STAT 512: Statistical Inference **Autumn 2022**

Quiz Session 10: Practice Final Problems

Yikun Zhang *December 7, 2022*

Problem 1 (Final problem in Autumn 2018). *Suppose that X, Y, Z are independent standard normal $N(0,1)$ random variables. Let*

$$U = X, \quad V = X + Y, \quad W = X + Y + Z.$$

(a) What is the distribution of the random vector (U, V, W)?

(b) What is the conditional distribution of $(V, W) \,|\, U$? Are V and W conditionally independent given U? (Justify your answer.)

(c) What is the conditional distribution of $(U, W) \,|\, V$? Are U and W conditionally independent given V? (Justify your answer.)

(d) What is the conditional distribution of $(U, V) \,|\, W$? Are U and V conditionally independent given W? (Justify your answer.)

Solution. (a) Note that $(X, Y, Z)^T \sim N_3(0, \boldsymbol{I}_3)$ and

$$\begin{pmatrix} U \\ V \\ W \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} := A \begin{pmatrix} X \\ Y \\ Z \end{pmatrix}.$$

By the linearity of multivariate normal distributions, the distribution of (U, V, W) is

$$\begin{pmatrix} U \\ V \\ W \end{pmatrix} \sim N_3\left(A \cdot \mathbf{0}, A\boldsymbol{I}_3 A^T\right) = N_3\left[\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 2 \\ 1 & 2 & 3 \end{pmatrix}\right].$$

(b) By the conditional normal distribution $X_1 | X_2 \sim N_{n_1}\left(\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(X_2 - \mu_2), \Sigma_{11,2}\right)$ with $\Sigma_{11,2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$ that we derived in the lecture, we know that

$$\begin{aligned} \begin{pmatrix} V \\ W \end{pmatrix} \,|\, U &\sim N_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 1 \\ 1 \end{pmatrix} 1^{-1}(U - 0), \begin{pmatrix} 2 & 2 \\ 2 & 3 \end{pmatrix} - \begin{pmatrix} 1 \\ 1 \end{pmatrix} 1^{-1}(1, 1)\right) \\ &= N_2\left(\begin{pmatrix} U \\ U \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}\right). \end{aligned}$$

Since $\text{Cov}(V, W | U) = 1 \neq 0$, V and W are not conditionally independent given U.

(c) Similarly, we know that

$$\begin{aligned} \begin{pmatrix} U \\ W \end{pmatrix} \,|\, V &\sim N_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 1 \\ 2 \end{pmatrix} 2^{-1}(V - 0), \begin{pmatrix} 1 & 1 \\ 1 & 3 \end{pmatrix} - \begin{pmatrix} 1 \\ 2 \end{pmatrix} 2^{-1}(1, 2)\right) \\ &= N_2\left(\begin{pmatrix} \frac{V}{2} \\ V \end{pmatrix}, \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & 1 \end{pmatrix}\right). \end{aligned}$$

Since $\text{Cov}(U, W | V) = 0$, U and W are conditionally independent given V.

(d) Analogously, we know that

$$\begin{aligned}\begin{pmatrix} U \\ V \end{pmatrix} \mid W &\sim N_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 1 \\ 2 \end{pmatrix} 3^{-1}(W-0), \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} - \begin{pmatrix} 1 \\ 2 \end{pmatrix} 3^{-1}(1,2) \right) \\ &= N_2\left(\begin{pmatrix} \frac{W}{3} \\ \frac{2W}{3} \end{pmatrix}, \begin{pmatrix} 2/3 & 1/3 \\ 1/3 & 2/3 \end{pmatrix} \right).\end{aligned}$$

Since $\mathrm{Cov}(U, V|W) = \frac{2}{3} \neq 0$, U and V are not conditionally independent given W. □

Problem 2 (Final problem in Autumn 2018)**.** *Two identical sealed envelopes are placed side-by-side on a table in front of you. One contains a real number a, the other contains a real number b, where a and b are unknown and fixed (nonrandom) with $a \neq b$. You choose one of the envelopes at random according to the toss of a fair coin and open it. Let X_1 denote the (random) number (either a or b) in the envelope that you choose and let X_2 denote the number in the second (unopened) envelope, which you give to your friend Tom. You open the first envelope and observe X_1. Consider the following two strategies:*

(I) Tell Tom to open the second envelope and reveal X_2. If $X_1 > X_2$, Tom pays you \$1; if $X_1 < X_2$, you pay Tom \$1.

(II) Switch envelopes with Tom and open the second envelope, revealing X_2. Now, if $X_1 > X_2$, you pay Tom \$1; if $X_1 < X_2$, Tom pays you \$1.

Question (a): Which of these two strategies is preferable, or are they equivalent? Justify your answer.

Now consider a third strategy, which combines (I) and (II) as follows:

(III) After observing X_1, you generate a standard normal random variable $Z \sim N(0,1)$ independently of the coin toss and use Z as a surrogate for X_2 as follows: if $X_1 > Z$, then you proceed according to Strategy (I); while if $X_1 < Z$, then you proceed according to Strategy (II).

Question (b): Which of these three strategies is preferable, or are they equivalent? Justify your answer by means of a quantitative argument (not hand-waving).

Solution. Without loss of generality, we assume that $a < b$.

(a) Note that the joint distribution of (X_1, X_2) is

$$\mathbb{P}(X_1 = a, X_2 = b) = \mathbb{P}(X_1 = b, X_2 = a) = \frac{1}{2}.$$

Hence,

$$\mathbb{P}\left(\text{Strategy (I) wins you \$1}\right) = \mathbb{P}\left(\text{Strategy (II) wins you \$1}\right) = \frac{1}{2},$$

so Strategies (I) and (II) are equivalent.

(b) We decompose the probability of winning \$1 in Strategy (III) as follows:

$$\begin{aligned}\mathbb{P}\left(\text{Strategy (III) wins you \$1}\right) &= \mathbb{P}(\text{Enter Strategy (I) when } X_1 = b) + \mathbb{P}(\text{Enter Strategy (II) when } X_1 = a) \\ &= \mathbb{P}(X_1 = b, X_1 > Z) + \mathbb{P}(X_1 = a, X_1 < Z) \\ &= \mathbb{P}(Z < b|X_1 = b) \cdot \mathbb{P}(X_1 = b) + \mathbb{P}(Z > a|X_1 = a) \cdot \mathbb{P}(X_1 = a) \\ &= \Phi(b) \cdot \frac{1}{2} + [1 - \Phi(a)] \cdot \frac{1}{2} \\ &= \frac{1}{2} + \frac{1}{2}\left[\Phi(b) - \Phi(a)\right]\end{aligned}$$

$$> \frac{1}{2},$$

where Φ is the CDF of $N(0,1)$ and $\Phi(b) > \Phi(a)$ by our assumption. Therefore, Strategy (III) is preferable in both cases. □

The following problems are partially adopted from quiz sessions developed by Apara Venkat and Zhen Miao for STAT 512 in Autumn 2021.

Problem 3. *Suppose that a density f_0 has its support on some fixed interval $I \subset (0,\infty)$, e.g., $I = (a,b)$ with $0 < a < b$, and we observe $X_1, ..., X_n$ from a distribution f_S satisfying*

$$f_S(x) = \frac{x f_0(x)}{\mu_0} \quad \text{with} \quad \mu_0 = \int_a^b t f_0(t)\, dt.$$

Assume that $\nu_0 = \int_a^b \frac{f_0(t)}{t} dt < \infty$ and we define

$$\mu_S = \int_0^\infty u f_S(u)\, du.$$

(a) Show that $\mu_S > \mu_0$.

(b) Using (a), prove that $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ is not a consistent estimator of μ_0.

(c) Define $Y_i = \frac{1}{X_i}$ for $i = 1, \dots, n$. Show that $1/\bar{Y}_n = n/\left(\sum_{i=1}^n \frac{1}{X_i}\right)$ is a consistent estimator of μ_0.

Proof. (a) By direct calculations,

$$\begin{aligned}
\mu_S &= \int_0^\infty u f_S(u) du \\
&= \int_0^\infty u \frac{u f_0(u)}{\mu_0} du \\
&= \frac{1}{\mu_0}\int_0^\infty u^2 f_0(u) du \\
&= \frac{1}{\mu_0}\mathbb{E}_{Z\sim f_0}\left[Z^2\right] \\
&> \frac{1}{\mu_0}\left[\mathbb{E}_{Z\sim f_0}(Z)\right]^2 \quad \text{(by Jensen's inequality)} \\
&= \mu_0,
\end{aligned}$$

where the equality does not hold because $g(z) = z^2$ is a nonlinear function and f_0 has its support as an interval.

(b) From law of large numbers, $\bar{X}_n \xrightarrow{p} \mu_S$. By (a), we know that $\bar{X}_n$ cannot be a consistent estimator of μ_0.

(c) We start by finding the mean of $Y = \frac{1}{X}$ as

$$\begin{aligned}
\mathbb{E}Y &= \int_0^\infty \frac{1}{x} f_S(x) dx \\
&= \int_0^\infty \frac{f_0(x)}{\mu_0} dx \\
&= \frac{1}{\mu_0}
\end{aligned}$$

By law of large numbers, $\bar{Y}_n \xrightarrow{p} \frac{1}{\mu_0}$. Finally, by continuous mapping theorem, we have $1/\bar{Y}_n \xrightarrow{p} \mu_0$. □

Problem 4. *Let $A, B, C \overset{i.i.d.}{\sim}$ Uniform$(0,1)$. Consider the polynomial $f(x) = Ax^2 + Bx + C$. What is the probability that $f(x)$ has real roots?*

Solution. Recall that a quadratic polynomial has real roots if and only if $B^2 \geq 4AC$. Therefore,

$$\begin{aligned}\mathbb{P}(f(x) \text{ has real roots}) &= \mathbb{P}(B^2 \geq 4AC) \\ &= \mathbb{P}(-2\log B \leq -\log 4 - \log A - \log C).\end{aligned}$$

Now, let $U = -\log B, V = -\log A, W = -\log C$. It can be shown that U, V, W are i.i.d. Exponential(1) because the CDF of U is

$$\mathbb{P}(U \leq u) = \mathbb{P}(-\log A \leq u) = \mathbb{P}(A \geq e^{-u}) = 1 - e^{-u} \quad \text{with} \quad u \in (0, \infty).$$

In addition, $2U \sim$ Exponential$(1/2), V + W \sim$ Gamma$(2,1)$. Thus, define $X = 2U, Y = V + W$. We have,

$$\begin{aligned}\mathbb{P}(f(x) \text{ has real roots}) &= \int_{y=\log 4}^{\infty} \int_{x=0}^{-\log 4+y} \frac{1}{2} e^{-x/2} dx y e^{-y} dy \\ &= \int_{y=\log 4}^{\infty} (1 - e^{(\log 4 - y)/2}) y e^{-y} dy \\ &= \int_{y=\log 4}^{\infty} (1 - 2e^{-y/2}) y e^{-y} dy \\ &= \int_{y=\log 4}^{\infty} (ye^{-y} - 2ye^{-3y/2}) dy\end{aligned}$$

Note that

$$\begin{aligned}\int te^{-kt} dt &= t \int e^{-kt} dt - \int -\frac{1}{k} e^{-kt} dt \\ &= -\frac{e^{-kt}}{k^2}(kt+1) + c \\ \implies \int_a^{\infty} te^{-kt} dt &= \frac{e^{ak}}{k^2}(ak+1).\end{aligned}$$

Therefore,

$$\begin{aligned}\mathbb{P}(f(x) \text{ has real roots}) &= \int_{y=\log 4}^{\infty} (ye^{-y} - 2ye^{-3y/2}) dy \\ &= \frac{e^{-\log 4}}{1^2}(\log 4 + 1) - 2\frac{e^{-3\log 4/2}}{(3/2)^2}\left(\frac{3}{2}\log 4 + 1\right) \\ &= \frac{1 + \log 4}{4} - \frac{1}{6}\left(\log 4 + \frac{2}{3}\right).\end{aligned}$$

□

Problem 5. *Let $X_1, \ldots, X_n$ be independent and identically distributed. Define $X_{(n)} = \max_{1 \leq i \leq n} X_i$.*

(a) Suppose $X_i \sim$ Beta$(1, \beta)$. Find a $\gamma \geq 0$ such that $n^{\gamma}(1 - X_{(n)})$ converges in distribution.

(b) Suppose $X_i \sim$ Exponential(1). Find a sequence of constants $\{a_n\}_{n=1}^{\infty}$ such that $X_{(n)} - a_n$ converges in distribution.

Note that both questions concern the existence of such values. We do not need to enumerate all possible values that satisfy the results.

Solution. (a) Note that the density of X_i is

$$p_X(x) = \frac{(1-x)^{\beta-1}}{B(1,\beta)} = \beta(1-x)^{\beta-1}$$

Now, let us find the distribution of $X_{(n)}$.

$$\begin{aligned}
P(X_{(n)} \le x) &= [P(X_1 \le x)]^n \\
&= \left[\int_0^x \beta(1-t)^{\beta-1} dt\right]^n \\
&= \left[\int_{1-x}^1 \beta u^{\beta-1} du\right]^n \\
&= \left[1-(1-x)^\beta\right]^n
\end{aligned}$$

Next, let us analyze the quantity in question. Assume $x \in (0,1)$.

$$\begin{aligned}
P(n^\gamma(1-X_{(n)}) \le x) &= P\left(1 - X_{(n)} \le \frac{x}{n^\gamma}\right) \\
&= P\left(X_{(n)} \ge 1 - \frac{x}{n^\gamma}\right) \\
&= 1 - P\left(X_{(n)} \le 1 - \frac{x}{n^\gamma}\right) \\
&= 1 - \left[1 - \frac{x^\beta}{n^{\beta\gamma}}\right]^n
\end{aligned}$$

First, we note that if $\gamma = 0$, then

$$\begin{aligned}
\lim_{n\to\infty} P(n^\gamma(1-X_{(n)}) \le x) &= \lim_{n\to\infty} 1 - \left[1 - x^\beta\right]^n \\
&= 1
\end{aligned}$$

Therefore, $X_{(n)} \xrightarrow{p} 1$. Think about why this makes sense. Second, we analyze for the general $\gamma \neq 0$. Assume the limit L exists

$$\begin{aligned}
L &= \lim_{n\to\infty} \left[1 - \frac{x^\beta}{n^{\beta\gamma}}\right]^n \\
\implies \log L &= \lim_{n\to\infty} \frac{\log\left(1 - \frac{x^\beta}{n^{\beta\gamma}}\right)}{\frac{1}{n}}
\end{aligned}$$

Note for all $\gamma > 0$, the limit is of the form, $0/0$. So we use L'Hospital's rule,

$$\begin{aligned}
\log L &= \lim_{n\to\infty} \frac{\frac{1}{\left(1-\frac{x^\beta}{n^{\beta\gamma}}\right)} \frac{\beta\gamma x^\beta}{n^{\beta\gamma+1}}}{-\frac{1}{n^2}} \\
&= \lim_{n\to\infty} -n^2 \frac{\beta\gamma x^\beta}{n^{\beta\gamma+1} - n x^\beta} \\
&= \lim_{n\to\infty} -\frac{\beta\gamma x^\beta}{n^{\beta\gamma-1} - \frac{x^\beta}{n}}
\end{aligned}$$

If $\beta\gamma = 1$, we have $n^{\beta\gamma-1} = 1$. This gives us

$$\begin{aligned} \log L &= -\beta\gamma x^{\beta} = -x^{\beta} \\ \implies L &= e^{-x^{\beta}} \\ \implies \lim_{n\to\infty} P(n^{\gamma}(1 - X_{(n)}) \leq x) &= 1 - e^{-x^{\beta}} \end{aligned}$$

This limit is actually the CDF of the Weibull distribution. This is a distribution that commonly appears in survival analysis and failure analysis in reliability engineering. It is also considered an *extreme-value distribution* which has connections with logit-choice models.[1]

We worked out a very similar problem on November 10, 2021. Compare the two results and see if you observe any similarities. It might help to notice that $\text{Beta}(1,1) \overset{d}{=} \text{Uniform}(0,1)$. As practice, you can work out what happens when $\gamma \neq 1/\beta$.

(b) First, let us find the distribution of $X_{(n)}$,

$$\begin{aligned} P(X_{(n)} \leq x) &= (P(X_1 \leq x))^n \\ &= \left(1 - e^{-x}\right)^n \\ \implies P(X_{(n)} - a_n \leq x) &= P(X_{(n)} \leq x + a_n) \\ &= \left(1 - e^{-x-a_n}\right)^n \end{aligned}$$

Thinking along the same lines as part (a),

$$\lim_{n\to\infty} P(X_{(n)} - a_n \leq x) = \lim_{n\to\infty} \left(1 - \frac{e^{-x}}{e^{a_n}}\right)^n$$

Note that this limit becomes well-defined when $e^{a_n} = n$ i.e., $a_n = \log n$. Then, we have

$$\begin{aligned} \lim_{n\to\infty} P(X_{(n)} - a_n \leq x) &= \lim_{n\to\infty} \left(1 - \frac{e^{-x}}{n}\right)^n \\ &= e^{-e^{-x}} \end{aligned}$$

This turns out to be the CDF of the Gumbel distribution.[2] This is also an extreme-value distribution and has connections with logistic regression (see the practice problem below).

As practice, try to work out what happens when $a_n = k \log n$, where $k > 1$ is an integer.

As more practice, consider the two independent random variables X, Y (with support all over $\mathbb{R}$) drawn from a Gumbel distribution with parameters $\mu_X, \mu_Y \in \mathbb{R}$ and $\beta > 0$. The density is given by

$$\begin{aligned} p_X(x) &= \frac{1}{\beta} e^{-\left(\frac{x-\mu_X}{\beta} + e^{-x}\right)} \\ p_Y(y) &= \frac{1}{\beta} e^{-\left(\frac{x-\mu_Y}{\beta} + e^{-y}\right)} \end{aligned}$$

Show that $X - Y$ follows a logistic distribution with parameters $(\mu_X - \mu_Y, \beta)$. A logistic distribution with parameters a, b has CDF

$$\frac{1}{1 + e^{-(x-a)/b}}, \quad x \in \mathbb{R}$$

[1] See https://en.wikipedia.org/wiki/Generalized_extreme_value_distribution

[2] This distribution is named after Emil Gumbel, who developed extreme value theory. Interestingly, he investigated political murders in Germany after losing his friend. Eventually the Nazis forced him out of this professorship at the University of Heidelberg.

You might recognize the CDF as the sigmoid function from your machine learning classes.

Also show that if $Z \sim \text{Exponential}(1)$, then $-\log Z$ follows a standard Gumbel distribution i.e., has parameters $\mu = 0, \beta = 1$. $\square$

Problem 6. *Let $X_1, \ldots, X_n \overset{i.i.d.}{\sim}$ Beta$(\theta, 1)$, whose density is $\theta x^{\theta-1}$ with $\theta > 0$ and $0 < x < 1$. Find the MLE of θ. Also find the MLE of $1/\theta$. Is the MLE of $1/\theta$ unbiased?*

Solution. We have the likelihood of the data,

$$\begin{aligned}\mathcal{L}(X_1, \ldots, X_n; \theta) &= \prod_{i=1}^{n} p_{X_i}(x_i; \theta) \\ &= \theta^n \left(\prod_{i=1}^{n} x_i \right)^{\theta-1}\end{aligned}$$

Therefore the log-likelihood is,

$$\begin{aligned}\ell(\theta) &= \log \mathcal{L}(X_1, \ldots, X_n; \theta) \\ &= n \log \theta + (\theta - 1) \sum_{i=1}^{n} \log x_i\end{aligned}$$

Now, we maximize this

$$\begin{aligned}\frac{d\ell(\theta)}{d\theta} &\overset{\text{set}}{=} 0 \\ \implies \frac{n}{\theta} + \sum_{i=1}^{n} \log x_i &= 0 \\ \implies \widehat{\theta}_{\text{MLE}} &= -\frac{n}{\sum_{i=1}^{n} \log x_i}\end{aligned}$$

The invariance property says that if $\widehat{\theta}$ is the MLE of θ, then the MLE of $\tau(\theta)$, where τ is some function, is $\tau(\widehat{\theta})$. Therefore, the MLE of $\frac{1}{\theta}$,

$$\frac{1}{\widehat{\theta}_{\text{MLE}}} = -\frac{\sum_{i=1}^{n} \log x_i}{n}$$

Next, we need to find $\mathbb{E}\left[\frac{1}{\widehat{\theta}_{\text{MLE}}}\right]$ and compare it with $\frac{1}{\theta}$.

To do this, we will start by finding $\mathbb{E}[\log X_1]$. There are few ways to do this. One way is to do a direct calculation of $\int_{x=0}^{1} \log x \cdot \theta x^{\theta-1} dx$. We will do it by finding the distribution of $-\log X_i$ as that gives us a useful result. Let $Y_i = -\log X_i = g(X_i)$. Then

$$\begin{aligned}p_Y(y) &= p_X(g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right| = p_X(e^{-y}) \left| \frac{d}{dy} e^{-y} \right| \\ &= \theta e^{-y(\theta-1)} e^{-y} \\ &= \theta e^{-\theta y}\end{aligned}$$

So $Y_i \sim \text{Exp}(\theta)$. Therefore, $\mathbb{E}[Y_i] = 1/\theta$. This gives,

$$\mathbb{E}\left[\frac{1}{\widehat{\theta}_{\text{MLE}}}\right] = \mathbb{E}\left[-\frac{\sum_{i=1}^{n} \log X_i}{n}\right]$$

$$= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[Y_i]$$
$$= \frac{1}{\theta}$$

So the MLE of $1/\theta$ is unbiased. □

Problem 7. *Let F denote the true distribution (CDF) of a random variable. Let $X_1, \ldots, X_n \overset{i.i.d.}{\sim} F$. Denote the empirical CDF by $\widehat{F}_n$, which places an equal point mass on each X_i. Suppose we want to estimate $\theta := T(F)$. We denote the plug-in estimate of θ as $\widehat{\theta}_n := T(\widehat{F}_n)$.*

Find the plug-in estimate for the following quantities:

(a) Mean, $\theta = \mathbb{E}[X]$

(b) Variance, $\theta = Var(X)$

(c) Skewness, $\theta = \frac{\mathbb{E}[X-\mu]^3}{\sigma^3}$

(d) The p-th quantile.

Solution. The strategy to find plug-in estimates is to first find the function $\theta = T(F)$. Once we find T, the plug-in estimate is as simple as replacing F with $\widehat{F}_n$ in T.

(a) Here the mean $\mathbb{E}[X] = \int x dF(x)$. Therefore, $T(F) = \int x dF(x)$. This gives us the following plug-in estimate

$$\begin{aligned}\widehat{\theta}_n &= T(\widehat{F}_n)\\ &= \int x d\widehat{F}_n(x)\\ &= \frac{1}{n}\sum_{i=1}^{n} X_i\end{aligned}$$

(b) We can write the variance as

$$\theta = \text{Var}(X) = \int x^2 dF(x) - \left[\int x dF(x)\right]^2$$

Therefore, $T(F) = \int x^2 dF(x) - \left[\int x dF(x)\right]^2$. This gives our plug-in estimate,

$$\begin{aligned}\widehat{\theta}_n &= T(\widehat{F}_n)\\ &= \int x^2 d\widehat{F}_n(x) - \left[\int x d\widehat{F}_n(x)\right]^2\\ &= \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \left(\overline{X}_n\right)^2\\ &= \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2\end{aligned}$$

Notice that this is different from the sample variance where instead of $1/n$, we have $1/(n-1)$.

(c) Once again we write θ as a function of the CDF,

$$\begin{aligned}\theta &= \frac{\mathbb{E}[X-\mu]^3}{\sigma^3} \\ &= \frac{\int (x - \int t dF(t))^3 dF(x)}{\left\{\int (x - \int t dF(t))^2 dF(x)\right\}^{3/2}} \\ &= T(F)\end{aligned}$$

So, the plug-in estimate is

$$\begin{aligned}\widehat{\theta}_n &= T(\widehat{F}_n) \\ &= \frac{\int (x - \int t d\widehat{F}_n(t))^3 d\widehat{F}_n(x)}{\left\{\int (x - \int t d\widehat{F}_n(t))^2 d\widehat{F}_n(x)\right\}^{3/2}}\end{aligned}$$

From previous problems, $\int x d\widehat{F}_n(x) = \overline{X}_n$, and $\widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \overline{X}_n)^2$ is the plug-in estimate for variance. Therefore,

$$\begin{aligned}\widehat{\theta}_n &= \frac{\int (x - \overline{X}_n)^3 d\widehat{F}_n(x)}{\widehat{\sigma}^{3/2}} \\ &= \frac{\sum_{i=1}^n (X_i - \overline{X}_n)^3}{n\widehat{\sigma}^{3/2}}\end{aligned}$$

(d) By definition, the p-th quantile is $\theta = F^{-1}(p)$. Therefore the plug-in estimate is $\widehat{\theta}_n = \widehat{F}_n^{-1}(p)$. □

Problem 8. *Let X, Y be two random variables. Define the correlation as $\rho = \frac{\mathbb{E}[(X-\mu_X)(Y-\mu_Y)]}{\sigma_X \sigma_Y}$. Define the joint distribution of (X, Y) as F. Find the plug-in estimate of ρ.*

Solution. First, let us find the mean and variance of the two random variables as a function of the CDF F:

$$\begin{aligned}\mu_X &= \mathbb{E}[X] = \int x dF(x,y) \\ &= T_1(F) \\ \mu_Y &= \mathbb{E}[Y] = \int y dF(x,y) \\ &= T_2(F) \\ \sigma_X^2 &= \text{Var}(X) = \int x^2 dF(x,y) - [\mu_X]^2 \\ &= T_3(F) - [T_1(F)]^2 \\ \sigma_Y^2 &= \text{Var}(X) = \int y^2 dF(x,y) - [\mu_Y]^2 \\ &= T_4(F) - [T_2(F)]^2\end{aligned}$$

Now, let us move to then numerator:

$$\begin{aligned}\mathbb{E}[(X-\mu_X)(Y-\mu_Y)] &= \int (x - T_1(F))(y - T_2(F)) dF(x,y) \\ &= \underbrace{\int xy dF(x,y)}_{=T_5(F)} - T_2(F)\int x dF(x,y) - T_1(F)\int y dF(x,y) + T_1(F)T_2(F)\end{aligned}$$

$$= T_5(F) - T_1(F)T_2(F)$$

Therefore,

$$\rho = \frac{T_5(F) - T_1(F)T_2(F)}{\Big[(T_3(F) - [T_1(F)]^2)(T_4(F) - [T_2(F)]^2)\Big]^{1/2}}$$

The plug-in estimate is

$$\widehat{\rho} = \frac{T_5(\widehat{F}_n) - T_1(\widehat{F}_n)T_2(\widehat{F}_n)}{\left[(T_3(\widehat{F}_n) - \left[T_1(\widehat{F}_n)\right]^2)(T_4(\widehat{F}_n) - \left[T_2(\widehat{F}_n)\right]^2)\right]^{1/2}}$$

From Problem 1,

$$\widehat{\rho} = \frac{T_5(\widehat{F}_n) - \overline{X}_n\overline{Y}_n}{\left[\frac{1}{n}\sum_{i=1}^n (X_i - \overline{X}_n)^2 \cdot \frac{1}{n}\sum_{i=1}^n (Y_i - \overline{Y}_n)^2\right]^{1/2}}$$

Finally, $T_5(\widehat{F}_n) = \frac{1}{n}\sum_{i=1}^n X_iY_i$. This gives,

$$\begin{aligned}\widehat{\rho} &= \frac{\frac{1}{n}\sum_{i=1}^n X_iY_i - \overline{X}_n\overline{Y}_n}{\left[\frac{1}{n}\sum_{i=1}^n (X_i - \overline{X}_n)^2 \cdot \frac{1}{n}\sum_{i=1}^n (Y_i - \overline{Y}_n)^2\right]^{1/2}} \\ &= \frac{\sum_{i=1}^n (X_i - \overline{X}_n)(Y_i - \overline{Y}_n)}{\left[\sum_{i=1}^n (X_i - \overline{X}_n)^2 \cdot \sum_{i=1}^n (Y_i - \overline{Y}_n)^2\right]^{1/2}}\end{aligned}$$

where showing the last equality needs some algebra. □