

Statistics 5444: Homework 2

For each homework assignment, turn in at the beginning of class on the indicated due date. Late assignments will only be accepted with special permission. Write each problem up *very* neatly (L^AT_EX is preferred). Show all of your work.

Problem 1

Consider the binomial sampling distribution $p(x) = \binom{N}{x} p^x (1-p)^{N-x}$. We learned from HW 1 that this interval given by

$$\hat{p} \pm 1.96 \times \sqrt{\frac{\hat{p}(1-\hat{p})}{N}}, \quad (1)$$

in practice doesn't hold up to what the theory suggests.

Let us consider a Bayesian method. Under the prior $p \sim \text{Beta}(\frac{1}{2}, \frac{1}{2})$, construct the Posterior distribution $p(p|X)$.

Generate a uniform spacing of p 's from .01 to .99, with a step size of .01. Under each value of p , generate a Binomial random number with $N = 30$. Now, sample from the posterior distribution (take 10,000 samples). Construct a 95% probability interval (Bayesians call these Credible Intervals). Of course these are not unique. Any interval containing 95% of the area under the posterior density is a 95% credible interval. I suggest you use the equal tail interval since it is the easiest one to construct. That is from your 10,000 samples, order them and use the 250th biggest sample as the lower bound and the 9750th biggest sample as the upper bound.

Since you know the actual value of p , you can confirm if this interval does in fact cover the *true* parameter.

Repeat this 10,000 times for each value of p , and keep track of the frequency of times each of 10,000 repetitions cover the known parameter. Show a plot of the behavior of the "true" coverage probabilities for the full range of p 's (0.01, 0.02, 0.03, ..., 0.99) (The y-axis will have frequency on it and the x-axis will be over p).

Repeat the exercise with $N = \{50, 100\}$.

Compare this with your findings in HW 1.

Problem 2

Consider the case where $x_i \sim N(\mu, \sigma^2)$. Denote $X = \{x_1, \dots, x_N\}$. In class we derived that the posterior predictive distribution $\tilde{x}|X, \sigma^2 \sim N(\bar{x}, (1 + \frac{1}{N})\sigma^2)$, under the reference prior $p(\mu) \propto 1$.

Part 1

Under the reference prior $p(\phi) \propto 1/\phi$, find $p(\phi|X)$ (you may just derive the kernel if you wish).

Part 2

Show that $\tilde{x}|X \sim t_{\nu-1}(\bar{x}, (1 + \frac{1}{N})s^2)$.

Problem 3

This next result is very important and can be useful in many situations. We will expand on this in future exercises. Let $x \sim N(\mu, \sigma^2\gamma)$. Parameterize $\phi = \frac{1}{\sigma^2}$. Hence

$$p(x|\mu, \sigma^2\gamma) \propto (\phi/\gamma)^{1/2} \exp\left(-\frac{1}{2}\phi\frac{(x-\mu)^2}{\gamma}\right).$$

Part 1

Letting $\phi \sim \text{Gamma}(\alpha, \beta)$ Find $p(x|\mu, \gamma) = \int p(x|\mu, \gamma/\phi)p(\phi)d\phi$.

Part 2

State what α and β must be for $p(x|\mu, \gamma)$ to have a Cauchy distribution and state the *shift* and the *scale* parameters (a Cauchy is a t-1 distribution).

Part 3

Recall that if $p(y) = \int p(y|\lambda)p(\lambda)d\lambda$, then we can generate $p(y)$ via the following algorithm:

$$\begin{aligned} \text{sample } \tilde{\lambda} &\leftarrow p(\lambda) \\ \text{sample } \tilde{y} &\leftarrow p(y|\tilde{\lambda}). \end{aligned}$$

The resulting \tilde{y} is a perfect sample from $p(y)$.

Use this method to simulate 1,000 Cauchy (t-1 distribution) random variables, using the result you obtained from part 1. Plot a histogram of the result and state the (min, max) from your 1,000 sample draws.

Problem 4

A basic property of the MLE is that it is invariant to transformations. For example, let $\eta = \tau(\theta)$, and let $L(\hat{\theta}|x) = \max_{\theta} L(\theta|x)$. Denote the likelihood function $L^*(\eta|x)$ as the likelihood function under the transformation $\eta = \tau(\theta)$. Letting $L^*(\hat{\eta}|x) = \max_{\eta} L^*(\eta|x)$ we have that $\hat{\eta} = \tau(\hat{\theta})$. This result holds for ALL functions $\tau(\cdot)$.

The Map (Maximum A-Posteriori) estimator is defined to be $\hat{\theta}$ such that $p(\hat{\theta}|x) = \max_{\theta} p(\theta|x)$. $\hat{\theta}$ can also be referred to as the posterior mode. Is the MAP estimator invariant to transformations? That is, if we let $\eta = \tau(\theta)$, and denote $p(\hat{\eta}|x) = \max_{\eta} p(\eta|x)$, is $\hat{\eta} = \tau(\hat{\theta})$? If so, prove it. If not, disprove.

Problem 5

The iterated expectation and variance formulas follow as

$$\begin{aligned} E[X] &= E[E[X|Y]], \quad \text{and} \\ V(X) &= V(E[X|Y]) + E[V(X|Y)] \end{aligned}$$

respectively. These are very useful in many cases. Recall, in class, we derived $p(\tilde{x}|X, \sigma^2) \sim N(\bar{x}, (1 + \frac{1}{n})\sigma^2)$. We will re-derive this result using the iterated formulas.

Part 1

Justify that under $p(\mu) \propto 1$, the predictive distribution $p(\tilde{x}|X, \sigma^2)$ is normally distributed.

Part 2

For using the iterated formulas above, choose and state the random variable that you are conditioning on. Derive the moments of $p(\tilde{x}|X, \sigma^2)$ using the iterated formulas.