02/22/22

- HW #4: March 8th 5pm
- Midterm #2: March 1st

- **PH p82** Recall the IQ example (similar but different!).
    - ⋆⋆ $X \sim \mathrm{N}(100, 225)$ for the general population.
    - ⋆⋆ Suppose that we sample n individuals from a particular town and estimate θ, the town-specific mean IQ score based on the sample of size n.
    - ⋆⋆ In fact, people in the town are extremely exceptional so $\theta_0 = 112$ and $\sigma^2 = 169$.
    - ⋆⋆ Consider $x_i \mid \theta \overset{iid}{\sim} \mathrm{N}(\theta, \sigma^2)$, where $\sigma^2 = 169$ but θ is unknown
    - ⋆⋆ Assume $\underline{\theta \sim \mathrm{N}(\mu_0, \tau_0^2)}$, where $\underline{\tau_0} = \sigma/\sqrt{\kappa_0}$ $\tau_0^2 = \frac{\sigma^2}{\kappa_0}$ $\kappa_0 \uparrow$, prior strong
    - ⋆⋆ For Bayesian inference, we lack the information about the town a natural choice of $\underline{\mu_0 = 100}$.

$$\delta_n^\pi = \left(\frac{n}{\sigma^2} + \frac{1}{\tau_0^2}\right)^{-1}\left(\frac{n\bar{x}}{\sigma^2} + \frac{\mu_0}{\tau_0^2}\right)$$

$$= \underbrace{\frac{\tau_0^2}{\tau_0^2 + \sigma^2/n}}\bar{x} + \frac{\sigma^2/n}{\tau_0^2+\sigma^2/n}\mu_0$$

$\tau_0^2 = \frac{\sigma^2}{k_0}$

$$= \underbrace{\frac{n}{n + k_0}}_{=w}\bar{x} + \underbrace{\frac{k_0}{k_0 + n}}_{=1-w}\mu_0$$

$E(\delta_n^\pi)$

$$= \frac{n}{n+k_0}\cdot\theta_0 + \frac{k_0}{k_0+n}\mu_0$$

$$= w\cdot\theta_0 + (1-w)\mu_0$$

$$MSE(\delta_n^\pi) = Var_{\theta_0}(\delta_n^\pi) + Bias^2(\delta_n^\pi)$$

$$= w^2\cdot\frac{\sigma^2}{n} + \left(w\theta_0 + (1-w)\mu_0 - \theta_0\right)^2$$

$$(1-w)\mu_0 - (1-w)\theta_0$$

$$= (1-w)(\mu_0 - \theta_0)$$

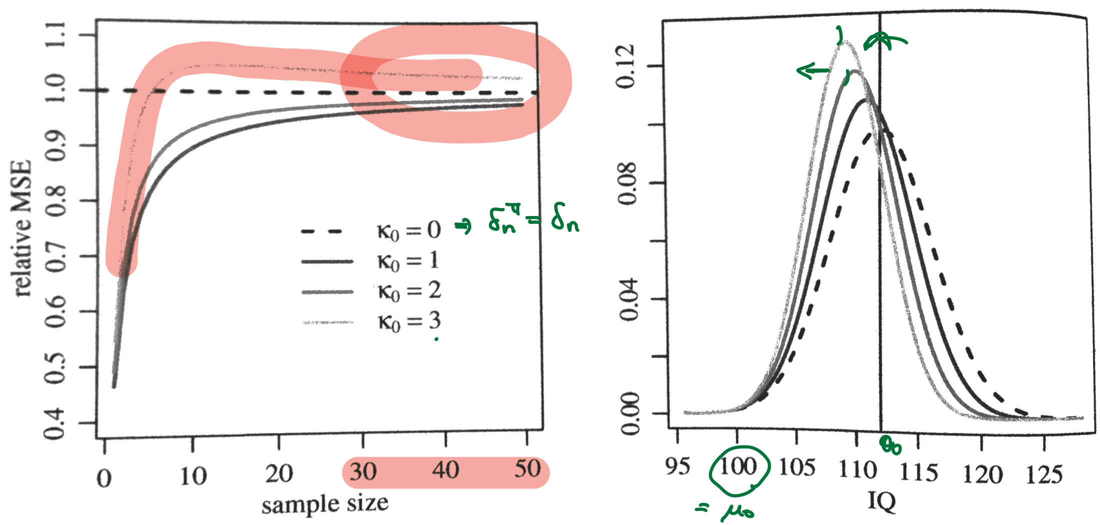
$$= w^2\cdot\frac{\sigma^2}{n} + \underline{(1-w)^2(\mu_0-\theta_0)^2}$$

$\delta_n = \bar{x}$ $\qquad MSE(\delta_n) = \frac{\sigma^2}{n}$

$$\frac{MSE(\delta_n \mid \theta_0)}{MSE(\delta_n^\pi \mid \theta_0)} \begin{cases} = 1 & \\ < 1 & \delta_n \text{ is better than } \delta_n^\pi \\ > 1 & \delta_n^\pi \text{ is better} \end{cases}$$

- **PH p82** Example: IQ Scores.
  - ⋆⋆ Let $\kappa_0 = \sigma^2/\tau^2$ and compare $\mathrm{MSE}(\delta_n^\pi \mid \theta_0)$ and $\mathrm{MSE}(\delta_n \mid \theta_0)$ by varying n and κ_0.
  - ⋆⋆ MSE errors and sampling distribution of different $\delta_n^\pi(x)$

For a given value of n = ?



- Comments on unbiasedness
  - ⋆⋆ No Bayes estimate with respect to the squared error loss can be unbiased, except in a case when its Bayes' risk is 0 (that is, the perfect estimation is possible).

    $\Leftrightarrow$ If $\delta^\pi(x)$ is unbiased for θ, then $\delta^\pi(x)$ is not Bayes under the squared error loss unless its Bayes risk is zero.

    For your practice, show this.

  - ⋆⋆ **No problem!** Even frequentist agree that insisting on unbiasedness can lead to bad estimators, and that in their quest to minimize the risk by trading off between variance and bias-squared a small dosage of bias can help.

† Interval Estimation (CR 5.5 and JB 4.3.2)

- $(1-\alpha)100\%$ confidence intervals (CI's)–Classical interval estimate

  ⋆⋆ Generate data from the assumed model many times and for each data set to exhibit the CI.

  ⋆⋆ Now, the proportion of CIs covering the unknown parameter "tends to" $1-\alpha$.

- We will construct $C_x \subset \Theta$ where θ should be with high probability.

  ⋆⋆ The distribution used to assess the credibility of an interval estimator is the posterior distribution.

† Credible Sets

- Credible Set: Assume the set C_x is a subset of Θ. Then C_x is a credible set with credibility $(1-\alpha)\cdot 100\%$ if

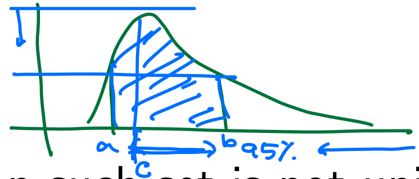
$$P^{\pi}(\theta \in C_x \mid x) = \mathrm{E}^{\pi}\{1(\theta \in C_x) \mid x\} = \int_{C_x} \pi(\theta \mid x) d\theta > 1-\alpha.$$

  ⋆⋆ If the posterior is discrete, then the integral becomes sum.

- Bayesian interpretation of a credible set C_x is natural: The probability of a parameter belonging to the set C_x is $1-\alpha$.

  ⋆⋆ The frequentist CI is random but our credible interval is fixed given data.

† Credible Sets (contd)



(c, ∞)

$Pr(\theta > c \mid x) = 1-\alpha$

- For a given posterior ~~function~~ distr. such set is not unique.

  ⋆⋆ *Q: How to choose one particular set?*

- For a given credibility level $(1-\alpha)100\%$, the shortest credible set is of interest.

- The size of a set is simply its total length if the parameter space Θ is one dimensional, total area, if Θ is two dimensional, and so on.

- To minimize the size, sets should correspond to highest posterior probability (density) areas.

† Credible Sets (contd)

- The $(1-\alpha)100\%$ HPD (high posterior density) credible set for parameter θ is a set C_x, subset of Θ of the form

$$C_x = \{\theta \in \Theta \mid \pi(\theta \mid x) \geq k_\alpha\},$$

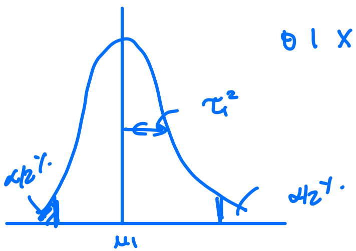
  where k_α is the largest constant for which

$$P^\pi(\theta \in C_x \mid x) \geq 1 - \alpha.$$

- Geometrically, if the posterior density is cut by a horizontal line at the height k_α, the set C is projection on the θ axis of the part of line inside the density, i.e., the part that lies below the density.

- See **Def 5.5.2**

- **Example 5.5.3** Consider $x \sim \mathrm{N}(\theta, \sigma^2)$. Consider $\theta \sim \mathrm{N}(0, \tau^2)$. Find the $100(1-\alpha)\%$ HPD credible interval.



$$\theta \mid x \sim \mathrm{N}\Big(\underbrace{\left(\tfrac{1}{\sigma^2} + \tfrac{1}{\tau^2}\right)^{-1} \cdot \left(x/\sigma^2 + \tfrac{0}{\tau^2} \right)}_{= \mu_1},\ \underbrace{\left(\tfrac{1}{\sigma^2} + \tfrac{1}{\tau^2}\right)^{-1}}_{= \tau_1^2} \Big)$$

$$\left(\mu_1 \pm z_{\alpha/2} \cdot \sqrt{\tau_1^2} \right) = (1-\alpha)100\%\ \text{CI.}$$

95%

$\star\star$ Find the $100(1-\alpha)\%$ HPD credible interval with $\pi(\theta) \propto 1$.

$$\theta \mid x \sim \mathrm{N}(x, \sigma^2)$$

$$\left(x \pm z_{\alpha/2} \cdot \sqrt{\sigma^2} \right)$$

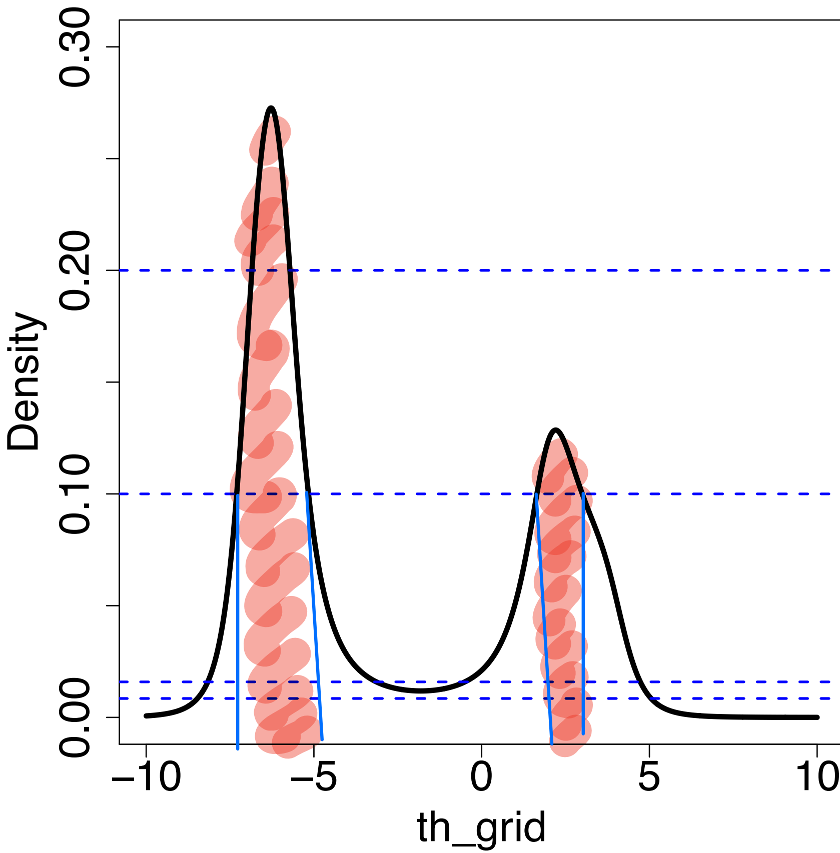
$\star\star$ Note that we can use improper priors in this setting and do not encounter the same difficulties as when testing the point-null hypothesis.

- **JB Example 10** (p141, with a slight change) Assume that four observations, $x_i = 2, -7, 4, -6,\ i = 1, \ldots, 4$ are sampled from Cauchy $C(\theta, 1)$ distribution with parameter of interest θ ($f(x \mid \theta) = 1/\{\pi(1 + (x - \theta)^2)\}$. Consider the flat prior $\pi(\theta) = 1$. Sketch the posterior.

proper!

$$\pi(\theta \mid x) \propto \prod_{i=1}^{n} \frac{1}{\pi(1+(x_i-\theta)^2)} \cdot 1$$

- **Example** (contd) The posterior is bimodal!



k_α

31.3%

$\alpha = 0.687$

$\rightarrow$ $\alpha = 0.358$ $\rightarrow$ 64.2%

- **Example** (contd) Four horizontal lines at levels $k = 0.008475$, 0.0159, 0.1, and 0.2 are shown. These lines determine four credible sets,
  - ⋆⋆ $k_{0.01} = 0.008475$: $[-8.498, 5.077]$ with $P^{\theta|X}(8.498 \leq \theta \leq 5.077) = 99\%$;
  - ⋆⋆ $k_{0.05} = 0.0159$: $[-8.189, -3.022] \cup [-0.615, 4.755]$ with posterior credibility of 95%;
  - ⋆⋆ $k = 0.1$: $[-7.328, -5.124] \cup [1.591, 3.120]$ with posterior credibility of 64.2%;
  - ⋆⋆ $k = 0.2$: $[-6.893, -5.667]$ with posterior credibility of 31.3%.

- **Example** (contd)
  - ⋆⋆ Observe for $\alpha = 0.05$ and 0.1, the credible intervals consist of two separate intervals.
  - ⋆⋆ This may indicate that the prior is not agreeing with the data (unimodal in the prior vs bimodal in data).
  - ⋆⋆ There is no frequentist counterpart for the CI for θ in the above model.

- **Example** Let $x \mid \theta$ be the shifted exponential with density

$$f(x \mid \theta) = \exp\{-(x-\theta)\}1(\theta \le x).$$

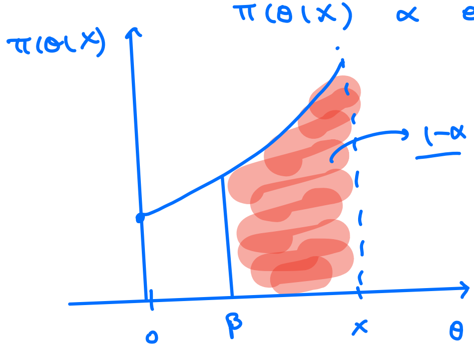
$$f(x|\theta) = e^{-x}\underbrace{e^{\theta}}_{=\eta} \cdot 1(\theta \le x) \qquad 1(\log(\eta) \le x)$$

Let θ be half-Cauchy,

$$\pi(\theta) = \frac{2}{\pi(1+\theta^2)}, \quad \overset{\log(\eta)}{\theta} > 0.$$

Find the posterior and show that $(1-\alpha)100\%$ HPD credible set is of the form $[\beta, x]$ for some $\beta \in (0, x)$.

$$\pi(\theta|x) \propto \exp(\theta) \cdot \frac{1}{(1+\theta^2)} \cdot 1(0 \le \theta \le x)$$



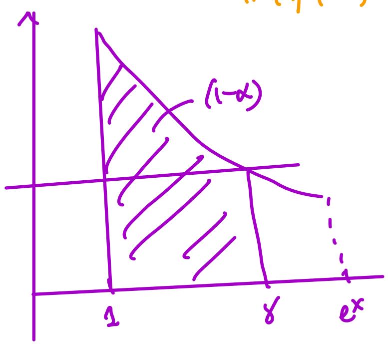
$$CI = [\beta, x]$$

$\theta = \log(\eta)$

- **Example** Let $\eta = e^{\theta}$ and find the posterior $\pi^\star(\eta \mid x)$. Show that $\pi^\star(\eta \mid x)$ is decreasing in η and that the credible set for η is of the form $[1, \gamma]$, for some $\gamma < e^x$.

$$\pi^*(\eta \mid x) \propto \underbrace{f(x \mid \eta)} \pi^*(\eta)$$

$$\cancel{\eta} \cdot 1(\log(\eta) \le x) \cdot \frac{2}{\pi(1+(\log\eta)^2)} \cdot \cancel{\frac{1}{\eta}} 1(\log\eta > 0)$$



$$\propto \frac{1}{(1+(\log\eta)^2)} \cdot 1(0 < \log\eta \le x)$$

$(1-\alpha)\,100\%$ CI $= [1, \gamma]$

$= (\log(1), \log(\gamma))$

$= 0, \log(\gamma)$

$\star\star$ Transform the interval of η back to the space of θ and observe $[\log 1, \log\gamma] = [0, \beta'] \neq [\beta, x]$.

- One undesirable property of credible sets is the lack of invariance with respect to monotone transformations.
- For a solution, read JB pages 144-145.
- A HPD credible sets can be found for multivariate cases. See JB p143

† Predictive Inference

- Predict a random variable $y \sim g(y \mid \theta)$ based on observations of $x \sim f(x \mid \theta)$.

  ⋆⋆ no need to be $g = f$

  ⋆⋆ easily can be extended to the case of $y \sim g(y \mid \theta, x)$.

- Find the predictive density of y given x, when the prior for θ is π,

$$p(y \mid x) = \int_{\Theta} g(y \mid \theta)\pi(\theta \mid x) d\theta.$$

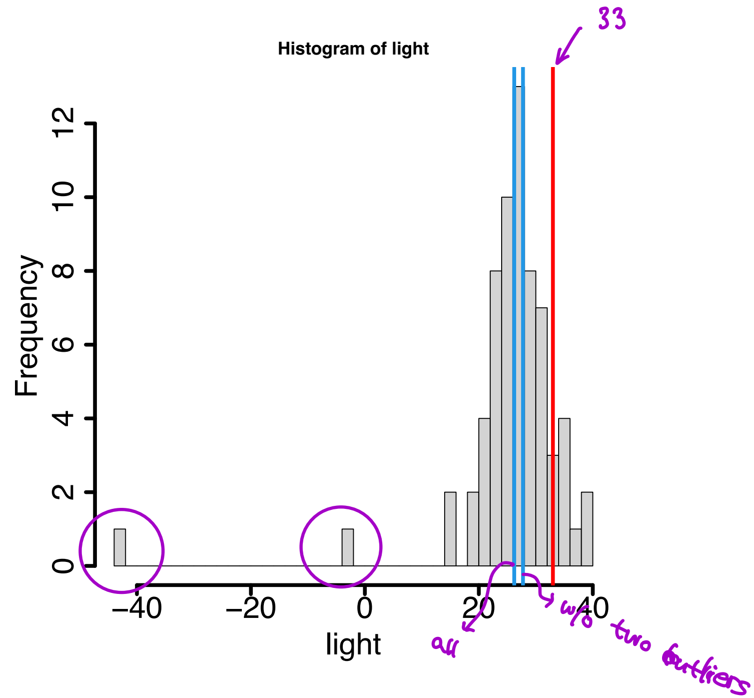
† Predictive Inference (contd)

- Point estimation: use the loss function and find the Bayes actions minimizing $\mathrm{E}(L(y,a) \mid x)$.

- Posterior predictive interval for y.

† Example: 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)

- Consider the normal model and assume that all 66 measurements are independent draws from $\mathrm{N}(\theta, \sigma^2)$.

  $\Leftrightarrow$ Assume $x_i \mid \theta, \sigma^2 \overset{iid}{\sim} \mathrm{N}(\theta, \sigma^2)$, $i = 1, \ldots, n$ with $n = 66$

  $\Rightarrow$ inferential goal: posterior inference for θ (so σ^2 is a nuisance parameter)

- Build a prior model for unknown random model parameters θ and σ^2.

  $\Leftrightarrow$ Consider a semi-conjugate prior distribution and let

$$\theta \sim \mathrm{N}(\mu, \tau^2) \quad \text{and} \quad \sigma^2 \sim \mathrm{IG}(a_0, b_0),$$

where μ, τ^2, a_0 and b_0 are fixed.

HW1 - Q11 - 1

$$\pi_1(\theta|\sigma^2) = N(\mu, \sigma^2/n_0)$$

$$\pi_2(\sigma^2) = IG(\nu/2, S_0^2/2)$$

$$\Rightarrow \pi(\theta, \sigma^2) = \pi_1(\theta|\sigma^2)\, \pi_2(\sigma^2)$$

$$\pi_1(\theta|\sigma^2, x) = N$$

$$\pi_2(\sigma^2|x) = IG$$

$$\pi(\theta, \sigma^2|x) \propto \overset{GG}{\underset{\|}{\prod_{i=1}^{n}}} \underbrace{f(x_i|\theta, \sigma^2)}_{N(x_i|\theta,\sigma^2)} \cdot \overbrace{\pi(\theta, \sigma^2)}^{= \overbrace{\pi_1(\theta)}^{N(\mu,\tau^2)} \overbrace{\pi_2(\sigma^2)}^{\to IG(a_0, b_0)}}$$

$$\propto (\sigma^2)^{-n/2} \cdot \exp\left(-\frac{s^2}{2\sigma^2} - \frac{n(\theta-\bar{x})^2}{2\sigma^2}\right)$$

$$\times \exp\left(-\frac{(\theta-\mu)^2}{2\tau^2}\right) \cdot (\sigma^2)^{-a_0-1} \exp\left(-\frac{b_0}{\sigma^2}\right)$$

① $$p(\theta|x, \sigma^2) \propto \exp\left[-\frac{1}{2}\left\{\left(\frac{n}{\sigma^2}+\frac{1}{\tau^2}\right)\theta^2 - 2\left(\frac{n\bar{x}}{\sigma^2}+\frac{\mu}{\tau^2}\right)\theta\right\}\right]$$

$$\Rightarrow \theta|x, \sigma^2 \sim N\left(\left(\frac{n}{\sigma^2}+\frac{1}{\tau^2}\right)^{-1}\left(\frac{n\bar{x}}{\sigma^2}+\frac{\mu}{\tau^2}\right), \left(\frac{n}{\sigma^2}+\frac{1}{\tau^2}\right)^{-1}\right)$$

② $$p(\sigma^2|x, \theta) \propto (\sigma^2)^{-(n/2+a_0)-1} \exp\left(-\frac{1}{\sigma^2}\left(\frac{s^2}{2} + \frac{n(\theta-\bar{x})^2}{2} + b_0\right)\right)$$

$$\Rightarrow \sigma^2|\theta, x \sim IG\left(\frac{n}{2}+a_0, \frac{s^2}{2} + \frac{n(\theta-\bar{x})^2}{2} + b_0\right)$$

† Example: Estimating the speed of light (contd)

- Use prior information and specify the values of the fixed hyper-parameter values.

```
> ## \theta ~ N(\mu, \tau2)
> hyper$mu <- 33
> hyper$tau2 <- 100
>
> ## \sig2 ~ IG(a0, b0)
> hyper$a0 <- 0.1
> hyper$b0 <- 0.1
```

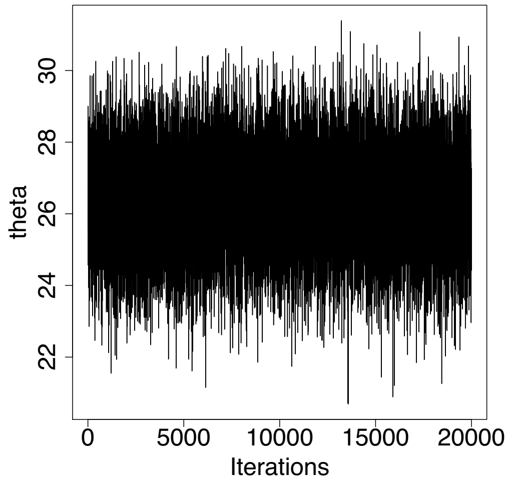
- Find the joint posterior distribution of all random parameters θ and σ^2.

- Find the posterior computation strategy.

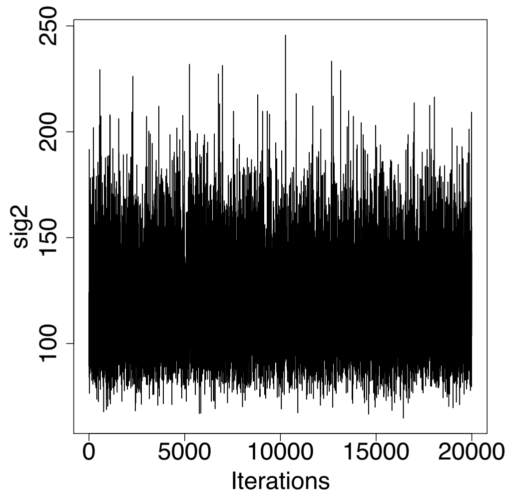
  ⋆⋆ Use the Gibbs sampler and derive the full conditional distributions.

† Example: Estimating the speed of light (contd)

- Check mixing and convergence of the Markov chain.



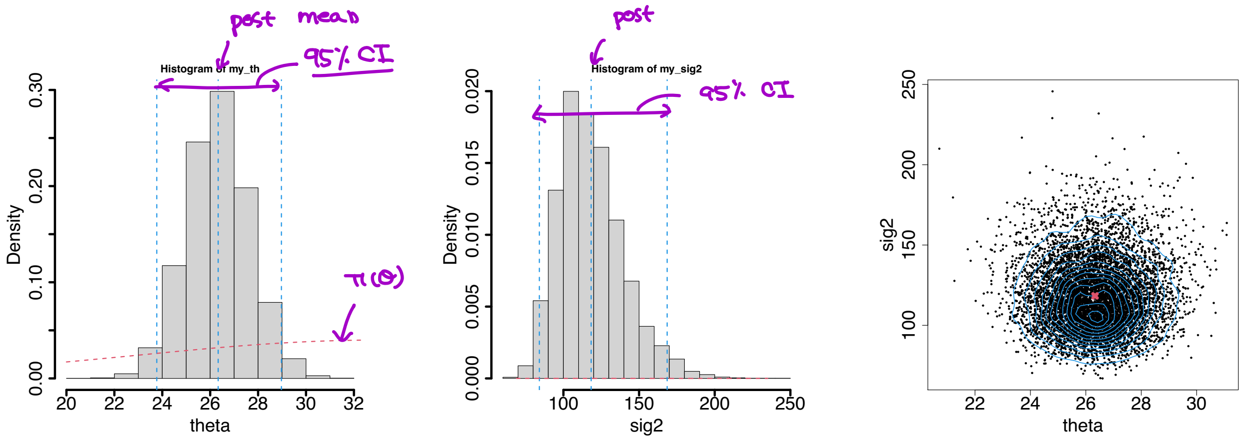
(a) θ



(b) σ^2

† 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)

- Posterior summary of θ and σ^2 (think about the implied loss function!)

```
> ### 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)

- Summary of the posterior predictive distribution of unobserved y

$$y^{(t)} \sim N(\underline{\theta^{(t)}}, \underline{\sigma^{2(t)}})$$

```
> ###%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
> #%%%## 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
>
```