Problem Set 5

Hoang Nguyen, Huy Nguyen

December 1, 2019

Problem 1.

1. From exercise 3 in Problem 2, If $Z \sim N(0,1)$, we get: $\mathbb{P}\Big[|\frac{\sqrt{n}(\bar{X_i}-\mu)}{\sigma}| \leqslant t\Big] = 2\mathbb{P}[Z \leqslant t] - 1$.
$X_1, X_2, ..., X_n$ 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, $\mathbb{P}\Big[|\frac{\sqrt{n}(\bar{X_n}-\lambda)}{\sqrt{\lambda}}| \leqslant t\Big] = 2\mathbb{P}\Big[\frac{\sqrt{n}(\bar{X_n}-\lambda)}{\sqrt{\lambda}} \leqslant t\Big] - 1 = 1-\alpha$.
Hence, $\mathbb{P}\Big[\frac{\sqrt{n}(\bar{X_n}-\lambda)}{\sqrt{\lambda}} \leqslant t\Big] = 1 - \frac{\alpha}{2}$. Therefore, $t = \phi^{-1}(1-\frac{\alpha}{2})$.
We have $= \mathbb{P}\Big[|\frac{\sqrt{n}(\bar{X_n}-\lambda)}{\sqrt{\lambda}}| \leqslant \phi^{-1}(1-\frac{\alpha}{2})\Big] \to (1-\alpha)$ as n tends to infinity. This equivalent to:

$$\mathbb{P}\Big[-\frac{\sqrt{n}(\bar{X_n}-\lambda)}{\sqrt{\lambda}} \leqslant \phi^{-1}(1-\frac{\alpha}{2}) \leqslant \frac{\sqrt{n}(\bar{X_n}-\lambda)}{\sqrt{\lambda}}\Big] \to 1-\alpha$$

$$\mathbb{P}\Big[\bar{X_n} - \frac{\phi^{-1}(1-\frac{\alpha}{2})\sqrt{\lambda}}{\sqrt{n}} \leqslant \lambda \leqslant \bar{X_n} + \frac{\phi^{-1}(1-\frac{\alpha}{2})\sqrt{\lambda}}{\sqrt{n}}\Big] = 1-\alpha$$

We know: $\bar{X_n} \xrightarrow{P} \mathbb{E}[\bar{X_n}] = \lambda$
Hence,

$$\mathbb{P}\Big[\bar{X_n} - \frac{\phi^{-1}(1-\frac{\alpha}{2})\sqrt{\bar{X_n}}}{\sqrt{n}} \leqslant \lambda \leqslant \bar{X_n} + \frac{\phi^{-1}(1-\frac{\alpha}{2})\sqrt{\bar{X_i}}}{\sqrt{n}}\Big] \geqslant 1-\alpha$$

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

2. *Theorem*: If bias $\to 0$ ans se (standard error) $\to 0$ as n $\to \infty$ then $\hat{\theta}$ is consistent, that is, $\hat{\theta} \xrightarrow{\mathbf{P}} \theta$.

$$\mathbb{V}[\bar{X_i}(1-\bar{X_i})] = \mathbb{V}[\frac{\sum_{i=1}^n X_i}{n}(1-\frac{\sum_{i=1}^n X_i}{n})] = \mathbb{V}[\frac{1}{n^2}\sum_{i=1}^n X_i\Big(n - \sum_{i=1}^n X_i\Big)]$$

$$= \frac{1}{n^4}\mathbb{V}[\sum_{i=1}^n X_i\Big(n - \sum_{i=1}^n X_i\Big)]$$

$$= \frac{1}{n^4}\mathbb{V}\Big[\sum_{i=1}^n X_i \sum_{i=1}^n (1-X_i)\Big]$$

$$= \frac{1}{n^4}\mathbb{V}\Big[\sum_{i=1}^n X_i(1-X_i) + \sum_{i\neq j} X_i(1-X_j)\Big]$$

$X_i \in 0,1 \Rightarrow \sum_{i=1}^n X_i(1-X_i) = 0$
Hence,

$$\mathbb{V}[\bar{X_i}(1-\bar{X_i})] = \frac{1}{n^4}\mathbb{V}\Big[\sum_{i\neq j} X_i(1-X_j)\Big]$$

By X_i is IID:

$$\mathbb{V}[\bar{X_i}(1-\hat{\bar{X_i}})] = \frac{1}{n^4}\sum_{i\neq j}\mathbb{V}[X_i]\mathbb{V}[1-X_i]$$

$$= \frac{1}{n^4} n(n-1)p(p-1)$$

The last equality tends to 0 as n $\to \infty$. So we have $\mathbb{V}[\bar{X}_i(1-\bar{X}_i)] \to 0$.
Hence, se$(\bar{X}_i(1-\bar{X}_i))$=$\sqrt{\mathbb{V}[\bar{X}_i(1-\bar{X}_i)]} \to 0$. (1)
By similar method, we can esaily get: $\mathbb{E}[\bar{X}_i(1-\bar{X}_i)] = \frac{1}{n^2}\sum_{i\neq j}\mathbb{E}[X_i(1-X_j)] = \frac{n(n-1)}{n^2}p(1-p) \to p(1-p)$ as n $\to \infty$.
Hence, bias$\Big(\bar{X}_i(1-\bar{X}_i)\Big) = \mathbb{E}[\bar{X}_i(1-\bar{X}_i)] - p(1-p) \to 0$ as n $\to \infty$ (2)
From (1), (2) and the theorem above, we get $\bar{X}_i(1-\bar{X}_i)$ is a consistent estimator of p(1-p).

3. We actually complete this exercise in previous solution.
bias$\Big(\bar{X}_i(1-\bar{X}_i)\Big) = \mathbb{E}[\bar{X}_i(1-\bar{X}_i)] - p(1-p) = \frac{n(n-1)}{n^2}p(1-p) - p(1-p)$.

4. To find an unbiased estimator, we have to find x such that $\frac{xn(n-1)}{n^2} = 1 \Rightarrow x = \frac{n}{n-1}$.
Hence, an unbiased estimator can be $\frac{n}{n-1}\bar{X}_i(1-\bar{X}_i)$

Problem 2.

1. $(\mathbb{N}, (Poiss(\lambda))_{\lambda>0})$. This paramter is identified.
2. $(\mathbb{R}_+, (Exp(\lambda))_{10>\lambda>0})$. This parameter is identified.
3. $(\mathbb{R}_+, (Uni(0,\theta))_{\theta>0})$. This parameter is identified.
4. $(\mathbb{R}, (N(\mu,\sigma^2))_{(\mu,\sigma^2)\in\mathbb{R}\times\mathbb{R}_+})$. These parater are identified.
5.

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

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 $\sim Ber(p)$ such that:

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

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 $(\{1,2,3,..,67\}, (Binominal(67, 1-e^{-500\lambda}))_{\lambda>0})$. 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 μ.

2. We have: $Z \backsim N(0,1)$
Hence, $\mathbb{P}[|Z| \leqslant t] = \mathbb{P}[-t \leqslant Z \leqslant t] = \phi(t) - \phi(-t) = \phi(t) - (1-\phi(t)) = 2\phi(t) - 1 = 2\mathbb{P}[Z \leqslant t] - 1$.

3. From part 1 we get:

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

from part 2 we get:

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

Substitution:

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

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

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

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

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

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

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

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

$$\mathbb{P}\Big[\bar{X}_i - \frac{1.96\sqrt{\bar{X}_i}}{\sqrt{n}} \leqslant \lambda \leqslant \bar{X}_i + \frac{1.96\sqrt{\bar{X}_i}}{\sqrt{n}}\Big] \geqslant 0.95$$

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

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

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

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

$$\mathbb{P}(M_n \leqslant t) = F(t) = \Big(\frac{t}{\theta}\Big)^n$$

Hence, the PDF of M_n is:

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

We can easily get:

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

By Markov's Inequality:

$$\mathbb{P}\Big[|M_n - \theta| > \epsilon\Big] \leqslant \mathbb{P}[M_n - \theta > \epsilon] \leqslant \frac{\mathbb{E}[M_n - \theta]}{\epsilon} = \frac{\mathbb{E}[M_n] - \theta}{\epsilon} \to 0$$

Hence, M_n converages in probility to θ.

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

$$P\Big[n(1-\frac{M_n}{\theta}) \leqslant t\Big] = \mathbb{P}\Big[M_n \geqslant \frac{(n-t)\theta}{n}\Big] = 1 - \Big(\frac{n-t}{n}\Big)^n \to 1 - e^{-t} \textbf{ as n} \to \infty$$

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

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

$$\mathbb{P}\Big[n(1-\frac{M_n}{\theta}) \leqslant t\Big] \to \mathbb{P}[X \leqslant t] = 1 - e^{-t}$$

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

$$\mathbb{P}\Big[n(1-\frac{M_n}{\theta}) \leqslant 3\Big] \to 0.95$$

which is:

$$\mathbb{P}\Big[\theta \leqslant \frac{nM_n}{n-3}\Big] \to 0.95$$

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

$$\mathbb{P}\Big[M_n \leqslant \theta \leqslant \frac{nM_n}{n-3}\Big] \to 0.95 \textbf{ as n} \to \infty$$

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

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

As described in the doc, $\boldsymbol{y}$ 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: ¡!– $-\sum_{w \in Vocab} y_w \log(\hat{y}_o) = -$¿

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

(b) we know this deravatives:

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

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

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

(c) similar to the equation above.

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