Variables aleatorias discretas en R

Sesión 7

Natalie Julian - www.nataliejulian.com

Estadística UC y Data Scientist en Zippedi Inc.

Contexto de una variable aleatoria

- **Experimento aleatorio** Etapa de observación de un fenómeno, cuyo resultado es incierto
- **Suceso elemental** Posible resultado del experimento aleatorio
- **Espacio muestral** Colección de todos los posibles valores de un experimento
- **Suceso** Subconjunto del espacio muestral, conjunto de sucesos elementales del experimento (posibles resultados)

Variables aleatorias discretas

Una variable aleatoria discreta representa la información de ocurrencia de uno o más eventos estudiados en un lapso de tiempo o en distintas realizaciones/ensayos.

Variables aleatorias discretas

Una variable aleatoria discreta X con posibles valores $\{x_1, x_2, ..., x_k\}$ se define a través de una función de probabilidad $p_i = P(X = x_i) \;\; i = 1, ...,$. Además, utilizando la función de probabilidad puede definirse su función de distribución acumulada:

Valor puntual	x_i	x_1	x_2	......	x_k
Función de probabilidad	$P(X = x)$	$P(X = x_1)$	$P(X = x_2)$	......	$P(X = x_k)$
Función de distribución	$\sum_{j=1}^{i} P(X = x_j)$	$\sum_{j=1}^{1} P(X = x_j)$	$\sum_{j=1}^{2} P(X = x_j)$	......	$\sum_{j=1}^{k} P(X = x_j)$

Ejemplos de variables aleatorias discretas

Algunas variables aleatorias discretas que son de particular interés modelar y estudiar en la actualidad:

- Acierto o no acierto de un jugador de Pasapalabra en una letra determinada
- Cantidad de matches que recibe una persona en Tinder
- Número de victorias de un equipo en un campeonato de League of Legends
- Cantidad de contagios de COVID-19 en un día
- Cantidad de meteoritos que cruzan la atmósfera terrestre por año

Generación de muestras aleatorias a través de su FDP

Utilizando la función de probabilidad de una variable aleatoria es posible obtener simulaciones independientes, es decir, muestras aleatorias de dicha distribución.

Distribución	Muestra aleatoria	Función de probabilidad	Función de distribución
Bernoulli(p)	`rbinom(n= ,size=1,prob=p)`	`dbinom(x= ,size=1,prob=p)`	`pbinom(q= ,size=1,prob=p)`
Binomial(m,p)	`rbinom(n= ,size=m,prob=p)`	`dbinom(x= ,size=m,prob=p)`	`pbinom(q= ,size=m,prob=p)`
Geométrica(p)	`rgeom(n= ,prob=p)`	`dgeom(x= ,prob=p)`	`pgeom(q= ,prob=p)`
Bineg(m,p)	`rnbinom(n= ,size=m,prob=p)`	`dnbinom(x= , size=m,prob=p)`	`pnbinom(q= ,size=m,prob=p)`
Poisson(λ)	`rpois(n= ,lambda)`	`dpois(x= ,lambda)`	`ppois(q= ,lambda)`
Hipgeo(m,n,k)	`rhyper(nn= ,m,n,k)`	`dhyper(x= ,m,n,k)`	`phyper(q= ,m,n,k)`

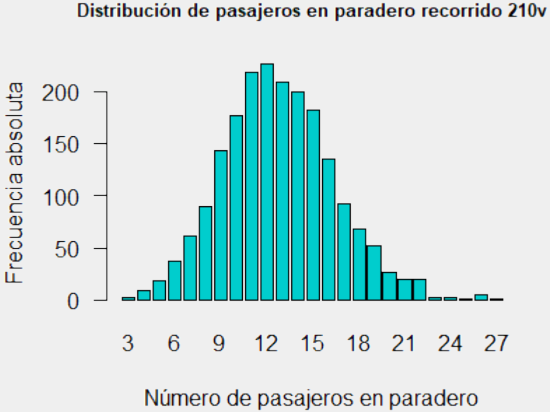
Ejemplo Poisson: Cantidad de pasajeros en un paradero

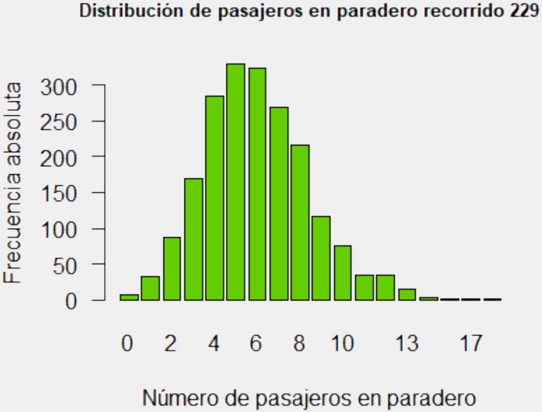
a) Sea X la cantidad de pasajeros que toman micro en el recorrido 210v del transantiago en un paradero en particular durante una hora. Se realizó un estudio, y se cree que $X \sim Poisson(13)$. Obtenga una muestra aleatoria de tamaño 40 de la cantidad de pasajeros que toman micro en el recorrido 210 del transantiago en un paradero en particular durante una hora.

b) En el mismo paradero, se midió además la cantidad de pasajeros que toman micro en el recorrido 229 en una hora, y se cree que $Y \sim Poisson(6)$. Obtenga una muestra aleatoria de X e Y de tamaño 2000. Muestre en un gráfico cómo se distribuye la cantidad de pasajeros en el paradero en una hora para el recorrido 210v y en otro gráfico, el recorrido 229.

c) Defina la suma de pasajeros como $Z = X + Y$, asumiendo que X e Y son independientes, ¿qué distribución tendría Z? ¿Con qué parámetro?

Efecto del parámetro λ



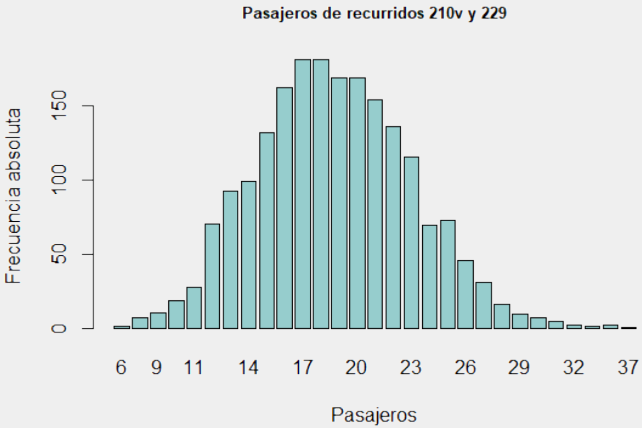



```
c(mean(muestraX), var(muestraX), mean(muestraY), var(muestraY))
[1] 12.785000 13.310430  5.971000  6.166242
```

Ejemplo Poisson: Cantidad de pasajeros en un paradero

c) Defina la suma de pasajeros como $Z = X + Y$, asumiendo que X e Y son independientes, ¿qué distribución tendría Z? ¿Con qué parámetro?

$Z = X + Y \sim Poisson(\lambda_1 + \lambda_2)$

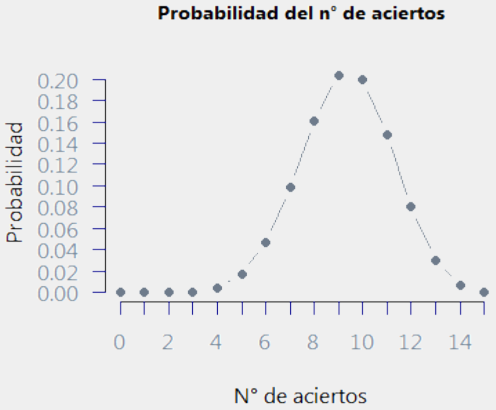


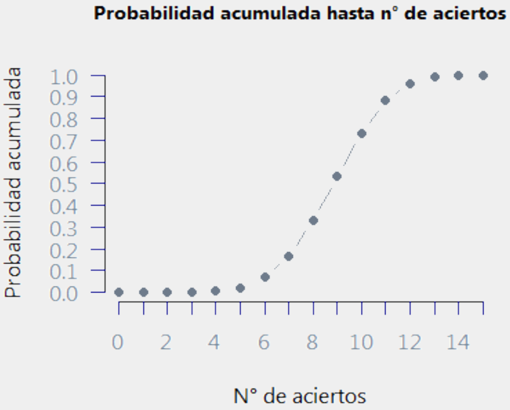

```
mean(Z);var(Z)
[1] 18.756
[1] 18.92893
```

Ejemplo Binomial: Aciertos en cuestionario de IQ

a) Un cuestionario de IQ contiene 15 preguntas. Definase el éxito como acertar en la pregunta. Si asume independencia entre las preguntas, se puede pensar que el puntaje total (suma de aciertos) distribuye binomial(15,p) pues cada acierto corresponde a un ensayo Bernoulli(p). Muestre en un gráfico la probabilidad de tener desde 0 aciertos hasta 15 aciertos considerando como probabilidad de acierto p=0.62.

Efecto del parámetro p





Ejemplo Binomial: Aciertos en cuestionario de IQ

b) Suponga que 73 personas rindieron el test. A través de variables aleatorias Bernoulli simule el puntaje total para las 73 personas usando p=0.62. Obtenga el promedio de los puntajes. Comente.

$E(Puntaje) = n \cdot p$

```
N=73

puntajes<-rep(NA, N)

for(i in 1:N){
  puntajes[i]<-sum(rbinom(n=15, size=1,prob=0.62))
}

puntajes
[1] 10 11  7  9 11  8  9  8  8 10 10 11  8 10  7  8  7  6 11
[20] 12  8 11 10 11 11 10  8  7 10  9 11 13 12 12 10 11 10 12
[39]  7 10 10 11 10 10 12  8 13  6 11 11  7 10  8  8  7  9  6
[58]  9 10  6 10 11 10  9 11  8 11  7 10 11 10 11  6

mean(puntajes); var(puntajes)
[1] 9.465753
[1] 3.280061

15*0.62; 15*0.62*(1-0.62)
[1] 9.3
[1] 3.534
```

Ejemplo hipergeométrica: Número de productos defectuosos

a) En la práctica de control de calidad, solo una parte de resultados del proceso de producción es muestreado y examinado, ya sea por el tiempo empleado en hacer esto o ya que la examinación puede resultar en un daño en el producto. Supongamos que existen n productos en un lote de producción el cual contiene k objetos defectuosos. Se toma una muestra de tamaño r de este lote. ¿Cuál es la probabilidad de que la muestra contenga exactamente m objetos defectuosos?

Ejemplo hipergeométrica: Número de productos defectuosos

a) En la práctica de control de calidad, solo una parte de resultados del proceso de producción es muestreado y examinado, ya sea por el tiempo empleado en hacer esto o ya que la examinación puede resultar en un daño en el producto. Supongamos que existen n productos en un lote de producción el cual contiene k objetos defectuosos. Se toma una muestra de tamaño r de este lote. ¿Cuál es la probabilidad de que la muestra contenga exactamente m objetos defectuosos? La probabilidad pedida corresponde a:

$$P(m) = \frac{\binom{k}{m}\binom{n-k}{r-m}}{\binom{n}{r}}.$$

Donde $P(m)$ corresponde a la probabilidad de tener exactamente m productos defectuosos, con m=1,...,k. k corresponde a la cantidad de objetos defectuosos. n corresponde al tamaño de la población completa (todos los productos). r corresponde al tamaño de la muestra (cantidad de productos en el lote examinado).

Ejemplo hipergeométrica: Número de productos defectuosos

b) Suponga que en un lote de producción hay 1000 productos de los cuales 60 son defectuosos y se toma un muestra de tamaño 200 del lote. Calcule la probabilidad de que hayan $m = 0, 1, 2, ..., 60$ productos defectuosos. Realícelo de forma manual y luego con la función correspondiente. Muestre en un gráfico las probabilidades obtenidas.

$E(X) = \frac{k \cdot r}{n}$