COSSAN TRAINING COURSE
on UNCERTAINTY QUANTIFICATION

Modelling of random phenomena: multivariate distributions

Edoardo Patelli

info@cossan.co.uk
www.cossan.co.uk

April 2020
Outline

1. Introduction

2. Statistical inference
   - Overview
   - Method of moments
   - Maximum Likelihood

3. Summary
Edoardo Patelli
Professor in risk and uncertainty quantification, University of Strathclyde, UK

- Project leader of the “COSSAN project”
- Head of Centre for Intelligent Infrastructure, Department of Civil and Environmental Engineering, University of Strathclyde
Pre-requisites

- Modelling of random phenomena: Random variables
- Modelling of random phenomena: main distribution type
- Modelling of random phenomena: multivariate distribution
Outline

1. Introduction

2. Statistical inference
   - Overview
   - Method of moments
   - Maximum Likelihood

3. Summary
Statistical Inference

Requires real-world data

- Speed of the wind (for the resistant design of a high-rise building)
- Number of vehicles (for design a intersection of roads)
- Rainfall intensities, flood levels, earthquake magnitudes...
- Laboratory data such as: strength of concrete, yield of steel, fatigue life...

Calculation of a single number, from a set of observational data, to represent the parameter of the underlying population
Point Estimation

Calculation of a single number, from a set of observational data, to represent the parameter of the underlying population $\hat{\theta}$ as the estimator of the parameter $\theta$ which ideally is:

1. **Unbiased**
2. **Consistent**
3. **Efficient**
4. **Sufficient**

It is seldom possible for an estimator to have all properties.
Point Estimation

Calculation of a single number, from a set of observational data, to represent the parameter of the underlying population $\hat{\theta}$ as the estimator of the parameter $\theta$ which ideally is:

1. Unbiased
2. Consistent
3. Efficient
4. Sufficient

It is seldom possible for an estimator to have all properties.
Point Estimation

Properties

1. **Unbiasedness**: \( \mathbb{E}[\hat{\theta}] = \theta \) (average value of the estimator, if repeated estimates are made will be equal to the parameter)

2. **Consistency**: \( \hat{\theta} \to \theta \) as \( n \to \infty \) (the error in the estimator decreases as the sample size increases)

3. **Efficiency**: \( \hat{\theta}_1 \) is more efficient than \( \hat{\theta}_2 \) if \( \text{Var}(\hat{\theta}_1) < \text{Var}(\hat{\theta}_2) \)

4. **Sufficiency**: The estimator utilizes all the information in a sample that is pertinent to the estimation of the parameter.
Parameter Estimation

Statistical inference
Calculation of parameter of a distribution from a set of observational data

- The data represents a sample from that population
- The value of the parameter evaluated on the basis of the sample values is an estimate of the true parameter.

Tools:
1. Method of moments
2. Maximum likelihood
Method of Moments

Technique for constructing estimators of the parameters based on matching the sample moments with the corresponding distribution moments.
Sample moments

Given a random sample of size $n$, with sample values $x_1, \ldots, x_n$,

- the **sample mean** $\bar{x}$ is defined as

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

- the **sample variance** $\bar{s}^2$ is defined as

$$\bar{s}^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$$
Sample moments

Given a random sample of size $n$, with sample values $x_1, \ldots, x_n$,

- the **sample mean** $\bar{x}$ is defined as
  
  $$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

- the **sample variance** $s^2$ is defined as
  
  $$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$$
Sample moments

Example 1: estimation of sample moments

Compute the mean and standard deviation estimation of the following set of samples:
\{4.52, 5.35, 6.02, 5.89, 5.03, 6.48, 6.30, 5.43, 7.88, 5.64, 6.85, 6.48, 4.64, 5.88, 6.88, 7.43, 4.06, 5.76, 5.89, 6.86\}

Results

\[ \bar{x} = 5.9635 \]
\[ s^2 = 0.9504 \]
Sample moments

Example 1: estimation of sample moments

Compute the mean and standard deviation estimation of the following set of samples:
\{4.52, 5.35, 6.02, 5.89, 5.03, 6.48, 6.30, 5.43, 7.88, 5.64, 6.85, 6.48, 4.64, 5.88, 6.88, 7.43, 4.06, 5.76, 5.89, 6.86\}

Results

\[\bar{x} = 5.9635\]
\[s^2 = 0.9504\]
Method of Moments

Procedure

1. Equate the first sample moment about the origin
   \[ M_1 = \frac{1}{n} \sum_{i=1}^{n} x_i = \bar{x} \] to the first theoretical moment \( E[X] \)

2. Equate the second sample moment about the mean
   \[ \sigma^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2 = \bar{\sigma}^2 \] to the second theoretical moment \( E[(X - \mu)^2] \)

3. Use higher moments \( E[(X - \mu)^k], k = 3, 4, \ldots \) until you have as many equations as the number of parameters

4. Solve for the parameters
Method of Moments
Example 2: Normal Distribution

Assuming that the previous sample set was obtained from a normal distribution with unknown parameters, what are the parameters $\mu$ and $\sigma$ of $N(\mu, \sigma^2)$?
Method of Moments

Example 3: Uniform Distribution

Assuming that the previous sample set was obtained from a uniform distribution with unknown parameters, what are the parameter $a$ and $b$ of $U(a, b)$?
Method of Moments

Example 3: Uniform Distribution

\( X \sim U(a, b) \). PDF: \( \frac{1}{b-a} \) in \( a \leq x \leq b \)

First moment: \( \mu_1 = E(X) = \int_a^b \frac{x}{b-a} \, dx = \frac{a+b}{2} \)

Second moment about the mean:
\[
\sigma^2 = E[(X - \mu)^2] = \int_a^b \left( \frac{x}{b-a} - \frac{a+b}{2} \right)^2 \, dx = \frac{1}{12} (b - a)^2
\]

Equating the first two moments with the sample moments and solving for \( a \) and \( b \)
Method of Moments

Example 3: Uniform Distribution

\[ X \sim U(a, b) \]. PDF: \( \frac{1}{b-a} \) in \( a \leq x \leq b \)

First moment: \( \mu_1 = E(X) = \int_a^b \frac{x}{b-a} \, dx = \frac{a+b}{2} \)

Second moment about the mean:
\[
\sigma^2 = E[(X - \mu)^2] = \int_a^b \left( \frac{x}{b-a} - \frac{a+b}{2} \right)^2 \, dx = \frac{1}{12}(b - a)^2
\]

Equating the first two moments with the sample moments and solving for \( a \) and \( b \)
Method of Moments

Example 3: Uniform Distribution (alternative procedure)

\( X \sim U(a, b) \) PDF: \( \frac{1}{b-a} \) in \( a \leq x \leq b \)

First moment: \( \mu_1 = E(X) = \int_a^b \frac{x}{b-a} \, dx = \frac{a+b}{2} \)

Second moment: \( \mu_2 = E(X^2) = \int_a^b \frac{x^2}{b-a} \, dx = \frac{a^2+ab+b^2}{3} \)

Equating the first two moments with the sample moments and solving for \( a \) and \( b \)

\( \hat{a} = \hat{\mu}_1 - \sqrt{3(\hat{\mu}_2 - \hat{\mu}_1^2)} \quad \hat{b} = \hat{\mu}_1 + \sqrt{3(\hat{\mu}_2 - \hat{\mu}_1^2)} \)
Maximum Likelihood

Identify the most likely value of $\theta$ parameter of a distribution $f(x; \theta)$ based on a sample set with $n$ values.

For independent and identically distributed sample of size $n$ (with sample values $x_1, \ldots, x_n$) the joint density function is:

$$f(x_1, x_2, \ldots, x_n | \theta) = f(x_1 | \theta) f(x_2 | \theta), \ldots, f(x_n | \theta)$$

The likelihood function $\mathcal{L}(\cdot)$ is given by

$$\mathcal{L}(\theta | x_1, \ldots, x_n) = f(x_1, x_2, \ldots, x_n | \theta) = \prod_{i=1}^{n} f(x_i; \theta)$$
Maximum Likelihood

Identify the most likely value of $\theta$ parameter of a distribution $f(x; \theta)$ based on a sample set with $n$ values.

For independent and identically distributed sample of size $n$ (with sample values $x_1, \ldots, x_n$) the joint density function is:

$$f(x_1, x_2, \ldots, x_n|\theta) = f(x_1|\theta)f(x_2|\theta), \ldots, f(x_n|\theta)$$

The likelihood function $\mathcal{L}(\cdot)$ is given by

$$\mathcal{L}(\theta|x_1, \ldots, x_n) = f(x_1, x_2, \ldots, x_n|\theta) = \prod_{i=1}^{n} f(x_i; \theta)$$
Maximum Likelihood

Identify the most likely value of $\theta$ parameter of a distribution $f(x; \theta)$ based on a sample set with $n$ values.

For independent and identically distributed sample of size $n$ (with sample values $x_1, \ldots, x_n$) the joint density function is:

$$f(x_1, x_2, \ldots, x_n | \theta) = f(x_1 | \theta)f(x_2 | \theta), \ldots, f(x_n | \theta)$$

The likelihood function $L(\cdot)$ is given by

$$L(\theta | x_1, \ldots, x_n) = f(x_1, x_2, \ldots, x_n | \theta) = \prod_{i=1}^{n} f(x_i; \theta)$$
Maximum Likelihood estimator

\[ \hat{\theta} \text{ that maximizes } \mathcal{L}(\cdot) \]

Obtained by differentiating the likelihood function

\[
\frac{d\mathcal{L}(\theta|x_1, \ldots, x_n)}{d\theta} = 0
\]

Since the likelihood function is a product, we can work with the logarithm (log-likelihood)

\[
\ln \mathcal{L}(\theta|x_1, \ldots, x_n) = \sum_{i=1}^{n} \ln f(x_i; \theta)
\]

(logarithm is monotonic and does not affect the optimisation)
Maximum Likelihood estimator

\( \hat{\theta} \) that maximizes \( \mathcal{L}(\cdot) \)

Obtained by differentiating the likelihood function

\[
\frac{d\mathcal{L}(\theta|x_1, \ldots, x_n)}{d\theta} = 0
\]

Since the likelihood function is a product, we can work with the logarithm (log-likelihood)

\[
\ln \mathcal{L}(\theta|x_1, \ldots, x_n) = \sum_{i=1}^{n} \ln f(x_i; \theta)
\]

(logarithm is monotonic and does not affect the optimisation)
Maximum Log-Likelihood

\[ \frac{d \ln \mathcal{L}(\theta \mid x_1, \ldots, x_n)}{d\theta} = 0 \]

For PDFs (or PMFs) defined by more than one parameter,

\[ \mathcal{L}(\theta_1, \ldots, \theta_m \mid x_1, \ldots, x_n) = \prod_{i=1}^{n} f(x_i; \theta_1, \ldots, \theta_m) \]

\[ \frac{\partial \ln \mathcal{L}(\theta_1, \ldots, \theta_m \mid x_1, \ldots, x_n)}{\partial \theta_j} = 0 \]
Maximum Log-Likelihood

Example

A random sample set \( x_1, x_2, x_n \) where \( x_i = 0 \) represents a randomly selected student that do not pass a test and \( x_i = 1 \) represents a randomly selected student that do pass a test

Assuming that \( X \sim \text{Bernoulli}(p) \), find the maximum likelihood estimator of \( p \)
Maximum Log-Likelihood

Example: solution

The PMF of each $x_i$ is: $f(x_i, p) = p^{x_i}(1 - p)^{1-x_i}$ (from Bernoulli)

Hence,

$$
\mathcal{L}(p) = \prod_{i=1}^{n} f(x_i|p) = p^{\sum_i x_i}(1 - p)^{n-\sum_i x_i}
$$

The log-likelihood is:

$$
\ln \mathcal{L}(p) = \sum_i x_i \ln(p) + (n - \sum_i x_i) \ln(1 - p)
$$
Maximum Log-Likelihood

Example: solution

The PMF of each $x_i$ is: $f(x_i, p) = p^{x_i}(1 - p)^{1-x_i}$ (from Bernoulli)

Hence,

$$L(p) = \prod_{i=1}^{n} f(x_i|p) = p^{\sum_i x_i}(1 - p)^{n-\sum_i x_i}$$

The log-likelihood is:

$$\ln L(p) = \sum_i x_i \ln(p) + (n - \sum_i x_i) \ln(1 - p)$$
Maximum Log-Likelihood

Example: solution

The maximum of the log-likelihood is identified as:

\[
\frac{\ln \mathcal{L}(p)}{dp} = \sum_i x_i \frac{p}{p} - (n - \sum_i x_i) \frac{1 - p}{1 - p} = 0
\]

That becomes:

\[
\sum x_i - p \sum x_i - np + p \sum x_i = 0
\]

and finally:

\[
\hat{p} = \frac{\sum x_i}{n}
\]
Maximum Log-Likelihood

Example: solution

The maximum of the log-likelihood is identified as:

\[
\ln L(p) \frac{d}{dp} = \sum_i x_i \frac{p}{p} - (n - \sum_i x_i) \frac{1}{1-p} = 0
\]

That becomes:

\[
\sum_i x_i - p \sum_i x_i - np + p \sum_i x_i = 0
\]

and finally:

\[
\hat{p} = \frac{\sum x_i}{n}
\]
Maximum Log-Likelihood

Example: solution

The maximum of the log-likelihood is identified as:

\[
\ln \mathcal{L}(p) \frac{dp}{dp} = \sum_i x_i \frac{p}{p} - \frac{(n - \sum_i x_i)}{1 - p} = 0
\]

That becomes:

\[
\sum_i x_i - p \sum_i x_i - np + p \sum_i x_i = 0
\]

and finally:

\[
\hat{p} = \frac{\sum x_i}{n}
\]
Maximum Log-Likelihood

Take away message

Used to estimate the parameters of a distribution

- Select a distribution (a PDF or PDF).
- Compute the log of such distribution (Log-Likelihood)
- Identify the maximum of the log-likelihood
- Estimate the parameter(s) of the distribution
Outline

1. Introduction

2. Statistical inference
   - Overview
   - Method of moments
   - Maximum Likelihood

3. Summary
Wrapping up

Summary

Statistical inference
- Method of moments
- Maximum likelihood

What next
- Uncertainty characterisation with Cossan software