Southern University of Science and Technology
Department of Mathematics

MA204: Mathematical Statistics

Tutorial 3: Examples/Solutions

A. Order Statistics

A.1 Definition

Let $X_1, \ldots, X_n$ be a random sample from a population with cdf $F(\cdot)$ and pdf $f(\cdot)$. Then, $X_{(1)} \leqslant \cdots \leqslant X_{(n)}$ are called the *order statistics*.

How to understand $X_{(1)} = \min(X_1, \ldots, X_n)$?

```
> x <- rnorm(4, mean=2, sd=0.2)                         min(x)
Sample 1: 1.792966 1.935051 2.205505 1.717259          1.717259
Sample 2: 1.736903 2.092398 1.955097 2.292306          1.736903
Sample 3: 3.100611 2.144726 1.855688 1.862706          1.855688
Sample 4: 2.087154 1.733278 2.185363 1.823968          1.733278
Sample 5: 1.821756 1.861677 1.785434 1.997736          1.785434
Sample 6: 2.209436 1.836078 2.413644 1.998032          1.836078
Sample 7: 1.862467 1.938282 2.011536 2.223131          1.862467
Sample 8: 1.927599 2.012464 2.003386 1.899905          1.899905
```

A.2 Single order statistic

Let $G_r(x)$ and $g_r(x)$ be the cdf and pdf of the r-th order statistic $X_{(r)}$, respectively. Then

$$\begin{aligned}
G_r(x) &= \sum_{i=r}^{n} \binom{n}{i} F^i(x)[1-F(x)]^{n-i} \\
&= \frac{1}{B(r, n-r+1)} \int_0^{F(x)} t^{r-1}(1-t)^{n-r}\,\mathrm{d}t, \\
g_r(x) &= \frac{n!}{(r-1)!(n-r)!} f(x) F^{r-1}(x)[1-F(x)]^{n-r}.
\end{aligned}$$

A.3 Multiple order statistics

Let $g_{r_1,\ldots,r_k}(x_1,\ldots,x_k)$ be the joint pdf of $X_{(r_1)},\ldots,X_{(r_k)}$ $(1 \leqslant r_1 \leqslant \cdots \leqslant r_k \leqslant n;\ 1 \leqslant k \leqslant n)$,

$$
\begin{aligned}
g_{r_1,\ldots,r_k}(x_1,\ldots,x_k) &= n!\left[\prod_{i=1}^{k} f(x_i)\right]\cdot\prod_{i=0}^{k}\left\{\frac{[F(x_{i+1})-F(x_i)]^{r_{i+1}-r_i-1}}{(r_{i+1}-r_i-1)!}\right\},\\
g_{1,\ldots,r}(x_1,\ldots,x_r) &= \frac{n!}{(n-r)!}f(x_1)\cdots f(x_r)[1-F(x_r)]^{n-r},\\
g_{r,s}(x,y) &= \frac{n!}{(r-1)!(s-r-1)!(n-s)!}f(x)f(y)\\
&\quad\times F^{r-1}(x)[F(y)-F(x)]^{s-r-1}[1-F(y)]^{n-s}.
\end{aligned}
$$

Example T3.1: Let $X_1,\ldots,X_n$ be independent exponential r.v.'s and $X_i \sim \text{Exponential}(\lambda_i)$. Show that

(i) $Y = \min(X_1,\ldots,X_n)$ is also an exponential r.v. with parameter $\lambda_1+\cdots+\lambda_n$.

(ii) For a fixed i, $\Pr(Y = X_i) = \Pr(X_i \leqslant X_j,\ \forall j = 1,\ldots,n; j \neq i) = \frac{\lambda_i}{\lambda_1+\cdots+\lambda_n}$.

Solution: (i) Although $X_1,\ldots,X_n$ are independent, they have different distributions, the formula $G_1(\cdot)$ on page 51 of Lecture Notes cannot be applied. Since

$$F_{X_i}(x) = (1-\mathrm{e}^{-\lambda_i x})\cdot I_{(0,\infty)}(x),$$

we have

$$
\begin{aligned}
\Pr(Y \leqslant y) &= 1-\Pr(Y > y)\\
&= 1-\Pr(\text{all } X_i > y)\\
&= 1-\prod_{i=1}^{n}\Pr(X_i > y) \quad \text{(since all } X_i \text{ are independent)}\\
&= 1-\prod_{i=1}^{n}[1-\Pr(X_i \leqslant y)]\\
&= 1-\prod_{i=1}^{n}[1-F_{X_i}(y)]\\
&= 1-\mathrm{e}^{-(\sum_{i=1}^{n}\lambda_i)y}.
\end{aligned}
$$

Thus, $Y \sim \text{Exponential}(\lambda_1 + \cdots + \lambda_n)$.

(ii) Let $Y_i = \min(X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n)$, $\lambda_{-i} = \sum_{j=1, j\neq i}^{n} \lambda_j$, and $Z_i = X_i - Y_i$, we have

$$\Pr(Y = X_i) = \Pr(X_i \leqslant X_j,\ \forall j \neq i) = \Pr(X_i \leqslant Y_i) = \Pr(Z_i \leqslant 0),$$

and

$$\begin{aligned}
F_{Z_i}(z) &= \Pr(X_i - Y_i \leqslant z) \\
&= \int_0^\infty \int_{x-z}^\infty f_{X_i, Y_i}(x, y)\,\mathrm{d}y\,\mathrm{d}x \qquad (x - y \leqslant z \Rightarrow x - z \leqslant y) \\
&= \int_0^\infty \int_{x-z}^\infty f_{X_i}(x) f_{Y_i}(y)\,\mathrm{d}y\,\mathrm{d}x \quad \text{(since } X_i \text{ and } Y_i \text{ are independent)} \\
&= -\int_0^\infty \int_z^{-\infty} f_{X_i}(x) f_{Y_i}(x-u)\,\mathrm{d}u\,\mathrm{d}x \quad (\text{let } y = x - u) \\
&= \int_{-\infty}^z \int_0^\infty f_{X_i}(x) f_{Y_i}(x-u)\,\mathrm{d}x\,\mathrm{d}u \\
&= \int_{-\infty}^z \int_0^\infty \lambda_i e^{-\lambda_i x} \lambda_{-i} e^{-\lambda_{-i}(x-u)}\,\mathrm{d}x\,\mathrm{d}u \\
&= \int_{-\infty}^z \frac{\lambda_i \lambda_{-i}}{\lambda_1 + \cdots + \lambda_n} e^{\lambda_{-i} u}\,\mathrm{d}u \\
&= \left[\frac{\lambda_i}{\lambda_1 + \cdots + \lambda_n} \mathrm{e}^{\lambda_{-i} u}\right]\Bigg|_{-\infty}^z \\
&= \frac{\lambda_i}{\lambda_1 + \cdots + \lambda_n} e^{\lambda_{-i} z}.
\end{aligned}$$

Therefore, $\Pr(Y = X_i) = \Pr(Z_i \leqslant 0) = F_{Z_i}(0) = \frac{\lambda_i}{\lambda_1 + \cdots + \lambda_n}$. ‖

B. Convergence in Distribution

B.1 Definition

Let the distribution function $F_n(x)$ of the r.v. X_n depends upon n ($n = 1, 2, \ldots$). If $F(x)$ is a distribution function of a r.v. X and

$$\lim_{n\to\infty} F_n(x) = F(x), \quad \forall x \text{ s.t. } F(x) \text{ is continuous},$$

then the sequence of r.v.'s $X_1, X_2, \ldots$ *converge in distribution* to X and denote as $X_n \xrightarrow{\mathrm{L}} X$.

B.2 **Theorem**

Let X_n, $n = 1, 2, \ldots$ have a moment generating function $M(t; n)$, $t \in (-h, h)$. If there exists a moment generating function $M(t)$ with respect to the distribution function $F(x)$ such that

$$M(t) = \lim_{n\to\infty} M(t; n), \quad \forall t \in (-h, h),$$

then $X_n \xrightarrow{\text{L}} X$.

Example T3.2: Let the sequence of r.v. $X_n \sim N(0, \frac{1}{n})$. Show that $X_n \xrightarrow{\text{L}} X$, where the cdf of X is $F(x) = \begin{cases} 1, & \text{if } x > 0 \\ 0, & \text{if } x < 0 \\ \frac{1}{2}, & \text{if } x = 0 \end{cases}$.

Solution: Let $\phi(s) = \exp(-s^2/2)/\sqrt{2\pi}$ denote the pdf of $N(0, 1)$, then the cdf of X_n is

$$F_n(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi/n}} \mathrm{e}^{-\frac{t^2}{2/n}}\,\mathrm{d}t = \int_{-\infty}^{\sqrt{n}x} \phi(s)\,\mathrm{d}s \quad (\text{let } s = \sqrt{n}t).$$

so that $\lim_{n\to\infty} F_n(x) = \lim_{n\to\infty} \int_{-\infty}^{\sqrt{n}x} \phi(s)\,\mathrm{d}s$

$$= \int_{-\infty}^{\lim_{n\to\infty}\sqrt{n}x} \phi(s)\,\mathrm{d}s$$

$$= \begin{cases} \displaystyle\int_{-\infty}^{\infty} \phi(s)\,\mathrm{d}s = 1, & \text{if } x > 0, \\ \displaystyle\int_{-\infty}^{-\infty} \phi(s)\,\mathrm{d}s = 0, & \text{if } x < 0, \\ \displaystyle\int_{-\infty}^{0} \phi(s)\,\mathrm{d}s = \frac{1}{2}, & \text{if } x = 0 \end{cases}$$

$$= F(x).$$

Therefore, $X_n \xrightarrow{\text{L}} X$. ‖

C. Central Limit Theorem (CLT)

If $X_1, \ldots, X_n$ be a sequence of i.i.d. r.v.'s with the mean μ and the variance σ^2, then the r.v. $\sqrt{n}(\overline{X}_n - \mu)/\sigma$ has a standard normal limiting distribution, where $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$.

$$\text{i.e., } \sqrt{n}(\overline{X}_n - \mu)/\sigma \xrightarrow{\text{L}} N(0, 1).$$

Example T3.3: Let X follow a negative binomial distribution, denoted by $X \sim \text{NBinomial}$ $(20, 0.7)$. (i) Exactly calculate the value of $\Pr(X = 12)$. (ii) Approximately calculate $\Pr(X = 12)$ by the Central Limit Theorem.

Solution: (i) The pmf of $X \sim \text{NBinomial}(r, p)$ is

$$\Pr(X = x) = \binom{x + r - 1}{x} p^r (1-p)^x, \qquad x = 0, 1, 2, \ldots$$

where $E(X) = r(1-p)/p$ and $\text{Var}(X) = r(1-p)/p^2$. We directly compute

$$\Pr(X = 12) = \binom{12 + 20 - 1}{12} \times 0.7^{20} \times 0.3^{12} = 0.0598.$$

(ii) Now,

$$E(X) = \frac{20 \times (1 - 0.7)}{0.7} = 8.5714 \quad \text{and} \quad \text{Var}(X) = \frac{20 \times (1 - 0.7)}{0.7^2} = 12.2449.$$

By CLT,

$$\frac{X - E(X)}{\sqrt{\text{Var}(X)}} \xrightarrow{\text{L}} Z \sim N(0, 1).$$

We obtain

$$\begin{aligned}
\Pr(X = 12) &= P(12 - 0.5 < X < 12 + 0.5) \\
&= \Pr\left(\frac{11.5 - 8.5714}{\sqrt{12.2449}} < \frac{X - 8.5714}{\sqrt{12.2449}} < \frac{12.5 - 8.5714}{\sqrt{12.2449}}\right) \\
&= \Pr(0.8369 < Z < 1.1227) \\
&= \Phi(1.1227) - \Phi(0.8369) \\
&= 0.8692 - 0.7987 = 0.0705.
\end{aligned}$$

The error is $0.0705 - 0.0598 = 0.0107$. And the percentage error is

$$\frac{0.0107}{0.0598} = 17.87\%. \qquad \|$$

Example T3.4: Let $X_i \overset{\text{i.i.d.}}{\sim} \text{Poisson}(0.03)$, $i = 1, \ldots, 50$ and $Y = \sum_{i=1}^{50} X_i$. Calculate $\Pr(Y \geqslant 3)$.

Solution: $E(X_i) = 0.03$ and $\mathrm{Var}(X_i) = 0.03$. By CLT,

$$\begin{aligned}
Z &= \sqrt{n} \cdot \frac{\overline{X} - E(X_i)}{\sqrt{\mathrm{Var}(X_i)}} \\
&= \sqrt{50} \cdot \frac{Y/50 - 0.03}{\sqrt{0.03}} \\
&= \frac{Y - 50 \times 0.03}{\sqrt{50 \times 0.03}} \\
&= \frac{Y - 1.5}{\sqrt{1.5}} \xrightarrow{\mathrm{L}} N(0,1).
\end{aligned}$$

Therefore,

$$\begin{aligned}
\Pr(Y \geqslant 3) &= \Pr\left(Z = \frac{Y - 1.5}{\sqrt{1.5}} \geqslant \frac{3 - 1.5}{\sqrt{1.5}}\right) \\
&= 1 - \Pr(Z < 1.224745) \\
&= 1 - \Phi(1.224745) \\
&= 1 - 0.8896643 \\
&= 0.1103357. \qquad \|
\end{aligned}$$