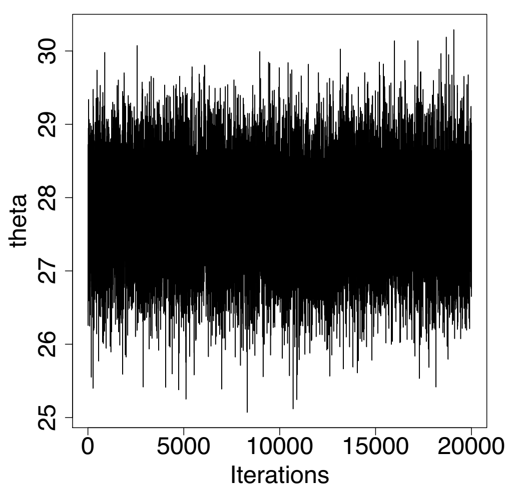
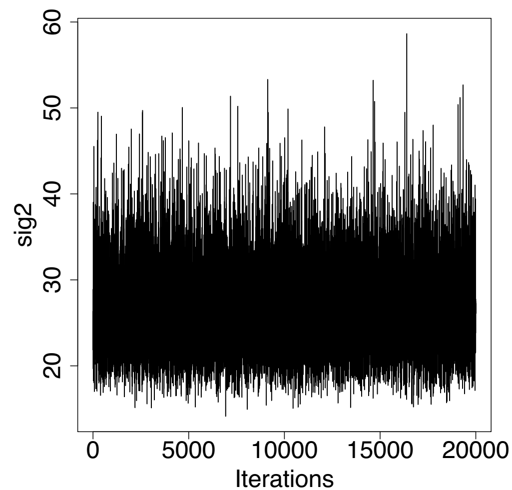
† Example: Estimating the speed of light (contd - redo)

- Remove the two outlying measurements and reanlyze the data with the same model.

- Check mixing and convergence of the Markov chain.



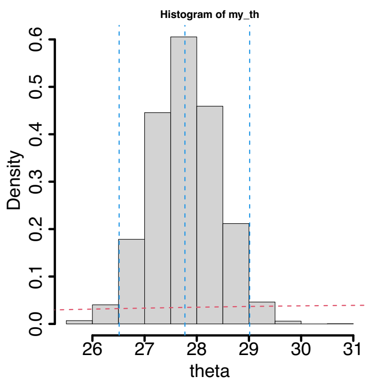
(a) θ



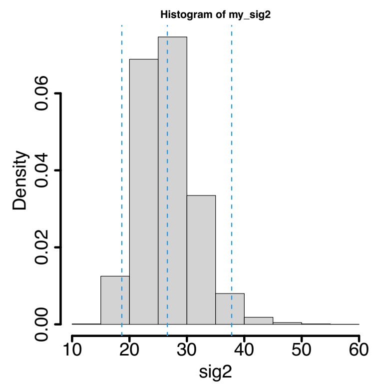
(b) σ^2

† Example: Estimating the speed of light (contd - redo)

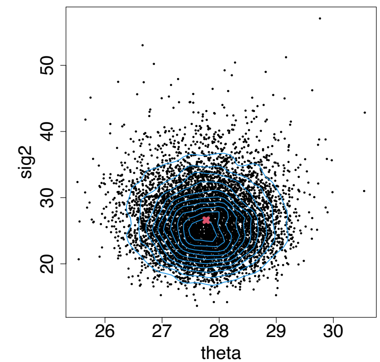
- Posterior summary of θ and σ^2



(a) θ



(b) σ^2



(c) Joint

† Example: Estimating the speed of light (contd - redo)

- Posterior summary of θ and σ^2

```
> ### summaries of the margianl posterior of theta
> post_m_th <- mean(my_th)
> post_sd_th <- sd(my_th)
> ci_th <- quantile(my_th, prob=c(0.025, 0.975))
> post_m_th
[1] 27.77308
> post_sd_th
[1] 0.6430802
> ci_th
    2.5%    97.5%
26.51241 29.01371
>
> ### summaries of the margianl posterior of sig2
> post_m_sig2 <- mean(my_sig2)
> post_sd_sig2 <- sd(my_sig2)
> ci_sig2 <- quantile(my_sig2, prob=c(0.025, 0.975))
> post_m_sig2
[1] 26.597
> post_sd_sig2
[1] 4.862329
> ci_sig2
    2.5%    97.5%
18.66827 37.80316
```

26.31

1.355

(23.67, 29.01)

119.01

21.49

84.55, 167.76

† Example: Estimating the speed of light (contd - redo)

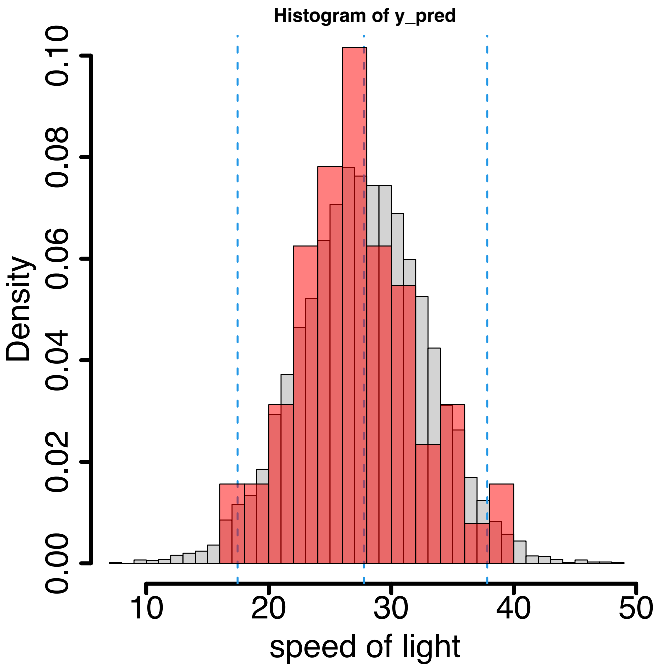
- Summary of the posterior predictive distribution of unobserved y

26.24

11.02

(4.34, 47.82)

```
> ###%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
> #%%%## preditive distribution
> y_pred <- rnorm(length(my_th), my_th, sqrt(my_sig2))
> mean(y_pred)
[1] 27.76714
> sd(y_pred)
[1] 5.173034
> quantile(y_pred, prob=c(0.025, 0.975))
    2.5%    97.5%
17.45357 37.83828
>
```



† **Hypothesis Testing** (CR Chapter 5 and JB Sec 4.3.3)

- Consider a statistical model $f(x \mid \theta)$ with $\theta \in \Theta$.

$\Theta = \{\theta_0, \theta_1\}$

- Specify

$\theta = \theta_0$ $\Theta_0 = \{\theta_0\}$

$\theta = \theta_1$ $\Theta_1 = \{\theta_1\}$

  ⋆⋆ Null hypothesis $H_0 : \theta \in \Theta_0$

  ⋆⋆ Alternative hypothesis $H_1 : \theta \in \Theta_1$

$\Theta_1 = \Theta_0^c$

  where $\Theta_0, \Theta_1 \subset \Theta$, $\theta \in \Theta_0 \cup \Theta_1 = \Theta$.

δ, d

- Consider a test procedure $\psi \in \mathcal{D} = \{0, 1\}$, where

  ⋆⋆ $\psi = 1$: conclude $H_0 : \theta \in \Theta_0$

  ⋆⋆ $\psi = 0$: conclude $H_1 : \theta \in \Theta_1$

- Consider the loss function proposed by Neyman-Pearson (a.k.a. the 0-1 loss function)

$$L(\theta, \psi) = \begin{cases} 1 & \text{if } \psi \neq \mathbb{I}_{\Theta_0}(\theta), \\ 0 & \text{otherwise.} \end{cases}$$

- Base the decision upon the posterior probability that the hypothesis is true.

- Recall: the *Bayesian decision* is

$$\psi^\pi(x) = \begin{cases} 1 & \text{if } \underline{P^\pi(\theta \in \Theta_0 \mid x)} > \underline{P^\pi(\theta \in \Theta_0^c \mid x)}, \\ 0 & \text{otherwise.} \end{cases}$$

⋆⋆ $P^\pi(\theta \in \Theta_0 \mid x) = P^\pi(H_0 \text{ is true} \mid x)$

$\Rightarrow$ we choose the hypothesis with the largest posterior probability.

$\pi(\theta_0) = 1 - \boxed{\pi(\theta_1)}$

† Testing Simple vs Simple

- Consider $H_0 : \theta = \theta_0$ vs $H_1 : \theta = \theta_1$, where θ_0 and θ_1 are preassigned values and $\Theta = \{\theta_0, \theta_1\}$.

$$
\begin{aligned}
P^\pi(H_0 \mid x) &= P^\pi(\theta = \theta_0 \mid x) \\
&= \frac{f(x \mid \theta_0)\pi(\theta = \theta_0)}{f(x \mid \theta_0)\pi(\theta = \theta_0) + f(x \mid \theta_1)\pi(\theta = \theta_1)}
\end{aligned}
$$

If $P^\pi(H_0 \mid x) > 0.5$, $\varphi = 1$

- Use the decision-theoretic approach as discussed in the previous slides.

- Consider the weighted 0-1 loss function

$$L(\theta, \psi) = \begin{cases} 0 & \text{if } \psi = \mathbb{I}_{\Theta_0}(\theta), \\ a_0 & \text{if } \underline{\theta} \in \underline{\Theta_0} \text{ and } \underline{\psi = 0}, \\ \underline{a_1} & \text{if } \underline{\theta} \notin \underline{\Theta_0} \text{ and } \underline{\psi = 1}. \end{cases}$$

$\rightarrow$ type I error

$\rightarrow$ type II error

★★ For a wrong answer under H_0, we lose by a_0.

★★ Suppose $\underline{a_0} > a_1$ (i.e., larger a_0/a_1) $\Rightarrow$ We lose more when we reject the true H_0 than when we reject the true H_1 (the more important a wrong answer under H_0 is relative to H_1).

★★ A large value of $\underline{a_0/a_1}$ guards against falsely rejecting H_0.

$\rho(\pi, \varphi \mid x)$, $\varphi \in \{0, 1\}$

$\rho(\pi, 1 \mid x) = \int_{\Theta} L(\theta, \varphi = 1)\ \pi(\theta \mid x)\, d\theta$

$\theta \in \Theta_0$ is correct

$= \int_{\Theta_1} a_1\ \pi(\theta \mid x)\, d\theta$

$= a_1 \cdot p^{\pi}(\theta \in \Theta_1 \mid x)$

$\rho(\pi, 0 \mid x) = a_0 \cdot p^{\pi}(\theta \in \Theta_0 \mid x)$

we let $\varphi = 1$ if $a_1\ p^{\pi}(\theta \in \underset{\Theta_0^c}{\underset{\shortparallel}{\Theta_1}} \mid x) < a_0\ p^{\pi}(\theta \in \Theta_0 \mid x)$

$a_1 \left(1 - p^{\pi}(\theta \in \Theta_0 \mid x)\right) < a_0\ p^{\pi}(\theta \in \Theta_0 \mid x)$

let $\varphi = 1$ if $\dfrac{a_1}{a_0 + a_1} < p^{\pi}(\theta \in \Theta_0 \mid x)$

$\dfrac{1}{\underbrace{a_0/a_1} + 1}$

- **Prop 5.2.2** The Bayesian estimator associated with a prior distribution π is

$$\psi^\pi(x) = \begin{cases} 1 & \text{if } P^\pi(\theta \in \Theta_0 \mid x) > \frac{a_1}{a_0+a_1}, \\ 0 & \text{otherwise.} \end{cases}$$

★★ Reject H_0 if the posterior probability of H_0 is too small.

★★ *How small?* smaller than $a_1/(a_0 + a_1)$

★★ The acceptance level $a_1/(a_0+a_1)$ is determined by the choice of the loss function.

★★ Note that ψ^π only depends on a_0/a_1 (rather than their actual values).

$\theta \in \mathbb{R} = \Theta$

- **Example 5.2.4** Consider $x \sim \mathrm{N}(\theta, \sigma^2)$ and $\theta \sim \mathrm{N}(\mu, \tau^2)$. Test $H_0 : \underbrace{\theta < 0}_{\Theta_0 = (-\infty, 0)}$ under the $a_0 - a_1$ loss.

$H_1 : \theta \geq 0 \iff \Theta_1 = \Theta_0^c = [0, \infty)$

$$p^\pi(H_0 \mid x) = \underbrace{p^\pi(\theta < 0 \mid x)}_{} = ?$$

Find $\pi(\theta \mid x) = N\Big(\underbrace{\left(\frac{1}{\tau^2} + \frac{1}{\sigma^2}\right)^{-1}\left(\frac{x}{\sigma^2} + \frac{\mu}{\tau^2}\right)}_{=\mu_1}, \underbrace{\left(\frac{1}{\tau^2} + \frac{1}{\sigma^2}\right)^{-1}}_{=\tau_1^2} \Big)$

$$= \Phi\left(\frac{-\mu_1}{\sqrt{\tau_1^2}}\right)$$

if $\Phi\left(\frac{-\mu_1}{\sqrt{\tau_1^2}}\right) > \frac{a_1}{a_0 + a_1}$, let $\varphi = 1$ (accept H_0)

otherwise, let $\varphi = 0$.

† Bayes Factor – another way to do a testing.

- **Def 5.2.5** The Bayes factor is the ratio of the posterior probabilities of the null and the alternative hypotheses over the ratio of the prior probabilities of the null and the alternative hypotheses , i.e.,

$$B_{01}^{\pi} = \frac{\text{posterior odds}}{\text{prior odds}} = \frac{P(\theta \in \Theta_0 \mid x)}{P(\theta \in \Theta_1 \mid x)} \Big/ \frac{\pi(\theta \in \Theta_0)}{\pi(\theta \in \Theta_1)}$$

$$\Rightarrow \underbrace{\frac{P(\theta \in \Theta_0 \mid x)}{P(\theta \in \Theta_1 \mid x)}}_{\text{posterior odds}} = \underbrace{B_{01}^{\pi}}_{\text{Bayes Factor}} \underbrace{\frac{\pi(\theta \in \Theta_0)}{\pi(\theta \in \Theta_1)}}_{\text{prior odds}}$$

⋆⋆ The Bayes factor evaluates the modification of the odds of Θ_0 against Θ_1 due to data (naturally compared to 1)

⋆⋆ The Bayes factor can be interpreted as how much the data favors $H_0 : \theta \in \Theta_0$ over $H_1 : \theta \in \Theta_1$.

• **JB Example 1 - p147** Recall the IQ test problem. The child taking the IQ test is to be classified as having below average IQ (less than 100) or above average IQ (greater than 100). Formally, test $H_0 : \theta \geq 100$ versus $H_1 : \theta < 100$. Recall $\theta \sim N(100, 225)$ and $\theta \mid x \sim N(110.39, 69.23)$. We find

$$P(\theta \geq 100) = 0.5, \qquad P(\theta < 100) = 0.5$$
$$\rightarrow P(\theta \geq 100 \mid x) = 0.894, \qquad P(\theta < 100 \mid x) = 0.106.$$

⋆⋆ In prior, H_0 and H_1 are viewed as equally plausible (the prior odds is 1).

⋆⋆ The Bayes factor $B_{01}^{\pi}(x) = (0.894/0.106)/(0.5/0.5) = 8.44$. That is, the odds of Θ_0 against Θ_1 increased by 8.44 times after observing data.

⋆⋆ In other words, the data is in favor of Θ_0.

† Bayes Factor for Simple Hypotheses

- Consider testing a simple null hypothesis against a simple alternative hypothesis; $\Theta_0 = \{\theta_0\}$ and $\Theta_1 = \{\theta_1\}$ where $\Theta = \{\theta_0, \theta_1\}$.

  $\pi(\theta) = \rho_0 \mathbb{1}(\theta = \theta_0) + (1-\rho_0)\mathbb{1}(\theta = \theta_1)$

- $\rho_0 = \pi(\theta \in \Theta_0)$ and $\pi(\theta \in \Theta_1) = 1 - \rho_0 = \rho_1$.

- Find
$$B_{01}^{\pi}(x) = \frac{f(x \mid \theta_0)}{f(x \mid \theta_1)}.$$

  $$\frac{P(\Theta_0 \mid x)}{P(\Theta_1 \mid x)} \Big/ \frac{Pr(\Theta_0)}{Pr(\Theta_1)} = \frac{\cancel{P(\theta=\theta_0)} \cdot f(x \mid \theta_0)}{\cancel{P(\theta=\theta_1)}\, f(x \mid \theta_1)} \Big/ \frac{\cancel{P(\theta=\theta_0)}}{\cancel{P(\theta=\theta_1)}} = \frac{f(x \mid \theta_0)}{f(x \mid \theta_1)}$$

  ⋆⋆ The Bayes factor does not depend on prior.

  ⋆⋆ A testing procedure solely based on the Bayes factor becomes the classical *likelihood ratio*.

$y = \beta_0 + \beta_1 x_1 + \varepsilon$

† Bayes Factor for a General Case

- $\rho_0 = \pi(\theta \in \Theta_0)$ and $\pi(\theta \in \Theta_1) = 1 - \rho_0 = \rho_1$. And let

$\int_{\Theta_0} \pi_0(\theta)\, d\theta = 1$

$$\theta \sim \begin{cases} \pi_0(\theta) & \text{if } \theta \in \Theta_0, \\ \pi_1(\theta) & \text{if } \theta \in \Theta_1. \end{cases}$$

⇒ Observe that $\pi(\theta) = \rho_0 \pi_0(\theta) + (1 - \rho_0)\pi_1(\theta)$.

$m(\theta) = \int_{\Theta} f(x|\theta)\, \pi(\theta)\, d\theta$

$= \rho \cdot \underbrace{\int_{\Theta_0} f(x|\theta)\, \pi_0(\theta)\, d\theta}_{= m_0(x)} + (1-\rho) \underbrace{\int_{\Theta_1} f(x|\theta)\, \pi_1(\theta)\, d\theta}_{= m_1(x)}$

$= \rho \cdot m_0(x) + (1-\rho)\, m_1(x)$

- We find

$$B_{01}^{\pi}(x) = \frac{\int_{\Theta_0} f(x \mid \theta)\pi_0(\theta)d\theta}{\int_{\Theta_1} f(x \mid \theta)\pi_1(\theta)d\theta} = \frac{m_0(x)}{m_1(x)}.$$

⋆⋆ Note that we have the marginals instead of the likelihoods.

⋆⋆ Observe $B_{01}^{\pi}(x)$ depends on *both* prior and data.

⋆⋆ When $H_0 : \theta = \theta_0$, observe $B_{01}^{\pi}(x) = \frac{f(x|\theta_0)}{m_1(x)}$.

$$B_{01}^{\pi}(x) = \frac{\text{posterior odds}}{\text{prior odds}}$$

$$P^{\pi}(\theta \in \Theta_0 \mid x)$$

$$\text{posterior odds} = \frac{\boxed{\dfrac{\int_{\Theta_0} p_0 \, f(x|\theta)\pi_0(\theta)\, d\theta}{m(\cancel{x})}}}{\dfrac{\int_{\Theta_1} (1-p_0)\, f(x|\theta)\pi_1(\theta)\, d\theta}{m(\cancel{x})}} = \frac{p_0 \cdot m_0(x)}{(1-p_0)\, m_1(x)}$$

$$= P^{\pi}(\theta \in \Theta_1 \mid x)$$

$$= \frac{p_0 \cdot m_0(x)}{(1-p_0)\, m_1(x)}$$

$$= \frac{\cancel{p_0}\, m_0(x)}{\cancel{(1-p_0)}\, m_1(x)} \Bigg/ \frac{\cancel{p_0}}{\underset{=(1-p_0)}{\cancel{p_1}}} = \frac{m_0(x)}{m_1(x)}$$

† How to connect the Bayes Factor to the decision theoretic testing procedure?

- Recall

$$\psi^\pi(x) = \begin{cases} 1 & \text{if } P^\pi(\theta \in \Theta_0 \mid x) > \frac{a_1}{a_0 + a_1}, \\ 0 & \text{otherwise.} \end{cases}$$

- This is equivalent to accepting H_0 when

$$\underline{B_{01}^\pi(x)} > \frac{a_1}{a_0} \Big/ \frac{\rho_0}{\rho_1} = \frac{a_1 \rho_1}{\underline{a_0 \rho_0}},$$

where $\rho_0 = \pi(\theta \in \Theta_0)$ and $\rho_1 = \pi(\theta \in \Theta_1)$.

$$B_{01}^{\pi} = \frac{P(\theta \in \Theta_0 \mid x)}{P(\theta \in \Theta_1 \mid x)} \Big/ \frac{P(\theta \in \Theta_0)}{\underbrace{P(\theta \in \Theta_1)}_{=(1-p_0)=p_1}}$$

$$P(\theta \in \Theta_0) = p_0$$

$$= \frac{p_1 \cdot P(\theta \in \Theta_0 \mid x)}{p_0 \cdot P(\theta \in \Theta_1 \mid x)}$$

$$= \frac{p_1 \cdot P(\theta \in \Theta_0 \mid x)}{p_0 \left(1 - P(\theta \in \Theta_0 \mid x)\right)}$$

$$\Rightarrow \quad B_{01}^{\pi} \, p_0 \left\{1 - P(\theta \in \Theta_0 \mid x)\right\} = p_1 \, P(\theta \in \Theta_0 \mid x)$$

$$\varphi = 1 \text{ if } \underline{P(\theta \in \Theta_0 \mid x)} = \underline{\frac{B_{01}^{\pi} p_0}{p_1 + B_{01}^{\pi} p_0} > \frac{a_1}{a_0 + a_1}}$$

- Many Bayesians consider the Bayes factor on its own ground (*outside a true decision-theoretic setting*).

- Jeffreys developed a scale to judge the evidence in favor of or against H_0 brought by the data,

  ∗ $\log_{10}(B_{01}^{\pi})$ varies between 0 and 0.5, the evidence against H_0 is *poor*,

  ∗ if it is between 0.5 and 1, it is *substantial*,

  ∗ if it is between 1 and 2, it is *strong*, and

  ∗ if it is above 2, it is *decisive*

  ⋆⋆ This gives guidelines for Bayesian testing of hypotheses depending on the value of log-Bayes factor.

- The exact bounds can be driven based on a loss function.

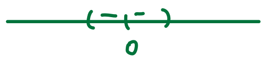
† Testing a point null hypothesis

- For continuous $\pi(\theta)$, $P^{\pi}(\theta = \theta_0) = 0$ & $P^{\pi}(\theta = \theta_0 \mid x) = 0$

- $H_0 : \theta = \theta_0$ will virtually never be the case that one seriously entertains the possibility that $\theta = \theta_0$ *exactly*.

  $\Rightarrow$ $H_0 : \theta = \theta_0$ *unrealistic!*

$\theta = 0$



- More reasonable that

  ⋆⋆ Define $H_0 : \theta \in \Theta_0 = (\theta_0 - \epsilon, \theta_0 + \epsilon)$

  ⋆⋆ Choose some constant $\epsilon > 0$ such that all θ in Θ_0 can be considered "indistinguishable" from θ_0.

  ⋆⋆ *indistinguishable?* For any $\theta \in (\theta_0 - \epsilon, \theta_0 + \epsilon)$, the observed likelihood function $f(x \mid \theta)$ is approximately constant.

† Testing a point null hypothesis (contd)

- Consider $H_0 : \theta = \theta_0$ vs $H_0 : \theta \neq \theta_0$.

- Consider the following prior

$$\theta \begin{cases} = \theta_0 & \text{with probability } \rho_0, \\ \sim g_1(\theta) & \text{with probability } \rho_1 = 1 - \rho_0, \end{cases}$$

where probability distribution $g_1(\theta)$ gives probability zero to the event $\theta = \theta_0$.

$\Rightarrow$ We rewrite

$$\pi(\theta) = \rho_0 \delta_{\theta_0} + (1 - \rho_0) g_1(\theta),$$

where δ_{θ_0} is the Dirac mass at θ_0.