STATS 200: Solutions to Homework 6

1. (a) The expectation of X is

$$\mathbb{E}[X] = \tfrac{2}{3}\theta \cdot 0 + \tfrac{1}{3}\theta \cdot 1 + \tfrac{2}{3}(1-\theta) \cdot 2 + \tfrac{1}{3}(1-\theta) \cdot 3 = \tfrac{7}{3} - 2\theta.$$

For an IID sample $X_1, \ldots, X_n$, equating $\frac{7}{3} - 2\theta$ with the sample mean $\bar{X}$ and solving for θ, the method-of-moments estimate is $\hat{\theta} = \frac{1}{2}(\frac{7}{3} - \bar{X})$. For the 10 given observations, $\hat{\theta} = 0.417$.

The variance of X is

$$\begin{aligned} \mathrm{Var}[X] &= \mathbb{E}[X^2] - \mathbb{E}[X]^2 \\ &= \tfrac{2}{3}\theta \cdot 0^2 + \tfrac{1}{3}\theta \cdot 1^2 + \tfrac{2}{3}(1-\theta) \cdot 2^2 + \tfrac{1}{3}(1-\theta) \cdot 3^2 - \left(\tfrac{7}{3} - 2\theta\right)^2 \\ &= \tfrac{2}{9} + 4\theta - 4\theta^2. \end{aligned}$$

Then $\mathrm{Var}[\hat{\theta}] = \frac{1}{4}\mathrm{Var}[\bar{X}] = \frac{1}{4n}\mathrm{Var}[X] = \frac{1}{4n}(\frac{2}{9} + 4\theta - 4\theta^2)$. An estimate of the standard error is $\sqrt{\frac{1}{4n}(\frac{2}{9} + 4\hat{\theta} - 4\hat{\theta}^2)}$, which for the 10 given observations is 0.173. (An alternative estimate of the standard error is given by $\frac{1}{4n}$ times the sample variance of $X_1, \ldots, X_n$, which for the 10 given observations is 0.171.)

(b) For an IID sample $X_1, \ldots, X_n$, let N_0, N_1, N_2, N_3 be the total numbers of observations equal to 0, 1, 2, and 3. Then the log-likelihood is

$$\begin{aligned} l(\theta) &= \log\left(\prod_{i=1}^{n}\left(\frac{2}{3}\theta\right)^{\mathbb{1}\{X_i=0\}}\left(\frac{1}{3}\theta\right)^{\mathbb{1}\{X_i=1\}}\left(\frac{2}{3}(1-\theta)\right)^{\mathbb{1}\{X_i=2\}}\left(\frac{1}{3}(1-\theta)\right)^{\mathbb{1}\{X_i=3\}}\right) \\ &= N_0 \log\frac{2}{3}\theta + N_1 \log\frac{1}{3}\theta + N_2 \log\frac{2}{3}(1-\theta) + N_3 \log\frac{1}{3}(1-\theta). \end{aligned}$$

To compute the MLE for θ, we set

$$0 = l'(\theta) = \frac{N_0}{\theta} + \frac{N_1}{\theta} - \frac{N_2}{1-\theta} - \frac{N_3}{1-\theta}$$

and solve for θ, yielding $\hat{\theta} = (N_0 + N_1)/(N_0 + N_1 + N_2 + N_3) = (N_0 + N_1)/n$. For the 10 given observations, $\hat{\theta} = 0.5$.

The total probability that $X = 0$ or $X = 1$ is θ, so $N_0 + N_1 \sim \text{Binomial}(n, \theta)$. Then $\text{Var}[\hat{\theta}] = \frac{1}{n^2}\text{Var}[N_0 + N_1] = \frac{\theta(1-\theta)}{n}$. (Alternatively, we may compute

$$\frac{\partial^2}{\partial\theta^2}\log f(x|\theta) = \begin{cases} -\frac{1}{\theta^2} & x = 0 \text{ or } x = 1 \\ -\frac{1}{(1-\theta)^2} & x = 2 \text{ or } x = 3, \end{cases}$$

so the Fisher information is $I(\theta) = -\mathbb{E}[\frac{\partial^2}{\partial\theta^2}\log f(X|\theta)] = \frac{1}{\theta(1-\theta)}$. This shows that the variance of $\hat{\theta}$ is approximately $\frac{\theta(1-\theta)}{n}$ for large n.) An estimate of the standard error is $\sqrt{\hat{\theta}(1-\hat{\theta})/n}$, which for the 10 given observations is 0.158.

(c)
```
X = c(3,0,2,1,3,2,1,0,2,1)
n = length(X)
B = 10000
theta_hat_star = numeric(B)
for (i in 1:B) {
  X_star = sample(X,n,replace=TRUE)
  theta_hat_star[i] = length(which(X_star <= 1))/n
}
print(sd(theta_hat_star))
```
We obtain a bootstrap-estimated standard error of 0.159.

2. (a) $\hat{p} \to p$ in probability, hence $\frac{\sqrt{n}(\hat{p}-p)}{\sqrt{\hat{p}(1-\hat{p})}} \to \mathcal{N}(0,1)$ in distribution by Slutsky's Lemma. So for large n

$$\mathbb{P}\left[-z(0.025) \leq \frac{\sqrt{n}(\hat{p}-p)}{\sqrt{\hat{p}(1-\hat{p})}} \leq z(0.025)\right] \approx 0.95.$$

We may rewrite the above as

$$\mathbb{P}\left[\hat{p} - \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}z(0.025) \leq p \leq \hat{p} + \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}z(0.025)\right] \approx 0.95,$$

so an approximate 95% confidence interval is $\hat{p} \pm \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}z(0.025)$.

(b) The following conditions are equivalent:

$$-\sqrt{p(1-p)}z(\alpha/2) \leq \sqrt{n}(\hat{p}-p) \leq \sqrt{p(1-p)}z(\alpha/2)$$
$$\Updownarrow$$
$$n(\hat{p}-p)^2 \leq p(1-p)z(\alpha/2)^2$$
$$\Updownarrow$$
$$(n + z(\alpha/2)^2)p^2 - (2n\hat{p} + z(\alpha/2)^2)p + n\hat{p}^2 \leq 0.$$

This occurs when p is between the two real roots of above quadratic equation, which are given by

$$\frac{2n\hat{p} + z(\alpha/2)^2 \pm \sqrt{(2n\hat{p} + z(\alpha/2)^2)^2 - 4(n + z(\alpha/2)^2)n\hat{p}^2}}{2(n + z(\alpha/2)^2)}.$$

Taking $\alpha = 0.05$ and simplifying the above, we obtain an approximate 95% confidence interval of

$$\frac{\hat{p} + \frac{z(0.025)^2}{2n} \pm z(0.025)\sqrt{\frac{\hat{p}(1-\hat{p})}{n} + \frac{z(0.025)^2}{4n^2}}}{1 + \frac{z(0.025)^2}{n}}.$$

(c)
```
ns = c(10,40,100)
ps = c(0.1,0.3,0.5)
B=100000
z = qnorm(0.975)
for (n in ns) {
  for (p in ps) {
    cover_A = numeric(B)
    cover_B = numeric(B)
    for (i in 1:B) {
      phat = rbinom(1,n,p)/n
      U = phat+z*sqrt(phat*(1-phat)/n)
      L = phat-z*sqrt(phat*(1-phat)/n)
      if (p <= U && p >= L) {
        cover_A[i] = 1
      } else {
        cover_A[i] = 0
      }
      U = (phat+z^2/(2*n)+z*sqrt(phat*(1-phat)/n+z^2/(4*n^2)))/(1+z^2/n)
      L = (phat+z^2/(2*n)-z*sqrt(phat*(1-phat)/n+z^2/(4*n^2)))/(1+z^2/n)
      if (p <= U && p >= L) {
        cover_B[i] = 1
      } else {
        cover_B[i] = 0
      }
    }
    print(c(n,p,mean(cover_A),mean(cover_B)))
  }
}
```

For the interval from part (a), we obtain the following coverage probabilities:

	$p = 0.1$	$p = 0.3$	$p = 0.5$
$n = 10$	0.65	0.84	0.89
$n = 40$	0.91	0.93	0.92
$n = 100$	0.93	0.95	0.94

For the interval from part (b), we obtain the following coverage probabilities:

	$p = 0.1$	$p = 0.3$	$p = 0.5$
$n = 10$	0.93	0.92	0.98
$n = 40$	0.94	0.94	0.96
$n = 100$	0.94	0.94	0.94

The intervals from part (b) are more accurate when n is small.

3. (a) The KL-divergence is given by

$$\begin{aligned} D_{\mathrm{KL}}(g(x)\|f(x|\lambda)) &= \mathbb{E}_g\left[\log \frac{g(X)}{f(X|\lambda)}\right] \\ &= \mathbb{E}_g\left[\log \frac{\frac{1}{\Gamma(2)}Xe^{-X}}{\lambda e^{-\lambda X}}\right] \\ &= \mathbb{E}_g[-\log\Gamma(2) + \log X - X - \log\lambda + \lambda X] \\ &= -\log\Gamma(2) - \log\lambda + \mathbb{E}_g[\log X] + (\lambda-1)\,\mathbb{E}_g[X] \\ &= -\log\Gamma(2) - \log\lambda + \psi(2) + 2(\lambda - 1). \end{aligned}$$

Setting the derivative with respect to λ equal to 0, this is minimized at $\lambda^* = 1/2$.

(b) By the Law of Large Numbers, $\bar{X} \to \mathbb{E}_g[X] = 2$ in probability, so $\hat{\lambda} = 1/\bar{X} \to 1/2$ in probability by the Continuous Mapping Theorem.

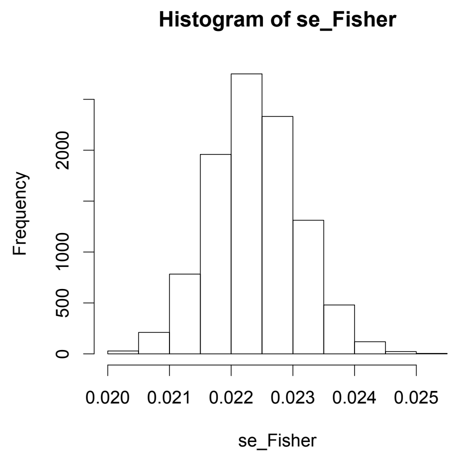
(c) The Fisher information in the exponential model is given by

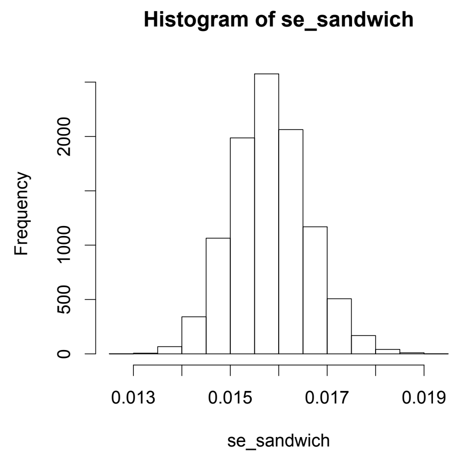
$$I(\lambda) = -\,\mathbb{E}_\lambda\left[\frac{\partial^2}{\partial\lambda^2}\log f(X|\lambda)\right] = -\,\mathbb{E}_\lambda\left[\frac{\partial^2}{\partial\lambda^2}(\log\lambda - \lambda X)\right] = 1/\lambda^2.$$

The corresponding plug-in estimate of the standard error is $\sqrt{\frac{1}{nI(\hat{\lambda})}} = \frac{1}{\bar{X}\sqrt{n}}$.

```
n = 500
B = 10000
lamb_hat = numeric(B)
se_Fisher = numeric(B)
se_sandwich = numeric(B)
for (i in 1:B) {
  X = rgamma(n,2,rate=1)
  lamb_hat[i] = 1/mean(X)
  se_Fisher[i] = 1/(mean(X)*sqrt(n))
```

```
  se_sandwich[i] = sd(X)/(mean(X)^2*sqrt(n))
}
print(mean(lamb_hat))
print(sd(lamb_hat))
hist(se_Fisher)
hist(se_sandwich)
```





The empirical mean and standard error of $\hat{\lambda}$ are 0.501 and 0.016; the mean is close to $\lambda^* = 0.5$ from part (a). The Fisher-information-based estimate of the standard error is incorrect—it estimates the standard error as approximately 0.022. The sandwich estimate of the standard error seems correct—it estimates the standard error as 0.016, with some variability in the third decimal place.

4. (a) We may estimate p by $\hat{p} = \bar{X}$, and q by $\hat{q} = \bar{Y}$. The plugin estimator for the log-odds-ratio is

$$\log\left(\frac{\hat{p}}{1-\hat{p}}\Big/\frac{\hat{q}}{1-\hat{q}}\right).$$

(b) Let

$$g(p,q) = \log\left(\frac{p}{1-p}\Big/\frac{q}{1-q}\right) = \log p - \log(1-p) - \log q + \log(1-q).$$

Applying a first-order Taylor expansion to g,

$$g(\hat{p},\hat{q}) \approx g(p,q) + \frac{\hat{p}-p}{p(1-p)} + \frac{\hat{q}-q}{q(1-q)}.$$

$\hat{p}$ and $\hat{q}$ are independent, and by the Central Limit Theorem, $\sqrt{n}(\hat{p}-p) \to \mathcal{N}(0, p(1-p))$ and $\sqrt{m}(\hat{q}-q) \to \mathcal{N}(0, q(1-q))$. Hence, for large m and n, $g(\hat{p},\hat{q})$ is approximately distributed as $\mathcal{N}(g(p,q), v)$ where

$$v = \frac{p(1-p)}{n} \times \frac{1}{p^2(1-p)^2} + \frac{q(1-q)}{m} \times \frac{1}{q^2(1-q)^2} = \frac{1}{np(1-p)} + \frac{1}{mq(1-q)}.$$

(c) Let $\hat{v} = \frac{1}{n\hat{p}(1-\hat{p})} + \frac{1}{m\hat{q}(1-\hat{q})}$ be the plugin estimate of v. As $m, n \to \infty$, $\hat{v}/v \to 1$ in probability, so by part (b) and Slutsky's lemma,

$$\mathbb{P}\left[-z(0.025) \leq \frac{g(\hat{p},\hat{q}) - g(p,q)}{\sqrt{\hat{v}}} \leq z(0.025)\right] \approx 0.95$$

for large m and n. Rearranging yields a 95% confidence interval for $g(p,q)$ given by

$$g(\hat{p},\hat{q}) \pm z(0.025)\sqrt{\hat{v}} = \log\left(\frac{\hat{p}}{1-\hat{p}}\Big/\frac{\hat{q}}{1-\hat{q}}\right) \pm z(0.025)\sqrt{\frac{1}{n\hat{p}(1-\hat{p})} + \frac{1}{m\hat{q}(1-\hat{q})}}.$$

Denoting this interval by $[L(\hat{p},\hat{q}), U(\hat{p},\hat{q})]$, we may exponentiate to obtain the confidence interval $[e^{L(\hat{p},\hat{q})}, e^{U(\hat{p},\hat{q})}]$ for the odds-ratio $\frac{p}{1-p}/\frac{q}{1-q}$.