# Estadística en Física Experimental

1$^{er}$ cuatrimestre de 2011

## Guía de Problemas No.5

### Función característica y generatriz - Teorema Central del Límite

1. Demuestre que:
   (a) La función característica de la gausiana es $\phi_X(t) = \exp(it\mu - t^2\sigma^2/2)$ y la de la poissoniana es $\phi_n(t) = \exp[\lambda(e^{it} - 1)]$.
   (b) La distribución poissoniana tiende a la gausiana en el límite $\mu \to \infty$. Para ello obtenga la función característica de $Y \equiv (n - E(n))/\sigma_n$, con $n$ poissoniana, y verifique la validez de $\lim_{\mu\to\infty} \phi_Y(t) = \phi_X(t)$, con $X(t)$ gausiana canónica.

2. **Ejercicio para entregar.** Estudie el grado de validez del teorema central del límite dibujando las distribuciones siguientes, y superponiendo sobre ellas la gaussiana con el $\mu$ y $\sigma$ correspondiente.
   (a) B(5,0.2), B(30,0.4)
   (b) P(4), P(10), P(40)

3. El teorema central del límite permite evaluar probabilidades binomiales sin necesidad de sumar muchos términos que involucran factoriales de grandes números, a partir de la distribución normal canónica $\Phi(x)$,
$$\sum_{k=a}^{b} \binom{n}{k} p^k q^{n-k} \;\simeq\; \Phi\left(\frac{b - np + \frac{1}{2}}{\sqrt{npq}}\right) - \Phi\left(\frac{a - np - \frac{1}{2}}{\sqrt{npq}}\right) \tag{1}$$
   Discuta el origen de esta fórmula y utilícela para calcular la probabilidad de aprobar un examen multiple choice con 100 preguntas de tres opciones cada una, si se contesta al azar y se aprueba con 4 (40 % de respuestas correctas). [Rta: 0.08]

4. Utilizando el teorema central del límite escribir un generador aproximado de números gausianos N(0,1), a partir de variables aleatorias independientes $\{X_i\}$ con distribución uniforme en [0,1], como una función $f(Z)$ siendo $Z = \sum_i^n X_i$.
   (a) Si se elige $n$=50, cuál debe ser $f(Z)$?
   (b) ¿En qué rango de la abscisa seguro falla la aproximación a la normal?
   (b) Genere de este modo 10000 números con la computadora, haga un histograma de su distribución, y grafique N(0,1) sobre éste.

5. A partir de la función característica para la distribución $\chi^2$ con un grado de libertad, $\phi(t)=1/\sqrt{1-2it}$, muestre que la esperanza y la varianza para el caso con $\nu$ grados de libertad valen $E(\chi^2_\nu)=\nu$ y $\mathrm{Var}(\chi^2_\nu)=2\nu$.

6. Se realiza un experimento para estudiar una cierta ley física, $y = f(x)$. Para 50 valores distintos $x_i$ se mide $y_i$, con errores gausianos de varianza $\sigma_i^2$, y se calcula la variable aleatoria $S = \sum_{i=1}^{50} [(y_i - f(x_i))/\sigma_i]^2$, la cual crece cuanto más difieran los datos experimentales $y_i$ de las respectivas predicciones teóricas $f(x_i)$.
(a) Bajo la suposición de que la ley física es válida, qué valor espera obtener para $S$?
(b) Utilizando la aproximación dada por el teorema central del límite, calcule dentro de qué rango espera encontrar a $S$ con un 95 % de probabilidad. Verifique la validez de esta aproximación utilzando la tabla de la función distribución $\chi^2_{(n)}$.
(c) Al realizarse el experimento se obtiene $S = 80$, y se concluye que la ley física no es válida. ¿Cuál es la probabilidad de que esta conclusión sea errónea y que un valor de $S \geq 80$ sea obtenido debido a una fluctuación estadística de los datos?

7. La función generatriz $\phi(\mathbf{t})$ correspondiente a la distribución multinormal $f(\mathbf{x})$ es:

$$f(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^n |\mathbf{V}|}} \exp\left[-\frac{1}{2}(\mathbf{x}-\mu)^T \mathbf{V}^{-1}(\mathbf{x}-\mu)\right] \quad \phi(\mathbf{t}) = \exp\left(i\,\mathbf{t}^\mathbf{T}\mu - \frac{1}{2}\,\mathbf{t}^\mathbf{T}\mathbf{V}\mathbf{t}\right) \qquad (2)$$

(a) Muestre que $\mu$ y $\mathbf{V}$ son en efecto la esperanza y la matriz de covarianza de la variable aleatoria multidimensional $\mathbf{x}$.
(b) Considere $m$ variables aleatorias $\mathbf{y}$, obtenidas como función lineal $\mathbf{y} = \mathbf{C}\mathbf{x}$ a partir de $n$ variables $\mathbf{x}$ con distribución conjunta multinormal, siendo $\mathbf{C}$ de $m \times n$. Muestre que $\mathbf{y}$ tiene también distribución multinormal con covarianza $\mathbf{V}' = \mathbf{C}\mathbf{V}\mathbf{C}^T$, lo cual coincide con la que se obtiene vía propagación de errores.
(c) Sea $\mathbf{S}$ la matriz ortogonal que diagonaliza $\mathbf{V}$ para $n$ arbitrario. Muestre que la distribución de $\mathbf{z} = \mathbf{S}\mathbf{x}$ corresponde a $n$ variables gaussianas independientes. A partir de ésto encuentre la distribución de la variable aleatoria $Q = (\mathbf{x}-\mu)^T \mathbf{V}^{-1}(\mathbf{x}-\mu)$ que aparece en el exponente de la multinormal.

8. Aplicando los resultados del ejercicio anterior para el caso de $\mathbf{x}$ bidimensional,
(a) Considere las elipses de covarianza encerradas dentro del rectángulo $(\pm k\sigma_1, \pm k\sigma_2)$ alrededor de $(\mu_1, \mu_2)$. Muestre que la probabilidad conjunta de que $(x_1, x_2)$ se encuentre dentro de una de estas elipses con $k$=1 es 39.3 %, independientemente del valor de la correlación $\rho$ (este resultado es el equivalentente al 68.3 % obtenido para el caso $n$=1)
(b) Cuánto debería ser $k$ para que la elipse corresponda a un nivel de confianza de 95 %? Verifique que este resultado puede obtenerse también analíticamente, además de usando las tablas.
(c) ¿Por qué tiene más sentido considerar la elipse como rango de confianza, que el propio rectángulo $(\pm k\sigma_1, \pm k\sigma_2)$?

9. Muestre que el promedio de $N$ variables con distribución de Cauchy tiene a su vez distribución de Cauchy. ¿Por qué falla en este caso el teorema central del límite?

10. (a) Encuentre la fórmula que relaciona la varianza de una variable aleatoria discreta $k$ con las derivadas primera y segunda de su función generatriz $G_k(z) = E(z^k)$.
(b) Halle $G_k(z)$ para la poissoniana $P_k(\mu)$, y verifique que reobtiene $E(k)$=Var$(k)$=$\mu$.
(c) Considere la suma $S_N = \sum X_i$ de $N$ variables aleatorias independientes $X_i$ de función generatriz $F_X(z)$, y supongamos que $N$ es a su vez una variable aleatoria de función generatriz $G(z)$. Muestre que la función generatriz de $S_N$ es $G(F_X(z))$.
(d) La distribución "Poisson compuesta" corresponde a $X_i$ y $N$ poissonianas, $P_X(\mu)$ y $P_N(\lambda)$. Halle su función generatriz y muestre que $E(S_N)$=$\lambda\mu$ y Var$(S_N)$=$\lambda\mu(1+\mu)$.
(e) Discuta un caso físico de aplicación de la distribución poisson compuesta y analice por qué tiene mayor varianza que una poissoniana de igual esperanza.