

Estadística 1 para administración

Medición de probabilidad
variables discretas 1

Clase 6



INTRODUCCIÓN DE LA CLASE 6

La distribución normal y T, vistas en clases anteriores corresponden a distribuciones en las cuales las variables han sido numéricas continuas, en las cuales si el valor de la variable aleatoria se incrementa en una cantidad, aunque sea pequeña, se esperaría que su probabilidad de ocurrencia se incremente o decremente, aunque sea en una cantidad mínima.

Pero en la práctica hay muchos otros fenómenos, eventos, o casos en los cuales las variables no son continuas sino discretas y no responden a este comportamiento. En estos casos las fórmulas de probabilidad y distribución vistas anteriormente para el cálculo de PDF o CDF no aplican. En esta clase veremos dos de estos casos conocidos como la distribución binomial y una variante de esta llamada la distribución de Bernoulli.

Resultado o resultados de aprendizaje que será abordado con el contenido de la clase:

Aplicar distribuciones de probabilidad con variables discretas y continuas en la solución de problemas, en distintas áreas del conocimiento, las medidas estadísticas y probabilísticas a partir de la presentación de tablas y gráficos sobre problemáticas de su área de conocimiento

Medición de probabilidad variables discretas 1

Empecemos por definir algunos conceptos:

¿Que un proceso de experimentación?

Definimos los procesos de experimentación como el proceso de medir u observar una actividad con el fin de recopilar datos. Supongamos que nuestro experimento de interés implica que un jugador de football profesional lanza tiros libres al arco. Cada tiro libre se consideraría un ensayo para el experimento. Para este experimento en particular, solo tenemos dos resultados posibles para cada ensayo, o el tiro libre entra o no entra al arco. A estos resultados se conocen como un éxito o fracaso. A este tipo de experimento llamaremos experimento binomial. Por otro lado, la distribución multinomial representa las probabilidades de varios posibles resultados a través de múltiples experimentos.

¿Qué es un ensayo en distribución binomial?

Los resultados de la distribución binomial se denominan como éxito o fracaso. La palabra éxito no necesariamente significa un resultado positivo. Es solo el resultado que nos interesa evaluar. Del mismo modo, la palabra fracaso no necesariamente significa un resultado negativo es solo un resultado que no nos interesa, o que no está mencionado explícitamente en los enunciados de los problemas.

Veamos algunos ejemplos:

- ✓ Comprobar si una parte manufacturada es defectuosa
- ✓ Observar la cantidad de defectos de control de calidad de producción

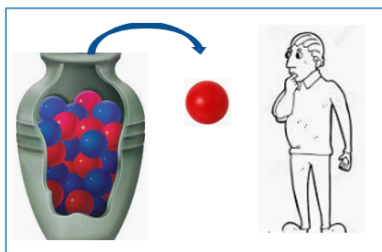
Distribución binomial

Por otro lado, decimos que la distribución binomial es una distribución con reemplazo. El concepto es el siguiente. Supongamos que se tiene un ánfora llena de bolas rojas y azules, y tenemos una persona sacando una bola a la vez. Intuitivamente diríamos que, si ya ha retirado 3 bolas rojas, la probabilidad de volver a sacar más bolas rojas disminuye. Este sería un ejemplo de una distribución sin sustitución. Por otro lado, una distribución en la que no importa la cantidad de bolas rojas o azules ya extraídas previamente, si la probabilidad de volver a obtener el mismo color no cambia, se dice que es “con reemplazo”.

Figura 1

Distribución binomial con reemplazo.

Creación del autor Alfonso Prado



De acuerdo con Webster(2020) Un experimento binomial tiene las siguientes características:

- ✓ El experimento consiste en un número fijo de ensayos denotado por n , este es un valor entero mayor que 0

- ✓ Cada ensayo tiene sólo dos resultados posibles, un éxito o un fracaso
- ✓ La probabilidad de éxito y la probabilidad de fracaso son constantes durante todo el experimento y es denotado por π
- ✓ Cada experimento indica la cantidad de éxitos que se desea obtener, denotado por x

Funciones matemáticas de la distribución binomial

Supongamos que la probabilidad de aprobar un examen de estadística es solo el 60%, por lo que la probabilidad de reprobado es del 40 por ciento. Esto representa un experimento binomial, con $p = 0,60$ (la probabilidad de un “éxito”) y $q = 0,40$ (la probabilidad de un “fracaso”).

Podemos calcular la probabilidad puntual de obtener x éxitos en n ensayos utilizando la función PMF como se muestra en la siguiente figura

Figura 2

Fórmula PMF de la distribución binomial Prado A. (2025)

Adaptación de la fórmula para la distribución binomial según Webster (2000) capítulo 5

$$P(x) = \frac{n!}{(n-x)!x!} p^x q^{n-x}$$

[Enlace relacionado](#)

[Explicación de la fórmula binomial](#)

Vale la pena recordar para este cálculo que $0!=1$ y que $x^0=1$

Así como en la clase 2 se mostró como obtener la media de un vector de datos, también se puede determinar la media de una distribución de probabilidad. La media aritmética de una distribución de probabilidad se llama el valor esperado $E(X)$, y se halla multiplicando n por la probabilidad como se muestra en la siguiente figura.

Figura 3

Media de la distribución de probabilidad discreta Prado A. (2025)

Adaptación de la fórmula para la media en distribución binomial según Webster (2000) capítulo 5

$$\mu = E(X) = n\pi$$

Donde:

$E(X)$ es el valor esperado o media

n es la cantidad de ensayos

π es la probabilidad de obtener una respuesta afirmativa

Y la varianza de la distribución se muestra en la siguiente figura

Figura 4

Varianza de la distribución binomial Prado A. (2025)

Adaptación de la fórmula para la varianza en distribución binomial según Webster (2000) capítulo 5

$$\sigma^2 = n\pi(1 - \pi)$$

Generalmente es mejor visualizar la probabilidad para entender la distribución de los datos. A diferencia de las distribuciones continuas en las que preferíamos representarlo mediante un diagrama de densidad, para las distribuciones binomiales usaremos un histograma.

Veamos un ejemplo concreto:

De acuerdo con un estudio de educación universitarias, el 40% de los estudiantes trabajan durante el verano para ganar dinero para su colegiatura del siguiente semestre.

Si 7 estudiantes se seleccionan de manera aleatoria, cual es la probabilidad de que 1,2,3,4,5,6 o 7 estudiantes trabajen.

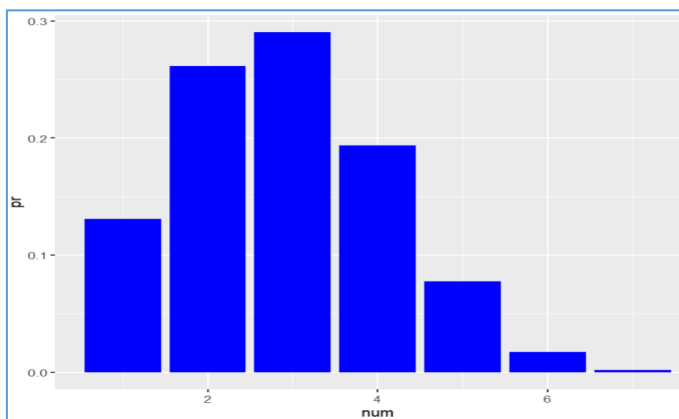
La solución consiste en encontrar 7 probabilidades binomiales para cada uno de los valores con lo cual obtenemos el siguiente resultado. Note que la suma de todas las

probabilidades es siempre igual a 1, lo cual es un requisito para todas las distribuciones de probabilidad.

Figura 5

Solución del problema

Creación autor Alfonso Prado



Por otro lado, es interesante mencionar como se aplica el teorema del límite central a este tipo de distribuciones. Recordando el teorema establece que a medida que n se vuelve más grande, la distribución de las medias muestrales se aproxima a una distribución normal con una media de μ y un error estándar de σ/\sqrt{n} . Para esto graficaremos 3 distribuciones binomiales con valores de cantidad de éxitos de 10, 100 y 1000, con una probabilidad $\pi=0.3$ y graficamos. Notaremos como la distribución se va aproximando a una distribución gaussiana.

El siguiente código nos permite visualizar la aplicación del teorema del limite central.

```
#Aproximación a la distribución normal  
  
df_10 <- data.frame(Exitos = rbinom(n=10000, size =10 ,prob=.3 ),Size=10)  
df_10  
df_100 <- data.frame(Exitos = rbinom(n=10000, size =100 ,prob=.3 ),Size=100)  
df_100  
df_1000 <- data.frame(Exitos = rbinom(n=10000, size =1000 ,prob=.3 ),Size=1000)  
df_1000
```

```

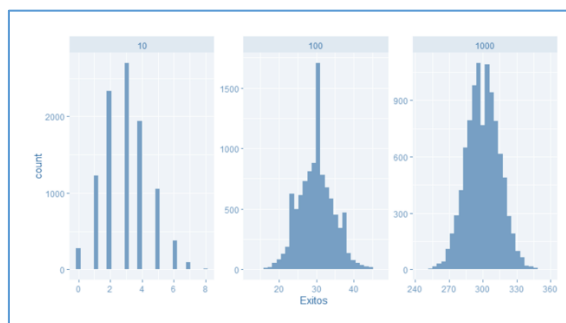
todo<- rbind(df_10, df_100, df_1000)
head(todo)
tail(todo)
ggplot (data=todo , aes(x=Exitos))+ geom_histogram() +
facet_wrap(~Size, scales="free")

```

Figura 6

Demostración del teorema del límite central en distribución binomial

Creación autor Alfonso Prado



Un caso especial de la distribución binomial es cuando $n=1$, esto es conocido como el ensayo de Bernoulli, si solo hay 1 ensayo con probabilidad de éxito p y probabilidad de fracaso $1-p$, esto se llama distribución de Bernoulli.

La siguiente figura muestra la probabilidad de éxito y fracaso derivadas de las fórmulas anteriores

Figura 7

Probabilidad de éxito (a) y fracaso (b) en un ensayo de Bernoulli Prado A. (2025)

Adaptación de la fórmula de distribución binomial para un único ensayo según Webster (2000) capítulo 5

<p>(a)</p> $P(X = 1) = \binom{1}{1} p^1 (1 - p)^{1-1} = p$	<p>(b)</p> $P(X = 0) = \binom{1}{0} p^0 (1 - p)^{1-0} = 1 - p$
------------------------------------------------------------	----------------------------------------------------------------

Donde:

$C(1,1) N$ representa la cantidad de experimentos de Bernoulli

n representa la cantidad de éxitos esperados

$P=\pi$

Funciones programáticas de la distribución binomial

De igual manera que vimos en la distribución normal, para el caso binomial tenemos las funciones que nos ayudan a calcular los resultados de un experimento puntual, que sería el PMF, o resultados acumulativos CDF. También de la misma forma la nomenclatura sigue los mismos lineamientos de la distribución normal, solo que en este caso sustituiremos el sufijo de norm por binom.

Enlace relacionado

Describe en detalle el uso de la funciones programáticas

Enlace

Tabla 1

Funciones programáticas en R

Creación del autor Alfonso Prado

Función de Distribución	Distrib. Binomial	
PMF	dbinom()	dbinom(x, size, prob, log = FALSE)
CDF	pbinom()	pbinom(q, size, prob, lower.tail = TRUE, log.p = FALSE)
Inverso CDF	qbinom()	qbinom(p, size, prob, lower.tail = TRUE, log.p = FALSE)
Función para obtener un set de datos de la distribución	Distrib. Binomial	
	rbinom()	rbinom(n, size, prob)

Donde:

x, q vector de cuantiles.

p vector de probabilidades.

n número de observaciones.

$prob$ probabilidad de éxito en cada prueba

size número de pruebas.

lower.tail TRUE (default), probabilidad de $P [X \leq x]$, FALSE $P [X > x]$

Función dbinom

Esta función devuelve la densidad de probabilidad (PMF) de la distribución binomial dada un determinado valor que representa que el número de resultados positivos que se desearía obtener, argumento x , en base a una cierta cantidad de ensayos dado por el argumento “size”, y cuando la probabilidad de éxito en cada ensayo es conocida y constante dado por el argumento “prob” .

Veamos un ejemplo

```
#Supongamos que los ítems producidos en una fábrica tienen una probabilidad de 0.005
#de ser defectuosos. Pero estos ítems se envían en cajas de cartón con 25 unidades
#La fábrica produce 1500 ítems diarios. ¿Cuál es la probabilidad de que una caja de
cartón elegida al azar contenga exactamente un ítem defectuoso?
#En este caso el valor de 1500 ítems no tiene ninguna importancia, porque se está
preguntando por caja

dbinom(x=1, size=25, prob=0.005)
[1] 0.1108317
```

Función pbinom

Esta función devuelve el valor de la función de densidad acumulada (CDF) de la distribución binomial dada una determinada variable aleatoria q que representa el valor discreto debajo del cual o sobre el cual se desea acumular las probabilidades puntuales. El número de ensayos dado por el argumento “size” y con una probabilidad de éxito en cada ensayo constante y definida por el argumento “prob”. A fin de determinar si la probabilidad se acumula hacia la cola izquierda o cola derecha, utilizaremos el argumento lower.tail. lower.tail es una variable lógica, si esta toma el valor de verdadero (TRUE)

indicaría que la acumulación es desde el valor discreto q hacia abajo ($P [X \leq x]$), caso contrario la acumulación es $P([X >x])$.

Veamos un ejemplo

```
#Un fabricante de discos USB externos para computadora entrega su producto en cajas
de 20 discos. La tasa de fallos en los discos es del 10% . se desea saber cuál es la
probabilidad de que 2 o más de los discos estén defectuosos y asumiendo esa
probabilidad cual variación que el cliente esperaría encontrar de una caja a otra?
n=20
prob=.1
x=3
p<- pbinom(x,n,prob=prob, lower.tail=FALSE)
var= 20*p*(1-p)
var
[1] 2.305535
```

Función qbinom

Esta función devuelve el valor de la función de densidad acumulativa inversa (CDF) de la distribución binomial dada una determinada variable aleatoria p que representa la probabilidad acumulada, con un número de ensayos definido por el argumento “size” y con probabilidad de éxito en cada ensayo dada por el argumento “prob”. El valor que devuelve es el p -ésimo cuantil de la distribución binomial, es decir el valor debajo del cual o sobre el cual se obtiene una probabilidad acumulada p . El argumento `lower.tail` es también usado en esta función y tiene igual significado que lo mencionado arriba.

Veamos un ejemplo

```
#Sobre el mismo ejemplo anterior de la fábrica
#Cuál es la cantidad de ensayos que tengo que hacer para obtener una probabilidad de
30% ,teniendo un universo=10 y cuando cada prueba tiene una probabilidad del 40%
qbinom(p=.3 , size=10 ,prob=.4)
[1] 3
```

Función rbinom

Esta función genera un vector de variables aleatorias distribuidas binomialmente dada una longitud de vector n , una cantidad de ensayos dada por el argumento “size” y una probabilidad de éxito en cada ensayo “prob”.

Veamos un ejemplo:

```
#Suponga que está a cargo del QC de una fábrica. La fábrica hace 150 ítems por día.
#Los ítems defectuosos deben ser re elaborados. Sabemos que hay una tasa de error
histórico del 5%. Queremos estimar cuantos ítems necesitaremos arreglar cada día esta
semana (laboral).
#Usamos rbinom para generar una serie con estas características, porque rbinom y no
rnorm, bueno el enunciado es que ya sea el item o está bien manufacturado o no, esto
hace que usemos la distribución binomial

rbinom (n=5 , size=150 , prob=.05)
#por que n=5? Es para poner en la misma unidad, si fabricamos 150 por semana,
queremos 5 muestras en la semana para tener una diaria, cada muestra representará la
cantidad de pruebas con ítems malos.
size y prob son obvios

#Veamos casos extremos(tendríamos que reparar casi todos)
rbinom (n=5 , size=150 , prob=.9)
#Para n=1 solo una prueba que es la distribución de Bernoulli
rbinom (n=1 , size=10 , prob=.5)
```

Visualización de la probabilidad acumulativa

La mejor forma de visualizar estas probabilidades acumulativas es a través de la función `geom_ribbon`. El siguiente código explica cómo lograrlo en base a las probabilidades obtenidas de la figura 5.

```
#Primero debemos obtener un dataset que contenga el valor de x y su probabilidad
```

```

df0 <- data.frame(x=NULL , prob=NULL)
for ( a in seq(from=0, to=7, by=1))
{
  print(a)
  prob= dbinom(a,7,0.4)
  dfline=data.frame(x=a, prob=prob)
  df0 <- rbind(df0, dfline)

}
df0
#La vista del dataset nos presenta las probabilidades puntuales para cada valor de x
#Ahora nos interesa saber la probabilidad acumulada para el valor de 4, implícitamente
se estaría usando lower.tail=TRUE
  x  prob
1 0 0.0279936
2 1 0.1306368
3 2 0.2612736
4 3 0.2903040
5 4 0.1935360
6 5 0.0774144
7 6 0.0172032
8 7 0.0016384

pbinom(4,7,0.4, lower.tail=TRUE)
[1] 0.903744
#Creamos un subset de las probabilidades obtenidas

dfsub <- subset(df0, x>= 0 & x<= 4)
View(dfsub)
  x  prob
1 0 0.0279936
2 1 0.1306368
3 2 0.2612736

```

```
4 3 0.2903040
```

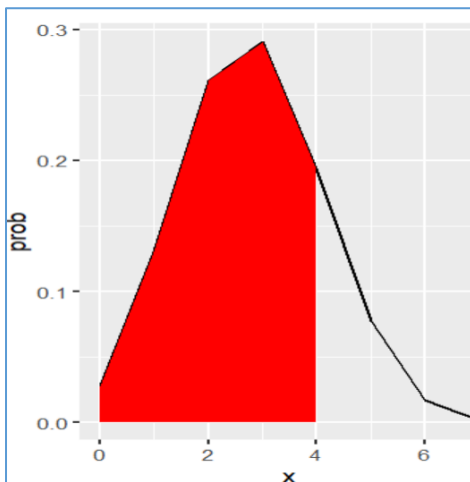
```
5 4 0.1935360
```

```
ggplot(data=df0, aes(x=x, y=prob)) +geom_line()+  
geom_ribbon(data=dfsub, aes(ymax=prob),ymin=0,fill="RED")
```

Figura 8

Probabilidad acumulativa en distribución binomial de obtener un número de éxitos menor o igual a 4 de 7 ensayos

Creación del autor Alfonso Prado



En otros casos se busca conocer la probabilidad acumulativa para un rango de valores, por ejemplo: De acuerdo con un estudio de educación universitarias, el 40% de los estudiantes trabajan durante el verano para ganar dinero para su colegiatura del siguiente semestre.

Si 7 estudiantes se seleccionan de manera aleatoria, cual es la probabilidad de que entre 3 y 5 estudiantes trabajen. Esto lo podemos mediante

una diferencia de probabilidades binomiales.

```
#Primero debemos obtener un dataset que contenga el valor de x y su probabilidad  
df0 <- data.frame(x=NULL , prob=NULL)  
for ( a in seq(from=0, to=7, by=1))  
{  
  print(a)  
  prob= dbinom(a,7,0.4)  
  dfline=data.frame(x=a, prob=prob)  
  df0 <- rbind(df0, dfline)  
}
```

```

df0
#La vista del dataset nos presenta las probabilidades puntuales para cada valor de x
#Ahora nos interesa saber la probabilidad acumulada para el valor de 4, implícitamente
se estaría usando lower.tail=TRUE
x  prob
1 0 0.0279936
2 1 0.1306368
3 2 0.2612736
4 3 0.2903040
5 4 0.1935360
6 5 0.0774144
7 6 0.0172032
8 7 0.0016384

dfsub <- subset(df0, x>= 3 & x<= 5)
dfsub7 <- subset(df0, x>= 5 )
View(dfsub)
ggplot(data=df0, aes(x=x, y=prob)) +geom_line()+
  geom_ribbon(data=dfsub, aes(ymax=prob),ymin=0,fill="RED")+
  geom_ribbon(data=dfsub7,aes(ymax=prob),ymin=0,fill="BLUE")

#podemos validarlo
m3 <- pbinom(2,7,0.4, lower.tail=FALSE)
m5<- pbinom(5,7,0.4, lower.tail=FALSE)
entre3_5 = m3-m5
entre3_5
[1] 0.5612544

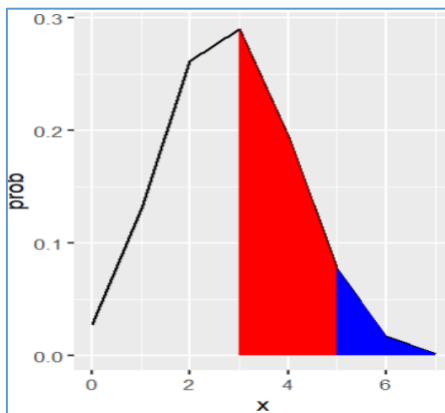
#Del grafico podemos concluir que lo que la probabilidad de que entre 3 y 5 estudiantes
trabajen durante el verano lo obtenemos mediante la resta de probabilidades de más de
3 estudiantes menos la probabilidad de más de 5 estudiantes. Según se muestra en la
siguiente figura

```

Figura 9

Visualización de rango de probabilidades binomiales del problema anterior

Creación del autor Alfonso Prado



REFERENCIAS CITADAS

Webster. A. , (2017), Estadística aplicada a los negocios y economía , Irwin- McGraw-HILL

Rumsay D.(2009), Statistics II For Dummies, Wiley Publishing Inc

The Data Visualisation Catalogue, <https://datavizcatalogue.com/>



La excelencia no se improvisa

síguenos

