Homework 4 Solutions
Statistics 200B
Due Feb. 28, 2019

1. Verify the statements made in class about the Fisher information matrix $I_n(\mu, \sigma)$ and its inverse when $X_1, \ldots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$.

   Solutions:

   The density of X_i, $i = 1, \ldots, n$ is

$$f(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}.$$

   For a single observation X, the log-likelihood is

$$\ell(\mu, \sigma) = \log L(\mu, \sigma) = \log f(X; \mu, \sigma) = -\log(\sqrt{2\pi}\sigma) - \frac{(X-\mu)^2}{2\sigma^2}.$$

   The first partial derivatives are

$$\frac{\partial \ell}{\partial \mu}(\mu, \sigma) = \frac{X-\mu}{\sigma^2}; \quad \frac{\partial \ell}{\partial \sigma}(\mu, \sigma) = -\frac{1}{\sigma} + \frac{(X-\mu)^2}{\sigma^3}.$$

   The Hessian matrix is

$$H = \begin{bmatrix} \frac{\partial^2 \ell}{\partial \mu^2}(\mu, \sigma) & \frac{\partial^2 \ell}{\partial \mu \partial \sigma}(\mu, \sigma) \\ \frac{\partial^2 \ell}{\partial \sigma \partial \mu}(\mu, \sigma) & \frac{\partial^2 \ell}{\partial \sigma^2}(\mu, \sigma) \end{bmatrix} = \begin{bmatrix} -1/\sigma^2 & -2(X-\mu)/\sigma^3 \\ -2(X-\mu)/\sigma^3 & 1/\sigma^2 - 3(X-\mu)^2/\sigma^4 \end{bmatrix}$$

   The fisher information matrix is

$$\begin{aligned} I(\mu, \sigma) &= \begin{bmatrix} 1/\sigma^2 & 2E(X-\mu)/\sigma^3 \\ 2E(X-\mu)/\sigma^3 & -1/\sigma^2 + 3E[(X-\mu)^2]/\sigma^4 \end{bmatrix} \\ (E(X-\mu) = 0, E[(X-\mu)^2] = \sigma^2) \quad &= \begin{bmatrix} 1/\sigma^2 & 0 \\ 0 & 2/\sigma^2 \end{bmatrix}. \end{aligned}$$

The joint fisher information matrix is

$$I_n(\mu,\sigma) = nI(\mu,\sigma) = \begin{bmatrix} n/\sigma^2 & 0 \\ 0 & 2n/\sigma^2 \end{bmatrix}.$$

Its inverse is

$$J_n(\mu,\sigma) = [I_n(\mu,\sigma)]^{-1} = \begin{bmatrix} \sigma^2/n & 0 \\ 0 & \sigma^2/2n \end{bmatrix}.$$

2. Let $X_1, \ldots, X_n \overset{iid}{\sim} Gamma(\alpha, \beta)$.

   (a) Find the MLE of β assuming α is known.

   (b) Find the Fisher information and construct an approximate 95% normal-based confidence interval for β.

   (c) When both α and β are unknown, there is no closed-form expression for the MLE. The file `berkeleyprecip.csv` on bSpace contains total monthly precipitation data for Berkeley, CA, going back to 1919. In `R`, calculate the total winter precipitation for each year (removing missing values) using

```
precip <- read.csv("berkeleyprecip.csv", header = TRUE)
precip[precip==-99999] <- NA # Missing values
winter.precip <- precip$DEC + precip$JAN + precip$FEB
winter.precip <- winter.precip[!is.na(winter.precip)]
```

   Numerically find the MLEs for α and β under the model that the values for each year are *iid* with distribution $Gamma(\alpha, \beta)$. Approximate the observed Fisher information matrix and use it to construct 95% normal-based confidence intervals for α and β. Hint: Look at the steps in `betaexample.R`. Turn in your MLEs and confidence intervals along with a comment about the numerical optimization: what evidence do you have about whether the algorithm found a global optimum?

Solution:

(a) The likelihood of β is

$$L_n(\beta) = \prod_{i=1}^{n} f(X_i; \beta|\alpha) = \prod_{i=1}^{n} \frac{X_i^{\alpha-1} e^{-X_i/\beta}}{\Gamma(\alpha)\beta^\alpha} = \frac{e^{-\sum_{i=1}^n X_i/\beta}}{[\Gamma(\alpha)]^n \beta^{n\alpha}} \left(\prod_{i=1}^n X_i\right)^{\alpha-1}.$$

The log-likelihood is

$$\ell_n(\beta) = \frac{-\sum_{i=1}^n X_i}{\beta} - n\log\Gamma(\alpha) - n\alpha\log\beta + (\alpha-1)\log\left(\prod_{i=1}^n X_i\right).$$

By taking the first derivative against β and setting it as 0, we have

$$\frac{d\ell_n}{d\beta}(\beta) = \frac{\sum_{i=1}^n X_i}{\beta^2} - \frac{n\alpha}{\beta} = 0 \Rightarrow \hat{\beta} = \frac{\sum_{i=1}^n X_i}{n\alpha}.$$

The second derivative is

$$\frac{d^2\ell_n}{d\beta^2}(\beta) = -\frac{2\sum_{i=1}^n X_i}{\beta^3} + \frac{n\alpha}{\beta^2},$$

$$\Rightarrow \frac{d^2\ell_n}{d\beta^2}(\hat{\beta}) = -\frac{2\sum_{i=1}^n X_i}{\hat{\beta}^3} + \frac{n\alpha}{\hat{\beta}^2} = -\frac{2(n\alpha)^3}{(\sum_{i=1}^n X_i)^2} + \frac{(n\alpha)^3}{(\sum_{i=1}^n X_i)^2} = -\frac{2(n\alpha)^3}{(\sum_{i=1}^n X_i)^2} < 0.$$

Because $\hat{\beta}$ is the unique point where the derivative is 0 and it is a local maximum, it is a global maximum. That is, $\hat{\beta}$ is the MLE.

(b) The fisher information of β (based on n observations) is

$$\begin{aligned} I_n(\beta) &= E_\beta\left[-\frac{d^2\ell_n}{d\beta^2}(\beta)\right] = E_\beta\left[\frac{2\sum_{i=1}^n X_i}{\beta^3} - \frac{n\alpha}{\beta^2}\right] \\ (E[X_i] = \alpha\beta) \quad &= \frac{2n\alpha\beta}{\beta^3} - \frac{n\alpha}{\beta^2} = \frac{n\alpha}{\beta^2}. \end{aligned}$$

An approximate 95% normal-based confidence interval for β is

$$\hat{\beta} \pm z_{0.025}\frac{1}{\sqrt{I_n(\hat{\beta})}} = \hat{\beta} \pm z_{0.025}\frac{\hat{\beta}}{\sqrt{n\alpha}} = \frac{\sum_{i=1}^n X_i}{n\alpha} \pm z_{0.025}\frac{\sum_{i=1}^n X_i}{(n\alpha)^{3/2}}.$$

(c) The `R` codes are as follows:

```
# Read in the data
precip <- read.csv("berkeleyprecip.csv", header = TRUE)
precip[precip==-99999] <- NA # Missing values
winter.precip <- precip$DEC + precip$JAN + precip$FEB
winter.precip <- winter.precip[!is.na(winter.precip)]
```

```
# Write a function for the negative log-likelihood
nll <- function(par, x, verbose = FALSE){
alpha <- par[1]; beta <- par[2] # unpack
ll <- sum(dgamma(x, shape=alpha, scale=beta, log=TRUE))
if (verbose) print(c(par, -ll))
return(-ll)
}

# Numerically minimize it
start <- c(alpha = 1, beta = 1) # starting values
eps <- 1e-10 # small value for lower bounds
op <- optim(par = start, fn = nll,
lower = rep(eps, 2), hessian = TRUE,
x = winter.precip, verbose = TRUE)
mle <- op$par

# Examine the nll at the min (evidence for local mimima)
alpha.test <- cbind(seq(mle[1]/2, mle[1]*2, length = 100), mle[2])
nll.alpha <- apply(alpha.test, 1, nll, x = winter.precip, verbose = FALSE)
plot(alpha.test[,1], nll.alpha, type = "l")

beta.test <- cbind(mle[1], seq(mle[2]/2, mle[2]*2, length = 100))
nll.beta <- apply(beta.test, 1, nll, x = winter.precip, verbose = FALSE)
plot(beta.test[,2], nll.beta, type = "l")

# Obtain the observed Fisher information matrix and CIs
I <- op$hessian # I is the observed Fisher information matrix
J <- solve(I) # no negative - already working with negative ll
se.hat <- sqrt(diag(J))

lower <- mle - qnorm(0.975)*se.hat
upper <- mle + qnorm(0.975)*se.hat

# Check if the MLEs are the global maximum (try different starting values)
check_global_max <- function(start) {
eps <- 1e-10 # small value for lower bounds
op <- optim(par = start, fn = nll,
lower = rep(eps, 2), method="L-BFGS-B",
```

```
x = winter.precip, verbose = FALSE)
return(op$par)
}

starts <- cbind(runif(50,0,1000), runif(50,0,1000))
mles <- apply(starts, 1, check_global_max)
# The mles are unique
```

The MLEs for α and β are

$$\hat{\alpha} = 5.26, \quad \hat{\beta} = 257.62.$$

The observed Fisher information matrix is

$$I_n^{obs}(\alpha, \beta) = \begin{bmatrix} 17.38 & 0.32 \\ 0.32 & 0.01 \end{bmatrix}.$$

95% normal-based CIs for α and β are

$$\hat{\alpha} \pm z_{0.025}\hat{se}(\hat{\alpha}) = [3.71, 6.81];$$

$$\hat{\beta} \pm z_{0.025}\hat{se}(\hat{\beta}) = [177.85, 337.40].$$

3. Let $X_1, \dots, X_n \overset{iid}{\sim} Unif(0, \theta)$. Show that the MLE is consistent. Hint: Let $Y = \max\{X_1, \dots, X_n\}$. For any c, $P(Y < c) = P(X_1 < c, X_2 < c, \dots, X_n < c) = P(X_1 < c)P(X_2 < c) \cdots P(X_n < c)$.

   Solutions:

   We already know that the MLE for θ is $\hat{\theta}_n = \max\{X_1, \dots, X_n\}$.

   Note that $\hat{\theta}_n \le \theta$, since $X_1, \dots, X_n \le \theta$.

   For any $\epsilon > 0$,

   $$P(|\hat{\theta}_n - \theta| > \epsilon) = P(\hat{\theta}_n < \theta - \epsilon) = [P(X_1 < \theta - \epsilon)]^n = \left(\frac{\theta - \epsilon}{\theta}\right)^n = \left(1 - \frac{\epsilon}{\theta}\right)^n \overset{n \to \infty}{\to} 0.$$

   So $\hat{\theta} \overset{P}{\to} \theta$, and $\hat{\theta}$ is consistent.

4. Let $X_1, \dots, X_n \overset{iid}{\sim} N(\theta, 1)$. Define $Y_i = I\{X_i > 0\}$. Let $\psi = P(Y_1 = 1)$.

   (a) Find the MLE of ψ.

(b) Find an approximate 95% confidence interval for ψ.

(c) Define $\tilde{\psi} = \frac{1}{n}\sum_{i=1}^{n} Y_i$. Show that $\tilde{\psi}$ is a consistent estimator of ψ.

(d) Compute the asymptotic relative efficiency of $\tilde{\psi}$ to $\hat{\psi}$. Hint: Use the delta method to get the standard error of the MLE. Then compute the standard error (i.e., the standard deviation) of $\tilde{\psi}$.

(e) Suppose that the data are not really normal. Show that $\hat{\psi}$ is not consistent. What, if anything, does $\hat{\psi}$ converge to?

Solution:

(a)

$$\psi = P(Y_1 = 1) = P(X_1 > 0) = P(X_1 - \theta > -\theta) = P(Z > -\theta) = 1 - \Phi(-\theta) = \Phi(\theta),$$

where $Z \sim N(0,1)$.
Since the MLE of θ is $\hat{\theta} = \bar{X}_n$, the MLE of ψ is

$$\hat{\psi} = 1 - \Phi(-\hat{\theta}) = 1 - \Phi(-\bar{X}_n) = \Phi(\bar{X}_n).$$

(b) Let $g(\theta) = 1 - \Phi(-\theta) = \Phi(\theta)$. Then $g'(\theta) = \phi(\theta)$.
Based on Delta method (approximated from Taylor expansion),

$$se(\hat{\psi}) \approx se(\hat{\theta})|g'(\theta)| = (1/\sqrt{n})\phi(\theta) = \phi(\theta)/\sqrt{n}.$$

Then the approximately estimated standard error of $\hat{\psi}_n$ is

$$\hat{se}(\hat{\psi}) = \phi(\hat{\theta})/\sqrt{n}.$$

(c) By the Weak Law of Large Numbers (WLLN),

$$\tilde{\psi} = \frac{1}{n}\sum_{i=1}^{n} Y_i \xrightarrow{P} E[Y_1] = P(Y_1 = 1) = \psi.$$

So $\tilde{\psi}$ is a consistent estimator of ψ.

(d) By Central Limit Theorem,

$$\sqrt{n}(\tilde{\psi} - \psi) \xrightarrow{D} N(0, V[Y_i])$$

Note that $Y_i \sim Bernoulli(p)$, where $p = P(X_i > 0) = \Phi(\theta)$. Hence $V[Y_i] = \Phi(\theta)[1 - \Phi(\theta)]$.

So the asymptotic variance of $\tilde{\psi}$ is

$$V[\tilde{\psi}] = \frac{\Phi(\theta)[1-\Phi(\theta)]}{n}.$$

From (b) the asymptotic variance of $\hat{\psi}$ is

$$V[\hat{\psi}] = \frac{[\phi(\theta)]^2}{n}.$$

Therefore, the asymptotic relative efficiency of $\tilde{\psi}$ to $\hat{\psi}$ is

$$ARE(\tilde{\psi},\hat{\psi}) = \frac{V[\tilde{\psi}]}{V[\hat{\psi}]} = \frac{\Phi(\theta)[1-\Phi(\theta)]}{[\phi(\theta)]^2}.$$

(e) Suppose that the data are not really normal, but still $E[X_i] = \theta$ and $V[X_i] = 1$. We can show that $\hat{\psi}$ is not consistent. A counter-example is as follows.

$$X_i = \begin{cases} \theta+1 & \text{with probability } 1/2 \\ \theta-1 & \text{with probability } 1/2 \end{cases}$$

Then

$$\psi = P(Y_i = 1) = P(X_i > 0) = \begin{cases} 0 & \text{if } \theta \le -1 \\ 1/2 & \text{if } -1 < \theta \le 1 \\ 1 & \text{if } \theta > 1 \end{cases}$$

When $\theta \le -1$,

$$\begin{aligned} P(|\hat{\psi}-\psi| > \epsilon) &= P(\Phi(\bar{X}_n) > \epsilon) = P(\bar{X}_n > \Phi^{-1}(\epsilon)) \\ &= P(\sqrt{n}(\bar{X}_n - \theta) > \sqrt{n}(\Phi^{-1}(\epsilon) - \theta)) \\ &= 1 - \Phi(\sqrt{n}(\Phi^{-1}(\epsilon) - \theta))0, \end{aligned}$$

for small $\epsilon > 0$ such that $\Phi^{-1}(\epsilon) - \theta < 0$.
Therefore, we have shown that $\hat{\psi}$ is not consistent in this case.
From (a), we have that $\hat{\psi} = \Phi(\bar{X}_n)$, and by WLLN, $\bar{X}_n \xrightarrow{P} \theta$. From Theorem 5.5 (Page 75 in the textbook),

$$\hat{\psi} = \Phi(\bar{X}_n) \xrightarrow{P} \Phi(\theta).$$

So we know that $\hat{\psi}$ always converges in probability to $\Phi(\theta)$. When $X_1,\ldots,X_n$ are not really normal, $\psi \neq \Phi(\theta)$. Then $\hat{\psi}^P\psi$ and thus $\hat{\psi}$ is not a consistent estimator of ψ.