STAT 457 Homework 05

Martha Eichlersmith

2019-11-16

Problem 1

Consider two urns each containing an unknown mixture of blue and white marbles. A random sample of size 18 (with replacement) is drawn from urn #1 and a random sample of size 6 (with replacement) is drawn from urn #2. Of the 18 selected marbles from urn #1, 14 are blue. The corresponding number of blue marbles from urn #2 is 2.

$$\begin{aligned} L(\pi|Y) &\propto \pi^y(1-\pi)^{n-y} = \pi^{14}(1-\pi)^4 \sim \text{Binomial}, n_\pi = 18,\ y_\pi = 14 \\ &\implies p(\pi \mid Y) \sim \text{Beta}(y+\alpha_0, n-y+\beta_0) \quad \text{where} \quad p(\pi) \sim \text{Beta}(\alpha_0, \beta_0) \end{aligned}$$

$$\begin{aligned} L(\psi|Y) &\propto \psi^y(1-\psi)^{n-y} = \psi^{2}(1-\psi)^4 \sim \text{Binomial}, n_\pi = 6,\ y_\pi = 2 \\ &\implies p(\psi \mid Y) \sim \text{Beta}(y+\alpha_0, n-y+\beta_0) \quad \text{where} \quad p(\psi) \sim \text{Beta}(\alpha_0, \beta_0) \end{aligned}$$

	Blue	White
Urn 1	14	4
Urn 2	2	4

Problem 1a

Let π denote the proportion of blue marbles in urn #1 and let ψ denote the corresponding proportion in urn #2. Under the (i) Haldane, (ii) flat and (iii) non-informative priors, compute $p\left(\ln\left[\frac{\pi}{1-\pi}\right] > \ln\left[\frac{\psi}{1-\psi}\right] \mid \text{data}\right)$ using the normal approximation.

$$p\left(\ln\left[\frac{\pi}{1-\pi}\right] > \ln\left[\frac{\psi}{1-\psi}\right] \mid \text{data}\right) = p\left(\ln\left[\frac{\pi}{1-\pi}\right] - \ln\left[\frac{\psi}{1-\psi}\right] > 0 \mid \text{data}\right)$$

$$\begin{aligned} p(\pi) &\sim \text{Beta}(\alpha_0, \beta_0) \\ p(\psi) &\sim \text{Beta}(\alpha_0, \beta_0) \\ p(\pi \mid Y) &\sim \text{Beta}(y_\pi+\alpha_0,\ n_\pi - y_\pi + \beta_0) \stackrel{\text{def}}{=} \text{Beta}(\alpha, \beta) \\ p(\pi \mid Y) &\sim \text{Beta}(y_\pi+\alpha_0,\ n_\pi - y_\pi + \beta_0) \stackrel{\text{def}}{=} \text{Beta}(\gamma, \delta) \\ \text{Normal Approx Mean} &= \ln\left(\frac{\alpha\cdot\delta}{\beta\cdot\gamma}\right) \\ \text{Normal Approx Variance} &= \frac{1}{\alpha} + \frac{1}{\beta} + \frac{1}{\gamma} + \frac{1}{\delta} \\ \text{Normal Approx} &\sim \mathcal{N}\left(\ln\left(\frac{\alpha\cdot\delta}{\beta\cdot\gamma}\right),\ \sqrt{\frac{1}{\alpha} + \frac{1}{\beta} + \frac{1}{\gamma} + \frac{1}{\delta}}\ \right) \end{aligned}$$

Normal Approx: Probability difference in logodds is greater than 0

	Haldane	Flat	Non-informative
Prior	Beta(0, 0)	Beta(1, 1)	Beta(.5, .5)
p-value	0.96994	0.96402	0.96706

Problem 1b

Repeat (1a) by drawing deviates from the appropriate beta distributions. Quantify the Monte Carlo error in your value.

Probability that the log differences are greater than 0

	Haldane Beta(0, 0)			Flat Beta(1, 1)			Non-informative Beta(.5, .5)		
Iterations	10000	1e+05	1e+06	10000	1e+05	1e+06	10000	1e+05	1e+06
p-value	0.97293	0.97254	0.97256	0.97303	0.97258	0.97265	0.97105	0.97251	0.97281
Standard Error	0.01137	0.00360	0.00114	0.01125	0.00359	0.00113	0.01136	0.00359	0.00113

Problem 1c

Compare your results in (1a) and (1b) to the p-value obtained via Fisher's exact test.

$$\text{Odds Ratio} > 1 \implies \frac{\pi}{1-\pi} > \frac{\psi}{1-\psi}$$

```
##
##  Fisher's Exact Test for Count Data
##
## data:  blue
## p-value = 0.06927
## alternative hypothesis: true odds ratio is greater than 1
## 95 percent confidence interval:
##  0.8606483       Inf
## sample estimates:
## odds ratio
##   6.334078
```

The Fisher Exact test p-value is significantly different from the methods used in (1a) and (1b).

Problem 1d

Add delinquency problem

Problem 2

Suppose a sample of size n is drawn at random and with replacement from some population. For large n the sample proportion ($\hat{p}$) is normally distributed with mean p and variance $\frac{p(1-p)}{n}$. Find the asymptotic distribution of $2\sin^{-1}\sqrt{\hat{p}}$ using the delta method.

$$\hat{p} \sim \mathcal{N}\left(p,\ \tfrac{p(1-p)}{n}\right)$$

$$\text{Let } g(\hat{p}) = 2\sin^{-1}\sqrt{\hat{p}}$$

$$\sqrt{n}\left(g(\hat{p}) - g(p)\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0,\ \sigma^2[g'(p)]^2\right)$$

$$\sigma^2 = \tfrac{p(1-p)}{n}$$

$$g'(p) = \frac{1}{\sqrt{1-p}\sqrt{p}}$$

$$[g'(p)]^2 = \frac{1}{(1-p)p}$$

$$\sigma^2[g'(p)]^2 = \frac{p(1-p)}{n} \cdot \frac{1}{(1-p)p} = \frac{1}{n}$$

$$\sqrt{n}\left(g(\hat{p}) - g(p)\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0,\ \tfrac{1}{n}\right)$$

$$g(\hat{p}) \xrightarrow{\mathcal{D}} \mathcal{N}\left(g(\hat{p}),\ 1\right)$$

$$2\sin^{-1}\sqrt{\hat{p}} \xrightarrow{\mathcal{D}} \mathcal{N}\left(2\sin^{-1}\sqrt{p}\ ,\ 1\right)$$

Problem 3

Let $x_1, \cdots, x_n$ be an iid sample from $\mathcal{N}(\theta, 1)$ and let $y_1, \cdots, y_n$ be an independent iid sample from $\mathcal{N}(\phi, 1)$. Derive the distribution of $\overline{x}/\overline{y}$ (where $\overline{y} \neq 0$) via the delta method.

$$
\begin{aligned}
\overline{x} &\sim \mathcal{N}(\theta, {}^1\!/\!{}_n) \\
\overline{y} &\sim \mathcal{N}(\phi, {}^1\!/\!{}_n) \\
\text{Let } h(x, y) &= x/y \\
\text{Let } h(B) &= \overline{x}/\overline{y} \\
\text{Let } h(\beta) &= \theta/\phi \\
\sqrt{n}\left(h(B) - h(\beta)\right) &\xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \nabla h(\beta)^T \cdot \Sigma \cdot \nabla h(\beta)\right) \\
\Sigma &= \begin{bmatrix} 1/n & 0 \\ 0 & 1/n \end{bmatrix} \\
\nabla h(\beta)^T &= \begin{bmatrix} \frac{\partial h}{\partial x} & \frac{\partial h}{\partial y} \end{bmatrix}_{\theta, \phi} \\
&= \begin{bmatrix} \frac{1}{y} & -\frac{x}{y^2} \end{bmatrix}_{\theta, \phi} \\
&= \begin{bmatrix} \frac{1}{\phi} & -\frac{\theta}{\phi^2} \end{bmatrix} \\
\nabla h(\beta)^T \cdot \Sigma \cdot \nabla &= \begin{bmatrix} \frac{1}{\phi} & -\frac{\theta}{\phi^2} \end{bmatrix} \begin{bmatrix} 1/n & 0 \\ 0 & 1/n \end{bmatrix} \begin{bmatrix} \frac{1}{\phi} \\ -\frac{\theta}{\phi^2} \end{bmatrix} \\
&= \begin{bmatrix} \frac{1}{n\phi} & -\frac{\theta}{n\phi^2} \end{bmatrix} \begin{bmatrix} \frac{1}{\phi} \\ -\frac{\theta}{\phi^2} \end{bmatrix} \\
&= \frac{1}{n}\left(\frac{1}{\phi^2} - \frac{\sigma^2}{\phi^4}\right) \\
\sqrt{n}\left({}^{\overline{x}}\!/\!{}_{\overline{y}} - {}^{\theta}\!/\!{}_{\phi}\right) &\xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \frac{1}{n}\left(\frac{1}{\phi^2} - \frac{\sigma^2}{\phi^4}\right)\right) \\
\frac{\overline{x}}{\overline{y}} &\xrightarrow{\mathcal{D}} \mathcal{N}\left(\frac{\theta}{\phi}, \ \frac{1}{\phi^2} - \frac{\sigma^2}{\phi^4}\right)
\end{aligned}
$$

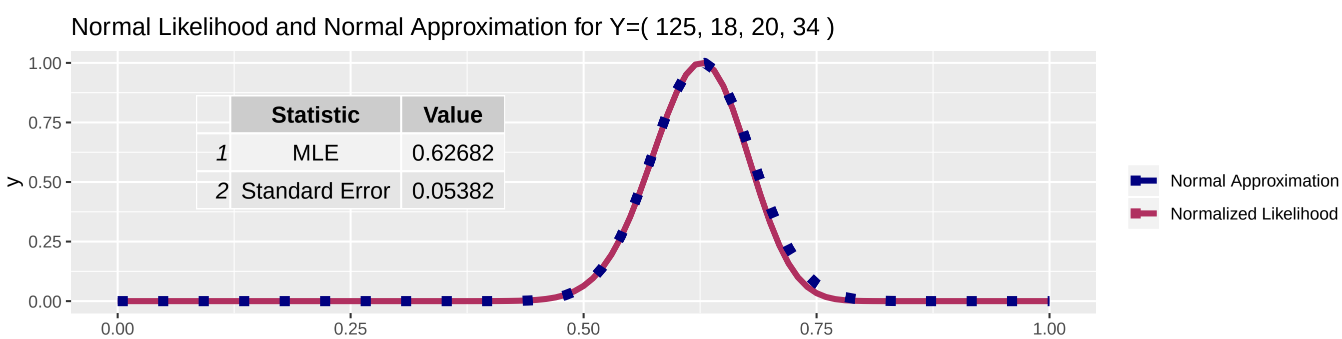
Problem 4

197 animals are distributed into four categories: $Y = (y_1, y_2, y_3, y_4)$ according to the genetic linkage model $\left(\frac{2+\theta}{4}, \frac{1-\theta}{4}, \frac{1-\theta}{4}, \frac{\theta}{4}\right)$. In HW#4 you derived the likelihood for the data $Y = (125, 18, 20, 34)$ and you derived the likelihood for the data $Y = (14, 0, 1, 15)$. In that homework, you also used Newton-Raphson algorithm to obtain the MLE $(\hat{\theta})$ of θ and the standard error of $\hat{\theta}$.

$$
\begin{aligned}
L(\theta \mid \boldsymbol{Y}) &= \frac{(y_1+y_2+y_3+y_4)!}{y_1!y_2!y_2!y_4!}\left(\frac{2+\theta}{4}\right)^{y_1}\left(\frac{1-\theta}{4}\right)^{y_2}\left(\frac{1-\theta}{4}\right)^{y_3}\left(\frac{\theta}{4}\right)^{y_4} \\
&\propto (2+\theta)^{y_1}\cdot(1-\theta)^{y_2+y_3}\cdot(\theta)^{y_4} \\
\ell(\theta \mid \boldsymbol{Y}) &\propto y_1\log(2+\theta) + (y_2+y_3)\log(1-\theta) + y_4\log(\theta) \\
\frac{\partial \ell}{\partial \theta} &= \frac{y_1}{2+\theta} - \frac{y_2+y_3}{1-\theta} + \frac{y_4}{\theta} \\
\frac{\partial^2 \ell}{\partial \theta^2} &= -\frac{y_1}{(2+\theta)^2} - \frac{y_2+y_3}{(1-\theta)^2} - \frac{y_4}{\theta^2} \\
\theta^{(i+1)} &= \theta^{(i)} - \frac{\frac{y_1}{2+\theta^{(i)}} - \frac{y_2+y_3}{1-\theta^{(i)}} + \frac{y_4}{\theta^{(i)}}}{-\frac{y_1}{(2+\theta^{(i)})^2} - \frac{y_2+y_3}{(1-\theta^{(i)})^2} - \frac{y_4}{(\theta^{(i)})^2}} \stackrel{\text{Newton-Raphson}}{\Longrightarrow} \hat{\theta} \\
s.e.(\hat{\theta}) &= \sqrt{1/\mathcal{I}(\theta)} \\
\mathcal{I}(\theta) &= \left[\frac{\partial^2 \ell}{\partial \theta^2}\right]_{\hat{\theta}} = -\frac{y_1}{(2+\hat{\theta})^2} - \frac{y_2+y_3}{(1-\hat{\theta})^2} - \frac{y_4}{\hat{\theta}^2}
\end{aligned}
$$

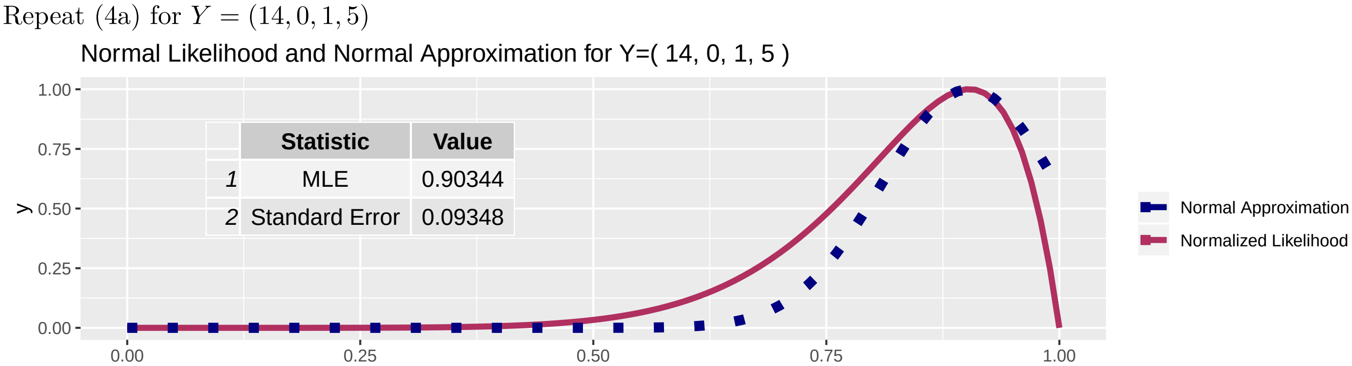
Problem 4a

Plot the normalized likelihood and the associated normal approximation in the same figure for the data $Y = (125, 18, 20, 34)$. Discuss the adequacy of the normal approximation.



Problem 4b

Repeat (4a) for $Y = (14, 0, 1, 5)$



Problem 5

Use Laplace's method (second order) to compute the posterior mean (under a flat prior) for the genetic linkage model for both data sets.