CONTENTS

CHAPTER	TOPIC	PAGE NUMBER
1	**Introduction to Python**	**1**
2	**Introduction to OpenAIUQ**	**1**
3	**Lecture O: Reliability, Risk and Resilience R^3**	**1**
4	**Lecture A1: Preliminary Data Analysis**	**1**
4.1	Dataset	1
4.2	Graphical Representation	5
4.2.1	Histogram	5
4.2.2	Quantiles	7
4.2.3	Cumulative relative frequency	10
4.2.4	Quantile plot	10
4.3	Numerical summaries of data	12
4.3.1	Measures of central tendency	15
	Sample mean	15
	Sample median	15
	Sample mode	16
4.3.2	Measures of dispersion	16
	Interquartile	16

Sample variance 16
Sample standard deviation 17
Sample coefficient of variation 17
4.3.3 Measures of asymmetry 17
Sample coefficient of skewness 17
Sample coefficient of kurtosis 18
4.4 Data Observed in pairs 21
4.4.1 Scatter Plot 21
4.4.2 Sample covariance 21
4.4.3 Sample linear coefficient of correlation 21

5 Lecture A1: Probability Theory 1
5.1 Random events 1
5.1.1 Sample space and events 1
5.1.2 Intersection and union of events 1
5.1.3 Venn diagram 1
5.2 Definition of Probability 2
5.2.1 Classical Definition 2
5.2.2 Frequentistic Definition 3
5.2.3 Bayesian Definition 4
5.3 Bayes Rule 4
5.4 Event Trees 5
5.5 Introduction to Bayesian Networks 5

6 Lecture A2: Random Variables 1
6.1 Random Variables 1
6.1.1 Discrete Random Variables 2
Probability Mass Function (PMF) 2
Cumulative Distribution Function (CDF) 3
6.1.2 Continuous Random Variables 4
Probability Density Function (PDF) 4

6.2 Descriptors of Random Variables 5
Moments 6
Mean or Expected Value 6
Variance 7
Standard deviation 7
Coefficient of variation 7
Higher-order moments 8
Measures of asymmetry: coefficient of skewness 9
Measures of "peak/tail": coefficient of kurtosis 10
6.2.1 Moment Generating function 11
6.2.2 Characteristic function 11
6.2.3 Cumulants 11
6.3 Transformation 11
6.3.1 Mean and variance 12
6.3.2 Density of a transformation 14

7 Lecture A2: Probability Distributions 1
7.1 Discrete distributions 2
7.1.1 Bernoulli distribution 2
7.1.2 Binomial distribution 2
7.1.3 Poisson distribution 2
7.2 Continuous distributions 2
7.2.1 Uniform distribution 2
7.2.2 Triangular distribution 2
7.2.3 Gaussian distribution 2
7.2.4 Lognormal distribution 2
7.2.5 Weibull distribution 2
7.2.6 Exponential distribution 2
7.3 Nonparametric distributions 2
7.3.1 Mixture of Gaussians 2
7.3.2 Kernel Density Estimation (KDE) 2

8 Lecture A2:Model Estimation **1**

8.1 Model Building 1

8.2 Selection of probability distribution function 1

8.3 Parameter estimation 2

8.3.1 Estimation of Distribution Parameters 4

8.3.2 Sampling variability 5

8.3.3 Estimators 6

8.3.4 Bias of an estimator 7

Sample Mean 7

Sample Variance 8

Corrected Sample Variance 8

8.3.5 Variance and standard error of an estimator 9

Standard Error of the Sample mean 9

8.3.6 Tradeoff bias-variance and Mean Square Error (MSE) of an estimator 11

8.3.7 Consistency 13

8.4 Confidence Intervals 13

8.5 Basics of Bootstrap 13

8.6 Analysis of Variance 13

8.7 Maximum Likelihood Estimation (MLE) 13

8.8 Basics of Information Theory and Minimum Divergence Estimation (MDE) 22

8.8.1 Shannon's Entropy 23

8.8.2 Differential Entropy 23

8.8.3 Relative Entropy (KL Divergence) 24

8.8.4 Minimum Divergence Estimation (MDE) 24

8.9 Method of the Moments 26

8.10 Maximum Entropy distribution 33

8.11 Kernel Density Maximum Entropy (KDME) 33

9 Lecture A3:Model Selection **1**

9.1 Model diagnostics 2

9.2 Graphical representation and visual inspection 2

9.2.1 Probability Plot 2
9.2.2 Q-Q Plot 4
9.2.3 Example 6
9.3 Information Theory for model selection 10
9.3.1 Tradeoff Bias-Variance 10
9.3.2 Akaike Information Criterion (AIC) for model selection 13
9.4 Basics of Machine Learning (ML) for model selection 14

10 Lecture A4:Multiple Random Variables 1
10.1 Multiple random variables 1
10.1.1 Conditional distributions 1
10.1.2 Independent random variables 1
10.1.3 Means and Moments 1
10.1.4 Covariances 1
10.1.5 Coefficient of Correlation 1
10.2 Multivariate Gaussian distribution 1
10.2.1 Multivariate Normal Standard distribution 5
10.2.2 Coordinate transformation toward normal standard PDF - Case of Uncorrelated Gaussian random variables 7
10.2.3 Coordinate transformation toward normal standard space - Case of correlated Gaussian random variables 9
Step 1. Coordinate transformation $\mathbf{x} \rightarrow \mathbf{z}$ 9
Step 2. Coordinate transformation $\mathbf{z} \rightarrow \mathbf{u}$ 11
Coordinate transformation $\mathbf{x} \rightarrow \mathbf{u}$ 14

11 Models of regression and multivariate analysis 1
11.1 Simple Linear Regression 1
11.2 Multiple Linear Regression 1
11.3 Multivariate Analysis 1

12 Basics of Stochastic Processes 1
12.1 Introduction 1

12.2 Statistical Indicators 1
12.3 Stationarity 1
12.4 Ergodicity 1
12.5 Power Spectral Density 1
12.6 Gaussian process 1
12.7 Barrier crossing problem 1

13 Analysis of Extreme events 1
13.1 Classical Extreme values: theory and methods 1
13.2 Threshld models 1
13.3 Dependent sequences 1
13.4 Extremes of non-stationary sequences 1
13.5 Multivariate extremes 1

14 Lecture C1: Bayesian Updating 1

15 Lecture B1:Structural Reliability 1
15.1 Introduction 1
15.2 Structural Reliability 2
15.2.1 Limit State 4
15.2.2 Joint PDF of the basic random variables $f_{\mathbf{X}}(\mathbf{x})$ 6
15.3 Isoprobabilistic Transformation 8

16 Lecture B1:Basics of MCS 1
16.1 Simulation of samples 1
16.1.1 One random variables 1
16.1.2 Multiple independent Random variables 2
16.2 Crude MCS 3
16.2.1 Computational cost and accuracy 7
16.3 Case of Gaussian uncorrelated 12
16.3.1 Original Space 13
16.3.2 Normal standard Space $\mathbf{u}$ 17

16.4 Case of Gaussian correlated 20
16.4.1 Original Space 22
16.4.2 Normal standard Space **u** 26

17 Lecture B2:Reliability Index 1
17.1 Introduction 1
17.2 Cornell Reliability index 2
17.2.1 Gaussian correlated 5
17.2.2 Geometric interpretation of Cornell Reliability index 6
17.2.3 Linear limit state in the normal standard space **u** 7
17.3 Hasofer-Lind Reliability index 9
17.3.1 Isoprobabilistic transformation 10
17.3.2 Design point 10
17.3.3 Reliability Index 12
17.3.4 First Order Reliability Method (FORM) 12
17.3.5 Design point, Optimization 13

18 Lecture B3:First Order Reliability Method (FORM) 1
18.1 Uncorrelated Gaussian 1
18.2 Correlated Gaussian 1
18.3 Uncorrelated non-Gaussian 1
18.4 Conditional methods 1
18.5 Nataf 1
18.6 Copula 1

19 Basics of Surrogate Modelling and Machine Learning for structural reliability 1
19.1 Second-Order Reliability Method 1
19.2 Response Surface Method 1
19.3 High-Dimensional Model Representation (HDMR) 1
19.4 Polynomial Chaos Expansion (PCE) 1
19.5 Support Vector Machines (SVM) 1

19.6 Gaussian Process and Kriging (GP) 1
19.7 Geometry of high-dimensional spaces 1
19.8 Secant Hyperplane Method 1

20 Lecture B4:System Reliability 1
20.1 series system 2
20.2 parallel system 2
20.3 general system 2
20.4 Reliability evaluation of series systems 2
20.4.1 MCS 2
20.4.2 FORM approximation 2
20.4.3 Simple Bounds 2
20.4.4 Ditlevsen Bounds 2
20.5 Reliability evaluation of parallel systems 2
20.5.1 MCS 2
20.5.2 FORM approximation 2
20.5.3 Simple Bounds 2
20.5.4 Ditlevsen Bounds 2
20.6 Event tree and fault tree 2
20.7 Bayesian Networks 2

21 Advanced Monte Carlo Simulation 1
21.1 Latin Hypercube Sampling (LHS) 1
21.2 Importance Sampling 1
21.3 Subset Simulation 1

22 Machine Learning 1
22.1 Introduction 1
22.2 Classification 1
22.3 Regression 1
22.4 Support Vector Machines 1
22.5 Decision Trees 1

22.6 Ensemble Learning and Random Forests 1
22.7 Dimensionality Reduction 1
22.8 Unsupervised Machine Learning 1

23 Neural Networks and Deep Learning 1
23.1 Introduction to Artificial Neural Networks 1
23.2 Training Deep Neural Networks 1
23.3 Computer Vision and Convolutional Neural Networks (CNN) 1
23.4 Recurrent Neural Networks (RNN) 1
23.5 Representation Learning and Generative Learning 1

24 Lecture C2:Basics of Decision Theory 1
24.1 Decision Theory 1
24.2 Decision with given information - prior analysis 1
24.3 Decision with additional information - posterior analysis 1
24.4 Decision with unknown information - pre-posterior analysis 1
24.5 Value of Information 1
24.6 Bayesian Networks 1

25 Reinforcement Learning 1

Chapter 1

Introduction to Python

Chapter 2

Introduction to OpenAIUQ

Chapter 3

Lecture O: Reliability, Risk and Resilience R^3

Chapter 4

Lecture A1: Preliminary Data Analysis

4.1 Dataset

To cope with uncertainty, first step is to obtain and investigate a *sample of data* and determine their range of definition, including *minimum* and *maximum* value of the data.

OpenAIUQ: for univariate data analysis we adopt the class *data1*. At first, we associate a *instance* to the object. The attributes of the class are shown in fig.4.1, and the methods in figs.4.2 and 4.3.

data1 – Attributes	
data	dataset
datamin	minimum
datamax	maximum
r	range
median	median
mean	sample mean
std	sample standard deviation
cov	sample coefficient of variation
g1	sample skewness
g2	sample kurtosis

data1 – Attributes	
Q	sample quantile
Q25	sample quantile 25%
Q50	sample quantile 50%
Q75	sample quantile 75%
iqr	interquartile
prob	sample probability

Figure 4.1: OpenAIUQ: class *data1*, attributes

data1 – Methods			
Method	**Description**	**Input**	**Output**
get_datamin	get the minimum of the data		self.datamin
get_datamax	get the maximum of the data		self.datamax
get_r	get the range		self.r
get_median	get the median		self.median
get_mean	get the sample mean		self.mean
get_std	get the sample standard deviation		self.std
get_cov	get the sample coefficient of variation		self.cov
get_skewness	get the sample skewness		self.g1
get_kurtosis	get the sample kurtosis		self.g2

Figure 4.2: OpenAIUQ: class *data1*, methods

data1 – Methods			
Method	**Description**	**Input**	**Output**
get_quantile	get the quantile of the data	p	self.Q
get_iqr	get the interquartile of the data		self.iqr
get_prob	get some probabilities of the sample	lambda0 method='**lower**, upper'	self.prob
set_range	set the domain of interest of the data	xiniz=0.80*datamin xfin=1.20*datamax size=100	self.xx
get_cdf_emp	get the empirical cdf:		self.xemp self.Femp
get_quantile_emp	get the empirical quantile function		self.Qemp
plot_cdf_emp	plot the empirical cdf	fignum figsize figdpi	figure
plot_quantile_emp	plot the quantile plot	fignum figsize figdpi	figure
get_summary	get the main statistics of the data		all the attributes

Figure 4.3: OpenAIUQ: class *data1*, methods (contd)

Case test 1, concrete20. Let us imagine now to have 20 data of concrete strength. The data present some kind of variability, see fig.4.4. This is expected, because the concrete has some natural variability.

Concrete strengths (Mpa)		
Data	Unordered	Ordered
1	35.8	24.4
2	39.2	27.6
3	34.6	27.8
4	27.6	27.9
5	37.1	28.5
6	33.3	30.1
7	32.8	30.3
8	34.1	31.7
9	27.9	32.2
10	24.4	32.8
11	27.8	33.3
12	33.5	33.5
13	35.9	34.1
14	39.7	34.6
15	28.5	35.8
16	30.3	35.9
17	31.7	36.8
18	32.2	37.1
19	36.8	39.2
20	30.1	39.7

Figure 4.4: Concrete strength, 20 data

First step is to detect the *minimum value* and the *maximum value* of the sample of data.

OpenAIUQ

- import the dataset from the file 'concrete20.dat'
- define d as instance of the class *auq.data1*
- minimum, maximum and range of the data are stored in the attributes *datamin*, *datamax* and r

```
#=======================================
#CONCRETE 20
#=======================================
dataset=np.loadtxt('concrete20.dat')

#d is the object collecting the dataset
d=auq.data1(dataset)

print('Data analysis')
print('min value: {:.3}'.format(d.datamin))
print('max value: {:.3}'.format(d.datamax))
print('range: {:.3}'.format(d.r))


----------------------------------
Data analysis
min value: 24.4
max value: 39.7
range: 15.3
----------------------------------
```

Therefore, in the *case test 1* we have

$$\begin{aligned} x_{min} &= 24.40\ MPa \\ x_{man} &= 39.70\ MPa \end{aligned} \tag{4.1}$$

while range of the dataset is

$$r = x_{max} - x_{min} = 39.70 - 24.40 = 15.30\ MPa \tag{4.2}$$

4.2 Graphical Representation

4.2.1 Histogram

- Define minimum value x_L and maximum value x_U of the bins of the histogram
- Choose the number of classes n_c (or equivalently, choose the width of the bins)
- Evaluate the number of observations in each class (or frequency)
- Evaluate the *relative frequency* of each class
- Evaluate the cumulative relative frequency

The number of classes n_c plays a key role in the histogram.

- Too few classes: omission of some important features of the data
- Too many classes: there may be high fluctuactions in the frequencies

Some rules of thumb to define the number of classes:

$$n_c = \sqrt{n}, \quad 5 \leq n_c \leq 25 \tag{4.3}$$

$$n_c = 1 + 3.3 \log_{10} n \tag{4.4}$$

Case test 1. Since in the dataset $x_{min} = 24.4$ and $x_{max} = 39.7$, given the sampling uncertainty, we choose

$$\begin{aligned} x_L &= 23 \ MPa \\ x_U &= 41 \ MPa \\ r &= x_U - x_L = 18 \ MPa \end{aligned} \tag{4.5}$$

This is a first guess. After observation of the data, we can keep these limits or changing them.

Number of classes:

$$n_c = \sqrt{n} = 4.47, \tag{4.6}$$

$$n_c = 1 + 3.3 \log_{10} n = 5.29 \approx 6 \tag{4.7}$$

and we choose $n_c = 6$ with

$$\Delta x = \frac{x_U - x_L}{n_c} = \frac{18}{6} = 3 \ MPa \tag{4.8}$$

OpenAIUQ:

Interval	Midpoint	Observations	Frequency	Cumulative Frequency
23-26	24.5	1	0.05	0.05
26-29	27.5	4	0.20	0.25
29-32	30.5	3	0.15	0.40
32-35	33.5	6	0.30	0.70
35-38	36.5	4	0.20	0.90
38-41	39.5	2	0.10	1.00

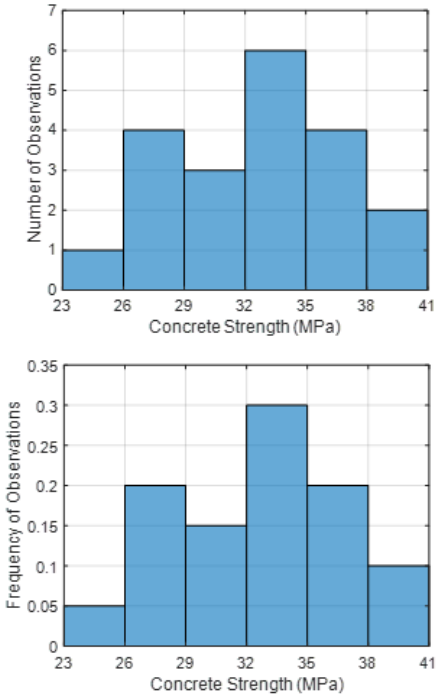


Figure 4.5: Concrete strength, 20 data, histogram

- choose a bin
- build the histograms applied to the dataset collected in the attribute *data* of the instance *d* of the object *auq.data1*

```
#Choose binning
bin1=[23,26,29,32,35,38,41]

fig1=plt.figure(num=1,figsize=(6,6),dpi=60)
#num: figure number
#figsize: figure dimension
#dpi

ax=fig1.add_subplot(1,1,1)
#plot=1 row, 1 column

n,bin,patches=ax.hist(d.data, bins=bin1,    histtype='bar',facecolor='blue',
    density=False, alpha=0.6, rwidth=0.95)
#bins: limits of the bins
#facecolor: color of the bar
```

```
#density: False, True
#alpha: transparency
#rwidth: distance between histograms

ax.set_title('Histogram $\Delta=3 MPa$')
ax.set_xlabel('Concrete Strength (MPa)')
ax.set_ylabel('Number of observations')
ax.grid()
ax.set_xlim(23,41)
ax.set_xticks(bin1);
```

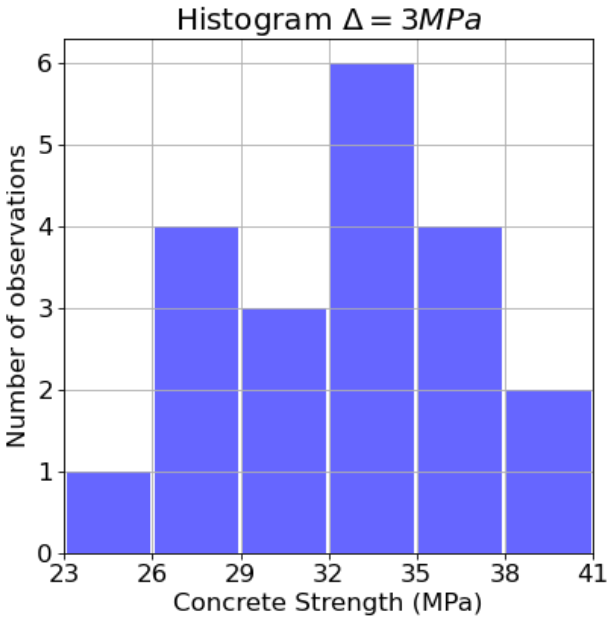


Figure 4.6: Concrete strength, 20 data, histogram, OpenAIUQ

4.2.2 Quantiles

Quantile is the p-th percentile (or quantile) is the value that is greater than p percent of the observations

If x_{65} is the quantile 0.65 (or percentile 65%) of a sample of data, this means that the 65% of the data are lower than x_{65}.

- We sort the data in ascending order

- In correspondence to the $i-$th data, we determine the $i-$th frequency

$$Q_i = \frac{i - 0.5}{n} \tag{4.9}$$

- we evaluate the quantile corresponding to the chosen frequency

Case test 1. For example, the data $i = 1$ and $i = 5$ are the lowest and the fifth lowest of the $n = 20$ ordered dataset, their frequencies are:

$$\begin{aligned} p_1 &= \frac{1 - 0.5}{20} = 0.025\ MPa \\ p_5 &= \frac{5 - 0.5}{20} = 0.225\ MPa \end{aligned} \tag{4.10}$$

Median Q_2: the middle terms of a set of data when the values are arranged in ascending order. If n is an even number, it is equal to the average of the two middle terms

Case test 1. We have $n = 20$ data (even number), therefore:

$$Median = \frac{x_{10} + x_{11}}{2} = \frac{32.8 + 33.3}{2} = 33.05\ MPa \tag{4.11}$$

Median Q_2: is the quantile 50% of the data

Lower quartile Q_1: the median of the lower half of the data, which is the 25% quantile

Upper quartile Q_3: the median of the upper half of the data, which is the 75% quantile

Concrete strengths (Mpa)		
Data	Ordered	Q
1	24.4	0.025
2	27.6	0.075
3	27.8	0.125
4	27.9	0.175
5	28.5	0.225
6	30.1	0.275
7	30.3	0.325
8	31.7	0.375
9	32.2	0.425
10	32.8	0.475
11	33.3	0.525
12	33.5	0.575
13	34.1	0.625
14	34.6	0.675
15	35.8	0.725
16	35.9	0.775
17	36.8	0.825
18	37.1	0.875
19	39.2	0.925
20	39.7	0.975

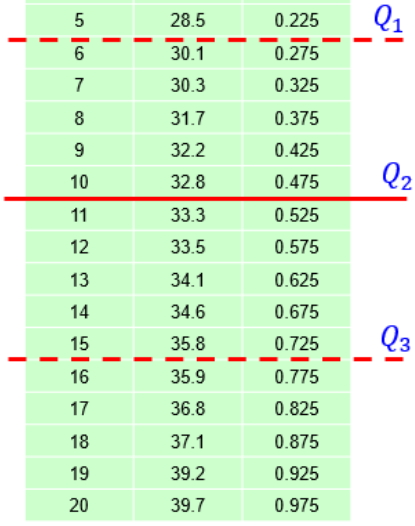


Figure 4.7: Concrete strength, 20 data, quantiles

Case test 1.

$$\begin{aligned} Q_1 &= \frac{x_5 + x_6}{2} = \frac{28.5 + 30.1}{2} = 29.30\ MPa \\ Q_2 &= \frac{x_{10} + x_{11}}{2} = \frac{32.8 + 33.3}{2} = 33.05\ MPa \\ Q_3 &= \frac{x_{15} + x_{16}}{2} = \frac{35.8 + 35.9}{2} = 35.85\ MPa \end{aligned} \tag{4.12}$$

Interquartile range iqr is defined as:

$$iqr = Q_3 - Q_1 \tag{4.13}$$

Case test 1.

$$iqr = Q_3 - Q_1 = 35.85 - 29.30 = 6.55 \tag{4.14}$$

4.2.3 Cumulative relative frequency

It is called also *empirical cdf*, in the plot we represent

- the quantiles in the $x-$axis
- the frequencies in the $y-$axis

OpenAIUQ:

- set the range of definition of the data ($x_L = 23\ MPa$ and $x_U = 41\ MPa$) through the method *set_range*
- build the empirical CDF (attributes *xemp* and *Femp*) through the method *get_cdf_emp*
- plot the empirical CDF through the method *plot_cdf_emp*

```
d.set_range(xiniz=23,xfin=41)
#set the range of interest of the variable

d.get_cdf_emp()
#evaluate the empirical cdf

d.plot_cdf_emp(fignum=2,figsize=(6,6),figdpi=60)
#plot the empirical cdf
```

4.2.4 Quantile plot

In the quantile plot we represent, see fig. 4.9

- the frequencies in the $x-$axis
- the quantiles in the $y-$axis

OpenAIUQ: represent the quantile plot through the method *plot_quantile_emp*; the method adopts the attributes *xemp* and *Femp* previously determined through the method *get_cdf_emp*

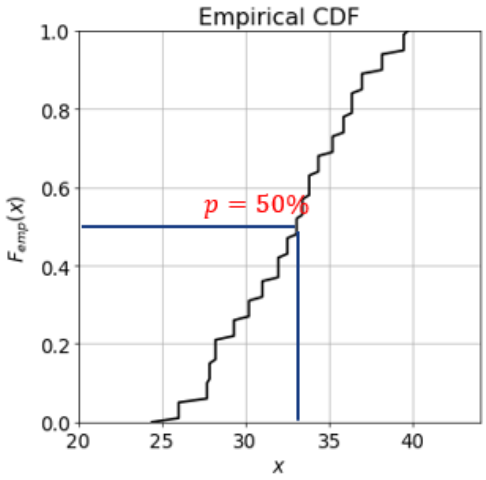


Figure 4.8: Concrete strength, 20 data, empirical CDF

```
d.plot_quantile_emp(fignum=3,figsize=(6,6),figdpi=60)
#plot the empirical quantile function
```

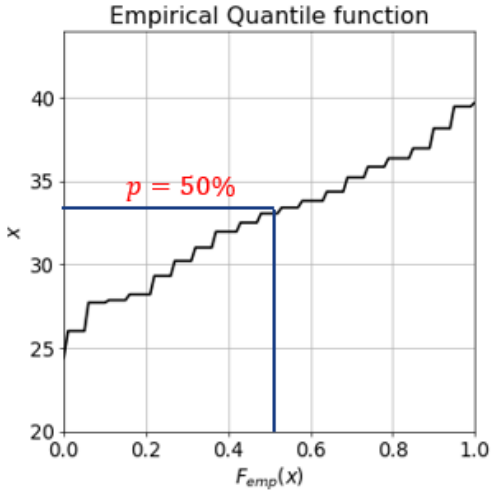


Figure 4.9: Concrete strength, 20 data, quantile plot

4.3 Numerical summaries of data

Case test 2. Let us consider now a set of 164 data of timber strength. The data present some kind of variability, see fig.4.10. This is expected, because wood is a natural material.

Timber strength [MPa]									
48.78	32.02	45.54	32.40	48.37	50.98	35.58	40.53	29.11	65.35
41.64	39.34	34.12	33.06	29.93	40.71	28.97	47.25	65.61	45.19
39.77	46.33	45.92	33.47	36.38	34.63	34.56	32.68	37.78	70.22
35.89	46.99	36.47	35.67	46.86	24.84	26.89	43.26	43.33	41.75
54.04	22.67	28.98	28.46	36.00	28.83	38.64	47.61	53.63	37.51
35.43	39.62	40.85	23.16	23.19	42.31	24.25	28.13	41.85	31.60
22.75	44.78	56.60	44.51	36.88	39.33	44.54	32.48	33.19	37.65
44.78	26.63	28.76	42.47	44.30	39.93	40.85	36.81	39.15	28.00
43.99	43.48	47.42	48.39	44.59	39.60	39.97	35.88	54.71	46.01
47.74	30.05	33.61	38.05	44.00	38.16	37.69	33.92	43.64	43.48
25.39	30.33	44.36	35.03	40.39	43.33	41.78	57.99	56.80	40.27
38.00	39.21	35.30	31.33	41.72	69.07	33.14	49.57	43.07	39.05
25.98	51.39	33.18	27.31	29.90	51.90	55.23	40.20	43.12	32.76
36.84	50.91	36.85	53.99	35.17	33.71	36.53	49.59	30.02	45.97
34.49	49.65	17.98	43.41	34.44	46.50	22.74	32.03	38.81	23.14
38.71	47.83	27.90	28.71	27.93	36.92	34.40	39.20	24.09	53.00
30.53	44.07	44.36	58.34						

Figure 4.10: Timber strength, 164 data

First step is to detect the *minimum value* and the maximum value of the sample of data. In such case we have

OpenAIUQ : the data are collected in the datafile 'timber164.dat'

```
#=======================================
#TIMBER 164
#=======================================
dataset=np.genfromtxt('timber164.dat')

#d is the object collecting the dataset
d=auq.data1(dataset)
data=d.data

print('Data analysis')
```

```
print('min value: {:.4}'.format(d.datamin))
print('max value: {:.4}'.format(d.datamax))
print('range: {:.4}'.format(d.r))
```

```
----------------------------------------
Data analysis
min value: 17.98
max value: 70.22
range: 52.24
----------------------------------------
```

Given the sampling uncertainty, we choose

$$\begin{aligned} x_0 &= 0\ MPa \\ x_n &= 80\ MPa \\ r &= x_n - x_0 = 80\ MPa \end{aligned} \tag{4.15}$$

Let us analyze two different binnings of the dataset

OpenAIUQ: in this case, we build a figure *fig* having two subplots. The two subplots are histograms with different binning. The plot has one row and two columns. The two subplots are detected respected by the object *axes* $ax1$ and $ax2$.

```
#Choose binning
bin1=np.arange(0,88,8)
bin2=np.arange(0,85,5)

fig1=plt.figure(num=1,figsize=(12,6),dpi=60)
#num: figure number
#figsize: figure dimension
#dpi

ax1=fig1.add_subplot(1,2,1)
ax2=fig1.add_subplot(1,2,2)
#plot=1 row, 2 columns

n,bin,patches=ax1.hist(d.data, bins=bin1, histtype='bar',facecolor='blue',
    density=False, alpha=0.6, rwidth=0.95)
n,bin,patches=ax2.hist(d.data, bins=bin2, histtype='bar',facecolor='blue',
    density=False, alpha=0.6, rwidth=0.95)
#bins: limits of the bins
```

```
#facecolor: color of the bar
#density: False, True
#alpha: transparency
#rwidth: distance between histograms

ax1.set_title('Histogram $\Delta=8 MPa$')
ax1.set_xlabel('Timber Strength (MPa)')
ax1.set_ylabel('Number of observations')
ax1.grid(True)
ax1.set_xlim(0,80)
ax1.set_xticks(bin1);

ax2.set_title('Histogram $\Delta=5 MPa$')
ax2.set_xlabel('Timber Strength (MPa)')
ax2.set_ylabel('Number of observations')
ax2.grid(True)
ax2.set_xlim(0,80)
ax2.set_xticks(np.arange(0,90,10));
```

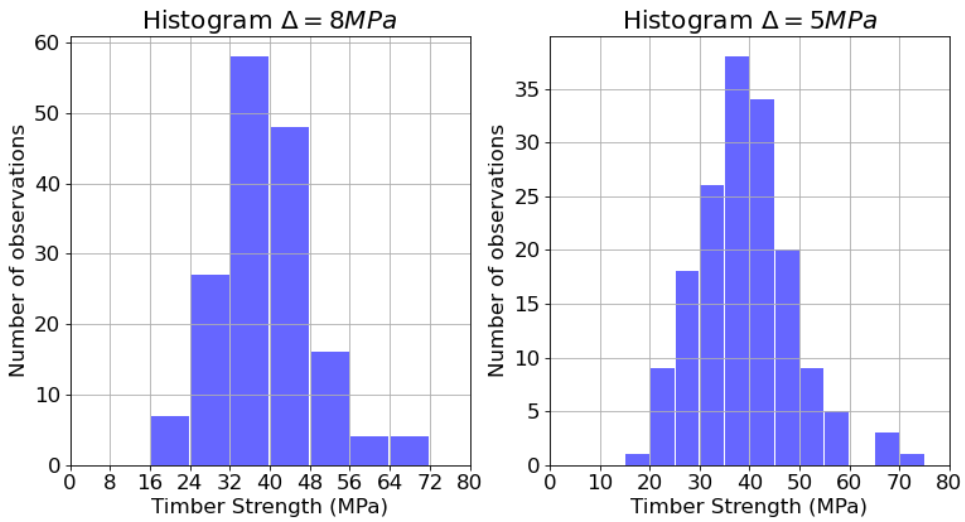


Figure 4.11: Timber strength, 164 data

The information conveyed by the histograms with different bin size looks similar. The diagrams are almost symmetrical with a peak in the class below 40 MPa and a decrease on both sides

4.3.1 Measures of central tendency

Sample mean

Generally the data tend to cluster around some values of variables. The sample mean is the arithmetic mean of the data:

$$m = \frac{1}{n}\sum_{i=1}^{n} x^{(i)} \tag{4.16}$$

The sample mean may sometimes be affected by unexpectedly high or low values, called outliers.

The outliers can be attributed to:

- the conditions are changed with respect to our assumption
- the data are generated by more than one process
- errors of instrumentation, measurement, observation or recording

The engineer needs to check if:

- outliers are erroneous
- outliers need to be included

Sample median

The sample median is the middle term of a set of data when the values are arranged in ascending order. If n is an even number, the median is equal to the average of the two midlle terms.

The median coincides with the quantile 50% Q_2 of the set of data

The sample median is robust to the outliers

Sample mode

The sample mode is the value that occurs most frequently, it is the value most likely to occur. It may be detected by the histogram. However, since the histograms are affected by the choice of the number of bin, often the evaluation of the mode is challenging from the knowledge of the data only.

4.3.2 Measures of dispersion

Interquartile

Interquartile range iqr is defined as:

$$iqr = Q_3 - Q_1 \tag{4.17}$$

where Q_1 and Q_3 are the lower quartile and the upper quartile, respectively. The interquartile is not sensitive to the outliers.

Sample variance

sample variance s^2 is the mean square deviation about the mean

$$s^2 = \frac{1}{n}\left[(x_1 - m)^2 + (x_2 - m)^2 + \ldots + (x_n - m)^2\right] = \frac{1}{n}\sum_{i=1}^{n}(x_i - m)^2 \tag{4.18}$$

The sample variance converges to the variance of the population when the number of samples tends to infinity, i.e. $s^2 \to \sigma^2$, when we replace in eq.(4.25) n with $n-1$

$$s^2 = \frac{1}{n-1}\left[(x_1 - m)^2 + (x_2 - m)^2 + \ldots + (x_n - m)^2\right] = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - m)^2 \tag{4.19}$$

This estimator is called *unbiased*

Sample standard deviation

sample standard deviation s is the root mean square deviation about the mean

$$s = \sqrt{\frac{1}{n-1}\left[(x_1 - m)^2 + (x_2 - m)^2 + \ldots + (x_n - m)^2\right]} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - m)^2} \quad (4.20)$$

The sample standard deviation is equal to the square root of the sample variance.

Sample coefficient of variation

sample coefficient of variation v is a dimensionless measure of dispersion

$$v = \frac{s}{m} \quad (4.21)$$

It is very useful for easy understanding of the degree of dispersion around the mean value. If for example we know that a sample has $v = 5\%$, this means that the dispersion around the mean value is low; conversely $v \geq 20\%$ is representative of high dispersion around the mean value.

4.3.3 Measures of asymmetry

Sample coefficient of skewness

Sample skewness g_1 is a dimensionless measure of the asymmetry of a set of data about its mean:

$$g_1 = \frac{1}{n}\frac{\sum_{i=1}^{n}(x_1 - m)^3}{s^3} \quad (4.22)$$

- $g_1 < 0$: the histogram has a longer tail on the left
- $g_1 < 0$: the histogram is symmetrical
- $g_1 > 0$: the histogram has a longer tail on the right

The skewness is sensitive to the number of data, it requires approximately more than $1,000$ data, to get reliable estimates.

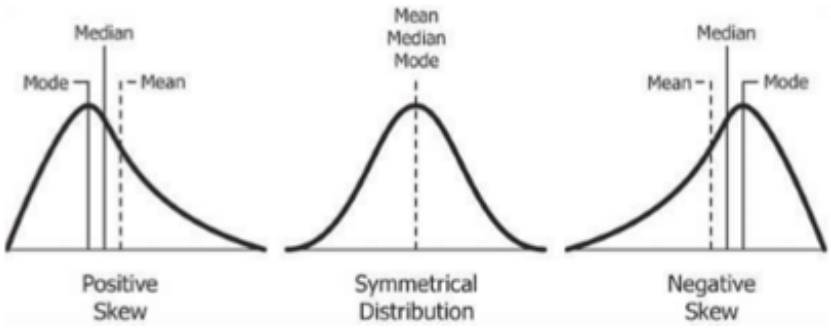


Figure 4.12: Skewness

Sample coefficient of kurtosis

Sample kurtosis g_2 is a dimensionless measure of the data defined as:

$$g_2 = \frac{1}{n} \frac{\sum_{i=1}^{n} (x_1 - m)^4}{s^4} \geq 0 \tag{4.23}$$

It provides information how heavily the tails of a distribution differ from the tails of a normal distribution. In other words, kurtosis identifies whether the tails of a given distribution contain extreme values (assuming as a reference the Gaussian distribution)

- $g_2 < 3$: the histogram has tails lower than Gaussian
- $g_2 = 3$: the histogram has tails like Gaussian
- $g_2 > 3$: the histogram has tails heavier than Gaussian

Sample *excess kurtosis* g_2' is defined as:

$$g_2' = g_2 - 3 \tag{4.24}$$

- $g_2' < 0$: the histogram has tails lower than Gaussian
- $g_2' = 0$: the histogram has tails like Gaussian
- $g_2' > 0$: the histogram has tails heavier than Gaussian

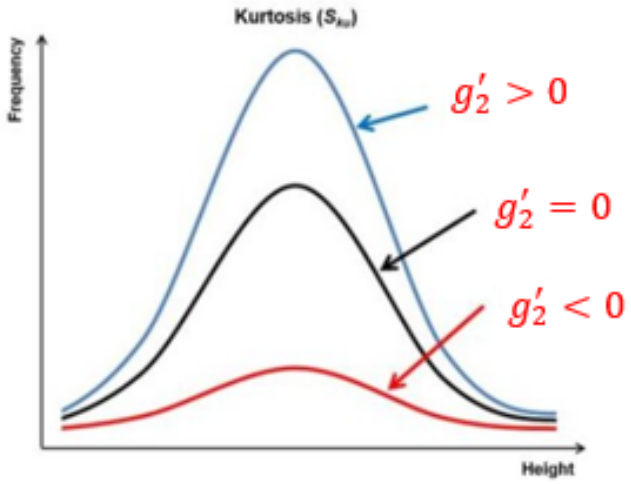


Figure 4.13: Sample kurtosis

The sample kurtosis is sensitive to the number of data, it requires approximately more than $10,000$ data, to get reliable estimates.

OpenAIUQ: the sample mean, standard deviation, coefficient of variation, skewness and kurtosis are the attributes *mean*, *std*, *cov*, *g1* and *g2*.

```
print('Data analysis')
print('Mean: {:.4} MPa'.format(d.mean))
print('Standard deviation: {:.3} MPa'.format(d.std))
print('Coefficient of variation: {:.3}'.format(d.cov))
print('Skewness: {:.3}'.format(d.g1))
print('Kurtosis: {:.3}'.format(d.g2))
```

```
----------------------------------------
Data analysis
Mean: 39.32 MPa
Standard deviation: 9.44 MPa
Coefficient of variation: 0.24
Skewness: 0.536
Kurtosis: 3.61
----------------------------------------
```

We have only 164 data, therefore it is expected that the estimates of sample skewness and kurtosis are not accurate enough. However, in first approximation, given the available information, it can be stated that:

- $v = 24\%$: there is high uncertainty around the mean value. This is expected, since the natural variability of the timber
- $g1 > 0$: the data has a longer tail on the right: this means that it is likely to have extreme events where the realizations of the strength are higher than the expected value
- $g2 > 3$:the tails of the distribution are heavier than Gaussian. The adoption of a Gaussian distribution would be conservative with respect to extreme events

4.4 Data Observed in pairs

Let us assume now that there are n pairs of observation: $(x_1^{(1)}, x_2^{(1)}), (x_1^{(2)}, x_2^{(2)}), \ldots, (x_1^{(n)}, x_2^{(n)})$

4.4.1 Scatter Plot

The **scatter plot** represents the observed pair of values.

- the values of the samples of x_1 in the $x-$axis
- the values of the samples of x_2 in the $y-$axis

4.4.2 Sample covariance

sample covariance $s_{X_1X_2}$ gives a *numerical summary* of the linear association between the two variables X_1 and X_2

$$s_{X_1X_2} = \frac{1}{n}\sum_{i=1}^{n}\left(x_{1i} - m_{X_1}\right)\left(x_{2i} - m_{X_2}\right) \tag{4.25}$$

The covariance is greater when there is a greater direct association between X_1 and X_2. The sample variance converges to the covariance of the population when the number of samples tends to infinity, i.e. $s_{X_1X_2} \to \sigma_{X_1X_2}$.

4.4.3 Sample linear coefficient of correlation

sample linear coefficient of correlation $r_{X_1X_2}$ is a dimensionless measure of linear association between the two variables X_1 and X_2

$$r_{X_1X_2} = \frac{s_{X_1X_2}}{s_{X_1}s_{X_2}} = \frac{\sum_{i=1}^{n}\left(x_{1i} - m_{X_1}\right)\left(x_{2i} - m_{X_2}\right)}{\sqrt{\sum_{i=1}^{n}\left(x_{1i} - m_{X_1}\right)^2}\sqrt{\sum_{i=1}^{n}\left(x_{2i} - m_{X_2}\right)^2}}, \quad -1 \le r_{X_1X_2} \le 1 \tag{4.26}$$

- $r_{X_1X_2} = 0$: uncorrelated variables

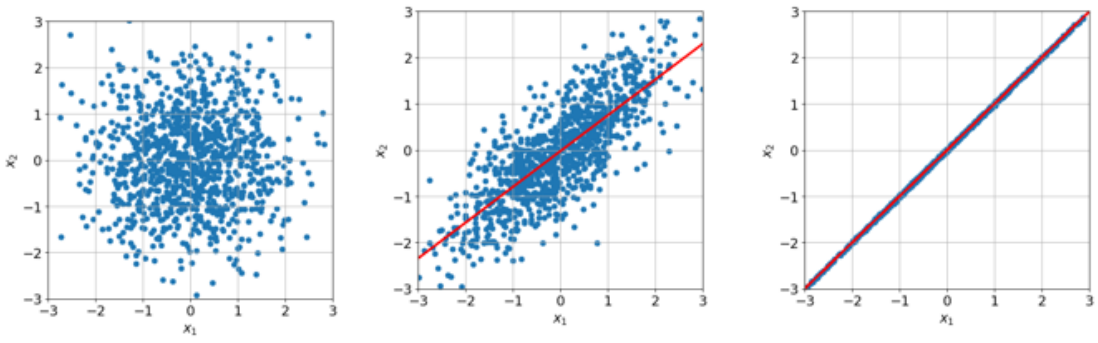


Figure 4.14: Correlation

- $r_{X_1X_2} = \pm 1$: perfectly correlated variables

Two variables are perfectly correlated only if all the points of a scatter plot lie on a straght line

$$x_{2i} = w_0 + w_1 x_{1i}, \qquad i = 1, 2, \ldots, n \tag{4.27}$$

OpenAIUQ: For multivariate data analysis we adopt the class *data2*. At first, we associate a *instance* to the object. The attributes of the class are shown in fig.4.15, and the methods in fig.4.16.

data2 – Attributes	
dataset	dataset
R	correlation
teta0	teta0 regression
teta1	teta1 regression

Figure 4.15: OpenAIUQ: class *data1*, attributes

Case test 3. Let us consider now a set of 40 data of concrete. The two columns collect respectively the concrete density (in kg/m^3), and the concrete strength (in MPa).

OpenAIUQ: Let us define two 1-dimensional data structures $d1$ and $d2$ through the class *auq.data1*, and the 2-dimensional data structure D through the class *auq.data2*.

```
#=====================================
#CONCRETE 40
```

data2 – Methods			
Method	**Description**	**Input**	**Output**
get_corr	evaluate sample correlation		self.R self.teta0 selfa.teta1
plot	plot scatter plot, with regression and histograms	numfig=1, figsize=(6,6), dpi=80, sec=[1,2], hist='yes/no', regression='yes/no'	figure

Figure 4.16: OpenAIUQ: class *data2*, methods

```
#=======================================
dataset=np.loadtxt('concrete40.dat')

#define the 2-dimensional data structure through auq.data2
D=auq.data2(dataset)

#Define the 1-dimensional data structures through auq.data1
d1=auq.data1(dataset[:,0])
d2=auq.data1(dataset[:,1])

print('d1= ',d1.data)
print()
print('d2= ',d2.data)
```

```
----------------------------------------
d1=  [2437. 2437. 2425. 2427. 2428. 2448. 2456. 2436. 2435. 2446. 2441. 2456.
 2444. 2447. 2433. 2429. 2435. 2471. 2472. 2445. 2436. 2450. 2454. 2449.
 2441. 2457. 2447. 2436. 2458. 2415. 2448. 2445. 2436. 2469. 2455. 2473.
 2488. 2454. 2427. 2411.]

d2=  [60.5 60.9 59.8 53.4 56.9 67.3 68.9 49.9 57.8 60.9 61.9 67.2 64.9 63.4
 60.5 68.1 68.3 65.7 61.5 60.  59.6 60.5 59.8 56.7 57.9 60.2 55.8 53.2
 61.1 50.7 59.  63.3 52.5 54.6 56.3 64.9 69.5 58.9 54.4 58.8]
----------------------------------------
```

OpenAIUQ: We read for the two objects d_1 and d_2 belonging to the class $auq.data1$ their values maximum and minimum

```
print('Data analysis')
```

```
print('min value d1: {:.4}'.format(d1.datamin))
print('max value d1: {:.4}'.format(d1.datamax))
print('range d1: {:.3}'.format(d1.r))

print('min value d2: {:.3}'.format(d2.datamin))
print('max value d2: {:.3}'.format(d2.datamax))
print('range d2: {:.3}'.format(d2.r))
```

```
----------------------------------------
Data analysis
min value d1: 2.411e+03
max value d1: 2.488e+03
range d1: 77.0

min value d2: 49.9
max value d2: 69.5
range d2: 19.6
----------------------------------------
```

Given the sampling uncertainty, we choose

$$\begin{aligned} x_{1,L} &= 2400\ kg/m^3, \quad x_{1,U} = 2500\ kg/m^3 \\ x_{2,L} &= 45\ N/mm^2, \quad\ \ x_{2,U} = 75\ N/mm^2 \end{aligned} \tag{4.28}$$

OpenAIUQ: plot the scatter plot with the detected limits: $2400 \leq x_1 \leq 2500\ kg/m^3, 45 \leq x_2 \leq 75\ MPa$

```
fig4=plt.figure(num=4,figsize=(6,6),dpi=80)
ax=fig4.add_subplot(1,1,1)
ax.scatter(d1.data,d2.data)
ax.grid(True)
ax.set_title('Concrete Data')
ax.set_xlabel('$Kg/m^3$')
ax.set_ylabel('$N/mm^2$')
ax.set_xlim(2400,2500)
ax.set_ylim(45,75);
ax.set_xticks([2400,2420,2440,2460,2480,2500]);
ax.set_yticks([45,50,55,60,65,70,75]);
```

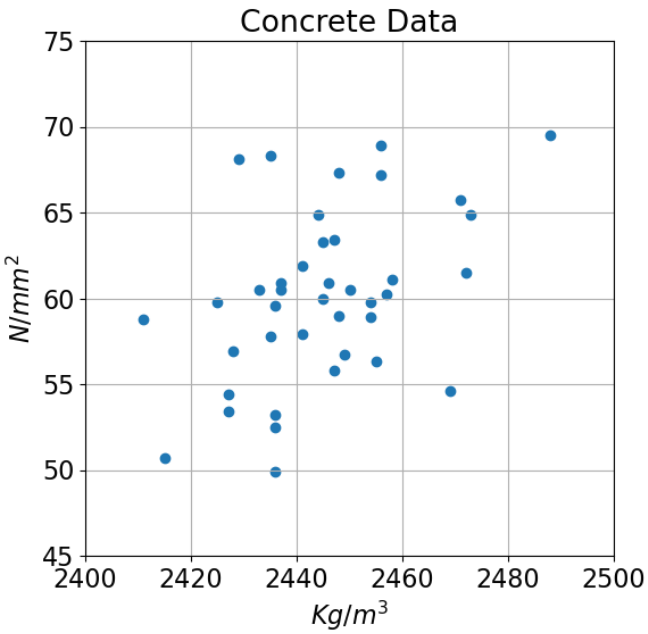


Figure 4.17: OpenAIUQ: concrete, 40 data, scatter plot

OpenAIUQ:

- determine the metrics of the marginal data (i.e. sample mean, standard deviation, coefficient of variation, skewness, kurtosis) collected in the attributes *mean, std, cov, g1, g2* of the instances *d1, d2* belonging to the class *auq.data1*
- determine the metrics of the joint data (i.e. the matrix of correlation) collected in the attribute R of the instance D belonging to the class *auq.data2*

```
#Evaluate the correlation
D.get_corr()

print('Data analysis')

print('Mean d1: {:.4} MPa'.format(d1.mean))
print('Standard deviation d1: {:.4} MPa'.format(d1.std))
print('Coefficient of variation d1: {:.4}'.format(d1.cov))
print('Skewness d1: {:.4}'.format(d1.g1))
print('Kurtosis d1: {:.4}'.format(d1.g2))
```

```
print('Mean d2: {:.4} MPa'.format(d2.mean))
print('Standard deviation d2: {:.4} MPa'.format(d2.std))
print('Coefficient of variation d2: {:.4}'.format(d2.cov))
print('Skewness d2: {:.4}'.format(d2.g1))
print('Kurtosis d2: {:.4}'.format(d2.g2))

print('Correlation R:')
print(D.R)
```

```
----------------------------------------
Data analysis
Mean d1: 2.445e+03 Kg/m3
Standard deviation d1: 15.79 Kg/m3
Coefficient of variation d1: 0.00646
Skewness d1: 0.3892
Kurtosis d1: 3.302

Mean d2: 60.1 MPa
Standard deviation d2: 4.95 MPa
Coefficient of variation d2: 0.0823
Skewness d2: 0.026
Kurtosis d2: 2.45

Correlation R:
[[1.         0.43642728]
 [0.43642728 1.        ]]
----------------------------------------
```

We have only 40 data, therefore it is expected that the estimates of sample skewness and kurtosis are not accurate enough. However, in first approximation, given the available information, it can be stated that:

- $v_{X_1} = 0.64\%, v_{X_2} = 8.23\%$: the concrete density has very low variability; the concrete strength has values quite realistic, as documented from the available literature (in the range $7 - 10\%$)
- $r_{X_1X_2} = 0.43$: the data are correlated. This looks realistic enough

Chapter 5

Lecture A1: Probability Theory

5.1 Random events

5.1.1 Sample space and events

5.1.2 Intersection and union of events

5.1.3 Venn diagram

5.2 Definition of Probability

The purpose of the **Probability theory** is to enable the quantitative assessment of probabilities.

5.2.1 Classical Definition

The **classical** probability definition originates from the days when the probability calculus was founded by Pascal and Fermat. The inspiration for this theory was found in the games of cards and dice

$$P(E) = \frac{n_E}{n_{tot}} \tag{5.1}$$

- n_E: number of *equally likely* ways by which an experiment may lead to E
- n_{tot}: number of equally likely ways in the experiment

Flipping a coin. According to the classical definition of probability, when flipping a coin the probability of achieving a head would be $1/2 = 0.5$, since

- $n_{head} = 1$, there is only one possible way to achieve a head in the experiment
- $n_{tot} = 2$, total number of equally likely ways in the experiment

Rolling a dice. According to the classical definition of probability, when throwing a dice the probability of achieving a three would be $1/6 = 0.1667$, since

- $n_3 = 1$, there is only one possible way to achieve a three in the experiment
- $n_{tot} = 6$, total number of equally likely ways in the experiment

With the classical definition of probability:

- The experiment does not need to be carried out as the answer is known in advance
- The classical theory gives no solution unless all the equally possible ways can be derived analytically

This probability is called **prior probability** because it is determined through *purely deductive reasoning*

But...

- What if the coin is biased in favor of heads?
- What is the probability that a hurricane will occur next year?
- What is the probability that a beam will collapse under a given load?

Notions such as *equally likely* cannot be used in this context

5.2.2 Frequentistic Definition

The **frequentistic** probability definition is the typical interpretation of probability of the experimentalist. It coincides with the relative frequency of occurrence of an event as observed in the experiment with n trials, with $n \to \infty$

$$P(E) = \lim_{n_{exp}\to\infty} \frac{n_E}{n_{exp}}, \qquad f(E) = \frac{n_E}{n_{exp}} \tag{5.2}$$

- n_E: number of experiments where E occurred
- n_{exp}: total number of experiments

Flipping a coin. According to the frequentistic definition of probability, which is the probability of achieving a head?

The experimentalist would run, say 1000 experiments. If:

- $n_{head} = 563$, number of times the experiment provided head
- $n_{tot} = 2$, total number of experiments

He/she would state:

$$P(E) \approx f(E) = \frac{563}{1000} = 56.3\% \tag{5.3}$$

This probability is called **posterior probability** because it derives from experiments

In the mind of a frequentist, probability is a *characteristic of nature*

Posterior probabilities require experiments where the outcomes can occur under somewhat uniform conditions.

But

- What is the probability that a highway system will meet the demands for the next 25 years?
- What is the probability of having a given soil below the foundation of a proposed structure?

Notions such as *subjective probability* can be useful in this context

5.2.3 Bayesian Definition

The **Bayesian** probability definition is the degree of belief of an event

$$P(E) = degree\ of\ belief\ that\ E\ will\ occur \tag{5.4}$$

The degree of belief is a reflection of the *state of mind* of the individual person in terms of experience, expertise and preferences. Thus, the Bayesian interpretation of probability is subjective.

Two different persons may assign different probabilities to a given event

This is different from the frequentistic interpretations, according to which probabilities are a characteristic of nature.

The degree of belief is denoted also **prior belief** or prior probability.

Modern structural reliability and risk analysis are based on the *Bayesian interpretation of probability*

5.3 Bayes Rule

The **Bayes rule** reads as

$$P\left(E_i|data\right) = \frac{P\left(data|E_i\right)P\left(E_i\right)}{\sum_{j=1}^{n} P\left(data|E_j\right)P\left(E_j\right)} \tag{5.5}$$

- $P\left(E_i|data\right)$: posterior probability
- $P\left(data|E_i\right)$: likelihood
- $P\left(E_i\right)$: prior probability

In eq.5.5 the denominator has just the role of a constant of normalization, since that the probabilities need to sum up to one. It follows that the posterior probability $P"\left(\theta\right)$ is proportional to the prior probability $P'\left(\theta\right)$ through the likelihood $L\left(\mathbf{z}\right)$

$$P"\left(\theta\right) \propto L\left(\mathbf{z}\right)P'\left(\theta\right) \tag{5.6}$$

Bayes rule has a very important role in engineering because it provides a method to *incorporate new information with previous data*.

By updating the prior probabilities, the engineer can assess the likelihood of design events by incorporating the additional information given by conditioned posterior probabilities.

5.4 Event Trees

5.5 Introduction to Bayesian Networks

Chapter 6

Lecture A2: Random Variables

6.1 Random Variables

Random variable is one variable whose value is uncertain or unpredictable or nondeterministic (e.g. concrete strength)

A random variable can be seen as a function defined on the sample space of an experiment, such that there is a probability associated with each occurrence in the sample space

More formally, probability theory aims at associating numbers to events, i.e. their probability of occurrence.

Let denote P this so-called *probability measure*. The collection F of possible events having well-defined probabilities is called the $\sigma - algebra$ associated with the sample space Ω.

The probability space constructed by means of this notions is denoted by (Ω, F, P)

A real random variable is a mapping

$$X : (\Omega, F, P) \rightarrow \Re \tag{6.1}$$

A random variable is denoted with capital letters, e.g. X. A realization of a random variable is denoted with small letters, e.g. x.

We distinguish between

- *continuous random variables* : can take any value in a given range
- *discrete random variables* : can take only discrete values.

A random variable is statistically specified by its probability distribution or probability law. This is specified using a mathematical function.

The cumulative distribution function (CDF) $F_X(x)$ of a random variable is the *probability of nonexceedance* of x

$$F_X(x) = P[X \leq x] \tag{6.2}$$

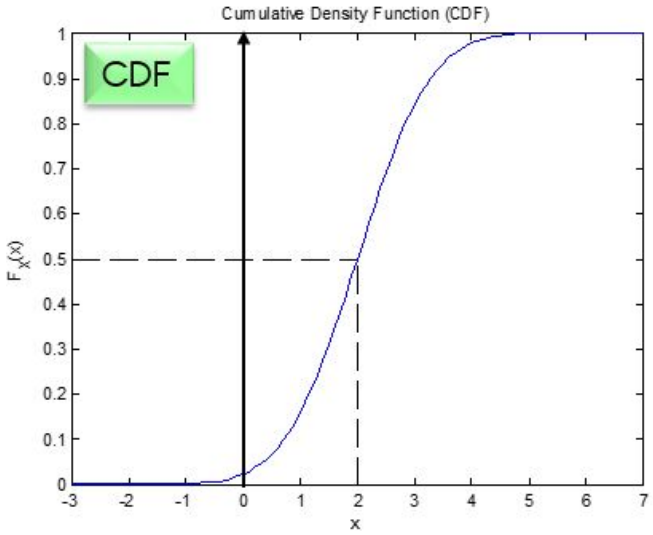


Figure 6.1: Cumulative Distribution Function (CDF)

The CDF is a monotonic (nondecreasing) continuous function

$$0 \leq F_X(x) \leq 1 \tag{6.3}$$

6.1.1 Discrete Random Variables

A *discrete random variable* can assume at most a countable set of values, like positive integers

Probability Mass Function (PMF)

The Probability Mass Function of a discrete random variable X gives the point probabilities of the values taken by X

$$p_X(x_i) = P[X = x_i] = p_i \tag{6.4}$$

The *Probability axioms* for a discrete random variable are:

- $0 \leq p_i \leq 1$, for all possible x_i
- $p_i = 0$, for all unrealizable x_i
- $\sum_i p_i = 1$, which is summed over all the possible x_i

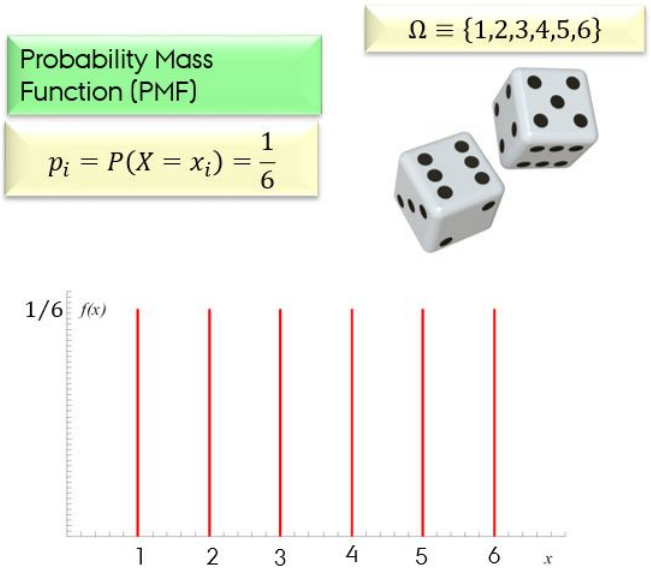


Figure 6.2: Probability Mass Function (PMF), dice

If one is certain that the outcome is x_0:

$$p_X(x_0) = P[X = x_0] = 1 \tag{6.5}$$

For mutually exclusive outcomes $x_1, x_2, \ldots, x_m$:

$$p_X(x_1 \cup x_2 \cup \ldots \cup x_m) = p_1 + p_2 + \ldots + p_m \tag{6.6}$$

Cumulative Distribution Function (CDF)

In the case of a discrete random variable, the CDF $F_X(x)$ is the *sum of the probabilities of all the possible values of* X *that are less or equal to the argument* x_i

$$F_X(x_i) = P[X \leq x_i] = \sum_{k=1}^{i} p_k \tag{6.7}$$

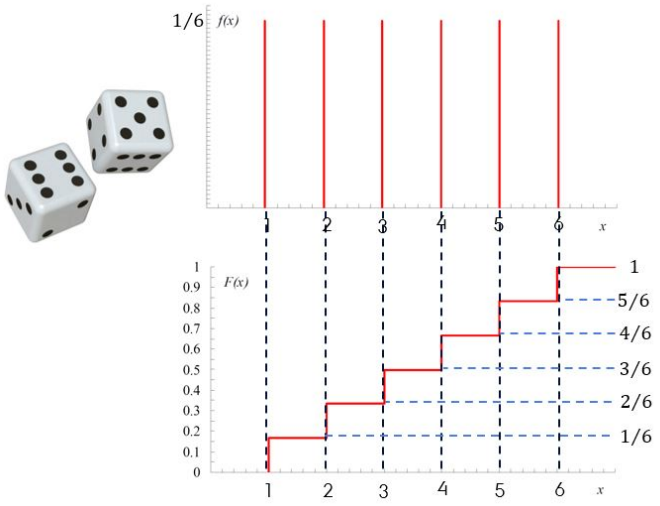


Figure 6.3: Cumulative Distribution Function (CDF), dice

For example, in the case of rolling a dice, the PMF is $p_1 = p_2 = p_3 = p_4 = p_5 = p_6 = \frac{1}{6}$, and the probability that X is not greater than 3 is

$$F_X(x_3) = P[X \leq x_3] = p_1 + p_2 + p_3 = \frac{1}{2} \tag{6.8}$$

6.1.2 Continuous Random Variables

A *continuous random variables* can take any value in a given range

Probability Density Function (PDF)

The probability density function (PDF) $f_X(x)$ is the probability law for a continuous random variable

As implied by its name, the PDF is not dimensionless and hence by itself does not represent a probability; it denotes an intensity of probability. The relationship between the PDF and the CDF is:

$$F_X(x) = \int_{-\infty}^{x} f_X(\xi)\, d\xi \tag{6.9}$$

$$f_X(x) = \frac{dF_X(x)}{dx} \tag{6.10}$$

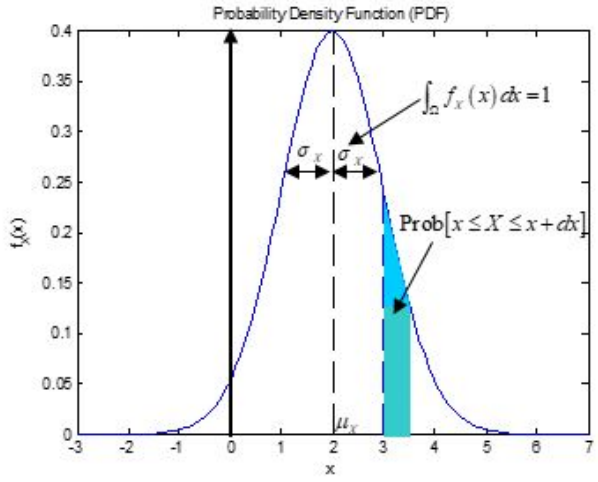


Figure 6.4: Probability Density Function (PDF)

Some properties:

$$f_X(x) \geq 0 \tag{6.11}$$

$$\int_\Omega f_X(x)\,dx = 1 \tag{6.12}$$

$$P[x_1 \leq X \leq x2] = F_X(x_2) - F_X(x_1) = \int_{x_1}^{x_2} f_X(x)\,dx = 1 \tag{6.13}$$

The PDF and the CDF of a random variable provide the probability model that describes a process subjected to uncertainty.

The probability model is often in the form of a parametric function $f_X(x;\theta)$. This represents a *parametric family* of PDFs: for each value of the parameters $\theta_1, \theta_2, ..., \theta_m$ we have a different distribution belonging to the family.

For example, consider the Gaussian distribution

$$f(x;\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right]$$

This represents a whole set of Gaussian distributions, but of course if we let $\mu = \mu_0, \sigma = \sigma_0$ we obtain a specific distribution$f(x|\mu_0,\sigma_0)$ belonging to the family

6.2 Descriptors of Random Variables

A random variable is fully characterized from a statistical point of view from the knowledge of its PDF (or PMF).

The descriptors summarize some important features of the behavior of the random variables, and they are relevant for engineering applications.

Some important descriptors are the *mean*, *variance*, *skewness* and *kurtosis*. These are population measures.

Given a set of data we have defined sample mean, sample variance, sample skewness and sample kurtosis. They derive from a sample (i.e. a realization of the random variable). They give information over the data, but they are not so useful for doing prediction over the *unseen data*.

The sample statistics and population statistics coincide when the sample size tends to infinity

Moments

It is common to represent the PDF through its moments defined as

Raw moment of order k

$$\mu_k = E\left[X^k\right] = \int x^k \; f_X\left(x\right) dx \tag{6.14}$$

Central moment of order k

$$\mu'_k = E\left[\left(X - \mu_X\right)^k\right] = \int \left(x - \mu_X\right)^k \; f_X\left(x\right) dx \tag{6.15}$$

Mean or Expected Value

The sample mean of a set of data is simply the average of the observed data

The expected value (or population mean) is the average value weighted according to the probability distribution

$$\mu_X = E\left[x\right] = \int_{-\infty}^{\infty} x \; f_X\left(x\right) \; dx \tag{6.16}$$

$\mu_X \equiv \mu_1$ is the mean value coinciding with the *moment of order 1*.

The mean is an important measure of the central tendency of the random variable

Variance

The main characteristic of a random phenomenon is its variability

The variance is the *central moment of order 2*:

$$Var\left[X\right] \equiv \sigma_X^2 = E\left[\left(X - \mu_X\right)^2\right] = E\left[X^2\right] - \mu_X^2 \tag{6.17}$$

$$Var\left[X\right] \equiv \sigma_X^2 = \mu_2 - \mu_1^2 \tag{6.18}$$

Thus, the variance is equal to the difference between the mean of the squares and the square of the mean

Standard deviation

The standard deviation is definite as the positive square root of the variance

$$\sigma_X = \sqrt{Var\left[X\right]} = \sqrt{E\left[X^2\right] - \mu_X^2} \tag{6.19}$$

It gives the *dispersion of the random variable around the mean value.*

It is measured in the same units as the variable and therefore it is more meaningful as a measure of dispersion than the variance, because it can be compared directly with the variance

Coefficient of variation

The coefficient of variation is definite as

$$\nu = \frac{\sigma}{\mu} \tag{6.20}$$

and it provides a dimensionless measure of dispersion.

In figs.6.5 and 6.6 it is shown respectively two distribution having $\nu = 0.10$ and $\nu = 0.20$, respectively

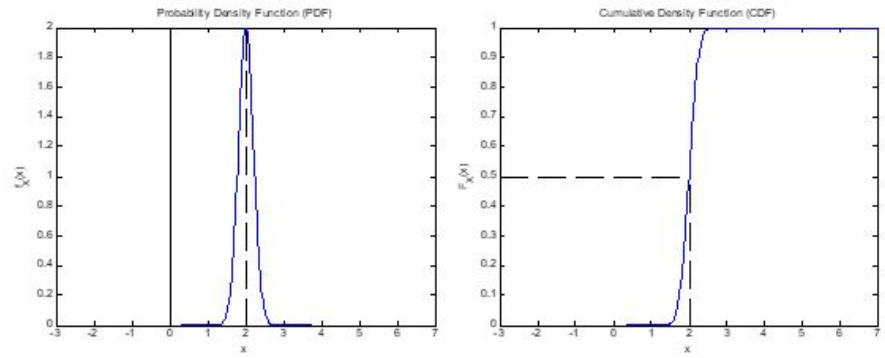


Figure 6.5: PDF and CDF, cov=10%

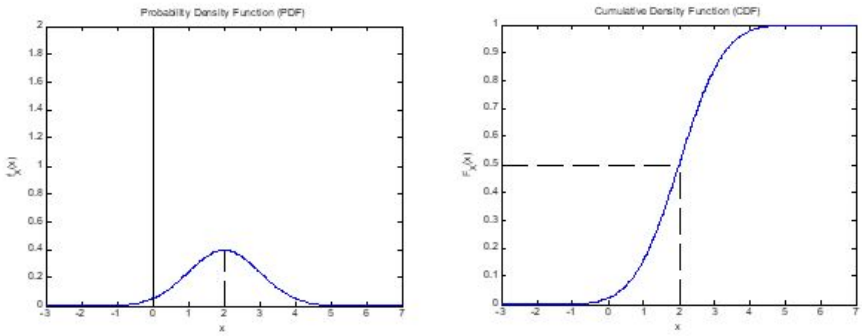


Figure 6.6: PDF and CDF, cov=20%

Higher-order moments

The higher-order moments take into account information about the tails of the distributions.

To explain this, it is noted that the moments of order k weight the PDF $f_X(x)$ with the polynomial x^k. This means that μ_1 weights the PDF with a linear function, μ_2 with a quadratic function, μ_3 with a cubic function, and so on.

In fig.6.7 it is seen that in the tails of the PDF, i.e. $x = 10$, the function $f(x) = x$ is equal to $f(x) = 10$, the quadratic function $f(x) = x^2$ has a value $f(x) = 100$, the cubic function $f(x) = x^3$ has a value $f(x) = 1,000$. The lower order moments put more moments over the central part, while higher-order moments put more weight over the tail region.

This is further confirmed in fig.6.8, where we represent the functions $(x - \mu)^k f(x),\ k = 2, 3, 4$, which are the integrand function of the central moments. In particular it is seen

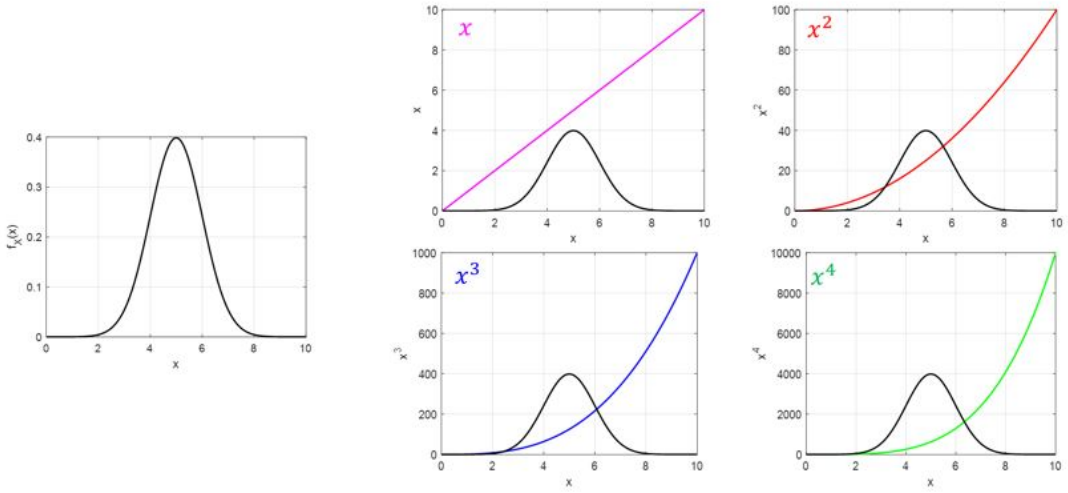


Figure 6.7: Higher-order moments

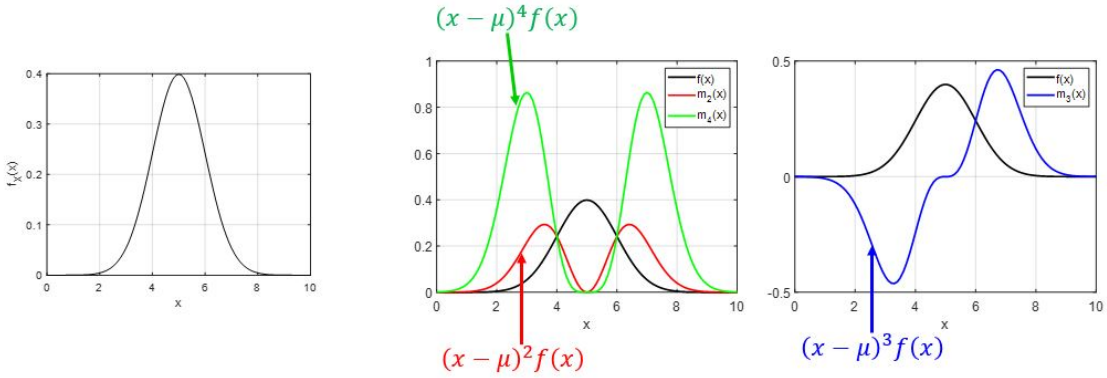


Figure 6.8: Moments of order 3 and 4

that the moment of order 3 describes the symmetry properties of the PDF. Moreover it is shown that the moment of order 4 put more weight on the tails with respect to the moment of order 2.

Measures of asymmetry: coefficient of skewness

The coefficient of skewness is estimated by γ_1, which is a dimensionless measures of the asymmetry of a set of data about its mean.

$$\gamma_1 = \frac{\mu_3'}{\sigma^3} = \frac{E\left[(x-\mu_X)^3\right]}{\sigma^3} \tag{6.21}$$

- $\gamma_1 < 0$: The PDF has its dominant tail on the left (negatively skewed)
- $\gamma_1 = 0$: The PDF is symmetrical
- $\gamma_1 > 0$: The PDF has its dominant tail on the right (positively skewed)

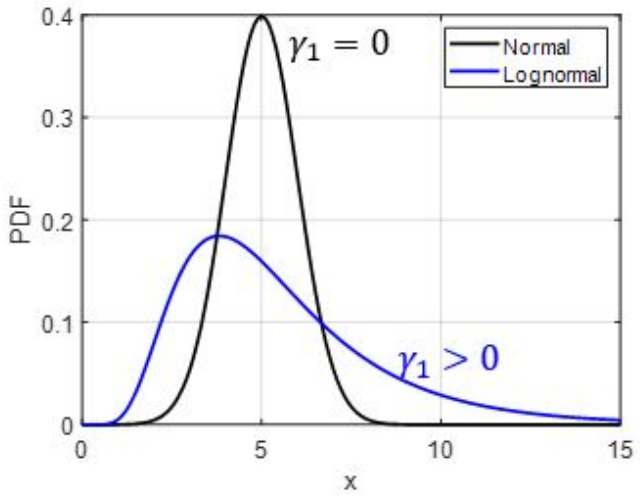


Figure 6.9: Skewness

Measures of "peak/tail": coefficient of kurtosis

The coefficient of Kurtosis is estimated by γ_2

$$Kurt[X] = \gamma_2 = \frac{\mu_4'}{\sigma^4} = \frac{E\left[(x-\mu_X)^4\right]}{\sigma^4} \tag{6.22}$$

The only data values (observed or observable) that contribute to kurtosis in any meaningful way are those outside the region of the peak; i.e., extreme values. Thus, it provides information through the heaviness of the tails. Since the area of a PDF is equal to unity

$$\int f(x)\,dx = 1$$

the shape of the distribution over the tails affects also the central region. For this reason the kurtosis provides information also how closely the data are distributed around the mode

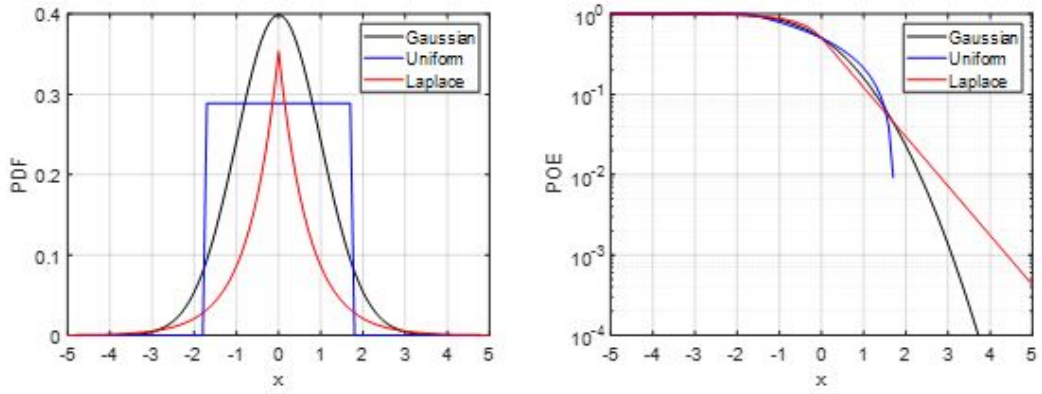


Figure 6.10: Kurtosis

The excess kurtosis γ_2' is defined as

$$\gamma_2' = Kurt[X] - 3 = \frac{\mu_4'}{\sigma^4} - 3 \tag{6.23}$$

- *Platykurtic*: $\gamma_2' \leq 0$, (e.g. uniform, Bernoulli). They has thinner tails than Gaussian and they are called sub-Gaussian.
- *Mesokurtic*: $\gamma_2' = 0$, (e.g. Gaussian)
- *Leptokurtic*: $\gamma_2' \geq 0$, (e.g. Students t, Laplace, exponential, Poisson, logistic). They have fatter tails than Gaussian and they are called super-Gaussian.

6.2.1 Moment Generating function

6.2.2 Characteristic function

6.2.3 Cumulants

6.3 Transformation

If X is a random variable, then

$$Y = g(X) \tag{6.24}$$

is a random variable, too.

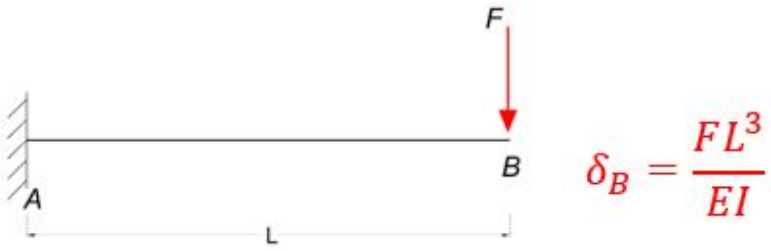


Figure 6.11: Deflection of a cantilever beam

The deflection of a cantilever beam is

$$\delta = \frac{FL^3}{3EI} \tag{6.25}$$

Two different cases can be considered:

1. $F(\equiv X)$ is a Gaussian random variable, (L, E, I) are deterministic $\rightarrow \delta(\equiv Y)$ is a (Gaussian) random variable
2. $E(\equiv X)$ is a Gaussian random variable, (F, L, I) are deterministic $\rightarrow \delta(\equiv Y)$ is a (non-Gaussian) random variable

6.3.1 Mean and variance

Given the linear transformation

$$Y = aX + b \tag{6.26}$$

with X being Gaussian (case 1), then:

- Y is Gaussian
- $\mu_Y = a\mu_X + b$
- $\sigma_Y^2 = a^2\sigma_X^2$

Linear transformations of Gaussian variables is a Gaussian variable, too.

Given the non-linear transformation

$$Y = g(X) \tag{6.27}$$

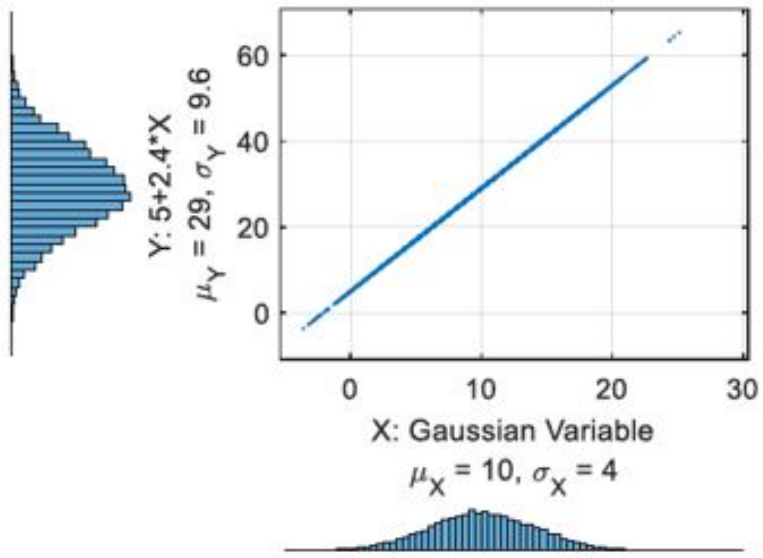


Figure 6.12: Linear Transformations

with X being Gaussian (case 2), nothing can be said in advance with respect to δ. This because of the non-linearity of the transformation, the parametric family of the density of X is not preserved

To represent the PDF of Y, a solution may be to develop a Taylor series expansion of $Y = g(X)$ around a point X_0

$$g(X) \approx g(X_0) + \frac{1}{1!}\frac{dg(X_0)}{dX}(X - X_0) + \frac{1}{2!}\frac{d^2g(X_0)}{dX^2}(X - X_0)^2 + \dots \tag{6.28}$$

It is therefore of particular interest the choice of the point of linearization X_0. If the mean value is chosen, i.e. $X_0 \equiv \mu_X$, then one has the *mean value approximation*. By developing first-order Taylor series expansion of $g(X)$ one has

$$g(X) \approx g(\mu_X) + g'(\mu_X)(X - \mu_X) \approx g'(\mu_X)X + b \equiv g_1(\mu_X) \tag{6.29}$$

where $g_1(\mu_X)$ means first-order approximation of $g(X)$ around the mean value μ_X. Therefore, we are approximating the non-linear function with a function around the mean value. Since $\widehat{Y} \equiv g_1(X)$ is a linear function of the Gaussian random variable X, then is:

- $g(X)$ is Gaussian
- $E[g(X)] = g'(\mu_X)X + b$
- $Var[g(X)] = [g'(\mu_X)]^2 \sigma_X^2$

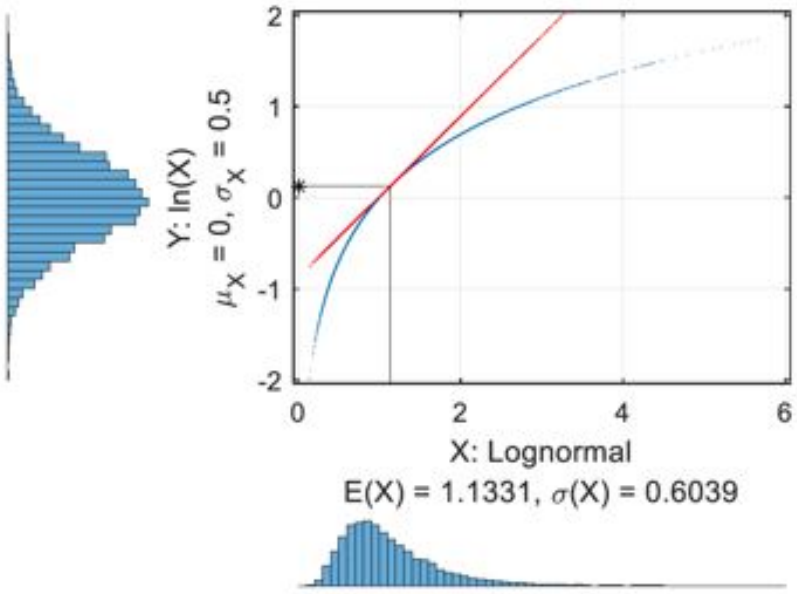


Figure 6.13: Nonlinear Transformation

However, the mean value approximation gives a good approximation only around the mean. The tails are badly approximated; to get a good approximation in the tail region a different linearization point needs to be chosen

6.3.2 Density of a transformation

Let $Y = g(X)$ the transformation of a random variable X

$$Y = g(X), \qquad X = g^{-1}(Y) = h(Y) \tag{6.30}$$

In such cases the **Equal Probability Formula** holds

$$P\left(x \leq X \leq x + dx\right) = P\left(y \leq Y \leq y + dy\right) \tag{6.31}$$

This means that the two distributions (evaluated in the two points x_1 and y_1) have the same probability p_1, as shown in fig.6.14. In other words, there is a mapping between x_1 and y_1, such that $F_X(x_1) = F_Y(y_1) = p_1$.

Eq.(6.31) in terms of PDF can be expressed as

$$f_X(x)dx = f_Y(y)dy \tag{6.32}$$

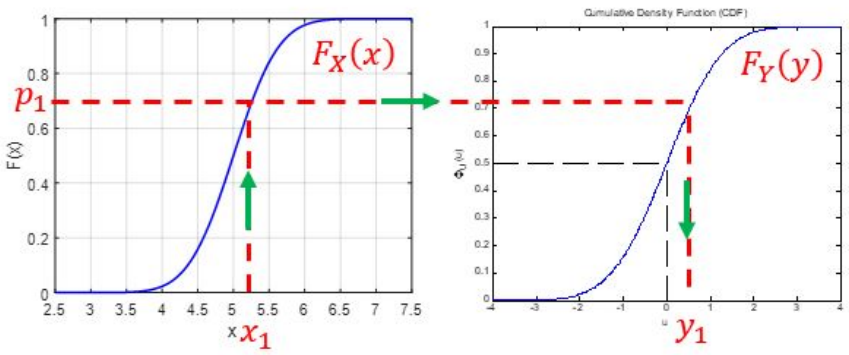


Figure 6.14: Equal probability formula

Thus, the PDF $f_Y(y)$ reads as

$$f_Y(y) = \frac{1}{\frac{dy}{dx}} f_X(x) \tag{6.33}$$

More formally, the Jacobian is defined as

$$J_{y,x}(x) = \frac{dg(x)}{dx} \tag{6.34}$$

and therefore, the density of the transformation (6.30) is

$$f_Y(y) = \frac{1}{|J_{y,x}(x)|} f_X\left[h\left(y\right)\right] \tag{6.35}$$

Linear example.

$$Y = aX + b, \qquad X = \frac{Y - b}{a} \tag{6.36}$$

Jacobian:

$$J_{Y,X} = \frac{dg(X)}{dX} = a \tag{6.37}$$

Density of the transformation:

$$f_Y(y) = \frac{1}{a} f_X\left(\frac{y - b}{a}\right) \tag{6.38}$$

Chapter 7

Lecture A2: Probability Distributions

7.1 Discrete distributions

7.1.1 Bernoulli distribution

7.1.2 Binomial distribution

7.1.3 Poisson distribution

7.2 Continuous distributions

7.2.1 Uniform distribution

7.2.2 Triangular distribution

7.2.3 Gaussian distribution

7.2.4 Lognormal distribution

7.2.5 Weibull distribution

7.2.6 Exponential distribution

7.3 Nonparametric distributions

7.3.1 Mixture of Gaussians

7.3.2 Kernel Density Estimation (KDE)

Chapter 8

Lecture A2:Model Estimation

8.1 Model Building

Model building may be seen to consist of five steps:

- Assessment and statistical quantification of the available data
- selection of distribution function
- Estimation of distribution parameters
- model verification and selection
- model updating

We have previously already discussed abut statistical quantification of the available data.

8.2 Selection of probability distribution function

In engineering application it is often the case that the available data is too sparse to be able to support/reject the hypothesis of a given probability distribution with a reasonable significance

Therefore it is necessary to use common sense, i.e.:

- First to consider physical reasons for selecting a given distribution
- Thereafter to check if the available data are in gross contradiction with the selected distribution

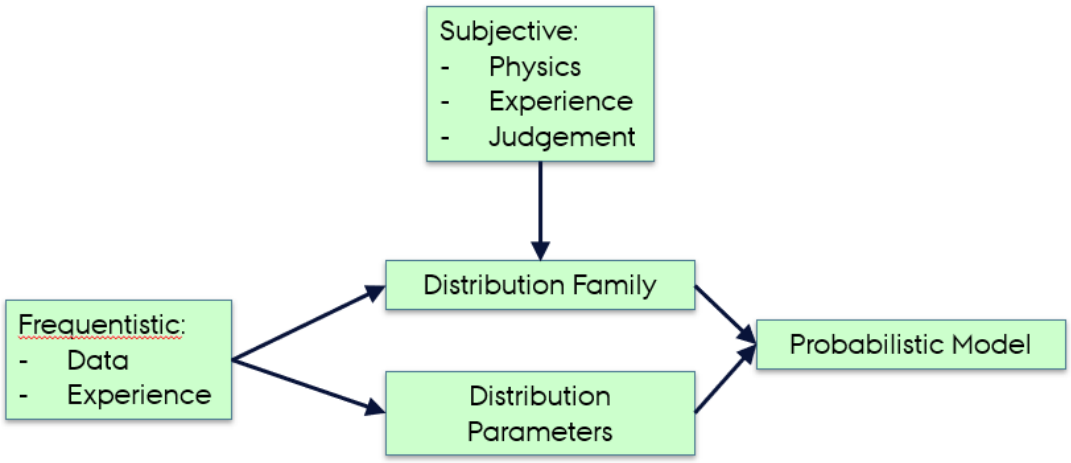


Figure 8.1: Model Building

All the models are wrong, but some are useful . Usefulness is the key metric when designing models

Distribution	PDF	Parameters	Support	Applications
Gumbel	$f(x;\mu,\alpha)=\frac{1}{\alpha}exp[-(z+exp\{-z\})]$ $z=\frac{x-\mu}{\alpha}$	μ *location* $\beta>0$ *scale*	$]-\infty,+\infty[$	Annual maximum wind pressure Snow load Temperature load
Weibull	$f(x;\lambda,k)=\frac{k}{\lambda}\left(\frac{x}{\lambda}\right)^{k-1}exp\left[-\left(\frac{x}{\lambda}\right)^{k}\right]$ $x\geq 0$	$\lambda\in(0,\infty)$ *scale* $k\in(0,\infty)$ *shape*	$[0,+\infty[$	Wave heights
Generalized Pareto	$f(x;\mu,\sigma,\xi)=\frac{1}{\sigma}(1+\xi z)^{-\frac{1}{1+\xi}}$ $z=\frac{x-\mu}{\sigma}$	$\mu\in(-\infty,\infty)$, *location* $\sigma\in(0,\infty)$ *scale* $\xi\in(-\infty,\infty)$ *shape*	$\xi\geq 0$: $x\geq\mu$ $\xi<0$: $\mu\leq x\leq\mu-\frac{\sigma}{\xi}$	Wave heights on shallow water

Figure 8.2: Extreme Loads

8.3 Parameter estimation

After we identify the families of distributions (e.g. Gaussian, Lognormal, Weibull, etc.) that could fit well for physical reasons the physical quantities, next step is the evaluation of its parameters, in order to target the particular distribution (i.e. with chosen value of the parameters, $\theta=\theta^*$)

We distinguish:

Distribution	PDF	Parameters	Support	Applications
Normal	$f(x;\mu,\sigma)=\frac{1}{\sigma\sqrt{2\pi}}exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]$	μ *location* $\sigma>0$ *scale*	$]-\infty,+\infty[$	Wind velocity Wave height
Lognormal	$f(x;\mu,\sigma)=\frac{1}{x\sigma\sqrt{2\pi}}exp\left[-\frac{(\log x-\mu)^2}{2\sigma^2}\right]$	μ *location* $\sigma>0$ *scale*	$[0,+\infty[$	Wind velocity Wave height
Weibull	$f(x;\lambda,k)=\frac{k}{\lambda}\left(\frac{x}{\lambda}\right)^{k-1}exp\left[-\left(\frac{x}{\lambda}\right)^k\right]$ $x\geq 0$	$\lambda\in(0,\infty)$ *scale* $k\in(0,\infty)$ *shape*	$[0,+\infty[$	Wind velocity Wave height

Figure 8.3: Fatigue analysis

Distribution	PDF	Parameters	Support	Applications
Normal	$f(x;\mu,\sigma)=\frac{1}{\sigma\sqrt{2\pi}}exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]$	μ *location* $\sigma>0$ *scale*	$]-\infty,+\infty[$	Not recommended for materials with high coeff. of variation
Lognormal	$f(x;\mu,\sigma)=\frac{1}{x\sigma\sqrt{2\pi}}exp\left[-\frac{(\log x-\mu)^2}{2\sigma^2}\right]$	μ *location* $\sigma>0$ *scale*	$[0,+\infty[$	
Weibull	$f(x;\lambda,k)=\frac{k}{\lambda}\left(\frac{x}{\lambda}\right)^{k-1}exp\left[-\left(\frac{x}{\lambda}\right)^k\right]$ $x\geq 0$	$\lambda\in(0,\infty)$ *scale* $k\in(0,\infty)$ *shape*	$[0,+\infty[$	Recommended when the largest defect is important

Figure 8.4: Material strength

- *Point Estimation*: the parameters of the distribution are deterministic (i.e. a number) and they are determined through the "Method of the Moments" (MM) or the Maximum Likelihood Estimation (MLE).
- *Bayesian Estimation*: the parameters of the distribution are modelled through random variables
- *Non probabilistic models*: the uncertainties and/or the parameters of the distributions are modelled through intervals

In this lecture our focus will be devoted mostly to the methods of **point estimation**.

In particular we can distinguish:

- *Parametric distributions*: the choice of the distribution follows
  1. Physical reasons
  2. statistical evidence

  Some common parametric distributions:

  - Normal
  - Lognormal
  - Uniform
  - Gamma
  - Extreme value
  - ....

- *Non parametric distributions*: the distributions are purely data-driven
  - Kernel Density Estimation (KDE)
  - Maximum Entropy (ME)
  - ....

  To this category belong also the *physical based data-driven models* (PB-ML), which integrate Physical-Based (PB) model with data-driven model (based on Machine Learning, ML)
  - Kernel Density Maximum Entropy (KDME)
  - Physics-Informed Neural Neutwork (PINN)
  - ....

8.3.1 Estimation of Distribution Parameters

Let us assume that we have identified a plausible family of probability distribution functions, as an example:

Normal Distribution:

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right] \tag{8.1}$$

Eq.(8.1) refers to a family of Gaussian distributions, since different choices of the parameters $\theta = (\theta_1, \theta_2) \equiv (\mu, \sigma)$ give rise to a specific distribution belonging to the family.

Weibull Distribution:

$$f_X(x) = \frac{k}{u-\epsilon}\left(\frac{x-\epsilon}{u-\epsilon}\right)^{k-1} \exp\left[-\left(\frac{x-\epsilon}{u-\epsilon}\right)^{k}\right] \tag{8.2}$$

Eq.(8.2) refers to a family of Weibull distributions, since different choices of the parameters $\theta = (\theta_1, \theta_2, \theta_3) \equiv (k, u, \epsilon)$ give rise to a specific distribution belonging to the family.

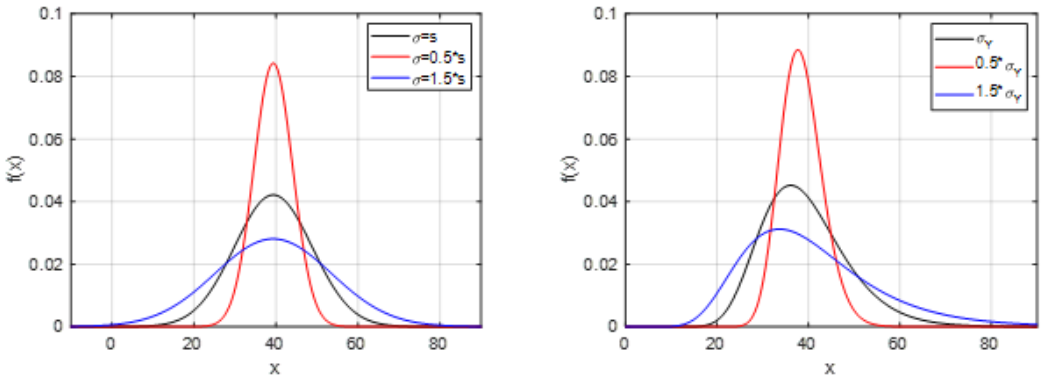


Figure 8.5: Family of distributions, left: Gaussian, right: Lognormal

In the figure on the left we see for example three distributions belonging to the family of Gaussian PDFs, (i) $\mu = 40, \sigma = 5$, (ii) $\mu = 40, \sigma = 10$, (iii) $\mu = 40, \sigma = 15$. Different parameters of the same family, provide different distributions.

8.3.2 Sampling variability

In any statistical analysis, estimates are the 'best guesses' of the truth given the available information.

It is assumed that the data come from an unknown target PDF $f(x)$, and our aim is to determine, from the available data, a *model* $f_m(x)$ as close as possible to the target $f(x)$.

It is implicit that other data, coming from the same unknown target PDF $f(x)$, equally representative of the true process being studied, *would have led to different estimates of the parameters of the model* $f_m(x)$.

It is important to complement the estimate of a model with measures of uncertainty due to sampling variability

To clarify this, assume now that we have a set of 16 data of timber strength, we will get a sample mean $m^{(1)} = 38.39MPa$. Now analyze a further set of 16 experimental tests and we get a sample mean $m^{(2)} = 40.08MPa$, and so on, $m^{(3)} = 37.12MPa, m^{(4)} = 36.21MPa, m^{(5)} = 37.50MPa, m^{(6)} = 41.30MPa, m^{(7)} = 35.99MPa$, etc., see fig.8.6.

Of course all these sample means $m^{(1)}, m^{(2)}, ..., m^{(7)}$ are representative of the same target unknown PDF $f(x)$. Therefore, if we had a dataset with infinite data (not just 16), then it is expected that the sample mean is converging to the *sample population* μ of $f(x)$.

	Timber strength [MPa]						
1	48.78	32.02	45.54	32.40	48.37	50.98	35.58
2	41.64	39.34	34.12	33.06	29.93	40.71	28.97
3	39.77	46.33	45.92	33.47	36.38	34.63	34.56
4	35.89	46.99	36.47	35.67	46.86	24.84	26.89
5	54.04	22.67	28.98	28.46	36.00	28.83	38.64
6	35.43	39.62	40.85	23.16	23.19	42.31	24.25
7	22.75	44.78	56.60	44.51	36.88	39.33	44.54
8	44.78	26.63	28.76	42.47	44.30	39.93	40.85
9	43.99	43.48	47.42	48.39	44.59	39.60	39.97
10	47.74	30.05	33.61	38.05	44.00	38.16	37.69
11	25.39	30.33	44.36	35.03	40.39	43.33	41.78
12	38.00	39.21	35.30	31.33	41.72	69.07	33.14
13	25.98	51.39	33.18	27.31	29.90	51.90	55.23
14	36.84	50.91	36.85	53.99	35.17	33.71	36.53
15	34.49	49.65	17.98	43.41	34.44	46.50	22.74
16	38.71	47.83	27.90	28.71	27.93	36.92	34.40
m	38.39	40.08	37.12	36.21	37.50	41.30	35.99

Figure 8.6: Timber, sampling variability

8.3.3 Estimators

Point estimation is the attempt to provide the single 'best' prediction of some quantity of interest.

An estimator (or statistic) is a method/function of random variables used to estimate the true parameters of the distribution. The particular value of an estimator for an observed set of data is the estimate. Randomness in the sampling process provides randomness in the estimator.

The observations $x^{(1)}, x^{(2)}, ..., x^{(n)}$ are realizations of independent and identically distributed (i.i.d.) random variables $X_1, X_2, ..., X_n$. All these random variables have the same target PDF $f(x)$. This means that $x^{(1)}$ is a realization of X_1 having PDF $f(x)$, $x^{(2)}$ is a realization of X_2 having the same PDF $f(x)$, and so on.

A point estimator or statistic $\hat{\theta_n}$ (n denotes the number of samples $x^{(1)}, x^{(2)}, ..., x^{(n)}$) of the m parameter(s) $\theta \equiv \theta_1, \theta_2, ..., \theta_m$ is a function of the observations, therefore it is also

a random variable

$$\hat{\theta_n} = g\left(X_1, X_2, ..., X_n\right) \tag{8.3}$$

The statistics of the estimator are distributed according to a certain probability law, called **sampling distribution**. It has therefore a PDF $f_\Theta(\theta)$.

8.3.4 Bias of an estimator

The **bias** of an estimator $\hat{\theta}_n$ measures the expected deviation from the true value of the parameter θ

$$Bias\left(\hat{\theta}_n\right) = E\left[\hat{\theta}_n - \theta\right] = E\left[\hat{\theta}_n\right] - \theta \tag{8.4}$$

where $E\left[\hat{\theta}_n\right]$ is the expected value of the random variable $\hat{\theta}_n$ having PDF $f_\Theta(\theta)$.

An estimator is **unbiased** if

$$Bias\left(\hat{\theta}_n\right) = 0 \tag{8.5}$$

An estimator is **asymptotically unbiased** if

$$\lim_{n\to\infty} Bias\left(\hat{\theta}_n\right) = 0 \tag{8.6}$$

Sample Mean

Consider a set of samples $x^{(1)}, x^{(2)}, ..., x^{(n)}$ that are i.i.d. according to a Gaussian distribution

$$f\left(x; \mu, \sigma\right) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\frac{\left(x-\mu\right)^2}{\sigma^2}\right] \tag{8.7}$$

An estimator of the mean parameter μ of the Gaussian distribution is the sample mean:

$$\hat{\mu}_n = \frac{1}{n}\sum_{i=1}^{n} x^{(i)} \equiv m \tag{8.8}$$

and it is unbiased:

$$\begin{aligned} Bias\left(\hat{\mu}_n\right) &= E\left[\hat{\mu}_n\right] - \mu = \\ &= E\left[\frac{1}{n}\sum_{i=1}^{n} x^{(i)}\right] - \mu = \\ &= \frac{1}{n}\sum_{i=1}^{n} E\left[x^{(i)}\right] - \mu = \\ &= \left(\frac{1}{n}\sum_{i=1}^{n} \mu\right) - \mu = \\ &= \mu - \mu = 0 \end{aligned} \tag{8.9}$$

Sample Variance

An estimator of the Gaussian parameter σ^2 is the sample variance

$$\hat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^{n}\left(x^{(i)} - \hat{\mu}_n\right)^2 \equiv s^2 \tag{8.10}$$

The expected value of this estimator is

$$E\left[\hat{\sigma}_n^2\right] = E\left[\frac{1}{n}\sum_{i=1}^{n}\left(x^{(i)} - \hat{\mu}_n\right)^2\right] = \frac{n-1}{n}\sigma^2 \tag{8.11}$$

and it is biased:

$$\begin{aligned} Bias\left(\hat{\sigma}_n^2\right) &= E\left[\hat{\sigma}_n^2\right] - \sigma^2 = \\ &= \frac{n-1}{n}\sigma^2 - \sigma^2 = -\frac{\sigma^2}{n} \end{aligned} \tag{8.12}$$

Corrected Sample Variance

A more common estimator of the Gaussian parameter σ^2 is the *corrected sample variance* $\hat{\sigma}_n^{\prime 2}$

$$\hat{\sigma}_n^{\prime 2} = \frac{1}{n-1}\sum_{i=1}^{n}\left(x^{(i)} - \hat{\mu}_n\right)^2 = \frac{n}{n-1}\hat{\sigma}_n^2 \tag{8.13}$$

The expected value of this estimator is

$$E\left[\hat{\sigma}_n^{'2}\right] = \frac{n}{n-1}\,\frac{n-1}{n}\sigma^2 = \sigma^2 \tag{8.14}$$

and it is unbiased:

$$Bias\left(\hat{\sigma}_n^{'2}\right) = E\left[\hat{\sigma}_n^{'2}\right] - \sigma^2 = \sigma^2 - \sigma^2 = 0 \tag{8.15}$$

8.3.5 Variance and standard error of an estimator

The **Variance of an estimator** measures the overall expected deviation -in a squared error sense- between the estimator $\hat{\theta}_n$ and its expected value $E\left[\hat{\theta}_n\right]$

$$Var\left(\hat{\theta}_n\right) = E\left[\left(\hat{\theta}_n - E\left[\hat{\theta}_n\right]\right)^2\right] \tag{8.16}$$

where $\hat{\theta}_n$ is a random variable of pdf $f_\Theta(\theta)$

The **standard error** of an estimator is defined as

$$SE\left(\hat{\theta}_n\right) = \sqrt{Var\left(\hat{\theta}_n\right)} \tag{8.17}$$

Assume to sample the dataset from the unknown underlying generating process $f(x)$ several times, say 30. We will obtain 30 different datasets and correspondingly 30 different values of the estimator $\hat{\theta}_n$. The variance (or standard error) of the estimator provides a measure of its variability with respect to the 30 datasets.

Standard Error of the Sample mean

Consider for example the standard error of the sample mean of the dataset of timber strength, where we have 7 different datasets (each one of them with 16 data), see fig.8.7

In the example, the SE of the sample mean is $SE(\hat{\mu}_7) = 1.83$. Note however that 7 data represent a dataset of small size.

More in general, the standard error of the sample mean is

$$SE\left(\hat{\mu}_n\right) = \sqrt{Var\left[\frac{1}{n}\sum_{i=1}^{n} x^{(i)}\right]} = \frac{\sigma}{\sqrt{n}} \tag{8.18}$$

	Timber strength [MPa]						
1	48.78	32.02	45.54	32.40	48.37	50.98	35.58
2	41.64	39.34	34.12	33.06	29.93	40.71	28.97
3	39.77	46.33	45.92	33.47	36.38	34.63	34.56
4	35.89	46.99	36.47	35.67	46.86	24.84	26.89
5	54.04	22.67	28.98	28.46	36.00	28.83	38.64
6	35.43	39.62	40.85	23.16	23.19	42.31	24.25
7	22.75	44.78	56.60	44.51	36.88	39.33	44.54
8	44.78	26.63	28.76	42.47	44.30	39.93	40.85
9	43.99	43.48	47.42	48.39	44.59	39.60	39.97
10	47.74	30.05	33.61	38.05	44.00	38.16	37.69
11	25.39	30.33	44.36	35.03	40.39	43.33	41.78
12	38.00	39.21	35.30	31.33	41.72	69.07	33.14
13	25.98	51.39	33.18	27.31	29.90	51.90	55.23
14	36.84	50.91	36.85	53.99	35.17	33.71	36.53
15	34.49	49.65	17.98	43.41	34.44	46.50	22.74
16	38.71	47.83	27.90	28.71	27.93	36.92	34.40
m	38.39	40.08	37.12	36.21	37.50	41.30	35.99

	Sample Mean [MPa]
1	38.39
2	40.08
3	37.12
4	36.21
5	37.50
6	41.30
7	35.99
s	1.83

Figure 8.7: Timber, standard error because of sampling variability

n	Sample Mean [MPa]	SE
16	38.39	2.17
32	39.23	1.56
48	38.53	1.30
64	37.95	1.10
80	37.86	0.95
96	38.43	0.90
112	38.08	0.83

Figure 8.8: Timber, standard error in terms of the sample size

being n the number of data.

Therefore, the SE decreases with the sample size. SE is a measure of the precision of the estimator. The approximation is reasonable for large n. With reference to the timber example, see fig.8.8. Here we assume first to have $n = 16$ data, and we get $m = 38.39, s = 2.17$, and coefficient of variation $v = s/m = 0.056$: this means that we have a sample mean $m = 38.39$ with error 5.65%. Later we repeat the same analysis with increasing values of sample size. We $n = 112$ data, and we obtain $m = 38.08, s = 0.83$, and coefficient of variation $v = s/m = 0.021$: this means that we have a sample mean $m = 38.08$ with error 2.18%.

8.3.6 Tradeoff bias-variance and Mean Square Error (MSE) of an estimator

Bias and variance measure two different sources of error in an estimator

Bias of an estimator $\hat{\theta}_n$ measures the expected deviation from the true value of the parameter θ

Variance measures the overall expected deviation -in a squared error sense- between the estimator and its expected value

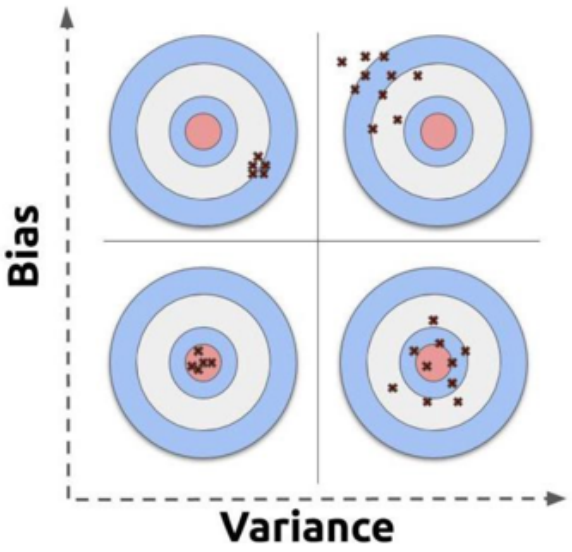


Figure 8.9: Tradeoff bias-variance

In fig.8.9, it can be seen:

- quadrant bottom left, *low-bias low-variance*: the estimator $\hat{\theta}_n$ converges to the true parameter θ and it is not strongly dependent upon the specific dataset;
- quadrant bottom right, *low-bias high-variance*: the estimator $\hat{\theta}_n$ converges to the the true parameter θ but it is strongly dependent upon the specific dataset;
- quadrant top left, *high-bias low-variance*: the estimator $\hat{\theta}_n$ does not converges to the true parameter θ, but it is not strongly dependent upon the specific dataset;
- quadrant top right, *high-bias high-variance*: the estimator $\hat{\theta}_n$ does not converge to the true parameter θ and it is strongly dependent upon the specific dataset

Of course the optimal situation would be an estimator 'low-bias low-variance', however bias and variance are two sources of error in an estimator, and these are typically conflicting each other, so that typically we have: (i) estimators 'low-bias high-variance', (ii)

estimators 'high-bias low-variance'. In the first case we know that the estimator is theoretically correct, but we don't know in advance if with the specific dataset the estimate is good enough. In the second case the estimator is not much dependent upon the dataset, but it could converge to a wrong estimate.

What happens when we meet two different estimators, one with more bias and one with more variance? How do we choose?

The **Mean Squared Error** (MSE) measures *the overall expected deviation -in a squared error sense- between the estimator and the true value of the parameter*

$$MSE\left[\left(\hat{\theta}_n - \theta\right)^2\right] = \left[Bias\left(\hat{\theta}_n\right)\right]^2 + Var\left(\hat{\theta}_n\right) \tag{8.19}$$

Low MSE implies that in any particular sample the estimate $\hat{\theta}_n$ is likely to be close to the true parameter value θ

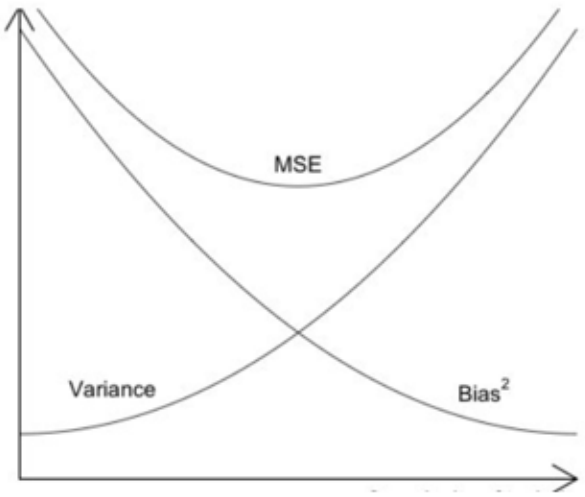


Figure 8.10: Tradeoff bias-variance

8.3.7 Consistency

8.4 Confidence Intervals

8.5 Basics of Bootstrap

8.6 Analysis of Variance

8.7 Maximum Likelihood Estimation (MLE)

A Random Variable is fully characterized from a probabilistic point of view from the knowledge of its Probability Density Function (PDF) $f(x;\theta)$, having m parameters collected in $\theta = \{\theta_1, \theta_2, ..., \theta_m\}$

Given a family of probability distributions $f(x;\theta)$, the likelihood function $L(\theta)$ gives *the probability of the observed data*

$$L(\theta) = L_n(\theta|\mathbf{x}) = f_{\mathbf{X}}(\mathbf{x};\theta) \tag{8.20}$$

where $f_{\mathbf{X}}(\mathbf{x};\theta)$ is the joint PDF of the random variables generating our dataset. Typically it is assumed that the data $x^{(1)}, x^{(2)}, ..., x^{(n)}$ are generated by the random variables $X_1, X_2, ..., X_n$ having all the same PDF $f(x;\theta)$, i.e. $f_{X_1}(x_1;\theta) \equiv f(x;\theta), f_{X_2}(x_2;\theta) \equiv f(x;\theta)$, and so on

For *independent and identically distributed random variables* the likelihood function is

$$L(\theta) = L_n(\theta|\mathbf{x}) = f_{X_1}(x_1;\theta) f_{X_2}(x_2;\theta) ... f_{X_n}(x_n;\theta) = \prod_{i=1}^{n} f(x_i;\theta) \tag{8.21}$$

The principle of Maximum Likelihood (MLE) is to adopt the model with greatest likelihood, since of all the models under consideration, this *assigns the highest probability to the observed data*.

The optimal parameters are defined through an optimization problem:

$$\theta^* = \max_{\theta} L(\theta) \tag{8.22}$$

while the maximum likelihood estimation MLE is

$$MLE = L(\theta^*) \tag{8.23}$$

Summarizing:

- Observe the data (data analysis)
- Choose a model (a family of probability distributions)
- The optimal parameters θ^* of the model maximize the likelihood of the parameters given the data

Usually, it is more convenient to take the logarithm and working with the log-likelihood function

$$l(\theta) = log\ L(\theta) = \sum_{i=1}^{n} log\ f\left(x^{(i)}; \theta\right) \tag{8.24}$$

This because:

- Since the logarithmic function is monotonic, the log-likelihood takes its maximum at the same point as the likelihood function
- The logarithm turns the product into a summation
- Many distributions are exponential, log and exponential cancel each other out, and logs allow to turn exponents in products. This may imply simpler equations to minimize

Often it is used the MLE estimator

$$\hat{l}(\theta) = \frac{1}{n} \sum_{i=1}^{n} log\ f\left(x^{(i)}; \theta\right) \tag{8.25}$$

which, in presence of samples of very great size, becomes

$$l(\theta) = E\left[log\ f\left(x^{(i)}; \theta\right)\right] \tag{8.26}$$

The MLE is largely used by the statisticians for some attractive properties. First, define now the **Fisher Information** as

$$I(\theta) = E\left[\left(\frac{\partial l(\theta)}{\partial \theta}\right)^2\right] = -E\left[\frac{\partial^2 l(\theta)}{\partial \theta^2}\right] \tag{8.27}$$

The variance of any unbiased estimator is bounded by the reciprocal of the Fisher information, known as **Cramer-Rao lower bound**

$$Var\left(\hat{\theta}\right) \geq \frac{1}{I\left(\theta\right)} \tag{8.28}$$

which shows that the precision to which we can estimate $\hat{\theta}$ is fundamentally limited by the Fisher information of the log-likelihood function.

Shortcomings

- *Maximum-likelihood estimators have no optimum properties for finite samples*
- Nothing can be said around the approximation over the tails

Advantages As the sample size increases to infinity, the MLE estimators have some attractive properties:

- Consistency: the sequence of MLEs converges in probability to the value of the parameters being estimated;
- Efficiency: MLE achieves asymptotically the Cramer-Rao Lower bound, this means that *no consistent estimator has lower MSE than MLE*

Case Test 2, Timber 164. Let us consider now the case of 164 data of timber strength.

Gaussian model

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right]$$

To find the parameters of the Gaussian distribution, we use the MLE.

$$\theta^* = \max_{\theta} l\left(\theta|\mathbf{x}\right) = \max_{\theta}\left[\sum_{i=1}^{n} log\ f\left(x^{(i)}|\theta\right)\right] =$$
$$= \max_{\mu,\sigma}\left[\sum_{i=1}^{164} log\ f^G\left(x^{(i)}|\mu,\sigma\right)\right] =$$
$$= \max_{\mu,\sigma}\left[log\left(\frac{1}{\sigma\sqrt{2\pi}}\exp\left[-\frac{\left(x^{(1)}-\mu\right)^2}{2\sigma^2}\right]\right) + \ldots + log\left(\frac{1}{\sigma\sqrt{2\pi}}\exp\left[-\frac{\left(x^{(164)}-\mu\right)^2}{2\sigma^2}\right]\right)\right] =$$
$$= \max_{\mu,\sigma}\left[log\left(\frac{1}{\sigma\sqrt{2\pi}}\exp\left[-\frac{(48.78-\mu)^2}{2\sigma^2}\right]\right) + \ldots + log\left(\frac{1}{\sigma\sqrt{2\pi}}\exp\left[-\frac{(58.34-\mu)^2}{2\sigma^2}\right]\right)\right]$$

and we obtain

$$\begin{cases} \hat{\mu} = 39.32MPa \\ \hat{\sigma} = 9.44MPa \end{cases}$$

The obtained optimal log-likelihood function is:

$$l\left(\hat{\mu},\hat{\sigma}|\mathbf{x}\right) = \sum_{i=1}^{164} log\ f^G\left(x^{(i)}|\hat{\mu},\hat{\sigma}\right) =$$

$$= log\left(\frac{1}{\hat{\sigma}\sqrt{2\pi}}\exp\left[-\frac{\left(x^{(1)}-\hat{\mu}\right)^2}{2\hat{\sigma}^2}\right]\right) + \ldots + log\left(\frac{1}{\hat{\sigma}\sqrt{2\pi}}\exp\left[-\frac{\left(x^{(164)}-\hat{\mu}\right)^2}{2\hat{\sigma}^2}\right]\right) =$$

$$= log\left(\frac{1}{\hat{\sigma}\sqrt{2\pi}}\exp\left[-\frac{\left(48.78-\hat{\mu}\right)^2}{2\hat{\sigma}^2}\right]\right) + \ldots + log\left(\frac{1}{\hat{\sigma}\sqrt{2\pi}}\exp\left[-\frac{\left(58.34-\hat{\mu}\right)^2}{2\hat{\sigma}^2}\right]\right) = -600.95$$

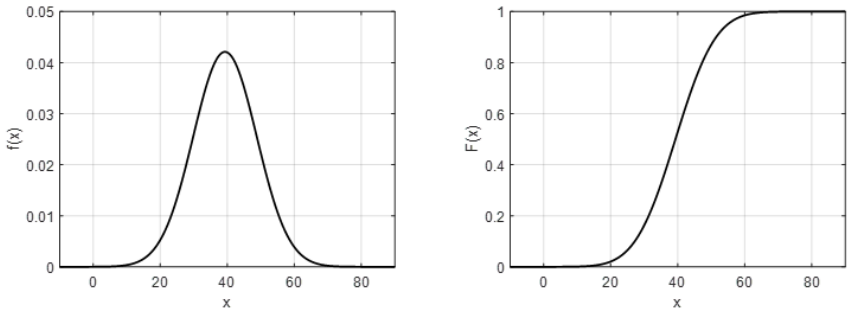


Figure 8.11: Timber, Gaussian distribution, MLE

OpenAIUQ :

- load the dataset collected in the file 'timber164.dat'
- associate the datasat to the object d, belonging to the class 'data1'

```
#=======================================
#TIMBER 164
#=======================================
dataset=np.loadtxt('timber164.dat')
#load the dataset saved in the file 'timber164.dat'

d=auq.data1(dataset)
#d is the object collecting the dataset
#the data are collected in 'd.data
```

- build the Gaussian variable 'x1', belonging to the class 'dist',
- define the domain of the variable through the method *set_range*
- evaluate the parameters of the Gaussian distribution through the method *MLEfit*

```
#=======================================
#Gaussian fit
#=======================================
x1=auq.dist('Gaussian')
#Define x1 as a Gaussian random variable

x1.set_range(xiniz=0,xfin=80,size=100)
#Define the domain of x1, ranging from x=0 to x=80
#this domain is only for representation
#the points are collected in the field 'x1.xx'

x1.MLEfit(d.data)
#Evaluate the parameters of x through MLE
#mean is collected in the field 'x1.mu'
#standard deviation is collected in the field 'x1.dev'
#the MLE is collected in the field 'x1.MLE'
```

Once the parameters are determined, we can evaluate the pdf and the cdf of x1

- build the pdf of the Gaussian variable 'x1', through the method *pdf*,
- build the cdf of the Gaussian variable 'x1', through the method *cdf*,

```
#=======================================
#Evaluate pdf
#=======================================
x1.pdf()
#evaluate the pdf of x1, 'Gaussian', of parameters 'x1.mu' and 'x1.dev'
#the domain is collected in the field 'x1.xx'
#the corresponding pdf is collected in the field 'x1.f'

#=======================================
#Evaluate cdf
#=======================================
x1.cdf()
#evaluate the cdf of x1, 'Gaussian', of parameters 'x1.mu' and 'x1.dev'
#the domain is collected in the field 'x1.xx'
#the corresponding pdf is collected in the field 'x1.F'
```

To plot fig.8.11 the following code is written:

- build the figure 'fig1', one row, two columns
- plot the pdf in the left panel, identified through the object 'ax1'
- plot the cdf in the right panel, identified through the object 'ax2'

```
#=======================================
#Define figure
#=======================================
fig1=plt.figure(num=1,figsize=(12,6),dpi=60)
#num: figure number
#figsize: fugure dimension
#dpi

ax1=fig1.add_subplot(1,2,1)
ax2=fig1.add_subplot(1,2,2)
#plot=1 row, 1 column

#=======================================
#Plot pdf
#=======================================
ax1.plot(x1.xx,x1.f,'r',lw=2)
ax1.set_xlabel('Timber Strength (MPa)')
ax1.set_ylabel('pdf')
ax1.grid(True)
ax1.set_xlim(0,80)
ax1.set_ylim(0)
ax1.set_xticks(np.arange(0,90,10));

#=======================================
#Plot cdf
#=======================================
ax2.plot(x1.xx,x1.F,'r',lw=2)
ax2.set_xlabel('Timber Strength (MPa)')
ax2.set_ylabel('cdf')
ax2.grid(True)
ax2.set_xlim(0,80)
ax2.set_ylim(0,1)
ax2.set_xticks(np.arange(0,90,10));
```

Later we choose the family of lognormal distributions

$$f(x) = \frac{1}{x\sigma_Y\sqrt{2\pi}} \exp\left[-\frac{1}{2}\frac{(\log x - \mu_Y)^2}{\sigma_Y^2}\right]$$

To find the parameters of the Lognormal distribution, we use the MLE.

$$\theta^* = \max_\theta l\left(\theta|\hat{\mathbf{x}}\right) = \max_\theta \left[\sum_{i=1}^{164} log\ f\left(x^{(i)}|\theta\right)\right] =$$
$$= \max_{\mu_Y,\sigma_Y} \left[\sum_{i=1}^{164} log\ f^{LN}\left(x^{(i)}|\mu_Y,\sigma_Y\right)\right] =$$
$$= \max_{\mu_Y,\sigma_Y} \left[log\left(\frac{1}{x\sigma_Y\sqrt{2\pi}} \exp\left[-\frac{\left(\log x^{(1)} - \mu_Y\right)^2}{2\sigma_Y^2}\right]\right) + \ldots + log\left(\frac{1}{x\sigma_Y\sqrt{2\pi}} \exp\left[-\frac{\left(\log x^{(164)} - \mu\right)^2}{2\sigma^2}\right]\right)\right.$$

and we obtain

$$\begin{cases} \hat{\mu}_Y = 3.64MPa \\ \hat{\sigma}_Y = 0.24MPa \end{cases}$$

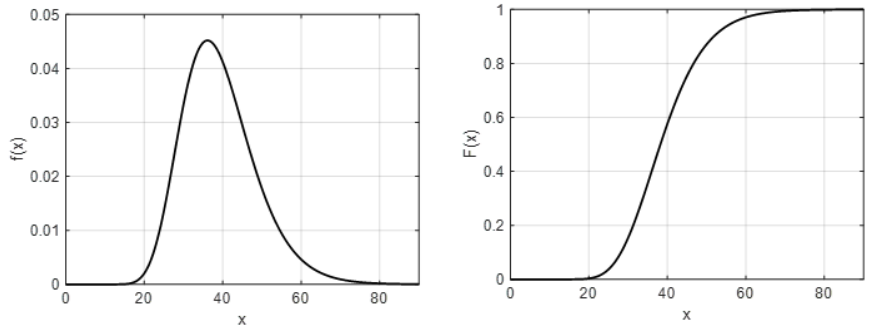


Figure 8.12: Timber, Lognormal distribution, MLE

OpenAIUQ:

- build the Lognormal variable 'x2', belonging to the class 'dist',
- define the domain of the variable
- evaluate the parameters of the Gaussian distribution through the method *MLEfit*

```
#=======================================
#Lognormal fit
#=======================================
x2=auq.dist('Lognormal')
#Define x2 as a Lognormal random variable

x2.set_range(xiniz=0,xfin=80,size=100)
#Define the domain of x2, ranging from x=0 to x=80
#this domain is only for representation
#the points are collected in the field 'x2.xx'

x2.MLEfit(d.data)
#Evaluate the parameters of x2 through MLE
#the parameters 'muY' and 'devY' are collected in 'x2.param'
#mean is collected in the field 'x2.mu'
#standard deviation is collected in the field 'x2.dev'
#the MLE is collected in the field 'x2.MLE'
```

Once the parameters are determined, we can evaluate the pdf and the cdf of $x2$

- build the pdf of the Lognormal variable 'x2', through the method *pdf*,
- build the cdf of the Lognormal variable 'x2', through the method *cdf*,

```
#=======================================
#Evaluate pdf
#=======================================
x2.pdf()
#evaluate the pdf of x2, 'Lognormal', of parameters 'x2.param'
#the domain is collected in the field 'x2.xx'
#the corresponding pdf is collected in the field 'x2.f'

#=======================================
#Evaluate cdf
#=======================================
x2.cdf()
#evaluate the cdf of x2, 'Lognormal', of parameters 'x2.param'
#the domain is collected in the field 'x2.xx'
#the corresponding cdf is collected in the field 'x2.F'
```

To plot fig.8.12 the following code is written:

- build the figure 'fig2', one row, two columns

- plot the pdf in the left panel, identified through the object 'ax1'
- plot the cdf in the right panel, identified through the object 'ax2'

```
#=======================================
#Define figure
#=======================================
fig2=plt.figure(num=1,figsize=(12,6),dpi=60)
#num: figure number
#figsize: fugure dimension
#dpi

ax1=fig2.add_subplot(1,2,1)
ax2=fig2.add_subplot(1,2,2)
#plot=1 row, 1 column

#=======================================
#Plot pdf
#=======================================
ax1.plot(x2.xx,x2.f,'r',lw=2)
ax1.set_xlabel('Timber Strength (MPa)')
ax1.set_ylabel('pdf')
ax1.grid(True)
ax1.set_xlim(0,80)
ax1.set_ylim(0)
ax1.set_xticks(np.arange(0,90,10));

#=======================================
#Plot cdf
#=======================================
ax2.plot(x2.xx,x2.F,'r',lw=2)
ax2.set_xlabel('Timber Strength (MPa)')
ax2.set_ylabel('cdf')
ax2.grid(True)
ax2.set_xlim(0,80)
ax2.set_ylim(0,1)
ax2.set_xticks(np.arange(0,90,10));
```

8.8 Basics of Information Theory and Minimum Divergence Estimation (MDE)

Main idea of the information theory is that the value of an event depends on the degree of surprise that the event carries

A likely event is not surprising, therefore it is not so interesting. Conversely it is much more informative to learn about an event unlikely to occur

- Probability to win the lottery: $p_1 = 0.001$
- Probability of not winning the lottery: $p_2 = 1 - p_1 = 1 - 0.001 = 0.999$

Of course the first event is more informative: if in a lottery there are 2 winning tickets, and no-one was sold, this is not informative event. If conversely, if I know that one winning ticket has been sold, this is informative.

Uniform distribution carries no information since it does not help to predict the result of a future test. *A dirac's delta* $f(x) = \delta(x - x0)$ *is highly informative* since it is possible to predict with almost absolute certainty that a future test will yield the value $x = x_0$. The amount of information carried by any other distribution is between the uniform distribution and Dirac's delta.

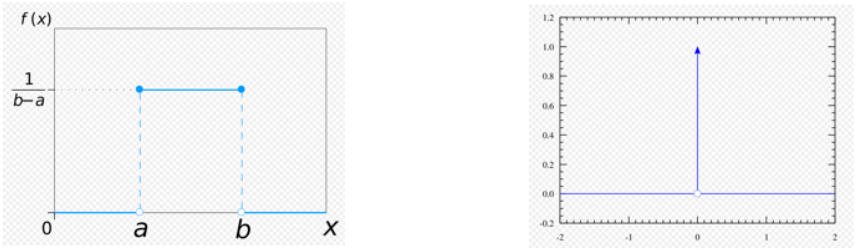


Figure 8.13: Information carried by distributions

Properties of information

- $I(p \to 0) \to \infty$: Rare events give high amount of information;
- $I(p) \geq 0$;
- $I(p = 1) = 0$: events certain do not provide information;
- $I(p_1 p_2) = I(p_1) + I(p_2)$

All these properties are owned by the following definition of **information**

$$I(p) = \log\left(\frac{1}{p}\right) = -\log(p) \tag{8.29}$$

8.8.1 Shannon's Entropy

Consider a discrete random variable X with probability distribution $p_1, p_2, ..., p_n$ evaluated at $x_1, x_2, ..., x_n$.

The **Shannon's entropy** functional is its *expected information content*

$$H(X) = E[I(X)] = E[-\log p(X)] = -\sum_{i=1}^{n} p_i \log p_i \tag{8.30}$$

The entropy is larger when the random variable is more random, i.e. more unpredictable. The entropy is a *measure of uncertainty*.

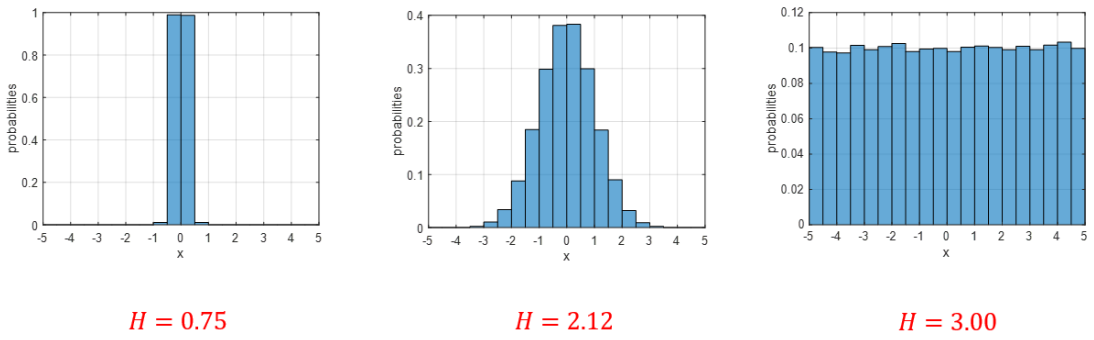


Figure 8.14: Shannon's entropy

In figure we see histograms drawn from datasets having same sample mean ($m = 0$) but different degree of uncertainty. It is seen that the entropy is greater in presence of increased uncertainty content.

8.8.2 Differential Entropy

Consider a continuous random variable X with PDF $f(x)$.

The differential entropy functional is

$$H(X) = -\int f(x) \log f(x)\, dx = -E[\log f(X)] \tag{8.31}$$

8.8.3 Relative Entropy (KL Divergence)

The Kullback-Leibler (KL) divergence (or relative entropy) measures the entropy difference between the PDFs $f_1(x)$ and $f_2(x)$ of two random variables X_1 and X_2.

$$D(X_1 \parallel X_2) = -\int f_1(x) \log \left[\frac{f_1(x)}{f_2(x)}\right] dx \tag{8.32}$$

The relative entropy functional can be considered *a measure of the distance between two distributions*

8.8.4 Minimum Divergence Estimation (MDE)

It is assumed that the sample data $x^{(1)}, x^{(2)}, ..., x^{(n)}$ are generated from an unknown target PDF $f(x)$. Our aim is approximating $f(x)$ with a model PDF $f_m(x;\theta)$, having m parameters collected in $\theta = \{\theta_1, \theta_2, ..., \theta_m\}$

Given a family of probability distributions $f(x;\theta)$, the KL Divergence $D(f \parallel f_m)$ gives *the lost information in representing the target PDF* $f(x)$ *with the model* $f_m(x;\theta)$

The principle of Minimum Divergence (MDE) is to adopt the model with minimum divergence, since of all the models under consideration, this *assigns the minimum lost information.*

The optimal parameters are defined through an optimization problem:

$$\begin{aligned}
\theta^* &= \min_\theta D\left[f(x) \parallel f_m(x);\theta)\right] = \\
&= \min_\theta \left[\int f(x) \log\left[\frac{f(x)}{f_m(x)}\right] dx\right] = \\
&= \min_\theta \left[\int f(x) \log f(x) dx - \int f(x) \log f_m(x;\theta)\, dx\right] = \\
&= \max_\theta \int f(x) \log\left[f_m(x;\theta)\right] dx = \\
&= \max_\theta E\left[\log f_m(x;\theta)\right] = \\
&\approx \max_\theta \left(\frac{1}{n}\sum_{i=1}^{n} \log\left[f_m\left(x^{(i)};\theta\right)\right]\right) = \\
&= \max_\theta \hat{l}(\theta)
\end{aligned} \tag{8.33}$$

Therefore, *the MDE is equivalent to the Maximum Likelihood Estimation (MLE) of the model.*

The optimal divergence D_m reads as

$$\begin{aligned}
D_m &= D\left(f \parallel f_m^*\right) \\
&= \left[\int f(x) \log f(x) dx - \int f(x) \log f_m(x;\theta^*)\, dx\right] \\
&= -H(f) - E\left[\log f_m(x;\theta^*)\right] \\
&\propto E\left[\log f_m(x;\theta^*)\right] \\
&\approx \frac{1}{n}\sum_{i=1}^{n} \log\left[f_m\left(x^{(i)};\theta^*\right)\right]
\end{aligned} \tag{8.34}$$

Summarizing:

- Observe the data (data analysis)
- Choose a model (a family of probability distributions)
- The optimal parameters θ^* of the model maximize the likelihood of the parameters given the data
- The MLE optimal parameters θ^* of the model minimize also the divergence, so the chosen distribution is optimal in the sense of the information theory

8.9 Method of the Moments

A Random Variable is fully characterized from a probabilistic point of view from the knowledge of its Probability Density Function (PDF) $f(x;\theta)$, whose m parameters are collected in the vector $\theta = \{\theta_1, \theta_2, ..., \theta_m\}$

However, in general, we know only the data (as discussed in the case of the MLE) or other derived quantities (e.g. moments) of the random variable

As discussed previously, to determine with good accuracy the tails of the distributions, it is necessary to know the higher-order moments $M \geq 2$. Therefore, it is crucial the knowledge of methods robust and effective for evaluating the PDF of a random variable from the knowledge of its moments

This is the classical **moment problem** of mathematical analysis:

Find $f(x)$ belonging to the class $f(x;\theta)$, given its moments

$$\begin{cases} \mu_0(\theta) = \int f(x;\theta)\,dx \\ \mu_1(\theta) = \int x f(x;\theta)\,dx \\ \mu_2(\theta) = \int x^2 f(x;\theta)\,dx \\ \mu_3(\theta) = \int x^3 f(x;\theta)\,dx \\ \mu_4(\theta) = \int x^4 f(x;\theta)\,dx \\ \dots \end{cases} \tag{8.35}$$

where $\mu_k(\theta)$ are the **population moments**.

If the distribution family were known, the population moments would be known in closed form in terms of the distribution parameters.

Gaussian distribution of parameters μ and σ, the moments are

$$\begin{cases} \mu_1 = E[X] = \mu \\ \mu_2 = E[X^2] = \mu^2 + \sigma^2 \\ \mu_3 = E[X^3] = \mu^3 + 3\mu\sigma^2 \\ \mu_4 = E[X^4] = \mu^4 + 6\mu^2\sigma^2 + \sigma^4 \\ \dots \end{cases} \tag{8.36}$$

The central moments are

$$\begin{cases} \mu'_1 = E\left[(X-\mu)\right] = 0 \\ \mu'_2 = E\left[(X-\mu)^2\right] = \sigma^2 \\ \mu'_3 = E\left[(X-\mu)^3\right] = \gamma_1\sigma^3 = 0 \\ \mu'_4 = E\left[(X-\mu)^4\right] = \gamma_2\sigma^4 = 3\sigma^4 \\ \dots \end{cases} \tag{8.37}$$

where γ_1 and γ_2 are the coefficient of skewness and kurtosis, respectively. For normal standard distribution, of parameters $\mu = 0$ and $\sigma = 1$ moments and central moments coincide

$$\begin{cases} \mu_1 = E\left[X\right] = 0 \\ \mu_2 = E\left[X^2\right] = 1 \\ \mu_3 = E\left[X^3\right] = \gamma_1 = 0 \\ \mu_4 = E\left[X^4\right] = \gamma_2 = 3 \\ \dots \end{cases} \tag{8.38}$$

Lognormal distribution of parameters μ_Y and σ_Y the moments are given by the following relation

$$\mu_k = E\left[X^k\right] = \exp\left(k\mu_Y + \frac{1}{2}k^2\sigma_Y^2\right) \tag{8.39}$$

and therefore the first moments are

$$\begin{cases} \mu_1 = E\left[X\right] = \exp\left(\mu_Y + \frac{1}{2}\sigma_Y^2\right) \\ \mu_2 = E\left[X^2\right] = \exp\left(2\mu_Y + 2\sigma_Y^2\right) \\ \mu_3 = E\left[X^3\right] = \exp\left(3\mu_Y + \frac{9}{2}\sigma_Y^2\right) \\ \mu_4 = E\left[X^4\right] = \exp\left(4\mu_Y + 8\sigma_Y^2\right) \\ \dots \end{cases} \tag{8.40}$$

But we have the data $x^{(1)}, x^{(2)}, ..., x^{(n)}$, from which we can derive the sample moments

$$\begin{cases} m_1 = \frac{1}{n}\sum_{i=1}^{n} x^{(i)} \\ m_2 = \frac{1}{n}\sum_{i=1}^{n} \left[x^{(i)}\right]^2 \\ m_3 = \frac{1}{n}\sum_{i=1}^{n} \left[x^{(i)}\right]^3 \\ m_4 = \frac{1}{n}\sum_{i=1}^{n} \left[x^{(i)}\right]^4 \\ \dots \end{cases} \tag{8.41}$$

Method of the moments

- Choose a family of distributions (e.g. Gaussian, Lognormal, etc.) of parameters $\theta = (\theta_1, \theta_2, ..., \theta_m)$
- Define the population moments $\mu_k(\theta)$ in terms of the (unknown) parameters θ
- Evaluate the sample moments m_k from the data
- By equating $\mu_k(\theta) = m_k$ find the parameters θ as solution of a system of non-linear equations
- Find the specific distribution of parameters $\hat{\theta_1}, \hat{\theta_2}, ..., \hat{\theta_m}$ belonging to the family

Case test2, timber 164. Consider now the case of the available 164 data of timber strength

Gaussian distribution.

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right]$$

To find the parameters μ and σ of the Gaussian distribution, we use the method of the moments.

In such case, the sample mean and the sample standard deviation are

$$\begin{cases} m = 39.32MPa \\ s = 9.44MPa \end{cases}$$

and the parameters of the Gaussian distribution are then

$$\begin{cases} \mu \equiv m = 39.32MPa \\ \sigma \equiv s = 9.44MPa \end{cases}$$

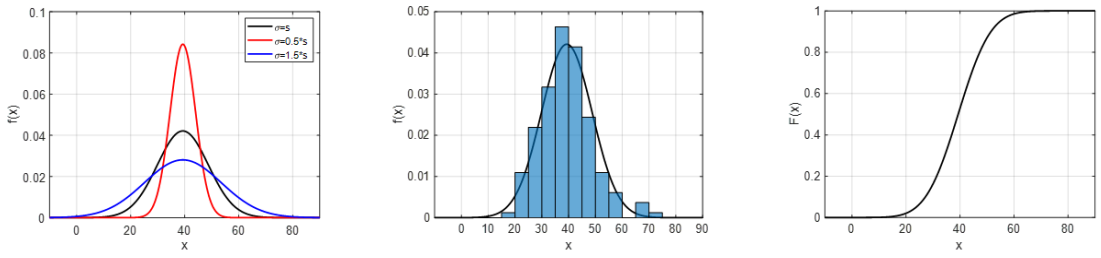


Figure 8.15: Timber, Gaussian distribution, Method of the moments

OpenAIUQ :

- load the dataset collected in the file 'timber164.dat'
- associate the dataset to the object d, belonging to the class 'data1'

```
#=======================================
#TIMBER 164
#=======================================
dataset=np.loadtxt('timber164.dat')
#load the dataset saved in the file 'timber164.dat'

d=auq.data1(dataset)
#d is the object collecting the dataset
#the data are collected in 'd.data
```

- build the Gaussian variable 'x1', belonging to the class 'dist',
- define the domain of the variable
- evaluate sample mean and sample standard deviation values are stored in 'd.mean' and 'd.std'
- evaluate the parameters of the Gaussian distribution through the method *Momentfit*

```
#=======================================
#Gaussian fit
#=======================================
x1=auq.dist('Gaussian')
#Define x1 as a Gaussian random variable

x1.set_range(xiniz=0,xfin=80,size=100)
#Define the domain of x1, ranging from x=0 to x=80
#this domain is only for representation
#the points are collected in the field 'x1.xx'

m=d.mean
s=d.std
#m=sample mean
#s=sample standard deviation

x1.Momentfit(m,s)
#Evaluate the parameters of x through Moment Method
#mean is collected in the field 'x1.mu'
#standard deviation is collected in the field 'x1.dev'
```

- build the pdf of the Gaussian variable 'x1', through the method *pdf*,
- build the cdf of the Gaussian variable 'x1', through the method *cdf*,

```
#=======================================
#Evaluate pdf
#=======================================
x1.pdf()
#evaluate the pdf of x1, 'Gaussian', of parameters 'x1.mu' and 'x1.dev'
#the domain is collected in the field 'x1.xx'
#the corresponding pdf is collected in the field 'x1.f'

#=======================================
#Evaluate cdf
#=======================================
x1.cdf()
#evaluate the cdf of x1, 'Gaussian', of parameters 'x1.mu' and 'x1.dev'
#the domain is collected in the field 'x1.xx'
#the corresponding pdf is collected in the field 'x1.F'
```

To plot fig.8.15 the following code is written:

- build the figure 'fig1', one row, two columns
- plot the pdf in the left panel, identified through the object 'ax1'
- plot the cdf in the right panel, identified through the object 'ax2'

```
#=======================================
#Define figure
#=======================================
fig1=plt.figure(num=1,figsize=(12,6),dpi=60)
#num: figure number
#figsize: fugure dimension
#dpi

ax1=fig1.add_subplot(1,2,1)
ax2=fig1.add_subplot(1,2,2)
#plot=1 row, 1 column

#=======================================
#Plot pdf
```

```
#=======================================
ax1.plot(x1.xx,x1.f,'r',lw=2)
ax1.set_xlabel('Timber Strength (MPa)')
ax1.set_ylabel('pdf')
ax1.grid(True)
ax1.set_xlim(0,80)
ax1.set_ylim(0)
ax1.set_xticks(np.arange(0,90,10));

#=======================================
#Plot cdf
#=======================================
ax2.plot(x1.xx,x1.F,'r',lw=2)
ax2.set_xlabel('Timber Strength (MPa)')
ax2.set_ylabel('cdf')
ax2.grid(True)
ax2.set_xlim(0,80)
ax2.set_ylim(0,1)
ax2.set_xticks(np.arange(0,90,10));
```

Lognormal distribution.

$$f(x) = \frac{1}{x\sigma_Y\sqrt{2\pi}} \exp\left[-\frac{1}{2}\frac{(\log x - \mu_Y)^2}{\sigma_Y^2}\right]$$

To find the parameters μ_Y and σ_Y of the Lognormal distribution, we use the Method of the Moments.

The sample mean, sample standard deviation and sample coefficient of variation are

$$\begin{cases} m = 39.32MPa \\ s = 9.47MPa \\ v = \frac{s}{m} = 0.24 \end{cases}$$

and the parameters of the Lognormal distribution are evaluated as

$$\begin{cases} \mu_Y = \log m - \frac{1}{2}\sigma_Y^2 = 3.64MPa \\ \sigma_Y = \sqrt{\log\left(1+v^2\right)} = 0.23MPa \end{cases}$$

Of course, to get a better approximation of the tails, it would be useful to consider also the higher-order moments, but their estimates are accurate only in presence of a very high number of sample data.

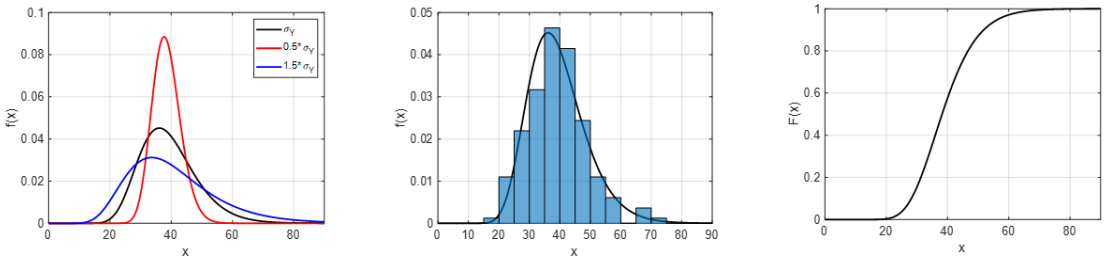


Figure 8.16: Timber, Lognormal distribution, Method of the moments

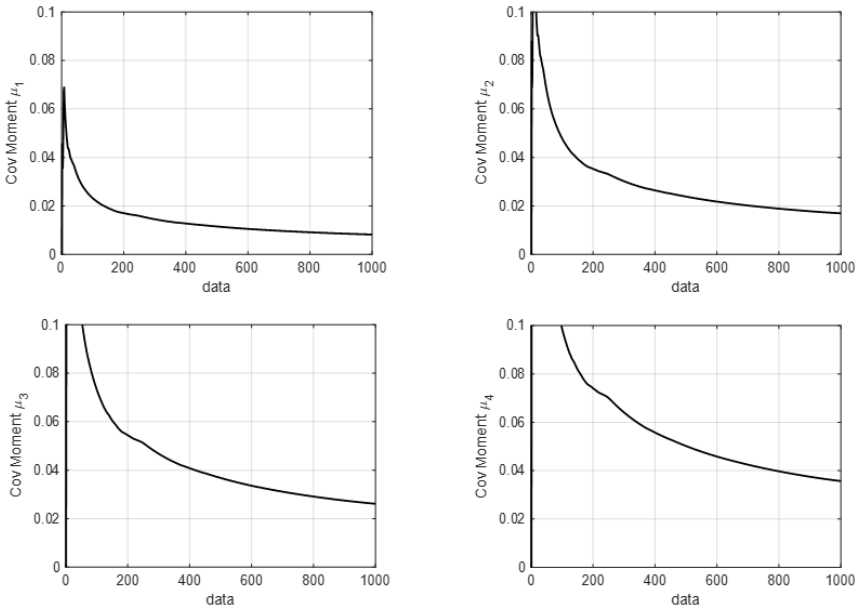


Figure 8.17: Sampling variability of sample moments

In fig.8.17 it is seen for example that 200 samples provide an estimate of the sample moment with a coefficient of variation below 5% only for the first two moments.

Advantages

- the estimators with the method of the moments are easy to be obtained and they are *consistent*;
- from a theoretical point of view, through inclusion of higher-order moments, it is possible to describe the tails of the distributions

Shortcomings

- the higher-order sample moments are estimator typically *unbiased*, but they have

very high variance

- if the sample size n is small (say $n < 30$ for the first two moments, or $n < 100$ if higher-order moments are considered) the sample moments provide an inaccurate estimate of the population moments
- The method of the moments does not give an estimate of the statistical uncertainties

8.10 Maximum Entropy distribution

8.11 Kernel Density Maximum Entropy (KDME)

Chapter 9

Lecture A3:Model Selection

Model building may be seen to consist of five steps:

- Assessment and statistical quantification of the available data
- selection of distribution function
- Estimation of distribution parameters
- model verification and selection
- model updating

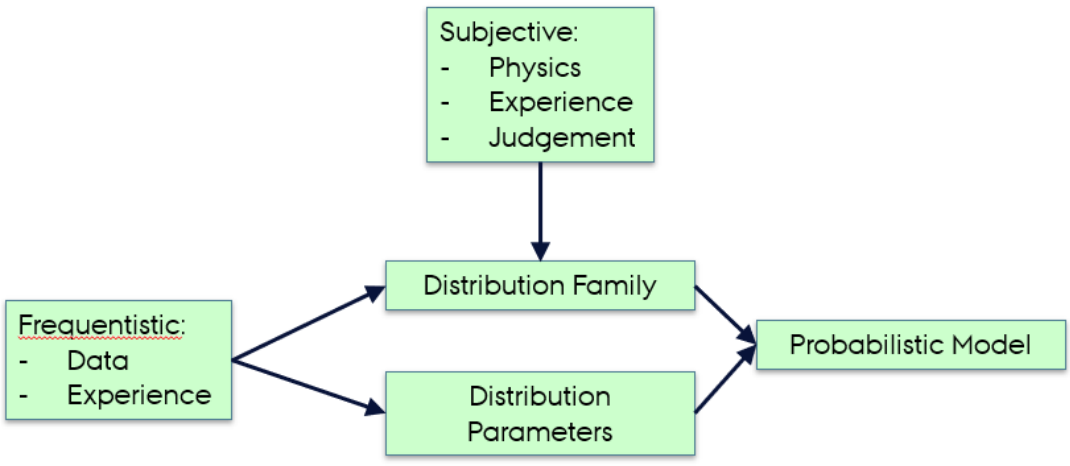


Figure 9.1: Model Building

We have previously already discussed:(i) data analysis (chapter 2), (ii) selection of distribution function (chapter 6), (iii) estimation of distribution parameters (chapter 6). In this chapter we will discuss about *model verification and selection.*

9.1 Model diagnostics

9.2 Graphical representation and visual inspection

9.2.1 Probability Plot

Assume now that we have a set of data $x^{(1)}, x^{(2)}, ..., x^{(n)}$. At first we order the data, and we obtain $x_{(1)}, x_{(2)}, ..., x_{(n)}$ such that $x_{(1)}$ is the minimum value, and $x_{(n)}$ is the maximum value of the dataset. The *empirical CDF* is defined as

$$F_{emp}(x_{(i)}) = \frac{i}{n+1} \tag{9.1}$$

so that the minimum value of the dataset has a probability

$$F_{emp}(x_{(1)}) = \frac{1}{n+1}$$

close to zero, while the maximum value has a probability

$$F_{emp}(x_{(n)}) = \frac{n}{n+1}$$

close to 1.

Assume now that we have chosen a family distribution and we have determined its parameters $\theta_1, \theta_2, ..., \theta_m$. Therefore, we have uniquely determined the PDF $f_m(x)$ and the CDF $F_m(x)$, where the pedix m here means 'model'. It is therefore possible to determine the CDF in all the points of the ordered dataset

$$F_m(x_{(i)}) = F\left(x_{(i)}|\theta_1, \theta_2, ..., \theta_m\right)$$

In the Probability Plot, for any point of the dataset, we represent in the x-axis the model CDF $F_m(x)$ and in the y-axis the empirical CDF $F_{emp}(x)$.

If the data are distributed along a line, then the fit of the model is reasonably good.

Case Test 1, Concrete20. Consider again the 20 sample data of concrete strength.

Data analysis. The sample mean and standard deviation are

$$\begin{cases} m = 32.07MPa \\ s = 4.04MPa \end{cases}$$

Concrete strengths (Mpa)			
Data	Ordered	F	F_{emp}
1	24.4	0.02	0.048
2	27.6	0.105	0.095
3	27.8	0.114	0.143
4	27.9	0.119	0.190
5	28.5	0.151	0.238
6	30.1	0.263	0.286
7	30.3	0.279	0.333
8	31.7	0.405	0.381
9	32.2	0.454	0.429
10	32.8	0.513	0.476
11	33.3	0.562	0.524
12	33.5	0.581	0.571
13	34.1	0.638	0.619
14	34.6	0.683	0.667
15	35.8	0.781	0.714
16	35.9	0.788	0.762
17	36.8	0.846	0.810
18	37.1	0.863	0.857
19	39.2	0.947	0.905
20	39.7	0.959	0.952

Figure 9.2: Concrete strength, 20 data

In fig.9.2 we represent in the second column the ordered data $x_{(1)}, x_{(2)}, ...x_{(20)}$ and in the last columns the empirical CDF, see eq.(9.1) so that

$$\begin{cases} F_{emp}(24.4) = \frac{1}{21} = 0.048 \\ F_{emp}(27.6) = \frac{2}{21} = 0.095 \\ ... \\ F_{emp}(39.7) = \frac{20}{21} = 0.952 \end{cases}$$

Gaussian distribution.

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right]$$

To find the parameters μ and σ of the Gaussian distribution, we use the method of the moments. The parameters are

$$\begin{cases} \mu \equiv m = 32.07MPa \\ \sigma \equiv s = 4.04MPa \end{cases}$$

and it is possible to determine the CDF $F_m(x)$ of the Gaussian model

$$\begin{cases} F_m(24.4) = 0.020 \\ F_m(27.6) = 0.105 \\ ... \\ F_m(39.7) = 0.959 \end{cases}$$

Probability Plot. In the probability plot we represent for each sample of the dataset the pair $[F_m(x_{(i)}), F_{emp}(x_{(i)})], i = 1, 2, ..., 20$ Although 20 data represent a sample of small

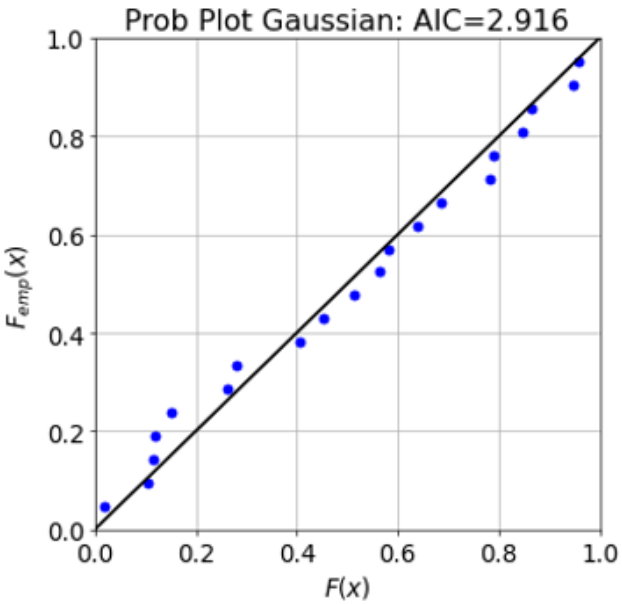


Figure 9.3: Concrete strength, probability plot

size, the probability plot shows that the Gaussian distribution could be a good fit.

9.2.2 Q-Q Plot

Assume now that we have a set of data $x^{(1)}, x^{(2)}, ..., x^{(n)}$. At first we order the data, and we obtain $x_{(1)}, x_{(2)}, ..., x_{(n)}$ such that $x_{(1)}$ is the minimum value, and $x_{(n)}$ is the maximum

value of the dataset. The *empirical quantile* coincides with the ordered data, that is

$$x_{emp,i} = x_{(i)} \tag{9.2}$$

Assume now that we have chosen a family distribution and we have determined its parameters $\theta_1, \theta_2, ..., \theta_m$. Therefore, we have uniquely determined the PDF $f_m(x)$, the CDF $F_m(x)$, and the quantile function $x_m(p) = F_m^{-1}(p)$ where the pedix m here means 'model', while p is a chosen value of probability. In particular we are interested in determining the quantiles corresponding to the probabilities defined by the empirical CDF, see eq.(9.1)

$$x_m\left[F_{emp}(x_{(i)})\right] = F_m^{-1}\left(F_{emp}\left(x_{(i)}\right)|\theta_1, \theta_2, ..., \theta_m\right)$$

In the Q-Q Plot, for any point of the dataset, we represent in the x-axis the quantile of the model $x_m(F_{emp})$ and in the y-axis the empirical quantile x_{emp}.

If the data are distributed along a line, then the fit of the model is reasonably good.

In general, *the Q-Q plot puts emphasis on the quality of the fit of the tails of the model.*

Case Test 1, Concrete20. Consider again the 20 sample data of concrete strength.

Concrete strengths (Mpa)				
Data	Ordered	F_{emp}	F^{-1}	$x_{(i)}$
1	24.4	0.048	25.91	24.4
2	27.6	0.095	27.37	27.6
3	27.8	0.143	28.34	27.8
4	27.9	0.190	29.12	27.9
5	28.5	0.238	29.78	28.5
6	30.1	0.286	30.37	30.1
7	30.3	0.333	30.92	30.3
8	31.7	0.381	31.44	31.7
9	32.2	0.429	31.93	32.2
10	32.8	0.476	32.42	32.8
11	33.3	0.524	32.9	33.3
12	33.5	0.571	33.39	33.5
13	34.1	0.619	33.89	34.1
14	34.6	0.667	34.4	34.6
15	35.8	0.714	34.95	35.8
16	35.9	0.762	35.54	35.9
17	36.8	0.810	36.21	36.8
18	37.1	0.857	36.98	37.1
19	39.2	0.905	37.96	39.2
20	39.7	0.952	39.41	39.7

Figure 9.4: Concrete strength, 20 data, table for Q-Q plot

Data analysis. In figure we represent in the second column the ordered data $x_{(1)}, x_{(2)}, ...x_{(20)}$,

in the third column the empirical CDF, and in the last columns the empirical quantile

$$\begin{cases} x_{emp,1} = 24.4MPa \\ x_{emp,2} = 27.6MPa \\ ... \\ x_{emp,20} = 39.7MPa \end{cases}$$

Gaussian distribution. The parameters of the Gaussian distribution are

$$\begin{cases} \mu \equiv m = 32.07MPa \\ \sigma \equiv s = 4.04MPa \end{cases}$$

where m and s are the sample mean and standard deviation of the sample, respectively. It is so possible to determine the CDF $F_m(x)$ of the Gaussian model and therefore its quantile $x_m(F)$

$$\begin{cases} x_{m,1} = F_m^{-1}(0.048) = 25.91 \\ x_{m,2} = F_m^{-1}(0.095) = 27.37 \\ ... \\ x_{m,20} = F_m^{-1}(0.952) = 39.41 \end{cases}$$

Q-Q Plot. In the Q-Q plot we represent for each sample of the dataset the pair $[x_{m,i}, x_{emp,i}], i = 1, 2, ..., 20$. Although 20 data represent a sample of small size, the Q-Q plot could trigger some doubts about the quality of the left tail.

9.2.3 Example

Case test 2, timber 164. Consider again the case of 164 data of timber strength. OpenAIUQ:

- load the dataset collected in the file 'timber164.dat'
- associate the dataset to the object d, belonging to the class 'data1'

```
#=======================================
#TIMBER 164
#=======================================
dataset=np.loadtxt('timber164.dat')
#load the dataset saved in the file 'timber164.dat'
```

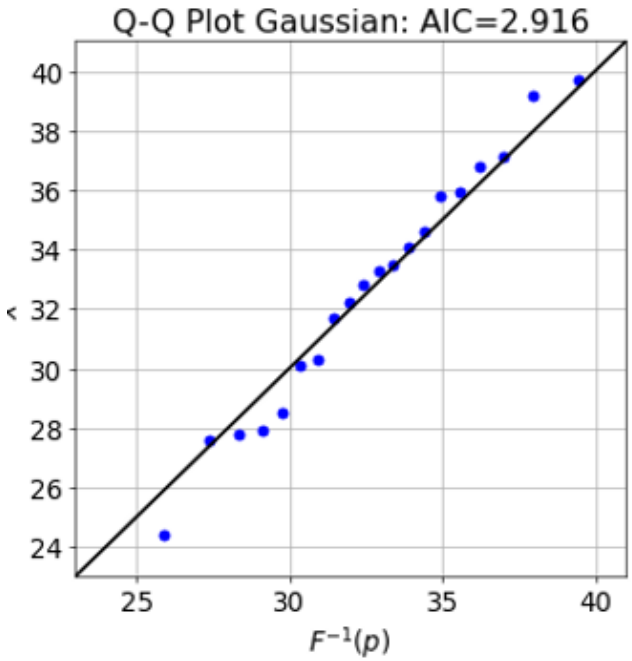


Figure 9.5: Concrete strength, Q-Q plot

```
d=auq.data1(dataset)
#d is the object collecting the dataset
#the data are collected in 'd.data
```

- build the Gaussian variable 'x1', belonging to the class 'dist',
- define the domain of the variable
- evaluate the parameters of the Gaussian distribution through the method *MLEfit*

```
#=======================================
#Gaussian fit
#=======================================
x1=auq.dist('Gaussian')
#Define x1 as a Gaussian random variable

x1.set_range(xiniz=0,xfin=80,size=100)
#Define the domain of x1, ranging from x=0 to x=80
#this domain is only for representation
```

```
#the points are collected in the field 'x1.xx'

x1.MLEfit(d.data)
#Evaluate the parameters of x through MLE
#mean is collected in the field 'x1.mu'
#standard deviation is collected in the field 'x1.dev'
#the MLE is collected in the field 'x1.MLE'
```

- build the probability plot 'fig1', belonging to the class 'distplot',
- plot in the same figure probability plot and Q-Q plot through the method *prob_plot_w2*

```
#========================================
#Probability plot
#========================================
fig1=auq.distplot(num=1,figsize=(12,6),dpi=60)
#fig1 is the object associated to the plot

fig1.prob_plot_w2(d.data,x1)
#include prob plot and q-q plot
#input: dataset collected in 'd.data' and the variable x1
```

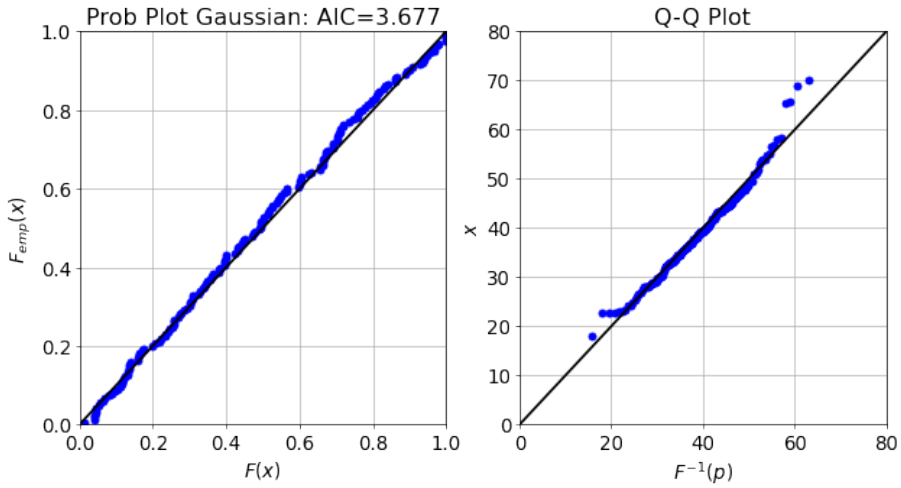


Figure 9.6: Timber strength, Gaussian fit, probability plot

The probability plot (left panel) shows that the Gaussian fit is good enough, but the Q-Q plot (right panel) shows that the approximation over the tails is not excellent.

The analysis is repeated with a Lognormal distribution.

- build the Lognormal variable 'x2', belonging to the class 'dist',
- define the domain of the variable
- evaluate the parameters of the Lognormal distribution through the method *MLEfit*
- build the probability plot 'fig2', belonging to the class 'distplot',
- plot in the same figure probability plot and Q-Q plot through the method *prob_plot_w2*

```
#=======================================
#Lognormal fit
#=======================================
x2=auq.dist('Lognormal')
#Define x2 as a Lognormal random variable

x2.set_range(xiniz=0,xfin=80,size=100)
#Define the domain of x1, ranging from x=0 to x=80
#this domain is only for representation
#the points are collected in the field 'x1.xx'

x2.MLEfit(d.data)
#Evaluate the parameters of x1 through MLE
#mean is collected in the field 'x2.mu'
#standard deviation is collected in the field 'x2.dev'
#the MLE is collected in the field 'x2.MLE'

#=======================================
#Probability plot
#=======================================
fig2=auq.distplot(num=2,figsize=(12,6),dpi=60)
#fig2 is the object associated to the plot

fig2.prob_plot_w2(d.data,x2)
#include prob plot and q-q plot
#input: dataset collected in 'd.data' and the variable x2
```

The probability plot (left panel) and the Q-Q plot show that the Lognormal fit outperforms the Gaussian distribution, especially at the tails of the distribution.

As a further test, we plot the CDF of the Gaussian fit. There is greater interest in knowing the left tail, since the timber strength is representative of a resistance, and the structural safety can be affected by low values of strength.

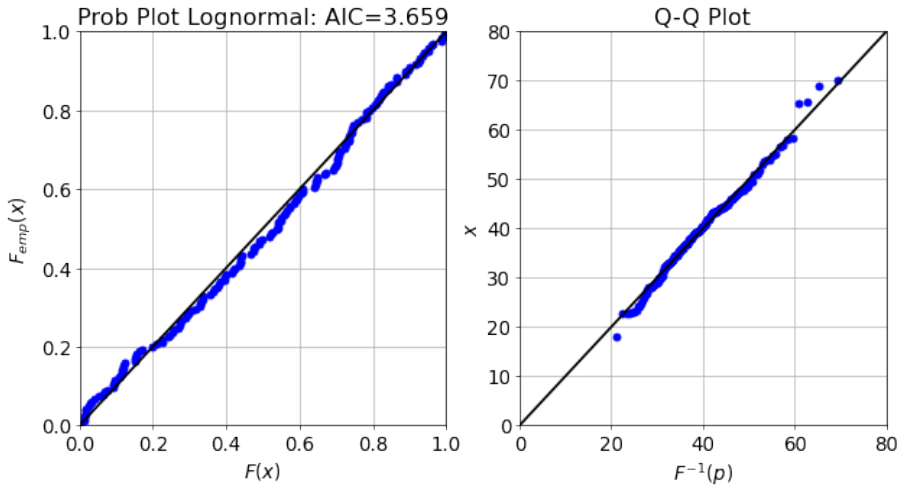


Figure 9.7: Timber strength, Lognormal fit, probability plot

The figure shows that in this case the Gaussian fit may accept some negative values of the timber strength, and this has no physical meaning.

Summary:

- The probability plots of Gaussian and Lognormal show that both distribution can be considered
- The Q-Q plots show that the Lognormal performs better than Gaussian, specifically with reference to the tails
- The Gaussian distribution can accept negative values of timber strength, and this has no physical meaning

Therefore, the Lognormal is chosen as optimal model.

9.3 Information Theory for model selection

It has been already discussed the tradeoff bias-variance for data-driven evaluation of the statistical estimators. Similar situation arises for model selection, as below discussed.

9.3.1 Tradeoff Bias-Variance

Bias is an error in the data-driven model $f_m(x)$ which is defined as the difference between the predicted values from the model and the actual values.

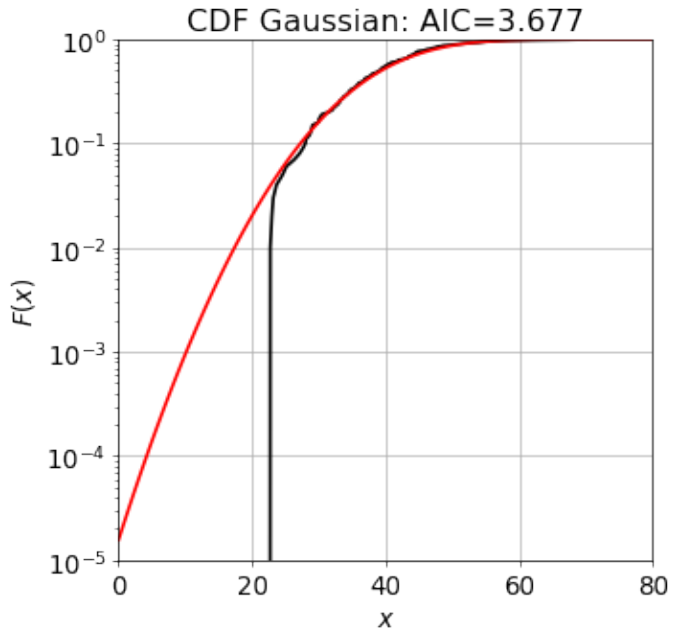


Figure 9.8: Timber strength, Gaussian fit, CDF in logscale

Variance is an other source of error in the data-driven model defined as the models sensitivity to the data.

Bias and variance come typically from the **model complexity**. As a consequence of the tradeoff bias-variance for model selection we have the following situations:

- **Underfitting**: the model is unable to fit the training data and as a result, will be unable to generalize new data.
- **Overfitting**: the model describe the training data too well; it will be unable to generalize new data.

In fig.9.9 we show a situation where we fit some data (black points) through three different models (linear, quadratic, higher-order polynomial) and we want to check which model predicts better the *unseen* data (red point).

Assume now that the true (unknown) function (generating the points) is quadratic, i.e.

$$f(x) = a_0 + a_1 x + a_2 x^2$$

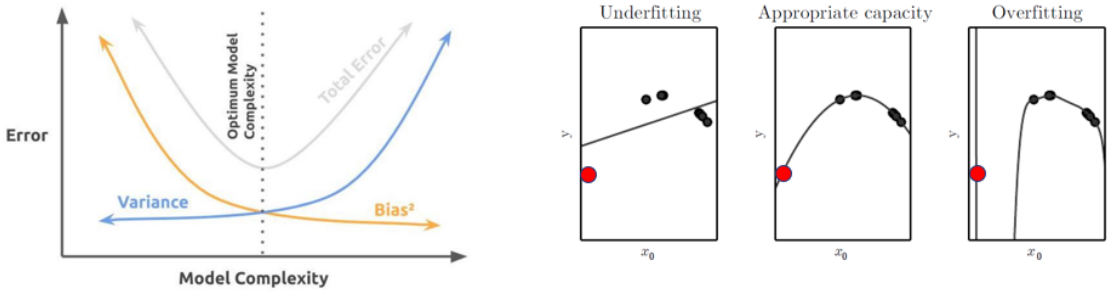


Figure 9.9: Tradeoff bias-variance, model complexity

We don't know $f(x)$, but we know 5 points (black points) generated: $P_1 \equiv (x_1, y_1), P_2 \equiv (x_2, y_2), P_3 \equiv (x_3, y_3), P_4 \equiv (x_4, y_4), P_5 \equiv (x_5, y_5)$

Let us try to approximate the true unknown function through:
Linear model $f_{m,1}(x)$

$$f_{m,1}(x) = w_0 + w_1 x$$

Quadratic model $f_{m,2}(x)$

$$f_{m,2}(x) = w_0 + w_1 x + w_2 x^2$$

Higher-order polynomial model $f_{m,3}(x)$

$$f_{m,3}(x) = w_0 + w_1 x + w_2 x^2 + w_3 x^3 + w_4 x^4$$

The following situation arises:

- Linear model $f_{m,1}(x)$, *high-bias low-variance*: the model is too simple, it is not capable to *interpolate* well the available data ('high-bias'), but its *generalization capabilities over unseen data* are poor too, so it is prone to *underfitting*
- higher-order polynomial model $f_{m,3}(x)$, *low-bias high-variance*: the model is too complex, it *interpolates* perfectly the available data, but its *generalization capabilities over unseen data* are poor, so it is prone to *overfitting*
- Quadratic model $f_{m,2}(x)$, *tradeoff bias-variance*: the model is simple enough, so may not be capable to *interpolate* perfectly the available data, but its *generalization capabilities over unseen data* are good enough

Occam's Razor: the simplest model capable to describe the available data, it is likely to be the right one.

9.3.2 Akaike Information Criterion (AIC) for model selection

Let us remember now that the *divergence* D_m of a model provides a information-related measure of distance between the target PDF $f(x)$ with the chosen model $f_m(x;\theta^*)$. But it does not take into account the model complexity.

AIC measures the lost information in representing the unknown PDF $f(x)$ with the chosen model $f_m(x;\theta^*)$.

$$AIC_m = \frac{m}{n} - D_m = \frac{m}{n} - \frac{1}{n}\sum \log\left[f_m\left(x^{(i)};\theta^*\right)\right] \tag{9.3}$$

where D_m is the divergence, while m is the number of parameters. Therefore, it allows to determine the optimal model, according to the information theory and incorporating issues of model complexity.

AIC for model selection

- Choose a first family model ($m = 1$) $f_1(x;\theta)$ (e.g. Gaussian, Lognormal, etc.)
- Evaluate the parameters $\hat{\theta}$ of the model through MLE and determine the specific distribution belonging to the chosen family $f_1(x) = f_1(x|\hat{\theta})$ fitting the data
- Determine the divergence D_1 and AIC_1
- repeat the same procedure for different families of distributions, $f_2(x;\theta), f_3(x;\theta), ...$
- the optimal distribution has the minimum AIC value

$$f_{opt}(x) = \min_m AIC_m = \min\{AIC_1, AIC_2, ...\}$$

Case Test2, Timber 164. Consider again the case of 164 data of timber strength.

- Choose as a first model ($m = 1$) the **Gaussian distribution** $f_1(x;\theta) \equiv f_G(x;\mu,\sigma)$
- Evaluate the parameters $\hat{\mu}, \hat{\sigma}$ through MLE and determine the specific Gaussian distribution $f_1(x) = f_G(x|\hat{\mu},\hat{\sigma})$ fitting the data
- Determine the divergence $D_1 \equiv D_G$ and $AIC_1 \equiv AIC_G$
- Choose as a second model ($m = 2$) the family of **Lognormal distributions** $f_2(x;\theta) \equiv f_{LN}(x;\mu_Y,\sigma_Y)$
- Evaluate the parameters $\hat{\mu}_Y, \hat{\sigma}_Y$ through MLE and determine the specific Lognormal distribution $f_2(x) = f_{LN}(x|\hat{\mu_Y},\hat{\sigma_Y})$ fitting the data
- Determine the divergence $D_2 \equiv D_{LN}$ and $AIC_2 \equiv AIC_{LN}$

$$\begin{cases} AIC_1 = 3.676 \\ AIC_2 = 3.658 \end{cases}$$

It follows that Lognormal is a better fit, according to the AIC criterion, since $AIC_2 < AIC_1$. This is confirmed by the probability plot and Q-Q plot of the two distributions, see fig.9.10.

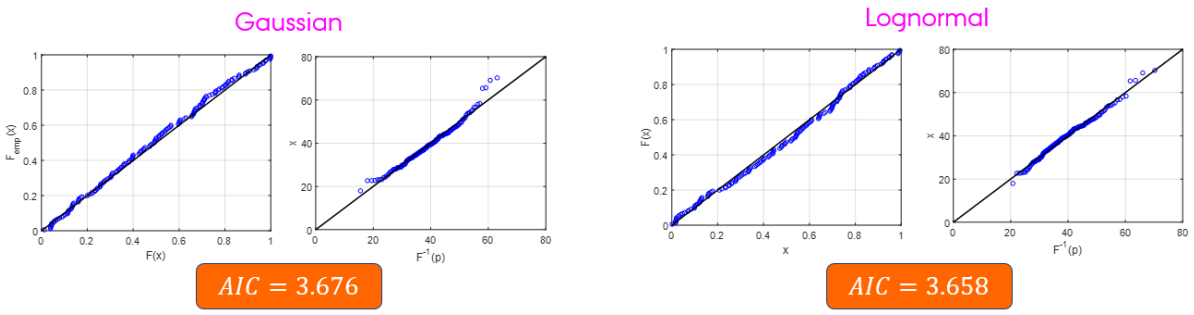


Figure 9.10: Timber strength, AIC and probability plot

9.4 Basics of Machine Learning (ML) for model selection

It is assumed that the data $x^{(1)}, x^{(2)}, ..., x^{(n)}$ are generated from an unknown target PDF $f(x)$. Our aim is evaluating the "best" model (over a selected set of models) fitting the data.

The main idea behind machine learning and *statistical learning theory* is to split the original dataset (n data) into:

- *training set* $\mathbf{x}_{tr}$: the n_{tr} data are used for finding the parameters of the distribution (typically n_{tr} is around the 70% of the original dataset)
- *test set* $\mathbf{x}_{ts}$: the remaining n_{ts} data (with $n_{tr} + n_{ts} = n$) are used for checking the prediction capabilities of the models

In this way the prediction capabilities over the unseen data are not estimated through theoretical considerations, but it is the dataset itself to provide information in regard

In the ML approach some assumptions are implicitly considered:

- all the n data are *independent and identically distributed*, and all of them are generated by the same target PDF $f(x)$

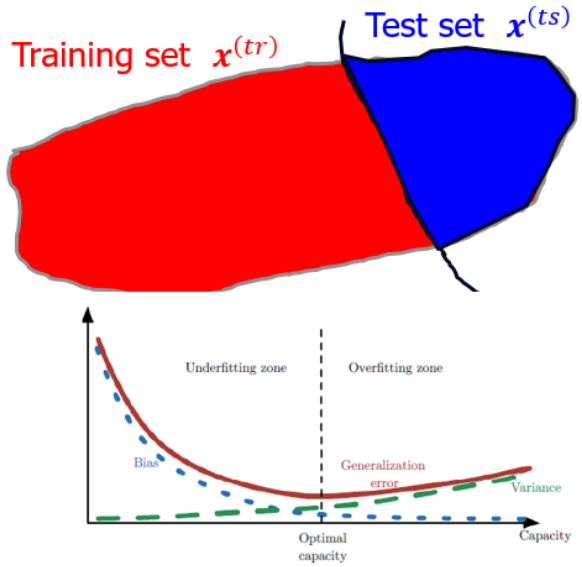


Figure 9.11: Training and test set

- the test set $\mathbf{x}_{ts}$ is removed from the dataset as a first step of the analysis; no 'contamination' between data of training set and test set should exist. This to avoid any kind of bias in the model selection. The parameters of all the different models (chosen for comparison) are evaluated on the same training set $\mathbf{x}_{tr}$

- the test set $\mathbf{x}_{ts}$ represents data that we will get in the future. Since the models are compared through the test set, it should be a representative sample (e.g. $n_{ts} \geq 30$)

Machine Learning for model selection

- Split the dataset (n data) in *training set* $\mathbf{x}_{tr}$ with n_{tr} data and *test set* $\mathbf{x}_{ts}$ with n_{ts} data

- Choose a first model ($m = 1$) $f_m(x;\theta)$ (e.g. Gaussian, Lognormal, etc.)

- Considering only the data of the training set $x_{tr}^{(1)}, x_{tr}^{(2)}, ..., x_{tr}^{(n_{tr})}$ evaluate the parameters of the model $\hat{\theta}_{tr}$ through MLE and determine the specific probability model $f_1(x) = f_m(x|\hat{\theta}_{tr})$ fitting the data

- Evaluate the *Divergence Training set*:

$$D_{1,tr} = D\left[f\left(x|x_{tr}\right) \parallel f_1\left(x|x_{tr}\right)\right] \propto -E\left[\log f_1\left(x|x_{tr}\right)\right] =$$
$$= -\frac{1}{n_{tr}} \sum_{i=1}^{n_{tr}} \log f_1\left(x_{tr}^{(i)}\right)$$

which represents the relative entropy (i.e. the 'distance' in terms of information theory) between the training set $\mathbf{x}_{tr}$, generated from the unknown target PDF $f(x)$, and the model $f_1(x)$ whose parameters are calibrated over the training set $\mathbf{x}_{tr}$

- Evaluate the *Divergence Test set*;

$$D_{1,ts} \equiv D_1 = D\left[f\left(x|x_{ts}\right) \parallel f_1\left(x|x_{tr}\right)\right] \propto -E\left[\log f_1\left(x|x_{tr}\right)\right] =$$
$$= -\frac{1}{n_{ts}} \sum_{i=1}^{n_{ts}} \log f_1\left(x_{ts}^{(i)}\right)$$

which represents the relative entropy (i.e. the 'distance' in terms of information theory) between the test set $\mathbf{x}_{ts}$, generated from the same unknown target PDF $f(x)$, and the model $f_1(x)$ whose parameters are calibrated over the training set $\mathbf{x}_{tr}$. Of course for model selection, the divergence test set $D_{m,ts}$ is to be analyzed.

- repeat the same procedure for different families of distributions, $f_2(x;\theta), f_3(x;\theta), ...$
- the optimal distribution has the minimum value of Divergence Test Set

$$f_{opt}(x) = \min_m D_m = \min\{D_1, D_2, ..., D_M\}$$

Case Test2, Timber 164.

- Split the dataset (n data) in *training set* (n_{tr} data) and *test set* (n_{ts} data)
- Choose as a first model ($m = 1$) the **Gaussian distribution** $f_1(x;\theta) \equiv f_G(x|\mu,\sigma)$
- Considering only the data of the training set $x_{tr}^{(1)}, x_{tr}^{(2)}, ..., x_{tr}^{(n_{tr})}$ evaluate the parameters $\hat{\mu}_{tr}, \hat{\sigma}_{tr}$ through MLE and determine the specific Gaussian distribution $f_1(x) = f_G(x|\hat{\mu}_{tr}, \hat{\sigma}_{tr})$ fitting the data
- Determine the divergence D_1 with respect to the test set $x_{ts}^{(1)}, x_{ts}^{(2)}, ..., x_{ts}^{(n_{ts})}$

$$D_1 = D\left[f\left(x|x_{ts}\right) \parallel f_1\left(x|x_{tr}\right)\right] = \quad \propto -E\left[\log f_1\left(x|x_{tr}\right)\right] =$$
$$= -\frac{1}{n_{ts}} \sum_{i=1}^{n_{ts}} \log f_1\left(x_{ts}^{(i)}\right)$$

- Choose as a second model ($m = 2$) the **Lognormal distribution** $f_2(x;\theta) \equiv f_{LN}(x;\mu_Y,\sigma_Y)$
- Considering only the data of the training set $x_{tr}^{(1)}, x_{tr}^{(2)}, ..., x_{tr}^{(n_{tr})}$ evaluate the parameters $\hat{\mu}_{Ytr}, \hat{\sigma}_{Ytr}$ through MLE and determine the specific Lognormal distribution $f_2(x) = f_{LN}(x|\hat{\mu}_{Ytr}, \hat{\sigma}_{Ytr})$ fitting the data

- Determine the divergence D_2 with respect to the test set $x_{ts}^{(1)}, x_{ts}^{(2)}, ..., x_{ts}^{(n_{ts})}$

$$D_2 = D\left[f\left(x|x_{ts}\right) \parallel f_2\left(x|x_{tr}\right)\right] = \propto -E\left[\log f_2\left(x|x_{tr}\right)\right] =$$
$$= -\frac{1}{n_{ts}} \sum_{i=1}^{n_{ts}} \log f_2\left(x_{ts}^{(i)}\right)$$

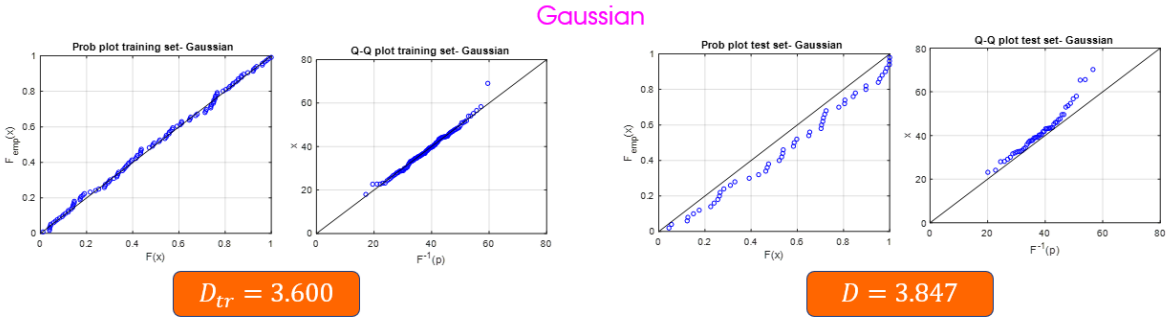


Figure 9.12: Timber strength, Machine Learning, Gaussian fit

In fig.9.12 we represent for the Gaussian fit the divergence over training set and test set. It is seen that for the Gaussian fit

$$\begin{cases} D_{1,tr} = 3.600 \\ D_{1,ts} = 3.847 \end{cases}$$

$D_{1,ts} > D_{1,tr}$ is fully expected since the parameters of the model have been trained over the training set. The same analysis is repeated for Lognormal distribution:

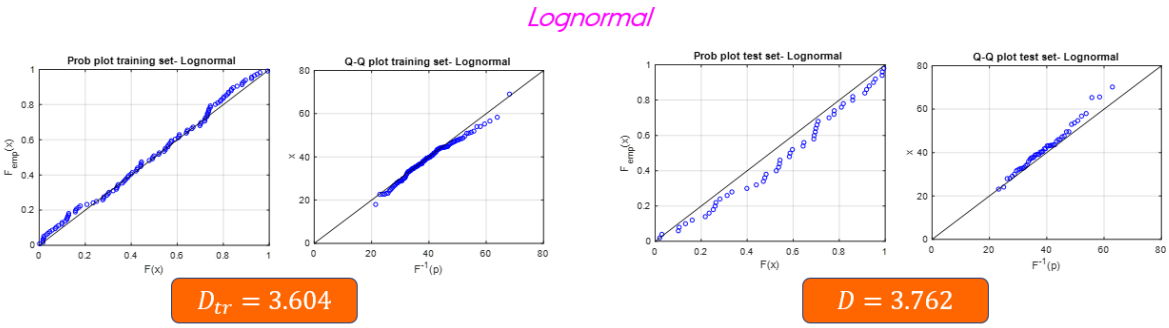


Figure 9.13: Timber strength, Machine Learning, Lognormal fit

$$\begin{cases} D_{2,tr} = 3.604 \\ D_{2,ts} = 3.762 \end{cases}$$

As before and expected, $D_{2,ts} > D_{2,tr}$. It is interesting to note that while $D_{1,tr} \approx D_{2,tr}$, their predictive capabilities appear different enough, $D_{2,ts} < D_{1,ts}$. This shows that also ML confirms the good performances of the Lognormal.

Summarizing:

- The probability plot and Q-Q plot show that the Lognormal is a better fit than the Gaussian
- The CDF in semilog-scale shows that the Gaussian model can accept negative values of the left tail, so it is not a good fit from a physical point of view
- The AIC (based on the information theory) shows that the Lognormal outperforms the Gaussian
- Machine Learning confirms that the predictive capabilities of the Lognormal are better than the Gaussian.

So, for this very specific example, the Lognormal is to be considered the preferred choice.

Chapter 10

Lecture A4:Multiple Random Variables

10.1 Multiple random variables

10.1.1 Conditional distributions

10.1.2 Independent random variables

10.1.3 Means and Moments

10.1.4 Covariances

10.1.5 Coefficient of Correlation

10.2 Multivariate Gaussian distribution

Consider now a vector of n Gaussian random variables $\mathbf{X} = \{X_1, X_2, ..., X_n\}$

The expected value $\mu_{\mathbf{X}}$ is

$$\mu_{\mathbf{X}} = E[\mathbf{X}] = \begin{Bmatrix} \mu_{X_1} \\ \mu_{X_2} \\ \cdots \\ \mu_{X_n} \end{Bmatrix} \tag{10.1}$$

The matrix of covariance $\mathbf{\Sigma_{XX}}$ is

$$\mathbf{\Sigma_{XX}} = E\left[(\mathbf{X}-\mu_{\mathbf{X}})(\mathbf{X}-\mu_{\mathbf{X}})^T\right] = \begin{bmatrix} \sigma_{X_1}^2 & \sigma_{X_1X_n} & ... & \sigma_{X_1X_2} \\ \sigma_{X_1X_2} & \sigma_{X_2}^2 & ... & \sigma_{X_2X_n} \\ ... & & & \\ \sigma_{X_1X_n} & \sigma_{X_2X_n} & ... & \sigma_{X_n}^2 \end{bmatrix} \tag{10.2}$$

The matrix of correlation $\mathbf{R}$ is

$$\mathbf{R} = \begin{bmatrix} 1 & \rho_{X_1X_2} & ... & \rho_{X_1X_n} \\ \rho_{X_1X_2} & 1 & ... & \rho_{X_2X_n} \\ ... & & & \\ \rho_{X_1X_n} & \rho_{X_2X_n} & ... & 1 \end{bmatrix} \tag{10.3}$$

The multivariate Gaussian distribution reads as

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)\sqrt{|\mathbf{\Sigma_{XX}}|}} \exp\left\{ -\frac{1}{2}(\mathbf{x}-\mu_{\mathbf{X}})^T \mathbf{\Sigma_{XX}^{-1}} (\mathbf{x}-\mu_{\mathbf{X}}) \right\} \tag{10.4}$$

In the following, we collect the standard deviations of the random variables in the vector $\sigma_{\mathbf{X}}$

$$\sigma_{\mathbf{X}} = \begin{Bmatrix} \sigma_{X_1} \\ \sigma_{X_2} \\ ... \\ \sigma_{X_n} \end{Bmatrix} \tag{10.5}$$

Define now the diagonal matrix $\mathbf{D}$ collecting the standard deviations

$$\mathbf{D} = \begin{bmatrix} \sigma_{X_1} & 0 & ... & 0 \\ 0 & \sigma_{X_2} & ... & 0 \\ ... & & & \\ 0 & 0 & ... & \sigma_{X_n} \end{bmatrix} \tag{10.6}$$

it is seen that the matrix of covariance can be described as

$$\mathbf{\Sigma_{XX}} = \mathbf{DRD} \tag{10.7}$$

Bivariate Gaussian distribution

Consider now a vector of two Gaussian random variables $\mathbf{X} = \{X_1, X_2\}$. The variable X_1 has mean value μ_{X_1}, standard deviation σ_{X_1}, while mean and standard of X_2 are μ_{X_2} and σ_{X_2}. It follows

$$\mu_{\mathbf{X}} = \begin{Bmatrix} \mu_{X_1} \\ \mu_{X_2} \end{Bmatrix}, \qquad \sigma_{\mathbf{X}} = \begin{Bmatrix} \sigma_{X_1} \\ \sigma_{X_2} \end{Bmatrix}, \qquad \mathbf{D} = \begin{bmatrix} \sigma_{X_1} & 0 \\ 0 & \sigma_{X_2} \end{bmatrix}$$

The matrix of correlation is

$$\mathbf{R} = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$$

and the matrix of correlation is

$$\mathbf{\Sigma_{XX}} = \begin{bmatrix} \sigma^2_{X_1} & \sigma_{X_1X_2} \\ \sigma_{X_1X_2} & \sigma^2_{X_2} \end{bmatrix} = \begin{bmatrix} \sigma^2_{X_1} & \rho\sigma_{X_1}\sigma_{X_2} \\ \rho\sigma_{X_1}\sigma_{X_2} & \sigma^2_{X_2} \end{bmatrix}$$

and it is also given as

$$\mathbf{\Sigma_{XX}} = \mathbf{DRD} = \begin{bmatrix} \sigma_{X_1} & 0 \\ 0 & \sigma_{X_2} \end{bmatrix} \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \begin{bmatrix} \sigma_{X_1} & 0 \\ 0 & \sigma_{X_2} \end{bmatrix}$$

Case 1: Bivariate Gaussian, uncorrelated, different variance

$$\mu_{\mathbf{X}} = \begin{Bmatrix} 1 \\ -1 \end{Bmatrix}, \qquad \sigma_{\mathbf{X}} = \begin{Bmatrix} 0.949 \\ 0.548 \end{Bmatrix}, \qquad \mathbf{R} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

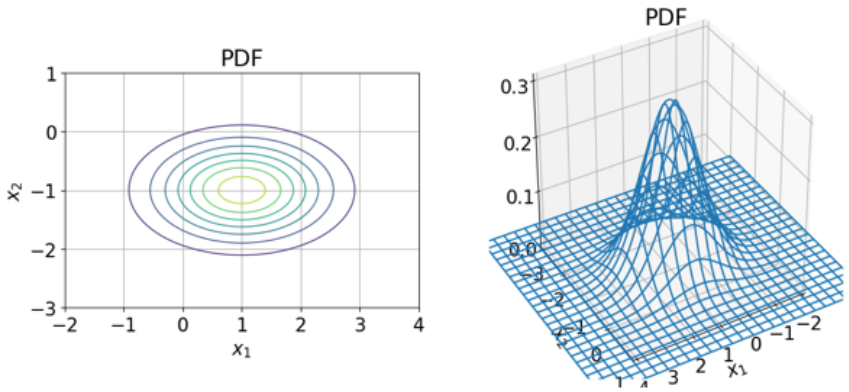


Figure 10.1: Gaussian, uncorrelated, different variance

In such case each contour level of the joint PDF is an ellipse:

- the center of the ellipse has coordinate (μ_{X_1}, μ_{X_2})

- the axes of the ellipse are parallel to the axes of the variables x_1 and x_2
- the semi-axes of the ellipse are proportional to the standard deviation σ_{X_1} and σ_{X_2}

Case 2: Bivariate Gaussian, uncorrelated, same variance

$$\mu_{\mathbf{X}} = \begin{Bmatrix} 1 \\ -1 \end{Bmatrix}, \qquad \sigma_{\mathbf{X}} = \begin{Bmatrix} 0.949 \\ 0.949 \end{Bmatrix}, \qquad \mathbf{R} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

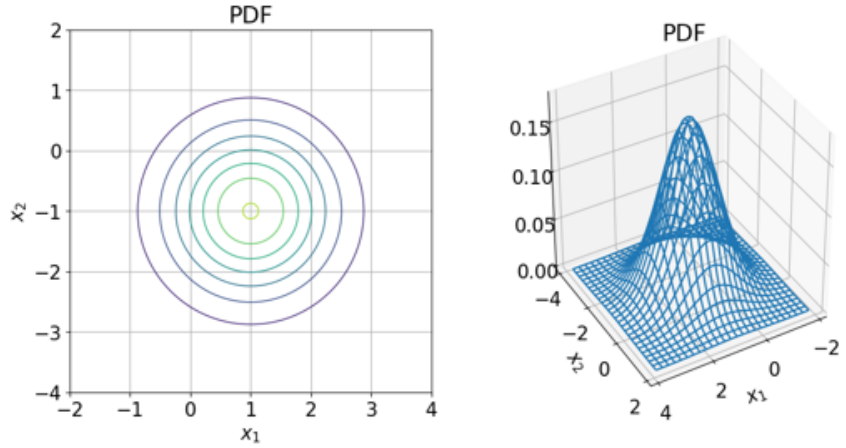


Figure 10.2: Gaussian, uncorrelated, different variance

In such case each contour level of the joint PDF is a circle:

- the center of the circle has coordinate (μ_{X_1}, μ_{X_2})
- there is polar symmetry with respect to the mean value $\mu_{\mathbf{X}}$
- the radius of the cirle is proportional to the standard deviation $\sigma_{X_1} \equiv \sigma_{X_2}$

Case 3: Bivariate Gaussian, correlated

$$\mu_{\mathbf{X}} = \begin{Bmatrix} 1 \\ -1 \end{Bmatrix}, \qquad \sigma_{\mathbf{X}} = \begin{Bmatrix} 0.949 \\ 0.548 \end{Bmatrix}, \qquad \mathbf{R} = \begin{bmatrix} 1 & 0.77 \\ 0.77 & 1 \end{bmatrix}$$

In such case each contour level of the joint PDF is an ellipse:

- the center of the ellipse has coordinate (μ_{X_1}, μ_{X_2})
- the axes of the ellipse are not parallel to the axes of the variables x_1 and x_2 and their slope is related to the correlation ρ
- the semi-axes of the ellipse are proportional to the standard deviation σ_{X_1} and σ_{X_2}

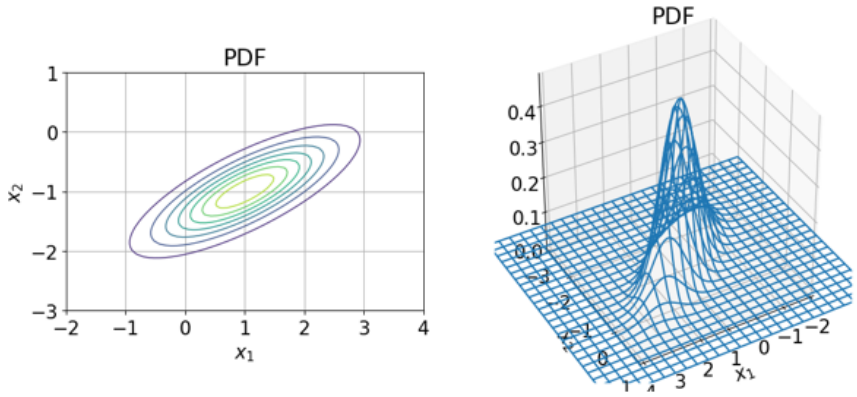


Figure 10.3: Gaussian, correlated, different variance

10.2.1 Multivariate Normal Standard distribution

A vector of n Gaussian random variables $\mathbf{U} = \{U_1, U_2, ..., U_n\}$ follow a multivariate normal standard distribution if:

- the random variables $U_1, U_2, ..., U_n$ follow a normal standard distribution
- they are uncorrelated

Therefore, $U_1, U_2, ..., U_n$ are *independent*.

The expected value $\boldsymbol{\mu}_\mathbf{U}$ and the standard devision $\boldsymbol{\sigma}_\mathbf{U}$ are

$$\mu_\mathbf{U} = E[\mathbf{U}] = \begin{Bmatrix} \mu_{U_1} \\ \mu_{U_2} \\ ... \\ \mu_{U_n} \end{Bmatrix} = \begin{Bmatrix} 0 \\ 0 \\ ... \\ 0 \end{Bmatrix}, \qquad \sigma_\mathbf{U} = \begin{Bmatrix} \sigma_{U_1} \\ \sigma_{U_2} \\ ... \\ \sigma_{U_n} \end{Bmatrix} = \begin{Bmatrix} 1 \\ 1 \\ ... \\ 1 \end{Bmatrix} \tag{10.8}$$

The matrix of covariance $\boldsymbol{\Sigma}_{\mathbf{UU}}$ and the matrix of correlation $\mathbf{R}$ coincice with the identity matrix

$$\boldsymbol{\Sigma}_{\mathbf{UU}} = E\left[\mathbf{U}\mathbf{U}^T\right] = \begin{bmatrix} \sigma^2_{U_1} & \sigma_{U_1U_2} & ... & \sigma_{U_1U_n} \\ \sigma_{U_1U_2} & \sigma^2_{U_2} & ... & \sigma_{U_2U_n} \\ ... & & & \\ \sigma_{U_1U_n} & \sigma_{U_2U_n} & ... & \sigma^2_{U_n} \end{bmatrix} = \begin{bmatrix} 1 & 0 & ... & 0 \\ 0 & 1 & ... & 0 \\ ... & & & \\ 0 & 0 & ... & 1 \end{bmatrix} \equiv \mathbf{R} \tag{10.9}$$

The multivariate normal standard distribution is

$$\varphi_n(\mathbf{u}) = \frac{1}{(2\pi)^{n/2}} \exp\left(-\frac{1}{2}\mathbf{u}^T\mathbf{u}\right) \tag{10.10}$$

Since $U_1, U_2, ..., U_n$ are independent

$$\varphi_n(\mathbf{u}) = \prod_{i=1}^{n} \varphi(u_i) = \varphi(u_1)\varphi(u_2)...\varphi(u_n) \tag{10.11}$$

where $\varphi(u)$ is the univariate normal standard PDF

$$\varphi(u) = \frac{1}{2\pi} \exp\left(-\frac{1}{2}u^2\right) \tag{10.12}$$

Bivariate Normal standard

$$\mu_{\mathbf{U}} = \begin{Bmatrix} 0 \\ 0 \end{Bmatrix}, \qquad \sigma_{\mathbf{U}} = \begin{Bmatrix} 1 \\ 1 \end{Bmatrix}, \qquad \mathbf{R} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

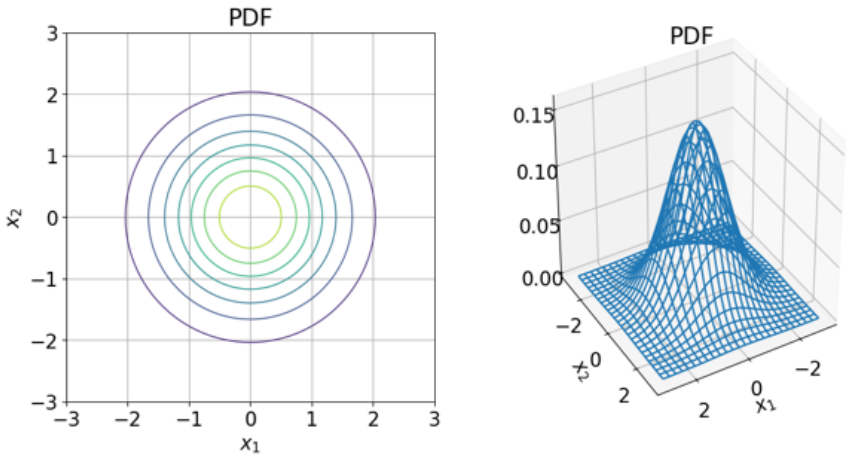


Figure 10.4: bivariate normal standard PDF

In such case each contour level of the joint PDF is a circle:

- the center is in the origin $(0, 0)$
- there is polar symmetry with respect to the origin
- the radius of the circle is $\sigma = 1$
- every diameter of the circle is a normal standard PDF

10.2.2 Coordinate transformation toward normal standard PDF - Case of Uncorrelated Gaussian random variables

Consider now a vector of n uncorrelated Gaussian random variables $\mathbf{X} = \{X_1, X_2, ..., X_n\}$ The expected value $\mu_{\mathbf{X}}$ and the matrix of covariance $\mathbf{\Sigma_{XX}}$ are

$$\mu_{\mathbf{X}} = E[\mathbf{X}] = \begin{Bmatrix} \mu_{X_1} \\ \mu_{X_2} \\ ... \\ \mu_{X_n} \end{Bmatrix}, \qquad \mathbf{\Sigma_{XX}} = \begin{bmatrix} \sigma^2_{X_1} & 0 & ... & 0 \\ 0 & \sigma^2_{X_2} & ... & 0 \\ ... & & & \\ 0 & 0 & ... & \sigma^2_{X_n} \end{bmatrix} \tag{10.13}$$

while the PDF is

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)\sqrt{|\mathbf{\Sigma_{XX}}|}} \exp\left\{ -\frac{1}{2} (\mathbf{x} - \mu_{\mathbf{X}})^T \mathbf{\Sigma_{XX}^{-1}} (\mathbf{x} - \mu_{\mathbf{X}}) \right\}$$

with

$$\mathbf{\Sigma_{XX}} = \mathbf{DD} = \begin{bmatrix} \sigma_{X_1} & 0 & ... & 0 \\ 0 & \sigma_{X_2} & ... & 0 \\ ... & & & \\ 0 & 0 & ... & \sigma_{X_n} \end{bmatrix} \begin{bmatrix} \sigma_{X_1} & 0 & ... & 0 \\ 0 & \sigma_{X_2} & ... & 0 \\ ... & & & \\ 0 & 0 & ... & \sigma_{X_n} \end{bmatrix}$$

Let us develop the following coordinate transformation

$$\mathbf{U} = \mathbf{D}^{-1} (\mathbf{X} - \mu_{\mathbf{X}}) \qquad \Longrightarrow \qquad \begin{cases} U_1 = \frac{X_1 - \mu_1}{\sigma_1} \\ U_2 = \frac{X_2 - \mu_2}{\sigma_2} \\ ... \\ U_n = \frac{X_n - \mu_n}{\sigma_n} \end{cases}, \tag{10.14}$$

The inverse transformation is

$$\mathbf{X} = \mu_{\mathbf{X}} + \mathbf{DU} \qquad \Longrightarrow \qquad \begin{cases} X_1 = \mu_1 + \sigma_1 U_1 \\ X_2 = \mu_2 + \sigma_2 U_2 \\ ... \\ X_n = \mu_n + \sigma_n U_n \end{cases} \tag{10.15}$$

Under this circumstance, the random variables $U_1, U_2, ..., U_n$ follow a multivariate normal standard PDF

$$\varphi_n(\mathbf{u}) = \frac{1}{(2\pi)^{n/2}} \exp\left(-\frac{1}{2} \mathbf{u}^T \mathbf{u} \right)$$

In the transformation (10.14)

- First we pursue a *translation*, such that $E[U_i] = E[X_i - \mu_i]/\sigma_i = 0$
- Later we pursue a *scale shift*, such that $Var[U_i] = \sigma_i^2/\sigma_i^2 = 1$

Case 1: Bivariate Gaussian, uncorrelated, different variance

$$\mu_{\mathbf{X}} = \begin{Bmatrix} 1 \\ -1 \end{Bmatrix}, \qquad \sigma_{\mathbf{X}} = \begin{Bmatrix} 0.949 \\ 0.548 \end{Bmatrix}, \qquad \mathbf{R} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

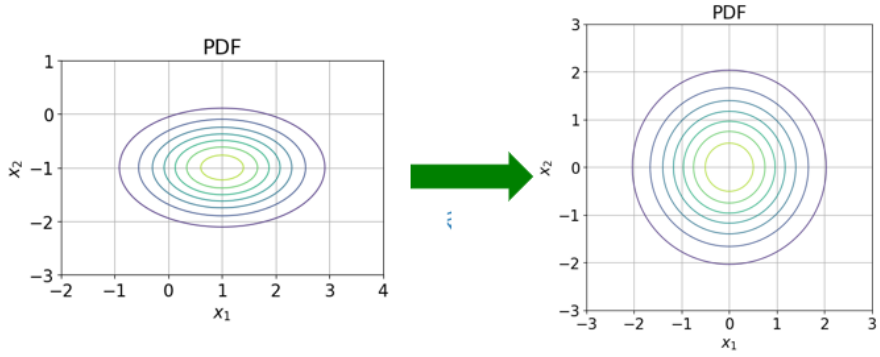


Figure 10.5: From bivariate Gaussian uncorrelated to bivariate normal standard

Original PDF $f_{X_1X_2}(x_1, x_2)$

- the contour level of PDF is an ellipse
- the center of the ellipse has coordinate $(1, -1)$
- the axes of the ellipse are parallel to the axes of the variables x_1 and x_2
- the semi-axes of the ellipse are proportional to the standard deviation σ_{X_1} and σ_{X_2}

Normal standard PDF $\varphi_2(u_1, u_2)$

- the contour level of PDF is a circle
- the center of the circle is the origin $(0, 0)$
- there is polar symmetry with respect to the origin
- the circle has radius $\sigma = 1$

10.2.3 Coordinate transformation toward normal standard space - Case of correlated Gaussian random variables

Consider now a vector of n correlated Gaussian random variables $\mathbf{X} = \{X_1, X_2, ..., X_n\}$

The expected value $\mu_{\mathbf{X}}$ and the matrix of covariance $\mathbf{\Sigma_{XX}}$ are

$$\mu_{\mathbf{X}} = E[\mathbf{X}] = \begin{Bmatrix} \mu_{X_1} \\ \mu_{X_2} \\ ... \\ \mu_{X_n} \end{Bmatrix}, \qquad \mathbf{\Sigma_{XX}} = \begin{bmatrix} \sigma^2_{X_1} & \sigma_{X_1X_2} & ... & \sigma_{X_1X_n} \\ \sigma_{X_1X_2} & \sigma^2_{X_2} & ... & \sigma_{X_2X_n} \\ ... & & & \\ \sigma_{X_1X_n} & \sigma_{X_2X_n} & ... & \sigma^2_{X_n} \end{bmatrix} \tag{10.16}$$

while the matrix of correlation $\mathbf{R}$ is

$$\mathbf{R} = \begin{bmatrix} 1 & \rho_{X_1X_2} & ... & \rho_{X_1X_n} \\ \rho_{X_1X_2} & 1 & ... & \rho_{X_2X_n} \\ ... & & & \\ \rho_{X_1X_n} & \rho_{X_2X_n} & ... & 1 \end{bmatrix}$$

The joint PDF is

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)\sqrt{|\mathbf{\Sigma_{XX}}|}} \exp\left\{ -\frac{1}{2} (\mathbf{x} - \mu_{\mathbf{X}})^T \mathbf{\Sigma_{XX}^{-1}} (\mathbf{x} - \mu_{\mathbf{X}}) \right\}$$

with

$$\mathbf{\Sigma_{XX}} = \mathbf{DRD} = \begin{bmatrix} \sigma_{X_1} & 0 & ... & 0 \\ 0 & \sigma_{X_2} & ... & 0 \\ ... & & & \\ 0 & 0 & ... & \sigma_{X_n} \end{bmatrix} \begin{bmatrix} 1 & \rho_{X_1X_2} & ... & \rho_{X_1X_n} \\ \rho_{X_1X_2} & 1 & ... & \rho_{X_2X_n} \\ ... & & & \\ \rho_{X_1X_n} & \rho_{X_2X_n} & ... & 1 \end{bmatrix} \begin{bmatrix} \sigma_{X_1} & 0 & ... & 0 \\ 0 & \sigma_{X_2} & ... & 0 \\ ... & & & \\ 0 & 0 & ... & \sigma_{X_n} \end{bmatrix}$$

The coordinate transformation is developed through 2 steps:

- Coordinate transformation $\mathbf{x} \to \mathbf{z}$ where $Z_1, Z_2, ..., Z_n$ are correlated normal standard random variables and whose joint PDF is $f_{\mathbf{Z}}(\mathbf{z}) \equiv \varphi_n(\mathbf{u}; \mathbf{R})$
- Coordinate transformation $\mathbf{z} \to \mathbf{u}$ where $U_1, U_2, ..., U_n$ are normal standard random variables and whose joint PDF is $\varphi_n(\mathbf{u})$

Step 1. Coordinate transformation $\mathbf{x} \to \mathbf{z}$

Let us develop the following coordinate transformation

$$\mathbf{Z} = \mathbf{D}^{-1}(\mathbf{X} - \mu_{\mathbf{X}}) \qquad \Longrightarrow \qquad \begin{cases} Z_1 = \frac{X_1 - \mu_1}{\sigma_1} \\ Z_2 = \frac{X_2 - \mu_2}{\sigma_2} \\ ... \\ Z_n = \frac{X_n - \mu_n}{\sigma_n} \end{cases}, \tag{10.17}$$

The inverse transformation is

$$\mathbf{X} = \mu_{\mathbf{X}} + \mathbf{DZ} \qquad \Longrightarrow \qquad \begin{cases} X_1 = \mu_1 + \sigma_1 Z_1 \\ X_2 = \mu_2 + \sigma_2 Z_2 \\ ... \\ X_n = \mu_n + \sigma_n Z_n \end{cases} \tag{10.18}$$

In the transformation (10.17)

- First we pursue a *translation*, such that $E[Z_i] = E[X_i - \mu_i]/\sigma_i = 0$
- Later we pursue a *scale shift*, such that $Var[Z_i] = \sigma_i^2/\sigma_i^2 = 1$
- Note that the correlations keep unchanged, so that $\rho_{Z_i Z_j} = \rho_{X_i X_j}$

Under this circumstance:

- the marginal PDFs of the random variables $Z_1, Z_2, ..., Z_n$ follow a *normal standard distribution*

$$\varphi(z) = \frac{1}{2\pi} \exp\left(-\frac{1}{2} z^2\right)$$

- the variables $Z_1, Z_2, ..., Z_n$ are correlated, so that the joint PDF is

$$f_{\mathbf{Z}}(\mathbf{z}; \mathbf{R}) = \frac{1}{(2\pi)^{n/2} \sqrt{|\mathbf{R}|}} \exp\left\{ -\frac{1}{2} \mathbf{z}^T \mathbf{R}^{-1} \mathbf{z} \right\} \tag{10.19}$$

Case 3: Bivariate Gaussian, correlated, tranformation $\mathbf{x} \to \mathbf{z}$

$$\mu_{\mathbf{X}} = \begin{Bmatrix} 1 \\ -1 \end{Bmatrix}, \qquad \sigma_{\mathbf{X}} = \begin{Bmatrix} 0.949 \\ 0.548 \end{Bmatrix}, \qquad \mathbf{R} = \begin{bmatrix} 1 & 0.77 \\ 0.77 & 1 \end{bmatrix}$$

$$\mu_{\mathbf{Z}} = \begin{Bmatrix} 0 \\ 0 \end{Bmatrix}, \qquad \sigma_{\mathbf{Z}} = \begin{Bmatrix} 1 \\ 1 \end{Bmatrix}, \qquad \mathbf{R} = \begin{bmatrix} 1 & 0.77 \\ 0.77 & 1 \end{bmatrix}$$

Original Gaussian PDF $f_{X_1 X_2}(x_1, x_2)$

- the contour level of PDF is an ellipse
- the center of the ellipse has coordinate $(1, -1)$
- the axes of the ellipse are not parallel to the axes of the variables x_1 and x_2

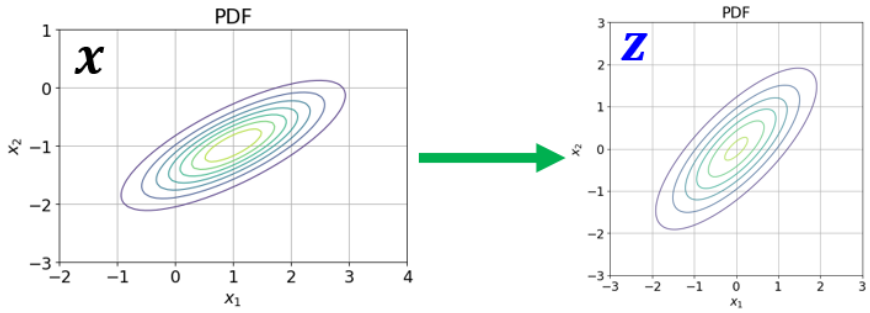


Figure 10.6: From bivariate Gaussian correlated to normalized Gaussian correlated

- the semi-axes of the ellipse are proportional to the standard deviation σ_{X_1} and σ_{X_2}

Mapped Gaussian PDF $f_{Z_1 Z_2}(z_1, z_2)$

- the contour level of PDF is an ellipse
- the center of the ellipse has coordinate $(0, 0)$
- the axes of the ellipse are not parallel to the axes of the variables z_1 and z_2
- the semi-axes of the ellipse are proportional to the standard deviation $\sigma = 1$

Step 2. Coordinate transformation z → u

First, we develop a Cholesky decomposition of the matrix of correlation $\mathbf{R}$

$$\mathbf{R} = \mathbf{L}\mathbf{L}^{\mathbf{T}} \tag{10.20}$$

$$\begin{bmatrix} 1 & \rho_{12} & \dots & \rho_{1n} \\ \rho_{12} & 1 & \dots & \rho_{2n} \\ \dots & & & \\ \rho_{1n} & \rho_{2n} & \dots & 1 \end{bmatrix} = \begin{bmatrix} L_{11} & 0 & \dots & 0 \\ L_{21} & L_{22} & \dots & 0 \\ \dots & & & \\ L_{n1} & L_{n2} & \dots & L_{nn} \end{bmatrix} \begin{bmatrix} L_{11} & L_{21} & \dots & L_{n1} \\ 0 & L_{22} & \dots & L_{n2} \\ \dots & & & \\ 0 & 0 & \dots & L_{nn} \end{bmatrix}$$

where $\mathbf{L}$ is lower triangular matrix. Let us develop the following coordinate transformation

$$\mathbf{Z} = \mathbf{L}\mathbf{U}, \qquad \Rightarrow \qquad \begin{Bmatrix} Z_1 \\ Z_2 \\ \dots \\ Z_n \end{Bmatrix} = \begin{bmatrix} L_{11} & 0 & \dots & 0 \\ L_{21} & L_{22} & \dots & 0 \\ \dots & & & \\ L_{n1} & L_{n2} & \dots & L_{nn} \end{bmatrix} \begin{Bmatrix} U_1 \\ U_2 \\ \dots \\ U_n \end{Bmatrix} \tag{10.21}$$

The inverse transformation is

$$\mathbf{U}=\mathbf{L}^{-1}\mathbf{Z}, \qquad \Rightarrow \qquad \begin{Bmatrix} U_1 \\ U_2 \\ ... \\ U_n \end{Bmatrix} = \begin{bmatrix} L_{11} & L_{21} & ... & L_{n1} \\ 0 & L_{22} & ... & L_{n2} \\ ... & & & \\ 0 & 0 & ... & L_{nn} \end{bmatrix} \begin{Bmatrix} Z_1 \\ Z_2 \\ ... \\ Z_n \end{Bmatrix} \tag{10.22}$$

In the transformation (10.21)

- we are doing a rotation of axes so that the axes of the ellipse are coinciding with the coordinate axes
- In the **u**-space the correlations are zero, $\rho_{ij}=0$

Under this circumstance the random variables $U_1, U_2, ..., U_n$ follow a multivariate normal standard PDF

$$\varphi(\mathbf{u})=\frac{1}{(2\pi)^{n/2}}\exp\left(-\frac{1}{2}\mathbf{u}^T\mathbf{u}\right)$$

Case 3: Bivariate Gaussian, correlated, tranformation z → u

$$\mu_{\mathbf{Z}}=\begin{Bmatrix}0\\0\end{Bmatrix}, \qquad \sigma_{\mathbf{U}}=\begin{Bmatrix}1\\1\end{Bmatrix}, \qquad \mathbf{R}=\begin{bmatrix}1 & 0.77\\ 0.77 & 1\end{bmatrix}$$

The Cholesky decomposition of the matrix of correlation **R** is

$$\mathbf{R}=\mathbf{L}\mathbf{L}^{\mathbf{T}}, \qquad \Rightarrow \qquad \begin{bmatrix}1 & \rho\\ \rho & 1\end{bmatrix} = \begin{bmatrix}L_{11} & 0\\ L_{21} & L_{22}\end{bmatrix}\begin{bmatrix}L_{11} & L_{21}\\ 0 & L_{22}\end{bmatrix} = \begin{bmatrix}L_{11}^2 & L_{11}L_{21}\\ L_{11}L_{21} & L_{21}^2+L_{22}^2\end{bmatrix}$$

It follows

$$\mathbf{L}=\begin{bmatrix}1 & 0\\ \rho & \sqrt{1-\rho^2}\end{bmatrix}$$

Let us develop the following coordinate transformation

$$\mathbf{Z}=\mathbf{L}\mathbf{U}, \qquad \Rightarrow \qquad \begin{Bmatrix}Z_1\\Z_2\end{Bmatrix} = \begin{bmatrix}1 & 0\\ \rho & \sqrt{1-\rho^2}\end{bmatrix}\begin{Bmatrix}U_1\\U_2\end{Bmatrix}, \qquad \Rightarrow \qquad \begin{cases}Z_1=U_1\\ Z_2=\rho U_1+\sqrt{1-\rho^2}U_2\end{cases}$$

The inverse transformation is

$$\mathbf{U}=\mathbf{L}^{-1}\mathbf{Z}, \qquad \Rightarrow \qquad \begin{cases}U_1=Z_1\\ U_2=\frac{Z_2-\rho Z_1}{\sqrt{1-\rho^2}}\end{cases}$$

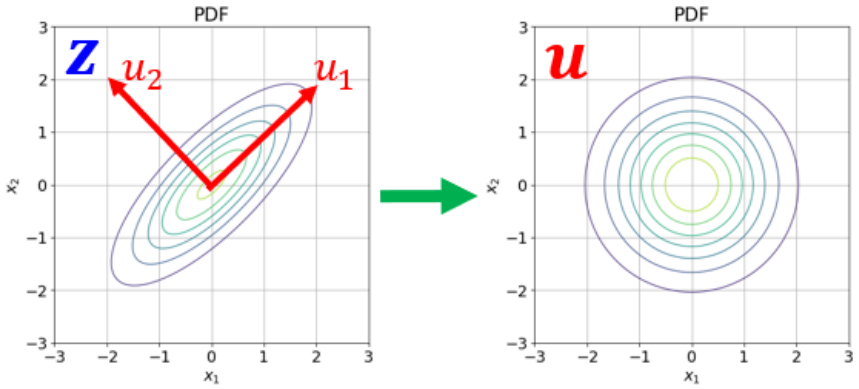


Figure 10.7: From joint normal Gaussian correlated variables to joint normal standard Gaussian

where

$$\mu_{\mathbf{U}} = \begin{Bmatrix} 0 \\ 0 \end{Bmatrix}, \qquad \sigma_{\mathbf{U}} = \begin{Bmatrix} 1 \\ 1 \end{Bmatrix}, \qquad \mathbf{R} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Joint PDF of normal standard correlated $f_{Z_1 Z_2}(z_1, z_2)$

- the contour level of PDF is an ellipse
- the center of the ellipse has coordinate $(0, 0)$
- the axes of the ellipse are not parallel to the axes of the variables z_1 and z_2
- the semi-axes of the ellipse are proportional to the standard deviation $\sigma = 1$

Mapped normal standard PDF $\varphi_2(u_1, u_2)$

- the contour level of PDF is a circle
- the center of the circle is the origin of coordinates $(0, 0)$
- the radius of the circle is $\sigma = 1$
- the PDF is polar symmetric with respect to the origin

Coordinate transformation $\mathbf{x} \to \mathbf{u}$

By substituting eq.(10.17) into eq.(10.21) we obtain

$$\mathbf{X} = \mathbf{DLU} + \mu \tag{10.23}$$

The inverse transformation is

$$\mathbf{U} = \mathbf{L}^{-1}\mathbf{D}^{-1}\left(\mathbf{X} - \mu\right) \tag{10.24}$$

Case 3: Bivariate Gaussian, correlated, tranformation $\mathbf{x} \to \mathbf{u}$

$$\mu_{\mathbf{X}} = \begin{Bmatrix} 1 \\ -1 \end{Bmatrix}, \qquad \sigma_{\mathbf{X}} = \begin{Bmatrix} 0.949 \\ 0.548 \end{Bmatrix}, \qquad \mathbf{R} = \begin{bmatrix} 1 & 0.77 \\ 0.77 & 1 \end{bmatrix}$$

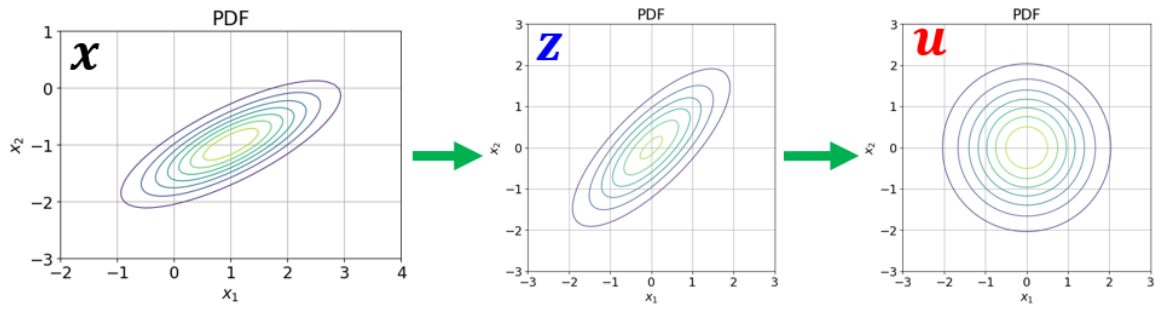


Figure 10.8: From joint Gaussian correlated variables to joint normal standard Gaussian

$$\mu_{\mathbf{U}} = \begin{Bmatrix} 0 \\ 0 \end{Bmatrix}, \qquad \sigma_{\mathbf{X}} = \begin{Bmatrix} 1 \\ 1 \end{Bmatrix}, \qquad \mathbf{R} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

and it is obtained through the transformation $\mathbf{x} \to \mathbf{u}$

$$\mathbf{X} = \mathbf{DLU} + \mu$$

$$\begin{Bmatrix} X_1 \\ X_2 \end{Bmatrix} = \begin{bmatrix} \sigma_{X_1} & 0 \\ 0\sigma_{X_2} & \end{bmatrix} \begin{bmatrix} 1 & 0 \\ \rho & \sqrt{1-\rho^2} \end{bmatrix} = \begin{Bmatrix} U_1 \\ U_2 \end{Bmatrix} + \begin{Bmatrix} \mu_{X_1} \\ \mu_{X_2} \end{Bmatrix}$$

Joint PDF of Gaussian correlated $f_{X_1X_2}(x_1, x_2)$

- the contour level of the PDF is an ellipse
- the center of the ellipse has coordinate (μ_{X_1}, μ_{X_2})
- the axes of the ellipse are not parallel to the axes of the variables x_1 and x_2

- the semi-axes of the ellipse are proportional to the standard deviations σ_{X_1} and σ_{X_2}

Joint PDF of normal standard correlated $f_{Z_1 Z_2}(z_1, z_2)$

- the contour level of PDF is an ellipse
- the center of the ellipse has coordinate $(0, 0)$
- the axes of the ellipse are not parallel to the axes of the variables z_1 and z_2
- the semi-axes of the ellipse are proportional to the standard deviation $\sigma = 1$

Mapped normal standard PDF $\varphi_2(u_1, u_2)$

- the contour level of PDF is a circle
- the center of the circle is the origin of coordinates $(0, 0)$
- the radius of the circle is $\sigma = 1$
- the PDF is polar symmetric with respect to the origin

Chapter 11

Models of regression and multivariate analysis

11.1 Simple Linear Regression

11.2 Multiple Linear Regression

11.3 Multivariate Analysis

Chapter 12

Basics of Stochastic Processes

12.1 Introduction

12.2 Statistical Indicators

12.3 Stationarity

12.4 Ergodicity

12.5 Power Spectral Density

12.6 Gaussian process

12.7 Barrier crossing problem

Chapter 13

Analysis of Extreme events

13.1 Classical Extreme values: theory and methods

13.2 Threshld models

13.3 Dependent sequences

13.4 Extremes of non-stationary sequences

13.5 Multivariate extremes

Chapter 14

Lecture C1: Bayesian Updating

Chapter 15

Lecture B1:Structural Reliability

15.1 Introduction

In the most general problem, we can define a set of basic random variables $x_1, x_2, ..., x_n$, which typically can give rise to:

- *Capacity* $R(\mathbf{x})$
- *Demand* $S(\mathbf{x})$

And define the following broad problem:

$$\begin{cases} R > S: & Safe \\ R \leq S: & Failure \end{cases}$$

But, *Capacity R and demand S are uncertain parameters!*

The failure probability with respect to a chosen limit state is

$$P_f = Prob\left[R \leq S\right] \tag{15.1}$$

The reliability is the complementary of the failure probability

$$Rel = 1 - P_f \tag{15.2}$$

The risk is given by the consequence C multiplied by the failure probability P_f

$$R = P_f \; C \tag{15.3}$$

Bar example. Assume that we have a bar in steel, whose yield strength is F_y, and subject to a load S. Let us assume that the area of the cross section is A, then

$$\begin{cases} R = AF_y : & \textit{Capacity} \\ S : & \textit{Demand} \end{cases}$$

where *Capacity* R and *Demand* S are random variables

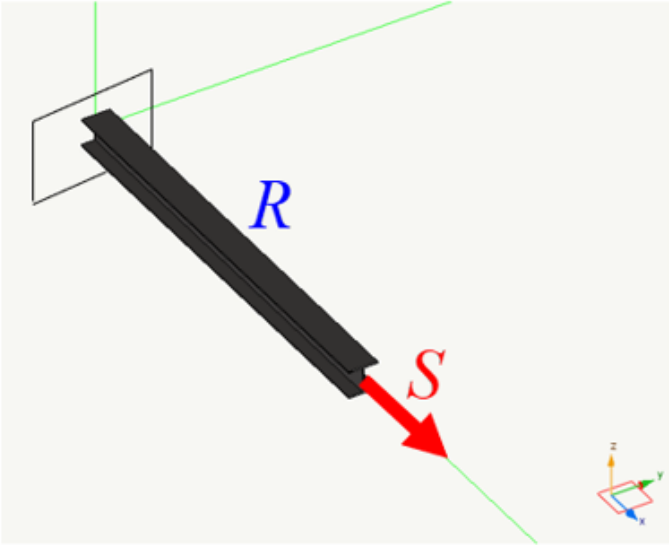


Figure 15.1: Bar example

- *Failure probability*: $P_f = Prob[AF_y \leq S]$
- *Reliability*: $Rel = 1 - P_f = Prob[AF_y > S]$
- *Risk*: $R = P_f\ C$

15.2 Structural Reliability

Given a set of basic random variables $\mathbf{x} \equiv \{x_1, x_2, ..., x_n\}$, the Limit State Function (LSF) is defined as

$$G(\mathbf{x}) = R(\mathbf{x}) - S(\mathbf{x}) \tag{15.4}$$

As a consequence, the failure probability is

$$P_f = Prob[R \leq S] = Prob[G \leq 0] \tag{15.5}$$

In particular, the space of the basic random variables is divided into two distinct domains, the safe set Ω_S and the failure set Ω_f, separated by the Limit State Surface (LSS) $G(\mathbf{x}) = 0$

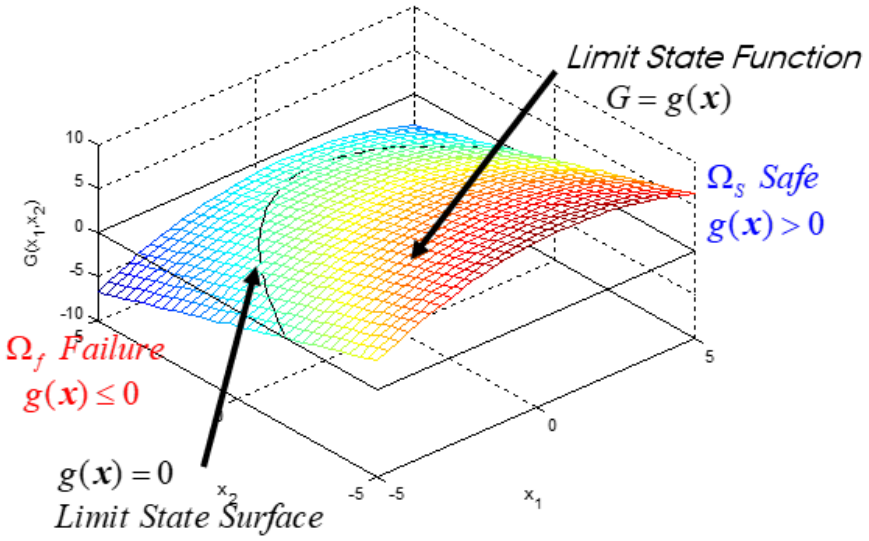


Figure 15.2: Limit State

$$\begin{cases} G(\mathbf{x}) > 0: & \Omega_S : Safe\ Set \\ G(\mathbf{x}) = 0: & Limit\ State \\ G(\mathbf{x}) < 0: & \Omega_f : Failure\ Set \end{cases}$$

It can be shown that the failure probability P_f is given as

$$P_f = Prob[G \le 0] = \int_{G(\mathbf{x}) \le 0} f_{\mathbf{X}}(\mathbf{x})\, d\mathbf{x} \tag{15.6}$$

and it depends on

- Limit state $G(\mathbf{X}) = 0$ (which defines the domain of integration)
- Joint PDF of the basic random variables $f_{\mathbf{X}}(\mathbf{x})$ (which defines the integrand function)

and in the following we will discuss more in detail these two contributions

15.2.1 Limit State

The failure probability P_f is given as

$$P_f = Prob\left[G \leq 0\right] = \int_{G(\mathbf{x})\leq 0} f_{\mathbf{X}}\left(\mathbf{x}\right) d\mathbf{x}$$

The **Limit State Surface (LSS)** $G(\mathbf{x}) = 0$ defines the boundary domain between safe set and failure set. Therefore it defines the domain of integration of the integral (15.6)

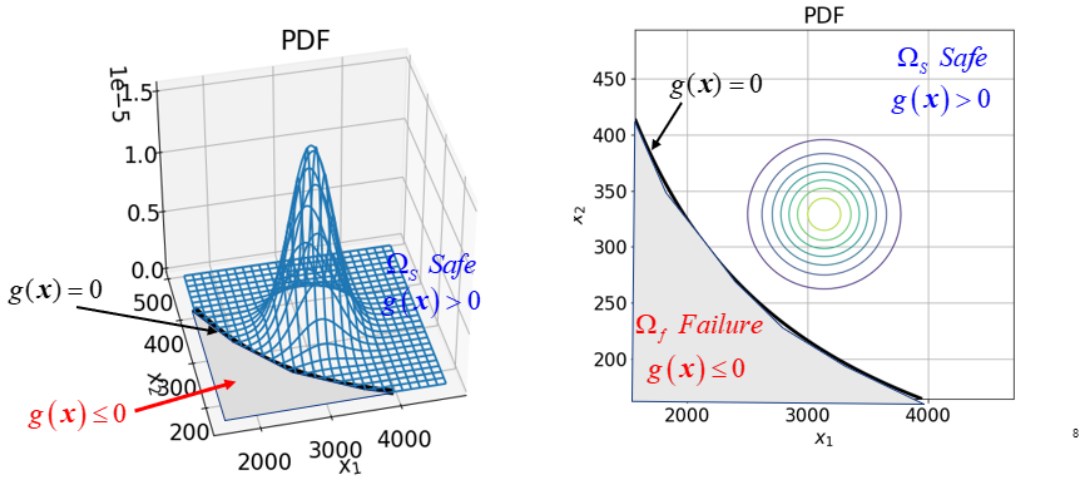


Figure 15.3: Limit State

and its knowledge is crucial for evaluating the failure probability P_f.

This means that if we have:

- a given joint distribution $f_{\mathbf{X}}(\mathbf{x})$ of the basic random variables $x_1, x_2, ..., x_n$
- two different limit states $G_1(\mathbf{x}) = 0$ and $G_2(\mathbf{x}) = 0$

then

$$P_{f,1} \neq P_{f,2}$$

Note that the LSS $G(\mathbf{x}) = 0$ is the intersection between the LSF $G(\mathbf{x})$ and the plane $G = 0$. This means that:

- to the same Limit State $G(\mathbf{x}) = 0$ can correspond infinite Limit State function $G_1(\mathbf{x}), G_2(\mathbf{x}), ...$

- The failure probability is affected by the LSS $G(\mathbf{x}) = 0$, and not from the specific form of the LSF $G(\mathbf{x})$

For example, the three limit state functions

- $G_1(R, S) = R - S$
- $G_2(R, S) = \frac{R}{S} - 1$
- $G_3(R, S) = R^3 - S^3 = (R - S)(R^2 + R\ S + S^2)$

differ each other, however they all have the same limit state equation $G(R, S) = 0$, (i.e. $R = S$), and therefore the 3 different LSF are all *representative of the same limit state* and consequently they provide the same failure probability P_f

Bar example. Consider now the case of the bar subjected to a tensile load S

We have the following random variables:

- Yield Strength F_y: it follows a Gaussian distribution, whose mean and standard deviation are $\mu_{F_y} = 329$ MPa and $\sigma_{F_y} = 32.90$ MPa, respectively
- Load S: it follows a Gaussian distribution, whose mean and standard deviation are $\mu_S = 650$ kN and $\sigma_S = 130$ kN, respectively

Choose a profile HE200A, whose area is $A = 5,383\ mm^2$, so that the capacity $R = AF_y$ follows a Gaussian distribution, whose mean and standard deviation are $\mu_S = 1,771$ kN and $\sigma_S = 177.10$ kN, respectively. Different realizations of the basic random variables can

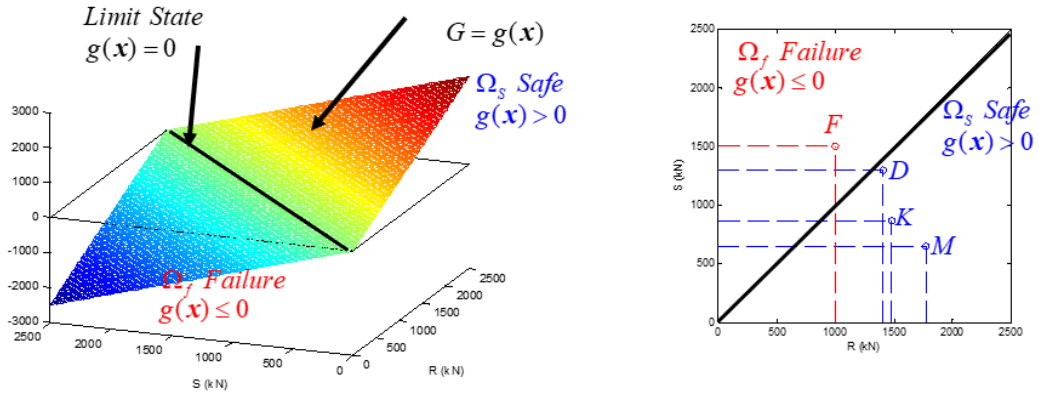


Figure 15.4: Limit State

give rise to a status of failure or safety.

- $M \equiv (R_m, S_m) = (1,771; 650)$ kN: R_m and S_m are the mean values of the 2 random variables F_y and S. M falls in the safe set Ω_S
- $K \equiv (R_k, S_k) = (1,480.3; 865)$ kN: R_k and S_k are the characteristics values of the 2 random variables F_y and S. K falls in the safe set Ω_S
- $D \equiv (R_d, S_d) = (1,409.8; 1,297.5)$ kN: R_d and S_d are the design values of the 2 random variables F_y and S. D falls in the safe set Ω_S
- $F = (1,000; 1,500)$ kN: specific realization of the 2 random variables F_y and S. F falls in the failure set Ω_f and it is representative of a failure scenario

15.2.2 Joint PDF of the basic random variables $f_{\mathbf{X}}(\mathbf{x})$

The failure probability P_f is given as

$$P_f = Prob\left[G \leq 0\right] = \int_{G(\mathbf{x})\leq 0} f_{\mathbf{X}}\left(\mathbf{x}\right) d\mathbf{x}$$

and it depends on both the limit state $G(\mathbf{x}) = 0$ (which defines the domain of integration), and also the joint PDF of the basic random variables $f_{\mathbf{X}}(\mathbf{x})$ (which defines the integrand function) This means that if we have:

- a chosen limit state $G(\mathbf{x}) = 0$
- two different distributions $f^{(1)}_{\mathbf{X}}(\mathbf{x})$ and $f^{(2)}_{\mathbf{X}}(\mathbf{x})$ of the same basic random variables $x_1, x_2, ..., x_n$

then

$$P_{f,1} \neq P_{f,2}$$

Case test 1, Bar example, profile HE140A. In this case $A_1 = 3,143\ mm^2$, $R = A_1 F_y$

$$R : \begin{pmatrix} \mu_R = 1,033kN \\ \sigma_R = 103.3kN \end{pmatrix}, \qquad S : \begin{pmatrix} \mu_S = 650kN \\ \sigma_S = 130kN \end{pmatrix}, \qquad P_f = 1.04 \times 10^{-2}$$

Case test 3, Bar example, profile HE200A. In this case $A_3 = 5,383\ mm^2$, $R = A_3 F_y$

$$R : \begin{pmatrix} \mu_R = 1,771kN \\ \sigma_R = 177.1kN \end{pmatrix}, \qquad S : \begin{pmatrix} \mu_S = 650kN \\ \sigma_S = 130kN \end{pmatrix}, \qquad P_f = 1.67 \times 10^{-7}$$

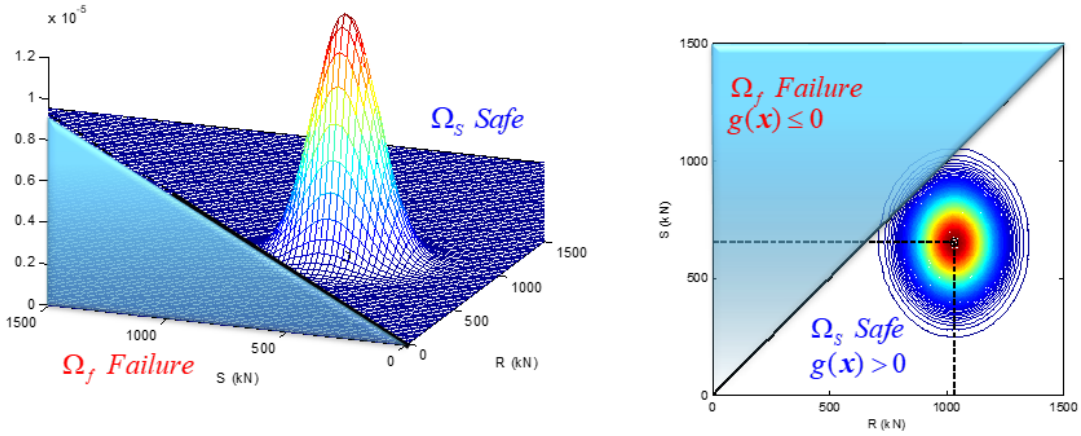


Figure 15.5: Limit State bar, profile HE140A

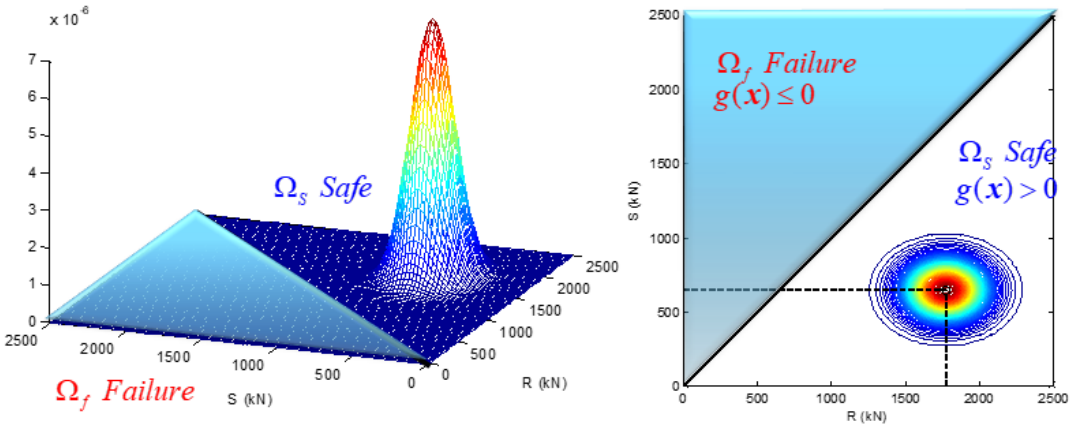


Figure 15.6: Limit State bar, profile HE200A

15.3 Isoprobabilistic Transformation

The failure probability P_f is given as

$$P_f = \int_{G(\mathbf{x})\leq 0} f_{\mathbf{X}}(\mathbf{x})\, d\mathbf{x}$$

and it depends on

- Limit state $G(\mathbf{X}) = 0$ (which defines the domain of integration)
- Joint PDF of the basic random variables $f_{\mathbf{X}}(\mathbf{x})$ (which defines the integrand function)

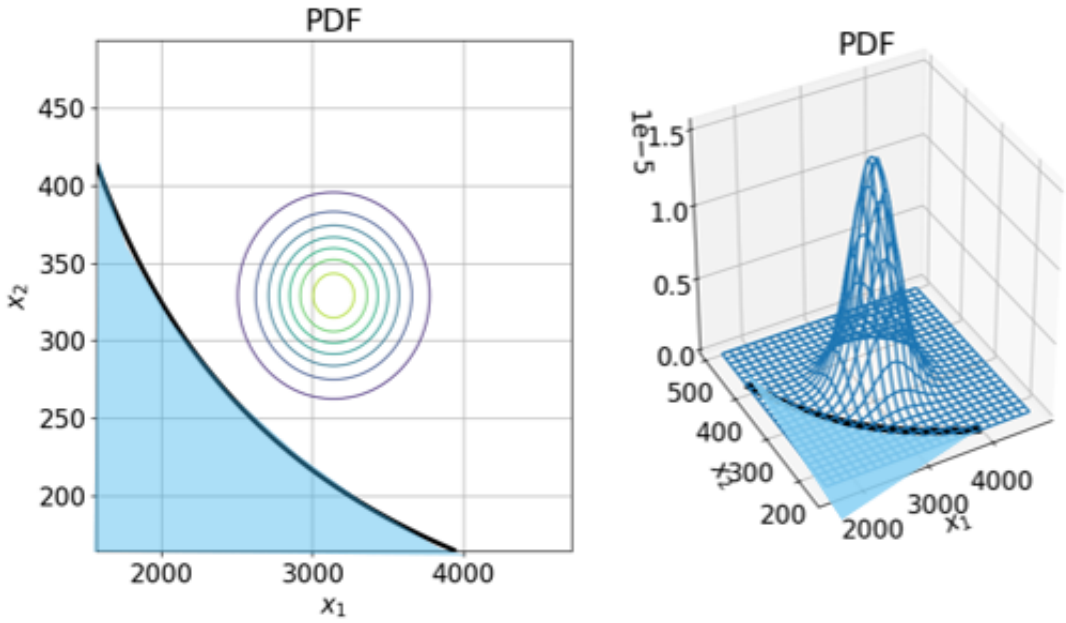


Figure 15.7: Reliability problem, x-space

The isoprobabilistic Transformation is a probability transformation from the *original space* $\mathbf{x}$ of the variables $X_1, X_2, ..., X_n$ toward the *normal standard space* $\mathbf{u}$ of $U_1, U_2, ..., U_n$

$$\mathbf{u} = \mathbf{T}(\mathbf{x}) \tag{15.7}$$

while the inverse transformation is

$$\mathbf{x} = \mathbf{T}^{-1}(\mathbf{u}) \tag{15.8}$$

In the normal standard space:

- the joint PDF $f_{\mathbf{X}}(\mathbf{x})$ of the basic random variables $X_1, X_2, ..., X_n$ is mapped into a multivariate normal standard PDF $\varphi_n(\mathbf{u})$ of the normal standard random variables $U_1, U_2, ..., U_n$
- the limit state $G(\mathbf{x}) = G(X_1, X_2, ..., X_n) = 0$ is mapped into the corresponding limit state $G(\mathbf{u}) = G(U_1, U_2, ..., U_n) = 0$

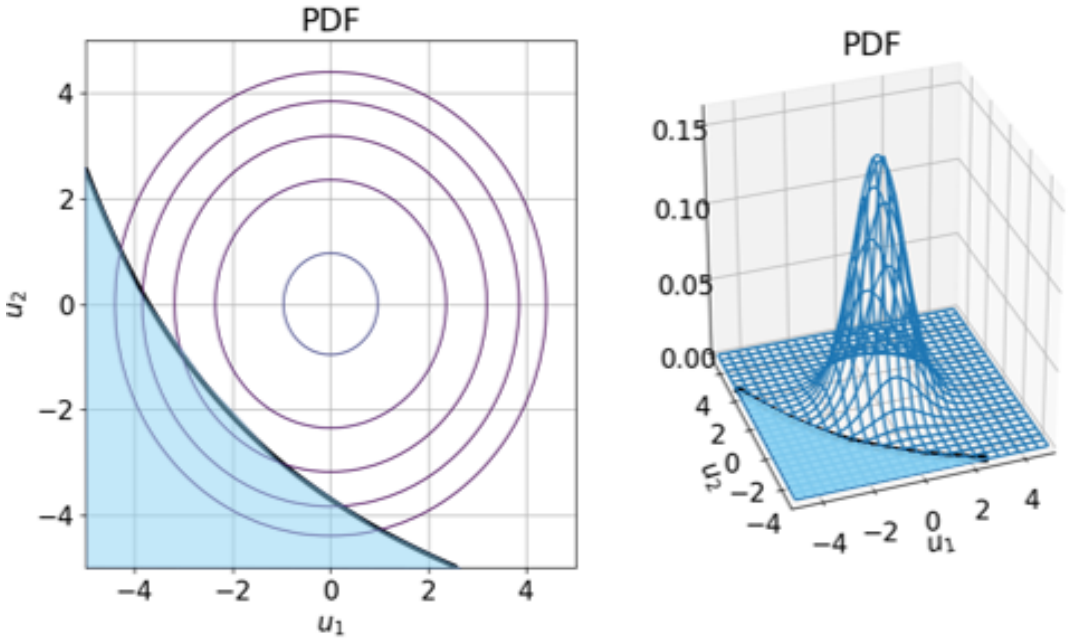


Figure 15.8: Reliability problem, u-space

In the **u**-space the failure probability P_f is given as

$$P_f = \int_{g(\mathbf{u}) \leq 0} \varphi_n(\mathbf{u})\, d\mathbf{u}$$

and it depends *only* on

- Limit state $g(\mathbf{u}) = 0$ (which defines the domain of integration)

Chapter 16

Lecture B1:Basics of MCS

16.1 Simulation of samples

16.1.1 One random variables

Assume that we have one random variable X having distribution $F_X(x)$.

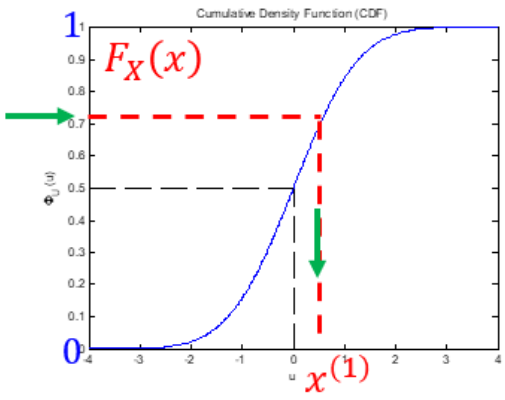


Figure 16.1: Simulation of samples from a distribution

Simulation of samples from one random variable:

1. $k = 1$
2. Generate a random number uniformly distributed $p^{(1)}$, with $0 \leq p^{(1)} \leq 1$

3. Find the quantile of X corresponding to $p^{(1)}$

$$x^{(1)} = F_X^{-1}\left[p^{(1)}\right]$$

4. $k = k + 1$

5. Generate a random number uniformly distributed $p^{(k)}$, with $0 \leq p^{(k)} \leq 1$

6. Find the quantile of X corresponding to $p^{(k)}$

$$x^{(k)} = F_X^{-1}\left[p^{(k)}\right]$$

7. repeat the steps 4-6 until needed

16.1.2 Multiple independent Random variables

Assume that we have two random variables X_1 and X_2 having distributions $F_{X_1}(x_1)$ and $F_{X_2}(x_2)$, respectively. The two random variables are collected in the vector $\mathbf{X} \equiv \{X_1, X_2\}$

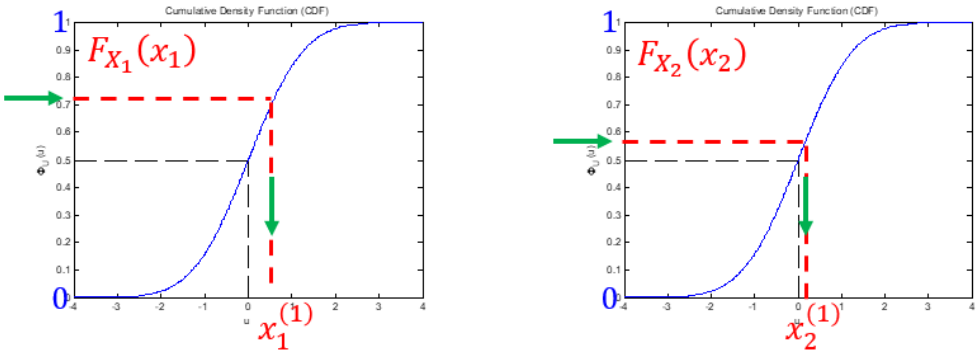


Figure 16.2: Simulation of samples from two independent random variables

Simulation of samples from multiple independent random variables:

1. $k = 1$

2. Generate a random number uniformly distributed $p_1^{(1)}$, with $0 \leq p_1^{(1)} \leq 1$

3. Find the quantile of X_1 corresponding to $p_1^{(1)}$

$$x_1^{(1)} = F_{X_1}^{-1}\left[p_1^{(1)}\right]$$

4. Generate a random number uniformly distributed $p_2^{(1)}$, with $0 \leq p_2^{(1)} \leq 1$
5. Find the quantile of X_2 corresponding to $p_2^{(1)}$

$$x_2^{(1)} = F_{X_2}^{-1}\left[p_2^{(1)}\right]$$

6. The first sample is $\mathbf{x}^{(1)} \equiv \left[x_1^{(1)}, x_2^{(1)}\right]$
7. $k = k + 1$
8. Generate a random number uniformly distributed $p_1^{(k)}$, with $0 \leq p_1^{(k)} \leq 1$
9. Find the quantile of X_1 corresponding to $p_1^{(k)}$

$$x_1^{(k)} = F_{X_1}^{-1}\left[p_1^{(k)}\right]$$

10. Generate a random number uniformly distributed $p_2^{(k)}$, with $0 \leq p_2^{(k)} \leq 1$
11. Find the quantile of X_2 corresponding to $p_2^{(k)}$

$$x_2^{(k)} = F_{X_2}^{-1}\left[p_2^{(k)}\right]$$

12. The $k-th$ sample is $\mathbf{x}^{(k)} \equiv \left[x_1^{(k)}, x_2^{(k)}\right]$
13. repeat the steps 7-12 until needed

16.2 Crude MCS

Given a set of basic random variables $X_1, X_2, ..., X_n$ collected in the vector $\mathbf{x}$ and having a joint PDF $f_{\mathbf{X}}(\mathbf{x})$, the failure probability P_f with respect to a given limit state function $G(\mathbf{x})$ is given as

$$P_f = \int_{G(\mathbf{x})\leq 0} f_{\mathbf{X}}(\mathbf{x})\, d\mathbf{x} \tag{16.1}$$

In the most general multidimensional case, this integral is extremely difficult to be solved. To this aim, let us introduce now the *indicator function* $I_G(\mathbf{x})$ defined as

$$I_G(\mathbf{x}) = \begin{cases} 1 & if\ G(\mathbf{x}) \leq 0 \\ 0 & otherwise \end{cases} \tag{16.2}$$

The eq.(16.1) becomes

$$P_f = \int I_G(\mathbf{x})\, f_{\mathbf{X}}(\mathbf{x})\, d\mathbf{x} = E[I_G(\mathbf{X})] \tag{16.3}$$

Therefore, the failure probability P_f, is equal to the expected value of the indicator function. When N samples $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, ..., \mathbf{x}^{(N)}$ are generated from the joint PDF $f_{\mathbf{X}}(\mathbf{x})$ the failure probability is estimated through its sample mean

$$P_f \approx \hat{P}_f = \frac{1}{N}\sum_{i=1}^{N} I_G\left(\mathbf{x}^{(i)}\right) \tag{16.4}$$

and this approximation is good for N big enough.

Crude MCS:

1. $k = 1, N = 0, N_f = 0$
2. Simulate a first sample $\mathbf{x}^{(1)}$ from the joint PDF $f_{\mathbf{X}}(\mathbf{x})$ of the basic random variables
3. Evaluate $G_1 = G[\mathbf{x}^{(1)}]$

$$\begin{cases} G[\mathbf{x}^{(1)}] \leq 0 & N = N+1, N_f = N_f + 1 \\ G[\mathbf{x}^{(1)}] > 0 & N = N+1 \end{cases}$$

4. $k = k + 1$
5. Simulate a sample $\mathbf{x}^{(k)}$, $k \geq 2$
6. Evaluate $G_k = G[\mathbf{x}^{(k)}]$

$$\begin{cases} G_k \leq 0 & N = N+1, N_f = N_f + 1 \\ G_k > 0 & N = N+1 \end{cases}$$

7. repeat steps 4-6 until needed
8. evaluate the failure probability

$$P_f = \frac{N_f}{N}$$

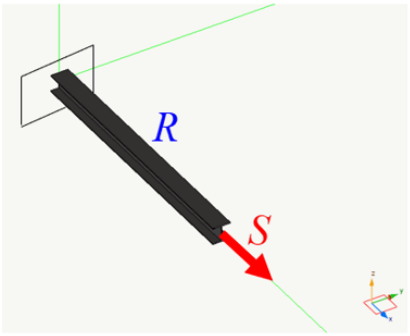
where N is the total number of simulated samples, while N_f is the number of failure events ($G \leq 0$)

Bar example. Consider now the case of a bar subjected to a tensile load S.

We have the following random variables:

- Yield Strength F_y: it follows a Gaussian distribution, whose mean and standard deviation are $\mu_{F_y} = 329$ MPa and $\sigma_{F_y} = 32.90$ MPa, respectively
- Load S: it follows a Gaussian distribution, whose mean and standard deviation are $\mu_S = 650$ kN and $\sigma_S = 130$ kN, respectively

Choose a profile HE140A, whose area is $A = 3,142\ mm^2$, so that the capacity $R = AF_y$ follows a Gaussian distribution, whose mean and standard deviation are $\mu_R = 1,033$ kN and $\sigma_R = 103.30$ kN, respectively. The failure probability is



Variable		Distribution	μ	v
	a		3,142	
x_1	F_y	Normal	329 MPa	0.10
x_2	S	Normal	650 kN	0.20

Figure 16.3: Bar example

$$P_f = Prob\left[R \leq S\right] = Prob\left[AF_y \leq S\right]$$

Let us define the limit state function as

$$G\left(A, F_y\right) = AF_y - S, \qquad \Rightarrow \qquad G\left(x_1, x_2\right) = Ax_1 - x_2$$

so that

$$P_f = Prob\left[G\left(x_1, x_2\right)\right] \leq 0$$

Crude MCS:

1. $k = 1, N = 0, N_f = 0$
2. Generate $F_y^{(1)}$
3. Generate $S^{(1)}$
4. the first sample is $\mathbf{x}^{(1)} \equiv \left[F_y^{(1)}; S^{(1)}\right]$
5. Evaluate $G_1 = G[\mathbf{x}^{(1)}] = AF_y^{(1)} - S^{(1)}$

$$\begin{cases} G[\mathbf{x}^{(1)}] \leq 0 & N = N + 1, N_f = N_f + 1 \\ G[\mathbf{x}^{(1)}] > 0 & N = N + 1 \end{cases}$$

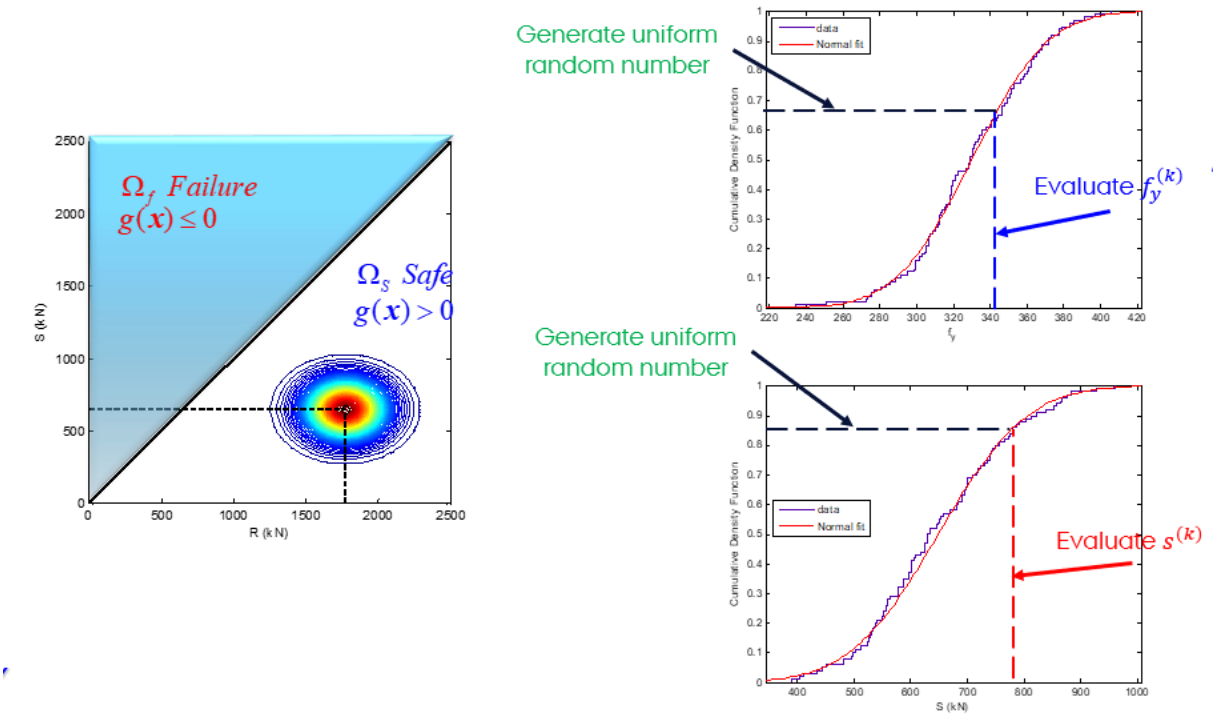


Figure 16.4: MCS,Bar example

6. $k = k + 1$
7. Generate $F_y^{(k)}$
8. Generate $S^{(k)}$
9. the $k - th$ sample is $\mathbf{x}^{(k)} \equiv \left[F_y^{(k)}; S^{(k)}\right]$
10. Evaluate $G_k = G[\mathbf{x}^{(k)}] = AF_y^{(k)} - S^{(k)}$

$$\begin{cases} G[\mathbf{x}^{(k)}] \le 0 & N = N + 1, N_f = N_f + 1 \\ G[\mathbf{x}^{(k)}] > 0 & N = N + 1 \end{cases}$$

11. repeat steps 6-10 until needed
12. evaluate the failure probability

$$P_f = \frac{N_f}{N}$$

16.2.1 Computational cost and accuracy

Key Question

- how many samples N we need to consider?
- which is the obtained accuracy for a given number of simulated samples?

Since we estimate the failure probability P_f through its sample mean $\hat{P}_f \approx P_f$, it comes natural to evaluate its coefficient of variation

$$\nu = \sqrt{\frac{1 - \hat{P}_f}{N\hat{P}_f}} \approx \sqrt{\frac{1}{N\hat{P}_f}} \tag{16.5}$$

where the last equation comes from the assumption that in structural reliability the failure probability is low, i.e. $P_f \ll 1$.

From eq.(16.5) it follows that

$$N \geq \frac{1}{P_f \nu^2} \tag{16.6}$$

But this means that *the Monte Carlo Simulation is computationally demanding!*

To show this, consider a coefficient of variation $\nu = 10\%$, from eq.(16.6) we obtain

$$N \geq \frac{1}{P_f(0.10)^2} = \frac{100}{P_f}$$

Therefore, for a failure probability of the order of $P_f = 10^{-5}$ we would need $N \geq 10^7$ samples! This number becomes not affordable, especially considering that in some problems, each analysis may require several hours (or even days).

Note also that the estimate of the failure probability depends upon the specific simulation, within the desired level of accuracy, provided by the coefficient of variation. Different MC runs, will provide different results (within the limits of the chosen accuracy, of course).

If we want to target a good estimate, we can:

- reduce the coefficient of variation (but this will require increased computational cost)
- run N_{MCS} different simulations, obtaining $P_{f,1}, P_{f,2}, ..., P_{f,N_{MCS}}$ and choose as estimation of the failure probability the sample mean of the simulations

$$P_{f.m} = \frac{1}{N_{MCS}} \sum_{i=1}^{N_{MCS}} P_{f,i}$$

In most cases, it is typically developed only one MCS with a coefficient of variation $\nu = 5\%$. This provides an estimate that is extremely reasonable for the aims of a reliability analysis. Note however that very often, MCS is considered as 'exact' solution and proposed as the benchmark solution. It would be more correct to define MCS as the 'reference' solution, being aware that it is evaluated within the level of accuracy provided by the chosen coefficient of variation.

Bar example, $\nu = 5\%$. Consider the bar problem, and choose a coefficient of variation $\nu = 5\%$. In this case

$$N \geq \frac{1}{P_f(0.05)^2} = \frac{400}{P_f}$$

Crude MCS:

1. $k = 1, N = 0, N_f = 0, \nu_{lim} = 0.05$
2. Generate $F_y^{(1)}$
3. Generate $S^{(1)}$
4. the first sample is $\mathbf{x}^{(1)} \equiv \left[F_y^{(1)}; S^{(1)}\right]$
5. Evaluate $G_1 = G[\mathbf{x}^{(1)}] = AF_y^{(1)} - S^{(1)}$

$$\begin{cases} G[\mathbf{x}^{(1)}] \leq 0 & N = N + 1, N_f = N_f + 1 \\ G[\mathbf{x}^{(1)}] > 0 & N = N + 1 \end{cases}$$

6. $k = k + 1$
7. Generate $F_y^{(k)}$
8. Generate $S^{(k)}$
9. the $k - th$ sample is $\mathbf{x}^{(k)} \equiv \left[F_y^{(k)}; S^{(k)}\right]$
10. Evaluate $G_k = G[\mathbf{x}^{(k)}] = AF_y^{(k)} - S^{(k)}$

$$\begin{cases} G[\mathbf{x}^{(k)}] \leq 0 & N = N + 1, N_f = N_f + 1 \\ G[\mathbf{x}^{(k)}] > 0 & N = N + 1 \end{cases}$$

11. Evaluate $p = \frac{N_f}{N}$
12. Evaluate $\nu = \sqrt{\frac{1-p}{Np}}$

13. repeat steps 6-12 until $\nu \leq \nu_{lim}$
14. evaluate the failure probability

$$P_f = \frac{N_f}{N}$$

The MCS analysis gives

$$P_{f,1} = 1.05 \times 10^{-2}, \qquad \nu_1 = 0.05, \qquad N_1 = 37,850$$

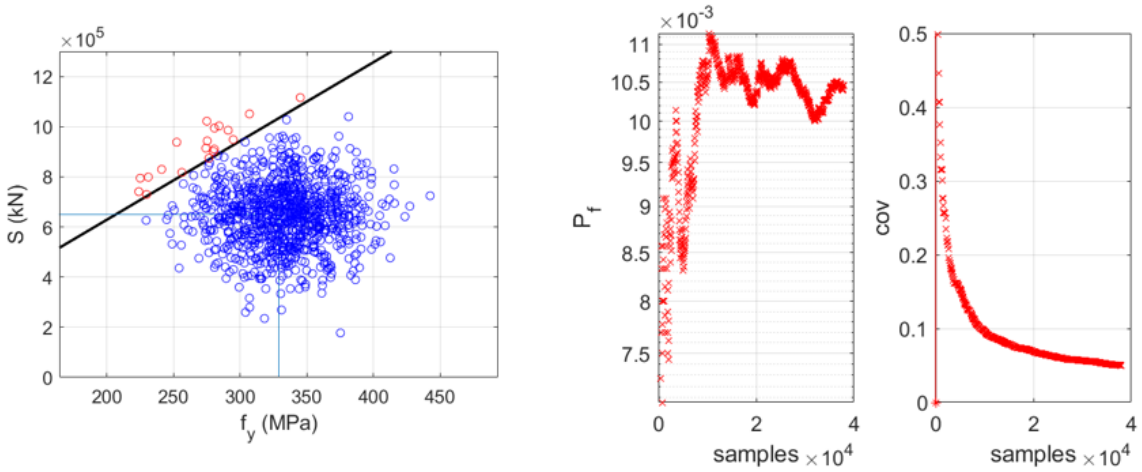


Figure 16.5: MCS,Bar example, $\nu = 5\%$, sample 1

To check the accuracy of the estimate, let us run further MC analyses:

$$P_{f,2} = 1.16 \times 10^{-2}, \qquad \nu_2 = 0.05, \qquad N_2 = 34,058$$

$$P_{f,3} = 0.97 \times 10^{-2}, \qquad \nu_3 = 0.05, \qquad N_3 = 40,764$$

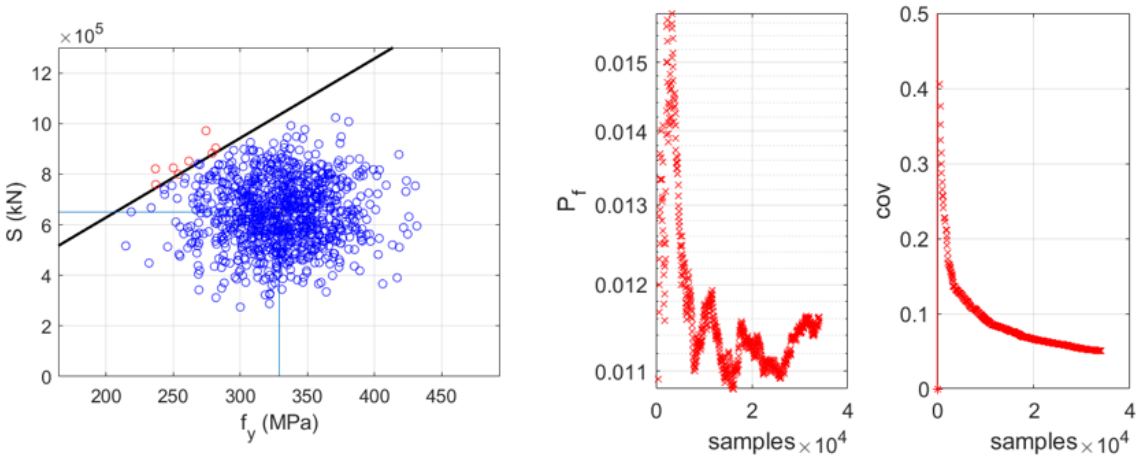


Figure 16.6: MCS,Bar example, $\nu = 5\%$, sample 2

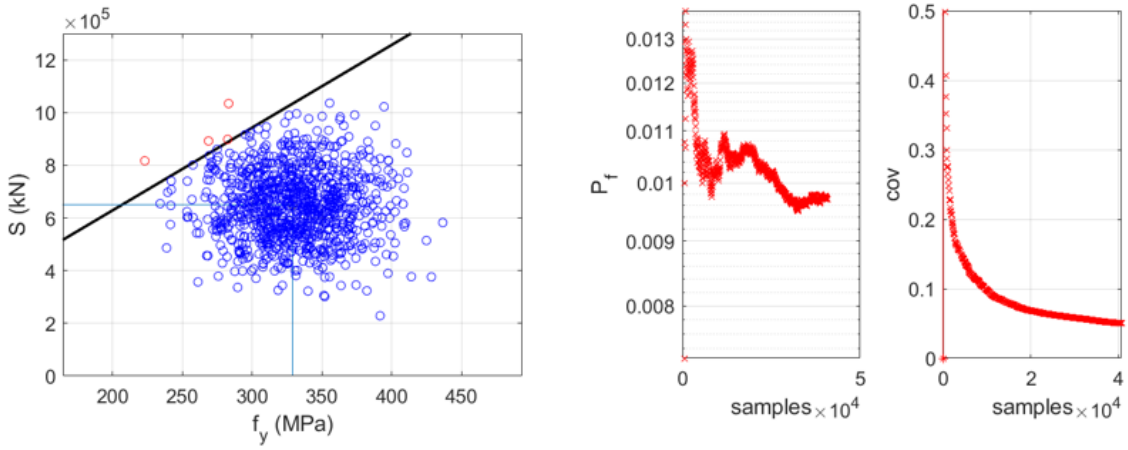


Figure 16.7: MCS,Bar example, $\nu = 5\%$, sample 3

If we run $N_{MCS} = 50$ different simulations we obtain 50 different estimates of the failure probability, all with a accuracy $\nu = 5\%$ and we can see the results show a certain dispersion

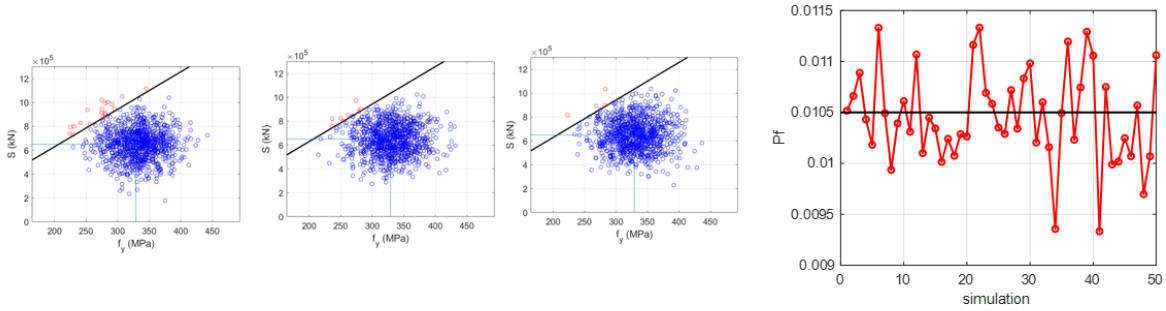


Figure 16.8: MCS,Bar example, $\nu = 5\%$, 50 simulations

around a mean value. Let us evaluate the average failure probability

$$P_f = \frac{1}{N_{MCS}} \sum_{i=1}^{N_{MCS}} P_{f,i} = 1.04 \times 10^{-2}$$

Profile	Area [mm^2]	$P_{f,MCS}$	ν	Samples
$HEA140$	3142	1.05×10^{-2}	0.05	37850
		1.16×10^{-2}	0.05	34058
		0.97×10^{-2}	0.05	40764
		$\mathbf{1.04 \times 10^{-2}}$	**0.05**	

Figure 16.9: MCS,Bar example, $\nu = 5\%$, summary of MCS

Bar example, $\nu = 1\%$. To check the accuracy of the analysis, consider the same bar example, this time with a lower coefficient of variation, i.e. $\nu = 1\%$. In this case it is expected that

$$N \geq \frac{1}{P_f (0.01)^2} = \frac{10,000}{P_f}$$

Let us run a first MCS simulation

$$P_{f,1} = 1.04 \times 10^{-2}, \qquad \nu_1 = 0.01, \qquad N_1 = 954,609$$

As expected, since the level of the failure probability is around $P_f \approx 10^{-2}$ (we already checked this by using a coefficient of variation $\nu = 5\%$), the required number of samples

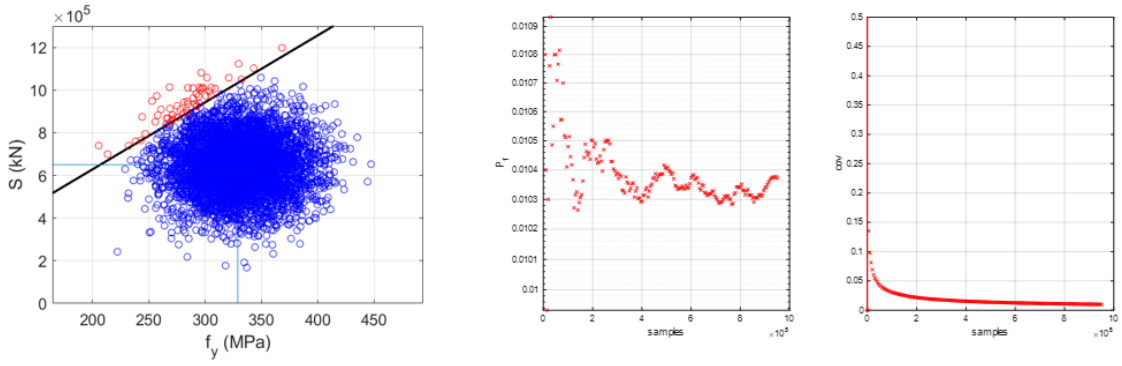


Figure 16.10: MCS,Bar example, $\nu = 1\%$, sample 1

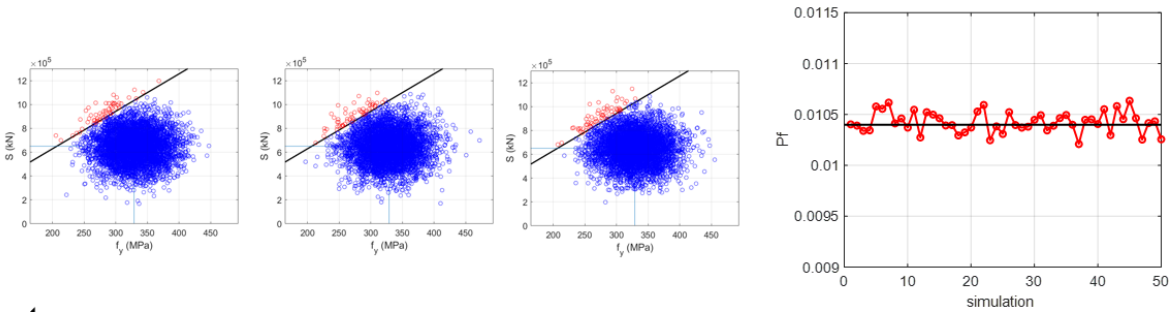


Figure 16.11: MCS,Bar example, $\nu = 1\%$, 50 simulations

to target a degree of accuracy $\nu = 1\%$ is approximately $N \approx 10^6$. Of course, also in this case we can run N_{MCS} different MC simulations, each of them with an accuracy of $\nu = 1\%$ and evaluating their average

$$P_{f,m} = \frac{1}{N_{MCS}} \sum_{i=1}^{N_{MCS}} P_{f,i} = 1.04 \times 10^{-2}$$

The analysis shows that a failure probability $P_f = 1.04 \times 10^{-2}$ is an excellent estimate. Note also that if we adopt a coefficient of variation $\nu = 1\%$, the oscillations around the mean value $P_{f,m}$ are lower. However lower coefficient of variation can require huge computational costs.

16.3 Case of Gaussian uncorrelated

Consider now the case that the basic random variables are n uncorrelated Gaussian random variables $\mathbf{X} = \{X_1, X_2, ..., X_n\}$

Profile	Area [mm²]	$P_{f,MCS}$	ν	Samples
$HEA140$	3142	1.04×10^{-2}	0.01	954609
		1.03×10^{-2}	0.01	960319
		1.05×10^{-2}	0.01	944835
		$\mathbf{1.04 \times 10^{-2}}$	**0.01**	

Figure 16.12: MCS,Bar example, $\nu = 1\%$, summary of MCS

The expected value $\mu_{\mathbf{X}}$, the standard deviation $\sigma_{\mathbf{X}}$, and the matrix of correlation $\mathbf{R}$ are

$$\mu_{\mathbf{X}} = \begin{Bmatrix} \mu_{X_1} \\ \mu_{X_2} \\ \dots \\ \mu_{X_n} \end{Bmatrix}, \qquad \sigma_{\mathbf{X}} = \begin{Bmatrix} \sigma_{X_1} \\ \sigma_{X_2} \\ \dots \\ \sigma_{X_n} \end{Bmatrix}, \qquad \mathbf{R} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \dots & & & \\ 0 & 0 & \dots & 1 \end{bmatrix}$$

while the matrix $\mathbf{D}$ collecting the standard deviations is

$$\mathbf{D} = \begin{bmatrix} \sigma_1 & 0 & \dots & 0 \\ 0 & \sigma_2 & \dots & 0 \\ \dots & & & \\ 0 & 0 & \dots & \sigma_n \end{bmatrix}$$

Let us develop the following coordinate transformation

$$\mathbf{U} = \mathbf{D}^{-1} \left(\mathbf{X} - \mu_{\mathbf{X}} \right) \qquad \Longrightarrow \qquad \begin{cases} U_1 = \frac{X_1 - \mu_1}{\sigma_1} \\ U_2 = \frac{X_2 - \mu_2}{\sigma_2} \\ \dots \\ U_n = \frac{X_n - \mu_n}{\sigma_n} \end{cases},$$

The inverse transformation is

$$\mathbf{X} = \mu_{\mathbf{X}} + \mathbf{D}\mathbf{U} \qquad \Longrightarrow \qquad \begin{cases} X_1 = \mu_1 + \sigma_1 U_1 \\ X_2 = \mu_2 + \sigma_2 U_2 \\ \dots \\ X_n = \mu_n + \sigma_n U_n \end{cases}$$

16.3.1 Original Space

The failure probability P_f with respect to a given limit state function $G(\mathbf{x})$ is

$$P_f = \int_{G(\mathbf{x}) \leq 0} f_{\mathbf{X}} \left(\mathbf{x} \right) d\mathbf{x}$$

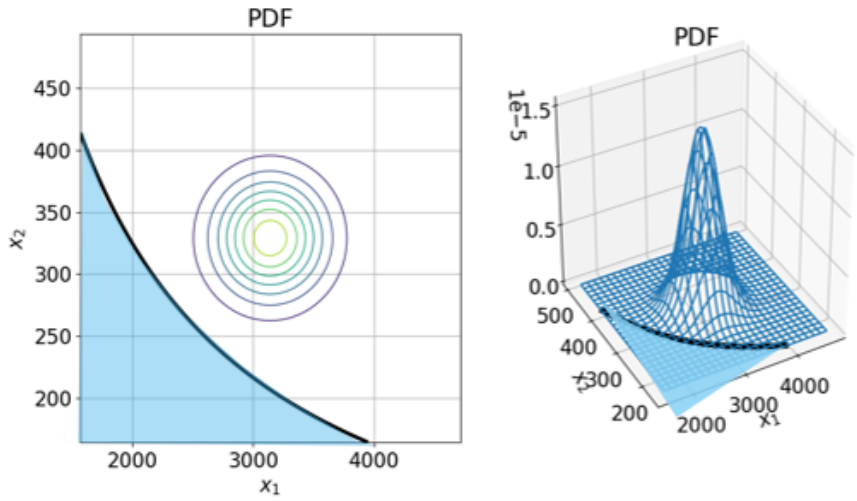


Figure 16.13: Original Space $\mathbf{x}$

Crude MCS:

1. $k = 1, N = 0, N_f = 0$
2. Simulate a first sample $\mathbf{u}^{(1)} \equiv \{u_1^{(1)}, u_2^{(1)}, ..., u_n^{(1)}\}$ from the multivariate normal standard PDF $\varphi_n(\mathbf{u})$
3. Evaluate the corresponding sample in the original space

$$\mathbf{x}^{(1)} = \mathbf{D}\mathbf{u}^{(1)} + \mu_{\mathbf{X}}$$

4. Evaluate $G_1 = G[\mathbf{x}^{(1)}]$

$$\begin{cases} G[\mathbf{x}^{(1)}] \leq 0 & N = N + 1, N_f = N_f + 1 \\ G[\mathbf{x}^{(1)}] > 0 & N = N + 1 \end{cases}$$

5. $k = k + 1$
6. Simulate a sample $\mathbf{u}^{(k)} \equiv \{u_1^{(k)}, u_2^{(k)}, ..., u_n^{(k)}\}$, $k \geq 2$
7. Evaluate the corresponding sample in the original space

$$\mathbf{x}^{(k)} = \mathbf{D}\mathbf{u}^{(k)} + \mu_{\mathbf{X}}$$

8. Evaluate $G_k = G[\mathbf{x}^{(k)}]$

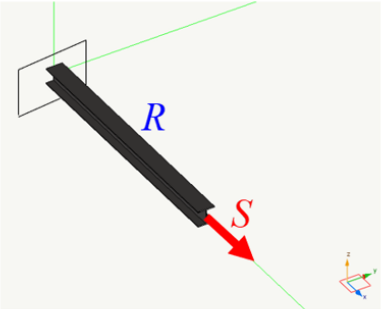
$$\begin{cases} G_k \leq 0 & N = N + 1, N_f = N_f + 1 \\ G_k > 0 & N = N + 1 \end{cases}$$

9. repeat steps 5-8 until needed

10. evaluate the failure probability

$$P_f = \frac{N_f}{N}$$

Case 1: Bar example, uncorrelated variables



Variable		Distribution	μ	v
	a		3,142	
x_1	F_y	Normal	329 MPa	0.10
x_2	S	Normal	650 kN	0.20

Figure 16.14: Bar example, bivariate Gaussian uncorrelated

$$\mu_{\mathbf{X}} = \begin{Bmatrix} 329 \\ 650,000 \end{Bmatrix}, \qquad \sigma_{\mathbf{X}} = \begin{Bmatrix} 32.90 \\ 130,000 \end{Bmatrix}, \qquad \mathbf{R} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \qquad A = 3,142\ mm^2$$

Limit State Function

$$G(x_1, x_2) = Ax_1 - x_2$$

Crude MCS:

1. $k = 1, N = 0, N_f = 0$

2. Simulate a first sample $\mathbf{u}^{(1)} \equiv \{u_1^{(1)}, u_2^{(1)}\}$ from the multivariate normal standard PDF $\varphi_2(u_1, u_2)$

3. Evaluate the corresponding sample in the original space

$$\begin{cases} x_1^{(1)} = \sigma_1 u_1^{(1)} + \mu_1 \\ x_2^{(1)} = \sigma_2 u_2^{(1)} + \mu_2 \end{cases}$$

4. Evaluate the limit state function

$$G_1 = G\left(x_1^{(1)}, x_2^{(1)}\right) = Ax_1^{(1)} - x_2^{(1)}$$

5. Check if $\mathbf{x}^{(1)} \equiv \left[x_1^{(1)}, x_2^{(1)}\right]$ is a failure event

$$\begin{cases} G\left(x_1^{(1)}, x_2^{(1)}\right) \leq 0 & N = N+1, N_f = N_f + 1 \\ G\left(x_1^{(1)}, x_2^{(1)}\right) > 0 & N = N+1 \end{cases}$$

6. $k = k + 1$

7. Simulate a sample $\mathbf{u}^{(k)} \equiv \{u_1^{(k)}, u_2^{(k)}\}$ from the multivariate normal standard PDF $\varphi_2(u_1, u_2)$, $k \geq 2$

8. Evaluate the corresponding sample in the original space

$$\begin{cases} x_1^{(1)} = \sigma_1 u_1^{(1)} + \mu_1 \\ x_2^{(1)} = \sigma_2 u_2^{(1)} + \mu_2 \end{cases}$$

9. Evaluate the limit state function

$$G_k = G\left(x_1^{(k)}, x_2^{(k)}\right) = Ax_1^{(k)} - x_2^{(k)}$$

10. Check if $\mathbf{x}^{(k)} \equiv \left[x_1^{(k)}, x_2^{(k)}\right]$ is a failure event

$$\begin{cases} G\left(x_1^{(k)}, x_2^{(k)}\right) \leq 0 & N = N+1, N_f = N_f + 1 \\ G\left(x_1^{(k)}, x_2^{(k)}\right) > 0 & N = N+1 \end{cases}$$

11. repeat steps 6-10 until needed

12. evaluate the failure probability

$$P_f = \frac{N_f}{N}$$

The estimated failure probability is

$$P_{f,MCS} = \int_{G(\mathbf{x})\leq 0} f_{\mathbf{X}}(\mathbf{x})\, d\mathbf{x} = 1.04 \times 10^{-2}$$

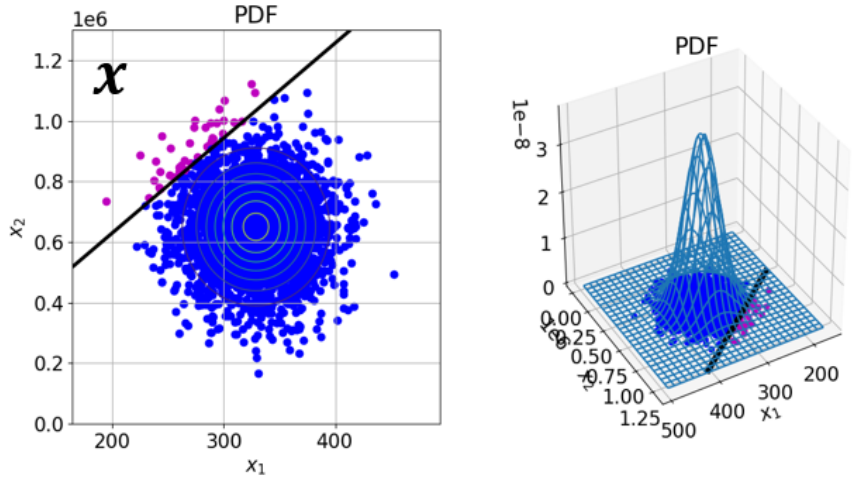


Figure 16.15: Bar example, bivariate Gaussian uncorrelated, MCS, original space $\mathbf{x}$

16.3.2 Normal standard Space u

After isoprobabilistic transformation, the failure probability P_f with respect to a given limit state function $G(\mathbf{x})$ is

$$P_f = \int_{g(\mathbf{u})\leq 0} \varphi_n(\mathbf{u})\, d\mathbf{u}$$

where $g(\mathbf{u})$ is the limit state function mapped in the normal standard space

$$g(\mathbf{u}) = g\left[\mathbf{D}^{-1}(\mathbf{x} - \mu_{\mathbf{X}})\right] \qquad \Rightarrow$$

$$g(u_1, u_2, ..., u_n) = g\left(\frac{x_1 - \mu_1}{\sigma_1}, \frac{x_2 - \mu_2}{\sigma_2}, ..., \frac{x_n - \mu_n}{\sigma_n}\right)$$

Crude MCS:

1. $k = 1, N = 0, N_f = 0$
2. Simulate a first sample $\mathbf{u}^{(1)} \equiv \{u_1^{(1)}, u_2^{(1)}, ..., u_n^{(1)}\}$ from the multivariate normal standard PDF $\varphi_n(\mathbf{u})$
3. Evaluate $g_1 = g[\mathbf{u}^{(1)}]$

$$\begin{cases} g[\mathbf{u}^{(1)}] \leq 0 & N = N + 1, N_f = N_f + 1 \\ g[\mathbf{u}^{(1)}] > 0 & N = N + 1 \end{cases}$$

4. $k = k + 1$

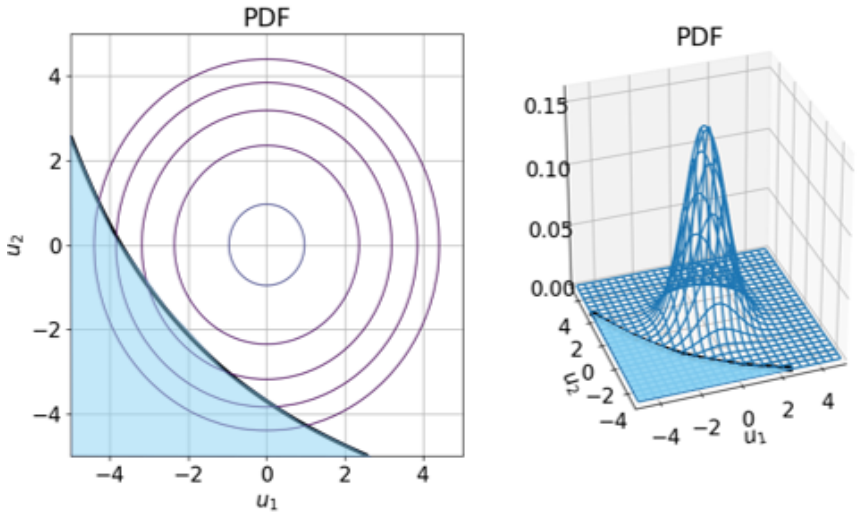


Figure 16.16: Normal standard Space **u**

5. Simulate a sample $\mathbf{u}^{(k)} \equiv \{u_1^{(k)}, u_2^{(k)}, ..., u_n^{(k)}\}$, $k \geq 2$
6. Evaluate $g_k = g[\mathbf{u}^{(k)}]$

$$\begin{cases} g[\mathbf{u}^{(k)}] \leq 0 & N = N + 1, N_f = N_f + 1 \\ g[\mathbf{u}^{(k)}] > 0 & N = N + 1 \end{cases}$$

7. repeat steps 4-6 until needed
8. evaluate the failure probability

$$P_f = \frac{N_f}{N}$$

Case 1: Bar example, uncorrelated variables

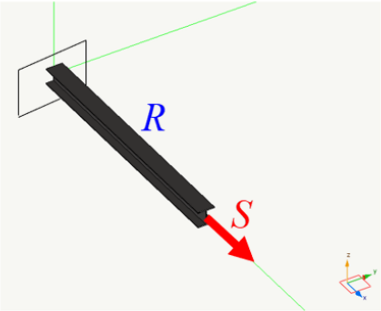
$$\mu_{\mathbf{X}} = \begin{Bmatrix} 329 \\ 650,000 \end{Bmatrix}, \qquad \sigma_{\mathbf{X}} = \begin{Bmatrix} 32.90 \\ 130,000 \end{Bmatrix}, \qquad \mathbf{R} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \qquad A = 3,142\ mm^2$$

Limit State Function **x**-space

$$G(x_1, x_2) = Ax_1 - x_2$$

Coordinate Transformation:

$$\begin{cases} U_1 = \frac{X_1 - \mu_1}{\sigma_1} \\ U_2 = \frac{X_2 - \mu_2}{\sigma_2} \end{cases}, \qquad \begin{cases} X_1 = U_1\sigma_1 + \mu_1 \\ X_2 = U_2\sigma_2 + \mu_2 \end{cases}$$



Variable		Distribution	μ	v
	a		3,142	
x_1	F_y	Normal	329 MPa	0.10
x_2	S	Normal	650 kN	0.20

Figure 16.17: Bar example, bivariate Gaussian uncorrelated

Limit State Function **u**-space

$$G(u_1, u_2) = A(u_1\sigma_1 + \mu_1) - (u_2\sigma_2 + \mu_2)$$

Crude MCS:

1. $k = 1, N = 0, N_f = 0$
2. Simulate a first sample $\mathbf{u}^{(1)} \equiv \{u_1^{(1)}, u_2^{(1)}\}$ from the multivariate normal standard PDF $\varphi_2(u_1, u_2)$
3. Evaluate the limit state function

$$g_1 = g\left(u_1^{(1)}, u_2^{(1)}\right)$$

$$\begin{cases} g\left(u_1^{(1)}, u_2^{(1)}\right) \leq 0 & N = N + 1, N_f = N_f + 1 \\ g\left(u_1^{(1)}, u_2^{(1)}\right) > 0 & N = N + 1 \end{cases}$$

4. $k = k + 1$
5. Simulate a sample $\mathbf{u}^{(k)} \equiv \{u_1^{(k)}, u_2^{(k)}\}$ from the multivariate normal standard PDF $\varphi_2(u_1, u_2)$, $k \geq 2$
6. Evaluate the limit state function

$$g_k = g\left(u_1^{(k)}, u_2^{(k)}\right)$$

$$\begin{cases} g\left(u_1^{(k)}, u_2^{(k)}\right) \leq 0 & N = N + 1, N_f = N_f + 1 \\ g\left(u_1^{(k)}, u_2^{(k)}\right) > 0 & N = N + 1 \end{cases}$$

7. repeat steps 4-6 until needed

8. evaluate the failure probability

$$P_f = \frac{N_f}{N}$$

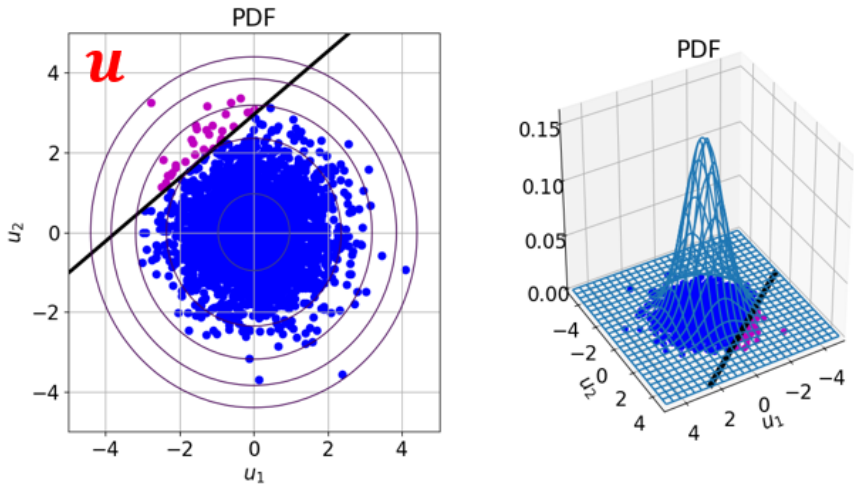


Figure 16.18: Bar example, bivariate Gaussian uncorrelated, MCS, normal standard space

The estimated failure probability is

$$P_{f,MCS} = \int_{g(\mathbf{u})\leq 0} \varphi_n(\mathbf{u})\, d\mathbf{u} = 1.04 \times 10^{-2}$$

16.4 Case of Gaussian correlated

Consider now the case that the basic random variables are n correlated Gaussian random variables $\mathbf{X} = \{X_1, X_2, ..., X_n\}$

The expected value $\mu_{\mathbf{X}}$, the standard deviation $\sigma_{\mathbf{X}}$, and the matrix of correlation $\mathbf{R}$ are

$$\mu_{\mathbf{X}} = \begin{Bmatrix} \mu_{X_1} \\ \mu_{X_2} \\ ... \\ \mu_{X_n} \end{Bmatrix}, \qquad \sigma_{\mathbf{X}} = \begin{Bmatrix} \sigma_{X_1} \\ \sigma_{X_2} \\ ... \\ \sigma_{X_n} \end{Bmatrix}, \qquad \mathbf{R} = \begin{bmatrix} 1 & \rho_{X_1X_2} & ... & \rho_{X_1X_n} \\ \rho_{X_1X_2} & 1 & ... & \rho_{X_2X_n} \\ ... & & & \\ \rho_{X_1X_n} & \rho_{X_2X_n} & ... & 1 \end{bmatrix}$$

while the matrix $\mathbf{D}$ collecting the standard deviations is

$$\mathbf{D} = \begin{bmatrix} \sigma_{X_1} & 0 & ... & 0 \\ 0 & \sigma_{X_2} & ... & 0 \\ ... & & & \\ 0 & 0 & ... & \sigma_{X_n} \end{bmatrix}$$

Step 1. Coordinate transformation $\mathbf{x} \to \mathbf{z}$ Let us develop the following coordinate transformation

$$\mathbf{Z} = \mathbf{D}^{-1}\left(\mathbf{X} - \mu_{\mathbf{X}}\right) \qquad \Longrightarrow \qquad \begin{cases} Z_1 = \frac{X_1 - \mu_1}{\sigma_1} \\ Z_2 = \frac{X_2 - \mu_2}{\sigma_2} \\ \ldots \\ Z_n = \frac{X_n - \mu_n}{\sigma_n} \end{cases},$$

The inverse transformation is

$$\mathbf{X} = \mu_{\mathbf{X}} + \mathbf{DZ} \qquad \Longrightarrow \qquad \begin{cases} X_1 = \mu_1 + \sigma_1 Z_1 \\ X_2 = \mu_2 + \sigma_2 Z_2 \\ \ldots \\ X_n = \mu_n + \sigma_n Z_n \end{cases}$$

Step 2. Coordinate transformation $\mathbf{z} \to \mathbf{u}$ First, we develop a Cholesky decomposition of the matrix of correlation $\mathbf{R}$

$$\mathbf{R} = \mathbf{L}\mathbf{L}^{\mathbf{T}}$$

$$\begin{bmatrix} 1 & \rho_{12} & \ldots & \rho_{1n} \\ \rho_{12} & 1 & \ldots & \rho_{2n} \\ \ldots & & & \\ \rho_{1n} & \rho_{2n} & \ldots & 1 \end{bmatrix} = \begin{bmatrix} L_{11} & 0 & \ldots & 0 \\ L_{21} & L_{22} & \ldots & 0 \\ \ldots & & & \\ L_{n1} & L_{n2} & \ldots & L_{nn} \end{bmatrix} \begin{bmatrix} L_{11} & L_{21} & \ldots & L_{n1} \\ 0 & L_{22} & \ldots & L_{n2} \\ \ldots & & & \\ 0 & 0 & \ldots & L_{nn} \end{bmatrix}$$

where $\mathbf{L}$ is lower triangular matrix. Let us develop the following coordinate transformation

$$\mathbf{Z} = \mathbf{L}\mathbf{U}, \qquad \Rightarrow \qquad \begin{Bmatrix} Z_1 \\ Z_2 \\ \ldots \\ Z_n \end{Bmatrix} = \begin{bmatrix} L_{11} & 0 & \ldots & 0 \\ L_{21} & L_{22} & \ldots & 0 \\ \ldots & & & \\ L_{n1} & L_{n2} & \ldots & L_{nn} \end{bmatrix} \begin{Bmatrix} U_1 \\ U_2 \\ \ldots \\ U_n \end{Bmatrix}$$

The inverse transformation is

$$\mathbf{U} = \mathbf{L}^{-1}\mathbf{Z}, \qquad \Rightarrow \qquad \begin{Bmatrix} U_1 \\ U_2 \\ \ldots \\ U_n \end{Bmatrix} = \begin{bmatrix} L_{11} & L_{21} & \ldots & L_{n1} \\ 0 & L_{22} & \ldots & L_{n2} \\ \ldots & & & \\ 0 & 0 & \ldots & L_{nn} \end{bmatrix} \begin{Bmatrix} Z_1 \\ Z_2 \\ \ldots \\ Z_n \end{Bmatrix}$$

Coordinate transformation $\mathbf{x} \to \mathbf{u}$ The isoprobabilistic transformation from the original $\mathbf{x}$-space to the normal standard space $\mathbf{u}$ is

$$\mathbf{X} = \mathbf{DLU} + \mu$$

The inverse transformation is

$$\mathbf{U} = \mathbf{L}^{-1}\mathbf{D}^{-1}\left(\mathbf{X} - \mu\right)$$

16.4.1 Original Space

The failure probability P_f with respect to a given limit state function $G(\mathbf{x})$ is

$$P_f = \int_{G(\mathbf{x})\leq 0} f_{\mathbf{X}}(\mathbf{x})\, d\mathbf{x}$$

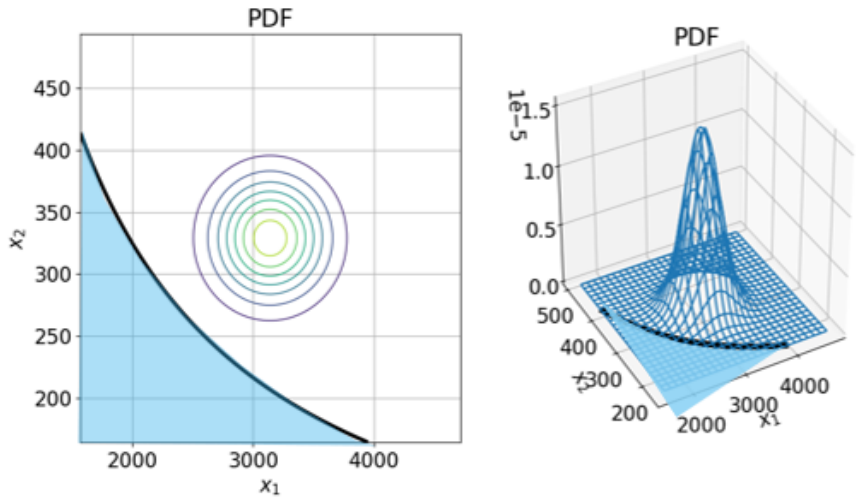


Figure 16.19: Original Space $\mathbf{x}$

Crude MCS:

1. $k = 1, N = 0, N_f = 0$
2. Simulate a first sample $\mathbf{u}^{(1)} \equiv \{u_1^{(1)}, u_2^{(1)}, ..., u_n^{(1)}\}$ from the multivariate normal standard PDF $\varphi_n(\mathbf{u})$
3. Evaluate the corresponding sample in the original space

$$\mathbf{x}^{(1)} = \mathbf{DLu}^{(1)} + \mu_{\mathbf{X}}$$

4. Evaluate $G_1 = G[\mathbf{x}^{(1)}]$

$$\begin{cases} G[\mathbf{x}^{(1)}] \leq 0 & N = N + 1, N_f = N_f + 1 \\ G[\mathbf{x}^{(1)}] > 0 & N = N + 1 \end{cases}$$

5. $k = k + 1$

6. Simulate a sample $\mathbf{u}^{(k)} \equiv \{u_1^{(k)}, u_2^{(k)}, ..., u_n^{(k)}\}$, $k \geq 2$
7. Evaluate the corresponding sample in the original space

$$\mathbf{x}^{(k)} = \mathbf{DLu}^{(k)} + \mu_{\mathbf{X}}$$

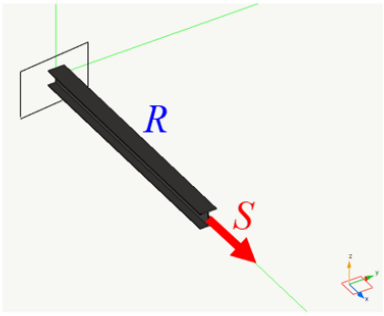
8. Evaluate $G_k = G[\mathbf{x}^{(k)}]$

$$\begin{cases} G_k \leq 0 & N = N + 1, N_f = N_f + 1 \\ G_k > 0 & N = N + 1 \end{cases}$$

9. repeat steps 5-8 until needed
10. evaluate the failure probability

$$P_f = \frac{N_f}{N}$$

Case 2: Bar example, $\rho = -0.7$



Variable		Distribution	μ	v
	a		3,142	
x_1	F_y	Normal	329 MPa	0.10
x_2	S	Normal	650 kN	0.20

Figure 16.20: Bar example, bivariate Gaussian

$$\mu_{\mathbf{X}} = \begin{Bmatrix} 329 \\ 650,000 \end{Bmatrix}, \qquad \sigma_{\mathbf{X}} = \begin{Bmatrix} 32.90 \\ 130,000 \end{Bmatrix}, \qquad \mathbf{R} = \begin{bmatrix} 1 & -0.70 \\ -0.70 & 1 \end{bmatrix} \qquad A = 3,142\, mm^2$$

Limit State Function

$$G(x_1, x_2) = Ax_1 - x_2$$

It is noted that in the $\mathbf{x}$-space the limit state is not affected by the presence of the correlation, while of course the joint PDF $f_{X_1X_2}(x_1, x_2)$ changes.

Step 1. Transformation $\mathbf{x} \to \mathbf{z}$:

$$\mathbf{z} = \mathbf{D}^{-1}(\mathbf{x} - \mu_{\mathbf{X}}) \qquad \Rightarrow \qquad \begin{cases} z_1 = \frac{x_1 - \mu_1}{\sigma_1} \\ z_2 = \frac{x_2 - \mu_2}{\sigma_2} \end{cases}$$

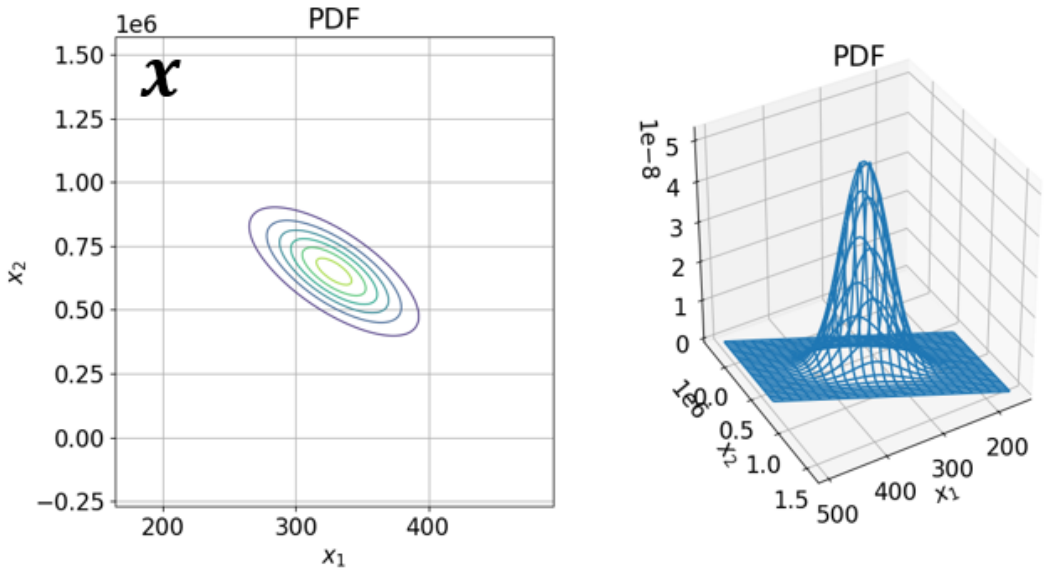


Figure 16.21: Bar example, bivariate Gaussian, $\rho = -0.7$, joint PDF $f_{X_1X_2}(x_1, x_2)$

The inverse transformation is

$$\mathbf{x} = \mu_{\mathbf{X}} + \mathbf{D}\mathbf{z} \qquad \Longrightarrow \qquad \begin{cases} x_1 = \mu_1 + \sigma_1 z_1 \\ x_2 = \mu_2 + \sigma_2 z_2 \end{cases}$$

Step 2. Coordinate transformation $\mathbf{z} \to \mathbf{u}$

First, we develop a Cholesky decomposition of the matrix of correlation $\mathbf{R}$

$$\mathbf{R} = \mathbf{L}\mathbf{L}^{\mathbf{T}}$$

where $\mathbf{L}$ is the lower triangular matrix

$$\mathbf{L} = \begin{bmatrix} 1 & 0 \\ \rho & \sqrt{1-\rho^2} \end{bmatrix}$$

Let us develop the following coordinate transformation

$$\mathbf{z} = \mathbf{L}\mathbf{u} \qquad \Rightarrow \qquad \begin{cases} z_1 = u_1 \\ z_2 = \rho u_1 + \sqrt{1-\rho^2} u_2 \end{cases}$$

The inverse transformation is

$$\mathbf{u} = \mathbf{L}^{-1}\mathbf{z} \qquad \Rightarrow \qquad \begin{cases} u_1 = u_1 \\ u_2 = \frac{z_2 - \rho z_1}{\sqrt{1-\rho^2}} \end{cases}$$

Isoprobabilistic transformation $\mathbf{x} \rightarrow \mathbf{u}$.

Let us develop the following coordinate transformation

$$\mathbf{x} = \mathbf{DLu} + \mu \qquad \Rightarrow \qquad \begin{cases} x_1 = \sigma_1 u_1 + \mu_1 \\ x_2 = \sigma_2 \left(\rho u_1 + \sqrt{1-\rho^2} u_2 \right) + \mu_2 \end{cases}$$

The inverse transformation is

$$\mathbf{u} = \mathbf{L}^{-1}\mathbf{D}^{-1}\left(\mathbf{x} - \mu\right) \qquad \Rightarrow \qquad \begin{cases} u_1 = \frac{x_1 - \mu_1}{\sigma_1} \\ u_2 = \frac{1}{\sqrt{1-\rho^2}} \left[\frac{x_2 - \mu_2}{\sigma_2} - \rho \frac{x_1 - \mu_1}{\sigma_1} \right] \end{cases}$$

Crude MCS:

1. $k = 1, N = 0, N_f = 0$
2. Simulate a first sample $\mathbf{u}^{(1)} \equiv \{u_1^{(1)}, u_2^{(1)}\}$ from the multivariate normal standard PDF $\varphi_2\left(u_1, u_2\right)$
3. Evaluate the corresponding sample in the original space

$$\begin{cases} x_1^{(1)} = \sigma_1 u_1^{(1)} + \mu_1 \\ x_2^{(1)} = \sigma_2 \left(\rho u_1^{(1)} + \sqrt{1-\rho^2} u_2^{(1)} \right) + \mu_2 \end{cases}$$

4. Evaluate the limit state function

$$G_1 = G\left(x_1^{(1)}, x_2^{(1)}\right) = A x_1^{(1)} - x_2^{(1)}$$

$$\begin{cases} G\left(x_1^{(1)}, x_2^{(1)}\right) \leq 0 & N = N + 1, N_f = N_f + 1 \\ G\left(x_1^{(1)}, x_2^{(1)}\right) > 0 & N = N + 1 \end{cases}$$

5. $k = k + 1$
6. Simulate a sample $\mathbf{u}^{(k)} \equiv \{u_1^{(k)}, u_2^{(k)}\}$ from the multivariate normal standard PDF $\varphi_2\left(u_1, u_2\right)$, $k \geq 2$
7. Evaluate the corresponding sample in the original space

$$\begin{cases} x_1^{(k)} = \sigma_1 u_1^{(k)} + \mu_1 \\ x_2^{(k)} = \sigma_2 \left(\rho u_1^{(k)} + \sqrt{1-\rho^2} u_2^{(k)} \right) + \mu_2 \end{cases}$$

8. Evaluate the limit state function

$$G_k = G\left(x_1^{(k)}, x_2^{(k)}\right) = Ax_1^{(k)} - x_2^{(k)}$$

$$\begin{cases} G\left(x_1^{(k)}, x_2^{(k)}\right) \leq 0 & N = N+1, N_f = N_f + 1 \\ G\left(x_1^{(k)}, x_2^{(k)}\right) > 0 & N = N+1 \end{cases}$$

9. repeat steps 5-8 until needed

10. evaluate the failure probability

$$P_f = \frac{N_f}{N}$$

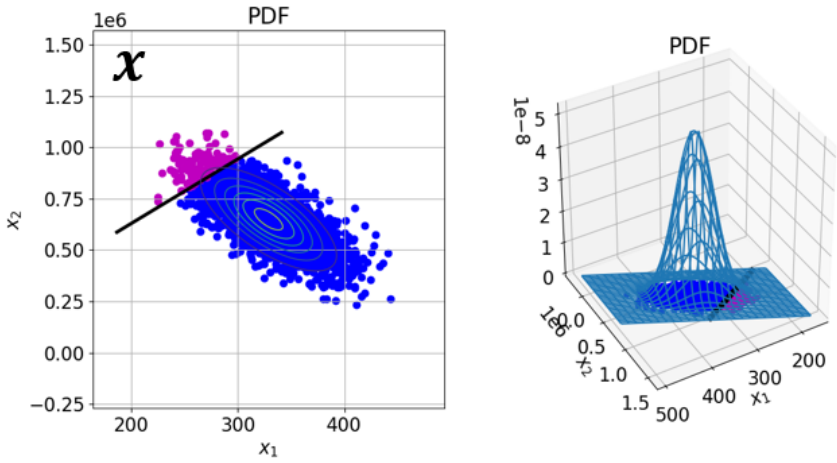


Figure 16.22: Bar example, bivariate Gaussian $\rho = -0.7$, MCS, original space $\mathbf{x}$

The estimated failure probability is

$$P_{f,MCS} = \int_{G(\mathbf{x})\leq 0} f_{\mathbf{X}}(\mathbf{x})\, d\mathbf{x} = 3.74 \times 10^{-2}$$

which is approximately four times the failure probability, with respect to the case of uncorrelated random variables.

16.4.2 Normal standard Space u

The MCS can be developed directly into the normal standard space, after isoprobabilistic transformation. The failure probability P_f with respect to a given limit state function $G(\mathbf{x})$ is

$$P_f = \int_{g(\mathbf{u})\leq 0} \varphi_n(\mathbf{u})\, d\mathbf{u}$$

where $g(\mathbf{u})$ is the limit state function mapped in the normal standard space

$$g(\mathbf{u}) = g\left[\mathbf{L}^{-1}\mathbf{D}^{-1}(\mathbf{x} - \mu_{\mathbf{X}})\right] \qquad \Rightarrow$$

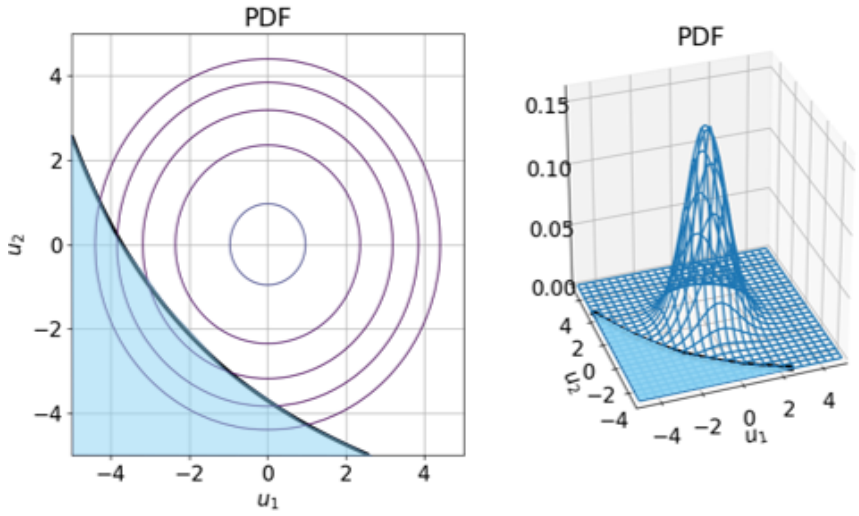


Figure 16.23: Normal standard Space $\mathbf{u}$

Crude MCS:

1. $k = 1, N = 0, N_f = 0$
2. Simulate a first sample $\mathbf{u}^{(1)} \equiv \{u_1^{(1)}, u_2^{(1)}, ..., u_n^{(1)}\}$ from the multivariate normal standard PDF $\varphi_n(\mathbf{u})$
3. Evaluate $g_1 = g[\mathbf{u}^{(1)}]$

$$\begin{cases} g[\mathbf{u}^{(1)}] \leq 0 & N = N + 1, N_f = N_f + 1 \\ g[\mathbf{u}^{(1)}] > 0 & N = N + 1 \end{cases}$$

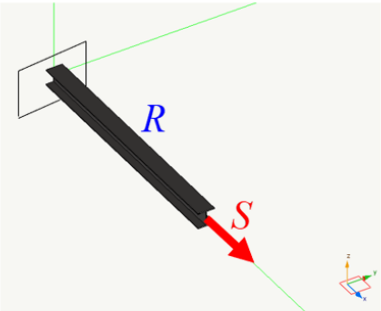
4. $k = k + 1$
5. Simulate a sample $\mathbf{u}^{(k)} \equiv \{u_1^{(k)}, u_2^{(k)}, ..., u_n^{(k)}\}$, $k \geq 2$
6. Evaluate $g_k = g[\mathbf{u}^{(k)}]$

$$\begin{cases} g[\mathbf{u}^{(k)}] \leq 0 & N = N + 1, N_f = N_f + 1 \\ g[\mathbf{u}^{(k)}] > 0 & N = N + 1 \end{cases}$$

7. repeat steps 4-6 until needed
8. evaluate the failure probability

$$P_f = \frac{N_f}{N}$$

Case 2: Bar example, $\rho = -0.7$



Variable		Distribution	μ	v
	a		3,142	
x_1	F_y	Normal	329 MPa	0.10
x_2	S	Normal	650 kN	0.20

Figure 16.24: Bar example, bivariate Gaussian

$$\mu_{\mathbf{X}} = \begin{Bmatrix} 329 \\ 650,000 \end{Bmatrix}, \qquad \sigma_{\mathbf{X}} = \begin{Bmatrix} 32.90 \\ 130,000 \end{Bmatrix}, \qquad \mathbf{R} = \begin{bmatrix} 1 & -0.70 \\ -0.70 & 1 \end{bmatrix} \qquad A = 3,142\, mm^2$$

Limit State Function

$$G(x_1, x_2) = Ax_1 - x_2$$

Isoprobabilistic transformation $\mathbf{x} \rightarrow \mathbf{u}$.

Let us develop the following coordinate transformation

$$\mathbf{x} = \mathbf{DLu} + \mu \qquad \Rightarrow \qquad \begin{cases} x_1 = \sigma_1 u_1 + \mu_1 \\ x_2 = \sigma_2 \left(\rho u_1 + \sqrt{1-\rho^2} u_2 \right) + \mu_2 \end{cases}$$

where

$$\mathbf{D} = \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix}, \qquad \mathbf{L} = \begin{bmatrix} 1 & 0 \\ \rho & \sqrt{1-\rho^2} \end{bmatrix}$$

The inverse transformation is

$$\mathbf{u} = \mathbf{L}^{-1}\mathbf{D}^{-1}(\mathbf{x} - \mu) \qquad \Rightarrow \qquad \begin{cases} u_1 = \frac{x_1 - \mu_1}{\sigma_1} \\ u_2 = \frac{1}{\sqrt{1-\rho^2}} \left[\frac{x_2 - \mu_2}{\sigma_2} - \rho \frac{x_1 - \mu_1}{\sigma_1} \right] \end{cases}$$

Limit state in the normal standard space $\mathbf{u}$

$$g(u_1, u_2) = A(\sigma_1 u_1 + \mu_1) - \left[\sigma_2\left(\rho u_1 + \sqrt{1-\rho^2}u_2\right) + \mu_2\right]$$

Crude MCS:

1. $k = 1, N = 0, N_f = 0$
2. Simulate a first sample $\mathbf{u}^{(1)} \equiv \{u_1^{(1)}, u_2^{(1)}\}$ from the multivariate normal standard PDF $\varphi_2(u_1, u_2)$
3. Evaluate the limit state function

$$g_1 = g\left(u_1^{(1)}, u_2^{(1)}\right)$$

$$\begin{cases} g\left(u_1^{(1)}, u_2^{(1)}\right) \leq 0 & N = N+1, N_f = N_f + 1 \\ g\left(u_1^{(1)}, u_2^{(1)}\right) > 0 & N = N+1 \end{cases}$$

4. $k = k + 1$
5. Simulate a sample $\mathbf{u}^{(k)} \equiv \{u_1^{(k)}, u_2^{(k)}\}$ from the multivariate normal standard PDF $\varphi_2(u_1, u_2)$, $k \geq 2$
6. Evaluate the corresponding sample in the original space
7. Evaluate the limit state function

$$g_k = g\left(u_1^{(k)}, u_2^{(k)}\right)$$

$$\begin{cases} g\left(u_1^{(k)}, u_2^{(k)}\right) \leq 0 & N = N+1, N_f = N_f + 1 \\ g\left(u_1^{(k)}, u_2^{(k)}\right) > 0 & N = N+1 \end{cases}$$

8. repeat steps 4-7 until needed
9. evaluate the failure probability

$$P_f = \frac{N_f}{N}$$

The estimated failure probability is

$$P_{f,MCS} = \int_{g(\mathbf{u})\leq 0} \varphi_2(\mathbf{u_1}, \mathbf{u_2})\, d\mathbf{u} = 3.74 \times 10^{-2}$$

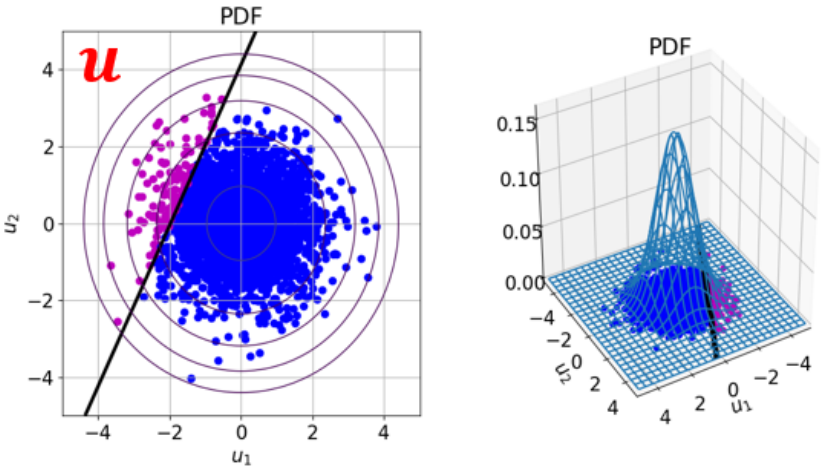


Figure 16.25: Bar example, bivariate Gaussian $\rho = -0.7$, MCS, normal standard space $\mathbf{u}$

Case 3: Bar example, $\rho = 0.7$

$$\mu_{\mathbf{X}} = \begin{Bmatrix} 329 \\ 650,000 \end{Bmatrix}, \qquad \sigma_{\mathbf{X}} = \begin{Bmatrix} 32.90 \\ 130,000 \end{Bmatrix}, \qquad \mathbf{R} = \begin{bmatrix} 1 & 0.70 \\ 0.70 & 1 \end{bmatrix} \qquad A = 3,142\ mm^2$$

Limit State Function

$$G(x_1, x_2) = Ax_1 - x_2$$

Limit state in the normal standard space $\mathbf{u}$

$$g(u_1, u_2) = A(\sigma_1 u_1 + \mu_1) - \left[\sigma_2 \left(\rho u_1 + \sqrt{1-\rho^2} u_2\right) + \mu_2\right]$$

The estimated failure probability is

$$P_{f,MCS} = \int_{g(\mathbf{u}) \leq 0} \varphi_2(\mathbf{u_1}, \mathbf{u_2})\, d\mathbf{u} = 2.09 \times 10^{-5}$$

The example shows that the presence of correlation can provide a failure probability more than 1,000 times different with respect to the case of uncorrelated variables. Therefore, the correlation, if available, cannot be neglected.

It is noted that the random variables X_1 and X_2 are Gaussian, while the Limit State Function $G(X_1, X_2)$ is their linear combination. Under these circumstances

- the limit state in the original space $\mathbf{x}$ is linear
- the limit state in the standard normal space $\mathbf{u}$ is linear too

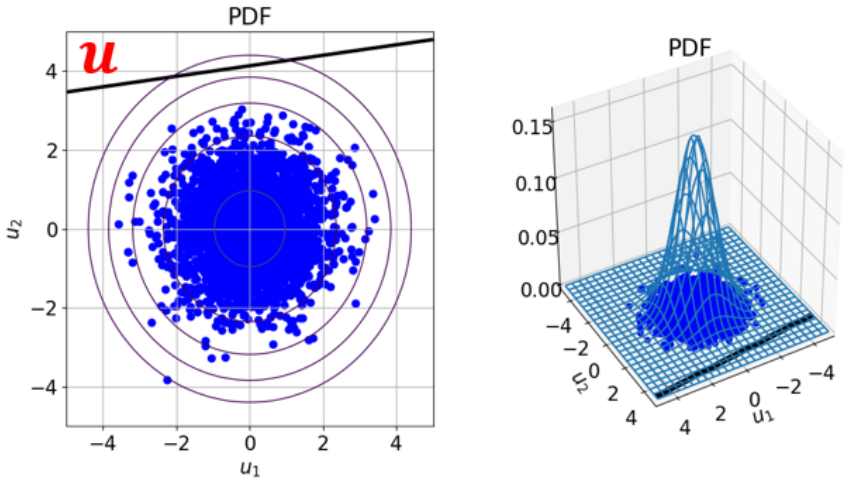


Figure 16.26: Bar example, bivariate Gaussian $\rho = 0.7$, MCS, normal standard space $\mathbf{u}$

- the correlation can significantly modify the failure probability
- in the normal standard space the effect of the correlation determine a limit state closer or far from the origin

Chapter 17

Lecture B2:Reliability Index

17.1 Introduction

In the most general problem, we can define a set of basic random variables $x_1, x_2, ..., x_n$, which typically can give rise to:

- *Capacity* $R(\mathbf{x})$
- *Demand* $S(\mathbf{x})$

And define the following broad problem:

$$\begin{cases} R > S: & Safe \\ R \leq S: & Failure \end{cases}$$

But, *Capacity R and demand S are uncertain parameters!*

The failure probability with respect to a chosen limit state is

$$P_f = Prob\left[R \leq S\right] \tag{17.1}$$

The reliability is the complementary of the failure probability

$$Rel = 1 - P_f \tag{17.2}$$

The risk is given by the consequence C multiplied by the failure probability P_f

$$R = P_f \; C \tag{17.3}$$

Bar example. Assume that we have a bar in steel, whose yield strength is F_y, and subject to a load S. Let us assume that the area of the cross section is A, then

$$\begin{cases} R = AF_y : & \textit{Capacity} \\ S : & \textit{Demand} \end{cases}$$

where *Capacity* R and *Demand* S are random variables

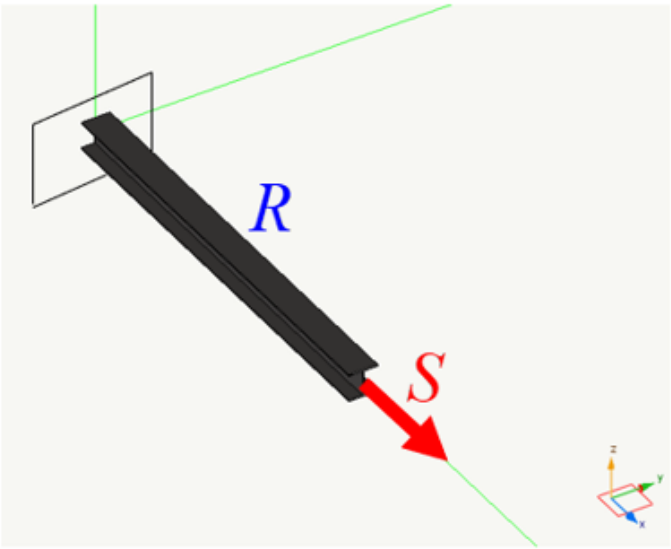


Figure 17.1: Bar example

- *Failure probability*: $P_f = Prob[AF_y \leq S]$
- *Reliability*: $Rel = 1 - P_f = Prob[AF_y > S]$
- *Risk*: $R = P_f\ C$

17.2 Cornell Reliability index

For this basic example, the Limit State Function (LSF) is defined as

$$G = R - S \tag{17.4}$$

and the failure probability is

$$P_f = Prob\left[R \leq S\right] = Prob\left[G \leq 0\right] \tag{17.5}$$

Let us assume now that:

- R is Gaussian
- S is Gaussian

Then, the limit state function is Gaussian too, its mean value μ_G and σ_G are

$$\mu_G = \mu_R - \mu_S \tag{17.6}$$

$$\sigma_G = \sqrt{\sigma_R^2 + \sigma_S^2} \tag{17.7}$$

The **Cornell Reliability index** is defined as

$$\beta_C = \frac{\mu_G}{\sigma_G} = \frac{\mu_R - \mu_S}{\sqrt{\sigma_R^2 + \sigma_S^2}} \tag{17.8}$$

The distance β_C, measured in terms of σ_G between μ_G and the limit state $G = 0$ gives a measure of the reliability

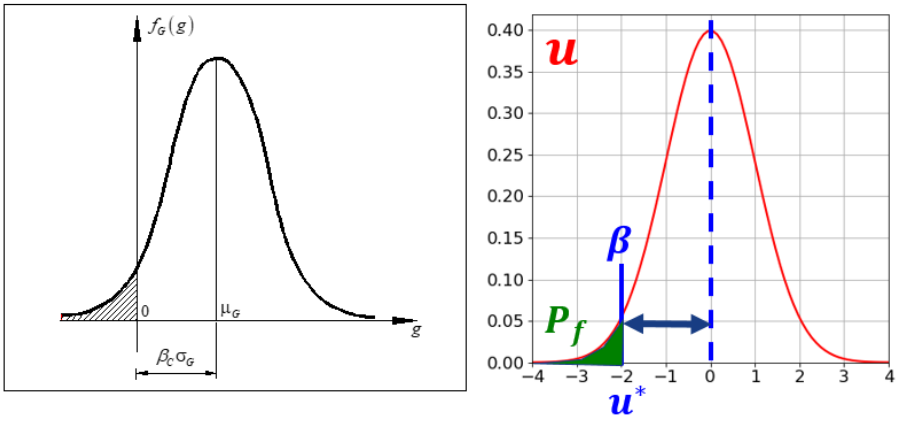


Figure 17.2: Cornell Reliability index β_C

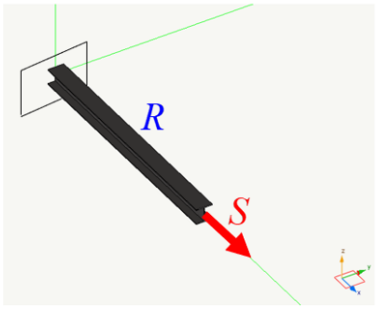
Let us develop a coordinate transformation from the Gaussian PDF $f_G(g)$ (having mean μ_G and standard deviation σ_G) toward the normal standard PDF

$$u = \frac{g - \mu_G}{\sigma_G} \qquad \Rightarrow \qquad u^* = \frac{0 - \mu_G}{\sigma_G} = -\beta$$

It follows

$$P_f = P\left[G \leq 0\right] = \Phi\left(\frac{0 - \mu_G}{\sigma_G}\right) = \Phi\left(-\beta\right) \tag{17.9}$$

The distance β_C, in the normal standard space, between the origin and the limit state $g = 0$ gives a measure of the reliability



Variable		Distribution	μ	σ	ν
	A		3,142		
x_1	F_y	Normal	329 MPa	32.90 MPa	0.10
x_2	S	Normal	650 kN	130 kN	0.20

Correlation		
	X_1	X_2
X_1	1	0
X_2	0	1

Figure 17.3: Bar example

Bar example.

Basic Random variables:

- $x_1 \equiv R$: Gaussian random variable, $\mu_R = 329A$ N, where A is the area of the section, and coefficient of variation $\nu_R = 0.10$
- $x_2 \equiv S$: Gaussian random variable, $\mu_S = 650$ kN, and coefficient of variation $\nu_R = 0.20$

Limit State Function:

$$G(R, S) = R - S$$

Cornell Reliability Index:

$$\beta = \frac{\mu_G}{\sigma_G} = \frac{\mu_R - \mu_S}{\sqrt{\sigma_R^2 + \sigma_S^2}}$$

Failure probability:

$$P_f = P[G \leq 0] = \Phi(-\beta)$$

As expected, it is seen that for increasing values of area of the profile:

- the reliability index β increases
- the failure probability P_f decreases

Profile	Area [mm²]	β	P_f
$HE140A$	3,142	2.310	1.04×10^{-2}
$HE180A$	4525	4.243	1.10×10^{-6}
$HE200A$	5,383	5.102	1.67×10^{-7}

Figure 17.4: Bar example, different areas

17.2.1 Gaussian correlated

The Cornell reliability index can be easily extended to multiple Gaussian random variables (independent or correlated)

$$G = a_0 + a_1 X_1 + a_2 X_2 + ... + a_n X_n = a_0 + \mathbf{a}^T \mathbf{x}$$

The limit state function is Gaussian too, its mean value μ_G and σ_G are

$$\mu_G = a_0 + \mathbf{a}^T \mu_{\mathbf{X}}$$

$$\sigma_G = \sqrt{\mathbf{a}^T \mathbf{\Sigma_X} \mathbf{a}}$$

The reliability index is

$$\beta_C = \frac{\mu_G}{\sigma_G} = \frac{a_0 + \mathbf{a}^T \mu_{\mathbf{X}}}{\sqrt{\mathbf{a}^T \mathbf{\Sigma_X} \mathbf{a}}}$$

Bar example.

Basic Random variables:

- $x_1 \equiv R$: Gaussian random variable, $\mu_R = 1,033$ kN, and coefficient of variation $\nu_R = 0.10$
- $x_2 \equiv S$: Gaussian random variable, $\mu_S = 650$ kN, and coefficient of variation $\nu_R = 0.20$
- the variables are assumed to be uncorrelated, $\rho = 0$

$$\mu_{\mathbf{X}} = \begin{Bmatrix} 1,033 \\ 650 \end{Bmatrix}, \qquad \sigma_{\mathbf{X}} = \begin{Bmatrix} 103.30 \\ 130 \end{Bmatrix}, \qquad \mathbf{R} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$\mathbf{\Sigma_X} = \mathbf{DRD} = \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix} = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}$$

Limit State Function:

$$G(x_1, x_2) = x_1 - x_2$$

Cornell Reliability Index:

$$\beta = \frac{\mu_G}{\sigma_G} = \frac{\mu_1 - \mu_2}{\sqrt{\sigma_1^2 + \sigma_2^2}} = 2.31$$

Failure probability:

$$P_f = P[G \leq 0] = \Phi(-\beta) = 1.04 \times 10^{-2}$$

17.2.2 Geometric interpretation of Cornell Reliability index

For the basic reliability problem, $G = R - S$, let us develop a isoprobabilistic transformation, from the original space $\mathbf{x} \equiv \{x_1, x_2\} \equiv \{R, S\}$ to the normal standard space $\mathbf{u} \equiv \{u_1, u_2\}$

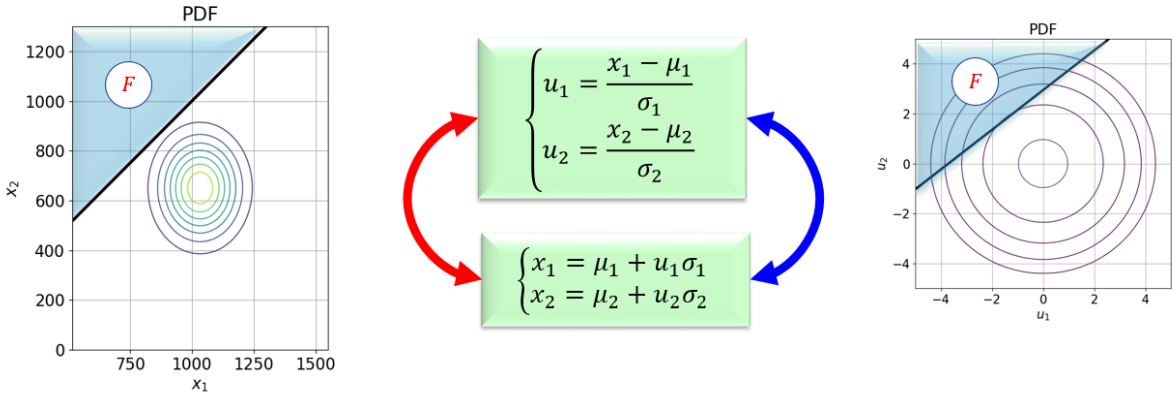


Figure 17.5: Isoprobabilistic transformation, bar example

Limit state function in the original space $\mathbf{x}$

$$G(x_1, x_2) = x_1 - x_2$$

Coordinate transformation

$$\begin{cases} u_1 = \frac{x_1 - u_1}{\sigma_1} \\ u_2 = \frac{x_2 - u_2}{\sigma_2} \end{cases} , \qquad \begin{cases} x_1 = u_1 + \mu_1\sigma_1 \\ x_2 = u_2 + \mu_2\sigma_2 \end{cases}$$

Limit state function in the normal standard space **u**

$$g(u_1, u_2) = \sigma_1 u_1 - \sigma_2 u_2 + (\mu_1 - \mu_2)$$

This is the equation of a plane in the space (u_1, u_2).

The *distance of the plane from the origin* is

$$\delta = \frac{\mu_1 - \mu_2}{\sqrt{\sigma_1^2 + \sigma_2^2}} \equiv \beta_C$$

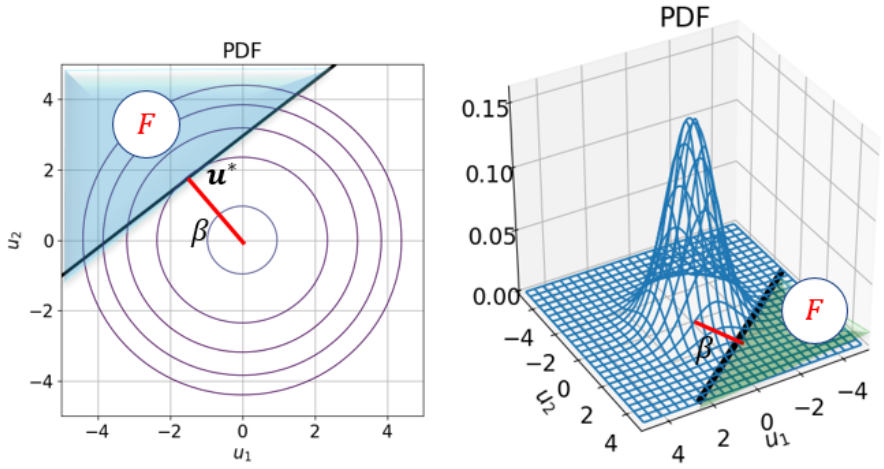


Figure 17.6: Basic reliability problem, limit state in normal standard space **u**

The design point $\mathbf{u}^*$ is *the point belonging to the limit state and closest to the origin of the standard normal space.*

The reliability index $\beta = \| \mathbf{u}^* \|$ is *the distance of the design point from the origin of the standard normal space.*

To see the relationship between this geometrical situation and the PDF $f_G(g)$ (see fig.17.2) below we represent the values of the Cornell reliability index for three different profiles

17.2.3 Linear limit state in the normal standard space u

Linear limit state in the normal standard space represent a *Gaussian Limit state function* $g(\mathbf{u})$.

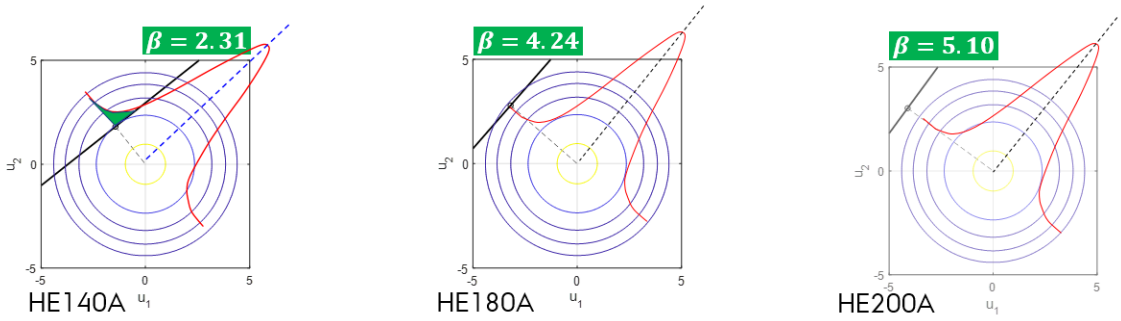


Figure 17.7: Cornell Reliability index, limit state in normal standard space **u**

Limit state function in the normal standard space **u**

$$g(\mathbf{u}) = b + (a_1u_1 + a_2u_2 + ... + a_nu_n) \tag{17.10}$$

Note that the coefficients a_i are the parameters of the hyperplane and they coincide with the gradient $\bigtriangledown g$

$$\begin{cases} a_1 = \frac{\partial g}{\partial u_1} \\ a_2 = \frac{\partial g}{\partial u_2} \\ ... \\ a_n = \frac{\partial g}{\partial u_n} \end{cases}$$

Reliability index β

$$\beta = \frac{b}{\|\mathbf{a}\|} \tag{17.11}$$

Design point $\mathbf{u}^*$

$$\mathbf{u}^* = \beta\alpha \tag{17.12}$$

where

$$\alpha = -\frac{\mathbf{a}}{\|\mathbf{a}\|} = -\frac{\bigtriangledown g}{\|\bigtriangledown g\|}$$

Bar example.

$$\mu = \begin{Bmatrix} 1,033 \\ 650 \end{Bmatrix}, \qquad \sigma = \begin{Bmatrix} 103.30 \\ 130 \end{Bmatrix},$$

Limit state function in the normal standard space **u**

$$g(u_1, u_2) = \sigma_1 u_1 - \sigma_2 u_2 + (\mu_1 - \mu_2)$$

Gradient **a**

$$\begin{cases} a_1 = \frac{\partial g}{\partial u_1} = \sigma_1 = 103.30 \\ a_2 = \frac{\partial g}{\partial u_2} = -\sigma_2 = -130 \end{cases} \quad \Rightarrow \quad \mathbf{a} = \begin{Bmatrix} 103.30 \\ -130 \end{Bmatrix}$$

Parameter α

$$\begin{cases} \alpha_1 = -\frac{a_1}{\|\mathbf{a}\|} = -\frac{\sigma_1}{\sqrt{\sigma_1^2+\sigma_2^2}} = -0.62 \\ \alpha_2 = -\frac{a_2}{\|\mathbf{a}\|} = -\frac{\sigma_2}{\sqrt{\sigma_1^2+\sigma_2^2}} = 0.78 \end{cases} \quad \Rightarrow \quad \alpha = \begin{Bmatrix} -0.62 \\ 0.78 \end{Bmatrix}$$

Reliability index β

$$\beta = \frac{b}{\|\mathbf{a}\|} = \frac{\mu_1 - \mu_2}{\sqrt{\sigma_1^2 + \sigma_2^2}} = 2.31 \tag{17.13}$$

Design point $\mathbf{u}^*$

$$\begin{cases} u_1^* = \beta\alpha_1 = -1.43 \\ u_2^* = \beta\alpha_2 = 1.80 \end{cases}$$

17.3 Hasofer-Lind Reliability index

The *Cornell Reliability index* can be applied when:

- all the basic random variables are Gaussian
- the Limit State Function is a their linear function

The Hasofer-Lind (HL) reliability index β is an extension of the Cornell reliability index β_C when the limit state is non-linear

- **Step 1**. Isoprobabilistic transformation toward the normal standard space **u**
- **Step 2**. Evaluation of the design point $\mathbf{u}^*$
- **Step 3**. Evaluation of the design point β

17.3.1 Isoprobabilistic transformation

The isoprobabilistic Transformation is a probability transformation from the *original space* $\mathbf{x}$ of the variables $X_1, X_2, ..., X_n$ toward the *normal standard space* $\mathbf{u}$ of $U_1, U_2, ..., U_n$

$$\mathbf{u} = \mathbf{T}(\mathbf{x}), \qquad \Rightarrow \qquad \mathbf{x} = \mathbf{T}^{-1}(\mathbf{u})$$

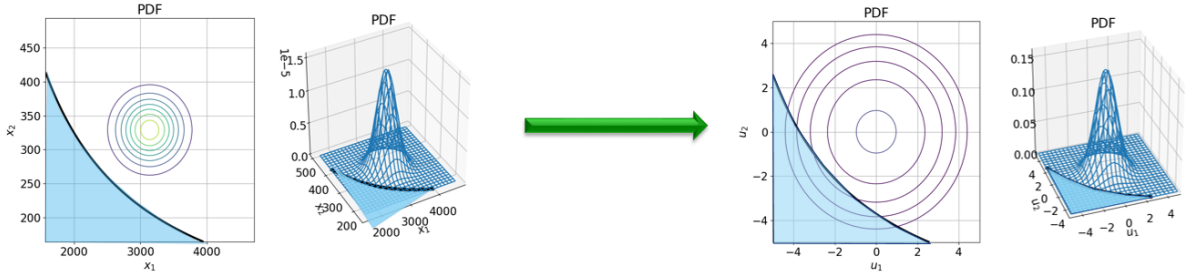


Figure 17.8: isoprobabilistic transformation

17.3.2 Design point

The Design point $\mathbf{u}^*$ is the point belonging to the limit state closest to the origin of the normal standard space

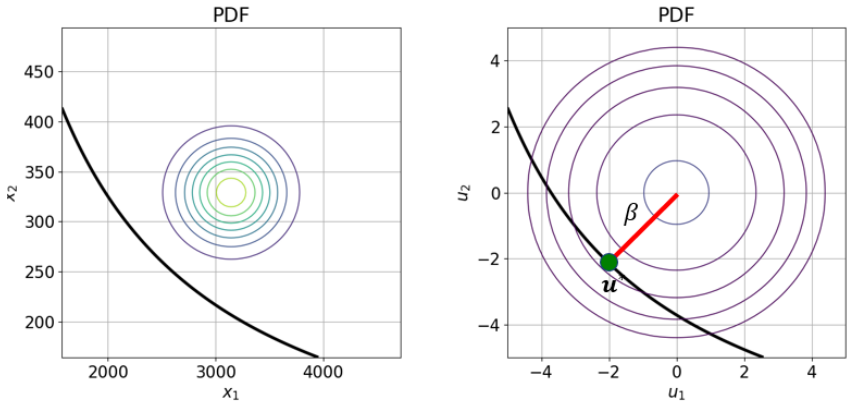


Figure 17.9: Design point

It is the point *belonging to the limit state with the maximum value of the multivariate normal standard distribution* and it represents the worst combination of the stochastic

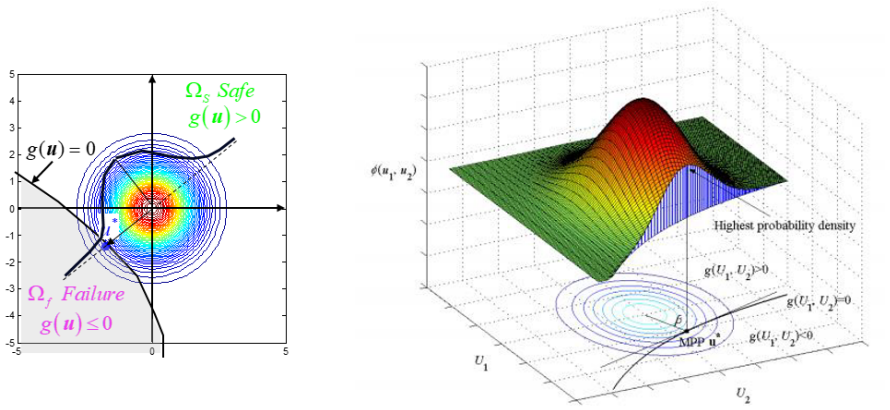


Figure 17.10: Design point, MPP

variables. For this reason sometimes (but not properly) it is also called the *Most Probable failure Point (MPP*

In view of this property, it is *the most likely realization of the random variables giving rise to failure*

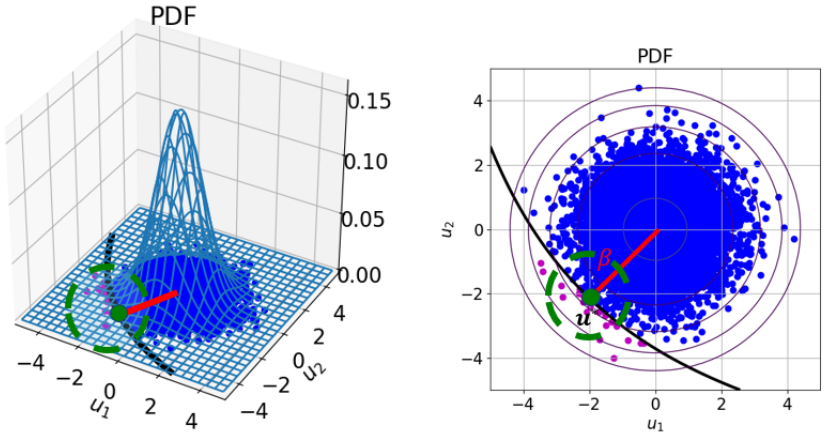


Figure 17.11: Design point, critical region

This means that most failure points *fall inside a narrow region around the design point*

17.3.3 Reliability Index

The reliability index $\beta = \|\mathbf{u}^*\|$ is the distance of the design point from the origin of the standard normal space

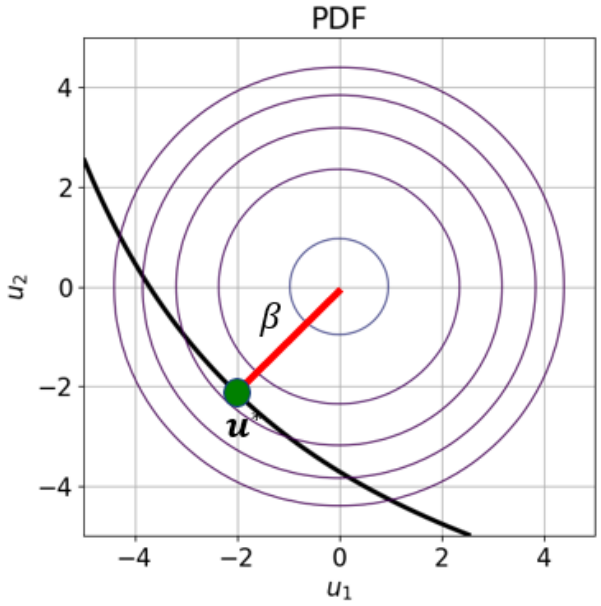


Figure 17.12: Reliability index

As discussed earlier, β is a measure of the reliability with respect to the chosen limit state

17.3.4 First Order Reliability Method (FORM)

The First Order Reliability Method (FORM) provides a linear approximation if the limit state around the design point It is obtained through a Taylor series expansion of the limit state around the design point

$$\begin{aligned} g(u_1, u_2) &= g(u_1^*, u_2^*) + \left.\frac{\partial g}{\partial u_1}\right|_{\mathbf{u}^*} (u_1 - u_1^*) + \left.\frac{\partial g}{\partial u_2}\right|_{\mathbf{u}^*} (u_2 - u_2^*) + \dots \\ &= a_1 (u_1 - u_1^*) + a_2 (u_2 - u_2^*) + \dots \\ &= b + \mathbf{a} \cdot \mathbf{u} \end{aligned}$$

and, by dividing for $\|\mathbf{a}\|$ becomes

$$g(\mathbf{u}) = \beta - \alpha \cdot \mathbf{u}$$

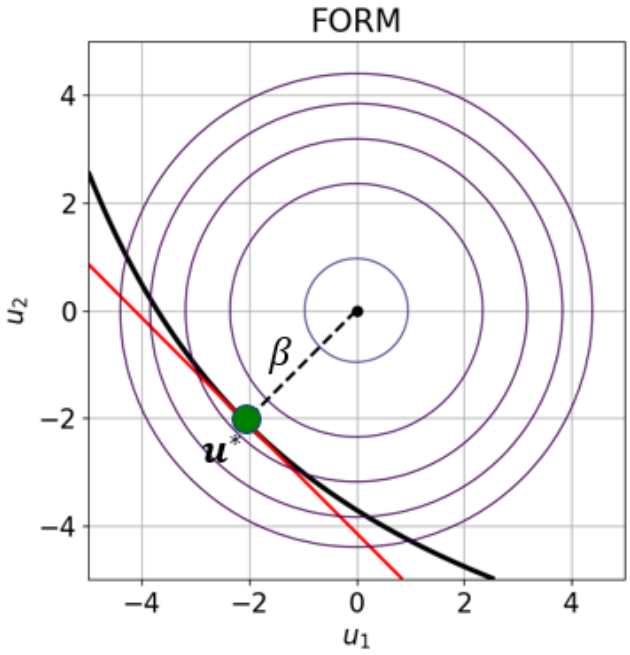


Figure 17.13: FORM

The failure probability provided by FORM is

$$P_{f,FORM} = \Phi(-\beta) \tag{17.14}$$

- in the most general case

  $$P_f \neq P_{f,FORM}$$

  and we cannot know in advance the degree of accuracy of the approximation

- when in the normal standard space the limit state is linear, then

  $$P_f \equiv P_{f,FORM}$$

17.3.5 Design point, Optimization

The design point is obtained as a solution of the following constrained optimization problem

$$\begin{cases} \beta_{HL} = & \min \beta(\mathbf{u}) = \|\mathbf{u}\| \\ s.t. & g(\mathbf{u}) = 0 \end{cases} \tag{17.15}$$

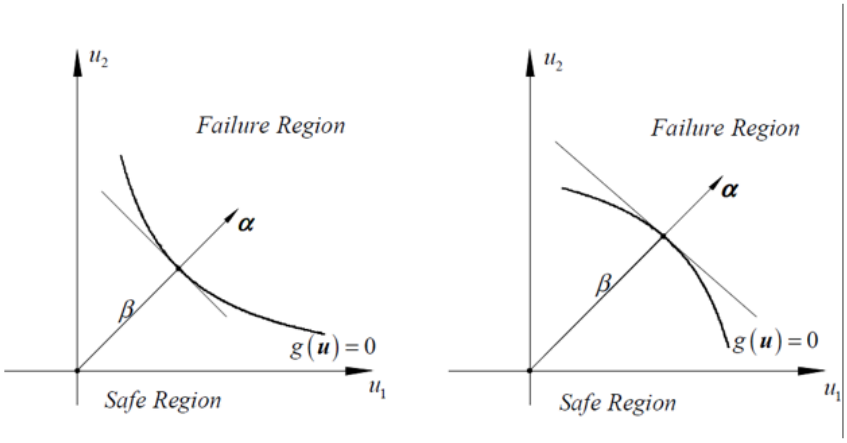


Figure 17.14: FORM error

Let us apply the Lagrangian

$$\min L(\mathbf{u}) = \min \beta(\mathbf{u}) + \lambda g(\mathbf{u})$$

Setting the derivatives

$$\begin{cases} \frac{\partial L(\mathbf{u})}{\partial u_i} = \frac{u_i}{\beta(\mathbf{u})} + \lambda \frac{\partial g(\mathbf{u})}{\partial u_i} \\ \frac{\partial L(\mathbf{u})}{\partial \lambda} = g(\mathbf{u}) = 0 \end{cases}$$

and the solution is

$$\mathbf{u}^* = -\lambda^* \beta(\mathbf{u}^*) \bigtriangledown g(\mathbf{u}^*) \tag{17.16}$$

It can be shown that

$$\lambda = \frac{\bigtriangledown g\left(\mathbf{u}^{(k)}\right) \cdot \mathbf{u}^{(k)} - g\left(\mathbf{u}^{(k)}\right)}{\left\| \bigtriangledown g\left(\mathbf{u}^{(k)}\right) \right\|^2} \tag{17.17}$$

In the algorithm of Hasofer-Lind an iterative procedure is adopted, where

$$\begin{aligned} \mathbf{u}^{(k+1)} &= \lambda^{(k)} \bigtriangledown g\left(\mathbf{u}^{(k)}\right) = \\ &= \bigtriangledown g\left(\mathbf{u}^{(k)}\right) \frac{\bigtriangledown g\left(\mathbf{u}^{(k)}\right) \cdot \mathbf{u}^{(k)} - g\left(\mathbf{u}^{(k)}\right)}{\left\| \bigtriangledown g\left(\mathbf{u}^{(k)}\right) \right\|^2} \end{aligned} \tag{17.18}$$

Hasofer-Lind, normal standard space

1. Define the limit state function $G(\mathbf{x})$

2. develop isoprobabilistic transformation
3. evaluate $g(\mathbf{u})$
4. $k = 0$
5. choose a first value $\mathbf{u}_0 = \mathbf{u}^{(0)}$
6. Evaluate $g\left(\mathbf{u}^{(0)}\right)$ and $\bigtriangledown g\left(\mathbf{u}^{(0)}\right)$
7. evaluate $\mathbf{u}^{(1)}$ using eq.(17.18)
8. evaluate $\beta^{(1)} = \left\|\mathbf{u}^{(1)}\right\|$
9. $k = k + 1$
10. Evaluate $g\left(\mathbf{u}^{(k)}\right)$ and $\bigtriangledown g\left(\mathbf{u}^{(k)}\right)$
11. evaluate $\mathbf{u}^{(k+1)}$ using eq.(17.18)
12. evaluate $\beta^{(k+1)} = \left\|\mathbf{u}^{(k+1)}\right\|$
13. Repeat 9-12 until convergence on β
14. Evaluate $\mathbf{u}^*$
15. Evaluate $x_i^* = \mu_i + \sigma_i u_i^*$

Step0:

- isoprobabilistic transformation

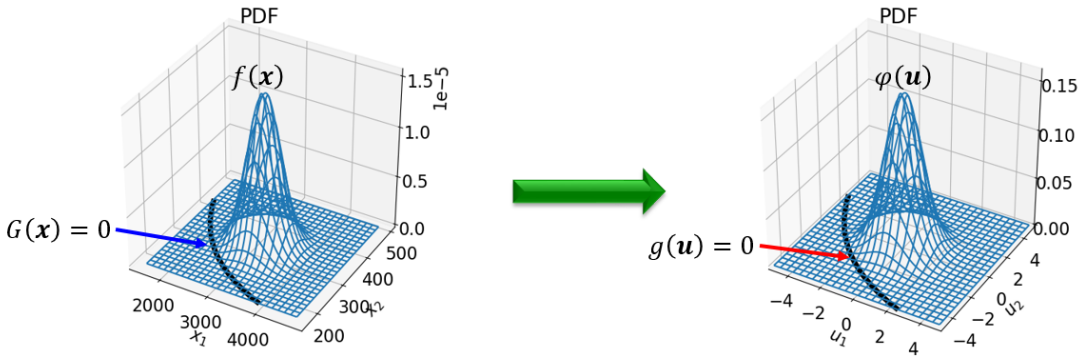


Figure 17.15: HL optimization, isoprobabilistic transformation

Step1:

- Choose a point $\mathbf{u}^{(0)}$
- find $\beta^{(0)} = \left\|\mathbf{u}^{(0)}\right\|$
- evaluate the plan of equation $\beta^{(0)} - \alpha^{(0)} \cdot \mathbf{u} = 0$

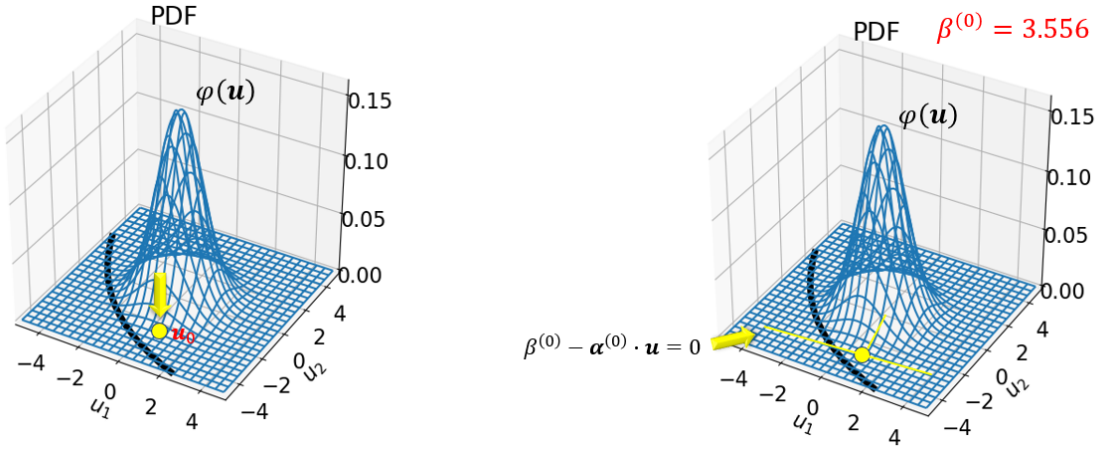


Figure 17.16: HL optimization, point $\mathbf{u}^{(0)}$

Step2:

- Evaluate

$$\mathbf{u}^{(1)} = \bigtriangledown g\left(\mathbf{u}^{(0)}\right) \frac{\bigtriangledown g\left(\mathbf{u}^{(0)}\right) \cdot \mathbf{u}^{(0)} - g\left(\mathbf{u}^{(0)}\right)}{\left\|\bigtriangledown g\left(\mathbf{u}^{(0)}\right)\right\|^2}$$

- find $\beta^{(1)} = \left\|\mathbf{u}^{(1)}\right\|$
- evaluate the plan of equation $\beta^{(1)} - \alpha^{(1)} \cdot \mathbf{u} = 0$
- evaluate $\Delta\beta = |\beta^{(1)} - \beta^{(0)}|$

Step3:

- Evaluate

$$\mathbf{u}^{(2)} = \bigtriangledown g\left(\mathbf{u}^{(1)}\right) \frac{\bigtriangledown g\left(\mathbf{u}^{(1)}\right) \cdot \mathbf{u}^{(1)} - g\left(\mathbf{u}^{(1)}\right)}{\left\|\bigtriangledown g\left(\mathbf{u}^{(1)}\right)\right\|^2}$$

- find $\beta^{(2)} = \left\|\mathbf{u}^{(2)}\right\|$
- evaluate the plan of equation $\beta^{(2)} - \alpha^{(2)} \cdot \mathbf{u} = 0$
- evaluate $\Delta\beta = |\beta^{(2)} - \beta^{(1)}|$

and so on until convergence

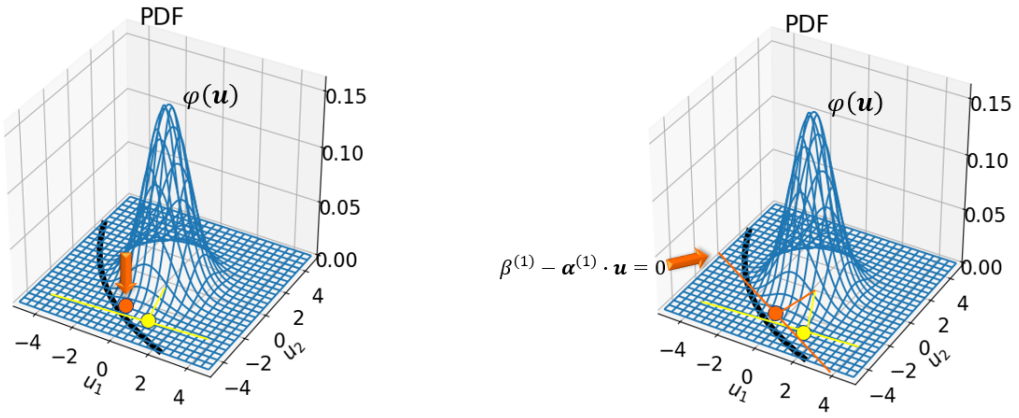


Figure 17.17: HL optimization, point $\mathbf{u}^{(1)}$

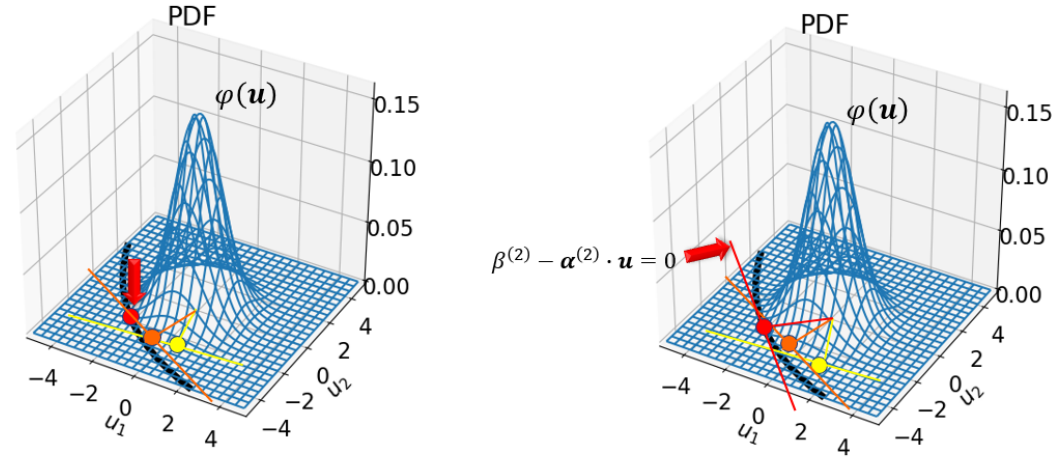


Figure 17.18: HL optimization, point $\mathbf{u}^{(2)}$

Chapter 18

Lecture B3:First Order Reliability Method (FORM)

18.1 Uncorrelated Gaussian

18.2 Correlated Gaussian

18.3 Uncorrelated non-Gaussian

18.4 Conditional methods

18.5 Nataf

18.6 Copula

Chapter 19

Basics of Surrogate Modelling and Machine Learning for structural reliability

19.1 Second-Order Reliability Method

19.2 Response Surface Method

19.3 High-Dimensional Model Representation (HDMR)

19.4 Polynomial Chaos Expansion (PCE)

19.5 Support Vector Machines (SVM)

19.6 Gaussian Process and Kriging (GP)

19.7 Geometry of high-dimensional spaces

19.8 Secant Hyperplane Method

Chapter 20

Lecture B4:System Reliability

20.1 series system

20.2 parallel system

20.3 general system

20.4 Reliability evaluation of series systems

20.4.1 MCS

20.4.2 FORM approximation

20.4.3 Simple Bounds

20.4.4 Ditlevsen Bounds

20.5 Reliability evaluation of parallel systems

20.5.1 MCS

20.5.2 FORM approximation

20.5.3 Simple Bounds

20.5.4 Ditlevsen Bounds

20.6 Event tree and fault tree

20.7 Bayesian Networks

Chapter 21

Advanced Monte Carlo Simulation

21.1 Latin Hypercube Sampling (LHS)

21.2 Importance Sampling

21.3 Subset Simulation

Chapter 22

Machine Learning

22.1 Introduction

22.2 Classification

22.3 Regression

22.4 Support Vector Machines

22.5 Decision Trees

22.6 Ensemble Learning and Random Forests

22.7 Dimensionality Reduction

22.8 Unsupervised Machine Learning

Chapter 23

Neural Networks and Deep Learning

23.1 Introduction to Artificial Neural Networks

23.2 Training Deep Neural Networks

23.3 Computer Vision and Convolutional Neural Networks (CNN)

23.4 Recurrent Neural Networks (RNN)

23.5 Representation Learning and Generative Learning

Chapter 24

Lecture C2:Basics of Decision Theory

24.1 Decision Theory

24.2 Decision with given information - prior analysis

24.3 Decision with additional information - posterior analysis

24.4 Decision with unknown information - pre-posterior analysis

24.5 Value of Information

24.6 Bayesian Networks

Chapter 25

Reinforcement Learning