

Homework 4

ALECK ZHAO

September 29, 2016

1. Consider a finite population of size N of objects, and suppose for each object i there are a pair of measurements (x_i, y_i) . Suppose we obtain a sample of size n of these pairs of values from this population, where we sample without replacement. Let \bar{X} and \bar{Y} be the sample means of the x -measurements and the y -measurements, respectively. We claimed in lecture that

$$\text{Cov}(\bar{X}, \bar{Y}) = \frac{\sigma_{sy}}{n} \left(1 - \frac{n-1}{N-1} \right). \quad (1)$$

Here σ_{xy} is the population covariance, defined by

$$\sigma_{xy} = \frac{1}{N} \sum_{i=1}^N (x_i - \mu_x)(y_i - \mu_y)$$

where μ_x and μ_y are the population means, respectively, of the x - and y -measurements. Prove that Equation 1 holds.

2. Let X_1, \dots, X_n be an iid sample from some distribution having mean μ and variance σ^2 . Recall that the CLT says that $\sqrt{n}(\bar{X} - \mu)$ has an approximate $N(0, \sigma^2)$ distribution as n tends to infinity. Let g be a smooth function with $g'(\mu) \neq 0$.
 - (a) Write down a first order Taylor's expansion with remainder for g about μ .
 - (b) Substitute \bar{X} for x in the expansion in (a) and use this to express $\sqrt{n}(g(\bar{X}) - g(\mu))$ as a sum of two terms, one involving $(\bar{X} - \mu)$ and the other involving $(\bar{X} - \mu)^2$.
 - (c) Explain why the linear term should have a distribution that is approximately $N(0, g'(\mu)^2 \sigma^2)$ as n tends to infinity. (Hint: If $Y \sim N(0, \sigma^2)$ what is the distribution of cY where c is a constant?)
 - (d) Explain why the quadratic term should tend to zero as n tends to infinity.
3. Suppose X_i are iid Poisson random variables with parameter λ .
 - (a) Determine the distribution of the sum $S = \sum_{i=1}^N X_i$. (Hint: consider the case $n = 2$ and then use induction.)
 - (b) Obtain the method of moments estimator for λ .
 - (c) Assuming n is large, obtain an approximate 95% confidence interval for λ . Your confidence interval should be in terms of your confidence level, your estimator, n , etc.
 - (d) Using R, generate a vector of 1000 Poisson random variables with parameter $\lambda = 4$ (you can use the `rpois` command for this). Now suppose you had this vector of data, but you did not, in fact, know λ . Generate a specific 95% confidence interval for λ based on this data.
 - (e) Now suppose that you don't see all of the data. What is the only piece of information from the sample that you need in order to generate the confidence interval in the previous part? Do you really need all n values from the sample?

4. Recall that the PDF for a gamma distribution is of the form

$$f(x; \alpha, \lambda) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}, x > 0$$

where α and λ are unknown positive parameters. Recall that the first two moments μ_1 and μ_2 can be expressed in terms of α and λ and are given by

$$\mu_1 = \frac{\alpha}{\lambda}$$

$$\mu_2 = \frac{\alpha(\alpha + 1)}{\lambda^2}$$

In lecture, we showed how this system of equations can be successfully inverted to express α, λ in terms of μ_1, μ_2 :

$$\alpha = \frac{\mu_1^2}{\mu_2 - \mu_1^2}$$

$$\lambda = \frac{\mu_1}{\mu_2 - \mu_1^2}$$

- (a) Suppose $\{X_1, \dots, X_n\}$ are an iid sample from the gamma distribution with parameters α and λ . In terms of these quantities, what is the exact mean and variance of $\hat{\mu}_1$ and $\hat{\mu}_2$? Also what is the exact covariance between $\hat{\mu}_1$ and $\hat{\mu}_2$?
 - (b) Using the delta method, what is the approximate mean and variance of $\hat{\alpha}$? Considering these expressions as functions of n , which contributes more toward the MSE of $\hat{\alpha}$: the squared bias or the variance, for large values of n ?
 - (c) Using the delta method, what is the approximate mean and variance of $\hat{\lambda}$? Considering these expressions as functions of n , which contributes more toward the MSE of $\hat{\lambda}$: the squared bias or the variance, for large values of n ?
 - (d) Using the delta method, what is the approximate covariance between $\hat{\alpha}$ and $\hat{\lambda}$? (Hint: use a linear approximation to relate $\hat{\alpha}$ to $\hat{\mu}_1$ and $\hat{\mu}_2$ and do the same for $\hat{\lambda}$, then compute the covariance between the approximations)
5. In lecture, we described how to use simulation to obtain approximate distributions of our estimators and to determine an estimated standard error. In this exercise, we address this. First, consider fixing the values of α, λ in our gamma distribution at $\alpha = 12$ and $\lambda = 4$, and also fixing n , the sample size, at $n = 200$. In R, complete the following steps:
- a) First, do the following for 10,000 trials. In each trial, sample the gamma distribution using a sample of size 200 (use the `rgamma()` command) and compute $\hat{\mu}_1$ and $\hat{\mu}_2$. Store the results from each trial as an entry in a vector. You will thus obtain a vector of 10,000 realizations of values of $\hat{\mu}_1$ and 10,000 realizations of $\hat{\mu}_2$.
 - b) Next, for each pair of $(\hat{\mu}_1, \hat{\mu}_2)$ values, compute the corresponding values of $\hat{\alpha}$ and $\hat{\lambda}$, and once again store each result as an entry in a vector. You will thus obtain a vector of 10,000 realizations of values of $\hat{\alpha}$ and 10,000 of $\hat{\lambda}$. These vectors of realizations represent independent samples from the sampling distributions of these quantities, and we can use them to understand these distributions.
 - c) Compute the sample mean and variance of the vector of $\hat{\mu}_1$ values and compare with the theoretical values you computed in the previous problem (Problem 2 (a)). Similarly, compute the sample mean and variance of the vector of $\hat{\mu}_2$ values and compare. Finally, compute the sample covariance between the vectors of $\hat{\mu}_1$ and $\hat{\mu}_2$ values and compare.
 - d) Compute the sample mean and variance of the vector of $\hat{\alpha}$ values and compare with the values in Problem 2 (b), and compute the sample mean and variance of the vector of $\hat{\lambda}$ values and compare with the values in Problem 2 (c).

- e) Compute the sample covariance between the vectors of $\hat{\alpha}$ and $\hat{\lambda}$ values and compare with the theoretical value in Problem 2 (d).
6. Recall that we said an estimator $\hat{\theta}_n$, which is based on a sample of size n , is *consistent* for a parameter θ_0 if for any $\varepsilon > 0$,

$$P\left(|\hat{\theta}_n - \theta_0| > \varepsilon\right) \rightarrow 0$$

as $n \rightarrow \infty$. Let $X_i, 1 \leq i \leq n$ be an iid collection of random variables with $E[X_i] = \mu$ and $\text{Var}(X_i) = \sigma^2$. Show that the sample mean \bar{X} is consistent for μ . (Hint: Use Chebyshev's inequality)

Chapter 8: Estimation of Parameters and Fitting of Probability Distributions

4. Suppose that X is a discrete random variable with

$$P(X = 0) = \frac{2}{3}\theta$$

$$P(X = 1) = \frac{1}{3}\theta$$

$$P(X = 2) = \frac{2}{3}(1 - \theta)$$

$$P(X = 3) = \frac{1}{3}(1 - \theta)$$

where $0 \leq \theta \leq 1$ is a parameter. The following 10 independent observations were taken from such a distribution: (3, 0, 2, 1, 3, 2, 1, 0, 2, 1).

- Find the method of moments estimate of θ .
 - Find an approximate standard error for your estimate.
5. Suppose that X is a discrete random variable with $P(X = 1) = \theta$ and $P(X = 2) = 1 - \theta$. Three independent observations of X are made: $x_1 = 1, x_2 = 2, x_3 = 2$.
- Find the method of moments estimate of θ .
 - What is the likelihood function?
 - What is the maximum likelihood estimate of θ ?
7. Suppose that X follows a geometric distribution,

$$P(X = k) = p(1 - p)^{k-1}$$

and assume an iid sample of size n .

- Find the method of moments estimate of p .
 - Find the MLE of p .
 - Find the asymptotic variance of the MLE.
16. Consider an iid sample of random variables with density function

$$f(x|\sigma) = \frac{1}{2\sigma} \exp\left(-\frac{|x|}{\sigma}\right)$$

- Find the method of moments estimate of σ .
- Find the MLE of σ .
- Find the asymptotic variance of the MLE.