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?

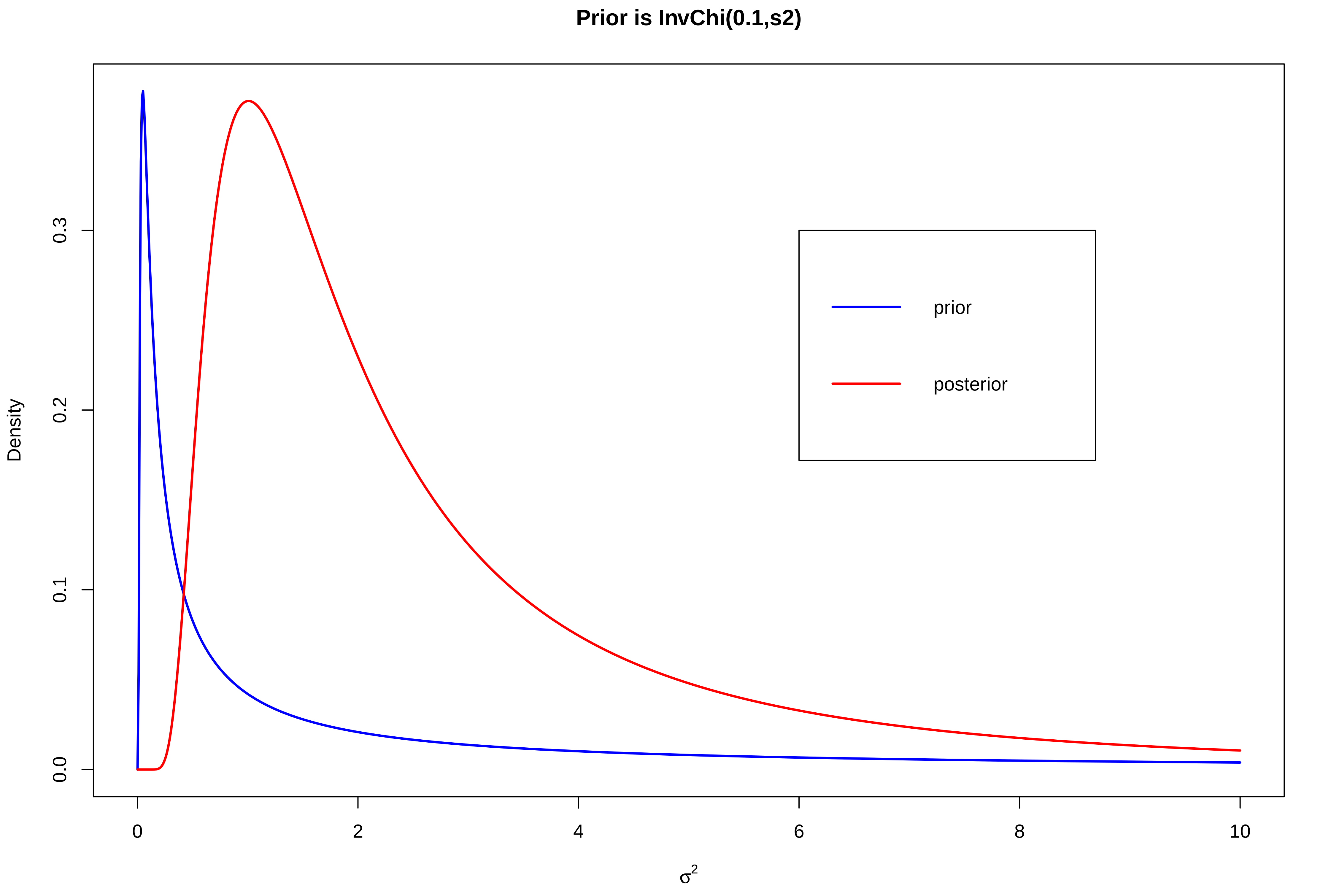
Reason 1: v_n becomes n

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,

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



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

Problem set No.2 - Problem 2 - Feel the Bern.

Problems of Chapter 6

6.1 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

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

7.1 Umbrella decision

(a) Let x_{11} be the binary variable indicating rain on the 11th day. From Problem 6.1, 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.
(b) Figure 15.1 shows how the optimal Bayesian decision varies for different combinations of the prior hyperparameters.
(c) Figure 15.2 shows how the optimal Bayesian decision varies for different combinations of the prior hyperparameters when $s = 16$ and $f = 64$.

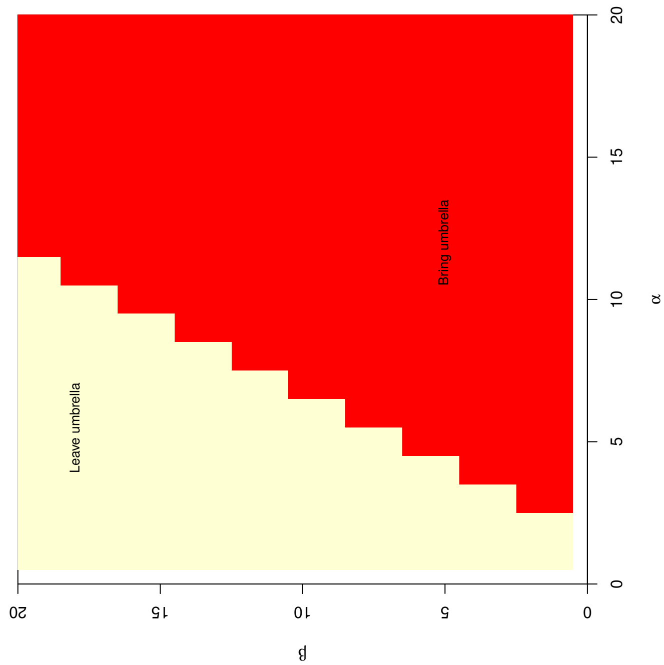


Fig. 15.1. How the Bayesian decision depends on the prior hyperparameters when $s = 2$ and $f = 8$

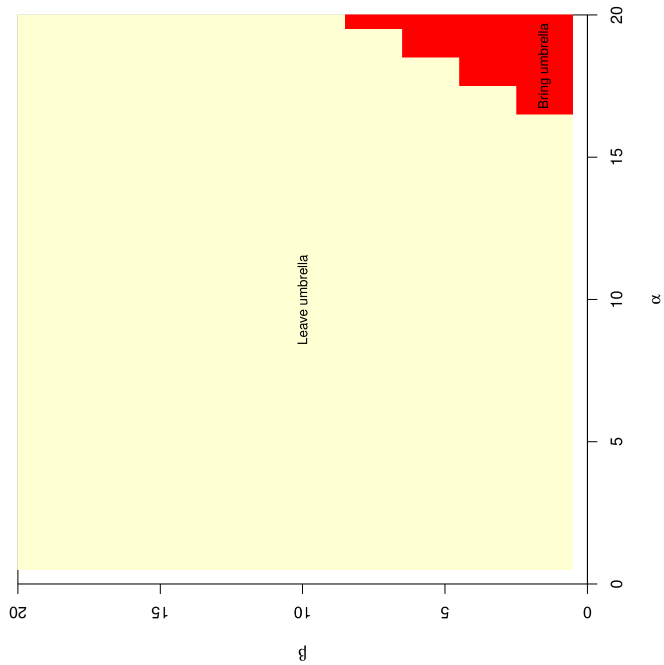


Fig. 15.2. How the Bayesian decision depends on the prior hyperparameters when $s = 16$ and $f = 64$