Advanced statistical methods

Frédéric Pascal

CentraleSupélec, Laboratory of Signals and Systems (L2S), France
frederic.pascal@centralesupelec.fr
http://fredericpascal.blogspot.fr

MSc in Data Sciences & Business Analytics
CentraleSupélec / ESSEC
Oct. 2nd - Dec. 20th, 2017

CentraleSupélec

Part A

Reminders of probability theory and mathematical statistics

Part A: Contents

Key references of Part A

- Jacod, Jean, and Philip Protter. *Probability essentials*. Springer Science & Business Media, 2004.
- Billingsley, Patrick. Convergence of probability measures. John Wiley & Sons, 2013. (More Advanced!!!)
- Support for the course of probabilities (in french) on my website: http://fredericpascal.blogspot.fr/p/teaching-activities.html

+ many many references...

Random Variables (r.v.) / Vectors (r.V.)

Notations

Let X (resp. $\mathbf{x}$) a random variable (resp. vectors). Denote by P or P_θ its probability :

- $P(X=x)$ or $P_\theta(X=x)$ for the discrete case
- $f(x)$ or $f_\theta(x)$ for the continuous case (with PDF)

Some other notations:

- $E[.]$ or $E_\theta[.]$ (resp. $V[.]$ / $V_\theta[.]$) stands for the statistical expectation (resp. the variance)
- i.i.d. $\rightarrow$ Independent (denoted $\perp$) and Identically Distributed, i.e. same distribution and $X \perp Y \iff$ for any measurable functions h and g, $E[g(X)h(Y)] = E[g(X)]E[h(Y)]$.
- n-sample $(X_1, \ldots, X_n) \iff X_1, \ldots, X_n$ are i.i.d.
- PDF, CDF and iff resp. means Probability Density Function, Cumulative Distribution Function and "*if and only if*"

Convergences

Multivariate case

Let $(\mathbf{x})_{n\in\mathbb{N}} \in \mathbb{R}^d$ a sequence of r.V. and $(\mathbf{x}) \in \mathbb{R}^d$ defined on the same probability space $(\Omega, \mathscr{A}, P)$, then

- Almost Sure CV: $\mathbf{x}_n \xrightarrow[n\to\infty]{a.s} \mathbf{x} \iff \exists N \in \mathscr{A}$ such that $P(N) = 0$ and $\forall \omega \in N^c, \lim_{n\to\infty} \mathbf{x}_n(\omega) = \mathbf{x}(\omega)$
- CV in probability: $\mathbf{x}_n \xrightarrow[n\to\infty]{P} \mathbf{x} \iff \forall \varepsilon > 0, \lim_{n\to\infty} P(\|\mathbf{x}_n - \mathbf{x}\| \geq \varepsilon) = 0$ where $\|\mathbf{x}\| = \left(\sum_{i=1}^d x_i^2\right)^{1/2}$ for $\mathbf{x} \in \mathbb{R}^d$.
  $\mathbf{x}_n \xrightarrow[n\to\infty]{P} \mathbf{x} \iff$ each component converges in probability.
- CV in $\mathscr{L}^p$: Let $p \in \mathbb{N}^*, \mathbf{x}_n \xrightarrow[n\to\infty]{\mathscr{L}^p} \mathbf{x} \iff (\mathbf{x})_{n\in\mathbb{N}}, \mathbf{x} \in \mathscr{L}^p$ and $E\left[\|\mathbf{x}_n - \mathbf{x}\|_{\mathscr{L}^p}^p\right] \xrightarrow[n\to\infty]{} 0$.

Convergence in distribution

- CV in distribution: $\mathbf{x}_n \xrightarrow[n\to\infty]{dist.} \mathbf{x}$ if for any continuous and bounded function g, one has $\lim_{n\to\infty} E\left[g(\mathbf{x}_n)\right] = E\left[g(\mathbf{x})\right]$.

⚠ The CV in distribution of a sequence of r.V. is stronger than the CV of each component!

How to characterise the CV in distribution?

Theorem (Levy continuity Theorem)

Let $\varphi_n(u) = E\left[\exp(iu^t\mathbf{x}_n)\right]$ and $\varphi(u) = E\left[\exp(iu^t\mathbf{x})\right]$ the characteristic functions of $\mathbf{x}_n$ and $\mathbf{x}$. Then,

$$\mathbf{x}_n \xrightarrow[n\to\infty]{dist.} \mathbf{x} \iff \forall u \in \mathbb{R}^d, \varphi_n(u) \xrightarrow[n\to\infty]{} \varphi(u).$$

Proposition (a.s., P, dist. convergences)

$\mathbf{x}_n \xrightarrow[n\to\infty]{} \mathbf{x} \Longrightarrow h(\mathbf{x}_n) \xrightarrow[n\to\infty]{} h(\mathbf{x})$, *if h is a continuous function*

Discussion on the cv hierarchy...

I. Random Variables / Vectors

II. Convergences

III. Essential theorems

- SLLN and CLT
- Slutsky theorem and the Delta-method

IV. Gaussian-related distributions

SLLN and CLT

Theorem (Strong (Weak) Low of Large Numbers)

Let $(\mathbf{x}_n)_{n\in\mathbb{N}^*}$ *a sequence of i.i.d. r.V. in* $\mathbb{R}^d$ *s.t.* $E[|\mathbf{x}_1|] < +\infty$. *Let* $\mu = E[\mathbf{x}_1]$ *the expectation of* $\mathbf{x}_1$. *Then,*

$$\bar{\mathbf{x}}_n = \frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i \xrightarrow[n\to\infty]{a.s,P} \mu.$$

Theorem (Central Limit Theorem)

Let $(\mathbf{x}_n)_{n\in\mathbb{N}^*}$ *a sequence of i.i.d. r.V. in* $\mathbb{R}^d$ *s.t.* $E[|\mathbf{x}_1|^2] < +\infty$. *Let* $\mu = E[\mathbf{x}_1]$ *and* $\mathbf{\Sigma} = E\left[\mathbf{x}_1\mathbf{x}_1^t\right] - E[\mathbf{x}_1]E[\mathbf{x}_1]^t$ *the covariance matrix of* $\mathbf{x}_1$. *Let* $\bar{\mathbf{x}}_n = \frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i$ *the empirical mean. Then,*

$$\sqrt{n}\left(\bar{\mathbf{x}}_n - \mu\right) \xrightarrow[n\to\infty]{dist.} \mathcal{N}\left(\mathbf{0},\mathbf{\Sigma}\right).$$

Slutsky theorem

Theorem (Slutsky theorem)

Let $(\mathbf{x}_n)_{n\in\mathbb{N}^*}$ a sequence of r.V. in $\mathbb{R}^d$ that cv in dist. to $\mathbf{x}$. Let $(\mathbf{y}_n)_{n\in\mathbb{N}^*}$ a sequence of r.V. in $\mathbb{R}^m$ (defined on the same proba. space as $(\mathbf{x}_n)_{n\in\mathbb{N}^*}$) that cv a.s. (or in P, or in dist.) towards a constant $\mathbf{a}$. Thus, the sequence $(\mathbf{x}_n, \mathbf{y}_n)_{n\in\mathbb{N}^*}$ cv in dist. towards $(\mathbf{x}, \mathbf{a})$, $(\mathbf{x}_n, \mathbf{y}_n) \xrightarrow[n\to\infty]{dist.} (\mathbf{x}, \mathbf{a})$

Remark (Important Applications of Slutsky (IAS))

Under previous assumptions, one has:

1. $\mathbf{x}_n + \mathbf{y}_n \xrightarrow[n\to\infty]{dist.} \mathbf{x} + \mathbf{a}$ *if* $m = d$
2. $\mathbf{x}_n \mathbf{y}_n \xrightarrow[n\to\infty]{dist.} a\mathbf{x}$ *if* $m = 1$
3. $\mathbf{x}_n / \mathbf{y}_n \xrightarrow[n\to\infty]{dist.} \mathbf{x}/a$ *if* $m = 1, a \neq 0$

Delta-method

Theorem (Delta-method)

Let $(\mathbf{x}_n)_{n\in\mathbb{N}^*}$ *a sequence of r.V. in* $\mathbb{R}^d$ *and* θ *a (deterministic) vector of* $\mathbb{R}^d$. *Let* $h:\mathbb{R}^d \mapsto \mathbb{R}^m$ *a function that is differentiable (at least) at point* θ. *Let us denote* $\dfrac{\partial h}{\partial \theta^t}(\theta)$ *the* $m\times d$ *matrix s.t.* $\left(\dfrac{\partial h_i}{\partial \theta_j}(\theta)\right)_{\substack{1\leq i\leq m\\ 1\leq j\leq d}}$ *and* $\dfrac{\partial h^t}{\partial \theta}(\theta) = \left(\dfrac{\partial h}{\partial \theta^t}(\theta)\right)^t$ *its transpose. Assume that* $\sqrt{n}(\mathbf{x}_n-\theta) \xrightarrow[n\to\infty]{dist.} \mathbf{x}$. *Then*

$$\sqrt{n}\,(h(\mathbf{x}_n)-h(\theta)) \xrightarrow[n\to\infty]{dist.} \frac{\partial h}{\partial \theta^t}(\theta)\,\mathbf{x}.$$

Particular case:

If $\mathbf{x}\sim\mathcal{N}(\mathbf{0},\mathbf{\Sigma})$, *then* $\sqrt{n}\,(h(\mathbf{x}_n)-h(\theta)) \xrightarrow[n\to\infty]{dist.} \mathcal{N}\left(\mathbf{0}, \dfrac{\partial h}{\partial \theta^t}(\theta)\,\mathbf{\Sigma}\,\dfrac{\partial h^t}{\partial \theta}(\theta)\right)$

I. Random Variables / Vectors

II. Convergences

III. Essential theorems

IV. Gaussian-related distributions

- Gamma and Beta distributions
- χ^2, Student-t and F- distributions
- Student Theorem

Gamma and Beta distributions

Definition (Gamma distribution)

Let $p > 0$ et $\lambda > 0$. A real-valued r.v. $X \sim \Gamma(p, \lambda)$ if its PDF is defined as

$$f(x) = \frac{\lambda^p}{\Gamma(p)} x^{p-1} \exp(-\lambda x) \mathbb{1}_{\mathbb{R}^+}(x),$$

where $\Gamma(x) = \int_0^{+\infty} t^{x-1} \exp(-t) dt$ for $x \in \mathbb{C}$ s.t. $\mathscr{R}e(x) > 0$. Also
$\Gamma(x+1) = x\Gamma(x)$ *($n \in \mathbb{N}^*, \Gamma(n) = (n-1)!$).*
If $X \sim \Gamma(p, \lambda)$ and $a > 0$, then $aX \sim \Gamma(p, \lambda/a)$

Proposition (Beta distributions)

1. *Let $Y \sim \Gamma(q, \lambda)$ and $X \sim \Gamma(p, \lambda)$ 2 independent r.v. Thus,*
   - $X + Y \sim \Gamma(p + q, \lambda)$,
   - *$X + Y$ and $\frac{X}{X+Y}$ (resp. $X + Y$ and $\frac{X}{Y}$) are independent*
   - *Distributions of $\frac{X}{X+Y}$ and $\frac{X}{Y}$ do NOT depend on λ. It resp. corresponds to **Beta distributions of 1^{st} and 2^{nd} kind**, denoted $\beta^1(p, q)$ and $\beta^2(p, q)$. PDF...*

Gamma and Beta distributions

Definition (Beta PDFs)

$$\begin{cases} \beta^1(p,q) & : \quad f(x) = \dfrac{x^{p-1}(1-x)^{q-1}}{\beta(p,q)} \mathbb{1}_{[0,1]}(x), \\ \\ \beta^2(p,q) & : \quad f(x) = \dfrac{x^{p-1}}{(1+x)^{p+q}\beta(p,q)} \mathbb{1}_{\mathbb{R}^+}(x), \end{cases}$$

with $\beta(p,q) = \dfrac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)}$.

Proposition

- *If* $U \sim \beta^1(p,q)$, $\frac{U}{1-U} \sim \beta^2(p,q)$,
- *If* $V \sim \beta^2(p,q)$, $\frac{V}{1+V} \sim \beta^1(p,q)$,
- *If* $V \sim \beta^2(p,q)$, $\frac{1}{V} \sim \beta^2(q,p)$.

χ^2, Student-t and Fisher (or F) distributions

Definition (χ^2 dist.)

Let $(X_n)_{n\in\mathbb{N}^*}$ a sequence of i.i.d. real-valued r.v. $\sim \mathcal{N}(0,1)$. Thus,

- $\sum_{i=1}^{k} X_i^2$ *follows a χ^2-distribution with k d.o.f., (denoted $\chi^2(k)$).*
- $X_1^2 \sim \Gamma\left(\frac{1}{2}, \frac{1}{2}\right)$ *and* $\sum_{i=1}^{k} X_i^2 \sim \Gamma\left(\frac{k}{2}, \frac{1}{2}\right)$

Definition (Student-t and F- distributions)

1. *If $X \sim \mathcal{N}(0,1)$, $Y \sim \chi^2(k)$, and X, Y independent, then $T = \dfrac{X}{\sqrt{(Y/k)}}$ follows a Student-t dist. with k d.o.f. (denoted $t(k)$).*
2. *If p and q are integers, if $X \sim \chi^2(p)$, $Y \sim \chi^2(q)$, and X, Y are independent, then $F = \dfrac{X/p}{Y/q}$ follows a F-dist. with p and q d.o.f., (denoted $F(p,q)$).*

Student Theorem

Theorem (Student theorem)

Let $(X_n)_{n\in\mathbb{N}^*}$ *a sequence a real-valued i.i.d. r.v.* $\sim \mathcal{N}(\mu,\sigma^2)$. *Then, one has:*

1. $\overline{X}_n = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} X_i \sim \mathcal{N}\left(\mu, \dfrac{\sigma^2}{n}\right)$.
2. $R_n = \displaystyle\sum_{i=1}^{n}\left(X_i - \overline{X}_n\right)^2 \sim \sigma^2\chi^2(n-1)$.
3. $\overline{X}_n$ *and* R_n *are independent.*
4. *If* $S_n = \sqrt{\dfrac{R_n}{n-1}}$, *then* $T_n = \dfrac{\sqrt{n}\left(\overline{X}_n - \mu\right)}{S_n} \sim t(n-1)$.

Proof

Some elements...

Some applications

Estimate unknown parameters??

A1 Mean estimation: $(X_1, \cdots, X_n) \overset{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$

- σ^2 known
- σ^2 unknown

A2 Variance estimation: $(X_1, \cdots, X_n) \overset{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$

- μ known
- μ unknown

A3 Variance comparison (test) between two independent samples: $(X_1, \cdots, X_n) \overset{iid}{\sim} \mathcal{N}(\mu_X, \sigma_X^2)$ and $(Y_1, \cdots, Y_n) \overset{iid}{\sim} \mathcal{N}(\mu_Y, \sigma_Y^2)$

- μ_X and μ_Y known
- μ_X and μ_Y unknown

Possible answers with confidence intervals

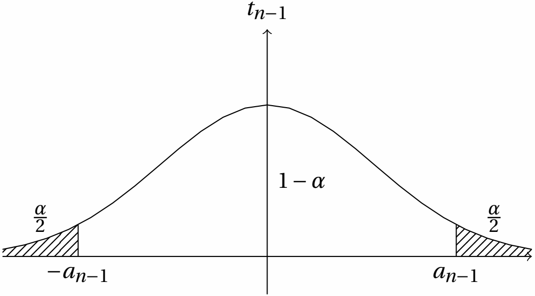
A1 Based on $\hat{\mu} = \bar{X}_n$...

- $I_n = \left[\bar{X}_n \pm \frac{1{,}96\sigma}{\sqrt{n}}\right]$ is an exact 95%-confidence interval
- $\tilde{I}_n = \left[\bar{X}_n \pm \frac{1{,}96\hat{\sigma}_n}{\sqrt{n}}\right]$ is an asymptotic 95%-confidence interval.

  OR use

$$T_n = \frac{\sqrt{n}(\bar{X}_n - \mu)}{S_n} \sim t(n-1) \Rightarrow \hat{I}_n = \left[\bar{X}_n \pm \frac{a_{n-1} S_n}{\sqrt{n}}\right]$$

  is an exact 95%-confidence interval



Possible answers with confidence intervals

A2 Based on ...

- $$R_n^* = \sum_{i=1}^{n} (X_i - \mu)^2 \sim \sigma^2 \chi^2(n) \Rightarrow I_n = \left[\frac{n\hat{\sigma}_n^2}{b_n}, \frac{n\hat{\sigma}_n^2}{a_n} \right]$$

  is an exact 95%-confidence interval with $\hat{\sigma}_n^2 = R_n^*/n$.

- $$R_n = \sum_{i=1}^{n} (X_i - X_n)^2 \sim \sigma^2 \chi^2(n-1) \Rightarrow \hat{I}_n = \left[\frac{(n-1)\hat{\sigma}_n^2}{b_{n-1}}, \frac{(n-1)\hat{\sigma}_n^2}{a_{n-1}} \right]$$

  is an exact 95%-confidence interval with $\hat{\sigma}_n^2 = R_n/(n-1)$

$\leadsto$ Loss when unknowns are present..., i.e. length of CI increases...

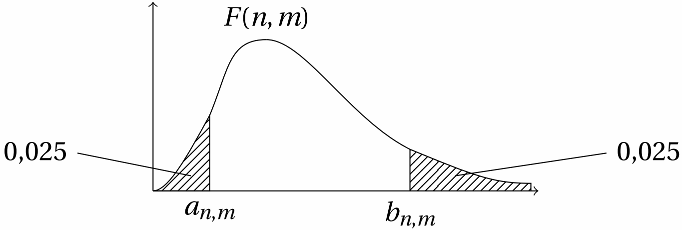
Possible answers with confidence intervals

A3 Based on ...

- $R^*_{n,X} = \sum_{i=1}^{n} (X_i - \mu_X)^2 \sim \sigma_X^2 \chi^2(n)$, $R^*_{m,Y} = \sum_{i=1}^{m} (Y_i - \mu_Y)^2 \sim \sigma_Y^2 \chi^2(m)$

$$\frac{R^*_{n,X}}{R^*_{m,Y}} \sim F(n,m) \Rightarrow \frac{\sigma_X^2}{\sigma_Y^2} \in \left[\frac{1}{b_{n,m}} \frac{\hat{\sigma}^2_{n,X}}{\hat{\sigma}^2_{m,Y}}, \frac{1}{a_{n,m}} \frac{\hat{\sigma}^2_{n,X}}{\hat{\sigma}^2_{m,Y}} \right]$$

with $\hat{\sigma}^2_{n,X} = R^*_{n,X}/n$ and $\hat{\sigma}^2_{m,Y} = R^*_{m,Y}/m$



- Same thing for μ_X and μ_Y unknown...