STAT 515

Homework #10, due Friday, Apr. 13 at 2:30pm

This homework must be submitted electronically to ANGEL. I strongly encourage the use of LaTeX.

Please make every assignment easier to grade by neatly organizing your writeup and clearly labeling your final answers when appropriate. Try using LaTeX!

1. We wish to approximate $\mu = P(X > 4.5)$ where $X \sim N(0, 1)$. Suppose that $q(x)$ is a normal density with mean k and variance 1, and suppose that $X_1, \ldots, X_n$ is a simple random sample from $q(\cdot)$.

   (a) Show that
   $$\tilde{\mu} = \frac{\frac{1}{n}\sum_{i=1}^{n} I\{X_i > 4.5\} \exp\{(X_i - k)^2/2 - X_i^2/2\}}{\frac{1}{n}\sum_{i=1}^{n} \exp\{(X_i - k)^2/2 - X_i^2/2\}}$$
   is a consistent estimator of μ. (To do this, it's enough to show that the true mean of the numerator divided by the true mean of the denominator equals μ.)

   (b) Based on samples of size 100,000 from $q(\cdot)$, try using $\tilde{\mu}$ several times for $k = 0$, $k = 4.5$, and some intermediate values of k. What value of k seems to give the most precise estimates?

   (c) Use the delta-method derivation
   $$\text{Var}\left(\frac{\frac{1}{n}\sum_i A_i}{\frac{1}{n}\sum_i B_i}\right) \approx \frac{1}{n\mu_B^2}\begin{bmatrix} 1 & \frac{-\mu_A}{\mu_B} \end{bmatrix}\begin{bmatrix} \sigma_A^2 & \sigma_{AB} \\ \sigma_{AB} & \sigma_B^2 \end{bmatrix}\begin{bmatrix} 1 \\ \frac{-\mu_A}{\mu_B} \end{bmatrix}$$
   to estimate the variances of your $\tilde{\mu}$ estimators from part (b). (Use sample estimates of μ_A, μ_B, and the covariance matrix.) Do the variance estimates correspond with your experience in part (b)?

   (d) Consider a modified estimator
   $$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} I\{X_i > 4.5\} \exp\{(X_i - k)^2/2 - X_i^2/2\},$$
   where once again $X_1, \ldots, X_n$ is a simple random sample from $q(\cdot)$. Verify that this estimator is a consistent estimator of μ. (Again, merely show that the true mean of each summand equals μ.) Using the best k you found earlier, compare the estimated variance of $\tilde{\mu}$ with the estimated variance of $\hat{\mu}$ (the latter should not be hard to find). Which estimator, $\tilde{\mu}$ or $\hat{\mu}$, appears to be more precise?

2. Suppose that X is a binomial random variable with parameters n and p, where $p = e^{\theta}/(1 + e^{\theta})$ for some real-valued parameter θ. The goal of this question will be to use ratio importance sampling to estimate the log-likelihood function $\ell(\theta) = \log P_{\theta}(X)$.

   (a) Show that the log-likelihood function may be written as
   $$\ell(\theta) = \theta X - \log c(\theta) + (\text{something not depending on } \theta),$$
   and find the normalizing function $c(\theta)$.

   (b) Fix some θ_0. Show that
   $$\ell(\theta) = \ell(\theta_0) + (\theta - \theta_0)X - \log E_{\theta_0}[\exp\{(\theta - \theta_0)Y\}],$$
   where the notation above means that Y has a binomial distribution according to θ_0 (and X is the data, as usual).

   (c) The equation of part (b) suggests a method for approximating $\ell(\theta) - \ell(\theta_0)$, which is a function that can be maximized to find the MLE of θ. Suppose that $n = 100$ and $X = 80$, then take a random sample $Y_1, \ldots, Y_m$ using $m = 10^6$ and $\theta_0 = 1$ to approximate the function $\ell(\theta) - \ell(\theta_0)$. On the same set of axes, plot both the true $\ell(\theta) - \ell(\theta_0)$ and your approximation. How does your approximate MLE compare with the true MLE?

   (d) Try the same technique as in part (d) but use $\theta_0 = 0$. What do you observe?