STA531 Midterm Exam 1 Solutions

1. Posterior consistency and asymptotic normality

    (a) The log likelihood is

    $$\ell(\theta; x_{1:n}) = \log p(x_{1:n}|\theta) = \log \theta^n e^{-\theta \sum x_i} = n \log \theta - \theta \sum_{i=1}^{n} x_i.$$

    The first and second derivatives are

    $$\frac{\partial}{\partial \theta}\ell = \frac{n}{\theta} - \sum_{i=1}^{n} x_i$$
    $$\frac{\partial^2}{\partial \theta^2}\ell = -\frac{n}{\theta^2} < 0.$$

    Since the second derivative is negative for all $\theta > 0$, the function is strictly concave. So, setting the derivative equal to zero and solving tells us that the maximum likelihood estimator is $\hat{\theta} = 1/\bar{x}$. The observed information is

    $$I(\theta; x_{1:n}) = -\frac{\partial^2}{\partial \theta^2}\ell = \frac{n}{\theta^2}.$$

    Therefore, the asymptotic normal approximation to the posterior is

    $$\theta|x_{1:n} \approx \mathcal{N}(\hat{\theta}, I(\hat{\theta}; x_{1:n})^{-1}) = \mathcal{N}(1/\bar{x}, 1/(n\bar{x}^2)).$$

    (b) The (strong) law of large numbers tells us that with probability one (i.e., almost surely), $\bar{x} \to \mathbb{E}_{\theta_0} X = 1/\theta_0$ as $n \to \infty$. Therefore, $\hat{\theta} = 1/\bar{x} \to \theta_0$ and $1/(\sqrt{n}\bar{x}) \to 0$ (a.s.) as $n \to \infty$. In other words, the mean of the asymptotic normal approximation is converging to θ_0, and the standard deviation of the asymptotic normal approximation is converging to zero. Hence, for any neighborhood of θ_0, as n grows, the amount of probability given to that neighborhood (by the normal approximation) will eventually go to one.

    To make this argument formal (which was certainly not required), one would have to argue that (i) convergence of the standard deviation to zero implies the claimed result about the probability in a neighborhood going to one (this can be done using Chebyshev's inequality), and (ii) the normal approximation is good enough (this can be done by stating the approximation more precisely in terms of convergence with respect to the total variation distance).

2. Posterior predictive checks

    (a) $p(\theta|x) \propto p(x|\theta)p(\theta) \propto \mathcal{N}(x|\theta, 1) = \mathcal{N}(\theta|x, 1)$.

    (b) The posterior predictive is $p(x^{\text{rep}}|x) = \int p(x^{\text{rep}}|\theta)p(\theta|x)d\theta$. Therefore, we can sample $X^{\text{rep}}|x$ by first drawing $\theta|x \sim \mathcal{N}(x, 1)$ and then drawing $X^{\text{rep}}|\theta \sim \mathcal{N}(\theta, 1)$. In other words, we can sample $X^{\text{rep}}|x$ by drawing $Y, Z \sim \mathcal{N}(0, 1)$ independently, and setting $X^{\text{rep}} = x + Y + Z$. Therefore, $X^{\text{rep}}|x \sim \mathcal{N}(x, 2)$ by the formula for linear combinations of independent normal random variables.

(c) The posterior predictive p-value is

$$\mathbb{P}(T(X^{\text{rep}}) \geq T(x) \mid x) = \mathbb{P}(X^{\text{rep}} \geq x \mid x) = 1/2$$

by part (b) since $X^{\text{rep}}|x$ is normal with mean x.

3. Modeling the data collection process

(a) $p(\theta|x, y_{\text{obs}}, I) = p(\theta|x, y_{\text{obs}})$.

(b) Here is one of many possible examples. Suppose $Y_1, \ldots, Y_n|\theta$ i.i.d. $\sim$ Bernoulli(θ), and given y, θ, suppose $I_j = Y_j$ for all $j = 1, \ldots, n$. Suppose there are no covariates x. Then the observed y's are all ones, however, all of the y's can be recovered perfectly from I. Consequently, $p(\theta|y_{\text{obs}})$ will concentrate at 1, but $p(\theta|y_{\text{obs}}, I)$ will concentrate at the true value θ_0.

4. Credible intervals and frequentist coverage

(a) First, note that since the function $f(x) = x^a$ is monotone increasing for any $a > 0$, we have that (i) $\theta_0 \leq x^{0.1}$ if and only if $\theta_0^{1/0.1} \leq x$, and (ii) $x^{0.9} \leq \theta_0$ if and only if $x \leq \theta_0^{1/0.9}$. Therefore, the frequentist coverage probability is

$$\begin{aligned}
\mathbb{P}(\theta_0 \in C(X) \mid \theta_0) &= \mathbb{P}(X^{0.9} \leq \theta_0 \leq X^{0.1} \mid \theta_0) \\
&= \mathbb{P}(\theta_0^{1/0.1} \leq X \leq \theta_0^{1/0.9} \mid \theta_0) \\
&= \int_{\theta_0^{1/0.1}}^{\theta_0^{1/0.9}} p(x|\theta_0)dx = \frac{1}{\theta_0}(\theta_0^{1/0.9} - \theta_0^{1/0.1}),
\end{aligned}$$

since $\theta_0^{1/0.9} < \theta_0$.

(b) We need to show that

$$\mathbb{P}(\theta < x^{0.9} \mid x) = \mathbb{P}(\theta > x^{0.1} \mid x) = 0.1.$$

First, the posterior is

$$p(\theta|x) \propto p(x|\theta)p(\theta) = \frac{1}{\theta}\mathbb{1}(0 < x < \theta)\mathbb{1}(0 < \theta < 1) \propto \frac{1}{\theta}\mathbb{1}(x < \theta < 1).$$

Since $\int_x^t (1/\theta)d\theta = \log\theta|_x^t = \log t - \log x$, then

$$p(\theta|x) = \frac{1/\theta}{-\log x}\mathbb{1}(x < \theta < 1),$$

and

$$\mathbb{P}(\theta < t \mid x) = \int_x^t p(\theta|x)d\theta = \frac{\log t - \log x}{-\log x} = 1 - \frac{\log t}{\log x}.$$

Therefore,

$$\mathbb{P}(\theta < x^{0.9} \mid x) = 1 - \frac{\log x^{0.9}}{\log x} = 0.1$$

and

$$\mathbb{P}(\theta > x^{0.1} \mid x) = \frac{\log x^{0.1}}{\log x} = 0.1.$$