**Stat 591 – Homework 02** ***Solutions***

1. Basic model is $X \mid \theta \sim \mathsf{Exp}(\theta)$ and the prior is $\mathsf{Gamma}(a, b)$; both the exponential and gamma are taken to have a *rate* parametrization.

   (a) When $X$ is censored, and we only observe $\{X < 100\}$ $(Z = 0)$ or $\{X \geq 100\}$ $(Z = 1)$, the sample is $Z$ is Bernoulli with parameter $e^{-100\theta}$. Since we observe $Z = 1$, the likelihood function is $e^{-100\theta}$ which, when combined with the prior, gives a gamma posterior, i.e., $\theta \mid (Z = 1) \sim \mathsf{Gamma}(a, b + 100)$. The posterior mean and variance are, respectively,

   $$\mathsf{E}(\theta \mid Z = 1) = \frac{a}{b + 100} \quad \text{and} \quad \mathsf{V}(\theta \mid Z = 1) = \frac{a}{(b + 100)^2}.$$

   (b) If the sample $X = 100$ is observed exactly, then we have the likelihood $\theta e^{-100\theta}$. Combined with the gamma prior, we get another gamma posterior, i.e., $\theta \mid (X = 100) \sim \mathsf{Gamma}(a + 1, b + 100)$. The posterior mean and variance are, respectively,

   $$\mathsf{E}(\theta \mid X = 100) = \frac{a + 1}{b + 100} \quad \text{and} \quad \mathsf{V}(\theta \mid X = 100) = \frac{a + 1}{(b + 100)^2}.$$

   (c) Notice that $\mathsf{V}(\theta \mid Z = 1) < \mathsf{V}(\theta \mid X = 100)$; that is, the posterior variance for the censored observation is smaller than that for the exact observation. This is counter-intuitive because one would expect the posterior variance to be smaller when a more informative sample is available. Although this is counter-intuitive, the mathematics aren't wrong. The connection we have between these posterior variances is the following:

   $$\mathsf{E}\{\mathsf{V}(\theta \mid X)\} + \mathsf{V}\{\mathsf{E}(\theta \mid X)\} = \mathsf{E}\{\mathsf{V}(\theta \mid Z)\} + \mathsf{V}\{\mathsf{E}(\theta \mid Z)\},$$

   where the equality is due to the fact that both the right- and left-hand sides equal the prior variance $\mathsf{V}(\theta) = a/b^2$. From this it is clear that the only connection is based on expectations, so for particular $X$ and $Z$, the counter-intuitive relationship between posterior variance is surely possible.

2. Let $(X_1, \ldots, X_n) \mid \mu \overset{\text{iid}}{\sim} \mathsf{N}(\mu, \sigma^2)$, with $\sigma^2$ known.

   (a) (Problem 2.3a in [GDS].) Goal is to test $H_0 : \mu \leq \mu_0$ vs. $H_1 : \mu > \mu_0$, where $\mu_0$ is some fixed value. The p-value for this test is

   $$\mathsf{pval}(H_0; z) = 1 - \Phi(Z) = \Phi(-z),$$

   where $z = n^{1/2}(\bar{x} - \mu_0)/\sigma$. If we take a uniform prior for $\mu$, i.e., the prior density is constant on $\mathbb{R}$, then it is easy to check (by symmetry of the normal PDF) that the posterior distribution of $\mu$ is $\mathsf{N}(\bar{x}, \frac{\sigma^2}{n})$. Therefore, the posterior probability of $H_0 = \{\mu \leq \mu_0\}$

   $$\Pi(H_0 \mid x) = \Pi(\mu \leq \mu_0 \mid x) = \Phi(-z).$$

   This is the same as the p-value.

(b) (Problem 2.21 in [GDS].) Now consider $H_0 : \mu = \mu_0$ vs. $H_1 : \mu \neq \mu_0$. This time, the p-value is

$$\mathsf{pval}(H_0; z) = 2\{1 - \Phi(|z|)\}.$$

For the posterior probability of $H_0$, we have

$$\Pi(H_0 \mid x) = \frac{\Pi(H_0) m_0(x)}{\Pi(H_0) m_0(x) + \Pi(H_1) m_1(x)} = \frac{m_0(x)}{m_0(x) + m_1(x)},$$

where $\Pi(H_0) = \Pi(H_1) = \frac{1}{2}$ are the prior probabilities for the two hypotheses, and $m_0(x)$ and $m_1(x)$ are the marginal likelihoods under $H_0$ and $H_1$, respectively. For $H_0$, the marginal likelihood is easy:

$$m_0(x) = L(\mu_0) = \Big(\frac{1}{2\pi\sigma^2}\Big)^{n/2} e^{-\frac{n}{2\sigma^2}[\hat\sigma^2 + (\bar x - \mu_0)^2]},$$

where $\hat\sigma^2 = \frac{1}{n}\sum_{i=1}^n (x_i - \bar x)^2$. For $H_1$, the marginal likelihood will require some integration:

$$\begin{aligned} m_1(x) &= \int L(\mu)\pi(\mu)\,d\mu \\ &= \Big(\frac{1}{n}\Big)^{1/2} \Big(\frac{1}{2\pi\sigma^2}\Big)^{(n-1)/2} e^{-n\hat\sigma^2/2\sigma^2} \int \Big(\frac{n}{2\pi\sigma^2}\Big)^{1/2} e^{-\frac{n}{2\sigma^2}(\bar x - \mu)^2} \pi(\mu)\,d\mu. \end{aligned}$$

Under $H_1$, the prior $\pi(\mu)$ is a $\mathsf{N}(\mu_0, \tau^2)$ density; note that the prior mean is the mean value under $H_0$. The integral above corresponds to the marginal density of $\bar X$ under the model $\bar X \mid \mu \sim \mathsf{N}(\mu, \frac{\sigma^2}{n})$ and $\mu \sim \mathsf{N}(\mu_0, \tau^2)$. This distribution is easy to find without integration: it's normal with mean

$$\mathsf{E}(\bar X) = \mathsf{E}\{\mathsf{E}(\bar X \mid \mu)\} = \mu_0$$

and variance

$$\mathsf{V}(\bar X) = \mathsf{V}\{\mathsf{E}(\bar X \mid \mu)\} + \mathsf{E}\{\mathsf{V}(\bar X \mid \mu)\} = \tau^2 + \sigma^2/n.$$

Therefore, the marginal likelihood under $H_0$ is

$$\begin{aligned} m_1(x) &= \Big(\frac{1}{n}\Big)^{1/2} \Big(\frac{1}{2\pi\sigma^2}\Big)^{(n-1)/2} e^{-n\hat\sigma^2/2\sigma^2} \Big(\frac{1}{2\pi(\tau^2 + \sigma^2/n)}\Big)^{1/2} e^{-\frac{n}{2(n\tau^2+\sigma^2)}(\bar x - \mu_0)^2} \\ &= \Big(\frac{1}{2\pi\sigma^2}\Big)^{n/2} e^{-n\hat\sigma^2/2\sigma^2} \Big(\frac{\sigma^2}{2\pi(n\tau^2 + \sigma^2)}\Big)^{1/2} e^{-\frac{n}{2(n\tau^2+\sigma^2)}(\bar x - \mu_0)^2}. \end{aligned}$$

Canceling out all the common factors in $m_0(x)$ and $m_1(x)$ we can finally write the posterior probability for $H_0$ as

$$\begin{aligned} \Pi(H_0 \mid x) &= \frac{e^{-\frac{n}{2\sigma^2}(\bar x - \mu_0)^2}}{e^{-\frac{n}{2\sigma^2}(\bar x - \mu_0)^2} + \big(\frac{\sigma^2}{2\pi(n\tau^2+\sigma^2)}\big)^{1/2} e^{-\frac{n}{2(n\tau^2+\sigma^2)}(\bar x - \mu_0)^2}} \\ &= \frac{1}{1 + \big(\frac{\sigma^2}{2\pi(n\tau^2+\sigma^2)}\big)^{1/2} e^{-\frac{n\tau^2}{n\tau^2+\sigma^2} z^2}}, \end{aligned}$$

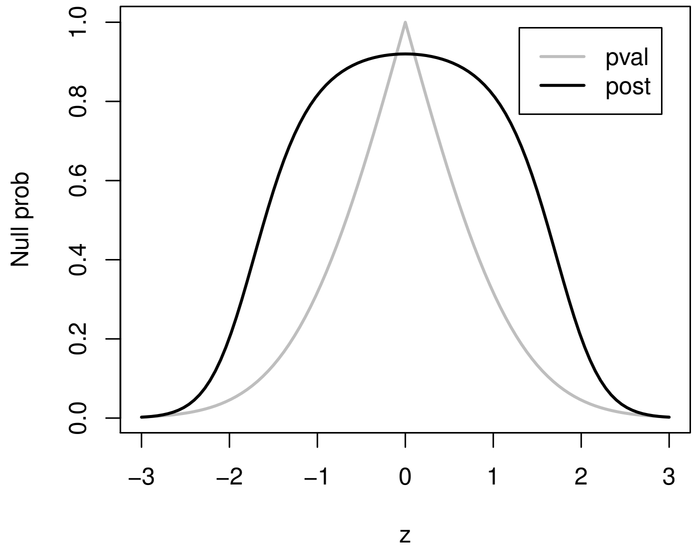


Figure 1: P-value and posterior probability of $H_0$ as a function of the z-score $z$.

where $z$ is as defined above. To see that this can differ from the p-value, we plot $\mathsf{pval}(H_0; z)$ and $\Pi(H_0 \mid z)$ as functions of $z$, where $n = 10$, $\sigma^2 = 1$, and $\tau^2 = 2$. You can see in Figure 1 that the p-value and posterior probability are drastically different at some intermediate values, e.g., between 1 and 2. In particular, it can happen that p-value $\leq 0.1$ and posterior $\geq 0.5$!

3. — (Problem 2.13bc in [GDS].) Suppose $X_1, \ldots, X_k \overset{\text{iid}}{\sim} \mathsf{Bin}(n, p)$ where both $n$ and $p$ are unknown. Here $n$ is the parameter of interest; $p$ is a nuisance parameter. Here we construct some likelihood functions for $n$. Note that the usual likelihood is

$$L(n, p) = \prod_{i=1}^{k} \binom{n}{X_i} p^{X_i} (1-p)^{n-X_i} = \prod_{i=1}^{n} \binom{n}{X_i} \cdot p^{k\bar{X}} (1-p)^{nk-k\bar{X}}.$$

(i) *Profile likelihood.* For a given $n$, it is easy to check that the conditional MLE $\hat{p}(n)$ is $\bar{X}/n$. Therefore, the profile likelihood is just the usual likelihood with $\hat{p}(n) = \bar{X}/n$ plugged in, i.e.,

$$L_{\text{prof}}(n) = L(n, \hat{p}(n)) = \prod_{i=1}^{n} \binom{n}{X_i} \left[\hat{p}(n)^{\hat{p}(n)} (1 - \hat{p}(n))^{1-\hat{p}(n)}\right]^{nk}.$$

(ii) *Conditional likelihood.* The conditional PMF of $(X_1, \ldots, X_k)$, given the sum $\sum_{i=1}^{k} X_i$, has distribution free of $p$. Let's check this. Let

$$g(x_1, \ldots, x_k; t) = \mathsf{P}(X_1 = x_1, \ldots, X_k = x_k \mid X_1 + \cdots + X_k = t).$$

In the calculation that follows, we consider only $(x_1, \ldots, x_k)$ and $t$ such that $\sum_i x_i = t$; in any other case, the conditional probability is zero. Then

$$\begin{aligned} g(x_1,\dots,x_k;t) &= \frac{\mathsf{P}(X_1=x_1,\dots,X_k=x_k, X_1+\cdots+X_k=t)}{\mathsf{P}(X_1+\cdots+X_k=t)} \\ &= \frac{\mathsf{P}(X_1=x_1,\dots,X_k=t-x_1-\cdots-x_{k-1})}{\mathsf{P}(X_1+\cdots+X_k=t)} \\ &= \frac{\prod_{i=1}^{k-1}\binom{n}{x_i}\binom{n}{t-x_1-\ldots-x_{k-1}}p^t(1-p)^{nk-t}}{\binom{nk}{t}p^t(1-p)^{nk-t}} \\ &= \frac{\prod_{i=1}^{k-1}\binom{n}{x_i}\binom{n}{t-x_1-\ldots-x_{k-1}}}{\binom{nk}{t}} \end{aligned}$$

Then the conditional likelihood is

$$L_{\text{cond}}(n) = \frac{\prod_{i=1}^{k}\binom{n}{X_i}}{\binom{nk}{X_1+\cdots+X_k}}.$$

(iii) If we consider a uniform prior on $p$, then the marginal likelihood is

$$L_{\text{marg1}}(n) = \int_0^1 L(n,p)\,dp = \prod_{i=1}^{n}\binom{n}{X_i}B(k\bar{X}+1, nk-k\bar{X}+1),$$

where $B(\cdot,\cdot)$ is the usual beta function.

(iv) If we consider Jeffreys prior for $p$, i.e., $p \sim \mathsf{Beta}(\frac{1}{2},\frac{1}{2})$, then the marginal likelihood is similar to that above:

$$L_{\text{marg2}}(n) = \int_0^1 L(n,p)\pi(p)\,dp = \prod_{i=1}^{n}\binom{n}{X_i}B(k\bar{X}+\tfrac{1}{2}, nk-k\bar{X}+\tfrac{1}{2}).$$

If we're given data $X = (17, 19, 21, 28, 30)$, so that $k = 5$ and $\bar{X} = 23$, we can plot the various likelihood functions for $n$. These are shown in Figure 2. There we see that the profile and conditional likelihoods are unbounded (at least on a range comparable to that shown), whereas the two genuine marginal likelihoods are bounded with peaks in the range 80–100.

— (Problem 2.14 in [GDS].)

(a) The vector $\mu$ is a location parameter in the $p$-variate normal.

(b) Clearly, the marginal distributions of $X$ and $\mu$ are $\mathsf{N}_p(\eta, \Gamma+\Sigma)$ and $\mathsf{N}_p(\eta,\Gamma)$, respectively, so all we need to do is see that the joint distribution is $2p$-variate normal. This can be seen by multiplying the marginal density for $\mu$ and the conditional density for $X$, given $\mu$. To figure out the off-diagonal elements of the joint covariance matrix, we need to know $\mathsf{C}(X,\mu)$. Since $X = \mu+\varepsilon$, we get

$$\mathsf{C}(X,\mu) = \mathsf{C}(\mu+\varepsilon,\mu) = \mathsf{V}(\mu) + \mathsf{C}(\varepsilon,\mu) = \mathsf{V}(\mu) = \Gamma.$$

(c) Look at the general results at, e.g., `http://en.wikipedia.org/wiki/Multivariate_normal_distribution#Conditional_distributions`

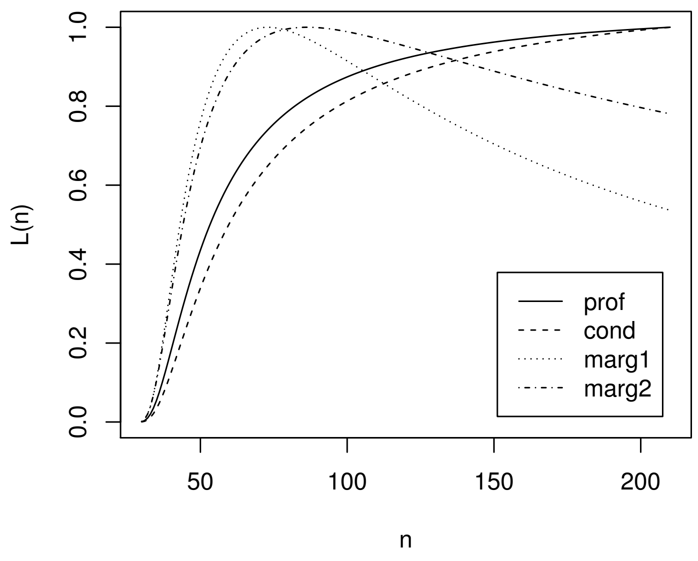


Figure 2: Various "marginal likelihoods" for $n$ in the binomial problem; these are normalized by dividing by the respective maxima over the shown range.

(d) The posterior mean for $\mu$ is

$$\eta_x := \mathsf{E}(\mu \mid x) = \Gamma(\Sigma + \Gamma)^{-1}x + \Sigma(\Sigma + \Gamma)^{-1}\eta,$$

and the posterior covariance/dispersion matrix is

$$\Gamma_x := \mathsf{V}(\mu \mid x) = \Gamma - \Gamma(\Sigma + \Gamma)^{-1}\Gamma.$$

A HPD credible region for $\mu$ would be one where the posterior density $\pi(\mu \mid x)$ is bigger than some constant. Clearly this is equivalent to a set where $(\mu - \eta_x)^\top \Gamma_x^{-1}(\mu - \eta_x) \leq c$. Since the quantity on the left-hand side of this inequality is a $\mathsf{ChiSq}(p)$ random variable, to make the credible region have posterior probability $1 - \alpha$, we should take $c$ to equal $\chi^2_p(\alpha)$, the $100(1 - \alpha)$ percentile of the $\mathsf{ChiSq}(p)$ distribution, i.e.,

$$\text{credible region} = \{m : (m - \eta_x)^\top \Gamma_x^{-1}(m - \eta_x) \leq \chi^2_p(\alpha)\}.$$

(e) If the prior for $\mu$ is uniform, i.e., the prior density is constant, then the posterior is still normal; this is an easy consequence of the symmetry of the normal density. In particular, under the uniform prior, the posterior is $\mathsf{N}_p(x, \Sigma)$. The rest of the problem is just like above.

4. (a) Just multiply the prior PDF and multinomial PMF. Easy to see that posterior is also of Dirichlet form, with $a_i' = a_i + x_i$, $i = 1, \ldots, 4$.

   (b) One can simulate from a Dirichlet distribution, say, $\theta \sim \mathsf{Dir}_4(a')$ by simulating four independent gamma random variables and dividing each by the sum. That is, take $\lambda_i \sim \mathsf{Gamma}(a_i', 1)$, $i = 1, \ldots, 4$, independent, and define

$$\theta = \Big(\frac{\lambda_1}{\lambda_+}, \frac{\lambda_2}{\lambda_+}, \frac{\lambda_3}{\lambda_+}, \frac{\lambda_4}{\lambda_+}\Big),$$

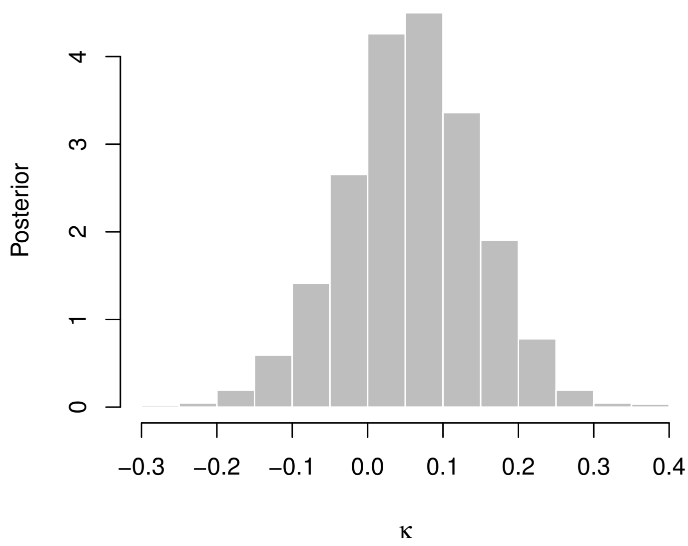


Figure 3: Histogram of a sample from $\kappa$ posterior.

where $\lambda_+ = \lambda_1 + \lambda_2 + \lambda_3 + \lambda_4$. Now, if $g(\theta)$ is a function of $\theta$, then we can simulate from the posterior distribution of $g(\theta)$ by sampling $\theta$ as described above, and then applying the $g$ function to the sampled $\theta$.

(c) We can use the recipe above to simulate from the posterior of $\theta$ and, consequently, the posterior of $\kappa$. A histogram of 3000 posterior samples, with $a = (1, 1, 1, 1)$, based on the data from Homework #1, is shown in Figure 3. This histogram looks similar to that for the bootstrap distribution of the MLE in Homework #1. The 90% equip-tailed credible interval is $(-0.085, 0.213)$, which is also similar to the corresponding bootstrap interval.

## R code for Problem 2.21 in [GDS].

```
n <- 10
sigma2 <- 1
tau2 <- 2
pval <- function(z) 2 * (1 - pnorm(abs(z)))
post <- function(z) {

  a <- sigma2 / (n * tau2 + sigma2)
  o <- 1 / (1 + sqrt(a / 2 / pi) * exp(-(a - 1) * z**2))
  return(o)

}
curve(pval, xlim=c(-3, 3), xlab="z", ylab="Null prob", col="gray", lwd=2)
curve(post, lwd=2, add=TRUE)
legend(x="topright", inset=0.05, lwd=2, col=c("gray", "black"), c("pval", "post"))
```

## R code for Problem 2.13c in [GDS].

```
x <- c(17, 19, 21, 28, 30)
k <- length(x)
sx <- sum(x)
Lprof <- function(n) {

  o <- 0 * n
  for(i in seq_along(n)) {

    phat <- sx / k / n[i]
    g <- sum(lchoose(n[i], x))
    o[i] <- g + sx * log(phat) + (n[i] * k - sx) * log(1 - phat)

  }
  return(exp(o))

}
Lcond <- function(n) {

  o <- 0 * n
  for(i in seq_along(n)) {

    g <- sum(lchoose(n[i], x))
    o[i] <- g - log(choose(n[i] * k, sx))

  }
  return(exp(o))

}
Lmarg1 <- function(n) {

  o <- 0 * n
  for(i in seq_along(n)) {

    g <- sum(lchoose(n[i], x))
    o[i] <- g + log(beta(sx + 1, n[i] * k - sx + 1))
```

```
  }
  return(exp(o))

}
Lmarg2 <- function(n) {

  o <- 0 * n
  for(i in seq_along(n)) {

    g <- sum(lchoose(n[i], x))
    o[i] <- g + log(beta(sx + 0.5, n[i] * k - sx + 0.5))

  }
  return(exp(o))

}
N <- max(x):(7 * max(x))
LprofN <- Lprof(N)
LcondN <- Lcond(N)
Lmarg1N <- Lmarg1(N)
Lmarg2N <- Lmarg2(N)
plot(0, 0, type="n", xlim=range(N), ylim=c(0,1), xlab="n", ylab="L(n)")
lines(N, LprofN / max(LprofN))
lines(N, LcondN / max(LcondN), lty=2)
lines(N, Lmarg1N / max(Lmarg1N), lty=3)
lines(N, Lmarg2N / max(Lmarg2N), lty=4)
legend(x="bottomright", inset=0.05, lty=1:4, c("prof", "cond", "marg1", "marg2"))
```

## R code for the Dirichlet-Multinomial example in Problem 4.

```
rdir <- function(a, dim) {

  if(length(a) != dim) stop("Mismatched dimensions!")
  V <- rgamma(dim, shape=a)
  return(V / sum(V))

}
g <- function(u) {

  o <- (u[1] + u[2]) * (u[1] + u[3]) + (u[3] + u[4]) * (u[2] + u[4])
  fn <- ((u[1] + u[4] - o) / (1 - o))
  return(fn)

}
x <- c(22, 15, 33, 30)
n <- sum(x)
a <- 1 + 0 * x
ax <- a + x
kappa <- numeric(3000)
for(m in 1:3000) kappa[m] <- g(rdir(ax, 4))
hist(kappa, freq=FALSE, col="gray", border="white", xlab=expression(kappa), ylab="Posterior")
cred.int <- as.numeric(quantile(kappa, c(0.05, 0.95)))
```