

Estadística en Física Experimental (1^{er} Cuatrimestre 2025)

Guía de Problemas N° 4 | Distribuciones multidimensionales – Covarianza y Correlación

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

2. La suma de dos variables aleatorias con distribución uniforme da lugar a una variable aleatoria con 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 es una variable aleatoria con distribución triangular:

$$f_U(t) = \begin{cases} t & 0 < t < 1 \\ 2 - t & 1 < t < 2 \\ 0 & \text{en otro caso} \end{cases}$$

(c) Encuentre la distribución de V y determine si U y V son independientes.

(d) Calcule la varianza de U via $\int_0^2 (t-1)^2 f_U(t) dt$.

(e) Confirme que se obtiene el mismo resultado usando la propiedad $\text{Var}(a_1 X_1 + a_2 X_2 + \dots + a_N X_N) = \sum_{i=1}^N a_i^2 \text{Var}(X_i) + 2 \sum_{i=1}^N \sum_{j>i}^N a_i a_j \text{Cov}(X_i, X_j)$.

3. La suma de gaussianas es gaussiana:

(a) Probar que 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 $(\mu_1 + \mu_2, \sqrt{\sigma_1^2 + \sigma_2^2})$.

(b) Ahora bien, si la suma de gaussianas es gaussiana, ¿cómo es que con la suma de dos gaussianas se consigue aproximar muy bien una distribución de Cauchy, que claramente no es gaussiana?

Nota: este ítem acompañaba a un ejercicio computacional, que ya no está en la guía, que proponía aproximar la distribución de Cauchy $f(x) = \frac{1}{\pi} \frac{1}{1+x^2}$ como $f(x) \approx p \cdot N(x|0, \sigma_1) + (1-p) \cdot N(x|0, \sigma_2)$

4. Muestre que el cociente $Z \equiv X/Y$ de dos variables independientes con distribución normal canónica tiene distribución de Cauchy, $f_Z(t) = 1/[\pi(1+t^2)]$.

5. 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.

Nota: este mapeo de $X, Y \rightarrow U, V$ se llama transformación de Box-Muller.

6. Para cada uno de los cuatro pares de datos de la tabla:

(a) Calcular la media muestral de X y de Y ,

(b) Calcular la varianza muestral de X y de Y ,

(c) Calcular la correlación entre X e Y : $\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)} \sqrt{\text{Var}(Y)}} = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}$;

(d) Graficar cada par de puntos.

I		II		III		IV	
x	y	x	y	x	y	x	y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89

Nota: estos son cuatro conjuntos de datos que F.J. Anscombe generó para mostrar que hacer buenos gráficos de los datos son una parte esencial del análisis de regresión lineal. F. J, Anscombe, (1973). "Graphs in Statistical Analysis". Am Stat, Vol. 27, No. 1, 17-21

7. Considerando las siguientes propiedades:

(a) La esperanza $E[\cdot]$ es un operador lineal, es decir,

$$E[aX + bY] = aE[X] + bE[Y]$$

(b) La covarianza $\text{Cov}(\cdot, \cdot)$ es un operador bilineal, es decir, es un operador lineal en ambos argumentos:

$$\text{Cov}(aX + bY, C) = a \text{Cov}(X, C) + b \text{Cov}(Y, C)$$

$$\text{Cov}(C, aX + bY) = a \text{Cov}(C, X) + b \text{Cov}(C, Y)$$

(c) La covarianza es simétrica:

$$\text{Cov}(X, Y) = \text{Cov}(Y, X)$$

(d) La covarianza de una variable consigo misma es su varianza:

$$\text{Cov}(X, X) = \text{Var}(X)$$

demuestre que

$$\begin{aligned} \text{Var} \left(\sum_{i=1}^N a_i X_i \right) &= \sum_{i=1}^N \sum_{j=1}^N a_i a_j \text{Cov}(X_i, X_j) \\ &= \sum_{i=1}^N a_i^2 \text{Var}(X_i) + 2 \sum_{i=1}^N \sum_{j>i}^N a_i a_j \text{Cov}(X_i, X_j) \end{aligned}$$

(o muestre para la suma de 2 o 3 variables).

8. Se tiene una variable aleatoria X con una varianza $\text{Var}(X)$ conocida. A partir de esta, se calculan dos variables: $Y = cX$ y $Z = (c + 1)X$, donde c es una constante. Finalmente, se calcula $D = Z - Y$. Se desea calcular la varianza de D . Proceda de dos maneras:

(a) Reescriba D en términos de X y calcule la varianza.

(b) Calcule la varianza de Y y Z y, a partir de estas, la varianza de D . ¿Qué sucede si ignora la covarianza entre Y y Z ?

(c) ¿Y si $D = Z + Y$?