Universidade Vigo

Departamento de Estatística e Investigación Operativa

Grao en Intelixencia Artificial
E.S.E. Informática
Ed. Politécnico
Campus de Ourense

Tel. 986 387 000
http://esei.uvigo.es

Práctica de Variables Aleatorias con R
Materia - Estatística. Curso 2022/2023

1 Introducción

El objetivo de esta práctica es calcular probabilidades, cuantiles, ..., características de los modelos de probabilidad, tanto discretos como continuos, vistos en clase y con R.

En R, las funciones relacionadas con las variables aleatorias llevan siempre la misma estructura, por ejemplo, si deseamos trabajar con la v.a. *Binomial*, al nombre *binom* tenemos que incluir el prefijo:

- *d* si deseamos obtener la función masa de probabilidad (*dbinom*),
- *p* si deseamos obtener la función de distribución $\longleftrightarrow$ masa de probabilidad acumulada (*pbinom*),
- *q* si deseamos obtener los cuantiles (*qbinom*),
- *r* si deseamos obtener valores aleatorios (*rbinom*).

2 Listado de funciones de R

Distribución	Probabilidad	Prob Acumulada	Cuantil	Valor aleatorio
Binomial	dbinom(x,size,prob)	pbinom(x,size,prob)	qbinom(p,size,prob)	rbinom(n,size,prob)
Bin. Negativa	dnbinom(x,size,prob)	pnbinom(x,size,prob)	qnbinom(p,size,prob)	rnbinom(n,size,prob)
Poisson	dpois(x,λ)	ppois(x,λ)	qpois(p,λ)	rpois(n,λ)
Uniforme	dunif(x,a,b)	punif(x,a,b)	qunif(p,a,b)	runif(n,a,b)
Exponencial	dexp(x,λ)	pexp(x,λ)	qexp(p,λ)	rexp(n,λ)
Normal	dnorm(x,μ,σ)	pnorm(x,μ,σ)	qnorm(p,μ,σ)	rnorm(n,μ,σ)
χ^2	dchisq(x,df)	pchisq(x,df)	qchisq(p,df)	rchisq(n,df)
$t-Student$	dt(x,df)	pt(x,df)	qt(p,df)	rt(n,df)
F	df(x,df1,df2)	pf(x,df1,df2)	qf(p,df1,df2)	rf(n,df1,df2)

Table 1: Listado de funciones de R

3 Variables aleatorias discretas

1. Sea X una variable aleatoria $Bi(25, 0.5)$, calcular:

   (a) Describe la variable aleatoria anterior, i.e. responde a: ¿qué tipo de v.aelatoria es? y ¿qué valores puede tomar con probabilidad positiva?

   (b) La función *dbinom(x)* devuelve la probabilidad de que la v.a. tome ese valor x. Comprueba que esta función de probabilidad es una función bien definida.

   (c) Representa gráficamente la función de probabilidad. Que conclusiones podemos sacar sobre la media, mediana y moda.

   (d) $P(X > 12)$.

   (e) $P(12 \leq X \leq 20)$.

   (f) $P(X \leq 20/X > 12)$; $P(X \leq 20/X > 19)$. Representa y compara la función de probabilidad de la variable original y la v. condincionada.

   (g) Los apartados anteriores usando la probabilidad individual y usando la probabilidad acumulada.

   (h) Los cuartiles y los deciles. A la vista de los deciles, ¿hai algún punto que tenga una probabilidad superior o igual a 0.1? Justifica la respuesta.

   (i) ¿Cuál es el valor mínimo que deja un *33.3%* de la distribución a su derecha?.

   (j) El rango Inter-cuartilico: $Q_3 - Q_1$. Compáralo con la varianza teórica.

(k) Genera una muestra aleatoria de tamaño 500.

(l) Calcula su media, varianza, mediana, moda teóricamente (sin usar las funciones de R).

(m) Calcula la media, varianza, mediana y moda de la muestra aleatoria obtenida anteriormente. Compara los valores.

(n) Genera, nuevamente, una muestra aleatoria de tamaño 10000. Calcula los valores resumen.

2. Sea X una variable aleatoria $Bi(100, 0.01)$, calcular:

   (a) $P(X > 12)$.

   (b) $P(12 \leq X \leq 20)$

   (c) Los cuartiles y los deciles.

   (d) Calcula nuevamente los valores anteriores pero usando la aproximación de la binomial a la poisson.

   (e) A través de un gráfico compara las probabilidades de la distribución binomial (eje X) con las probabilidades de la distribución de poisson (eje Y) correspondiente. ¿Como sería el gráfico si la aproximación entre estas dos distribuciónes fuese perfecta?

3. Sea X una variable aleatoria $BN(10, 0.3)$, calcular:

   (a) Probabilidad de que el número de fallos sea menor de 10, es decir, $P(X < 10)$.

   (b) $P(X \leq 20/X \geq 12)$; $P(X \leq 8)$.

   (c) Los apartados anteriores usando la probabilidad individual y usando la probabilidad acumulada.

   (d) Cuantos intentos tengo que realizar para asegurarme con un 95% que obtengo los 10 exitos.

   (e) Los cuartiles y los deciles.

   (f) Genera una muestra aleatoria de tamaño 500.

   (g) Calcula su media, varianza, mediana, moda teóricamente (sin usar R).

   (h) Calcula la media, varianza, mediana y moda de la muestra aleatoria obtenida anteriormente. Compara los valores.

   (i) Genera, nuevamente, una muestra aleatoria de tamaño 10000. Calcula los valores resumen.

   (j) Representa gráficamente la distribución de frecuencias de los 10000 datos anteriores. Compara con la distribución teórica de la variable.

4. Sea X una variable aleatoria $Pois(\lambda = 10)$, calcular:

   (a) $P(X < 10)$.

   (b) $P(X \leq 20/X \geq 12)$; $P(X \leq 8)$.

   (c) Los apartados anteriores usando la probabilidad individual y usando la probabilidad acumulada.

   (d) Los cuartiles y los deciles.

   (e) Genera una muestra aleatoria de tamaño 500.

   (f) Calcula su media, varianza, mediana, moda teóricamente (sin usar R).

   (g) Calcula la media, varianza, mediana y moda de la muestra aleatoria obtenida anteriormente. Compara los valores.

   (h) Genera, nuevamente, una muestra aleatoria de tamaño 10000. Calcula los valores resumen.

   (i) Representa gráficamente la distribución de frecuencias de los 10000 datos anteriores. Compara con la distribución teórica de la variable.

   (j) Comparar gráficamente, la probabilidad de los 40 primeros valores de la distribución de Poisson con su aproximación por la distribución normal.

Universidade Vigo

Departamento de Estatística e Investigación Operativa

Grao en Intelixencia Artificial
E.S.E. Informática
Ed. Politécnico
Campus de Ourense

Tel. 986 387 000
http://esei.uvigo.es

Práctica de Variables Aleatorias con R
Materia - Estatística. Curso 2022/2023

1 Introducción

El objetivo de esta práctica es calcular probabilidades, cuantiles, ..., características de los modelos de probabilidad, tanto discretos como continuos, vistos en clase y con R.

En R, las funciones relacionadas con las variables aleatorias llevan siempre la misma estructura, por ejemplo, si deseamos trabajar con la v.a. *Binomial*, al nombre *binom* tenemos que incluir el prefijo:

- *d* si deseamos obtener la función masa de probabilidad (*dbinom*),
- *p* si deseamos obtener la función de distribución $\longleftrightarrow$ masa de probabilidad acumulada (*pbinom*),
- *q* si deseamos obtener los cuantiles (*qbinom*),
- *r* si deseamos obtener valores aleatorios (*rbinom*).

2 Listado de funciones de R

Distribución	Probabilidad	Prob Acumulada	Cuantil	Valor aleatorio
Binomial	dbinom(x,size,prob)	pbinom(x,size,prob)	qbinom(p,size,prob)	rbinom(n,size,prob)
Bin. Negativa	dnbinom(x,size,prob)	pnbinom(x,size,prob)	qnbinom(p,size,prob)	rnbinom(n,size,prob)
Poisson	dpois(x,λ)	ppois(x,λ)	qpois(p,λ)	rpois(n,λ)
Uniforme	dunif(x,a,b)	punif(x,a,b)	qunif(p,a,b)	runif(n,a,b)
Exponencial	dexp(x,λ)	pexp(x,λ)	qexp(p,λ)	rexp(n,λ)
Normal	dnorm(x,μ,σ)	pnorm(x,μ,σ)	qnorm(p,μ,σ)	rnorm(n,μ,σ)
χ^2	dchisq(x,df)	pchisq(x,df)	qchisq(p,df)	rchisq(n,df)
$t-Student$	dt(x,df)	pt(x,df)	qt(p,df)	rt(n,df)
F	df(x,df1,df2)	pf(x,df1,df2)	qf(p,df1,df2)	rf(n,df1,df2)

Table 2: Listado de funciones de R

3 Variables aleatorias continuas

1. Sea X una variable aleatoria $U(-1,1)$, calcular:
   (a) $P(X < 0.1)$.
   (b) $P(X \leq 0.2/X \geq 0)$.
   (c) Los cuartiles y los deciles.
   (d) Genera una muestra aleatoria de tamaño 500.
   (e) Calcula su media, varianza, mediana, moda teóricamente (sin usar R).
   (f) Calcula la media, varianza, mediana y moda de la muestra aleatoria obtenida anteriormente. Compara los valores.
   (g) Genera, nuevamente, una muestra aleatoria de tamaño 10000. Calcula los valores resumen.
   (h) Representa gráficamente la distribución de frecuencias de los 10000 datos anteriores. Compara con la distribución teórica de la variable.
2. Sea X una variable aleatoria exponencial de medida 10, calcular:
   (a) $P(X < 10)$.
   (b) $P(X \leq 20/X \geq 12)$; $P(X \leq 8)$. ¿Que conclusión podemos sacar de esta igualdad?.

(c) Los cuartiles y los deciles.
(d) Genera una muestra aleatoria de tamaño 500.
(e) Calcula su media, varianza, mediana, moda teóricamente (sin usar R).
(f) Calcula la media, varianza, mediana y moda de la muestra aleatoria obtenida anteriormente. Compara los valores.
(g) Genera, nuevamente, una muestra aleatoria de tamaño 10000. Calcula los valores resumen.
(h) Representa gráficamente la distribución de frecuencias de los 10000 datos anteriores. Compara con la distribución teórica de la variable.

3. Sea X una variable aleatoria $N(\mu = 10, \sigma = 2)$, calcular:

(a) $P(X > 10)$ y $P(X \geq 10)$. Justifica la igualdad de estos resultados.
(b) $P(X < 13.91993)$. Sea $Y = \frac{X-10}{2}$, calcular su media y desviación típica. Como Y sigue una distribución *Normal*, obtener $P(Y < 1.959964)$. Justifica la igualdad de estos resultados.
(c) $P(X \leq 20/X \geq 12)$; $P(X \leq 8)$.
(d) Los cuartiles y los deciles.
(e) Genera una muestra aleatoria de tamaño 500.
(f) Calcula su media, varianza, mediana, moda teóricamente (sin usar R).
(g) Calcula la media, varianza, mediana y moda de la muestra aleatoria obtenida anteriormente. Compara los valores.
(h) Genera, nuevamente, una muestra aleatoria de tamaño 10000. Calcula los valores resumen.
(i) Representa gráficamente la distribución de frecuencias de los 10000 datos anteriores. Compara con la distribución teórica de la variable.

4. Un fabricante produce envases con una capacidad media de 600 cm^3 con una desviación típica de 0.1 cm^3. En sus contratos de venta se estipula una cláusula por la que se acepta una devolución de piezas defectuosas siempre que su capacidad no esté comprendida entre *600 - L* y *600 + L*. Calcular cual debe de ser el valor de L para que el porcentaje de envases rechazados no supere el 8%. Suponer distribución normal.

3.1 Variables relacionadas con la Normal

5. Calcular:

(a) $P(\chi^2_{150} \geq 126)$
(b) $P(40 \leq \chi^2_{65} \leq 50)$
(c) $P(\chi^2_{220} \geq 260)$
(d) El valor de a tal que $P(\chi^2_{100} \geq a) = 0.6$
(e) Dados 10 v.a. independientes e idénticamente distribuidas Y_i con $Y_i \sim N(0,5)$ para $i = 1, 2, ..., 10$, calcular
- $P(\sum_{i=1}^{10} Y_i^2 \leq 600)$
- $P(\frac{1}{10}\sum_{i=1}^{10} Y_i^2 \geq 12.175)$
- El número a tal que $P(\sqrt{\frac{1}{10}\sum_{i=1}^{10} Y_i^2} \geq a) = 0.5$

6. Para la distribución $t - student$ calcular

(a) $P(t_8 \leq 0.25)$, $P(t_8 \leq 0.1625)$
(b) $P(t_{95} \geq -1.645)$
(c) El valor de a que verifica $P(t_{90} > a) = 0.95$

7. Para la distribución F calcular

(a) $P(F_{6;7} \leq 4.25)$
(b) $P(F_{6;7} \geq 0.245)$
(c) El valor de a que verifica $P(F_{6;7} > a) = 0.975$
(d) Encontrar $t^2_{0.975;5}$ y $F_{0.975;1;5}$. Comprobar que $(t^2_{0.975;5})^2 = F_{0.975;1;5}$. Usando las definiciones de ambas distribuciones demostrar este resultado: $t^2_{\alpha;n} = F_{\alpha;1;n}$