Lecture 3: Statistical estimation and inference

Matthieu Stigler
Matthieu.Stigler@gmail.com

December 15, 2008

convergence of a random variable

Definition (Convergence)

X_n converges to X if $\lim_{n\to\infty} p(|\hat{\theta}_n - \theta| > \epsilon) = 0 \qquad \forall \epsilon > 0$.

X is a variable or a constant!

Theorem (Weak law of large numbers)

$lim_{n\to\infty} \Pr(|\bar{X}_n - \mu| < \varepsilon) = 1$

convergence in distribution

Definition

X_n converge in distribution/law to X:
$lim_{n\to\infty} \Pr(X_n < a) = \Pr(X < a) \qquad \forall a$

Notation

$X_n \xrightarrow{\mathcal{L}} X$

χ^2 distribution:

Definition (Chi square law)

If $X_i iid \sim \mathcal{N}(0,1)$ and $Q = \sum_{i=1}^k X_i^2$ then $Q \sim \chi_k^2$

Proposition (Convergence of chi 2)

$lim_{k\to\infty}\chi_k^2 \sim \mathcal{N}(k, 2k)$

Student distribution:

Definition (Student distribution:)

If $X \sim \mathcal{N}(0,1)$ and $Y \sim \chi^2(k)$ then $Z = \frac{X}{\sqrt{Y/\nu}} \sim t(k)$. (provided X and Y are independant):

Proposition (Convergence of student dis)

$lim_{k \rightarrow \infty} t(k) \sim \mathcal{N}(0,1)$

Fisher distribution:

Definition (Fisher distribution:)

If $X \sim \chi^2_k$ and $Y \sim \chi^2(l)$ then $\frac{X/k}{Y/l} \sim F_{k,l}$. (provided X and Y are independant):

central limit theorem

Theorem (central limit theorem)

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

Outline

Often, we are interested in estimating some values:

- Some usual numbers
  - Number of inhabitants
  - Proportion of right-hand writers
- Parameters of a distribution
  - Expectation
  - Variance
  - Min or max
  - Median or any quantile
- Parameters of a model

Estimation and inference

But with estimation we will always obtain a value!

- Does this value have a sens?
- How sure am I about this value?
- How is this value affected if I add one observation?

Estimation and inference

Statistical philosophy:
consider the experiment / sample as a result of random variables.

So our estimation is also random!!

Estimation and inference

So we now define an estimator:

Notation

We write the true value θ

Definition

An estimator is a function of the observed values.
$\hat{\theta} = f(x_1, x_2, \ldots, x_n)$

Definition

The value obtained from the estimator is called an estimate and is written: $\hat{\theta}$

Estimator

As the observations are considered as random variables, the estimator is also a random variable, it has furthermore:

- A distribution
- An expectation
- A variance

Outline

Estimation and inference

If we know or assume the distribution of the data, it is possible to derive the distribution of our estimator.

Example

Take the estimator of the sum: $\theta = \sum x_i$ and assume X follows a Poisson distribution: $x_i \sim \mathcal{P}(\lambda)$. As $X + Y$ has still a Poisson distribution, $\hat{\theta} \sim \mathcal{P}(n\lambda)$

Without knowledge or assumption about the observations, how to know the distribution of the estimate, as we have only one realization of it?

- Asymptotic theory: when $n \to \infty$
- Bootstrap

A (prematurated) intro to bootstrap

Bootstrap principle: As you consider the observations as random, try to have another set of these observations.

- Natural sciences: make same experiment again
- Economics: resample

Resampling:

1. Compute $\hat{\theta}$ on $X = \{x_i, x_2, \ldots, x_n\}$
2. Resample with replacement: obtain new $X^* = \{x_i^*, x_2^*, \ldots, x_n^*\}$
3. Compute new $\hat{\theta}^*$ on X^*
4. Repeat 2 and 3 10000 times and obtain 10000 $\hat{\theta}^{i*}$
5. Operation 4 gives you the *bootstrap distribution*

Bias of an estimator

We will see first a propriety related to the expectation of the estimator:

Definition

$Bias(\hat{\theta}) \equiv E[\hat{\theta}] - \theta$

Definition

θ is an unbiased estimator: $\Leftrightarrow Bias(\hat{\theta}) = 0 \Leftrightarrow E[\hat{\theta}] - \theta$

Efficiency 1

After having spoken about the expectation of the estimator, we speak about its variance:

Definition

Let $\hat{\theta}_1$ and $\hat{\theta}_2$ be two **unbiased** estimators. $\hat{\theta}_1$ is relatively more efficient than $\hat{\theta}_2$ if $\mathrm{Var}(\hat{\theta}_1) < \mathrm{Var}(\hat{\theta}_1)$

Efficiency 2: Cramer-Rao Bound

The variance of any estimator has a lower bound: it can't be lower than a quantity obtained from the Cramer-Rao Bound.

Proposition

$\text{Var}\left(\widehat{\theta}\right) \geq \frac{1}{\mathcal{I}(\theta)}$

Definition

An **unbiased** estimator $\hat{\theta}$ is efficient if it reaches the Cramer-Rao boud, ie: $\text{Var}(\hat{\theta}) = \frac{1}{\mathcal{I}(\theta)}$

That is, no **unbiased** estimator can have a smaller variance!

Remember:

> *That is, no* ***unbiased*** *estimator can have a smaller variance!*

But it is possible that a biased estimator has a lower variance!

So how to compare?

Mean square error MSE

Define the Mean square error (MSE):

Definition

$$\text{MSE}(\hat{\theta}|\theta) = \mathbb{E}\left((\hat{\theta}-\theta)^2\right)$$

It can be better understood/interpreted from:

Proposition

$$\text{MSE}(\hat{\theta}|\theta) = \textit{Bias}(\hat{\theta})^2 + \text{Var}(\hat{\theta})$$

Hence it allows to compare biased estimators.

Robustness

What happens if we add some new observations, potentially false/extrem (outliers)?

Definition

An estimator is robust if it is not attracted by extreme values.

compare: mean and median

Outline

All the proprieties were discussed based on the finite sample/exact distribution.
What happens if my sample is growing? Do I get to the true value if I have more and more informations?

Study the asymptotic proprieties: $n \to \infty$

Consistency 1

Remember:

Definition

X_n converges to X if $\lim_{n \to \infty} p(|\hat{\theta}_n - \theta| > \epsilon) = 0 \qquad \forall \epsilon > 0$.

X can be a random variable or a constant.
We are interested here in the convergence to the true value (constant).

Definition

$\hat{\theta}_n$ is convergent/consistent if
$\lim_{n \to \infty} p(|\hat{\theta}_n - \theta| > \epsilon) = 0 \qquad \forall \epsilon > 0$.

Consistency 2

Notation

$\hat{\theta} \xrightarrow{p} \theta$ *or* $\operatorname{plim} \hat{\theta} = \theta$

Relation between unbiasedness and consistency of an estimator?

No one!
Estimator can be:

- Consistent and unbiased (best!)
- Consistent but biased (often)
- Unbiased but not consistent

Consistency 3

Nevertheless:

Proposition

$$\left\{ \mathrm{E}[\hat{\theta}] = \theta \quad \textbf{et} \quad \lim_{n\to\infty} \mathrm{Var}[\hat{\theta}] = 0 \right\} \Rightarrow \hat{\theta} \xrightarrow{p} \theta$$

Recall that: $\mathrm{MSE}(\hat{\theta}|\theta) = Bias(\hat{\theta})^2 + \mathrm{Var}(\hat{\theta})$

Proposition

$$\lim_{n\to\infty} MSE(\hat{\theta}|\theta) = 0 \Rightarrow \hat{\theta} \xrightarrow{p} \theta$$

Often, the finite sample distribution of the estimator is unknown, unless we make assumptions about the distribution of the observations.
But we can know sometimes its asymptotic distribution!

Definition

$F_n \xrightarrow{d} F$

Example

By the central limit theorem: $\bar{X} \xrightarrow{d} \mathcal{N}(0, \frac{\sigma^2}{n})$

This result is independant on the distribution of the observations (no assumption needed).

Outline

We will see how to apply this:

- Variance
- Expectation

Compare the properties of two estimators of var 1/n and 1/n-1

- -bias: S^2_{n-1} unbiased
- -var: S^2_n has a smaller variance
- -MSE: S^2_n has a smaller MSE
- -Consistency: both are convergent!
- -Distribution: $(n-1)\frac{s^2}{\sigma^2} \sim \chi^2_{n-1}$

Convergence of estimator for variance

Theorem

$s_n^2 \xrightarrow{p} \sigma^2$

Proof.

Rewrite: $s_n^2 = \frac{1}{n}\sum_{i=1}^n (y_i - \overline{y})^2 = \left(\frac{1}{n}\sum_{i=1}^n y_i^2\right) - \overline{y}^2$

Then study:

- $\overline{y}^2 \xrightarrow{p} \mu^2$ by Slutzky theorem
- $\frac{1}{n}\sum_{i=1}^n (y_i - \overline{y})^2 \xrightarrow{p} \mathrm{E}[x^2] = \mu^2 + \sigma^2$

So $s_n^2 \xrightarrow{p} \mu^2 + \sigma^2 - \mu^2 = \sigma^2$

□

Study the properties of the mean as estimator of the expected value:

- -unbiased
- -not robust
- -convergent by law of large numbers
- -asymptotically normal by central limit theorem

Outline

$$\left[\overline{X} - 2\frac{\sigma(X)}{\sqrt{n}}; \overline{X} + 2\frac{\sigma(X)}{\sqrt{n}}\right]$$

Exercises

Show that $E(s_n^2) = \frac{n-1}{n} V(X)$
Idea: use that $\frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{x})^2$ can be rewritten as: (try also to prove it!)
$= \left(\frac{1}{n} \sum_{i=1}^{n} x_i^2\right) - \overline{x}^2$

Exercise 2: Show that $E(s_n^2) = V(X)$ if μ is known!
Idea: same as before, but starting from: $\frac{1}{n} \sum_{i=1}^{n} (x_i - \mu)^2$ and μ is here a constant!

You will always need for that: recall $\text{Var}(X) = E[X^2] - E[X]^2$)

Exercises 2

Defining:

$Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$

(recall that Z then follows asymptotically a standard normal!)

We want to obtain a confidence interval of the form:

$P(-b \leq \mu \leq b) = 1 - \alpha = 0.95.$

And the result is:

$P\left(\bar{X} - a\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + a\frac{\sigma}{\sqrt{n}}\right)$

where $a = \Phi^{-1}(0.975) = 1.96$,

Why? Show it!