Sta532 Hw5

Xu Chen 2/19/2020

Exercise 1

(a) Let $X \sim \chi^2(k)$, then we have

$$\begin{aligned}
M_X(t) = \mathsf{E}[\exp(tx)] &= \int_0^\infty e^{tx} \frac{1}{2^{k/2}\Gamma(k/2)} x^{k/2-1} e^{-x/2} \mathrm{d}x \\
&= \int_0^\infty \frac{1}{2^{k/2}\Gamma(k/2)} \underbrace{x^{k/2-1} e^{-(1/2-t)x}}_{\text{Gamma kernel } (t<1/2)} \mathrm{d}x \\
&= \frac{1}{2^{k/2}(1/2-t)^{k/2}} \int_0^\infty \frac{(1/2-t)^{k/2}}{\Gamma(k/2)} x^{k/2-1} e^{-(1/2-t)x} \mathrm{d}x \\
&= (1-2t)^{-k/2}, \quad t \in (-\infty, 1/2)
\end{aligned}$$

Let $X \sim \mathsf{Exp}(\lambda)$, then we have

$$\begin{aligned}
M_X(t) = \mathsf{E}[\exp(tx)] &= \int_0^\infty e^{tx} \lambda e^{-\lambda x} \mathrm{d}x \\
&= \frac{\lambda}{\lambda - t} \int_0^\infty (\lambda - t) e^{-(\lambda - t)x} \mathrm{d}x \\
&= \frac{\lambda}{\lambda - t}, \quad t \in (-\infty, \lambda)
\end{aligned}$$

Let $X \sim \mathsf{Ga}(\alpha, \beta)$, then we have

$$\begin{aligned}
M_X(t) = \mathsf{E}[\exp(tx)] &= \int_0^\infty e^{tx} \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} \mathrm{d}x \\
&= \frac{\beta^\alpha}{(\beta-t)^\alpha} \int_0^\infty \frac{(\beta-t)^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-(\beta-t)x} \mathrm{d}x \\
&= (1-t/\beta)^{-\alpha}, \quad t \in (-\infty, \beta)
\end{aligned}$$

(b) As $\{Y_i\}_{i=1}^n$ are i.i.d., the MGF of $Y = \sum_{i=1}^n Y_i$ is

$$M_Y(t) = \mathsf{E}\left[\exp\left(t \sum_{i=1}^n Y_i\right)\right] = \prod_{i=1}^n \mathsf{E}\left[\exp\left(tY_i\right)\right]$$

Therefore, if $\{Y_i\}_{i=1}^n \overset{\text{iid}}{\sim} \chi^2(k)$,

$$M_Y(t) = (1-2t)^{-nk/2}, \quad t \in (-\infty, 1/2)$$

and hence $Y \sim \chi^2(nk)$.

If $\{Y_i\}_{i=1}^n \overset{\text{iid}}{\sim} \mathsf{Exp}(\lambda)$,

$$M_Y(t) = \frac{\lambda^n}{(\lambda - t)^n}, \quad t \in (-\infty, \lambda)$$

and hence $Y \sim \mathsf{Ga}(n, \lambda)$.

(c) The MGF of $\chi^2(k)$ is

$$M_X(t) = \left(\frac{1/2}{1/2 - t}\right)^{k/2}$$

Comparing it with the MGF of $\mathsf{Ga}(\alpha, \beta)$, we know that $\chi^2(k)$ and $\mathsf{Ga}(k/2, 1/2)$ is the same distribution.

Exercise 2

Let W_1 and W_2 denote the lengths of approximate normal interval and Chebyshev interval, respectively. Then $R = W_1/W_2 = z_{1-\alpha/2}\sqrt{\alpha}$. A figure of R as a function of α is shown below.

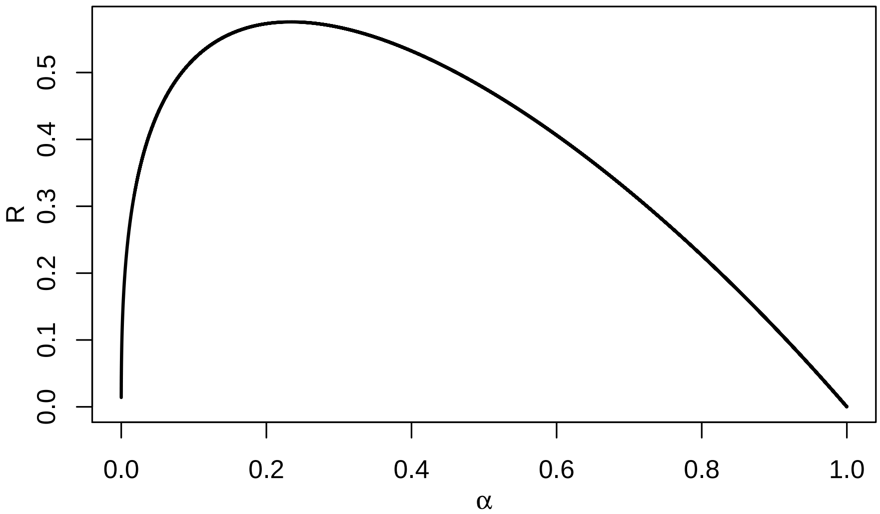

Figure 1: Relative length of intervals as a function of α.

Using simulations, the coverage probabilities (CP) are as follows.

	Normal		Laplace		Beta	
n	1	10	1	10	1	10
approx. normal	0.200	0.199	0.303	0.207	0.076	0.189
Chebyshev	0.737	0.736	0.794	0.747	0.846	0.728

Table 1: Coverage probabilities via simulation with nominal value $1 - \alpha = 0.2$.

Several conclusions can be made out of the figure and table above. First, approximate normal intervals are narrower than Chebyshev intervals. Because of its loose bounds, the CPs of Chebyshev intervals are much larger than the nominal values across all simulation scenarios. However, the CPs of approximate normal intervals are typically close to the nominal values. Second, CPs of both intervals become closer to the nominal values as sample size increases. Third, as the normal assumption is violated for both Laplace and Beta distributions, the CPs of approximate normal intervals become less accurate. This situation exacerbates for Beta distribution with shape parameters 0.1 and 0.5. This distribution has a bowl shape density with support on $(0, 1)$ which is far from a normal distribution and the CPs for approximate normal interval is clearly less than 0.2 with only one sample. However, for all scenarios, the CPs become close to the nominal value when sample size increases to 10. Therefore, z-intervals are robust based on our simulations.

Exercise 3

(a) Since $\{Y_i\}_{i=1}^n \overset{\text{iid}}{\sim} \mathsf{N}(0, \sigma^2)$, we have

$$\mathsf{E}[\overline{Y^2}] = \frac{1}{n}\sum_{i=1}^{n} \mathsf{E}[Y_i^2] = \mathsf{E}[Y_i^2] = \mathsf{E}[Y_i]^2 + \mathsf{Var}[Y_i] = \sigma^2$$

$$\mathsf{Var}[\overline{Y^2}] = \frac{1}{n}\mathsf{Var}[Y_i^2]$$

Since $Y_i \sim \mathsf{N}(0, \sigma^2)$, we have $Y_i/\sigma \sim \mathsf{N}(0, 1)$ and hence $(Y_i/\sigma)^2 \sim \chi^2(1)$. Therefore,

$$\mathsf{Var}[(Y_i/\sigma)^2] = 2$$

and we have $\mathsf{Var}[Y_i^2] = 2\sigma^4$. Therefore, $\mathsf{Var}[\overline{Y^2}] = 2\sigma^4/n$.

(b) We claim that $\overline{Y^2}$ converges to σ^2 in probability. As $\mathsf{E}[\overline{Y^2}] = \sigma^2$, we have

$$\mathsf{Pr}\left(|\overline{Y^2} - \sigma^2| > \varepsilon\right) \leqslant \frac{\mathsf{Var}[\overline{Y^2}]}{\varepsilon^2} = \frac{2\sigma^4}{n\varepsilon^2} \to 0 \quad \text{as } n \to +\infty \text{ for any } \varepsilon > 0$$

by Chebyshev's inequality.

(c) Let $Z_i = (Y_i/\sigma)^2$. Then we know that $\{Z_i\}_{i=1}^n \overset{\text{iid}}{\sim} \chi^2(1)$ and hence

$$\begin{aligned} M_{\overline{Y^2}}(t) &= \mathsf{E}[\exp(t/n\sum_{i=1}^{n} Y_i^2)] \\ &= \mathsf{E}[\exp(t\sigma^2/n\sum_{i=1}^{n} Z_i)] \\ &= (\mathsf{E}[\exp(t\sigma^2/nZ_i)])^n \\ &= (1 - 2\sigma^2/nt)^{-n/2} \end{aligned}$$

Comparing it with the result from problem 1 part (a), we know that $\overline{Y^2} \sim \mathsf{Ga}\left(\frac{n}{2}, \frac{n}{2\sigma^2}\right)$. By CLT, since Y_i are i.i.d. with finite mean and variance, we have $\overline{Y^2} \sim \mathsf{N}(\sigma^2, 2\sigma^4/n)$ asymptotically.

Exercise 4

(a) As $\sum_{i=1}^n w_i = 1$, we have

$$\mathsf{E}[\bar{Y}_w] = \sum_{i=1}^n w_i \mathsf{E}[Y_i] = \sum_{i=1}^n w_i \mu = \mu$$
$$\mathsf{Var}[\bar{Y}_w] = \sum_{i=1}^n w_i^2 \mathsf{Var}[Y_i] = \sigma^2 \sum_{i=1}^n w_i^2 a_i$$

(b) Two approaches are provided. One uses Cauchy-Schwarz inequality and the other uses Lagrange multiplier. The goal is to minimize $\mathsf{Var}[\bar{Y}_w] = \sigma^2 \sum_{i=1}^n w_i^2 a_i$ with respect to $(w_1, \ldots, w_n)$.

(1) Note that, by Cauchy-Schwarz inequality,

$$\left(\sum_{i=1}^n (w_i\sqrt{a_i})^2\right)\left(\sum_{i=1}^n (1/\sqrt{a_i})^2\right) \geqslant \left(\sum_{i=1}^n w_i\sqrt{a_i} \times 1/\sqrt{a_i}\right)^2 = \left(\sum_{i=1}^n w_i\right)^2 = 1$$

Hence, we have $\sum_{i=1}^n w_i^2 a_i \geqslant 1/\sum_{i=1}^n a_i^{-1}$ and the equation holds if and only if $w_i\sqrt{a_i} = k/\sqrt{a_i}$. Since we have $\sum_{i=1}^n w_i = 1$, the minimum of variance is achieved when $w_i = a_i^{-1}/\sum_{i=1}^n a_i^{-1}$.

(2) Let $w = (w_1, \ldots, w_n)$. We can construct the Lagrange function as

$$\mathcal{L}(w, \lambda) = \sum_{i=1}^n w_i^2 a_i + \lambda\left(\sum_{i=1}^n w_i - 1\right)$$

Setting all partial derivatives to 0, we have

$$\frac{\partial \mathcal{L}}{\partial w_i} = 2w_i a_i + \lambda = 0, \quad i = 1, \ldots, n$$
$$\frac{\partial \mathcal{L}}{\partial \lambda} = \sum_{i=1}^n w_i - 1 = 0$$

The solution is $w_i = a_i^{-1}/\sum_{i=1}^n a_i^{-1}, \quad i = 1, \ldots, n$ and $\lambda = -2/\sum_{i=1}^n a_i^{-1}$.

(c) We have $\mathsf{E}[\bar{Y}_w] = \mu$,

$$\mathsf{Var}[\bar{Y}_w] = \sigma^2 \sum_{i=1}^n \frac{a_i^{-2} a_i}{(\sum_{i=1}^n a_i^{-1})^2} = \sigma^2 \frac{\sum_{i=1}^n a_i^{-1}}{(\sum_{i=1}^n a_i^{-1})^2} = \frac{\sigma^2}{\sum_{i=1}^n a_i^{-1}}$$

Therefore, if $\sum_{i=1}^n a_i^{-1} \to +\infty$ as $n \to +\infty$, then we have $\bar{Y}_w$ converges in probability to μ. This result makes intuitive sense as it requires that the variance of Y_i cannot be too large (growing with i) such that a sample Y_i contains almost no information of its mean. For example, if $a_i = i^2$, then $\sum_{i=1}^n a_i^{-1} = \pi^2/6$ and we will not have WLLN.

Exercise 5

(a) Define $\hat{F}(y) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}(Y_i \leqslant y)$ where $\mathbb{1}(\cdot)$ is an indicator function. To show it is unbiased and consistent, let $Z_i = \mathbb{1}(Y_i \leqslant y)$. Then we have

$$Z_i = \begin{cases} 1 & \text{if } Y_i \leqslant y \\ 0 & \text{otherwise} \end{cases}$$

Therefore, since $Y_i \overset{\text{iid}}{\sim} F$, we have $Z_i \overset{\text{iid}}{\sim} \mathsf{Ber}(F(y))$.

$$\begin{aligned} \mathsf{E}[\hat{F}(y)] &= \frac{1}{n}\sum_{i=1}^{n} \mathsf{E}[Z_i] = F(y) \\ \mathsf{Var}[\hat{F}(y)] &= \frac{1}{n^2}\sum_{i=1}^{n} \mathsf{Var}[Z_i] = \frac{F(y)(1-F(y))}{n} \end{aligned}$$

Therefore, $\hat{F}(y)$ is unbiased and consistent as its variance goes to 0 as $n \to +\infty$.

(b) By CLT, we have

$$\hat{F}(y) \xrightarrow{\text{d}} \mathsf{N}\left(F(y), \frac{F(y)(1-F(y))}{n}\right)$$

as $n \to +\infty$ and hence

$$\frac{\hat{F}(y) - F(y)}{\sqrt{\frac{F(y)(1-F(y))}{n}}} \xrightarrow{\text{d}} \mathsf{N}(0,1)$$

Since $\hat{F}(y)$ converges in probability to $F(y)$, we have

$$\begin{aligned} \frac{\hat{F}(y) - F(y)}{\sqrt{\frac{F(y)(1-F(y))}{n}}} \times \frac{\sqrt{\frac{F(y)(1-F(y))}{n}}}{\sqrt{\frac{\hat{F}(y)(1-\hat{F}(y))}{n}}} &\xrightarrow{\text{d}} \mathsf{N}(0,1) \\ \frac{\hat{F}(y) - F(y)}{\sqrt{\frac{\hat{F}(y)(1-\hat{F}(y))}{n}}} &\xrightarrow{\text{d}} \mathsf{N}(0,1) \end{aligned}$$

Therefore, an approximate 95% CI of $F(y)$ is

$$\left(\hat{F}(y) - z_{0.975}\sqrt{\frac{\hat{F}(y)(1-\hat{F}(y))}{n}}, \hat{F}(y) + z_{0.975}\sqrt{\frac{\hat{F}(y)(1-\hat{F}(y))}{n}}\right)$$

Remark 1. We should note here that $F(y)$ is unknown and we hope to perform inference on this unknown quantity. A confidence interval can be constructed using $\hat{F}(y)$. The estimator $\hat{F}(y)$ is a known quantity which is a summary of data. Therefore, the confidence interval cannot contain $F(y)$.