STATISTICS 225 – Fall 2018

Homework 3 - solutions

1. **Project proposals.** Feedback provided separately.

2. **Simulation inference** - Apologies for confusion caused by an error of mine. I intended to say that $V_\theta = 15^2$ so that the standard deviation (σ below) was 15.

   (a) The width of the 95% confidence interval is $3.92\sigma/\sqrt{n}$. Taking $\sigma = 15$ and setting equal to 10 yields $n = 34.57$. Thus require 35 or more observations. Setting equal to 5 yields $n = 138.30$ so that 139 or more observations are required.

   (b) We have for the mean that $Var(\bar{y}) = \sigma^2/n$. The normal median is $F^{-1}(0.5) = \mu$ and $f(F^{-1}(0.5)) = \frac{1}{\sqrt{2\pi}\sigma}$. Then according to the result the median is asymptotical normal with mean μ and variance $\pi\sigma^2/(2n)$. The median has larger variance (by a factor of $\pi/2 = 1.57$. The mean is the most efficient estimator in the normal case, it uses the information in all of the samples. The median does not use information in the tails as well. (Of course if the data are not normal, then there are no guarantees that the mean is a useful measure; the median may provide a better estimate of the distribution's location). For the median the width of the 95% confidence interval is $3.92\sqrt{\pi/2}\sigma/\sqrt{n}$. Setting width to 5 yields $n \geq 218$.

   (c) For the 2.5%ile we first note that $F^{-1}(.025) = \mu - 1.96\sigma$ and then note that $f(F^{-1}(0.025)) = \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2}(-1.96)^2}$. Plugging in with $\sigma = 15$ and width 5 yields $n \geq 988$. By symmetry the answer for the 2.5%ile and the 97.5%ile are the same.

   (d) Our posterior inference for θ might be reported by a posterior interval centered at the estimated posterior mean (estimated by simulations) with width determined by the posterior variance V (standard deviation ($\sqrt{V}$). The argument in the book is that we'd like to have the posterior inference centered at the true posterior mean but only have our simulation-based estimate. This estimate has variance V/S where S is the number of simulations. Thus the total uncertainty taking into account the simulation uncertainty is $V(1 + 1/S)$.

3. **t-inference using a Metropolis algorithm** - My code is provided on the website. I found that scaling with c=.25 (so s.d.=.5) seemed to work well and accept between 35 and 40 percent of the draws. Posterior inference for $\mu = (2.5\% = 1.87, 25\% = 2.34, 50\% = 2.63, 75\% = 2.92, 97.5\% = 3.65)$ and for $\sigma = (2.5\% = 1.07, 25\% = 1.39, 50\% = 1.62, 75\% = 1.89, 97.5\% = 2.49)$. I must confess that I don't recall the values of μ and σ that were used to generate the data. (Sorry.) In one small sample there is no guarantee that posterior distribution will be centered at the true values ... but we'd hope the true values would be in the 95% posterior interval.

4. **t-inference using a Gibbs sampler** - My code is provided on the website. I had to run the Gibbs sampler for 10000 steps to be satisfied. Even then some of the V_i's have not converged. Some of the techniques in Chapter 12 could help with this. Results are very close to the above. Posterior inference for $\mu = (2.5\% = 1.75, 25\% = 2.34, 50\% = 2.63, 75\% = 2.93, 97.5\% = 3.53)$ and for $\sigma = (2.5\% = 1.03, 25\% = 1.38, 50\% = 1.60, 75\% = 1.86, 97.5\% = 2.54)$.