Lecture 9. Dependence measures

Enkelejd Hashorva & Pavel Ievlev
Université de Lausanne

17 November, 2025

Leargning objectives

- Introduce the popular dependence measures: Pearson's correlation, Kendall's tau and Spearman's rho and discuss their properties
- Introduce the correlation order
- Focus on the dependence measures for Gaussian/elliptical/Archimedean copulas

Pearson's linear correlation

Definition 1 (Pearson's correlation).

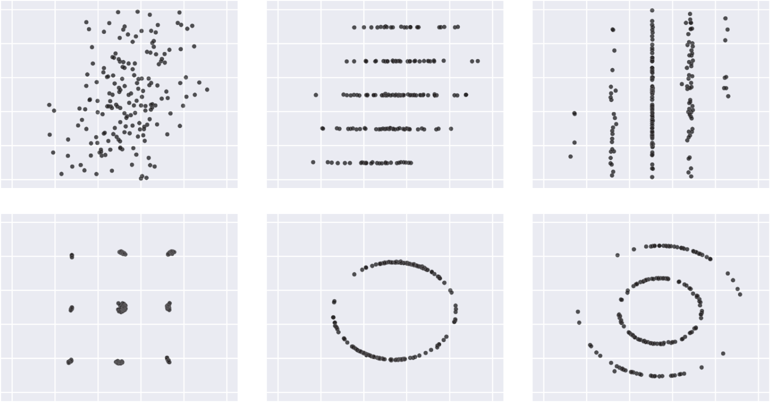
Let X, Y be two random variables with finite and positive variances. The **Pearson's correlation** between X and Y is defined as

$$\rho(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)}\sqrt{\mathrm{Var}(Y)}}.$$

Properties:

- $\rho(aX+b, cY+d) = \mathrm{sign}(ac)\,\rho(X,Y)$
- $\rho(X,Y) \in [-1,1]$.
- $\rho(X,Y) = 1$ (resp. -1) if and only if there exists a comonnotonic (resp. countermonotonic) linear relationship $Y = aX + b$ with $a > 0$ (resp. $a < 0$).
- Pearson's correlation is **not** invariant under strictly increasing transformations of X and Y.

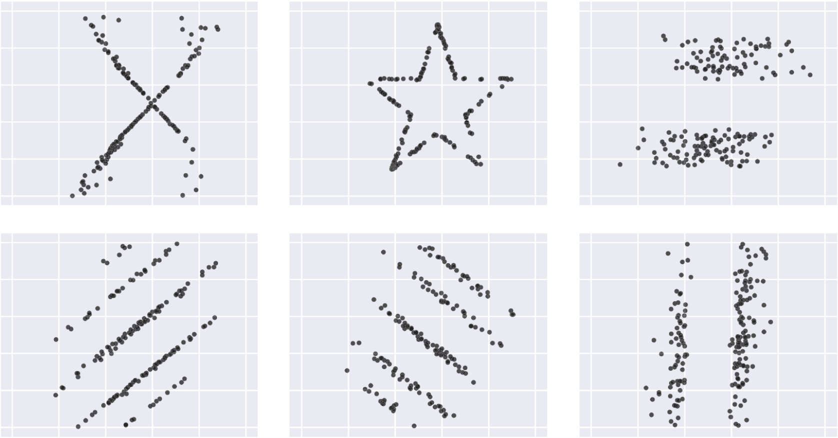
Different data with the same ρ



All datasets[1] have $\overline{x}_n = 54.02$, $\overline{y}_n = 47.83$, $\sigma_x = 14.52$, $\sigma_y = 15.26$, $\rho = 0.32$.

[1] See `http://dx.doi.org/10.1145/3025453.3025912` or `https://en.wikipedia.org/wiki/Anscombe's_quartet`.

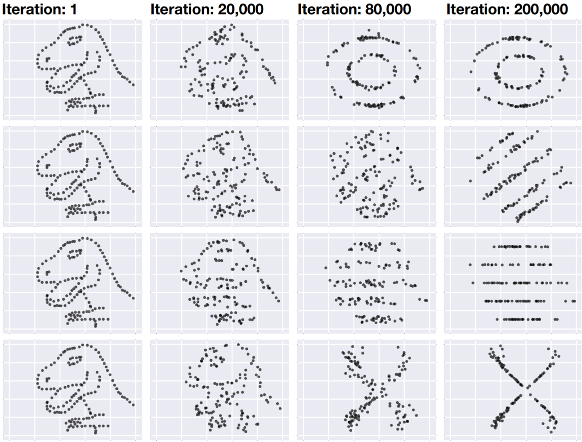
More datasets with the same parameters



All datasets[2] have $\overline{x}_n = 54.02$, $\overline{y}_n = 47.83$, $\sigma_x = 14.52$, $\sigma_y = 15.26$, $\rho = 0.32$.

[2] See `http://dx.doi.org/10.1145/3025453.3025912` or `https://en.wikipedia.org/wiki/Anscombe's_quartet`.

Construction from initial dinosaur dataset



See `http://dx.doi.org/10.1145/3025453.3025912`.

Spearman's rank correlation coefficient

Definition 2.
Let (X, Y) be a random vector with continuous marginals F_1, F_2. The **Spearman's rank correlation coefficient** between X and Y is defined as the Pearson's correlation of their coupla:

$$\rho_S(X, Y) = \rho(F_1(X), F_2(Y)).$$

Properties of Spearman's rank correlation

- ρ_S only depends on the copula of (X, Y).
- Hence, ρ_S inherits invariance properties of the copula:

$$\rho_S(g_1(X), g_2(Y)) = \pm\rho_S(X, Y)$$

  with $+$ if g_1, g_2 have the same monotonicity and $-$ otherwise.
- In particular, $\rho_S(aX + b, cY + d) = \operatorname{sign}(ac)\,\rho_S(X, Y)$.
- Although ρ only makes sense for random variables with finite variances, ρ_S is well-defined for any random variables with continuous marginals.

Equivalent formulas for Spearman's rho

- Let (X', Y') be a random vector independent of (X, Y), with the **same marginals**, but **independent components**. Define $\Delta_X = X - X'$ and $\Delta_Y = Y - Y'$. Then

$$\rho_S(X, Y) = 3\,\mathbb{E}\{\text{sign}(\Delta_X \Delta_Y)\}.$$

- Let (U', V') be two **independent** $\text{Unif}(0, 1)$ random variables. Then

$$\rho_S(X, Y) = 12\,\mathbb{E}\{C(U', V')\} - 3 = 12 \int_{[0,1]^2} C(u, v)\,du\,dv - 3.$$

- Applying integration by parts formula to the last integral yields

$$\rho_S(X, Y) = 12\,\mathbb{E}\{UV\} - 3 \quad \text{where} \quad (U, V) \sim C.$$

Kendall's tau

Definition 3 (Kendall's tau).

Let (X_1, Y_1) and (X_2, Y_2) be two **independent copies** of (X, Y). Denote the differences by $\Delta_X = X_1 - X_2$ and $\Delta_Y = Y_1 - Y_2$. The **Kendall's tau** between X and Y is defined as

$$\tau(X, Y) = \mathbb{P}\left\{\Delta_X \Delta_Y > 0\right\} - \mathbb{P}\left\{\Delta_X \Delta_Y < 0\right\}.$$

Two equivalent formulas:

- $\tau(X, Y) = \mathbb{E}\{\text{sign}(\Delta_1 \Delta_2)\}$
- $\tau(X, Y) = 2\,\mathbb{P}\left\{\Delta_X \Delta_Y > 0\right\} - 1$

Kendall's tau invariance property

Theorem 4.
Let (X, Y) be a random vector with continuous marginals and copula C. Then Kendall's tau is invariant under strictly increasing transformations of X and Y.

Proof: Let g be an increasing function. Then

$$\Delta_X > 0 \iff \Delta_{g(X)} > 0, \quad \Delta_Y > 0 \iff \Delta_Y > 0.$$

Hence, for g_1, g_2 increasing we have

$$\tau(X, Y) = \tau(g_1(X), g_2(Y)).$$

Kendall's tau in terms of copula

Theorem 5.
The Kendall's tau of a random vector (X, Y) depends only on its copula C and is given by

$$\tau(X, Y) = 4 \int_{[0,1]^2} C(u, v)\, dC(u, v) - 1.$$

Proof: By monotone invariance theorem,

$$\tau(X, Y) = \tau(U, V), \quad \text{where} \quad U = F_1(X), \quad V = F_2(Y).$$

Hence, τ only depends on the copula C of (X, Y). Next,

$$\mathbb{P}\{\Delta_U \Delta_V > 0\} = 2\,\mathbb{P}\{U_1 < U_2, V_1 < V_2\} = 2 \int_{[0,1]^2} C(u, v)\, dC(u, v).$$

It remains to use $\tau(U, V) = 2\,\mathbb{P}\{\Delta_U \Delta_V > 0\} - 1$.

Example: Kendall's tau for C_I, C_U and C_L

- If $(U, V) \sim C_I$, then

$$\tau(U, V) = 4\,\mathbb{E}\{UV\} - 1 = 4 \cdot \frac{1}{4} - 1 = 0.$$

- If $(U, V) \sim C_U$, then $U = V$ a.s. and thus

$$\tau(U, V) = 4\,\mathbb{E}\{\min\{U, V\}\} - 1 = 4 \cdot \frac{1}{2} - 1 = 1.$$

- If $(U, V) \sim C_L$, then $U = 1 - V$ a.s. and thus

$$\begin{aligned}\tau(U, V) &= 4\,\mathbb{E}\{(U + V - 1)_+\} - 1 \\ &= 4\,\mathbb{E}\{(U + (1 - U) - 1)_+\} - 1 = 4 \cdot 0 - 1 = -1.\end{aligned}$$

Compare Kendall's tau and Spearman's rho

$$\tau(C) = 4\int_{[0,1]^2} C(u,v)\, dC(u,v) - 1 \quad \text{vs} \quad \rho_S(C) = 12\int_{[0,1]^2} C(u,v)\, du\, dv - 3.$$

$$\tau(C) = 4\,\mathbb{E}\{C(U,V)\} - 1 \quad \text{vs} \quad \rho_S(C) = 12\,\mathbb{E}\{C(U',V')\} - 3,$$

Important difference: Note that $\rho_S(C)$ is linear in C whereas $\tau(C)$ is not. Hence,

$$\rho_S(\theta C_1 + (1-\theta)C_2) = \theta\,\rho_S(C_1) + (1-\theta)\,\rho_S(C_2),$$

whereas $\tau(\theta C_1 + (1-\theta)C_2)$ is not equal to the corresponding linear combination of $\tau(C_1)$ and $\tau(C_2)$.

τ and ρ_S for elliptical random vectors

Let (X, Y) be an elliptical random vector with correlation parameter ρ. Similarly to the Gaussian case, we can represent (X, Y) as

$$(X, Y) \stackrel{d}{=} R\,(U_1, \rho U_1 + \sqrt{1-\rho^2}U_2),$$

where $R > 0$ a.s. is a random radius independent of $(U_1, U_2) \sim \mathbb{S}^1$.

Then, it can be shown that

$$\tau(X, Y) = \frac{2}{\pi}\arcsin(\rho) \quad \text{and} \quad \rho_S(X, Y) = \frac{6}{\pi}\arcsin\left(\frac{\rho}{2}\right).$$

In particular, these formulas are valid if $(X, Y) \sim N_2(\boldsymbol{\mu}, \Sigma)$.

Kendall's tau for Archimedean copulas

The following theorem is given without proof.

Theorem 6.
If ψ is a differentiable Archimedean generator, then $\tau(C_\psi) = 1 + 4\int_0^1 \frac{\psi(t)}{\psi'(t)}\,dt.$

Example: $\psi(x) = (-\ln x)^\theta$ is the generator of the Gumbel copula C_θ. Then,

$$\tau(C_\theta) = 1 + 4\int_0^1 \frac{(-\ln t)^\theta}{-\theta(-\ln t)^{\theta-1}\ \frac{1}{t}}\,dt = 1 - \frac{4}{\theta}\int_0^\infty u e^{-u}\,du = 1 - \frac{4}{\theta}.$$

Correlation order

Definition 7 (Correlation order).
Let $\boldsymbol{X} \sim F$ and $\boldsymbol{Y} \sim G$ be two random vectors. We say that $\boldsymbol{X}$ is smaller than $\boldsymbol{Y}$ in the **correlation order**, denoted by $\boldsymbol{X} \preceq_{\text{corr}} \boldsymbol{Y}$, if

$$F(\boldsymbol{x}) \leq G(\boldsymbol{x}) \quad \text{for all} \quad \boldsymbol{x} \in \mathbb{R}^d.$$

Example: If $\boldsymbol{X}$ and $\boldsymbol{Y}$ are bivariate Gaussian random vectors with same marginals and correlations ρ_X and ρ_Y, respectively, then

$$\boldsymbol{X} \preceq_{\text{corr}} \boldsymbol{Y} \iff \rho_X \leq \rho_Y.$$

Dependence measures and correlation order

Theorem 8.
Let $\boldsymbol{X}$ and $\boldsymbol{Y}$ be two random vectors with continuous marginals. If $\boldsymbol{X} \preceq_{\text{corr}} \boldsymbol{Y}$, then

$$\mu(\boldsymbol{X}) \leq \mu(\boldsymbol{Y}) \quad \textit{for} \quad \mu \in \{\rho_S, \tau\}$$

and with $\mu = \rho$ if ρ exists.

Questions/exercises

- How to estimate Kendall's tau and Spearman's rho from data?
- We know that $\rho \in [-1, 1]$. Check that the same is true for ρ_S and τ.
- Explain why ρ, ρ_S and τ agree in the edge cases:

$$\rho = \pm 1 \iff \rho_S = \pm 1 \iff \tau = \pm 1.$$

- If $\boldsymbol{X}$ and $\boldsymbol{Y}$ are bivariate vectors with the same marginals, is it true that

$$\boldsymbol{X} \preceq_{\text{corr}} \boldsymbol{Y} \iff \overline{F}_{\boldsymbol{X}} \leq \overline{F}_{\boldsymbol{Y}} \quad ?$$