Regresión Avanzada ITAM
27 de enero de 2024 2024

Tarea 2.

Fecha de entrega: 6 de febrero 2024.

1. **Estimando una media Poisson usando una inicial discreta.** Supongan que son dueños de una compañía de transporte con una flota grande de camiones. Las descomposturas ocurren aleatoriamente en el tiempo y supóngase que el número de descomposturas durante un intervalo de t días sigue un distribución Poisson con media λt. El parámetro λ es la tasa de descompostura diaria. Los posibles valores para λ son 0.5, 1, 1.5, 2,2.5 y 3, con respectivas probabilidades 0.1, 0.2, 0.3, 0.2, 0.15 y 0.05. Si uno observa y descomposturas, entonces la probabilidad posterior de λ es proporcional a

$$g(\lambda)\exp(-t\lambda)(t\lambda)^y,$$

donde g es la distribución inicial.

   a. Si 12 camiones se descomponen en un periodo de 6 días, encontrar la probabilidad posterior para las diferentes tasas.
   b. Encontrar la probabilidad de que no haya descomposturas durante la siguiente semana. *Hint: Si la tasa es λ, la probabilidad condicional de no descomposturas durante un periodo de 7 días está dado por* $\exp(-7\lambda)$. *Se puede calcular esta probabilidad predictiva multiplicando la lista de probabilidades condicionales por las probabilidades posteriores de λ y encontrando la suma de los productos*

2. **Estimando una proporción y predicción de una muestra futura**. Un estudio reporta sobre los efectos de largo plazo de exposición a bajas dosis de plomo en niños. Los investigadores analizaron el contenido de plomo en la caída de los dientes de leche. De los niños cuyos dientes tienen un contenido de plomo mayor que 22.22 ppm, 22 eventualmente se graduaron de la preparatoria y 7 no. Supongan que su densidad inicial para p, la proporción de todos tales niños que se graduaron de preparatoria es $beta(1,1)$, y posterior es $beta(23,8)$.

   a. Encontrar un intervalo estimado de 90 % para p.
   b. Encontrar la probabilidad de que p exceda 0.6.

3. **Estimando una media normal posterior con una inicial discreta**. Supongamos que están interesados en estimar el promedio de caida de lluvia por año μ en (cm) para una ciudad grande del Centro de México. Supongan que la caída anual individual $y_1,\ldots,y_n$ son obtenidas de una población que se supone $\mathcal{N}(\mu,100)$. Antes de recolectar los datos, supongan que creen que la lluvia media puede estar en los siguiente valores con respectivas probabilidades

μ	20	30	40	50	60	70
$g(\mu)$	0.1	0.15	0.25	0.25	0.15	0.1

a. Supongan que se observan los totales de caída de lluvia 38.6, 42.4, 57.5, 40.5, 51.7, 67.1, 33.4, 60.9, 64.1, 40.1, 40.7 y 6.4. Calcular la media.

b. Calcular la función de verosimilitud utilizando como estadística suficiente la media $\bar{y}$.

- Calcular las probabilidades posteriores para μ
- Encontrar un intervalo de probabilidad de 80 % para μ.

4. **Modelo muestral Cauchy**. Supongan que se observa una muestra aleatoria $y_1, \ldots, y_n$ de una densidad Cauchy con parámetro de localización θ y parámetro de escala 1. Si una inicial uniforme se considera para θ, entonces la densidad posterior, ¿cuál es? Supongan que se observan los datos 0,10,9,8,11,3,3,8,8,11.

   a. Calcula un grid para θ de -2 a 12 en pasos de 0.1
   b. Calcula la densidad posterior en este grid.
   c. Grafica la densidad y comenten sobre sus características principales.
   d. Calcula la media posterior y desviación estándar posterior.

5. **Robustez Bayesiana** Supongan que están a punto de lanzar una moneda que creen que es honesta. Si p denota la probabilidad de obtener sol, entonces su mejor creencia es que $p = 0.5$ Adicionalmente, creen que es altamente probable que la moneda sea cercana a honesta, lo que cuantifican como $P(0.44 \leq p \leq 0.56) = 0.9$. Consideren las siguientes dos iniciales para p:

   P1 $p \sim beta(100, 100)$

   P2 $p \sim 0.9beta(500, 500) + 0.1beta(1, 1)$

   a. Simular 1000 valores de cada densidad inicial P1 y P2. Resumiendo las muestras simuladas, mostrar que ambas iniciales concuerdan con las creencias iniciales acerca de la probabilidad p del lanzamiento de moneda.
   b. Supongan que lanzan la moneda 100 veces y obtienen 45 soles. Simular 1000 valores de las distribuciones posteriores P1 y P2, y calcular intervalos de probabilidad del 90 %.
   c. Supongan que sólo observan 30 soles de los 100 lanzamientos. Nuevamente simular 1000 valores de las dos posteriores y calcular intervalos de probabilidad del 90 %.
   d. Viendo los resultados de (b) y (c), comentar sobre la robustez de la inferencia con respecto a la elección de la densidad inicial en cada caso.

6. **Aprendiendo de datos agrupados**. Supongan que manejan en carretera y típicamente manejan a una velocidad constante de 70km/h. Un día, rebasan un carro y son rebasados por 17 carros. Supongan que las velocidades son distribuídas $\mathcal{N}(\mu, 100)$. Si rebasan s carros y son rebasados por f,

   a. ¿Cuál es la verosimilitud de μ?
   b. Asignando una densidad inicial plana para μ, si $s = 1$ y $f = 17$, graficar la densidad posterior de μ.
   b. Usando la densidad encontrada en (a), encontrar la media posterior de μ.

c. Encontrar la probabilidad de que la velocidad promedio de los carros exceda 80 km/h.

7. **Problema de Behrens-Fisher**. Supongan que se observan dos muestras normales independientes, la primera se distribuye de acuerdo a una $\mathcal{N}\left(\mu_1, \sigma_1^2\right)$ y la segunda de acuerdo a $\mathcal{N}\left(\mu_2, \sigma_2^2\right)$. Denoten la primera muestra por $x_1, \ldots, x_m$ y la segunda muestra por $y_1 \ldots, y_n$ Supongan también que los parámetros $\boldsymbol{\theta} = (\mu_1, \sigma_1^2, \mu_2, \sigma_2^2)$ tienen la distribución inicial vaga dada por:

$$g(\boldsymbol{\theta}) \propto \frac{1}{\sigma_1^2 \sigma_2^2}$$

   a. Encontrar la densidad posterior. Mostrar que los vectores (μ_1, σ_1^2) y (μ_2, σ_2^2) tienen distribuciones posteriores independientes.
   b. Describir cómo simular la densidad posterior conjunta de $\boldsymbol{\theta}$.
   c. Los siguientes datos dan la longitud de la mandíbula en mm para 10 chacales machos y 10 chacales hembras en la colección del Museo Británico. Usando simulación, encontrar la densidad posterior de la diferencia en la longitud media de las mandíbulas entre los sexos. ¿Hay suficiente evidencia para concluir que los machos tienen una longitud promedio mayor que las hembras?

Machos	120	107	110	116	114	111	113	117	114	112
Hembras	110	111	107	108	110	105	107	106	111	111

8. **Estimando los parámetros de una densidad Poisson/Gamma**. Supongamos que $y_1, \ldots, y_n$ es una muestra aleatoria de una densidad Poisson/Gamma:

$$f(y|a, b) = \frac{\Gamma(y+a)}{\Gamma(a) y!} \frac{b^a}{(b+1)^{y+a}}$$

donde $a \geq 0$, $b \geq 0$. Esta densidad es un modelo apropiado para conteos que muestran más dispersión que la que predice un modelo Poisson. Supongamos que (a, b) tiene asignada la inicial no informativa proporcional a $1/(ab)$. Si transformamos a los parámetros $\theta_1 = \log(a)$ y $\theta_2 = log(b)$, la densidad posterior es proporcional a

$$g(\theta_1, \theta_2) \propto \prod_{i=1}^{n} \frac{\Gamma(y_i + a}{\Gamma(a) y_i!} \frac{b^a}{(b+1)^{y_i+a}},$$

donde $a = \exp(\theta_1)$ y $b = \exp(\theta_2)$. Usa este marco para modelar los datos obtenidos por Gilchrist (1984), en los que una serie de 33 trampas de insectos fueron puestas sobre varias dunas de arena y se registra el número de diferentes insectos atrapados sobre un tiempo fijo. El número de inssectos en las trampas se muestran a continuación:

2	5	0	2	3	1	3	4	3	0	3
2	1	1	0	6	0	0	3	0	1	1
5	0	1	2	0	0	2	1	1	1	0

Calculando la densidad posterior sobre una retícula, simular 1000 extracciones de la densidad conjunta posterior de (θ_1, θ_2). De la muestra simulada, encontrar intervalos estimados de 90 % para los parámetros a y b.