Guia 15 Opcional

Agustín Muñoz González

15/6/2020

Preparamos el entorno

```
rm(list=ls())
library(ggplot2)
```

1. El objetivo de este ejercicio es utilizar la Ley de los Grandes Números para aproximar integrales de funciones continuas en intervalos finitos. Procuraremos aproximar la probabilidad de que una variable Z ~ N (0, 1) tome valores en un intervalo [a, b]. Es decir, queremos aproximar numéricamente la siguiente integral

$$\int_a^b \frac{1}{2\pi}\exp(-\frac{x^2}{2})dx.$$

Para ello, consideremos la siguiente propuesta: tomamos U 1 , U 2 , . . . U n variables i.i.d. U i ~ U[a, b] independientes y calculamos

$$\lim_{n\to\infty}\frac{\sum_i \frac{1}{2\pi}\exp(-\frac{\mathcal{U}_i^2}{2})}{n}.$$

(a) Aproximamos la integral mediante

$$(b-a)\frac{\sum_i \frac{1}{2\pi}\exp(-\frac{\mathcal{U}_i^2}{2})}{n},$$

para los siguientes valores de a y b, y para n = 100, 1000 y 50000.

i. a = -1.96 y b = 1.96.
ii. a = -2 y b = 1.
iii. a = 0 y b = 2.34.

Resolución:

Defino una función que aproxime la integral.

```
aprox_int=function(a,b,n){
  variables=runif(n,a,b)
  1/n*(b-a)*sum(1/(2*pi)*exp(-variables^2/2))
}
#otra forma
aprox_int2=function(a,b,n){
  variables=runif(n,a,b)
  (b-a)*mean(1/(2*pi)*exp(-variables^2/2))
}
```

Veamos los valores que queremos.

i. a = -1.96 y b = 1.96.

```
aprox_int(-1.96,1.96,100)
```

```
## [1] 0.3740844
```

```
aprox_int(-1.96,1.96,1000)
```

```
## [1] 0.3776741
```

```
aprox_int(-1.96,1.96,50000)
```

```
## [1] 0.3799009
```

ii. a = -2 y b = 1.

```
aprox_int(-2,1,100)
```

```
## [1] 0.3164519
```

```
aprox_int(-2,1,1000)
```

```
## [1] 0.3232446
```

```
aprox_int(-2,1,50000)
```

```
## [1] 0.327178
```

iii. a = 0 y b = 2.34.

```
aprox_int(0,2.34,100)
```

```
## [1] 0.1974555
```

```
aprox_int(0,2.34,1000)
```

```
## [1] 0.1980485
```

```
aprox_int(0,2.34,50000)
```

```
## [1] 0.1956737
```

(b) Otro modo de aproximar la probabilidad de que una variable Z ~ N (0, 1) tome valores en un intervalo [a, b]. Sería generar variables Z 1 , Z 2 , . . . Z n i.i.d. Z i ~ N [0, 1] y calcular

$$\frac{1}{n}\sum_i I_{(a,b)}(Z_i).$$

Explique por que es razonable este métdodo. Y aproxime la probabilidad para los siguientes valores de a y b, y para n = 100, 1000 y 50000.

i. a = -1.96 y b = 1.96.
ii. a = -2 y b = 1.
iii. a = 0 y b = 2.34.

Resolución:

Este método es razonable porque lo que estamos haciendo es considerar nuevas variables aleatorias

$$Y_i = I_{Z_i \in (a,b)} = I_{(a,b)}(Z_i) \sim \mathcal{B}(1,p), \text{ con } p = P(Z_i \in (a,b))).$$

Luego por la LGN

$$\overline{Y_n} = \frac{1}{n}\sum_i I_{(a,b)}(Z_i) \xrightarrow[n\to\infty]{P} E(Y_1) = P(Z \in (a,b))).$$

Defino primero la función que aproxima la proba.

```r
# una forma
aprox_prob=function(a,b,n){
  variables=rnorm(n,0,1)
  mean(a<=variables & variables <= b)
}
# otra forma:
aprox_prob2=function(a,b,n){
  variables=rnorm(n,0,1)
  1/n*sum(indicadora(a,b,variables))
}
```

Veamos los valores que nos piden.

i. a = -1.96 y b = 1.96.

```r
a = -1.96
b = 1.96
aprox_prob(a,b,100)
```

```
## [1] 0.94
```

```r
aprox_prob(a,b,1000)
```

```
## [1] 0.957
```

```r
aprox_prob(a,b,50000)
```

```
## [1] 0.9497
```

ii. a = -2 y b = 1.

```r
a = -2
b = 1
aprox_prob(a,b,100)
```

```
## [1] 0.85
```

```r
aprox_prob(a,b,1000)
```

```
## [1] 0.807
```

```r
aprox_prob(a,b,50000)
```

```
## [1] 0.81912
```

iii. a = 0 y b = 2.34.

```r
a = 0
b = 2.34
aprox_prob(a,b,100)
```

```
## [1] 0.6
```

```r
aprox_prob(a,b,1000)
```

```
## [1] 0.497
```

```r
aprox_prob(a,b,50000)
```

```
## [1] 0.49284
```

ME DIERON COSAS DISTINTAS A LAS DEL ITEM (a)

(c) Comparar los resultados obtenidos en a) y b) con aquellos que calcularía utilizando con R la función Φ.

Resolución:

Usaremos la función $pnorm(x, \mu, \sigma) = P(N < x)$.

i. a = -1.96 y b = 1.96.

```
a = -1.96
b = 1.96
pnorm(b,0,1)-punif(a,0,1)
```

```
## [1] 0.9750021
```

ii. a = -2 y b = 1.

```
a = -2
b = 1
pnorm(b,0,1)-punif(a,0,1)
```

```
## [1] 0.8413447
```

iii. a = 0 y b = 2.34.

```
a = 0
b = 2.34
pnorm(b,0,1)-punif(a,0,1)
```

```
## [1] 0.9903581
```

ME DIERON COSAS DISTINTAS A LAS DE LOS OTROS ITEMS ¿?

2. Cuando se realiza una encuesta con preguntas delicadas, donde la gente tiene cierta resistencia a responder, se suele usar una método indirecto para hacer la pregunta. Supongamos que estamos interesados en conocer (o estimar) la proporción de mujeres que se realizaron un aborto. Un modo de realizar la encuesta sería el siguiente: a cada mujer le damos un dado y una moneda y le decimos que tire el dado. Si el resultado del dado es 3, 4, 5, o 6 la mujer responde la verdad y si sale 1 o 2 tira la moneda; si sale cara responde SI y si sale ceca responde NO. Obviamente, el encuestador no ve el resultado del dado ni el de la moneda y por lo tanto no sabría si la respuesta del encuestado es la verdad o es producto del azar. A partir de este mecanismo quisieramos poder estimar la verdadera proporción de mujeres que se hicieron un aborto que llamaremos p, es decir p =Probabilidad de que una mujer elegida al azar se haya realizado un aborto.

(a) Calcular (en función de p) la probabilidad de que una persona elegida al azar responda afirmativamente a la pregunta (P (SI)). Ayuda: condicione al resultado del dado.

Resolución:

(b) A partir de encuestar a n personas con este mecanismo, como estimaría la probabilidad de que alguien responda que SI, es decir P (SI)? Mas precisamente, si pensamos en n variables i.i.d W 1 , . . . , W n donde W i = 1 si la mujer responde SI y 0 en caso contario, que cuenta deberá hacer para estimar la proporción de gente que responde SI, P(SI).

Resolución:

(c) Usando a) y b), proponga una cuenta para estimar p, la probabilidad de que una mujer elegida al azar se haya realizado un aborto. Llamemos $\hat{p}$ a esta cuenta.

Resolución:

(d) Para convencernos que este mecanismo resulta efectivo, realice la siguiente simulación. Supongamos que el verdadero valor de p = 0.1 es decir la verdadera proporción de mujeres que se realizó un aborto. Simule n = 100 respuestas, W 1 , . . . , W n según el mecánismo de respuesta propuesto y compare con la verdadera proporción p. Repita esto N rep = 1000 veces.

Resolución:

(e) Si p = 0.1, ¿ qué distribución tiene W i ? Usando el Teorema Central del Límite, aproxime la probabilidad P $(0 < \hat{p} < 0.2)$.

Resolución:

(f) Supongamos como en el ejercicio anterior que p = 0.1 pero hacemos la encuenta directa y nadie miente. En ese caso, estimaríamos a p simplemente contando el porcentaje de gente que contesta si. Es decir, tendríamos X 1 , . . . , X n variables aleatorias i.i.d Bi(1, p) y estimaríamos p con $\overline{X}_n = \frac{1}{n}\sum_i X_i$. Usando el Teorema Central del Límite, aproxime la probabilidad $P(0 < \overline{X}_n < 0.2)$.

Resolución:

(g) Explique la relación o las diferencias entre los dos item anteriores.

Resolución:

3. En este ejercicio estudiaremos la distribución del promedio $\overline{X}_n$ de variables X 1 , . . . , X n (i.i.d.) pero con distribución asimétrica. Consideremos X con distribución LogNormal(μ, σ^2), es decir su densidad es

$$f_X(x) = \frac{1}{x\sigma\sqrt{(2\pi)}}$$

En tal caso, la esperanza y varianza de X están dadas por

$$E(X) = e^{\mu+\sigma^2/2} \text{ y } Var(X) = (e^{\sigma^2} - 1)e^{2\mu+\sigma^2},$$

repectivamente. El nombre de la distributicón proviene del siguiente hecho:

$$X \sim LogNormal(\mu, \sigma^2) \text{ si y solo sí } log(X) \sim N(\mu, \sigma^2)$$

(a) Consideremos una variable X_1 con distribución LogNormal(0,2). Grafiquemos su densidad. Indique el valor de E(X_1) y el de var(X_1).

Resolución:

(b) Generaremos datos correspondientes a una muestra X 1 , . . . , X n con distribución LogNormal(0,2) y computamos

$$T_n = \frac{\overline{X}_n - E(X_1)}{(Var(\overline{X}_n))^{1/2}} = \frac{\overline{X}_n - E(X_1)}{(Var(\overline{X}_1)/n)^{1/2}}.$$

Replicamos N rep = 1000 veces, para diferentes valores de n: n = 30, 100, 500, 1000, obteniendo 4 conjuntos de datos. Cada uno de ellos contienen los N rep = 1000 valores de T_n obtenidos para n = 30, 100, 500, 1000, respectivamente.

i. Realice un histograma para conjunto de T n obtenidos. ¿Qué características tienen estos histogramas? ¿Qué se observa? Notemos que para poder comparar los histogramas de los distintos conjuntos de datos será necesario representarlos en la misma escala, tanto para el eje horizontal como para el vertical.

ii. Realice también los boxplots para comparar los distintos conjuntos de datos en un mismo gráfico. ¿Qué observa?

Resolución: