Decision Theory & Bayesian Inference

R Project No. 2

Marie Ternes, Patrick Finke and Xiaotong Meng

January 28, 2020

Overview

Outline

Probability model for passing speed

- To study the safety of pedestrians, we aim to model the passing speed of the cars at a yellow traffic light by a gamma distributed variable $V \sim \mathcal{G}(\alpha, \beta)$.
- From studies on traffic behavior, we can assume that $\alpha = 0.55$ and then estimate β by Bayes estimator under LINEX loss, where two different prior distribution are proposed.

Outline

Simulate random sampling from uniform distribution

We implement the Linear Conguential Generator

$$x_{i+1} = (ax_i + b) \bmod m,$$

where m is the modulus, b the step size, a is the multiplier, to generate the random number. In particular we choose the values

$$a = 171, b = 0 \text{ and } m = 30269.$$

Then we simulate the random sample from $\mathcal{U}(0, 1)$ as

$$u_i = \frac{x_i}{m},$$

Performance: slower than the built-in one by a factor of about 10.50, because of R's poor performance when using loops

Approximation of π

We approximate π using the coin-toss experiment and the relation

$$\frac{\#(\text{tosses in the circle})}{\#(\text{total tosses})} = \frac{\pi}{4} \quad \Leftrightarrow \quad \pi = \frac{4 \cdot \#(\text{tosses in the circle})}{\#(\text{total tosses})}.$$

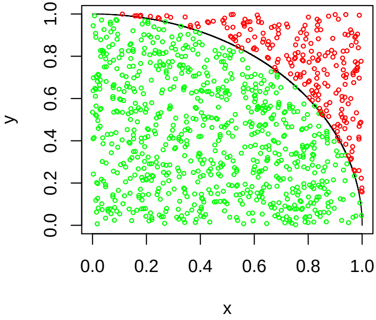
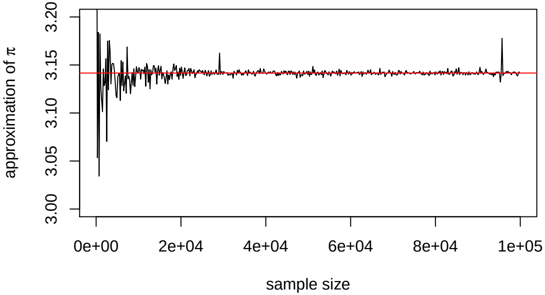


Figure: Coin-toss experiment with samples size n = 1000

Convergence of coin-toss



sample size	10	100	1,000	10,000	100,000	1,000,000
approximation	3.2000	3.1600	3.1120	3.1404	3.1409	3.1413
absolute error	0.0584	0.0184	0.0296	0.0012	0.0007	0.0002

Outline

Generating Random Samples from Posterior Distributions

- $n = 50$ observations of measured velocity (in mph) passing yellow traffic light
- $V_1, ..., V_n \overset{iid}{\sim} \mathcal{G}(\alpha, \beta)$
- Joint distribution / Likelihood:

$$\begin{aligned} f^V(y|\alpha,\beta) &= \prod_{i=1}^{n} \frac{\beta^\alpha}{\Gamma(\alpha)} y_i^{\alpha-1} e^{-\beta y_i} \\ &= \frac{\beta^{n\alpha}}{\Gamma(\alpha)^n} \Big(\prod_{i=1}^{n} y_i \Big)^{\alpha-1} e^{-\beta \sum_{i=1}^{n} y_i}. \end{aligned}$$

- $\alpha = 0.55$

Prior and Posterior Distributions (1)

Prior (1)

- $\beta \sim \mathcal{G}(2, 0.2)$, with density

$$\pi_1(\beta) = 0.04 \cdot \beta \; e^{-0.2\beta}, \quad \beta > 0.$$

Posterior (1)

- $\pi_1(\beta|y) \propto \beta^{n\alpha+1} \; e^{-(\sum_{i=1}^n y_i + 0.2)\beta}$
- Thus, $\beta|y \sim \mathcal{G}(n\alpha + 2, \sum_{i=1}^n y_i + 0.2) = \mathcal{G}(29.5, 755.15)$.

Prior and Posterior Distributions (2)

Prior (2)

- β follows a distribution given by the continuous p.d.f

$$\pi_2(\beta) \propto e^{-0.2\beta}\left(4\,\sin(2\beta+1)^2 + 1 + \frac{\beta e^{-15(\beta-15)}}{2(1+e^{-15(\beta-15)})^2}\right), \quad \beta > 0.$$

Posterior (2)

- $\pi_2(\beta|y) \propto \beta^{n\alpha}\ e^{-(\sum_{i=1}^n y_i)\beta} \cdot e^{-0.2\beta}\left(4\,\sin(2\beta+1)^2+1+\frac{\beta e^{-15(\beta-15)}}{2(1+e^{-15(\beta-15)})^2}\right)$
- Normalizing constant is not known.

Rejection Method

Rejection Method Algorithm

1. Generate Y from $g(y)$: $Y = y$
2. Generate $U \sim \mathcal{U}(0, 1)$: $U = u$
3. If $uMg(y) \leq f(y)$, set $\beta = y$;
   else return to step 1

- Our distribution "g" is the truncated **(at 0)** two-parameter Laplace (λ, μ) distribution

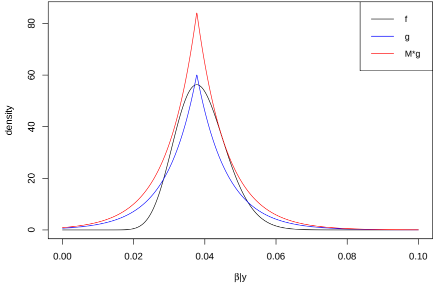
$$\widetilde{g}(y|\lambda, \mu) = \frac{\frac{\lambda}{2} e^{-\lambda|y-\mu|}}{1 - G(0)}, \qquad y > 0,$$

  where $G(y)$ is the cdf of the Laplace (λ, μ) distribution.
- f is either posterior $\pi_1(\beta|y)$ or $\pi_{2u}(\beta|y)$

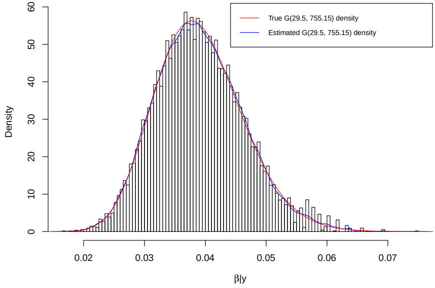
Inverse Function

$$\widetilde{G}^{-1}(u|\lambda,\mu) = \begin{cases} \mu + \frac{1}{\lambda}\left[\ln\{2(1-G(0))u + e^{-\lambda\mu}\}\right] & \text{if } u < \frac{0.5-0.5e^{-\lambda\mu}}{1-G(0)} \\ \mu - \frac{1}{\lambda}\left[\ln\{2-2(1-G(0))u - e^{-\lambda\mu}\}\right] & \text{else} \end{cases}$$

Under Prior (1): Rejection Method

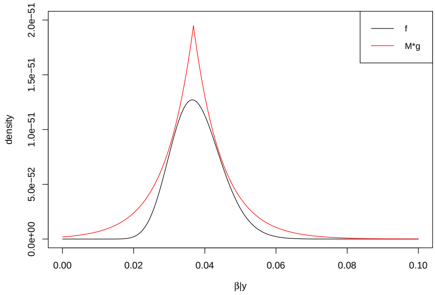


(a) Plot of $f \sim \mathcal{G}(29.5, 755.15)$, $g \sim$ trunc. Laplace$(120, 0.0377)$ and Mg with $M = 1.4$.

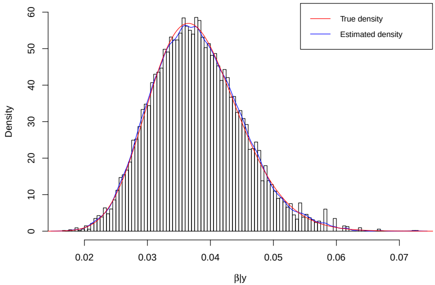


(b) Histogram of 100,000 random samples, estimated density and true density

Under Prior (2): Rejection Method



(a) Plot of $f = \pi_{2u}(\beta|y)$ and $Mg = M\widetilde{g}(y|\lambda, \mu)$ with $M = 3.1 \times 10^{-53}$ and trunc. Laplace $(125, 0.0368)$.



(b) Histogram of 100,000 random samples, estimated density and true density

Comparison Posterior Distribution

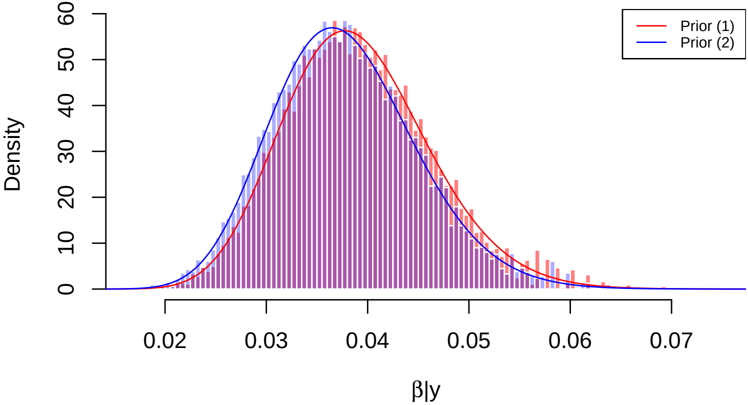


Figure: Histograms and true posterior densities under the Gamma prior (1) and the second prior (2).

Comparison Posterior Distributions

	Prior (1)	Prior (2)
$\widetilde{g}(y\|\lambda, \mu)$	$\lambda = 120,\ \mu = 0.0377$	$\lambda = 125,\ \mu = 0.0368$
M	1.4	3.1×10^{-53}
C	1.000036	4.479645×10^{52}
Acceptance rate	71.43%	72.01%
Number of simulations	100,000	100,000

- In shape, both posteriors are very similar to each other.
- Posterior density of the second prior is shifted slightly to the left in comparison to the first prior.
- It seems like that the influence of the different priors on the posteriors is not too strong.

Outline

Bayes Estimators

LINEX Loss:

$$L_c(d \mid y) = e^{c(\beta - d)} - c(\beta - d) - 1$$
$$\text{where } \beta, d > 0,\ 0 \neq c \in \{\pm 5, \pm 15\}$$

Posterior Expected Loss:

$$r_\pi^*(d \mid y) = \mathbb{E}[e^{c(\beta - d)} - c(\beta - d) - 1 \mid y]$$

Bayes Estimator d^*:

$$r_\pi^*(d^* \mid y) = \inf_d r_\pi^*(d \mid y)$$

$$0 \overset{!}{=} \frac{\partial}{\partial d} r_\pi^*(d \mid y) = \mathbb{E}[-c e^{c(\beta - d)} + c \mid y]$$
$$\Leftrightarrow d^* = \frac{1}{c} \log(\mathbb{E}[e^{c\beta} \mid y])$$

$$\frac{\partial^2}{\partial d^2} r_\pi^*(d \mid y) = \mathbb{E}[c^2 e^{c(\beta - d)} \mid y] > 0,\ c \neq 0$$

Bayes Estimators - Posterior (1)

Posterior under Prior (1):

$$\beta \mid y \sim \mathrm{G}(\alpha' = 29.5,\ \beta' = 755.15)$$

Exact Bayes Estimator:

$$\mathbb{E}[e^{c\beta} \mid y] = (1 - \frac{c}{\beta'})^{-\alpha'},\ \ c < \beta'$$

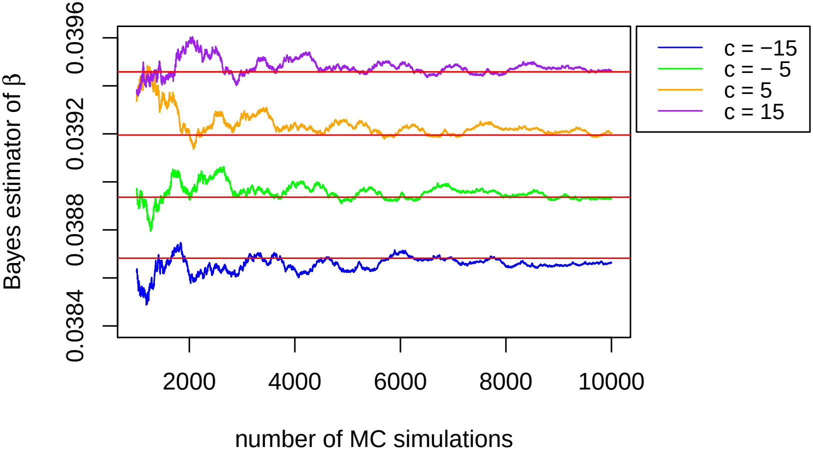
$$\Rightarrow d^* = -\frac{\alpha'}{c} \log(1 - \frac{c}{\beta'})$$

Bayes Estimator by MC integration:

$$\mathbb{E}[c^{e\beta} \mid y] = \int_0^\infty e^{c\beta} \pi_1(\beta \mid y)\, \mathrm{d}\beta \approx \frac{1}{n} \sum_{i=1}^{n} e^{c\beta_i}$$

where $\beta_i \sim \pi_1(\,\cdot \mid y)$ and $n = 10,000$

Bayes Estimators - Posterior (1) Convergence



Bayes Estimators - Posterior (2)

Posterior under Prior (2):

$$\beta \mid y \sim \pi_2(\beta \mid y) \propto \beta^{n\alpha} e^{-(\sum_{i=1}^n y_i + 0.2)\beta} \left(4\sin(2\beta+1)^2 + 1 + \frac{\beta e^{-15(\beta-15)}}{2(1+e^{-15(\beta-15)})^2}\right)$$

Bayes Estimator by Importance Sampling:

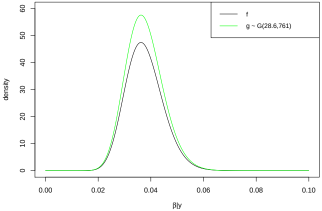
1. generate $\beta_1, \ldots, \beta_n$ from $g \sim G(\widetilde{\alpha}, \widetilde{\beta})$, $n = 10,000$
2. compute weights $w(\beta_i) = \frac{e^{c\beta_i}\pi_2(\beta_i|y)}{g(\beta_i)}$
3. approximate
   $\mathbb{E}_{\pi_2}[e^{c\beta} \mid y] = \int_0^\infty 1 \cdot e^{c\beta}\pi_2(\beta \mid y)\,\mathrm{d}\beta \approx \frac{1}{n}\sum_{i=1}^n 1 \cdot \frac{e^{c\beta_i}\pi_2(\beta_i|y)}{g(\beta_i)}$

Bayes Estimator by MC integration:

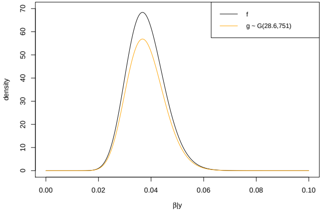
$$\mathbb{E}[e^{c\beta} \mid y] = \int_0^\infty e^{c\beta}\pi_2(\beta \mid y)\,\mathrm{d}\beta \approx \frac{1}{n}\sum_{i=1}^n e^{c\beta_i}$$

where $\beta_i \sim \pi_2(\,\cdot \mid y)$ and $n = 10,000$

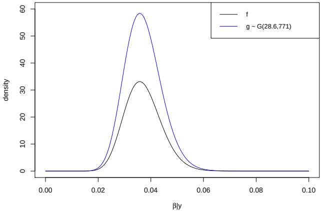
Bayes Estimators - Posterior (2) Importance Distribution



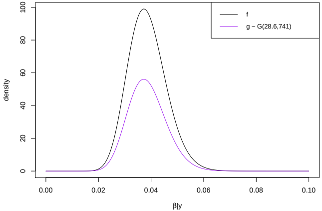
(a) c = -5



(b) c = 5



(c) c = -15



(d) c = 15

Bayes Estimators - Posterior (2) Importance Weights

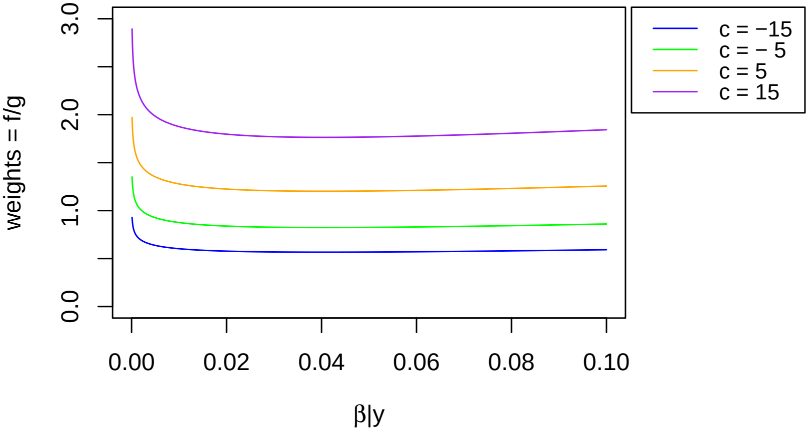
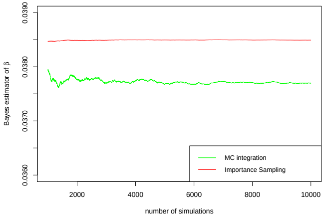
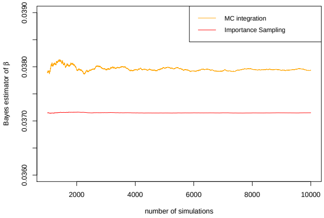


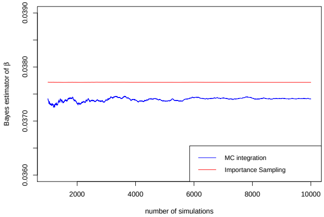
Figure: Importance weights $w(x) = \frac{f(x)}{g(x)}$.



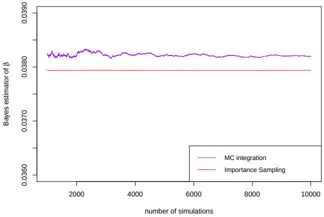
(a) c = -5



(b) c = 5

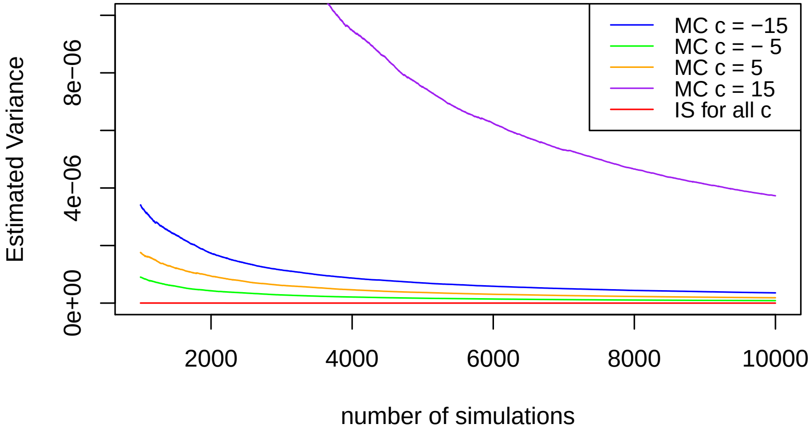


(c) c = -15



(d) c = 15

Bayes Estimators - Posterior (2) Variance



Bayes Estimators - CI and Bayes Estimators

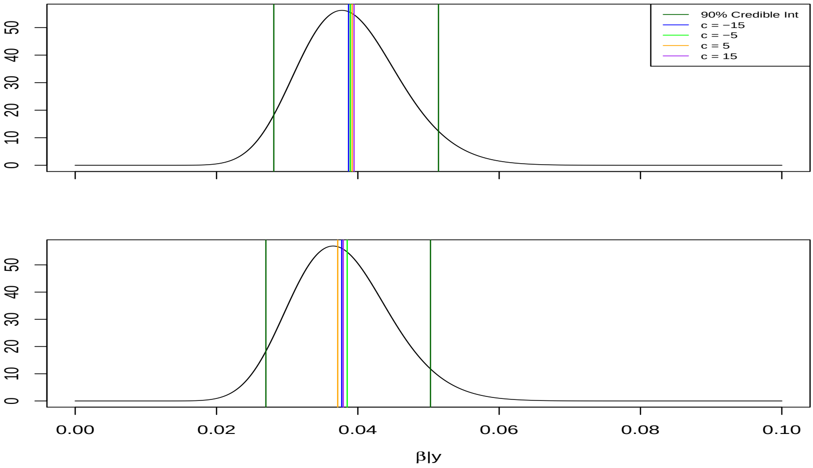


Figure: Posterior (1) top, Posterior (2) bottom

Outline

Conclusion

- The two posterior for β are very similar in shape and scale.
- The posterior density under the second prior is shifted slightly towards the left
    - Bayes estimators are slightly smaller
    - CI are shifted to the left (but same in length)
- The final velocity model will have similar characteristics regardless of the choice of the β prior and c, because of the slight difference.

Thank you for your attention