Statistics problems: team contest

May 2019

1. Consider the simple linear regression

$$Y_i = \alpha + \beta X_i + \epsilon_i, \qquad i = 1, \ldots, 2n.$$

(i) Suppose that only the first half of Y's are observed, i.e. only Y_i, $i = 1, \ldots, n$ are observed and Y_i, $i = n+1, \ldots, 2n$ are missing, while all X's are observed. What would you suggest for the estimation of β? Are there assumptions you need to make for your estimator to be valid?

(ii) Suppose for the second half of Y's, their absolute values are observed, i.e. we observe $|Y_i|$, $i = n+1, \ldots, 2n$. What would you suggest for the estimation of β? Are there assumptions you need to make for your estimator to be valid? Any good properties for your estimator?

2. Suppose that $X = (X_1, \ldots, X_n)'$ is an observation from the n-dimensional multivariate normal distribution $N_n(\theta, I)$ with unknown parameter $\theta \in R^n$, that is, X_i's are independent of each other with $X_i \sim N(\theta_i, 1)$ for $i = 1, \ldots, n$. (i). Derive the maximum likelihood estimator (MLE) of $\|\theta\|^2 = \sum_{i=1}^n \theta_i^2$.

(ii). Show that the MLE is a biased estimator.

(iii). Find the distribution of the MLE and describe how to use this distribution to construct exact confidence intervals.

3. Suppose that $X_1, \cdots, X_n$ is a sample of size n from the Student-t distribution $t_\nu(\mu, 1)$ with known degrees of freedom $\nu \geq 1$, unit scale, and known center μ. The Student-t distribution $t_\nu(\mu, 1)$ has density function of the form

$$f_X(x; \mu, \nu) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\pi\nu}\,\Gamma(\frac{\nu}{2})} \left(1 + \frac{(x-\mu)^2}{\nu}\right)^{-\frac{\nu+1}{2}}, \quad (x \in \mathcal{R}^1).$$

(i). Write $X_i = \mu + U_i$, where U_i are independently and identically distributed (iid) with $t_\nu(0, 1)$ for $i = 1, \cdots, n$. Find the conditional distribution of U_1 given $U_i - U_1 = X_i - X_1$ for $i = 2, \cdots, n$.

(ii). Describe a method to construct confidence intervals by making use of the above result, and argue for its efficiency and coverage probability.

(iii). Use the limiting case of $\nu \to \infty$ to verify your answers.