14.380: Statistics

Problem Set 3

due Tuesday, October 1, 2019

You should hand in the solution for problem 4 to David. Problems 1- 3 are for practice.

1. Let $X_1, \dots, X_n$ be a random sample from a Poisson distribution with parameter λ

$$P\{X = j\} = \frac{e^{-\lambda}\lambda^j}{j!} \quad j = 0, 1, \dots$$

   (a) Find the MLE of λ and its asymptotic distribution.

   (b) Assume that we are interested in estimating the probability of a count of zero $\theta = P\{X = 0\} = \exp\{-\lambda\}$. Find the MLE of θ and its asymptotic distribution. *Hint:* you may use the delta-method.

   (c) Is the MLE of θ you derived in (b) unbiased? Derive a theoretical formula for the first order bias. Describe a bootstrap bias-correction you may do here.

   (d) (Computer exercise). Simulate a sample of size $n = 20$ from a population with $\lambda = 1$. Implement your initial estimator $\widehat{\theta}$. Perform a bias correction you described with the number of bootstrapped samples $B = 20, 50, 100$. Call the bias-corrected estimator $\widetilde{\theta}$. Store the results. Repeat the whole procedure 1000 times. Average $\widehat{\theta}$ and $\widetilde{\theta}$ over repetitions. What can you say about biases of the two estimators?

   (e) Now consider a question of variance estimation. Find the asymptotic distribution of estimator $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})^2$. You may use the following facts about Poisson distribution: $EX = \lambda, Var(X) = \lambda, E(X - EX)^3 = \lambda, E(X - EX)^4 = \lambda(1 + 3\lambda)$.

(f) Notice that since $Var(X) = \lambda$, the estimator in (a) is the MLE for variance in this model. How would you compare asymptotic efficiency of estimators in (a) and (d)?

2. Assume one observes random variables $\{X_{i,t}, t = 1, 2, i = 1, ..., n\}$ which are independent of each other and come from the following model:

$$X_{i,t} = \mu_i + \varepsilon_{i,t}, \text{ where } \varepsilon_{i,t} \sim N(0, \sigma^2)$$

The unknown parameter here is $(\sigma^2, \mu_1, \mu_2, ..., \mu_n)$. This is the simplest panel data, and you can treat this situation as you observe each entity (i) for two periods $t = 1$ and $t = 2$, keeping in mind that each entity has its own unknown mean μ_i. This is also known as fixed effects model.

(a) Write down the likelihood function.

(b) Find MLE for the unknown parameters.

(c) Is the estimator for μ_i unbiased? Consistent as $n \to \infty$?

(d) Is the MLE for σ^2 unbiased? Consistent as $n \to \infty$?

(e) Why does asymptotic MLE theory fail to work in this case?

The described problem is known as the incidental parameter problem and is extremely important for panel data analysis.

3. Let $X_1, \ldots, X_n$ be a sample from the following distributions. In each case find the asymptotic variance of the MLE.

(a) $f(x \mid \theta) = \theta x^{\theta-1}, \quad 0 < x < 1, \quad 0 < \theta < \infty.$

(b) $f(x \mid \theta) = \frac{1}{\theta} \exp\left\{-\frac{x}{\theta}\right\}, \quad 0 < x < \infty, \quad 0 < \theta < \infty.$

4. *(Required problem)* Suppose that income Y is distributed as a Pareto distribution: $f(y) = \alpha y^{-(\alpha+1)}$ for $1 \leq y$, with $\alpha > 1$.

(a) It is quite common to not observe all incomes, but only those that are higher than some threshold (so-called truncated variables). Assume that you observe only those individuals with an income greater than or equal to \$9,000, and their income is described by a random variable Y^*. How is Y^* distributed?

(b) Your have a sample of size N drawn from the population of persons with incomes greater than or equal to \$9,000. What is the MLE of α?

(c) What is asymptotic distribution of the estimator in (b)?

(d) Suppose you believe that the mean of the Pareto distribution out of which you draw an observation is affected linearly by a variable w, that is, $E(Y_i|w_i) = \beta_1 + \beta_2 w_i$. Assume that you have a sample of (Y_i^*, w_i) of size N. Explain how you would estimate the parameters β_1 and β_2 (it is enough to set out the optimization problem that the parameters should solve, you do not need to solve for them directly). *Hint: calculate the mean of the Pareto distribution. How is it related to* α*?*