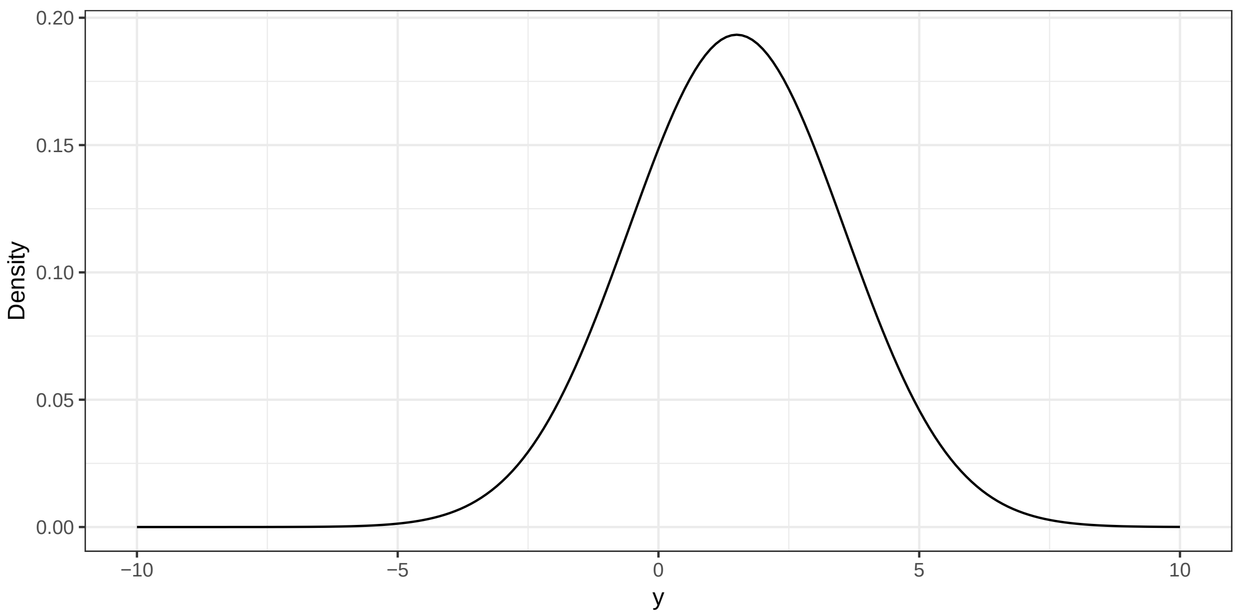
Bayesian Homework 1

Penn ID: ngraetz

February 13, 2019

Chapter 1, Question 1a

$$
\begin{aligned}
P(y) &= P(y|\theta = 1)P(\theta = 1) + P(y|\theta = 2)P(\theta = 2) && (1)\\
&= N(1, \sigma^2)(0.5) + N(2, \sigma^2)(0.5) && (2)\\
&= N(1, 4)(0.5) + N(2, 4)(0.5) && (3)
\end{aligned}
$$



Chapter 1, Question 1b

$$
\begin{aligned}
P(\theta = 1|y = 1) &= \frac{P(\theta = 1, y = 1)}{P(\theta = 1, y = 1)P(\theta = 2, y = 1)} && (4)\\
&= \frac{P(\theta = 1)P(y = 1|\theta = 1)}{P(\theta = 1)P(y = 1|\theta = 1) + P(\theta = 2)P(y = 1|\theta = 2)} && (5)\\
&= \frac{P(\theta = 1)P(y = 1|\theta = 1)}{P(\theta = 1)P(y = 1|\theta = 1) + P(\theta = 2)P(y = 1|\theta = 2)} && (6)\\
&= \frac{(0.5)N(1|1, 4)}{(0.5)N(1|2, 4) + (0.5)N(1|2, 4)} && (7)\\
&= 0.53 && (8)
\end{aligned}
$$

```
## [1] 0.53
```

Chapter 1, Question 1c

The posterior density for θ, $P(\theta|y) = P(y|\theta)P(\theta)$, approaches the prior $P(\theta)$ as $\sigma \to \infty$ (the variation in the data gets larger, i.e. the data provide no useful information). Conversely, as $\theta \to 0$, the posterior density for θ becomes completely concentrated at 1.

Chapter 1, Question 9

variable	value
total_patients	37.00
patients_waiting	9.00
total_wait_time	88.00
time_waiting_per_patient	2.38
time_waiting_per_waiting_patient	9.78
closing_time	432.00
sim	1.00

variable	lower	median	upper
total_patients	31.48	42.00	55.05
patients_waiting	0.00	5.00	14.52
total_wait_time	0.00	19.50	90.10
time_waiting_per_patient	0.00	0.49	2.13
time_waiting_per_waiting_patient	1.00	4.00	9.18
closing_time	420.00	427.50	440.52

The first table shows the results from a single simulation. The lower table summarizes all measures over 100 simulations. Upper and lower indicate the 95% predictive intervals for each measure based on 100 simulations.

Chapter 2, Question 5a

$$\begin{aligned}
P(y = k) &= \int_0^1 P(y = k|\theta)d\theta && (9)\\
&= \int_0^1 \binom{n}{k}\theta^k(1-\theta)^{n-k}d\theta && (10)\\
&= \binom{n}{k}\frac{\Gamma(k+1)\Gamma(n-k+1)}{\Gamma(n+2)} && (11)\\
&= \frac{1}{n+1} && (12)
\end{aligned}$$

Here I use the fact that the beta density has an intergral of 1 and simplify with $\Gamma(x) = (x-1)!$.

Chapter 2, Question 5b

To demonstrate that the posterior mean $\frac{\alpha+y}{\alpha+\beta+n}$ lies between $\frac{y}{n}$ and $\frac{\alpha}{\alpha+\beta}$, we can use a weight τ and show it is between 0 and 1 in the following:

$$\frac{\alpha+y}{\alpha+\beta+n} = \frac{y}{n} + \tau(\frac{\alpha}{\alpha+\beta} - \frac{y}{n}) \tag{13}$$

$$\frac{\alpha+y}{\alpha+\beta+n} - \frac{y}{n} = \tau(\frac{\alpha}{\alpha+\beta} - \frac{y}{n}) \tag{14}$$

$$\frac{n\alpha-\alpha y-\beta y}{(\alpha+\beta+n)n} = \tau\frac{n\alpha-\alpha y-\beta y}{(\alpha+\beta)n} \tag{15}$$

$$\tau = \frac{\alpha+\beta}{\alpha+\beta+n} \tag{16}$$

As n is a positive integer, we know that this is always between 0 and 1, indicating that the posterior mean is a weighted average of the data and the prior mean.

Chapter 2, Question 5c

Using a uniform distribution for θ ($\alpha = \beta = 1$), the prior variance is $\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} = \frac{1}{12}$. The posterior variance for $Beta(\alpha+y, \beta+n-y)$ is given below.

$$var(p(\theta|y)) = \frac{(\alpha+y)(\beta+n-y)}{(\alpha+\beta+n)^2(\alpha+\beta+n+1)} \tag{17}$$

$$var(p(\theta|y)) = \frac{(1+y)(1+n-y)}{(2+n)^2(3+n)} \tag{18}$$

$$= \frac{1+y}{2+n}\frac{1+n-y}{2+n}(\frac{1}{3+n} \tag{19}$$

Because the first two terms sum to 1, their product is at most $0.5 * 0.5 = 0.25$. Because $n \geq 1$, the last term must be less than or equal to $\frac{1}{4}$. So the maximum of the posterior variance is $\frac{1}{16}$, which is less than $\frac{1}{12}$.

Chapter 2, Question 5d

If $y = 1$, $n = 1$, and the prior distribution of θ is $Beta(\alpha = 1, \beta = 5)$, then the prior variance is $\frac{1*5}{(1+5)^2(1+5+1)} = 0.0198$. The posterior density is $Beta(y+1, n-y+1)$, so the posterior variance is:

$$var(p(\theta|y)) = \frac{(\alpha+y)(\beta+n-y)}{(\alpha+\beta+n)^2(\alpha+\beta+n+1)} \tag{20}$$

$$= 0.0255 \tag{21}$$

The posterior variance is greater than the prior variance.

Chapter 2, Question 8a

$$\theta|y \sim N(\frac{\frac{1}{40^2}180 + \frac{n}{20^2}150}{\frac{1}{40^2} + \frac{n}{20^2}}, \frac{1}{\frac{1}{40^2} + \frac{n}{20^2}}) \tag{22}$$

Chapter 2, Question 8b

$$\hat{y}|y \sim N(\frac{\frac{1}{40^2}180 + \frac{n}{20^2}150}{\frac{1}{40^2} + \frac{n}{20^2}}, \frac{1}{\frac{1}{40^2} + \frac{n}{20^2}} + 20^2) \tag{23}$$

Chapter 2, Question 8c

Plugging in above, the 95% posterior interval for $\theta|\bar{y} = 150, n = 10$ is:

$$150.7 \pm 1.96(6.25) = [138, 163] \tag{24}$$

The 95% posterior predictive interval for $\hat{y}|\bar{y} = 150, n = 10$ is:

$$150.7 \pm 1.96(20.95) = [110, 192] \tag{25}$$

Chapter 2, Question 8d

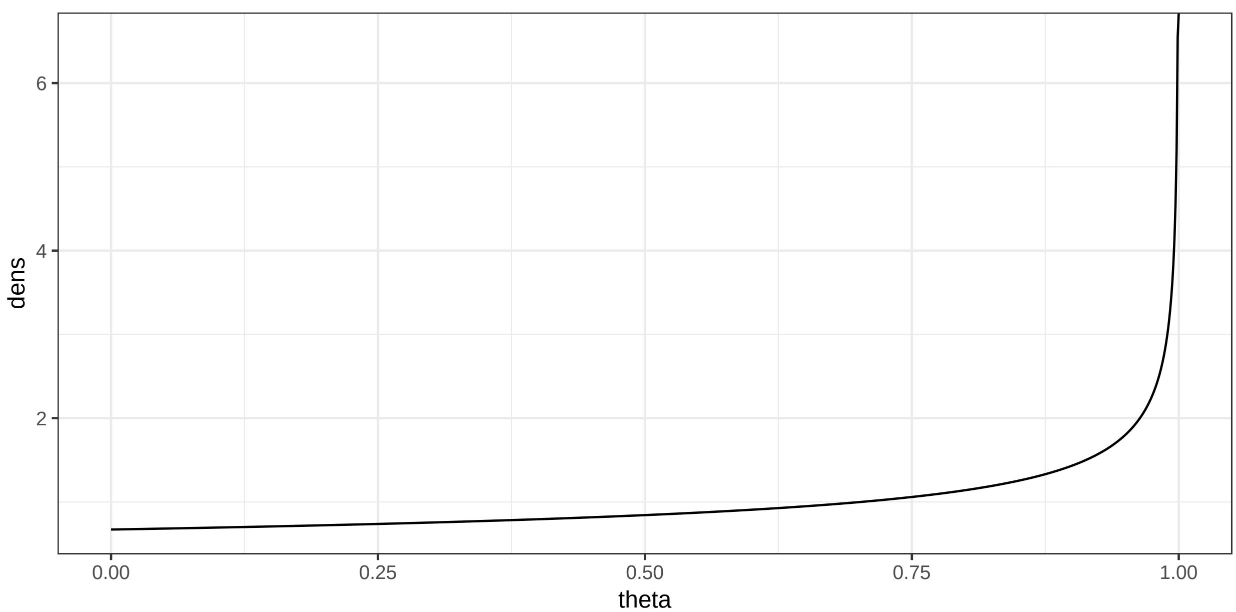
We can similarly plug in $n = 100$ to get an interval for $\theta|\bar{y} = 150, n = 100$, $[146, 154]$, and $\hat{y}|\bar{y} = 150, n = 100$, $[111, 189]$.

Chapter 2, Question 9a

$$\alpha + \beta = \frac{E(\theta)(1 - E(\theta))}{var(\theta)} - 1 = 1.67 \tag{26}$$

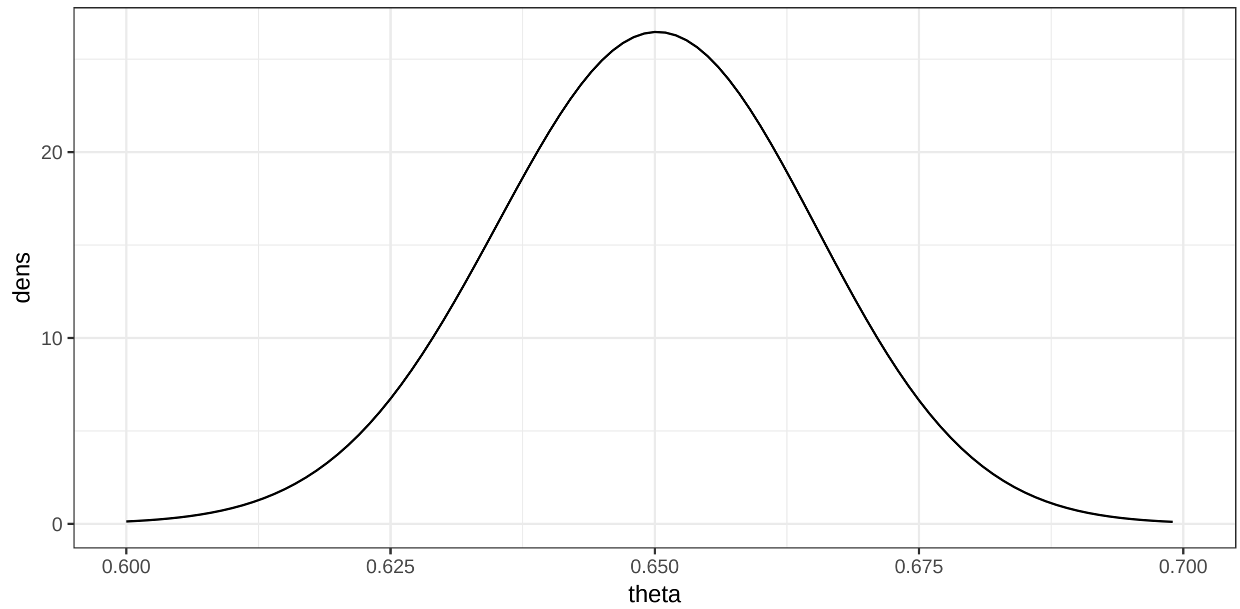
$$\alpha = (\alpha + \beta)(E(\theta)) = 1 \tag{27}$$

$$\beta = (\alpha + \beta)(1 - E(\theta)) = 0.67 \tag{28}$$



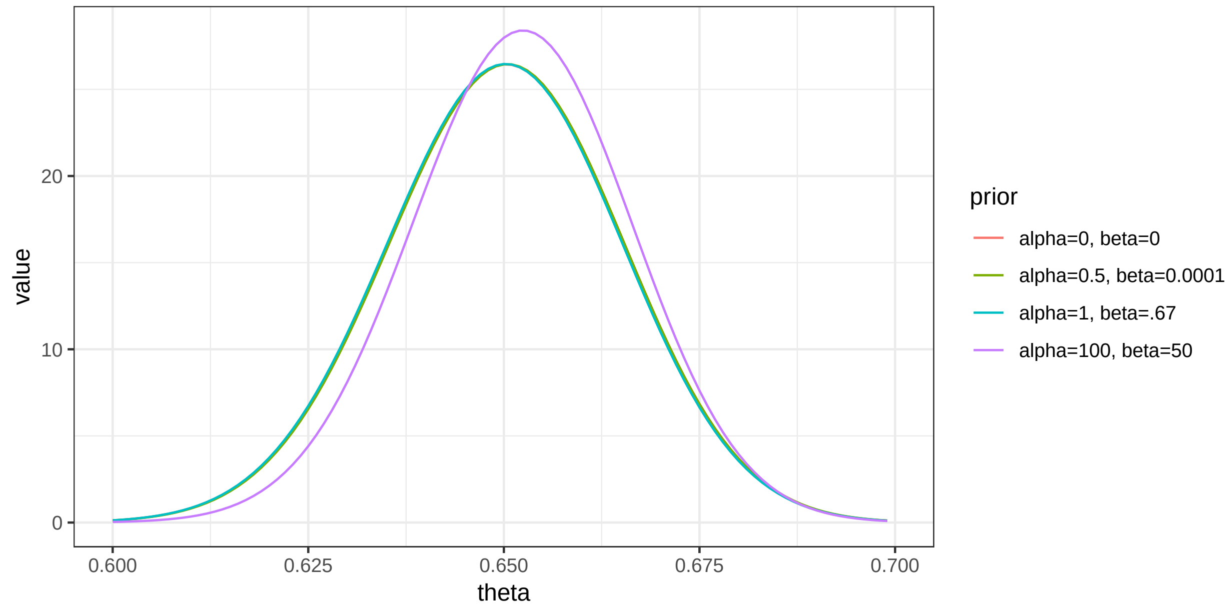
Chapter 2, Question 9b

If $n = 1000$ and $y = 650$, then the posterior is also Beta: $p(\theta|y) = Beta(\alpha + y, \beta + n - y) = Beta(\alpha + 650, \beta + 350) = Beta(651, 350.67)$. The mean and standard deviation of this posterior distribution are $E(\theta|y) = 0.65$ and $sd(\theta|y) = 0.015$. The data clearly wipe out any influence of the prior distribution of θ.

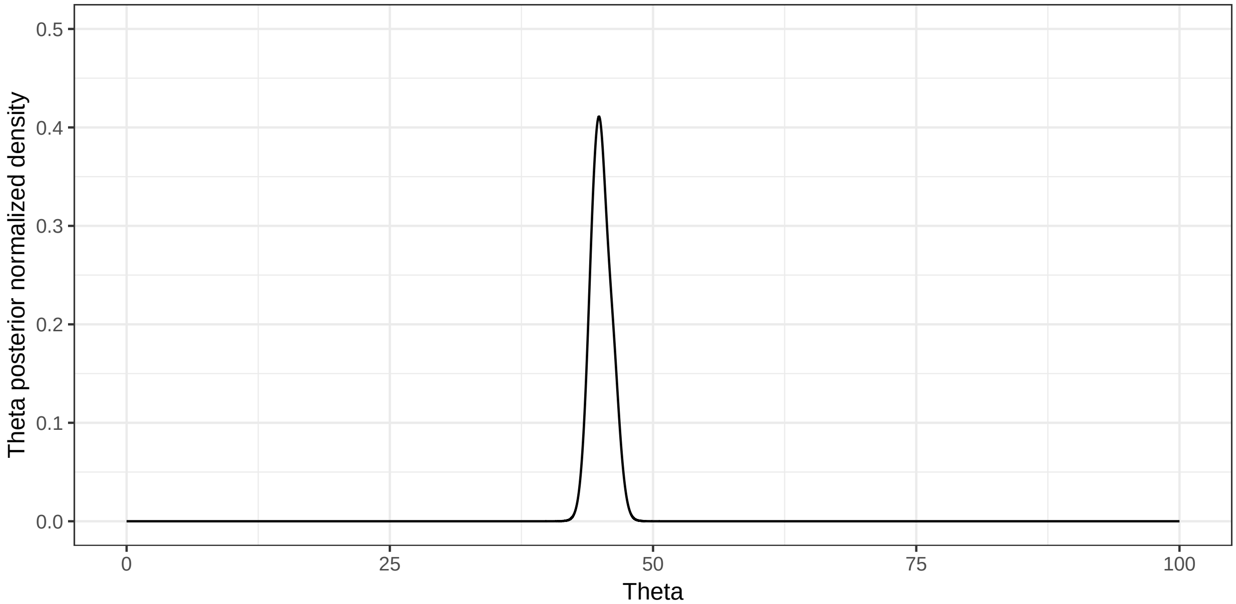


Chapter 2, Question 9c

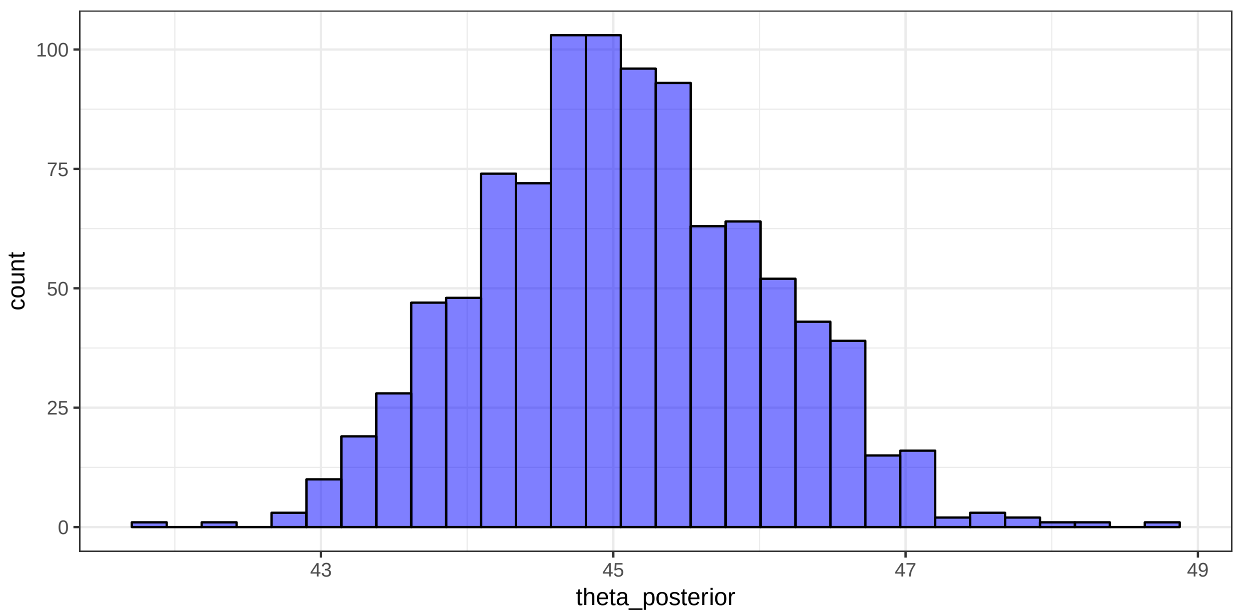
The posterior is very robust to changes in the prior because we have so much data - this is swamping out the prior.



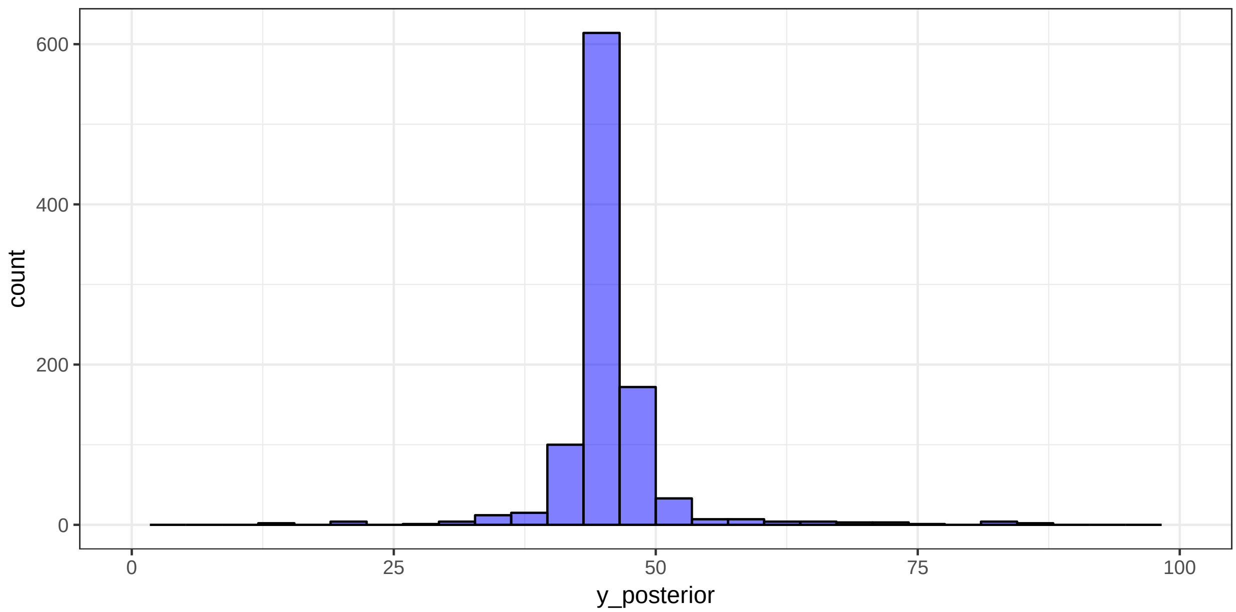
Chapter 2, Question 11a



Chapter 2, Question 11b



Chapter 2, Question 11c



Chapter 2, Question 13a

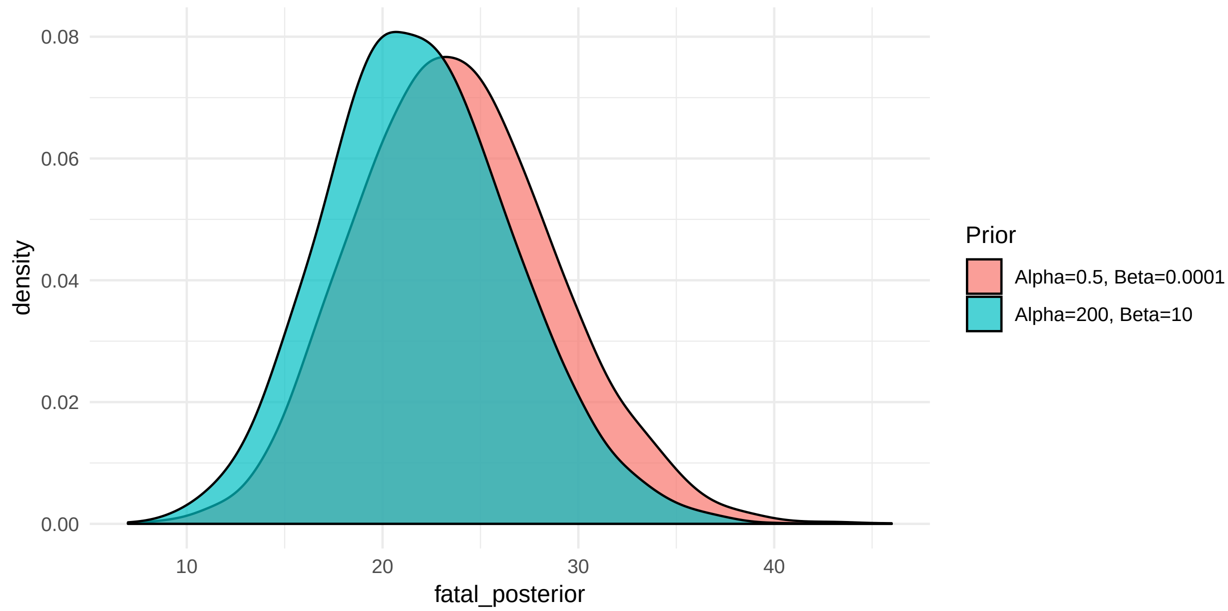
Assume the number of fatal accidents each year, y, are independent with a Poisson distribution. The model for the data is:

$$y_i|\theta\ Poisson(\theta) \tag{29}$$

Using the conjugate family of distribution, we can say the prior distribution for θ follows a Gamma distribution with hyperparameters α and β. This means that the posterior distribution for θ is:

$$P(\theta|\mathbf{y}) \sim Gamma(\sum_i y_i + \alpha, n + \beta) \tag{30}$$

I simulate 1000 draws from this posterior distribution. I then used each draw of my posterior distribution for θ to draw from a Poisson distribution to calculate my predictive posterior distribution for y^*. I tested the set of hyperpriors that defined the non-informative prior distribution for θ as well as a wider prior.



The 95% credible interval for y^* given the non-informative prior is 15 to 34 fatal accidents.

Chapter 2, Question 16a

$$p(y) = \int p(y|\theta)p(\theta)d\theta \tag{31}$$

We know that y is binomially distributed given unknown θ, and the prior for θ is $Beta(\alpha, \beta)$. We can integrate this over the domain of θ, 0 to 1, to calculate the marginal distribution of y (unconditional on θ).

$$= \int_0^1 \binom{n}{y} \theta^y (1-\theta)^{n-y} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha-1}(1-\theta)^{\beta-1} d\theta \tag{32}$$

Pull out terms not conditional on θ.

$$= \frac{\Gamma(n+1)}{\Gamma(y+1)\Gamma(n-y+1)} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\theta)} \int_0^1 \theta^{y+\alpha-1}(1-\theta)^{n-y+\beta-1} d\theta \tag{33}$$

$$= \frac{\Gamma(n+1)}{\Gamma(y+1)\Gamma(n-y+1)} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\theta)} \frac{\Gamma(y+\alpha)\Gamma(n-y+\beta}{\Gamma(n+\alpha+\beta)} \tag{34}$$

$p(y)$ is the beta-binomial density.

Chapter 2, Question 16b

Only looking at the terms in $p(y)$ that depend on y:

$$\frac{\Gamma(\alpha+y)\Gamma(\beta+n-y)}{\Gamma(y+1)\Gamma(n-y+1)} \tag{35}$$

If $\alpha = \beta = 1$, then this expression evaluates to 1.

$$\frac{\Gamma(1+y)\Gamma(1+n-y)}{\Gamma(y+1)\Gamma(n-y+1)} \tag{36}$$

Therefore if $\alpha = \beta = 1$, then $p(y)$ is constant across y.

Chapter 3, Question 3a

The data are distributed:

$$p(y|\mu_c, \mu_t, \sigma_c, \sigma_t) = \prod_{i=1}^{32} N(y_{c,i}|\mu_c, \sigma_c^2) * \prod_{i=1}^{36} N(y_{t,i}|\mu_t, \sigma_t^2) \tag{37}$$

So the posterior distribution is:

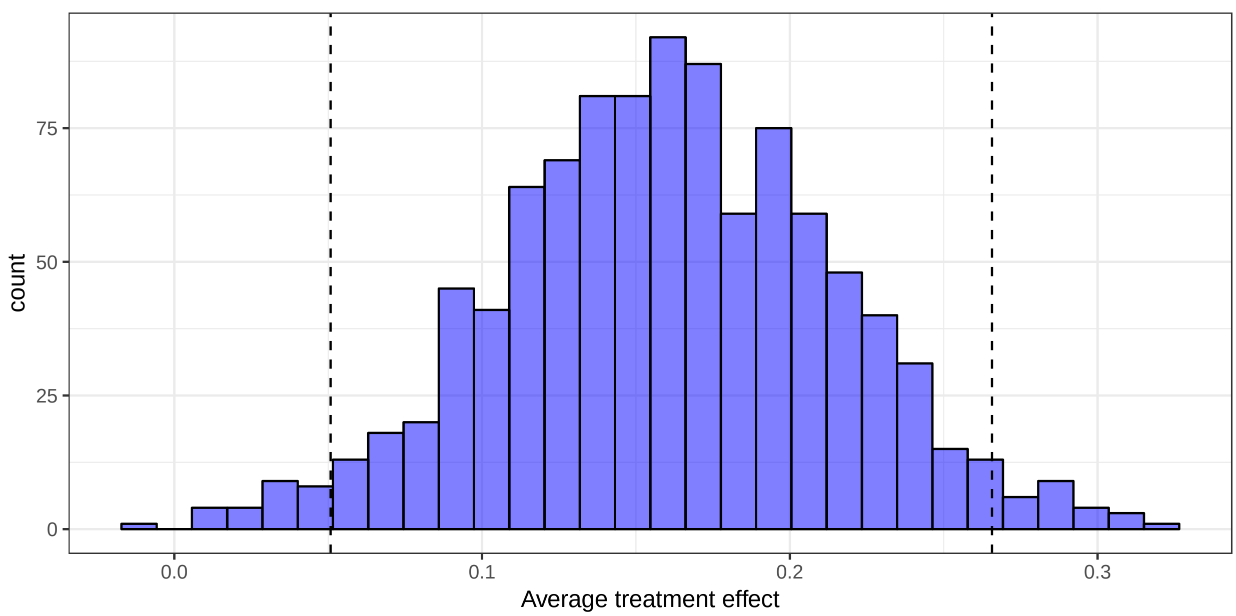
$$\begin{aligned} p(\mu_c, \mu_t, log(\sigma_c), log(\sigma_t)|y) &= p(\mu_c, \mu_t, log(\sigma_c), log(\sigma_t))p(y|\mu_c, \mu_t, log(\sigma_c), log(\sigma_t)) & (38)\\ &= p(\mu_c, \log\sigma_c|y)p(\mu_c, \log\sigma_c|y)) & (39) \end{aligned}$$

Considering (μ_c, σ_c) and (μ_t, σ_t) independently, we have the marignal posterior densities for μ_c and μ_t:

$$\mu_c|y \sim t_{31}(\mu_c, \frac{\sigma_c^2}{32}) \tag{40}$$

$$\mu_t|y \sim t_{35}(\mu_t, \frac{\sigma_t^2}{36}) \tag{41}$$

Chapter 3, Question 3b



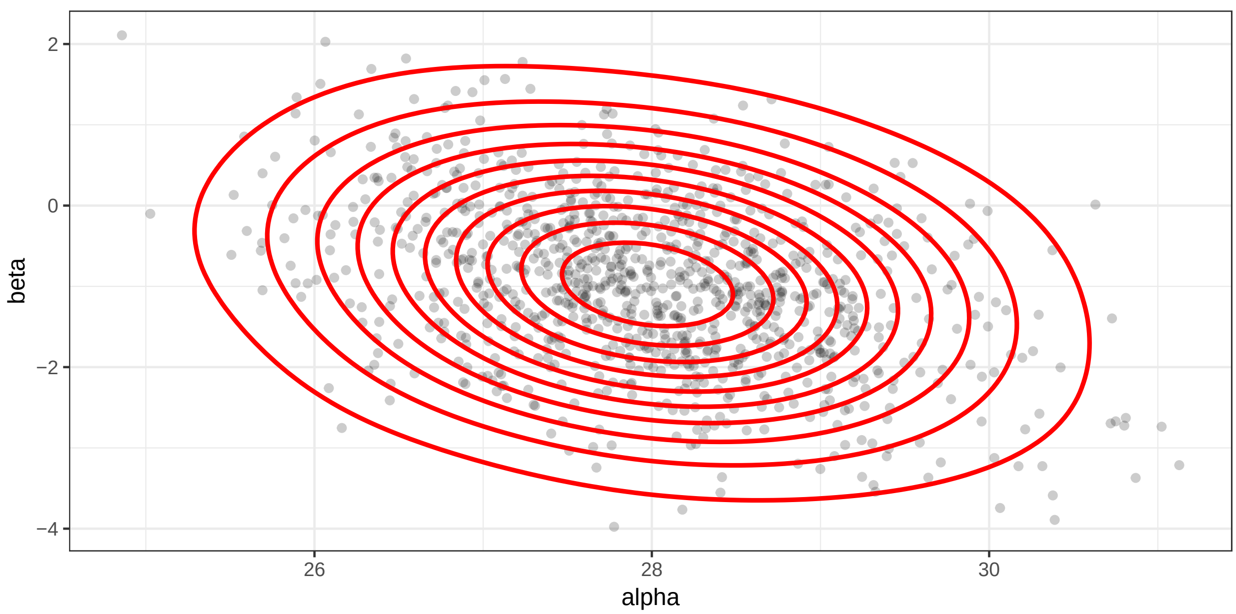
I draw from both posterior densities 1000 times to create 1000 draws of the posterior quantity $\mu_t - \mu_c$ (the treatment effect). The 95% posterior interval for the treatment effect is [0.05,0.27].

Chapter 3, Question 12a

We could use an independent uniform distribution for α and β: $p(\alpha, \beta) \propto 1$. This is an improper prior because the domain of α is $0 \to \infty$, which will not integrate to 1 so is not a proper probability density. But we will check that our posterior is proper.

Chapter 3, Question 12b

For an informative prior, we might want to assume that (α, β) follows a multivariate Normal distribution centered around -1 for α and 28 for β. There would be some negative correlation between the two terms, because in our linear model the slope will necessarily be lower if the intercept is higher.



Chapter 3, Question 12c

The posterior density for (α, β) is:

$$p(\alpha, \beta|\boldsymbol{y}) \propto \prod_{t=1}^{10} (\alpha + \beta t)^{y_t} e^{-(\alpha+\beta t)} \tag{42}$$

The sufficient statistics are the ordered pairs of fatal accidents and year (i.e. the entire dataset).

Chapter 3, Question 12d

The posterior density is a polynomial in α and β multiplied by the same exponential. Expanding this out would yield a summation where each term is a polynomial of α and β multiplied by the same exponential, which would all by Gamma densities. This would be integrable and thus a proper posterior density.

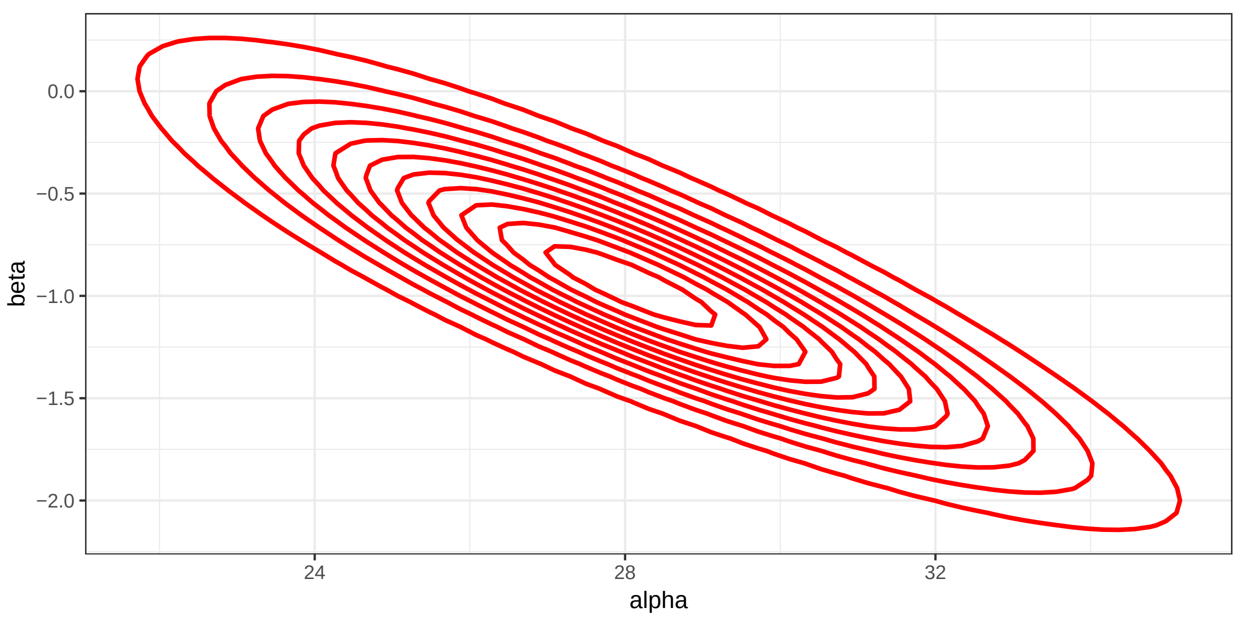
Chapter 3, Question 12e

```
##
```

```
## Call:
## lm(formula = fatal ~ t, data = d)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -4.576 -2.320 -1.761  3.273  5.818
##
## Coefficients:
##             Estimate Std. Error t value   Pr(>|t|)
## (Intercept)  27.9455     2.3656  11.813 0.00000242 ***
## t            -0.9212     0.4431  -2.079     0.0712 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.025 on 8 degrees of freedom
## Multiple R-squared:  0.3508, Adjusted R-squared:  0.2696
## F-statistic: 4.322 on 1 and 8 DF,  p-value: 0.07123
```

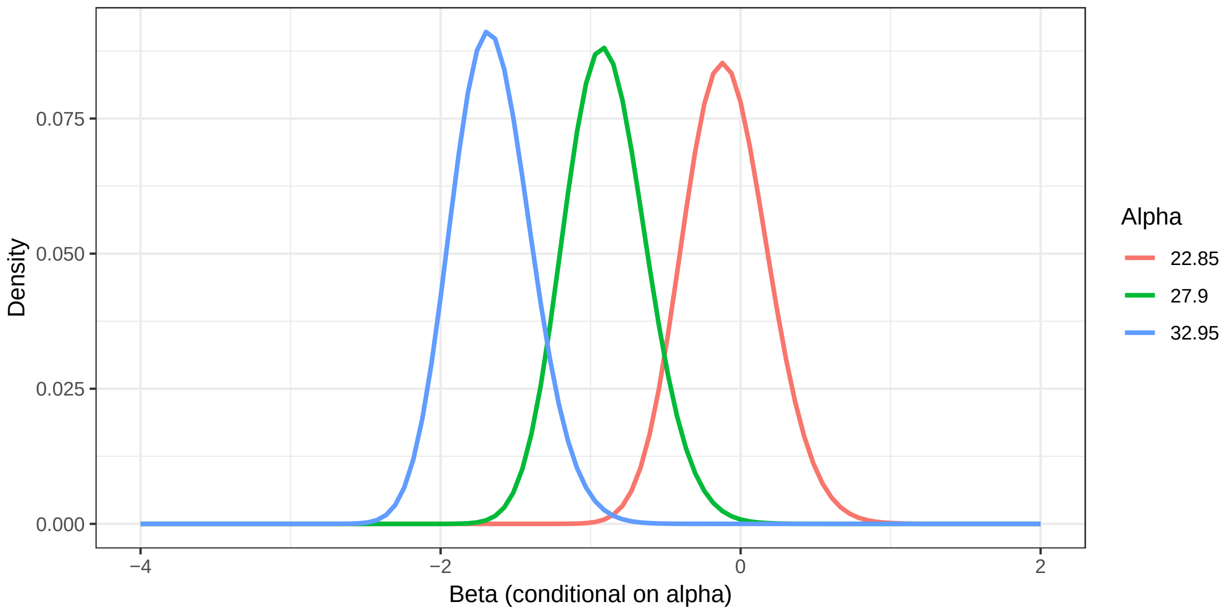
The estimates from our linear model are $\alpha = 27.95$ and $\beta = -0.92$. These crude estimates can provide good general starting locations for our grid sampling of the joint posterior density. I use 18 to 38 for α and -4 to 2 for β.

Chapter 3, Question 12f

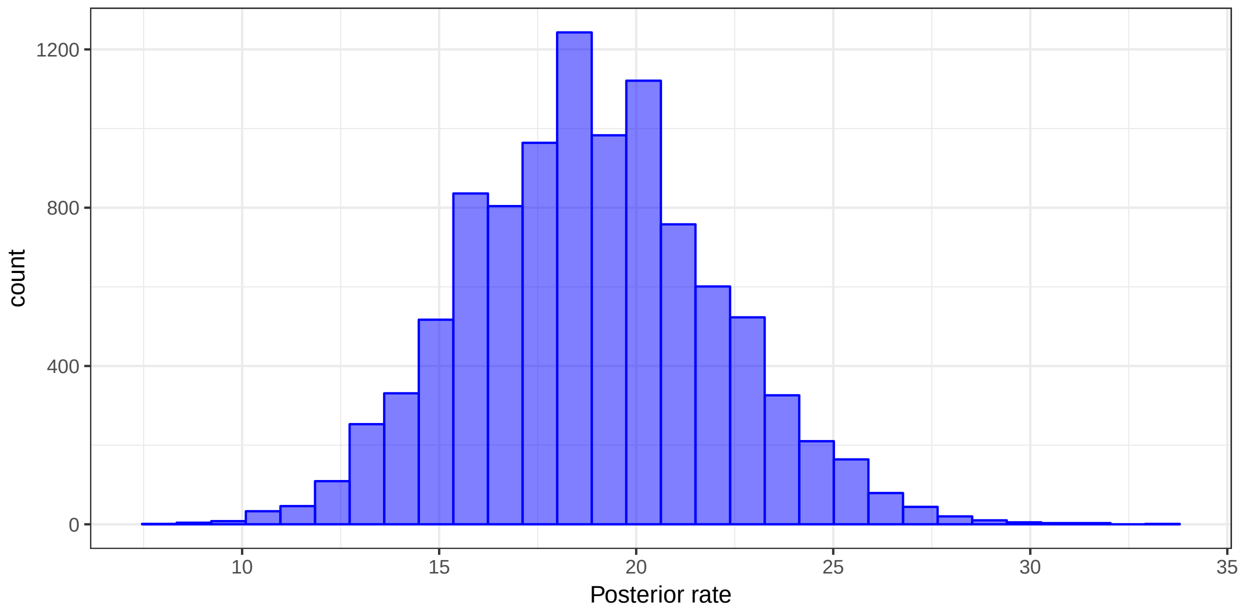


Chapter 3, Question 12g

I calculate the marginal density for α and the corresponding conditional density for β using the joint samples from my grid above just to compare how they co-vary.



I then calculate 1000 simulations of the expected number of fatal accidents in 1986 (θ in the data Poisson) by plugging in my 1000 draws of (α, β) to: $\alpha + 1986 * \beta$. This predicted posterior rate is plotted below.



Chapter 3, Question 12h

I take 1000 samples from a Poisson distribution using my 1000 draws of the posterior density for θ generated above.

The 95% predictive interval for the number of fatal accidents in 1986 is 9 to 30.

Chapter 3, Question 12i

My hypothetical informative prior is different than the posterior obtained under my non-informative prior. One reason is that my guess of the variance was quite far off, which does seem harder to provide robust prior information on than the mean.