STAT 30400 : Distribution Theory

Fall 2019

•

Homework 1

•

Solutions by

Jinhong Du

12243476

STAT 30400, Homework 1

1. (10 pts) Let $F_1, F_2, \ldots, F_n, \cdots$ and F be distribution functions with corresponding left-continuous inverses $F_1^-, F_2^-, \ldots, F_n^-, \cdots$ and F^-. Show that

$$\lim_{n\to\infty} F_n(x) = F(x) \text{ for all continuity points } x \text{ of } F$$

if and only if

$$\lim_{n\to\infty} F_n^-(u) = F^-(u) \text{ for all continuity points } u \text{ of } F^-$$

Proof. First we have the following fact.

Let B be either $\mathbb{R}$ or $(0,1)$. Let f be a non-decreasing mapping from B into $\mathbb{R}$. Then $\mathcal{D}_f = \{x \in B : f \text{ is discontinuous at } x\}$ is at most countable and $\mathcal{D}_f^c$ is a dense set. To see this, first look at a bounded interval $[n, n+1)$ $\forall\, n \in \mathbb{Z}$. If $f(n+1) \neq f(n)$, let D_n denote the set of points at which f has a discontinuity. For each positive integer m, let $D_{n,m}$ denote the set of points $x \in [n, n+1)$ such that f has a jump of at least $\frac{1}{m}(f(n+1) - f(n))$ at x and let $N_{n,m}$ denote the number of elements in $D_{n,m}$. Note that

$$D_n = \bigcup_{m=1}^{\infty} D_{n,m},$$

we have $N_{n,m} \leq m$. It follows that the number of points of discontinuity is bounded by $\sum_n \sum_{m=1}^{\infty} m$. The result follows. Then $\mathcal{D}_F$ and $\mathcal{D}_{F^-}$ are both countable.

$\Longrightarrow$

Suppose that $u \in (0,1)$ and that w is a continuity point of F with $F^-(u) > w$. Since

$$F^-(u) > w \iff F(w) < u$$

and $\lim\limits_{n\to\infty} F_n(x) = F(x)$ for all continuity points x of F, we have $\lim\limits_{n\to\infty} F_n(w) < u$. Therefore, $\exists\, N_1 = N_1(w) \in \mathbb{Z}_+$, s.t. $\forall\, n > N_1$, $F_n(w) < u$, i.e. $F_n^-(u) > w$.

Suppose that $u \in (0,1)$ and that y is a continuity point of F with $F^-(u+) < y$. Since

$$F^-(u) \leq y \iff F(y) \geq u$$

Similarly, we have $\lim\limits_{n\to\infty} F_n(y) \geq u$. Therefore, $\exists\, N_2 = N_2(y) \in \mathbb{Z}_+$, s.t. $\forall\, n > N_2$, $F_n(y) \geq u$, i.e. $F_n^-(u) \leq y$.

Since the set $\mathcal{D}_F$ of jumps of the distribution function F is at most countable, the complement of countable set $\mathbb{R} \setminus \mathcal{D}_F$ is dense. $\forall\, \epsilon > 0$, which makes $w = F^-(u) - \epsilon$, $y = F^-(u+) + \epsilon \in \mathbb{R} \setminus \mathcal{D}_F$, $\exists\, N = \max\{N_1, N_2\}$, s.t. $\forall\, n > N$,

$$F^-(u) - \epsilon \leq F_n^-(u) \leq F^-(u+) + \epsilon.$$

Since $\mathbb{R} \setminus \mathcal{D}_F$ is dense, such ϵ can be arbitrarily small. Let $\epsilon \to 0$, we have

$$F^-(u) \leq \liminf_n F_n^-(u) \leq \limsup_n F_n^-(u) \leq F^-(u+),$$

which implies

$$\lim_n F_n^-(u) = F^-(u) \text{ for all continuity points } u \text{ of } F^-.$$

$\Longleftarrow$

Solution (cont.)

Suppose that $x \in \mathbb{R}$ and that w' is a continuity point of F^- with $F(x) < w'$. Since

$$F(u) < w' \iff F^-(w') > x$$

and $\lim_{n\to\infty} F_n^-(x) = F^-(x)$ for all continuity points u of F^-, we have $\lim_{n\to\infty} F_n^-(w') > x$. Therefore, $\exists\ N_3 = N_3(w') \in \mathbb{Z}_+$, s.t. $\forall\ n > N_3$, $F_n^-(w') > x$, i.e. $F_n(x) < w'$.

Suppose that $x \in \mathbb{R}$ and that y' is a continuity point of F^- with $F(x-) \geq y'$. Since

$$F(x) \geq y' \iff F^-(y') \leq x$$

Similarly, we have $\lim_{n\to\infty} F_n^-(y') \leq x$. Therefore, $\exists\ N_4 = N_4(y') \in \mathbb{Z}_+$, s.t. $\forall\ n > N_4$, $F_n^-(y') \leq x$, i.e. $F_n(x) \geq y'$.

Since the set $\mathcal{D}_{F^-}$ of jumps of the distribution function F^- is at most countable, the complement of countable set $(0,1) \setminus \mathcal{D}_{F^-}$ is dense. $\forall\ \epsilon' > 0$, which makes $w' = F(x) + \epsilon$, $y' = F(x-) - \epsilon \in (0,1) \setminus \mathcal{D}_{F^-}$, $\exists\ N' = \max\{N_3, N_4\}$, s.t. $\forall\ n > N'$,

$$F(x-) - \epsilon' \leq F_n(x) \leq F(x) + \epsilon'.$$

Since $(0,1) \setminus \mathcal{D}_F$ is dense, such ϵ' can be arbitrarily small. Let $\epsilon' \to 0$, we have

$$F(x-) \leq \liminf_n F_n(x) \leq \limsup_n F_n(x) \leq F(x),$$

which implies

$$\lim_n F_n(x) = F(x) \text{ for all continuity points } x \text{ of } F.$$

$\square$

2. (10 pts) Let X be a random variable with df F. Find the distribution functions of $|X|$ and X^+, where X^+ is equal to X if $X > 0$ and is equal to 0 otherwise.

$$\begin{aligned}
F(x) &= \mathbb{P}(X \leq x) \\
F_{|X|}(x) &= \mathbb{P}(|X| \leq x) \\
&= \mathbb{P}(-x \leq X \leq x)\mathbb{1}_{\{x\geq 0\}} \\
&= [F(x) - F((-x)-)]\mathbb{1}_{\{x\geq 0\}} \\
F_{X^+}(x) &= \mathbb{P}(X^+ \leq x) \\
&= [\mathbb{P}(X \leq 0) + \mathbb{P}(X \leq x | X > 0)\mathbb{P}(X > 0)]\mathbb{1}_{\{x\geq 0\}} \\
&= F(x)\mathbb{1}_{\{x\geq 0\}}
\end{aligned}$$

3. (10 pts) Suppose that $X_1, \ldots, X_n$ are independent and that each has distribution function F. Let $X_{(k)}$ be the k-th order statistic. Show that $X_{(k)}$ has the distribution function

$$G_k(x) = \sum_{j=k}^{n} \binom{n}{j} [F(x)]^j [1 - F(x)]^{n-j}.$$

Proof.

$$\begin{aligned}
G_k(x) &= \mathbb{P}(X_{(k)} < x) \\
&= \mathbb{P}(X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(k)} < x, X_{(k)} \leq X_{(k+1)} \leq \cdots \leq X_{(n)}) \\
&= \sum_{j=k}^{n} \mathbb{P}(X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(j)} \leq x < X_{(j+1)} \leq \cdots \leq X_{(n)}) \\
&= \sum_{j=k}^{n} \binom{n}{j} \mathbb{P}(X_1 \leq x, X_2 \leq x, \cdots, X_j \leq x, x < X_{j+1}, \cdots, x \leq X_n) \\
&= \sum_{j=k}^{n} \binom{n}{j} \mathbb{P}(X_1 \leq x) \cdots \mathbb{P}(X_j) \leq x) \mathbb{P}(x < X_{j+1}) \cdots \mathbb{P}(x \leq X_n) \\
&= \sum_{j=k}^{n} \binom{n}{j} [F(x)]^j [1 - F(x)]^{n-j}
\end{aligned}$$

where the fourth equality comes from the fact that the event $X_{(1)} \leq \cdots \leq X_{(j)} \leq x < X_{(j+1)} \leq \cdots \leq X_{(n)}$ is simply the event that j observations fall in the interval $(-\infty, x]$ and $n - j$ observations fall in the interval $(x, +\infty)$. $\square$

4. (10 pts) The *Levy distance* $d(F, G)$ between two distribution functions is the infimum of those ϵ such that

$$G(x-\epsilon)-\epsilon \leq F(x) \leq G(x+\epsilon)+\epsilon, \qquad \forall x.$$

(a) Verify that this is a metric on the set of distribution functions.

Let $\mathcal{D}$ be the set of distribution functions. For any $F, G\mathcal{D}$, the Levy distance is given by

$$d(F,G) = \inf_{\epsilon} \mathcal{E}(F,G)$$

where $\mathcal{E}(F,G) = \{\epsilon | G(x-\epsilon)-\epsilon \leq F(x) \leq G(x+\epsilon)+\epsilon,\ \forall x.\}$.

1. **Nonnegativity**

Suppose $F \neq G$. since G is non-decreasing, we have $\forall\ \epsilon < 0$,

$$G(x-\epsilon)-\epsilon \geq G(x+\epsilon)-\epsilon > G(x+\epsilon)+\epsilon$$

thus $\epsilon < 0$ cannot satisfy the conditon that $G(x-\epsilon)-\epsilon \leq F(x) \leq G(x+\epsilon)+\epsilon,\ \forall x.$, i.e., $d(F,G) \geq 0$. Analogously, $d(F,F) \geq 0$.

Next we show that $d(F,G) \neq 0$. Otherwise, if $d(F,G) = 0$, then either $0 \in \mathcal{E}(F,G)$ or $\exists\ \{\epsilon_n \in \mathcal{E}(F,G),\ n = 1, 2, \cdots\}$, s.t. $\epsilon_n \downarrow 0$. In the former case, we have $F(x) = G(x)$, which is a contradiction. In the latter case, it implies $G(x-) \leq F(x) \leq G(x),\ \forall\ x$. Since $G(x) - F(x) \geq 0$ and $\int_{\mathbb{R}} [G(x) - F(x)] \mathrm{d}x = 0$, we have $G(x) = F(x)$, almost everywhere in $\mathbb{R}$. Then, $\forall\ x$ such that $G(x-) \leq F(x) < G(x)$, $\exists\ \{\delta_n,\ n = 1, 2, \cdots\}$, s.t. $\delta_n \downarrow 0$ and $F(x+\delta_n) = G(x+\delta_n)$ (otherwise $F(x) < G(x)$ at a small right neighborhood of x, which has positive measurement.). Since F and G are right-continuous function, it implies $F(x) = G(x)$ as $\delta_n \downarrow 0$, which is a contradiction again. Therefore, $d(F,G) > 0$ for $F \neq G$.

From the above argument, if $d(F,G) = 0$, then $F = G$. Apparently $0 \in \mathcal{E}(F,F)$, so $d(F,F) = 0$. Therefore, $d(F,G) = 0 \iff F = G$.

2. **Symmetry**

$\forall\ \epsilon \in \mathcal{E}(F,G), \forall\ x \in \mathbb{R}$, $G(x-\epsilon)-\epsilon \leq F(x) \leq G(x+\epsilon)+\epsilon$. Let $y = x-\epsilon$, from the left hand side, we have $G(y) \leq F(y+\epsilon)+\epsilon$, $\forall\ y$. Let $y = x+\epsilon$, from the left hand side, we have $F(y-\epsilon)-\epsilon \leq G(y)$, $\forall\ y$. Therefore, $F(x-\epsilon)-\epsilon \leq G(x) \leq F(x+\epsilon)+\epsilon,\ \forall x$, i.e., $\epsilon \in \mathcal{E}(G,F)$, i.e., $\mathcal{E}(F,G) \subseteq \mathcal{E}(G,F)$. Similarly, $\mathcal{E}(F,G) \supseteq \mathcal{E}(G,F)$. Then, $\mathcal{E}(F,G) = \mathcal{E}(G,F)$, which means $d(E,F) = d(F,E)$.

3. **Triangle Inequality**

For any $F, G, H\mathcal{D}$, $\forall\ \epsilon_1 \in \mathcal{E}(F,G), \epsilon_2 \in \mathcal{E}(G,H)$, we have $\forall\ x \in \mathbb{R}$,

$$G(x-\epsilon_1)-\epsilon_1 \leq F(x) \leq G(x+\epsilon_1)+\epsilon_1, \tag{1}$$

$$H(x-\epsilon_2)-\epsilon_2 \leq G(x) \leq H(x+\epsilon_2)+\epsilon_2. \tag{2}$$

Substituting (2) into (1),

$$H(x-(\epsilon_1+\epsilon_2))-(\epsilon_1+\epsilon_2) \leq F(x) \leq H(x+(\epsilon_1+\epsilon_2))+(\epsilon_1+\epsilon_2),$$

i.e., $\epsilon_1 + \epsilon_2 \in \mathcal{E}(F,H)$. Therefore,

$$\mathcal{E}(F,G) + \mathcal{E}(G,H) \triangleq \{\epsilon_1+\epsilon_2 | \epsilon_1 \in \mathcal{E}(F,G), \epsilon_2 \in \mathcal{E}(G,H)\} \subseteq \mathcal{E}(F,H). \tag{3}$$

Next we prove the fact that $\inf(\mathcal{E}(F,G)+\mathcal{E}(G,H)) = \inf \mathcal{E}(F,G) + \inf \mathcal{E}(G,H)$ given every set has finite infimum. $\forall\ \epsilon > 0$, by definition of the infimum, there exists x in $\mathcal{E}(F,G)$ and y in $\mathcal{E}(G,H)$ such that

$$x < \inf \mathcal{E}(F,G) + \frac{\epsilon}{2}, \qquad y < \inf \mathcal{E}(G,H) + \frac{\epsilon}{2}.$$

Solution (cont.)

By summing,

$$x + y < \inf \mathcal{E}(F, G) + \inf \mathcal{E}(G, H) + \epsilon.$$

Let $z = x + y$. Then, for all $z \in \inf(\mathcal{E}(F, G) + \mathcal{E}(G, H))$, $z < \inf \mathcal{E}(F, G) + \inf \mathcal{E}(G, H) + \epsilon$. Since $\inf \mathcal{E}(F, G) \leq x$, $\inf \mathcal{E}(G, H) \leq y$, we also have $\inf \mathcal{E}(F, G) + \inf \mathcal{E}(G, H) \leq z$. Let $\epsilon \to 0$, we have

$$\inf(\mathcal{E}(F, G) + \mathcal{E}(G, H)) = \inf \mathcal{E}(F, G) + \inf \mathcal{E}(G, H). \tag{4}$$

From (3) and (4), we have

$$\begin{aligned} \inf \mathcal{E}(F, G) + \inf \mathcal{E}(G, H) &= \inf(\mathcal{E}(F, G) + \mathcal{E}(G, H)) \\ &\geq \inf \mathcal{E}(F, H), \end{aligned}$$

i.e.,

$$d(F, H) \leq d(F, G) + d(G, H).$$

So, the Levy distance is a metric.

(b) Let F be a continuous distribution function, and $(F_n)_n$ a sequence of distribution functions. Show that $F_n(x)$ converges to $F(x)$ for any x if and only if $d(F_n, F) \to 0$.

Proof. $\Longrightarrow$

$\forall\ \epsilon > 0$, since F is continuous, we can choose $x_0, y_0 \in \mathbb{R}$ such that $F(x_0) = \epsilon$ and $F(y_0) = 1 - \epsilon$.
(1) For $x > y_0$, since $F_n \to F$, we have that for y_0, $\exists\ N_y \in \mathbb{Z}_+$, s.t. $\forall\ n > N_y$, $|F_n(y_0) - F(y_0)| < \epsilon$. So,

$$F(x) \leq 1 = F(y_0) + \epsilon \leq F_n(y_0) + 2\epsilon \leq F_n(x + 3\epsilon) + 3\epsilon$$

and

$$F(x) \geq 1 - \epsilon = F(y_0) \geq F_n(y_0) - \epsilon \geq F_n(x) - 3\epsilon \geq F_n(x - 3\epsilon) - 3\epsilon,$$

where the third inequality comes from that for $x > y_0$, $F_n(x) - 2\epsilon \leq 1 - 2\epsilon \leq F_n(y_0)$ since $F_n(y_0) \geq F(y_0) - \epsilon = 1 - 2\epsilon$.
(2) For $x < x_0$. For x_0, $\exists\ N_0 \in \mathbb{Z}_+$, s.t. $\forall\ n > N_0$, $|F_n(x_0) - F(x_0)| < \epsilon$. So,

$$F(x) \geq 0 = F(x_0) - \epsilon \geq F_n(x_0) - 2\epsilon \geq F_n(x) - 2\epsilon \geq F_n(x - 3\epsilon) - 3\epsilon,$$

and

$$F(x) \leq \epsilon = F(x_0) \leq F_n(x_0) + \epsilon \leq F_n(x) + 3\epsilon \leq F_n(x + 3\epsilon) + 3\epsilon$$

where the third inequality comes from that for $x < x_0$, $F_n(x) + 2\epsilon \geq 2\epsilon \geq F_n(x_0)$ since $F_n(x_0) \leq F(x_0) + \epsilon = 2\epsilon$.
(3) Let $x_0 < x_1 < \cdots < x_m = y_0$ such that $x_{i+1} - x_i < \epsilon$ $(i = 1, \ldots, m-1)$. For x_i, $\exists\ N_i \in \mathbb{Z}_+$, s.t. $\forall\ n > N_i$, $|F_n(x_i) - F(x_i)| < \epsilon$. For $x \in [x_i, x_{i+1}]$,

$$F(x) \leq F(x_{i+1}) \leq F(x_i) + \epsilon \leq F_n(x_i) + 2\epsilon \leq F_n(x) + 2\epsilon \leq F_n(x + 3\epsilon) + 3\epsilon$$

Solution (cont.)

and

$$F(x) \geq F(x_i) \geq F_n(x_{i+1}) - \epsilon \geq F_n(x) - \epsilon \geq F_n(x - 3\epsilon) - 3\epsilon$$

Then $\exists\ N = \max\{N_0, \dots, N_m, N_y\}$, s.t.

$$F_n(x - 3\epsilon) - 3\epsilon \leq F(x) \leq F_n(x + 3\epsilon) + 3\epsilon, \qquad \forall\ x \in \mathbb{R},$$

which implies that $3\epsilon \in \mathcal{E}(F_n, F)$ and $d(F_n, F) = \inf \mathcal{E}(F_n, F) \leq 3\epsilon$. Let $\epsilon \to 0$, we have $d(F_n, F) \to 0$.

$\Longleftarrow$

Since $d(F_n, F) \to 0$, $\forall\ \epsilon > 0$, $\exists\ N \in \mathbb{Z}_+$, s.t. $\forall n > N$, $d(F_n, F) < \epsilon$. Then $\forall\ x \in \mathbb{R}$,

$$F(x - \epsilon) - \epsilon \leq F_n(x) \leq F(x + \epsilon) + \epsilon,$$

Since F is a continuous function, let $\epsilon \to 0$, we have

$$F(x) \leq \liminf_n F_n(x) \leq \limsup_n F_n(x) \leq F(x),$$

So

$$\lim_{n \to \infty} F_n(x) = F(x),$$

i.e. $F_n(x)$ converges to $F(x)$ for any x. $\square$

5. (10 pts) Suppose Z is a standard normal variable, with density $\phi(z)$.

(a) Show that as $z \to \infty$,

$$\mathbb{P}(Z \geq z) = (1 + o(1))\frac{\phi(z)}{z}$$

where $o(1)$ denotes a quantity that tends to 0 as $z \to \infty$. (Hint: integration by parts).

$$\begin{aligned}
\mathbb{P}(Z \geq z) &= \int_z^{+\infty} \phi(x)\mathrm{d}x \\
&= \int_z^{+\infty} \left(-\frac{1}{x}\right)\left(-x\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}\right)\mathrm{d}x \\
&= \int_z^{+\infty} \left(-\frac{1}{x}\right)\mathrm{d}\left(\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}\right) \\
&= \left(-\frac{1}{x}\cdot\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}\right)\Bigg|_z^{+\infty} - \int_z^{+\infty}\frac{\phi(x)}{x^2}\mathrm{d}x \\
&= \frac{\phi(z)}{z} - \int_z^{+\infty}\frac{\phi(x)}{x^2}\mathrm{d}x
\end{aligned}$$

Since

$$\begin{aligned}
\lim_{z\to\infty}\frac{-\int_z^{+\infty}\frac{\phi(x)}{x^2}\mathrm{d}x}{\frac{\phi(z)}{z}} &= \lim_{z\to\infty}\frac{\frac{\phi(z)}{z^2}}{\frac{-z^2\phi(z)-\phi(z)}{z^2}} \\
&= \lim_{z\to\infty}\frac{1}{-z^2-1} \\
&= 0,
\end{aligned}$$

$$-\int_z^{+\infty}\frac{\phi(x)}{x^2}\mathrm{d}x = o(1)\frac{\phi(z)}{z}.$$

So as $z \to \infty$,

$$\mathbb{P}(Z \geq z) = (1 + o(1))\frac{\phi(z)}{z}.$$

(b) Let q_α be the $(1-\alpha)$ quantile of Z. Show that, as $\alpha \to 0$,

$$q_\alpha = \sqrt{2\log(\frac{1}{\alpha}) - \log(\log(\frac{1}{\alpha})) - \log(4\pi) + o(1)},$$

where $o(1)$ denotes a quantity that tends to 0 as $\alpha \to 0$.

Since q_α is the $(1-\alpha)$ quantile of Z, we have

$$\mathbb{P}(Z \geq q_\alpha) = \alpha,$$

i.e.,

$$\begin{aligned}(1+o(1))\frac{\phi(q_\alpha)}{\alpha} &= \alpha \\ \frac{(1+o(1))}{\alpha} &= \sqrt{2\pi}e^{\frac{1}{2}q_\alpha^2}q_\alpha.\end{aligned}$$

Squaring both sides,

$$\frac{(1+o(1))}{\alpha^2} = 2\pi e^{q_\alpha^2}q_\alpha^2.$$

Let $x = q_\alpha^2$,

$$\frac{(1+o(1))}{\alpha^2} = 2\pi e^x x.$$

Take the log transform,

$$\log(1+o(1)) + 2\log\frac{1}{\alpha} = \log(2\pi) + \log x + x, \tag{1}$$

then the right side is dominated by the term x as $x \to \infty$ (equivalently, $\alpha \to 0$). Suppose that $x = 2\log\frac{1}{\alpha} - \log\left(2\log\frac{1}{\alpha}\right) - \log(2\pi) + g(\alpha)$ such that $\lim_{\alpha\to 0+}\frac{g(\alpha)}{2\log\frac{1}{\alpha}} = 0$ and substitute it in (1), we have

$$\log(1+o(1)) = \log\left(2\log\frac{1}{\alpha} - \log\left(2\log\frac{1}{\alpha}\right) - \log(2\pi) + g(\alpha)\right) - \log\left(2\log\frac{1}{\alpha}\right) + g(\alpha).$$

Since as $\alpha \to 0$,

$$\begin{aligned}&\log(1+o(1)) \to 0 \\ &\log\left(2\log\frac{1}{\alpha} - \log\left(2\log\frac{1}{\alpha}\right) - \log(2\pi) + g(\alpha)\right) - \log\left(2\log\frac{1}{\alpha}\right) \to 0,\end{aligned}$$

we have $g(\alpha) \to 0$, i.e. $g(\alpha) = o(1)$ where $o(1)$ denotes a quantity that tends to 0 as $\alpha \to 0$. Therefore, as $\alpha \to 0$,

$$x = 2\log\frac{1}{\alpha} - \log\left(2\log\frac{1}{\alpha}\right) - \log(2\pi) + o(1).$$

For $\alpha < \frac{1}{2}$, $q_\alpha > 0$. So as $\alpha \to 0$,

$$\begin{aligned}q_\alpha &= \sqrt{2\log(\frac{1}{\alpha}) - \log(2\log(\frac{1}{\alpha})) - \log(2\pi) + o(1)} \\ &= \sqrt{2\log(\frac{1}{\alpha}) - \log(\log(\frac{1}{\alpha})) - \log(4\pi) + o(1)}.\end{aligned}$$