

Statistics and Data Analysis

Evaluation Exercise for Parameter Estimation

Let us consider the following pdf:

$$f(x) = 4x^2 \left((3 - 3x) + \frac{2}{3}\rho(4x - 3) \right)$$

with x limited to the range $[0, 1]$

1. Check analytically that $f(x)$ is a properly normalized pdf. Find analytical expressions for the *distribution* (or *population*) mean and variance, as a function of ρ .
2. Produce a Monte Carlo sample following the pdf $f(x)$, assuming $\rho = 0.75$, with $N = 1000$ entries.
3. Compute the *sample* mean and variance of the Monte Carlo sample from exercise 2 and compare them to the values obtained in exercise 1. (Remember that the sample mean and variance are estimators of the distribution mean and variance, respectively). Compute also the variance of the mean estimator.
4. The expression of the distribution mean obtained in exercise 1 can be written as $\langle x \rangle = a + b\rho$. Generate 5 Monte Carlo samples, for $\rho = 0, 0.25, 0.5, 0.75, 1.0$ respectively, each with N entries, and use the estimated means (and associated uncertainties) to estimate the values of a and b , and their covariance matrix, using the least-squares method. Do it analytically (taking into account that $\langle x \rangle$ is linear in a and b) and by numerical minimization of the χ^2 expression. Compare both results between them and with the values obtained in exercise 1.
5. Since the sample mean, \bar{x} , is an estimator of the population mean, we can use the relation between the population mean and ρ from exercise 1 to define an estimator for ρ , $\hat{\rho}_1$:
 $\bar{x} = a + b\hat{\rho}_1$
 - (a) Find analytical expressions for the variance and bias of $\hat{\rho}_1$.
 - (b) For $\rho = 0.75$, compute numerically the variance of $\hat{\rho}_1$, using: $n = 100$ independent Monte Carlo simulations of N entries each.
 - (c) Compare the values obtained in parts 5a and 5b.
6. Let $\hat{\rho}_2$ be the maximum likelihood estimator of ρ for $\mathcal{L}(\rho|\mathbf{x}) = \prod_{i=1}^N f(x_i|\rho)$. Assuming one of the Monte Carlo samples generated in exercise 5 is the result from the real experiment, compute:
 - (a) The estimated value $\hat{\rho}_2$
 - (b) The variance of the estimator $\hat{\rho}_2$ using the $\log \mathcal{L}_{\min} + 1/2$ rule, compare it with the values obtained in exercise 5.
 - (c) The 95% confidence level interval for the estimated $\hat{\rho}_2$ value