Guía de ejercicios

Clase 1

1. SeanX, Y dos v.a. i.i.d. $\mathcal{N}(0,1)$. Encontrar la expresión de la función de densidad de probabilidad conjunta.
   (a) ¿Cuál es la probabilidad de que $X > 0.7$ y $Y < 0.4$ simultáneamente?
   (b) Simular en Python

Para los ejercicios que siguen, simular las transformaciones y verificar los resultados obtenidos.

Método de la transformada inversa

1. Sea F una función de distribución de la forma:

$$F(x) = 1 - e^{-\sqrt{t/60}}\mathbf{I}\{t > 0\}$$

   Hallar una función h tal que

$$X = h(U) \sim F, \text{ con } U \sim \mathcal{U}(0,1)$$

2. Se arroja un dado equilibrado dos veces, sean X e Y los resultados de cada tiro, respectivamente. Hallar la función de probabilidad de $W = X + Y$. Sugerencia: hacer una tabla con los valores posibles de X e Y analizar que ocurre con W en cada caso.

Transformación de variables

1. Sea $X \sim \mathcal{N}(0,1)$. Hallar la función de densidad de $Y = X^2$. Identificar a que variable corresponde (nos va a ser de mucha utilidad en algunas clases!).
2. Sean $X_1, \dots X_n$ variables aleatorias i.i.d (independientes e idénticamente distribuidas) con distribución exponencial de parámetro λ. Hallar la distribución de $Y = \min(X_1, \dots, X_n)$. *Sugerencia: usar el método de sucesos equivalentes y pensar qué significa que* $\min(X_1, \dots X_n) > y$.

Clase 2

Transformación de variables parte2

1. Sean X e Y dos variables aleatorias independientes con funciones de densidad marginales:

$$f_X(x) = xe^{-x^2/2}\mathbf{1}\{0 \leq x\}, \qquad f_Y(y) = \frac{1}{2\pi}\mathbf{1}\{0 < y < 2\pi\}$$

   Sea $(U, V) = (X\cos(Y), X\sin(Y))$. Hallar las función de densidad conjunta del vector aleatorio (U, V).

2. (Extra) Sean X e Y dos variables independientes, cada una con distribución uniforme en el intervalo (0,1). Hallar la función de densidad conjunta de $U = \min(X, Y)$ y $V = \max(X, Y)$. *Sugerencia: este ejercicio se puede encarar por el método de sucesos equivalentes o por el método del Jacobiano generalizado.

3. (Extra) Sean $X \sim \mathcal{P}(\lambda)$ y $Y \sim \mathcal{P}(\mu)$ dos variables aleatorias independientes con distribución de Poisson. Probar utilizando la función generadora de momentos que $W = X + Y \sim \mathcal{P}(\lambda + \mu)$

Variables condicionadas

1. En un cierto día, una fábrica produjo 100 televisores, cada uno de los cuales está fallado con una probabilidad de 0.1. Antes de salir al mercado, cada televisor es sometido a ciertas pruebas, de forma tal que si estaba fallado, la falla se detecta con probabilidad 0.8. Sean X la cantidad de televisores fallados e Y la cantidad de televisores con fallas detectadas. Hallar la función de probabilidad de $Y|X = x$.

2. Sean $X \sim \mathcal{P}(\lambda)$ y $Y \sim \mathcal{P}(\mu)$ dos variables independientes. Hallar la distribución de $X|(X+Y) = m$. *Sugerencia: usar el resultado del ejercicio 5. de Transformaciones de variables.*

3. Sean X e Y dos variables con función de densidad conjunta $f_{X,Y}(x, y) = \frac{5}{8\pi}e^{-\frac{25}{32}(x^2 - \frac{6}{5}xy + y^2)}$. Hallar la función de densidad condicional de Y dada $X = x$. *Sugerencia: Qué distribución (conjunta) tienen X e Y?*

4. La velocidad del viento (X) y el promedio de ozono en la atmósfera (Y), son dos variables aleatorias con función de densidad conjunta

   $f_{X,Y}(x, y) = \lambda e^{-\lambda x}\mu e^{-\mu x}y^{-(1+\mu)}I\{x > 0, y > e^{-x}\}$

   Hallar la función de densidad de $Y|X = x$. Motivación del ejercicio

5. (Mezcla) En un sistema electrónico se debe determinar si se ha enviado señal o no. Se transmite señal $(S = 1)$ con probabilidad 0.6. Además, por fabricación, el medio introduce un ruido (N) con distribución normal de media nula y varianza 0.8, independiente de lo que se , de forma tal que

se recibe $X = S + N$, con $S = \{0, 1\}$. Hallar la probabilidad de haber enviado una señal sabiendo que se recibió $X = 0.7532$.

Clase 3

Esperanza condicional

1. Calcular la recta de regresión y esperanza condicional de Y dada X, para las variables de los ejercicios 1, 2 y 3 del tema variables condicionadas
2. Sean X, Y dos variables aleatorias tales que $X \sim \mathcal{U}(0, 10)$, e $Y|X = x \sim \mathcal{E}\left(\frac{1}{2+3x}\right)$. Hallar la varianza de Y.

Clase 4

Estadística Bayesiana

1. La cantidad de accidentes semanales en una planta industrial tiene una distribución de Poisson de media μ. En una muestra de 100 semanas se observaron las frecuencias:

Cant. de accidentes	0	1	2	3	4	5
Frecuencia	10	29	25	17	13	6

   A priori, μ tiene una distribución exponencial de media 2. En virtud de la información muestral stimar la probabilidad de que en la semana del 18 de diciembre de 2021 no ocurra ningún accidente en la mencionada planta.
2. Dada una muestra de tamaño n de una población X tal que $X|M = \mu \sim \mathcal{N}(\mu, \sigma_X^2)$, hallar el estimator MAP de M si se considera a priori que $M \sim \mathcal{N}(\mu_0, \sigma_\mu^2)$.
3. Se realiza un estudio para estimar la proporción de residentes en una ciudad y en sus suburbios que están a favor de la construcción de una planta de energía nuclear. Para ello, se entrevistaron 100 personas, de las cuales 58 dijeron estar a favor. Hallar el estimador Bayesiano correspondiente al error cuadrático para la proporción de residentes a favor de construir la planta de energía nuclear. Suponer una distribucón a priori Beta(2,5) para el parámetro. ¿Qué ocurre si considerábamos una Beta(5,2)? Analizar que significa cada una de esas dstribuciones a priori.

Estimador de máxima verosimilitud

1. Una moneda tiene una probabilidad de cara p, $p = \{2/5; 4/5\}$. En 10 lanzamientos de la moneda se observaron exactamente 3 caras. Estimar por máxima
   verosimilitud la probabilidad de que en otros tres lanzamientos se observe exactamente una cara.

2. Sea $\underline{X} = (X_1, \ldots, X_n)$ una muestra aleatoria de una variable X con distribución normal de media μ y varianza σ^2 desconocidas.

   (a) A partir del teorema de factorización encuentre el estadístico para $\theta = [\mu, \sigma^2]$.

   (b) Hallar el estimador de máxima verosimilitud para ambos parámetros.

   (c) Simular n realizaciones de una variable con distribución normal de media 5 y desvío 3, para $n = 10, 100, 1000, 10000$ y evaluar $\hat{\mu}$ y $\hat{\sigma}^2$. Qué puede concluir?

1. Hallar el estimador de máxima verosimilitud para la media de una variable con distribución exponencial.

Clase 5

Estimación no paramétrica

1. Cargar los datos del archivo 'winequality-red.csv'. Realizar un histograma con los datos de la columna 'density'.
2. Superponer al histograma la densidad estimada por kernel, utilizando un kernel Gaussiano.
3. Simular 1600 realizaciones de una normal. Utilizar como parámetros la media y desvío estándar muestrales de la columna 'density'. Graficar el histograma y la densidad estimada por kernel y comparar con los puntos 1 y 2. ¿Qué puede decir respecto de la variable 'density'?

Intervalos de confianza

1. Un emisor transmite una señal de valor μ. El receptor recibe mediciones $X_i = \mu + \epsilon_i$ donde ϵ_i son los errores de medición, que pueden considerarse variables aleatorias independientes con distribución $\mathcal{N}(0, 3)$.

   (a) Hallar un intervalo de confianza de nivel 0.95 para la señal transmitida μ basado en una muestra aleatoria de tamaño 9.

   (b) Si en una muestra de tamaño 9 se obtuvo un promedio de 3.48, hallar el intervalo de confianza de nivel 0.95 basado en la muestra observada.

   (c) Calcular el número mínimo de mediciones que debe recibir el receptor para poder construir un intervalo de confianza de nivel 0.95 para μ cuya longitud sea la mitad de la del intervalo hallado en el ítem anterior.

2. El punto de ebullición del agua (en grados Celcuis) es una variable aleatoria con distribución normal. En un laboratorio se realizaron 16 experimentos

independientes y se registraron los valores del punto de ebullición del agua, obteniendo un promedio de 95.67 y un desvío muestral estándar de 5.965.

(a) En base a la información muestral, construir un intervalo de confianza de nivel 0.95 para el punto medio de ebullición del agua.

(b) Si los verdaderos parámetros de la distribución son $\mu = 96$ y $\sigma = 5$, verificar que el inervalo hallado es un intervalo de confianza de nivel 0.95.

3. La duración (en horas) de cada lámpara de un lote es una variable aleatoria con distribución exponencial. Se pusieron a prueba 24 lámparas y se observó que el promedio de sus duraciones fue 24480 horas. En base a la información muestral, hallar una cota inferior de confianza de nivel 0.9 para la media de la duración de las lámparas del lote.

4. Sea X una variable aleatoria cuya función de densidad es de la forma

$$f_\theta(x) = \frac{3x^2}{\theta^3}\boldsymbol{I}\{0 < x < \theta\}, \quad \theta > 0.$$

Hallar una cota superior de confianza de nivel 0.95 pata θ basada en la muestra aleatoria: 2.85, 2.44, 3.93, 3.83. **Sugerencia:** encararlo similar al ejemplo de la v.a. uniforme visto en clase.

5. En el archivo `'Islander_data.csv'` se tiene información referente a un estudio acerca de un experimento en los efectos de un medicamento contra la ansiedad en un test de memoria cuando se expone a la persona ante recuerdos felices y tristes . Hallar un intervalo de confianza para la media de la diferencia de tiempos de reacción del test (columna 'Diff').

6. Se realiza un estudio para estimar la proporción de residentes en una ciudad y en sus suburbios que están a favor de la construcción de una planta de energía nuclear. ¿Qué tan grande debe ser una muestra, si se quiere una confianza de al menos 95%, de que la estimación estará a una distancia menor que 0.04 de la proporción real de residentes que están a favor de la construcción de la planta?

Clase 6

Test de hipótesis

1. De un experimento en los efectos de un medicamento ara la ansiedad se midió el puntaje de un test de memoria antes y después de tomar el medicamento.

   (a) A partir de los datos que se encuentran en el archivo `Islander_data.csv`, diseñar un test de hipótesis para decidir el tiempo medio de respuesta es diferente antes y después de tomar el medicamento.

(b) Hallar el p-valor.

2. Sea $\underline{X}$ una muestra aleatoria de tamaño n de una población con distribución uniforme sobre el intervalo $(0, \theta)$.

   (a) Diseñar un test de hipótesis de nivel 0.01 para decidir entre las hipótesis $H_0 : \theta = 3$ vs. $H_1 : \theta \neq 3$.

   (b) Si en una observación de tamaño $n = 10$ se observó que $\max(x_1, \dots, x_n) = 2.75$, ¿qué concluye?

3. En un concurso con 10 participantes, cada participante prueba una muestra de 3 vasos de bebida. Dos de los 3 vasos contienen la misma bebida marca a, mientras que el vaso restante contiene la bebida marca b.

   (a) Queremos determinar si la gente realmente puede discriminar la bebida b con un nivel de significancia de 5%.

   (b) ¿Cuál es la mínima cantidad de personas que deben identificar la bebida b para concluir que, en general, la bebida b es *claramente identificable* con respecto a la bebida a?

4. Se realiza un estudio para estimar la proporción de residentes en una ciudad y en sus suburbios que están a favor de la construcción de una planta de energía nuclear. En una muestra de 100 personas, se observó que la proporción de individuos que se encuentran a favor fue de 0.62.

   (a) ¿Puede decirse con un nivel de significación de 0.01 que la mayor parte de la población está a favor de la construcción de la planta nuclear?

   (b) Hallar el p-valor del test.

   (c) Hallar la cantidad de muestras necesarias para que la probabilidad de equivocarse cuando $p = 0.6$ sea como máximo 0.1.