Inicio: XX de XX de 202X, HH:00
Límite de entrega: HH+1:00

Nombre y apellidos : JUANITA ZURITA DNI : 271828

Probabilidad e Inferencia Paramétrica

Enunciado

- Escribe nombre, DNI y fecha en un folio en blanco. Si tienes algún inconveniente, escribe rápidamente al profesor. Copia el número de la pregunta que estás resolviendo y la respuesta. No copies el enunciado.
- Aségurate de responder a todas las preguntas.
- Firma el documento anterior, pon encima de ese folio tu DNI o tu carnet de estudiante de modo que se pueda ver el texto y la firma, saca una foto de todo ello. Las siguientes aplicaciones pueden ser útiles: "herramienta de escaneo de google drive", `adobe scan` ó `camscanner` desde el móvil, o `webcamoid` desde PC.
- Si el formato obtenido no es pdf, convierte el archivo a pdf, utilizando por ejemplo la opción "imprimir a archivo".
- Sube el archivo pdf a la tarea moodle.
- Graba un archivo de vídeo en el que respondas a la última pregunta.
- Entrega el vídeo o vídeos en ...

Preguntas

Question 1
En un edificio de 43 viviendas, el gasto mensual de agua corriente de cada casa se puede modelizar mediante la misma distribución, uniforme continua en el intervalo $[20, 44]$. Los gastos tienen una correlación del 10%.

- Calcula la esperanza del gasto de agua conjunto de todas las viviendas del edificio.
- ¿Se puede aproximar el gasto de agua conjunto de todas las viviendas del edificio usando el teorema central del límite? Si es así, estima la desviación típica del gasto conjunto.

La portería pasa a la comunidad gastos de electricidad y calefacción que siguen una distribución normal multivariable. El gasto de electricidad medio es de 20 euros, y tiene una desviación estándar de 16 euros. El gasto de calefacción medio en enero es de 60 euros, y tiene una desviación estándar de 25 euros. Ambos gastos tienen una correlación del 50%.

- Calcula la probabilidad de que el gasto de electricidad más calefacción de la portería sea mayor de 90 euros. ¿Cuál sería esa probabilidad si los gastos en electricidad y calefacción en la portería fueran independientes?

Explanation: El gasto medio de cada vivienda es $\frac{20+44}{2}$ y su varianza es $\frac{(44-20)^2}{12}$.

- El gasto medio total es $\frac{43(20+44)}{2} = 1376.0$. Es indiferente si los gastos de cada vivienda son independientes o no.
- Como los gastos tienen una correlación del 10%, no son independientes, y no se puede aplicar el teorema central del límite.
- El vector (GE, GC) = (gasto electricidad, gasto calefacción) de la portería se distribuye según una normal 2D (con vector de medias $(20, 60)$ y matriz de covarianzas $array([[256., 200.], [200., 625.]])$, aunque eso no es importante para este ejercicio). La suma $GE+GC$ es una normal unidimensional, con media $E[GE+GC] = E[GE]+E[GC] = 20+60$. Para la suma $GE + GC$, usamos la fórmula $Var[GE + GC] = Var[GE] + Var[GC] + 2Cov[GE, GC] = 16^2 + 25^2 + 2\sigma_{GE}\sigma_{GC}50/100 = 1281.0$. La probabilidad pedida es $P(GE + GC > 90) = 0.3899691741401351$, mientras que si los gastos fueran independientes sería 0.3680929160228634. Se puede calcular con `scipy.stats`:

  ```
  suma = st.norm(loc=20+60,scale=np.sqrt(1281.0))
  prob_porteria = 1-suma.cdf(90)
  ```

Question 2
Para una operación de reparaciones en alta mar, cierto sistema crítico S_1 está formado por 33 componentes que deben funcionar todos correctamente para que la misión termine con éxito. La probabilidad de que cada componente funcione es 0.99 y los componentes se averían independientemente.

- ¿Cuál es la probabilidad de que el sistema S_1 funcione?
- Como pensamos que la probabilidad de fallo de S_1 es demasiado grande, se instala otro sistema auxiliar S_2, con otro enfoque totalmente diferente no basado en componentes, que tiene una probabilidad de funcionar del 40%. Calcula la probabilidad de que funcione alguno de los dos sistemas y la misión se pueda completar.
- Si el sistema funcionó, calcula la probabilidad de que S_1 fallase, y la activación de S_2 haya salvado la misión.

Explanation:

- S_1 funciona si cada componente funciona y cada uno es independiente, luego $P(S_1 \text{ funciona}) = P(\text{componente 1 funciona y} \dots \text{componente 33 funciona}) = P(\text{componente 1 funciona}) \cdot \ldots \cdot P(\text{componente 33 funciona}) = 0.99^{33} = 0.7177305325982749$.
- Usamos el teorema de la probabilidad total: $P(\text{algún sistema funciona}) = P(S_1 \text{ funciona}) \cdot 1 + P(S_1 \text{ NO funciona}) \cdot P(S_2 \text{ funciona}) = 0.8306383195589649$.
- $P(S_1 \text{ falló}|\text{algún sistema funciona}) = \frac{P(\text{algún sistema funcionó, pero } S_1 \text{ falló})}{P(\text{algún sistema funciona})} = \frac{P(S_2 \text{ funcionó, y } S_1 \text{ falló})}{P(\text{algún sistema funciona})} = 0.1359289407941587$.

Question 3
Según cierto dietista, la proporción en la flora bacteriana de actinobacteria sigue una distribución normal de media 0.05 y desviación típica 0.01.
La dieta del submarino podría provocar trastornos digestivos a las personas con menos de un 2% de actinobacteria. Y si más de 2 personas tienen estos problemas, puede comprometer la misión.

- Si en el submarino hay una tripulación de 30 personas, calcula la probabilidad de que más de 2 personas tengan problemas digestivos.
- En cada misión, la proporción de actinobacteria de cada tripulante puede cambiar, de modo que tenemos la misma probabilidad de trastornos digestivos. ¿Qué distribución de probabilidad sigue el número de misiones hasta que una misión se vea comprometida por esta causa por primera vez? Encuentra los percentiles 5% y 95% para este número de misiones.

Explanation:

- Según el modelo, el tripulante i-ésimo tiene problemas digestivos si su proporción de actinobacteria A_i es menor que 2%. Por ejemplo, normalizamos A_i: $P(A_i < 2\%) = P(\frac{A_i - 0.05}{0.01} < \frac{0.02-0.05}{0.01}) = P(Z < \frac{0.02-0.05}{0.01}) = 0.0013498980316300933$. Es decir, asociada a la cantidad aleatoria A_i, que sigue una distribución continua de tipo normal, nos quedamos con una propiedad binaria: problemas digestivos o no, que sigue una distribución de Bernouilli.

  El número de personas con problemas digestivos D sigue una distribución binomial con $n = 30$ y $p = 0.0013498980316300933$, porque estamos contando la cantidad de tripulantes que tienen problemas y cada tripulante es independiente. Nos piden $P(D > 2) = 1 - P(D \leq 2) = 9.717660913621806e - 06$.

- El número de misiones hasta un "fallo de misión" sigue una distribución geométrica, porque cada misión es independiente y tiene una probabilidad de fallo de misión igual a $9.717660913621806e - 06$. Calculamos los percentiles con `scipy.stats`: `G = st.geom(p=9.717660913621806e-06); percentil5=G.ppf(0.05); percentil95=G.ppf(0.95)` . Los percentiles son $p_{0.05} = 5279.0, p_{0.95} = 308276.0$.

Atención: La respuesta al siguiente ejercicio debe ser en *formato de vídeo*, como se indicó en la convocatoria de examen.

Question 4
En un trayecto en ferry el número de episodios de mareo en promedio es $N \cdot T \cdot \lambda$, donde

- N: pasaje
- T: tiempo de viaje
- λ: tasa de mareos, que depende de las condiciones de la mar

Cada vez que comienza el trayecto, el número λ es desconocido. Se asume que λ no cambia durante el trayecto. Una persona se puede marear más de una vez, y contaría como más de un episodio de mareo.
Tras tres horas de viaje con un pasaje de 100 personas, se han mareado 9 personas.

- Estima por máxima verosimilitud la tasa de mareos para ese viaje.
- Usando la estimación anterior, calcula la probabilidad de que en la siguiente hora de viaje se produzcan más de 10 mareos.
- Realiza una estimación bayesiana para λ, calcula el intervalo delimitado por los percentiles 5% y 95% para λ.

Explanation:

- Entendemos que los mareos son un proceso de Poisson con una tasa que se mide en mareos por persona y hora, ya que es igual de probable a priori que se maree una persona u otra en cualquier franja horaria con la misma longitud. La estimación de máxima verosimilitud es el número total de mareos 9 entre la "longitud del intervalo", que en este caso, para ponerlo en las mismas unidades, son 300 personas*hora. El resultado es 0.03.
- En la siguiente hora de navegación, contamos el total de mareos TM para un total de 100 personas*hora: sigue una distribución de Poisson con parámetro μ igual al producto de la tasa por 100 personas*hora: 3.0. Con `scipy.stats`: `TM = st.poisson(mu=0.03*100); pmasde10 = 1-M.cdf(10)`. La probabilidad pedida es 0.0002923369506473428.
- Para la tasa de un proceso de Poisson, elegimos la familia conjugada Gamma, y para el prior, escogemos por ejemplo el prior de Jeffreys, Gamma($\alpha = 1/2, \beta = 0$), que actualizamos con los datos disponibles de 9 mareos en 300 personas*hora: Gamma($\alpha = 1/2 + 9, \beta = 300$).

  Con `scipy.stats` podemos calcular por ejemplo la probabilidad de que λ sea mayor que $\frac{1}{10}$ o los percentiles 0.05 y 0.95: `tasa_bayes = st.gamma(a=0.5+9,scale=1/300); percentil5=tasa_bayes.ppf(0.05); percentil95=tasa_bayes.ppf(0.95); plambdamenor = tasa_bayes.cdf(0.1)`. Los percentiles son $p_{0.05} = 0.016861688439765075, p_{0.95} = 0.05023921200941027$, la probabilidad de que λ sea mayor que $\frac{1}{10}$ es 0.9999961301736994.