Solution to assignment 2

Problem 3.2 (Total 15')

The posterior distribution of θ, $P(\theta|\sum_{i=1}^{100} y_i = 57) = Beta(a+57, b+100-57)$. The contour plot is given in Figure 1(a). A high degree of certainty (95% or above) could only be achieved by high prior expectation (large value on θ_0). Lower values of $n_0(< 10)$ result in generally 70% or more certainty of the statement that $\theta > 0.5$. So someone would believe that $\theta > 0.5$, only if his or her prior expectation, θ_0 is large. For lower prior expectation, the certainty decreases as n_0 increases.

Problem 3.3 (Total 15')

(a) Since $P(\theta_A, \theta_B) = P(\theta_A)P(\theta_B)$ is assumed, θ_A and θ_B ban be considered separately.

- $\theta_A|y_A \sim \text{Gamma}(120 + \sum y_{A_i}, 10 + n_A) = \text{Gamma}(237, 20)$
- $\theta_B|y_B \sim \text{Gamma}(12 + \sum y_{B_i}, 10 + n_B) = \text{Gamma}(125, 14)$

$E(\theta_A) = 237/20 = 11.85, Var(\theta_A) = 237/20^2 = 0.5925$

$E(\theta_B) = 125/14 = 8.9286, Var(\theta_B) = 125/14^2 = 0.6378$

The 2.5% and 97.5% quantiles of θ_A are 10.38924 and 13.40545, hence 95% CI of θ_A is $[10.38924, 13.40545]$, similarly, that of θ_B is $[7.432064, 10.56031]$.

(b) $\theta_B|y_B \sim \text{Gamma}(12 \times n_0 + \sum y_{B_i}, n_0 + n_B) = \text{Gamma}(113 + 12n_0, 13 + n_0)$.

$E(\theta_B) = \frac{113+12n_0}{13+n_0}$, Plot of $(n_0, E(\theta_B))$ is given in Figure 1(b), from which we can see that a high prior belief (40 or above) is needed to make the posterior mean of θ_B and θ_A are close.

(c) Since "Type B mice are related to type A mice", knowledge about population A would help to improve our knowledge of population B. Hence independent prior is not a good choice based on this assumption.

Problem 3.9 (Total 15')

(a) Assume that prior of θ is $p(\theta)$, then posterior distribution of θ is

$$p(\theta|y) \propto p(\theta)p(y|\theta) \propto p(\theta)\theta^{2a}e^{-\theta^2 y^2}.$$

This means that the conjugate prior distribution have to include terms like $\theta^{c_1}e^{-c_2\theta^2}$. Actually, Galenshore distribution meets this kind of demand, so we take $p(\theta) = \text{Galenshore}(a, \theta)$. Plots of Galenshore distribution are given in Figure 1(c).

(b) Let $p(\theta) \propto \theta^{2k-1}e^{-m^2\theta^2}$, then

$$p(\theta|y_1,\cdots,y_n) \propto p(\theta)p(y_1,\cdots,y_n|\theta) \propto \theta^{2(na+k)-1}e^{-\theta^2(\sum_{i=1}^n y_i^2+m^2)}$$

Hence $\theta|y_1,\cdots,y_n \sim \text{Galenshore}(na+k, \sqrt{\sum_{i=1}^n y_i^2+m^2})$.

(c) From (b), we have $\frac{p(\theta_a|y_1,\cdots,y_n)}{p(\theta_b|y_1,\cdots,y_n)} = (\frac{\theta_a}{\theta_b})^{2na}e^{-\sum_{i=1}^n y_i^2(\theta_a^2-\theta_b^2)}\frac{p(\theta_a)}{p(\theta_b)}$, so $\sum_{i=1}^n y_i^2$ is a sufficient statistic.

(d) From (b), we have $E[\theta|y_1,\cdots,y_n] = \frac{\Gamma(na+k+0.5)}{\Gamma(na+k)\sqrt{\sum_{i=1}^n y_i^2+m^2}}$.

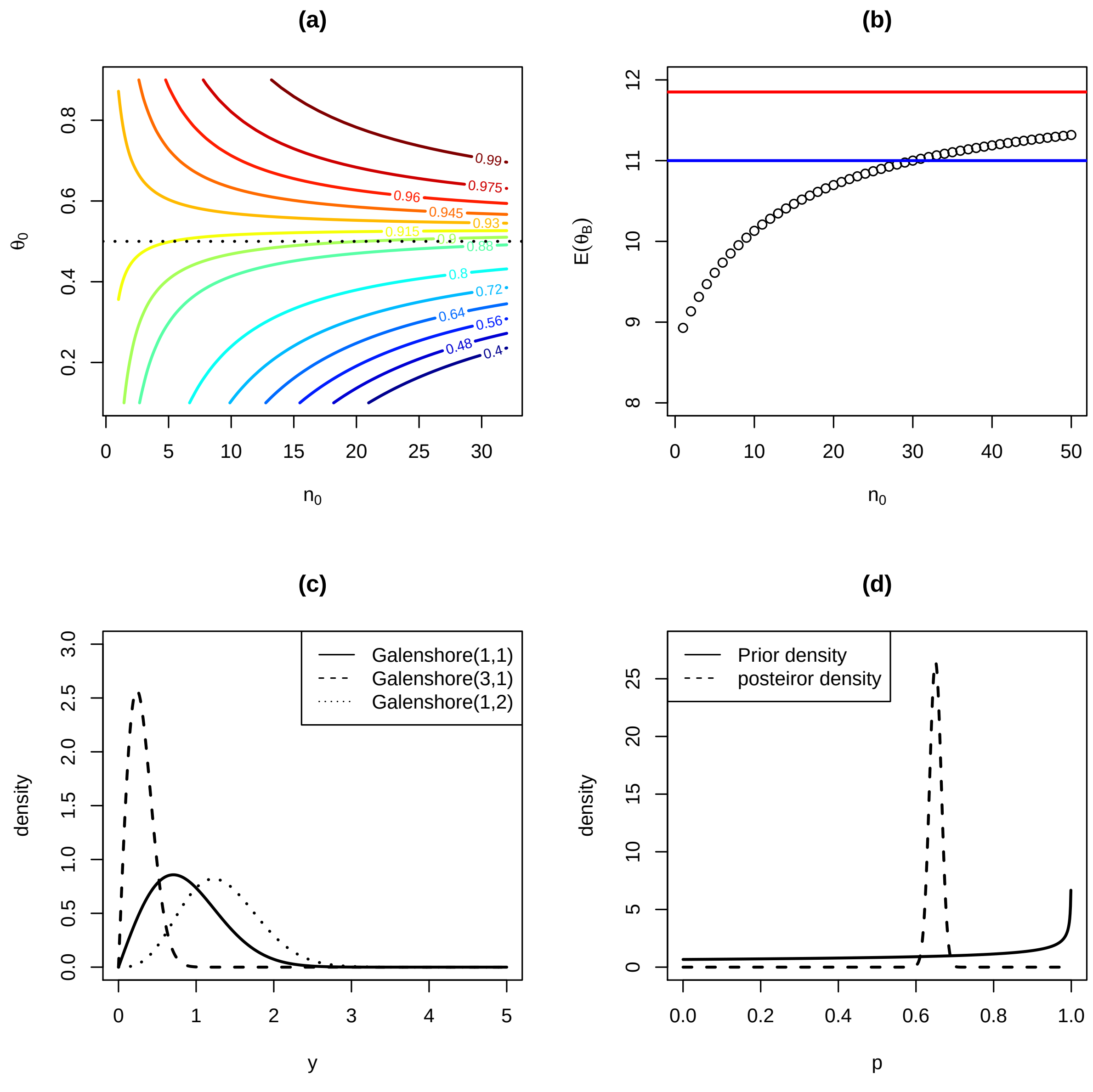

Figure 1: Plots in this assignment. (a) for problem 3.2; (b) for problem 3.3(b); (c) for problem 3.9(a); (d) for problem (1). In (b) the red line is the mean of θ_A.

(e) Let $d=\sqrt{\sum_{i=1}^n y_i^2+m^2}$,$c=na+k$ then

$$\begin{aligned}
p(\tilde{y}|y_1,\cdots,y_n) &= \int p(\tilde{y}|\theta)p(\theta||y_1,\cdots,y_n)d\theta \\
&= \int \frac{2}{\Gamma(a)}\theta^{2a}\tilde{y}^{2a-1}e^{-\theta^2\tilde{y}^2}\frac{2}{\Gamma(c)}d^{2c}\theta^{2c-1}e^{-d^2\theta^2}d\theta \\
&= \frac{4d^{2c}\tilde{y}^{2a-1}}{\Gamma(a)\Gamma(c)}\int \theta^{2(a+c)-1}e^{-\theta^2(\tilde{y}^2+d^2)}d\theta \\
&= \frac{4d^{2c}\tilde{y}^{2a-1}}{\Gamma(a)\Gamma(c)}\frac{\Gamma(a+c)}{2}(\tilde{y}^2+d^2)^{-(a+c)} \\
&= \frac{2\Gamma(a+c)}{\Gamma(a)\Gamma(c)}d^{2c}\tilde{y}^{2a-1}(\tilde{y}^2+d^2)^{-(a+c)}
\end{aligned}$$

Other five problems (Total 55', 15' for (4) and 10' for others)

(1) Set the prior distribution be beta(α,β) with mean 0.6 and sd 0.3, so, we have $\frac{\alpha}{\alpha+\beta}=0.6$, $\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}=0.3^2$, solving these two equations gives $\alpha=1,\beta=2/3$.
Because 650 of 1000 californian support the death penalty, then we get the data that $X=650$ successes in 1000 trials, hence the likelihood is $P(X=650|p,n=1000)=C_{1000}^{650}p^{650}(1-p)^{350}$.
Then $P(p|X=650)\propto P(p)P(X=650|p,n=1000)\propto p^{651-1}(1-p)^{1052/3-1}\sim \text{Beta}(651,1052/3)$.
The plots of prior density and posterior density are given in Figure 1(d).

(2) Assume coin C is chose, and the data is $X=2$ tails in $n=2$ trails.
$P(C=C_1|X=2)=\frac{P(C=C_1)P(X=2|C=C_1)}{P(X=2)}=\frac{0.4^2\times0.5}{0.4^2\times0.5+0.6^2\times0.5}=\frac{4}{13}$.
$P(C=C_2|X=2)=1-P(C=C_1|X=2)=\frac{9}{13}$.
Let N be the number of additional spins until a head shows up,

$$\begin{aligned}
E(N|X=2) &= \sum_{k=1}^{\infty}kP(N=k|X=2) \\
&= \sum_{k=1}^{\infty}k[P(N=k,C=C_1|X=2)+P(N=k,C=C_2|X=2)] \\
&= \sum_{k=1}^{\infty}k[P(N=k|C=C_1)P(C_1|X=2)+P(N=k|C=C_2)P(C_2|X=2)] \\
&= E[N|C_1]P(C_1|X=2)+E[N|C_2]P(C_2|X=2) \\
&= \frac{1}{0.6}\frac{4}{13}+\frac{1}{0.4}\frac{9}{13} \\
&= 175/78.
\end{aligned}$$

Note: $N|C_i(i=1,2)\sim$ Geometric distribution(p_i), with $p_1=0.6,p_2=0.4$.

(3a) The prior of θ is $N(180,40^2)$, and the likelihood function is

$$\begin{aligned}
f(x_1,\cdots,x_n|\theta) &= \prod_{i=1}^n f(x_i|\theta)=\prod_{i=1}^n\frac{1}{20\sqrt{2\pi}}e^{-\frac{1}{800}(x_i-\theta)^2} \\
&\propto e^{-\frac{n\bar{x}}{400}\theta-\frac{n}{800}\theta^2}
\end{aligned}$$

without considering the normalized constant, the likelihood function a function of θ is proportional to $N(\bar{x}, \frac{400}{n}) = N(150, \frac{400}{n})$.

Then, the posterior distribution.

$$\begin{aligned} f(\theta|x_1, \cdots, x_n) &= \prod_{i=1}^{n} f(x_i|\theta) f(\theta) \\ &= \prod_{i=1}^{n} \frac{1}{20\sqrt{2\pi}} e^{-\frac{1}{800}(x_i-\theta)^2} \frac{1}{40\sqrt{2\pi}} e^{-\frac{1}{3200}(\theta-180)^2} \\ &\propto e^{-(\frac{n}{800}+\frac{1}{3200})\theta^2 + 2(\frac{180}{3200}+\frac{150n}{800})\theta} \\ &\sim N(\frac{180+600n}{1+4n}, \frac{1600}{1+4n}) \end{aligned}$$

(3b) If $n = 10$, $f(\theta|x_1, \cdots, x_n) \sim N(150.732, 39.024)$. Hence the 95% posterior interval for θ is $150.732 \pm 1.96 \times \sqrt{39.024} = [138.488, 162.976]$.

(3c) Plots are given in Figure 2. We can find that the prior is much flatter than the posterior, because the standard deviation of the prior is much bigger than that of posterior, as the posterior is a combination of information from the prior and likelihood. Besides, as n increases, the posterior and likelihood function get closer, since the information comes from likelihood dominates that from the prior.

(4a) It is a multinomial distribution, which is a generalization of the binomial distribution. The likelihood function is given by

$f(n_1 = 10, n_2 = 10, n_3 = 10, n_4 = 20, n_5 = 10, n_6 = 40|p_1, \cdots, p_6, N = 100)$
$= \frac{100!}{10!10!10!20!10!40!} p_1^{10} p_2^{10} p_3^{10} p_4^{20} p_5^{10} p_6^{40}$

(4b) Assume the prior of $\{p_1, \cdots, p_6\}$ is $f(p_1, \cdots, p_6)$, then the posterior distribution of $\{p_1, \cdots, p_6\}$ is

$$\begin{aligned} f(p_1, \cdots, p_6|x_1, \cdots, x_100) &\propto f(x_1, \cdots, x_{100}|p_1, \cdots, p_6) f(p_1, \cdots, p_6) \\ &\propto p_1^{10} p_2^{10} p_3^{10} p_4^{20} p_5^{10} p_6^{40} P(p_1, \cdots, p_6) \end{aligned}$$

This means that the conjugate prior distribution has to include terms like $\prod_{i=1}^{6} p_i^{a_i}$, which suggests us take Dirichlet distribution as the prior distribution. Hence $f(p_1, \cdots, p_6) = \text{Dir}(\alpha)$, where $\alpha = (\alpha_1, \cdots, \alpha_6) > 0$, that is, $f(p_1, \cdots, p_6) = \frac{\Gamma(\sum_{i=1}^{6}\alpha_i)}{\prod_{i=1}^{6}\Gamma(\alpha_i)} \prod_{i=1}^{6} p_i^{a_i - 1}$.

(4c) Since we do not have any prior information about p, so we take $f(p_1, \cdots, p_6) = 1$, that is the Dirichlet distribution with $\alpha = (1, 1, 1, 1, 1, 1)$.

(4d) From (b), we have $f(p_1, \cdots, p_6|x_1, \cdots, x_{100}) \propto p_1^{10} p_2^{10} p_3^{10} p_4^{20} p_5^{10} p_6^{40} \sim \text{Dir}(\alpha)$, with $\alpha = (11, 11, 11, 21, 11, 41)$.

(4e) The posterior point estimate could be the posterior mean, and it is $(11, 11, 11, 21, 11, 41)/106 = (0.1037736, 0.1037736, 0.1037736, 0.1981132, 0.1037736, 0.3867925)$, and the corresponding uncertainty is the posterior variance:$(1045, 1045, 1045, 1785, 1045, 2665)/(107 \times 106^2) = (0.0008692021, 0.0008692021, 0.0008692021, 0.0014847137, 0.0008692021, 0.0022166734)$.

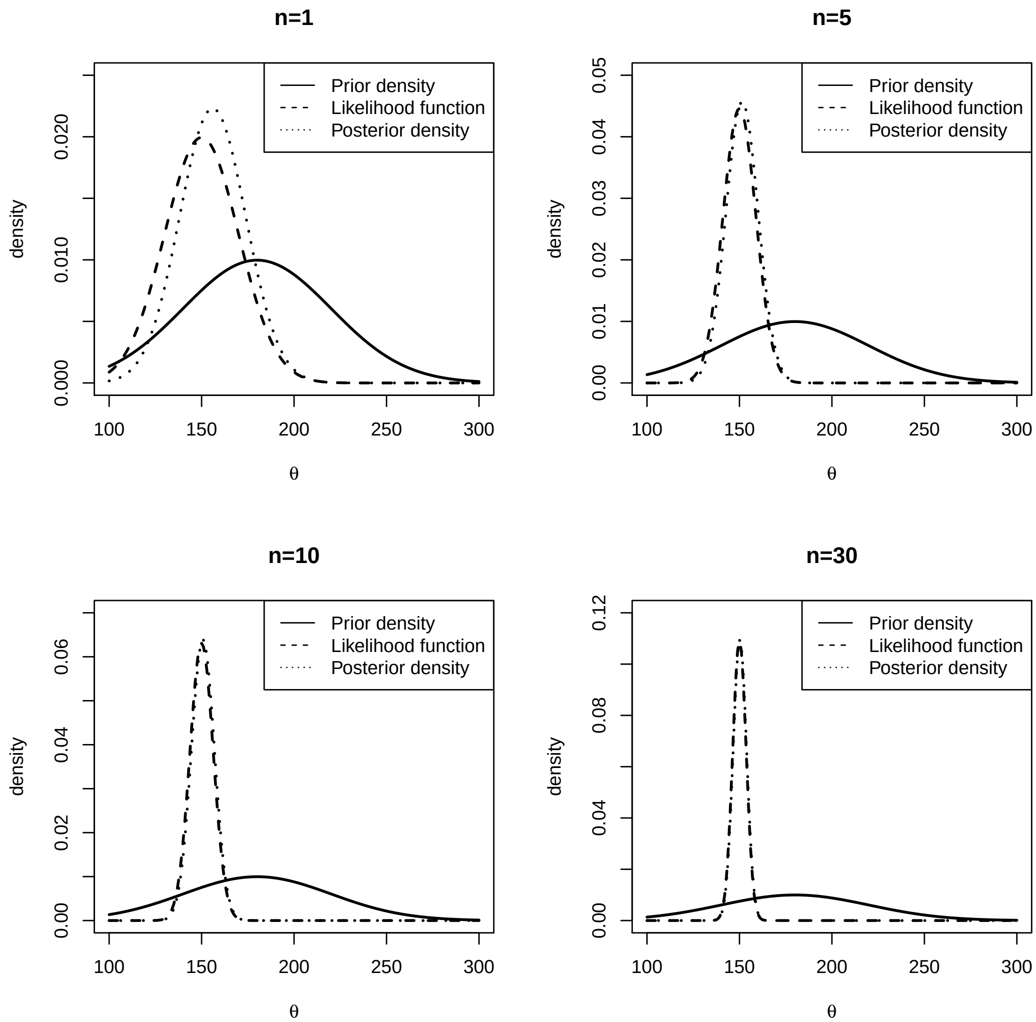

Figure 2: Prior density, Likelihood function and Posterior density in Problem (3c).

(5a) Since we assume a uniform prior on $(\mu_c, \log\sigma_c)$, the joint prior of (μ_c, σ_c^2) is $\propto \frac{1}{\sigma_c^2}$, and the likelihood function is

$$\begin{aligned} f(x_1,\cdots,x_n|\mu_c,\sigma_c^2) &= \prod_{i=1}^n f(x_i|\mu_c,\sigma_c^2) = \prod_{i=1}^n \frac{1}{\sigma_c\sqrt{2\pi}} e^{-\frac{1}{2\sigma_c^2}(x_i-\mu_c)^2} \\ &\propto \sigma_c^{-n} e^{-\frac{1}{2\sigma_c^2}(\sum_{i=1}^n x_i^2 - n(\bar{x})^2 + n(\bar{x}-\mu_c)^2)} \\ &= \sigma_c^{-n} e^{-\frac{1}{2\sigma_c^2}[(n-1)s_c^2 + n(\bar{x}-\mu_c)^2]} \end{aligned}$$

where $s_c^2 = \frac{1}{n-1}\sum_{i=1}^n (x_i - \bar{x})^2$. Hence the joint posterior of (μ_c, σ_c^2) is,

$$\begin{aligned} f(\mu_c,\sigma_c^2|x_1,\cdots,x_n) &\propto f(\mu_c,\sigma_c^2) f(x_1,\cdots,x_n|\mu_c,\sigma_c^2) \\ &= \sigma_c^{-(n+2)} e^{-\frac{1}{2\sigma_c^2}[(n-1)s_c^2 + n(\bar{x}-\mu_c)^2]} \end{aligned}$$

Hence, the marginal posterior of μ_c is

$$\begin{aligned} f(\mu_c|x_1,\cdots,x_n) &= \int f(\mu_c,\sigma_c^2|x_1,\cdots,x_n) d\sigma_c^2 \\ &\propto \int \sigma_c^{-(n+2)} e^{-\frac{1}{2\sigma_c^2}[(n-1)s_c^2 + n(\bar{x}-\mu_c)^2]} d\sigma_c^2 \\ &= A^{-n/2} \int y^{n/2-1} e^{-y} dy \quad (A = (n-1)s_c^2 + n(\bar{x}-\mu_c)^2) \\ &\propto [(n-1)s_c^2 + n(\bar{x}-\mu_c)^2]^{-n/2} \\ &\propto [1 + \frac{n(\bar{x}-\mu_c)^2}{(n-1)s_c^2}]^{-n/2} \end{aligned}$$

Since $\int y^{n/2-1}e^{-y}dy$ does not depend on μ_c. So, $f(\mu_c|x_1,\cdots,x_n) \sim t_{n-1}(\bar{x}, s_c^2/n)$. Put in $\bar{x} = 1.013, s_c^2 = 0.24^2, n = 32$ we have $t_{n-1}(\bar{x}, s_c^2/n) = t_{31}(1.013, 0.0018)$.

Similarly, we get the posterior density of μ_t, $f(\mu_t|y_1,\cdots,y_m) \sim t_{35}(1.173, 0.0011)$.

(5b)

$$f(\mu_c,\mu_t|x_1,\cdots,x_n,y_1,\cdots,y_m) = f(\mu_c|x_1,\cdots,x_n) f(\mu_t|y_1,\cdots,y_m)$$

Let $\mu = \mu_t - \mu_c, \mu_c = \mu_c$, we have $|\frac{d\mu d\mu_c}{d\mu_t d\mu_c}| = 1$, and

$$f(\mu,\mu_c|x_1,\cdots,x_n,y_1,\cdots,y_m) = f(\mu_c|x_1,\cdots,x_n) f(\mu_c+\mu|y_1,\cdots,y_m)$$

So, $f(\mu|x_1,\cdots,x_n,y_1,\cdots,y_m) = \int f(\mu_c|x_1,\cdots,x_n) f(\mu_c+\mu|y_1,\cdots,y_m) d\mu_c$, which is intractable.

We should resort to sampling method: (1) Sample μ_t from $f(\mu_t|y_1,\cdots,y_m)$ by $t_{35}\sqrt{s_t^2/m} + \bar{y}$ (2) Sample μ_c from $f(\mu_c|x_1,\cdots,x_n)$ by $t_{31}\sqrt{s_c^2/m} + \bar{x}$ (3) Get the empirical distribution of $\mu_t - \mu_c$; where t_n is a random number from t distribution with degree of freedom n. The histogram of $\mu_t - \mu_c$ is given in Figure 3 with R-code under it.

2.5% and 97.5% quantiles of $\mu_t - \mu_c$ is 0.05312562, and 0.2697418 respectively.

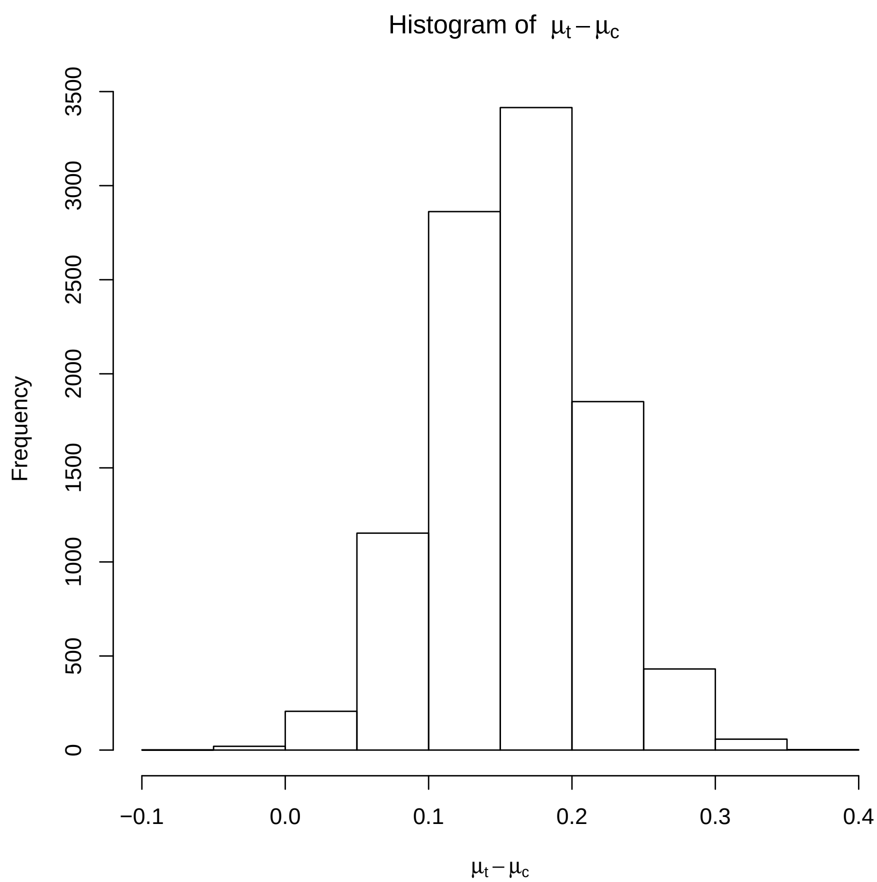

Figure 3: Histogram of $\mu_t - \mu_c$. R-Code of sampling μ_c and μ_t: n=10000(sample size) (1) Sample μ_c: st=rt(n,31)(sample 10000 samples from t_{31}), $\mu_c = st * \sqrt{0.0018} + 1.013$(samples of μ_c); (2) Sample μ_t: st=rt(n,35), $\mu_t = st * \sqrt{0.0011} + 1.173$)