Adicionales de la Parte I

1. Sea $X_1, \ldots, X_n$ una m.a. de una distribución $F(x, \theta)$ y sea $T(\mathbf{X})$ un estadístico suficiente para θ. Si $\Phi(\mathbf{X})$ es el test más potente de nivel α para
$$H : \theta = \theta_1 \qquad \text{vs} \qquad K : \theta = \theta_2$$
mostrar que $\Phi(\mathbf{X})$ está basado en $T(\mathbf{X})$, o esa depende de la muestra solo a través de $T(\mathbf{X})$.

2. Sea $X \sim C(\theta, 1)$, mostrar que el test
$$\Phi(X) = \begin{cases} 1 & \text{si } 1 < X < 3 \\ 0 & \text{en otro caso} \end{cases}$$
es el mejor para su nivel para testear
$$H : \theta = 0 \qquad \text{vs} \qquad K : \theta = 1.$$

3. Sean $Y_1, \ldots, Y_n$ v.a. independientes con distribución $\mathcal{U}(0.\theta)$. Muestre que el test dado por
$$\Phi(Y) = \begin{cases} 1 & \text{si} \quad Y^{(n)} > \theta_0 \ \text{ o } \ Y^{(n)} \leq \theta_0 \alpha^{\frac{1}{n}} \\ 0 & \text{si} \quad \theta_0 \alpha^{\frac{1}{n}} < Y^{(n)} \leq \theta_0 \end{cases},$$
es UMP de nivel α para testear $H_0 : \theta = \theta_0$ vs $H_0 : \theta \neq \theta_0$

4. Sean $X_1, \ldots, X_n$ es una m.a. de una distribución $\mathcal{U}(\theta, \theta + 1)$,
   (a) Hallar la distribución del estadístico suficiente $T = (T_1, T_2) = (X^{(1)}, X^{(n)})$.
   (b) Mostrar que existe un test UMP de nivel α para $H_0 : \theta \leq \theta_0$ vs $H_1 : \theta > \theta_0$. Y tiene la forma,
$$\Phi_0(t_1, t_2) = \begin{cases} 0 & \text{si } t_1 < k \text{ y } t_2 < 1 \\ 1 & \text{en otro caso} \end{cases},$$
   (c) Hallar el valor de k para obtener un test de nivel α.

5. Probar que las siguientes familias son de cociente de verosimilitud monótono:
   (a) $\mathcal{U}(\theta, \theta + 1)$
   (b) Las familias exponenciales a un parámetro con densidad o probabilidad
$$p(x|\theta) = a(\theta) h(x) \exp(c(\theta) t(x))$$
   siempre que $c(\theta)$ sea no decreciente. Luego las familias: binomial, binomial negativa, Poisson, Gamma, Beta y Normal (estas 3 ultimas considerando uno de los parámetros fijos) son de CVM.
   (c) $\mathcal{H}(n, m, M)$ donde m es el único parámetro desconocido, siendo
$$P_X(x) = \frac{\begin{pmatrix} m \\ x \end{pmatrix} \begin{pmatrix} M - m \\ n - x \end{pmatrix}}{\begin{pmatrix} M \\ n \end{pmatrix}}$$
   y $x = \max\{0, m + n - M\}, \ldots, \min\{m, n\}$

(d) La familia doble exponencial

$$f(x|\alpha,\beta) = \frac{1}{2\beta}\exp\left\{-\frac{|x-\alpha|}{\beta}\right\}$$

con β desconocido.

(e) La familia exponencial negativa

$$f(x|\alpha) = \exp\{-(x-\alpha)\}\, \mathrm{I}_{(\alpha,\infty)}(x)$$

C) Tests del cociente de máxima verosimilitud.

1. Se diseñó un sistema de riego de manera que el tiempo promedio de activación sea a lo sumo 25 segundos. Se lo probó 11 veces, obteniéndose los siguientes tiempos de activación:

   27 41 22 27 23 35 30 24 27 28 22

   Se supone que el tiempo de activación es una v.a. con distribución normal.

   (a) A nivel 0.05, ¿contradicen estos datos las especificaciones del sistema?

   (b) Acotar el valor p.

2. (a) Sea $X_1,\ldots,X_n$ una m.a. de una distribución $N(\mu_1,\sigma^2)$ e $Y_1,\ldots,Y_m$ una m.a. de una distribución $N(\mu_2,\sigma^2)$ independiente de la anterior. Si llamamos $\Delta = \mu_1-\mu_2$, probar que el test de CMV de nivel α para $H_0:\Delta=0$ vs $H_1:\Delta\neq 0$ rechaza H_0 si y sólo si

   $$\frac{\left|\bar{X}-\bar{Y}\right|}{S_p\sqrt{\frac{1}{n}+\frac{1}{m}}} > t_{n+m-2,\alpha/2}.$$

   (b) Un terreno se divide en 20 lotes semejantes. Se pone a prueba un nuevo fertilizante (I) colocándolo por aspersión en 10 lotes elegidos al azar. En los otros 10 lotes se utiliza como control un fertilizante de uso corriente (II). En cada parcela se planta el mismo número de plantas de tomate, y se observa el rendimiento (en kg/ha) en cada lote, obteniéndose

   $$\begin{array}{lllll} \bar{X} = 130 & & s_X = 10 & & \text{para el fertilizante I} \\ \bar{Y} = 120 & & s_Y = 8 & & \text{para el fertilizante II} \end{array}$$

   Se supone que los datos provienen de muestras aleatorias con distribución normal y varianzas iguales.

   i. Testear a nivel 0.05 si hay diferencia entre los rendimientos.

   ii. Acotar el valor p. ¿Qué decisión se hubiera tomado a nivel 0.01?

3. Sea $X_1,\ldots,X_n$ una m.a. de una distribución $N(\mu_1,\sigma_1^2)$ e $Y_1,\ldots,Y_m$ una m.a. de una distribución $N(\mu_2,\sigma_2^2)$ independiente de la anterior. Probar que el test de CMV de nivel α para $H_0:\sigma_1^2/\sigma_2^2=1$ vs $H_1:\sigma_1^2/\sigma_2^2\neq 1$ rechaza H_0 si y sólo si

   $$\frac{s_1^2}{s_2^2} > k_1 \text{ ó } \frac{s_1^2}{s_2^2} < k_2$$

   para ciertas constantes k_1 y k_2 adecuadas. (Por simplicidad, se toma $k_1 = f_{n-1,m-1,\alpha/2}$ y $k_2 = f_{n-1,m-1,1-\alpha/2}$.)

4. El riesgo de inversiones alternativas es evaluado generalmente por la varianza de las ganancias asociadas a cada inversión. La distribución de las ganancias anuales de dos inversiones alternativas A y B se supone normal. En base a la información de 10 años pasados en el caso de la inversión A y de 15 años en el caso de la inversión B, se calcularon las varianzas estimadas de las ganancias para las dos inversiones, obteniéndose $s_A^2 = 3.2$ y $s_B^2 = 7.5$. ¿Puede decirse a nivel 0.05 que la inversión B es tan riesgosa como la A?¿Y a nivel 0.10?

5. Se tomaron aleatoriamente 8 pares de pollos hermanos y se eligió al azar dentro de cada par cual iba a la muestra A y cual a la B. Los pollos de la muestra A han sido criados en encierro, mientras que los de la muestra B al aire libre. Los datos siguientes corresponden al peso de los pollos de 7 semanas:

Muestra A	255	481	360	368	425	283	311	368
Muestra B	226	425	311	311	255	340	311	284

Suponiendo normalidad de los datos,

(a) ¿existe alguna diferencia en el peso medio a las 7 semanas entre los métodos de crianza? Testear a nivel 0.01.

(b) Acotar el valor p.

D) Tests con nivel de significación asintótico.

1. Se supone que en cierta población de insectos la cantidad de hembras es igual a la de machos. Para poner a prueba esta creencia, un entomólogo tomó una muestra de 100 de estos insectos, resultando machos 43 de ellos.

   (a) Utilizando un test de nivel asintótico, decidir si hay suficiente evidencia a nivel 0.05 para rechazar la suposición. Calcular el valor p aproximado.

   (b) Si el porcentaje de hembras fuera 0.55, ¿cuál sería la probabilidad de tomar una decisión incorrecta?

2. El número de personas que llega a la boletería de una estación de trenes entre las 14 y 14:30 hs. tiene distribución de Poisson. El jefe de la estación está evaluando la posibilidad de abrir otra ventanilla en ese horario y considera que es necesario sólo si la media de arribos es superior a 20 individuos. Además, quiere que la probabilidad de abrir una nueva ventanilla cuando en realidad no es necesario sea a lo sumo 0.01. Durante 45 días se observa el número de personas que arriba en dicho horario y se calcula una media muestral de 21 individuos.

   (a) Proponer un test para este problema. ¿Qué decisión toma en base a estos datos? ¿Cuál es el valor p aproximado?

   (b) ¿Cuál es la probabilidad de dejar una sola boletería cuando en realidad la media verdadera es de 22 personas?

   (c) Si se hubiera querido que la probabilidad calculada en (b) fuera a lo sumo 0.05, ¿qué tamaño muestral debería haberse tomado?

3. Se ha tomado una muestra aleatoria de tamaño 30 de los días que demora una empresa en instalar una línea telefónica a partir del momento de su solicitud. Los datos obtenidos fueron los siguientes:

$$\begin{array}{ccccccccccccccc} 3, & 4, & 5, & 3, & 2, & 7, & 6, & 4, & 6, & 4, & 7, & 2, & 3, & 4, & 6, \\ 7, & 3, & 5, & 6, & 6, & 4, & 5, & 5, & 7, & 3, & 2, & 2, & 4, & 3, & 5. \end{array}$$

La compañía, por contrato, debe tener un valor medio de demora no mayor que 4 días. El gobierno le inició juicio a la compañía porque cree que ésta no cumplió con el contrato. El juez considera que su veredicto debe ser hecho de manera tal que la probabilidad de fallar contra la compañía cuando ésta ha cumplido con lo estipulado en el contrato sea aproximadamente 0.05.

   (a) Suponiendo que la distribución de los tiempos de espera es $\mathcal{E}(\lambda)$, encontrar el test asintótico en el cual el juez debe basar su decisión y el veredicto al que arribaría con la muestra dada.
   (b) Expresar la función de potencia del test en términos de una función de distribución conocida.
   (c) ¿Qué tamaño de muestra debería utilizarse si se quiere que la probabilidad de condenar a la compañía sea 0.99 cuando el valor medio de la demora es de 5 días?

4. Un dado se tira 600 veces y se obtienen los siguientes resultados:

Número	**1**	**2**	**3**	**4**	**5**	**6**
Frecuencia	87	96	108	89	122	98

¿Puede decirse que el dado no es balanceado? Calcular el valor p aproximado.

E) Tests IUMP.

1. Generalización de Neyman-Pearson: sean $(f_k(x))_{k=0}^{n}$ funciones integrables en $I\!R$ cualquier función integrable de la forma

$$\Phi_0(X) = \begin{cases} 1 & \text{si} \quad f_0(x) > \sum_{i=1}^{n} k_i f_i(x) \\ \gamma(x) & \text{si} \quad f_0(x) = \sum_{i=1}^{n} k_i f_i(x) \\ 0 & \text{si} \quad f_0(x) < \sum_{i=1}^{n} k_i f_i(x) \end{cases},$$

$0 \leq \gamma(x) \leq 1$. Sea $\Phi_0(X)$ maximiza $\int \Phi(X)\ f_0(x)\ dx$ entre las funciones Φ, $0 \leq \Phi \leq 1$, tales que:

$$\int \Phi(x) f_j(x)\, dx = \int \Phi_0(x) f_j(x)\, dx, \quad j = 1, \ldots, n.$$

Si $k_j \geq 0$, $j = 1, \ldots, n$, entonces maximiza entre las funciones tales que:

$$\int \Phi(x) f_j(x)\, dx \leq \int \Phi_0(x) f_j(x)\, dx.$$

2. Sean $X_1, \ldots, X_n$ es una m.a. de una distribución Paretto, o sea la densidad de X_1 esta dada por

$$f(x, \theta) = \theta x^{-\theta-1} \mathrm{I}_{(1,+\infty)}(x) \quad 0 < \theta$$

Hallar el test IUMP para:

(a) $H_0 : \theta = \theta_0$ vs $H_1 : \theta \neq \theta_0$

(b) $H_0 : \theta \leq \theta_0$ vs $H_1 : \theta > \theta_0$

3. Sean $X_1, \ldots, X_n$ es una m.a. de una distribución $\mathcal{P}(\lambda)$ hallar el test IUMP para:

   (a) $H_0 : \lambda = \lambda_0$ vs $H_1 : \lambda \neq \lambda_0$

   (b) $H_0 : \lambda \leq \lambda_0$ vs $H_1 : \lambda > \lambda_0$

4. Sean $\mathbf{X}$ e $\mathbf{Y}$ v.a. distribución conjunta

$$f_{\mathbf{XY}}(x, y) = \lambda\mu \exp\{-\lambda x - \mu y\} \mathrm{I}_{(0,+\infty)}(x)\, \mathrm{I}_{(0,+\infty)}(y)$$

   Hallar un test insesgado UMP de nivel 0.2 para:

   (a) $H_0 : \lambda \leq \mu + 1$ vs $H_1 : \lambda > \mu + 1$

   (b) $H_0 : \lambda = \mu$ vs $H_1 : \lambda \neq \mu$

   (c) $H_0 : \lambda \leq 2\mu$ vs $H_1 : \lambda < 2\mu$

5. Sean $\mathbf{X}$ e $\mathbf{Y}$ v.a. independientes con distribución geométrica

$$p_{\mathbf{XY}}(x, y) = (1 - \theta_1)(1 - \theta_2)\theta_1^x \theta_2^y$$

   $x \in I\!N \cup \{0\}$ $y \in I\!N \cup \{0\}$ siendo $0 < \theta_1 < 1$ $0 < \theta_2 < 1$ Hallar un test UMP de nivel 0.2 para testear:

   (a) $H_0 : \theta_1 \leq \theta_2$ vs $H_1 : \theta_1 > \theta_2$

   (b) $H_0 : \theta_1 = \theta_2$ vs $H_1 : \theta_1 \neq \theta_2$

   Para qué funciones $\varphi(\theta_1, \theta_2)$ los métodos para tests UMP insesgados garantizan la existencia de un test UMP insesgado para testear $H_0 : \varphi(\theta_1, \theta_2) = 0$ vs $H_1 : \varphi(\theta_1, \theta_2) \neq 0$?

6. Sean $\mathbf{X}_1, \ldots, \mathbf{X}_n$ y $\mathbf{Y}_1, \ldots, \mathbf{Y}_m$ m.a. independientes de distribuciones $N(\mu, \sigma^2)$ y $N(\eta, \sigma^2)$ respectivamente. Hallar un test UMP insesgado de nivel α para testear $H_0 : \mu \leq \eta$ vs $H_1 : \mu > \eta$