# STA 431s13 Assignment Four[1]

For the SAS question, please bring your log and list files to the quiz. Do not write anything on the printouts except your name and student number. The other questions are just practice for the quiz on Feb. 1st, and are not to be handed in.

1. Let $X_1, \ldots, X_n$ be a random sample from a continuous distribution with density
$$f(x;\theta) = \frac{1}{\theta^{1/2}\sqrt{2\pi}}\, e^{-\frac{x^2}{2\theta}},$$
where the parameter $\theta > 0$. Propose a reasonable estimator for the parameter $\theta$, and use the Law of Large Numbers to show that your estimator is consistent.

2. Let $X_1, \ldots, X_n$ be a random sample from a distribution with expected value $\mu$ and variance $\sigma_x^2$. Independently of $X_1, \ldots, X_n$, let $Y_1, \ldots, Y_n$ be a random sample from a distribution with the same expected value $\mu$ and variance $\sigma_y^2$. Let Let $T_n = \alpha\overline{X}_n + (1-\alpha)\overline{Y}_n$, where $0 \leq \alpha \leq 1$. Is $T_n$ always a consistent estimator of $\mu$? Answer Yes or No and show your work.

3. Let $X_1, \ldots, X_n$ be a random sample from a Gamma distribution with $\alpha = \beta = \theta > 0$. That is, the density is
$$f(x;\theta) = \frac{1}{\theta^\theta \Gamma(\theta)} e^{-x/\theta} x^{\theta-1},$$
for $x > 0$. Let $\widehat{\theta} = \overline{X}_n$. Is $\widehat{\theta}$ a consistent estimator of $\theta$? Answer Yes or No and prove your answer.

4. Let $X_1, \ldots, X_n$ be a random sample from a distribution with mean $\mu$ and variance $\sigma^2$. Prove that the sample variance $S^2 = \frac{\sum_{i=1}^n (X_i - \overline{X})^2}{n-1}$ is consistent for $\sigma^2$.

5. Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be a random sample from a bivariate distribution with $E(X_i) = \mu_x$, $E(Y_i) = \mu_y$, $Var(X_i) = \sigma_x^2$, $Var(Y_i) = \sigma_y^2$, and $Cov(X_i, Y_i) = \sigma_{xy}$. Show that the sample covariance $S_{xy} = \frac{\sum_{i=1}^n (X_i - \overline{X})(Y_i - \overline{Y})}{n-1}$ is a consistent estimator of $\sigma_{xy}$.

[1] Copyright information is at the end of the last page.

6. Independently for $i = 1, \ldots, n$, let

$$Y_i = \beta X_i + \epsilon_i,$$

where $E(X_i) = \mu$, $E(\epsilon_i) = 0$, $Var(X_i) = \sigma^2_X$, $Var(\epsilon_i) = \sigma^2_\epsilon$, and $\epsilon_i$ is independent of $X_i$. The variables $X_i$ and $Y_i$ are both observable.

   (a) Let

$$\widehat{\beta}_1 = \frac{\sum_{i=1}^n X_i Y_i}{\sum_{i=1}^n X_i^2}.$$

   - Is $\widehat{\beta}_1$ a consistent estimator of $\beta$? Answer Yes or No and prove your answer.
   - Does it matter if $\mu = 0$?

   (b) Let

$$\widehat{\beta}_2 = \frac{\sum_{i=1}^n Y_i}{\sum_{i=1}^n X_i}.$$

   - Is $\widehat{\beta}_2$ a consistent estimator of $\beta$? Answer Yes or No and justify your answer.
   - Does it matter if $\mu = 0$?

7. Men and women are calling a technical support line according to independent Poisson processes with rates $\lambda_1$ and $\lambda_2$ per hour. Data for 144 hours are available, but unfortunately the sex of the caller was not recorded. All we have is the number of callers for each hour, which is distributed Poisson$(\lambda_1 + \lambda_2)$.

   (a) The parameter in this problem is $\boldsymbol{\theta} = (\lambda_1, \lambda_2)'$. Try to find the MLE analytically. Show your work. Are there any points in the parameter space where both partial derivatives are zero?

   (b) Is the parameter identifiable? Answer Yes or No and prove your answer. If the answer is no, all you need is a simple numerical example of two different parameter vectors that yield the same probability distribution for the sample data.

8. Let $X_1, \ldots, X_n$ be a random sample from a distribution with density

$$f(x; \theta) = \frac{1}{\theta} e^{-x/\theta},$$

for $x > 0$. Is the parameter identifiable? Answer Yes or No and prove your answer. If the answer is no, all you need is a simple numerical example of two different parameter vectors that yield the same probability distribution for the sample data.

9. Here is a multivariate regression model with no intercept and no measurement error. Independently for $i = 1, \ldots, n$,

$$\mathbf{Y}_i = \boldsymbol{\beta}\mathbf{X}_i + \boldsymbol{\epsilon}_i$$

where

$\mathbf{Y}_i$ is an $q \times 1$ random vector of observable response variables, so the regression can be multivariate; there are $q$ response variables.

$\mathbf{X}_i$ is a $p \times 1$ multivariate normal observable random vector; there are $p$ explanatory variables. $\mathbf{X}_i$ has expected value zero and variance-covariance matrix $\boldsymbol{\Phi}$, a $p \times p$ symmetric and positive definite matrix of unknown constants.

$\boldsymbol{\beta}$ is a $q \times p$ matrix of unknown constants. These are the regression coefficients, with one row for each response variable and one column for each explanatory variable.

$\boldsymbol{\epsilon}_i$ is the error term of the latent regression. It is an $q \times 1$ multivariate normal random vector with expected value zero and variance-covariance matrix $\boldsymbol{\Psi}$, a $q \times q$ symmetric and positive definite matrix of unknown constants. $\boldsymbol{\epsilon}_i$ is independent of $\mathbf{X}_i$.

(a) Calculate the variance-covariance matrix of the observable variables. It's a partitioned matrix. Show your work.

(b) Write down the moment structure equations. These are matrix equations.

(c) Does this problem pass the test of the Parameter Count Rule? Answer Yes or No and give the numbers.

(d) Are the parameters of this model identifiable? Answer Yes or No and prove your answer. If the answer is no, all you need is a simple numerical example of two different parameter vectors that yield the same probability distribution for the sample data.

10. Independently for $i = 1, \ldots, n$, let

$$
\begin{aligned}
Y_i &= \beta X_i + \epsilon_i \\
W_{i,1} &= X_i + e_{i,1} \\
W_{i,2} &= X_i + e_{i,2},
\end{aligned}
$$

where

- $X_i$ is a normally distributed *latent* variable with mean zero and variance $\phi > 0$
- $\epsilon_i$ is normally distributed with mean zero and variance $\psi > 0$
- $e_{i,1}$ is normally distributed with mean zero and variance $\omega_1 > 0$
- $e_{i,2}$ is normally distributed with mean zero and variance $\omega_2 > 0$
- $X_i$, $\epsilon_i$, $e_{i,1}$ and $e_{i,2}$ are all independent of one another.

(a) What is the parameter vector $\boldsymbol{\theta}$ for this model?

(b) Does this problem pass the test of the Parameter Count Rule? Answer Yes or No and give the numbers.

(c) Calculate the variance-covariance matrix of the observable variables. Show your work.

(d) Is the parameter vector identifiable? Answer Yes or No and prove your answer.

11. Recall the SAT data from Assignment 3. The data are given in the file `opensat.data`. There is a link on the course web page in case the one in this document does not work.

    Using SAS `proc calis`, carry out an unconditional regression in which the explanatory variables are Verbal SAT and Math SAT, and the response variable is GPA. You are just fitting one model, and it is saturated. Bring your log file and your list file to the quiz. You may be asked for numbers from your printouts, and you may be asked to hand them in. **There must be no warnings, error messages or notes about missing data on your log file.**

---