STAT 457 Homework 03

Martha Eichlersmith

2019-10-25

Problem 1

Consider an *iid* sample of size n from the $\mathcal{N}(\mu, \sigma^2)$ distribution, where σ^2 is **known**. Derive the posterior distribution of μ under the prior $\mathcal{N}(\mu_0, \sigma_0^2)$.

$$
\begin{aligned}
Y \mid \mu, \sigma^2 &\sim \mathcal{N}(\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x-\mu)}{2\sigma^2}\right\} \\
\mu \mid Y &\sim \mathcal{N}(\mu_0, \sigma_0^2) = \frac{1}{\sqrt{2\pi\sigma_0^2}} \exp\left\{-\frac{(\mu-\mu_0)}{2\sigma_0^2}\right\} \\
p(\mu \mid \boldsymbol{Y}) &\propto p(\boldsymbol{Y} \mid \mu) \cdot p(\mu) \\
&\propto \prod_{i=1}^n \exp\left\{-\frac{(x_i-\mu)^2}{2\sigma^2}\right\} \exp\left\{-\frac{(\mu-\mu_0)^2}{2\sigma_0^2}\right\} \\
&= \exp\left\{-\frac{1}{2}\left(\frac{1}{\sigma^2}\sum_{i=1}^n (x_i-\mu)^2 + \frac{1}{\sigma_0^2}(\mu-\mu_0)^2\right)\right\} \\
&= \exp\left\{-\frac{1}{2}\left(\frac{1}{\sigma^2}\sum_{i=1}^n \left(x_i^2 - 2x_i\mu + \mu^2\right) + \frac{1}{\sigma_0^2}\left(\mu^2 - 2\mu\mu_0 + \mu_0^2\right)\right)\right\} \\
&= \exp\left\{-\frac{1}{2}\frac{1}{\sigma^2\sigma_0^2}\left(\sigma_0^2\sum_{i=1}^n x_i^2 - 2\sigma_0^2\mu n\bar{x} + n\mu^2\sigma_0^2 + \sigma^2\mu^2 - 2\sigma^2\mu\mu_0 + \sigma^2\mu_0^2\right)\right\} \\
&= \exp\left\{-\frac{1}{2}\frac{1}{\sigma^2\sigma_0^2}\left(\mu^2(\sigma^2 + n\sigma_0^2) - 2\mu(\mu_0\sigma^2 + \sigma_0^2 n\bar{x}) + (\mu_0^2\sigma^2 + \sigma_0^2\sum_{i=1}^n x_i^2)\right)\right\} \\
&\propto \exp\left\{-\frac{1}{2}\left[\mu^2\left(\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}\right) - 2\mu\left(\frac{\mu_0}{\sigma_0^2} + \frac{n\bar{x}}{\sigma^2}\right) + k\right]\right\} \qquad \text{where } k \text{ is some normalizing constant} \\
&= \exp\left\{-\frac{1}{2}\left(\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}\right)\left[\mu^2\frac{\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}}{\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}} - 2\mu\frac{\frac{\mu_0}{\sigma_0^2} + \frac{n\bar{x}}{\sigma^2}}{\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}} + k\right]\right\} \\
&\propto \exp\left\{-\frac{1}{2}\left(\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}\right)\left[\mu - \frac{\frac{\mu_0}{\sigma_0^2} + \frac{n\bar{x}}{\sigma^2}}{\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}}\right]^2\right\} \\
&\sim \mathcal{N}(\hat{\mu}, \hat{\sigma}_\mu^2) \\
\hat{\mu} &= \left(\frac{\mu_0}{\sigma_0^2} + \frac{n\bar{x}}{\sigma^2}\right) \Big/ \left(\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}\right) \\
\hat{\sigma}_\mu^2 &= \left(\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}\right)^{-1}
\end{aligned}
$$

Problem 2

Charles Darwin in his book *The Effects of Cross and Self Fertilization in the Vegetable Kingdom* examines the advantages of cross-fertilization in plants. He reports the following data on the difference in heights in corn plants between crossed and self-fertilized plants. The data are in eights of an inch.

```
plants <- c(49, -67, 8, 16, 6, 23, 28, 41, 14, 29, 56, 24, 75, 60, -48)
```

Problem 2a

Suppose the data are normally distributed with mean μ and variance σ^2 (**unknown**). Under the non informative prior, derive the posterior of μ. What is your best guess for μ? Obtain a 95% confidence interval for μ and interpret this interval. Compute $p(\mu > 0 \mid Y)$.

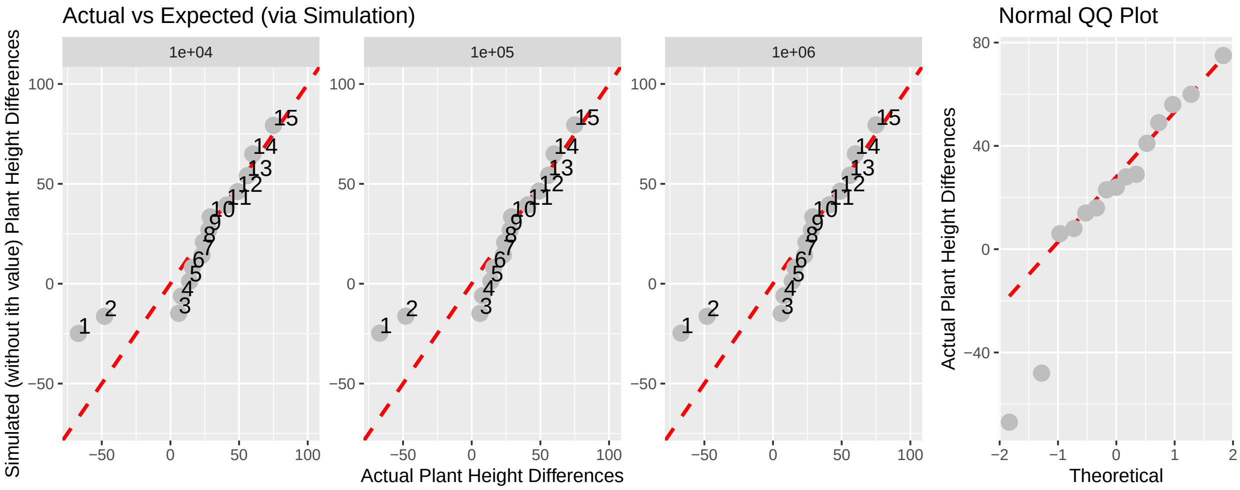
$$
\begin{aligned}
L(\mu, \sigma^2 \mid \boldsymbol{Y}) &= (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right\} \\
&\propto \sigma^{-n} \exp\left\{-\frac{1}{2\sigma^2}\left[(n-1)s^2 + n(\bar{x}-\mu)^2\right]\right\} \quad \text{where } s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2 \\
p(\mu, \sigma^2 \mid \boldsymbol{Y}) &\propto L(\mu, \sigma^2 \mid \boldsymbol{Y})p(\mu)p(\sigma^2) \quad \text{Let } p(\sigma^2) \propto \sigma^{-2} \text{ and } p(\mu) = c \\
&\propto \sigma^{-(n+2)} \exp\left\{-\frac{1}{2\sigma^2}\left[(n-1)s^2 + n(\bar{x}-\mu)^2\right]\right\} \\
\text{We know: } &\int_0^\infty x^{-b-1}\exp\left\{-a/x^2\right\}dx = \tfrac{1}{2}a^{-b/2}\,\Gamma\left(\tfrac{b}{2}\right) \\
p(\mu \mid \boldsymbol{Y}) &= \int_0^\infty p(\mu, \sigma^2)d\sigma^2 \\
&= \tfrac{1}{2}\left\{\tfrac{1}{2}\left[(n-1)s^2 + n(\mu-\bar{x})^2\right]\right\}^{-\frac{(n+1)}{2}}\Gamma\left(\tfrac{n+1}{2}\right) \\
&= \left(\tfrac{1}{2}\right)s^{-(n+1)}\left(\tfrac{1}{2}(n-1)\right)^{-\frac{(n+1)}{2}}\ \Gamma\left(\tfrac{n+1}{2}\right)\left\{1 + \tfrac{1}{n-1}\left(\tfrac{\mu-\bar{x}}{s/\sqrt{n}}\right)^2\right\}^{-\frac{(n+1)}{2}} \\
\text{Let } t &= \tfrac{\mu-\bar{x}}{s/\sqrt{n}} \text{ and so } \tfrac{d}{dt}\mu = \tfrac{s}{\sqrt{n}} \\
\implies p(t \mid \boldsymbol{Y}) &= \left(\tfrac{1}{2}\right)s^{-(n+1)}\left(\tfrac{1}{2}(n-1)\right)^{-\frac{(n+1)}{2}}\ \Gamma\left(\tfrac{n+1}{2}\right)\left\{1 + \tfrac{1}{n-1}t^2\right\}^{-\frac{(n+1)}{2}}\left(\tfrac{s}{\sqrt{n}}\right) \\
&\propto \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \cdot \frac{1}{\sqrt{n\pi}}\left(1 + \tfrac{1}{n-1}t^2\right)^{-\frac{(n+1)}{2}} \\
&\sim t \text{ distribuiton with } n \text{ degrees of freedom} \\
\implies p(\mu \mid \boldsymbol{Y}) &= \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)}\frac{s/\sqrt{n}}{\sqrt{n\pi}\left[1+\mu^2/n\right]^{(n+1)/2}} + \bar{x}
\end{aligned}
$$

t.test:

mu.hat	CI.lower	CI.upper	P(mu > 0 | Y)
20.93333	0.03119	41.83547	0.97515

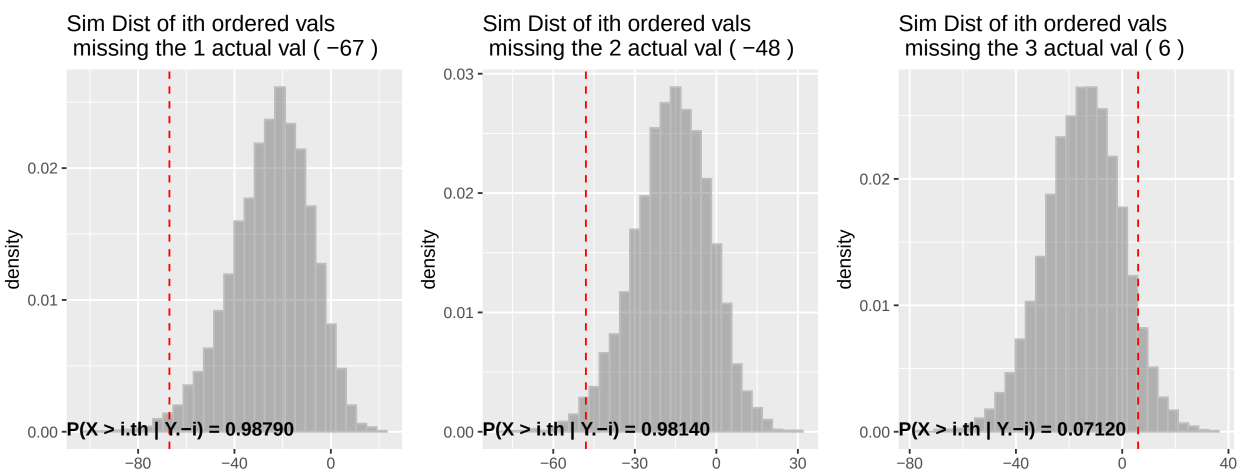
Problem 2b

Sort data and throw out the i^{th} value. Simulate from the posterior and sort data again. Plot the mean of the simulated i^{th} data and the actual i^{th} data.



The points that stray from the line and might be considered as outliers are the 1st, 2nd, and 3rd ordered values. The normal QQ plot also suggests that the 1st and 2nd values are outliers.

Looking at simulated distribution after taking out i^{th} value



The first and second actual values (-67, -48) are in the bottom 2% for the 1st and 2nd order statistics, respectively so they may be considered as outliers. The third actual value (6) would not be considered an outlier since it is in the top 7.12% of values expected for the third ordered value.

Problem 2c (Extra Credit)

Unfortunately, did not have time to attempt extra credit.

Problem 3

The following data, taken from Wallace (1980), represent the hours of post-operative pain relief for subjects receiving one of two drugs:

```
drug1 <- c(2, 4, 4, 5, 6, 8, 13)
drug2 <- c(0, 0, 0, 1, 1, 2, 2, 2 ,3, 3 ,3, 4, 8)
```

Problem 3a

Assuming that $\sigma_1^2 = \sigma_2^2$, compute the posterior distribution of $\mu_1 - \mu_2$. Obtain a 95% credible interval. Compute $p(\mu_1 - \mu_2 > 0 \mid Y)$.

CI.lower	CI.upper	CI.length	P(mu.1-mu.2 > 0 | Y)
1.07882	6.45964	5.38082	0.99565

Problem 3b

Assume that $\sigma_1^2 \neq \sigma_2^2$. Via simulation, repeat 3a.

Iterations	CI.lower	CI.upper	CI.length	P(mu.1-mu.2 > 0 | Y)
n=1e+04	0.16040	7.32940	7.16899	0.97950
n=1e+05	0.19597	7.30583	7.10987	0.97909
n=1e+06	0.20442	7.33582	7.13140	0.97956
n=1e+07	0.20668	7.33355	7.12687	0.97960

Problem 3c

Again, assume that $\sigma_1^2 \neq \sigma_2^2$. Use exact Behrens-Fisher repeat 3a.

CI.lower	CI.upper	CI.length	P(mu.1-mu.2 > 0 | Y)
0.20600	7.33246	7.12646	0.97959

Using the Patil approximation to the B-F distribution, repeat 3a.

f1	f2	a	b	qt.0.975.b.	se.patil
1.45118	0.79410	1.00741	6.65197	2.38994	1.50035

CI.lower	CI.upper	CI.length	P(mu.1-mu.2 > 0 | Y)
0.18348	7.35498	7.17150	0.97901

Problem 3d

Repeat 3a using the Welch t approximation.

CI.lower	CI.upper	CI.length	P(mu.1-mu.2 > 0 | Y)
0.36312	7.17534	6.81223	0.98305

Problem 3e

Compare results form parts a, b, c, and d. Of particular inters in the location and length of the intervals.

	CI.lower	CI.upper	CI.length	P(mu.1-mu.2 > 0 | Y)
t.test Common Variance	1.0788210	6.459640	5.380820	0.9956547
Uncommon Variance Simulation, n=1e+04	0.1604037	7.329398	7.168994	0.9795000
Uncommon Variance Simulation, n=1e+05	0.1959653	7.305832	7.109867	0.9790900
Uncommon Variance Simulation, n=1e+06	0.2044185	7.335820	7.131401	0.9795570
Uncommon Variance Simulation, n=1e+07	0.2066790	7.333548	7.126869	0.9795987
Behrens-Fisher	0.2060013	7.332460	7.126459	0.9795935
Patil approximation to B-F	0.1834802	7.354981	7.171501	0.9790120
t.test Welch (uncommon variance)	0.3631176	7.175344	6.812226	0.9830465