Problems and suggested solution

Question 1

(a) [**4 Points**] Consider a random variable X with distribution function F given by

$$F(x) = \begin{cases} 0 & ; \quad x < 0 \\ 0.5 & ; \quad 0 \le x < 3 \\ 1 - \frac{1}{x^2}; & \quad x \ge 3. \end{cases}$$

Compute $\mathrm{VaR}_\alpha(X)$, $\mathrm{AVaR}_\alpha(X)$ and $\mathrm{ES}_\alpha(X)$ at level $\alpha = 0.8$.

Solution:

It holds that

$$\mathrm{VaR}_u(X) = \begin{cases} 0, & u \in (0, 0.5] \\ 3, & u \in (0.5, 8/9] \\ \frac{1}{\sqrt{1-u}}, & u \in (8/9, 1). \end{cases}$$

Hence,

$$\mathrm{VaR}_{0.8}(X) = 3.$$

$$\begin{aligned} \mathrm{AVaR}_{0.8}(X) &= \frac{1}{1-0.8} \int_{0.8}^{1} \mathrm{VaR}_u(X) \, \mathrm{d}u \\ &= 5\Big[3(8/9 - 0.8) + \int_{8/9}^{1} \frac{1}{\sqrt{1-u}} \, \mathrm{d}u\Big] \\ &= 4/3 + 5 \int_0^{1/9} \frac{1}{\sqrt{u}} \, \mathrm{d}u \\ &= 4/3 + 5\Big[2\sqrt{u}\Big]_0^{1/9} \\ &= 4/3 + 5 \cdot 2/3 \\ &= 14/3. \end{aligned}$$

Finally,

$$\begin{aligned} \mathrm{ES}_{0.8}(X) &= \mathbb{E}[X \mid X \ge \mathrm{VaR}_{0.8}(X)] \\ &= \frac{1}{1-0.5} \int_{0.5}^{1} \mathrm{VaR}_u(X) \, \mathrm{d}u \\ &= 2\Big[3(8/9 - 0.5) + \int_{8/9}^{1} \frac{1}{\sqrt{1-u}} \, \mathrm{d}u\Big] \\ &= 7/3 + 2 \cdot 2/3 \\ &= 11/3. \end{aligned}$$

(b) For a random variable L with distribution function F_L, we can define

$$\mathrm{VaR}_1(L) := \sup\{x \in \mathbb{R} : F_L(x) < 1\} \in \mathbb{R} \cup \{\infty\}.$$

That means $\mathrm{VaR}_1(L)$ corresponds to the right endpoint of F_L.

In this exercise, we are going to show that VaR_1 is a coherent risk measure on the space $\mathcal{L}$ of essentially bounded random variables, that is, $\mathcal{L} = \{L \mid \mathrm{VaR}_1(L) < \infty\}$.

(i) [**2 Points**] First show that for any random variable $L \in \mathcal{L}$ it holds that

$$\mathrm{VaR}_1(L) = \sup_{\alpha \in (0,1)} \mathrm{AVaR}_\alpha(L). \tag{1}$$

Hint: You may use without proof that for any $L \in \mathcal{L}$, the quantile function $(0,1] \ni \alpha \to \mathrm{VaR}_\alpha(L)$ is left-continuous and increasing.

Solution:

For any $\alpha \in (0,1)$ and any random variable $L \in \mathcal{L}$ it holds that

$$\mathrm{VaR}_\alpha(L) \leq \mathrm{AVaR}_\alpha(L) = \frac{1}{1-\alpha}\int_\alpha^1 \mathrm{VaR}_u(L)\,\mathrm{d}u \leq \mathrm{VaR}_1(L).$$

Due to the hint, it holds that $\mathrm{VaR}_\alpha(L) \to \mathrm{VaR}_1(L)$ as $\alpha \to 1$. Hence, due to the above inequality, also $\mathrm{AVaR}_\alpha(L) \to \mathrm{VaR}_1(L)$. Since $\mathrm{AVaR}_\alpha(L)$ is increasing in α, (1) holds.

(ii) [**4 Points**] Use the representation (1) to prove that VaR_1 is a coherent risk measure on $\mathcal{L}$.
Hint: You may use any properties of AVaR_α established in the lecture.

Solution:

Since VaR_1 is a supremum of coherent risk measures, it is itself coherent.

Indeed, let's consider all properties, assuming that $L, L_1, L_2 \in \mathcal{L}$:

Monotonicity: Let $L_1 \leq L_2$ almost surely. Then $\mathrm{AVaR}_\alpha(L_1) \leq \mathrm{AVaR}_\alpha(L_2)$. Hence, with (1)

$$\mathrm{VaR}_1(L_1) \leq \mathrm{VaR}_1(L_2)$$

Translation property: For a random variable L and $m \in \mathbb{R}$ it holds that $\mathrm{AVaR}_\alpha(L+m) = \mathrm{AVaR}_\alpha(L) + m$. Hence,

$$\mathrm{VaR}_1(L+m) = \sup_{\alpha \in (0,1)} \mathrm{AVaR}_\alpha(L+m) = \sup_{\alpha \in (0,1)} \mathrm{AVaR}_\alpha(L) + m = \mathrm{VaR}_1(L) + m.$$

Subadditivity: Let L_1, L_2 be random variables. Since AVaR_α is subadditive, it holds that

$$\begin{aligned}
\mathrm{VaR}_1(L_1 + L_2) &= \sup_{\alpha \in (0,1)} \mathrm{AVaR}_\alpha(L_1 + L_2) \\
&\leq \sup_{\alpha \in (0,1)} \mathrm{AVaR}_\alpha(L_1) + \mathrm{AVaR}_\alpha(L_2) \\
&\leq \sup_{\alpha \in (0,1)} \mathrm{AVaR}_\alpha(L_1) + \sup_{\alpha \in (0,1)} \mathrm{AVaR}_\alpha(L_2) \\
&= \mathrm{VaR}_1(L_1) + \mathrm{VaR}_1(L_2)
\end{aligned}$$

Positive homogeneity: Let L be a random variable and $\lambda > 0$. Since AVaR_α is positive homogeneous, it holds that

$$\mathrm{VaR}_1(\lambda L) = \sup_{\alpha \in (0,1)} \mathrm{AVaR}_\alpha(\lambda L) = \sup_{\alpha \in (0,1)} \lambda \mathrm{AVaR}_\alpha(L) = \lambda \mathrm{VaR}_1(L).$$

(iii) **[1 Point]** Briefly discuss the adequacy of VaR_1 as a risk measure in quantitative risk management and regulation.

Solution:

- (The coherence is a desirable property.)
- However, in QRM, most loss distributions of interest are heavy tailed. In particular, that means that VaR_1 is infinite for those distributions. Hence, this risk measure is not very informative, since it cannot not distinguish well between different distributions.
- The above point can also be formulated differently: The space $\mathcal{L}$ is too restrictive to be considered relevant in QRM.
- (Moreover, even for bounded random variables, VaR_1 is very conservative.)

Question 2

(a) (i) **[1 Point]** Provide the definition of an ARCH(p) process.

(ii) **[2 Points]** Is an ARCH(p) process a white noise process? Justify your answer.

(iii) **[1 Point]** Is an ARCH(p) process a *strict* white noise process? Justify your answer.

(iv) **[1 Point]** Discuss the adequacy of a strict white noise process to model financial log-returns, referring to the relevant stylized facts of financial log-returns.

Solution:

(i) $(X_t)_{t\in\mathbb{Z}}$ is an ARCH(p) process if it satisfies

$$X_t = \sigma_t Z_t$$
$$\sigma_t^2 = \alpha_0 + \sum_{k=1}^{p} \alpha_k X_{t-k}^2,$$

where $(Z_t)_{t\in\mathbb{Z}}$ is a strict white noise process with mean 0, variance 1, Z_t is independent from $\mathcal{F}_{t-1} = \sigma(X_{t-1}, X_{t-2}, \ldots)$ and $\alpha_0 > 0$, $\alpha_k \geq 0$, $k = 1, \ldots, p$.

Moreover, to make it stationary, we need to impose that

$$\sum_{k=1}^{p} \alpha_k < 1.$$

(ii) We need to check two things: (1) $\mathbb{E}[X_t] = 0$, and (2) $\mathrm{Cov}(X_t, X_{t+h}) = 0$ for all $t \in \mathbb{Z}$, $h \neq 0$.

Define $\mathcal{F}_{t-1} = \sigma(X_{t-1}, X_{t-2}, \ldots)$. Then

$$\mathbb{E}[X_t] = \mathbb{E}[\mathbb{E}[X_t|\mathcal{F}_{t-1}]] = \mathbb{E}[\sigma_t \underbrace{\mathbb{E}[Z_t|\mathcal{F}_{t-1}]}_{=0}] = 0$$

For the covariance, suppose without loss of generality that $h > 0$. Then

$$\begin{aligned}\mathrm{Cov}(X_t, X_{t+h}) &= \mathbb{E}[X_t X_{t+h}] = \mathbb{E}[\mathbb{E}[\sigma_t Z_t \sigma_{t+h} Z_{t+h}|\mathcal{F}_{t+h-1}]] \\ &= \mathbb{E}[\sigma_t Z_t \sigma_{t+h} \underbrace{\mathbb{E}[Z_{t+h}|\mathcal{F}_{t+h-1}]}_{=0}] = 0\end{aligned}$$

(iii) For a strict white noise process, $(X_t)_{t\in\mathbb{Z}}$ would need to be iid. However, it holds that

$$\mathbb{E}[X_t^2|\mathcal{F}_{t-1}] = \sigma_t^2 = \alpha_0 + \sum_{k=1}^{p} \alpha_k X_{t-k}^2.$$

That means X_t^2 depends on $\mathcal{F}_{t-1}$ if $\alpha_1 > 0$ or $\ldots$ or $\alpha_p > 0$. So in general, an ARCH(p) process is not a strict white noise process.

On the other hand, for $\alpha_1 = \cdots = \alpha_p = 0$ it trivially becomes strictly stationary since then $X_t = \alpha_0^{1/2} Z_t$.

(iv) One of the most important stylized facts of financial log-returns are their volatility clusters and – more generally – the time varying volatility. This cannot be replicated by a strict white noise process. Hence, such a process cannot be recommended to model financial log-returns.

More generally, a strict white noise process cannot satisfy properties (U1), (U2), (U4) and (U5).

(b) Suppose you have a portfolio with d assets $X_1, \dots, X_d$, which all have variance 1.

(i) [**1 Point**] What does it mean that the vector $(X_1, \dots, X_d)$ has an exchangeable distribution?
Solution:

It means that for any permutation $\pi\colon \{1, \dots, d\} \to \{1, \dots, d\}$ the distribution of $(X_1, \dots, X_d)$ coincides with the distribution of $(X_{\pi(1)}, \dots, X_{\pi(d)})$.

(ii) [**1 Point**] Show that exchangeability implies that all pairs (X_i, X_j), $1 \le i < j \le d$, have the same Pearson correlation $\rho \in [-1, 1]$.
Solution:

Let $\rho := \mathrm{cov}(X_1, X_2)$. The exchangeability implies that for any permutation $\pi\colon \{1, \dots, d\} \to \{1, \dots, d\}$ we have

$$(X_1, X_2) \overset{(d)}{=} (X_{\pi(1)}, X_{\pi(2)}).$$

Therefore,

$$\rho = \mathrm{corr}(X_{\pi(1)}, X_{\pi(2)}).$$

Moreover, for any $i < j$ there is a permutation π such that $(\pi(1), \pi(2)) = (i, j)$.

(iii) [**2 Points**] Show that necessarily $\rho \ge -1/(d-1)$.
Solution:

Since the variances are 1, the correlation matrix coincides with the covariance matrix. .
We can calculate the variance of $X_1 + \cdots + X_d$ and use that the variance is non-negative.

$$0 \le \mathrm{Var}(X_1 + X_2 + \cdots + X_d) = \sum_{i,j=1}^{d} \mathrm{Cov}(X_i, X_j) = \sum_{i,j=1}^{d} \mathrm{corr}(X_i, X_j) = d(1 + \rho(d-1))$$

This is equivalent to

$$\rho \ge -\frac{1}{d-1}$$

(iv) [**3 Points**] Show that for $\rho = -1/(d-1)$, the distribution of $(X_1, \dots, X_d)$ cannot have a joint density.
Hint: Start by calculating the variance $\mathrm{Var}(X_1 + X_2 + \cdots + X_d)$.
Solution:

If $\rho = -1/(d-1)$, then

$$\mathrm{Var}(X_1 + X_2 + \cdots + X_d) = \sum_{i,j=1}^{d} \mathrm{Cov}(X_i, X_j) = d(1 + \rho(d-1)) = 0.$$

Hence, $X_1 + X_2 + \cdots + X_d$ must be constant almost surely. Therefore, $(X_1, \ldots, X_d)$ only attains values on a $(d-1)$-dimensional sub-vectorspace of $\mathbb{R}^d$.

Since any such subspace has d-dimensional Lebesgue measure 0, the distribution cannot have a Lebesgue density.

Question 3

(a) **[3 Points]** Let $X_1, X_2, \dots$ be iid random variables with distribution F. Let $M_n = \max\{X_1, \dots, X_n\}$ and suppose that $x_F = \sup\{x \in \mathbb{R} : F(x) < 1\} < \infty$.

Show that M_n converges in distribution to x_F.

Solution:

We need to show that the distribution function F_{M_n} of M_n converges to the step function

$$G(x) = \begin{cases} 0, & x < x_F \\ 1, & x \geq x_F \end{cases}$$

for all its continuity points of G, that is, for all $x \in \mathbb{R}$, $x \neq x_F$.

For any $x \in \mathbb{R}$ it holds that

$$F_{M_n}(x) = \mathbb{P}[X_1 \leq x \text{ and } X_2 \leq x, \dots, \text{ and } X_n \leq x] = \Big(F(x)\Big)^n.$$

If $x < x_F$, then $F(x) \in [0, 1)$. Therefore, it holds that

$$\lim_{n\to\infty} \Big(F(x)\Big)^n = 0.$$

On the other hand for $x > x_F$ it holds that $F(x) = 1$. Hence,

$$\lim_{n\to\infty} \Big(F(x)\Big)^n = 1.$$

(b) Let F be the distribution function of a uniform distribution on $[0, 1]$.

(i) **[2 Points]** Does F belong to the maximum domain of attraction, $\mathrm{MDA}(H_\xi)$, of a standard GEV distribution H_ξ? If yes, determine the parameter ξ.

Solution:

We may use a characterisation result from the lecture for $\xi < 0$: $F \in \mathrm{MDA}(H_\xi)$ if and only if $x_F < \infty$ and $1 - F(x_F - 1/x) = x^{1/\xi} L(x)$ where L is a slowly varying function.

It holds that $x_F = 1 < \infty$ and

$$1 - F(x_F - 1/x) = 1 - (1 - 1/x) = x^{-1}.$$

Since any constant is a slowly varying function, it holds that $F \in \mathrm{MDA}(H_\xi)$ for $\xi = -1$.

(ii) **[1 Point]** Calculate the excess distribution function $F_u(x) = \mathbb{P}[X - u \leq x \mid X > u]$, $0 \leq u < x_F$, $x \in [0, x_F - u)$.

Solution:

For $0 \le x \le x_F$, $u < x_F$ we get

$$\mathbb{P}[X - u \le x \mid X > u] = \frac{\mathbb{P}[u < X \le x + u]}{\mathbb{P}[X > u]} = \frac{x}{1-u}.$$

(iii) **[2 Points]** Does there exist a parameter $\xi \in \mathbb{R}$ and a function $\beta \colon \mathbb{R} \to (0, \infty)$ such that

$$\lim_{u \uparrow x_F} \sup_{x > 0} |F_u(x) - G_{\xi, \beta(u)}(x)| = 0,$$

where $G_{\xi,\beta}$ denotes the cumulative distribution function of a GPD? If yes, for which ξ and β does this hold?

Solution:

Pickands–Balkema–de Haan Theorem implies that there exists a measurable function $\beta \colon \mathbb{R} \to (0, \infty)$ such that

$$\lim_{u \uparrow x_F} \sup_{x > 0} |F_u(x) - G_{\xi, \beta(u)}(x)| = 0, \tag{2}$$

if and only if $F \in \mathrm{MDA}(H_\xi)$.

We have shown in (c) that $F \in \mathrm{MDA}(H_{-1})$, thus (1) holds for $\xi = -1$ and for some function $\beta(u)$.

Moreover, $G_{-1,\beta}(x) = x/\beta$. Hence, the assertion holds for $\beta(u) = 1 - u$.

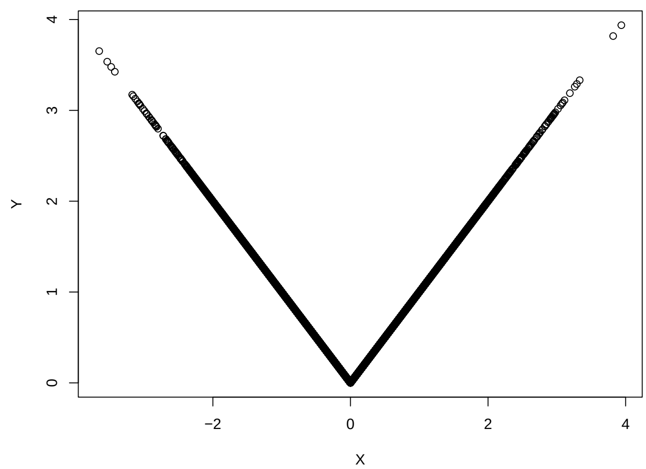
Question 4

(a) Consider the random vector $(Z, |Z|)$, where Z follows a standard normal distribution.

(i) [**2 Points**] Sketch a scatterplot of a random sample from the distribution of $(Z, |Z|)$ and also a scatterplot of a random sample from the underlying copula of $(Z, |Z|)$.

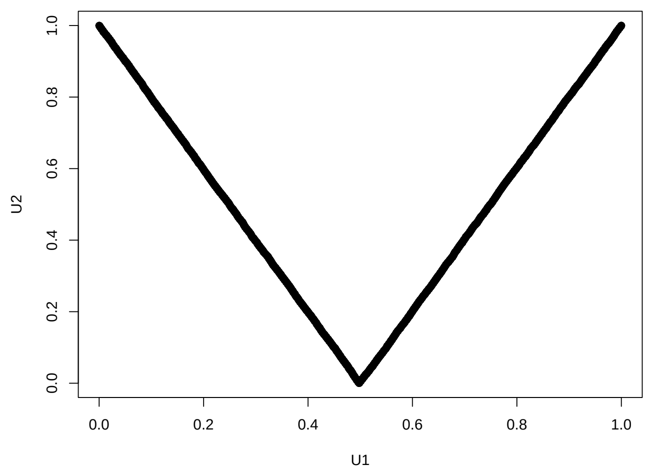
Solution:

A scatterplot of a random sample from $(Z, |Z|)$ has the following shape:



(The qualitative V-shape of the graph is important here! Of course, it needs to be symmetric around the line $x = 0$ and the slopes need to be ± 1.)

A scatterplot of a random sample from the copula of $(Z, |Z|)$ has the following form:



(It is important that the qualitative V-shape does not change, but the symmetry axis must be correct ay $u_1 = 0.5$. Moreover, the slopes need to be ± 2.)

(ii) **[2 Points]** Compute the Pearson correlation of Z and $|Z|$. Are Z and $|Z|$ independent? Provide an argument for your claim.

Solution:

Since $Z \stackrel{(d)}{=} -Z$ it holds that $(Z, |Z|) \stackrel{(d)}{=} (-Z, |-Z|)$. Hence,

$$\mathrm{Cov}(Z, |Z|) = \mathrm{Cov}(-Z, |-Z|) = \mathrm{Cov}(-Z, |Z|) = -\mathrm{Cov}(Z, |Z|).$$

Hence, $\mathrm{Cov}(Z, |Z|) = 0$.

Z and $|Z|$ are not independent, even though they are uncorrelated. One possible argument is that the copula of $(Z, |Z|)$ does not correspond to the independence copula, as can be seen from (a).

(iii) **[1 Point]** How would the scatterplot of a random sample from the copula change if we considered (Z, Z^2)? Provide a brief argument for your claim.

Solution:

The scatterplot would not change. The reason is that $(Z, |Z|)$ and (Z, Z^2) have the same copula. Indeed, we only apply a strictly increasing transformation to the second component. (The map $x \mapsto x^2$ is strictly increasing on $[0, \infty)$.) And copulas are invariant under strictly increasing transformations of each component.

(b) Suppose you have iid observations $(X_1, Y_1), \ldots, (X_n, Y_n)$ of a two-dimensional random vector (X, Y).

(i) **[1 Point]** Provide a formula for the sample version of Kendall's tau, $r_\tau(n)$.

Solution:

$$\begin{aligned} r_\tau(n) &= \binom{n}{2}^{-1} \sum_{1 \le i < j \le n} \operatorname{sign}\big((X_i - X_j)(Y_i - Y_j)\big) \\ &= \frac{1}{n(n-1)} \sum_{1 \le i \neq j \le n} \operatorname{sign}\big((X_i - X_j)(Y_i - Y_j)\big) \end{aligned}$$

(ii) **[1 Point]** Suppose that $n = 4$ and your random sample is

$$\{(-1, 3), (0, 0), (2, 2), (3, 1)\}. \tag{3}$$

Compute $r_\tau(n)$ on this sample.

Solution:

$$\begin{aligned} r_\tau(n) &= \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1, j \neq i}^{n} \operatorname{sign}\big((X_i - X_j)(Y_i - Y_j)\big) \\ &= \frac{1}{4 \cdot 3}(-3 + 1 - 1 - 1) \\ &= -\frac{1}{3} \end{aligned}$$

(iii) **[2 Points]** Suppose you have a new data point (X_{n+1}, Y_{n+1}). Describe a condition under which this new data point influences the sample version of Kendall's tau maximally. Provide an appropriate example for the concrete sample in (3).

Solution:

The new data point (X_{n+1}, Y_{n+1}) must be either *concordant* with all data points $(X_1, Y_1), \dots, (X_n, Y_n)$ or *discordant* with all data points $(X_1, Y_1), \dots, (X_n, Y_n)$.

Equivalently, for an equivalent characterisation of concordance with all data points, $(X_1, Y_1), \dots, (X_n, Y_n)$, consider the componentwise order-statistics $X_{(1)} \le \dots \le X_{(n)}$ and $Y_{(1)} \le \dots \le Y_{(n)}$. Then simultaneous concordance can be expressed as

$$\Big(X_{n+1} > X_{(n)} \text{ and } Y_{n+1} > Y_{(n)}\Big) \quad \text{or} \quad \Big(X_{n+1} < X_{(1)} \text{ and } Y_{n+1} < Y_{(1)}\Big).$$

On the other hand, simultaneous discordance can be expressed as

$$\Big(X_{n+1} < X_{(1)} \text{ and } Y_{n+1} > Y_{(n)}\Big) \quad \text{or} \quad \Big(X_{n+1} > X_{(n)} \text{ and } Y_{n+1} < Y_{(1)}\Big).$$

Possible appropriate examples are: For concordance $(5, 5)$ or $(-2, -2)$. For discordance $(-2, 5)$ or $(5, -2)$.

(iv) **[1 Point]** What is the maximal influence of a new data point (X_{n+1}, Y_{n+1}) on the sample version of Kendall's tau? That is, what is an upper bound for the difference

$$|r_\tau(n) - r_\tau(n+1)|,$$

which holds for any data set $(X_1, Y_1), \dots, (X_n, Y_n), (X_{n+1}, Y_{n+1})$?

Decide whether it is either

$$2 \quad \text{or} \quad \frac{4}{n+1} \quad \text{or} \quad \frac{4}{(n+1)^2}.$$

Write down the solution on your submission sheet (answers on this sheet will not be counted). You do not need to provide an explanation.

Solution:

$$\frac{4}{n+1}$$

The derivation is as follows (not necessary to get the point!): We can write

$$r_\tau(n) = \frac{1}{\binom{n}{2}} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \operatorname{sign}\Big((X_i - X_j)(Y_i - Y_j)\Big).$$

Hence, we have

$$\begin{aligned}
r_\tau(n+1) &= \frac{1}{\binom{n+1}{2}}\Big[\binom{n}{2} r_\tau(n) + \sum_{i=1}^n \operatorname{sign}\Big((X_i - X_{n+1})(Y_i - Y_{n+1})\Big)\Big] \\
&= \frac{2}{(n+1)n}\Big[\frac{n(n-1)}{n} r_\tau(n) + \sum_{i=1}^n \operatorname{sign}\Big((X_i - X_{n+1})(Y_i - Y_{n+1})\Big)\Big] \\
&= \frac{n-1}{n+1} r_\tau(n) + \frac{2}{(n+1)n}\sum_{i=1}^n \operatorname{sign}\Big((X_i - X_{n+1})(Y_i - Y_{n+1})\Big).
\end{aligned}$$

This yields

$$\begin{aligned}
|r_\tau(n+1) - r_\tau(n)| &= \Big| r_\tau(n)\left[\frac{n-1}{n+1} - 1\right] + \frac{2}{(n+1)n}\sum_{i=1}^n \operatorname{sign}\Big((X_i - X_{n+1})(Y_i - Y_{n+1})\Big)\Big| \\
&= \Big| -\frac{2}{n+1} r_\tau(n) + \frac{2}{(n+1)n}\sum_{i=1}^n \operatorname{sign}\Big((X_i - X_{n+1})(Y_i - Y_{n+1})\Big)\Big| \\
&= \frac{2}{n+1}\underbrace{\Big|\frac{1}{n}\sum_{i=1}^n \operatorname{sign}\Big((X_i - X_{n+1})(Y_i - Y_{n+1})\Big) - r_\tau(\tau)\Big|}_{\le 2} \\
&\le \frac{4}{n+1}.
\end{aligned}$$

(v) [**1 Point**] Compare and possibly contrast your previous result to Pearson's correlation coefficient.

Solution:

In contrast to Kendall's tau, a new data point for Pearson's correlation coefficient can change the empirical correlation coefficient almost arbitrarily. That is, no matter what the correlation on the existing sample $(X_1, Y_1), \ldots, (X_n, Y_n)$ is, the correlation on $(X_1, Y_1), \ldots, (X_n, Y_n), (X_{n+1}, Y_{n+1})$ can be any number in the interval $(-1, 1)$.
(Hence, the correct answer for Pearson's correlation coefficient is 2.)

Question 5

In this exercise, any scoring function $S(x, y)$, $x, y \in \mathbb{R}$, has the interpretation that $x \in \mathbb{R}$ is a forecast and $y \in \mathbb{R}$ an observation.

(a) **[1 Point]** In a comparative backtest you consider $\mathrm{VaR}_{0.95}$-forecasts from two different models, M1 and M2. You compute average scores on a validation window for the following scoring functions:

$$S_1(x, y) = |x - y|$$
$$S_2(x, y) = (x - y)^2$$
$$S_3(x, y) = (1_{\{y \le x\}} - 0.95)(x - y)$$

The results are summarized in the following table.

	M1	M2
S_1	3.88	3.23
S_2	7.49	6.91
S_3	2.54	3.02

Which model would you prefer and why?

Solution:

The model M1 is preferable. We should evaluate forecasts using strictly consistent scoring functions. Only S_3 is strictly consistent for $\mathrm{VaR}_{0.95}$. And a smaller average score indicates a better forecast performance.

(b) **[1 Point]** Why is it important to evaluate forecasts for a distribution-based risk measure using a strictly consistent scoring function for this risk measure?

Solution:

This is important because strictly consistent scores incentivize truthful forecasting. That is, a correctly specified forecast achieves a smaller expected score than any other forecast.

(c) **[2 Points]** Show that for any strictly convex function $\phi\colon \mathbb{R} \to \mathbb{R}$ with strictly positive second derivative, the score

$$S(x, y) = -\phi(x) + \phi'(x)(x - y), \quad x, y \in \mathbb{R},$$

is strictly consistent for the mean functional on the class of distributions with a finite mean.

Solution:

For any distribution F with finite mean and any forecast $x \in \mathbb{R}$, we can compute the expected score

$$\bar{S}(x, F) := \mathbb{E}_{Y\sim F}[S(x, Y)] = -\phi(x) + \phi'(x)(x - \mathbb{E}_{Y\sim F}[Y]).$$

We need to show that $\bar{S}(x, F)$ is strictly minimized for $x = \mathbb{E}_{Y\sim F}[Y]$.

The first order condition is

$$\frac{\partial}{\partial x}\bar{S}(x, F) = \phi''(x)(x - \mathbb{E}_{Y\sim F}[Y]) = 0.$$

Since $\phi''(x) > 0$ this condition holds if and only if $x = \mathbb{E}_{Y\sim F}[Y]$.

Moreover, since $\phi''(x) > 0$, we have

$$\frac{\partial}{\partial x}\bar{S}(x,F) \begin{cases} < 0, & x < \mathbb{E}_{Y\sim F}[Y] \\ > 0, & x > \mathbb{E}_{Y\sim F}[Y]. \end{cases}$$

This yields that $\bar{S}(x,F)$ has a strict global minimum at $x = \mathbb{E}_{Y\sim F}[Y]$.

(d) **[4 Points]** Consider an AR(1) process $(Y_t)_{t\in\mathbb{N}}$ of the form

$$Y_0 = 0 \quad \text{and} \quad Y_t = \theta Y_{t-1} + u_t, \quad t \geq 1,$$

where $|\theta| < 1$ and $\mathbb{E}[u_t \mid Y_{t-1}] = 0$.

Two competing forecasts are available, $X_t = 0$ and $X_t^* = \theta Y_{t-1}$, where $t \geq 1$.

If $S\colon \mathbb{R}\times\mathbb{R} \to \mathbb{R}$ is a strictly consistent scoring function for the mean functional, is it possible to establish an inequality between the expected scores $\mathbb{E}[S(X_t, Y_t)]$ and $\mathbb{E}[S(X_t^*, Y_t)]$? Explain your answer.

Hint: Start by computing $\mathbb{E}[Y_t]$ and $\mathbb{E}[Y_t \mid Y_{t-1}]$.

Solution:

It holds that

$$\begin{aligned} X_t &= \mathbb{E}[Y_t] \\ X_t^* &= \mathbb{E}[Y_t \mid Y_{t-1}] \end{aligned}$$

That means both forecasts are calibrated mean forecasts. However, strictly consistent scoring functions prefer the more informative forecast, which is X_t^*.

Hence,

$$\mathbb{E}[S(X_t^*, Y_t)] \leq \mathbb{E}[S(X_t, Y_t)].$$