Pontificia Universidad Católica de Chile
Facultad de Matemáticas
Departamento de Estadística
EYP1016-1 – Introducción a la Estadística

Guía 3

Introducción a la Estadística - Laboratorio (01-2024)

Profesor: Cristian Capetillo Constela
Ayudante: Andrés Díaz

1. Para cada una de las distribuciones expuestas abajo, 1) genere $n = 1000$ valores, 2) utilizando la función *cumsum* calcule el promedio para cada tamaño muestral, 3) grafique estos promedios para cada tamaño muestral, y 4) sobreponga una línea horizontal en el verdadero valor de la media.

   a) $X \sim \mathcal{N}(0, 1)$ ($\mathbb{E}[X] = 0$).
   b) $X \sim \mathcal{P}oi(2)$ ($\mathbb{E}[X] = 2$).
   c) $X \sim \mathcal{G}a(1, 5)$ ($\mathbb{E}[X] = 1/5$).
   d) $X \sim \mathcal{U}(4, 8)$ ($\mathbb{E}[X] = 6$).

2. Para cada una de las distribuciones expuestas abajo, 1) genere $n = 1000$ valores, 2) calcule la probabilidad de que X sea menor o igual a su esperanza, 3) utilizando la función *cumsum* calcule la probabilidad anterior para cada tamaño muestral y 4) grafique estas probabilidades para cada tamaño muestral.

   a) $X \sim \mathcal{N}(0, 1)$ ($\mathbb{E}[X] = 0$).
   b) $X \sim \mathcal{P}oi(2)$ ($\mathbb{E}[X] = 2$).
   c) $X \sim \mathcal{G}a(1, 5)$ ($\mathbb{E}[X] = 1/5$).
   d) $X \sim \mathcal{U}(4, 8)$ ($\mathbb{E}[X] = 6$).

3. En cierta sector de una avenida principal de una ciudad, las cámaras instaladas pueden monitorear la distancia que recorrieron los vehículos y cuánto se demoraron en recorrerla. La distancia en metros (X) se supone ser una variable aleatoria $\mathcal{N}(100, 25)$, mientras que el tiempo en segundos (Y), independiente de la distancia, se supone $\mathcal{G}a(4, 1)$. Se quieren calcular distintas probabilidades y métricas en torno a la velocidad de los vehículos (X/Y). Para ello, simule 3000 pares (X, Y) y responda.

   a) Entregue un histograma de las velocidades simuladas.
   b) Calcule la probabilidad aproximada de que un vehículo conduzca a $10m/s$.
   c) Quizás para aproximar la probabilidad, 3000 es demasiado. ¿Cuántos datos serían suficiente para estabilizarla? Grafique la probabilidad para cada tamaño muestral.

4. Simule 1000 veces una muestra aleatoria $\mathcal{G}a(3, 2)$ de tamaño $n = 100$ y responda.

   a) Calcule el promedio en cada una de las muestras.
   b) Para cada una de las muestras, calcule $\frac{\overline{x}-1.5}{\sqrt{0.75/n}}$ y cree un histograma de aquello.

c) ¿Se cumple el Teorema del Límite Central? Grafique la curva $\mathcal{N}(0,1)$ para comprobarlo.

5. En el contexto de la revisión de proyectos con impacto ambiental, la tabla 1 contiene las frecuencias relativas para la cantidad de extensiones de plazo solicitadas por los titulares de tales proyectos.

Tabla 1: Cantidad de extensiones solicitadas por los titulares en la revisión de proyectos con impacto ambiental.

Extensiones	0	1	2	3	4	5	6
Frecuencia absoluta	1	2	4	8	7	4	1
Frecuencia relativa	0.037	0.0741	0.1481	0.2963	0.2593	0.1481	0.037

Se quiere ajustar un modelo $\mathcal{P}oi(\lambda)$ sobre los datos. Para ello, responda.

a) Grafique la función de masa empírica, esto es, las frecuencias relativas.

b) Calcule, mediante la función *optim*, la estimación máximo verosímil de λ, $\widehat{\lambda}$.

c) Grafique sobre lo obtenido en (a) la función de masa $\mathcal{P}oi(\widehat{\lambda})$.

d) Suponiendo que la estimación es cierta. ¿Cuál es la probabilidad de que un titular solicite 4 extensiones o más?

e) Es sabido que el EMV para lambda es el promedio de los datos. Utilizando el TLC, y recordando que la esperanza y varianza Poisson coinciden, calcule un intervalo de 95 % de confianza para $\widehat{\lambda}$.

6. Genere una muestra aleatoria de tamaño $n = 50$ de la distribución $\mathcal{G}eom(\theta)$, con $\theta = 0.3$. Luego, responda.

a) Grafique la función de log-verosimilitud para θ.

b) Calcule la estimación máximo verosímil de θ basada en su muestra.

c) Compárelo con el valor real de θ a través de la diferencia absoluta.

7. Simule 1000 veces una muestra aleatoria de tamaño $n = 100$ de la distribución $\mathcal{G}eom(\theta)$ con $\theta = 0.3$ y responda

a) Calcule la estimación máximo verosímil de θ para cada una de las muestras.

b) Cree un histograma con los resultados obtenidos en (a). ¿Se asemeja a una distribución Normal centrada en 0.3?.

8. Genere una muestra aleatoria de tamaño $n = 500$ de la distribución $\mathcal{G}a(\alpha, \beta)$ con $\alpha = \beta = 2$ y responda.

a) Programe la función de log-verosimilitud de α y β.

b) Calcule la estimación máximo verosímil de ambos parámetros, basada en su muestra.

c) Compare lo anterior con los valores reales de los parámetros.

d) Realice la transformación α/β y compárela (a través de su diferencia absoluta) con el promedio muestral y con la esperanza teórica (1).

9. En el contexto de una competencia de tiro al arco, se sabe que los tiros de una arquera en particular tienen cierto comportamiento aleatorio. Más específicamente, se sabe que sus desvíos horizontales en centímetros (X) distribuyen $\mathcal{N}(0, 15)$, mientras que de forma independiente sus desvíos verticales (Y) distribuyen $\mathcal{N}(0, 8)$. Interesa estudiar su puntería, esto es, la distancia entre sus tiros y el centro. Para ello, responda.

   a) Genere valores aleatorios (usted debe decidir cuántos, justificando su respuesta) de los desvíos verticales y horizontales.

   b) Realice la transformación $\sqrt{X^2 + Y^2}$ para obtener las distancias de los tiros a su centro. Y realice un histograma de su resultado.

   c) ¿Cuál es la probabilidad de que la distancia entre un tiro de la arquera y el centro del objetivo sea mayor a 5 centímetros?