Solutions for Homework 11

(1) Recall that $\tau = \mu + \sigma c$ where $c = \Phi^{-1}(.95) = 1.645$. The mle is $\hat{\tau} = \hat{\mu} + \hat{\sigma} c = 4.18$. The delta method se (formula from last homework) is

$$\frac{\hat{\sigma}}{\sqrt{n}}\sqrt{1 + \frac{c^2}{2}} = .557.$$

The bootstrap code is:

```
x <- c(3.23,-2.50,1.88,-0.68,4.43,0.17,
      1.03,-0.07,-0.01,0.76,1.76,3+.18,
      0.33,-0.31,0.30,-0.61,1.52,5.43,
      1.54,2.28,0.42,2.33,-1.03,4.00,0.39)
n <- length(x)
B <- 10000
mu.hat <- mean(x)
sigma.hat <- sqrt(sum( (x-mu.hat)^2)/n)
cc <- qnorm(.95)
tau.hat <- mu.hat + sigma.hat*cc
print(tau.hat)
se <- (sigma.hat/sqrt(n))*sqrt(1 + cc^2/2)  ### delta method
print(se)
tau <- rep(0,B)
for(i in 1:B){
     xx <- rnorm(n,mu.hat,sigma.hat)
     mu.hat.star <- mean(xx)
     sigma.hat.star <- sqrt(sum( (xx-mu.hat.star)^2)/n)
     tau[i] <- mu.hat.star + sigma.hat.star*cc
     }
se <- sqrt(var(tau))
print(se)
```

You should get a se close to the delta method. I got se = .551.

(2) $\theta = g(\mu) = e^{\mu}$ so $g'(\mu) = e^{\mu}$. Therefore, $se(\hat{\theta}) = se(\hat{\mu})e^{\hat{\mu}} = e^{\overline{X}}/\sqrt{n}$. For my data I get $\hat{\theta} = 145.014$ and:

method	se	confidence interval
delta method	14.50	(116,174)
parametric bootstrap	14.48	(120,176)
nonparametric bootstrap	16.01	(117,180)

Both the parametric and nonparametric bootstraps produce estimates of the sampling distribution (i.e. histograms of the bootstrap replications) that are close to the true sampling distribution (as computed by simulation). Here is the code I used:

```
n         <- 100
x         <- rnorm(n,5,1)
mu.hat    <- mean(x)
theta.hat <- exp(mu.hat)
se        <- exp(mu.hat)/sqrt(n)
print(se)
ci <- c(theta.hat - 2*se, theta.hat + 2*se)
print(ci)

B            <- 10000
theta.nonpar <- rep(0,B)
theta.par    <- rep(0,B)
for(i in 1:B){
      xx              <- rnorm(n,mu.hat,1)
      theta.par[i]    <- exp(mean(xx))
      xx              <- sample(x,size=n,replace=T)
      theta.nonpar[i] <- exp(mean(xx))
      }

se.par    <- sqrt(var(theta.par))
se.nonpar <- sqrt(var(theta.nonpar))
print(se.par)
print(se.nonpar)
```

```
print(quantile(theta.par,c(.025,.975)))
print(quantile(theta.nonpar,c(.025,.975)))
hist(theta.par,main="parametric bootstrap")
hist(theta.nonpar,main="nonparametric bootstrap")
theta.grid <- seq(90,220,length=1000)
plot(theta.grid,dnorm(theta.grid,theta.hat,se),type="l")

### Get true distribution of theta.hat
n <- 100
theta.hat <- rep(0,B)
for(i in 1:B){
      x            <- rnorm(n,5,1)
      theta.hat[i] <- exp(mean(x))
      }
hist(theta.hat,main="True Sampling Distribution")
```

(3a) Let G be the cdf of $\hat{\theta}$. Then

$$\begin{aligned} G(c) &= P(\hat{\theta} \leq c) \\ &= P(\max\{X_1, \ldots, X_n\} \leq c) \\ &= P(X_1 \leq c, \ldots, X_n \leq c) \\ &= \prod_i P(X_i \leq c) \\ &= P(X_1 \leq c)^n \\ &= \left(\frac{c}{\theta}\right)^n. \end{aligned}$$

Thus, the density of $\hat{\theta}$ is $g(c) = \theta^{-n} n c^{n-1}$. When $\theta = 1$, $g(c) = nc^{n-1}$. So, $\hat{\theta} \sim \text{Beta}(n, 1)$. The distribution of $\hat{\theta}$ from the parametric bootstrap mimics g well. The nonparametric bootstrap does poorly.

(3b) In the parametric bootstrap we draw $X_1, \ldots, X_n \sim \text{Unif}(0, \hat{\theta})$. Since we are drawing from a continuous distribution, $P(\hat{\theta}^* = c)$ for any c is 0. For the nonparametric bootstrap, we will get $\hat{\theta}^* = \hat{\theta}$ anytime $X_1^*, \ldots, X_n^*$ includes the maximum of the original sample. The chance that X_i^* is equal to the maximum of of the original sample is $1/n$. The chance it is not equal to the maximum of of the original sample is thus $1 - (1/n)$. Therefore, the chance that the bootstrap sample omits the maximum is $(1-(1/n))^n$. Hence, the chance that the bootstrap sample includes the maximum is $1-(1-(1/n))^n$. Now $(1-(1/n))^n \to e^{-1}$ so $P(\hat{\theta}^* = \hat{\theta}) \to 1-e^{-1} = .63$.

(4) $\hat{p}_1 = X_1/n_1 = 3/5$ and $\hat{p}_2 = X_2/n_2 = 4/5$.
(4a) $\hat{\tau} = \hat{p}_2 - \hat{p}_1 = .8 - .6 = 0.2$. The likelihood is

$$L(p_1,p_2) = p_1^{X_1}(1-p_1)^{n_1-X_1} p_2^{X_2}(1-p_2)^{n_2-X_2}.$$

The matrix H of second derivatives is

$$H = \begin{bmatrix} -\frac{X_1}{p_1^2} - \frac{1-X_1}{(1-p_1)^2} & 0 \\ 0 & -\frac{X_2}{p_2^2} - \frac{1-X_2}{(1-p_2)^2} \end{bmatrix}.$$

Since $E(X_1) = n_1 p_1$ and $E(X_2) = n_2 p_2$, the Fisher information matrix is

$$I(p_1,p_2) = E(-H) = \begin{bmatrix} \frac{n_1}{p_1(1-p_1)} & 0 \\ 0 & \frac{n_2}{p_2(1-p_2)} \end{bmatrix}.$$

Now $\tau = g(p_1,p_2) = p_2 - p_1$ and the gradient of g is

$$\nabla g = \begin{pmatrix} -1 \\ 1 \end{pmatrix}.$$

By the delta method, the estimated standard error of $\hat{\tau}$ is

$$\hat{se} = \sqrt{(\nabla g)^T I^{-1}(\hat{p}_1,\hat{p}_2)(\nabla g)} = \left\{ \frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2} \right\}^{1/2}.$$

The standard error is .08944. A 90 per cent interval is $.8 \pm 1.645(.08944) = (0.05287982, 0.3471202)$.

(4b) The bootstrap code is:

```
B <- 10000
tau.boot <- rep(0,B)
for(i in 1:B){
     xx1 <- rbinom(1,n1,p1.hat)
     xx2 <- rbinom(1,n2,p2.hat)
     tau.boot[i] <- (xx2/n2)-(xx1/n1)
     }
```

This gives a standard error of 0.08899104 and the bootstrap 90 per cent confidence interval is (.04,.34).

(4c) To draw from the posterior:

```
B <- 10000
p1 <- rbeta(B,x1+1,n1-x1+1)
p2 <- rbeta(B,x2+1,n2-x2+1)
tau <- p2 - p1
```

The posterior mean is .19 and the interval is (.05,.34).

(4d) $\psi = g(p_1, p_2)$ has gradient

$$\nabla g = \begin{pmatrix} \frac{\partial \psi}{p_1} \\ \frac{\partial \psi}{p_2} \end{pmatrix} = \nabla g = \begin{pmatrix} \frac{1}{p_1(1-p_1)} \\ -\frac{1}{p_2(1-p_2)} \end{pmatrix}.$$

The delta method gives

$$\hat{se} = \sqrt{(\nabla g)^T I^{-1}(\hat{p}_1, \hat{p}_2)(\nabla g)} = \left\{ \frac{1}{n_1 \hat{p}_1 (1-\hat{p}_1)} + \frac{1}{n_2 \hat{p}_2 (1-\hat{p}_2)} \right\}^{1/2}.$$

We find that $\hat{\psi} = 0.2924813$ and the standard error is 0.4564355. The 90 per cent interval is $\hat{\psi} \pm 1.645(0.4564355) = (-.46, 1.04)$.

(4e) Draws from the posterior are obtained by

```
psi <- log(p1/(1-p1))/log(p2/(1-p2))
```

The posterior mean is .32 and the 90 per cent interval is (-0.05, 0.76).

(5) Let $\sigma_n = \sigma/\sqrt{n}$. The likelihood is

$$\mathcal{L}(\mu) \propto \prod_i e^{-(X_i-\mu)^2/(2\sigma^2)} \propto \exp\left\{ -\frac{(\mu - \overline{X})^2}{2\sigma_n^2} \right\}.$$

The posterior is

$$\begin{aligned} f(\mu | x^n) &\propto \exp\left\{ -\frac{(\mu - \overline{X})^2}{2\sigma_n^2} \right\} \exp\left\{ -\frac{(\mu - a)^2}{2b^2} \right\} \\ &\propto \exp\left\{ -\frac{1}{2}\left[\frac{(\mu - \overline{X})^2}{2\sigma_n^2} + \frac{(\mu - a)^2}{2b^2} \right] \right\} \end{aligned}$$

$$
\begin{aligned}
&\propto \exp\left\{-\frac{1}{2}\left[\frac{\mu^2 - 2\overline{X}\mu}{2\sigma_n^2} + \frac{\mu^2 - 2a\mu}{2b^2}\right]\right\} \\
&\propto \exp\left\{-\frac{1}{2}\left[\mu^2\left(\frac{1}{\sigma_n^2} + \frac{1}{b^2}\right) - 2\mu\left(\frac{\overline{X}}{\sigma_n^2} + \frac{a}{b^2}\right)\right]\right\} \\
&\propto \exp\left\{-\frac{1}{2}\left(\frac{1}{\sigma_n^2} + \frac{1}{b^2}\right)\left[\mu^2 - 2\mu\frac{\left(\frac{\overline{X}}{\sigma_n^2} + \frac{a}{b^2}\right)}{\left(\frac{1}{\sigma_n^2} + \frac{1}{b^2}\right)}\right]\right\} \\
&\propto \exp\left\{-\frac{1}{2}\left(\frac{1}{\sigma_n^2} + \frac{1}{b^2}\right)\left[\mu - \frac{\left(\frac{\overline{X}}{\sigma_n^2} + \frac{a}{b^2}\right)}{\left(\frac{1}{\sigma_n^2} + \frac{1}{b^2}\right)}\right]^2\right\}.
\end{aligned}
$$

(6) The posteriors are Beta distributions. The parameters are obtained by adding 2 to the first parameter and 8 to the second. The posterior can be obtained using the dbeta function.

(7a) The likelihood is

$$\mathcal{L}(\lambda) = e^{-n\lambda}\lambda^{\sum_i X_i}.$$

The prior is

$$f(\lambda) \propto \lambda^{\alpha-1}e^{-\lambda/\beta}.$$

The posterior is

$$f(\lambda|x^n) \propto \lambda^{\sum_i X_i + \alpha - 1}\exp\left\{-\lambda\left[n + \frac{1}{\beta}\right]\right\}.$$

So the posterior is Gamma($\alpha + \sum_i X_i, \beta/(n\beta + 1)$).

(7b) The Fisher information is $I(\lambda) \propto 1/\lambda$. Jeffreys' prior is $f(\lambda) \propto I^{1/2}(\lambda) = \lambda^{-1/2}$. The posterior is

$$f(\lambda|x^n) \propto e^{-n\lambda}\lambda^{\sum_i X_i - (1/2)}.$$

which is Gamma($(1/2) + \sum_i X_i, 1/n$).