Problem Set 5

Hoang Nguyen, Huy Nguyen

December 19, 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:

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

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

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. 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:
$\hat{\mu_1} = \bar{X_{n1}}$, $\hat{\mu_2} = \bar{Y_{n2}}$, $\hat{\sigma_1}^2 = \frac{1}{n1}\sum_{i=1}^{n1}(X_i - \hat{\mu_1})^2$ and $\hat{\sigma_2}^2 = \frac{1}{n1}\sum_{i=1}^{n2}(Y_i - \hat{\mu_2})^2$.

2. We have: $\frac{n_1\hat{\sigma_1}^2}{\sigma_1^2} = \frac{n_1\frac{1}{n1}\sum_{i=1}^{n_1}(X_i-\mu_1)^2}{\sigma_1^2} = \sum_{i=1}^{n_1}(\frac{X_i-\mu_1}{\sigma_1})^2$. Since $\frac{X_i-\mu_1}{\sigma_1} \sim N(0,1)$, hence $\frac{n_1\hat{\sigma_1}^2}{\sigma_1^2} \sim \chi^2_{n1}$.
By similar method, we get $\frac{n_2\hat{\sigma_2}^2}{\sigma_2^2} \sim \chi^2_{n2}$

3. From previous result, $\frac{n_1\hat{\sigma_1}^2}{\sigma_1^2} + \frac{n_2\hat{\sigma_2}^2}{\sigma_2^2} \sim \chi^2_{n1} + \chi^2_{n2} \sim \chi^2_{n1+n2}$.

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. From the previous question, we get $\Delta = \hat{\mu_1} - \hat{\mu_2} = \bar{X_{n1}} - \bar{X_{n2}} \sim N(\mu_1 - \mu_2, \sigma_1^2 + \sigma_2^2)$.
Under H_0: $\Delta \sim N(0, \sqrt{\sigma_1^2 + \sigma_2^2})$. Hence, $\frac{\Delta}{\sigma_1^2+\sigma_2^2} \to \frac{\Delta}{\sqrt{\hat{\sigma_1}^2+\hat{\sigma_2}^2}} \sim N(0,1)$.
Let's denote test statistic $T = \Big|\frac{\Delta}{\sqrt{\hat{\sigma_1}^2+\hat{\sigma_2}^2}}\Big|$. We now $\mathbb{P}[T > c] = \alpha$, hence $c = 1 - \phi^{-1}(\frac{\alpha}{2})$.
Conclusion: If $T > c$, we reject H_0, otherwise, we fail to reject H_0.

6. In this case, from the previous qeuestion, we have $T = \Big|\frac{\Delta}{\sqrt{\hat{\sigma_1}^2+\hat{\sigma_2}^2}}\Big| = \frac{8.43-8.07}{\sqrt{0.22^2+0.17^2}} = 1.295$. Morever, $c = \phi^{-1}(1-\frac{\alpha}{2}) = \phi^{-1}(0.975) = 1.96$. Because $T < c$, we fail to reject H_0 which means we can conclude two machines are significantly identical.
p-value=$\mathbb{P}[|N(0,1)| > 1.295] = 1 - \mathbb{P}[|N(0,1)| \leqslant 1.295] = 1 - (2\mathbb{P}[N(0,1) \leqslant 1.295] - 1) = 2 - 2\phi(1.295) = 0.3$.

Problem 3. Let denote the function $\mathbb{R}^2 \longrightarrow \mathbb{R} : g(\mu, \sigma^2) = \mu - \sqrt{\sigma^2}$. We have:

$$\frac{\partial g}{\partial \mu} = 1$$

$$\frac{\partial}{\partial \sigma^2} = -\frac{1}{2\sigma}$$

Let denote $\hat{\mu} = \bar{X_n}$ and $\hat{\sigma}^2 = \frac{1}{n}\sum(X_i - \hat{\mu})^2$. Hence, by this result:

$$\sqrt{n}(\hat{\mu} - \mu) \longrightarrow N(0, \sigma^2).$$

$$\sqrt{n}(\hat{\sigma^2} - \sigma^2) \longrightarrow N(0, 2\sigma^4)$$

Hence $(\hat{\mu}, \hat{\sigma^2})$ is asymptotically normal with covariance matrix:

$$\sum = \begin{bmatrix} \sigma^2 & 0 \\ 0 & 2\sigma^4 \end{bmatrix}$$

Using Delta method we have:

$$\sqrt{n}(g(\hat{\mu}, \hat{\sigma}^2) - g(\mu, \sigma^2)) \longrightarrow N(0, \nabla^T_{\mu,\sigma^2}\Sigma\nabla_{\mu,\sigma^2}))$$

Where $\nabla_{\mu,\sigma^2} = \nabla g(\mu, \sigma^2) = [1 \ -\frac{1}{2\sigma}]^T$. Hence $\nabla^T_{\mu,\sigma^2}\Sigma\nabla_{\mu,\sigma^2} = [1 \ -\frac{1}{2\sigma}] \begin{bmatrix} \sigma^2 & 0 \\ 0 & 2\sigma^4 \end{bmatrix} \begin{bmatrix} 1 \\ -\frac{1}{2\sigma} \end{bmatrix} = \frac{3}{2}\sigma^2.$

Hence $\sqrt{n}(g(\hat{\mu}, \hat{\sigma}^2) - g(\mu, \sigma^2)) \longrightarrow N(0, \frac{3}{2}\sigma^2))$ which equivalents to $\sqrt{n}\frac{\sqrt{\frac{2}{3}}}{\sigma}(g(\hat{\mu}, \hat{\sigma}^2) - g(\mu, \sigma^2)) \longrightarrow N(0, 1).$
Computing second order Delta method, we have:

$$ng(\hat{\mu}, \hat{\sigma}^2)^T \frac{2}{3\sigma^2} g(\hat{\mu}, \hat{\sigma}^2) \longrightarrow \chi^2_1$$

Let denote c is $(1-\alpha)$-quantile of Kai-square with 1 degree of fredom. Because this is one-sided test statistic, if $T > c$ we reject null hypothesis, othewise, we fail to reject null hypothesis.
In our case, under null hypothesis, $T = ng(\hat{\mu}, \hat{\sigma}^2)^T \frac{2}{3\sigma^2} g(\hat{\mu}, \hat{\sigma}^2) > 100 \times (2.41 - \sqrt{5.2}) \times \frac{2}{3\times 5.2} \times (2.41 - \sqrt{5.2}) = 0.215$ and $c = 3.841$

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: $\mathbb{P}[M_n \leqslant t] = \Big(\frac{t}{\theta}\Big)^n$. 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:

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

(c) similar to the equation above.

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