MA 578 HW2 Solutions

1 (BDA 2.10)

The derivations here are analytical, but you can get the same results numerically by summing enough terms and controlling the error.

```r
X <- 203
p <- 1 / 100
# check details below
rx <- 1:(X - 1)
C <- -log(p) - sum((1 - p) ^ rx / rx)
mean_posterior <- (1 / p - 1 - sum((1 - p) ^ rx)) / C
sd_posterior <- sqrt(((1 - p) / p ^ 2 - sum(rx * (1 - p) ^ rx)) / C - mean_posterior ^ 2)
```

(a)

With $p = 1/100$ and $X = 203$, the posterior is

$$\mathbb{P}(N \mid X) = \frac{p(1-p)^{N-1} I(N \geq X)/N}{\sum_{\tilde{N}=X}^{\infty} p(1-p)^{\tilde{N}-1}/\tilde{N}} = \frac{(1-p)^N/N}{\sum_{\tilde{N}=X}^{\infty}(1-p)^{\tilde{N}}/\tilde{N}} I(N \geq X) = \frac{(1-p)^N I(N \geq X)}{NC},$$

where the normalizing constant C is

$$C = \sum_{\tilde{N}=X}^{\infty} \frac{(1-p)^{\tilde{N}}}{\tilde{N}} = \sum_{\tilde{N}=1}^{\infty} \frac{(1-p)^{\tilde{N}}}{\tilde{N}} - \sum_{\tilde{N}=1}^{X-1} \frac{(1-p)^{\tilde{N}}}{\tilde{N}} = -\log p - \sum_{\tilde{N}=1}^{X-1} \frac{(1-p)^{\tilde{N}}}{\tilde{N}} = 0.04658.$$

(b)

The posterior mean is

$$\mathbb{E}[N \mid X] = \frac{1}{C} \sum_{N=X}^{\infty} N \frac{(1-p)^N}{N} = \frac{1}{C}\left(\frac{1}{p} - \sum_{N=0}^{X-1}(1-p)^N\right) = 279.0885.$$

Since

$$\mathbb{E}[N^2 \mid X] = \frac{1}{C} \sum_{N=X}^{\infty} N^2 \frac{(1-p)^N}{N} = \frac{1}{C}\left(\frac{1-p}{p^2} - \sum_{N=1}^{X-1} N(1-p)^N\right),$$

then $\mathrm{SD}[N \mid X] = \sqrt{\mathbb{E}[N^2 \mid X] - \mathbb{E}[N \mid X]^2} = 79.9646$.

(c)

If we had K observations $X_1, \ldots, X_K$ with $X^* = \max_{k=1,\ldots,K}\{X_k\}$, the likelihood would be $\mathbb{P}(X \mid N) = I(N \geq X^*)/N^K$, which would suggest a conjugate prior $\mathbb{P}(N) \propto N^{-\alpha}$, a discrete *Pareto* distribution. A flat prior would correspond to $\alpha = 0$, while Jeffreys prior, $\mathbb{P}(N) \propto 1/N$, has $\alpha = 1$.

The posterior is

$$\mathbb{P}(N \mid X) = \frac{I(N \geq X^*)N^{-\alpha-K}}{\sum_{\tilde{N}=X^*}^{\infty} \tilde{N}^{-\alpha-K}} = \frac{1}{C} I(N \geq X^*) N^{-\alpha-K},$$

where the normalizing constant C is now

$$C = \sum_{\tilde{N}=X^*}^{\infty} \tilde{N}^{-\alpha-K} = \zeta(\alpha+K) - \sum_{\tilde{N}=1}^{X^*-1} \tilde{N}^{-\alpha-K},$$

with $\zeta(s) = \sum_{k=1}^{\infty} k^{-s}$ the Riemann zeta function.

If we denote $F_s(x) = \sum_{k=x}^{\infty} k^{-s} = \zeta(s) - \sum_{k=1}^{x-1} k^{-s}$, then $C = F_{\alpha+K}(X^*)$. The posterior mean is then

$$\mathbb{E}[N \mid X] = \frac{1}{C} \sum_{\tilde{N}=X^*}^{\infty} \tilde{N}^{-\alpha-K+1} = \frac{F_{\alpha+K-1}(X^*)}{F_{\alpha+K}(X^*)},$$

while the variance is

$$\mathrm{Var}[N \mid X] = \mathbb{E}[N^2 \mid X] - \mathbb{E}[N \mid X]^2 = \frac{F_{\alpha+K-2}(X^*)}{F_{\alpha+K}(X^*)} - \left(\frac{F_{\alpha+K-1}(X^*)}{F_{\alpha+K}(X^*)}\right)^2.$$

Back to our case, we have $K = 1$ and $X = 203$. A flat prior would result in an improper posterior since $\zeta(1) = \infty$. Setting $\alpha = 1$ for Jeffreys prior would get us propriety since $C = F_2(X) = \pi^2/6 - \sum_{\tilde{N}=1}^{X-1} \tilde{N}^{-2} < \infty$, but unfortunately $\mathbb{E}[N \mid X] = \mathrm{Var}[N \mid X] = \infty$. We would need at least two more observations to have finite posterior mean and variance. Other possible non-informative prior would be uniform but under a reasonably large upper bound N_0, $\mathbb{P}(N) = I(1 \le N \le N_0)/N_0$.

2 (BDA 2.13)

Here we set Jeffreys prior, $\theta \sim \mathsf{Gamma}(\alpha, \beta)$ with $\alpha = 1/2$ and $\beta = 0$ (improper). From our discussions in lecture, we know that if $Y_i \mid \theta \overset{\text{ind}}{\sim} \mathsf{Po}(\theta X_i)$, $i = 1, \ldots, n$, then the posterior is $\theta \mid Y \sim \mathsf{Gamma}(\alpha + n\bar{Y}, \beta + n\bar{X})$ and the posterior predictive at $\tilde{X}$ is

$$\tilde{Y} \mid Y \sim \mathsf{NegBinom}\left(\alpha + n\bar{Y}, \frac{\tilde{X}}{\tilde{X} + \beta + n\bar{X}}\right).$$

```
# Setup: data and prior parameters
airline <- data.frame(
              year = 1976:1985,
              fatal = c(24, 25, 31, 31, 22, 21, 26, 20, 16, 22),
              deaths = c(734, 516, 754, 877, 814, 362, 764, 809, 223, 1066),
              deathrate = c(.19, .12, .15, .16, .14, .06, .13, .13, .03, .15)
           )
alpha <- .5; beta <- 0 # Jeffreys prior
quants <- c(0.025, .975) # 95% interval
```

(a) and (b)

The variable of interest here is fatal accidents, but the exposures change. In (a) we just have unit exposures, while in (b) we use number of passenger miles flown.

```
# (a)
a <- alpha + sum(airline$fatal); b <- beta + nrow(airline)
xpred <- 1 # one year
pred.int.a <- qnbinom(quants, a, 1 - xpred / (b + xpred))
message(paste0("95% predictive: (", paste(pred.int.a, collapse=", "), ")"))
```

```
95% predictive: (14, 34)
```

```
# (b)
x <- airline$deaths / airline$deathrate # in miles
a <- alpha + sum(airline$fatal); b <- beta + sum(x)
xpred <- 8e11 / 1e8 # death rate is per 1e8 miles
```

```
pred.int.b <- qnbinom(quants, a, 1 - xpred / (b + xpred))
message(paste0("95% predictive: (", paste(pred.int.b, collapse=", "), ")"))
```

```
95% predictive: (22, 46)
```

(c) and (d)

Similar to (a) and (b) above, but the variable of interest is now passenger deaths.

```
# (c)
a <- alpha + sum(airline$deaths); b <- beta + nrow(airline)
xpred <- 1 # one year
pred.int.c <- qnbinom(quants, a, 1 - xpred / (b + xpred))
message(paste0("95% predictive: (", paste(pred.int.c, collapse=", "), ")"))
```

```
95% predictive: (638, 747)
```

```
# (d)
x <- airline$deaths / airline$deathrate # in miles
a <- alpha + sum(airline$deaths); b <- beta + sum(x)
xpred <- 8e11 / 1e8 # death rate is per 1e8 miles
pred.int.d <- qnbinom(quants, a, 1 - xpred / (b + xpred))
message(paste0("95% predictive: (", paste(pred.int.d, collapse=", "), ")"))
```

```
95% predictive: (904, 1034)
```

(e)

The more reasonable setting for the Poisson model would be to account for miles flown since longer flights should have a higher risk of an accident. While it would make more sense to consider fatal accidents as "rare" events, the usual interpretation of Poisson events, if we use exposures in terms of passenger miles flown then we have to adjust event rates and have passenger deaths as the variable of interest.

3

(a)

With $\lambda = \log(\theta/(1-\theta))$ we have $\theta = e^\lambda/(1+e^\lambda)$ and so, by change of variables,

$$\mathbb{P}(\lambda) = \left|\frac{d\theta}{d\lambda}\right|\mathbb{P}(\theta) \propto \frac{e^\lambda}{(1+e^\lambda)^2}\left(\frac{e^\lambda}{1+e^\lambda}\right)^{\alpha-1}\left(1-\frac{e^\lambda}{1+e^\lambda}\right)^{\beta-1} = \frac{e^{\alpha\lambda}}{(1+e^\lambda)^{\alpha+\beta}}.$$

(b)

Up to a constant we have

$$\ell(\lambda) = \log \mathbb{P}(X \mid \lambda) = \lambda X - n\log(1+e^\lambda)$$

and so $\ell'(\lambda) = -ne^\lambda/(1+e^\lambda)$, $I(\lambda) = \mathbb{E}_{X\mid\lambda}[-\ell''(\lambda)] = e^\lambda/(1+e^\lambda)^2$, and thus Jeffreys prior is

$$\mathbb{P}(\lambda) \propto I(\lambda)^{1/2} = \frac{e^{\lambda/2}}{1+e^\lambda},$$

that is, $\lambda \sim \text{Beta-logit}(1/2, 1/2)$. Since Jeffreys prior is invariant to reparameterizations, $\theta = \text{logit}^{-1}(\lambda) \sim \text{Beta}(1/2, 1/2)$.