20.4 Combined Bayes-Frequentist Estimation

May 29, 2018

PRESENTER : JongYun Kim

UNIST Autonomous System LAB
Address. 112-#810, 50 UNIST-gil, Ulju-gun, Ulsan, 44919, South Korea
Tel. +82 52 217 2368 **Web.** https://sites.google.com/site/aslunist/

Introduction

Bayesian estimates

- Immune from selection bias
- A data-based model selection rule has no effect on the likelihood function or posterior distribution

...

In high dimensions

- To set proper prior becomes difficult and possibly dangerous in higher dimensions
- A chosen prior has to apply to the entire parameters and not just the part we are interested in

Introduction to a Bayes-frequentist estimation technique like Tweedie's rule in Empirical Bayes methods

Problem Definition

First, suppose $\mathcal{F} = \{f_\alpha(x)\}$ is a multi-dimensional family of $\mathcal{F} = \{f_\mu(x); x \in \mathcal{X}, \mu \in \Omega\}$

Note that we now have different notations.

We want to estimate

$$\theta = t(\alpha)$$

A prior $g(\alpha)$ yields a posterior expectation as follows

$$\hat{\theta} = E\{t(\alpha)|x\}$$

$$\Downarrow$$

Q : How accurate is $\widehat{\theta}$?

- If we believe the prior, the posterior distribution has the exact answer.
- But, the prior reflects mathematical convenience and a desire to be uninformative in most cases of high Dim.
- There's a danger of circular reasoning in using a self-selected prior distribution to compute the accuracy of its own estimator

An Alternative Way : Frequentist Accuracy

$$\hat{\theta} = E\{t(\alpha)|x\}$$ ➡ **Q : How accurate is $\widehat{\theta}$?**

Calculating frequentist accuracy of $\widehat{\theta}$ might be applicable.

Note : Although $\hat{\theta}$ is a Bayes estimate, we consider $\hat{\theta}$ is a function of x . And then suppose that the prior is unavailable or uncertain in order to put it on frequentist calculations.

Let the family be a *p*-parameter exponential family. And then we get $f_\alpha(x) = e^{\alpha' x - \psi(\alpha)} f_0(x)$

Now we obtain the frequentist accuracy at the following theorem.

Theorem 20.4 $$\widehat{\mathrm{se}}_{\mathrm{delta}}\left\{\hat{\theta}\right\} = \left(\mathrm{Cov}_x' \, V_{\hat{\alpha}} \, \mathrm{Cov}_x\right)^{1/2}$$

Where, $V_{\hat{\alpha}}$ *is* V_α *evaluated at the MLE* $\hat{\alpha}$

with given $V_\alpha = \mathrm{cov}_\alpha(x)$: $p \times p$ covariance matrix of x

$\mathrm{Cov}_x = \mathrm{cov}\{\alpha, t(\alpha)|x\}$: the posterior covariance given x between θ and α

Bayesian Accuracy

Now, we are going to compute Bayesian accuracy and compare the two accuracies.

In order to generate posterior distribution of α given x, suppose we've employed an MCMC or Gibbs sampling algorithm : $\alpha^{(1)}, \alpha^{(2)}, \ldots, \alpha^{(B)}$

The usual estimate for the expectation of θ given x is

$$\hat{\theta} = \frac{1}{B} \sum_{b=1}^{B} t\left(\alpha^{(b)}\right)$$

Then the Bayesian accuracy is calculated as follows. (ignore the i notation for now..)

$$\widehat{\mathrm{se}}_{\mathrm{Bayes}}\left(\hat{\theta}_i\right) = \left[\frac{1}{B} \sum_{b=1}^{B} \left(\theta_i - \hat{\theta}_i\right)^2\right]^{1/2}$$

Comparison with an Example

To get the two accuracies, we consider the `diabetes` data of Section 20.1.

$x_i{}'$: the transpose of the vector of ith patient the ith raw of $\boldsymbol{X}$

$\boldsymbol{X}$: the 442×10 matrix of predictions

y : response vector of progression scores (Note : rescaled to have variance 1 in the normal regression model)

$$\boldsymbol{y} \sim \mathcal{N}_n(\boldsymbol{X}\beta, \boldsymbol{I})$$

The prior is given : $g(\beta) = ce^{-\lambda\|\beta\|_1}$

Note that $B = 10000$ samples for an MCMC algorithm.

A similar expression of covariance between θ and α is given :

$$\mathrm{Cov}_x = \frac{1}{B}\sum_{b=1}^{B}\left(\alpha^{(b)} - \alpha^{(\cdot)}\right)\left(t^{(b)} - t^{(\cdot)}\right)$$

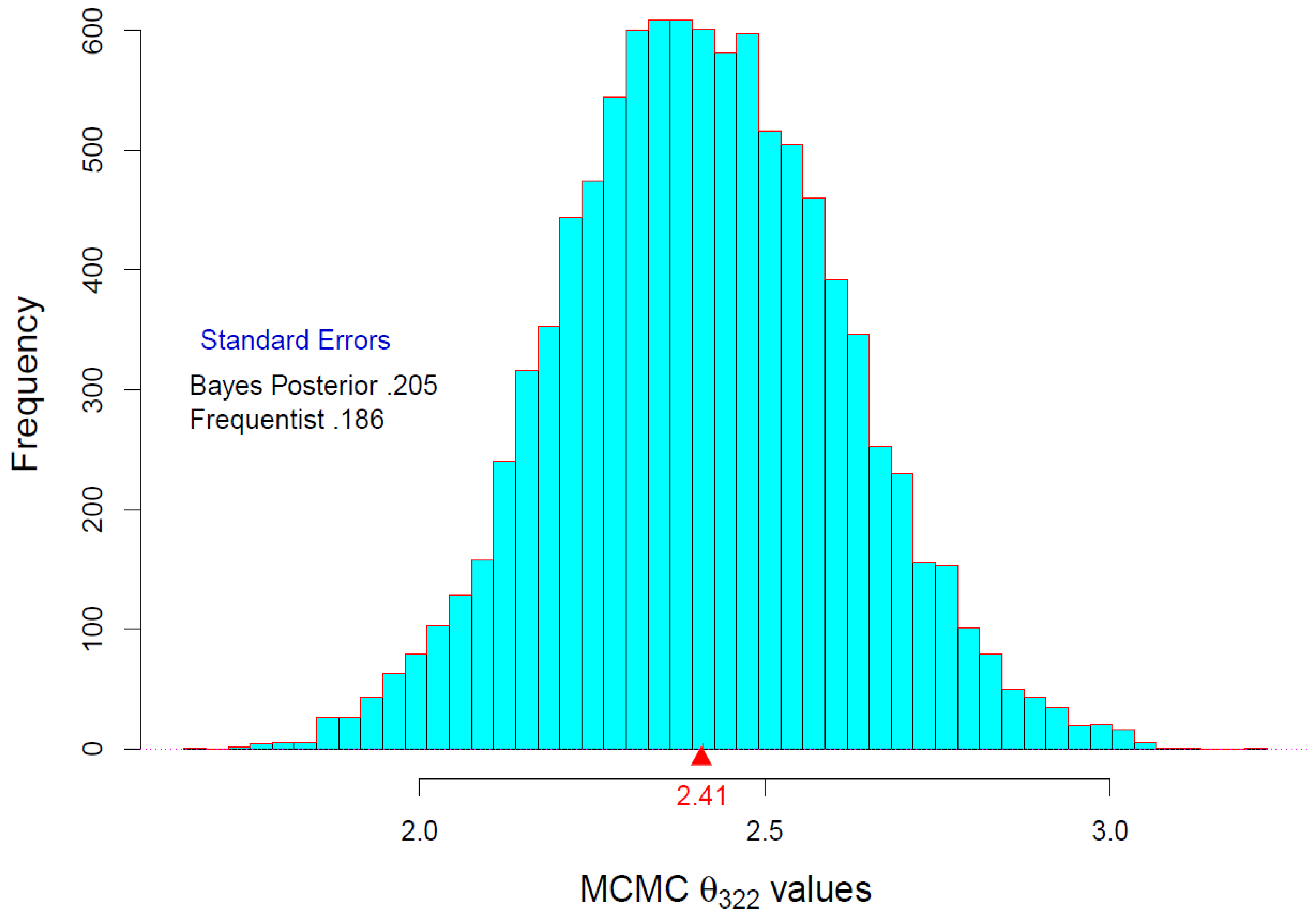
where, $t^{(b)} = t(\alpha^{(b)})$, $t^{(\cdot)} = \sum_b t^{(b)}/B$, and $\alpha^{(\cdot)} = \sum_b \alpha^{(b)}/B$

Comparison with an Example(1)

- The point estimate $\hat{\theta}_i$ equaled 2.41
- Bayes and frequentist standard error estimates are as follows

$$\widehat{se}_{Bayes} = 0.203 \quad and \quad \widehat{se}_{delta} = 0.186$$

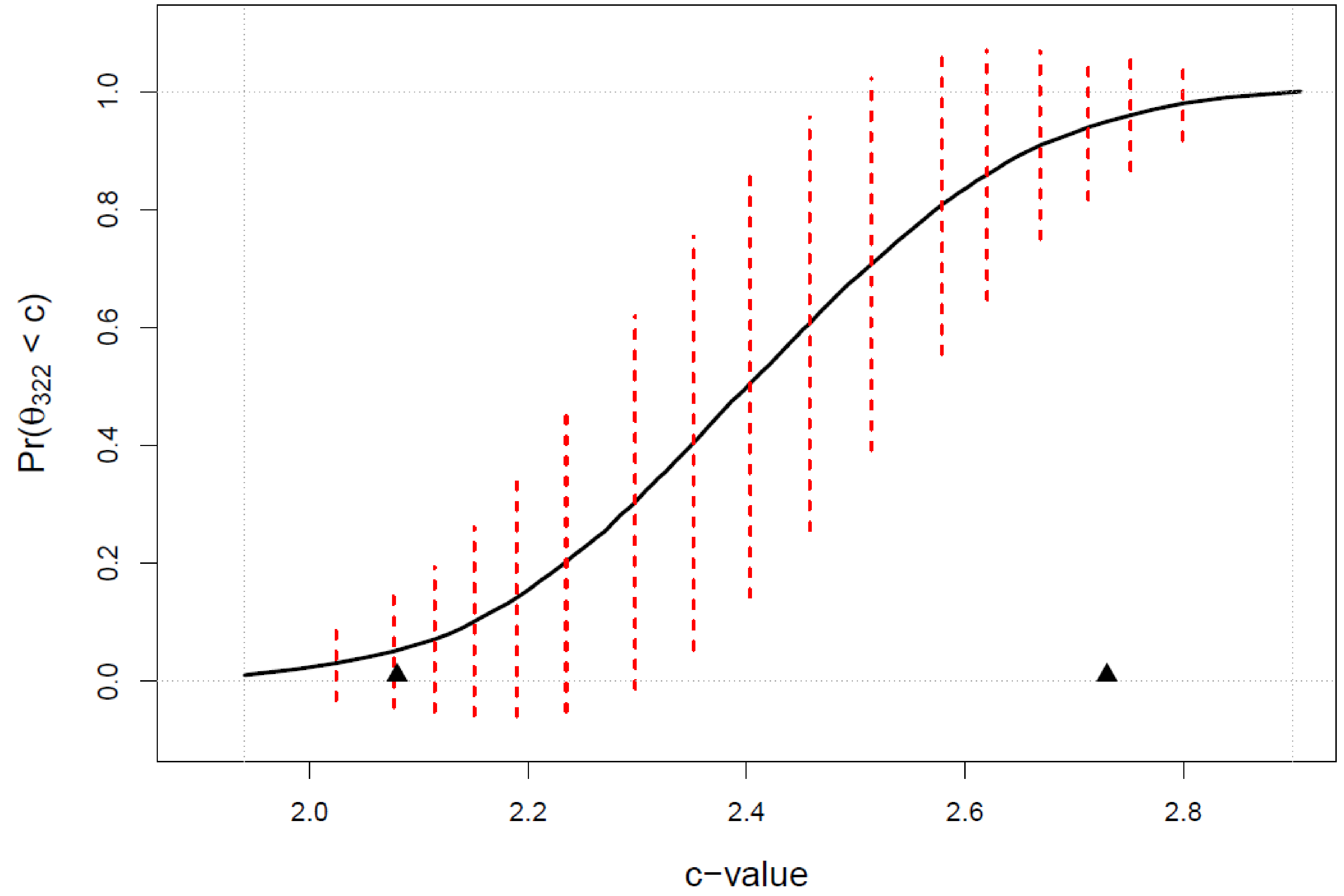
- The figure shows the 10,000 MCCM replications for $\hat{\theta}_i^{(b)} = x_i'\beta$ for patient $i = 322$.
- The frequentist standard error is 9% smaller in this case; even smaller for all 422 patients. (averagely 5%)



Comparison with an Example(2)

Let's consider cdf of θ_{322} given y, when $t\left(c,\beta^{(b)}\right)=\begin{cases}1 & \text{if } x_{322}'\beta^{(b)}\le c\\ 0 & \text{if } x_{332}'\beta^{(b)}> c\end{cases}$

$$\mathrm{cdf}(c)=\frac{1}{B}\sum_{b=1}^{B} t\left(c,\beta^{(b)}\right)$$

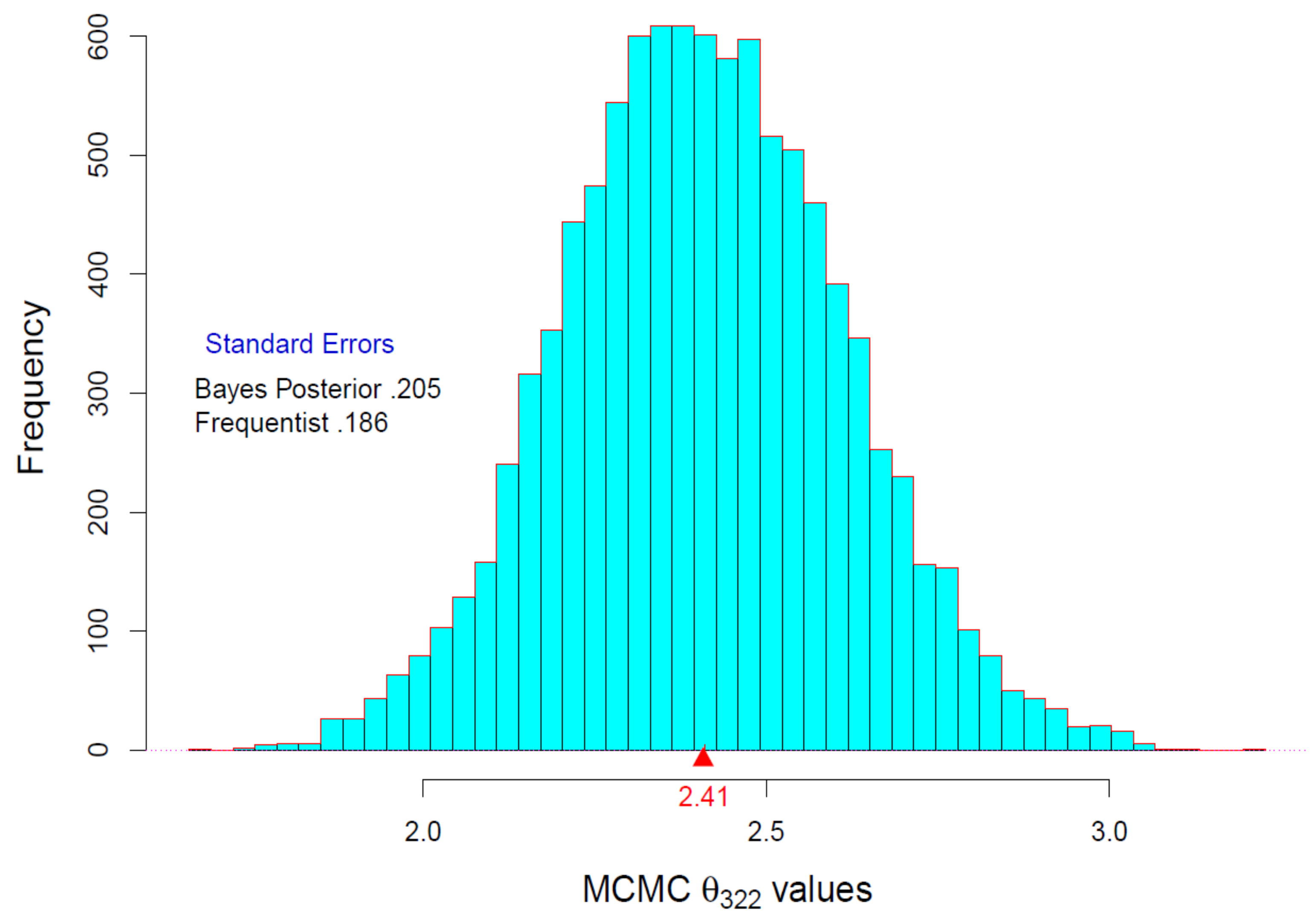


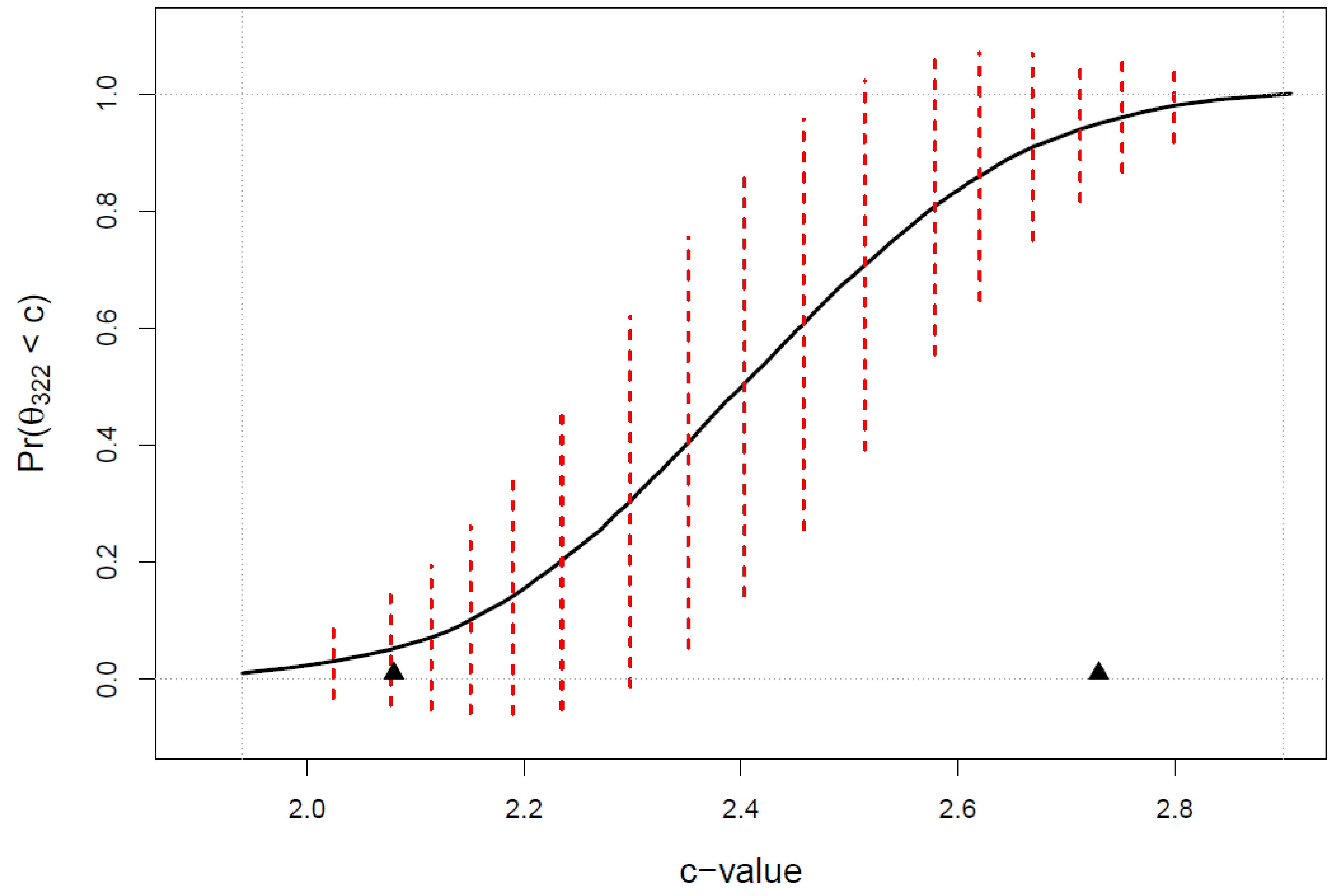
The solid curve is the posterior cdf of θ_{322}. Vertical red bars indicate $\pm$ one frequentist standard error, as obtained from Theorem 20.4. Black triangles are endpoints of the 0.90 central credible interval.

Regardless of belief of the prior, Thm.20.4 is available.

- se $= 0.687 \pm 0.325$ at $c = 2.5$.
- The central 90% credible interval is (2.08, 2.73)
- The interval has standard errors about 0.185 for each end point (28% of the interval length)
- In a new study, the result might vary much, even ignoring selection bias

Conclusion





- Bayesian calculations encourage a disregard for model selection effects. This can be dangerous in objective Bayes settings where one can't rely on genuine prior experience.
- Theorem 20.4 serves as a frequentist checkpoint, offering some reassurance as in the left figure, or sounding a warning as in the right figure.

THANK YOU

Q&A