Biost/Stat 578A
Emerson, Spr 13
April 10, 2013

**Homework #1**: Written problems to be submitted electronically by 11:59 pm on Tuesday, April 16.

1. Suppose $Y_1, \ldots, Y_n$ are independent random variables distributed according to an exponential distribution $Y_i \sim \mathcal{E}(\theta)$, with $E[Y_i] = \theta$.
   a. What is the asymptotically efficient estimator $\hat{\theta}$ of $\theta$?
   b. What is the exact sampling distribution of $\hat{\theta}$?
   c. What is the approximate sampling distribution of $\hat{\theta}$ based on regular asymptotic theory?
   d. Suppose a parametric Bayesian chooses a conjugate prior with mean $\zeta$ and variance $\tau^2$. What is that conjugate prior, and what is the resulting posterior distribution of $\theta$ conditional on $\vec{Y}$? Why is it valid to consider this posterior distribution as conditional on $\hat{\theta}$? Write a function in R that will compute the cumulative distribution function for that posterior distribution.
   e. Consider instead a distribution-free Bayesian approach in which the posterior distribution of $\theta$ is computed conditional on $\hat{\theta}$, using the approximate distribution of $\hat{\theta}$ from part c (with the known standard error) as the sampling distribution and as prior distribution for $\theta$ a normal distribution with mean $\zeta$ and variance $\tau^2$. Write a function in R that will compute the cumulative distribution function for that posterior distribution.
   f. Consider a modified distribution-free Bayesian approach that again uses the approximate distribution from part c, but now uses an estimated standard error that is held constant as $\theta$ varies. Again use a normal prior distribution. What is the resulting posterior distribution of $\theta$ conditional on $\hat{\theta}$? Write a function in R that will compute the cumulative distribution function for that posterior distribution.
   g. Write an R function `simBayesExponential (nSim, n, zeta, tauSqr, theta0)` that simulates `nSim` datasets and computes the posterior probability $Pr(\theta < \theta_0|\hat{\theta})$ under each of the methods in parts d, e, and f. The simulated datasets should be created by
      - For each dataset, randomly sample a value of $\theta$ from the conjugate prior distribution having mean $\zeta$ and variance $\tau^2$ as specified in function arguments `zeta` and `tauSqr`.
      - Generate data for `n` observations of $Y_i$ using the corresponding value of $\theta$.
      - Compute $\hat{\theta}$ and an estimated standard error of $\hat{\theta}$ using distribution-free methods (i.e., using the sample variance).
      - Compute the posterior probability $Pr(\theta < \theta_0|\hat{\theta})$ for the value of $\theta_0$ specified by `theta0` using the formulas derived in parts d, e, and f. When using the methods for part f, use the standard error estimate in the approximate distribution for $\hat{\theta}$.
   h. Using the R function, produce plots comparing the agreement between the various methods of computing posterior probabilities. On the x-axis, plot the "true" posterior probability computed under the parametric model, and on the y-axis plot the difference between each of the distribution-free calculations and the true value. Produce plots for 1000 simulations using $n = 10$, 25, and 100. You might consider $\theta_0 = \zeta$ for various choices of $\zeta$ and $\tau^2$.
2. Suppose $Y_1, \ldots, Y_n$ are independent random variables distributed according to an Bernoulli distribution $Y_i \sim \mathcal{B}(1, \theta)$, with $E[Y_i] = \theta$. Repeat all parts of problem 1 to produce `simBayesBernoulli (nSim, n, zeta, tauSqr, theta0)` .
3. Suppose $Y_1, \ldots, Y_n$ are independent random variables distributed according to an Poisson distribution $Y_i \sim \mathcal{P}(\theta)$, with $E[Y_i] = \theta$. Repeat all parts of problem 1 to produce `simBayesPoisson (nSim, n, zeta, tauSqr, theta0)` .