**Stat 591 – Homework 03** ***Solutions***

1. (a) With a flat prior on $\theta$, it is easy to check that the posterior distribution for $\theta$ is $\mathsf{N}(\bar{X}, n^{-1})$. That is, $\mathsf{E}(\theta \mid X_1, \ldots, X_n) = \bar{X}$ and $\mathsf{V}(\theta \mid X_1, \ldots, X_n) = n^{-1}$. To prove that the posterior is consistent at $\theta^\star$ under the iid $\mathsf{N}(\theta^\star, 1)$ model, follow the approach discussed in class. That is, for any $\varepsilon > 0$, we have

$$\begin{aligned}
\Pi(|\theta - \theta^\star| > \varepsilon \mid X_1, \ldots, X_n) & \\
&\leq \varepsilon^{-2}\mathsf{E}\{(\theta - \theta^\star)^2 \mid X_1, \ldots, X_n\} \\
&= \varepsilon^{-2}\big[\mathsf{V}(\theta \mid X_1, \ldots, X_n) + \{\mathsf{E}(\theta \mid X_1, \ldots, X_n) - \theta^\star\}^2\big] \\
&= \varepsilon^{-2}\big[n^{-1} + (\bar{X} - \theta^\star)^2\big].
\end{aligned}$$

Clearly, $n^{-1} \to 0$ and, by the law of large numbers, $(\bar{X} - \theta^\star)^2 \to 0$ in probability under the iid $\mathsf{N}(\theta^\star, 1)$ model. Therefore, the posterior probability above converges to zero in probability, so the posterior is consistent at $\theta^\star$.

(b) R codes for my simulation are given in Figure 1, and the plot summarizing the results is shown in Figure 2. Observe that, for large enough $n$, all three quantiles of $Q_n$ are close to 0, which means that the distribution of $Q_n$ is concentrated at 0, hence $Q_n \to 0$ in probability.

2. Fix $\theta_i$ as the "true" parameter value, i.e., the parameter value that corresponds to the data $X_1, \ldots, X_n$ in the frequentist iid model. The goal is to show that $\Pi(\theta = \theta_r \mid X_1, \ldots, X_n)$ converges in $\mathsf{P}_{\theta_i}$-probability to zero for each $r \neq i$. Since the parameter space is discrete, it suffices to prove that

$$q_n(r) := \Pi(\theta = \theta_r \mid X_1, \ldots, X_n)/\Pi(\theta = \theta_i \mid X_1, \ldots, X_n)$$

converges to zero in $\mathsf{P}_{\theta_i}$-probability. Since posterior is proportional to prior times likelihood, we can see that

$$q_n(r) = \frac{\pi_r \prod_{j=1}^n f_{\theta_r}(X_j)}{\pi_i \prod_{j=1}^n f_{\theta_i}(X_j)} = e^{\log(\pi_r/\pi_i) + nZ_{n,r}},$$

where

$$Z_{n,r} = \frac{1}{n}\sum_{j=1}^n \log\frac{f_{\theta_r}(X_j)}{f_{\theta_i}(X_j)}.$$

Under the iid $f_{\theta_i}$ model, the term $Z_{n,r}$ is an average of iid random variables, with expectation

$$\mathsf{E}_{\theta_i}\Big[\log\frac{f_{\theta_r}(X)}{f_{\theta_i}(X)}\Big] = \int \log\frac{f_{\theta_r}(x)}{f_{\theta_i}(x)} f_{\theta_i}(x)\,dx.$$

This integral is the minus Kullback–Leibler divergence of $f_{\theta_r}$ from $f_{\theta_i}$. Since the Kullback–Leibler divergence is positive for $r \neq i$, the expectation above must be negative. Then, by the law of large numbers, $Z_{n,r}$ will be negative for large enough $n$. But $q_n(r)$ has a multiplier "$n$" on $Z_{n,r}$, so if $Z_{n,r}$ is eventually negative, $nZ_{n,r} \to -\infty$ in probability. Then, by continuous mapping theorem, $q_n(r) \to 0$ in probability, which is what we wanted to show. That is, $\Pi(\theta = \theta_r \mid X_1, \ldots, X_n) \to 0$ in probability under the iid $\mathsf{P}_{\theta_i}$ model.

```
N <- seq(1000, 150000, len=50)
eps <- 0.01
qtile <- matrix(0, nrow=length(N), ncol=3)
for(i in 1:length(N)) {

  s <- 1 / sqrt(N[i])
  xbar <- rnorm(3000, 0, s)
  pr <- function(x) 1 - (pnorm(eps, x, s) - pnorm(-eps, x, s))
  Q <- sapply(xbar, pr)
  qtile[i,] <- as.numeric(quantile(Q, c(0.5, 0.75, 0.95)))

}
plot(N, qtile[,3], ylim=c(0,1), type="l", xlab="n", ylab="50-75-95 Quantiles")
lines(N, qtile[,2])
lines(N, qtile[,1])
```

Figure 1: R code for the posterior consistency simulation in Problem 1b.

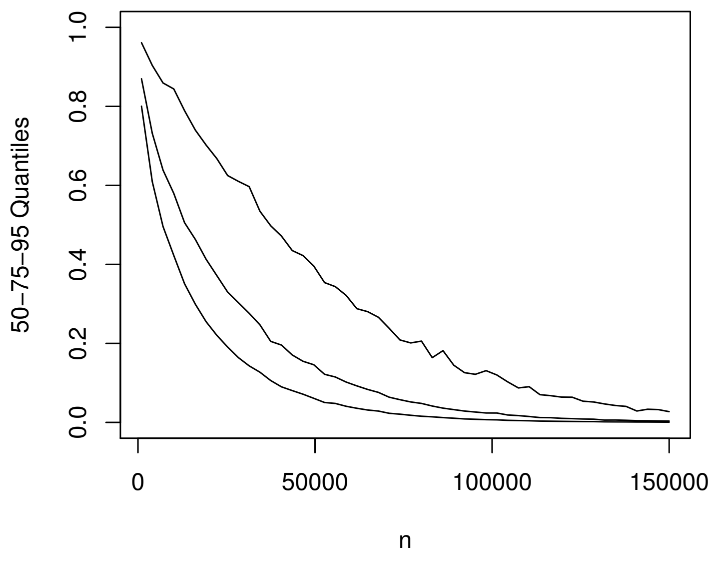


Figure 2: Plot of the 50th, 75th, and 95th quantiles of the distribution of $Q_n$ as a function of $n$; here $\varepsilon = 0.01$, $\theta^\star = 0$, and 3000 Monte Carlo samples are used for each $n$.

3. Following the hint provided, and using the fact that $Q_n \leq 1$, the posterior probability $Q_n$ can be bounded by $Q_n \cdot I_{A_n} + I_{A_n^c}$. By definition of $A_n$, it follows from consistency of the MLE $\hat{\theta}_n$ that $I_{A_n^c} \to 0$ in probability under the iid $\mathsf{P}_{\theta^\star}$ model. So, it remains to show that $Q_n \cdot I_{A_n^c}$ converges to zero. For a data sequence in $A_n$, we know that $\hat{\theta}_n$ and $\theta^\star$ are within $\varepsilon/2$ units from one another. So, for such a data sequence, if $\theta$ is more than $\varepsilon$ units from $\theta^\star$, then $\theta$ must be more than $\varepsilon/2$ units from $\hat{\theta}_n$. Technically, this requires a triangle inequality type argument, but it's easy to check by drawing a picture. Therefore,

$$\Pi(|\theta - \theta^\star| > \varepsilon \mid X_1, \ldots, X_n) I_{A_n} \leq \Pi(|\theta - \hat{\theta}_n| > \varepsilon/2 \mid X_1, \ldots, X_n).$$

Since we assume the posterior normality result holds, the right-hand side above can be approximated by a normal probability, i.e.,

$$\Pi(|\theta - \hat{\theta}_n| > \varepsilon/2 \mid X_1, \ldots, X_n) \approx \int_{x:|x|>\sqrt{nI(\theta^\star)}\varepsilon/2} \mathsf{N}(x \mid 0, 1)\, dx.$$

Since the range of integration is approaching $\varnothing$ as $n \to \infty$ (and since the normal density is integrable), the integral itself must be converging to zero (in probability under the $\mathsf{P}_{\theta^\star}$ model). Since both terms in the decomposition of $Q_n$ are vanishing, we conclude that the posterior normality implies posterior consistency.

There is an alternative argument I didn't think of till after I posted the hint. Let $\pi_n(\theta)$ and $\tilde{\pi}_n(\theta)$ denote the actual posterior and the $\mathsf{N}(\hat{\theta}_n, [nI(\theta^\star)]^{-1})$ approximation, respectively. Then, for the same $Q_n$ as above, we can write

$$\begin{aligned} Q_n &= \int_{\theta:|\theta-\theta^\star|>\varepsilon} \pi_n(\theta)\, d\theta \\ &= \int_{\theta:|\theta-\theta^\star|>\varepsilon} [\pi_n(\theta) - \tilde{\pi}_n(\theta) + \tilde{\pi}_n(\theta)]\, d\theta \\ &\leq \int |\pi_n(\theta) - \tilde{\pi}_n(\theta)|\, d\theta + \int_{\theta:|\theta-\theta^\star|>\varepsilon} \tilde{\pi}_n(\theta)\, d\theta. \end{aligned}$$

The first term in the upper bound above goes to zero in probability by Theorem 4.2 in [GDS]; to see this, just standardize $\theta$ as $Z = [nI(\theta^\star)]^{1/2}(\theta - \hat{\theta}_n)$. For the second term, do the same standardizing to get

$$\int_{z:|\hat{\theta}_n-\theta^\star+z[nI(\theta^\star)]^{-1/2}|>\varepsilon} \mathsf{N}(z \mid 0, 1)\, dz.$$

Since $\hat{\theta}_n \to \theta^\star$ in probability, it is clear that the range of the above integration collapses to $\varnothing$. It follows that the integral also converges to zero in probability and, therefore, so does $Q_n$, as was to be shown.

4. From the definition of "$L_n(\theta)$" in [GDS], we have

$$\frac{L_n(\theta) - L_n(\theta^\star)}{n} = \frac{1}{n}\sum_{i=1}^n \log \frac{f_\theta(X_i)}{f_{\theta^\star}(X_i)} = (\bar{X} - \theta^\star)(\theta - \theta^\star) - \frac{(\theta - \theta^\star)^2}{2}.$$

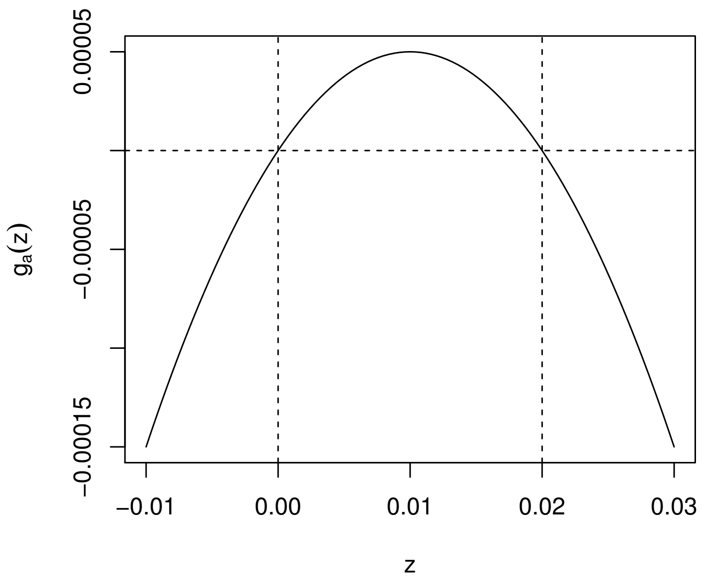


Figure 3: Plot of the $g_a(z)$ function for $a = 0.01$ in Problem 4.

For a constant $a \in \mathbb{R}$, define a function $g_a(z) = az - z^2/2$. Using the quadratic formula, the only zeros of $g_a(z)$ are $z = 0$ and $z = 2a$; in particular, by looking at the derivative of $g_a$, you can see that if $|z| > 2|a|$, then $g_a(z) < 0$; a plot of this function for $a = 0.01$ is given in Figure 3. Given an arbitrary $\delta > 0$, the law of large numbers says that, with probability 1, there exists $N$ such that, if $n \geq N$, then $|\bar{X} - \theta^\star| < \delta/2$. Therefore, if $n \geq N$ and $|\theta - \theta^\star| > \delta$, then

$$\sup_{\theta : |\theta - \theta^\star| > \delta} \Big[\frac{L_n(\theta) - L_n(\theta^\star)}{n}\Big] = g_{\delta/2}(\delta) < 0.$$

This verifies Condition (A4) in [GDS] for the normal case.

5. The posterior normality result says that $\theta$, given $X$, is approximately (4-dimensional) normal with mean $\hat{\theta}_n = X/n$ and covariance matrix proportional to the inverse Fisher information. All these quantities were computed in Homework #1. Since $\kappa$ is a smooth function of $\theta$, the posterior distribution of $\kappa$, given $X$, is also normal by the Delta Theorem. For the given data, the normal approximation for the posterior distribution of $\kappa$ is exactly the same as the approximate sampling distribution of the MLE $\hat{\kappa}$. So, since $n = 100$ is relatively large, it is not surprising that the bootstrap, the MLE normal approximation, and the posterior normality approximation are all essentially the same.