Introduction to Probability and Statistics Welcome Tutorial :-)

Tutorial 9

Lecturer: Ming Gao

DaSE @ ECNU

Nov. 7, 2023

1. (1). Finding the pdf of $X_{(1)}$:
$X_{(1)}$ is the minimum value among n independent and identically distributed (i.i.d.) continuous random variables $X_1, X_2, \cdots, X_n$. Its cumulative distribution function (CDF) can be expressed as:

$$F_{X_{(1)}}(x) = P(X_{(1)} \leq x) = 1 - P(X_{(1)} > x)$$

Since $X_{(1)}$ is the minimum, all X_i need to be greater than x for $X_{(1)} > x$:

$$P(X_{(1)} > x) = P(X_1 > x, X_2 > x, \cdots, X_n > x)$$

As the X_i are independent, we have:

$$P(X_{(1)} > x) = P(X_1 > x) \cdot P(X_2 > x) \cdots P(X_n > x)$$

Since they are identically distributed, each $P(X_i > x)$ is equal to $1 - F_X(x)$. Therefore:

$$P(X_{(1)} > x) = (1 - F_X(x))^n$$

Hence:

$$F_{X_{(1)}}(x) = 1 - (1 - F_X(x))^n$$

To find the probability density function, differentiate the CDF:

$$f_{X_{(1)}}(x) = \frac{d}{dx} F_{X_{(1)}}(x) = \frac{d}{dx}\left[1 - (1 - F_X(x))^n\right]$$
$$= n \cdot (1 - F_X(x))^{n-1} \cdot f_X(x)$$

(2). Finding the pdf of $X_{(n)}$: $X_{(n)}$ is the maximum value among n independent and identically distributed continuous random variables. Its CDF can be expressed as:

$$F_{X_{(n)}}(x) = P(X_{(n)} \le x)$$

Since $X_{(n)}$ is the maximum, all X_i must be less than or equal to x for $X_{(n)} \le x$:

$$P(X_{(n)} \le x) = P(X_1 \le x, X_2 \le x, \cdots, X_n \le x)$$

As the X_i are independent, we have:

$$P(X_{(n)} \leq x) = P(X_1 \leq x) \cdot P(X_2 \leq x) \cdots P(X_n \leq x)$$

Again, since they are identically distributed, each $P(X_i \leq x)$ is equal to $F_X(x)$. Therefore:

$$F_{X_{(n)}}(x) = [F_X(x)]^n$$

To find the probability density function, differentiate the CDF:

$$f_{X_{(n)}}(x) = \frac{d}{dx} F_{X_{(n)}}(x) = \frac{d}{dx}[F_X(x)]^n$$

$$= n \cdot [F_X(x)]^{n-1} \cdot f_X(x)$$

In summary, the pdf of $X_{(1)}$ is $f_{X_{(1)}}(x) = n \cdot (1 - F_X(x))^{n-1} \cdot f_X(x)$, and the pdf of $X_{(n)}$ is $f_{X_{(n)}}(x) = n \cdot [F_X(x)]^{n-1} \cdot f_X(x)$.

2. Use $f_X(x) = 1/\theta, F_X(x) = x/\theta, 0 < x < \theta$. Let $Y = X_{(n)}, Z = X_{(1)}$. Then, from Theorem 5.4.6,

$$f_{Z,Y}(z,y) = \frac{n!}{0!(n-2)!0!} \cdot \frac{1}{\theta} \cdot \frac{1}{\theta} \cdot \left(\frac{z}{\theta}\right)^0 \cdot \left(\frac{y-z}{\theta}\right)^{n-2} \cdot \left(1 - \frac{y}{\theta}\right)^0$$

$$= \frac{n(n-1)}{\theta^n} \cdot (y-z)^{n-2}, \quad 0 < z < y < \theta$$

Now let $W = Z/Y, Q = Y$. Then $Y = Q, Z = WQ$, and $|J| = q$. Therefore

$$f_{W,Q}(w,q) = \frac{n(n-1)}{\theta^n}(q - wq)^{n-2} \cdot q$$
$$= \frac{n(n-1)}{\theta^n}(1-w)^{n-2} \cdot q^{n-1},$$

where $0 < w < 1, 0 < q < \theta$

The joint pdf factors into functions of w and q, and, hence, W and Q are independent.

3. $f_{X_{(i)}|X_{(j)}}(u \mid v) = f_{X_{(i)},X_{(j)}}(u,v)/f_{X_{(j)}}(v)$. Consider two cases, depending on which of i or j is greater. Using the formulas from Theorems 5.4.4 and 5.4.6, and after cancellation, we obtain the following.
(i) If $i < j$,

$$\begin{aligned} f_{X_{(i)}|X_{(j)}}(u \mid v) &= \frac{(j-1)!}{(i-1)!(j-i-1)!} \cdot f_X(u) \cdot F_X^{i-1}(u) \\ &\cdot [F_X(v) - F_X(u)]^{j-i-1} \cdot F_X^{1-j}(v) \\ &= \frac{(j-1)!}{(i-1)!(j-i-1)!} \cdot \frac{f_X(u)}{F_X(v)} \cdot \left[\frac{F_X(u)}{F_X(v)}\right]^{i-1} \\ &\cdot \left[1 - \frac{F_X(u)}{F_X(v)}\right]^{j-i-1}, \quad u < v. \end{aligned}$$

This is the pdf of the i th order statistic from a sample of size $j-1$, from a population with pdf given by the truncated distribution, $f(u) = f_X(u)/F_X(v)$, $u < v$.

(ii) If $j < i$ and $u > v$,

$$f_{X_{(i)}|X_{(j)}}(u \mid v) = \frac{(n-j)!}{(n-1)!(i-1-j)!} f_X(u)\left[1-F_X(u)\right]^{n-i}\left[F_X(u)-F_X(v)\right]^{i-1-j} \left[1-F_X(v)\right]^{j-n}$$

$$= \frac{(n-j)!}{(i-j-1)!(n-i)!} \frac{f_X(u)}{1-F_X(v)} \left[\frac{F_X(u)-F_X(v)}{1-F_X(v)}\right]^{i-j-1} \left[1-\frac{F_X(u)-F_X(v)}{1-F_X(v)}\right]^{n-i}$$

This is the pdf of the $(i-j)$ th order statistic from a sample of size $n-j$, from a population with pdf given by the truncated distribution, $f(u) = f_X(u)/(1-F_X(v))\,, u > v$.

4. From the CLT we have, approximately, $\bar{X}_1 \sim \mathrm{n}\left(\mu, \sigma^2/n\right), \bar{X}_2 \sim \mathrm{n}\left(\mu, \sigma^2/n\right)$. Since $\bar{X}_1$ and $\bar{X}_2$ are independent, $\bar{X}_1 - \bar{X}_2 \sim \mathrm{n}\left(0, 2\sigma^2/n\right)$. Thus, we want

$$\begin{aligned}
.99 &\approx P\left(\left|\bar{X}_1 - \bar{X}_2\right| < \sigma/5\right) \\
&= P\left(\frac{-\sigma/5}{\sigma/\sqrt{n/2}} < \frac{\bar{X}_1 - \bar{X}_2}{\sigma/\sqrt{n/2}} < \frac{\sigma/5}{\sigma/\sqrt{n/2}}\right) \\
&\approx P\left(-\frac{1}{5}\sqrt{\frac{n}{2}} < Z < \frac{1}{5}\sqrt{\frac{n}{2}}\right),
\end{aligned}$$

where $Z \sim \mathrm{n}(0,1)$. Thus we need $P(Z \geq \sqrt{n}/5(\sqrt{2})) \approx .005$. From Table 1, $\sqrt{n}/5\sqrt{2} = 2.576$, which implies $n = 50(2.576)^2 \approx 332$

5. We know that $\sigma^2_{\bar{X}} = 9/100$. Use Chebyshev's Inequality to get

$$P(-3k/10 < \bar{X} - \mu < 3k/10) \geq 1 - 1/k^2.$$

We need $1 - 1/k^2 \geq .9$ which implies $k \geq \sqrt{10} = 3.16$ and $3k/10 = .9487$. Thus

$$P(-.9487 < \bar{X} - \mu < .9487) \geq .9$$