CV 510$^{0}_{1}$
Modeling, Uncertainty, and Data for Engineers
(July – Nov 2025)

Dr. Prakash S Badal

1

Flow

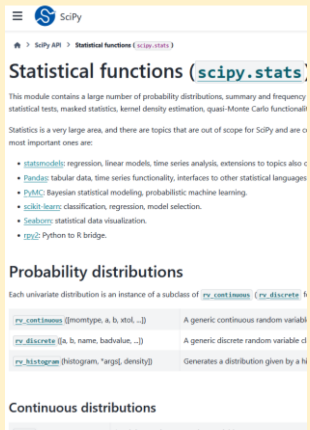
- Revise
  - Variance, covariance, correlation
  - Gaussian
  - reading z-table
  - Expectation
- Parametric distributions
  - Gaussian, Uniform, exponential, lognormal, Gumbel
- Propagation laws

2

2

Requested practice?

- Practice `scipy.stats` for the following:
  - Using distributions → Shifting, scaling
  - Generating random numbers
  - Fitting distributions



SciPy

SciPy API > Statistical functions (scipy.stats)

Statistical functions (scipy.stats)

This module contains a large number of probability distributions, summary and frequency
statistical tests, masked statistics, kernel density estimation, quasi-Monte Carlo functionalit

Statistics is a very large area, and there are topics that are out of scope for SciPy and are c
most important ones are:

- statsmodels: regression, linear models, time series analysis, extensions to topics also c
- Pandas: tabular data, time series functionality, interfaces to other statistical languages
- PyMC: Bayesian statistical modeling, probabilistic machine learning.
- scikit-learn: classification, regression, model selection.
- Seaborn: statistical data visualization.
- rpy2: Python to R bridge.

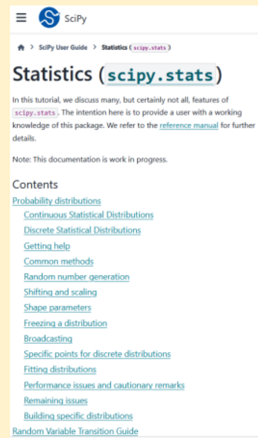
Probability distributions

Each univariate distribution is an instance of a subclass of rv_continuous (rv_discrete f

rv_continuous ([momtype, a, b, xtol, ...])	A generic continuous random variabl
rv_discrete ([a, b, name, badvalue, ...])	A generic discrete random variable cl
rv_histogram (histogram, *args[, density])	Generates a distribution given by a hi

Continuous distributions

alpha	An alpha continuous random variable.



SciPy

SciPy User Guide > Statistics (scipy.stats)

Statistics (scipy.stats)

In this tutorial, we discuss many, but certainly not all, features of scipy.stats. The intention here is to provide a user with a working knowledge of this package. We refer to the reference manual for further details.

Note: This documentation is work in progress.

Contents

3

3

Revise

- Two structural engineers, Alice and Bob, independently measure the wind speed in m/s at the top of each tower.

Alice's reading $X \sim \mathcal{N}(\mu = 40, \sigma = 5)$

Bob's reading $Y \sim \mathcal{N}(\mu = 42, \sigma = 6)$

Their readings are correlated with $\rho = 0.8$

1. What's the probability that Alice's reading exceeds 50?
2. What's the covariance of X and Y?
3. If Alice and Bob average their measurements, $W = (X + Y)/2$, what are $\mathrm{E}[W]$ and $\mathrm{Var}[W]$?
4. If design wind speed is 55 m/s, what's the prob. that the average exceeds 55 m/s?

4

Revise

5

5

scipy

- [Ref.] https://docs.scipy.org/doc/scipy/reference/stats.html
- Tutorial https://docs.scipy.org/doc/scipy/tutorial/stats.html
- Probability distributions https://docs.scipy.org/doc/scipy/tutorial/stats/probability_distributions.html

- Run examples

  - Common Methods
  - Random number generator
  - Shifting-scaling: loc, scale
  - Fitting distributions

6

6

Check-in with teach-book

https://mude.citg.tudelft.nl/book/2024/

- Q1 Topics (Chapters 5, 6, 2, and 3)
  - Q1C5 Univariate continuous distribution
    - Q1C5.1 PDF/CDF
    - Q1C5.2 Empirical Distributions
    - Q1C5.3 Parametric Distributions
    - Q1C5.4 Fitting a Distribution
  - Q1C6 Multivariate Distributions (briefly)
  - Q1C2 Propagation of Uncertainty
  - Q1C3 Observation Theory: least-sq., Hyp. Test, Conf. Intervals
- Q2 Topics (Chapter 7 and 8)
  - Q2C7 Extreme value theory: GEV, return period, POT
  - Q2C8 Risk and decision making (CBA)
- Fundamental Concepts
  - Chapter 6, 7, 8, 9. Probability basics, rv, z- and t-tables
- Programming
  - Fundamental Concepts → Chapter 10

7

7

Module Overview

B.1

Uncertainty & Estimation

B.1.1

Probabilistic basics and RVs
- Probability laws
- Random variables
- Discrete, continuous RVs
- PMF/PDF/CDF
- Expectation
- Variance/cov./corr.

Q1C5, Funda6-7-8

B.1.2

Core distributions
- Gaussian
- Uniform
- Exponential
- Lognormal
- Symmetric/skewness
- *Tails*: thick/light

Q1C5, Funda9-10

B.1.3

Propagation laws & least-sq.
- Uncertainty propagation
- Least-square errors
- Parameter estimation

Q1C2-C3, Funda9-10

8

Gaussian

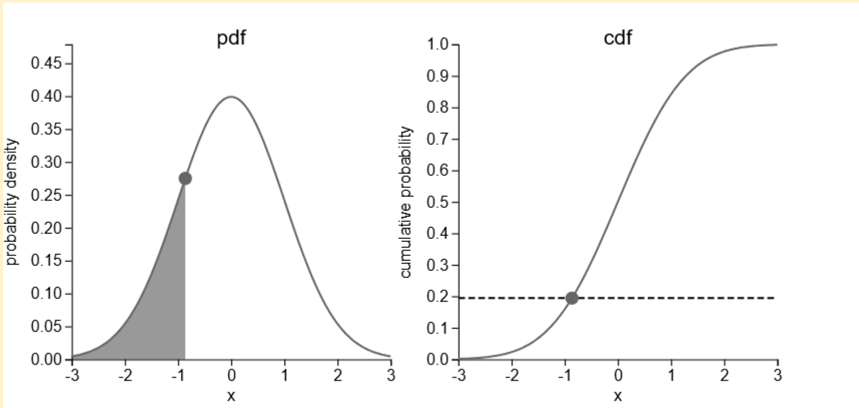


Fig. 5.1 Interactively visualize the relationship between the PDF and the CDF (the integral of the PDF value). #

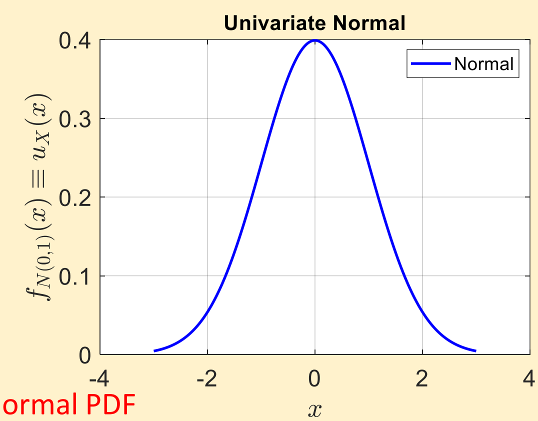
Teach-book Q1-C5 **Fig. 5.1** (interactive)

9

9

Gaussian

- Gaussian



Univariate normal PDF

$$N(\mu,\sigma) = f_X(x) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]; \qquad u(x) \equiv \varphi(x) = \frac{1}{\sqrt{2\pi}}\exp\left[-\frac{x^2}{2}\right]$$

No closed-form equation for Normal CDF

10

10

Gaussian

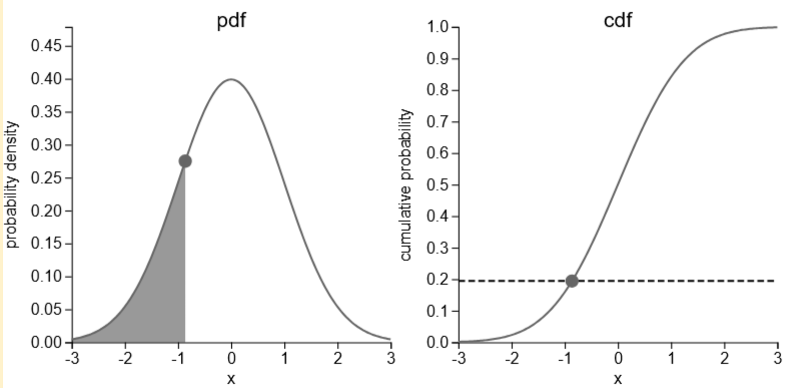


Fig. 5.1 Interactively visualize the relationship between the PDF and the CDF (the integral of the PDF value). #

Teach-book Q1-C5 **Fig. 5.1** (interactive)

11

11

Standard Normal Probabilities

Z-table

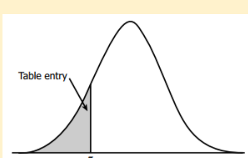


Table entry for z is the area under the standard normal curve to the left of z.

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
−3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
−3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
−3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
−3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
−2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
−2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
−2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
−2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
−2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
−2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
−2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
−2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
−2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
−2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
−1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
−1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
−1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
−1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
−1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
−1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
−1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
−1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
−1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
−1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
−0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
−0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
−0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
−0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
−0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
−0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
−0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
−0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
−0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
−0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641

12

Gaussian

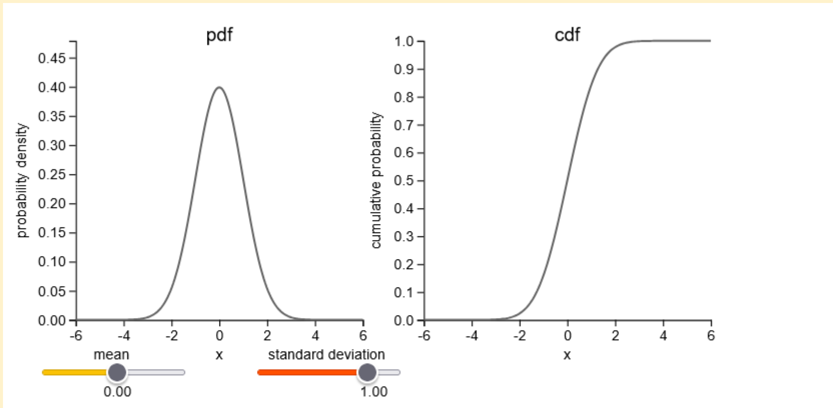


Fig. 5.3 Interactively change the mean and standard deviation of the Gaussian distribution to visualize the effect on the PDF and CDF.

Scale and location parameters
Teach-book Q1-C5 **Fig. 5.3** (interactive)

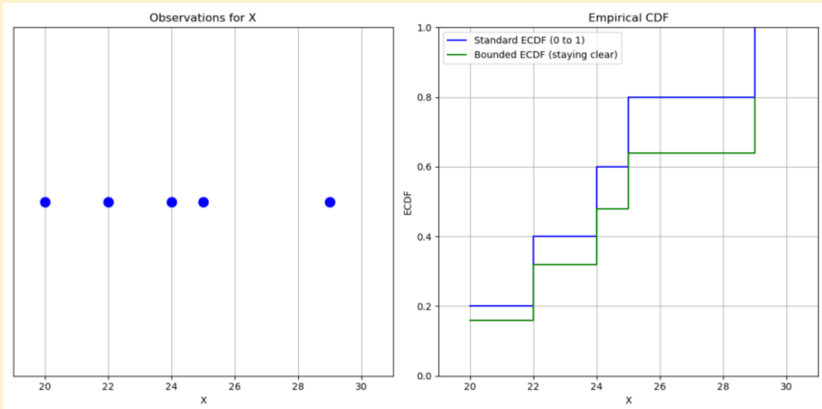
Inverse CDF

- For designing a structure, we often need

a value that is not exceeded with more than p probability:

$$x = F^{-1}(p)$$

Empirical distribution



Standard ECDF: $F_n(x) = \frac{i}{n}$ goes from 0 to 1

Bounded ECDF: $F_{n,b}(x) = \frac{i}{n+1}$ stays clear of 0 and 1

useful in Q-Q/probability plotting, avoids $-\infty$ or $+\infty$.

15

15

Uniform Distribution

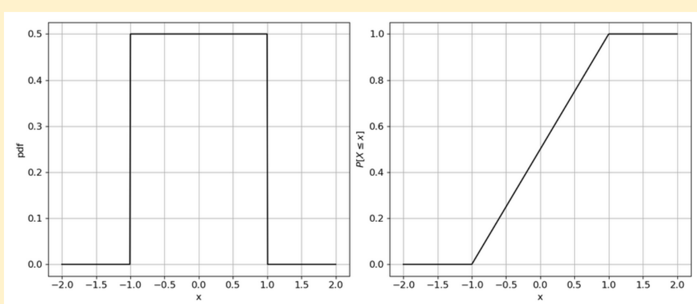


Fig. 5.11 Uniform distribution function: (left) PDF, and (right) CDF.

PDF

$$f_X(x) = \begin{cases} \frac{1}{b-a}, & a \le x \le b \\ 0 & \text{otherwise} \end{cases}$$

CDF

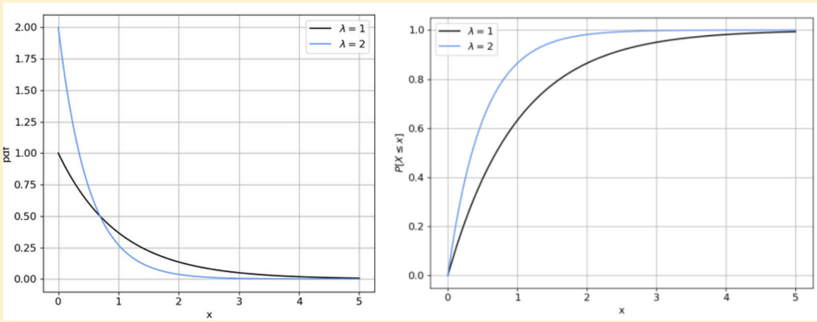
$$F_X(x) = \begin{cases} 0 & \text{for } x < a \\ \frac{x-a}{b-a} & \text{for } a \le x \le b \\ 1 & \text{for } x > b \end{cases}$$

Properties $\mathrm{E}[X] = \frac{1}{2}(a+b); \quad \mathrm{Var}[X] = \frac{1}{12}(b-a)^2$

16

16

Exponential Distribution



PDF

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0, \lambda > 0 \\ 0 & \text{otherwise} \end{cases}$$

CDF

$$F_X(x) = 1 - e^{\lambda x}, \qquad x > 0$$

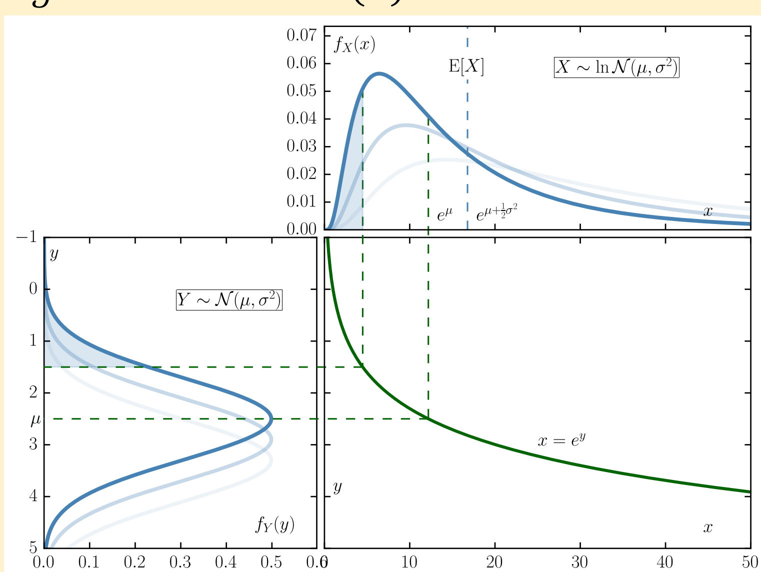
Properties $\mathrm{E}[X] = \dfrac{1}{\lambda}; \qquad \mathrm{Var}[X] = \dfrac{1}{\lambda^2}$

17

17

Lognormal

- $X \sim Lognormal \Leftrightarrow Y = \ln(X) \sim Normal$



18

18

lognormal distribution

- Notation: $\mathcal{LN}(\lambda, \zeta)$

$f_Y(y) =$

$$\frac{1}{\sqrt{2\pi}\zeta y}\exp\left[-\frac{1}{2}\left(\frac{\ln y - \lambda}{\zeta}\right)^2\right], \qquad 0 < y$$

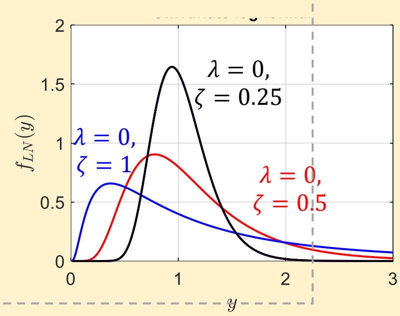
$\mu_X = \mu_{\ln Y} = \lambda$ and $\sigma_X = \sigma_{\ln Y} = \zeta$

$\mu_Y = \exp\left(\lambda + \frac{1}{2}\zeta^2\right)$ and c.o.v. $\delta_Y = \sqrt{\exp(\zeta^2) - 1}$

The inverse relations are

$\lambda = \ln \mu_Y - \frac{1}{2}\zeta^2$ and

$\zeta = \sqrt{\ln(1 + \delta_Y^2)}$



19

Gumbel distributions

- When we are interested in

the smallest

or

the largest of a set of rv's,

e.g., a chain of links: smallest strength.

Flood level under a bridge: highest flood level during its lifetime.

$Y_1 = \quad \min(X_1, X_2, \ldots, X_n)$,

$Y_n = \quad \max(X_1, X_2, \ldots, X_n)$.

20

20

Gumbel distributions

$$Y_1 = \min(X_1, X_2, \dots, X_n)\,,$$
$$Y_n = \max(X_1, X_2, \dots, X_n)\,.$$

The CDF of Y_1 is (smallest extreme):

$F_{Y_1}(y) =$ $\Pr(Y_1 \leq y) = 1 - \Pr(Y_1 > y) = 1 - \prod_{i=1}^{n} \Pr(X_i > y)$

CDF $F_{Y_1}(y) = 1 - [1 - F_X(y)]^n$

PDF $f_{Y_1}(y) = 1 - n f_X(y)[1 - F_X(y)]^{n-1}$

21

21

Gumbel distributions

$$Y_1 = \min(X_1, X_2, \dots, X_n)\,,$$
$$Y_n = \max(X_1, X_2, \dots, X_n)\,.$$

The CDF of Y_n is (largest extreme):

$F_{Y_n}(y) =$ $\Pr(Y_n \leq y) = \prod_{i=1}^{n} \Pr(X_i \leq y)$

CDF $F_{Y_n}(y) = [F_X(y)]^n$

PDF $f_{Y_n}(y) = n f_X(y)[F_X(y)]^{n-1}$

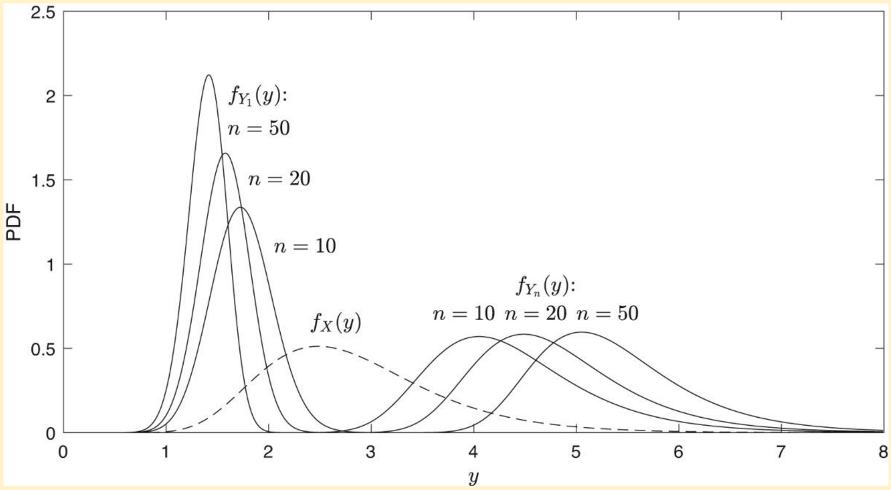
When distributions are known -> above are exact

Often asymptotic distribution is needed -> Hence, GEV

22

22

Gumbel distributions



23

23

Gumbel distributions

Positive tail

PDF

$$f_{Y_n}(y) = \frac{1}{\beta} e^{-\left(\frac{y-\mu}{\beta} + e^{-\left(\frac{y-\mu}{\beta}\right)}\right)}$$

CDF

$$F_{Y_n}(y) = e^{-e^{-\left(\frac{y-\mu}{\beta}\right)}}$$

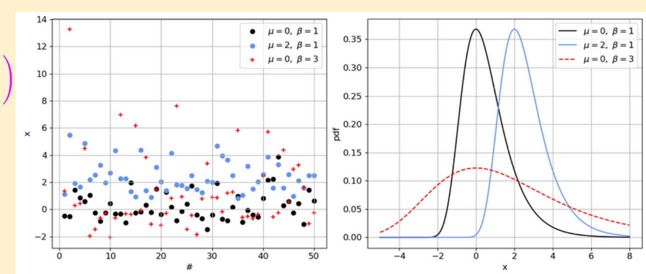


Fig. 5.14 Gumbel distribution function: (left) random samples, and (right) PDF.

Negative tail

PDF

$$f_{Y_1}(y) = \frac{1}{\beta} e^{\left(\frac{y-\mu}{\beta} - e^{\left(\frac{y-\mu}{\beta}\right)}\right)}$$

CDF

$$F_{Y_1}(y) = e^{-e^{\left(\frac{y-\mu}{\beta}\right)}}$$

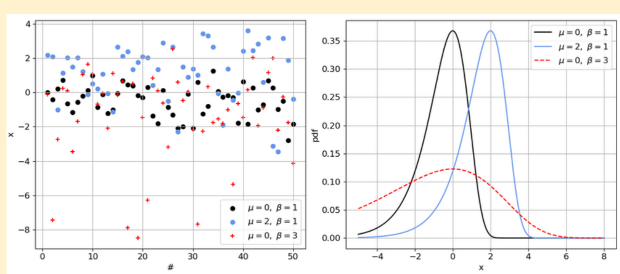
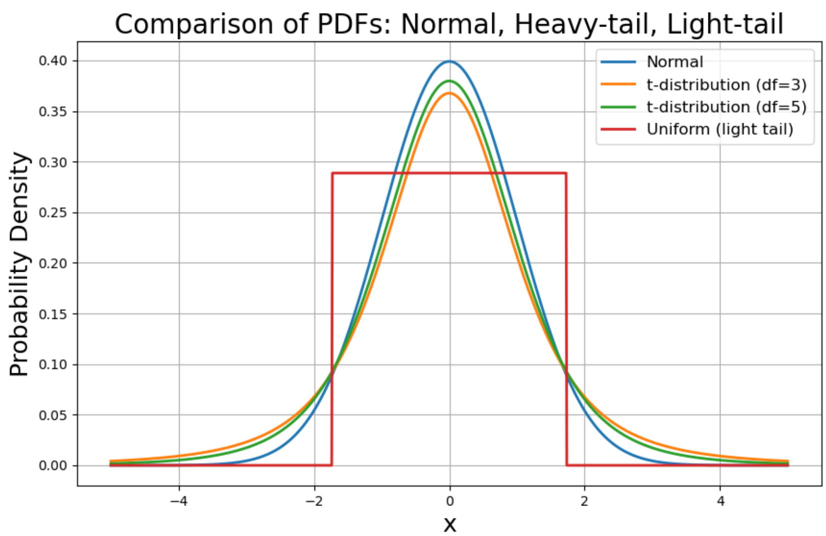


Fig. 5.16 Left-tailed Gumbel distribution function: (left) random samples, and (right) PDF.

24

Core distributions

- *Tails*: thick/light



25

B.1.3 Propagation laws & least-squares (Q1C2)

26

Uncertainty classification

27

Sources of Uncertainty

- Inherent uncertainty: concrete strength; irreducible
- Statistical uncertainty: due to the lack of data; reducible
- Characteristics of system: material properties, dimensions
- Demand placed on the system
- Mathematical models used to analyze structure's behavior
- Measurements uncertainty: while assessing structure's health
- Probabilistic models used to describe uncertain quantities
- Human error

28

28

Categories of Uncertainty

In classical statistics,

- Aleatory
  - Irreducible and inherent variability, e.g., tossing a fair coin
  - Treated using probability distributions
  - Covered today
- Epistemic
  - Reducible uncertainty; due to lack of knowledge/data/better model, e.g., earthquake on an unmapped fault, e.g., Latur 1993
  - Treated using confidence interval
  - Covered later in the course

29

29

Categories of Uncertainty

- Wind speed on the surface of a building:
  - Aleatory or epistemic?
- *Case 1: When there is no predictive wind model and only long-term statistical data is used?*
- *Case 2: When wind fields are studied using precise computational fluid dynamics?*

Transition from Aleatory to epistemic

In MUDE, prob. dist. Models to capture all uncertainty.

Above classification can be useful for interpreting results

30

30

Propagation Laws

31

Mean and Variance propagation laws

- A random variable, $\mathrm{X} \sim \mathcal{N}(40,5)$

If we define, another random variable, $Y = 3X + 10$, what are its mean and standard deviation (stdev) and type.

Normal
$\mu_Y = 3\mu_X + 10 = 130$
$\sigma_Y^2 = 9\sigma_X^2 = 225 \Rightarrow \sigma_Y = 15$

Recall:

Black (observed)	→	Red (centered; loc)	→	Blue (standardized; *scale*)

32

Mean and Variance propagation laws

- More generally, if we have $X_1, X_2, \dots, X_n$ with means $\mu_1, \mu_2, \dots, \mu_n$ and stdevs $\sigma_1, \sigma_2, \dots, \sigma_n$.
- For a $Y = q(\mathbf{X}) = q(X_1, X_2, \dots, X_n)$, what are its expected value and variance?

If $Y = a_1X_1 + a_2X_2 + c$, with a_i and c deterministic const.

$\mathrm{E}[Y] = \mathrm{E}(a_1X_1 + a_2X_2 + c) = a_1\mathrm{E}(X_1) + a_2\mathrm{E}(X_2) + c$

$= a_1\mu_1 + a_2\mu_2 + c$

$\mathrm{Var}[Y] = \mathrm{E}[(Y - \mu_Y)^2]$

$= \mathrm{E}[\{(a_1X_1 + a_2X_2 + c) - (a_1\mu_1 + a_2\mu_2 + c\ )\}^2]$

$= \mathrm{E}[\{(a_1X_1 - a_1\mu_1) + (a_2X_2 - a_2\mu_2)\}^2]$

Mean and Variance propagation laws

If $Y = a_1X_1 + a_2X_2 + c$, with a_i and c deterministic const.

$\mathrm{E}[Y] = a_1\mu_1 + a_2\mu_2 + c$

$\mathrm{Var}[Y] = \quad \mathrm{E}[(Y - \mu_Y)^2]$

$= \mathrm{E}[\{(a_1X_1 + a_2X_2 + c) - (a_1\mu_1 + a_2\mu_2 + c\ )\}^2]$

$= \mathrm{E}[\{(a_1X_1 - a_1\mu_1) + (a_2X_2 - a_2\mu_2)\}^2]$

$= \mathrm{E}[(a_1X_1 - a_1\mu_1)^2 + (a_2X_2 - a_2\mu_2)^2 + 2(a_1X_1 - a_1\mu_1)(a_2X_2 - a_2\mu_2)]$

$= a_1^2\mathrm{E}[(X_1 - \mu_1)^2] + a_2^2\mathrm{E}[(X_2 - \mu_2)^2] + 2a_1a_2E[2(X_1 - \mu_1)(X_2 - \mu_2)]$

$= a_1^2\sigma_1^2 + a_2^2\sigma_2^2 + 2a_1a_2\mathrm{Cov}[X_1, X_2]$

Propagation: μ and σ

- Two structural engineers, Alice and Bob, independently measure the wind speed in m/s at the top of each tower.

Alice's reading $X \sim \mathcal{N}(\mu = 40, \sigma = 5)$

Bob's reading $Y \sim \mathcal{N}(\mu = 42, \sigma = 6)$

Their readings are correlated with $\rho = 0.8$

1. What's the probability that Alice's reading exceeds 50?
2. What's the covariance of X and Y?
3. If Alice and Bob average their measurements, $W = (X + Y)/2$, what are $\mathrm{E}[W]$ and $\mathrm{Var}[W]$?
4. If design wind speed is 55 m/s, what's the prob. that the average exceeds 55 m/s?


©wikimedia

35

Propagation: mean and uncertainty

Alice's reading $X \sim \mathcal{N}(\mu = 40, \sigma = 5)$

Bob's reading $Y \sim \mathcal{N}(\mu = 42, \sigma = 6)$

Their readings are correlated with $\rho = 0.8$

$W = (X + Y)/2,$

$\mathrm{E}[W] = 41$

$\mathrm{Var}[W] = \frac{1}{4}(\sigma_X^2 + \sigma_Y^2 + 2\rho_{XY}\sigma_X\sigma_Y) = 27.25$

$\sigma_W = 5.22$

$$\Pr(W > 55) = \Phi\left(\frac{41 - 55}{5.22}\right) = \Phi(-2.68) = 0.0037$$

1. If design wind speed is 55 m/s, what's the prob. that the average exceeds 55 m/s?

36

36

Standard Normal Probabilities

Z-table

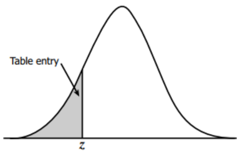


Table entry for z is the area under the standard normal curve to the left of z.

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
−3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
−3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
−3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
−3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
−2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
−2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
−2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
−2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
−2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
−2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
−2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
−2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
−2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
−2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
−1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
−1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
−1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
−1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
−1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
−1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
−1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
−1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
−1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
−1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
−0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
−0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
−0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
−0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
−0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
−0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
−0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
−0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
−0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
−0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641

37

Mean and Variance propagation laws

If $Y = a_1X_1 + a_2X_2 + \cdots + a_nX_n + c$, with a_i & c det. consts.

$\mathrm{E}[Y] = a_1X_1 + a_2X_2 + \cdots + a_nX_n + c$

$\mathrm{Var}[Y] = \quad \mathrm{E}[(Y - \mu_Y)^2]$

$= \mathrm{E}[\{(a_1X_1 - a_1\mu_1) + \cdots + (a_nX_n - a_n\mu_n)\}^2]$

$= \sum_{i=1}^{n} a_i^2\sigma_i^2 + 2\sum_{1\le i<j<n} a_ia_j\mathrm{Cov}[X_i, X_j]$

38

38

Mean and Variance propagation laws

- If $Y = g(X)$ is a nonlinear function of X. Find $\mathrm{E}[Y]$ & $\mathrm{Var}(Y)$.

$\mathrm{E}[Y] = \mathrm{E}[g(X)]$

Taylor series expansion:

$$g(X) = g(\mu_X) + \left(\frac{\partial g}{\partial x}\right)_{\mu_X} (X - \mu_X) + \frac{1}{2!}\left(\frac{\partial^2 g}{\partial x^2}\right)_{\mu_X} (X - \mu_X)^2 + \mathrm{H.O.T.}$$

$$\mathrm{E}[Y] \cong \mathrm{E}\left(g(\mu_X) + \left(\frac{\partial g}{\partial x}\right)_{\mu_X} (X - \mu_X) + \frac{1}{2!}\left(\frac{\partial^2 g}{\partial x^2}\right)_{\mu_X} (X - \mu_X)^2\right)$$

$$= g(\mu_X) + 0 + \frac{1}{2!}\left(\frac{\partial^2 g}{\partial x^2}\right)_{\mu_X} \mathrm{E}[(X - \mu_X)^2]$$

$\mathrm{E}[Y] \cong g(\mu_X)$ First-order mean approximation

$\mathrm{E}[Y] \cong g(\mu_X) + \frac{1}{2}\left(\frac{\partial^2 g}{\partial x^2}\right)_{\mu_X} \sigma_X^2$ Second-order mean approximation

39

39

Mean and Variance propagation laws

- If $Y = g(X)$ is a nonlinear function of X. Find $\mathrm{E}[Y]$ & $\mathrm{Var}(Y)$.

Taylor series expansion:

$$g(X) = g(\mu_X) + \left(\frac{\partial g}{\partial x}\right)_{\mu_X} (X - \mu_X) + \frac{1}{2!}\left(\frac{\partial^2 g}{\partial x^2}\right)_{\mu_X} (X - \mu_X)^2 + \mathrm{H.O.T.}$$

$\mathrm{Var}[Y] = \mathrm{E}[(Y - \mu_Y)^2] \cong \left(\left(\frac{\partial g}{\partial x}\right)_{\mu_X}\right)^2 \sigma_X^2$ First-order var. approx.

40

40

Example

- If $Y = \ln X$, what are expectation and variance of Y? Use Taylor series approximations.

$\mathrm{E}[Y] \cong \ln \mu_X$ first order mean approximation

$\mathrm{E}[Y] \cong \ln \mu_X - \frac{\delta_X^2}{2}$ second order mean approximation

$\mathrm{Var}[Y] \cong \delta_X^2$ first order variance approximation

$\mathrm{E}[Y] \cong g(\mu_X) + \frac{1}{2}\left(\frac{\partial^2 g}{\partial x^2}\right)_{\mu_X} \sigma_X^2$ Second-order mean approximation

$\mathrm{Var}[Y] = \mathrm{E}[(Y - \mu_Y)^2] \cong \left(\left(\frac{\partial g}{\partial x}\right)_{\mu_X}\right)^2 \sigma_X^2$ First-order var. approx.

41

41

Questions, comments, or concerns?

42

42