Pontificia Universidad Católica de Chile
Facultad de Matemáticas
Departamento de Estadística

Temporada Académica de Verano (enero 2021)

Interrogación 3 - Pauta

Pregunta 1 *Teorema del Limite Central* **[1.0 Punto]**

Como han de saber, estos últimos días el valor del Dólar estadounidense (US$) ha subido / bajado "aleatoriamente". Suponga que el comportamiento del valor del dólar los próximos 30 días será de tipo discreto: baja $7 pesos o sube $10 pesos con igual probabilidad. ¿Cuál es la probabilidad (aproximada) que obtenga una utilidad en 30 días más sobre $50 pesos por cada dólar invertido hoy?

Hint: El soporte de la suma es $\{-159, -142, -125, \ldots, 28, 45, 62, 79, \ldots, 232, 249\}$.

Solución

Definamos X_1, X_2, ..., X_{30} la subida o bajada en pesos del valor del dólar los próximos 30 días.

Por enunciado se tiene que:

$$\textbf{[0.2 ptos]} \quad \mu_X = 1.5 \quad \text{y} \quad \sigma_X = 8.5 \quad \textbf{[0.3 ptos]}$$

Como no se indica grado de dependencia entre subidas o bajadas diarias y las probabilidades permanecen constantes, entonces podemos asumir un comportamiento iid.

Se pide

$$P\left(\sum_{i=1}^{30} X_i > 50\right) = 1 - P\left(\sum_{i=1}^{30} X_i \leq 50\right)$$

Alternativa 1: Como la utilidad entorno a $50 solo puede tomar valores: $45 ó $62 pesos, entonces

$$\begin{aligned} P\left(\sum_{i=1}^{30} X_i > 50\right) &= 1 - P\left(\sum_{i=1}^{30} X_i \leq 50\right) = 1 - P\left(\sum_{i=1}^{30} X_i \leq 45\right) = 1 - P\left(\sum_{i=1}^{30} X_i < 62\right) \\ &\approx \texttt{1-pnorm(45 + 8.5, mean = 1.5 * 30, sd = 8.5 * sqrt(30))} \\ &= 0.4275661 \quad \textbf{[0.5 ptos]} \end{aligned}$$

Alternativa 2: Simular y entregar un resultado empírico ó calcular el exacto.

```
y = c()
for(i in 1:1000000){
n = 30
x = rbinom(n, size = 1, prob = 1/2)
x[x==0] = -7
x[x==1] = +10
y[i] = sum(x)
}
mean(y>50)
[1] 0.427688
```

[0.5 ptos]

Pregunta 2 *Covarianza Combinación Lineal* **[1.0 Punto]**

Un análisis a los postulantes a ING-UC muestra que los puntajes NEM se comportan de acuerdo a una distribución cuya media es de [`mu1`] puntos con una desviación estándar de 50. Por otra parte, los puntajes en la PDT-Lenguaje tiene una media [`mu2`] punto y desviación estándar 80, y los puntajes en la PDT-Matemáticas presentan una media de [`mu3`] puntos y desviación estándar de 30.

Además se tiene que la correlación entre los puntajes:

- NEM vs PDT-Leng es igual a [`rho1`]
- NEM vs PDT-Matemáticas es igual a [`rho2`]
- PDT-Leng vs PDT-Matemáticas es igual a [`rho3`]

Calcule la covarianza entre el puntaje NEM y el promedio de los puntajes PDT.

Solución

Se pide

$$\begin{aligned}\text{Cov}(\text{NEM}, 0.5 \cdot \text{PDT-L} + 0.5 \cdot \text{PDT-M}) &= 0.5 \cdot \text{Cov}(\text{NEM, PDT-L}) + 0.5 \cdot \text{Cov}(\text{NEM, PDT-M}) \quad \textbf{[0.5 ptos]} \\ &= 0.5 \cdot 50 \cdot 80 \cdot \texttt{rho1} + 0.5 \cdot 50 \cdot 30 \cdot \texttt{rho2} \quad \textbf{[0.5 ptos]}\end{aligned}$$

En R:

```
## Ejemplo
rho1 = 0.6
rho2 = 0.7
0.5*50*80*rho1+0.5*50*30*rho2
[1] 1725
```

Pregunta 3 *Método Delta* **[1.0 Punto]**

Suponga que un atleta lanzador de jabalina luego de estudiar su técnica modeló la velocidad de tiro V (en `m/s`) como una variable Uniforme $(23, [\texttt{B}])$ y el ángulo de tiro T (en radianes) como una variable aleatoria Beta(30, [`C`]) con soporte en $[0,\ \pi/2]$.

Suponga que la distancia (en metros) medida desde el atleta hasta donde aterriza la jabalina está dada por la siguiente ecuación

$$D = \frac{V^2 \cdot \sin(2 \cdot T)}{\texttt{g}},$$

donde $\texttt{g} = 9.8\ \texttt{m/s}^2$.

Bajo el supuesto que V y T son independientes, obtenga la desviación estándar aproximada de 1er orden para D.

Solución

Aproximado por Taylor se tiene que

$$D \approx \frac{\mu_V^2 \cdot \sin(2 \cdot \mu_T)}{\texttt{g}} + (V - \mu_V) \cdot \left(\frac{2\,\mu_V \cdot \sin(2 \cdot \mu_T)}{\texttt{g}}\right) + (T - \mu_T) \cdot \left(\frac{\mu_V^2 \cdot \cos(2 \cdot \mu_T) \cdot 2}{\texttt{g}}\right) \quad \textbf{[0.4 ptos]}$$

Aplicando operador varianza

$$\text{Var}(D) \approx \sigma_V^2 \cdot \left(\frac{2\,\mu_V \cdot \sin(2 \cdot \mu_T)}{\texttt{g}}\right)^2 + \sigma_T^2 \cdot \left(\frac{\mu_V^2 \cdot \cos(2 \cdot \mu_T) \cdot 2}{\texttt{g}}\right)^2 \quad \textbf{[0.4 ptos]}$$

+ [0.2 ptos] por resultado correcto.

En R:

```
## Ejemplo
g = 9.8

B = 25
C = 6
a = 23
b = B
mu.V = (a+b)/2
sigma.V = sqrt((b-a)^2/12)

a = 0
b = pi/2
q = 30
r = C
mu.T = a+q*(b-a)/(q+r)
sigma.T = sqrt(q*r*(b-a)^2/((q+r)^2*(q+r+1)))

sqrt(sigma.V^2*(2*mu.V*sin(2*mu.T)/g)^2+sigma.T^2*(mu.V^2*cos(2*mu.T)*2/g)^2)
[1] 9.898876
```

Pregunta 4 *Min y Max* **[1.0 Punto]**

La I_2 de hace una semana tenían 12 preguntas, al descargar al azar por cada alumno el tiempo (en minutos) de permanencia en una pregunta, se observó un comportamiento $\text{Beta}(1, r)$ con soporte $[0,\, 20]$ y una media de 8 minutos. Bajo el supuesto, que los alumnos responden la prueba de manera independiente, calcule es la probabilidad que al observa dos de estos tiempos tomados al azar, el máximo sea mayor al doble del mínimo.

Solución

Sea X el tiempo de permanencia en una pregunta tomada al azar de un alumno entre las 12 que tenía la I_2.

Del enunciado se tiene que

$$X \sim \text{Beta}(1,\, 1.5), \quad x \in [0,\, 20]$$

al despejar r de la siguiente igualdad (Formulario):

$$\mu_X = 8 = 0 + \frac{1}{(1+r)} \cdot (20 - 0) \to r = 1.5\,. \quad \textbf{[0.4 ptos]}$$

Por formulario también tenemos que la distribución conjunta entre el min y max cuando $n = 2$ está dada por

$$f_{Y_1,Y_2}(u, v) = 2 \cdot f_X(u) \cdot f_X(v), \quad u \leq v \quad \textbf{[0.2 ptos]}$$

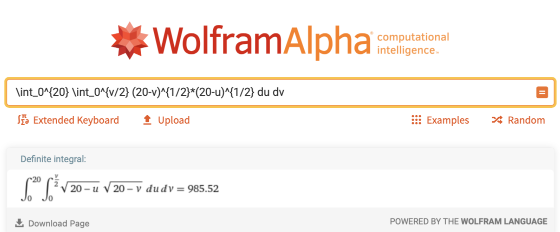
Se pide

$$\begin{aligned}
P(Y_2 > 2 \cdot Y_1) &= \int_0^{20} \int_0^{v/2} 2 \cdot \frac{1}{B(1, 1.5)} \cdot \frac{\sqrt{20-u}}{20^{1.5}} \cdot \frac{1}{B(1, 1.5)} \cdot \frac{\sqrt{20-v}}{20^{1.5}}\, du\, dv \\
&= 2 \cdot \left(\frac{1}{B(1, 1.5)} \cdot \frac{1}{20^{1.5}}\right)^2 \int_0^{20} \int_0^{v/2} \sqrt{20-u} \cdot \sqrt{20-u}\, du\, dv \quad \textbf{[0.2 ptos]} \\
&= 2 \cdot \left(\frac{1}{B(1, 1.5)} \cdot \frac{1}{20^{1.5}}\right)^2 \cdot 985.52 \\
&= 0.0005625 \cdot 985.52 = 0.554355 \quad \textbf{[0.2 ptos]}
\end{aligned}$$

En R:

```
a = 0
b = 20
q = 1
mu = 8
## mu = a + q * (b-a)/(q+r)
r = ((mu-a)/(q*(b-a)))^(-1)-q

## f(u,v) = 2*f(u)*f(v), u < v --> ¿P(V > 2*U)?
2*(1/(beta(1,1.5)*20^1.5))^2 * 985.52
[1] 0.554355
```



Pregunta 5 *Transformación* **[1.0 Punto]**

Un análisis a las notas finales de un curso, muestra un comportamiento Gamma (k, ν), pero como esta distribución tiene como soporte los $\mathbb{R}^+$ y las notas se encuentra en el intervalo $[1-7]$, se propone una Gamma (k, ν) truncada cuya función de densidad está dada por:

$$f_X(x) = \frac{1}{c} \cdot \frac{\nu^k}{\Gamma(k)} \cdot x^{k-1} \cdot e^{-\nu x}, \quad 1 \leq x \leq 7$$

$c = \text{pgamma }(7, \text{shape} = k, \text{rate} = \nu) - \text{ pgamma }(1, \text{shape} = k, \text{rate} = \nu)$

Suponga que $k =$ [k] y $\nu =$ [nu]$\cdot$ ¿Cuál es la probabilidad que un alumno haya reprobado?

Hint: Un alumno reprueba si NF < 3.95.

Solución

Nos pide

$$\begin{aligned}
P(X < 3.95) &= \int_1^{3.95} \frac{1}{c} \cdot \frac{\nu^k}{\Gamma(k)} \cdot x^{k-1} \cdot e^{-\nu x}\, dx \\
&= \frac{1}{c} \cdot \int_1^{3.95} \frac{\nu^k}{\Gamma(k)} \cdot x^{k-1} \cdot e^{-\nu x}\, dx \\
&= \frac{1}{c} \cdot \left\{ \texttt{pgamma(3.95, shape = } k\texttt{, rate = } \nu\texttt{) - pgamma(1, shape = } k\texttt{, rate =}\nu\texttt{)} \right\} \quad \textbf{[0.5 ptos]} \\
&= \frac{\texttt{pgamma(3.95, shape = } k\texttt{, rate = } \nu\texttt{) - pgamma(1, shape = } k\texttt{, rate =}\nu\texttt{)}}{\texttt{pgamma(7.00, shape = } k\texttt{, rate = } \nu\texttt{) - pgamma(1, shape = } k\texttt{, rate =}\nu\texttt{)}} \quad \textbf{[0.5 ptos]}
\end{aligned}$$

En R:

```
## Ejemplo
k = 7
nu = 2.9
C = pgamma(7.00, shape = k, rate = nu) - pgamma(1, shape = k, rate = nu)
(pgamma(3.95, shape = k, rate = nu) - pgamma(1, shape = k, rate = nu))/C
[1] 0.9366225
```

Pregunta 6 $Z = g(X, Y)$ **[1.0 Punto]**

Debido al toque de queda que rige actualmente en Santiago (entre 10.00 pm y 5.00 am), las propinas hoy en día son más generosa. Suponga que en un restaurante en particular la propina se puede modelar a partir de la siguiente relación entre el tiempo de servicio y el tiempo de espera de atención :

$$\text{Propina} = 10000 \cdot \frac{X}{Y}$$

Si $X \sim$ Exponencial(1) e $Y \sim$ Exponencial(2) de manera independiente, ¿cuál es la probabilidad que una propina sea mayor a $30.000 pesos?

Solución

Por independencia se tiene que

$$f_{X,Y}(x, y) = 2\,e^{-2\,y}\,e^{-x}, \quad x > 0 \quad \text{e} \quad y > 0 \quad \textbf{[0.4 ptos]}$$

Se pide $P\left(10000 \cdot \frac{X}{Y} > 30000\right) = P(X > 3 \cdot Y)$.

Alternativa 1:

$$\begin{aligned} P(X > 3 \cdot Y) &= \int_0^\infty \int_0^{x/3} 2\,e^{-2\,y}\,e^{-x}\,dy\,dx = \int_0^\infty (1 - e^{2x/3})\,e^{-x}\,dx \quad \textbf{[0.4 ptos]} \\ &= \int_0^\infty e^{-x}\,dx - \frac{3}{5} \cdot \int_0^\infty \frac{5}{3}\,e^{5\,x/3}\,dx = 1 - \frac{3}{5} \cdot 1 = 0.4 \quad \textbf{[0.2 ptos]} \end{aligned}$$

Alternativa 2: Definamos $Z = X/Y$ y por formulario tenemos que su función de densidad es

$$f_Z(z) = \int_0^\infty 2 \cdot e^{-z\,y} \cdot e^{-2\,y} \cdot |y|\,dy = \frac{2\,\Gamma(2)}{(z+2)} \int_0^\infty \frac{(z+2)^2}{\Gamma(2)}\,y^{2-1}\,e^{-(z+2)\,y}\,dy = \frac{2}{(z+2)^2} \cdot 1, \quad z > 0 \quad \textbf{[0.4 ptos]}$$

Luego

$$P(Z > 3) = \int_3^\infty \frac{2}{(z+2)^2}\,dz = -\frac{2}{(z+2)}\bigg|_3^\infty = 0.4 \quad \textbf{[0.2 ptos]}$$

Pregunta 7 $Y = g(X)$, *TCL, EMV* **[1.0 Punto]**

Considere una muestra iid de tamaño [n] proveniente de una distribución Gamma([k] , ν). ¿Cuál es la probabilidad (exacta o aproximada) que el estimador máximo verosímil de ν sea mayor a $\dfrac{3\,\nu}{4}$? Considere para su calculo que ν es igual a [nu] .

Solución

Como k es conocido, se tiene que el EMV de ν está dado por $\hat{\nu} = \dfrac{k}{\overline{X}}$ y se pide $P\left(\hat{\nu} > \dfrac{3\,\nu}{4}\right)$.

Alternativa 1: Como $\overline{X} \sim \text{Gamma}(k \cdot n,\, \nu \cdot n)$. [Resultado exacto visto en clase]. **[0.5 ptos]**

$P\left(\hat{\nu} > \dfrac{3\,\nu}{4}\right) = P\left(\overline{X} < \dfrac{4\,k}{3\,\nu}\right) =$ `pgamma(4 * k / (3 * nu), shape = k * n, rate = nu * n)` **[0.5 ptos]**

Alternativa 2: Como $\overline{X} \overset{\text{aprox}}{\sim} \text{Normal}\left(\frac{k}{\nu},\, \sqrt{\frac{k}{\nu^2\,n}}\right)$. [Aproximación Normal por TCL] **[0.5 ptos]**

$P\left(\hat{\nu} > \dfrac{3\,\nu}{4}\right) = P\left(\overline{X} < \dfrac{4\,k}{3\,\nu}\right) =$ `pnorm(4 * k / (3 * nu), mean = k / n, sd = sqrt(k / n) / nu))` **[0.5 ptos]**

Alternativa 3: Como $\hat{\nu} \overset{\text{aprox}}{\sim} \text{Normal}\left(\nu,\, \sqrt{\frac{\nu^2}{n\,k}}\right)$. [Distribución asintótica de los EMV] **[0.5 ptos]**

$P\left(\hat{\nu} > \dfrac{3\,\nu}{4}\right) = 1 -$ `pnorm(3 * nu / 4, mean = nu, sd = nu / sqrt(k * n))` **[0.5 ptos]**

En R:

```
## Ejemplo
n = 12
k = 4
nu = 4.3
pgamma(k/(0.75*nu), shape = k*n, rate = nu*n)
[1] 0.9838376
pnorm(k/(0.75*nu), k/nu, sqrt(k/n)/nu))
[1] 0.9895393
1-pnorm(0.75*nu, nu, sqrt(nu^2/(k*n)))
[1] 0.9583677
```

Pregunta 8 *Min - ECM* **[1.0 Punto]**

Considere una muestra aleatoria iid proveniente de una población cuya distribución es Exponencial(ν) trasladada en α, es decir, con la siguiente función de densidad:

$$f_X(x) = \nu e^{-\nu(x-\alpha)}$$

con $x > \alpha, \nu > 0$ y $\alpha \in \mathbb{R}$.

Un buen estimador de α es $\hat{\alpha} = \text{mín}\,\{X_1, X_2, \ldots, X_n\}$.

Si $n =$ `[n]`, $\nu =$ `[nu]` y $\alpha =$ `[alpha]`, calcule el error cuadrático medio de $\hat{\alpha}$.

Solución

Por formulario y resultado visto el modulo de ejercicios el martes se tiene que

$$\hat{\alpha} = \text{mín}\{X_1, \ldots, X_n\} \sim \text{Exponencial}(n\,\nu) \text{ trasladada en } \alpha \quad \textbf{[0.3 ptos]}$$

Luego

$$\textbf{[0.2 ptos]} \quad \text{E}(\hat{\alpha}) = \frac{1}{n\,\nu} + \alpha \quad \text{y} \quad \text{Var}(\hat{\alpha}) = \frac{1}{n^2\,\nu^2} \quad \textbf{[0.2 ptos]}$$

Por lo tanto

$$\text{ECM}(\hat{\nu}) = \frac{1}{n^2\,\nu^2} + \left(\frac{1}{n\,\nu} + \alpha - \alpha\right)^2 = \frac{2}{n^2\,\nu^2} \quad \textbf{[0.3 ptos]}$$

En R:

```
## Ejemplo
n = 39
nu = 0.6
a = 5
2/(n*nu)^2
[1] 0.003652568
```

Hace unos días la zona norte-centro del país sintió un movimiento telúrico de gran intensidad y nos recordó que en esta época de COVID, en nuestro país en cualquier momento tendremos un nuevo terremoto. La base `Sismos2021.xlsx` contiene la información actualizada hasta ayer de los sismos que han sido registrados por el centro de sismología este año.

Esta base contiene el tiempo en segundos transcurridos desde el ultimo movimiento registrado, su ubicación medida con latitud-longitud, profundidad en km, magnitud en escala Richter y una referencia del epicentro con respecto a un hito geográfico. En la 2da hora del archivo se encuentra esta misma descripción de variables.

A continuación responda las últimas dos preguntas:

```
library(readxl)
Base = as.data.frame(read_excel("Sismos2021.xlsx"))
library("fitdistrplus")
```

Pregunta 9 ***fitdist()*** **[1.0 Punto]**

Utilizando la función `fitdist()` del paquete `"fitdistrplus"`, ajuste una distribución Exponencial, Gamma y Weibull a los tiempos entre sismos por el método de máxima verosimilitud.

Responda:

1. Modelo Exponencial: $P(\text{Tiempo} > 20) = 0.7593375$ **[0.25 ptos]**
2. Modelo Gamma: $P(\text{Tiempo} > 20) = 0.749438$ **[0.25 ptos]**
3. Modelo Weibull: $P(\text{Tiempo} > 20) = 0.7431949$ **[0.25 ptos]**
4. ¿Cuál ajusta mejor al resultado empírico? Gamma **[0.25 ptos]**

Solución

En R:

```
X = Base$Tiempo
par1 = fitdist(data = X, distr = "exp", method = "mle")$estimate
1-pexp(20, rate = par1)
[1] 0.7593375
par2 = fitdist(data = X, distr = "gamma", method = "mle")$estimate
1-pgamma(20, shape = par2[1], rate = par2[2])
[1] 0.749438
par3 = fitdist(data = X, distr = "weibull", method = "mle")$estimate
1-pweibull(20, shape = par3[1], scale = par3[2])
[1] 0.7431949
mean(X>20)
[1] 0.7487805
## Mejor ajuste Gamma
```

Pregunta 10 *fitdist()* **[1.0 Punto]**

Solución

La variable Magnitud tiene un comportamiento correspondiente a una distribución trasladada. Estime el trasladamiento por el mínimo, luego ajuste un modelo Exponencial y Gamma por el método de los momentos. Utilice la función del paquete y responda:

1. Modelo Exponencial Trasladado: $P(\text{Magnitud} \leq 4) = 0.894109$ **[0.4 ptos]**
2. Modelo Gamma Trasladado: $P(\text{Magnitud} \leq 4) = 0.9059824$ **[0.4 ptos]**
3. ¿Cuál ajusta mejor al resultado empírico? Gamma **[0.2 ptos]**

Solución

En R:

```
X = Base$Magnitud
a = min(X)
par1 = fitdist(data = X-a, distr = "exp", method = "mme")$estimate
pexp(4-a, rate = par1)
[1] 0.894109
par2 = fitdist(data = X-a, distr = "gamma", method = "mme")$estimate
pgamma(4-a, shape = par2[1], rate = par2[2])
[1] 0.9059824
mean(X<=4)
[1] 0.9146341
## Mejor ajuste Gamma
```