Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

ETH RISK CENTER

LECTURE NOTES

Fundamentals of Probability and Statistics
Chapter 1: Lecture 2

Spring 2018

Written by marco **br**occardo
IBK, ETHZ

3 Random Variables

A random variable is a *numerical representation* of a random phenomenon. In particular, the random variable is defined as a one-to-one mapping of its sample space on the real line, Figure 5. Random variables are convenient in many engineering applications since outcomes of engineering problems are often given in numerical values (e.g. magnitude of an earthquake, peak-ground acceleration, yielding point, etc.). Random variables can also have a pure conventional mapping. For example, the sample space of the state of a building after a seismic event can be defined as follow $S \in \{D_0, D_1, D_2, D_3, C\}$, where D_0 is the no damage state, D_1 is light damage state, D_2 the moderate damage state, D_3 the heavy damage state, and C the collapse state. A possible description of this random phenomenon in terms of a random variable X is: $X = 0$ for D_0, $X = 1$ for D_1, $X = 2$ for D_2, $X = 3$ for D_3, and $X = 4$ for C. Observe that this mapping is totally arbitrary and it can be defined differently by a different analyst. **Notation:** we use UPPERCASE letter to denote a random variable, e.g. X. We use italic lowercase to denote the sample points (outcomes), e.g. $x_1, x_2, ..., x_n, ...$ are the possible outcomes of X. In this section, there are two exceptions to this rule, that are the letters N and M. These are used to indicate the upper bound of a sequence of natural numbers characterized by the indexes m and n.

3.1 Probability Distributions

3.1.1 Discrete Random Variables

Given a discrete random variable X with possible outcomes $x_1, x_2, ...x_n...$, the probability mass function, PMF, is defined as

$$p_X(x_n) = P(X = x_n). \tag{26}$$

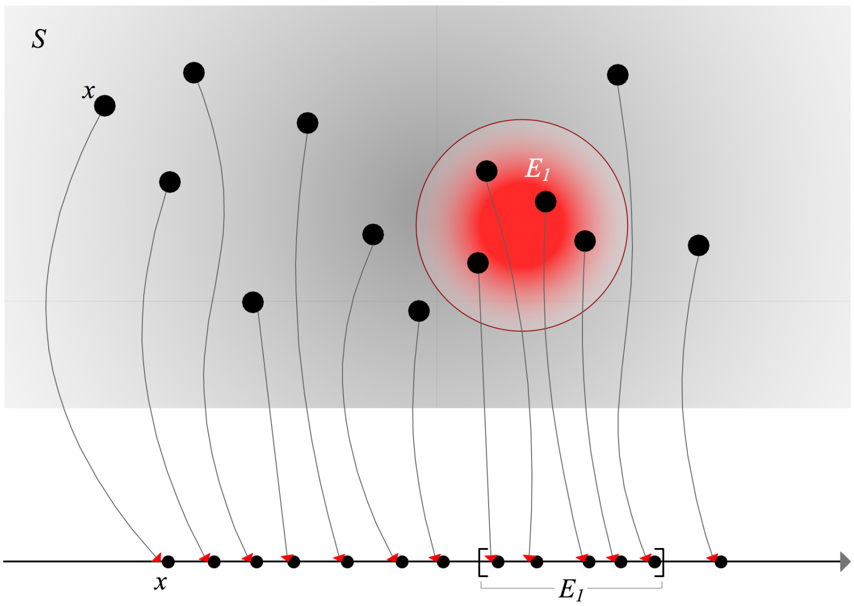


Figure 5: Random variable, mapping.

Observe that $p_X(x) = 0$ for $x \neq x_n$ and $p_X(x_n) = P(X = x_n)$ for any n. Moreover, $x_1, x_2, ..., x_n, ...$ is a set of mutually exclusive and collectively exhaustive events; then,

$$\sum_n p_X(X = x_n) = 1, \tag{27}$$

and $0 \leq p_X(x) \leq 1$. Given the PMF, the probability of any event can be determined by rules of probability theory. For example, the probability that X belongs to the interval $(a, b]$ is given by

$$P(a < X \leq b) = \sum_{a < x_n \leq b} p_X(x_n). \tag{28}$$

The probability distribution of a random variable can be expressed also in terms of the cumulative distribution function, CDF, defined as

$$F_X(x) = \sum_{x'_n \leq x} p_X(x'_n). \tag{29}$$

Example II

Suppose that you design a block of six identical buildings in a seismic area. Given an earthquake ground motion the probability of damage of one building is denoted with p. Since the buildings are close, it can be assumed that they will experience the same ground motion. It is also assumed that given the ground motion, the damaging events are statistically independent, i.e. if we know the ground motion, knowledge of the state of one building does not affect the probability of damage of another one. Observe that this is equivalent of assuming that the structural uncertainties are statistically independent.

We are interested in determining the probability distribution of the number of damage buildings. Let us define a random variable, X, representing this random phenomenon with possible outcomes $0, 1, 2, ..., 6$. A possible way to have x damage buildings out of 6 is to consider the first x buildings to be damaged and the remaining $(6 - x)$ not damaged. Given the independence assumption, this is equivalent to write $p^x \times (1 - p)^{6-x}$. However, this combination is not the only possible one, and there are other ways of selecting x item out of n. In particular, the total number of combinations is given by the binomial factor

$$\binom{6}{x} = \frac{6!}{x!(6 - x)!}; \tag{30}$$

then, we can write

$$p(x) = \binom{6}{x} p^x (1 - p)^{(n-x)}, \text{ for } x = 0, 1, 2, ..., 6. \tag{31}$$

The (31) is known as the binomial distribution. Later in this chapter, we will examine this distribution in greater details. Verify that $\sum_{x=0}^{6} p(x) = 1$.

Problems-III

i Compute the PMF of X given $p = 0.15$, what is the most likely number of buildings that are damaged?

ii What is the probability that two or more buildings are damaged?

iii Suppose you are the owner of the first building and you are in vacation. Suppose that you receive a call saying that there was an earthquake and at least 2 buildings are damaged. Given this information, what is the probability that your building is damaged?

3.1.2 Continuous Random Variables

A continuous random variable X has an infinite number of possible outcomes. It follows that we can not list out the probability associated with a specific sample point x, since x is one of the infinite number of sample points. To overcome this issue, we characterize probability within an interval. Then, we define the probability density function (PDF), $f_X(x)$, as a non-negative function such that

$$f_X(x) = \frac{P(x < X \leq x + dx)}{dx}, \tag{32}$$

where dx is a differential element of infinitesimal length.

Common pitfalls: given a sample point x, often students confuse the probability with the probability DENSITY of x. Notice the difference between the two. The PDF gives the probability per unit length! Moreover, observe that $f_X(x) \geq 0$ does not have an upper bound, so it is possible that $f(x) > 1$ for some x.

Given the definition (32), the probability that X is within the interval $(l, u]$ is written as

$$P(l < X \leq u) = \int_l^u f_X(x) dx. \tag{33}$$

Moreover, $f_X(x)$ must satisfy

$$\int_{-\infty}^{\infty} f_X(x) dx = 1. \tag{34}$$

An alternative description for the probability distribution is the cumulative distribution function (CDF), $F_X(x)$, defined as

$$F_X(x) = \int_{-\infty}^{x} f_X(x') dx'. \tag{35}$$

The CDF is related to the PDF by the following relationship

$$f_X(x) = \frac{dF(x)}{dx}; \tag{36}$$

moreover, the CDF must satisfy the following relationships: $F_X(-\infty) = 0$, $F_X(\infty) = 1$ and $F_X(l) \leq F_X(u)$ for $l \leq u$.

Example III

Suppose that you design an important structure which is located 10 km from the Hayward fault[4], Figure 6. We would like to derive the PDF and the CDF of the epicentral distance r, from the design site to the epicenter of an earthquake. r can be defined as random variable, R, because

[4]click these links to see simulations of earthquakes occurring in the Hayward fault for different epicenters: San Pablo Bay epicenter, Oakland epicenter, Fremont epicenter.

the epicenter occurs randomly along the fault. In this example, it is easier to define first the CDF, $F(r)$, and then derive the PDF with (36). Given that all the locations along the fault are equally likely to be an epicenter we can write

$$F_R(r) = P(R \leq r) = \frac{\text{length of segment with distance} \leq r}{\text{length of the fault}}. \tag{37}$$

Given the geometry of the site (Figure 6), the numerator can be written as $2\sqrt{r^2 - 10^2}$, and $10 \leq r \leq \sqrt{40^2 + 10^2} = 41.23$. Thus,

$$\begin{aligned} F_R(r) &= 0, \text{ for } r < 10 \\ &= \frac{2\sqrt{r^2 - 10^2}}{80}, \text{ for } 10 \leq r \leq 41.23 \\ &= 1 \text{ for } r > 41.23. \end{aligned} \tag{38}$$

The PDF is obtained by differentiation of the (38) as

$$\begin{aligned} f_R(r) &= 0, \text{ for } r < 10 \\ &= \frac{2r}{80\sqrt{r^2 - 10^2}}, \text{ for } 10 \leq r \leq 41.23 \\ &= 0 \text{ for } r > 41.23. \end{aligned} \tag{39}$$

Figure 7 shows the CDF and PDF of R.

Problems-IV

i Which conclusions can you draw from this PDF? Where the earthquake is more likely to occur ? (Give just a rough description, e.g. far from the site, near the site, etc.)

ii Now suppose that you do not know the exact location of the building, can you write a parametric code that, given a location, automatically give you the PDF and the CDF of R? (Assume the site is within the extension of the fault).

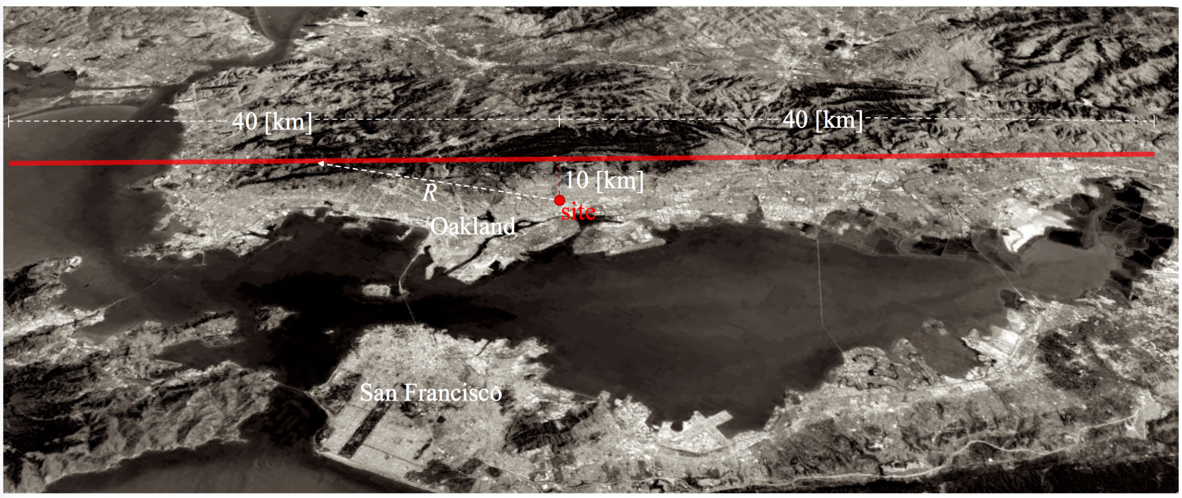


Figure 6: Hayward fault. Source: Google earth.

iii Suppose that your site is located in Sion, Switzerland. In this case there is not a clear fault geometry, so the epicenter can occur anywhere randomly within 80 [km] from the site. Assume that all outcome points have equal likelihood. What is the CDF and PDF of R? Make a plot of the functions.

iv Compare this last PDF with (39), which conclusions can you draw?

3.1.3 Mixed Random variables

In some engineering application we deal with random variables that are of mixed type. In particular, these can take both finite probabilities on specific sample points and probability densities function in one or more intervals. Such variables can be encompassed within the continuous random variable probability density functions by using the Dirac[5] delta function, $\delta(x)$, here defined as

$$\delta(x-a) \quad = \quad \lim_{\sigma \to 0} \frac{1}{\sigma\sqrt{\pi}} e^{-(x-a)^2/\sigma^2}. \tag{40}$$

(41)

Given the above definition it is easy to verify that

$$\int_{-\infty}^{\infty} \delta(x-a)dx = 1. \tag{42}$$

A little more thoughts (which are however beyond the purpose of this class) are required to understand the following equation

$$f(a) = \int_{-\infty}^{\infty} f(x)\delta(x-a)dx. \tag{43}$$

Roughy speaking the above equation essentially define a spike of unit area at a specific point a.

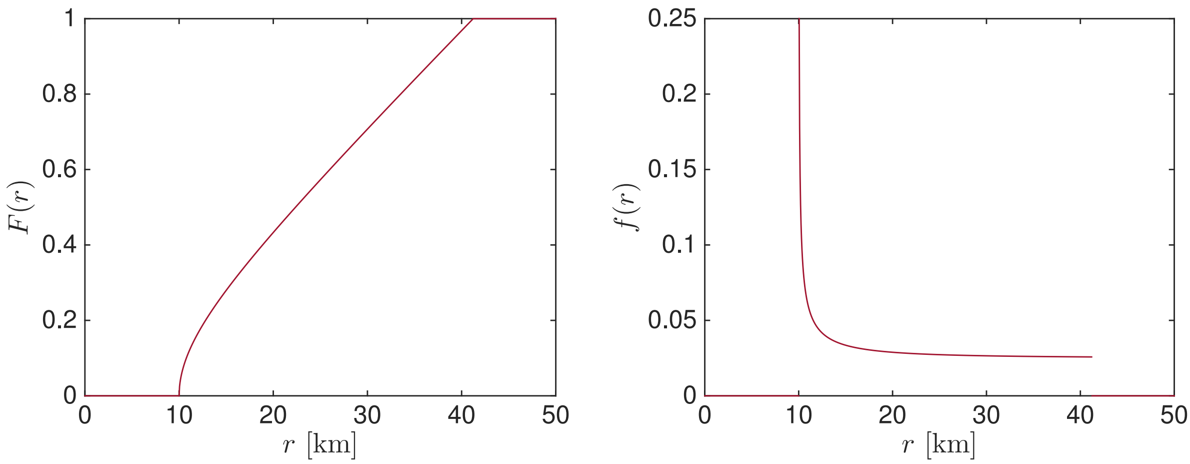


Figure 7: CDF and PDF of the epicentral distance for the given site

[5]Paul Adrien Maurice Dirac (*1902 †1984) was an English physicist who made fundamental contributions in quantum mechanics and electrodynamics. He is regarded as one of the most important physicist of the 20th century. Observe that the history of delta function is precedent than Dirac, and its origin can be found in the work of Jean-Baptiste Joseph Fourier (*1749 †1827). In particular in the famous treatise *Théorie analytique de la chaleur*, where for the first time it was presented what we know now as Fourier integral theorem.

The generalized PDF of a mixed random variable X, with probability mass, $p(x_n)$ at discrete points $x_1, x_2, ..., x_N$ and a probability density function $\bar{f}(x)$ for $x \in (-\infty, \infty)$, can be written as

$$f_X(x) = \bar{f}_X(x) + \sum_{n=1}^{N} p_{X_n}(x_n)\delta(x - x_n) \tag{44}$$

Example IV

Different researchers argue that the epicentral distance is not the "most appropriate" site-source distance to consider in seismic risk analysis. In fact, it would be more realistic to represent the earthquake as a finite rupture rather than a single point. In this case, a "more appropriate" site-source distance is given by the distance from the site to the nearest point on the finite rupture. In earthquake engineering, this distance is known as Joyner-Boore distance, r_{JB}. Let s denote the length of the rupture; moreover, assume that $s \leq l/2 = 40$[km], where l is the total length of the fault. It follows that in case the rupture extends between San Pablo Bay and Oakland the r_{JB} is the distance from the site to the right end of the rupture. Conversely, if the rupture extends between Oakland and Fremont the r_{JB} is the distance from the site to the left end of the rupture. Moreover, if the rupture crosses the foot of the normal from the site, then $r_{JB} = d = 10$[km] (where d is the perpendicular distance between the site to the fault). Since the location of the rupture is unknown, this is a random phenomenon. Therefore, it is convenient to define the random variable R_{JB}. Given the geometry of the problem (Figure 8, first subplot), the realizations r_{jb} of R_{JB} are belonging to the following interval $[d, \sqrt{d^2 + (l/2 - s)^2}]$ Figure 8, third subplot).

A convenient way to solve this problem is to fix a reference point along the rupture s. Let us define this reference point as the right end of the rupture. It follows that the sample space of the reference point is the interval $[s, l]$ of length $l - s$ (Figure 8 second subplot). Within this interval the realizations of the event $\{R_{JB} \leq r_{JB}\}$ occurs when the reference point belong to the interval $[l/2 - \sqrt{r^2 - d^2}, l/2 + \sqrt{r^2 - d^2} + s]$ of length $2\sqrt{r^2 - d^2} + s$ (Figure 8 last subplot). Given the equal likelihood of the points within the two intervals $P(R_{JB} \leq r_{JB})$ is equal to the ratio of the length of the two intervals, i.e.

$$\begin{aligned} F_{R_{JB}}(r) &= 0, \text{ for } r < 10 \\ &= \frac{2\sqrt{r^2 - 10^2} + s}{l - s}, \text{ for } 10 \leq r \leq \sqrt{100 + (40 - s)^2} \\ &= 1 \text{ for } r > \sqrt{100 + (40 - s)^2}. \end{aligned} \tag{45}$$

Observe that while $F_{R_{JB}}(R = 10) = 0$ in Example III,

$$F_{R_{JB}}(R_{JB} = 10) = \frac{s}{l - s}, \tag{46}$$

which is the finite probability that the rupture passes through the foot of the normal from the site. This events occurs when the reference point fell in the interval $[l/2, l/2 + s]$. $F_{R_{JB}}(r_{JB})$ has a discontinuity at $R_{JB} = d$, which makes R_{JB} a mixed random variable with a probability mass $s/(l - s)$ at $r = d$ and probability density $\bar{f}_{R_{JB}}(r_{JB}) = dF(r_{JB})/dr_{JB}$ in the interval $[d, \sqrt{d^2 + (l/2 - s)^2}]$. Owning Eq. (44)

$$f_{R_{JB}}(r_{JB}) = \frac{2r}{(80 - s)\sqrt{r^2 - 100}} + \delta(r_{JB} - d)\frac{s}{l - s}, \text{ for } 10 \leq r \leq \sqrt{100 + (40 - s)^2} \tag{47}$$

One can easy verify that for $s = 0$, Eq. (45) and Eq. (47) are equal to Eq. (38) and (39). Figure 9 show the CDF and PDF for different s lengths.

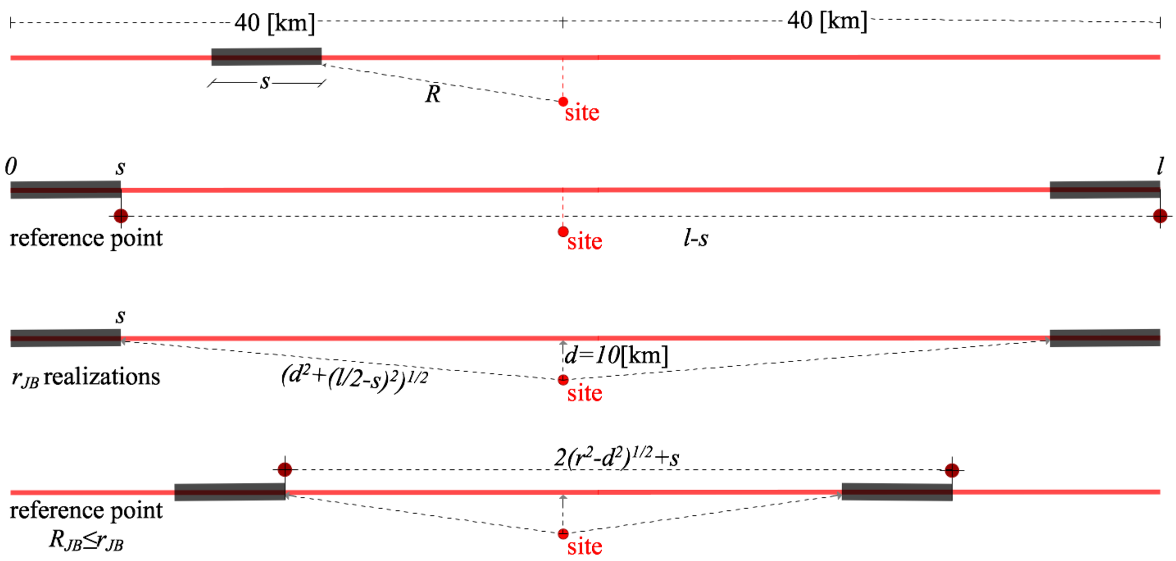


Figure 8: Hayward fault. Example IV

4 Multiple random variables

A lot of engineering problems deal with more than one random variable at time. For example, safety of structural components involve analysis of both load and a structural capacity, which are usually random variables. The fundamental question when we deal with problems involving multiple random variables is: if we know individually the PMF or PDF of single random variables, do we have complete information about the problem? The answer is in general NO! What we are missing is the statistical relationship between the random variables. This relationship is completely defined in terms of a joint probability distribution.

Given two random variables, the joint PDF is defined as

$$f_{XY}(x,y) = \frac{P(x < X \leq x+dx \cap y < Y \leq y+dy)}{dxdy}. \tag{48}$$

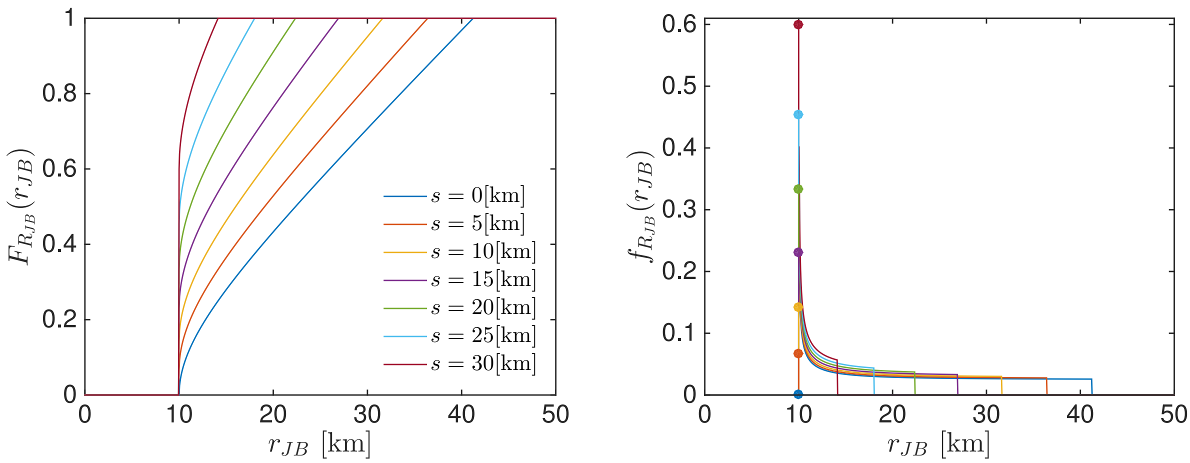


Figure 9: CDF and PDF of the Joyner-Boore distance for the given site. Mixed distribution, the stems at $R_{JB} = d$ indicate the discrete probability mass $P(R_{JB} = d) = s/(l-s)$.

Similarly to single random variables we can find the probability of falling in a given region, e.g. $x \in (l, u], y \in (\bar{l}, \bar{u}]$

$$P(l < X \leq u \cap \bar{l} < Y \leq \bar{u}) = \int_{\bar{l}}^{\bar{u}} \int_{l}^{u} f_{XY}(x, y) dx dy. \tag{49}$$

Moreover, $f_{XY}(x, y)$ must satisfy

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{XY}(x, y) dx dy = 1. \tag{50}$$

Similarly to the single random variable the joint CDF can be written as

$$F_{XY}(x, y) = P(X \leq x \cap Y \leq y) = \int_{-\infty}^{y} \int_{-\infty}^{x} f_{XY}(x', y') dx' dy', \tag{51}$$

and it can be related to joint PDF by the following

$$f_{XY}(x, y) = \frac{\partial^2 F(x, y)}{\partial x \partial y}. \tag{52}$$

Moreover, we have $F_{XY}(x, -\infty) = F_{XY}(-\infty, y) = F_{XY}(-\infty, -\infty)$=0, and $F_{XY}(\infty, \infty) = 1$.

Given the joint distribution of X and Y, one can obtain the the PDF of X, or Y, alone as follow

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x, y) dy, \text{ and } f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x, y) dx. \tag{53}$$

These operations are named marginalization, and $f_X(x)$ and $f_Y(y)$ are called the marginal distributions of X and Y. Figure 10-a shows this operation.

The above expressions can be extended to any number of random variables. For the sake of simplicity when dealing with multiple random variables we use the vectorial notations $\boldsymbol{X} = [X_1, ..., X_N]^T$ (i.e. a random vector) and $\boldsymbol{x} = [x_1, ..., x_N]^T$ (i.e a sample point). Given this notation, the joint PDF and CDF are written as $f_{\boldsymbol{X}}(\boldsymbol{x})$ and $F_{\boldsymbol{X}}(\boldsymbol{x})$. Then, the properties described for the bivariate PDF and CDF can be extended for the multivariate distribution function. In particular,

$$\underbrace{\int_{-\infty}^{\infty}}_{N \text{ folds}} f_{\boldsymbol{X}}(\boldsymbol{x}) d\boldsymbol{x} = 1; \tag{54}$$

moreover, given a vector partition, $\boldsymbol{X} = (\boldsymbol{X}_1, \boldsymbol{X}_2)^T$, where $\boldsymbol{X}_1 = [X_1, ..., X_M]^T$ and $\boldsymbol{X}_2 = [X_{M+1}, ..., X_N]^T$ we can obtain the PDF of $\boldsymbol{X}_1$, or $\boldsymbol{X}_2$, alone as follow:

$$f_{\boldsymbol{X}_1}(\boldsymbol{x}_1) = \underbrace{\int_{-\infty}^{\infty}}_{N-M+1 \text{ folds}} f_{\boldsymbol{X}}(\boldsymbol{x}_1, \boldsymbol{x}_2) d\boldsymbol{x}_2, \text{ and } f(\boldsymbol{x}_2) = \underbrace{\int_{-\infty}^{\infty}}_{M \text{ folds}} f_{\boldsymbol{X}}(\boldsymbol{x}_1, \boldsymbol{x}_2) d\boldsymbol{x}_1, \tag{55}$$

4.1 Conditional Distribution and Statistical Independence

For two continuous random variables, X and Y, we define the conditional distribution of X given Y as

$$\begin{aligned} f_{X|Y}(x|y) &= \frac{P(x < X \leq x + dx | y < Y \leq y + dy)}{dx} \\ &= \frac{P(x < X \leq x + dx \cap y < Y \leq y + dy)}{P(y < Y \leq y + dy)} \frac{1}{dx} \\ &= \frac{f(x, y)}{f(y)}. \end{aligned} \tag{56}$$

Figure 10-b shows this concept. One can observe that the joint PDF $f_{XY}(x, y)$ can be expressed as a product of one conditional and one marginal PDF, i.e.

$$f_{XY}(x, y) = f_{X|Y}(x|y)f_Y(y) = f_{Y|X}(x|y)f_Y(y). \tag{57}$$

Two random variables X and Y are statistically independent if

$$f_{X|Y}(x|y) = f_X(x), \text{ or } f_{Y|X}(y|x) = f_Y(y). \tag{58}$$

It follows that for two statistically independent random variables knowledge of the marginals is sufficient to describe the joint probability distribution, i.e. $f_{XY}(x, y) = f(x)f(y)$. Conditional distribution for multiple random variables are defined by generalizing the concepts introduced in the previous Section, e.g. the conditional distribution of $\boldsymbol{X}_2$ given $\boldsymbol{X}_1 = \boldsymbol{x}_1$ is

$$f_{\boldsymbol{X}_2|\boldsymbol{X}_1}(\boldsymbol{x}_2|\boldsymbol{x}_1) = \frac{f_{\boldsymbol{X}}(\boldsymbol{x})}{f_{\boldsymbol{X}_1}(\boldsymbol{x}_1)}, \tag{59}$$

which leads to the following rule

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = f_{\boldsymbol{X}_2|\boldsymbol{X}_1}(\boldsymbol{x}_2|\boldsymbol{x}_1)f_{\boldsymbol{X}_1}(\boldsymbol{x}_1), \tag{60}$$

and more in general

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = f(\boldsymbol{x}_N|\boldsymbol{x}_1, ..., \boldsymbol{x}_{N-1})f(\boldsymbol{x}_{N-1}|x_1, ..., x_{N-2})...f(x_2|x_1)f(x_1), \tag{61}$$

The set of random variables $\boldsymbol{X}$ are jointTLY statistically independent if

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = f(x_1)f(x_2)...f(x_N), \text{ for } x_1, x_2, ..., x_N. \tag{62}$$

Please note that in (61) and (62) for clarity of notation we drop the conventional subscript $\cdot_{\boldsymbol{X}}$. The reader should consider it implicit.

Common pitfalls: it is not uncommon to relate (wrongly) joint statistically independence to pairwise independence. It is important to note that pairwise independence DOES NOT guarantee independence as specified by the above condition.

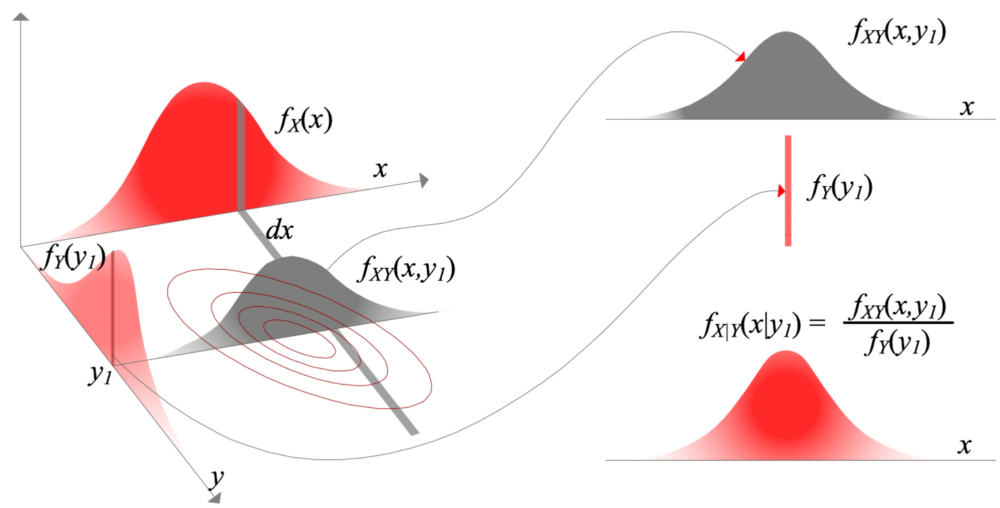


Figure 10: CDF and PDF of the epicentral distance for the given site

Example V

The Californian Memorial Stadium, Figure 11-a, is the stadium of the University of California, Berkeley football team. The stadium is a US historic landmark, opened in 1923, which currently seats circa 63,000 people. The Hayward fault crosses the infrastructure, Figure 11-b. The stadium has been recently retrofitted, by splitting it into two components to allow a relative permanent displacement in case of an earthquake. In the following Example, we compute the probability that the stadium will have a relative permanent displacement. We assume that there will be such displacement in the case the rupture will cross (even partially) the Stadium. We denote this event as E_s. Moreover, we denote with l_1 the distance from the left end of the Hayward fault to the left end of the Memorial Stadium, with $l_2 = l - l_1$, with w the width of the stadium, and with s the unknown rupture length. Given this, we consider the rupture length as a random phenomenon, and therefore we define the random variable S. Following the same geometrical considerations of Example -III and -IV, we can first define the conditional probability distribution for a given length s, $P(E_s|S = s)$, and then using the total probability theorem to compute $P(E_s)$ as

$$P(E_s) = \int_s P(E_s|S = s) f_S(s) ds.$$

In this Example V, we compute the conditional probability density $P(E|S = s)$; then, in Chapter 3 we will define $f_S(s)$ and compute $P(E_s)$. Following Figure (12), we can identify three cases:
Case-I:

$$P(E|S = s) = \frac{w + s}{l_1 + l_2 - s}, \; 0 < s \leq l_1,$$

Case-II:

$$P(E|S = s) = \frac{w + l_1}{l_1 + l_2 - s}, \; l_1 < s \leq l_2 - w,$$

Case-III:

$$P(E|S = s) = 1, \; s > l_2 - w.$$

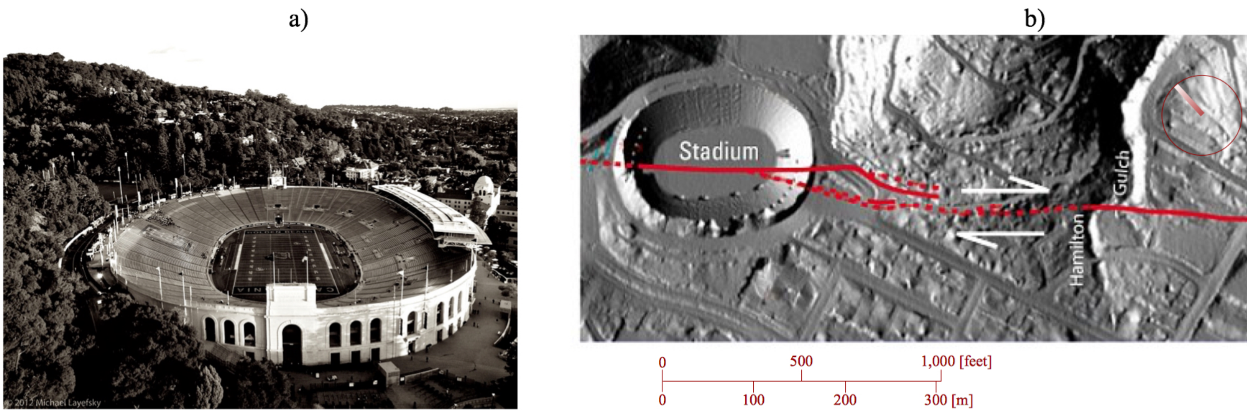


Figure 11: a- Memorial Stadium, Source: flickr, Michael Layefsky. b- Hayward fault crossing the stadium; Source: USGS

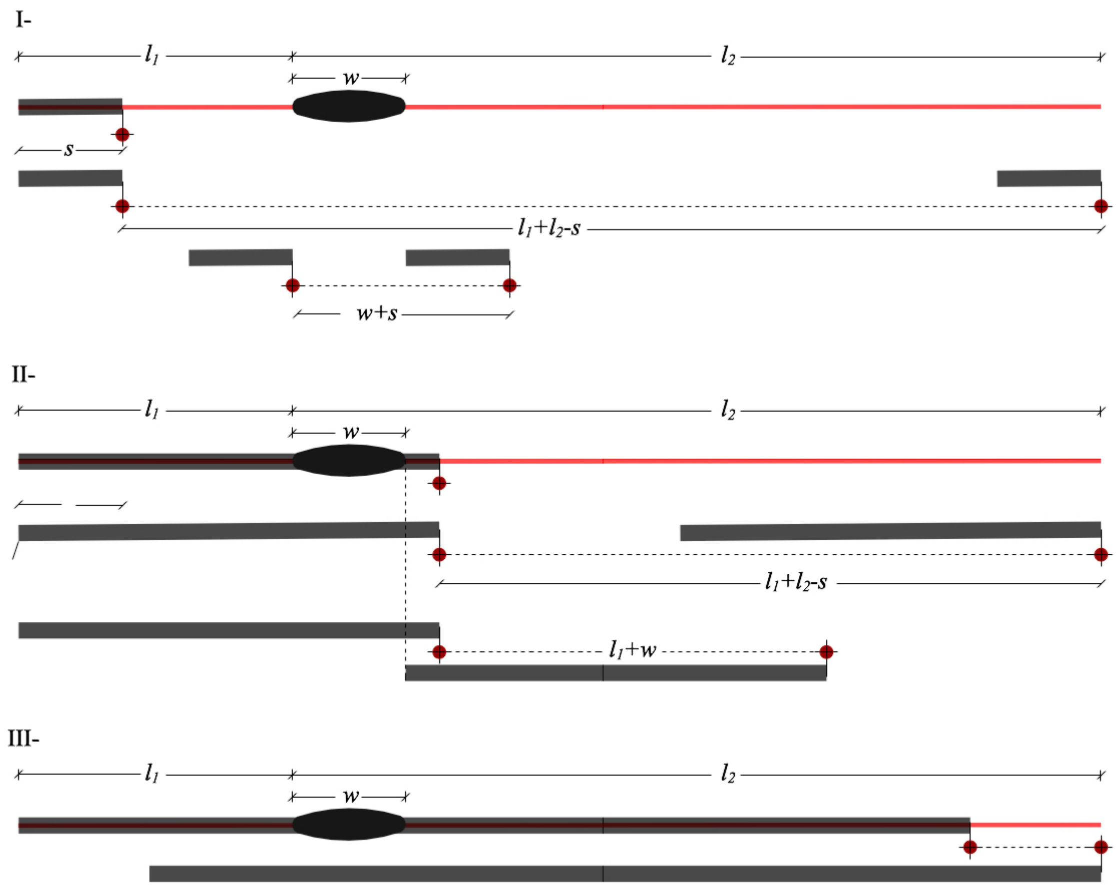


Figure 12: Example 5, case I- $0 < s \leq l_1$, case II- $l_1 < s \leq l_2 - w$, case III- $l_2 - w < s$

5 Partial Description of Random Variables, Moments and Expectation

Given a random variable, X, the complete information about this variable is given by its PDF. However, oftentimes we would like to have summary values which partially describe X. The most common partial descriptions of a random variable are: mean, variance (or standard deviation), coefficient of variation, median and mode. The mean is defined as

$$\mu_x = \int_{-\infty}^{\infty} x f_X(x) dx. \tag{63}$$

The variance is defined as

$$\sigma_x^2 = \int_{-\infty}^{\infty} (x - \mu_x)^2 f_X(x) dx, \tag{64}$$

and the standard deviation as $\sigma_x = \sqrt{\sigma_x^2}$. The coefficient of variation (c.o.v) is defined as

$$\delta_x = \frac{\sigma_x}{|\mu_x|}, \text{ for } \mu_x \neq 0. \tag{65}$$

The median of a distribution, $\tilde{x}$, is defined as the "middle" point of the distribution, i.e.

$$\int_{-\infty}^{\tilde{x}} f_X(x) dx = 0.5. \tag{66}$$

The mode of the distribution, is defined as the point/s (or the interval) with the highest probability density

$$\text{mode} = \arg\max_x f_X(x). \tag{67}$$

5.1 Expectation

Consider a function $g(X)$ of random variable X. The mathematical expectation of $g(X)$ is defined by

$$E[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x) dx. \tag{68}$$

Given a vector of random variables $\boldsymbol{X}$ and a multivariate function $g(\boldsymbol{X})$ the mathematical expectation is simply

$$E[g(\boldsymbol{X})] = \underbrace{\int_{-\infty}^{\infty}}_{N \text{ folds}} g(\boldsymbol{x}) f_{\boldsymbol{X}}(\boldsymbol{x}) d\boldsymbol{x}. \tag{69}$$

Observe that the integral operation is linear; thus, given two functions $g_1(\boldsymbol{X})$ and $g_2(\boldsymbol{X})$, one can write

$$E[g_1(\boldsymbol{X}) + g_2(\boldsymbol{X})] = E[g_1(\boldsymbol{X})] + E[g_2(\boldsymbol{X})] \tag{70}$$

5.2 Moments of Random Variable

Consider the function $g(X) = X^k$ for a random variable X, where $k \in \mathbb{N}$. Then, the expectation

$$E[X^k] = \int_{-\infty}^{\infty} x^k f_X(x) dx \tag{71}$$

is known also as the k-moment of X. In particular, the first moment $E[X] = \mu_x$ describes the barycenter of the distribution, the second moment $E[X^2]$ is a measure of the dispersion, and the third and fourth moments are useful in computing measures of skewness and flatness of the distribution.

Consider $g(\boldsymbol{X}) = (X - \mu_x)^k$; then, the expectation

$$E[X^k] = \int_{-\infty}^{\infty} (x - \mu_x)^k f_X(x) dx \tag{72}$$

is the k-th central moment of X. For $k = 2$ we have $E[(X - \mu_x)^2] = \text{Var}[X] = \sigma^2$ is the central moment of inertia of the distribution. One can use the linear property of expectation to write

$$E[(X - \mu_x)^2] = E[X^2 - 2\mu_x X + \mu_x^2] = E[X^2] - E^2[X] \tag{73}$$

Continuing our analogy with mechanics, we observe that (73) is the equivalent of the parallel axis theorem. Similarly, the third central moment can be written as

$$E[(X - \mu_x)^3] = E[X^3] - 3E[X^2]E[X] + 2E^3[X] \tag{74}$$

Observe that $(X - \mu_x)^3$ retains the sign of the deviation from the mean; it follows that the third moment provides a good measure of the non-symmetry of the distribution. Usually, this is expressed by the coefficient of skewness defined by

$$\gamma_1 = \frac{E[(X - \mu_x)^3]}{\sigma^3}. \tag{75}$$

One can observe that for $\gamma_1 = 0$ the distribution is symmetric, whereas $\gamma_1 > 0$ indicates skewness of the distribution to the right, and $\gamma_1 < 0$ indicates skewness to the left. Finally, the coefficient of kurtosis is written as

$$\gamma_2 = \frac{E[(X - \mu_x)^4]}{\sigma^4}. \tag{76}$$

Kurtosis derives from Greek $\kappa\upsilon\rho\tau\grave{o}s$ which means "curved"; in fact, it is a measure of "flatness" of the distribution shape.

Consider the function $g(X) = \exp(\omega X)$ of random variable X. The mathematical expectation

$$E[\exp(\omega X)] = \int_{-\infty}^{\infty} \exp(\omega x) f_X(x) dx = M(\omega), \tag{77}$$

is named moment generating function. Note that this is essentially the same as the definition of the Laplace transform of a function $f(x)$, except that we are using ω instead of $-\omega$. The (77) is a convenient expression since

$$\frac{d^k M(\omega)}{d\omega^k} = \int_{-\infty}^{\infty} x^k \exp(\omega x) f_X(x) dx, \tag{78}$$

and

$$\left.\frac{d^k M(\omega)}{d\omega^k}\right|_{\omega=0} = \int_{-\infty}^{\infty} x^k f_X(x) dx = E[X^k]. \tag{79}$$

Similar expression can be derived in case of a random vector $\boldsymbol{X}$. In this case given $\boldsymbol{\omega} = [\omega_1, ..., \omega_N]$, one can write

$$M(\boldsymbol{\omega}) = E[\exp(\boldsymbol{\omega X})] = \underbrace{\int_{-\infty}^{\infty}}_{N \text{ folds}} \exp(\boldsymbol{\omega x}) f_X(\boldsymbol{x}) d\boldsymbol{x}, \tag{80}$$

and

$$\frac{\partial^{k_1+k_2+...+k_N}}{\partial\omega_1^{k_1}, \partial\omega_2^{k_2}, ..., \partial\omega_N^{k_N}} M(\boldsymbol{\omega}) = E[X_1^{k_1}, X_2^{k_2}, ..., X_N^{k_N}]. \tag{81}$$

5.3 Joint Moments of Random Variables

Given the random vector X, the joint moment of order $k_1 + ... + k_N$ is defined as the expectation

$$E[X_1^{k_1} X_2^{k_2} ... X_N^{k_N}] = \underbrace{\int_{-\infty}^{\infty}}_{N \text{ folds}} x_1^{k_1} x_2^{k_2} ... x_N^{k_N} f_{\boldsymbol{X}}(\boldsymbol{x}) d\boldsymbol{x}. \tag{82}$$

The joint central moment of order $k_1 + ... + k_N$ is defined by

$$E[(X_1 - \mu_1)^{k_1}(X_2 - \mu_2)^{k_2}...(X_N - \mu_N)^{k_N}] = \underbrace{\int_{-\infty}^{\infty}}_{N \text{ folds}} (x_1 - \mu_{x_1})^{k_1}(x_2 - \mu_{x_2})^{k_2}...(x_N - \mu_{x_N})^{k_N} f(\boldsymbol{x})d\boldsymbol{x}, \tag{83}$$

where $\mu_n = E[X_n]$ is the mean of X_n. Among these moments the most important are the lowest pair-wise central moments, i.e.,

$$E[(X_n - \mu_n),(X_m - \mu_m)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x_n - \mu_n)(x_m - \mu_m)f_X(x_n, x_m)dx_n dx_m. \tag{84}$$

The (84) is known as covariance; therefore, we can write

$$\begin{aligned} \mathrm{Cov}[X_n, X_m] &= E[(X_n - \mu_n)(X_m - \mu_m)] && (85)\\ &= E[X_n X_m] - E[X_n]\mu_m - \mu_n E[X_m] + E[X_n]E[X_m] && (86)\\ &= E[X_n X_m] - E[X_n]E[X_m]; && (87) \end{aligned}$$

moreover, if $X_m \perp\!\!\!\perp X_n \Rightarrow \mathrm{Cov}[X_n, X_m] = 0$ since $E[XY] = E[X]E[Y]$. An important property of the covariance is that its absolute value is bounded by square root of the product of variances, i.e.

$$|\mathrm{Cov}[X_n X_m]| \leq \sqrt{\mathrm{Var}[X_n]\mathrm{Var}[X_m]}. \tag{88}$$

The correlation coefficient between random variables X_m and X_n is defined as

$$\rho_{mn} = \frac{\mathrm{Cov}[X_m X_n]}{\sigma_{X_m}\sigma_{X_n}}. \tag{89}$$

Observe that $-1 \leq \rho \leq 1$. This coefficient provides a dimensionless measure of linear dependence between the two variables. It is easy to show that when $|\rho| = 1$ there is perfect linear dependence.

Common pitfalls: it is not uncommon among student to reverse the following condition $X \perp\!\!\!\perp Y \Rightarrow \rho_{XY} = 0$. Please, NEVER do this. In fact $\rho_{XY} = 0$ or equivalently $\mathrm{Cov}[XY] = 0$ indicates ONLY that there is not linear dependence. For example two random variables can have nonlinear dependence which leads to $\rho = 0$ but this does not mean that they are statistically independent! For example, let's investigate the correlation between the kinetic energy, E_K, of a particle of mass m and its velocity V. Suppose that the velocity is a random variable that has a symmetric distribution (i.e. $E[V^3] = 0$) with 0 mean (i.e. E[V]=0). It follows that the kinetic energy is a random variable since $E_k = 1/2mV^2$. Clearly V and K are perfectly correlated, i.e. knowledge of V is sufficient to know deterministically Y. It follows that V and K are statistically dependent. However, $\mathrm{Cov}[VK] = mE[VK] - E[V]E[K] = 1/2mE[V^3] = 0$. Thus, keep in mind $\rho_{XY} = 0 \not\Rightarrow X \perp\!\!\!\perp Y$!!

6 The Normal Distribution

The Gaussian (from Carl Friedrich Gauss [6] normal distribution is a probability density function for a random variable that appears frequently in engineering applications.

The probability density function is

$$f_X(x;\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right], \quad -\infty \leq x \leq \infty, \tag{90}$$

where μ is the location parameter (mean), and σ the scale parameter (standard deviation). A standard normal random variable, Z, is a normal random variable with zero mean and standard deviation one. It follows that its PDF is:

$$f_Z(z) = \varphi(z) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}z^2\right], \quad -\infty \leq z \leq \infty, \tag{91}$$

and CDF

$$F_Z(z) = \Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}z'^2\right] dz', \quad -\infty \leq z \leq \infty. \tag{92}$$

Observe that there is no close form solution for the CDF of the normal distribution. Moreover given the symmetry of the distribution $\Phi(-z) = 1 - \Phi(z)$. Since there is no close form solution the CDF is evaluated numerically and its values are reported in the normal distribution table, Figure 13. Alternatively, nowadays, any software can give directly the values of CDF.

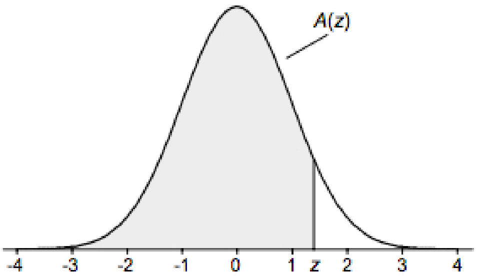
To use the normal distribution table with a classical normal random variable X, we should implement the following transformation:

$$Z = \frac{X-\mu}{\sigma}; \tag{93}$$

then,

$$P(X < x) = P\left(\frac{X-\mu}{\sigma} \leq \frac{x-\mu}{\sigma}\right) = P\left(Z \leq \frac{x-\mu}{\sigma}\right) = \Phi\left(\frac{x-\mu}{\sigma}\right). \tag{94}$$

[6] Johann Carl Friedrich Gauss (*1777 †1855) was probably the greatest mathematician of modern history. Gauss had an unprecedented and extraordinary influence in several fields of mathematics and science. Oftentimes, he is remembered as the *Princeps mathematicorum* (the primer mathematician)



$A(z)$ is the integral of the standardized normal distribution from $-\infty$ to z (in other words, the area under the curve to the left of z). It gives the probability of a normal random variable not being more than z standard deviations above its mean. Values of z of particular importance:

z	$A(z)$	
1.645	0.9500	Lower limit of right 5% tail
1.960	0.9750	Lower limit of right 2.5% tail
2.326	0.9900	Lower limit of right 1% tail
2.576	0.9950	Lower limit of right 0.5% tail
3.090	0.9990	Lower limit of right 0.1% tail
3.291	0.9995	Lower limit of right 0.05% tail

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998
3.6	0.9998	0.9998	0.9999							

Figure 13: Standard Normal CDF