03/07

1. Final: March 17th 9 am — March 18th 12pm

2. Midterm 2 $\longrightarrow$

• **Example 1** (contd) We rewrite the model using the normal-scale mixture representation of a t-distribution;

$$
\begin{aligned}
y_i \mid \mu, V_i &\overset{indep}{\sim} \mathrm{N}(\mu, V_i),\ i = 1, \ldots, n, \\
V_i \mid \sigma^2 &\overset{iid}{\sim} \text{Inv-}\chi^2(\nu, \sigma^2), \\
\pi(\mu, \sigma^2) &\propto 1/\sigma^2,
\end{aligned}
$$

where ν is fixed.

$\{\mu, \sigma^2; V_1, \ldots, V_n\}$

⋆⋆ The joint posterior is

$$
\underline{p(\mu, \sigma^2, V_i \mid y_1, \ldots, y_n)} \propto \overbrace{\frac{1}{\sigma^2}}^{\pi(\mu,\sigma^2)} \prod_{i=1}^{n} \overbrace{\frac{1}{\sqrt{2\pi V_i}} \exp\left\{-\frac{(y_i - \mu)^2}{2V_i}\right\}}^{= f(y_i \mid \mu, V_i)}
$$

$$
\times \prod_{i=1}^{n} \underbrace{\frac{(\nu\sigma^2/2)^{\nu/2}}{\Gamma(\nu/2)} V_i^{-\nu/2} \exp\left(-\frac{\nu\sigma^2}{2V_i}\right)}_{p(V_i \mid \sigma^2)}.
$$

$$\pi(\mu \mid \text{---}) \quad \propto \quad \exp\left\{ - \sum_{i=1}^{n} \frac{(y_i - \mu)^2}{2 V_i} \right\}$$

$$\propto \quad \exp\left[- \frac{1}{2} \left\{ \left(\sum_{i=1}^{n} \frac{1}{V_i} \right) \mu^2 - 2 \sum_{i=1}^{n} \frac{y_i}{V_i} \cdot \mu \right\} \right]$$

$$\Rightarrow \quad \pi(\mu \mid V_i, y) \quad = \quad N\left(\left(\sum_{i=1}^{n} \frac{1}{V_i} \right)^{-1} \left(\sum_{i=1}^{n} \frac{y_i}{V_i} \right), \ \left(\sum_{i=1}^{n} \frac{1}{V_i} \right)^{-1} \right)$$

$$\pi(\sigma^2 \mid \text{---}) \quad \propto \quad (\sigma^2)^{-1 + n\nu/2} \quad \exp\left(- \sigma^2 \cdot \sum_{i=1}^{n} \frac{\nu}{2V_i} \right)$$

$$\Rightarrow \quad \pi(\sigma^2 \mid \text{---}) \quad = \quad Ga\left(\frac{n\nu}{2}, \quad \sum_{i=1}^{n} \frac{\nu}{2V_i} \right)$$

$$\pi(V_i \mid \text{---}) \quad \propto \quad (V_i)^{-1/2 - \nu/2} \quad \exp\left(- \frac{(y_i - \mu)^2}{2V_i} - \frac{\nu \sigma^2}{2 V_i} \right)$$

$$\Rightarrow \quad \pi(V_i \mid \text{---}) = IG\left(\left(\frac{\nu}{2} + \frac{1}{2} \right), \quad \frac{1}{2} \left((y_i - \mu)^2 + \nu \cdot \sigma^2 \right) \right)$$

$i = 1, \ldots, n$

- **Example 1 Model 2** (contd)

  ⋆⋆ Then the full conditionals are

$$p(\mu \mid -) \propto \exp\left\{-\sum_{i=1}^{n}\frac{(y_i-\mu)^2}{2V_i}\right\}$$

$$\Rightarrow \quad \mu \mid - \sim \mathrm{N}\left((\sum_{i=1}^{n}\frac{1}{V_i})^{-1}\sum_i \frac{y_i}{V_i}, (\sum_{i=1}^{n}\frac{1}{V_i})^{-1}\right)$$

$$p(\sigma^2 \mid -) \propto (\sigma^2)^{-1+n\nu/2}\exp\left(-\sum_{i=1}^{n}\frac{\nu\sigma^2}{2V_i}\right)$$

$$\Rightarrow \quad \sigma^2 \mid - \sim \mathrm{Gamma}\left(\frac{n\nu}{2}, \sum_{i=1}^{n}\frac{\nu}{2V_i}\right)$$

- **Example 1 Model 2** (contd)
  - ⋆⋆ (contd) Then the full conditionals are

$$p(V_i \mid -) \propto V_i^{-\nu/2-1/2} \exp\left\{-\frac{(y_i-\mu)^2}{2V_i} - \frac{\nu\sigma^2}{2V_i}\right\}$$

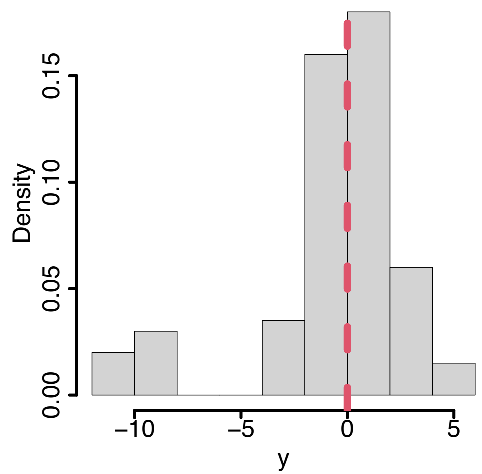
$$\Rightarrow \quad V_i \mid - \overset{indep}{\sim} \mathrm{IG}\left(\frac{\nu+1}{2}, \frac{(y_i-\mu)^2+\nu\sigma^2}{2}\right), \quad i=1,\ldots,n$$

  - ⋆⋆ It is straightforward to perform the Gibbs sampler on V, μ and σ^2 in the augmented model.
  - ⋆⋆ More importantly, the simulations for μ and σ^2 under the augmented model represent the posterior distribution of μ and σ^2 under the original t model.

- Simulated data for **Example 1**

  ⋆⋆ Simulate data

$$
\begin{aligned}
y_i &\overset{iid}{\sim} \mathrm{N}(0,4), i = 1, \ldots, 90, \quad \texttt{good obs} \\
y_i &\overset{iid}{\sim} \mathrm{N}(-10,1), i = 91, \ldots, 100. \quad \texttt{bad obs}
\end{aligned}
$$



- Simulated data for **Example 1** (contd)

Consider the following models

⋆⋆ Model A:

$t_\nu(\mu, \sigma^2)$

$E(y) = \mu$

$Var(y) = \frac{\nu}{\nu-2} \cdot \sigma^2$ if $\nu > 2$

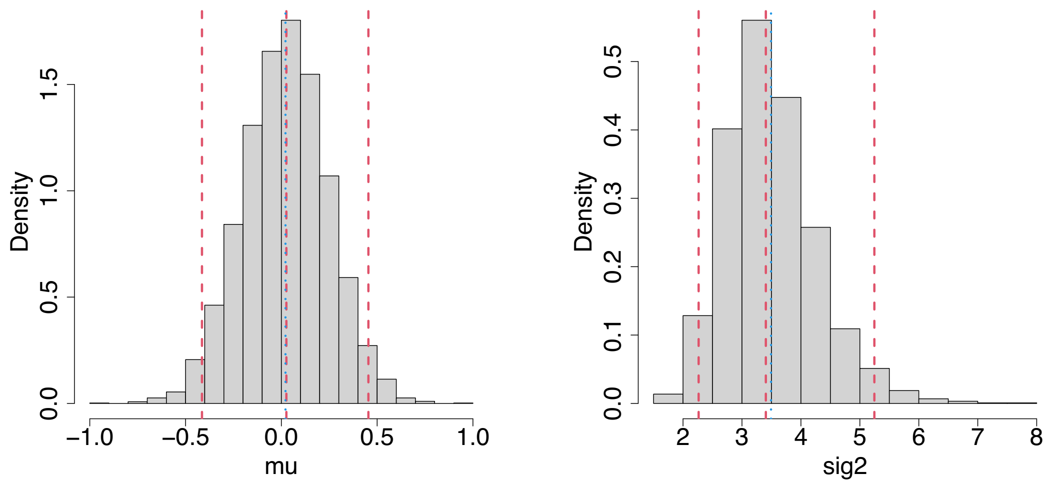
$$\begin{aligned} y_i \mid \mu, V_i &\overset{indep}{\sim} \mathrm{N}(\mu, V_i), i = 1, \ldots, n, \\ V_i \mid \sigma^2 &\overset{iid}{\sim} \text{Inv-}\chi^2(\nu, \sigma^2), \\ \pi(\mu, \sigma^2) &\propto 1/\sigma^2, \end{aligned}$$

where $\nu = 3$ is fixed.

⋆⋆ Model B:

$$\begin{aligned} y_i \mid \mu, \sigma^2 &\overset{iid}{\sim} \mathrm{N}(\mu, \sigma^2), i = 1, \ldots, n, \\ \pi(\mu, \sigma^2) &\propto 1/\sigma^2. \end{aligned}$$

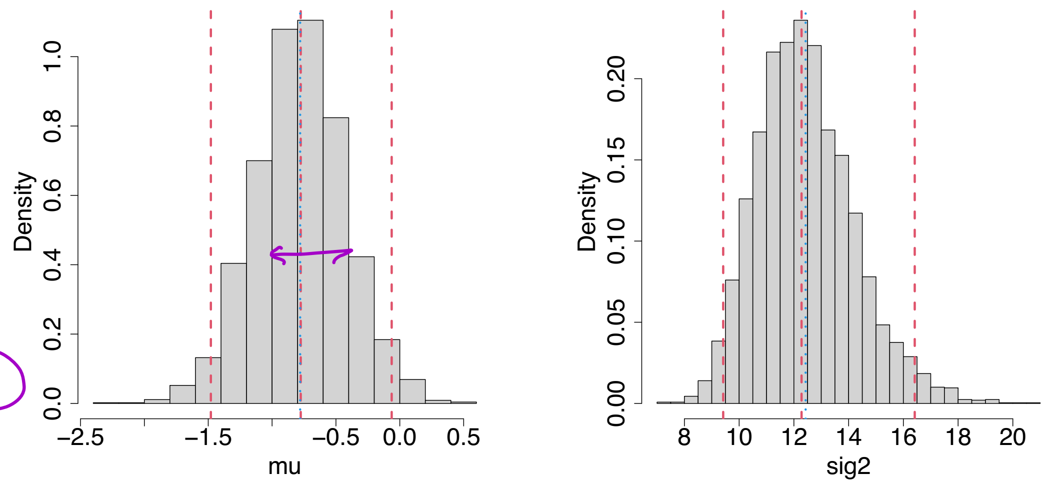
- Model A:
  - ⋆⋆ post. mean $\hat{\mu} = 0.022$ with 95% CI $(-0.414, 0.454)$
  - ⋆⋆ post. mean $\hat{\sigma}^2 = 3.493$ with 95% CI $(2.061, 4.578)$.



- Model B:

  ⋆⋆ post. mean $\hat{\mu} = -0.78$ with 95% CI $(-1.479, -0.064)$

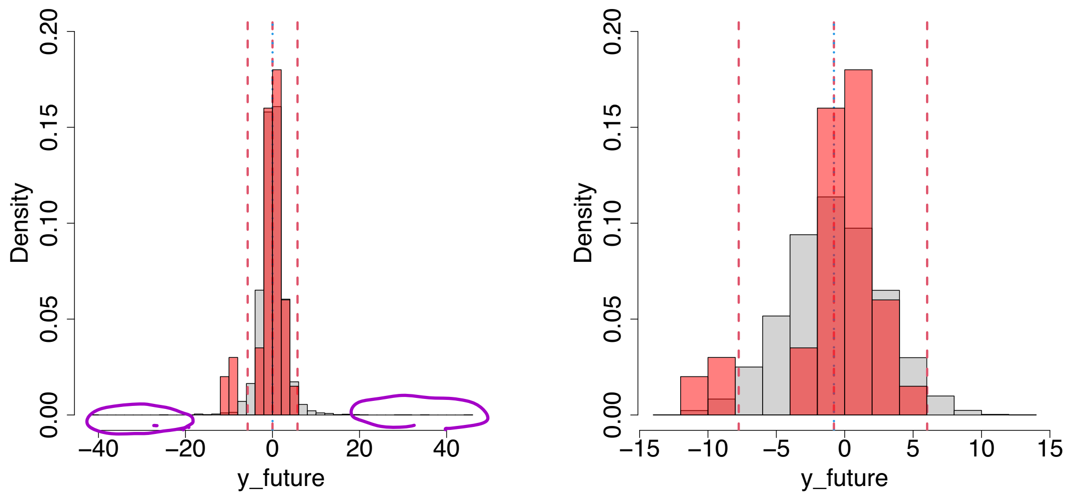
  ⋆⋆ post. mean $\hat{\sigma}^2 = 12.429$ with 95% CI $(9.417, 16.415)$.



- Predictive distribution

  ⋆⋆ Model A: post. pred. mean $\hat{y}^{\text{NEW}} = 0.009$ with 95% posterior predictive interval $(-5.710, 5.747)$

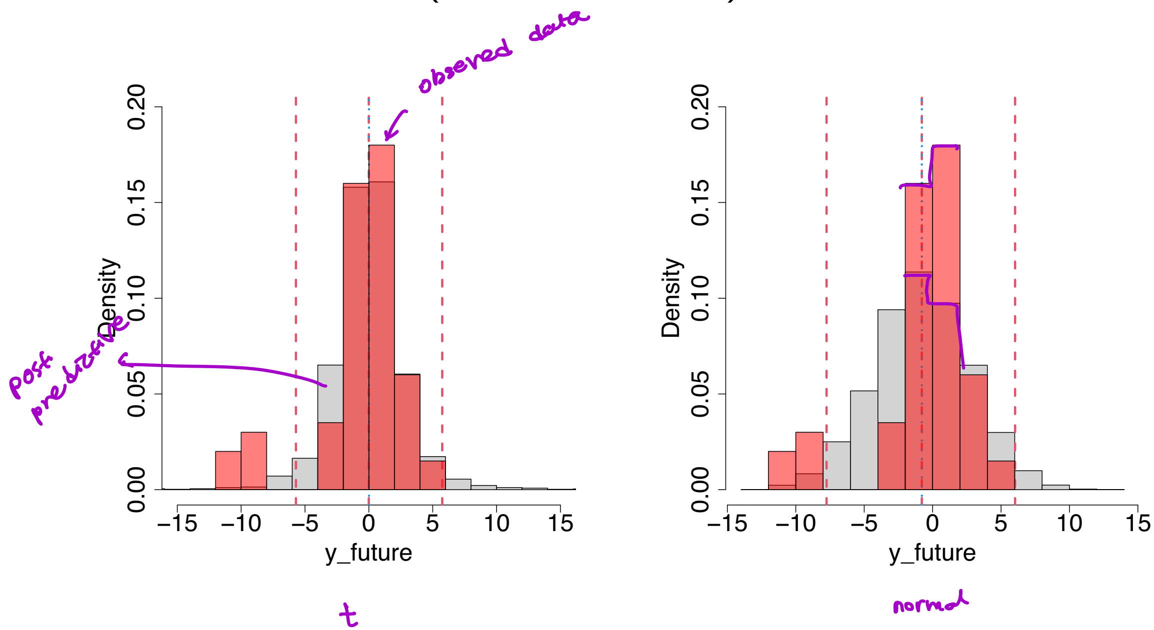
  ⋆⋆ Model B: post. pred. mean $\hat{y}^{\text{NEW}} = -0.789$ with 95% posterior predictive interval $(-7.750, 6.026)$



- Predictive distribution (contd) : zoom-in

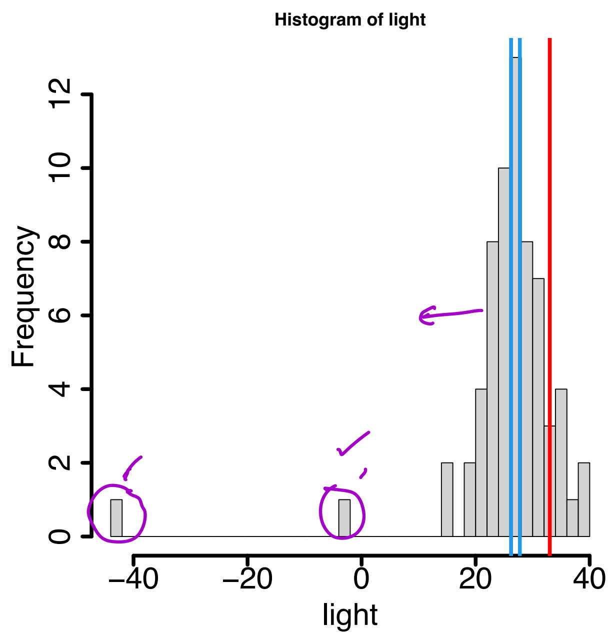
  ★★ Model A: post. pred. mean $\hat{y}^{NEW} = 0.009$ with 95% posterior predictive interval $(-5.710, 5.747)$

  ★★ Model B: post. pred. mean $\hat{y}^{NEW} = -0.789$ with 95% posterior predictive interval $(-7.750, 6.026)$



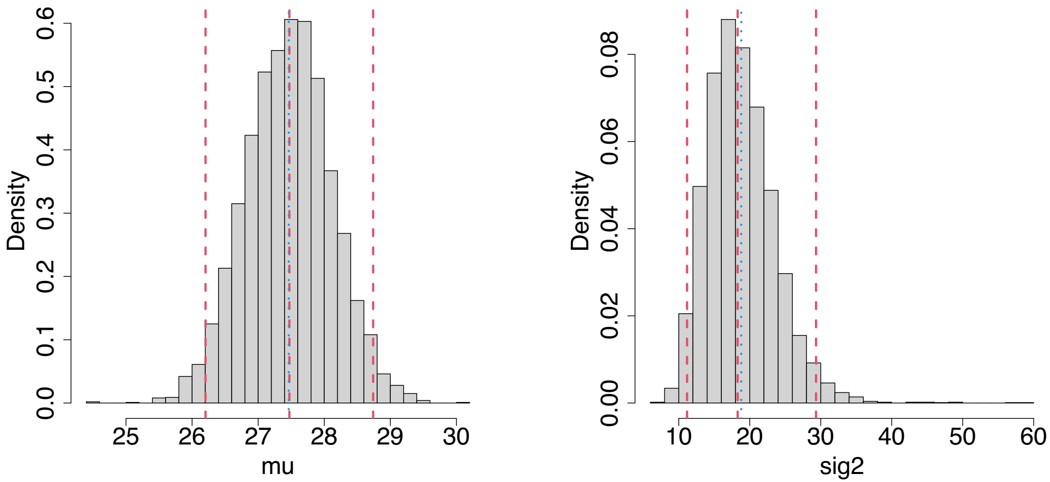
† Example (revisit): Estimating the speed of light (BDA p 66)

- Simon Newcomb set up an experiment in 1882 to measure the speed of light. Newcomb measured the amount of time required for light to travel a distance of 7442 meters. He made 66 measurements. Consider the problem of estimating the speed of light.



† Example: Estimating the speed of light (contd)

- Use a t-model; Posterior summary of θ and σ^2

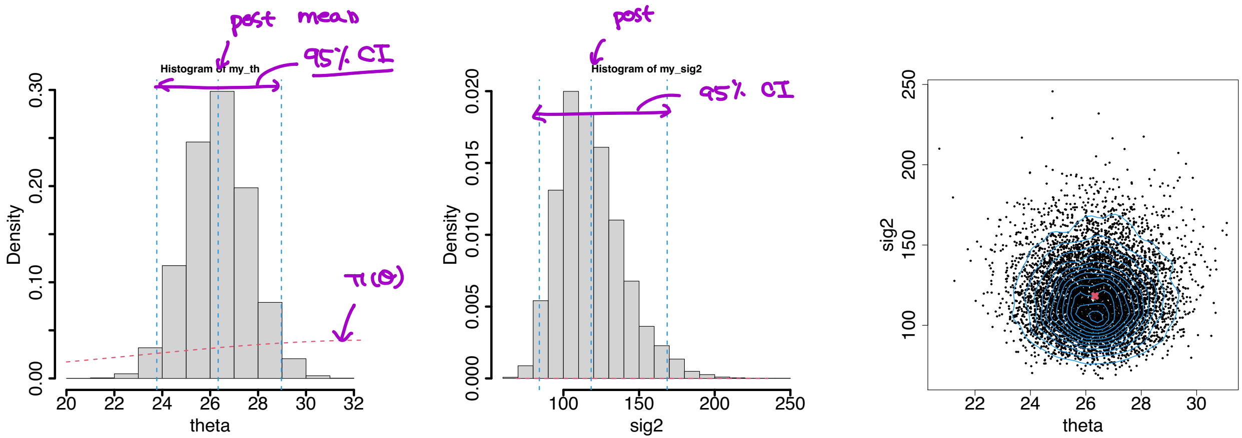


(a) θ (b) σ^2

normal model

† Example: Estimating the speed of light (contd)

- Posterior summary of θ and σ^2



(a) θ (b) σ^2 (c) Joint

† Example: Estimating the speed of light (contd)

- Use a t-model; Posterior summary of θ and σ^2

t-model

27.463

(26.2, 28.74)

```
> print(round(quantile(SAVE_MCMC_sam$mu[inf_sam], prob=c(0.025, 0.5, 0.975)), 3))
  2.5%    50%  97.5%
26.203 27.472 28.740
> print(round(mean(SAVE_MCMC_sam$mu[inf_sam]), 3))
[1] 27.463
>
> print(round(quantile(SAVE_MCMC_sam$sig2[inf_sam], prob=c(0.025, 0.5, 0.975)), 3))
  2.5%    50%  97.5%
11.169 18.298 29.337
> print(round(mean(SAVE_MCMC_sam$sig2[inf_sam]), 3))
[1] 18.801
>
```

(a) t-model

normal

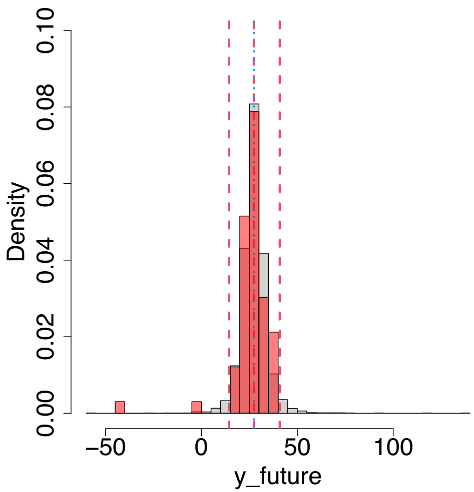
26.307

(23.67, 29.01)

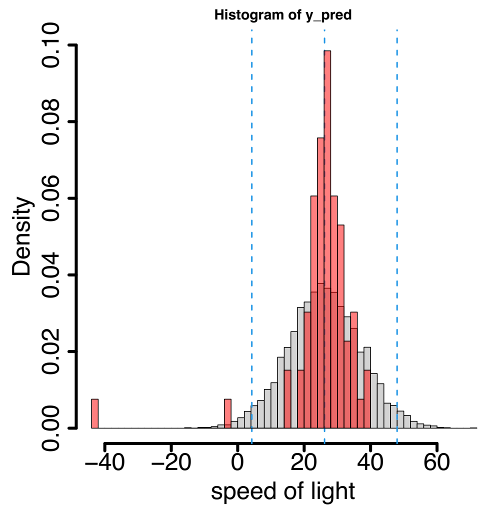
```
> ### summaries of the margianl posterior of theta
> post_m_th <- mean(my_th)
> post_sd_th <- sd(my_th)
> ci_th <- quantile(my_th, prob=c(0.025, 0.975))
> post_m_th
[1] 26.30754
> post_sd_th
[1] 1.355212
> ci_th
    2.5%    97.5%
23.66675 29.01357
>
> ### summaries of the margianl posterior of sig2
> 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] 119.0088
> post_sd_sig2
[1] 21.49393
> ci_sig2
     2.5%     97.5%
 84.55515 167.76078
>
```

† Example: Estimating the speed of light (contd)

- Use a t-model; Summary of the posterior predictive distribution of unobserved y



(a) t-model

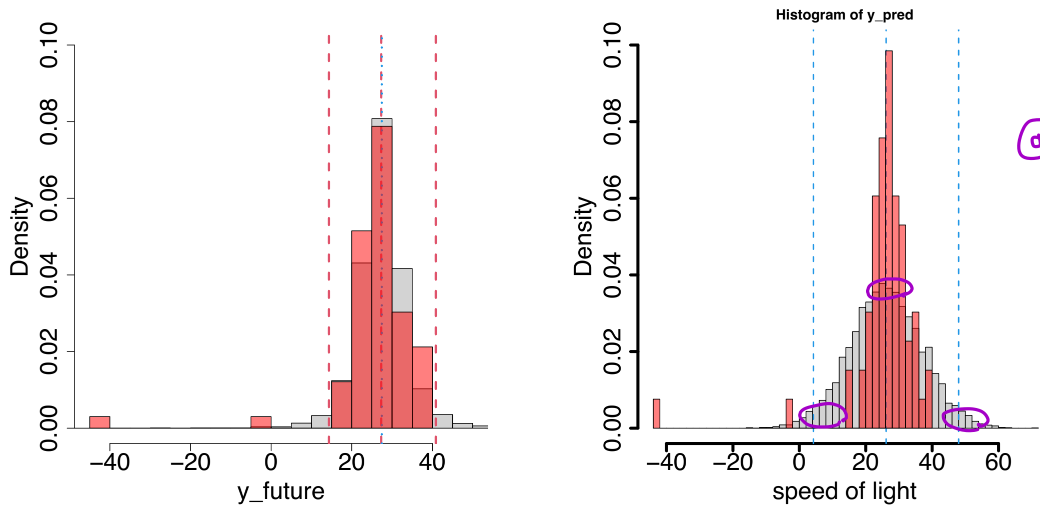


(b) normal

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

- Use a t-model; Summary of the posterior predictive distribution of unobserved y

Red: observed



(a) t-model (b) normal

σ^2

† Example: Estimating the speed of light (contd)

- Use a t-model; Summary of the posterior predictive distribution of unobserved y

```
> print(round(quantile(y_pred, prob=c(0.025, 0.5, 0.975)), 3))
  2.5%    50%  97.5%
14.321 27.317 40.842
> print(round(mean(y_pred), 3))
[1] 27.441
```

(a) t-model

```
> ###%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
> #%%%## preditive distribution
> y_pred <- rnorm(length(my_th), my_th, sqrt(my_sig2))
> mean(y_pred)
[1] 26.24271
> sd(y_pred)
[1] 11.01951
> quantile(y_pred, prob=c(0.025, 0.975))
     2.5%     97.5%
 4.343315 47.816016
>
```

(b) normal

- **Example 1** (contd) More examples?

  ⋆⋆ $y \mid \theta \sim \text{Bin}(n, \theta)$ and $\theta \sim \text{Be}(\alpha, \beta)$ (Beta-Binomial Mixture) where θ is an auxiliary variable.

  $\Rightarrow y \mid \alpha, \beta \sim \text{Beta-Binom}(n, \alpha, \beta)$

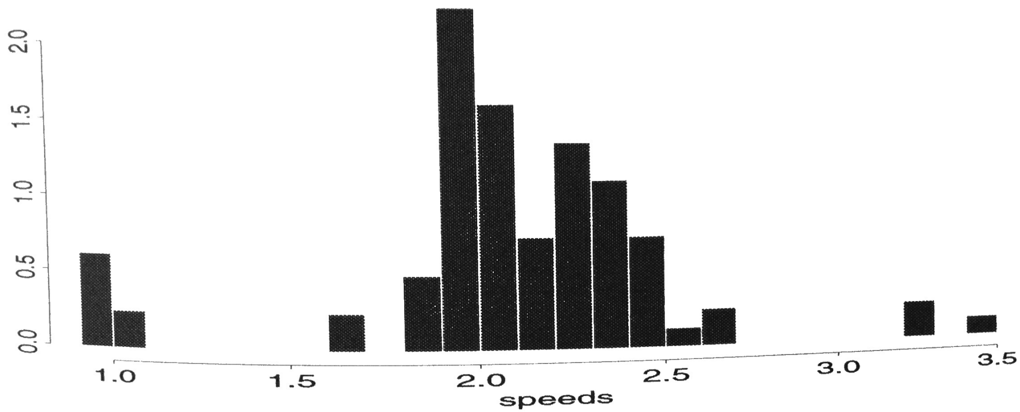
  See also **Example 6.3.4**.

  ⋆⋆ $y \mid \theta \sim \text{Poi}(\theta)$ and $\theta \sim \text{Gamma}(r, \frac{1-p}{p})$ (Gamma-Poisson Mixture) where θ is an auxiliary variable.

  $\Rightarrow y \mid r, p \sim \text{Neg-Binom}(r, p)$ where r: $\#$ of failures and p: success probability.

• **Example 7.1.2** (I changed a bit, especially notation) The dataset consists in 82 observations of galaxy velocities.

⋆⋆ Histogram of the galaxy dataset of Roeder (1992)



⋆⋆ For astrophysical reasons, the distribution of this dataset can be represented as a mixture of normal distributions. Suppose the number of components is k (fixed).

HW#2- Q1

$$X_i \mid \theta, \sigma^2 \overset{iid}{\sim} N(\theta, \sigma^2)$$

$$\pi(\theta) = \sum_{\ell=1}^{K} w_\ell \cdot \phi(\theta \mid \mu_\ell, \tau^2)$$

- **Example 7.1.2** (contd) Recall a mixture model with k components:

If $\lambda_j = 1$, $y_j \sim N(\mu_1, \sigma^2)$
If $\lambda_j = \ell$, $y_j \sim N(\mu_\ell, \sigma^2)$

$$y_j \stackrel{iid}{\sim} \sum_{\ell=1}^{k} p_\ell \mathrm{N}(\mu_\ell, \sigma^2), \quad j = 1, \ldots, J(= 82).$$

$y_j \mid \lambda_j \sim N(\mu_{\lambda_j}, \sigma^2)$

The mixture model can be represented as follows;

⋆⋆ We introduce auxiliary variables $\lambda_j \in \{1, \ldots, k\}$.

⋆⋆ We assume $p(\lambda_j = \ell) = p_\ell$, independence between λ_j.

⋆⋆ Given λ_j, we write the distribution of y_j

$$\Rightarrow \quad y_j \mid \boldsymbol{\mu}, \sigma^2, \lambda_j = \ell \sim \mathrm{N}(\mu_\ell, \sigma^2).$$

$p(y) = \sum_{\ell=1}^{k} P_r(\lambda = \ell) \cdot P_r(y \mid \lambda = \ell)$

- **Example 7.1.2** (contd) Let's develop the model further.
  - ⋆⋆ The likelihood

$$y_j \mid \lambda_j, \mu, \sigma \sim \mathrm{N}(\mu_{\lambda_j}, \sigma^2).$$

  - ⋆⋆ (prior) Let $p(\lambda_j = \ell \mid p) = p_\ell$, independence between λ_j.
  - ⋆⋆ (prior) Let $p = (p_1, \ldots, p_k) \sim \mathrm{Dir}(\alpha_1, \ldots, \alpha_k)$ with fixed α.
  - ⋆⋆ (prior) Let $\mu_\ell \overset{iid}{\sim} \mathrm{N}(\bar{\mu}, \tau^2)$ with fixed $\bar{\mu}$ and τ^2, and $\sigma^2 \sim \mathrm{IG}(a, b)$ with fixed a and b.
  - ⇒ We have random parameters $\boldsymbol{\theta} = (\{\lambda_j\}_{j=1}^{\cancel{n}\,J}, p, \{\mu_\ell\}_{\ell=1}^{k}, \sigma^2)$.
  - ⇒ Without λ_j, the likelihood evaluation becomes messy. But the likelihood evaluation conditional on λ_j is so simple! We will simulate $\boldsymbol{\theta}$ through MCMC.

$$p(\theta \mid y) \propto p(y \mid \theta)\; p(\lambda, p, \mu, \sigma^2)$$

$$\propto \prod_{j=1}^{J} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(y_j - \mu_{\lambda_j})^2\right)$$

$$\times \prod_{j=1}^{J} p_{\lambda_j} \cdot \prod_{l=1}^{k} p_l^{\alpha_l - 1} \cdot \prod_{l=1}^{k} \exp\left(-\frac{1}{2\tau^2}(\mu_l - \bar{\mu})^2\right)$$

$$\times (\sigma^2)^{-a-1} \exp\left(-\frac{b}{\sigma^2}\right)$$

① update λ_j, $j=1,\dots,J$

$l=1,\dots,k$

$$p(\lambda_j = l \mid \text{—}) \propto \underbrace{\exp\left(-\frac{1}{2\sigma^2}(y_j - \mu_l)^2\right) \cdot p_l}_{= z_l}$$

$$p(\lambda_j = l \mid \text{—}) = \frac{z_l}{\sum_{l'=1}^{k} z_{l'}} = \tilde{z}_l$$

$$\lambda_j \sim \text{Multinomial}(1, \tilde{z}_l, l=1,\dots,k)$$

R. sample ((1:k), 1, T, z)

$(z_1, \dots, z_k)$

② $p = (p_1, \dots, p_k)$ $\quad \lambda_j \in \{1, \dots, k\}$ $\quad S_l = \{ j ; \lambda_j = l \}$

$$p(p \mid \text{—}) \propto \prod_{j=1}^{J} p_{\lambda_j} \cdot \prod_{l=1}^{k} p_l^{\alpha_l - 1}$$

$$= \prod_{l=1}^{k} p_l^{|S_l|} \cdot \prod_{l=1}^{k} p_l^{\alpha_l - 1}$$

$$= \prod_{l=1}^{k} p_l^{|S_l| + \alpha_l - 1}$$

$$p = (p_1, \dots, p_k) \mid \text{—} \sim \text{Dir}(|S_l| + \alpha_l)$$

③ μ_l, $l=1,\dots,k$

$$p(\mu_l \mid \text{—}) \propto \prod_{j \in S_l} \exp\left(-\frac{(y_j - \mu_l)^2}{2\sigma^2}\right) \cdot \exp\left(-\frac{(\mu_l - \bar{\mu})^2}{2\tau^2}\right)$$

$$\propto \exp\left[-\frac{1}{2} \left\{ \left(\frac{|S_\ell|}{\sigma^2} + \frac{1}{\tau^2} \right) \mu_\ell^2 - 2 \left(\frac{\sum_{j \in S_\ell} y_j}{\sigma^2} + \frac{\mu}{\tau^2} \right) \mu_\ell \right\} \right]$$

$$\Rightarrow \mu_\ell \mid \text{——} \sim N\left(\left(\frac{|S_\ell|}{\sigma^2} + \frac{1}{\tau^2} \right)^{-1} \left(\frac{\sum_{j \in S_\ell} y_j}{\sigma^2} + \frac{\mu}{\tau^2} \right), \left(\frac{|S_\ell|}{\sigma^2} + \frac{1}{\tau^2} \right)^{-1} \right)$$

④ σ^2.

$$p(\sigma^2 \mid \text{—}) \propto (\sigma^2)^{-J/2} \cdot \exp\left(- \frac{\sum_{j=1}^{J} (y_j - \mu_{z_j})^2}{2\sigma^2} \right) \cdot (\sigma^2)^{-a-1} e^{-\frac{b}{\sigma^2}}$$

$$\Rightarrow \sigma^2 \mid \text{—} \sim IG\left(\frac{J}{2} + a,\ b + \frac{\sum_{j=1}^{J} (y_j - \mu_{z_j})^2}{2} \right)$$

- **Example 7.1.2** (contd) We first write the joint posterior of $\boldsymbol{\theta}$.

$$
\begin{aligned}
\pi(\boldsymbol{\theta} \mid y) &\propto \prod_{\ell=1}^{k} \pi(\mu_\ell)\ \pi(\sigma^2)\ \pi(p) \prod_{j=1}^{J} \pi(\lambda_j)\ p(y_j \mid \lambda_j, \mu, \sigma) \\
&\propto \underbrace{\exp\left\{-\sum_{\ell=1}^{k} \frac{(\mu_\ell - \bar{\mu})^2}{2\tau^2}\right\}}_{\pi(\mu_\ell)} \underbrace{(\sigma^2)^{-a-1} \exp\left(-\frac{b}{\sigma^2}\right)}_{\pi(\sigma^2)} \\
&\quad \times \underbrace{\prod_{\ell=1}^{k} p_\ell^{\alpha_\ell - 1}}_{\pi(p)} \underbrace{\prod_{j=1}^{J} p_{\lambda_j}}_{\pi(\lambda_j)} \underbrace{\prod_{j=1}^{J} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(y_j - \mu_{\lambda_j})^2}{2\sigma^2}\right\}}_{p(y_j \mid \lambda_j, \mu, \sigma)}
\end{aligned}
$$

⋆⋆ We use the Gibbs sampler to simulate θ. We first drive the full conditionals.

- **Example 7.1.2** (contd) the full conditionals

We use S_ℓ to denote the set of y_j having λ_j,
$S_\ell = \{j : \lambda_j = \ell, j = 1, \ldots, J\}$. Also, let $\bar{y}_\ell = \frac{\sum_{j \in S_\ell} y_j}{|S_\ell|}$.

⋆⋆ μ_ℓ, $\ell = 1, \ldots, k$.

$$p(\mu_\ell \mid \lambda, \sigma^2, y) \propto \exp\left\{-\frac{(\mu_\ell - \bar{\mu})^2}{2\tau^2} - \sum_{j \in S_\ell} \frac{(y_j - \mu_\ell)^2}{2\sigma^2}\right\}.$$

$$\Rightarrow \mu_\ell \mid \lambda, \sigma^2, y \sim \mathrm{N}\left(\left(\tfrac{1}{\tau^2} + \tfrac{|S_\ell|}{\sigma^2}\right)^{-1}\left(\tfrac{\bar{\mu}}{\tau^2} + \tfrac{\bar{y}_\ell}{\sigma^2/|S_\ell|}\right), \left(\tfrac{1}{\tau^2} + \tfrac{|S_\ell|}{\sigma^2}\right)^{-1}\right).$$

⋆⋆ σ^2

$$p(\sigma^2 \mid \lambda, \mu, y) \propto (\sigma^2)^{-a-1} \exp\left(-\frac{b}{\sigma^2}\right)(\sigma^2)^{-J/2} \exp\left\{-\sum_{j=1}^{J} \frac{(y_j - \mu_{\lambda_j})^2}{2\sigma^2}\right\}.$$

$$\Rightarrow \mu_\ell \mid \lambda, \sigma^2, y \sim \mathrm{IG}\left(a + \tfrac{J}{2}, b + \sum_{j=1}^{J} \tfrac{(y_j - \mu_{\lambda_j})^2}{2}\right).$$

- **Example 7.1.2** (contd) the full conditionals

$\star\star$ $p = (p_1, \ldots, p_k)$

$$p(p \mid \lambda) \propto \prod_{\ell=1}^{k} p_\ell^{\alpha_\ell - 1} \prod_{\ell=1}^{k} p_\ell^{|S_\ell|}.$$

$\Rightarrow$ $p \mid \lambda \sim \mathrm{Dir}(\alpha_1 + |S_1|, \ldots, \alpha_k + |S_k|)$.

$\star\star$ λ_j, $j = 1, \ldots, J$

$$p(\lambda_j = \ell \mid \mu, \sigma^2, y) \propto p_\ell \exp\left\{-\frac{(y_j - \mu_\ell)^2}{2\sigma^2}\right\}.$$

$\Rightarrow$ No standard form. So we sample on the grid of $(1, \ldots, k)$.

- **Example 7.1.2** (contd) Hyperparameters

  ⋆⋆ $k = 4$

  ⋆⋆ $\bar{\mu} = 2.08$ and $\tau^2 = 10$

  ⋆⋆ $a = 1$ and $b = 0.01$

  ⋆⋆ $\alpha_1 = \ldots = \alpha_k = 1$

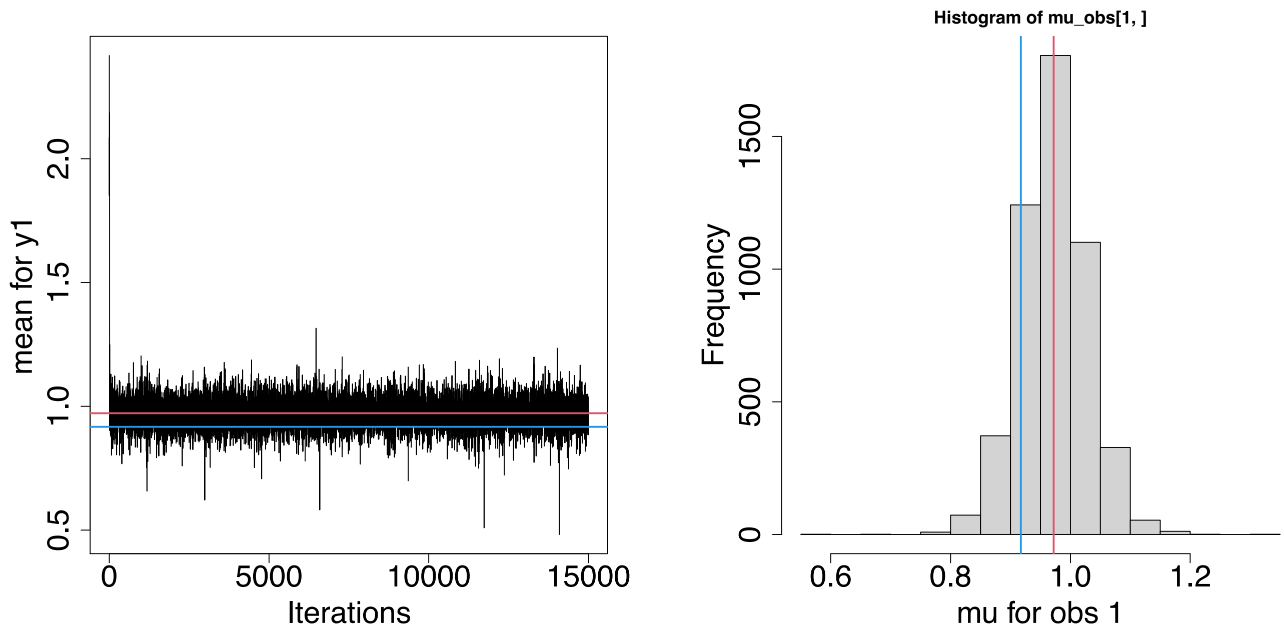
∗ run MCMC

```
n_sam <- 15000
n_burn_in <- 5000
n_thin <- 2
```

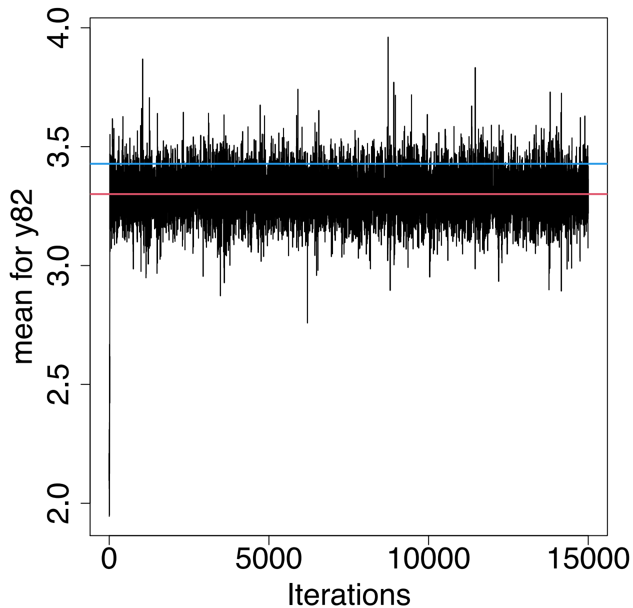
For details, see my code (posted on the course webpage).

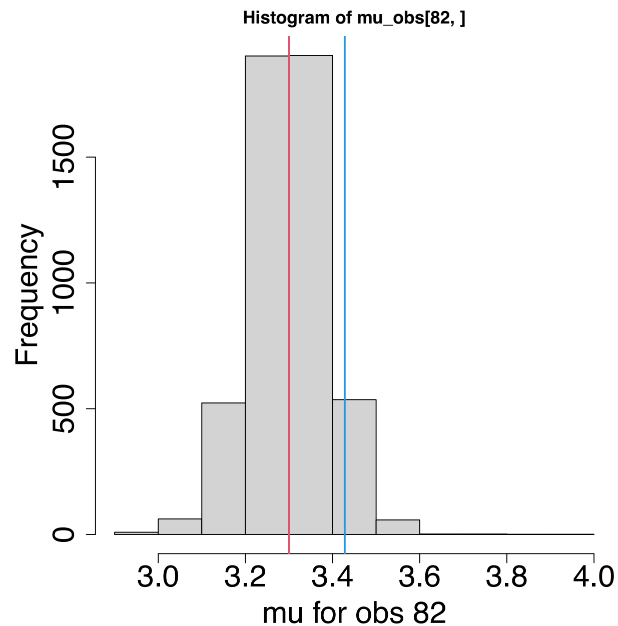
- **Example 7.1.2** (contd) $y_1 = 0.9172$ (blue, smallest), posterior mean for y_1=0.9716 (red).

μ_{x_1}

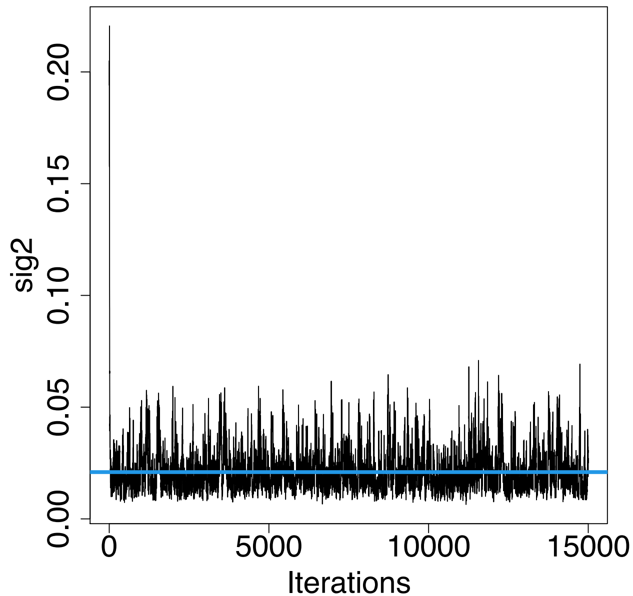


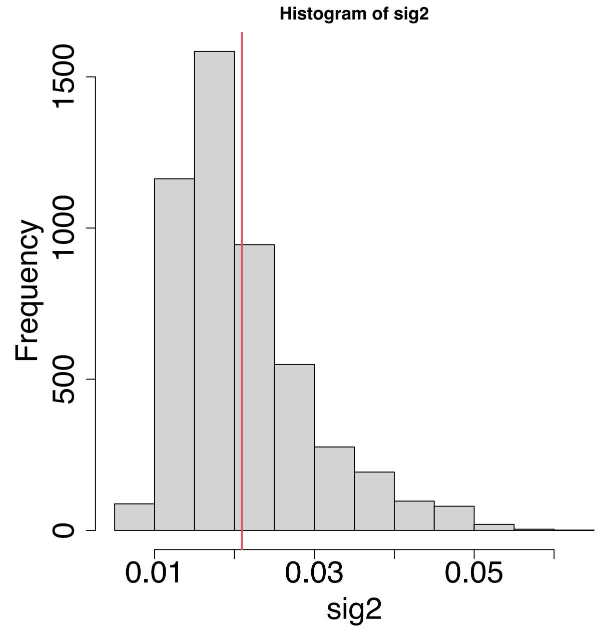
- **Example 7.1.2** (contd) $y_{82} = 3.4279$ (blue, largest), posterior mean for y_{82}=3.30 (red).





- **Example 7.1.2** (contd) σ^2

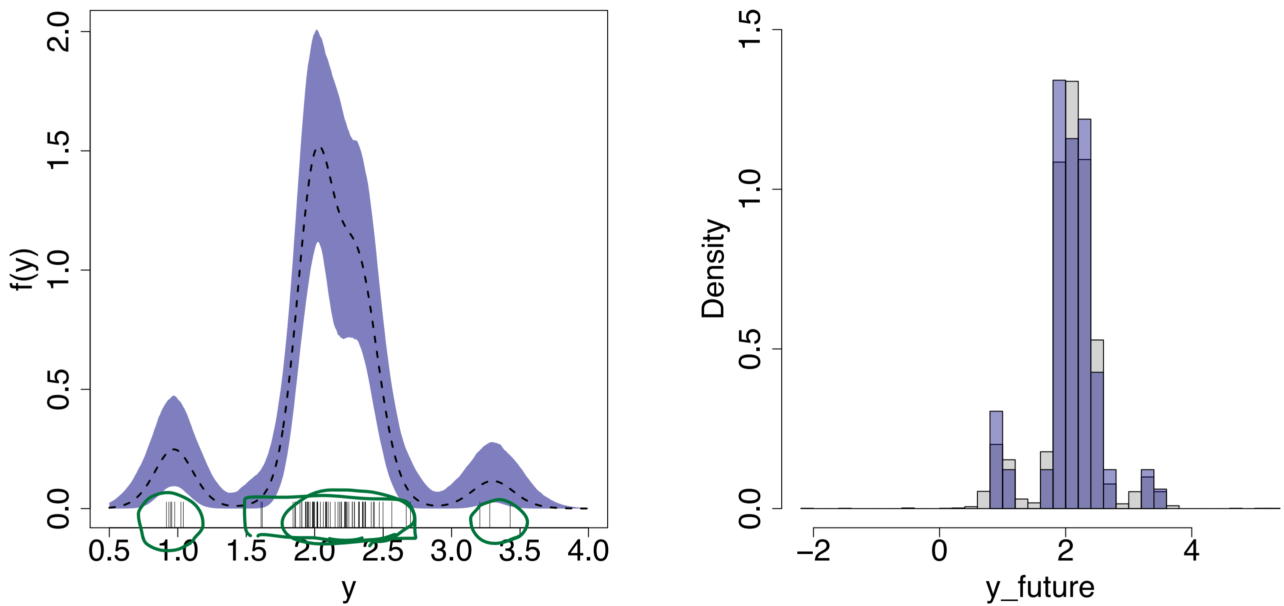




- **Example 7.1.2** (contd) σ^2

$z^{(b)}$ $\quad P(z^{(b)} = l) = p_l^{(b)}$

$y^{(b)} \sim N(\mu^{(b)}_{z^{(b)}}, \sigma^{2,(b)})$



(a) $\hat{f}(y)$

(b) $p(y^{\text{new}} \mid \boldsymbol{y})$

$f(y) = \sum_{l=1}^{k} p_l N(y \mid \mu_l, \sigma^2)$

$\approx \frac{1}{B} \sum_{b=1}^{B} \sum_{l=1}^{k} p_l^{(b)} N(y \mid \mu_l^{(b)}, \sigma^{2,(b)})$