Pontificia Universidad Católica de Chile
Facultad de Matemáticas
Departamento de Estadística

Primer Semestre 2021

Interrogación 2 - Pauta

Pregunta 1

Suponga que el precio (en dólares) de la libra de Salmón varía diariamente de manera independiente, según una variable aleatoria Weibull. ¿Cuál es la probabilidad que en los próximos 10 días hábiles, en al menos un `[x]` % de los días el precio supere los US$5.00 por libra?

Considere para su respuesta la siguiente información:

- El 10 % de los precios del último mes fueron inferiores a US$3.88.
- El 10 % de los precios del último mes fueron superiores a US$5.19.

Solución

Definamos como X_1, X_2, $\ldots$, X_{10} al precio de la libra de salmón durante 10 días hábiles consecutivos, los cuales se comportan según enunciado como variables aleatorias independientes con distribución Weibull(η, β).

Para efectos de este ejercicios nos indican que consideremos además la siguiente información del último mes:

$$x_{0.10} = 3.88 \quad \text{y} \quad x_{0.90} = 5.19$$

Estos percentiles se relaciona teóricamente con η y β de la siguiente manera:

$$\ln(x_p) = \ln(\eta) + \frac{1}{\beta} \cdot \ln[-\ln(1-p)]$$

Resolviendo se tiene que

$$\eta = 4.797398 \quad \text{y} \quad \beta = 10.60301 \quad \textbf{[0.3 Ptos.]}$$

Definamos ahora la variable aleatoria Y como el número de días en que el precio supera los US$5.00:

$$Y \sim \text{Binomial}(10,\, p), \quad \textbf{[0.20 Ptos.]}$$

con

$$p = \texttt{1 - pweibull(5, shape = beta, scale = eta)} = 0.2121394 \quad \textbf{[0.20 Ptos.]}$$

Se pide

$$P\left(Y \geq \frac{\texttt{x} \cdot 10}{100}\right) = \texttt{1 - pbinom(x/10-1, size = 10, prob = p)} \quad \textbf{[0.30 Ptos.]}$$

x	Respuesta
30	0.35908569
40	0.14321634
50	0.04149758

Pregunta 2

Este jueves la selección chilena de futbol reanuda su participación en las clasificatorias del próximo mundial de fútbol y como protocolo a los jugadores les realizan constantemente exámenes PCR.

Suponga que uno de los jugadores no se ha portado muy bien el ultimo tiempo ya que rompió la burbuja sanitaria y tuvo contacto estrecho con una persona con Covid.

Suponga que este jugador contrajo el virus y ya tiene un PCR positivo. Para "limpiar" su imagen tiene pensado realizarse varios PCR hasta que tenga uno con resultado negativo.

¿Cuántos PCR adicionales se esperaría que debiera realizarse? Considere que la prueba PCR tiene un sensibilidad, capacidad de detectar Covid en una persona con el virus, del `[q]` %.

Solución

Definamos como X al número de PCR adicionales que el jugador (que tiene Covid) debe realizase hasta obtener un PCR negativo:

$$X \sim \text{Geométrica}(p),$$

con p la probabilidad que un PCR salga negativo si es que la persona tiene Covid.

Del enunciado se tiene $p = 1 - \texttt{[q]}/100$ **[0.5 Ptos.]** y se pide $\text{E}(X) = \dfrac{1}{p}$. **[0.5 Ptos.]**

q	p	Resp
85	0.15	6.6667
86	0.14	7.1429
87	0.13	7.6923
88	0.12	8.3333
89	0.11	9.0909
90	0.10	10.0000
91	0.09	11.1111
92	0.08	12.5000
93	0.07	14.2857
94	0.06	16.6667
95	0.05	20.0000

Nota: Si el alumno descuenta 1 PCR asignar puntaje completo

Pregunta 3

Chile es un país exportador de Paltas Hass a Estados Unidos (avocados).

El proceso de recepción en EE.UU. es muy exigente y simple de explicar: De un embarque se extraen `[m]` paltas al azar, las cuales son examinadas secuencialmente, si hay dos o más que no cumplan con los requerimientos, el embarque completo es rechazado.

Suponga que usted, exportador de Paltas, ha enviado un embarque tal, que hay una probabilidad del `[q]` % que cualquier palta no cumpla con los requerimientos.

¿Cuál es la probabilidad que en el proceso de recepción su embarque sea rechazado antes de completar la revisión de las paltas seleccionadas?

Solución

Sea X el número de paltas que no cumplen con los requisitos de exportación, entre las primeras $\texttt{[m]} - 1$ revisiones:

$$X \sim \text{Binomial}\,(n = \texttt{[m]} - 1 \textbf{ [0.2 Ptos.]},\, p = \texttt{[q]}/100 \textbf{ [0.3 Ptos.]})$$

Se pide

$$P(X \geq 2) = \texttt{1 - pbinom(1, size = m-1, prob = q/100)} \quad \textbf{[0.5 Ptos.]}$$

Como alternativa se podría haber definido la variable aleatoria Y como número de revisiones hasta obervar la 2da palta que no cumple los requisitos para exportación:

$$Y \sim \text{Bin-Neg}(k = 2,\, p = \texttt{[q]}/100) \textbf{ [0.3 Ptos.]}$$

Se pide

$$P(Y \leq \texttt{[m]} - 1 \textbf{ [0.2 Ptos.]}) = \texttt{pnbinom(m-1-2, size = 2, prob = q/100)} \quad \textbf{[0.5 Ptos.]}$$

```
## Ejemplo ##

m = 15
q = 5
1-pbinom(1, size = m-1, prob = q/100)
[1] 0.1529856
pnbinom(m-1-2, size = 2, prob = q/100)
[1] 0.1529856
```

Pregunta 4.1

Durante las ultimas semanas muchas familias que no habían solicitado ayuda estatal durante la pandemia comenzaron a inscribirse para ver si cumplén con los requisitos para el nuevo aporte estatal desde el IFE ampliado.

Las autoridades esperan 500 mil nuevas postulaciones, pero desconocen cuántas de ellas cumplirán con los requisitos. Suponga que a partir de una muestra aleatoria de tamaño `[n]` entre las postulaciones recibidas hasta el momento, hay `[x]` que cumplen requisito.

¿Cuántos nuevas postulaciones esperaría que cumplan los requisitos?

Pregunta 4.2

Durante las ultimas semanas muchas familias que no habían solicitado ayuda estatal durante la pandemia comenzaron a inscribirse para ver si cumplén con los requisitos para el nuevo aporte estatal desde el IFE ampliado.

Las autoridades esperan 500 mil nuevas postulaciones, pero desconocen cuántas de ellas NO cumplirán con los requisitos. Suponga que a partir de una muestra aleatoria de tamaño `[n]` entre las postulaciones recibidas hasta el momento, hay `[x]` que NO cumplen requisito.

¿Cuántos nuevas postulaciones esperaría que NO cumplan los requisitos?

Solución

Definamos cómo X el número de familias que "cumplen" o "NO cumplen" requisito en la muestra:

$$X \sim \text{Hipergeométrica}(n = \texttt{[n]},\, N = 500000,\, m) \quad \textbf{[0.5 Ptos.]}$$

Se pide el valor de m, el cuál lo podemos obtener a partir de la siguiente relación vista en clases:

$$N = \frac{m \cdot \texttt{[n]}}{\texttt{[x]}} \to m = \frac{N \cdot \texttt{[x]}}{\texttt{[n]}} \quad \textbf{[0.5 Ptos.]}$$

```
## Ejemplo ##

n = 50
x = 20
N = 500000
m = N * x / n
m
[1] 200000
```

Nota: En esta pregunta se consideró un margen de error $\pm n$.

Pregunta 5

Se ha ideado un sistema eficiente de ingreso a los Campus. Este permite determinar si un estudiante cumple con los protocolos para acceder a éste (registrar o validar el pase de movilidad, vacunación, test KOR, etc), de tal forma que en una estación de acceso el tiempo que toma por cada estudiante, se comporta de acuerdo a una variable exponencial con un tiempo medio de [mu] min.

Considerando que no hay tiempos muertos en la estación de acceso (siempre hay estudiantes esperando para el ingreso), ¿Cuál es la probabilidad que entre 9.00 y 10.00 a.m. **revisen** más de 9 estudiantes por esta estación?

Solución

Definamos X_t al número de alumnos revisados en t minutos y como el tiempo entre revisiones es Exponencial(ν), con $\nu = \dfrac{1}{[\text{mu}]}$, entonces

$$X_t \sim \text{Poisson}(\nu\, t) \quad \textbf{[0.5 Ptos.]}$$

Se pide

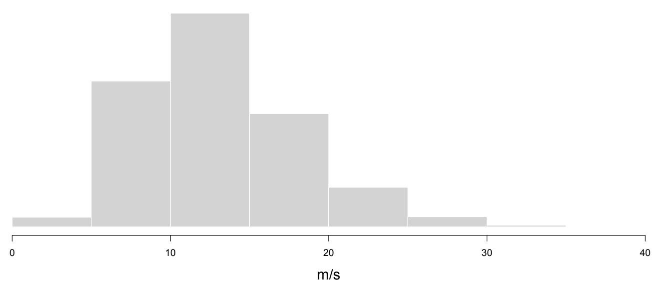
$$P(X_{60} > 9) = \texttt{1 - ppois(9, lambda = 60/[mu])} \quad \textbf{[0.5 Ptos.]}$$

mu	Resp
4	0.9301
5	0.7576
6	0.5421
7	0.3563
8	0.2236
9	0.1374
10	0.0839
11	0.0515
12	0.0318

Nota: Si el alumno calculó $P(X_{60} \geq 9)$ descontar 0.2 puntos, si calculó $P(X_{60} \leq 9)$ o $P(X_{60} < 9)$ descontar 0.5 puntos.

Pregunta 6

Un parque eólico es una central eléctrica donde la producción de la energía eléctrica se consigue a partir de la fuerza del viento, mediante aerogeneradores que aprovechan las corrientes de aire. Un análisis del comportamiento del viento en el Parque Eólico Pinón Blanco, IX Región, se presenta a continuación:



La media de observada fue de `[mu]` m/s y el coeficiente de asimetría resultó igual a `[theta]`.

Bajo es supuesto que las mediciones se comportan como una variable aleatoria Gamma, ¿cuál es la probabilidad que la velocidad de viento en la zona supere los 20 m/s?

Solución

Definamos como X la velocidad del viento en el Parque Eólico.

$$X \sim \text{Gamma}(k,\, \nu)$$

Igualando

$$\mu_X = \frac{k}{\nu} = \texttt{[mu]} \quad \text{y} \quad \theta_X = \frac{\dfrac{k\,(k+1)\,(k+2)}{\nu^3} - 3\cdot\dfrac{k\,(k+1)}{\nu^2}\cdot\dfrac{k}{\nu} + 2\,\dfrac{k^3}{\nu^3}}{\dfrac{k\,\sqrt{k}}{\nu^3}} = \frac{2}{\sqrt{k}} = \texttt{[theta]} \quad \textbf{[0.5 Ptos.]}$$

Se tiene que

$$k = \frac{4}{\texttt{[theta]}^2} \quad \text{y} \quad \nu = \frac{4}{\texttt{[theta]}^2 \cdot \texttt{[mu]}} \quad \textbf{[0.3 Ptos.]}$$

Se pide

$P(X > 20) =$ `1 - pgamma(20, shape = 4/(theta*theta), rate = 4/(theta*theta*mu))` **[0.2 Ptos.]**

```
## Ejemplo ##

theta = 0.9
mu = 14
k = (2/theta)^2
nu = k/mu
1-pgamma(20, shape = k, rate = nu)
[1] 0.1614129
```

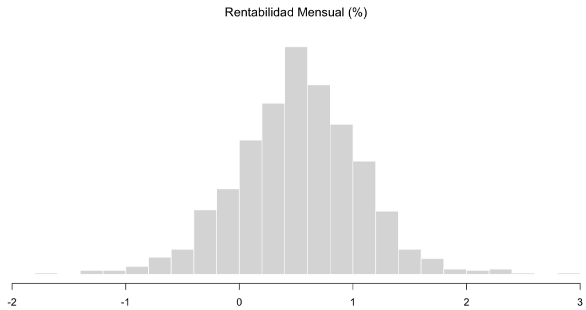
Nota: El coeficiente de asimetría se puede obtener a partir de los 1ros tres momentos de una Gamma(k, ν):

$$\mu_X = \frac{k}{\nu}, \quad E(X^2) = \frac{k\,(k+1)}{\nu^2}, \quad E(X^3) = \frac{k\,(k+1)\,(k+2)}{\nu^3}$$

o consultando "Univariate Distribution Relationships" en el módulo Enlace en la página principal de Canvas.

Pregunta 7

La siguiente figura muestra la rentabilidad mensual porcentual de un fondo mutuo, cuyo comportamiento puede ser modelado por una distribución logística.



La media de estos datos fue de `[mu]` %, con un coeficiente de variación igual a `[delta]`.

¿Cuál es la probabilidad que una inversión en este fondo al menos entregue una rentabilidad mensual del 1 %?

Solución

Definamos como X la rentabilidad mensual porcentual, la cuál distribuye Logistica(μ, σ).

Del formulario tenemos que:

$$\mu = \texttt{[mu]} \quad \text{y} \quad \sigma = \frac{\texttt{[delta]} \cdot \texttt{[mu]}\sqrt{3}}{\pi} \quad \textbf{[0.5 Ptos.]}$$

Se pide

$$\textbf{[0.3 Ptos.]} \quad P(X > 1) = \texttt{1 - plogis(1, location = } \mu\texttt{, scale = } \sigma\texttt{)} \quad \textbf{[0.2 Ptos.]}$$

```
## Ejemplo ##

mu = 0.52
delta = 0.96
sigma = mu*delta*sqrt(3)/pi
1 - plogis(1, location = mu, scale = sigma)
[1] 0.1488008
```

Pregunta 8.1

El archivo `ENS Muestra.xlsx` contiene una muestra aleatoria de 350 casos con indicadores de salud. Considerando solo las personas de género "Femenino", ajuste una distribución normal bivariada para COLESTEROL e IMC y responda:

(a) Proporción de personas en la muestra que tiene un COLESTEROL > 180 y $20 <$ IMC < 30.

(b) Según el modelo estimado, ¿cuál sería la probabilidad del evento consultado en (a)?

Solución

```
Base = rio::import("ENS_Muestra.xlsx")
Data = dplyr::filter(Base, SEXO == "Femenino")[c("COLESTEROL", "IMC")]
mean(Data$COLESTEROL > 180 & (Data$IMC >20 & Data$IMC < 30))
[1] 0.3830846
```

[0.5 Ptos.]

```
sigma = cov(Data)
mu = apply(Data,2,mean)
mvtnorm::pmvnorm(lower = c(180, 20), upper = c(Inf, 30), mean = mu, sigma = sigma)[1]
[1] 0.371005
```

[0.5 Ptos.]

Pregunta 8.2

El archivo `ENS Muestra.xlsx` contiene una muestra aleatoria de 350 casos con indicadores de salud. Considerando solo las personas de género "Masculino", ajuste una distribución normal bivariada para COLESTEROL e IMC y responda:

(a) Proporción de personas en la muestra que tiene un COLESTEROL > 170 y $25 <$ IMC < 30.

(b) Según el modelo estimado, ¿cuál sería la probabilidad del evento consultado en (a)?

Solución

```
Base = rio::import("ENS_Muestra.xlsx")
Data = dplyr::filter(Base, SEXO == "Masculino")[c("COLESTEROL", "IMC")]
mean(Data$COLESTEROL > 170 & (Data$IMC >25 & Data$IMC < 30))
[1] 0.2214765
```

[0.5 Ptos.]

```
sigma = cov(Data)
mu = apply(Data,2,mean)
mvtnorm::pmvnorm(lower = c(170, 25), upper = c(Inf, 30), mean = mu, sigma = sigma)[1]
[1] 0.1742347
```

[0.5 Ptos.]

Pregunta 9

Durante el Cyber Day muchas tiendas han incrementado sus ventas. A partir de datos recolectados hasta el momento se tiene que las visitas por hora a la página de una tienda de retail se comporta como una variable aleatoria Poisson (λ) y que las ventas condicionadas a x visitas se comporta como una variable aleatoria Gamma(x, $1/x$).

Determine el coeficiente de variación de las ventas incondicional a las visitas y evalúe para $\lambda =$ `[lambda]`.

Solución

Definamos como X las visitas por hora a la pagina e Y las ventas.

Del enunciado se tiene que

$$X \sim \text{Poisson}(\lambda) \quad \text{e} \quad Y \,|\, X = x \sim \text{Gamma}(x,\, 1/x)$$

Se pide $\delta_Y = \dfrac{\sigma_Y}{\mu_Y}$. **[0.2 Ptos.]**

Por teorema de esperanzas iteradas tenemos que

$$\mu_Y = \text{E}[\text{E}(Y \,|\, X)] = \text{E}(X^2) = \lambda^2 + \lambda \quad \textbf{[0.2 Ptos.]}$$

y

$$\begin{aligned}
\sigma_Y^2 &= \text{E}[\text{Var}(Y \,|\, X)] + \text{Var}[\text{E}(Y \,|\, X)] \\
&= \text{E}(X^3) + \text{Var}(X^2) \\
&= \text{E}(X^3) + \text{E}(X^4) - \left[\text{E}(X^2)\right]^2 \\
&= (\lambda^3 + 3\,\lambda^2 + \lambda) + (\lambda^4 + 6\,\lambda^3 + 7\,\lambda^2 + \lambda) - (\lambda^2 + \lambda)^2 \quad \textbf{[0.6 Ptos.]}
\end{aligned}$$

Primeros cuatro momentos teóricos de X usan WolframAlpha:

Infinite sum:

$$\sum_{x=0}^{\infty} \frac{x\left(\lambda^x e^{-\lambda}\right)}{x!} = \lambda$$

Infinite sum:

$$\sum_{x=0}^{\infty} \frac{x^2\left(\lambda^x e^{-\lambda}\right)}{x!} = \lambda\,(\lambda + 1)$$

Infinite sum:

$$\sum_{x=0}^{\infty} \frac{x^3\left(\lambda^x e^{-\lambda}\right)}{x!} = \lambda\left(\lambda^2 + 3\,\lambda + 1\right)$$

Infinite sum:

$$\sum_{x=0}^{\infty} \frac{x^4\left(\lambda^x e^{-\lambda}\right)}{x!} = \lambda\left(\lambda^3 + 6\,\lambda^2 + 7\,\lambda + 1\right)$$

```
## Ejemplo ##

lambda = 80
sqrt((lambda^3+3*lambda^2+lambda) + (lambda^4+6*lambda^3
+7*lambda^2+lambda)-(lambda^2+lambda)^2)/(lambda^2+lambda)
[1] 0.2496835
```

Pregunta 10

Suponga que el número de clientes que atiende un cajero durante una hora en un supermercado distribuye Poisson, donde la tasa esperada de clientes en cada hora se comporta como una variable aleatoria Exponencial(ν).

Si $\nu =$ `[nu]`, determine la probabilidad que en una hora el cajero atienda a 3 clientes.

Solución

Sea Y el número de clientes que llegan por hora y X la tasa esperada en cada hora.

$$X \sim \text{Exponencial}(\nu) \quad \text{e} \quad Y \,|\, X = x \sim \text{Poisson}(x)$$

Se pide $p_Y(3)$.

Por Teorema de Probabilidades Totales se tiene que

$$\begin{aligned}
p_Y(y) &= \int_{-\infty}^{\infty} p_{Y\,|\,X=x}(y) \cdot f_X(x)\,dx = \int_0^{\infty} \frac{x^y\,e^{-x}}{y!} \cdot \nu\,e^{-\nu\,x}\,dx \\
&= \frac{\nu}{(\nu+1)^{y+1}} \int_0^{\infty} \frac{(\nu+1)^{y+1}}{\Gamma(y+1)}\,x^{(y+1)-1}\,e^{-(\nu+1)\,x}\,dx \quad \textbf{[0.5 Ptos.]} \\
&= \frac{\nu}{(\nu+1)^{y+1}} \cdot 1, \quad \text{por área bajo su soporte de una Gamma}(y+1,\,\nu+1) \\
&= \frac{\nu}{(\nu+1)^{y+1}}, \quad y \in \mathbb{N}_0 \quad \textbf{[0.5 Ptos.]}
\end{aligned}$$

```
## Ejemplo ##

nu = 2
y = 3
nu/(1+nu)^(y+1)
[1] 0.02469136
```

Pregunta 11

La densidad conjunta de dos variables aleatorias continuas X e Y, esta dada por

$$f_{X,Y}(x,y) = \frac{\nu^3 x}{2} \cdot \exp\left\{-\frac{(\nu\, x^2 + y)}{x}\right\}$$

con $x > 0$, $y > 0$ y $\nu > 0$.

El mejor predictor de Y en función de X es:

Solución

Se pide $\text{E}(Y \,|\, X)$.

Por Teorema de Probabilidades Totales se tiene que

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy = \frac{\nu^2\, x^2\, e^{-\nu\, x}}{2} \int_0^{\infty} \frac{1}{x}\, e^{-y/x}\, dx = \frac{\nu^2\, x^2\, e^{-\nu\, x}}{2} \cdot 1$$

Luego

$$f_{Y\,|\,X=x}(y) = \frac{f_{X,Y}(x,y)}{f_X(x)} = \frac{1}{x}\, e^{-y/x}, \quad x > 0 \quad \textbf{[0.5 Ptos.]}$$

Por lo tanto

$$Y \,|\, X = x \sim \text{Exponencial}(1/x) \to \text{E}(Y \,|\, X) = X \quad \textbf{[0.5 Ptos.]}$$

Pregunta 12

Este fin de semana se apreciaron largas filas en los centros de vacunación (vacunatorios). Un vacunatorio de Providencia disponía de vacunas AstraZeneca para los Hombres y Sinovac para las Mujeres.

Si la tasa de ingreso al vacunatorio (para el proceso de vacunación) se comporta de acuerdo a una distribución Poisson con tasa igual a `[lambda]` personas por hora, y 2/3 de ellas corresponden a Hombres, determine el tiempo esperado que transcurrirá entre el ingreso del duodécimo Hombre y decimoquinto Hombre.

Solución

Sea X_t el número de hombres que ingresan em t horas.

$$X_t \sim \text{Poisson}\big((2/3) \cdot \texttt{[lambda]} \cdot t\big) \quad \textbf{[0.3 Ptos.]}$$

Por lo tanto, el tiempo T entre el duodécimo Hombre y decimoquinto distribuye Gamma$(3,\, 2 \cdot \texttt{[lambda]}/3)$. **[0.5 Ptos.]**

Se pide $\text{E}(T) = \dfrac{3}{2 \cdot \texttt{[lambda]}/3}$. **[0.2 Ptos.]**

lambda	Resp
3	1.5000
4	1.1250
5	0.9000
6	0.7500
7	0.6429
8	0.5625
9	0.5000