STA 313F2004 Assignment 1

Do this assignment in preparation for the quiz on Friday, Sept. 24th. The hand-written parts are practice for the quiz, and are not to be handed in. The computer parts may be handed in, so please begin the computer part of each question on a separate piece of paper.

1. Let the random vector $\mathbf{X} = (X_1, \ldots, X_n)'$ have density $f(\mathbf{x})$. For a general function g, use the rule

$$E[g(\mathbf{X})] = \int \cdots \int g(\mathbf{x}) f(\mathbf{x}) d\mathbf{x}.$$

as if it were a definition. Prove

   (a) $E[a] = a$, where a is a constant.

   (b) $E[aX] = aE[X]$, where a is a constant.

   (c) $E[\sum_{i=1}^n X_i] = \sum_{i=1}^n E[X_i]$

   (d) If X and Y are independent, $E[XY] = E[X]E[Y]$. Draw an arrow to the place in your answer where you use independence, and write "This is where I use independence."

In all the remaining questions about variances and covariances, you should just use the linear properties of expectation that you have proved in Question 1, and avoid using integrals.

2. Denoting $E[X]$ by μ_X, define the variance $V(V) = E[(X - \mu_X)^2]$. Show that $V(X) = E[X^2] - (E[X])^2$.

3. Define the covariance of X and Y by $Cov[X, Y] = E[(X - \mu_X)(Y - \mu_Y)]$. Show that $Cov[X, Y] = E[XY] - E[X]E[Y]$.

4. In the following, X and $X_1, \ldots, X_n$ are random variables, while a, b and $a_1, \ldots, a_n$ are fixed constants. For each pair of statements below, one is true and one is false (that is, not true in general). State which one is true, and prove it. Zero marks if you prove both statements are true, even if one of the proofs is correct.

   (a) $V(aX) = aV(X)$ or $V(aX) = a^2V(X)$

   (b) $V(aX + b) = a^2V(X) + b^2$ or $V(aX + b) = a^2V(X)$

   (c) $V(a) = 0$ or $V(a) = a^2$

   (d) $V(\sum_{i=1}^n a_iX_i) = \sum_{i=1}^n a_i^2V(X_i)$ or $V(\sum_{i=1}^n a_iX_i) = \sum_{i=1}^n a_i^2V(X_i) + \sum_{i=1}^n \sum_{j \neq i} a_ia_jCov(X_i, X_j)$

5. Let $X_1, \ldots, X_n$ be independent and identically distributed random variables (the standard model of a random sample with replacement). Denoting $E(X_i)$ by μ and $V(X_i)$ by σ^2,

   (a) Show $E[\overline{X}] = \mu$; that is, the sample mean is unbiased for μ.

   (b) Find $V(\overline{X})$.

   (c) Show $E[S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \overline{X})^2 = \sigma^2$; that is, the sample variance is unbiased for σ^2.

6. For each of the following distributions, try to derive a general expression for the Maximum Likelihood Estimator (MLE). Then, use R's `nlm` function to obtain the MLE numerically for the data supplied for the problem. Finally, if you were able to derive an explicit formula for the MLE, compute it with R to check your answer. The data are in a separate HTML document, because it saves a lot of effort to copy and paste rather than typing the data in by hand, and PDF documents can contain invisible characters that mess things up.

   (a) $p(x) = \theta(1-\theta)^{x-1}$ for $x = 1, 2, \ldots$, where $0 < \theta < 1$.

   (b) $f(x) = \frac{\alpha}{x^{\alpha+1}}$ for $x > 1$, where $\alpha > 0$.

   (c) $f(x) = \frac{1}{\pi[1+(x-\theta)^2]}$ for x real, where $-\infty < \theta < \infty$.

   (d) $f(x) = \frac{1}{2}e^{-|x-\theta|}$ for x real, where $-\infty < \theta < \infty$.

   (e) $f(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}x^{\alpha-1}(1-x)^{\beta-1}$ for $0 < x < 1$, where $\alpha > 0$ and $\beta > 0$.

For the computer part of this problem set, bring to the quiz one sheet of printed output for each of the 5 distributions. The five sheets should be separate, because you may hand only one of them in. Each printed page should show the following, *in this order.*

1. Definition of the function that computes the likelihood, or log likelihood, or minus log likelihood or whatever.
2. How you got the data into R – probably a `scan` statement.
3. Listing of the data for the problem.
4. The `nlm` statement and resulting output.
5. Calculation of the MLE using your explicit formula, if this is possible.

Don't hand-write notes or formulas on your printouts. Also, you are expected to do the computer part of this assignment mostly on your own. You may compare answers and general strategies, but **do not, under any circumstances, show anyone your code or look at anyone else's code** until after the quiz.