Estadística en Física Experimental (1[er] cuatrimestre de 2015)

Guía de Problemas N° 6 | Función característica y generatriz – Teorema central del límite

La función característica.

1. Demuestre que:
   (a) La función característica de la gaussiana $N(\mu, \sigma)$, 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 gaussiana en el límite $\lambda \to \infty$. Para ello obtenga la función característica de $Y \equiv (n - E(n))/\sigma_n$, con n poissoniana, y verifique la validez de $\lim_{\lambda\to\infty}\phi_Y(t) = \phi_X(t)$, con $X(t)$ gaussiana canónica.

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

3. Una distribución multinormal

$$f_X(\underline{\mathrm{x}}) = \frac{1}{\sqrt{(2\pi)^n|\mathbb{V}|}} \exp\left[-\frac{1}{2}\left(\underline{\mathrm{x}} - \underline{\mu}\right)^T \mathbb{V}^{-1}\left(\underline{\mathrm{x}} - \underline{\mu}\right)\right]$$

tiene función característica $\phi(\underline{t})$ dada por

$$\phi(\underline{t}) = \exp\left(i\,\underline{t}^T\underline{\mu} - \frac{1}{2}\,\underline{t}^T\mathbb{V}\,\underline{t}\right)$$

Muestre que $\underline{\mu}$ y $\mathbb{V}$ son en efecto la esperanza y la matriz de covarianza de la variable aleatoria multidimensional $\underline{\mathrm{x}}$. $|\mathbb{V}|$ indica el determinante de $\mathbb{V}$.
Ayuda: alcanza con mostrar que μ_l es la esperanza de x_l (la componente l-ésima de $\underline{\mathrm{x}}$) y que $\mathbb{V}_{kl}$ (el elemento kl de la matriz $\mathbb{V}$) es la covarianza de x_k con x_l.

Teorema central del límite.

4. **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 μ y σ correspondiente.
   (a) $\mathrm{B}_k(5{,}0.2)$, $\mathrm{B}_k(30{,}0.4)$
   (b) $\mathrm{P}_n(4)$, $\mathrm{P}_n(10)$, $\mathrm{P}_n(40)$

5. 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 acumulativa normal canónica $\Phi(x)$,

$$\sum_{k=a}^{b} B_k(n,p) = \sum_{k=a}^{b}\binom{n}{k}p^k q^{n-k} \simeq \Phi\left(\frac{b - np + \frac{1}{2}}{\sqrt{npq}}\right) - \Phi\left(\frac{a - np - \frac{1}{2}}{\sqrt{npq}}\right)$$

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: 0.0966 con la suma exacta, y 0.0951 con la fórmula aproximada.]

6. Utilizando el teorema central del límite escribir un generador aproximado de números gaussianos 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?
   (c) Genere de este modo 10000 números con la computadora, haga un histograma de su distribución, y grafique N(0,1) sobre éste.

7. 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 gaussianos de varianza σ_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? [Rta: $p \simeq 0.5\%$]

(d) Se repite otras 14 veces el experimento (c/u con 50 mediciones), obteniendo en todos los casos $S < 80$. ¿Cuál es la probabilidad en este caso de que la medición $S \geq 80$ se debiera a una fluctuación? [Rta: $p \simeq 7\%$]

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

9. *Modelos de Incerteza.*

   (a) Considere la medición de un observable Z, que es sensible a dos tipos de factores aleatorios: X_i e Y_i, donde cada una de estas variables pueden tener cualquier distribución, y digamos que las esperanzas de los X_i son controladas por el experimentador y las de Y_i son constantes. O sea $Z \sim f_Z(X_1, X_2, \ldots; Y_1, Y_2, \ldots)$. Probar que las fluctuaciones de Z son aproximadamente gaussianas si el número de factores aleatorios es muy grande.

   Comentario: la distinción entre X_i e Y_i es irrelevante para probar el ejercicio, pero está más cerca de lo que ocurre en la práctica. Por ejemplo en la medición de la ley de Ohm $V = IR$, Z sería la caída de tensión en una resistencia, X_1 sería la corriente inyectada, X_2 el valor nominal de la resistencia, y los Y_i serían por ejemplo: la impedancia de entrada del voltímetro, la tensión de la fuente de corriente, los potenciales de contacto, la velocidad de discipación de temperatura en la resistencia (que puede depender del flujo de aire), la temperatura de los cables, etc.

   (b) Considere un observable que depende de variables aleatorias de la forma $Z = X_1 X_2 \ldots$, donde las variables aleatorias X_i solo pueden tomar valores positivos. Mostrar que Z fluctua con un error lognormal, es decir $\delta Z \sim \exp(\alpha + \beta Y)$, con Y una distribución normal.

***La función generatriz*.**

10. Encuentre la fórmula que relaciona la varianza de una variable aleatoria discreta k con las derivadas primera y segunda de su función generatriz: $G_k(z) = E(z^k) = \sum_k z^k P_k$.

11. Halle la función generatriz $G_k(z)$ para la poissoniana $P_k(\mu)$, y verifique que reobtiene $E(k)$=Var(k)=μ.

12. Considerar la diferencia entre dos variable aleatorias $D = X_1 - X_2$, donde ambas variables tienen la misma función generatriz $G_X(z)$.

    (a) Probar que la función generatriz de la diferencia es $G_D(z) = G_X(z) G_X(1/z)$.

    (b) Encuentre la esperanza de D.

13. Considere la suma $S_N = \sum_i^N X_i$ de N variables aleatorias independientes X_i (idénticamente distribuidas) de función generatriz $G_X(z)$, y supongamos que N es a su vez una variable aleatoria de función generatriz $G_N(z)$.

    (a) Muestre que la función generatriz de S_N es $G_{S_N}(z) = G_N(G_X(z))$.
    Ayuda: puede ser útil la propiedad 1c de la guía 4.

    (b) La distribución "Poisson compuesta" corresponde a X_i y N poissonianas, $P_X(\mu)$ y $P_N(\lambda)$.

       i. Halle su función generatriz y muestre que $E(S_N)$=$\lambda\mu$ y Var(S_N)=$\lambda\mu(1+\mu)$.

       ii. Discuta un caso físico de aplicación de la distribución poisson compuesta y analice por qué tiene mayor varianza que una poissoniana de igual esperanza.