Homework 2 Hanze Yao

3032093286

1、 I assume all 100 booklets' weight $\overset{iid}{\sim} N(1, 0.0025)$

Then because the weight of those 100 booklets are independent, the total weight of them also satisfies normal distribution which is $N(100, 0.25)$

then use normalization

$$\frac{N(100, 0.25) - 100}{\sqrt{0.25}} \sim N(0, 1)$$

$$\therefore \frac{100.4 - 100}{\sqrt{0.25}} = 0.8$$

$\therefore$ P (100 booklets weigh more than 100.4 ounces)

$= 1 - \underset{(\Phi(0.8))}{78.81\%} = 21.19\%$

2、(1) $E(\bar{X}_n^* \mid X_1 \cdots X_n) = E(\frac{1}{n}\sum_{i=1}^n X_i^* \mid X_1 \cdots X_n)$

$= (X_1^* \mid X_1 \cdots X_n) = E_{ECDF}(X_1^*)$ because bootstrap samples are independent from $X_1 \cdots X_n$ $= \sum_{i=1}^n X_i P(X_1^* = X_i)$

$= \sum_{i=1}^n (X_i \times \frac{1}{n}) = \frac{1}{n}\sum_{i=1}^n X_i = \bar{X}_n$ but sample from $X_1 \cdots X_n$

(2) $V(\bar{X}_n^* \mid X_1 \cdots X_n) = V(\frac{1}{n}\sum_{i=1}^n X_i^* \mid X_1 \cdots X_n)$

$= V(\frac{1}{n}\sum_{i=1}^n X_i^*) = \frac{1}{n^2} \times V(\sum_{i=1}^n X_i^*) = \frac{1}{n^2} \times n\, V(X_1^*) = \frac{1}{n} V(X_1^*)$

$= \frac{1}{n} E[X_1^* - E(X_1^*)]^2 = \frac{1}{n} \times \sum_{i=1}^n [X_i - E(X_1^*)]^2 \times P(X_1^* = X_i)$

$= \frac{1}{n} \times \sum_{i=1}^n (X_i - \bar{X}_n)^2 \times \frac{1}{n} = \frac{1}{n^2}\sum_{i=1}^n (X_i - \bar{X}_n)^2$

(3) $\underbrace{E(\bar{X}_n^*) = E(E(\bar{X}_n^* \mid X_1 \cdots X_n)) = E(\bar{X}_n) = E(X_1)}_{\text{if } X_1 \cdots X_n \text{ are i.i.d}}$

if $E(X_1) = \mu$ then $E(\bar{X}_n^*) = E(\bar{X}_n) = E(X_1) = \mu$

(4) $V(\bar{X}_n^*) = V(E(\bar{X}_n^* \mid X_1 \cdots X_n)) + E(V(\bar{X}_n^* \mid X_1 \cdots X_n))$

$= V(\bar{X}_n) + E(\frac{1}{n^2}\sum_{i=1}^n (X_i - \bar{X}_n)^2)$

if $X_1 \cdots X_n$ iid and $Var(X_i) = \sigma^2$

then : $= \frac{\sigma^2}{n} + \frac{1}{n^2} E(\sum_{i=1}^n (X_i - \bar{X}_n)^2)$

$= \frac{\sigma^2}{n} + \frac{1}{n^2}(\sum_{i=1}^n E(X_i^2) - n E(\bar{X}_n^2))$

$= \frac{\sigma^2}{n} + \frac{1}{n^2}(n(V(X_i) + (EX_i)^2) - n(V(\bar{X}_n) + E^2(\bar{X}_n)))$

$= \frac{\sigma^2}{n} + \frac{1}{n^2}(n(\sigma^2 + \mu^2) - n(\frac{\sigma^2}{n} + \mu^2))$

$= \frac{\sigma^2}{n} + \frac{1}{n^2}(n\sigma^2 + n\mu^2 - \sigma^2 - n\mu^2) = \frac{\sigma^2}{n} + \frac{n-1}{n^2}\sigma^2 = \frac{2n-1}{n^2}\sigma^2$

3. $L_n(\theta) = \prod_{i=1}^{n} P_\theta(X = x_i) = \theta^{\sum_{i=1}^{n} x_i} (1-\theta)^{n - \sum_{i=1}^{n} x_i}$

$l_n(\theta) = \log\left[\theta^{\sum_{i=1}^{n} x_i} (1-\theta)^{n - \sum_{i=1}^{n} x_i}\right]$

$= \sum_{i=1}^{n} x_i (\log\theta) + \left(n - \sum_{i=1}^{n} x_i\right) \log(1-\theta)$

$\frac{\partial}{\partial\theta} l_n(\theta) = \sum_{i=1}^{n} x_i \times \frac{1}{\theta} + \left[\left(n - \sum_{i=1}^{n} x_i\right) \times \frac{1}{1-\theta} \times (-1)\right] = 0$

$(\theta - 1) \sum_{i=1}^{n} x_i + \theta\left(n - \sum_{i=1}^{n} x_i\right) = 0$

$n\theta = \sum_{i=1}^{n} x_i \qquad \theta = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{X}_n$

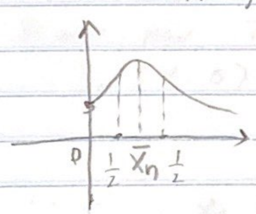
when $\theta < \bar{X}_n$: $\frac{\partial}{\partial\theta} l_n(\theta) = \frac{1}{\theta}\sum_{i=1}^{n} x_i + \frac{1}{\theta - 1}\left(n - \sum_{i=1}^{n} x_i\right)$

$= \frac{-1}{\theta(\theta-1)}\sum_{i=1}^{n} x_i + \frac{n}{\theta-1} = \frac{n\theta - \sum_{i=1}^{n} x_i}{\theta(\theta-1)}$

∵ $0 < \theta \le \frac{1}{2}$ ∴ $\theta(\theta-1) < 0$ ∵ $\theta < \bar{X}_n$ ∴ $\frac{\partial}{\partial\theta} l_n(\theta) > 0$

∴ using the same method, get when $\theta > \bar{X}_n$, $\frac{\partial}{\partial\theta} l_n(\theta) < 0$

∴ The graph of $l_n(\theta)$ looks approximately like this.



∴ When $\bar{X}_n \ge \frac{1}{2}$, then MLE of θ is $\frac{1}{2}$

when $\bar{X}_n < \frac{1}{2}$, then MLE of θ is $\bar{X}_n$

4. Now we only have one x. So, $L(\theta) = f(x;\theta)$

So, the MLE of θ will maximize $L(\theta)$ at different X value respectively.

∴

x	$\max f(x;\theta)$	θ with $\max f(x;\theta)$
0	$\frac{1}{3}$	1
1	$\frac{1}{3}$	1
2	$\frac{1}{4}$	2 and 3
3	$\frac{1}{2}$	3
4	$\frac{1}{4}$	3

∵ $\theta \in \{1, 2, 3\}$ ∴ the θ with $\max f(x;\theta)$ at each x value is the MLE of θ.

5. when $X_1, \dots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$, the density for $X_1 \dots X_n$ is

$f(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ then $\log[f(x;\mu,\sigma^2)] = -\frac{(x-\mu)^2}{2\sigma^2} - \log(\sqrt{2\pi}\sigma)$

$= -\frac{(x-\mu)^2}{2\sigma^2} - \log\sigma - \log\sqrt{2\pi}$

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

$\frac{\partial^2}{\partial\sigma\partial\mu}(\log f) = \frac{\partial^2}{\partial\mu\partial\sigma}\log f(x;\mu,\sigma) = \frac{\partial^2}{\partial\mu\partial\sigma}\left(-\frac{(x-\mu)^2}{2\sigma^2}-\log\sigma-\log\sqrt{2\pi}\right) = -\frac{2(x-\mu)}{\sigma^3}$

$\frac{\partial}{\partial\sigma^2}\log f(x;\mu,\sigma) = \frac{\partial^2}{\partial\sigma^2}\left(-\frac{(x-\mu)^2}{2\sigma^2}-\log\sigma-\log\sqrt{2\pi}\right) = -\frac{3(x-\mu)^2}{\sigma^4}+\frac{1}{\sigma^2}$

Then $I_n(\theta) = nI_1(\theta) = n\begin{pmatrix} -E_\theta(-\frac{1}{\sigma^2}) & -E_\theta(-\frac{2(x-\mu)}{\sigma^3}) \\ -E_\theta(-\frac{2(x-\mu)^2}{\sigma^3}) & -E_\theta(\frac{1}{\sigma^2}-\frac{3(x-\mu)^2}{\sigma^4}) \end{pmatrix}$

$= n\begin{bmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & -\frac{1}{\sigma^2}+\frac{3}{\sigma^2}=\frac{2}{\sigma^2} \end{bmatrix} = \begin{bmatrix} \frac{n}{\sigma^2} & 0 \\ 0 & \frac{2n}{\sigma^2} \end{bmatrix}$

because $E(X_i-\mu) = 0$

because $E(X_i-\mu)^2 = Var(X_i) = \sigma^2$

Because now $I_n(\theta)$ is a 2×2 matrix

$J_n(\mu,\sigma) = I_n(\mu,\sigma)^{-1} = \begin{pmatrix} \frac{\sigma^2}{n} & 0 \\ 0 & \frac{\sigma^2}{2n} \end{pmatrix}$

6. $\because X_1 \dots X_n \overset{iid}{\sim} Unif(0,\theta)$

$\therefore f(x;\theta) = \frac{1}{\theta}I\{0<x<\theta\}$

$\therefore L_n(\theta) = \frac{1}{\theta^n}\prod_{i=1}^n I\{0<X_i<\theta\}$ To maximizes $L_n(\theta)$ here

I need to get the smallest θ possible.

$\because \prod_{i=1}^n I\{0<X_i<\theta\} = \prod_{i=1}^n(0<X_{(1)}<X_{(2)}\cdots<X_{(n)}<\theta)$

$\therefore$ the minimum θ will be $X_{(n)}$ ($\max(X_i)$)

To show that the MLE of θ is consistent, I need to show that $\hat\theta_n \xrightarrow{P} \theta$ then I need to show that:

$P\{|\hat\theta_n-\theta|>\varepsilon\} \xrightarrow{p} 0$ for any $\varepsilon>0$

$\because \hat\theta_n$ is the minimum θ possible $\therefore \hat\theta_n-\theta\le 0$

$\therefore P\{\theta-\hat\theta_n>\varepsilon\} = P\{\hat\theta_n<\theta-\varepsilon\} = P\{\max(X_i)<\theta-\varepsilon\}$

$= [P\{X_1<\theta-\varepsilon\}]^n = [\frac{\theta-\varepsilon}{\theta}]^n = (1-\frac{\varepsilon}{\theta})^n$ $\because 0<1-\frac{\varepsilon}{\theta}<1$

$\therefore (1-\frac{\varepsilon}{\theta})^n \longrightarrow 0$ as $n\to\infty$ $\therefore P\{\theta-\hat\theta_n>\varepsilon\}\to 0$ as $n\to\infty$ for any $\varepsilon>0$

$\therefore \hat\theta_n \xrightarrow{P} \theta$

7 (a) $\psi = P(Y_1 = 1) = P(X_1 > 0) = P(X_1 - \theta > -\theta)$

$\because X_1 \sim N(\theta, 1) \quad \therefore X_1 - \theta \sim N(0,1)$

$\therefore \psi = 1 - \Phi(-\theta) = \Phi(\theta)$

$\because X_1 \cdots X_n \overset{iid}{\sim} N(\theta, 1)$

$\therefore l_n(\theta) = \log\left(\left(\frac{1}{\sqrt{2\pi}}\right)^n e^{-\frac{\sum_{i=1}^n (X_i-\theta)^2}{2}}\right)$

$= n\log\left(\frac{1}{\sqrt{2\pi}}\right) - \frac{\sum_{i=1}^n (X_i-\theta)^2}{2}$

$\frac{\partial}{\partial\theta} l_n(\theta) = -\frac{1}{2} \times \sum_{i=1}^n 2(X_i - \theta) \times (-1) = \sum_{i=1}^n (X_i - \theta) = 0$

$\sum_{i=1}^n X_i - n\theta = 0 \quad n\theta = \sum_{i=1}^n X_i \quad \theta = \frac{1}{n}\sum_{i=1}^n X_i = \bar{X}_n$

$\therefore$ MLE of θ is $\bar{X}_n$

$\therefore$ MLE of ψ is $\Phi(\bar{X}_n)$

(b) $\because \Phi(x)$ is the CDF of standard normal distribution.

$\therefore \Phi'(x)$ is the PDF of standard normal distribution

$\therefore \Phi'(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$

$\because \frac{\hat{\psi} - \psi}{se(\hat{\psi})} \xrightarrow{D} N(0,1) \quad \therefore \frac{\Phi(\bar{X}_n) - \psi}{se(\hat{\psi})} \xrightarrow{D} N(0,1)$

$se(\hat{\psi}) \approx se(\hat{\theta})|\Phi'(\hat{\theta})| = se(\bar{X}_n)|\Phi'(\hat{\theta})| = \sqrt{\frac{1}{n}}\,|\Phi'(\hat{\theta})|$

$= \frac{\Phi'(\hat{\theta})}{\sqrt{n}} = \frac{1}{\sqrt{n}} \times \frac{1}{\sqrt{2\pi}} e^{-\frac{\hat{\theta}^2}{2}} = \frac{e^{-\frac{\bar{X}_n^2}{2}}}{\sqrt{2n\pi}}$

$\therefore \frac{\Phi(\bar{X}_n) - \psi}{\frac{e^{-\frac{\bar{X}_n^2}{2}}}{\sqrt{2n\pi}}} \xrightarrow{D} N(0,1)$

CI: $Z_{0.025} \le \frac{\Phi(\bar{X}_n) - \psi}{\frac{e^{-\frac{\bar{X}_n^2}{2}}}{\sqrt{2n\pi}}} \le -Z_{0.025}$

$C_n: \Phi(\bar{X}_n) \pm Z_{0.025} \frac{e^{-\frac{\bar{X}_n^2}{2}}}{\sqrt{2n\pi}}$

(c) Use Weak law of large numbers

$\tilde{\psi} = \frac{1}{n}\sum_{i=1}^n Y_i \xrightarrow{P} E(Y_1) = E(I\{X_1 > 0\}) = P\{X_1 > 0\} = \psi$

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

(d) $\sqrt{n}(\tilde{\psi} - \psi) \xrightarrow{D} N(0, n\,Var(\tilde{\psi}))$ $\quad n\,Var(\tilde{\psi}) = n \times \frac{1}{n^2} \times n \times Var(Y_i)$

$= Var(Y_i)$

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

$\because Y_i = I\{X_i > 0\}$ $\quad \therefore Y_i \sim Ber(P\{X_i > 0\})$ $\therefore Y_i \sim Ber(\Phi(\theta))$

$\therefore V(Y_i) = \Phi(\theta)(1-\Phi(\theta))$

$\therefore V(\tilde{\psi}) = \dfrac{\Phi(\theta)(1-\Phi(\theta))}{n}$

$\because se(\hat{\psi}) \approx \dfrac{\Phi'(\theta)}{\sqrt{n}}$ $\quad \therefore V(\hat{\psi}) \approx \dfrac{(\Phi'(\theta))^2}{n}$

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

$$\frac{V(\tilde{\psi})}{V(\hat{\psi})} = \frac{\Phi(\theta)(1-\Phi(\theta))}{(\Phi'(\theta))^2} = \frac{\Phi(\theta)(1-\Phi(\theta))}{\left(\frac{e^{-\frac{\theta^2}{2}}}{\sqrt{2\pi}}\right)^2}$$

$$= \frac{\Phi(\theta)(1-\Phi(\theta))}{\frac{e^{-\theta^2}}{2\pi}} = 2\pi e^{\theta^2}\Phi(\theta)(1-\Phi(\theta))$$

e) $\because E(X_i) = \theta$, $Var(X_i) = 1$ but X_i is not normal distribution

$\therefore$ One possibility is that $X_i = \begin{cases} \theta+1 & \frac{1}{2} \\ \theta-1 & \frac{1}{2} \end{cases}$ (Probability)

Prove: $E(X_i) = (\theta+1)\times\frac{1}{2} + (\theta-1)\times\frac{1}{2} = \theta$

$Var(X_i) = E(X_i^2) - (EX_i)^2 = (\theta+1)^2\times\frac{1}{2} + (\theta-1)^2\times\frac{1}{2} - \theta^2 = 1$

$\therefore$ This example is valid

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

if $\theta \le -1$

then: $P(|\hat{\psi} - \overset{=0}{\psi}| > \varepsilon) = P(|\hat{\psi}| > \varepsilon) = P(\Phi(\bar{X}_n) > \varepsilon)$

$= P(\bar{X}_n > \Phi^{-1}(\varepsilon))$ according to Central Limit Theorem

$= P(\sqrt{n}(\bar{X}_n - \theta) > \sqrt{n}(\Phi^{-1}(\varepsilon) - \theta))$

$= 1 - \Phi(\sqrt{n}(\Phi^{-1}(\varepsilon) - \theta))$

when ε is a small positive number $\Phi^{-1}(\varepsilon)$ will be negative

then $\sqrt{n}(\Phi^{-1}(\varepsilon) - \theta) \to -\infty$ as $n \to \infty$

then $\Phi(\sqrt{n}(\Phi^{-1}(\varepsilon) - \theta)) \to 0$ as $n \to \infty$

then $1 - \Phi(\sqrt{n}(\Phi^{-1}(\varepsilon) - \theta)) \to 1$ as $n \to \infty$

$\therefore \hat{\psi} \overset{P}{\nrightarrow} \psi$

$\because \hat{\psi} = \Phi(\bar{X}_n)$, $\bar{X}_n \xrightarrow{P} E(X_1) = \theta$ according to weak law of large numbers $\therefore \Phi(\bar{X}_n) \xrightarrow{P} \Phi(\theta)$ which is $\hat{\psi} \xrightarrow{P} \Phi(\theta)$

$\therefore \hat{\psi}$ always $\xrightarrow{P} \Phi(\theta)$ no matter $X_1 \cdots X_n$ are normal or not.

8. (a) $\because \alpha$ is known $\therefore f(x;\beta) = \dfrac{\beta^{\alpha} x^{\alpha-1} e^{-\beta x}}{\Gamma(\alpha)}$

$$l_n(\beta) = \log\left(\prod_{i=1}^{n} \frac{\beta^{\alpha} x_i^{\alpha-1} e^{-\beta x_i}}{\Gamma(\alpha)}\right) = \log\left[\left(\frac{1}{\Gamma(\alpha)}\right)^n \beta^{n\alpha}\right] + (\alpha-1)\sum_{i=1}^{n} \log X_i - \beta \sum_{i=1}^{n} X_i$$

$$= -n\log(\Gamma(\alpha)) + n\alpha \log\beta + (\alpha-1)\sum_{i=1}^{n} \log X_i - \beta \sum_{i=1}^{n} X_i$$

$$\frac{\partial}{\partial\beta} l_n(\beta) = \frac{n\alpha}{\beta} - \sum_{i=1}^{n} X_i = 0 \iff \beta = \frac{\alpha}{\frac{1}{n}\sum_{i=1}^{n} X_i} = \frac{\alpha}{\bar{X}_n} > 0$$

. It satisfies the range of β. $\therefore \beta = \frac{\alpha}{\bar{X}_n}$ will maximize $l_n(\beta)$

$\therefore$ The MLE of β is $\frac{\alpha}{\bar{X}_n}$.

(b) $\because$ Gamma distribution belongs to exponential family

$$\therefore I_n(\beta) = -E_\beta\left(\frac{\partial^2}{\partial\beta^2} l_n(\beta)\right) = -E_\beta\left((-1) \times n\alpha \times \frac{1}{\beta^2}\right) = -E_\beta\left(-\frac{n\alpha}{\beta^2}\right)$$

$$= E_\beta\left(\frac{n\alpha}{\beta^2}\right) = \frac{n\alpha}{\beta^2}$$

$$V(\hat{\beta}_n) = \frac{1}{I_n(\hat{\beta}_n)} = \frac{\hat{\beta}_n^2}{n\alpha} = \frac{\left(\frac{\alpha}{\bar{X}_n}\right)^2}{n\alpha} = \frac{\alpha^2}{n\alpha \bar{X}_n^2} = \frac{\alpha}{n\bar{X}_n^2}$$

$$\therefore \frac{\beta - \hat{\beta}_n}{\sqrt{V(\hat{\beta}_n)}} = \frac{\beta - \frac{\alpha}{\bar{X}_n}}{\sqrt{\frac{\alpha}{n\bar{X}_n^2}}} \xrightarrow{D} N(0,1)$$

$\therefore$ The 95% confidence interval for β is:

$$\frac{\alpha}{\bar{X}_n} \pm Z_{1-\frac{0.95}{2}} \sqrt{\frac{\alpha}{n\bar{X}_n^2}} \Rightarrow \frac{\alpha}{\bar{X}_n} \pm Z_{0.025} \sqrt{\frac{\alpha}{n\bar{X}_n^2}}$$

(c) First, delete the code in betaexample.R that generates the 100 x samples $\sim$ Beta(5, 4). Then add the code in this question on top to import the precipitation data of Berkeley. Second, because in this question it says "Assume the values for each year are iid with distribution Gamma(α, β). So change dbeta into dgamma. Finally, change all $x = x$ in the code into x = winter.precip. After all the changes above, I can run the code.

	MLE α	MLE β	
I get	6.204	0.004651	by typing "mle" in the Console

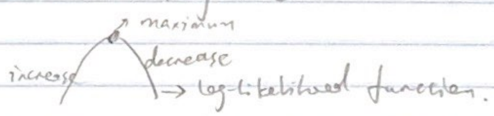
The fisher information matrix is $\begin{pmatrix} 14.514 & -18128.85 \\ -18128.85 & 26322409.88 \end{pmatrix}$ by typing "op $ hessian" in the Console.

	α	β
By typing "lower" in the console I get	4.800	0.003608
By typing "upper" in the Console I get	7.608	0.005694

So the 95% CI for α is (4.800, 7.608)

So the 95% CI for β is (0.03608, 0.005694)

The evidence for this algorithm to find a global optimum is that the second derivative of log-likelihood function to α and β are both negative.



$\therefore$ Second derivative must be negative for α and β.

Note: The plots and output in R are on the next page.