

Probability & Statistics for EECS:

Homework #13

Due on May 14, 2023 at 23:59

Name: **Penghao Wang**
Student ID: 2021533138

Problem 1

- (a) As for the distribution of $N - 1$, as X_1, X_2, \dots i.i.d. $\text{Expo}(1)$, we have that X_1, X_2, \dots are independent, and each X has probability of $\frac{1}{e}$ to exceed 1. As to the definition of Geometric distribution, which is the number of trials to get the first success, we have that $N - 1$ follows Geometric distribution with parameter $\frac{1}{e}$. So we get that

$$N - 1 \sim \text{Geom}\left(\frac{1}{e}\right),$$

then with the property of Geometric distribution, we have that

$$E(N) = E(N - 1) + 1 = \frac{1 - 1/e}{1/e} + 1 = e - 1 + 1 = e.$$

So in conclusion, we have that the distribution is

$$N - 1 \sim \text{Geom}\left(\frac{1}{e}\right)$$

and the expectation is

$$E(N) = e.$$

- (b) As for the

$$\min\{n : X_1 + X_2 + \dots + X_n \geq 10\},$$

it is obvious that this can be considered as Poisson process as we calculate the sum of X_i and observe them until sum exceeds 10, which is the same as Poisson process is the number of arrivals until the time exceeds 10, where the arrival interval follows $\text{Expo}(1)$. So we can consider X_1, X_2, \dots, X_n as the interarrival times in a Poisson process with rate 1. where the range of the time is $[0, 10)$. Then we have that

$$M - 1 \sim \text{Pois}(10)$$

and we have that with the property of Poisson distribution, the $E(M)$ is

$$E(M) = E(M - 1 + 1) = E(M - 1) + 1 = 10 + 1 = 11$$

So we have that the distribution is

$$M - 1 \sim \text{Pois}(10)$$

and the expectation is

$$E(M) = 11.$$

- (c) As for the \bar{X}_n , we have that

$$\bar{X}_n = \frac{(X_1 + X_2 + \dots + X_n)}{n} = \frac{X_1}{n} + \frac{X_2}{n} + \dots + \frac{X_n}{n}$$

, As we have that

$$X_1, X_2, \dots \text{i.i.d.} \text{Expo}(1),$$

we then have that

$$\frac{X_1}{n}, \frac{X_2}{n}, \dots \sim \text{Expo}(n)$$

Then we have that

$$\bar{X}_n \sim \text{Gamma}(n, n)$$

As for the approximate distribution of \bar{X}_n for n large, with the center limit theorem, we have that when n is large, the distribution will be approximately normal distribution with the same mean and variance

as the origin distribution. As we have that the origin distribution has mean of 1 and variance of 1, we have that the approximate distribution is normal distribution

$$\bar{X}_n \sim N(1, \frac{1}{n}).$$

So we have the exact distribution is Gamma distribution

$$\bar{X}_n \sim \text{Gamma}(n, n)$$

and the approximate distribution is normal distribution

$$\bar{X}_n \sim N(1, \frac{1}{n}).$$

Problem 2

To show that the inequality

$$P(|\frac{1}{n} \sum_{i=1}^n X_i - \mu| \geq \varepsilon) \leq 2\exp(-\frac{2n\varepsilon^2}{(b-a)^2}).$$

holds, we use the Hoeffding Lemma + Chernoff Inequality, the Hoeffding Lemma inequality is

$$E(e^{\lambda x}) \leq e^{\frac{1}{8}\lambda^2(b-a)^2},$$

the Chernoff Inequality is

$$P(X \geq a) \leq \frac{E(e^{tX})}{e^{ta}}.$$

Proof are as follows:

As $E(X_i) = \mu$, and that $a \leq X_i \leq b$, we then have that

$$X_i - \mu \leq b - \mu$$

$$\mu - X_i \leq \mu - a.$$

Then as for $s > 0$, we have that

$$\begin{aligned} P(\frac{1}{n} \sum_{i=1}^n X_i - \mu \geq \varepsilon) &= P(s(\frac{1}{n} \sum_{i=1}^n X_i - \mu) \geq \varepsilon s) \\ &= P(e^{s(\frac{1}{n} \sum_{i=1}^n X_i - \mu)} \geq e^{\varepsilon s}) \end{aligned}$$

with Chernoff Inequality, we have that

$$\begin{aligned} &\leq e^{-s\varepsilon} E(e^{s(\frac{1}{n} \sum_{i=1}^n X_i - \mu)}) \\ &= e^{-s\varepsilon} \prod_{i=1}^n E(e^{\frac{s}{n}(X_i - \mu)}) \end{aligned}$$

with Hoeffding Lemma, we have that

$$\begin{aligned} &\leq e^{-s\varepsilon} \prod_{i=1}^n e^{\frac{1}{8}(\frac{s}{n})^2(b-a)^2} \\ &= e^{-s\varepsilon + \frac{1}{8}(\frac{s}{n})^2(b-a)^2 n} \\ &= e^{-s\varepsilon + \frac{s^2}{8n}(b-a)^2} \end{aligned}$$

Then we try to find the minimum of $-s\varepsilon + \frac{s^2}{8n}(b-a)^2$ the to verify if the inequality satisfy. We have the derivative is

$$\frac{\partial \frac{s^2(b-a)^2}{8n} - s\varepsilon}{\partial s} = \frac{s(b-a)^2}{4n} - \varepsilon$$

Then as we have that

$$s = \frac{4n\varepsilon}{(b-a)^2} \quad \text{that we will get the minimum}$$

we get that

$$e^{-s\varepsilon + \frac{s^2}{8n}(b-a)^2} = e^{-\frac{2n\varepsilon^2}{(b-a)^2}}$$

So we get that

$$P(\frac{1}{n} \sum_{i=1}^n X_i - \mu \geq \varepsilon) \leq e^{-\frac{2n\varepsilon^2}{(b-a)^2}},$$

then the same, we will get that

$$P(\mu - \frac{1}{n} \sum_{i=1}^n X_i \leq -\varepsilon) \leq e^{-\frac{2n\varepsilon^2}{(b-a)^2}},$$

So we will get that

$$P(|\frac{1}{n} \sum_{i=1}^n X_i - \mu| \geq \varepsilon) = P(\mu - \frac{1}{n} \sum_{i=1}^n X_i \leq -\varepsilon) + P(\frac{1}{n} \sum_{i=1}^n X_i - \mu \geq \varepsilon) \leq 2e^{-\frac{2n\varepsilon^2}{(b-a)^2}}$$

So we proved

$$P(|\frac{1}{n} \sum_{i=1}^n X_i - \mu| \geq \varepsilon) \leq 2e^{-\frac{2n\varepsilon^2}{(b-a)^2}}.$$

Problem 3

From the problem, we first get that the r.v. X has expectation μ and that variance σ^2 , we denote that t is a variable that satisfy $t > 0$, we then have that

$$\begin{aligned} P(X - \mu \geq a) &= P(X - \mu + t \geq a + t) \\ &= P((X - \mu + t)^2 \geq (a + t)^2) \end{aligned}$$

Then we use the markov inequality, we have that as $(a + t)^2 > 0$ we get that

$$\begin{aligned} P((X - \mu + t)^2 \geq (a + t)^2) &\leq \frac{E((X - \mu + t)^2)}{(a + t)^2} \quad \text{With Markov inequality} \\ &= \frac{E(X^2 + t^2 + \mu^2 - 2\mu X + 2tX - 2\mu t)}{(a + t)^2} \\ &= \frac{E(X^2) + t^2 + \mu^2 - 2\mu E(X) + 2tE(X) - 2\mu t}{(a + t)^2} \end{aligned}$$

As we have that

$$\text{Var}(X) = E(X^2) - E(X)^2 = \sigma^2$$

Then the $E(X^2) = \sigma^2 + \mu^2$ So we get that

$$\begin{aligned} P((X - \mu + t)^2 \geq (a + t)^2) &\leq \frac{\sigma^2 + \mu^2 + t^2 + \mu^2 - 2\mu^2 + 2t\mu - 2\mu t}{(a + t)^2} \\ &= \frac{\sigma^2 + t^2}{(a + t)^2} \end{aligned}$$

Then as for the t in $\frac{\sigma^2 + t^2}{(a + t)^2}$, it is a variable, then we can find the minimum of the $\frac{\sigma^2 + t^2}{(a + t)^2}$, can verify if it satisfy the equality we have

$$\frac{\partial \frac{\sigma^2 + t^2}{(a + t)^2}}{\partial t} = \frac{2at - 2\sigma^2}{(a + t)^3}$$

where the $t = \frac{\sigma^2}{a}$, when we get minimum and the minimum is that

$$\frac{2at - 2\sigma^2}{(a + t)^3} = \frac{\sigma^2}{a^2 + \sigma^2}$$

which is exactly the bound.

So we get that

$$P((X - \mu + t)^2 \geq (a + t)^2) \leq \frac{\sigma^2}{a^2 + \sigma^2}.$$

As $P((X - \mu - t)^2 \geq (a - t)^2)$ is the same with that $P(X - \mu \geq a)$, we get that

$$P(X - \mu \geq a) \leq \frac{\sigma^2}{a^2 + \sigma^2}.$$

Problem 4

From the problem, we know that the r.v. X_i is that given the value of common mean, the X_i are normal and independent, with known variances $\sigma_1^2, \dots, \sigma_n^2$. then as we have that a normal prior Θ of mean x_0 and variance σ_0^2 , we have that the prior satisfy that

$$\Theta \sim N(x_0, \sigma_0^2).$$

And the $X_i|\Theta$ satisfy that

$$f_{X_i|\Theta}(x_i|\theta) = \frac{1}{\sqrt{2\pi\sigma_i^2}} e^{-\frac{(x_i-\theta)^2}{2\sigma_i^2}}.$$

Then we need to find the posterior PDF of Θ . We have that with the Bayes' theorem, we get that as X is a collection of X_1, \dots, X_n , we have that

$$\begin{aligned} f_{\Theta|X}(\theta|x) &= \frac{f_{X|\Theta}(x|\theta)f_{\Theta}(\theta)}{f_X(x)} \\ &= \frac{f_{(X_1, \dots, X_n)|\Theta}((x_1, \dots, x_n)|\theta)f_{\Theta}(\theta)}{f_{(X_1, \dots, X_n)}((x_1, \dots, x_n))} \end{aligned}$$

As the X_i are independent, we get that

$$= \frac{f_{(X_1, \dots, X_n)|\Theta}((x_1, \dots, x_n)|\theta)f_{\Theta}(\theta)}{f_{(X_1, \dots, X_n)}((x_1, \dots, x_n))}$$

As the Denominator is a Integral which is a constant, we turn to use the fraction form as the left side is a valid PDF, we get that

$$f_{\Theta|X}(\theta|x) \propto f_{X|\Theta}(x|\theta)f_{\Theta}(\theta)$$

The right side can be write as

$$\frac{1}{\sqrt{2\pi\sigma_0^2}} e^{-\frac{(\theta-x_0)^2}{2\sigma_0^2}} \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma_i^2}} e^{-\frac{(x_i-\theta)^2}{2\sigma_i^2}} \propto e^{-\frac{(\theta-x_0)^2}{2\sigma_0^2}} e^{-\sum_{i=1}^n \frac{(x_i-\theta)^2}{2\sigma_i^2}}$$

which equals

$$e^{-\sum_{i=0}^n \frac{(x_i-\theta)^2}{2\sigma_i^2}} \propto e^{-\sum_{i=0}^n \frac{\theta^2}{2\sigma_i^2} + 2\sum_{i=0}^n \frac{\theta x_i}{2\sigma_i^2}}.$$

Then we try to simplify it to the form of Normal distribution, we have that

$$\begin{aligned} f_{\Theta|X}(\theta|x) &\propto e^{-\sum_{i=0}^n \frac{\theta^2}{2\sigma_i^2} + 2\sum_{i=0}^n \frac{\theta x_i}{2\sigma_i^2}} \\ &\propto e^{-\sum_{i=0}^n \frac{1}{2\sigma_i^2} (\theta - \frac{\sum_{i=0}^n \frac{x_i}{\sigma_i^2}}{\sum_{i=0}^n \frac{1}{\sigma_i^2}})^2} \quad (\text{Constant are abandon.}) \\ &= e^{-(\theta - \frac{\sum_{i=0}^n \frac{x_i}{\sigma_i^2}}{\sum_{i=0}^n \frac{1}{\sigma_i^2}})^2 / \frac{1}{\sum_{i=0}^n \frac{1}{\sigma_i^2}}} \end{aligned}$$

Which is in the form of Normal distribution. So we get that

$$\Theta|X \sim N\left(\frac{\sum_{i=0}^n \frac{x_i}{\sigma_i^2}}{\sum_{i=0}^n \frac{1}{\sigma_i^2}}, \frac{1}{\sum_{i=0}^n \frac{1}{\sigma_i^2}}\right).$$

Then as for the PDF, we have that

$$f_{\Theta|X}(\theta|x) = \frac{1}{\sqrt{2\pi \sum_{i=0}^n \frac{1}{\sigma_i^2}}} e^{-\left(\theta - \frac{\sum_{i=0}^n \frac{x_i}{\sigma_i^2}}{2 \sum_{i=0}^n \frac{1}{2\sigma_i^2}}\right)^2 / \frac{1}{\sum_{i=0}^n \frac{1}{2\sigma_i^2}}} = \frac{1}{\sqrt{2\pi \sum_{i=0}^n \frac{1}{\sigma_i^2}}} e^{-\left(\theta - \frac{\sum_{i=0}^n \frac{x_i}{\sigma_i^2}}{\sum_{i=0}^n \frac{1}{\sigma_i^2}}\right)^2 / \frac{2}{\sum_{i=0}^n \frac{1}{\sigma_i^2}}}.$$

Problem 5

(a) As for the n independent random variables X_1, X_2, \dots, X_n , we have that

$$X_i \sim \text{Expo}(\theta).$$

Then with the definition of Maximum Likelihood Estimation, we get that the likelihood function is that

$$\log[f_X(X_1, \dots, X_n; \theta)] = \log \prod_{i=1}^n f_{X_i}(X_i; \theta) = \sum_{i=1}^n \log[f_{X_i}(X_i; \theta)].$$

As we have $X_i \sim \text{Expo}(\theta)$, we get that

$$f_{X_i}(X_i, \theta) = \theta e^{-\theta x_i}.$$

So the likelihood function is that

$$\sum_{i=1}^n \log[\theta e^{-\theta x_i}] = n \log \theta - \theta \sum_{i=1}^n x_i.$$

Then we calculate the derivative of the likelihood function with respect to θ , we get that

$$\frac{\partial \log[f_X(X_1, \dots, X_n; \theta)]}{\partial \theta} = \frac{n}{\theta} - \sum_{i=1}^n x_i.$$

Then we get that the maximum likelihood estimation of θ is that

$$\hat{\theta} = \frac{n}{\sum_{i=1}^n x_i}.$$

(b) As for the n independent random variables X_1, X_2, \dots, X_n , we have that

$$X_i \sim N(\mu, \nu).$$

Then with parameter vector $\theta = (\mu, \nu)$, with the definition of Maximum Likelihood Estimation, we get that the likelihood function is that

$$\log[f_X(X_1, \dots, X_n; \mu, \nu)] = \log \prod_{i=1}^n f_{X_i}(X_i; \mu, \nu) = \sum_{i=1}^n \log[f_{X_i}(X_i; \mu, \nu)].$$

As we have $X_i \sim N(\mu, \nu)$, we get that

$$f_{X_i}(X_i, \mu, \nu) = \frac{1}{\sqrt{2\pi\nu}} e^{-\frac{(x_i - \mu)^2}{2\nu}}.$$

So the likelihood function is that

$$\sum_{i=1}^n \log\left[\frac{1}{\sqrt{2\pi\nu}} e^{-\frac{(x_i - \mu)^2}{2\nu}}\right] = n \log \frac{1}{\sqrt{2\pi\nu}} - \frac{1}{2\nu} \sum_{i=1}^n (x_i - \mu)^2.$$

Then we calculate the derivative of the likelihood function with respect to μ , we get that

$$\frac{\partial \log[f_X(X_1, \dots, X_n; \mu, \nu)]}{\partial \mu} = \frac{1}{\nu} \sum_{i=1}^n (x_i - \mu).$$

Then we get that the maximum likelihood estimation of μ is that

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n x_i.$$

Then we calculate the derivative of the likelihood function with respect to ν , we get that

$$\frac{\partial \log[f_X(X_1, \dots, X_n; \mu, \nu)]}{\partial \nu} = -\frac{n}{2\nu} + \frac{1}{2\nu^2} \sum_{i=1}^n (x_i - \mu)^2.$$

Then we get that the maximum likelihood estimation of ν is that

$$\hat{\nu} = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2 = \frac{1}{n} \sum_{i=1}^n \left(x_i - \frac{1}{n} \sum_{j=1}^n x_j\right)^2.$$

So in conclusion we get that the maximum likelihood estimation of $\theta = (\mu, \nu)$ is that

$$\hat{\theta} = (\hat{\mu}, \hat{\nu}) = \left(\frac{1}{n} \sum_{i=1}^n x_i, \frac{1}{n} \sum_{i=1}^n \left(x_i - \frac{1}{n} \sum_{j=1}^n x_j\right)^2\right).$$