# Probability & Statistics for EECS: Homework #13

Due on May 14, 2023 at 23:59

Name: **Zhou Shouchen**
Student ID: 2021533042

# Problem 1

Since $X_1, X_2, \cdots$ are i.i.d. Expo(1), so $f_{X_1}(x) = f_{X_2}(x) = \cdots = e^{-x}, x > 0$.

(a) $P(X_n \geq 1) = \int_1^{+\infty} e^{-x}dx = e^{-1}$
From the definition of $N = min\{n : X_n \geq 1\}$, so we can get that $N \sim FS(\frac{1}{e})$.
So $E(N) = \frac{1}{\frac{1}{e}} = e$.

So above all, the distribution of $N$ is $FS(\frac{1}{e})$.
And $E(N) = e$.

(b) From the Poisson process with rate $\lambda = 1$, we can get that:
Let $X_i$ be the arriving interval, so $X_i \sim Expo(1)$.
Suppose that the time starts at time 0, since $M = min\{n : X_1 + \cdots + X_n \geq 10\}$.
Which means that $M - 1$ is the number of people arrival in the time interval $[0, 10)$.
Let $Y_i$ be the number of arrivals in the interval $[0, 10)$, since the interval's length is 10,
so $Y_i \sim Pois(1 \cdot 10) \sim Pois(10)$.

And from the defination of $M$, we can get that $M - 1$ is the number of arrivals in the interval $[0, 10)$,
i.e. $M - 1 \sim Y_i \sim Pois(10)$.
So $E(M - 1) = 10$.
So $E(M) = E(M - 1) + 1 = 11$.

So above all, the distribution of $M - 1$ is $Pois(10)$.
And $E(M) = 11$.

(c) Since $X_1, \cdots, X_n$ are i.i.d. $Expo(1)$ with finate mean $\mu = E(X_i) = 1$,
and finate variance $\sigma^2 = Var(X_i) = 1$.
For the exact distribution,
Let $F_X(x)$ be the CDF of $X \sim Expo(1)$, then $F_X(x) = 1 - e^{-x}$.
Then the CDF of $Y = \frac{1}{n}X$ is $F_Y(y) = P(Y \leq y) = P(\frac{1}{n}X \leq y) = P(X \leq ny) = F_X(ny) = 1 - e^{-ny}$.
So $Y \sim Expo(n)$.
i.e. $\frac{1}{n}X \sim Expo(n)$.
So $\frac{1}{n}X_i$ are i.i.d. $Expo(n)$.

From the theorem we have learned, since $\frac{1}{n}X_i$ are i.i.d. $Expo(n)$, so we can get that

$$\sum_{i=1}^{n} \frac{1}{n}X_i \sim Gamma(n, n)$$

i.e. the exact distribution of $\bar{X_n} = \dfrac{X_1 + \cdots + X_n}{n} = \sum\limits_{i=1}^{n} \frac{1}{n}X_i$ is

$$\bar{X_n} \sim Gamma(n, n)$$

As for the approximate distribution,
for $n$ is large, from the Central Limit Theorem, we can get that:
$\bar{X_n} = \frac{1}{n}\sum\limits_{i=1}^{n} X_i$ is approximately normal with mean $\mu$ and variance $\frac{\sigma^2}{n} = \frac{1}{n}$.

Problem 1 continued on next page...

i.e. $\bar{X_n}$ is approximate to $N(1, \frac{1}{n})$.

So above all, the exact distribution of $\bar{X_n}$ is $Gamma(n, n)$.
And for $n$ is large, the approximate distribution of $\bar{X_n}$ is $N(1, \frac{1}{n})$.

# Problem 2

(a) We know that $X_1, X_2, \cdots$ are i.i.d. with $E(X_i) = \mu, a \le X_i \le b$.
Let $X = \dfrac{1}{n}\sum\limits_{i=1}^{n} X_i$.

From the Chernoff Inequality, we can get that:
since $\epsilon \ge 0$, so $\forall t > 0$,

$$\begin{aligned}
P(X - \mu \ge \epsilon) &\le \frac{E(e^{t(X-\mu)})}{e^{t\epsilon}} \\
&= \frac{E(e^{t((\frac{1}{n}\sum\limits_{i=1}^{n} X_i)-\mu)})}{e^{t\epsilon}} \\
&= \frac{E(\prod\limits_{i=1}^{n} e^{t(\frac{1}{n}(X_i-\mu))})}{e^{t\epsilon}}
\end{aligned}$$

Let $Y_i = \dfrac{1}{n}(X_i - \mu)$
Since $X_1, \cdots, X_n$ are independent, so $e^{t(\frac{1}{n}(X_i-\mu))} = e^{tY_i}$ are independent.
So the origin inequality can be written as:

$$\begin{aligned}
P(X - \mu \ge \epsilon) &\le \frac{\prod\limits_{i=1}^{n} E(e^{t(\frac{1}{n}(X_i-\mu))})}{e^{t\epsilon}} \\
&= \frac{\prod\limits_{i=1}^{n} E(e^{tY_i})}{e^{t\epsilon}}
\end{aligned}$$

And $E(Y_i) = E(\dfrac{1}{n}(X_i - \mu)) = 0$, and $\dfrac{a-\mu}{n} \le Y_i \le \dfrac{b-\mu}{n}$.
So from Hoeffding Lemma, we can get that:

$$E(e^{tY_i}) \le e^{\frac{1}{8}t^2(\frac{b-a}{n})^2}$$

And since $e^{tY_i} > 0$, so $E(e^{tY_i}) > 0$.
So

$$\begin{aligned}
\frac{\prod\limits_{i=1}^{n} E(e^{tY_i})}{e^{t\epsilon}} &\le \frac{(e^{\frac{1}{8}t^2(\frac{b-a}{n})^2})^n}{e^{t\epsilon}} \\
&= \frac{e^{\frac{1}{8n}t^2(b-a)^2}}{e^{t\epsilon}} \\
&= e^{\frac{1}{8n}t^2(b-a)^2 - t\epsilon}
\end{aligned}$$

Since $e^x$ is strictly increasing, so $e^{\frac{1}{8n}t^2(b-a)^2 - t\epsilon}$ is strictly increasing.
To get the minimum of $e^{\frac{1}{8n}t^2(b-a)^2 - t\epsilon}$, we can just get the minimum of $\frac{1}{8n}t^2(b-a)^2 - t\epsilon$.
With the knowledge of quadratic function, we can get that
when $t = \dfrac{\epsilon}{2 \cdot \frac{(b-a)^2}{8n}} = \dfrac{4n\epsilon}{(b-a)^2}$,
the minimum of $= e^{\frac{1}{8n}t^2(b-a)^2 - t\epsilon}$ is that

$$\begin{aligned}
& e^{\frac{1}{8n}(\frac{4n\epsilon}{(b-a)^2})^2(b-a)^2 - \frac{4n\epsilon}{(b-a)^2}\epsilon} \\
&= e^{\frac{-2n\epsilon^2}{(b-a)^2}}
\end{aligned}$$

So we can get that

$$P(X-\mu \geq \epsilon) \leq e^{\frac{-2n\epsilon^2}{(b-a)^2}}$$

Similarly, with the same method, we can get that

$$P(X-\mu \leq -\epsilon) = P(\mu - X \geq \epsilon) \leq e^{\frac{-2n\epsilon^2}{(b-a)^2}}$$

So combine them, we can get that

$$P(|X-\mu| \geq \epsilon) = P(X-\mu \geq \epsilon) + P(X-\mu \leq -\epsilon) \leq 2e^{\frac{-2n\epsilon^2}{(b-a)^2}}$$

And put $X = \frac{1}{n}\sum\limits_{i=1}^{n} X_i$ into it, we can get that

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

So above all, the Hoeffding bound

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

holds.

# Problem 3

Let $Y = X - \mu$, then $P(X - \mu \geq a) = P(Y \geq a)$
So $\forall t \geq 0$, we have $a + t \geq 0$, so
$P(Y \geq a) = P((Y + t) \geq (a + t)) \leq P((Y + t)^2 \geq (a + t)^2)$.
From Marcov's Inequality, we can get that when $\forall a > 0$

$$P(|X| \geq a) \leq \frac{E|X|}{a}$$

So

$$\begin{aligned} P((Y+t)^2 \geq (a+t)^2) &\leq \frac{E((Y+t)^2)}{(a+t)^2} \\ &= \frac{E(Y^2) + 2tE(Y) + t^2}{(a+t)^2} \end{aligned}$$

Since $Y = X - \mu$, so $E(Y) = E(X) - \mu = 0$, $Var(Y) = Var(X) = \sigma^2$.
And since $Var(Y) = E(Y^2) - (E(Y))^2$, so we can get that $E(Y^2) = \sigma^2$.
So

$$\frac{E(Y^2) + 2tE(Y) + t^2}{(a+t)^2} = \frac{\sigma^2 + t^2}{(a+t)^2}$$

Let $f(t) = \dfrac{\sigma^2 + t^2}{(a+t)^2}$, then $f'(t) = \dfrac{2(at - \sigma^2)}{(a+t)^3}$.
And $f''(t) = \dfrac{2(a^2 - 2at + 3\sigma^2)}{(a+t)^4}$.
And when $f'(t) = 0$, we can get that $t = \dfrac{\sigma^2}{a}$.
And at this time, $f''(t) = \dfrac{2(a^2 - 2a \cdot \frac{\sigma^2}{a} + 3\sigma^2)}{(a + \frac{\sigma^2}{a})^4} = \dfrac{2(a^2 + \sigma^2)}{(a + \frac{\sigma^2}{a})^4} > 0$.
So $\{f(t)\}_{min} = f(\dfrac{\sigma^2}{a}) = \dfrac{\sigma^2(a^2 + \sigma^2)}{(a^2 + \sigma^2)^2} = \dfrac{\sigma^2}{a^2 + \sigma^2}$.
So

$$P(Y \geq a) = P((Y+t)^2 \geq (a+t)^2) \leq \frac{E((Y+t)^2)}{(a+t)^2} \leq \frac{\sigma^2}{a^2 + \sigma^2}$$

i.e.

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

So above all, for any $a \geq 0$, the one-side Chebyshev Inequality

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

have been proved.

# Problem 4

With the Bayes Inference.
The prior distribution is $\Theta \sim N(x_0, \sigma_0^2)$.

And the observations are independent normals, i.e. $X_i|\Theta \sim N(\theta, \sigma_i^2)$.
So $f_{X_i|\Theta}(x_i|\theta) = \frac{1}{\sqrt{2\pi}\sigma_i} e^{-\frac{(x_i-\theta)^2}{2\sigma_i^2}}$.

So the posterior PDF of $\Theta|\mathbf{X}$ is that

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

With Bayes' Rule, we can get that

$$= \frac{f_{\mathbf{X}|\Theta}(\mathbf{x}|\theta) f_\Theta(\theta)}{f_{\mathbf{X}}(\mathbf{x})}$$

Since $X_i$ are independent, so $f_{\mathbf{X}|\Theta}(\mathbf{x}|\theta) = \prod\limits_{i=1}^{n} f_{X_i|\Theta}(x_i|\theta)$.
Also, with LOTP, we can get that $f_{\mathbf{X}}(\mathbf{x}) = \int_{-\infty}^{+\infty} f_{\mathbf{X}|\Theta}(\mathbf{x}|\theta) f_\Theta(\theta) d\theta$.
Which must be a formula without $\theta$, so it could be regarded as a constant.

Since $f_{\Theta|\mathbf{X}}(\theta|\mathbf{x})$ is a valid PDF, so we can ignore its constant part.
i.e.

$$\frac{f_{\mathbf{X}|\Theta}(\mathbf{x}|\theta) f_\Theta(\theta)}{f_{\mathbf{X}}(\mathbf{x})} \propto \prod_{i=1}^{n} f_{X_i|\Theta}(x_i|\theta) f_\Theta(\theta)$$

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

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

$$= exp(-\sum_{i=0}^{n} \frac{(x_i-\theta)^2}{2\sigma_i^2})$$

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

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

Let $A = \sum\limits_{i=0}^{n} \frac{1}{2\sigma_i^2}$, $B = \sum\limits_{i=0}^{n} \frac{x_i}{\sigma_i^2}$, then we can get that

$$= exp(-A\theta^2 + B\theta)$$

$$= exp(-A(\theta - \frac{B}{2A})^2 + \frac{B^2}{4A})$$

$$\propto exp(-A(\theta - \frac{B}{2A})^2)$$

$$= exp(\frac{-(\theta - \frac{B}{2A})^2}{2 \cdot \frac{1}{2A}})$$

Since $f_{\Theta|\mathbf{X}}(\theta|\mathbf{x})$ is a valid PDF, so we can ignore its constant part.
From the part without constant, we can get that

$$f_{\Theta|\mathbf{X}}(\theta|\mathbf{x}) \propto exp(\frac{-(\theta-\frac{B}{2A})^2}{2\cdot\frac{1}{2A}})$$

i.e.

$$\Theta|\mathbf{X} \sim N(\frac{B}{2A},\frac{1}{2A})$$

put $A=\sum_{i=0}^n\frac{1}{2\sigma_i^2}$, $B=\sum_{i=0}^n\frac{x_i}{\sigma_i^2}$ into it, we can get that

$$\Theta|\mathbf{X} \sim N(\frac{\sum\limits_{i=0}^n\frac{x_i}{\sigma_i^2}}{\sum\limits_{i=0}^n\frac{1}{\sigma_i^2}},\frac{1}{\sum\limits_{i=0}^n\frac{1}{\sigma_i^2}})$$

So the PDF of the posterior distribution of $\Theta$ is that

$$\begin{aligned}
f_{\Theta|\mathbf{X}}(\theta|\mathbf{x}) &= \frac{1}{\sqrt{2\pi}\frac{1}{\sqrt{\sum\limits_{i=0}^n\frac{1}{\sigma_i^2}}}}\cdot\exp(-\frac{(\theta-\frac{\sum\limits_{i=0}^n\frac{x_i}{\sigma_i^2}}{\sum\limits_{i=0}^n\frac{1}{\sigma_i^2}})^2}{2\cdot\frac{1}{\sum\limits_{i=0}^n\frac{1}{\sigma_i^2}}}) \\
&= \frac{\sqrt{\sum\limits_{i=0}^n\frac{1}{\sigma_i^2}}}{\sqrt{2\pi}}\cdot\exp(-\frac{((\sum\limits_{i=0}^n\frac{1}{\sigma_i^2})\cdot\theta-\sum\limits_{i=0}^n\frac{x_i}{\sigma_i^2})^2}{2\cdot\sum\limits_{i=0}^n\frac{1}{\sigma_i^2}})
\end{aligned}$$

So above all, the posterior PDF of $\Theta$ is that

$$f_{\Theta|\mathbf{X}}(\theta|\mathbf{x}) = \frac{\sqrt{\sum\limits_{i=0}^n\frac{1}{\sigma_i^2}}}{\sqrt{2\pi}}\cdot\exp(-\frac{((\sum\limits_{i=0}^n\frac{1}{\sigma_i^2})\cdot\theta-\sum\limits_{i=0}^n\frac{x_i}{\sigma_i^2})^2}{2\cdot\sum\limits_{i=0}^n\frac{1}{\sigma_i^2}})$$

# Problem 5

(a) Since $X_i$ are independent $Expo(\theta)$, so $f_{X_i}(x_i;\theta) = \theta e^{-\theta x_i}$.
And since $X_i$ are independent, so $f_{\mathbf{X}}(\mathbf{x};\theta) = \prod\limits_{i=1}^{n} f_{X_i}(x_i;\theta)$.
i.e. the maximum likelihood function is that

$$\hat{\theta}_n = \arg\max_{\theta} f_{\mathbf{X}}(\mathbf{x}) = \arg\max_{\theta} \prod_{i=1}^{n} f_{X_i}(x_i) = \arg\max_{\theta} \prod_{i=1}^{n} \theta e^{-\theta x_i}$$

We can also write it by taking log to the right-hand side because the log-function is strictly increasing, i.e.

$$\hat{\theta}_n = \arg\max_{\theta} \sum_{i=1}^{n} (\ln\theta - \theta x_i)$$

Let $g(\theta) = \sum\limits_{i=1}^{n}(\ln\theta - \theta x_i)$, then we can get that
$g'(\theta) = \sum\limits_{i=1}^{n}(\frac{1}{\theta} - x_i) = \frac{n}{\theta} - \sum\limits_{i=1}^{n} x_i$.
When $g'(\theta) = 0$, we can get that $\theta = \frac{n}{\sum\limits_{i=1}^{n} x_i}$.
And since $g''(\theta) = -\frac{n}{\theta^2} < 0$, so $\theta = \frac{n}{\sum\limits_{i=1}^{n} x_i}$ is the maximum point.
i.e. $\hat{\theta}_n = \frac{n}{\sum\limits_{i=1}^{n} x_i}$.
So above all, the MLE of $\theta$ is that $\hat{\theta}_n = \frac{n}{\sum\limits_{i=1}^{n} x_i}$.

(b) Since $X_i$ are independent $N(\mu,\nu)$, so $f_{X_i}(x_i;\theta) = \frac{1}{\sqrt{2\pi}\nu} e^{-\frac{(x_i-\mu)^2}{2\nu^2}}$.
And since $X_i$ are independent, so $f_{\mathbf{X}}(\mathbf{x};\theta) = \prod\limits_{i=1}^{n} f_{X_i}(x_i;\theta)$.
i.e. the maximum likelihood function is that

$$\hat{\theta}_n = \arg\max_{\theta} f_{\mathbf{X}}(\mathbf{x}) = \arg\max_{\theta} \prod_{i=1}^{n} f_{X_i}(x_i) = \arg\max_{\theta} \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\nu} e^{-\frac{(x_i-\mu)^2}{2\nu^2}}$$

We can also write it by taking log to the right-hand side because the log-function is strictly increasing, i.e.

$$\hat{\theta}_n = \arg\max_{\theta} = \sum_{i=1}^{n} (-\ln(\sqrt{2\pi}\nu) - \frac{(x_i-\mu)^2}{2\nu^2})$$

Let $g(\mu,\nu) = \sum\limits_{i=1}^{n}(-\ln(\sqrt{2\pi}\nu) - \frac{(x_i-\mu)^2}{2\nu^2}) = -n\ln(\sqrt{2\pi}\nu) - \frac{\sum\limits_{i=1}^{n}(x_i-\mu)^2}{2\nu^2}$,
then we can get that $\frac{\partial g(\mu,\nu)}{\partial\mu} = \frac{\sum\limits_{i=1}^{n}(x_i-\mu)}{\nu^2}$
Let $\frac{\partial g(\mu,\nu)}{\partial\mu} = 0$, we can get that $\mu = \frac{\sum\limits_{i=1}^{n} x_i}{n}$.

And $\dfrac{\partial g(\mu,\nu)}{\partial \nu} = -\dfrac{n}{\nu} + \dfrac{\sum\limits_{i=1}^{n}(x_i-\mu)^2}{\nu^3}$.

Let $\dfrac{\partial g(\mu,\nu)}{\partial \nu} = 0$, we can get that $\nu^2 = \dfrac{\sum\limits_{i=1}^{n}(x_i-\mu)^2}{n}$.

And check when $\mu = \dfrac{\sum\limits_{i=1}^{n} x_i}{n}$, $\nu = \sqrt{\dfrac{\sum\limits_{i=1}^{n}(x_i-\mu)^2}{n}}$ is the MLE.

$\dfrac{\partial^2 g(\mu,\nu)}{\partial \mu^2} = -\dfrac{n}{\nu^2} < 0$.

$\dfrac{\partial^2 g(\mu,\nu)}{\partial \nu^2} = \dfrac{n}{\nu^2} - 3\dfrac{\sum\limits_{i=1}^{n}(x_i-\mu)^2}{\nu^4}$.

put $\nu^2 = \dfrac{\sum\limits_{i=1}^{n}(x_i-\mu)^2}{n}$ into it, we can get that $\dfrac{\partial^2 g(\mu,\nu)}{\partial \nu^2} = \dfrac{n}{\nu^2} - \dfrac{3n}{\nu^2} = -\dfrac{2n}{\nu^2} < 0$.

$\dfrac{\partial^2 g(\mu,\nu)}{\partial \mu \partial \nu} = \dfrac{\partial^2 g(\mu,\nu)}{\partial \nu \partial \mu} = \dfrac{2(n\mu - \sum\limits_{i=1}^{n} x_i)}{\nu^3}$.

put $\mu = \dfrac{\sum\limits_{i=1}^{n} x_i}{n}$ into it, we can get that $\dfrac{\partial g(\mu,\nu)}{\partial \mu \partial \nu} = \dfrac{2(n\mu - \sum\limits_{i=1}^{n} x_i)}{\nu^3} = 0$.

So $\hat{\mu}_n = \dfrac{\sum\limits_{i=1}^{n} x_i}{n}$, $\hat{\nu}_n = \sqrt{\dfrac{\sum\limits_{i=1}^{n}(x_i-\hat{\mu}_n)^2}{n}}$ is the maximum point.

i.e. $\hat{\theta}_n = (\hat{\mu}_n, \hat{\nu}_n) = (\dfrac{\sum\limits_{i=1}^{n} x_i}{n}, \sqrt{\dfrac{\sum\limits_{i=1}^{n}(x_i-\hat{\mu}_n)^2}{n}})$.

So above all, the MLE of $\theta = (\mu,\nu)$ is that $\hat{\theta}_n = (\hat{\mu}_n, \hat{\nu}_n) = (\dfrac{\sum\limits_{i=1}^{n} x_i}{n}, \sqrt{\dfrac{\sum\limits_{i=1}^{n}(x_i-\hat{\mu}_n)^2}{n}})$.