Statistics for Applications

Sorbonne Data Analytics

Jose Angel Garcia Sanchez

Université Paris 1 Panthéon-Sorbonne

November 2025

Outline

UNIVERSITÉ PARIS 1
PANTHÉON SORBONNE

Goals:

- To give you a solid introduction to the mathematical theory behind statistical methods
- To provide theoretical guarantees for the statistical methods that you may use for certain applications

At the end of this class, you will be able to:

1. From a real-life situation, formulate a statistical problem in mathematical terms
2. Select appropriate statistical methods for your problem
3. Understand the implications and limitations of various methods

Why Statistics?

Real-world applications across history and disciplines:

Hydrology (Netherlands, 10th century, building dams and dykes)

- Should be high enough for most floods
- Should not be too expensive (high)

Insurance

- Given your driving record, car information, coverage
- What is a fair premium?

Clinical trials

- A drug is tested on 100 patients; 56 were cured and 44 showed no improvement
- Is the drug effective?

What is common to all these examples?

What is common to all these examples?

RANDOMNESS

Associated questions:

- Notion of average ("fair premium", ...)
- Quantifying chance ("most of the floods", ...)
- Significance, variability, ...

Probability

Probability studies randomness

Sometimes, the physical process is completely known: dice, cards, roulette, fair coins, ...

Example: Rolling 1 die

- Alice gets 1€ if # of dots ≥ 3
- Bob gets 2€ if # of dots ≤ 2
- Who do you want to be: Alice or Bob?

Example: Rolling 2 dice

- Choose a number between 2 and 12
- Win 100€ if you chose the sum of the 2 dice
- Which number do you choose?

Well known random process from physics: 1/6 chance of each side, dice are independent. We can deduce the probability of outcomes, and expected amounts. This is probability.

How about more complicated processes?

- Need to estimate parameters from data. This is statistics
- Sometimes real randomness (random student, biased coin, measurement error, ...)
- Sometimes deterministic but too complex phenomenon: statistical modeling

Key Principle

Complicated process = Simple process + random noise

(Good) Modeling consists in choosing (plausible) simple process and noise distribution.

Statistics vs. Probability

Probability:

Previous studies showed that the drug was 80% effective. Then we can anticipate that for a study on 100 patients, in average 80 will be cured and at least 65 will be cured with 99.99% chances.

Statistics:

Observe that 78/100 patients were cured. We (will be able to) conclude that we are 95% confident that for other studies the drug will be effective on between 69.88% and 86.11% of patients.

Heuristics (1)

Study Question

"A neonatal right-side preference makes a surprising romantic reappearance later in life."

- Let p denote the proportion of couples that turn their head to the right when kissing
- Let us design a statistical experiment and analyze its outcome
- Observe n kissing couples and collect the value of each outcome (say 1 for RIGHT and 0 for LEFT)
- Estimate p with the proportion $\hat{p}$ of RIGHT
- Study: "Human behaviour: Adult persistence of head-turning asymmetry" (Nature, 2003): $n = 124$, 80 to the right so $\hat{p} = \frac{80}{124} = 64.5\%$

UNIVERSITÉ PARIS 1
PANTHÉON SORBONNE

Heuristics (2)

Back to the data:

- 64.5% is much larger than 50% so there seems to be a preference for turning right
- What if our data was RIGHT, RIGHT, LEFT ($n = 3$). That's 66.7% to the right. Even better?
- Intuitively, we need a large enough sample size n to make a call. How large?

Key Question

We need mathematical modeling to understand the accuracy of this procedure

Heuristics (3)

Formally, this procedure consists of doing the following:

- For $i = 1, \ldots, n$, define $R_i = 1$ if the ith couple turns to the right, $R_i = 0$ otherwise
- The estimator of p is the sample average

$$\hat{p} = \bar{R}_n = \frac{1}{n} \sum_{i=1}^{n} R_i$$

What is the accuracy of this estimator?

In order to answer this question, we propose a statistical model that describes/approximates well the experiment.

Heuristics (4)

Coming up with a model consists of making assumptions on the observations R_i, $i = 1, \ldots, n$ in order to draw statistical conclusions.

Here are the assumptions we make:

1. Each R_i is a random variable
2. Each of the r.v. R_i is Bernoulli with parameter p
3. $R_1, \ldots, R_n$ are mutually independent

Heuristics (5)

Let us discuss these assumptions:

1. **Randomness** is a way of modeling lack of information; with perfect information about the conditions of kissing (including what goes in the kissers' mind), physics or sociology would allow us to predict the outcome.

2. **Bernoulli distribution:** The R_i's are necessarily Bernoulli r.v. since $R_i \in \{0, 1\}$. They could have different parameters $R_i \sim \text{Ber}(p_i)$ for each couple but we don't have enough information to estimate the p_i's accurately. So we assume: $p_i = p$ for all i.

3. **Independence** is reasonable (people were observed at different locations and different times).

UNIVERSITÉ PARIS 1
PANTHÉON SORBONNE

Two Important Tools: LLN & CLT

Let $X, X_1, X_2, \dots, X_n$ be i.i.d. r.v., $\mu = \mathbb{E}[X]$ and $\sigma^2 = \mathrm{Var}[X]$.

Laws of Large Numbers (weak and strong):

$$\bar{X}_n := \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{\mathbb{P},\text{a.s.}} \mu$$

Central Limit Theorem:

$$\sqrt{n}\frac{\bar{X}_n - \mu}{\sigma} \xrightarrow{(d)} N(0,1)$$

(Equivalently, $\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{(d)} N(0, \sigma^2)$)

UNIVERSITÉ PARIS 1 PANTHÉON SORBONNE

Consequences (1)

- The LLN's tell us that

$$\bar{R}_n \xrightarrow{\mathbb{P}\text{,a.s.}} p$$

- Hence, when the size n of the experiment becomes large, $\bar{R}_n$ is a good (say "consistent") estimator of p
- The CLT refines this by quantifying how good this estimate is

Consequences (2)

Let $\Phi(x)$: cdf of $N(0,1)$; $\Phi_n(x)$: cdf of $\sqrt{n}\frac{\bar{R}_n - p}{\sqrt{p(1-p)}}$.

CLT: $\Phi_n(x) \approx \Phi(x)$ when n becomes large. Hence, for all $x > 0$,

$$\mathbb{P}\left[|\bar{R}_n - p| \leq x\sqrt{\frac{p(1-p)}{n}}\right] \approx 2\Phi(x) - 1$$

UNIVERSITÉ PARIS 1
PANTHÉON SORBONNE

Consequences:

- Approximation on how $\bar{R}_n$ concentrates around p
- For a fixed $\alpha \in (0,1)$, if $q_{\alpha/2}$ is the $(1-\alpha/2)$-quantile of $N(0,1)$, then with probability $\approx 1-\alpha$ (if n is large enough!),

$$p \in \left[\bar{R}_n - q_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}, \bar{R}_n + q_{\alpha/2}\sqrt{\frac{p(1-p)}{n}} \right]$$

Consequences (4)

- Note that no matter the (unknown) value of p,

$$p(1-p) \leq 1/4$$

- Hence, roughly with probability at least $1-\alpha$,

$$p \in \left[\bar{R}_n - \frac{q_{\alpha/2}}{2\sqrt{n}}, \bar{R}_n + \frac{q_{\alpha/2}}{2\sqrt{n}}\right]$$

- In other words, when n becomes large, the interval

$$\left[\bar{R}_n - \frac{q_{\alpha/2}}{2\sqrt{n}}, \bar{R}_n + \frac{q_{\alpha/2}}{2\sqrt{n}}\right]$$

  contains p with probability $\approx 1-\alpha$
- This interval is called an **asymptotic confidence interval** for p

Consequences (5)

In the kiss example, we get:

$$0.645 \pm \frac{1.96}{2\sqrt{124}} = [0.56, 0.73]$$

If the extreme ($n = 3$) case we would have $[0.10, 1.23]$ but CLT is not valid! Actually we can make exact computations!

Another Useful Tool: Hoeffding's Inequality

What if n is not so large?

Hoeffding's Inequality (i.i.d. case)

Let n be a positive integer and $X, X_1, \ldots, X_n$ be i.i.d. r.v. such that $X \in [a, b]$ a.s. ($a < b$ are given numbers). Let $\mu = \mathbb{E}[X]$. Then, for all $\lambda > 0$,

$$\mathbb{P}[|\bar{X}_n - \mu| \geq \lambda] \leq 2e^{-\frac{2n\lambda^2}{(b-a)^2}}$$

Consequence: For $\alpha \in (0,1)$, with probability $\geq 1 - \alpha$,

$$\bar{R}_n - \sqrt{\frac{\log(2/\alpha)}{2n}} \leq p \leq \bar{R}_n + \sqrt{\frac{\log(2/\alpha)}{2n}}$$

This holds even for small sample sizes n.

UNIVERSITÉ PARIS 1 PANTHÉON SORBONNE

Review of Different Types of Convergence (1)

Let $(T_n)_{n\geq 1}$ a sequence of r.v. and T a r.v. (T may be deterministic).

Almost surely (a.s.) convergence:

$$T_n \xrightarrow{\text{a.s.}} T \text{ iff } \mathbb{P}[\{\omega : T_n(\omega) \to T(\omega)\}] = 1$$

Convergence in probability:

$$T_n \xrightarrow{\mathbb{P}} T \text{ iff } \mathbb{P}[|T_n - T| \geq \lambda] \to 0, \ \forall \lambda > 0$$

Review of Different Types of Convergence (2)

Convergence in L^p ($p \geq 1$):

$$T_n \xrightarrow{L^p} T \text{ iff } \mathbb{E}[|T_n - T|^p] \to 0$$

Convergence in distribution:

$$T_n \xrightarrow{(d)} T \text{ iff } \mathbb{P}[T_n \leq x] \to \mathbb{P}[T \leq x]$$

for all $x \in \mathbb{R}$ at which the cdf of T is continuous.

Remark

These definitions extend to random vectors (i.e., random variables in $\mathbb{R}^d$ for some $d \geq 2$).

Important characterizations of convergence in distribution

The following propositions are equivalent:

(i) $T_n \xrightarrow{(d)} T$

(ii) $\mathbb{E}[f(T_n)] \to \mathbb{E}[f(T)]$, for all continuous and bounded function f

(iii) $\mathbb{E}[e^{ixT_n}] \to \mathbb{E}[e^{ixT}]$, for all $x \in \mathbb{R}$

Review of Different Types of Convergence (4)

Important properties:

- If $(T_n)_{n\geq 1}$ converges a.s., then it also converges in probability, and the two limits are equal a.s.
- If $(T_n)_{n\geq 1}$ converges in L^p, then it also converges in L^q for all $q \leq p$ and in probability, and the limits are equal a.s.
- If $(T_n)_{n\geq 1}$ converges in probability, then it also converges in distribution
- If f is a continuous function:

$$T_n \xrightarrow{\text{a.s.}/\mathbb{P}/(d)} T \Rightarrow f(T_n) \xrightarrow{\text{a.s.}/\mathbb{P}/(d)} f(T)$$

UNIVERSITÉ PARIS 1 PANTHÉON SORBONNE

Review of Different Types of Convergence (5)

Limits and operations

One can add, multiply, ... limits almost surely and in probability. If $U_n \xrightarrow{\text{a.s.}/\mathbb{P}} U$ and $V_n \xrightarrow{\text{a.s.}/\mathbb{P}} V$, then:

- $U_n + V_n \xrightarrow{\text{a.s.}/\mathbb{P}} U + V$
- $U_n V_n \xrightarrow{\text{a.s.}/\mathbb{P}} UV$
- If in addition, $V \neq 0$ a.s., then $\frac{U_n}{V_n} \xrightarrow{\text{a.s.}/\mathbb{P}} \frac{U}{V}$

Warning

In general, these rules do not apply to convergence in distribution unless the pair (U_n, V_n) converges in distribution to (U, V).

Another Example (1)

Scenario:

- You observe the times between arrivals of the T at Kendall: $T_1, \ldots, T_n$
- You assume that these times are:
  - Mutually independent
  - Exponential random variables with common parameter $\lambda > 0$
- You want to estimate the value of λ, based on the observed arrival times

Another Example (2)

Discussion of the assumptions:

Mutual independence of $T_1, \ldots, T_n$: plausible but not completely justified (often the case with independence).

$T_1, \ldots, T_n$ **are exponential r.v.:** lack of memory of the exponential distribution:

$$\mathbb{P}[T_1 > t + s | T_1 > t] = \mathbb{P}[T_1 > s], \ \forall s, t \geq 0$$

Also, $T_i > 0$ almost surely!

Same parameter: in average all the same inter-arrival time. True only for limited period (rush hour $\neq$ 11pm).

Another Example (3)

- Density of T_1: $f(t) = \lambda e^{-\lambda t}, \forall t \geq 0$
- $\mathbb{E}[T_1] = \frac{1}{\lambda}$
- Hence, a natural estimate of $\frac{1}{\lambda}$ is

$$\bar{T}_n := \frac{1}{n} \sum_{i=1}^{n} T_i$$

- A natural estimator of λ is

$$\hat{\lambda} := \frac{1}{\bar{T}_n}$$

Another Example (4)

- By the LLN's,

$$\bar{T}_n \xrightarrow{\text{a.s.}/\mathbb{P}} \frac{1}{\lambda}$$

- Hence,

$$\hat{\lambda} \xrightarrow{\text{a.s.}/\mathbb{P}} \lambda$$

- By the CLT,

$$\sqrt{n}\left(\bar{T}_n - \frac{1}{\lambda}\right) \xrightarrow{(d)} N(0, \lambda^{-2})$$

- How does the CLT transfer to $\hat{\lambda}$? How to find an asymptotic confidence interval for λ?

UNIVERSITÉ PARIS 1
PANTHÉON SORBONNE

The Delta Method

Let $(Z_n)_{n\geq 1}$ sequence of r.v. that satisfies

$$\sqrt{n}(Z_n - \theta) \xrightarrow{(d)} N(0, \sigma^2)$$

for some $\theta \in \mathbb{R}$ and $\sigma^2 > 0$ (the sequence $(Z_n)_{n\geq 1}$ is said to be asymptotically normal around θ).

Let $g : \mathbb{R} \to \mathbb{R}$ be continuously differentiable at the point θ.

Then:

- $(g(Z_n))_{n\geq 1}$ is also asymptotically normal
- More precisely,

$$\sqrt{n}(g(Z_n) - g(\theta)) \xrightarrow{(d)} N(0, (g'(\theta))^2\sigma^2)$$

Consequence of the Delta Method (1)

- $\sqrt{n}(\hat{\lambda} - \lambda) \xrightarrow{(d)} N(0, \lambda^2)$
- Hence, for $\alpha \in (0, 1)$ and when n is large enough,

$$|\hat{\lambda} - \lambda| \leq \frac{q_{\alpha/2}\lambda}{\sqrt{n}}$$

- Can $\left[\hat{\lambda} - \frac{q_{\alpha/2}\lambda}{\sqrt{n}}, \hat{\lambda} + \frac{q_{\alpha/2}\lambda}{\sqrt{n}}\right]$ be used as an asymptotic confidence interval for λ?
- No! It depends on λ, ...

UNIVERSITÉ PARIS 1
PANTHÉON SORBONNE

Two ways to overcome this issue:

1. **Solve for λ:**

$$|\hat{\lambda} - \lambda| \leq \frac{q_{\alpha/2}\lambda}{\sqrt{n}} \Leftrightarrow \lambda\left(1 - \frac{q_{\alpha/2}}{\sqrt{n}}\right) \leq \hat{\lambda} \leq \lambda\left(1 + \frac{q_{\alpha/2}}{\sqrt{n}}\right)$$

$$\Leftrightarrow \hat{\lambda}\left(1 + \frac{q_{\alpha/2}}{\sqrt{n}}\right)^{-1} \leq \lambda \leq \hat{\lambda}\left(1 - \frac{q_{\alpha/2}}{\sqrt{n}}\right)^{-1}$$

Hence, $\left[\hat{\lambda}\left(1 + \frac{q_{\alpha/2}}{\sqrt{n}}\right)^{-1}, \hat{\lambda}\left(1 - \frac{q_{\alpha/2}}{\sqrt{n}}\right)^{-1}\right]$ is an asymptotic confidence interval for λ.

2. **A systematic way:** Slutsky's theorem.

Slutsky's Theorem

Slutsky's Theorem

Let (X_n), (Y_n) be two sequences of r.v., such that:

(i) $X_n \xrightarrow{(d)} X$

(ii) $Y_n \xrightarrow{\mathbb{P}} c$

where X is a r.v. and c is a given real number. Then,

$$(X_n, Y_n) \xrightarrow{(d)} (X, c)$$

In particular:

- $X_n + Y_n \xrightarrow{(d)} X + c$
- $X_n Y_n \xrightarrow{(d)} cX$

UNIVERSITÉ PARIS 1 PANTHÉON SORBONNE

Consequence of Slutsky's Theorem (1)

- Thanks to the Delta method, we know that

$$\sqrt{n}\frac{\hat{\lambda}-\lambda}{\lambda} \xrightarrow{(d)} N(0,1)$$

- By the weak LLN,

$$\hat{\lambda} \xrightarrow{\mathbb{P}} \lambda$$

- Hence, by Slutsky's theorem,

$$\sqrt{n}\frac{\hat{\lambda}-\lambda}{\hat{\lambda}} \xrightarrow{(d)} N(0,1)$$

- Another asymptotic confidence interval for λ is

$$\left[\hat{\lambda} - \frac{q_{\alpha/2}\hat{\lambda}}{\sqrt{n}}, \hat{\lambda} + \frac{q_{\alpha/2}\hat{\lambda}}{\sqrt{n}}\right]$$

Consequence of Slutsky's Theorem (2)

Remark:

- In the first example (kisses), we used a problem dependent trick: "$p(1-p) \leq 1/4$"
- We could have used Slutsky's theorem and get the asymptotic confidence interval

$$\left[\bar{R}_n - q_{\alpha/2}\sqrt{\frac{\bar{R}_n(1-\bar{R}_n)}{n}}, \bar{R}_n + q_{\alpha/2}\sqrt{\frac{\bar{R}_n(1-\bar{R}_n)}{n}}\right]$$

UNIVERSITÉ PARIS 1
PANTHÉON SORBONNE

Summary

Key concepts covered:

- Statistical thinking: from real-world problems to mathematical formulations
- Difference between probability and statistics
- Practical examples: kissing study and metro arrivals
- Mathematical tools: LLN, CLT, Hoeffding's inequality
- Types of convergence and their properties
- Delta method and Slutsky's theorem
- Confidence intervals and their construction

Next Steps

Apply these theoretical tools to more complex statistical problems and understand their limitations.

Thank you for your attention!

Questions?