EXERCISE SET NO. 2
BAYESIAN LEARNING

1a) $X_1, \dots, X_n \mid \theta, \sigma^2 \sim N(\theta, \sigma^2)$

θ known

$\sigma^2 \sim \mathrm{Inv}\chi^2(\nu_0, \sigma_0^2)$

Posterior: (implicit conditioning on θ)

$$p(\sigma^2 \mid x_1, \dots, x_n) \propto p(x_1, \dots, x_n \mid \sigma^2)\, p(\sigma^2)$$

$$= \prod_{i=1}^{n} \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left(-\frac{1}{2\sigma^2}(x_i - \theta)^2\right) \cdot p(\sigma^2)$$

$$\left[\text{define } s^2 = \frac{\sum_{i=1}^{n} (x_i - \theta)^2}{n} \right]$$

$$\propto \frac{1}{(\sigma^2)^{n/2}} \exp\left(-\frac{ns^2}{2\sigma^2}\right) \boxed{\frac{\exp\left(-\frac{\nu_0 \sigma_0^2}{2\sigma^2}\right)}{(\sigma^2)^{\nu_0/2+1}}}$$

Density (pdf) of $\mathrm{Inv}\chi^2(\nu_0, \sigma_0^2)$ prior.

$$= \frac{\exp\left(-\frac{ns^2 + \nu_0\sigma_0^2}{2\sigma^2}\right)}{(\sigma^2)^{(n+\nu_0)/2+1}}$$

So $\sigma^2 \mid x_1, \dots, x_n, \theta \sim \mathrm{Inv}\chi^2(\nu_n, \sigma_n^2)$

$$\nu_n = \nu_0 + n$$

$$\sigma_n^2 = \frac{ns^2 + \nu_0\sigma_0^2}{\nu_0 + n}$$

1b) $s^2 = \dfrac{\sum_{i=1}^{3}(x_i-\theta)^2}{3} = \dfrac{(0.6-1)^2+(3.2-1)^2+(1.2-1)^2}{3} = 1.68$

Non-informative : $v_0 \to 0$

why is this non-informative?

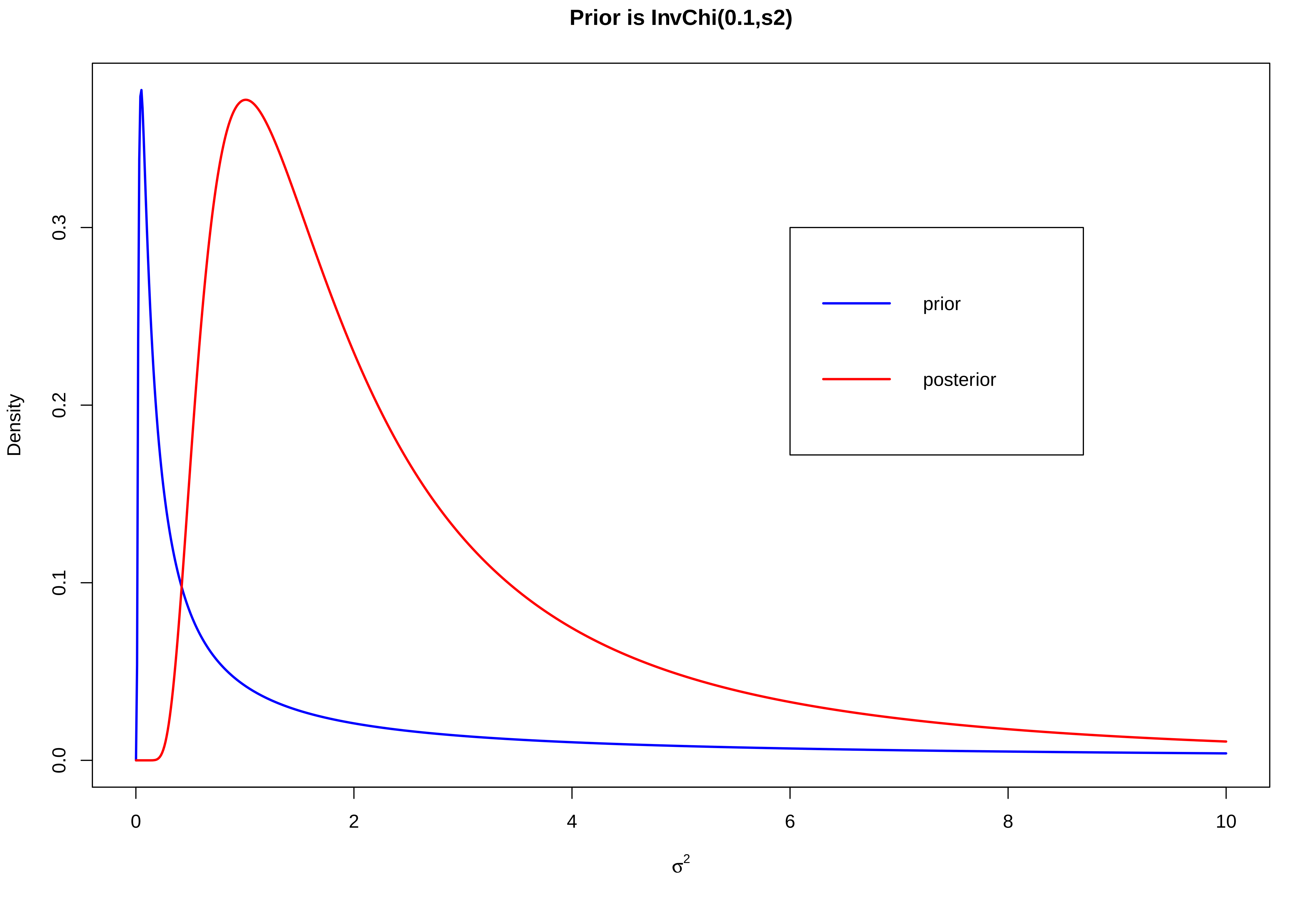
Reason 1: v_n becomes n

Reason 2: $\text{Inv}\chi^2(v_0, \sigma_0^2)$ becomes $\frac{1}{\sigma^2}$
when $v_0 \to 0$.

Note that as $v_0 \to 0$ the posterior approaches the $\text{Inv}\chi^2(n, s^2)$ density.

So,

$\sigma^2 | x_1, x_2, x_3 \sim \text{Inv}\chi^2(3, 1.68)$



Prior is InvChi(0.1,s2)
Density
0.0
0.1
0.2
0.3
0
2
4
6
8
10
σ²
prior
posterior

2a Prediction of Bernoulli data

The predictive distribution of x_{n+1} given the first n trials ($x_{1:n}$) is

$$\begin{aligned}
p(x_{n+1}|x_{1:n}) &= \int p(x_{n+1}|\theta)p(\theta|x_{1:n})d\theta && x_{n+1} \text{ is indep. of } x_{1:n} \text{ given } \theta \\
&= \int \theta^{x_{n+1}}(1-\theta)^{1-x_{n+1}}p(\theta|x_{1:n})d\theta && \theta|x_{1:n} \sim \text{Beta}(\alpha+s,\beta+f) \\
&= \int \theta^{x_{n+1}}(1-\theta)^{1-x_{n+1}}\frac{\Gamma(\alpha+\beta+n)}{\Gamma(\alpha+s)\Gamma(\beta+f)}\theta^{\alpha+s-1}(1-\theta)^{\beta+f-1}d\theta \\
&= \frac{\Gamma(\alpha+\beta+n)}{\Gamma(\alpha+s)\Gamma(\beta+f)}\int \theta^{x_{n+1}+\alpha+s-1}(1-\theta)^{1-x_{n+1}+\beta+f-1}d\theta \\
&= \frac{\Gamma(\alpha+\beta+n)}{\Gamma(\alpha+s)\Gamma(\beta+f)}\frac{\Gamma(x_{n+1}+\alpha+s)\Gamma(1-x_{n+1}+\beta+f)}{\Gamma(1+\alpha+\beta+n)} \\
&= \frac{\Gamma(x_{n+1}+\alpha+s)\Gamma(1-x_{n+1}+\beta+f)}{\Gamma(\alpha+s)\Gamma(\beta+f)(\alpha+\beta+n)} && \text{using } \Gamma(y+1)=y\Gamma(y)
\end{aligned}$$

So,

$$p(x_{n+1}=1|x_{1:n}) = \frac{\Gamma(1+\alpha+s)}{\Gamma(\alpha+s)(\alpha+\beta+n)} = \frac{(\alpha+s)\Gamma(\alpha+s)}{\Gamma(\alpha+s)(\alpha+\beta+n)} = \frac{\alpha+s}{\alpha+\beta+n}$$

and therefore [since $p(x_{n+1}=0|x_{1:n}) = 1 - p(x_{n+1}=1|x_{1:n})$]

$$p(x_{n+1}=0|x_{1:n}) = \frac{\beta+f}{\alpha+\beta+n}.$$

The predictive distribution is therefore

$$x_{n+1}|x_{1:n} \sim \text{Bern}\left(\frac{\alpha+s}{\alpha+\beta+n}\right).$$

2b Umbrella decision

(a) Let x_{11} be the binary variable indicating rain on the 11th day. From Problem 2a, the predictive distribution for the $(n+1)$th Bernoulli trial is

$$x_{n+1}|x_{1:n} \sim \text{Bern}\left(\frac{\alpha+s}{\alpha+\beta+n}\right).$$

and the predictive probability for rain is therefore here

$$\Pr(x_{11}=1|x_{1:10}) = \frac{1+2}{1+1+10} = 0.25.$$

The expected utility from the decision to bring the umbrella is then

$$\mathrm{EU}_{\mathrm{bring}} = \Pr(\mathrm{sunny}) \cdot \mathrm{U}(\mathrm{bring}, \mathrm{sunny}) + \Pr(\mathrm{rain}) \cdot \mathrm{U}(\mathrm{bring}, \mathrm{rain}) = 0.75 \cdot 20 + 0.25 \cdot 10 = 17.5$$

and the expected utility of leaving the umbrella at home is

$$\mathrm{EU}_{\mathrm{leave}} = \Pr(\mathrm{sunny}) \cdot \mathrm{U}(\mathrm{leave}, \mathrm{sunny}) + \Pr(\mathrm{rain}) \cdot \mathrm{U}(\mathrm{leave}, \mathrm{rain}) = 0.75 \cdot 50 + 0.25 \cdot (-50) = 25.0.$$

The expected utility is therefore maximized by leaving the umbrella at home. This is the Bayesian decision.

2c Figure 15.1 shows how the optimal Bayesian decision varies for different combinations of the prior hyperparameters.

Figure 15.2 shows how the optimal Bayesian decision varies for different combinations of the prior hyperparameters when $s = 16$ and $f = 64$.

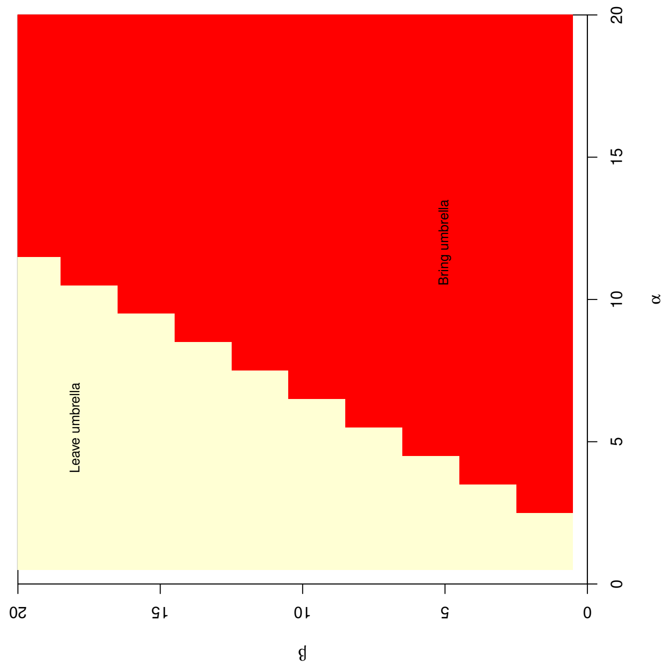


Fig. 15.1. How the Bayesian decision depends on the prior hyperparameters when $s = 2$ and $f = 8$

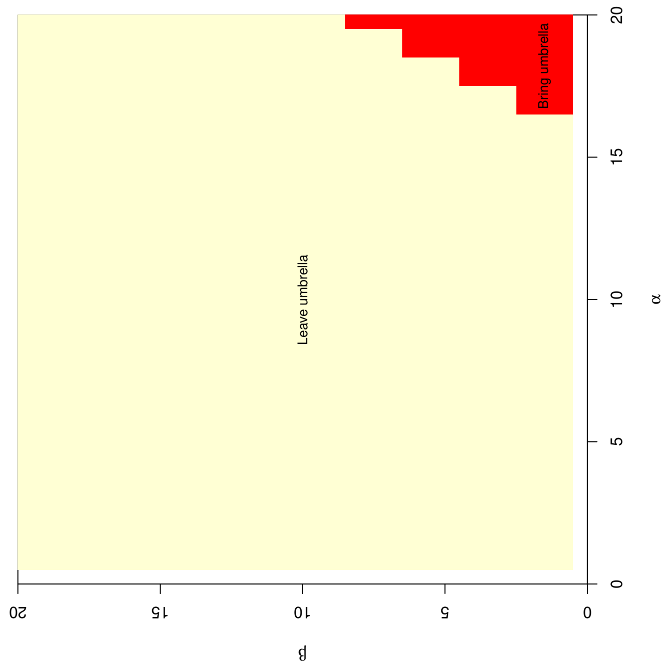


Fig. 15.2. How the Bayesian decision depends on the prior hyperparameters when $s = 16$ and $f = 64$

Solution 3a): The predictive distribution of x_6:

$$x_6|x_{1:5} \sim \mathcal{N}\left(\mu_n, \sigma^2 + \tau_n^2\right)$$

as shown in Lecture 4, slide 6. Here μ_n and ${\tau_n}^2$ are the posterior mean and variance of θ, which were derived in Lecture 2. So,

$$\tau_n^2 = \frac{1}{\frac{1}{\tau_0^2} + \frac{n}{\sigma^2}} = \frac{1}{\frac{1}{50^2} + \frac{5}{25^2}} = 119,$$

$$\mu_n = w\bar{x} + (1-w)\,\mu_0$$

$$\text{with } w = \frac{\frac{n}{\sigma^2}}{\frac{1}{\tau_0^2} + \frac{n}{\sigma^2}} = \frac{\frac{5}{25^2}}{\frac{1}{50^2} + \frac{5}{25^2}} = 0.95$$

$$\text{so } \mu_n = 0.95 \cdot 320.4 + 0.05 \cdot 200 = 315.$$

So,

$$x_6|x_{1:5} \sim \mathcal{N}\left(315, 25^2 + 119\right) = \mathcal{N}\left(315, 27.3^2\right)$$

Solution 3b): The expected utility when there is no campaign is

$$E\left[U\left((p-q)\,x_6\right)|x_{1:5}\right] = E\left[1 - \exp\left(-5x_6/1000\right)|x_{1:5}\right] = 1 - E\left[\exp\left(S_1\right)\right],$$

where S_1 is a normal random variable with mean $-5 \cdot 315/1000 = -1.575$ and standard deviation $5 \cdot 27.3/1000 = 0.1365$. So the expected utility is

$$1 - E\left[\exp\left(S_1\right)\right] = 1 - \exp\left(-1.575 + 0.1365^2/2\right) = 0.7911.$$

The expected utility when there is a campaign is

$$E\left[U\left(1.2\,(p-q)\,x_6 - 300\right)|x_{1:5}\right] = E\left[1 - \exp\left(-\left(1.2 \cdot 5x_6 - 300\right)/1000\right)|x_{1:5}\right] = 1 - E\left[\exp\left(S_2\right)\right],$$

where S_2 is a normal random variable with mean $-\left(1.2 \cdot 5 \cdot 315 - 300\right)/1000 = -1.59$ and standard deviation $1.2 \cdot 5 \cdot 27.3/1000 = 0.1638$. So the expected utility is

$$1 - E\left[\exp\left(S_2\right)\right] = 1 - \exp\left(-1.59 + 0.1638^2/2\right) = 0.7933.$$

Since the expected utility of running the campaign is higher, this is what the company should do.

Solution 4a): Let $y_i \overset{iid}{\sim} Poi\left(\theta\right)$ be the number of fatal accidents for year $i = 1, \ldots, 10$. Then,

$$p\left(y_i|\theta\right) = \frac{1}{y!}\theta^{y_i}\exp\left(-\theta\right).$$

Use a conjugate gamma prior for $\theta \sim Gamma\left(\alpha, \beta\right)$ so that $p\left(\theta\right) \propto \theta^{\alpha-1}\exp\left(-\beta\theta\right)$. Now

$$p\left(\theta|y\right) \propto p\left(\theta\right)\prod_{i=1}^{10} p\left(y_i|\theta\right) \propto \theta^{\alpha-1}\exp\left(-\beta\theta\right)\theta^{\sum_{i=1}^{10} y_i}\exp\left(-10\theta\right)$$

$$\propto \theta^{\alpha+10\bar{y}-1}\exp\left(-\left(\beta+10\right)\theta\right).$$

So, we get the posterior $\theta|y \sim Gamma\left(\alpha + 10\bar{y}, \beta + 10\right)$. We compute $\bar{y} = 23.8$ and a non-informative (non-proper) gamma prior is obtained from $\alpha = 0,\ \beta = 0$, so that $\theta|y \sim Gamma\left(238, 10\right)$. To get the 95% predictive bands we can use either normal approximation or simulation. For the normal approximation, we need to compute the predictive mean and variance for a new observation

$$\tilde{y} : E\left(\tilde{y}|y\right) = E\left[E\left(\tilde{y}|\theta, y\right)|y\right] = E\left(\theta|y\right) = \frac{238}{10} = 23.8,$$

$$\begin{aligned} Var\left(\tilde{y}|y\right) &= E\left[Var\left(\tilde{y}|\theta, y\right)|y\right] + Var\left[E\left(\tilde{y}|\theta, y\right)|y\right] \\ &= E\left[\theta|y\right] + Var\left[\theta|y\right] = \frac{238}{10} + \frac{238}{10^2} = 26.18, \end{aligned}$$

where we have used the formulas for means and variances of conditional distributions on page 21 in the coursebook, and that $E(X) = Var(X) = \theta$ for $X \sim Poi(\theta)$ and $E(X) = \alpha/\beta$ and $Var(X) = \alpha/\beta^2$ for $X \sim Gamma\left(\alpha, \beta\right)$. A normal approximation of the posterior is

$$\tilde{y}|y \sim N\left(E\left(\tilde{y}|y\right), Var\left(\tilde{y}|y\right)\right) = N\left(23.8, 26.18\right).$$

A 95% predictive interval is

$$23.8 \pm 1.96 \cdot \sqrt{26.18} \Rightarrow 13.77 < \tilde{y} < 33.83.$$

Alternatively, using simulation, a predictive interval can be computed by repeatedly simulating $\theta^{(j)}$ from $\theta|y$ and then $\tilde{y}^{(j)} \sim \tilde{y}|\theta^{(j)}$ for $j = 1, \ldots, 1000$ and extracting the 2.5th and 97.5th percentile from the samples. The following code does this in R:

```
theta = rgamma(1000,238,10)
y = rpois(1000,theta)
quantile(y,probs=c(0.025,0.975))
```