

## Problem Set 5

Hoang Nguyen, Huy Nguyen

December 2, 2019

### Problem 1.

1. From exercise 3 in Problem 2, If  $Z \sim N(0, 1)$ , we get:  $\mathbb{P}\left[\left|\frac{\sqrt{n}(\bar{X}_i - \mu)}{\sigma}\right| \leq t\right] = 2\mathbb{P}[Z \leq t] - 1$ .  
 $X_1, X_2, \dots, X_n$  follow Poisson distribution. Hence,  $\mathbb{E}[X_i] = \lambda$  and  $\mathbb{V}[X_i] = \lambda$   
 From CTL:  $\frac{\sqrt{n}(\bar{X}_n - \lambda)}{\sqrt{\lambda}} \sim N(0, 1)$ . Hence,  $\mathbb{P}\left[\left|\frac{\sqrt{n}(\bar{X}_n - \lambda)}{\sqrt{\lambda}}\right| \leq t\right] = 2\mathbb{P}\left[\frac{\sqrt{n}(\bar{X}_n - \lambda)}{\sqrt{\lambda}} \leq t\right] - 1 = 1 - \alpha$ .  
 Hence,  $\mathbb{P}\left[\frac{\sqrt{n}(\bar{X}_n - \lambda)}{\sqrt{\lambda}} \leq t\right] = 1 - \frac{\alpha}{2}$ . Therefore,  $t = \phi^{-1}(1 - \frac{\alpha}{2})$ .  
 We have  $= \mathbb{P}\left[\left|\frac{\sqrt{n}(\bar{X}_n - \lambda)}{\sqrt{\lambda}}\right| \leq \phi^{-1}(1 - \frac{\alpha}{2})\right] \rightarrow (1 - \alpha)$  as  $n$  tends to infinity. This equivalent to:

$$\mathbb{P}\left[-\frac{\sqrt{n}(\bar{X}_n - \lambda)}{\sqrt{\lambda}} \leq \phi^{-1}(1 - \frac{\alpha}{2}) \leq \frac{\sqrt{n}(\bar{X}_n - \lambda)}{\sqrt{\lambda}}\right] \rightarrow 1 - \alpha$$

$$\mathbb{P}\left[\bar{X}_n - \frac{\phi^{-1}(1 - \frac{\alpha}{2})\sqrt{\lambda}}{\sqrt{n}} \leq \lambda \leq \bar{X}_n + \frac{\phi^{-1}(1 - \frac{\alpha}{2})\sqrt{\lambda}}{\sqrt{n}}\right] \rightarrow 1 - \alpha$$

We know:  $\bar{X}_n \xrightarrow{P} \mathbb{E}[\bar{X}_n] = \lambda$

Hence,

$$\mathbb{P}\left[\bar{X}_n - \frac{\phi^{-1}(1 - \frac{\alpha}{2})\sqrt{\bar{X}_n}}{\sqrt{n}} \leq \lambda \leq \bar{X}_n + \frac{\phi^{-1}(1 - \frac{\alpha}{2})\sqrt{\bar{X}_n}}{\sqrt{n}}\right] \geq 1 - \alpha$$

$$\Rightarrow L = [\bar{X}_n - \frac{\phi^{-1}(1 - \frac{\alpha}{2})\sqrt{\bar{X}_n}}{\sqrt{n}}, \bar{X}_n + \frac{\phi^{-1}(1 - \frac{\alpha}{2})\sqrt{\bar{X}_n}}{\sqrt{n}}]$$

2. From previous result, if  $\lambda_0 \in L$ , we do not reject  $H_0$ . Otherwise, there is evidence to reject  $H_0$ .

### Problem 2.

1. Take derivative of Gaussian expression respect to  $\mu_1, \sigma_1, \mu_2, \sigma_2$  and set it equals to 0, we get:

$$\hat{\mu}_1 = \bar{X}_{n1}, \hat{\mu}_2 = \bar{Y}_{n2}, \hat{\sigma}_1^2 = \frac{1}{n_1} \sum_{i=1}^{n_1} (X_i - \mu_1)^2 \text{ and } \hat{\sigma}_2^2 = \frac{1}{n_2} \sum_{i=1}^{n_2} (Y_i - \mu_2)^2.$$

2. We have:  $\frac{n_1 \hat{\sigma}_1^2}{\sigma_1^2} = \frac{n_1 \frac{1}{n_1} \sum_{i=1}^{n_1} (X_i - \mu_1)^2}{\sigma_1^2} = \sum_{i=1}^{n_1} \left(\frac{X_i - \mu_1}{\sigma_1}\right)^2$ . Since  $\frac{X_i - \mu_1}{\sigma_1} \sim N(0, 1)$ , hence  $\frac{n_1 \hat{\sigma}_1^2}{\sigma_1^2} \sim \chi_{n_1}^2$ .

By similar method, we get  $\frac{n_2 \hat{\sigma}_2^2}{\sigma_2^2} \sim \chi_{n_2}^2$

3. From previous result,  $\frac{n_1 \hat{\sigma}_1^2}{\sigma_1^2} + \frac{n_2 \hat{\sigma}_2^2}{\sigma_2^2} \sim \chi_{n_1}^2 + \chi_{n_2}^2 \sim \chi_{n_1 + n_2}^2$ .

4. From CTT,  $\bar{X}_{n1} \sim N(\mu_1, \sigma_1^2)$ ,  $\bar{X}_{n2} \sim N(\mu_2, \sigma_2^2)$ . Hence,  $\Delta = \hat{\mu}_1 - \hat{\mu}_2 = \bar{X}_{n1} - \bar{X}_{n2} \sim N(\mu_1 - \mu_2, \sigma_1^2 + \sigma_2^2)$ .

5.

$$\mathbb{P}(N(\mu, \sigma^2) > 0) = \mathbb{P}\left(N(0, 1) > \frac{-\mu}{\sigma^2}\right) = \phi\left(\frac{\mu}{\sigma^2}\right)$$

Hence, the statistical model is:  $(\{0, 1\}, (Ber(\phi(\frac{\mu}{\sigma^2})))_{(\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}_+})$ . This model depends on  $\frac{\mu}{\sigma^2} \Rightarrow$  these parameters are not identified.

6. Same for 3.

7. Let  $X \sim Exp(\lambda) \Rightarrow \mathbb{P}(X > 20) = e^{-20\lambda}$ . Hence, the statistical model is:

$$(\{0, 1\}, (Ber(e^{-20\lambda}))_{\lambda > 0})$$

This parameter is identified.

8. Let  $X \sim \text{Ber}(p)$  such that:

$$\begin{cases} X_i = 1 & \text{if machine } i \text{ has timelife less than 500 days} \\ X_i = 0 & \text{otherwise} \end{cases} \quad (1)$$

Hence:

$$p = \mathbb{P}(X_i = 1) = 1 - e^{-500\lambda}$$

The number of machines that have stopped working before 500 days is a binominal random variable with parameter  $(67, 1 - e^{-500\lambda})$

The statistical model is  $(\{1, 2, 3, \dots, 67\}, (\text{Binominal}(67, 1 - e^{-500\lambda}))_{\lambda > 0})$ . This parameter is identified.

**Problem 3.**

1. By central limit theorem (CLT), we have:

$$\frac{\sqrt{n}(\bar{X}_i - \mu)}{\sigma} \sim (N(0, 1))$$

Hence,  $(a_n)_{n \in \mathbb{N}}$  can be  $\frac{\sqrt{n}}{\sigma}$  and  $(b_n)_{n \in \mathbb{N}}$  can be  $\mu$ .

2. We have:  $Z \sim N(0, 1)$

Hence,  $\mathbb{P}[|Z| \leq t] = \mathbb{P}[-t \leq Z \leq t] = \phi(t) - \phi(-t) = \phi(t) - (1 - \phi(t)) = 2\phi(t) - 1 = 2\mathbb{P}[Z \leq t] - 1$ .

3. From part 1 we get:

$$\frac{\sqrt{n}(\bar{X}_i - \mu)}{\sigma} \sim (N(0, 1))$$

from part 2 we get:

$$\mathbb{P}[|Z| \leq t] = 2\mathbb{P}[Z \leq t] - 1$$

Substitution:

$$\mathbb{P}\left[\left|\frac{\sqrt{n}(\bar{X}_i - \mu)}{\sigma}\right| \leq t\right] = 2\mathbb{P}[Z \leq t] - 1$$

We have  $2\mathbb{P}[Z \leq t] - 1 = 0.95 \Rightarrow t = \phi^{-1}\left(\frac{0.95+1}{2}\right) = 1.96$ .

Hence,

$$\mathbb{P}\left[\left|\frac{\sqrt{n}(\bar{X}_i - \mu)}{\sigma}\right| \leq 1.96\right] = 0.95$$

$$\mathbb{P}\left[-\frac{\sqrt{n}(\bar{X}_i - \mu)}{\sigma} \leq 1.96 \leq \frac{\sqrt{n}(\bar{X}_i - \mu)}{\sigma}\right] = 0.95$$

Because  $X_i$  is Poisson random variable with parameter  $\lambda$ , so  $\mu = \lambda$  and  $\sigma = \sqrt{\lambda}$   
We get:

$$\mathbb{P}\left[-\frac{\sqrt{n}(\bar{X}_i - \lambda)}{\sqrt{\lambda}} \leq 1.96 \leq \frac{\sqrt{n}(\bar{X}_i - \lambda)}{\sqrt{\lambda}}\right] = 0.95$$

$$\mathbb{P}\left[\bar{X}_i - \frac{1.96\sqrt{\lambda}}{\sqrt{n}} \leq \lambda \leq \bar{X}_i + \frac{1.96\sqrt{\lambda}}{\sqrt{n}}\right] = 0.95$$

We know:  $\bar{X}_i \xrightarrow{P} \mathbb{E}[\bar{X}_i] = \lambda$

Hence,

$$\mathbb{P}\left[\bar{X}_i - \frac{1.96\sqrt{\bar{X}_i}}{\sqrt{n}} \leq \lambda \leq \bar{X}_i + \frac{1.96\sqrt{\bar{X}_i}}{\sqrt{n}}\right] \geq 0.95$$

$$\Rightarrow L = \left[\bar{X}_i - \frac{1.96\sqrt{\bar{X}_i}}{\sqrt{n}}, \bar{X}_i + \frac{1.96\sqrt{\bar{X}_i}}{\sqrt{n}}\right]$$

4. We can easily see  $\min(X_i) \leq \bar{X}_i \leq \max(X_i)$ . Hence, a new interval can be:

$$L = \left[\min(X_i) - \frac{1.96\sqrt{\bar{X}_i}}{\sqrt{n}}, \max(X_i) + \frac{1.96\sqrt{\bar{X}_i}}{\sqrt{n}}\right]$$

**Problem 4.** We have  $X_i$  is IID. Hence,  $\mathbb{P}(M_n \leq t) = \prod_{i=1}^n \mathbb{P}(X_i \leq t)$ .  
By uniform distribution, the CDF of  $M_n$ :

$$\mathbb{P}(M_n \leq t) = F(t) = \left(\frac{t}{\theta}\right)^n$$

Hence, the PDF of  $M_n$  is:

$$f(t) = \frac{dF}{dt} = n\theta^{-n}t^{n-1}$$

We can easily get:

$$\mathbb{E}[M_n] = \int_0^\theta tn\theta^{-n}t^{n-1}dt = \frac{n}{n+1}\theta \rightarrow \theta \text{ as } n \rightarrow \infty$$

By Markov's Inequality:

$$\mathbb{P}[|M_n - \theta| > \epsilon] \leq \mathbb{P}[M_n - \theta > \epsilon] \leq \frac{\mathbb{E}[M_n - \theta]}{\epsilon} = \frac{\mathbb{E}[M_n] - \theta}{\epsilon} \rightarrow 0$$

Hence,  $M_n$  converges in probability to  $\theta$ .

2. From part 1 we get:  $M_n: \mathbb{P}[M_n \leq t] = \left(\frac{t}{\theta}\right)^n$ . Hence, CDF of  $n(1 - \frac{M_n}{\theta})$  is:

$$P\left[n\left(1 - \frac{M_n}{\theta}\right) \leq t\right] = \mathbb{P}\left[M_n \geq \frac{(n-t)\theta}{n}\right] = 1 - \left(\frac{n-t}{n}\right)^n \rightarrow 1 - e^{-t} \text{ as } n \rightarrow \infty$$

Hence,  $n(1 - \frac{M_n}{\theta})$  converges in distribution to an exponential random variable with parameter 1.

3. Let  $A$  is an exponential random variable with parameter 1. Because  $n(1 - \frac{M_n}{\theta})$  converges in distribution to  $X$ , we have:

$$\mathbb{P}\left[n\left(1 - \frac{M_n}{\theta}\right) \leq t\right] \rightarrow \mathbb{P}[X \leq t] = 1 - e^{-t}$$

$1 - e^{-t} = 0.95 \Rightarrow t = 3$ . We have:

$$\mathbb{P}\left[n\left(1 - \frac{M_n}{\theta}\right) \leq 3\right] \rightarrow 0.95$$

which is:

$$\mathbb{P}\left[\theta \leq \frac{nM_n}{n-3}\right] \rightarrow 0.95$$

On the other hand, we always have  $\theta \geq M_n$  (uniform distribution). Hence, we get:

$$\mathbb{P}\left[M_n \leq \theta \leq \frac{nM_n}{n-3}\right] \rightarrow 0.95 \text{ as } n \rightarrow \infty$$

We conclude  $L = \left[M_n, \frac{nM_n}{n-3}\right] = \left[M_n, M_n + \frac{3M_n}{n-3}\right]$ .

4.  $bias(M_n) = \mathbb{M}_X - \theta = \frac{n}{n+1}\theta - \theta \neq 0$ . Hence,  $M_n$  is biased.

As described in the doc,  $\mathbf{y}$  is a one-hot vector with a 1 for the true outside word  $o$ , that means  $y_i$  is 1 if and only if  $i = o$ . so the proof could be below:

$$\begin{aligned} - \sum_{w \in Vocab} y_w \log(\hat{y}_w) &= -[y_1 \log(\hat{y}_1) + \dots + y_o \log(\hat{y}_o) + \dots + y_w \log(\hat{y}_w)] \\ &= -y_o \log(\hat{y}_o) \\ &= -\log(\hat{y}_o) \\ &= -\log P(O = o | C = c) \end{aligned}$$

\*\* (b) \*\* we know this derivatives:

$$\because J = CE(y, \hat{y}) \hat{y} = softmax(\theta) \therefore \frac{\partial J}{\partial \theta} = (\hat{y} - y)^T$$

$y$  is a column vector in the above equation. So, we can use chain rules to solve the derivative:

$$\begin{aligned}
\frac{\partial J}{\partial v_c} &= \frac{\partial J}{\partial \theta} \frac{\partial \theta}{\partial v_c} \\
&= (\hat{y} - y) \frac{\partial U^T v_c}{\partial v_c} \\
&= U^T (\hat{y} - y)^T
\end{aligned}$$

\*\* (c) \*\* similar to the equation above.

$$\begin{aligned}
\frac{\partial J}{\partial v_c} &= \frac{\partial J}{\partial \theta} \frac{\partial \theta}{\partial U} \\
&= (\hat{y} - y) \frac{\partial U^T v_c}{\partial U} \\
&= v_c (\hat{y} - y)^T
\end{aligned}$$