Estadística en Física Experimental (1er cuatrimestre de 2017)

Guía de Problemas N° 9 | Tests de Hipótesis

1. Se desea testear si un conjunto de n mediciones $\{x_i, y_i \pm \sigma_i\}$, con errores gaussianos en y, es consistente con cierta predicción teórica $y = f(x)$. Para ello se construye el estadístico $S = \sum_i [(y_i - f(x_i))/\sigma_i]^2$, cuya distribución es χ^2_n si la hipótesis es correcta (¿por qué?), y se calcula $p \equiv P(\chi^2 > S)$, denominada la "probabilidad χ^2" del resultado.
   (a) Muestre que si X es una variable aleatoria con densidad de probabilidad $f_X(t)$, entonces la variable Y, definida como $Y \equiv \int_{-\infty}^{x} f_X(t)\, dt$, tiene distribución uniforme en [0,1].
   (b) Usando 1a encuentre cual es la distribución de la variable aleatoria valor-P bajo la suposición de que la teoría es correcta, y muestre que el valor esperado de esta probabilidad es 50%.
   (c) Se decide usar como criterio para rechazar la hipótesis que p sea menor que 0.05. Exprese en palabras cuál es el suceso cuya probabilidad es 5%.
   (d) ¿Qué le sugiere o que concluye si el resultado del test hubiera sido $p \geq 99\%$?

2. El bosón W es una partícula elemental, intermediaria de la fuerza débil y descubierta en 1983. Su masa ha sido medida por diversos experimentos, con distintos niveles de precisión. Los datos de la edición 2000 del Review of Particle Properties son, en MeV:

Masa W	Experimento
80.482 ± 0.091	DØ
80.423 ± 0.112	Aleph
80.38 ± 0.12	Opal
80.61 ± 0.15	L3
80.41 ± 0.18	CDF

   Se desea testear si los resultados son consistentes entre sí. Escriba un test a partir de lo discutido en el problema anterior y calcule el valor-P. [Rta: P=0.1172]

3. Considere dos muestras independientes $X_1, \ldots, X_m$ e $Y_1, \ldots, Y_n$ con distribuciones gausianas $N(\mu_X, \sigma^2)$ y $N(\mu_Y, \sigma^2)$, μ_X, μ_Y y σ desconocidas. Si quisiéramos testear a un nivel de significancia α_0 la hipótesis $H_0 :\ \mu_X \leq \mu_Y$ contra $H_1 :\ \mu_X > \mu_Y$, se puede demostrar que el test de mayor potencia, derivado del de cociente de verosimilitudes (LRT), corresponde a rechazar la hipótesis H_0 cuando $U \geq T_{m+n-2,(1-\alpha_0)}$, siendo

$$U = (\bar{X} - \bar{Y})\sqrt{\frac{m+n-2}{(1/m + 1/n)(s_X^2 + s_Y^2)}}, \quad \text{con} \quad s_X^2 = \sum_i^m (X_i - \bar{X})^2 \quad \text{y} \quad s_Y^2 = \sum_i^n (Y_i - \bar{Y})^2$$

   y $T_{m+n-2,(1-\alpha_0)}$ el cuantil $1 - \alpha_0$ de la distribución t-Student con $n + m - 2$ grados de libertad.

   Se mide el módulo de Young Y de un conjunto de aceros templados (A) y sin templar (B) disponibles en el mercado, para determinar si Y es mayor para los últimos. Se obtiene A: 0.8 0.9 1.05 1.2 1.3 1.3; y B: 1.0 1.4 1.7 1.9 2.3.

   Escriba la hipótesis nula y la alternativa, indique la zona crítica para una significancia $\alpha = 0.05$, determine el valor-P de las mediciones y obtenga sus conclusiones. [Rta: P=0.0148]

4. Debido a dudas sobre la *gaussianidad* de los datos del ejercicio anterior, se les aplica el test de Wilcoxon.
   (a) Indique la zona crítica y el valor-P de las mediciones. [Rta: P=0.0274]
   (b) Si bien el test-t tiene más potencia, es también más dependiente de la suposición de normalidad. Para estudiar la robustez de ambos suponga que se agrega una medición sin sentido a la muestra B (20, por ejemplo, un 'outlier'). Muestre que Wilcoxon esencialmente no cambia, mientras el test-t ahora da acuerdo con la hipótesis nula. Discuta por qué ocurre ésto, teniendo en cuenta que, aunque poco creíble, el dato agregado refuerza en realidad la evidencia de mayor Y para aceros templados.

5. Se desea determinar si los siguientes tres conjuntos de mediciones independientes son consistentes. Considere (1) normalidad de los datos; (2) de ignorancia de su distribución (test de Kruskal-Wallis).
   A: 6.4 6.8 7.2 8.3 8.4 9.1 9.4 9.7
   B: 2.5 3.7 4.9 5.4 5.9 8.1 8.2
   C: 1.3 4.1 4.9 5.2 5.5 8.2

6. Los individuos con una cierta enfermedad pueden presentar un síntoma X y un síntoma Y. Para confirmar la hipótesis que al ocurrir uno hay alta probabilidad que ocurra el otro, un investigador examina 20 pacientes. Encuentra que hay 11 que no presentan el síntoma X, y de ellos 9 tampoco presentan el síntoma Y. Por otro lado, entre los 9 que presentan X, 6 presentan Y. Los datos apuntan entonces en la dirección de la hipótesis, pero es necesario establecer si la medición obtenida puede ser fruto de una fluctuación estadística. Para ello se aplica el test exacto de Fisher, con una significancia de 0.05.

   (a) Escriba la tabla de contingencia de las mediciones y dé los otros 8 posibles resultados del experimento. Calcule sus probabilidades y sus grados de asimetría, $|a/(a+b) - c/(c+d)|$. A partir de ésto encuentre las regiones críticas con nivel de significación 0.05 para ambos tests unilaterales y para el test bilateral.

	X	$\bar{X}$	
Y	a	b	n_Y
$\bar{Y}$	c	d	$n_{\bar{Y}}$
	n_X	$n_{\bar{X}}$	n

   (b) Discuta si corresponde un test unilateral o bilateral, y muestre que el resultado permite rechazar la hipótesis nula con el grado de significación establecido.

   (c) Calcule el valor-P para el test unilateral y bilateral [rtas: 0.03989 y 0.06478]. Note que si el test hubiera sido bilateral no se podría haber establecido correlación al 95% CL.

   (d) Proponga como debería haber sido un problema para que corresponda hacer el test bilateral.

7. *Tests de Kolmogorov y de Cramer-von Misses.* Se desea investigar la hipótesis de que una muestra $\{x_i\}$, con $x_i \in \mathcal{R}$ e i=1,n, proviene de una cierta distribución $F(x)$. Para ello se ordena la muestra en orden creciente $x_i < x_{i+1}$ y se define la distribución $S : \mathcal{R} \to [0,1]$

$$S(x) = \begin{cases} 0 & x < x_1 \\ i/n & x_i \leq x < x_{i+1} \qquad i = 1, \ldots, n-1 \\ 1 & x_n \leq x \end{cases}$$

   El test de Kolmogorov cuantifica la discrepancia entre datos y $F(x)$ tomando la máxima diferencia entre $S(x)$ y $F(x)$, $T_K \equiv \max|S(x)\text{-}F(x)|$, mientras que Cramer-vonMisses usa el valor cuadrático medio de esta diferencia $T_C \equiv \int (S(x) - F(x))^2 f(x)\, \mathrm{d}x$, con $f(x)=\frac{\mathrm{d}F}{\mathrm{d}x}$.

   Considere una muestra de 150 mediciones. Usando la tabla respectiva, determine en que rango deben estar las variables aleatorias T_K y T_C para decidir rechazar la hipótesis con un 99% de confianza.

8. Dos experimentos miden el mismo observable obteniéndo los siguientes resultados ordenados en forma creciente:
   Experimento A: 81 82 87 93 102 104 108 112 116 122 125 131 131 133 134 139 139 142 144 146 152 156 182 202 206 216 226 270
   Experimento B: 50 64 68 76 79 83 88 96 97 98 99 103 105 107 113 114 115 126 128 130 132 138 150 169 171
   Se desea investigar si ambos experimentos son consistentes. Muestre que el run test no encuentra incompatibilidad, dando una probabilidad de 0.21 (¿qué quiere decir ésto?) mientras que los tests de Kolmogorov y de Wilcoxon permiten rechazar la hipótesis de consistencia con un CL de 98% y 99%, respectivamente.

9. Las estrellas en formación en galaxias jóvenes se pueden clasificar en 3 grupos según su velocidad de rotación: (1) lentas, (2) normales, (3) rápidas. Una hipótesis es que la velocidad angular está relacionada con las características del disco de acreción, ya que hay estrellas jóvenes: (a) sin disco, (b) con disco cercano, (c) con disco lejano. Se midieron para ello 189 estrellas en el cluster NGC2264 obteniéndo la siguiente tabla:

	(1)	(2)	(3)	Tot
(a)	20	15	12	47
(b)	24	27	32	83
(c)	14	22	23	59
Tot	58	64	67	189

   Los valores totales en la última fila y la última columna se obtuvieron sumando las respectivas filas o columnas de la tabla de 3×3. Llamemos p_{ix}, p_i, p_x, con $i = 1, 2, 3$ y $x = a, b, c$, a las probabilidades que una estrella sea del tipo ix, i o x, y sea H_0: $p_{ix} = p_i\, p_x$, $i = 1, 2, 3$ y $x = a, b, c$

   Considerando que todos los n_{ix} son suficientemente grandes tal que vale la aproximación poisson$\to$gauss, diseñe un test de H_0 basado en la estadística χ^2 y determine el valor-P correspondiente a los datos obtenidos. Justifique cuidadosamente el número de grados de libertad usados para obtener el valor-P, y explique en palabras que concluye de su resultado.

10. Sea $\{X_i\}$ una muestra de una distribución uniforme en $[0, \theta]$. Se desea testear $H_0 : \theta \leq 3$ vs $H_1 : \theta > 3$.

   (a) Muestre que para cada nivel de significancia α existe un test UMP que especifica como zona crítica $\max(X_1, \ldots, X_n) \geq c$. Exprese c en función de α.

   (b) Para un dado n y α haga un sketch a mano alzada de la función de potencia del test.

   (c) Repita los dos items anteriores si la hipótesis a testear hubiera sido $H_0 : \theta \geq 3, H_1 : \theta < 3$.

11. La eficiencia de un detector en condiciones normales es 50%, y para verificar que no caiga significativamente se testea periódicamente $H_0 : \varepsilon = 1/2$ vs $H_1 : \varepsilon < 1/2$ realizando 40 ensayos y midiendo la fracción de fallos. Suponga que para $n = 40$ ya es válida la aproximación normal $\hat{\varepsilon} \sim N(\varepsilon, \sqrt{\varepsilon(1-\varepsilon)/n})$.

   (a) Detemine el criterio de rechazo para un nivel de significación 0.05

   (b) Supóngase que si la eficiencia baja de 33% los datos se vuelven básicamente inusables. Bajo esta consideración los investigadores necesitan una alta probabilidad de detectar una disminución de tal magnitud, si ésta se produce. ¿Cuál es la potencia de nuestro test si el valor real de la eficiencia es $\varepsilon = {}^1\!/_3$?

   (c) A la vista de la importancia que adopta no superar el límite de ${}^1\!/_3$, ¿qué tamaño muestral deberíamos elegir para garantizar que la potencia del test sea 0.97? [Rta: 106]