Estadística en Física Experimental

2$^{\text{er}}$ cuatrimestre de 2011

Guía de Problemas No.4

Distribuciones multidimensionales - Covarianza y Correlación

1. Demuestre las siguientes propiedades que involucran a la matriz de covarianza.
   (a) $\mathrm{Cov}(X,Y) = E(XY) - E(X)E(Y)$
   (b) $\mathrm{Var}(aX+bY) = a^2\mathrm{Var}(X) + b^2\mathrm{Var}(Y) + 2ab\,\mathrm{Cov}(X,Y)$
   (c) $\mathrm{Var}(a_1X_1 + a_2X_2 + ... + a_NX_N) = \sum_{i=1}^{N} a_iX_i + 2\sum_{\{i;j=1:i<j\}}^{N} a_ia_j\mathrm{Cov}(X,Y)$

2. Se realizan N mediciones X_i de una variable aleatoria cuya distribución tiene esperanza μ y varianza σ^2 (es decir N repeticiones del mismo experimento). Muestre que el error del promedio $\bar{X} = \sum_i X_i/N$ es $\sigma/\sqrt{N}$ si las mediciones son independientes. Discuta la diferencia entre el error de cada medición y el error del promedio. ¿Cuánto vale el error del promedio si las N mediciones están completamente correlacionadas?

3. Sea X una variable aleatoria con densidad de probabilidad simétrica alrededor de cero. Muestre que X e $Y{=}X^2$, pese a no ser independientes, tienen correlación nula.

4. Muestre que la suma de dos distribuciones uniformes es la distribución triangular:
   (a) Si X e Y son variables independientes con distribución uniforme en [0,1], halle la distribución conjunta $g(U,V)$ de $U \equiv X+Y$ y $V \equiv X-Y$.
   (b) Tomando la correspondiente distribución marginal, muestre que U tiene densidad de probabilidad $f_U(t) = t$, $0{<}t{<}1$, y $f_U(t) = 2-t$, $1{<}t{<}2$.
   (c) Encuentre la distribución de V y determine si U y V son independientes.
   (d) Calcule la varianza de U via $\int_0^2 (t-1)^2 f_U(t)\mathrm{d}t$ y compárela con la que se obtiene usando la fórmula de propagación de errores.

5. Demuestre que la suma de gaussianas es gaussiana: si X e Y son variables independientes con distribución normal de parámetros (μ_1,σ_1) y (μ_2,σ_2), entonces $Z = X+Y$ es una gaussiana de parámetros $(\mu_1+\mu_2, \sqrt{\sigma_1^2+\sigma_2^2})$. Ahora bien, si la suma de gaussianas es gaussiana, cómo es que en la guía 3 mostró que con la suma de dos gausianas se consigue aproximar una distribución de Cauchy, que claramente no es gausiana?

6. Muestre que el cociente $Z \equiv X/Y$ de dos variables independientes con distribución normal canónica tiene distribución de Cauchy, $f_Z(t) = 1/[\pi(1+t^2)]$. ¿Qué ocurre en este caso si se utiliza la fórmula de propagación de errores para obtener $\mathrm{Var}(Z)$?

7. Sea (X, Y) una variable aleatoria bidimensional donde $X=\{x_1, x_2\}$ e $Y=\{y_1, y_2\}$, y considere la siguiente asignación de probabilidades: $P(x_1, y_1) = P(x_2, y_2) = p$ y $P(x_1, y_2) = P(x_2, y_1) = q$. Indique qué condiciones deben satisfacer p y q para que:
(a) P sea una medida de probabilidad.
(b) X e Y sean independientes.

8. Sean X e Y dos variables independientes con distribución uniforme en [0,1], a partir de las cuales se definen $U = \sqrt{-2a \ln X} cos(2\pi Y)$ y $V = \sqrt{-2a \ln X} sin(2\pi Y)$. Encuentre la distribución conjunta $g(U, V)$, identifique qué distribución es, indique el significado del parámetro a y determine si U y V son independientes.

9. *Método de Monte Carlo.* Para generar números pseudoaleatorios con una distribución arbitraria $f(t)$, en un dominio $[t_0, t_1]$ en el que f está acotada (siendo f_0 su valor máximo) se procede de la siguiente manerageneran dos números al azar y y z con distribución uniforme en $[0, 1]$. A partir de y y z se determinan $u = t_0 + y(t_1 - t_0)$ y $v = f_0 z$. Muestre que u tiene distribución uniforme en $[t_0, t_1]$ y v tiene distribución uniforme en $[0, f_0]$.
(b) Si $v \leq f(u)$, se incluye $x = u$ en la muestra de números generados, sino se descarta. Se repite el procedimiento hasta obtener la cantidad de números deseada. Demuestre que la densidad de probabilidad de x es f.
(c) ¿Es necesario conocer la constante de normalización de f para utilizar este procedimiento?
(d) Muestre que la fracción de números que se incluye en la muestra, respecto del total de números generados, depende del área bajo la curva de f. ¿Qué concluye acerca de la eficiencia del método?

10. **Ejercicio para entregar (Licenciatura: sólo el punto d).** *Generación al azar de variables multidimensionales por el método de Monte Carlo.* Se requiere simular numéricamente el comportamiento de una esfera de fluído (inicialmente en reposo) con una densidad

$$\rho(r) = \frac{\rho_0}{1 + (\frac{r}{r_e})^2} \quad (0 < r < r_e)$$

donde ρ_0 es la densidad central y r_e el radio de la esfera. Para ello se generan como condiciones iniciales las posiciones $\overrightarrow{r}$ de 10000 elementos de fluído con una distribución proporcional a la masa (i.e., la probabilidad de encontrar un elemento de fluído en un determinado dV alrededor de $\overrightarrow{r}$ es proporcional a ρdV), y luego se calcula la evolución temporal de las mismas integrando las ecuaciones de la hidrodinámica.Utilice el método descripto en el ejercicio 9 para generar las condiciones iniciales del experimento en coordenadas esféricas (considere $r_e = 1$). Para ello:
(a) Escriba la densidad de probabilidad conjunta de r, θ y φ (las coordenadas esféricas usuales).
(b) Calcule las distribuciones de probabilidad marginales de cada una de las coordenadas

por separado. ¿Son estas independientes?
(c) Genere 10000 valores al azar de cada variable, usando el método de Monte Carlo. Grafique los histogramas de las distintas variables, y superponga las distribuciones correspondientes.
(d) Haciendo un cambio de variables calcule la densidad de probabilidad conjunta de R, ϕ y z (las coordenadas cilíndricas usuales). Encuentre la distribución marginal de R y compárela con el correspondiente histograma.

11. La *Distribución Multinormal* es la generalización a n dimensiones de la gaussiana y, al igual que ésta, juega un rol preponderante en probabilidades y estadística. Dadas n variables aleatorias correlacionadas $\{x_i\}$, con esperanza $E(x_i) = \mu_i$ y matriz de covarianza V, ésto es Cov(x_i,x_j)=V_{ij}, se dice que su densidad de probabilidad conjunta $f(\mathbf{x})$ es multinormal si todas las distribuciones marginales $f(x_i)$ y todas las distribuciones condicionales unidimensionales $f(x_i|x_j, j \neq i)$ son gaussianas. La densidad de probabilidad conjunta $f(\mathbf{x})$ viene dada por

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

donde $\mathbf{x}$, $\mathbf{t}$, μ son matrices columna de $1\times n$, $\mathbf{x^T}$, $\mathbf{t^T}$, $\mu^\mathbf{T}$, las respectivas traspuestas de $n\times 1$, y $\mathbf{V}$ es cuadrada definida positiva de $n\times n$, con $|\mathbf{V}| \equiv \det(\mathbf{V})$.

(a) Verifique que para $n = 1$, $f(\mathbf{x})$ es una gaussiana.

(b) Para el caso $n = 2$, la matriz de covarianza depende de 3 parámetros (¿por qué?). Elijamos σ_1, σ_2 y el coeficiente de correlación ρ, ésto es, $V_{11} = \sigma_1^2$, $V_{22} = \sigma_2^2$ y $V_{12} = \rho\,\sigma_1\sigma_2$. Muestre entonces que

$$f(x_1, x_2) = (2\pi\sigma_1\sigma_2\sqrt{1-\rho^2})^{-1} \exp\,(-\frac{Q}{2})$$

con

$$Q = \frac{1}{1-\rho^2}\left[\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2 - 2\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right)\right]$$

(c) Compruebe que cuando $\rho = 0$, $f(x_1, x_2) = N(\mu_1, \sigma_1)N(\mu_2, \sigma_2)$. Esto es, para la multinormal, correlación nula implica que las variables son independientes.

(d) Muestre que la distribución marginal $f(x_2)$ es la gaussiana $N(\mu_2, \sigma_2)$, independientemente del valor del nivel de correlación ρ.

(e) Una manera de visualizar la forma de $f(x_1, x_2)$ es dibujar curvas de nivel de f en el plano x_1, x_2. Considere las correspondientes a $Q = 1$, y muestre que son elipses centradas en (μ_1, μ_2), denominadas *elipses de covarianza*. Para $\mu_1 = \mu_2 = 0$, verifique que éstas están contenidas en el rectángulo $(\pm\sigma_1, \pm\sigma_2)$, que son tangentes a dicho rectángulo en los puntos $(\sigma_1, \rho\,\sigma_2)$ y $(\rho\,\sigma_1, \sigma_2)$ y que su eje principal forma un ángulo ϕ con el eje x_1 dado por $\tan 2\phi = 2\rho\,\sigma_1\sigma_2\,/(\sigma_1^2 - \sigma_2^2)$.

(f) Dibuje a mano alzada elipses de covarianza con distintos ρ para el caso $\sigma_1 = \sigma_2$. Discuta

la diferencia entre tomar como error para x_1 el rango máximo cubierto por la elipse sobre el eje x, o el segmento entre los puntos de intersección de la elipse con el eje x.

(g) Muestre que $f(x_2|x_1)$ es gaussiana, con $N(\mu_2 + \rho(\sigma_2/\sigma_1)(x_1 - \mu_1), \sigma_2\sqrt{1-\rho^2})$. Discuta cómo varía la esperanza de x_2 en función de x_1 según el signo de ρ, y analice cómo varía el ancho de la distribución condicional con el grado de correlación. Interprete estos resultados cortando con líneas x=cte las elipses dibujadas a mano alzada en el item anterior. ¿Qué ocurre en el caso límite $\rho = 1$?