Lecture 10. Non-parametric statistics

Enkelejd Hashorva & Pavel Ievlev
Université de Lausanne

24 November, 2025

Leargning objectives

- Random and pseudo random samples.
- Non-parametric estimation.
- How to estimate dependence measures from data?

Random sample

Definition 1 (Random sample).
A **random sample** is a collection

$$\mathcal{D}_n = \{\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n\}$$

of n **independent** and **identically distributed** (i.i.d.) random vectors. We abbreviate this as $\mathcal{D} \sim F$, where F is the common distribution function of the vectors $\boldsymbol{X}_i$.

Random sample has to be distinguished from **observations**, which are the **actual values** taken by the random vectors in the sample in a particular experiment:

$$\mathscr{D}_n = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}.$$

Properties and non-examples

Properties of random samples:

- If $f : \mathbb{R}^d \to \mathbb{R}^m$ is a **deterministic** function and $\mathcal{D}_n = \{\boldsymbol{X}_1, \dots, \boldsymbol{X}_n\}$ is a random sample, then $f(\mathcal{D}_n) := \{f(\boldsymbol{X}_1), \dots, f(\boldsymbol{X}_n)\}$ is a random sample.
- Joint distribution function of a random sample is the product of the marginal distribution functions: $F_{\mathcal{D}_n}(\boldsymbol{x}_1, \dots, \boldsymbol{x}_n) = \prod_{i=1}^n F(\boldsymbol{x}_i)$.

Non-examples:

- A **time series** $\{X_t\}_{t \in \mathbb{N}}$ where dependence between observations exists is **not** a random sample.
- If we apply a **random transformation** to a random sample, the resulting collection is **not necessarily** a random sample.

Empirical distribution function

Given a random sample $\mathcal{D}_n \sim F$, we can estimate the joint distribution function F by the empirical distribution function:

$$\hat{F}_n(\boldsymbol{x}) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}\{\boldsymbol{X}_i \leq \boldsymbol{x}\}.$$

Similarly, we can estimate the marginal distribution functions F_j by their empirical counterparts:

$$\hat{F}_{j,n}(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}\{X_{i,j} \leq x\}, \quad j = 1, \ldots, d.$$

By the **Glivenko-Cantelli theorem**, the empirical distribution functions converge uniformly (almost surely) to the true distribution functions: $\hat{F}_n \to F$.

Monte-Carlo estimators

Empirical distribution functions are examples of **Monte-Carlo estimators**.

Assume that we have a random sample $\mathcal{D}_n \sim F$ and we want to estimate the quantity θ which **can be represented** as an expectation:

$$\theta = \mathbb{E}\left\{g(\boldsymbol{X})\right\} \quad \text{with some function } g : \mathbb{R}^d \to \mathbb{R}.$$

Then, the **Monte-Carlo estimator** of θ is given by:

$$\hat{\theta}_n = \frac{1}{n}\sum_{i=1}^{n} g(\boldsymbol{X}_i).$$

The key idea is to swap expectation by empitical average: $\mathbb{E} \rightsquigarrow \frac{1}{n}\sum_i$.

Properties of Monte-Carlo estimators

- Monte-Carlo estimators are **unbiased**:

$$\mathbb{E}\left\{\hat{\theta}_n\right\} = \theta \quad \text{for all } n.$$

- Monte-Carlo estimators are **strongly consistent** by SLLN:

$$\hat{\theta}_n \overset{a.s.}{\to} \theta \quad \text{as } n \to \infty.$$

- Monte-Carlo estimators are **asymptotically normal** by CLT:

$$\sqrt{n}(\hat{\theta}_n - \theta) \overset{d}{\to} \mathcal{N}(0, \sigma^2) \quad \text{where} \quad \sigma^2 = \operatorname{Var} g(\boldsymbol{X}).$$

- Some authors use the term **Monte-Carlo method** to refer to more

Remarks about extensions of MC estimation method

Some authors use the term **Monte-Carlo method** to refer to more general situations. For example,

$$\theta = \frac{\mathbb{E}\left\{g(X)\right\}}{\mathbb{E}\left\{h(X)\right\}} \quad \text{can be estimated by} \quad \hat{\theta}_n = \frac{\sum_{i=1}^n g(X_i)}{\sum_{i=1}^n h(X_i)}.$$

Such "extended" Monte-Carlo estimators are often still consistent and asymptotically normal[1], but they are **biased** in general.

[1] By the **Delta method**

Remarks about extensions of the MC estimation method

Note that the MC estimators as we defined them can be written as integrals with respect to the empirical distribution function:

$$\hat{\theta}_n = \int g(\boldsymbol{x})\, d\hat{F}_n(\boldsymbol{x}) = \frac{1}{n}\sum_{i=1}^{n} g(\boldsymbol{X}_i).$$

This motivates the following extension of the MC estimators:

$$\text{if } \theta = T(F) \text{ for some functional } T : \{\text{distribution functions}\} \to \mathbb{R},$$

$$\text{then we can estimate } \theta \text{ by } \hat{\theta}_n = T(\hat{F}_n).$$

The ratio estimator from the previous slide is an example of such an extension. Such extended MC estimators are also **not unbiased** in general, but they are often consistent and asymptotically normal.

Remark about extensions of the MC estimation method

Throughout this course, "MC estimators" will always refer to the basic version defined previously, i.e.

$$\hat{\theta}_n = \frac{1}{n}\sum_{i=1}^{n} g(\boldsymbol{X}_i).$$

Note also what estimating $\theta = T(F)$ by $\hat{\theta}_n = T(\hat{F}_n)$ can be thought of as **plug-in estimation**: we replace the unknown distribution function F by its empirical counterpart $\hat{F}_n$.

Motivation of pseudo random samples

To study the dependence structure of a random vector $\boldsymbol{X}$, we often need to remove the marinal information by applying the Smirnov/Rosenblatt transform:

$$\boldsymbol{X} \rightsquigarrow \boldsymbol{U} = \big(F_1(X_1), \dots, F_d(X_d)\big).$$

Given a random sample $\mathcal{D}_n = \{\boldsymbol{X}_1, \dots, \boldsymbol{X}_n\}$, we can apply the above transform to each vector in the sample and the result will still be a random sample. However, in practice the marginal distribution functions $F_1, \dots, F_d$ are **unknown** and have to be estimated from the data. Replacing F_j by its empirical counterpart $\hat{F}_{j,n}$ leads to the notion of a **pseudo random sample**.

Pseudo random sample

Definition 2 (Pseudo random sample).

A **pseudo random sample** $\mathcal{U}_n$ is a collection of vectors obtained by applying the empirical distribution functions $\hat{F}_{1,n}, \dots, \hat{F}_{d,n}$ to the components of the vectors in a random sample $\mathcal{D}_n = \{\boldsymbol{X}_1, \dots, \boldsymbol{X}_n\}$:

$$\mathcal{U}_n = \left\{ \left(\hat{F}_{1,n}(X_{i,1}), \dots, \hat{F}_{d,n}(X_{i,d}) \right) : i = 1, \dots, n \right\}.$$

Important remarks:

- Note that $\mathcal{U}_n$ is **not** a random sample since the vectors in $\mathcal{U}_n$ are **dependent** though the empirical distribution functions.
- Since $\hat{F}_{j,n} \approx F_j$ for large n, we expect that $\hat{F}_{j,n}(X_j) \approx \text{Unif}(0,1)$.
- We also expect that $(\hat{F}_{1,n}(X_1), \dots, \hat{F}_{d,n}(X_d)) \overset{d}{\approx} C_F$, where C_F is the copula of F.

Remarks about pseudo-random samples

Assume for simplicity that $d = 2$ and consider the pseudo random sample $\mathcal{D}_n = \{(X_1, Y_1), \ldots, (X_n, Y_n)\}$. Let $\hat{U}_{ni} = \hat{F}_{1,n}(X_i)$ and $\hat{V}_{n,i} = \hat{F}_{2,n}(Y_i)$ be the components of the vectors in the pseudo random sample.

Important remarks:

- $(\hat{U}_{n,i}, \hat{V}_{n,i})$ are identically distributed.
- $\hat{U}_{n,i}$ and $\hat{V}_{n,i}$ are **not** independent.
- $\hat{U}_{n,i}$ and $\hat{V}_{n,i}$ are **not** uniformly distributed on $(0, 1)$, but they *almost* are: they have the **discrete** distribution with mass $1/n$ at points $1/n,\ 2/n,\ \ldots,\ 1$.
- Since they are **not** independent, standard i.i.d. asymptotic results like CLT, WLLN and SLLN cannot be applied without further justification.

Estimation of Kendall's τ

Recall that Kendall's τ was defined as

$$\tau = \mathbb{P}\{\Delta_X \Delta_Y > 0\} - \mathbb{P}\{\Delta_X \Delta_Y < 0\}.$$

Let us use this formula to build a Monte Carlo estimator of τ from a random sample $\mathcal{D}_n$. Consider all possible pairs of observations (X_i, Y_i) and (X_j, Y_j) with $i \neq j$. We say that the pair is **concordant** if $(X_i - X_j)(Y_i - Y_j) > 0$ and **discordant** otherwise. Then, estimating $\mathbb{P}\{\dots\}$ by empirical probabilities we obtain

$$\hat{\tau}_n = \frac{\text{number of concordant pairs} - \text{number of discordant pairs}}{\binom{n}{2}}.$$

Estimation of Kendall's τ

Recall that Kendall's τ was defined as

$$\tau = \mathbb{P}\left\{\Delta_X \Delta_Y > 0\right\} - \mathbb{P}\left\{\Delta_X \Delta_Y < 0\right\}.$$

Let us use this formula to build a Monte Carlo estimator of τ from a random sample $\mathcal{D}_n$. Consider all possible pairs of observations (X_i, Y_i) and (X_j, Y_j) with $i \neq j$. We say that the pair is **concordant** if $(X_i - X_j)(Y_i - Y_j) > 0$ and **discordant** otherwise. Then, estimating $\mathbb{P}\left\{\ldots\right\}$ by empirical probabilities we obtain

$$\hat{\tau}_n = \frac{n_C - n_D}{n(n-1)/2} = 1 - \frac{4n_D}{n(n-1)}$$

$$= 1 - \frac{4}{n(n-1)} \sum_{1 \leq i < j \leq n} \mathbb{1}\{(X_i - X_j)(Y_i - Y_j) < 0\}$$

Estimation of ρ_S

Recall that one of the equivalent formulas for Spearman's ρ_S was

$$\rho_S = 12\,\mathbb{E}\left\{UV\right\} - 3, \quad \text{where } (U, V) \sim C_F.$$

Let us use this formula to build a Monte Carlo estimator of ρ_S from a random sample $\mathcal{D}_n$. Using the pseudo random sample $\mathcal{U}_n$, we can estimate $\mathbb{E}\left\{UV\right\}$ by the empirical average:

$$\hat{\rho}_{S,n} = 12 \cdot \frac{1}{n} \sum_{i=1}^{n} \hat{U}_{n,i} \hat{V}_{n,i} - 3.$$

Estimation of λ

Recall that one of the equivalent formulas for the upper tail dependence coefficient λ was

$$\lambda = \lim_{u \uparrow 1} \frac{\mathbb{P}\{V > u,\ U > u\}}{1-u}, \quad \text{where } (U,V) \sim C.$$

Let us use this formula to build a Monte Carlo estimator of λ from a random sample $\mathcal{D}_n$. Swapping the limit by just high quantile level and replacing U and V by their pseudo random sample counterparts, we obtain

$$\hat{\lambda}_n(u) = \frac{1}{n(1-u)} \sum_{i=1}^{n} \mathbb{1}\{\hat{U}_{n,i} > u,\ \hat{V}_{n,i} > u\}$$

$$= \frac{1}{n(1-u)} \sum_{i=1}^{n} \mathbb{1}\{X_i > X_{[nu]:n},\ Y_i > Y_{[nu]:n}\}$$

The choice of u is a hard problem in practice!

Questions/exercises

- Is it okay to use MC estimators on pseudo random samples?
- How to construct a confidence interval for $\hat{F}$?
- Describe the MC estimator of $\mathbb{P}\{(X, Y) \in A\}$ if $A \subset \mathbb{R}^2$ in simple words.