriba. Entonces $m = 6, \mu = 7/2 = 3,5$, y $\sigma^2 = (6^2 - 1)/12 = 35/12$.

En R

Desde la consola se puede elegir un entero aleatoriamente con la función *sample*. La sintaxis general para simular una variable aleatoria uniforme discreta es *sample(x, size, replace = TRUE)*.

El argumento x indentifica los números de los cuales se toma la muestra aleatoriamente. Si x es un número, entonces el muestreo se hace de 1 a x . El argumento *size* indica qué tan grande debería ser el tamaño de la muestra, y *replace* indica si los números deberían o no ser reemplazados en la urna después de haber sido muestreados. La opción default es *replace = FALSE* pero para uniformes discretos los valores muestreados deberían ser reemplazados. A continuación se muestran algunos ejemplos

Ejemplos

- Lanzar un dado 3000 veces: *sample(6, size = 3000, replace = TRUE)*
- Elegir 27 números aleatoriamente entre 30 y 70: *sample(30 : 70, size = 27, replace = TRUE)*
- Lanzar una moneda 1000 veces: *sample(c("A", "S"), size = 1000, replace = TRUE)*

2.3. La Distribución Binomial

La distribución binomial se basa en un ensayo de Bernoulli, en el cuál sólo hay dos posibles resultados: éxito (S) y fracaso (F). Se realiza el experimento y sea

Si la probabilidad de éxito es p entonces la probabilidad de error debe ser $1 - p = q$ y la función de masa de probabilidad de X es

$$f_x(x) = p^x(1-p)^{1-x}, \quad x = 0, 1.$$

Es fácil calcular $\mu = EX = p$ y $EX^2 = p$ de manera que $\sigma^2 = p - p^2 = p(1-p)$.

2.3.1. El Modelo Binomial

El modelo binomial está definido por 3 propiedades:

- Los ensayos de Bernoulli se realizan n veces.
- Los ensayos son independientes
- La probabilidad de éxito p no cambia en cada intento

Si X cuenta el número de éxitos en los n ensayos independientes, entonces la función de masa de probabilidad de X es

$$f_X(x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0, 1, 2, \dots, n.$$

Se dice que X tiene una distribución binomial y se escribe $X \sim \text{binom}(size = n, prob = p)$. Es claro que $f_X(x) \geq 0$ para toda x en el conjunto soporte porque el valor es el producto de números no negativos. Después se comprueba que $\sum f(x) = 1$:

$$\sum_{x=0}^n \binom{n}{x} p^x (1-p)^{n-x} = [p + (1-p)]^n = 1^n = 1$$

Luego se halla la media:

$$\mu = \sum_{x=0}^n x \binom{n}{x} p^x (1-p)^{n-x} \tag{2.1}$$

$$= \sum_{x=1}^n x \frac{n!}{x!(n-x)!} p^x q^{n-x} \tag{2.2}$$

$$= n \cdot p \sum_{x=1}^n \frac{(n-1)!}{(x-1)!(n-x)!} p^{x-1} q^{n-x} \tag{2.3}$$

$$= np \sum_{x-1=0}^{n-1} \binom{n-1}{x-1} p^{(x-1)} (1-p)^{(n-1)-(x-1)} \tag{2.4}$$

$$= np. \tag{2.5}$$

De manera similar, se muestra que $EX(X-1) = n(n-1)p^2$. Por lo tanto

$$\sigma^2 = EX(X-1) + EX - [EX]^2 \quad (2.6)$$

$$= n(n-1)p^2 + np - (np)^2 \quad (2.7)$$

$$= n^2p^2 - np^2 + np - n^2p^2 \quad (2.8)$$

$$= np - np^2 = np(1-p) \quad (2.9)$$

Ejemplo 3.5 Hay una familia con cuatro hijos. Cada hijo puede ser un niño (H) o una niña (M). Por simplicidad supóngase que $P(M) = P(H) = 1/2$ y que los géneros de los hijos son independientes. Si X es la cuenta de los H entonces $X \sim \text{binom}(\text{size} = 4, \text{prob} = 1/2)$. Además, $P(X = 2)$ es

$$f_X(2) = \binom{4}{2}(1/2)^2(1/2)^2 = \frac{6}{2^4}$$

El número promedio de niños es $4(1/2) = 2$ y la varianza de X es $4(1/2)(1/2) = 1$

En R

Las funciones correspondientes en R para la función de masa de probabilidad y la función de distribución acumulada son *dbinom* y *pbinom*, respectivamente. A continuación se muestra su uso con ejemplos.

Ejemplo 3.6. Se lanzan 12 dados simultáneamente. Sea X la cantidad de seises que aparecen. Se desea encontrar la probabilidad de encontrar 7, 8, o 9 seises. Sea $S = \{\text{obtener un 6 en una lanzada}\}$, entonces $P(S) = 1/6$ y las lanzadas constituyen ensayos de Bernoulli; por lo tanto $X \sim \text{binom}(\text{size} = 12, \text{prob} = 1/6)$ y la tarea es encontrar $P(7 \leq X \leq 9)$ lo cuál es

$$P(7 \leq X \leq 9) = \sum_{x=7}^9 \binom{12}{x} (1/6)^x (5/6)^{12-x}$$

De nuevo, un método para resolver este problema sería generar una tabla de masa de probabilidad y agregar las columnas relevantes. Sin embargo, un método alternativo es notar que $P(7 \leq X \leq 9) = P(X \leq 9) - P(X \leq 6) = F_X(9) - F_X(6)$

```
> pbinom(9, size=12, prob=1/6)-pbinom(6,size=12,prob=1/6)
```

```
[1] 0.001291758
```

```
> diff(pbinom(c(6,9),size=12,prob=1/6))
```

```
[1] 0.001291758
```

Como se puede observar, arrojan el mismo resultado.

Ejemplo 3.7. Al lanzar una moneda 3 veces X es el número de Águilas observadas. Para este experimento $X \sim \text{binom}(\text{size} = 3, \text{prob} = 1/2)$ a partir de lo cuál se obtiene la siguiente tabla de masa de probabilidad:

$x = \# \text{de Águilas}$	0	1	2	3
$f(x) = P(X = x)$	1/8	3/8	1/8	1/8

El siguiente objetivo es escribir a la función de distribución acumulada de X de manera explícita. El primer caso es sencillo: es imposible que X sea negativa, por lo que si $X < 0$ entonces se obtendría $P(X \leq x) = 0$. Ahora se elige un valor x que satisfaga $0 \leq x \leq 1$, por ejemplo, $x = 0.3$. La única manera de que $X \leq x$ pudiera pasar sería si $X = 0$, por lo tanto, $P(X \leq x)$ debería ser $P(X = 0)$ y lo mismo aplica para cualquier $0 \leq x \leq 1$. Similarmente, para cualquier $1 \leq x \leq 2$, por ejemplo, $x = 1.73$, el evento $\{X \leq x\}$ es exactamente el evento $\{X = 0 \text{ o } X = 1\}$. Consecuentemente, $P(X \leq x)$ debería ser $P(X = 0 \text{ o } X = 1) = P(X = 0) + P(X = 1)$. Continuando por el mismo camino, se pueden averiguar los valores de $F_X(x)$ para todas las posibles entradas $-\infty < x < \infty$, y resumir dichas observaciones con la siguiente función definida a trozos:

funcion a trozos

En particular, la función de distribución acumulada de X está definida para la recta real \mathbb{R} . La función de distribución acumulada es continua por la derecha y no decreciente. Una gráfica de la función $\text{binom}(\text{size} = 3, \text{prob} = 1/2)$ se muestra en la Figura 3.1.

Ejemplo 3.8. Otra manera de hacer el Ejemplo 3.7 es con la familia de paqueterías *distr*. Usan un acercamiento orientado a objetos con variables aleatorias, es decir, una variable aleatoria se almacena en un objeto X , y entonces las preguntas acerca de la variable aleatoria se traducen a funciones que involucran a X . Las variables aleatorias con distribuciones del paquete *base* se especifican escribiendo la letra inicial del nombre de la distribución en mayúscula.

```
> library(distr)
> X<-Binom(size=3,prob=1/2)
> X
```

```
Distribution Object of Class: Binom
size: 3
prob: 0.5
```

Analógico a la función *dbinom* para X es la función $d(X)$, y a *pbinom* es la función $p(X)$. Comparando:

```
> d(X)(1)
[1] 0.375
> p(X)(2)
[1] 0.875
```

Las variables aleatorias definidas con el paquete *distr* pueden generar gráficas de las funciones de masa de probabilidad y de distribución acumulada.
meter graficas

2.4. Funciones Genradoras de Momentos

2.4.1. El Operador de Expectación

Anteriormente se vio que cada función de masa de probabilidad (cuya serie converge) tiene dos números importantes asociados:

$$\mu = \sum_{x \in S} x f_X(x), \quad \sigma^2 = \sum_{x \in S} (x - \mu)^2 f_X(x)$$

Intuitivamente, para observaciones repetidas de X se esperaría que la media de la muestra fuera aproximadamente igual a μ a medida que el tamaño de la muestra crece sin frontera. Por esta razón se llama a μ el **valor esperado** de X y se escribe $\mu = EX$, donde E es el operador de expectación.

Definición. Más generalmente, dada una función g se defina el valor esperado de $g(X)$ mediante

$$Eg(X) = \sum_{x \in S} g(X) f_X(x)$$

siempre que $\sum_x |g(x)| f(x)$ sea convergente. Se dice que $Eg(x)$ existe.

En esta notación la varianza es $\sigma^2 = E(X - \mu)^2$ y se puede deducir la identidad

$$E(X - \mu)^2 = EX^2 - (EX)^2$$

Intuitivamente, para observaciones repetidas de X se esperaría que la media muestral de los valores de $g(X)$ se aproximara a $Eg(X)$ a medida que el tamaño de la muestra incremente sin frontera.

Más aún, si se espera que el valor de $g(X)$ sea cercano a $Eg(X)$ en promedio, ¿dónde se esperaría encontrar el promedio de $3g(X)$? Sólo podría ser en $3Eg(X)$. El siguiente teorema hace la idea más precisa.

Proposición. Para cualesquiera g y h funciones, cualquier variable aleatoria X , y cualquier constante c :

- 1. $Ec = c$
- 2. $E[c \cdot g(X)] = cEg(X)$
- 3. $E[g(X) + h(X)] = Eg(X) + Eh(X)$

siempre que $Eg(X)$ y $Eh(X)$ existan.

Desmotración. Surge directamente de la definición. Por ejemplo,

$$E[c \cdot g(X)] = \sum_{x \in S} c \cdot g(x) f_X(x) = c \cdot \sum_{x \in S} g(x) f_X(x) = cEg(X)$$

2.4.2. Funciones Generadoras de Momentos

Definición. Dada una variable aleatoria X , su función generadora de momento se define como

$$M_X(t) = Ee^{tX} = \sum_{x \in S} f_X(x)$$

siempre que la serie sea convergente para todo t en una vecindad de 0 (es decir, para todo $-\epsilon < t < \epsilon$, para algún $\epsilon > 0$).

Hay que notar que para cualquier función generadora de momento M_X ,

$$M_X(0) = Ee^{0 \cdot X} = E1 = 1$$

Ahora se calculará la función generadora de momentos para las dos distribuciones introducidas anteriormente.

Ejemplo 3.12. Encontrar la función generadora de momentos para $X \sim \text{disunif}(m)$

Ya que $f(x) = 1/m$, la FGM toma la forma de

$$M(t) = \sum_{x=1}^m e^{tx} \frac{1}{m} = \frac{1}{m}(e^t + e^{2t} + \dots + e^{mt}), \text{ para cualquier } t$$

Ejemplo 3.13. Encontrar la FGM para $X \sim \text{binom}(\text{size} = n, \text{prob} = p)$

$$M_X(t) = \sum_{x=0}^n e^{tx} \binom{n}{x} p^x (1-p)^{n-x} \quad (2.10)$$

$$= \sum_{x=0}^{n-x} \binom{n}{x} (pe^t)^x q^{n-x} \quad (2.11)$$

$$= (pe^t + q)^n, \text{ para cualquier } t \quad (2.12)$$

Aplicaciones.

A continuación se discutirán 3 aplicaciones de las funciones generadoras de momento. La primera es el hecho de que una función generadora de momento puede ser usada para identificar de manera precisa la distribución de probabilidad que la generó, que se sostiene en el siguiente teorema:

Teorema. La función generadora de momento, si existe en una vecindad de cero, determina una distribución de probabilidad de manera única.

Ejemplo 3.15. Una variable aleatoria tiene una función generadora de momento

$$M_X(t) = (0,3 + 0,7e^t)^{13}$$

Entonces $X \sim \text{binom}(\text{size} = 13, \text{prob} = 0,7)$.

Una función generadora de momento también se conoce como una Transformada de Laplace.

¿Por qué se le llama Función Generadora de Momento?

Esta pregunta está relacionada con la segunda aplicación poderosa de las funciones generadoras de momentos. Muchos de los modelos que se estudian tienen una función generadora simple, que permite determinar la media, la varianza e incluso momentos más altos muy rápido. A continuación el porqué. Ya se sabe que

$$M(t) = \sum_{x \in S} e^{tx} f(x)$$

Derivando con respecto a t se obtiene

$$M'(t) = \frac{d}{dt} \sum_{x \in S} e^{tx} f(x) = \sum_{x \in S} \frac{d}{dt} (e^{tx} f(x)) = \sum_{x \in S} x e^{tx} f(x)$$

evaluando en $t = 0$

$$M'(0) = \sum_{x \in S} x e^0 f(x) = \sum_{x \in S} x f(x) = \mu = EX$$

De manera similar, $M''(t) = \sum x^2 e^{tx} f(x)$ por lo tanto $M''(0) = EX^2$, y en general, se puede ver que

$$M_X^r(0) = EX^r = \text{momento } r\text{-ésimo de } X \text{ al rededor del origen.}$$

Estos también son conocidos como momentos crudos y son denotados por μ'_r . Adicionalmente, también existen los momentos centrales μ_r definidos por

$$\mu_r = E(X - \mu)^r, \quad r = 1, 2, \dots$$

Ejemplo 3.16. Sea $X \sim \text{binom}(\text{size} = n, \text{prob} = p)$ con $M(t) = (q + pe^t)^n$. Anteriormente se calculó la media y la varianza de una variable aleatoria binomial por medio de las series binomiales. Pero es interesante ver qué fácil es encontrar la media y la varianza con la función generadora de momento.

$$M'(t) = n(q + pe^t)^{n-1}pe^t|_{t=0} \quad (2.13)$$

$$= n \cdot 1^{n-1}p \quad (2.14)$$

$$= np \quad (2.15)$$

Y

$$M''(0) = n(n-1)[q + pe^t]^{n-2}(pe^t)^2 + n[q + pe^t]^{n-1}pe^t|_{t=0} \quad (2.16)$$

$$EX^2 = n(n-1)p^2 + np \quad (2.17)$$

Por lo tanto

$$\sigma^2 = EX^2 - (EX)^2 \quad (2.18)$$

$$= n(n-1)p^2 + np - n^2p^2 \quad (2.19)$$

$$= np - np^2 = npq \quad (2.20)$$

que resulta mucho más fácil de calcular.

Como observación, y recordando que ciertas funciones pueden ser representadas por medio de expansiones en series de Taylor al rededor de un punto a , de la forma

$$f(x) = \sum_{r=0}^{\infty} \frac{f^{(r)}(a)}{r!} (x-a)^r, \text{ para todo } |x-a| < R$$

donde R es el radio de convergencia de la serie. Combinando ambos resultados, se dice que si una función generadora de momentos existe para todo t en el intervalo $(-\epsilon, \epsilon)$ se puede reescribir como

$$M_X(t) = \sum_{r=0}^{\infty} \frac{EX^r}{r!} t^r, \text{ para toda } |t| < \epsilon$$

En R

El paquete *distrEx* provee un operador de expectación que puede ser usado en variables aleatorias que han sido definidas por medio de *distr*:

```
> X<-Binom(size=3,prob=0.45)
> library(distrEx)
> E(X)
[1] 1.35
> E(3*X+4)
[1] 8.05
```

Para variables aleatorias discretas con conjunto soporte finito, la expectación es calculada mediante la suma directa. En el caso de que la variable aleatoria tenga conjunto soporte infinito, entonces la expectación no se calcula directamente con la suma, sino que se estima primero generando una muestra aleatoria del modelo y después se calcula una media muestral de la función de interés.

Hay métodos para otros parámetros de la población:

```
> var(X)
[1] 0.7425
> sd(X)
[1] 0.8616844
```

2.5. La Distribución Empírica

Haz un experimento n veces y observa los n valores x_1, x_2, \dots, x_n de una variable aleatoria X . Por simplicidad en la mayoría de casos será conveniente imaginar que los valores observados son distintos, pero la aplicación funciona incluso si los valores observados se repiten.

Definición. La función de distribución acumulada empírica F_n es la distribución de probabilidad que sitúa la masa de probabilidad $1/n$ en cada uno de los valores x_1, x_2, \dots, x_n . Dicha función toma la forma

$$f_X(x) = \frac{1}{n}, \quad x \in \{x_1, x_2, \dots, x_n\}$$

Si el valor x_i se repite k veces, la masa en x_i se acumula a k/n .

La media de la distribución empírica es

$$\mu = \sum_{x \in S} x f_X(x) = \sum_{i=1}^n x_i \cdot \frac{1}{n}$$

que resulta ser la media muestral \bar{x} . La varianza de la distribución empírica es

$$\sigma^2 = \sum_{x \in S} (x - \mu)^2 f_X(x) = \sum_{i=1}^n (x_i - \bar{x})^2 \cdot \frac{1}{n}$$

y esta última cantidad luce muy similar a una ya conocida, a saber, la varianza muestral.

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

En R

La distribución empírica no está disponible directamente como lo está la base de las otras distribuciones de probabilidad, sin embargo hay varias fuentes disponibles.

Dado un vector de datos con valores x , se puede ver la función de distribución acumulada empírica por medio de la función `ecdf`:

```
> x<-c(4,7,9,11,12)
> ecdf(x)
```

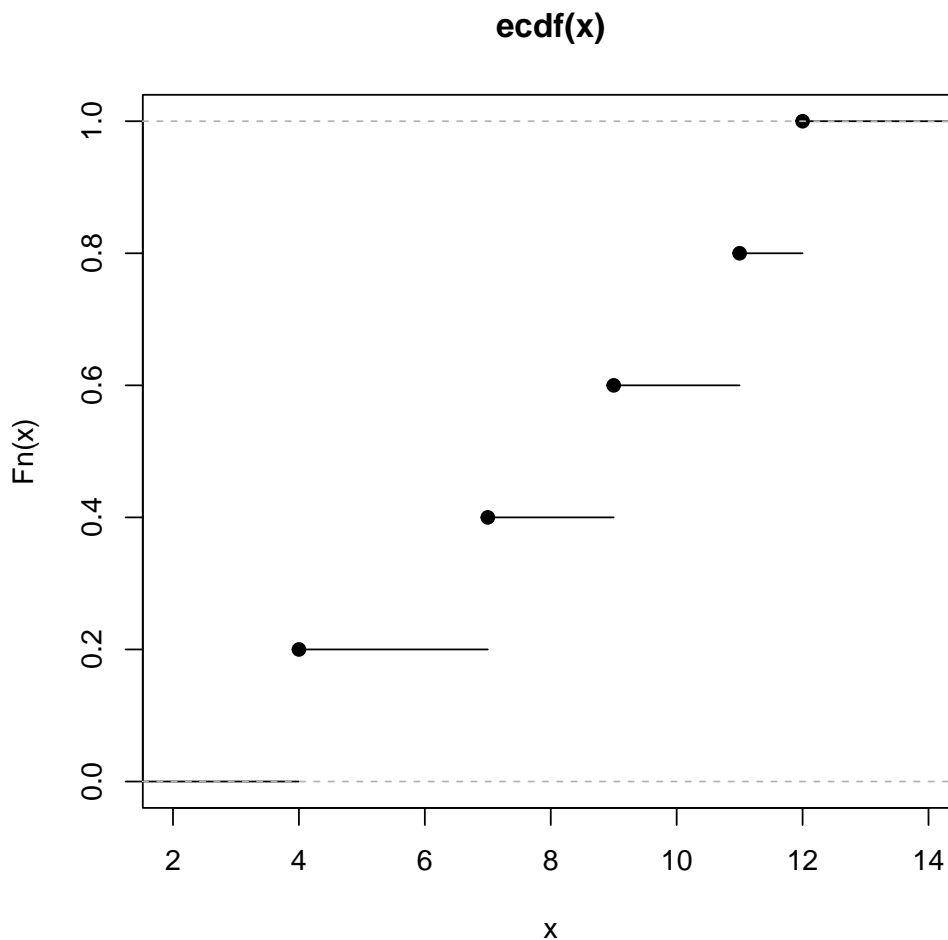
Empirical CDF

Call: `ecdf(x)`

```
x[1:5] =      4,      7,      9,     11,     12
```

Lo anterior muestra que el valor retornado de `ecdf(x)` no es un número sino una función. La función `ecdf` no es usualmente usada por sí misma en esta forma. Más comúnmente se usa como un paso intermedio en un cálculo más complicado, por ejemplo, en la prueba de hipótesis, o en el remuestreo. Sin embargo, es instructivo ver cómo se ve la función `ecdf`, y hay un método especial para graficar objetos `ecdf`.

```
> plot(ecdf(x))
```



La gráfica es una función continua por la derecha con saltos exactamente en los puntos x . No hay valores repetidos de x por lo tanto todos los saltos son iguales a $1/5$.

La función de distribución de probabilidad empírica usualmente no es de interés por sí misma, pero si se quisiera se podría definir una función que sirviera como la función de distribución de probabilidad empírica:

```
> epdf<-function(x) function(t){sum(x%in%t)/length(x)}
> x<-c(0,0,1)
> epdf(x)(0)

[1] 0.6666667
```

Para simular de la distribución empírica soportada en el vector x , se usa la función *sample*.

```
> x<-c(0,0,1)
> sample(x,size=7,replace=TRUE)

[1] 0 1 1 0 1 1 0

>
```

Como se mencionó anteriormente, la distribución empírica es significativa en mayor medida debido a cómo y dónde aparece en más aplicaciones sofisticadas. Se explorarán algunas de estas en capítulos más adelante.

2.6. Otras Distribuciones Discretas

Las distribuciones uniforme discreta y binomial son populares, simples y forman las bases para muchas otras distribuciones más complicadas. Pero los modelos particulares binomial y uniforme sólo son aplicables a un rango limitado de problemas. En esta sección se introducen situaciones para las cuales se necesitan más de lo que las distribuciones binomial y uniforme ofrecen.

2.6.1. Ensayos de Bernoulli Dependientes

La distribución Hypergeométrica

Hay una urna con 7 bolas blancas y 5 bolas negras. El experimento aleatorio consiste en elegir 4 de las bolas de manera aleatoria, sin remplazo. Entonces la probabilidad de observar 3 bolas blancas (y por tanto 1 negra) sería

$$P(3B, 1N) = \frac{\binom{7}{3}\binom{5}{1}}{\binom{12}{4}}$$

Más generalmente, se muestrea sin remplazo K veces de una urna con M bolas blancas y N bolas negras. Sea X el número de bolas blancas en la urna. La función de masa de probabilidad de X es

$$f_X(x) = \frac{\binom{M}{x}\binom{N}{K-x}}{\binom{M+N}{K}}$$

Se dice que X tiene distribución hypergeométrica y se escribe $X \sim \text{hyper}(m = M, n = N, k = K)$.

El conjunto soporte para la distribución hypergeométrica es un poco confuso. Es tentador decir que x debería ir desde 0 (sin bolas blancas en la muestra) hasta K (sin bolas negras en la muestra), pero no funciona si $K > M$, porque es imposible tener más bolas blancas en la muestra de las que había originalmente en la urna. Se tiene el mismo problema si $K > N$. Lo bueno es que la mayoría de ejemplos que se estudia tiene $K \leq M$ y $K \leq N$ y se tomará el conjunto soporte como $x = 0, 1, \dots, K$.

Se puede demostrar que la media y la varianza son

$$\mu = K \frac{M}{M+N}, \sigma^2 = K \frac{MN}{(M+N)^2} \frac{M+N-K}{M+N-1}$$

Las funciones asociadas en R para la función de masa de probabilidad y la función de distribución acumulada son *dhyper*(x, m, n, k) y *phyper*, respectivamente. Hay dos funciones más: *qhyper* y *rhyp*, que se explicará a continuación.

Ejemplo 3.19. En un cargamento de 250 procesadores hay 17 defectuosos. Un consultor de control de calidad elige aleatoriamente 5 procesadores para inspeccionarlos y determinar si son o no defectuosos. Sea X el número de procesadores defectuosos en la muestra.

1. Encontrar la probabilidad de que el consultor elija 3 procesadores defectuosos de la muestra, es decir, hallar $P(X = 3)$.

Solución: Se sabe que $X \sim \text{hyper}(m = 17, n = 233, k = 5)$ Entonces la probabilidad será

$$f_X(3) = \frac{\binom{17}{3}\binom{233}{2}}{\binom{250}{5}}$$

en R simplemente es

```
> dhyper(3,m=17,n=233,k=5)
```

```
[1] 0.002351153
```

2. Encontrar la probabilidad de que máximo 2 procesadores elegidos sean defectuosos, es decir, $P(X \leq 2)$.

```
> phyper(2,m=17,n=233,k=5)
```

```
[1] 0.9975771
```

Por lo tanto $P(X \leq 2) \approx 0.9975771$.

3. Encontrar $P(X > 1)$.

```
> phyper(1,m=17,n=233,k=5,lower.tail = FALSE)
```

```
[1] 0.03863065
```

Muestreo con y sin Reemplazo

Supóngase que se tiene una urna con M bolas blancas y N bolas negras. Se toma una muestra de tamaño n de la urna, y sea X el contador del número de bolas blancas en la muestra. Si se muestrea sin reemplazo, entonces $X \sim \text{hyper}(m = M, n = N, k = n)$ con media y varianza

$$\mu = n \frac{M}{M+N} \quad (2.21)$$

$$\sigma^2 = n \frac{MN}{(M+N)^2} \frac{M+N-n}{M+N-1} \quad (2.22)$$

$$= n \frac{M}{M+N} \left(1 - \frac{M}{M+N}\right) \frac{M+N-n}{M+n-1} \quad (2.23)$$

Por otra parte, si se muestrea con reemplazo, entonces $X \sim \text{binom}(\text{size} = n, \text{prob} = M/(M+N))$ con media y varianza

$$\mu = n \frac{M}{M+N} \quad (2.24)$$

$$\sigma^2 = n \frac{M}{M+N} \left(1 - \frac{M}{M+N}\right) \quad (2.25)$$

Se puede ver que ambos procedimientos de muestreo tienen la misma media, y el método con la varianza más grande es con reemplazo. El factor por el cuál difieren las varianzas,

$$\frac{M+N-n}{M+N-1}$$

se llama corrección de población finita. Para un tamaño de muestra fijo n , mientras $M, N \rightarrow \infty$ es claro que la corrección tiende a 1, es decir, para poblaciones de tamaño infinito los esquemas de muestreo son esencialmente iguales para la varianza y la media.

2.6.2. Distribución de la variable Tiempo de Espera

Otra clase importante de problemas está asociada con la cantidad de tiempo que le toma ocurrir a un evento de interés. Por ejemplo, se podría lanzar una moneda repetidamente hasta que se observara un Sol. Se podría lanzar una bola de papel hasta atinarle a un bote de basura.

La Distribución Geométrica

Se conducen ensayos de Bernoulli repetidamente, notando los éxitos y los fracasos. Sea X el número de fracasos antes de un éxito. Si $P(E) = p$ entonces X tiene una función de masa de probabilidad

$$f_X(x) = p(1-p)^x, \quad x = 0, 1, 2, \dots$$

¿Por qué? Se dice que X tiene una distribución geométrica y se escribe $X \sim \text{geom}(\text{prob} = p)$. Las funciones asociadas en R son *dgeom*(x, prob), *pgeom*, *qgeom*, y *rhyper*, las cuales calculan la función de masa de probabilidad, la función de distribución acumulada, la función cuantil, y simulan variables aleatorias, respectivamente.

De nuevo, es claro que $f(x) \geq 0$ y se puede encontrar que $\sum f(X) = 1$

$$\sum_{x=0}^{\infty} p(1-p)^x = p \sum_{x=0}^{\infty} q^x = p \frac{1}{1-q} = 1$$

Se encontrará en la siguiente sección que la media y la varianza son

$$\mu = \frac{1-p}{p} = \frac{q}{p}, \quad \sigma^2 = \frac{q}{p^2}$$

Ejemplo 3.20. Un jugador de futbol americano, consiguió anotar el 81.2% de los goles de campo que intentó. Asumiendo que sus intentos de gol de campo sucesivos son aproximadamente ensayos de Bernoulli, encontrar la probabilidad de que el jugador falle al menos 5 goles de campo antes de su primer éxito.

Solución: Si X es el número de goles fallados antes de su primer gol anotado, entonces $X \sim \text{geom}(\text{prob} = 0.812)$ y se busca $P(X \geq 5) = P(X > 4)$. En R se puede calcular mediante

```
> pgeom(4, prob=0.812, lower.tail=FALSE)
```

```
[1] 0.0002348493
```

Nota: Algunos libros usan una definición un poco distinta de la distribución geométrica. Consideran los ensayos de Bernoulli y hacen Y el contador del número de intentos hasta el éxito, de tal manera que Y tiene función de masa de probabilidad

$$f_Y(y) = p(1-p)^{y-1}, \quad y = 1, 2, 3, \dots$$

Cuando dicen "distribución geométrica.^{es} a esta a la que se refieren. No es difícil ver que las dos definiciones están relacionadas. De hecho, si X denota la distribución geométrica establecida en este texto, y Y la establecida en algún otro, entonces $Y = X + 1$. Consecuentemente, ellos obtienen $\mu_Y = \mu_X + 1$ y $\sigma_Y^2 = \sigma_X^2$.

La Distribución Binomial Negativa

Se puede generalizar el problema y considerar el caso donde se espera más de un éxito. Suponéngase que se conducen ensayos de Bernoulli repetidamente, notando los respectivos éxitos y fracasos. Sea X el número de fracasos antes de r éxitos. Si $P(E) = p$ entonces X tiene función de masa de probabilidad

$$f_X(x) = \binom{r+x-1}{r-1} p^r (1-p)^x, \quad x = 0, 1, 2, \dots$$

Se dice que X tiene una distribución binomial negativa y se escribe $X \sim \text{nbinom}(size = r, prob = p)$. Las funciones asociadas en R son `dbinom(x, size, prob)`, `pnbinom`, `qnbinom`, `rnbinom`, las cuales calculan la función de probabilidad, la función cuantil, y simula variables aleatorias, respectivamente.

Como siempre, debe ser claro que $f_X(x) \geq 0$ y el hecho de que $\sum f_X(x) = 1$ se deriva de una generalización de las series geométricas por medio de la expansión en series de Maclaurin:

$$\frac{1}{1-t} = \sum_{k=0}^{\infty} t^k, \quad \text{para } -1 < t < 1, \quad \text{y} \quad (2.26)$$

$$\frac{1}{(1-t)^r} = \sum_{k=0}^{\infty} \binom{r+k-1}{r-1} t^k \quad \text{para } -1 < t < 1 \quad (2.27)$$

Por lo tanto

$$\sum_{x=0}^{\infty} f_X(x) = p^r \sum_{x=0}^{\infty} \binom{r+x-1}{r-1} q^x = p^r (1-q)^{-r} = 1$$

ya que $|q| = |1-p| < 1$.

Ejemplo 3.22. Se lanza una moneda repetidamente y sea X el contador del número de Soles hasta que aparecen siete Águilas. ¿Cuál es $P(X = 5)$?

Solución: Se sabe que $X \sim \text{nbinom}(size = 7, prob = 1/2)$.

$$P(X = 5) = f_X(5) = \binom{7+5-1}{7-1} (1/2)^7 (1/2)^5 = \binom{11}{6} 2^{-12}$$

Y se puede calcular con R con

```
> dbinom(5, size=7, prob=0.5)
```

```
[1] 0.1640625
```

Ahora se calculará la función generadora de momentos de $X \sim \text{nbinom}(size = r, prob = p)$:

$$M_X(t) = \sum_{x=0}^{\infty} e^{tx} \binom{r+x-1}{r-1} p^r q^x \quad (2.28)$$

$$= p^r \sum_{x=0}^{\infty} \binom{r+x-1}{r-1} [qe^t]^x \quad (2.29)$$

$$= p^r (1 - qe^t)^{-r}, \quad \text{siempre que } |qe^t| < 1 \quad (2.30)$$

de manera que

$$M_X(t) = \left(\frac{p}{1-qe^t}\right)^r, \quad \text{para } qe^t < 1$$

Se puede ver que $qe^t < 1$ cuando $t < -\ln(1-p)$.

Sea $X \sim \text{nbinom}(size = r, prob = p)$ con $M(t) = p^r (1 - qe^t)^{-r}$. Se establecieron anteriormente los valores de la media y la varianza. Ahora se pueden encontrar directamente con las herramientas mostradas:

$$M'(t) = p^r(-r)(1 - qe^t)^{-r-1}(-qe^t) \quad (2.31)$$

$$= r q e^t p^r (1 - qe^t)^{-r-1} \quad (2.32)$$

$$= \frac{r q e^t}{1 - qe^t} M(t), \text{ por tanto} \quad (2.33)$$

$$M'(0) = \frac{r q}{1 - q} \cdot 1 = \frac{r q}{p} \quad (2.34)$$

Por lo tanto $\mu = r q / p$. Después se encuentra EX^2 :

$$M''(0) = \frac{r q e^t (1 - qe^t) - r q e^t (-qe^t)}{(1 - qe^t)^2} M(t) + \frac{r q e^t}{1 - qe^t} M'(t)|_{t=0} \quad (2.35)$$

$$= \frac{r q p + r q^2}{p^2} \cdot 1 + \frac{r q}{p} \left(\frac{r q}{p} \right) \quad (2.36)$$

$$= \frac{r q}{p^2} + \left(\frac{r q}{p} \right)^2 \quad (2.37)$$

Finalmente se dice que $\sigma^2 = M''(0) - [M'(0)]^2 = r q / p^2$.

Ejemplo 3.23. Una variable aleatoria tiene función generadora de momentos

$$M_X(t) = \left(\frac{0.19}{1 - 0.81e^t} \right)^{31}$$

Entonces $X \sim \text{nbinom}(size = 31, prob = 0.19)$.

Es importante notar que como con una distribución geométrica, algunos libros usan una definición un poco distinta de la distribución Binomial Negativa. Consideran los ensayos de Bernoulli, con Y el número de ensayos hasta r éxitos, de tal manera que Y tiene función de masa de probabilidad

$$f_Y(y) = \binom{y-1}{r-1} p^r (1-p)^{y-r}, \quad y = r, r+1, r+2, \dots$$

De nuevo, no es difícil ver que si X denota nuestro Binomio Negativo y Y el suyo, entonces $Y = X + r$. Consecuentemente, tienen $\mu_Y = \mu_X + r$ y $\sigma_Y^2 = \sigma_X^2$.

2.6.3. Procesos de Llegada

La Distribución de Poisson

Esta es una distribución asociada con "eventos raros", por razones que serán claras a continuación. Los eventos pueden ser:

- Accidentes de tráfico
- Errores de transcripción
- Clientes llegando a un banco

Sea λ el número promedio de eventos en un intervalo de tiempo $[0, 1]$. Sea X la variable aleatoria que cuenta el número de eventos ocurriendo en ese intervalo. Entonces, bajo ciertas condiciones razonables se puede mostrar que

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

Se usa la notación $X \sim \text{pois}(\lambda)$. Las funciones asociadas en R son `dpois(x, lambda)`, `ppois`, `qpois`, y `rpois` las cuales indican la función de masa de probabilidad, la función de distribución acumulada, la función cuantil, y una simulación de variables aleatorias, respectivamente.

¿Cuáles son las condiciones razonables? Divide al intervalo $[0, 1]$ en intervalos de longitud $1/n$. Un proceso de Poisson satisface las siguientes condiciones:

- La probabilidad de que ocurra un evento en un subintervalo particular es λ/n .
- La probabilidad de que ocurran 2 o más eventos en cualquier intervalo es 0.
- Los eventos que suceden en subintervalos disjuntos son independientes.

Si X cuenta el número de eventos en el intervalo $[0, t]$ y λ es el número promedio que ocurre en el tiempo unitario, entonces $X \sim \text{pois}(\lambda t)$, es decir,

$$P(X = x) = e^{-\lambda t} \frac{(\lambda t)^x}{x!}, \quad x = 0, 1, 2, 3, \dots$$

Ejemplo 3.25. En promedio, 5 carros llegan a un autolavado cada hora. Sea X el contador del número de carros que llegan de las 10 AM a las 11 AM. Entonces $X \sim \text{pois}(\text{lambda} = 5)$. Además, $\mu = \sigma^2 = 5$. ¿Cuál es la probabilidad de que ningún carro llegue al autolavado durante este periodo?

Solución: La probabilidad de que ningún carro llegue es

$$P(X = 0) = e^{-5} \frac{5^0}{0!} \approx 0,0067$$

Ejemplo 3.26. Supóngase que el autolavado abre a las 8AM y cierra a las 6PM, y sea Y el número de clientes que llegan en este periodo. Ya que este periodo cubre un total de 10 horas, $Y \sim \text{pois}(\text{lambda} = 5 * 10 = 50)$. ¿Cuál es la probabilidad de que lleguen entre 48 y 50 clientes?

Solución: Se busca $P(49 \leq Y \leq 50) = P(X \leq 50) - P(X \leq 47)$

```
> diff(ppois(c(47,50),lambda=50))
```

```
[1] 0.1678485
```

2.7. Funciones de Variables Aleatorias Discretas

Se ha cosntruído un gran catálogo de distribuciones discretas, pero las herramientas de esta sección proporcionarán la habilidad para considerar muchas más. Dada una variable aleatoria X y dada una función h , se puede considerar $Y = h(X)$. Ya que los valores de X se determinan por probabilidad, también los de Y . La pregunta es, ¿cuál es la función de masa de probabilidad de la variable aleatoria Y ? La respuesta claramente depende de h . En el cado de que h sea uno a uno, la solución puede ser encontrada por sustitución.

Ejemplo 3.27. Sea $X \sim \text{nbinom}(\text{size} = r, \text{prob} = p)$. En la sección 5.6 se vió que X representa el número de fracasos hasta alcanzar r éxitos en una secuencia de ensayos de Bernoulli. Suponiendo ahora que el interés está en contar el número de ensayos (éxitos y fracasos) hasta encontrar el r^{vo} éxito, que será denotado por Y . En dada realización del experimento, el número de fracasos (X) y el número de éxitos (r), juntos, comprenderán el número total de intentos (Y), es decir, $X + r = Y$. Se puede definir h como $h(x) = x + r$ de modo que $Y = h(X)$, y se puede notar que h es lineal, y por lo tanto uno a uno. Finalmente, X toma los valores $0, 1, 2, \dots$ lo que implica que el conjunto soporte de Y sería $\{r, r+1, r+2, \dots\}$. Resolviendo para X se obtiene $X = Y - r$. Examinando la función de masa de probabilidad de X

$$f_X(x) = \binom{r+x-1}{r-1} p^r (1-p)^x$$

sustituyendo $x = y - r$ se obtiene

$$f_Y(y) = f_X(y - r) \quad (2.38)$$

$$= \binom{r + (y - r) - 1}{r - 1} p^r (1-p)^{y-r} \quad (2.39)$$

$$= \binom{y-1}{r-1} p^r (1-p)^{y-r}, \quad y = r, r+1, \dots \quad (2.40)$$

Incluso cuando la función h no es uno a uno, aún se puede encontrar la función de masa de probabilidad de Y simplemente acumulando para cada y la probabilidad de todas las x que son mapeadas a esa y .

Proposición. Sea X una variable aleatoria discreta con función de masa de probabilidad f_x con conjunto soporte S_X . Sea $Y = h(X)$ para alguna función h . Entonces Y tiene función de probabilidad f_Y definida por

$$f_Y(y) = \sum_{\{x \in S_X | h(x)=y\}} f_X(x)$$

Ejemplo 3.28. Sea $X \sim \text{binom}(\text{size} = 4, \text{prob} = 1/2)$, y sea $Y = (X - 1)^2$. Considera la siguiente tabla:

x	0	1	2	3	4
$f_X(x)$	1/16	1/4	6/16	1/4	1/16
$y = (x - 2)^2$	1	0	1	4	9

De esta tabla se puede ver que el conjunto soporte S_Y de Y es $S_Y = \{0, 1, 4, 9\}$. También se puede ver que $h(x) = (x - 1)^2$ no es uno a uno en el conjunto soporte de X , porque tanto $x = 0$ como $x = 2$ estan mapeados por h a $y = 1$. Sin embargo, se puede ver que $Y = 0$ soo cuando $X = 1$, lo cuál tiene una probabilidad de 1/4; por lo tanto, $f_Y(0)$ debería ser igual a 1/4. Un acercamiento similar funciona para $y = 4$ y $y = 9$. Y $Y = 1$

y	0	1	4	9
$f_X(x)$	1/4	7/16	1/4	1/16

exactamente cuando $X = 0$ o $X = 2$, cuya probabilidad total es $7/16$. En resumen, la función de masa de probabilidad puede ser escrita como:

Hay que notar que no hay un nombre especial para la distribución de Y , solo es un ejemplo de lo que se hace cuando la transformación de una variable aleatoria no es uno a uno. El método es el mismo para problemas más complicados.

Proposición. Si X es una variable aleatoria con $EX = \mu$ y $Var(X) = \sigma^2$, entonces la media y la varianza de $Y = mX + b$ es

$$\mu_Y = m\mu + b, \sigma_Y^2 = m^2\sigma^2, \sigma_Y = |m|\sigma$$

Capítulo 3

Distribuciones Continuas

El enfoque en el capítulo anterior fue respecto a las variables aleatorias cuyo conjunto soporte puede ser escrito en una lista de valores (finitos o infinitos contables), tales como el número de éxitos en una secuencia de ensayos de Bernoulli. Ahora es momento de pasar a las variables cuyo conjunto soporte consta de un rango de valores, a saber, un intervalo (a, b) . Dado que la cantidad de estos números es infinita, es imposible anotar cada uno de ellos en una lista.

Este capítulo inicia con las variables aleatorias continuas y las funciones de densidad de probabilidad y funciones de distribución acumuladas asociadas. La distribución uniforme continua se remarca, junto con la distribución Gaussiana o normal. Algunos detalles matemáticos pavimentan el camino para un catálogo de modelos.

En este capítulo se aprenderá:

- Cómo elegir un modelo continuo razonable bajo una variedad de circunstancias físicas
- Correspondencia básica entre variables continuas contra discretas
- Las herramientas generales para la manipulación de variables aleatorias continuas, integración, etc.
- Algunos detalles sobre un par de modelos continuos, y una exposición de varios otros
- Cómo hacer variables aleatorias continuas nuevas partiendo de antiguas

3.1. Variables Aleatorias Continuas

3.1.1. Funciones de Densidad de Probabilidad

Las variables aleatorias continuas tienen conjuntos de soporte que se ven como

$$S_X = [a, b] \text{ o } (a, b)$$

o uniones de intervalos de la forma anterior. Algunos ejemplos de variables aleatorias que a menudo se consideran continuas son:

- El peso o estatura de un individuo
- Otras medidas físicas tales como la longitud o el tamaño de un objeto, y
- Duración de tiempo (usualmente)

Toda variable aleatoria continua X tiene una función de densidad de probabilidad f_X asociada que satisface 3 propiedades básicas:

1. $f_X(x) > 0$ para $x \in S_X$,
2. $\int_{x \in S_X} f_X(x) dx = 1$, y
3. $P(X \in A) = \int_{x \in A} f_X(x) dx$, para un evento $A \subset S_X$.

Además, se puede decir lo siguiente acerca de las variables aleatorias continuas:

- Usualmente, el conjunto A en 3 toma la forma de un intervalo, por ejemplo, $A = [c, d]$ en cuyo caso

$$P(X \in A) = \int_c^d f_X(x) dx$$

- La probabilidad de que X caiga en un intervalo dado es simplemente el área bajo la curva de f_X sobre el intervalo.
- Ya que el área de una línea $x = c$ en el plano es cero, $P(X = c) = 0$ para cualquier valor c . En otras palabras, la probabilidad de que X sea igual a algún valor particular c es cero, y esto es cierto para cualquier número c . Más aún, cuando $a < b$ todas las siguientes probabilidades son las mismas:

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

- La función de distribución de probabilidad f_X a veces puede ser mayor que 1. Esto es un contraste con el caso discreto; cada valor distinto de cero de una función de masa de probabilidad es una probabilidad que está restringida a encontrarse en el intervalo $[0, 1]$.

En el capítulo 5 se introdujo la función de distribución acumulada F_X . Recordando, esta se define por $F_X(t) = P(X \leq t)$, para $-\infty < t < \infty$. Mientras que en el caso discreto, dicha función es inmanejable, en el caso continuo tiene una forma relativamente conveniente:

$$F_X(t) = P(X \leq t) = \int_{-\infty}^t f_X(x)dx, \quad -\infty < t < \infty$$

También, para cualquier función de distribución acumulada F_X lo siguiente es verdadero

- F_X es no decreciente, es decir, $t_1 < t_2$ implica $F_X(t_1) < F_X(t_2)$.
- F_X es continua. Hay que notar la distinción con el caso discreto: las funciones de distribución acumuladas de las variables aleatorias discretas no son continuas, sólo son continuas por la derecha.
- $\lim_{t \rightarrow -\infty} F_X(t) = 0$ y $\lim_{t \rightarrow \infty} F_X(t) = 1$.

Hay una relación útil entre la función de distribución acumulada y la función de distribución de probabilidad en el caso continuo. Considérese la derivada de F_X :

$$F'_X(t) = \frac{d}{dt} F_X(t) = \frac{d}{dt} \int_{-\infty}^t f_X(x)dx = f_X(t),$$

con la última igualdad cierta por el Teorema Fundamental del Cálculo. En resumen, $(F_X)' = f_X$ en el caso continuo.

3.1.2. Expectación de Variables Aleatorias Continuas

Para una variable aleatoria continua X el valor esperado de $g(X)$ es

$$Eg(X) = \int_{x \in S} g(x)f_X(x)dx,$$

siempre que la integral $\int_S |g(x)|f(x)dx$ converja. Un ejemplo importante es la media μ , también conocida como EX :

$$\mu = EX = \int_{x \in S} xf_X(x)dx,$$

siempre que $\int_S |x|f(x)dx$ sea finita. También está la varianza

$$\sigma^2 = E(X - \mu)^2 = \int_{x \in S} (x - \mu)^2 f_X(x)dx,$$

la cual puede ser calculada con la ecuación alternativa $\sigma^2 = EX^2 - (EX)^2$. Adicionalmente, está la desviación estándar $\sigma = \sqrt{\sigma^2}$. La función generadora de momento está dada por

$$M_X(t) = Ee^{tX} = \int_{-\infty}^{\infty} e^{tx} f_X(x)dx,$$

siempre que la integral exista para todo t en una vecindad de $t = 0$.

Ejemplo 4.1. Sea X una variable aleatoria continua con función de distribución de probabilidad

$$f_X(x) = 3x^2, \quad 0 \leq x \leq 1.$$

Más adelante se verá que f_X pertenece a la familia de distribuciones Beta. Es fácil ver que $\int_{-\infty}^{\infty} f(x)dx = 1$

$$\int_{-\infty}^{\infty} f_X(x)dx = \int_0^1 3x^2 dx \quad (3.1)$$

$$= x^3 \Big|_{x=0}^1 \quad (3.2)$$

$$= 1^3 - 0^3 \quad (3.3)$$

$$= 1 \quad (3.4)$$

Dicho lo anterior, se puede encontrar $P(0,14 \leq X \leq 0,71)$.

$$P(0,14 \leq X \leq 0,71) = \int_{0,14}^{0,71} 3x^2 dx \quad (3.5)$$

$$= x^3 \Big|_{x=0,14}^{0,71} \quad (3.6)$$

$$= 0,71^3 - 0,14^3 \quad (3.7)$$

$$\approx 0,355167 \quad (3.8)$$

De manera similar se puede calcular la media y la varianza

$$\mu = \int_{-\infty}^{\infty} x f_X(x) dx = \int_0^1 x \cdot 3x^2 dx \quad (3.9)$$

$$= \frac{3}{4} x^4 \Big|_{x=0}^1 \quad (3.10)$$

$$= \frac{3}{4} \quad (3.11)$$

Tal vez lo mejor sería calcular la varianza con la ecuación $\sigma^2 = EX^2 - \mu^2$

$$EX^2 = \int_{-\infty}^{\infty} x^2 f_X(x) dx = \int_0^1 x^2 3x^2 dx \quad (3.12)$$

$$= \frac{3}{5} x^5 \Big|_{x=0}^1 \quad (3.13)$$

$$= \frac{3}{5} \quad (3.14)$$

por lo tanto $\sigma^2 = 3/5 - (3/4)^2 = 3/80$.

Ejemplo 4.2. El siguiente ejemplo presenta un conjunto soporte sin frontera para trabajar con integración impropia. Sea X la variable aleatoria con función de distribución de probabilidad

$$f_X(x) = \frac{3}{x^4}, \quad x > 1$$

Se puede mostrar que $\int_{-\infty}^{\infty} f(x)dx = 1$:

$$\int_{-\infty}^{\infty} f_X(x)dx = \int_1^{\infty} \frac{3}{x^4} dx \quad (3.15)$$

$$= \lim_{t \rightarrow \infty} \int_1^t \frac{3}{x^4} dx \quad (3.16)$$

$$= \lim_{t \rightarrow \infty} 3 \frac{1}{-3} x^{-3} \Big|_{x=1}^t \quad (3.17)$$

$$= - \left(\lim_{t \rightarrow \infty} \frac{1}{t^3} - 1 \right) \quad (3.18)$$

$$= 1 \quad (3.19)$$

Calculando $P(3,4 \leq X \leq 7,1)$:

$$P(3,4 \leq X \leq 7,1) = \int_{3,4}^{7,1} 3x^{-4} dx \quad (3.20)$$

$$= 3 \frac{1}{-3} x^{-3} \Big|_{x=3,4}^{7,1} \quad (3.21)$$

$$= -1(7,1^{-3} - 3,4^{-3}) \quad (3.22)$$

$$\approx 0,0226487123 \quad (3.23)$$

De nuevo, se calcula la media y la varianza.

$$\mu = \int_{-\infty}^{\infty} x f_X(x) dx = \int_1^{\infty} x \cdot \frac{3}{x^4} dx \quad (3.24)$$

$$= 3 \frac{1}{-2} x^{-2} \Big|_{x=1}^{\infty} \quad (3.25)$$

$$= -\frac{3}{2} \lim_{t \rightarrow \infty} \frac{1}{t^2} - 1 \quad (3.26)$$

$$= \frac{3}{2} \quad (3.27)$$

Y usando la ecuación $\sigma^2 = EX^2 - \mu^2$:

$$EX^2 = \int_{-\infty}^{\infty} x^2 f_X(x) dx = \int_1^{\infty} x^2 \cdot \frac{3}{x^4} dx \quad (3.28)$$

$$= 3 \frac{1}{-1} x^{-1} \Big|_{x=1}^{\infty} \quad (3.29)$$

$$= -3 \left(\lim_{t \rightarrow \infty} \frac{1}{t^2} - 1 \right) \quad (3.30)$$

$$= 3, \quad (3.31)$$

resultando entonces $\sigma^2 = 3 - (3/2)^2 = 3/4$.

En R

Existen utilidades para calcular probabilidades y expectativas para variables aleatorias continuas generales, pero es más conveniente encontrar un modelo construido, de ser posible. A veces no es posible. Se muestra cómo hacerlo por el camino largo, y cómo hacerlo usando la paquetería *distr*.

Ejemplo 4.3. Sea X la variable con función de distribución de probabilidad $f(x) = 3x^2$ para $0 < x < 1$. Hallar $P(0,14 \leq X \leq 0,71)$.

```
> f<-function(x) 3*x^2
> integrate(f, lower=0.14, upper=0.71)
```

```
0.355167 with absolute error < 3.9e-15
```

Se puede integrar la función $xf(x) = 3 * x^3$ desde cero hasta uno para obtener la media, y usar la ecuación $\sigma^2 = EX^2 - (EX)^2$ para la varianza.

Ejemplo 4.4. Sea X la variable con función de distribución de probabilidad $f(x) = 3/x^4, x > 1$. Se puede integrar la función $xf(x) = 3/x^3$ desde cero hasta infinito para obtener la media de X .

```
> g<-function(x) 3/x^3
> integrate(g, lower=1, upper=Inf)
```

```
1.5 with absolute error < 1.7e-14
```

Ejemplo 4.5. Rehaciendo el Ejemplo 4.1 con la paquetería *distr*. El método es similar al mostrado en el capítulo 5. Se define una variable absolutamente aleatoria continua:

```
> library(distr)
> f<-function(x) 3*x^2
> X<-AbscontDistribution(d=f, low1=0,up1=1)
> p(X)(0.71)-p(X)(0.14)

[1] 0.355167
```

Ahora usando la paquetería para los valores de expectación *distrEx*:

```
> library(distrEx)
> E(X)

[1] 0.7496337

> var(X)

[1] 0.03768305

> 3/80

[1] 0.0375
```

Comparando estas respuestas con las encontradas en el Ejemplo 4.1. ¿Por qué son diferentes? Porque la paquetería *distrEx* recurre a métodos numéricos cuando se encuentra con modelos que no reconoce. Esto significa que las respuestas obtenidas a partir de los cálculos pueden no ser iguales a los valores teóricos.

3.2. La Distribución Uniforme Continua

Una variable aleatoria X con la distribución uniforme continua en el intervalo (a, b) tiene función de distribución de probabilidad

$$f_X(x) = \frac{1}{b-a}, \quad a < x < b$$

La función asociada en R es *dunif*(*min* = *a*, *max* = *b*). Se escribe $X \sim \text{unif}(\text{min} = a, \text{max} = b)$. Debido a la forma particularmente simple de esta función de distribución se puede escribir explícitamente una fórmula para la función de distribución acumulada F_X :

$$F_X(t) = \begin{cases} 0, & \text{si } t < a \\ \frac{t-a}{b-a}, & \text{si } a \leq t \leq b \\ 1, & \text{si } t \geq b. \end{cases}$$

La distribución uniforme continua es la analogía continua de la distribución uniforme discreta; es usada para modelar experimentos cuyo resultado es un intervalo de números que son igualmente probables en el sentido de que cualquier par de intervalos de igual longitud en el conjunto soporte tienen la misma probabilidad asociada a ellos.

Ejemplo 4.6. Elige un número entre $[0, 1]$ de manera aleatoria, y sea X el número elegido. Entonces $X \sim \text{unif}(\text{min} = 0, \text{max} = 1)$.

La media de $X \sim \text{unif}(\text{min} = a, \text{max} = b)$ es relativamente simple de calcular:

$$\mu = EX = \int_{-\infty}^{\infty} x f_X(x) dx \quad (3.32)$$

$$= \int_a^b x \frac{1}{b-a} dx, \quad (3.33)$$

$$= \frac{1}{b-a} \frac{x^2}{2} \Big|_{x=a}^b \quad (3.34)$$

$$= \frac{1}{b-a} \frac{b^2 - a^2}{2} \quad (3.35)$$

$$= \frac{b+a}{2} \quad (3.36)$$

3.3. La Distribución Normal

Se dice que X tiene una distribución normal si tiene función de distribución de probabilidad de la forma

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \quad -\infty < x < \infty$$

Se escribe $X \sim \text{norm}(\text{mean} = \mu, \text{sd} = \sigma)$, y la función asociada en R es $\text{dnorm}(x, \text{mean} = 0, \text{sd} = 1)$

La curva con forma de campana, o distribución normal es también conocida como la distribución Gaussiana. Esta distribución es por mucho la más importante de las distribuciones, continuas o discretas. El modelo normal aparece en la teoría de todo tipo de fenómenos, desde la manera en que las partículas de humo se disipan, hasta el viaje que realiza una botella en el océano, o el ruido blanco de la radiación cósmica de fondo.

Cuando $\mu = 0$ y $\sigma = 1$ se dice que la variable aleatoria tiene una distribución normal estándar y se escribe típicamente $Z \sim \text{norm}(\text{mean} = 0, \text{sd} = 1)$. La letra minúscula ϕ se usa para denotar la función de distribución de probabilidad normal estándar y la mayúscula Φ para la función de distribución acumulada: para $-\infty < z < \infty$

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \text{ y } \int_{-\infty}^t \phi(z) dz$$

Proposición. Si $X \sim \text{norm}(\text{mean} = \mu, \text{sd} = 1)$ entonces

$$Z = \frac{X-\mu}{\sigma} \sim \text{norm}(\text{mean} = 0, \text{sd} = 1)$$

La función generadora de momento de $Z \sim \text{norm}(\text{mean} = 0, \text{sd} = 1)$ es relativamente fácil de deducir:

$$M_Z(t) = \int_{-\infty}^{\infty} e^{tz} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz \quad (3.37)$$

$$= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}(z^2 + 2tz + t^2) + \frac{t^2}{2}\right\} dz \quad (3.38)$$

$$= e^{t^2/2} \left(\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-[z-(-t)]^2/2} dz \right) \quad (3.39)$$

y la cantidad en paréntesis es el área total bajo la densidad $\text{norm}(\text{mean} = -t, \text{sd} = 1)$ que es uno. Por lo tanto,

$$M_Z(t) = e^{-t^2/2}, \quad -\infty < t < \infty$$

Ejemplo 4.7 La función generadora de momento de $X \sim \text{norm}(\text{mean} = \mu, \text{sd} = \sigma)$ no es difícil ya sea porque

$$Z = \frac{X-\mu}{\sigma} \text{ o reescribiendo, } X = \sigma Z + \mu$$

Por lo tanto:

$$M_X(t) = Ee^{tX} = Ee^{t(\sigma Z + \mu)} = Ee^{\sigma t Z} e^{t\mu} = e^{t\mu} M_Z(\sigma t),$$

Y se sabe que $M_Z(t) = e^{-t^2/2}$, por lo tanto al sustituir se obtiene

$$M_X(t) = e^{t\mu} e^{(\sigma t)^2/2} = \exp\{\mu t + \sigma^2 t^2/2\}$$

para $-\infty < t < \infty$

El mismo argumento anterior muestra que si X tiene una función generadora de momento $M_X(t)$ entonces la función generadora de $Y = a + bX$ es

$$M_Y(t) = e^{ta} M_X(bt)$$

Ejemplo 4.8. La regla 69-95-99.7. Cuando una distribución empírica es aproximadamente una curva de campana hay proporciones específicas de las observaciones que se sitúan en distancias variables de la media. Se puede ver de donde salen estas y obtener proporciones más precisas con lo siguiente:

```
> pnorm(1:3)-pnorm(-(1:3))
```

```
[1] 0.6826895 0.9544997 0.9973002
```

Ejemplo 4.9. Un experimento aleatorio consiste en una persona tomando un test de coeficiente intelectual, y sea X el puntaje obtenido en el test. Los puntajes en estos tests están típicamente estandarizados para tener una media de 100 y una desviación estándar de 15. ¿Cuál es la probabilidad $P(85 \leq X \leq 115)$

Solución: los límites 85 y 115 se sitúan exactamente una desviación estándar abajo y arriba respectivamente de la media 100. Por lo tanto la respuesta es aproximadamente 68 %.

3.3.1. Cuantiles Normales y la Función Cuantil

Hasta ahora se han dado dos valores y la tarea ha sido encontrar el área bajo la función de distribución de probabilidad entre ambos valores. En esta sección, se va de reversa: dada un área, se busca encontrar los valores extremos de esa área.

Ejemplo 4.10 Asumiendo que el modelo de IQ del ejemplo anterior, ¿cuál es el IQ más bajo posible que una persona puede tener y estar en el 1 % superior del puntaje de IQ?

Solución: Si una persona está en el 1 % superior, entonces eso significa que el 99 % de la gente tiene un IQ más bajo. Entonces, en otras palabras, se busca un valor x tal que $F(x) = P(X \leq x)$ satisface $F(x) = 0,99$, o incluso otra manera de decirlo es que se busca resolver la ecuación $F(x) - 0,99 = 0$. A continuación se mostrará cómo hacerlo de la manera larga. Se define la función $g(x) = F(x) - 0,99$, y luego se busca la raíz de g con la función *uniroot*. Usa procedimientos numéricos para encontrar la raíz entonces se necesita dar un intervalo de x valores en el cuál busque la raíz. La raíz debería estar en algún lugar entre dos y tres desviaciones estándar por encima de la media (que es 100).

```
> g<-function(x) pnorm(x,mean=100,sd=15)-0.99
> uniroot(g,interval = c(130,145))

$root
[1] 134.8952

$f.root
[1] -4.873083e-09

$iter
[1] 6

$init.it
[1] NA

$estim.prec
[1] 6.103516e-05
```

La respuesta se muestra en *\$root*, que es aproximadamente 134.8952, que significa que una persona con este IQ o una mayor se encuentra en el 1 % superior de todos los puntajes de IQ.

La discusión en el ejemplo anterior se centró en la búsqueda de un valor x que resolviera una ecuación $F(x) = p$, para alguna probabilidad p dada, o dicho de otra manera, se buscó F^{-1} , la función inversa de la función de distribución acumulada de X , evaluada en p . Esto es tan importante que requiere una definición propia.

Definición. La función cuantil de una variable aleatoria X es la inversa de su función de distribución acumulada:

$$Q_X(p) = \min\{x : F(x) \geq p\}, \quad 0 < p < 1.$$

A continuación algunas propiedades de las funciones cuantiles:

1. La función cuantil está definida y es finita para todo $0 < p < 1$.
2. Q_X es continua por la izquierda. Para variables aleatorias discretas es una función a trozos, y para variables aleatorias continuas es una función continua.
3. En el caso continuo la gráfica de Q_X puede ser obtenida reflejando la gráfica de F_X con respecto a la línea $y = x$. En el caso discreto, antes de reflejar es necesario conectar los puntos para deshacerse de los saltos, borrar las líneas horizontales de manera que solo quede líneas verticales, y finalmente intercambiar los círculos abiertos con los puntos sólidos.
4. Los dos límites

$$\lim_{p \rightarrow 0^+} Q_X(p) \text{ y } \lim_{p \rightarrow 1^-} Q_X(p)$$

siempre existe, pero puede ser infinito (es decir, a veces $\lim_{p \rightarrow 0} Q(p) = -\infty$ y/o $\lim_{p \rightarrow 1} Q(p) = \infty$)

Como era de esperarse, la distribución normal estándar es un caso muy especial y tiene su propia notación especial.

Definición. Para $0 < \alpha < 1$, el símbolo z_α denota la única solución de la ecuación $P(Z > z_\alpha) = \alpha$, donde $Z \sim \text{norm}(\text{mean} = 0, \text{sd} = 1)$. Puede ser calculada en una de dos maneras equivalentes: *qnorm*($1 - \alpha$) y *qnorm*($\alpha, \text{lower.tail} = \text{FALSE}$).

En R Las funciones cuantiles están definidas para todas las distribuciones básicas con el prefijo q en el nombre de la distribución, excepto para la función de distribución acumulada empírica cuya función cuantil es exactamente la función $Q_X(p) = \text{quantile}(x, probs = p, type = 1)$.

Ejemplo 4.11. Retomando el ejemplo donde se busca $Q_X(0.99)$, donde $X \sim \text{norm}(mean = 100, sd = 15)$. Es muy sencillo realizarlo con R

```
> qnorm(0.99, mean=100, sd=15)
```

```
[1] 134.8952
```

que es igual al obtenido con la función *uniroot*.

Ejemplo 4.12. Encontrar los valores $z_{0.025}, z_{0.01}$

```
> qnorm(c(0.025, 0.01, 0.005), lower.tail=FALSE)
```

```
[1] 1.959964 2.326348 2.575829
```

3.4. Funciones de Variables Aleatorias Continuas

El objetivo de esta sección es determinar la distribución de $U(x) = g(X)$ basada en la distribución de X . En el caso discreto todo lo que se necesitaba era sustituir $x = g^{-1}(u)$ en la función de masa de probabilidad de X (a veces acumulando masa de probabilidad a lo largo de procedimiento). Sin embargo, en el caso continuo se necesitan procedimientos más sofisticados.

3.4.1. El Método de Función de Distribución de Probabilidad

Proposición. Sea X la variable con función de distribución de probabilidad f_X y sea g una función que es uno a uno con inversa diferenciable g^{-1} . Entonces la función de distribución de probabilidad de $U = g(X)$ está dada por

$$f_U(u) = f_X[g^{-1}(u)] \left| \frac{d}{du} g^{-1}(u) \right|$$

La ecuación anterior es matemáticamente correcta pero no tiene un sentido claro. Es mejor reescribirla en una manera más intuitiva

$$f_U(u) = f_X(x) \left| \frac{dx}{du} \right|$$

Ejemplo 4.13. Se tiene $X \sim \text{norm}(mean = \mu, sd = \sigma)$, y $Y = e^X$. ¿Cuál es la función de distribución de Y ?

Solución: Hay que notar primero que $e^x > 0$ para cualquier x , por tanto el conjunto soporte de Y es $(0, \infty)$. Ya que la transformación es monótona, se puede resolver $y = e^x$ para x para obtener $x = \ln y$, resultando $dx/dy = 1/y$. Por lo tanto, para cualquier $y > 0$

$$f_Y(y) = f_X(\ln y) \cdot \left| \frac{1}{y} \right| = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(\ln y - \mu)^2}{2\sigma^2}\right\} \cdot \frac{1}{y}$$

donde se ha prescindido del valor absoluto ya que $y > 0$. Se dice que la variable aleatoria Y tiene una distribución logarítmica normal.

Ejemplo 4.14. Se tiene $X \sim \text{norm}(mean = 0, sd = 1)$, y $Y = 4 - 3X$. ¿Cuál es la función de distribución de probabilidad de Y ?

El conjunto soporte de X es $(-\infty, \infty)$, y mientras x va desde $-\infty$ hasta ∞ , la cantidad $y = 4 - 3x$ también recorre el mismo intervalo. Resolviendo para x en la ecuación $y = 4 - 3x$ se obtiene $x = -(y - 4)/3$ resultando $dx/dy = -1/3$. Y como

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \quad -\infty < x < \infty$$

se tiene

$$f_Y(y) = f_X\left(\frac{y-4}{3}\right) \cdot \left| -\frac{1}{3} \right|, \quad -\infty < y < \infty \quad (3.40)$$

$$= \frac{1}{3\sqrt{2\pi}} e^{-(y-4)^2/2 \cdot 3^2}, \quad -\infty < y < \infty \quad (3.41)$$

Se tiene entonces que la función de distribución de probabilidad de Y tiene la forma $norm(mean = 4, sd = 3)$. De hecho, se puede usar un argumento idéntico para demostrar el siguiente hecho:

Si $X \sim norm(mean = \mu, sd = \sigma)$ y si $Y = a + bX$ para a, b constantes, con $b \neq 0$, entonces $Y \sim norm(mean = a + b\mu, sd = |b|\sigma)$.

Es importante notar que a veces es más fácil posponer la resolución de la transformación inversa $x = x(u)$. En vez de eso, se deja la transformación en la forma $u = u(x)$ y se calcula la derivada de la transformación original

$$du/dx = g'(x)$$

Una vez que se conoce esta, se puede obtener la función de distribución de probabilidad de U con

$$f_U(u) = f_X(x) \left| \frac{1}{du/dx} \right|$$

En muchos casos hay cancelaciones y el trabajo es menor. Por supuesto, no siempre es verdadero que

$$\frac{dx}{du} = \frac{1}{du/dx}$$

pero en el ejemplo anterior, esta ecuación funciona.

En el caso cuando g no es monótona no se puede aplicar la anterior proposición directamente. En vez de eso, se descompone el conjunto soporte de X en trozos de manera que g sea monótona en cada uno. Se aplica la proposición en cada trozo, y se suman los resultados.

3.4.2. El método de Función de Distribución Acumulada

Se sabe que $f_X = F'_X$ en el caso continuo. Partiendo de la ecuación $F_Y(y) = P(Y \leq y)$, se puede sustituir $g(X)$ por Y , y luego resolver para X para obtener $P[X \leq g^{-1}(y)]$, que es otra manera de escribir $F_X[g^{-1}(y)]$. Derivar esta última cantidad con respecto a y resultará en la función de distribución de probabilidad de Y .

Ejemplo 4.15. Suponiendo $X \sim unif(min = 0, max = 1)$ y $Y = -\ln X$. ¿Cuál es la función de distribución de probabilidad de Y ?

El conjunto soporte de X es $(0, 1)$, y y recorre $(0, \infty)$ mientras x va de 0 a 1, por lo que el conjunto soporte de Y es $S_Y = (0, \infty)$. Para cualquier $y > 0$ se considera

$$F_Y(y) = P(Y \leq y) = P(-\ln X \leq y) = P(X \geq e^{-y}) = 1 - P(X < e^{-y})$$

Donde las últimas igualdades surgen porque la función exponencial es monótona. Además, dado que X es continua, las dos probabilidades $P(X < e^{-y})$ y $P(X \leq e^{-y})$ son iguales, por lo tanto

$$1 - P(X < e^{-y}) = 1 - P(X \leq e^{-y}) = 1 - F_X(e^{-y})$$

Ahora recordando que la función de distribución acumulada de una variable aleatoria $unif(min = 0, max = 1)$ satisface $F(u) = u$ se puede decir que

$$F_Y(y) = 1 - F_X(e^{-y}) = 1 - e^{-y}, \text{ para } y > 0$$

Por tanto se ha encontrado la ecuación para la función de distribución acumulada de Y ; para obtener la función de distribución de probabilidad f_Y solo es necesario derivar F_Y :

$$f_Y(y) = \frac{d}{dy}(1 - e^{-y}) = 0 - e^{-y}(-1)$$

o $f_Y(y) = e^{-y}$ para $y > 0$. Esto resulta ser un miembro de la familia de distribuciones exponenciales.

Ejemplo 4.17. La Transformación Integral de Probabilidad. Dada una variable aleatoria continua X con función de distribución acumulada F_X estrictamente creciente, y Y una variable aleatoria definida por $Y = F_X(X)$. Entonces la distribución de Y es $unif(min = 0, max = 1)$.

Demostración. Se emplea el método de función de distribución acumulada. Primero hay que notar que el conjunto soporte de Y es $(0, 1)$. Entonces para cualquier $0 < y < 1$,

$$F_Y(y) = P(Y \leq y) = P(F_X(X) \leq y)$$

Ahora, ya que F_X es estrictamente creciente, tiene una inversa bien definida F_X^{-1} . Por lo tanto,

$$P(F_X(X) \leq y) = P(X \leq F_X^{-1}(y)) = F_X[F_X^{-1}(y)] = y.$$

En resumen, se ha visto que $F_Y(y) = y$, $0 < y < 1$. Pero esto es exactamente la función de distribución acumulada de una variable aleatoria *unif*($min = 0, max = 1$).

La Transformación Integral de Probabilidad es verdadera para todas las variables continuas con función de distribución acumulada continua, no solo para aquellas con función de distribución acumulada estrictamente creciente. La transformación no es verdadera para las variables aleatorias discretas, o para variables aleatorias continuas con algún componente discreto (es decir, con saltos en su función de distribución acumulada).

Ejemplo 4.18. Se tiene $Z \sim norm(mean = 0, sd = 1)$ y $U = Z^2$. ¿Cuál es la función de distribución de probabilidad de U ?

Primero hay que notar que $Z^2 \geq 0$, y por lo tanto el conjunto soporte de U es $[0, \infty)$. Y para cualquier $u \geq 0$,

$$F_U(u) = P(U \leq u) = P(Z^2 \leq u)$$

Pero $Z^2 \leq u$ ocurre si y solo si $-\sqrt{u} \leq Z \leq \sqrt{u}$. La última probabilidad anterior es simplemente el área bajo la función de distribución de probabilidad normal estándar de $-\sqrt{u}$ a \sqrt{u} , y ya que ϕ es simétrica al rededor de 0, se tiene que

$$P(Z^2 \leq u) = 2P(0 \leq Z \leq \sqrt{u}) = 2[F_Z(\sqrt{u}) - F_Z(0)] = 2\Phi(\sqrt{u}) - 1$$

ya que $\Phi(0) = 1/2$. Para encontrar la función de distribución de probabilidad de U se deriva la función de distribución acumulada recordando que $\Phi' = \phi$,

$$f_U(u) = (2\Phi(\sqrt{u}) - 1)' = 2\phi(\sqrt{u}) \cdot \frac{1}{2\sqrt{u}} = u^{1/2}\phi(\sqrt{u})$$

Sustituyendo,

$$f_U(u) = u^{-1/2} \frac{1}{\sqrt{2\pi}} e^{-(\sqrt{u})^2/2} = (2\pi u)^{-1/2} e^{-u}, \quad u > 0$$

Este resultado más adelante se verá y se llama distribución chi cuadrada con 1 grado de libertad.

En R

La paquetería *distr* tiene funcionalidad para investigar transformaciones de distribuciones univariadas. Hay resultados exactos para transformaciones ordinarias de las distribuciones estándar, y *distr* saca ventaja de estos en muchos casos. Por ejemplo, la paquetería *distr* puede manejar la transformación en el Ejemplo 4.12 de manera bastante útil:

```
> library(distr)
> X<-Norm(mean=0,sd=1)
> Y<-4-3*X
> Y
```

```
Distribution Object of Class: Norm
mean: 4
sd: 3
```

Por lo tanto *distr* sabe que una transformación lineal de una variable aleatoria normal es de nuevo normal, e incluso sabe cuales son la media y la desviación estándar correctas. Pero es imposible para *distr* saber todo, y no tarda en encontrarse con tales situaciones, por ejemplo, con la función e^x

```
> Y<-exp(X)
> Y
```

```
Distribution Object of Class: AbscontDistribution
```

El resultado es un objeto de la clase *AbscontDistribution*, la cuál es una de las clases que *distr* usa para denotar distribuciones generales que no reconoce (resulta que Z tiene una distribución logarítmica normal). Una descripción simplificada del proceso que sigue *distr* cuando encuentra una transformación $Y = g(X)$ que no reconoce es:

1. Aleatoriamente generar muchas copias X_1, X_2, \dots, X_n de la distribución de X ,
2. Calcular $Y_1 = g(X_1), Y_2 = g(X_2), \dots, Y_n = g(X_n)$ y almacenarlos para su uso posterior,
3. Calcular la función de distribución de probabilidad, la función de distribución acumulada, cuantiles, y generar variables aleatorias usando los valores simulados de Y .

Mientras la transformación sea suficientemente buena, tal como una transformación lineal, la exponencial, el valor absoluto, etc, las funciones d-p-q son calculadas analíticamente basadas en las funciones d-p-q asociadas a X . Pero si se intenta una transformación más compleja entonces R lanza una advertencia:


```
> W<-sin(exp(X)+27)
> W
```

Distribution Object of Class: AbscontDistribution

La advertencia confirma que las funciones d-p-q no son calculadas analíticamente, sino que se basan en los valores simulados aleatoriamente de Y . La naturaleza de la simulación aleatoria significa que se pueden obtener diferentes respuestas para la misma pregunta, por ejemplo, al calcular $P(W \leq 0,5)$ usando la W del ejemplo anterior, después definiendo W de nuevo, y calculando la supuesta misma $P(W \leq 0,5)$.

```
> p(W)(0.5)

[1] 0.5793242
```

```
> W<-sin(exp(X)+27)
> p(W)(0.5)

[1] 0.5793242
```

Sorprendentemente las respuestas no son las mismas. Más aún, si se repitiera el proceso se obtendría una respuesta más para $P(W \leq 0,5)$.

Pero las respuestas fueron cercanas. Y las X generadas aleatoriamente no fueron las mismas por lo que no debería ser una sorpresa que las W calculadas no fueran las mismas. Esto funciona como una advertencia para recordar que las transformaciones complicadas calculadas por R son solo aproximadas y pueden fluctuar ligeramente debido a la naturaleza de la manera en que son calculadas.

3.5. Otras Distribuciones Continuas

3.5.1. Distribuciones de Espera de Tiempo

En algunos experimentos, la variable aleatoria medida es el tiempo hasta el que cierto evento ocurre. Por ejemplo, un especialista de control de calidad puede estar probando un producto manufacturado para ver cuánto tiempo tarda en fallar. Un experto en eficiencia puede estar grabando el tráfico de clientes en una tienda para asignar horarios más eficientes al personal.

La Distribución Exponencial

Se dice que X tiene una distribución exponencial y se escribe $X \sim \exp(\text{rate} = \lambda)$.

$$f_X(x) = \lambda e^{-\lambda x}, \quad x > 0$$

Las funciones asociadas son $dexp(x, \text{rate} = 1)$, $pexp$, $qexp$ y $rexp$, las cuales calculan la función de distribución de probabilidad, la función de distribución acumulada, la función cuantil, y genera variables aleatorias, respectivamente.

El parámetro λ mide la tasa de llegadas y debe ser positiva. La función de distribución acumulada está dada por la ecuación

$$F_X(t) = 1 - e^{-\lambda t}, \quad t > 0$$

La media es $\mu = 1/\lambda$ y la varianza es $\sigma^2 = 1/\lambda^2$.

La distribución exponencial está cercanamente relacionada a la distribución de Poisson. Si los clientes llegan a una tienda de acuerdo con un proceso de Poisson con tasa λ y si Y cuenta el número de clientes que llegan en el intervalo de tiempo $[0, t)$, entonces $Y \sim \text{pois}(\text{lambda} = \lambda t)$. Ahora considerando otra cuestión: un reloj empieza en 0 y se detiene en el momento en que llega el primer cliente. Sea X la longitud de este intervalo aleatorio. Entonces $X \sim \exp(\text{rate} = \lambda)$. Obsérvese la siguiente cadena de igualdades

$$P(X > t) = P(\text{primera llegada después del tiempo } t) \quad (3.42)$$

$$= P(\text{ningún evento en } [0, t)), \quad (3.43)$$

$$= P(Y = 0) \quad (3.44)$$

$$= e^{-\lambda t} \quad (3.45)$$

donde la última línea es la función de masa de probabilidad de Y evaluada en $y = 0$. En otras palabras, $P(X \leq t) = 1 - e^{-\lambda t}$, que es exactamente la función de distribución acumulada de una distribución $\exp(\text{rate} = \lambda)$.

Se dice que la distribución exponencial no tiene memoria porque las variables aleatorias exponenciales olvidan qué tan viejas son en cada momento. Es decir, la probabilidad de que se deba esperar a un cliente adicional 5 horas en llegar, dado que ya se ha esperado 7 horas, es exactamente la probabilidad de que se necesite esperar 5 horas por un cliente en primer lugar. Matemáticamente, para cualquier $s, t > 0$,

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

La Distribución Gamma Esta es una generalización de la distribución exponencial. Se dice que X tiene una distribución gamma y se escribe $X \sim \text{gamma}(\text{shape} = \alpha, \text{rate} = \lambda)$. Tiene función de distribución de probabilidad

$$f_X(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}, \quad x > 0$$

Las funciones asociadas en R son `dgamma(x, shape, rate = 1)`, `pgamma`, `qgamma` y `rgamma`, las cuales corresponden a la función de distribución de probabilidad, la función de distribución acumulada, función cuantil, y simula variables aleatorias, respectivamente. Si $\alpha = 1$ entonces $X \sim \exp(\text{rate} = \lambda)$. La media es $\mu = \alpha/\lambda$ y la varianza es $\sigma^2 = \alpha/\lambda^2$.

Hay que recordar que si X mide la longitud de tiempo hasta que el primer evento ocurre en un proceso de Poisson con tasa λ entonces $X \sim \exp(\text{rate} = \lambda)$. Si Y mide la longitud de tiempo hasta que el evento α^{esimo} ocurre entonces $Y \sim \text{gamma}(\text{shape} = \alpha, \text{rate} = \lambda)$. Cuando α es un entero esta distribución también es conocida como la distribución Erlang.

Ejemplo 4.29. En un autolavado, dos clientes llegan por hora, en promedio. Se decide medir cuánto tiempo toma hasta que llega el tercer cliente. Si Y denota este tiempo aleatorio entonces $Y \sim \text{gamma}(\text{shape} = 3, \text{rate} = 1/2)$.

3.5.2. Las Distribuciones Chi Cuadrada, t Student y F Snedecor

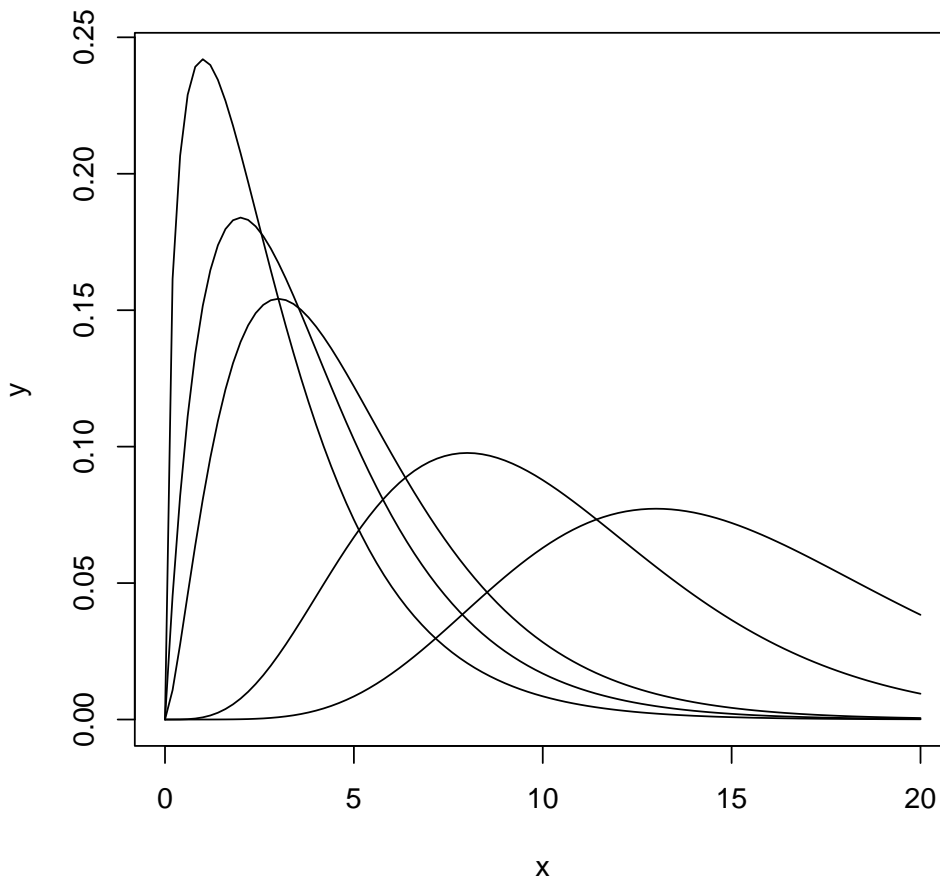
Distribución Chi cuadrada.

Una variable aleatoria X con función de distribución de probabilidad

$$f_X(x) = \frac{1}{\Gamma(p/2)2^{p/2}} x^{p/2-1} e^{-x/2}, \quad x > 0$$

se dice que tiene una distribución Chi cuadrada con p grados de libertad. Se escribe $X \sim \text{chisq}(df = p)$. Las funciones asociadas en R son `dchisq(x, df)`, `pchisq`, `qchisq` y `rchisq`, las cuales corresponden a la función de distribución de probabilidad, la función de distribución acumulada, función cuantil, y simula variables aleatorias, respectivamente. En una notación obvia se puede definir $\chi_\alpha^2(p)$ como un número en el eje x tal que hay exactamente α área bajo la curva $\text{chisq}(df = p)$ a su derecha. El código para producir esta gráfica es el siguiente:

```
> curve(dchisq(x, df=3), form=0, to=20, ylab="y")
> ind<-c(4, 5, 10, 15)
> for(i in ind) curve(dchisq(x, df=i), 0, 20, add=TRUE)
```



A continuación se mencionan algunos datos importantes acerca de la distribución Chi cuadrada.

1. Si $Z \sim \text{norm}(\text{mean} = 0, \text{sd} = 1)$, entonces $Z^2 \sim \text{chisq}(\text{df} = 1)$. Este hecho es importante cuando se trata de encontrar la distribución de la varianza muestral, S^2 .
2. La distribución chi cuadrada se soporta en el eje positivo x , con una distribución inclinada a la derecha.
3. La distribución $\text{chisq}(\text{df} = p)$ es la misma que la distribución $\text{gamma}(\text{shape} = p/2, \text{rate} = 1/2)$.
4. La función generadora de momentos de $X \sim \text{chisq}(\text{df} = p)$ es

$$M_X(t) = (1 - 2t)^{-p/2}, \quad t < 1/2$$

Distribución t Student

Se dice que una variable X con distribución

$$f_X(x) = \frac{\Gamma[(r+1)/2]}{\sqrt{r\pi}\Gamma(r/2)} \left(1 + \frac{x^2}{r}\right)^{-(r+1)/2}, \quad -\infty < x < \infty$$

tiene una distribución t Student, con r grados de libertad, y se escribe $X \sim t(\text{df} = r)$. Las funciones asociadas en R son `dt.pr`, `qt` y `rt`, las cuales corresponden a la función de distribución de probabilidad, la función de distribución acumulada, función cuantil, y simula variables aleatorias, respectivamente.

Distribución F Snedecor

Se dice que una variable aleatoria X con función

$$f_X(x) = \frac{\Gamma[(m+n)/2]}{\Gamma(m/2)\Gamma(n/2)} \left(\frac{m}{n}\right)^{m/2} x^{m/2-1} \left(1 + \frac{m}{n}x\right)^{-(m+n)/2}, \quad x > 0$$

tiene una distribución F Snedecor con (m, n) grados de libertad. Se escribe $X \sim f(\text{df1} = m, \text{df2} = n)$. Las funciones asociadas en R son `df(x, df1, df2)`, `pf`, `qf` y `rf`, las cuales corresponden a la función de distribución de probabilidad, la función de distribución acumulada, función cuantil, y simula variables aleatorias, respectivamente. Se define $F_\alpha(m, n)$ como el número en el eje x tal que hay exactamente α área bajo la curva $f(\text{df1} = m, \text{df2} = n)$ a su derecha.

A continuación algunas notas acerca de la distribución F

1. Si $X \sim f(\text{df1} = m, \text{df2} = n)$ y $Y = 1/X$, entonces $Y \sim f(\text{df1} = n, \text{df2} = m)$. Históricamente este hecho fue especialmente conveniente. Anteriormente, los estadistas usaban tablas impresas para sus cálculos

estadísticos. Ya que las tablas de F eran simétricas en m y n , eso significaba que los publicadores podrían cortar el tamaño de sus tablas a la mitad.

2. Si $X \sim t(df = r)$, entonces $X^2 \sim f(df1 = 1, df2 = r)$.

3.5.3. Otras Distribuciones Populares

La Distribución de Cauchy

Esta es un caso especial de la distribución t de Student. Tiene función de distribución

$$f_X(x) = \frac{1}{\beta\pi} [1 + (\frac{x-m}{\beta})^2]^{-1}, \quad -\infty < x < \infty$$

Se escribe $X \sim \text{cauchy}(\text{location} = m, \text{scale} = \beta)$. La función asociada en R es `dcauchy(x, location = 0, scale = 1)`.

Es fácil ver que una distribución $\text{cauchy}(\text{location} = 0, \text{scale} = 1)$ es la misma que $t(df = 1)$. La distribución de cauchy parece una distribución normal pero con colas pesadas. La media y la varianza no existen, es decir, son infinitas. La media está representada por el parámetro *location*, y el parámetro *scale* influencia la anchura de la distribución al rededor de su media.

La Distribución Beta

Esta es una generalización de la distribución uniforme continua.

$$f_X(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1}, \quad 0 < x < 1$$

Se escribe $X \sim \text{beta}(\text{shape1} = \alpha, \text{shape2} = \beta)$. La función asociada en R es `dbeta(x, shape1, shape2)`. La media y la varianza son

$$\mu = \frac{\alpha}{\alpha+\beta} \text{ y } \sigma^2 = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

Esta distribución aparece en muchas estadísticas Bayesianas porque es un buen modelo para las predicciones acerca de una proporción de una población $p, 0 \leq p \leq 1$.

La Distribución Logística

$$f_X(x) = \frac{1}{\sigma} \exp(-\frac{x-\mu}{\sigma}) [1 + \exp(-\frac{x-\mu}{\sigma})]^{-1}, \quad -\infty < x < \infty$$

Se escribe $X \sim \text{logis}(\text{location} = \mu, \text{scale} = \sigma)$. La función asociada en R es `dlogis(x, location = 0, scale = 1)`. La distribución logística aparece en ecuaciones diferenciales como un modelo para el crecimiento poblacional bajo ciertas condiciones. La media es μ y la varianza es $\pi^2 \sigma^2 / 3$.

La Distribución Logarítmica Normal

Esta es una distribución derivada de la distribución normal. Si $U \sim \text{norm}(\text{mean} = \mu, \text{sd} = \sigma)$, entonces $X = e^U$ tiene función de distribución

$$f_X(x) = \frac{1}{\sigma x \sqrt{2\pi}} \exp[-\frac{(\ln x - \mu)^2}{2\sigma^2}], \quad 0 < x < \infty$$

Se escribe $X \sim \text{lnorm}(\text{meanlog} = \mu, \text{sdlog} = \sigma)$. La función asociada en R es `dlnorm(x, meanlog = 0, sdlog = 1)`. Hay que notar que el conjunto soporte está concentrado en el eje x positivo; la distribución está cargada a la derecha con una cola pesada.

La Distribución Weibull

Esta tiene función de distribución

$$f_X(x) = \frac{\alpha}{\beta} (\frac{x}{\beta})^{\alpha-1} \exp(-\frac{x}{\beta})^\alpha, \quad x > 0$$

Se escribe $X \sim \text{weibull}(\text{shape} = \alpha, \text{scale} = \beta)$. La función asociada en R es `dweibull(x, shape, scale = 1)`.

En R

Hay un soporte de momentos y función generadora de momentos para algunas distribuciones continuas incluidas en la paquetería *actuar*. La convención es m en frente del nombre de la distribución para momentos crudos, y mgf en frente del nombre de la distribución para la función generadora de momentos. Las siguientes distribuciones están soportadas: gamma, Gaussiana inversa, chi cuadrada no centrada, exponencial, y uniforme.

Ejemplo 4.22. Calcular los primeros cuatro momentos crudos para $X \sim \text{gamma}(\text{shape} = 13, \text{rate} = 1)$ y graficar la función generadora de momentos.

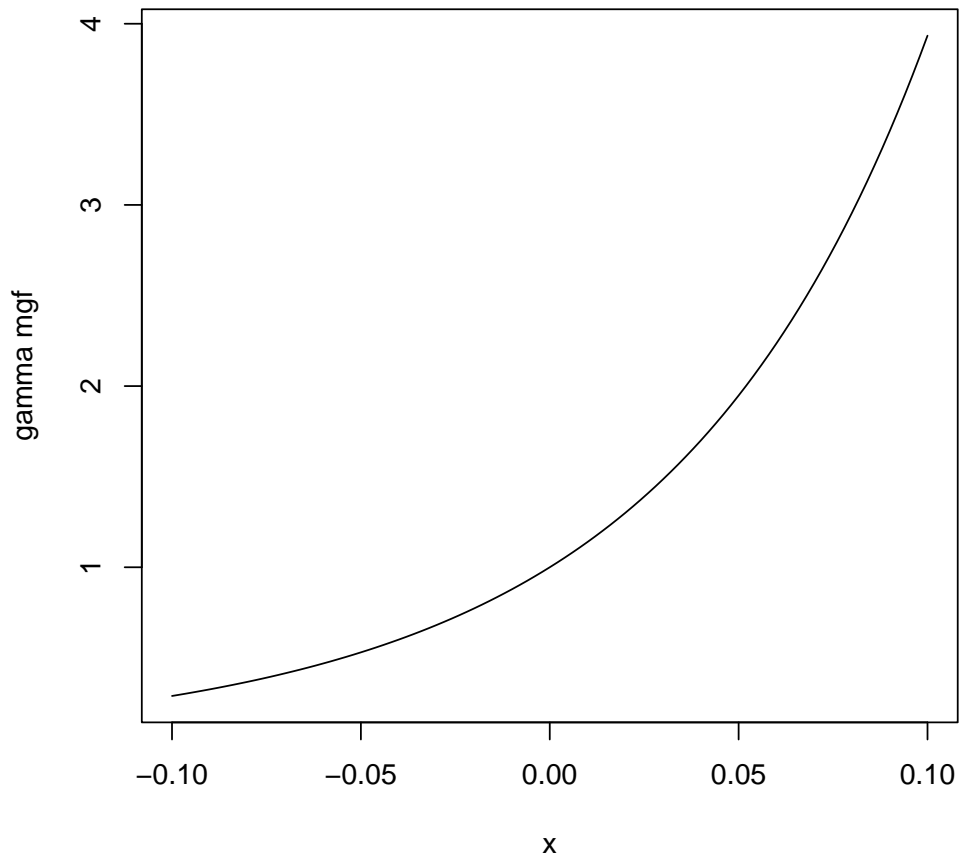
Se carga la paquetería *actuar* y se usan las funciones *mgamma* y *mgfgamma*:

```
> library(actuar)
> mgamma(1:4, shape=13, rate=1)
```

```
[1] 13 182 2730 43680
```

Para la gráfica se puede usar la función de la siguiente forma:

```
> plot(function(x){  
+   mgfgamma(x,shape=13,rate=1)  
+ }, from =-0.1, to= 0.1, ylab="gamma mgf")
```



Capítulo 4

Distribuciones de Muestras

Este capítulo es el puente entre la probabilidad y la estadística descriptiva, estudiadas en los capítulos anteriores.

El enlace es el siguiente: Se presenta una población acerca de la cuál se desea aprender. Y mientras sería deseable examinar cada miembro de la población, resulta ser difícil o imposible hacerlo, por tanto, se busca recolectar una muestra en vez de eso. El método basta, siempre que la muestra sea representativa de la población. Una buena manera de lograr esto es tomar una muestra aleatoriamente de la población.

Suponiendo que se ha recolectado una muestra aleatoria, la siguiente tarea es dar sentido a los datos porque la lista completa de la información de la muestra es engorrosa, difícil de manejar. Se resumen los datos con estadística descriptiva, una cantidad calculada a partir de los datos. Pero la muestra era aleatoria, por tanto, significa que la estadística será aleatoria también. ¿Cómo está distribuida la estadística?

La distribución de probabilidad asociada con la población (de la cuál se tomó la muestra) es llamada la distribución de población, y la distribución de probabilidad asociada con nuestra estadística es su distribución muestral; claramente, ambas están interrelacionadas. Para aprender acerca de la distribución de la población, es necesario conocer todo acerca de la distribución muestral. Ese es el objetivo del capítulo.

Primero se introduce la noción de muestras aleatorias simples y se catalogan algunas de sus propiedades matemáticas más convenientes. Después uno se concentra en lo que pasa en el caso especial de muestrear a partir de la distribución normal (la cual, de nuevo, tiene varias propiedades matemáticas convenientes), y en particular, se conocen las distribuciones muestrales \bar{X} y S^2 . Después se explora lo que sucede con la distribución muestral de \bar{X} cuando la población no es normal y se prueba uno de los teoremas más importantes en estadística, el Teorema del Límite Central.

Teniendo dicho teorema, se investigan las distribuciones muestrales de varias estadísticas populares, tomando ventaja de aquellas con formas más manejables. Se termina el capítulo con una exploración de estadísticas cuyas distribuciones muestrales no son manejables con facilidad, y para lograr este objetivo se usarán métodos de simulación que son usados en capítulos previos.

Se busca enseñar

- La noción de población versus muestra aleatoria simple, parámetro versus estadística, y distribución de población versus distribución muestral.
- Las distribuciones muestrales clásicas de una y dos muestras estadísticas estándar.
- Cómo generar una distribución muestral simulada cuando la estadística es muy desordenada
- El Teorema del Límite Central
- Algunos conceptos básicos relacionados a la utilidad de la distribución muestral, tales como el sesgo y la varianza

4.1. Muestras Aleatorias Simples

Definición. Si X_1, X_2, \dots, X_n son independientes con $X_i \sim f$ para $i = 1, 2, \dots, n$, entonces se dice que X_1, X_2, \dots, X_n son independientes e igualmente distribuidas de la población f o alternativamente se dice que X_1, X_2, \dots, X_n son una muestra aleatoria simple de tamaño n , denotada por $SRS(n)$, de la población f .

Proposición. Sea X_1, X_2, \dots, X_n una $SRS(n)$ de una distribución de población con media μ y desviación estándar finita σ . Entonces la media y desviación estándar de \bar{X} están dadas por las ecuaciones $\mu_{\bar{X}} = \mu$ y $\sigma_{\bar{X}} = \sigma/\sqrt{n}$

Proposición. Sea X_1, X_2, \dots, X_n una $SRS(n)$ de una distribución de población con función generadora de momentos $M(t)$. Entonces la función generadora de momentos de \bar{X} está dada por

$$M_{\bar{X}}(t) = [M(\frac{t}{n})]^n$$

Demostración. De la definición:

$$M_{\bar{X}}(t) = Ee^{t\bar{X}} \quad (4.1)$$

$$= Ee^{t(X_1, X_2, \dots, X_n)/n} \quad (4.2)$$

$$= Ee^{tX_1/n} e^{tX_2/n} \dots e^{tX_n/n} \quad (4.3)$$

y dado que X_1, X_2, \dots, X_n son independientes, lo que permite distribuir la expectación entre cada término en el producto, que es

$$Ee^{tX_1/n} Ee^{tX_2/n} \dots Ee^{tX_n/n}$$

El último paso es reconocer que cada término en el último producto es exactamente $M(t/n)$

4.2. Muestreo de una Distribución Normal

4.2.1. La Distribución de la Media Muestral

Proposición. Sea X_1, X_2, \dots, X_n una S $RS(n)$ de una distribución $norm(mean = \mu, sd = \sigma)$. Entonces la media muestral \bar{X} tiene una distribución muestral $norm(mean = \mu, sd = \sigma/\sqrt{n})$.

Demostración La media y la desviación estándar de \bar{X} surge directamente de la proposición anterior. Recordando que función generadora de momentos $norm(mean = \mu, sd = \sigma)$ tiene la forma

$$M(t) = \exp\{\mu t + \sigma^2 t^2/2\}$$

Luego se usa una proposición para encontrar

$$M_{\bar{X}}(t) = [M(\frac{t}{n})]^n \quad (4.4)$$

$$= [\exp\{\mu(t/n) + \sigma^2(t/n)^2/2\}]^n \quad (4.5)$$

$$= \exp\{n[\mu(t/n) + \sigma^2(t/n)^2/2]\} \quad (4.6)$$

$$= \exp\{\mu t + (\sigma/\sqrt{n})^2 t^2/2\} \quad (4.7)$$

y se reconoce que esta última cantidad como la función generadora de momentos de la distribución $norm(mean = \mu, sd = \sigma/\sqrt{n})$.

4.2.2. La distribución de la Varianza Muestral

Teorema. Sea X_1, X_2, \dots, X_n con una S $RS(n)$ de una distribución $norm(mean = \mu, sd = \sigma)$ y sean

$$\bar{X} = \sum_{i=1}^n X_i \text{ y } S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

Entonces

1. \bar{X} y S^2 son independientes, y
2. La varianza muestral reescalada

$$\frac{(n-1)}{\sigma^2} S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\sigma^2}$$

tiene una distribución muestral $chisq(df = n - 1)$.

4.2.3. Estadística de la Distribución de t Student

Proposición. Sea X_1, X_2, \dots, X_n una S $RS(n)$ de una distribución $norm(mean = \mu, sd = \sigma)$. Entonces la cantidad

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}}$$

tiene una distribución muestral $t(df = n - 1)$

Demostración. Dividiendo el numerador y el denominador por σ y reescribiendo

$$T = \frac{\frac{\bar{X} - \mu}{S/\sqrt{n}}}{\frac{S/\sigma}{\sqrt{\frac{(n-1)S^2}{\sigma^2}/(n-1)}}}$$

Ahora sea

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \text{ y } V = \frac{(n-1)S^2}{\sigma^2}$$

tal que

$$T = \frac{Z}{\sqrt{V/r}}$$

conde $r = n - 1$

Se sabe que $Z \sim norm(mean = 0, sd = 1)$ y que $V \sim chisq(df = n - 1)$. Más aún, dado que se está muestreando a partir de una distribución normal, \bar{X} y S^2 son independientes y también lo son Z y V . En resumen, la distribución de T es la misma que la distribución de la cantidad $Z/\sqrt{V/r}$, conde $Z \sim norm(mean = 0, sd = 1)$ y $V \sim chisq(df = r)$ son independientes. Esta es la definición de la distribución de t Student.

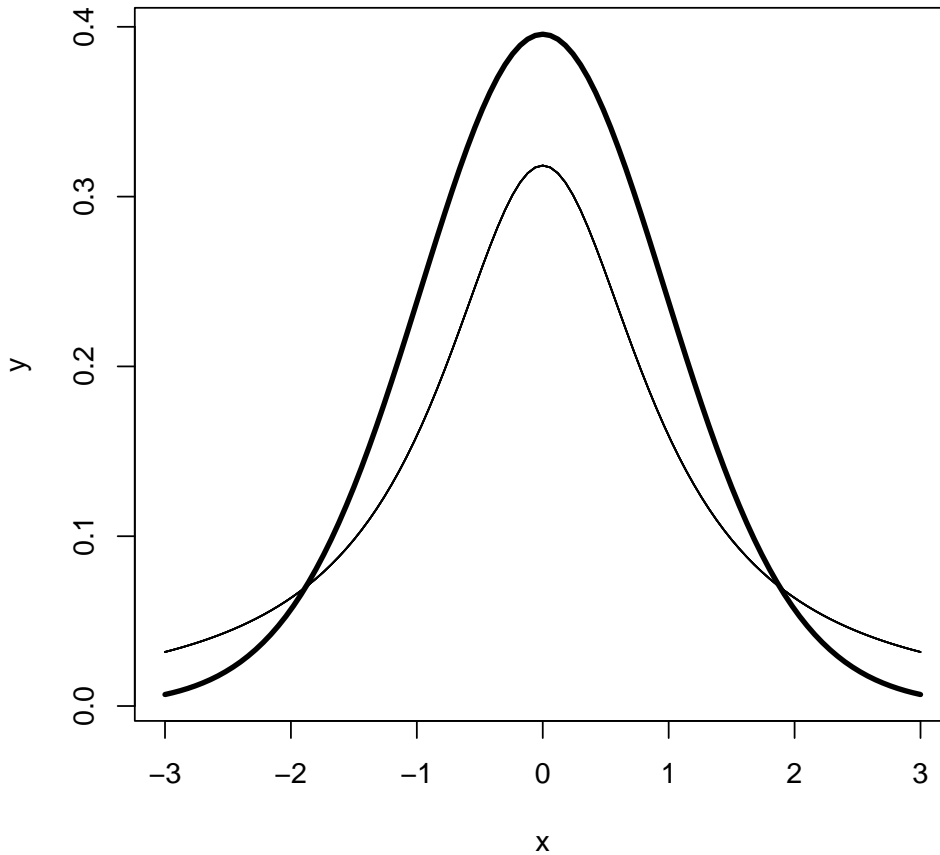
Esta distribución fue publicada por primera vez en 1900 por W.S. Gosset bajo el pseudónimo Student, y la distribución ha sido llamada consecuentemente distribución t Student. La función de distribución de probabilidad de T puede ser derivada explícitamente y tiene la forma

$$f_X(x) = \frac{\Gamma[(r+1)/2]}{\sqrt{r\pi}\Gamma(r/2)} \left(1 + \frac{x^2}{r}\right)^{-(r+1)/2}, \quad -\infty < x < \infty$$

Cualquier variable aleatoria X con la función de distribución anterior tiene una distribución t Student con r grados de libertad, y se escribe $X \sim t(df = r)$. La forma de la función de distribución es similar a la normal, pero con las colas considerablemente más pesadas. Como con la distribución normal, hay 4 funciones en R asociadas con la distribución t, a saber, dt , pt , qt , y rt , las cuales calcular la función de distribución de probabilidad, la función de distribución acumulada, la función cuantil, y genera variables aleatorias, respectivamente.

Para graficar dicha curva el código es:

```
> curve(dt(x,df=30),from=-3, to=3, lwd=3, ylab="y")
> ind<-c(1,2,3,5,10)
> for (i in ind) curve(dt(x,df=i),-3,3,add=TRUE)
```



Ejemplo 4.1. Encontrar $t_{0,01}(df = 23)$ con la función cuantil

```
> qt(0.01,df=23,lower.tail = FALSE)
```

```
[1] 2.499867
```

Hay algunas cosas a notar acerca de la distribución $t(df = r)$.

1. La distribución $t(df = 1)$ es la misma que la distribución *cauchy*(*location* = 0, *scale* = 1).
2. La desviación estándar de $t(df = r)$ está indefinida (es decir, infinita) a menos que $r > 2$. Cuando r es mayor que 2, la desviación estándar es siempre mayor que uno, pero decrece a 1 mientras $r \rightarrow \infty$.
3. Mientras $r \rightarrow \infty$, la distribución $t(df = r)$ se acerca a la distribución *norm*(*mean* = 0, *sd* = 1).

4.3. El Teorema del Límite Central

En esta sección se estudia la distribución de la media muestral cuando la distribución subyacente no es normal. Cuando X_1, X_2, \dots, X_n es una *SR* $S(n)$ de una distribución *norm*(*mean* = μ , *sd* = σ) entonces $\bar{X} \sim \text{norm}(\text{mean} = \mu, \text{sd} = \sigma/\sqrt{n})$. En otras palabras, se puede decir que cuando la población subyacente es normal entonces la distribución muestral de Z definida por

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

es *norm*(*mean* = 0, *sd* = 1).

Sin embargo, hay muchas poblaciones que no son normales, y los estadistas a menudo se encuentran muestreando tales poblaciones. ¿Qué se puede decir en este caso? La respuesta está contenida en el siguiente teorema.

Teorema. El Teorema del Límite Central. Sea X_1, X_2, \dots, X_n una *SR* $S(n)$ de una distribución de población con media μ y desviación estándar finita σ . Entonces la distribución muestral de

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

se acerca a la distribución $norm(mean = 0, sd = 1)$ cuando $n \rightarrow \infty$.

Se está suponiendo que X_1, X_2, \dots, X_n son distribuidas idéntica e independientemente; además \bar{X} tiene media μ y desviación estándar σ/\sqrt{n} , por tanto ya se sabe que Z tiene media 0 y desviación estándar 1. La belleza de este teorema es que aborda la forma de la distribución Z cuando el tamaño muestral es grande.

Hay que notar que la forma de la distribución de población subyacente no se menciona en el teorema anterior; de hecho, el resultado es verdadero para cualquier población que se comporta bien lo suficiente para tener una desviación estándar finita. En particular, si la población tiene distribución normal entonces se sabe que la distribución de \bar{X} (y Z por extensión) es exactamente normal, para cada n .

¿Qué tan grande es "suficientemente grande"? Es aquí que la forma de la distribución de población subyacente juega un rol. Para poblaciones con distribuciones que son aproximadamente simétricas y con forma de montículo, las muestras pueden necesitar ser solo de tamaño cuatro o cinco, mientras que para poblaciones altamente sesgadas o con colas pesadas las muestras pueden necesitar ser mucho más grandes para la distribución de las medias muestrales para empezar a mostrar una forma de campana. A pesar de esto, para una distribución de población dada (con desviación estándar finita) la aproximación tiende a ser mejor para tamaños muestrales mayores.

4.4. Distribuciones Muestrales de Estadísticas de Dos Muestras

A menudo hay dos poblaciones bajo consideración, y a veces es de interés comparar las propiedades entre grupos. Para hacerlo se toman muestras independientes de cada población y se calculan estadísticas muestrales respectivas para comparación. En algunos casos simples la distribución muestral de la comparación es conocida y fácil de manejar; tales casos son el objeto de esta sección.

4.4.1. Diferencia de Medias Muestrales Independientes

Proposición. Sea X_1, X_2, \dots, X_{n_1} una $SRS(n_1)$ de una distribución $norm(mean = \mu_X, sd = \sigma_X)$ y sea Y_1, Y_2, \dots, Y_{n_2} una $SRS(n_2)$ de una distribución $norm(mean = \mu_Y, sd = \sigma_Y)$. Suponiendo que X_1, X_2, \dots, X_{n_1} y Y_1, Y_2, \dots, Y_{n_2} son muestras independientes. Entonces la cantidad

$$\frac{\bar{X} - \bar{Y} - (\mu_X - \mu_Y)}{\sqrt{\sigma_X^2/n_1 + \sigma_Y^2/n_2}}$$

tiene una distribución muestral $norm(mean = 0, sd = 1)$. Equivalentemente $\bar{X} - \bar{Y}$ tiene distribución muestral $norm(mean = \mu_X - \mu_Y, sd = \sqrt{\sigma_X^2/n_1 + \sigma_Y^2/n_2})$.

Demostración. Se sabe que \bar{X} es $norm(mean = \mu_X, sd = \sigma_X/\sqrt{n_1})$ y también se sabe que \bar{Y} es $norm(mean = \mu_Y, sd = \sigma_Y/\sqrt{n_2})$. Y dado que las muestras X_1, X_2, \dots, X_{n_1} y Y_1, Y_2, \dots, Y_{n_2} son independientes, entonces también lo son \bar{X} y \bar{Y} . La distribución de su diferencia es por lo tanto normal.

Incluso si la distribución de una o ambas muestras no es normal, la cantidad de la ecuación anterior será aproximadamente normal siempre que ambas muestras sean grandes.

Para el caso especial de $\mu_X = \mu_Y$ se ha demostrado que

$$\frac{\bar{X} - \bar{Y}}{\sqrt{\sigma_X^2/n_1 + \sigma_Y^2/n_2}}$$

tiene distribución muestral $norm(mean = 0, sd = 1)$, o en otras palabras, $\bar{X} - \bar{Y}$ tiene distribución muestral $norm(mean = 0, sd = \sqrt{\sigma_X^2/n_1 + \sigma_Y^2/n_2})$.

4.4.2. Diferencia de Proporciones Muestrales Independientes

Proposición. Sea X_1, X_2, \dots, X_{n_1} una $SRS(n_1)$ de una distribución $binom(size = 1, prob = p_1)$ y sea Y_1, Y_2, \dots, Y_{n_2} una $SRS(n_2)$ de una distribución $binom(size = 1, prob = p_2)$. Supóngase que X_1, X_2, \dots, X_{n_1} y Y_1, Y_2, \dots, Y_{n_2} son muestras independientes. Definiendo

$$\hat{p}_1 = \frac{1}{n_1} \sum_{i=1}^{n_1} X_i \text{ y } \hat{p}_2 = \frac{1}{n_2} \sum_{j=1}^{n_2} Y_j$$

Entonces la distribución muestral de

$$\frac{\hat{p}_1 - \hat{p}_2 - (p_1 - p_2)}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}}$$

se acerca a una distribución $norm(mean = 0, sd = 1)$ cuando $n_1, n_2 \rightarrow \infty$. En otras palabras, la distribución muestral de $\hat{p}_1 - \hat{p}_2$ es aproximadamente

$$\text{norm}(\text{mean} = p_1 - p_2, \text{sd} = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}})$$

siempre que n_1 y n_2 sean lo suficientemente grandes.

Demostración. Se sabe que \hat{p}_1 es aproximadamente normal para n_1 suficientemente grande por el teorema del límite central, y se sabe que \hat{p}_2 es aproximadamente normal para n_2 suficientemente grande, también por el teorema del límite central. Además, \hat{p}_1 y \hat{p}_2 son independientes ya que son derivados de muestras independientes. Y una diferencia de distribuciones normales independientes es normal.

4.4.3. Tasa de Varianzas Muestrales Independientes

Proposición. Sea X_1, X_2, \dots, X_{n_1} una $S\ RS(n_1)$ de una distribución $\text{norm}(\text{mean} = \mu_X, \text{sd} = \sigma_X)$ y sea Y_1, Y_2, \dots, Y_{n_2} una $S\ RS(n_2)$ de una distribución $\text{norm}(\text{mean} = \mu_Y, \text{sd} = \sigma_Y)$. Supóngase que X_1, X_2, \dots, X_{n_1} y Y_1, Y_2, \dots, Y_{n_2} son muestras independientes. Entonces la tasa

$$F = \frac{\sigma_Y^2 S_X^2}{\sigma_X^2 S_Y^2}$$

tiene una distribución muestral $t(df1 = n_1 - 1, df2 = n_2 - 1)$.

Demostración. Se sabe que $(n_1 - 1)S_X^2/\sigma_X^2$ tiene distribución $\text{chisq}(df = n_1 - 1)$ y $(n_2 - 1)S_Y^2/\sigma_Y^2$ tiene distribución $\text{chisq}(df = n_2 - 1)$. Ahora, escribiendo

$$F \frac{\sigma_Y^2 S_X^2}{\sigma_X^2 S_Y^2} = \frac{(n_1 - 1)S_X^2/(n_1 - 1)}{(n_2 - 1)S_Y^2/(n_2 - 1)} \cdot \frac{1/\sigma_X^2}{1/\sigma_Y^2}$$

multiplicando y dividiendo el numerador con $n_1 - 1$ y haciendo lo mismo para el denominador con $n_2 - 1$. Ahora se pueden reagrupar los términos en

$$F = \frac{\frac{(n_1 - 1)S_X^2}{\sigma_X^2}/(n_1 - 1)}{\frac{(n_2 - 1)S_Y^2}{\sigma_Y^2}/(n_2 - 1)}$$

y se reconoce a F como la tasa de distribuciones chisq independientes, cada una dividida por su respectivo numerador $df = n_1 - 1$ y el denominador $df = n_2 - 1$ grados de libertad. Esta es, de hecho, la distribución F de Snedecor.

Para el caso especial $\sigma_X = \sigma_Y$ se ha demostrado que

$$F = \frac{S_X^2}{S_Y^2}$$

tiene distribución muestral $f(df1 = n_1, df2 = n_2 - 1)$.

4.5. Distribuciones Muestrales Simuladas

Algunas comparaciones son interesantes, pero su distribución muestral no es tan sencilla de describir analíticamente. ¿Qué se hace entonces?

Resulta que no se necesitan formas analíticas exactas de la distribución muestral; a veces es suficiente aproximarla con una distribución simulada. En esta sección se mostrará cómo. Hay que notar que R está particularmente bien equipado para calcular distribuciones muestrales simuladas.

4.5.1. El Rango Intercuartil

```
> iqrs<-replicate(100, IQR(rnorm(100)))
```

Se puede pedir la media de los valores simulados

```
> mean(iqrs)
```

```
[1] 1.324773
```

y se puede ver la desviación estándar

```
> sd(iqrs)
```

```
[1] 0.165943
```

Ahora se mostrará cómo graficar los valores simulados

4.5.2. La Desviación Absoluta Mediana

```
> mads<-replicate(100,mad(rnorm(100)))
```

Se puede calcular la media de los valores simulados.

```
> mean(mads)
```

```
[1] 0.9871807
```

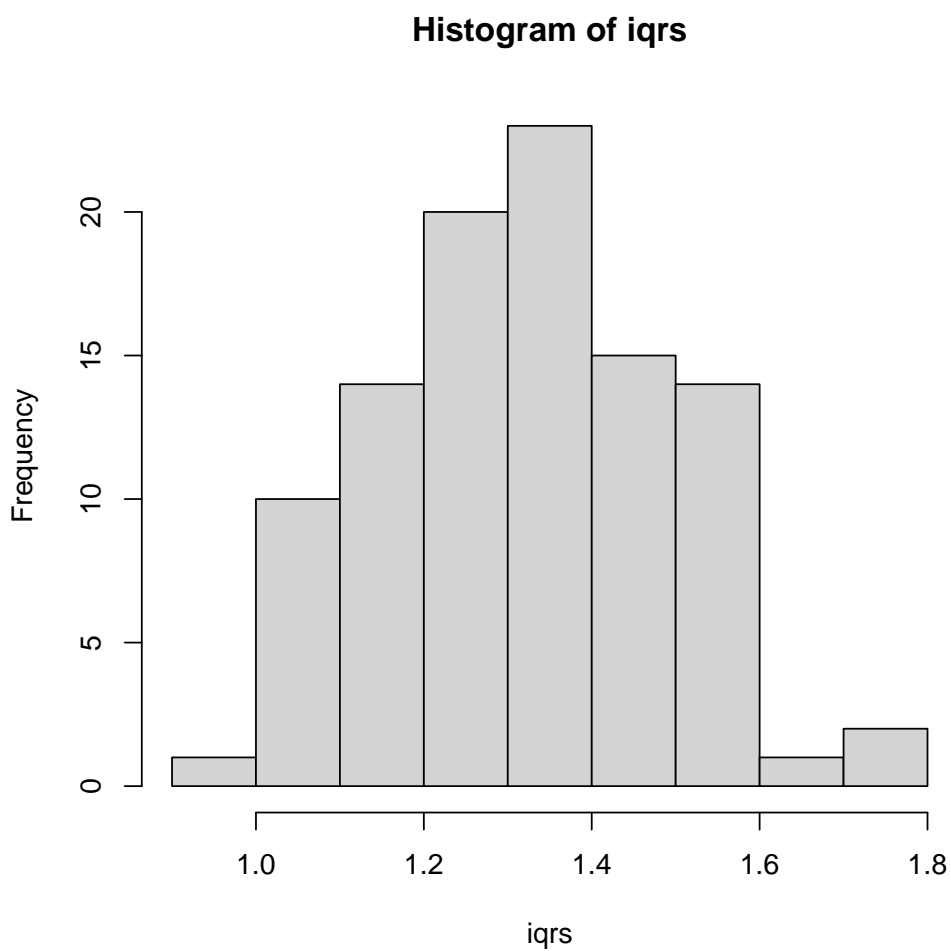
Y se puede ver la desviación estándar

```
> sd(mads)
```

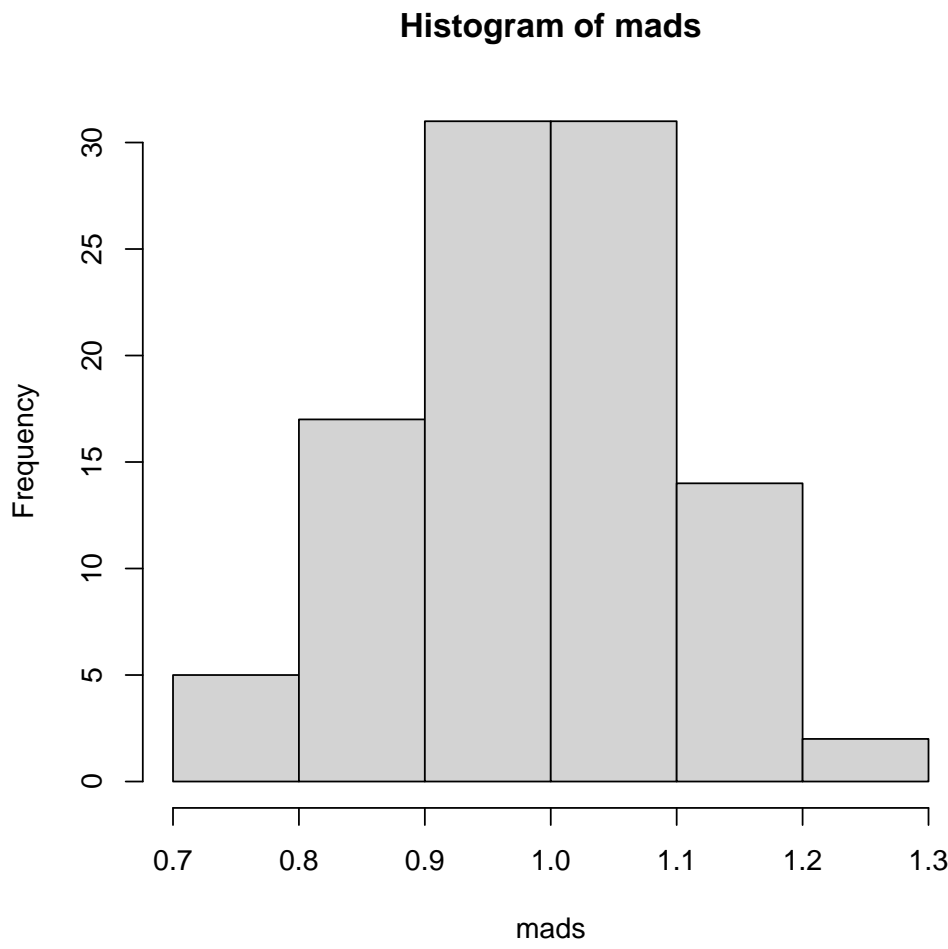
```
[1] 0.1144789
```

Los histogramas de frecuencia de *iqrs* y de *mads* se grafican de la siguiente manera:

```
> hist(iqrs)
```



```
> hist(mads)
```



Capítulo 5

Pruebas de Hipótesis de Una Muestra

Cualquiera que sea la ocupación de uno, a menudo se tiene que evaluar si algo diferente y nuevo ha sucedido. A veces se inicia con una población de la que se conoce bastante (como su media y desviación estándar) y se selecciona una muestra. ¿Es dicha muestra como el resto de la población, o representa algo fuera de lo ordinario?

Para responder a esa pregunta, se mide cada muestra individual y se calcula la estadística de la muestra. Después se comparan esas estadísticas con los parámetros de la población. ¿Son los mismos? ¿son diferentes? ¿es la muestra extraordinaria en algún sentido? El uso apropiado de la estadística ayuda a tomar la decisión.

A veces, sin embargo, no se conocen los parámetros de la población de donde vino la muestra. ¿Qué sucede en esos casos? En este capítulo, se discuten técnicas estadísticas y funciones en R para lidiar con ambos casos.

5.1. Hipótesis, Pruebas y Errores

Una hipótesis es una suposición acerca de la manera en que algo funciona. Es una explicación tentativa de un proceso, ya sea que el proceso ocurra en la naturaleza o en un laboratorio.

Antes de estudiar y medir a los individuos en una muestra, un investigador formula hipótesis que predican cómo deben ser los datos.

Generalmente, una hipótesis predice que los datos no mostrarán nada nuevo o fuera de lo ordinario. Esto es llamado la hipótesis nula H_0 . De acuerdo con la hipótesis nula, si los datos se desvían de la norma en cualquier manera, esa desviación es debida estrictamente a la probabilidad. Otra hipótesis, la hipótesis alternativa H_1 , explica las cosas de manera diferente. De acuerdo con la hipótesis alternativa, los datos muestran algo importante.

Después de reunir los datos, depende del investigador tomar una decisión. El investigador debe decidir si rechazar la hipótesis nula o no rechazarla.

En la prueba de hipótesis se formulan hipótesis alternativas y nulas, se reúnen datos, y se decide si rechazar o no rechazar la hipótesis nula.

Nada en la lógica implica aceptar cualquier hipótesis. Ni tampoco implica tomar ninguna decisión acerca de la hipótesis alternativa. Se trata de rechazar o no rechazar la hipótesis nula H_0 .

A pesar de la decisión rechazar-no rechazar, un error puede suceder. Un tipo de error ocurre cuando se cree que los datos muestran algo importante y se rechaza H_0 , pero en realidad los datos son debidos a la probabilidad. Esto se llama **Error Tipo 1**. Al principio de un estudio, el investigador propone los criterios para rechazar H_0 . Cuando esto se hace, también se propone la probabilidad de un error tipo 1. Esta probabilidad se llama α .

El otro tipo de error ocurre cuando no se rechaza H_0 y los datos realmente son debidos a algo fuera de lo ordinario. Por uno u otra razón, el investigador se perdió de algo. Eso se llama **Error Tipo 2**. La probabilidad de que sucede se llama β .

5.2. Pruebas de Hipótesis y Distribuciones Muestrales

En el capítulo anterior se discutieron las distribuciones muestrales. Una distribución muestral, recordando, es un conjunto de todos los posibles valores de una estadística para un tamaño muestral dado.

También se discutió el Teorema del límite central. Este teorema dice que la distribución muestral de la media se aproxima a la distribución normal si el tamaño de la muestra es grande (para propósitos prácticos, al menos de 30). Esto funciona ya sea si la población está distribuida normalmente o no. Si la población tiene una distribución normal, la distribución muestral es normal para cualquier tamaño de muestra. Aquí hay otros dos puntos del teorema del límite central:

- La media de la distribución muestral de la media es igual a la media de la población. Es decir

$$\mu_{\bar{X}} = \mu$$

- El error estándar de la media (la desviación estándar de la distribución muestral) es igual a la desviación estándar de la población dividida por la raíz cuadrada del tamaño de la muestra.

$$\sigma_{\bar{X}} = \sigma/\sqrt{N}$$

La distribución muestral de la media figura prominentemente en el tipo de prueba de hipótesis que se discute en este capítulo. Teóricamente, cuando se prueba una hipótesis nula contra una hipótesis alternativa, cada hipótesis corresponde a una distribución muestral separada.

La Figura 5.1 lo ilustra. La figura muestra dos distribuciones normales. Estas están colocadas arbitrariamente. Cada distribución normal representa una distribución muestral de la media. La que está en la izquierda representa la distribución de posibles medias muestrales si la hipótesis nula es verdadera. La que está en la derecha representa la distribución de posibles medias muestrales si la hipótesis alternativa es verdadera.

METER FIGURA

Por supuesto que cuando se realiza una prueba de hipótesis, nunca se sabe qué distribución produce el resultado. Se trabaja con una media muestral, un punto en el eje horizontal. La decisión rechazar o no rechazar se reduce a decidir a cuál distribución pertenece la media muestral. Se elige un valor crítico, un criterio de decisión. Si la media muestral está a un lado del valor crítico, se rechaza H_0 . De lo contrario, no se rechaza.

En este sentido, la figura también muestra α y β . Estas, como se mencionó ya, son las probabilidades de los errores de decisión. El área que corresponde a α está en la distribución H_0 . Representa la probabilidad de que una media muestral provenga de la distribución de H_0 , pero es algo muy extremo rechazar H_0 .

Donde se coloca el valor crítico determina α . En la mayoría de las pruebas de hipótesis, se coloca α en 0.05. Esto significa que se está dispuesto a tolerar un error del tipo 1, 5 por ciento de las veces. Gráficamente, el valor crítico corta 5 por ciento del área de la distribución muestral. Por cierto, si se está hablando del 5 por ciento del área que se encuentra en la cola derecha de la distribución, se está hablando del 5 por ciento superior. Si es el 5 por ciento en la cola izquierda, se habla del 5 por ciento inferior.

El área que corresponde a β se encuentra en la distribución H_1 . Esta área representa la probabilidad de que una media muestral proviene de la distribución de H_1 , pero está lo suficientemente cerca al centro de la distribución de H_0 , tanto que no se rechaza H_0 , aunque se debería. No se elige β . El tamaño de esta área depende de la separación entre las medias de dos distribuciones, y esto depende del proceso investigado.

Estas distribuciones muestrales son apropiadas cuando el trabajo de uno corresponde a las condiciones del teorema del límite central: si se conoce que la población con la que se está trabajando es normal o si se tiene una muestra grande.

5.3. Trabajando de nuevo con Z

A continuación se presenta un ejemplo de una prueba de hipótesis que involucra una muestra de una población normalmente distribuida. Debido a que la población tiene la distribución normal, cualquier muestra resulta en una distribución muestral normalmente distribuida. Debido a que es una distribución normal, se usan las puntuaciones z en la prueba de hipótesis:

$$z = \frac{\bar{X} - \mu}{\sigma/\sqrt{N}}$$

Debido a que se usan las puntuaciones z en la prueba de hipótesis, la puntuación z se llama la estadística de la prueba.

Supóngase que se piensa que la gente que vive en un código postal particular tiene IQ mayor que el promedio. Se toma una muestra de nueve personas de ese código postal, se les dan exámenes de IQ, se tabulan los resultados, y se calculan las estadísticas. Para la población de puntajes de IQ, $\mu = 100$ y $\sigma = 15$.

Las hipótesis son

$$H_0 : \mu_{CP} \leq 100 \quad (5.1)$$

$$H_1 : \mu_{CP} > 100 \quad (5.2)$$

Luego se asume que $\alpha = 0.05$. Esta es el área sobreada en la cola de la distribución de H_0 en la Figura 5.1. ¿A qué se debe el \leq en H_0 ? Se usa este símbolo porque se rechazará H_0 sólo si la media muestral es mayor que el valor hipotetizado. Cualquier otro resultado es evidencia en favor de no rechazar H_0 .

Suponiendo que la media muestral es 108.67. ¿Se puede rechazar H_0 ?

La prueba implica convertir 108.67 en un puntaje estándar en la distribución muestral de la media:

$$z = \frac{\bar{X} - \mu}{\sigma/\sqrt{N}} = \frac{108,67 - 100}{15/\sqrt{9}} = \frac{8,67}{5} = 1,73$$

¿Es el valor de la estadística de prueba suficientemente grande para rechazar H_0 con $\alpha = 0,05$? Lo es. El valor crítico - el valor de z que corta 5 por ciento del área en una distribución normal estándar - es 1.645. El valor calculado, 1.73, excede 1.645, así que se encuentra en la región de rechazo. La decisión es rechazar H_0 .

Esto significa que si H_0 es verdadera, la probabilidad de obtener un valor estadístico de la prueba que es al menos así de grande es menor que 0.05. La evidencia fuerte está en favor de rechazar H_0 .

En lenguaje estadístico, cualquier ocasión que se rechaza H_0 , se dice que el resultado es estadísticamente significativo.

Este tipo de prueba se llama de una cola, porque su región de rechazo se encuentra en una cola de la distribución muestral.

Una prueba de hipótesis puede ser de una cola en la otra dirección. Suponiendo que se tienen razones para creer que la gente que vive en ese código postal tiene IQ menor que el promedio, entonces las hipótesis en ese caso son:

$$H_0 : \mu_{CP} \geq 100 \quad (5.3)$$

$$H_1 : \mu_{CP} < 100 \quad (5.4)$$

Para esta prueba de hipótesis, el valor crítico de la estadística de la prueba es -1.645 si $\alpha = 0,05$.

Una prueba de hipótesis puede ser de dos colas, lo que significa que la región de rechazo se encuentra en ambas colas de la distribución muestral de H_0 . Esto sucede cuando las hipótesis se ven así:

$$H_0 : \mu_{CP} = 100 \quad (5.5)$$

$$H_1 : \mu_{CP} \neq 100 \quad (5.6)$$

En este caso, la hipótesis alternativa sólo especifica que la media es diferente del valor de la hipótesis nula, sin decir si es mayor o menor. La Figura 5.2 muestra cómo se ve la región de rechazo de dos colas para $\alpha = 0,05$. El 5 por ciento se divide igualmente entre las colas izquierda y derecha.

PONER FIGURA

Para una distribución normal estándar, la puntuación z que corta el 2.5 por ciento en la cola derecha es 1.96. La puntuación z que corta el 2.5 por ciento en la cola izquierda es -1.96. La puntuación z en el ejemplo anterior, 1.73, no excede 1.96. La decisión, en el caso de dos colas, es no rechazar H_0 .

De aquí surge un punto importante. Una prueba de hipótesis de una cola puede rechazar H_0 , mientras que una prueba de dos colas aplicada a los mismos datos puede no rechazarla. Una prueba de dos colas indica que se está buscando una diferencia entre la media muestral y la media de la hipótesis nula, pero no sabiendo en qué dirección. Una prueba de una cola muestra que se tiene una buena idea de cómo debe ser la diferencia. Para propósitos prácticos, esto significa que se debe intentar tener conocimiento suficiente para ser capaz de especificar una prueba de una cola: esto proporciona una mejor probabilidad de rechazar H_0 cuando es debido.

En R

Una función en R llamada `z.test()` sería muy útil para hacer el tipo de prueba que se discute en esta sección. El problema es que dicha función no existe en R. Aunque se puede encontrar alguna en otras paqueterías, es fácil crear una.

Se inicia creando una función con nombre `z.test`. El primer argumento es el vector de datos, el segundo es la media de la población, y el tercero es la varianza de la población.

Luego, se crea un vector que almacenará la probabilidad de una cola de la puntuación z que se calculará.

Después se calcula la puntuación z y se redondea a tres puntos decimales. Sin el redondeo, R podría calcular muchos puntos decimales, pero el resultado sería muy tedioso.

Finalmente, se calcula la probabilidad de una cola, y de nuevo se redondea a 3 decimales.

Se usa la función `abs()` porque una hipótesis alternativa puede especificar un valor por debajo de la media, y los datos podrían resultar en una puntuación z negativa.

```
> z.test=function(x,mu,popvar){
+   one.tail.p<-NULL
+   z.score <-round((mean(x)-mu)/(popvar/sqrt(length(x))),3)
+   one.tail.p <-round(pnorm(abs(z.score),lower.tail=FALSE),3)
+   cat(" z =",z.score,"\n",
+       "probabilidad de una cola =", one.tail.p,"\n",
+       "probabilidad de dos colas =", 2*one.tail.p)}
> datos_IQ<-c(100,101,104,109,125,116,105,108,110)
> z.test(datos_IQ,100,15)
```

```

z = 1.733
probabilidad de una cola = 0.042
probabilidad de dos colas = 0.084

```

La última línea de comandos consiste en imprimir el resultado.

5.4. Prueba de hipótesis con t

En el ejemplo anterior, se trabaja con puntos de IQ. La población de puntos de IQ es una distribución normal con una media y desviación estándar bien conocidas. Por lo tanto, se puede trabajar con el teorema del límite central y describir la distribución muestral como una distribución normal. Se puede usar z como la prueba estadística.

Sin embargo, en el mundo real, no se tiene el lujo de trabajar con poblaciones bien definidas. Usualmente se tienen muestras pequeñas, y típicamente se mide algo que no es tan bien conocido como el IQ. Más aún, a menudo no se conocen los parámetros de la población, y tampoco se sabe si la población tiene distribución normal.

Cuando este es el caso, se usan los datos de la muestra para estimar la desviación estándar de la población, y se trata la distribución muestral de la media como un miembro de una familia de distribuciones llamada la distribución t . Se usa t como una prueba estadística. En el capítulo anterior se introdujo esta distribución y se mencionó que se distinguen los miembros de esta familia por medio de un parámetro llamado grados de libertad (df).

La ecuación para la prueba estadística es

$$t = \frac{\bar{X} - \mu}{s/\sqrt{N}}$$

Es útil pensar en df como el denominador de la estimación de la varianza de la población. Para las pruebas de hipótesis en esta sección, es decir $N - 1$, donde N es el número de puntos en la muestra. Entre mayor sea df , más se parece la distribución t a la distribución normal.

Ejemplo 5.2. Una empresa de tecnología que produce microrobots asegura que, en promedio, en cada unidad producida, hay cuatro defectos. Un grupo de consumidores cree que el promedio es mayor. Dicho grupo toma una muestra de 9 microrobots y encuentra un promedio de 7 defectos, con una desviación estándar de 3.12. La prueba de hipótesis es

$$H_0 : \mu \leq 4 \quad (5.7)$$

$$H_1 : \mu > 4 \quad (5.8)$$

$$\alpha = 0,05 \quad (5.9)$$

La ecuación es

$$t = \frac{\bar{X} - \mu}{s/\sqrt{N}} = \frac{7-4}{3,12/\sqrt{9}} = 2,88$$

¿Se puede rechazar H_0 ? A continuación la respuesta con R

En R

Afortunadamente existe la función `t.test` en R. Entonces, sólo hay que introducir el vector de datos del número de defectos por producto, obtenidos de la muestra:

```

> robots_defectuosos<-c(3,6,9,9,4,10,6,4,12)
> t.test(robots_defectuosos,mu=4,alternative="greater")

```

One Sample t-test

```

data: robots_defectuosos
t = 2.8823, df = 8, p-value = 0.01022
alternative hypothesis: true mean is greater than 4
95 percent confidence interval:
 5.064521      Inf
sample estimates:
mean of x
 7

```

El segundo argumento de la función indica que se está probando contra una media hipotetizada de 4, y el tercer argumento indica que la hipótesis alternativa es que la media verdadera es mayor que 4.

El output provee el valor t y el valor p bajo muestra que se puede rechazar la hipótesis nula con $\alpha = 0,05$.

5.5. Prueba de una Varianza

Hasta ahora se ha discutido pruebas de una muestra para medias. Sin embargo, también se pueden probar hipótesis acerca de varianzas.

Este tema a veces resulta en un contexto de manufactura. Supóngase que una empresa que fabrica micro-robots produce una parte que tiene que tener una cierta longitud con una variabilidad muy pequeña. Se puede tomar una muestra de partes, medirlas, encontrar la variabilidad muestral, y realizar una prueba de hipótesis contra la variabilidad deseada.

La familia de distribuciones para la prueba se llama chi cuadrada. Recordando del capítulo anterior, su símbolo es χ^2 , y el parámetro df distingue un miembro de la familia de otro. Además, esta distribución no es como las anteriores, ya que pueden presentar un sesgo, y ninguna de ellas puede tener un valor menor a cero.

La ecuación para la prueba estadística es

$$\chi^2 = \frac{(N-1)s^2}{\sigma^2}$$

N es el número de puntajes en la muestra, s^2 es la varianza de la muestra, y σ^2 es la varianza de la población especificada en H_0 .

Con esta prueba, se tiene que asumir que lo que se está midiendo tiene una distribución normal.

Supóngase que un proceso que la empresa de microrrobots realiza tiene que tener, como máximo, una desviación estándar de 1.5 pulgadas para su longitud. Después de medir una muestra de 10 partes, se encuentra una desviación estándar de 1.80 pulgadas.

Las hipótesis son:

$$H_0 : \sigma^2 \leq 2,25 \quad (5.10)$$

$$H_1 : \sigma^2 > 2,25 \quad (5.11)$$

$$\alpha = 0,05 \quad (5.12)$$

Trabajando con la ecuación

$$\chi^2 = \frac{(N-1)s^2}{\sigma^2} = \frac{(10-1)(1,80)^2}{(1,5)^2} = 12,96$$

¿Se puede rechazar H_0 ?

En R

Hasta este punto, uno podría pensar que la función `chisq.test()` respondería la pregunta. Aunque R provee esta función, no es apropiado usarla aquí. Esta se usa para otro tipo de pruebas, que no se manejarán en este libro.

En vez de eso, se debe usar una función llamada `varTest`, que se encuentra en la paquetería `EnvStats`.

Se crea un vector que contenga las diez medidas descritas en el ejemplo:

```
> micro_datos<-c(12.43,11.71,14.41,11.05,9.53,11.66,9.33,11.71,14.35,13.81)
```

Y después se introduce el comando con los parámetros

```
> library(EnvStats)
> varTest(micro_datos,alternative="greater",conf.level=0.95,sigma.squared=2.25)
```

```
$statistic
Chi-Squared
12.9812
```

```
$parameters
df
9
```

```
$p.value
[1] 0.163459
```

```
$estimate
variance
3.245299
```

```

$null.value
variance
  2.25

$alternative
[1] "greater"

$method
[1] "Chi-Squared Test on Variance"

$data.name
[1] "micro_datos"

$conf.int
      LCL      UCL
1.726327      Inf
attr(,"conf.level")
[1] 0.95

attr(,"class")
[1] "htestEnvStats"

```

El primer argumento es el vector de datos. El segundo especifica la hipótesis alternativa que la varianza verdadera es mayor que la varianza hipotética, y el tercero es el nivel de confianza ($1 - \alpha$), y el cuarto es la varianza hipotética.

Entre otras estadísticas, el resultado muestra la chi cuadrada (12.9812) y el valor p (0.163459). El valor p es mayor que 0.05, por lo tanto, no se puede rechazar la hipótesis nula.

Capítulo 6

Pruebas de Hipótesis de Dos Muestras

En una gran variedad de campos, la necesidad a menudo está en comprar una muestra con otra. A veces las muestras son independientes, y a veces están relacionadas de alguna manera. Cada muestra viene de una población separada. El objetivo es decidir si estas poblaciones son diferentes entre sí.

Usualmente, esto involucra pruebas de hipótesis acerca de las medias poblacionales. También se pueden probar hipótesis acerca de las varianzas poblacionales. En este capítulo, se muestra como llevar a cabo estas pruebas, usando R.

6.1. Hipótesis para 2

Como en el caso de una muestra del capítulo anterior, la prueba de hipótesis con dos muestras inicia con una hipótesis nula H_0 y una hipótesis alternativa H_1 . La hipótesis nula especifica que cualesquiera diferencias que se aprecien entre ambas muestras son debidas estrictamente a la probabilidad. La hipótesis alternativa dice, en efecto, que cualesquiera diferencias que se aprecien son reales y no seon debidas a la probabilidad.

Es posible tener una prueba de una cola, en la cual la hipótesis alternativa especifique la dirección de la diferencia entre las dos medias, o una prueba de dos colas en la que la hipótesis alternativa no especifique la dirección de la diferencia.

Para una prueba de una cola, las hipótesis se ven así:

$$H_0 : \mu_1 - \mu_2 = 0 \quad (6.1)$$

$$H_1 : \mu_1 - \mu_2 > 0 \quad (6.2)$$

o así:

$$H_0 : \mu_1 - \mu_2 = 0 \quad (6.3)$$

$$H_1 : \mu_1 - \mu_2 < 0 \quad (6.4)$$

Para una prueba de dos colas, las hipótesis son

$$H_0 : \mu_1 - \mu_2 = 0 \quad (6.5)$$

$$H_1 : \mu_1 - \mu_2 \neq 0 \quad (6.6)$$

El cero en la hipótesis es el caso típico. Es posible, sin embargo, probar para cualquier valor.

Para realizar la prueba, primero se elige un α , la probabilidad del error del tipo 1 con la que se está dispuesto a lidiar. Después se calcula la media y la desviación estándar de cada muestra, se resta una de la otra, y se usa la ecuación para convertir el resultado en una prueba estadística. Se compara la estadística de la prueba con una distribución muestral de dicha estadística. Si se encuentra en la región de rechazo que especifica α se rechaza H_0 . De lo contrario, no se rechaza.

6.2. Pruebas con Z

Dado que el teorema del límite central dice que la distribución muestral es aproximadamente normal para muestras grandes o para muestras pequeñas de poblaciones normalmente distribuidas, se usa el puntaje z como prueba estadística. Otra manera de decir usar el puntaje z como prueba estadística es realizar una prueba z ". Aquí la ecuación:

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sigma_{\bar{x}_1 - \bar{x}_2}}$$

Por lo tanto $(\mu_1 - \mu_2)$ representa la diferencia entre las medias en H_0 .

Esta ecuación convierte la diferencia entre medias muestrales en un puntaje estándar. Compárese el puntaje estándar con una distribución normal estándar - una distribución normal con $\mu = 0$ y $\sigma = 1$. Si el puntaje está en la zona de rechazo definida por α_1 se rechaza H_0 . De lo contrario, no se rechaza.

Se usa esta ecuación cuando se conoce el valor de σ_1^2 y σ_2^2 .

Ejemplo 6.1. Una nueva técnica de entrenamiento se diseña para incrementar el IQ. Se toma una muestra de nueve personas y se les entrena con dicha técnica. Se toma otra muestra de nueve personas y se les da un entrenamiento no especial. Supóngase que la media muestral para la muestra de la nueva técnica es 110.222, y para la muestra sin entrenamiento especial es 101. La prueba de hipótesis es

$$H_0 : \mu_1 - \mu_2 \leq 0 \quad (6.7)$$

$$H_1 : \mu_1 - \mu_2 > 0 \quad (6.8)$$

Y se coloca $\alpha = 0,05$.

Se sabe que el IQ tiene una desviación estándar de 15, y se asume que la desviación estándar sería la misma en la población de gente entrenada con la nueva técnica. Por supuesto, la población no existe. La suposición es que si existiera, debería tener el mismo valor de desviación estándar que la población de IQ regular. ¿La media de esta población teórica tiene el mismo valor que la población regular? H_0 dice que sí. H_1 indica que es mayor.

La prueba estadística es

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sigma_{\bar{x}_1 - \bar{x}_2}} = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}}} = \frac{5,8}{4,53} = 1,28$$

Con $\alpha = 0,05$, el valor crítico de z - el valor que corta el 5 por ciento superior del área bajo la distribución normal estándar - es 1.645. (Es posible usar la función *qnorm*). El valor calculado de la prueba estadística es menor que el valor crítico, entonces la decisión es no rechazar H_0 .

En R

Como en el caso para una muestra, R no provee una función para una prueba z . Si esta función existiera, funcionaría de la siguiente manera:

```
> z.test2=function(x,y,popsd1,popsd2){
+   one.tail.p<-NULL
+   std.error<-sqrt((popsd1^2/length(x)+popsd2^2/length(y)))
+   z.score<-round((mean(x)-mean(y))/std.error,3)
+   one.tail.p<-round(pnorm(abs(z.score),lower.tail=FALSE),3)
+   cat(" media 1 =",mean(x), " ", "media 2 =", mean(y),"\n",
+   "error estándar =", std.error, "\n",
+ }
```

Los dos primeros argumentos son los vectores de datos, y los últimos dos son las desviaciones estándar de las poblaciones. Se inicializa el vector que contendrá la probabilidad de una cola. Luego se calcula el error estándar de la diferencia de ambas medias y también el puntaje z .

Finalmente, se calcula la probabilidad de una cola.

6.3. Prueba t para dos Muestras

El ejemplo en la sección anterior involucra una situación que se encuentra raramente - varianzas poblacionales conocidas. Si se conoce la varianza poblacional, es probable que se conozca la media. Si se conoce la media, probablemente no sea necesario realizar la prueba de hipótesis.

No conocer las varianzas elimina la posibilidad de usar el teorema del límite central. Esto significa que no se puede usar la distribución normal como una aproximación de la distribución muestral de la diferencia de las medias. En vez de eso, se usa la distribución t . Se debe realizar una prueba t .

Varianzas poblacionales desconocidas llevan a dos posibilidades para la prueba de hipótesis. Una posibilidad es que aunque las varianzas sean desconocidas, se tienen razones para asumir que son iguales. La otra posibilidad es que no se pueden asumir iguales.

6.3.1. Varianzas Iguales

Cuando no se conoce la varianza poblacional, se usa la varianza muestral para estimarla. Si se tienen 2 muestras, el promedio de las dos varianzas muestrales puede considerarse como la estimación.

Unir varianzas muestrales para estimar una varianza poblacional se conoce como agrupamiento. Con 2 varianzas muestrales, se hace de la siguiente manera:

$$s_p^2 = \frac{(N_1-1)s_1^2 + (N_2-1)s_2^2}{(N_1-1) + (N_2-1)}$$

En esta ecuación, s_p^2 representa el estimado de agrupamiento. Hay que notar que el denominador de este estimado es $(N_1 - 1) + (N_2 - 1)$, que son los grados de libertad.

La ecuación para calcular t es

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{s_p \sqrt{\frac{1}{N_1} + \frac{1}{N_2}}}$$

Ejemplo 6.2. La empresa que fabrica microrrobots está intentando elegir entre dos máquinas para producir un componente para su nuevo microrrobot. La velocidad es esencial, por lo que la compañía tiene cada máquina produciendo diez copias del componente, y mide el tiempo de cada producción. Las hipótesis son

$$H_0 : \mu_1 - \mu_2 = 0 \quad (6.9)$$

$$H_1 : \mu_1 - \mu_2 \neq 0 \quad (6.10)$$

Eligen $\alpha = 0,05$. Esta es una prueba de dos colas porque no saben de antemano qué máquina puede ser más veloz.

Obtuvieron que la máquina uno tiene un tiempo de producción medio de 23.00 mientras que la máquina 2 tiene 20.00. Sus desviaciones estándar son 2.71 y 2.79 respectivamente y sus tamaños muestrales son de 10.

El estimado de agrupamiento de σ^2 es

$$s_p^2 = \frac{(N_1 - 1)s_1^2 + (N_2 - 1)s_2^2}{(N_1 - 1) + (N_2 - 1)} = \frac{(10 - 1)(2,72)^2 + (10 - 1)(2,79)^2}{(10 - 1) + (10 - 1)} \quad (6.11)$$

$$= \frac{(9)(2,71)^2 + (9)(2,79)^2}{(9) + (9)} = \frac{66 + 70}{18} = 7,56 \quad (6.12)$$

El estimado de σ es 2.75, la raíz cuadrada de 7.56.

La prueba estadística es

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{s_p \sqrt{\frac{1}{N_1} + \frac{1}{N_2}}} = \frac{(23 - 20)}{2,75 \sqrt{\frac{1}{10} + \frac{1}{10}}} = \frac{3}{1,23} = 2,44$$

Para esta prueba estadística, $df = 18$, el denominador del estimado de varianza. En una distribución t con 18 grados de libertad (df), el valor crítico es 2.10 para la cola del lado derecho (superior) y -2.10 para la cola inferior. El valor calculado de la prueba estadística es mayor que 2.10, por lo que la decisión es rechazar H_0 . Los datos proveen evidencia de que la máquina 2 es significativamente más rápida que la máquina 1.

En R

Aquí hay un par de vectores para la muestra de datos en el ejemplo anterior.

```
> maquina1<-c(24.58, 22.09, 23.70, 18.89, 22.02, 28.71, 24.44,
+ 20.91, 23.83, 20.83)
> maquina2 <- c(21.61, 19.06, 20.72, 15.77, 19, 25.88, 21.48,
+ 17.85, 20.86, 17.77)
```

R provee dos maneras de realizar la prueba t . Ambas involucran la función `t.test()`.

Trabajando con 2 vectores

A continuación se muestra cómo probar hipótesis con dos vectores y suponiendo que ambas varianzas son iguales:

```
> t.test(maquina1,maquina2,var.equal=TRUE,alternative="two.sided",mu=0)
```

Two Sample t-test

```

data:  maquina1 and maquina2
t = 2.4396, df = 18, p-value = 0.02528
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.4164695 5.5835305
sample estimates:
mean of x mean of y
      23      20

```

El argumento *alternative = two.sided* refleja el tipo de hipótesis alternativa especificada en el ejemplo, y el último argumento indica la diferencia hipotética entre medias.

El valor *t* y el valor *p* bajo indican que se puede rechazar la hipótesis nula. La máquina 2 es significativamente más rápida que la máquina 1.

Trabajando con un data frame y una fórmula

La otra manera de llevar a cabo esta prueba es creando un data frame y usando la fórmula *prod.time ~ machine*.

La ecuación expresa la idea de que el tiempo de producción depende de la máquina que se usa. Aunque no es necesario hacer esta prueba de esta manera, es una buena idea para acostumbrarse a las fórmulas.

La primera acción es crear un data frame de formato largo. Primero se crea un vector para las 20 veces de producción. Luego se crea un vector de los nombres de las dos máquinas. Posteriormente se convierte dicho vector en un *vvector* de 10 repeticiones de *maquina1* seguido por otro *vecotr* de 10 repeticiones de *maquina2*.

```

> prod.time<-c(maquina1,maquina2)
> maquina<-c("maquina1","maquina2")
> maquina<-rep(maquina, times=c(10,10))
> robots.frame<-data.frame(maquina,prod.time)
> head(robots.frame)

```

```

  maquina prod.time
1 maquina1    24.58
2 maquina1    22.09
3 maquina1    23.70
4 maquina1    18.89
5 maquina1    22.02
6 maquina1    28.71

```

Finalmente la función *t.test* es

```
> with(robots.frame,t.test(prod.time~maquina, var.equal=TRUE, alternative="two.sided", mu=0))
```

Two Sample t-test

```

data:  prod.time by maquina
t = 2.4396, df = 18, p-value = 0.02528
alternative hypothesis: true difference in means between group maquina1 and group maquina2 is not equal
95 percent confidence interval:
 0.4164695 5.5835305
sample estimates:
mean in group maquina1 mean in group maquina2
      23      20

```

Lo que produce el mismo resultado que la versión de dos vectores.

6.3.2. Varianzas desiguales

El caso de varianzas no iguales representa un reto. Como suele suceder, cuando las varianzas no son iguales, la distribución *t* con $(N_1 - 1) + (N_2 - 1)$ grados de libertad no es una aproximación a la distribución muestral tan cercana como a los estadistas les gustaría.

Los estadistas atacan este reto reduciendo los grados de libertad. Para completar la reducción, usan una ecuación que depende de la desviación estándar muestral y los tamaños de las muestras.

Dado que las varianzas no son iguales, un estimado de agrupamiento no es apropiado. Por lo que se calcula la prueba *t* en una manera distinta:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}}}$$

Se evalúa la prueba estadística contra un miembro de la familia de distribuciones t que tiene los grados de libertad reducidos.

Aquí se muestra lo que produce *t.test* para el ejemplo de los microrrobots si se asume que las varianzas no son iguales:

```
> with(robots.frame, t.test(prod.time~maquina,
+                             var.equal=FALSE,
+                             alternative="two.sided",
+                             mu=0))
```

Welch Two Sample t-test

data: prod.time by maquina

t = 2.4396, df = 17.985, p-value = 0.02529

alternative hypothesis: true difference in means between group maquina1 and group maquina2 is not equal

95 percent confidence interval:

0.4163193 5.5836807

sample estimates:

mean in group maquina1 mean in group maquina2

23

20

Se puede apreciar la ligera reducción de grados de libertad. Las varianzas son muy cercanas que muy poco cambia.