Estadística en Física Experimental

Guía de Problemas No.9 (2do cuatr 2009)

Tests de hipótesis

1. a. Considere dos muestras independientes $X_1, \ldots, X_m$ e $Y_1, \ldots, Y_n$ con distribuciones gausianas $N(\mu_1, \sigma^2)$ y $N(\mu_2, \sigma^2)$, μ_1, μ_2 y σ desconocidas. Queremos testear a un nivel de significancia α_0 la hipótesis $H_0 : \mu_1 \leq \mu_2$ contra $H_1 : \mu_1 > \mu_2$. Muestre que el test de mayor potencia, el de cociente de verosimilitudes (LRT), corresponde a rechazar la hipótesis H_0 cuando $U \geq T_{m+n-2,(1-\alpha_0)}$, siendo
$$U = (\bar{X} - \bar{Y})\sqrt{\frac{m+n-2}{(1/m+1/n)(S_X^2 + S_Y^2)}} \qquad \text{con } S_X^2 = \sum_i^m (X_i - \bar{X})^2 \ \text{ y } \ S_Y^2 = \sum_i^n (Y_i - \bar{Y})^2$$
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.
b. Se mide la 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 primeros. Se obtiene A: 1.0 1.4 1.7 1.9 2.3 y B: 0.8 0.9 1.05 1.2 1.3 1.3. Escriba las hipótesis nula y 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].

2. a. Debido a dudas sobre la gausianidad de los datos, se aplica el test de Wilcoxon a los datos del problema anterior. 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 tambien 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 porqué ocurre ésto, teniendo en cuenta que, aunque poco creíble, el dato agregado refuerza en realidad la evidencia de mayor Y para aceros templados.

3. El test-t del problema 1 supone igual varianza para ambas muestras. Apoyan los datos esta suposición, con un CL de 95%?

4. 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 $H_0 : \theta \geq 3, H_1 : \theta < 3$.

5. 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 del 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 se debería elegir para garantizar que la potencia del test sea 0.97? [Rta: 106]

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 fluctuacion 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 de 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
$\bar{Y}$	c	d

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. Exprese como deberia haber sido la hipótesis para que corresponda hacer el test bilateral.

7. Se desea determinar si los siguientes tres conjuntos de mediciones independientes son consistentes. Considere (1) normalidad de los datos; (2) ignorancia sobre 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

8. 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 la estadística $S = \sum_i [(y_i - f(x_i))/\sigma_i]^2$, cuya distribución es χ^2_n si la hipótesis es correcta (porqué?), y se calcula $p \equiv P(\chi^2 > S)$, denominada la "probabilidad χ^2" del resultado (con qué otro nombre conocemos este número?).
a. Muestre que si x es una variable aleatoria con densidad de probabilidad $f_x(t)$, entonces la variable y, definida como $y \equiv \int_{-\infty}^{x} f_x(t)\,\mathrm{d}t$, tiene distribución uniforme en [0,1] (primer parcial, de nuevo).
b. Usando (a) encuentre cual es la distribución de la variable aleatoria p bajo la suposición 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 cual es el suceso cuya probabilidad es 5%.

9. El bosón W es una partícula subatómica, 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 $\pm$ 0.091	DØ
80.423 $\pm$ 0.112	Aleph
80.38 $\pm$ 0.12	Opal
80.61 $\pm$ 0.15	L3
80.41 $\pm$ 0.18	CDF

a. Analice si los resultados son consistentes y muestre que el resultado combinado es 80.458±0.054. Note que, como era de esperar, el error obtenido es más pequeño que el menor de los errores individuales
b. ¿Cuál es el máximo error que debería tener un nuevo experimento, si tomamos como criterio para que valga la pena realizarlo que disminuya el error global en al menos 20%?

10. Se desea investigar si los datos experimentales representados en la figura (a), y disponibles en la página web de la materia, son consistentes con una dependencia lineal.
a. Muestre que mientras que el test χ^2 sugiere que la hipótesis debe ser aceptada, el test runs permite rechazarla con un nivel de confianza de 99%. Verifique a su vez este resultado utilizando la aproximación gaussiana para la distribución del número de runs, ésto es, $P(r \leq r_a) = \Phi[(r_a + 0.5 - E(r))/\sqrt{\mathrm{Var}(r)}]$
b. Los datos de este problema fueron generados con fluctuaciones gaussianas con σ=0.3 alrededor de la cuadrática $y(x)$=1+x+0.3x^2. Grafique conjuntamente la parábola, la recta ajustada y los datos, y discuta porqué el test runs resulta más efectivo en este caso que el χ^2.

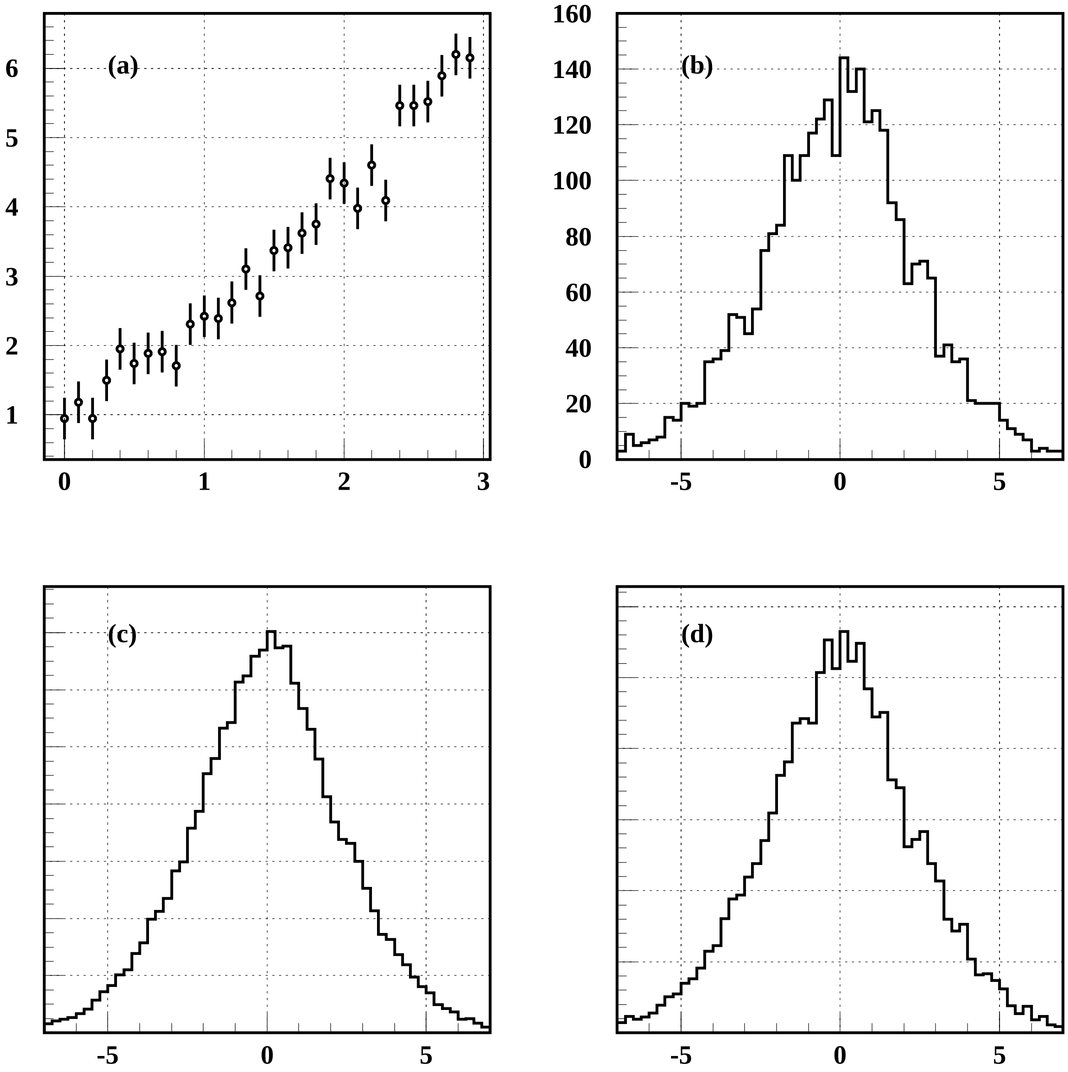
11. Sobre un conjunto de datos se realiza el test χ^2 y el test runs, obteniéndose χ^2_a y r_a, con respectivas probabilidades $P_1 = P(\chi^2 > \chi^2_a)$ y $P_2 = P(r \leq r_a)$.
a. Muestre que y_1=-2 ln P_1 tiene distribución χ^2_2. Discuta bajo que condiciones también y_2 = -2 ln P_2 es $y_2 \sim \chi^2_2$. ¿Qué distribución tendrá entonces y=-2 ln$(P_1 P_2)$, dado que ambos tests son independientes?
b. Si los resultados numéricos son $P_1 = 0.04$ y $P_2 = 0.04$, cuál es el valor-P de ambos tests combinados?

12. Dos experimentos miden una misma variable aleatoria (la masa invariante en MeV de pares de fotones producidos en colisiones de antineutrinos con núcleos pesados), obteniéndose 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.

Ejercicios para entregar (sólo estudiantes de doctorado)

13. El propósito de este ejercicio es comparar distintos tests de hipótesis y mostrar cómo incrementan su sensibilidad al aumentar la estadística disponible. Se cuenta con una muestra $\{x_i\}$, i=1,100000, disponibles en la página web de la materia, y deseamos testear si proviene de una distribución N(0,2.5)
a. Las figuras (b), (c) y (d) representan histogramas de $\{x_i\}$ con distintas cantidades de eventos. Por simple inspección ocular, ordénelas según el número de entradas.
b. Considere solamente los primeros 3000 elementos de $\{x_i\}$:
(*i*) presente los datos en un histograma de 56 bines de tamaño 0.25 entre -7.0 y 7.0. Muestre que al aplicar sobre éste el test χ^2 se obtiene una probabilidad de 10% (qué quiere decir ésto?).
(*ii*) Aplique sobre los mismos primeros 3000 datos *sin binear* los tests de Kolmogorov y de Cramer-vonMisses y verifique que, a diferencia del test χ^2, éstos permiten rechazar la hipótesis a un nivel de confianza de 98%.
c. Repita (a) pero tomado cantidades crecientes de eventos, y observe como la significancia de los tests crece rapidamente, permitiendo rechazar la hipótesis básicamente sin lugar a dudas. Repita a su vez los tests para menos de 3000 eventos, y observe que en este caso los datos devienen compatibles con la hipótesis propuesta. i

14. a. Usando el programa escrito para el ejercicio 2 de la guía 6, calcule el valor de χ^2 respecto de la recta teórica para cada una de las 1000 repeticiones del experimento, preséntelos en un histograma y grafique sobre él la distribución teórica correspondiente.
b. Repita el punto anterior, pero con el χ^2 respecto de la recta ajustada. Haga además dos histogramas de las probabilidades χ^2, uno usando el número correcto de grados de libertad ν, y uno usando ν=11, el número de puntos. Discuta los resultados obtenidos.


(a)
(b)
(c)
(d)
0
1
2
3
1
2
3
4
5
6
0
20
40
60
80
100
120
140
160
-5
0
5