Stat 591 – Homework 04 *Solutions*

1. Let $X \sim \mathsf{Bin}(n, \theta)$; the goal is to choose the prior $\pi(\theta)$ such that the marginal distribution for X is $\mathsf{Unif}\{0, 1, \ldots, n\}$. That is,

$$\mathsf{P}(X = x) = \int_0^1 f_\theta(x)\pi(\theta)\, d\theta = \binom{n}{x} \int_0^1 \theta^x (1-\theta)^{n-x} \pi(\theta)\, d\theta$$

should equal $(n+1)^{-1}$ for each $x \in \{0, 1, \ldots, n\}$. A first guess is to take $\pi(\theta) = 1$, uniform. In this case,

$$\mathsf{P}(X = x) = \binom{n}{x} \int_0^1 \theta^x (1-\theta)^{n-x}\, d\theta = \binom{n}{x} B(x+1, n-x+1),$$

where $B(a, b)$ is the usual beta function. If we recall the definition of the beta function for non-negative integer arguments,

$$B(a, b) = \frac{(a-1)!(b-1)!}{(a+b-1)!},$$

then we see that

$$\mathsf{P}(X = x) = \frac{n!}{x!(n-x)!} \cdot \frac{x!(n-x)!}{(n+1)!} = \frac{1}{n+1}.$$

Therefore, the uniform prior for θ results in a uniform marginal distribution for X.

2. (a) The argument here is similar that given in class to justify the claim that "Jeffreys prior for a transformed parameter is the same as the transformation of Jeffreys prior." The key idea is an application of the chain rule from calculus. For a function $h(x, \theta)$ that's differentiable in θ, define a prior

$$\pi(\theta) = \Big\{ \mathsf{V}_\theta \Big(\frac{\partial}{\partial \theta} h(X, \theta) \Big) \Big\}^{1/2}.$$

Let $\eta = g(\theta)$ be smooth one-to-one reparametrization. The goal is to show that the prior for η does not depend on whether we specify it directly or if we transform the prior for θ above following the rules of probability. The prior we get for η by direct application of the formula is

$$\pi_{\text{formula}}(\eta) = \Big\{ \mathsf{V}_{g^{-1}(\eta)} \Big(\frac{\partial}{\partial \eta} h(X, g^{-1}(\eta)) \Big) \Big\}^{1/2}.$$

The prior for η we get by transforming $\pi(\theta)$ is

$$\begin{aligned}
\pi_{\text{trans}}(\eta) &= \pi(g^{-1}(\eta)) \Big| \frac{dg^{-1}(\eta)}{\eta} \Big| \\
&= \Big\{ \mathsf{V}_{g^{-1}(\eta)} \Big(\frac{\partial}{\partial \theta} h(x, \theta) \Big|_{\theta = g^{-1}(\eta)} \Big) \Big\}^{1/2} \Big| \frac{dg^{-1}(\eta)}{\eta} \Big| \\
&= \Big\{ \Big| \frac{dg^{-1}(\eta)}{\eta} \Big|^2 \mathsf{V}_{g^{-1}(\eta)} \Big(\frac{\partial}{\partial \theta} h(x, \theta) \Big|_{\theta = g^{-1}(\eta)} \Big) \Big\}^{1/2} \\
&= \Big\{ \mathsf{V}_{g^{-1}(\eta)} \Big(\frac{\partial}{\partial \theta} h(x, \theta) \Big|_{\theta = g^{-1}(\eta)} \frac{dg^{-1}(\eta)}{\eta} \Big) \Big\}^{1/2} \\
&= \Big\{ \mathsf{V}_{g^{-1}(\eta)} \Big(\frac{\partial}{\partial \eta} h(X, g^{-1}(\eta)) \Big) \Big\}^{1/2}.
\end{aligned}$$

The second-to-last line is based on the property $\mathsf{V}(aY) = a^2\mathsf{V}(Y)$, and the last line is a result of the chain rule. The last expression is exactly $\pi_{\text{formula}}(\eta)$, so we're done.

(b) Take $h(x, \theta) = \log f_\theta(x)$, the log density.

3. (a) With the new "$a = mp$" parametrization, the mean vector is p and the covariance matrix has entries

$$C_{ij} = \begin{cases} \frac{p_i(1-p_i)}{m+1} & \text{if } i = j \\ \frac{p_i p_j}{m+1} & \text{if } i \neq j. \end{cases}$$

The entries in the covariance matrix are decreasing in m, so m controls how concentrated, or "precise," the Dirichlet distribution. Hence the name "precision parameter."

(b) i. When $m \to \infty$, all terms in the matrix approach 0, which indicates that the Dirichlet distribution is rather concentrated (i.e., informative) when $m \to \infty$. When $m \to 0$, the variance-covariance terms all increase, suggesting that the limiting Dirichlet distribution is diffuse (i.e., "non-informative").

ii. The limits of the marginal mean and variance follows immediately from the formulas above. The fact that the limiting marginals are Bernoulli, together with the constraint that $\theta_1 + \cdots + \theta_4 = 1$, suggests that the limiting Dirichlet distribution must be concentrated at the corners of the simplex. That is, a sample from the limiting Dirichlet can only be $(1, 0, 0, 0)$, $(0, 1, 0, 0)$, $(0, 0, 1, 0)$, or $(0, 0, 0, 1)$.

(c) This is a simple modification to the previous code. Since the sample size $n = 100$ is relatively large, it's unlikely that this minor change to the posterior will have any effect; that is, I expect that the posterior summaries here and in previous homework are indistinguishable.

4. (a) The normal likelihood can be written as

$$L(\mu, \sigma^2) = (1/\sigma^2)^{n/2} e^{-D/\sigma^2},$$

where $D = \frac{1}{2}\{(n-1)s^2 + n(\mu - \bar{x})^2\}$, where $s^2 = \frac{1}{n-1}\sum_{i=1}^n (x_i - \bar{x})^2$. With prior $\pi(\mu, \sigma^2) \propto 1/\sigma^2$, the posterior density satisfies

$$\pi(\mu, \sigma^2 \mid x) \propto (1/\sigma^2)^{n/2+1} e^{-D/\sigma^2}.$$

The right-hand side above is proportional to the density of a known distribution, namely, the normal inverse gamma distribution.[1] But it's not so important to recognize the distribution here.

(b) For the marginal posterior distribution of μ, we need to integrate out σ^2 from the posterior $\pi(\mu, \sigma^2 \mid x)$. The key here is that the right-hand side in the above

[1] `http://en.wikipedia.org/wiki/Normal-inverse-gamma_distribution`

display is, as a function σ^2, proportional to an inverse gamma density which has form

$$\frac{b^a}{\Gamma(a)}\Big(\frac{1}{x}\Big)^{a+1} e^{-b/x}.$$

Therefore, if we integrate over σ^2 in the above display, then we get

$$\int_0^\infty (1/\sigma^2)^{n/2+1} e^{-D/\sigma^2}\, d\sigma^2 = \frac{\Gamma(n/2)}{D^{n/2}},$$

and the marginal posterior density must satisfy

$$\pi(\mu \mid x) \propto \frac{\Gamma(n/2)}{D^{n/2}} \propto \Big(\frac{1}{(n-1)s^2 + n(\mu - \bar{x})^2}\Big)^{n/2}.$$

With a little reflection (and some help from wikipedia!), you'll see that the expression on the right-hand side above, as a function of μ, is proportional to a location-scale transformation of Student-t density with $n-1$ degrees of freedom; that is, given x, the distribution of $n^{1/2}(\mu - \bar{x})/s$ is $\mathsf{t}(n-1)$.

(c) Since the marginal posterior distribution of μ is a familiar one, no hard work is needed to get the credible upper bound. That is,

$$\bar{\mu}_\alpha(x) = \bar{x} + s n^{-1/2} t^\star_\alpha,$$

where $t^\star_\alpha$ is the upper α quantile of the $\mathsf{t}(n-1)$ distribution.

(d) The coverage probability for $\bar{\mu}_\alpha(X)$ is

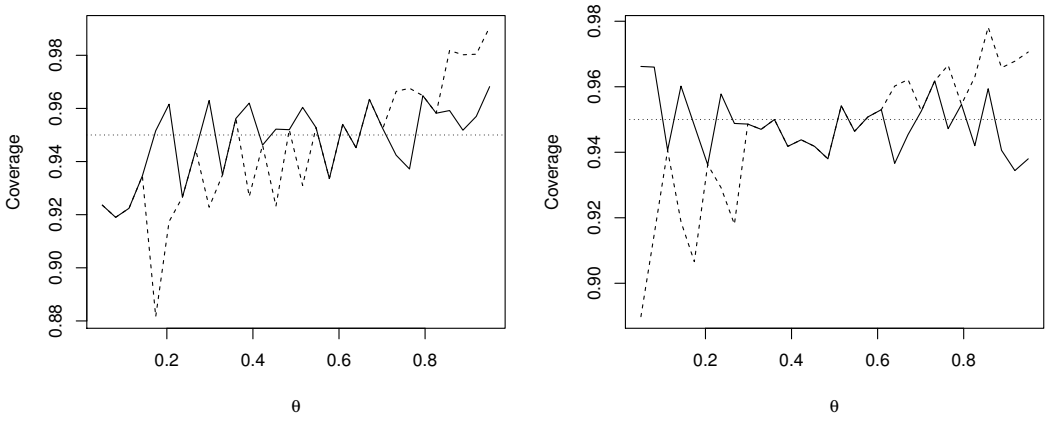
$$\mathsf{P}_{(\mu,\sigma^2)}\{\bar{X} + S n^{-1/2} t^\star_\alpha \geq \mu\} = \mathsf{P}_{(\mu,\sigma^2)}\Big\{\frac{n^{1/2}(\bar{X} - \mu)}{S} \geq -t^\star_\alpha\Big\}.$$

Since the sampling distribution of $n^{1/2}(\bar{X} - \mu)/S$ is $\mathsf{t}(n-1)$, the probability above is $1 - \alpha$, as was to be shown.

5. (a) It is easy to check that Jeffreys prior for the success probability in a binomial problem corresponds to a $\mathsf{Beta}(\frac{1}{2}, \frac{1}{2})$ prior. Since beta is a conjugate prior, the posterior is also beta and, in particular, $\theta \mid X \sim \mathsf{Beta}(\frac{1}{2} + X, \frac{1}{2} + n - X)$.

(b) The quantity $\bar{\theta}_\alpha(X)$ is the upper α quantile of the $\mathsf{Beta}(\frac{1}{2} + X, \frac{1}{2} + n - X)$ distribution. This can be computed via the `qbeta` function in R.

(c) Coverage probabilities of upper 95% confidence limits based on Jeffreys prior and on the classical Wald test are are plotted as functions of θ in Figure 1. Code used to generate these simulation results is give in Figure 2. Here the coverage probability for Jeffreys interval is generally closer to the target coverage 0.95 than Wald's interval.



(a) $n = 50$ (b) $n = 100$

Figure 1: Coverage probabilities for upper 95% confidence limits based on Jeffreys prior (solid) and the classical Wald test (dashed) for two values of n.

```
M <- 5000
N <- c(50, 100)
Theta <- seq(0.05, 0.95, len=30)
b.cvg <- w.cvg <- matrix(0, nrow=length(N), ncol=length(Theta))
for(i in seq_along(N)) {

  for(j in seq_along(Theta)) {

    X <- rbinom(M, N[i], Theta[j])
    b <- qbeta(0.95, 0.5 + X, 0.5 + N[i] - X)
    w <- X / N[i] + 1.65 * sqrt(X * (N[i] - X) / N[i]**3)
    b.cvg[i,j] <- mean(b >= Theta[j])
    w.cvg[i,j] <- mean(w >= Theta[j])

  }

}
plot(Theta, b.cvg[1,], type="l", ylim=range(b.cvg[1,], w.cvg[1,]),
     xlab=expression(theta), ylab="Coverage")
lines(Theta, w.cvg[1,], lty=2)
abline(h=0.95, lty=3)
plot(Theta, b.cvg[2,], type="l", ylim=range(b.cvg[2,], w.cvg[2,]),
     xlab=expression(theta), ylab="Coverage")
lines(Theta, w.cvg[2,], lty=2)
abline(h=0.95, lty=3)
```

Figure 2: R code for the binomial coverage probability simulation.