# Solution to Question 1

(a) See solution of problem 11 of the exercises, in the 'Exercises Solns.pdf' file. Combining the $n = 10$ observations with sample mean $\bar{x}$ with the prior mixture distribution, we obtain the following posterior mixture distributions and summaries:

| | Posterior mixture | Mean | SD | $Pr(\mu > 2.5)$ |
|---|---|---|---|---|
| No data | $0.2\,N(3.3, 0.37^2) + 0.8\,N(1.1, 0.47^2)$ | 1.540 | 0.989 | 0.198 |
| $\bar{x} = 2.0$ | $0.028\,N(2.549, 0.240^2) + 0.972\,N(1.720, 0.262^2)$ | 1.743 | 0.296 | 0.018 |
| $\bar{x} = 2.3$ | $0.249\,N(2.722, 0.240^2) + 0.751\,N(1.926, 0.262^2)$ | 2.125 | 0.430 | 0.216 |
| $\bar{x} = 2.4$ | $0.423\,N(2.780, 0.240^2) + 0.577\,N(1.995, 0.262^2)$ | 2.327 | 0.463 | 0.387 |
| $\bar{x} = 2.5$ | $0.615\,N(2.838, 0.240^2) + 0.385\,N(2.064, 0.262^2)$ | 2.540 | 0.452 | 0.584 |
| $\bar{x} = 2.8$ | $0.939\,N(3.011, 0.240^2) + 0.061\,N(2.270, 0.262^2)$ | 2.966 | 0.300 | 0.935 |

(b) Figure 1 gives a plot of the prior density and posterior densities for these values of $\bar{x}$, with left mode increasing in $\bar{x}$.

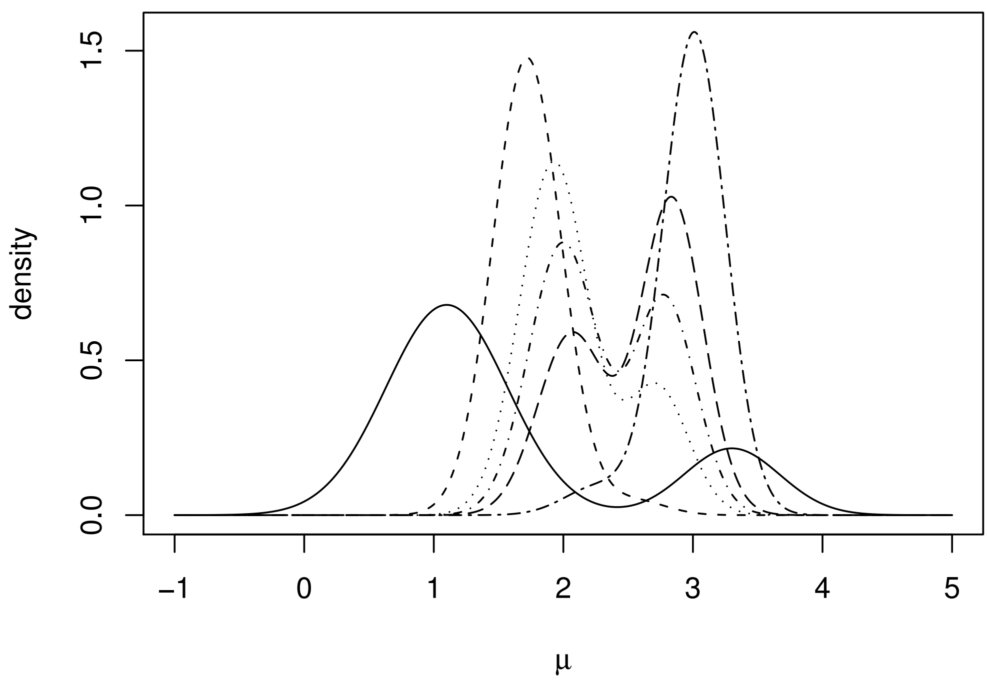


Figure 1: Prior density and posterior densities for $\bar{x} = 2.0, 2.3, 2.4, 2.5, 2.8$

(c) The prior density and most of the posterior densities are bimodal, except those with a very low weight for one of the components ($\bar{x} = 2.0$ and $\bar{x} = 2.8$). As $\bar{x}$ increases, the posterior mean and $Pr(\mu > 2.5|\boldsymbol{x})$ increase and the posterior standard deviation increases and then decreases (and is largest when $\bar{x} = 2.4$). The prior standard deviation is much larger than any of the posterior standard deviations. The prior probability $Pr(\mu > 2.5)$ lies between the posterior probabilities for $\bar{x} = 2.0$ and $\bar{x} = 2.3$.

(d) Let us define the first component distribution as the one with the larger mean. Note that with this prior this also corresponds to the component distribution with the smaller variance. Figure 2 shows the dependence of the posterior weight of the first component on the sample mean. Low values of $\bar{x}$ are more consistent with the component distribution with the smaller prior mean value (component 2). Also, as $\bar{x}$ increases, so does the (posterior) probability of the component with the larger prior mean value (component 1). However, and rather surprisingly, around $\bar{x} = 16.5$, the component with the smaller prior mean becomes more likely (again). Recall that

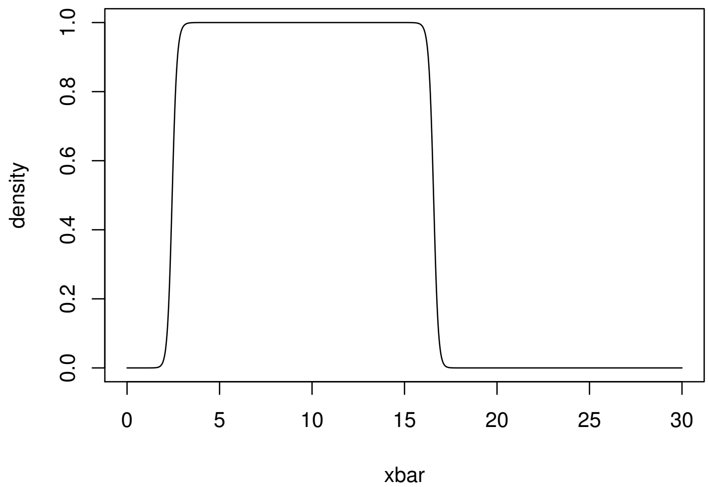


Figure 2: Plot of $p_1^*$ for various values of $\bar{x}$

$$(p_1^*)^{-1} - 1 = \frac{p_2\sqrt{D_1 d_2}}{p_1\sqrt{D_2 d_1}} \exp\left\{\frac{1}{2}\left[D_2 B_2^2 - d_2 b_2^2 - D_1 B_1^2 + d_1 b_1^2\right]\right\}.$$

The dependence of $p_1^*$ on $\bar{x}$ is only through the terms in $B_1$ and $B_2$. Now

$$\begin{aligned} g(\bar{x}) \equiv D_2 B_2^2 - D_1 B_1^2 &= \frac{(d_1 b_1 + n\tau\bar{x})^2}{d_1 + n\tau} - \frac{(d_2 b_2 + n\tau\bar{x})^2}{d_2 + n\tau} \\ &= n^2\tau^2\left(\frac{1}{d_2 + n\tau} - \frac{1}{d_1 + n\tau}\right)\bar{x}^2 \\ &\quad + 2n\tau\left(\frac{d_2 b_2}{d_2 + n\tau} - \frac{d_1 b_1}{d_1 + n\tau}\right)\bar{x} \\ &\quad + \frac{d_2^2 b_2^2}{d_2 + n\tau} - \frac{d_1^2 b_1^2}{d_1 + n\tau}. \end{aligned}$$

Without loss of generality, we will assume that $d_1 > d_2$, as is the case for the numbering of components we've used in these solutions – if not then renumber the components $1 \to 2$ and $2 \to 1$. The coefficient of $\bar{x}^2$ is positive, and therefore $\exists\, k$ such that $g(\bar{x})$ is increasing in $|\bar{x} - k|$, that is, $g(\bar{x})$ increases as $\bar{x}$ increases (for $\bar{x} > k$) and as $\bar{x}$ decreases (for $\bar{x} < k$). As $p_1^*$ is a decreasing function of $g(\bar{x})$, we have that $p_1^*$ is decreasing in $|\bar{x} - k|$. Thus, $\exists\, k_1 < k_2$ and $\epsilon > 0$ such that

$$p_1^* < \epsilon \quad \text{if } \bar{x} < k_1 \text{ or } \bar{x} > k_2 \quad (d_1 > d_2).$$

In other words, the posterior weight of the first component distribution will be (arbitrarily) small if $\bar{x}$ is sufficiently small or sufficiently large ($d_1 > d_2$) – as seen in Figure 2.

# Solution to Question 2

(a) A normal Q-Q plot of the data is given in Figure 3. The points follow the straight line fairly closely and so the normality assumption for the data is reasonably plausible.

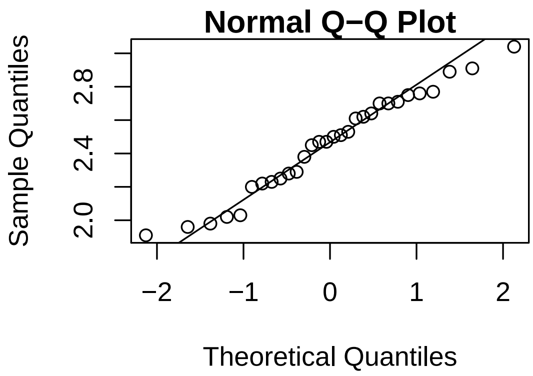


Figure 3: Normal Q-Q plot of the hepatitis data

(b) If $\begin{pmatrix} \mu \\ \tau \end{pmatrix} \sim NGa(b, c, g, h)$ then the univariate marginal distributions are $\mu \sim t_{2g}\{b, h/(gc)\}$, $\tau \sim Ga(g, h)$ and $\sigma \sim \text{Inv-Chi}(g, h)$. Therefore

$$E(\mu) = b = 2.6, \qquad SD(\mu) = \sqrt{\frac{2gh}{gc(2g-2)}} = \sqrt{\frac{h}{c(g-1)}} = 0.3162.$$

Also

$$E(\tau) = \frac{g}{h} = 12.5, \qquad SD(\tau) = \frac{\sqrt{g}}{h} = 5.5902$$

and

$$E(\sigma) = \frac{\sqrt{h}\,\Gamma\left(g - \frac{1}{2}\right)}{\Gamma(g)} = 0.3065$$

$$SD(\sigma) = \sqrt{\frac{h}{g-1} - E(\sigma)^2} = 0.0777.$$

(c) The posterior distribution is $(\mu, \tau)^T | \boldsymbol{x} \sim NGa(B, C, G, H)$ where

$$B = \frac{bc + n\bar{x}}{c+n} = 2.46387, \qquad C = c + n = 31,$$

$$G = g + \frac{n}{2} = 20, \qquad H = h + \frac{cn(\bar{x}-b)^2}{2(c+n)} + \frac{ns^2}{2} = 1.76697.$$

Therefore

$$E(\mu|\boldsymbol{x}) = B = 2.46387, \qquad SD(\mu|\boldsymbol{x}) = \sqrt{\frac{H}{C(G-1)}} = 0.05476.$$

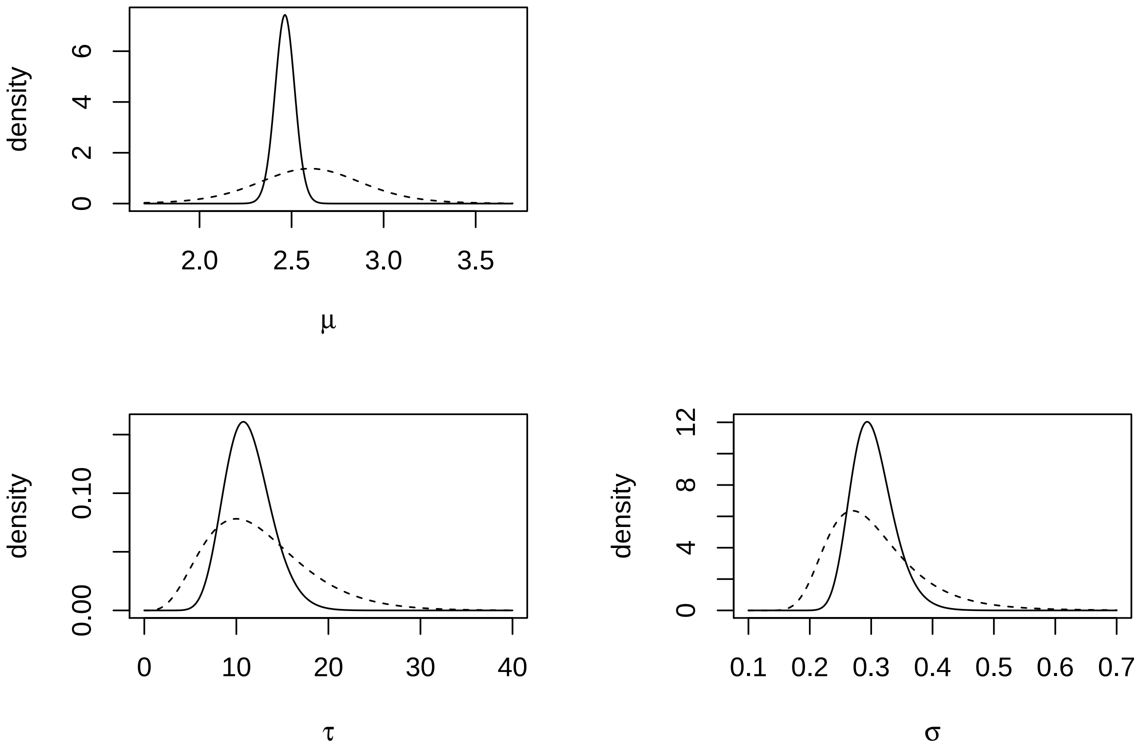


Figure 4: Prior (dashed) and posterior (solid) densities for $\mu$, $\tau$ and $\sigma$

Also

$$E(\tau|\boldsymbol{x}) = \frac{G}{H} = 11.3252, \qquad SD(\tau|\boldsymbol{x}) = \frac{\sqrt{G}}{H} = 2.5324$$

and

$$E(\sigma|\boldsymbol{x}) = \frac{\sqrt{H}\,\Gamma\left(G-\frac{1}{2}\right)}{\Gamma(G)} = 0.3029, \qquad SD(\sigma|\boldsymbol{x}) = \sqrt{\frac{H}{G-1} - E(\sigma|\boldsymbol{x})^2} = 0.0349.$$

(d) Plots of the (marginal) prior and posterior densities for $\mu$, $\tau$ and $\sigma$ are given in Figure 4. The contour plot of the (joint) prior and posterior densities for $(\mu, \tau)$ is shown in Figure 5.

(e) A plot of the 80%, 90% and 95% prior and posterior confidence regions for $(\mu, \tau)^T$ is given in Figure 6.

(f) Changes in beliefs:

$\mu$: its mean, mode and standard deviation have decreased

$\tau$: its mean has decreased (slightly); its mode has increased (slightly); its standard deviation has more than halved

$\sigma$: its mode has increased slightly; its standard deviation has decreased

$(\mu, \tau)$: the contours are less "triangular" and more "elliptical"

$(\mu, \tau)$: the confidence regions are much smaller

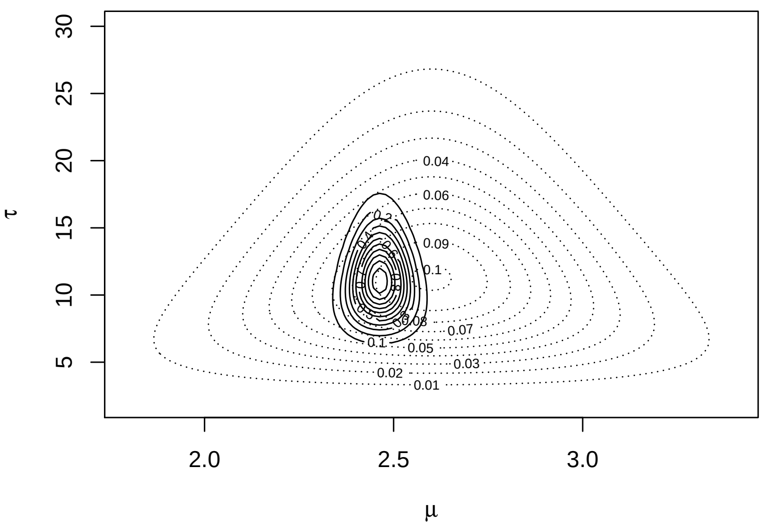


Figure 5: Contour plot of the prior (dashed) and posterior (solid) densities for $(\mu, \tau)^T$

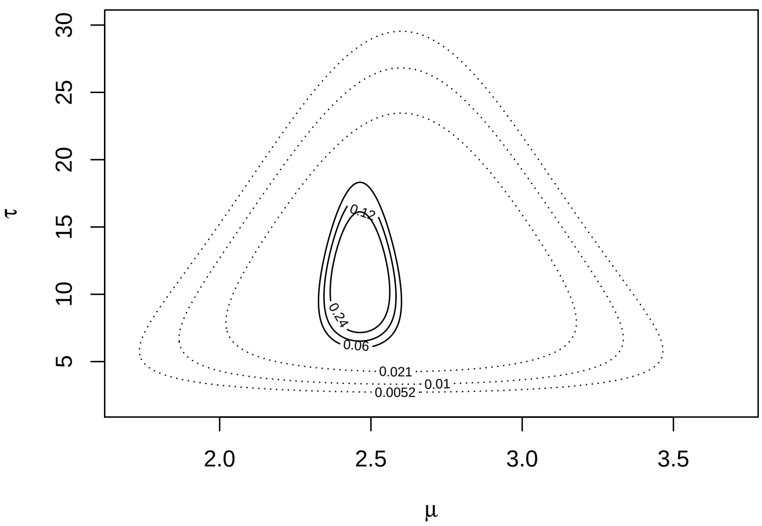


Figure 6: 95%, 90% and 80% prior (dashed) and posterior (solid) confidence regions for $(\mu, \tau)^T$

(g) The probabilities are $Pr(\mu > 2.7) = 0.36551$ and $Pr(\mu > 2.7|\boldsymbol{x}) = 0.000036$, calculated using `1-pgt(2.7,2*g,b,h/(g*c))` and `1-pgt(2.7,2*G,B,H/(G*C))`. There is a large difference between these probabilities and so the data have been very informative.

(h) The predictive distribution of $\bar{Y}$ can be found using similar arguments to those in section 2.4. We know that $\overline{Y}|\mu, \tau \sim N\{\mu, 1/(m\tau)\}$ and so

$$\overline{Y}|\boldsymbol{x}, \tau \sim N\left(B,\, \frac{1}{m\tau} + \frac{1}{C\tau}\right) \equiv N\left(B,\, \frac{C+m}{Cm\tau}\right),$$

from which we obtain

$$\overline{Y}|\boldsymbol{x} \sim t_{2G}\left(B,\, \frac{H(C+m)}{GCm}\right).$$

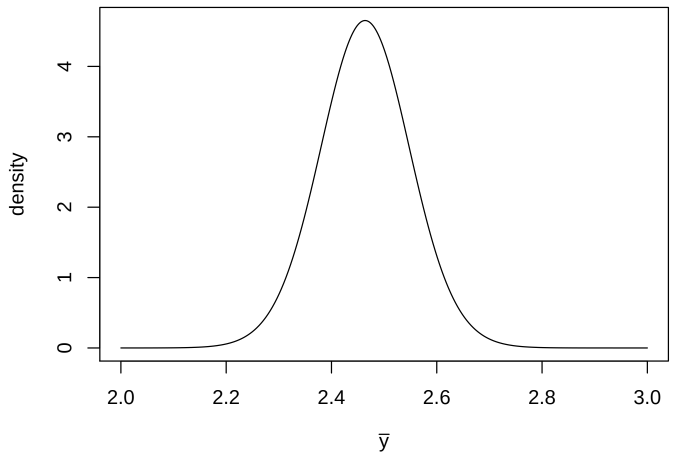


Figure 7: Predictive density of $\overline{Y}$, the mean of a future random sample of size $m = 20$

(i) When $m = 20$, we have $\overline{Y}|\boldsymbol{x} \sim t_{40}(2.46387, 0.00726)$. A plot of this predictive density is given in Figure 7.

The 95% prediction interval has limits `qgt(0.025,2*G,B,H*(C+m)/(G*C*m))` and `qgt(0.975,2*G,B,H*(C+m)/(G*C*m))`, and so the interval is $(2.29163, 2.63612)$.

(j) As $mV\tau|\tau \sim \chi^2_{m-1}$, we have that $V|\tau \sim Ga\left(\frac{m-1}{2}, \frac{m\tau}{2}\right)$. Therefore the predictive density for $V$ is, for $v > 0$

$$
\begin{aligned}
f(v|\boldsymbol{x}) &= \int f(v|\tau)\,\pi(\tau|\boldsymbol{x})\,d\tau \\
&= \int_0^\infty \frac{\left(\frac{m\tau}{2}\right)^{\frac{m-1}{2}} v^{\frac{m-1}{2}-1} e^{-m\tau v/2}}{\Gamma(\frac{m-1}{2})} \times \frac{H^G \tau^{G-1} e^{-H\tau}}{\Gamma(G)}\,d\tau \\
&= \frac{\left(\frac{m}{2}\right)^{\frac{m-1}{2}} H^G v^{\frac{m-1}{2}-1}}{\Gamma(\frac{m-1}{2})\Gamma(G)} \int_0^\infty \tau^{\frac{m-1}{2}+G-1} e^{-(mv/2+H)\tau}\,d\tau \\
&= \frac{\left(\frac{m}{2}\right)^{\frac{m-1}{2}} H^G v^{\frac{m-1}{2}-1}}{\Gamma(\frac{m-1}{2})\Gamma(G)} \times \frac{\Gamma(\frac{m-1}{2}+G)}{\left(\frac{mv}{2}+H\right)^{\frac{m-1}{2}+G}} \\
&= \frac{\Gamma(\frac{m-1}{2}+G)}{\Gamma(\frac{m-1}{2})\Gamma(G)} \times \frac{\left(\frac{m}{2H}\right)^{\frac{m-1}{2}} v^{\frac{m-1}{2}-1}}{\left(1+\frac{mv}{2H}\right)^{\frac{m-1}{2}+G}} \\
&= \frac{1}{B(\frac{m-1}{2}, G)} \left(\frac{m}{2H}\right)^{\frac{m-1}{2}} v^{\frac{m-1}{2}-1} \left(1+\frac{mv}{2H}\right)^{-(\frac{m-1}{2}+G)}.
\end{aligned}
$$

This is the density of a scaled–F distribution with $\nu_1 = m-1$, $\nu_2 = 2G$ and $a = \frac{(m-1)H}{mG}$. Therefore

$$V|\boldsymbol{x} \sim \frac{(m-1)H}{mG}\,F_{m-1,2G}.$$

(k) When $m = 20$, we have $V|\boldsymbol{x} \sim 0.08388\,F_{19,40}$ and so the 95% equi-tailed prediction interval has limits
`((m-1)*H/(m*G))*qf(0.025,m-1,2*G)` and `((m-1)*H/(m*G))*qf(0.975,m-1,2*G)`, and so the interval is $(0.03600, 0.17501)$.

A 95% confidence interval for $S = \sqrt{V}$, the standard deviation of this future sample, is

$$(\sqrt{0.03600}, \sqrt{0.17501}) = (0.1896, 0.4183).$$