Instituto Politécnico Nacional
"La Técnica al Servicio de la Patria"

Laboratorio de Ciberseguridad

Centro de Investigación en Computación(CIC)
Instituto Politécnico Nacional - México.

Probability, Random Processes and Inference

Dr. Ponciano Jorge Escamilla Ambrosio
pescamilla@cic.ipn.mx
http://www.cic.ipn.mx/~pescamilla/

Course Content

1.4. General Random Variables
- 1.4.1. Continuous Random Variables and PDFs
- 1.4.2. Cumulative Distribution Function
- 1.4.3. Normal Random Variables
- 1.4.4. Joint PDFs of Multiple Random Variables
- 1.4.5. Conditioning
- 1.4.6. The Continuous Bayes' Rule
- 1.4.7. The Strong Law of Large Numbers

General Random Variables

- Continuous random variables
  - The velocity of a vehicle traveling along the highway
- Continuous random variables can take on any real value in an interval.
  - possibly of infinite length, such as $(0,\infty)$ or the entire real line.
- In this section the concepts and method for discrete r.v.s, such as expectation, PMF, and conditioning, for their continuous counterparts are introduced.

Probability Density Function

- Continuous random variable. A random variable is called **continuous** if there exists a non negative function f_X, called the **probability density function** of X, or PDF, such that:

$$\mathbf{P}(X \in B) = \int_B f_X(x)\,dx$$

- For every subset *B* of the real line

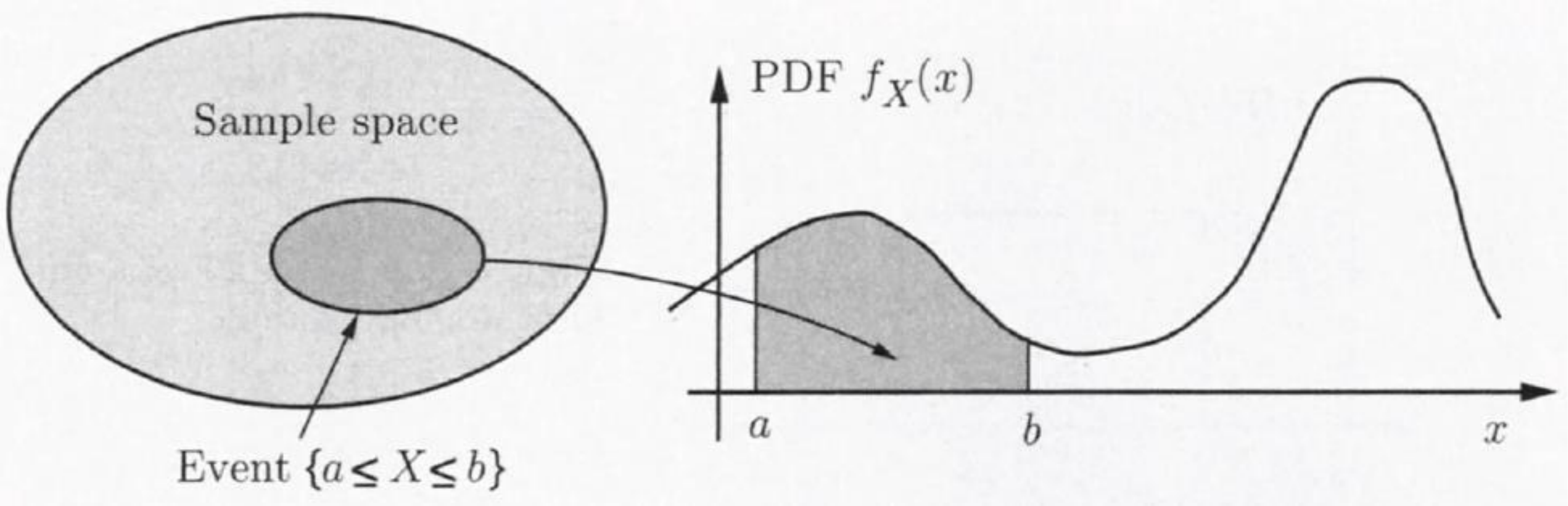
Probability Density Function

- The probability that the value of X falls within an interval is:

$$\mathbf{P}(a \le X \le b) = \int_a^b f_X(x)\, dx$$

which can be interpreted as the area under the graph of the PDF.

Probability Density Function



$$\mathbf{P}(a \leq X \leq b) = \int_a^b f_X(x)\, dx$$

Probability Density Function

- For any single value *a*, we have:

$$\mathbf{P}(X = a) = \int_a^a f_X(x)\,dx = 0$$

- For this reason, including or excluding the endpoints of an interval has no effect on its probability:

$$\mathbf{P}(a \le X \le b) = \mathbf{P}(a < X < b) = \mathbf{P}(a \le X < b) = \mathbf{P}(a < X \le b)$$

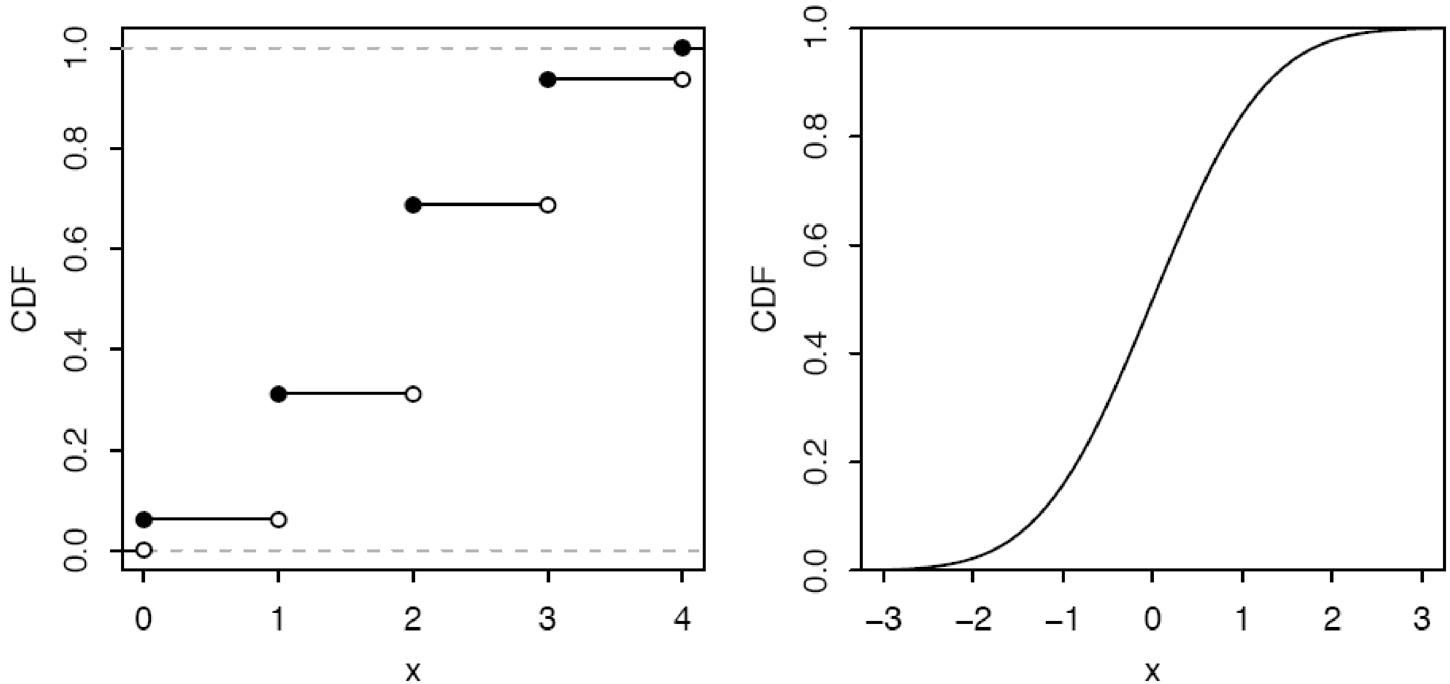
Probability Density Function

- To qualify as a PDF, a function f_X must be:
  - nonnegative, i.e., $f_X(x) \geq 0$ for every x,
  - have the normalisation property:

$$\int_{-\infty}^{\infty} f_X(x)\,dx = \mathbf{P}(-\infty < X < \infty) = 1$$

- Graphically, this means that the entire area under the graph of the PDF must be equal to 1.

Discrete vs. continuous r.v.s.



Recall that for a discrete r.v., the CDF jumps at every point in the support, and is flat everywhere else. In contrast, for a continuous r.v. the CDF increases smoothly.

Discrete vs. continuous r.v.s.

- For a continuous r.v. X with CDF, $F_X(x)$, the probability density function (PDF) of X is the derivative $f_X(x)$ of the CDF, given by $f_X(x) = F'_X(x)$. The support of X, and of its distribution, is the set of all x where $f_X(x) > 0$.

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

- The PDF represents the "density" of probability at the point x.

Probability Density Function

- To get from the PDF back to the CDF we apply:

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\, dt$$

$$\int_{-\infty}^{x} f(t)dt = F(x) - F(-\infty) = F(x)$$

- Thus, analogous to how we obtained the value of a discrete CDF at *x* by summing the PMF over all values less than or equal to *x*; here we integrate the PDF over all values up to *x*, so the CDF is the accumulated area under the PDF.

Probability Density Function

- Since we can freely convert between the PDF and the CDF using the inverse operations of integration and differentiation, both the PDF and CDF carry complete information about the distribution of a continuous r.v.
- **Thus the PDF completely specifies the behavior of continuous random variables.**

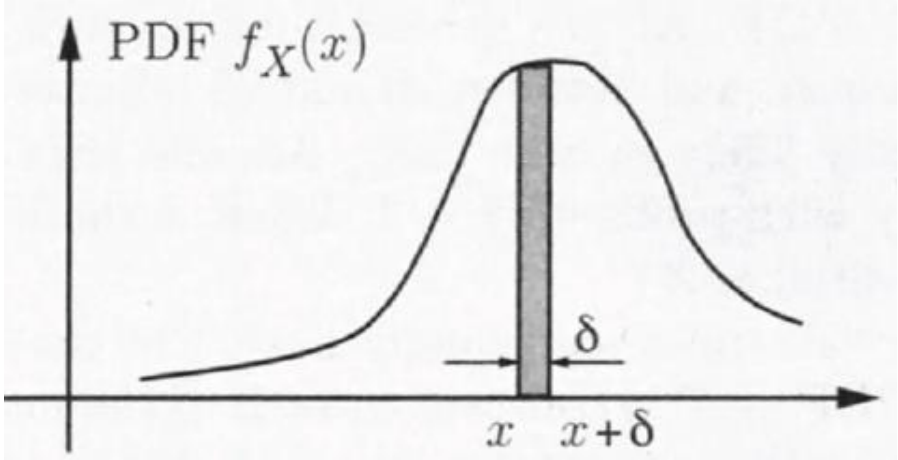
Probability Density Function

- For an interval $[x, x+\delta]$ with very small length δ, we have:

$$\mathbf{P}([x, x+\delta]) = \int_x^{x+\delta} f_X(t)\,dt \approx f_X(x)\cdot\delta$$

So we can view $f_X(x)$ as the "probability mass per unit length" near x.

Even though a PDF is used to calculate event probabilities, $f_X(x)$ is not the probability of any particular event. In particular, it is not restricted to be less than or equal to one.



Probability Density Function

- An important way in which continuous r.v.s differ from discrete r.v.s is that for a continuous r.v. X, P(X = *x*) = 0 for all *x*. This is because P(X = *x*) is the height of a jump in the CDF at *x*, but the CDF of X has no jumps! Since the PMF of a continuous r.v. would just be 0 everywhere, we work with a PDF instead.

$$\mathbf{P}(X = a) = \int_a^a f_X(x)\,dx = 0$$

Probability Density Function

- The PDF is analogous to the PMF in many ways, but there is a key difference: for a PDF f_X , the quantity $f_X(x)$ is not a probability, and in fact it is possible to have $f_X(x) > 1$ for some values of x. To obtain a probability, we need to integrate the PDF.
- In summary:
  - **To get a desired probability, integrate the PDF over the appropriate range.**

Examples of PDFs

- The **Logistic** distribution has CDF:

$$F(x) = \frac{e^x}{1+e^x}, \quad x \in \mathbb{R}$$

- To get the PDF, we differentiate the CDF, which gives:

$$f(x) = \frac{e^x}{(1+e^x)^2}, \quad x \in \mathbb{R}$$

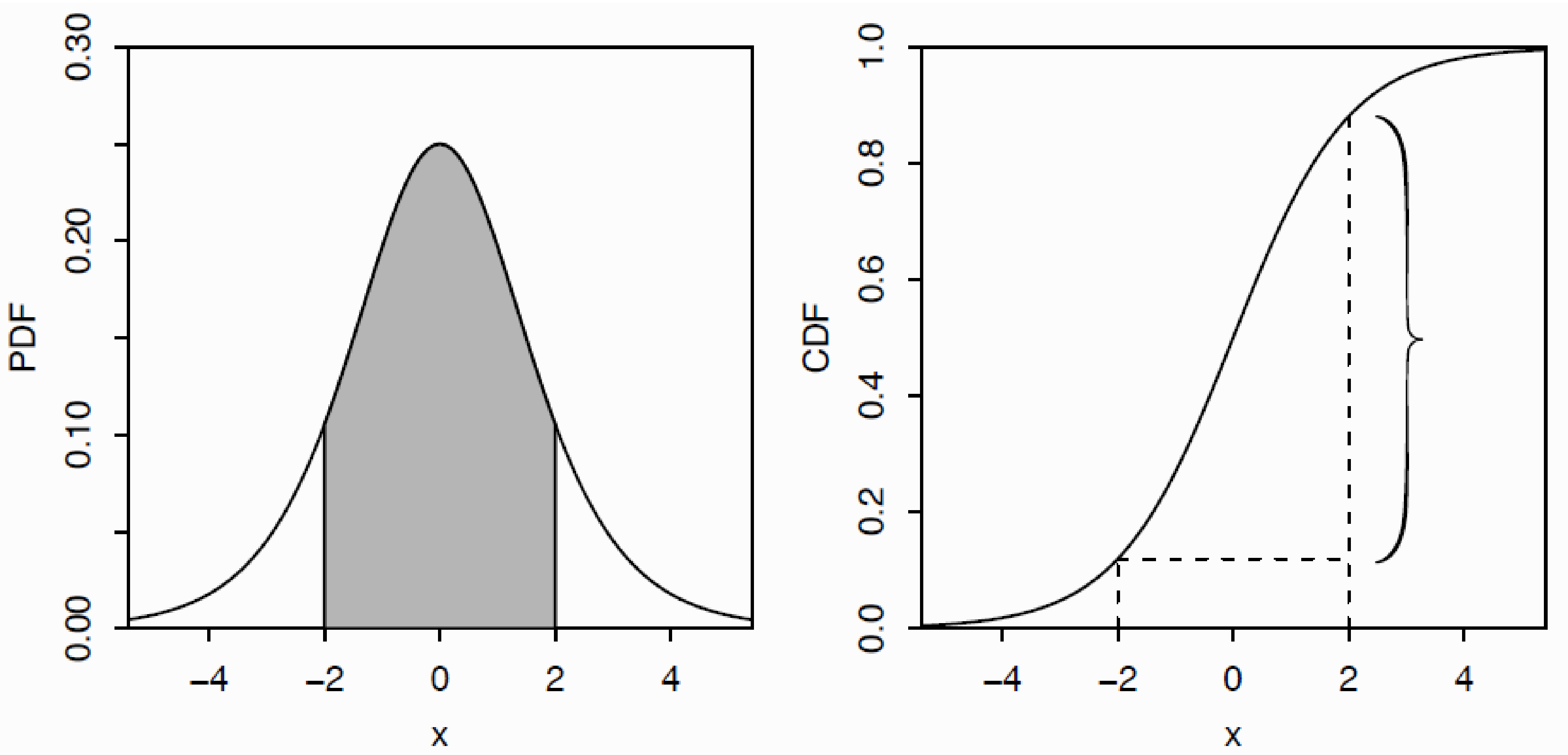
- Example:

$$P(-2 < X < 2) = \int_{-2}^{2} \frac{e^x}{(1+e^x)^2} dx = F(2) - F(-2) \approx 0.76$$

$$\int_{-2}^{2} \frac{e^x}{(1+e^x)^2} dx = \int_{1+e^{-2}}^{1+e^2} \frac{1}{u^2} du = \left(-\frac{1}{u}\right)\Big|_{1+e^{-2}}^{1+e^2} \approx 0.76$$

$u = 1 + e^x$, so $du = e^x dx$

Examples of PDFs



Logistic PDF and CDF. The probability $P(-2 < X < 2)$ is indicated by the shaded area under the PDF and the height of the curly brace on the CDF.

Examples of PDFs

- The **Rayleigh** distribution has CDF:

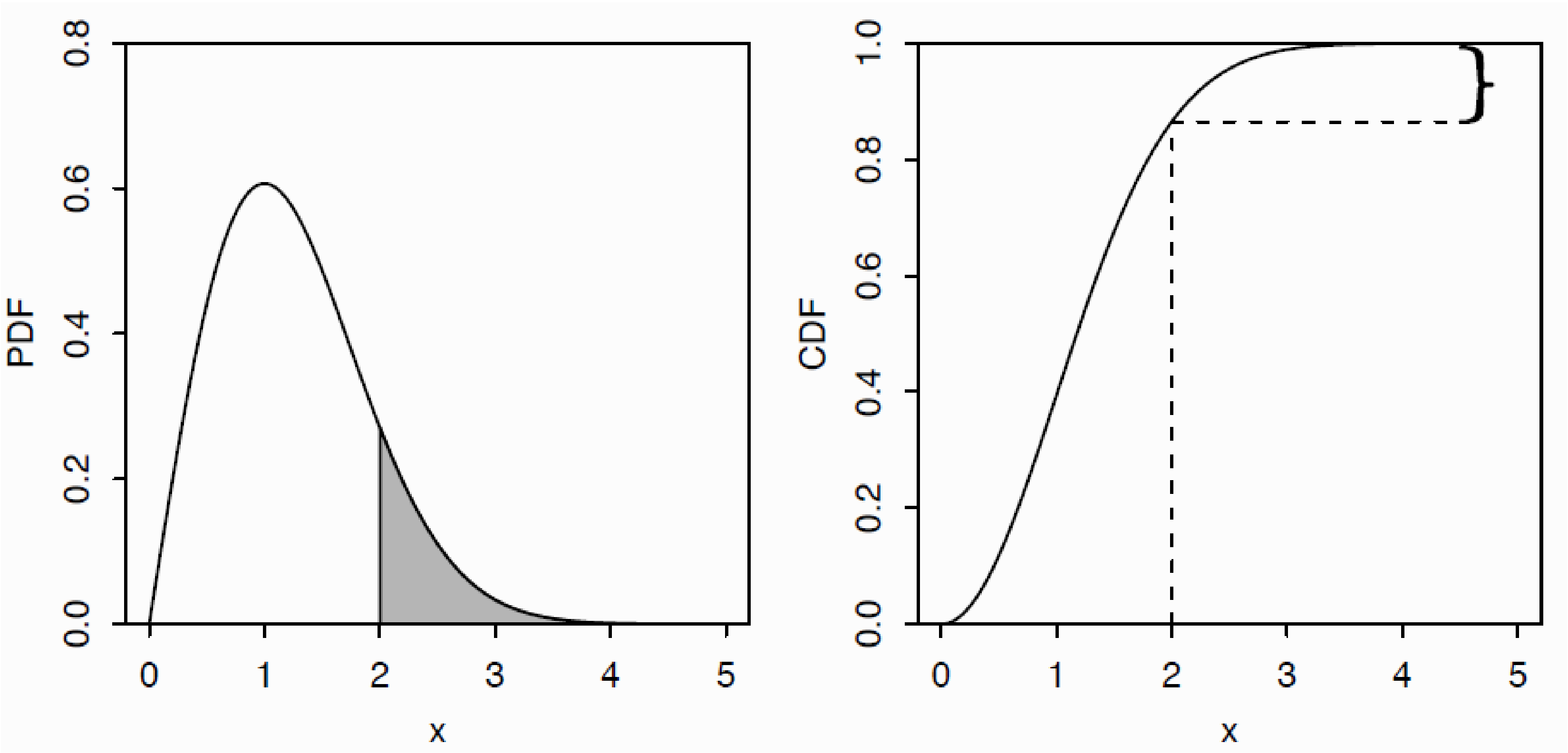
$$F(x) = 1 - e^{-x^2/2}, \quad x > 0$$

- To get the PDF, we differentiate the CDF, which gives:

$$f(x) = xe^{-x^2/2}, \quad x > 0$$

- Example: $P(X > 2) = \int_2^{\infty} xe^{-x^2/2} dx = 1 - F(2) \approx 0.14$

Examples of PDFs



Rayleigh PDF and CDF. The probability $P(X > 2)$ is indicated by the shaded area under the PDF and the height of the curly brace on the CDF.

Examples of PDFs

- A continuous r.v. X is said to have Uniform distribution on the interval (*a*, *b*) if its PDF is:

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{if } a < x < b, \\ 0 & \text{otherwise.} \end{cases}$$

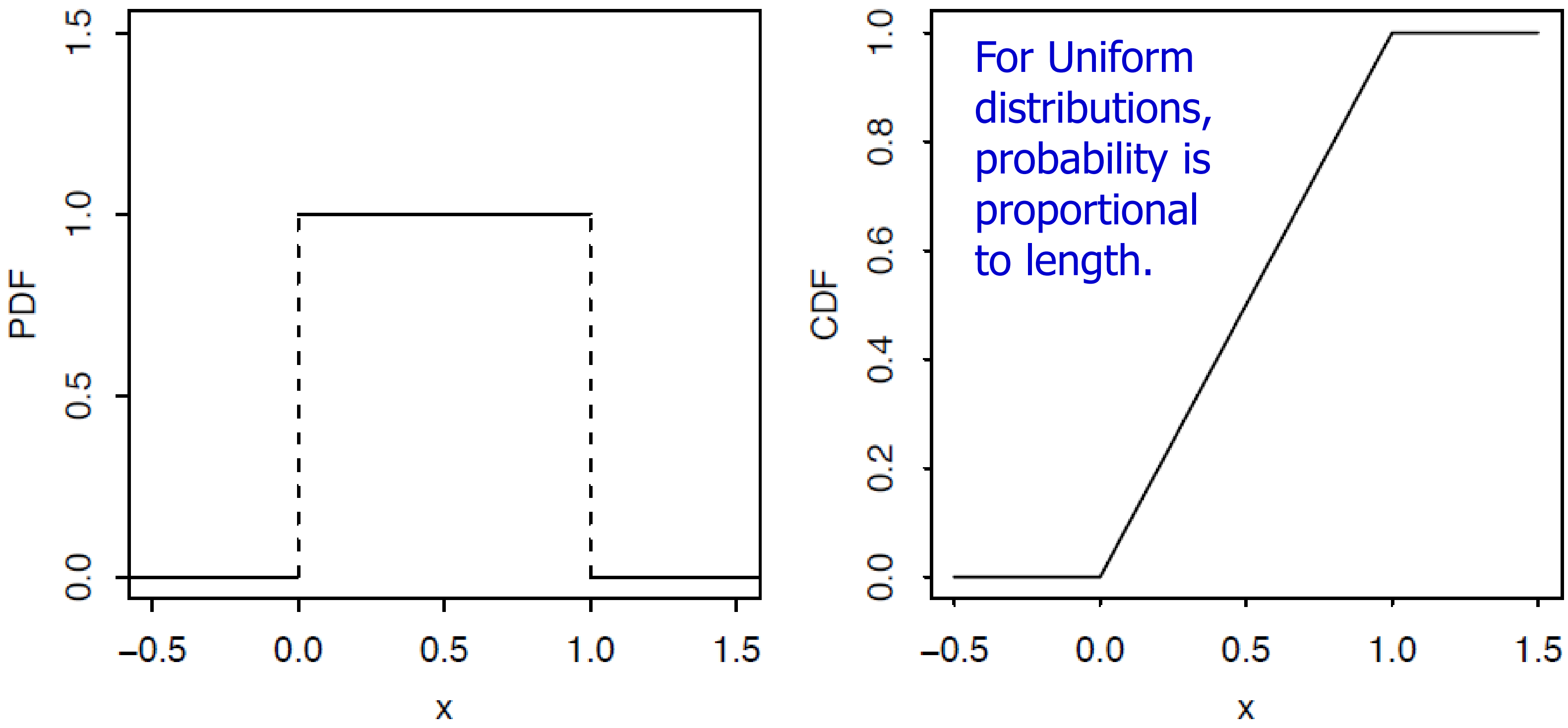
- The CDF is the accumulated area under the PDF:

$$F(x) = \begin{cases} 0 & \text{if } x \leq a, \\ \frac{x-a}{b-a} & \text{if } a < x < b, \\ 1 & \text{if } x \geq b. \end{cases}$$

Examples of PDFs

- We denote this by X ~ Unif(a, b).
- The Uniform distribution that we will most frequently use is the Unif(0, 1) distribution, also called the standard Uniform.
- The Unif(0, 1) PDF and CDF are particularly simple: $f(x) = 1$ and $F(x) = x$ for $0 < x < 1$.
- For a general Unif(a, b) distribution, the PDF is constant on (a, b), and the CDF is ramp-shaped, increasing linearly from 0 to 1 as x ranges from a to b.

Examples of PDFs



Unif(0, 1) PDF and CDF

PDF Properties

Summary of PDF Properties

Let X be a continuous random variable with PDF f_X.

- $f_X(x) \geq 0$ for all x.
- $\int_{-\infty}^{\infty} f_X(x)\,dx = 1$.
- If δ is very small, then $\mathbf{P}\big([x, x+\delta]\big) \approx f_X(x) \cdot \delta$.
- For any subset B of the real line,

$$\mathbf{P}(X \in B) = \int_B f_X(x)\,dx.$$

Expected Value and Variance of a Continuous r.v.

- The **expected value** or **expectation** or **mean** of a continuous r.v. X is defined by:

$$\mathbf{E}[X] = \int_{-\infty}^{\infty} x f_X(x)\, dx$$

- This sis similar to the discrete case except that the PMF is replaced by the PDF, and summation is replaced by integration.
- Its mathematical properties are similar to the discrete case.

Expected Value and Variance of a Continuous r.v.

- If X is a continuous random variable with given PDF, then any real-valued function Y = g(X) of X is also a random variable.
  - Note that Y can be a continuous r.v., but Y can also be discrete, e.g., $g(x) = 1$ for $x > 0$ and $g(x) = 0$, otherwise.
- In either case, the mean of g(X) satisfies the expected value rule:

$$\mathbf{E}[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x)\, dx$$

Expected Value and Variance of a Continuous r.v.

- The nth **moment** of a continuous r.v. X is defined as E[X^n], the expected value of the random variable X^n.
- The variance of X denoted as var(X), is defined as the expected value of the random variable (X - E[X^n])^2 :

$$\operatorname{var}(X) = \mathbf{E}\left[\left(X - \mathbf{E}[X]\right)^2\right] = \int_{-\infty}^{\infty} \left(x - \mathbf{E}[X]\right)^2 f_X(x)\, dx$$

$$0 \le \operatorname{var}(X) = \mathbf{E}[X^2] - \left(\mathbf{E}[X]\right)^2$$

If $Y = aX + b$, where a and b are given scalars, then

$$\mathbf{E}[Y] = a\mathbf{E}[X] + b, \qquad \operatorname{var}(Y) = a^2 \operatorname{var}(X)$$

Expected Value and Variance of a Continuous r.v.

- Example. Consider a uniform PDF over an interval [a, b], its expectation is given by:

$$\begin{aligned} \mathbf{E}[X] &= \int_{-\infty}^{\infty} x f_X(x)\,dx = \int_a^b x \cdot \frac{1}{b-a}\,dx \\ &= \frac{1}{b-a} \cdot \frac{1}{2} x^2 \Big|_a^b \\ &= \frac{1}{b-a} \cdot \frac{b^2 - a^2}{2} \\ &= \frac{a+b}{2}, \end{aligned}$$

Expected Value and Variance of a Continuous r.v.

- Its variance is given as:

$$\mathbf{E}[X^2] = \int_a^b \frac{x^2}{b-a}\,dx = \frac{1}{b-a}\int_a^b x^2\,dx$$

$$= \frac{1}{b-a}\cdot\frac{1}{3}x^3\Big|_a^b = \frac{b^3-a^3}{3(b-a)}$$

$$= \frac{a^2+ab+b^2}{3}.$$

$$\text{var}(X) = \mathbf{E}[X^2] - \big(\mathbf{E}[X]\big)^2 = \frac{a^2+ab+b^2}{3} - \frac{(a+b)^2}{4} = \frac{(b-a)^2}{12}$$

Expected Value and Variance of a Continuous r.v.

- The **exponential** continuous random variable has PDF:

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

where λ is a positive parameter characterising the PDF, with

$$\int_{-\infty}^{\infty} f_X(x)\,dx = \int_0^{\infty} \lambda e^{-\lambda x}\,dx = -e^{-\lambda x}\Big|_0^{\infty} = 1$$

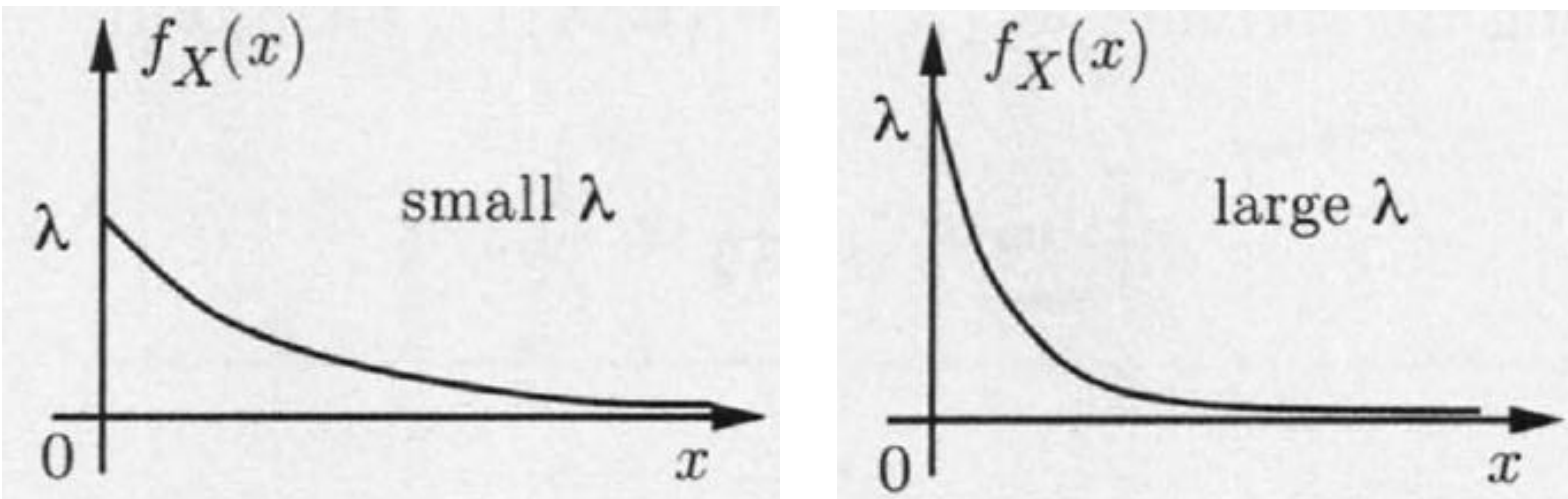
Expected Value and Variance of a Continuous r.v.

- The probability that X exceeds a certain value decreases exponentially. This is, for any $a \geq 0$, we have:

$$\mathbf{P}(X \geq a) = \int_a^\infty \lambda e^{-\lambda x}\, dx = -e^{-\lambda x}\Big|_a^\infty = e^{-\lambda a}$$

- An exponential random variable can be a good model for the amount of time until an incident of interest takes place.
  - a message arriving at a computer, some equipment breaking down, a light bulb burning out, etc.

Expected Value and Variance of a Continuous r.v.



The PDF $\lambda e^{-\lambda x}$ of an exponential random variable.

Expected Value and Variance of a Continuous r.v.

- The mean of the exponential r.v. X is calculated by:

$$
\begin{aligned}
\mathbf{E}[X] &= \int_0^\infty x\lambda e^{-\lambda x}\,dx \\
&= (-xe^{-\lambda x})\Big|_0^\infty + \int_0^\infty e^{-\lambda x}\,dx \\
&= 0 - \frac{e^{-\lambda x}}{\lambda}\Big|_0^\infty \\
&= \frac{1}{\lambda}.
\end{aligned}
$$

Expected Value and Variance of a Continuous r.v.

- The variance of the exponential r.v. X is calculated by:

$$\begin{aligned}\mathbf{E}[X^2] &= \int_0^\infty x^2 \lambda e^{-\lambda x}\, dx \\ &= \left(-x^2 e^{-\lambda x}\right)\Big|_0^\infty + \int_0^\infty 2x e^{-\lambda x}\, dx \\ &= 0 + \frac{2}{\lambda}\mathbf{E}[X] \\ &= \frac{2}{\lambda^2}.\end{aligned}$$

$$\operatorname{var}(X) = \mathbf{E}[X^2] - \big(\mathbf{E}[X]\big)^2 \quad \Rightarrow \quad \operatorname{var}(X) = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$$

Cumulative Distribution Functions

- The cumulative distribution function, CDF, of a random variable X is denoted as F_X and provides the probability P(X ≤ *x*). In particular for every *x* we have:

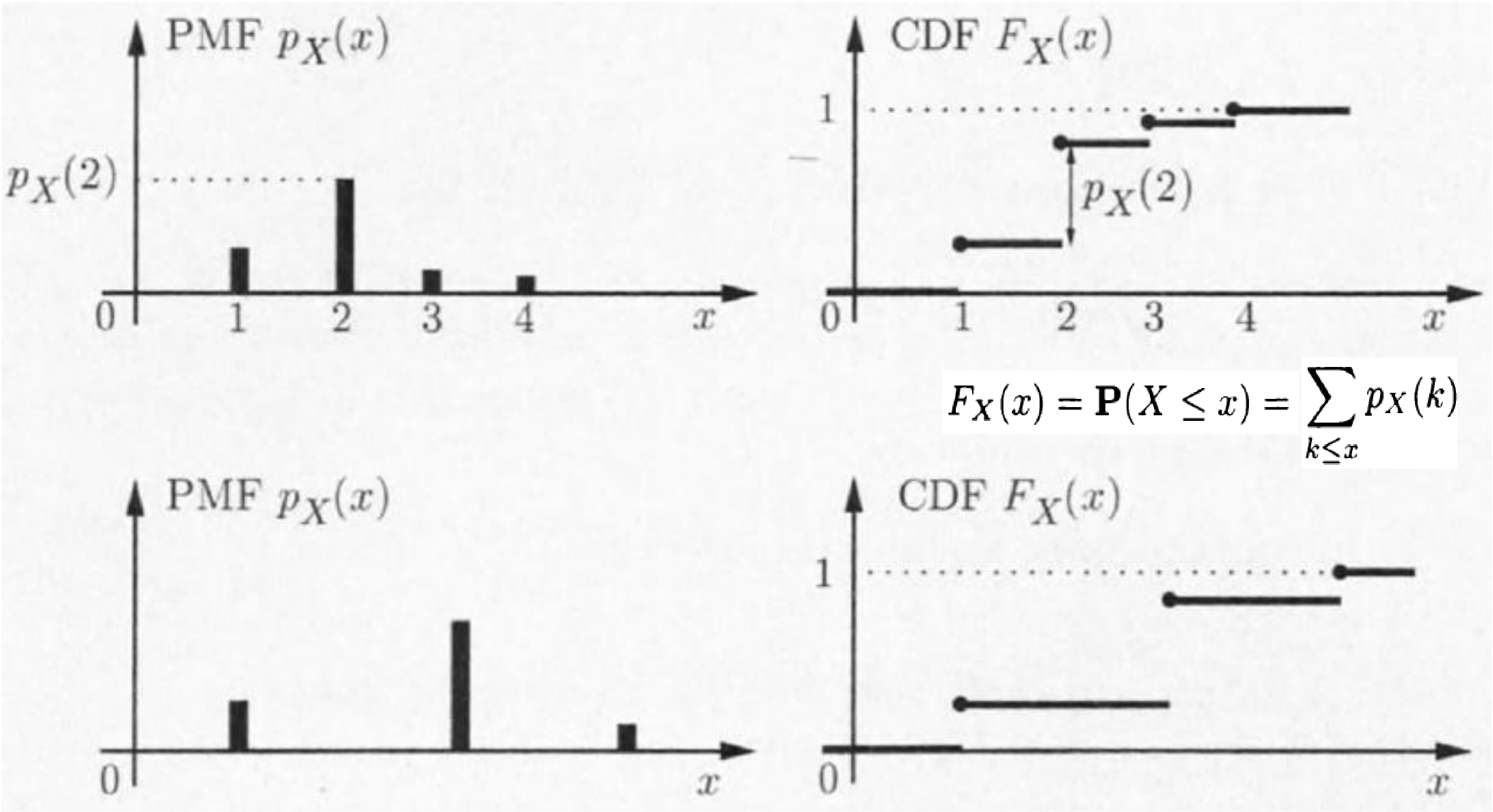
$$F_X(x) = \mathbf{P}(X \leq x) = \begin{cases} \sum_{k \leq x} p_X(k), & \text{if } X \text{ is discrete,} \\ \int_{-\infty}^{x} f_X(t)\,dt, & \text{if } X \text{ is continuous.} \end{cases}$$

The CDF $F_X(x)$ “accumulates” probability “up to” the value of *x*.

Cumulative Distribution Functions

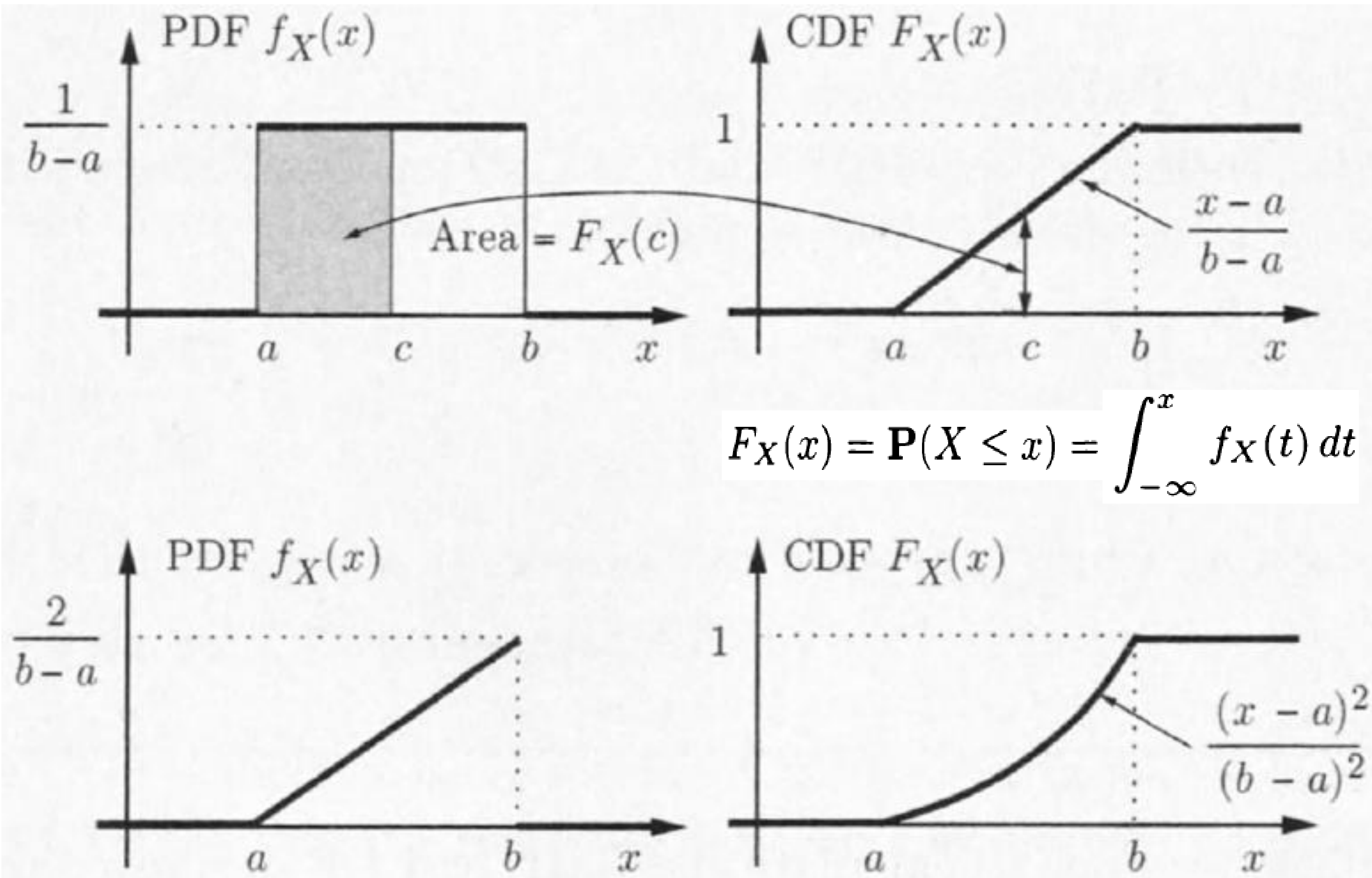
- Any random variable associated with a given probability model has CDF, regardless of whether it is discrete or continuous.
  - $\{X \leq x\}$ is always an event and therefore has well-defined probability.

Cumulative Distribution Functions



$$F_X(x) = \mathbf{P}(X \leq x) = \sum_{k \leq x} p_X(k)$$

Cumulative Distribution Functions



$$F_X(x) = \mathbf{P}(X \leq x) = \int_{-\infty}^{x} f_X(t)\, dt$$

Cumulative Distribution Functions

Properties of a CDF

The CDF F_X of a random variable X is defined by

$$F_X(x) = \mathbf{P}(X \leq x), \qquad \text{for all } x,$$

and has the following properties.

- F_X is monotonically nondecreasing:

$$\text{if } x \leq y, \text{ then } F_X(x) \leq F_X(y).$$

- $F_X(x)$ tends to 0 as $x \to -\infty$, and to 1 as $x \to \infty$.
- If X is discrete, then $F_X(x)$ is a piecewise constant function of x.
- If X is continuous, then $F_X(x)$ is a continuous function of x.

Cumulative Distribution Functions

- If X is discrete and takes integer values, the PMF and the CDF can be obtained from each other by summing or differencing:

$$F_X(k) = \sum_{i=-\infty}^{k} p_X(i),$$

$$p_X(k) = \mathbf{P}(X \leq k) - \mathbf{P}(X \leq k-1) = F_X(k) - F_X(k-1),$$

for all integers k.

- If X is continuous, the PDF and the CDF can be obtained from each other by integration or differentiation:

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\,dt, \qquad f_X(x) = \frac{dF_X}{dx}(x).$$

(The second equality is valid for those x at which the PDF is continuous.)

Normal Random Variables

- A continuous random variable X is **normal** or **Gaussian** or **normally distributed** if it has PDF of the form:

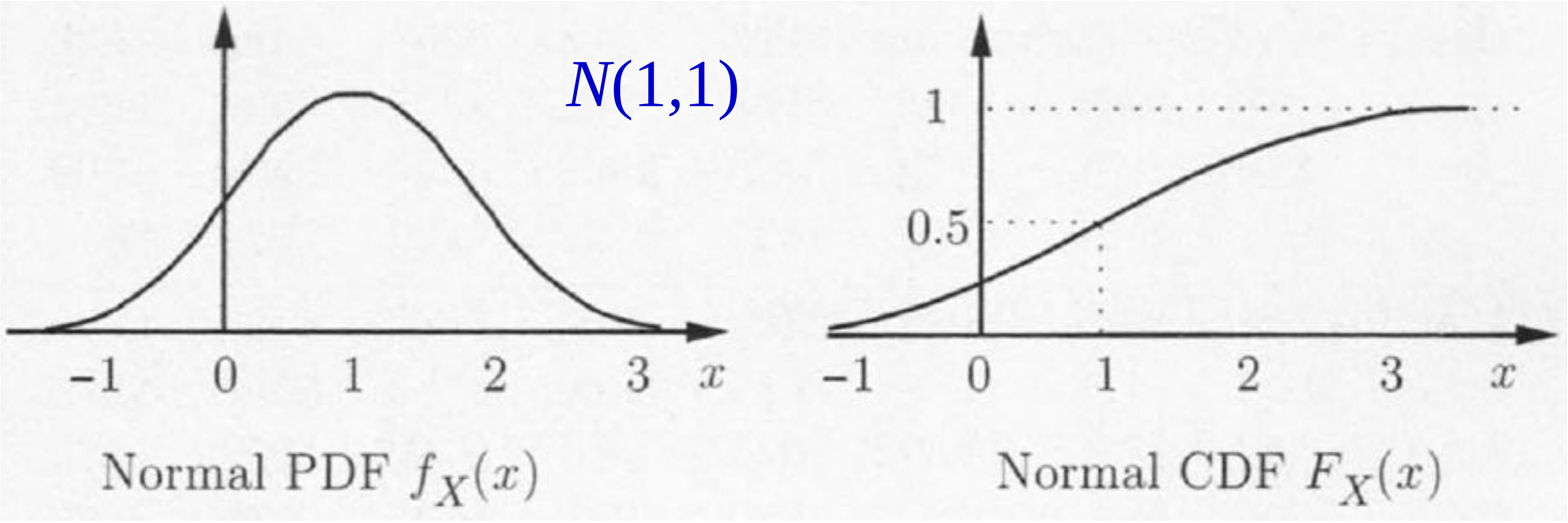
$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2} \quad \text{for } -\infty < x < \infty$$

where μ and σ are two scalar parameters characterising the PDF (abbreviated $N(\mu, \sigma^2)$, and referred to as normal density function), with σ assumed positive.

Normal Random Variables

- It can be verified that the normalisation property holds:

$$\frac{1}{\sqrt{2\pi}\,\sigma}\int_{-\infty}^{\infty} e^{-(x-\mu)^2/2\sigma^2}\,dx = 1$$



Normal Random Variables

- If X is $N(\mu, \sigma^2)$, then: $E(X) = \mu$

Proof: The PDF is symmetric about $x = \mu$.

- If X is $N(\mu, \sigma^2)$, then: $\mathrm{Var}(X) = \sigma^2$

Proof:

$$\mathrm{Var}(X) = E((X-\mu)^2) = \int_{-\infty}^{\infty} (x-\mu)^2 \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2} dx$$

Substituting $u = (x-\mu)/\sigma$ and integrating by parts, we get

$$\mathrm{Var}(X) = \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} u^2 e^{-u^2/2} du = \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} u \cdot u e^{-u^2/2} du$$

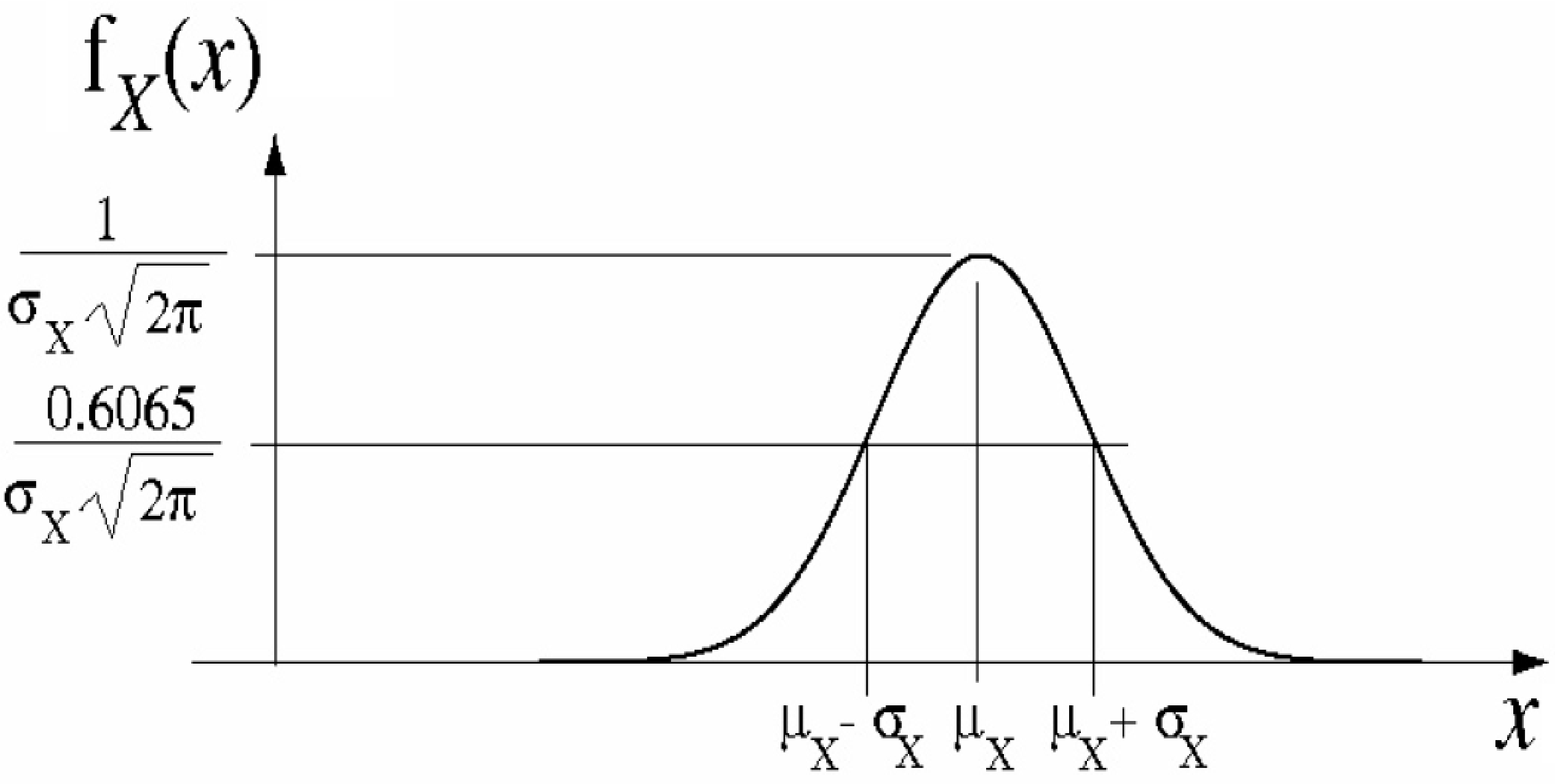
$$= \frac{\sigma^2}{\sqrt{2\pi}} \left(-\left[u e^{-u^2/2} \right]_{-\infty}^{\infty} + \int_{-\infty}^{\infty} e^{-u^2/2} du \right) = \frac{\sigma^2}{\sqrt{2\pi}} \left(0 + \sqrt{2\pi} \right) = \sigma^2$$

Normal Random Variables

- Its maximum value occurs at the mean value of its argument.
- It is symmetrical about the mean value.
- The points of maximum absolute slope occur at one standard deviation above and below the mean.
- Its maximum value is inversely proportional to its standard deviation.
- The limit as the standard deviation approaches zero is a unit impulse.

$$\delta(x-\mu_x) = \lim_{\sigma_X \to 0} \frac{1}{\sigma_X \sqrt{2\pi}} e^{-(x-\mu_X)^2/2\sigma_X^2}$$

Normal Random Variables



Linear Function of a Normal Random Variable

- If X is a normal r.v. with mean μ and variance σ^2, and if $a \neq 0$, b are scalars, then the random variable:

$$Y = aX + b$$

is also normal, with mean and variance:

$$E[Y] = a\mu + b, \qquad \text{var}(Y) = a^2\sigma^2$$

Standard Normal Random Variables

- A normal random variable Y with zero mean and unit variance, $N(0, 1)$, is said to be a standard normal. Its PDF and CDF are denoted by φ and Φ, respectively:

$$\varphi(y) = \frac{1}{\sqrt{2\pi}} e^{-y^2/2}, \quad -\infty < y < \infty$$

$$\Phi(y) = P(Y \le y) = P(Y < y) = \int_{-\infty}^{y} \varphi(t)dt = \int_{-\infty}^{y} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$$

Standard Normal Random Variables

- The PDF of a normal r.v. cannot be integrated in terms of the common elementary functions, and therefore the probabilities of X falling in various intervals are obtained from tables or by computer.
- Example, the Standard Normal Table.
- The table only provides the values of $\Phi(y)$ for $y \geq 0$, because the omitted values can be calculated using the symmetry of the PDF.

$$\Phi(-y) = 1 - \Phi(y), \qquad \text{for all } y.$$

Standard Normal Random Variables

The standard normal table. The entries in this table provide the numerical values of $\Phi(y) = \mathbf{P}(Y \leq y)$, where Y is a standard normal random variable, for y between 0 and 3.49. For example, to find $\Phi(1.71)$, we look at the row corresponding to 1.7 and the column corresponding to 0.01, so that $\Phi(1.71) = .9564$. When y is negative, the value of $\Phi(y)$ can be found using the formula $\Phi(y) = 1 - \Phi(-y)$.

	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319

Standard Normal Random Variables

1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

Standard Normal Random Variables

- It would be overwhelming to construct tables for all μ and σ values required in application.
  - Standardise the r.v.
- Let X be a normal (Gaussian) random variable with mean μ and variance σ^2 values. We standardise X by defining a new random variable Y given by:

$$Y = \frac{X - \mu}{\sigma}$$

Standard Normal Random Variables

- Since Y is a linear function of X, it is normal, This means:

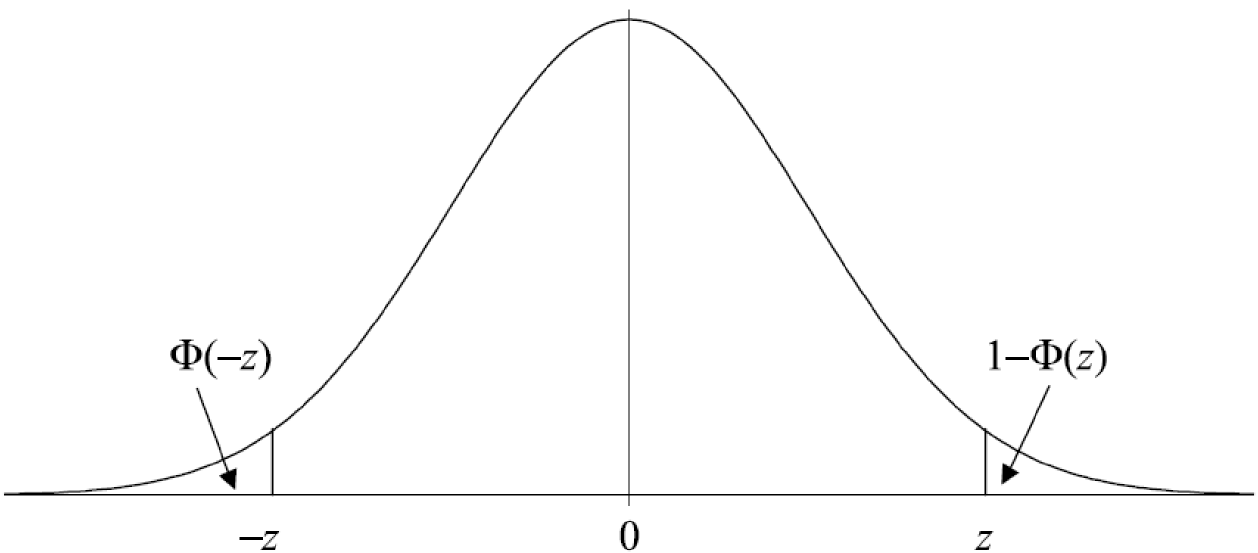
$$\mathbf{E}[Y] = \frac{\mathbf{E}[X] - \mu}{\sigma} = 0 \qquad \text{var}(Y) = \frac{\text{var}(X)}{\sigma^2} = 1$$

- Thus, Y is a standard normal random variable.
  - This allows us to calculate the probability of any event defined in terms of X by redefining the event in terms of Y, and then using the standard normal table.

Standard Normal Random Variables

- Example 1:

$$\Phi(-0.5) = \mathbf{P}(Y \leq -0.5) = \mathbf{P}(Y \geq 0.5) = 1 - \mathbf{P}(Y < 0.5)$$
$$= 1 - \Phi(0.5) = 1 - .6915 = 0.3085.$$



Standard Normal Random Variables

- Example 2: The annual snowfall at a particular geographic location is modelled as a normal random variable with a mean $\mu = 60$ inches and a standard deviation of $\sigma = 20$. What is the probability that this year's snowfall will be at least 80 inches?

Standard Normal Random Variables

- Solution:

Let X be the snow accumulation, viewed as a normal random variable, and let

$$Y = \frac{X-\mu}{\sigma} = \frac{X-60}{20},$$

be the corresponding standard normal random variable. We have

$$\mathbf{P}(X \geq 80) = \mathbf{P}\left(\frac{X-60}{20} \geq \frac{80-60}{20}\right) = \mathbf{P}\left(Y \geq \frac{80-60}{20}\right) = \mathbf{P}(Y \geq 1) = 1-\Phi(1),$$

where Φ is the CDF of the standard normal. We read the value $\Phi(1)$ from the table:

$$\Phi(1) = 0.8413,$$

so that

$$\mathbf{P}(X \geq 80) = 1 - \Phi(1) = 0.1587.$$

Standard Normal Random Variables

- Example 3: (Height Distribution of Men). Assume that the height X, in inches, of a randomly selected man in a certain population is normally distributed with $\mu = 69$ and $\sigma = 2.6$. Find

1. $P(X < 72)$,
2. $P(X > 72)$,
3. $P(X < 66)$,
4. $P(|X - \mu| < 3)$.

Standard Normal Random Variables

- The table gives $\Phi(z)$ only for $z \geq 0$, and for $z < 0$ we need to make use of the symmetry of the normal distribution. This implies that, for any z, $P(Z < -z) = P(Z > z)$. Thus, solution:

1. $P(X < 72) = P((X-\mu)/\sigma < (72-69)/2.6) \approx P(Z < 1.15) = \Phi(1.15) \approx 0.875$.
2. $P(X > 72) = P((X-\mu)/\sigma > (72-69)/2.6) \approx P(Z > 1.15) = 1 - P(Z \leq 1.15) = 1 - \Phi(1.15) \approx 1 - 0.875 = 0.125$.
3. $P(X < 66) = P((X-\mu)/\sigma < (66-69)/2.6) \approx P(Z < -1.15) = P(Z > 1.15) = 0.125$.
4. $P(|X-\mu| < 3) = P(|(X-\mu)/\sigma| < 3/2.6) \approx P(|Z| < 1.15) = 1 - [P(Z < -1.15) + P(Z > 1.15)] = 2\Phi(1.15) - 1 \approx 0.75$.

Standard Normal Random Variables

CDF Calculation for a Normal Random Variable

For a normal random variable X with mean μ and variance σ^2, we use a two-step procedure.

(a) "Standardize" X, i.e., subtract μ and divide by σ to obtain a standard normal random variable Y.

(b) Read the CDF value from the standard normal table:

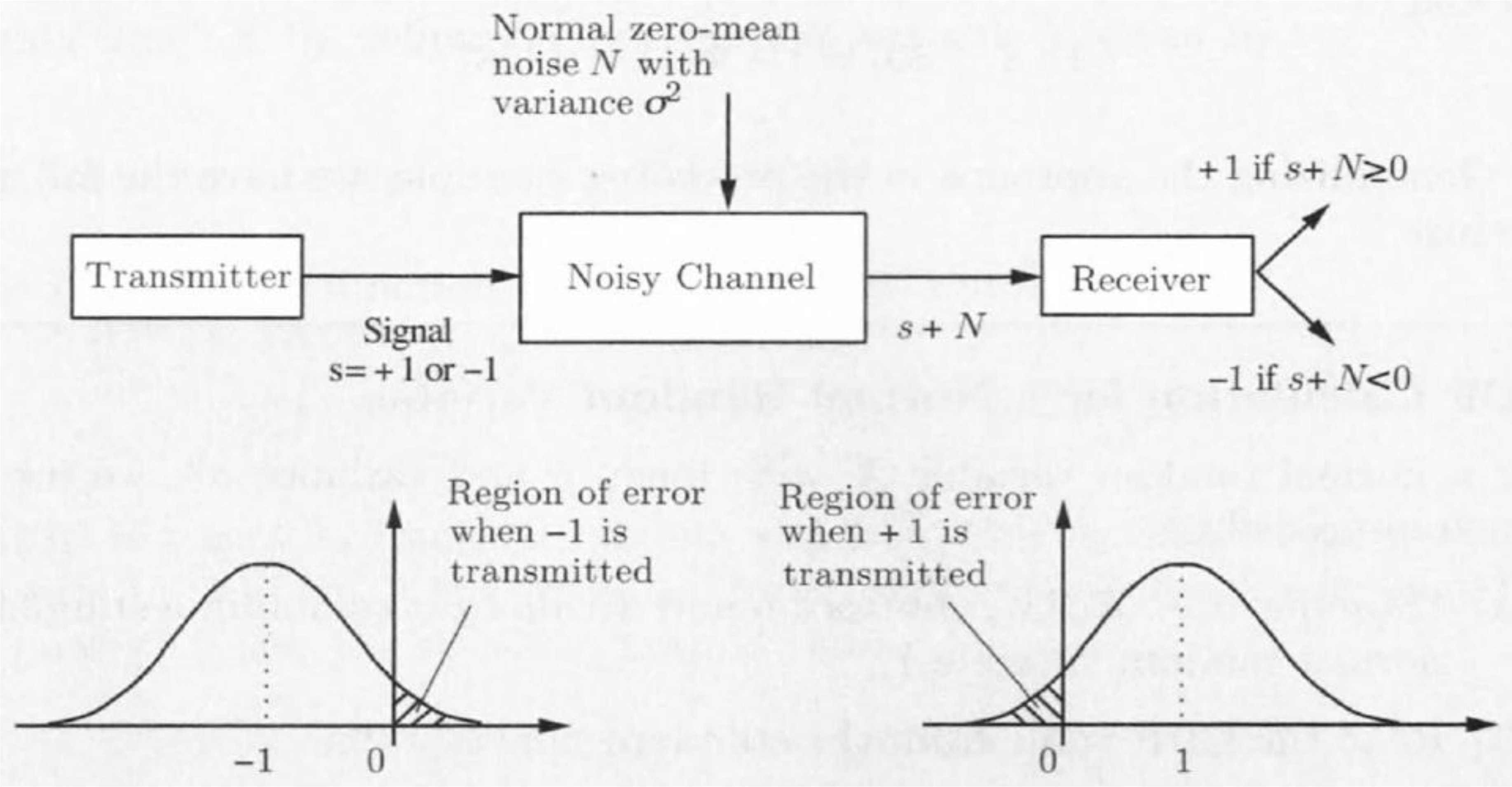
$$\mathbf{P}(X \le x) = \mathbf{P}\left(\frac{X-\mu}{\sigma} \le \frac{x-\mu}{\sigma}\right) = \mathbf{P}\left(Y \le \frac{x-\mu}{\sigma}\right) = \Phi\left(\frac{x-\mu}{\sigma}\right).$$

Standard Normal Random Variables

- Normal r.v.s. are often used in signal processing and communications engineering to model noise and unpredictable distortions of signals.
- Example:

Example 3.8. Signal Detection. A binary message is transmitted as a signal s, which is either -1 or $+1$. The communication channel corrupts the transmission with additive normal noise with mean $\mu = 0$ and variance σ^2. The receiver concludes that the signal -1 (or $+1$) was transmitted if the value received is < 0 (or ≥ 0, respectively); see Fig. 3.11. What is the probability of error?

Standard Normal Random Variables



Standard Normal Random Variables

- Solution:

An error occurs whenever -1 is transmitted and the noise N is at least 1 so that $s+N = -1+N \geq 0$, or whenever $+1$ is transmitted and the noise N is smaller than -1 so that $s+N = 1+N < 0$. In the former case, the probability of error is

$$\mathbf{P}(N \geq 1) = 1 - \mathbf{P}(N < 1) = 1 - \mathbf{P}\left(\frac{N-\mu}{\sigma} < \frac{1-\mu}{\sigma}\right)$$
$$= 1 - \Phi\left(\frac{1-\mu}{\sigma}\right) = 1 - \Phi\left(\frac{1}{\sigma}\right).$$

In the latter case, the probability of error is the same, by symmetry. The value of $\Phi(1/\sigma)$ can be obtained from the normal table. For $\sigma = 1$, we have $\Phi(1/\sigma) = \Phi(1) = 0.8413$, and the probability of error is 0.1587.

Standard Normal Random Variables

- Three important benchmarks for the Normal distribution are the probabilities of falling within one, two, and three standard deviations of the mean. The 68-95-99.7% rule tells us that these probabilities are what the name suggests.
- (68-95-99.7% rule). If X ~ N(μ, σ^2), then:

$$P(|X-\mu|<\sigma)\approx 0.68$$
$$P(|X-\mu|<2\sigma)\approx 0.95$$
$$P(|X-\mu|<3\sigma)\approx 0.997$$

Standardising →

$$P(|Z|<1)\approx 0.68$$
$$P(|Z|<2)\approx 0.95$$
$$P(|Z|<3)\approx 0.997$$

Standard Normal Random Variables

- Three important benchmarks for the Normal distribution are the probabilities of falling within one, two, and three standard deviations of the mean. The 68-95-99.7% rule tells us that these probabilities are what the name suggests.
- (68-95-99.7% rule). If X ~ N(μ, σ^2), then:

$$P(|X-\mu|<\sigma)\approx 0.68$$
$$P(|X-\mu|<2\sigma)\approx 0.95$$
$$P(|X-\mu|<3\sigma)\approx 0.997$$

Standardising →

$$P(|Z|<1)\approx 0.68$$
$$P(|Z|<2)\approx 0.95$$
$$P(|Z|<3)\approx 0.997$$

Joint PDF of Multiple Random Variables

- Two continuous random variables associated with the same experiment are **jointly continuous** and can be described in terms of a **joint PDF** $f_{X,Y}$ if $f_{X,Y}$ is a nonnegative function that satisfies:

$$\mathbf{P}((X,Y) \in B) = \iint_{(x,y)\in B} f_{X,Y}(x,y)\,dx\,dy$$

for every subset B of the two-dimensional plane.

- The notation means that the integration is carried out over the set B.

Joint PDF of Multiple Random Variables

- In the particular case where B is a rectangle of the form B = $\{(x, y) \mid a \leq x \leq b, c \leq y \leq d\}$, we have:

$$\mathbf{P}(a \leq X \leq b,\, c \leq Y \leq d) = \int_c^d \int_a^b f_{X,Y}(x, y)\, dx\, dy$$

- If B is the entire two-dimensional plane, then we obtain the normalisation property:

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dx\, dy = 1$$

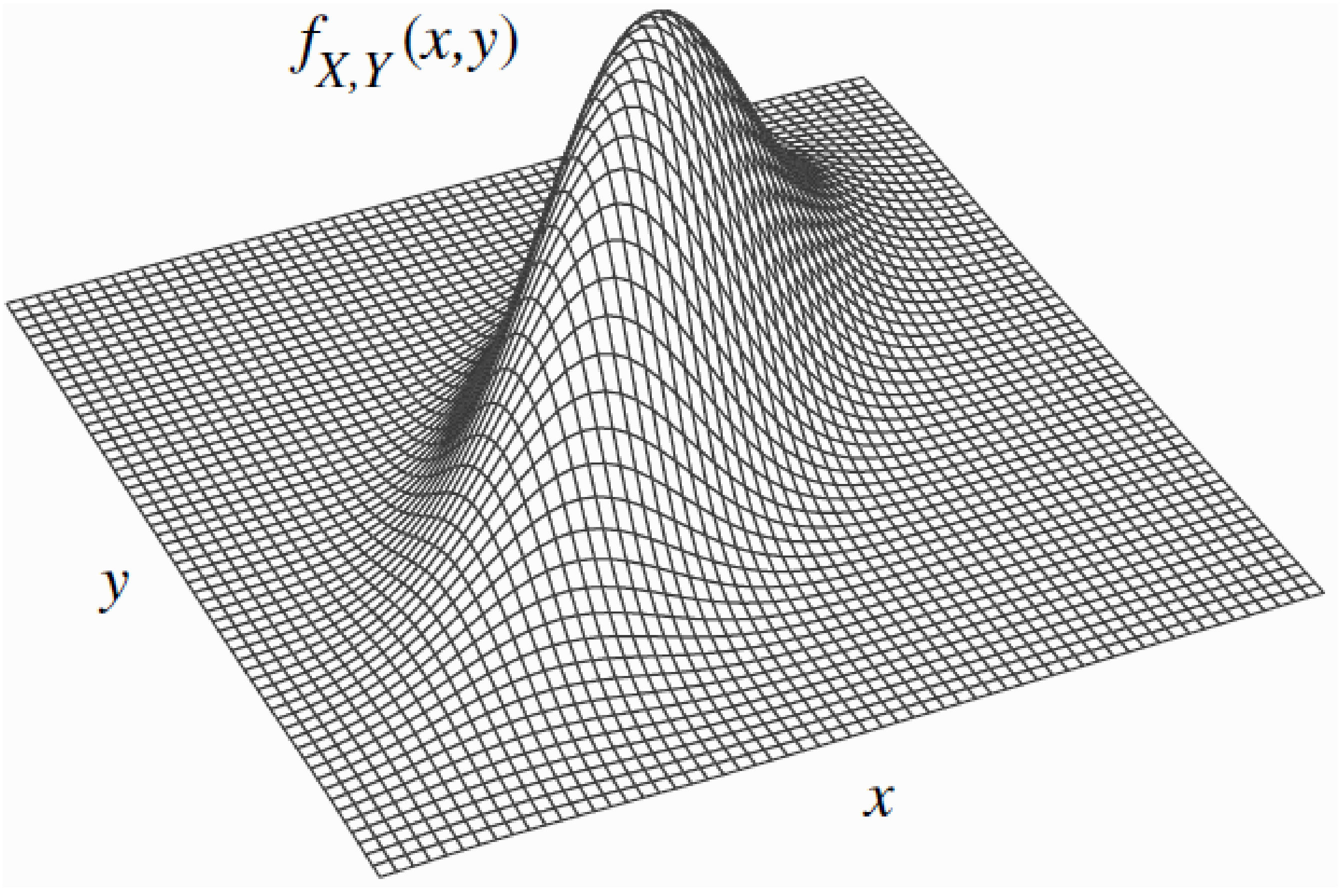
Joint PDF of Multiple Random Variables

- To interpret the joint PDF, we let δ be a small positive number and consider the probability of a small rectangle. Then we have:

$$\mathbf{P}(a \le X \le a+\delta,\, c \le Y \le c+\delta) = \int_c^{c+\delta} \int_a^{a+\delta} f_{X,Y}(x,y)\,dx\,dy \approx f_{X,Y}(a,c) \cdot \delta^2$$

so we can view $f_{X,Y}(a, c)$ as the **probability per unit area** in the vicinity of (a, c).

Joint PDF of Multiple Random Variables



Joint PDF of Multiple Random Variables

- The joint PDF contains all relevant probabilistic information on the random variables X, Y, and their dependencies.
- Therefore, the joint PDF allow us to calculate the probability of any event that can be defined in terms of these two random variables.

Marginals

- Marginal PDF. For continuous r.v.s X and Y with joint PDF $f_{X,Y}$, the **marginal** PDF of X is:

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dy$$

- Similarly, the marginal PDF of Y is:

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dx$$

Marginals

- Marginalisation works analogously with any number of variables. For example, if we have the joint PDF of X, Y, Z,W but want the joint PDF of X,W, we just have to integrate over all possible values of Y and Z:

$$f_{X,W}(x,w) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y,Z,W}(x,y,z,w)dydz$$

  - Conceptually this is very easy—just integrate over the unwanted variables to get the joint PDF of the wanted variables—but computing the integral may or may not be difficult.

Marginals

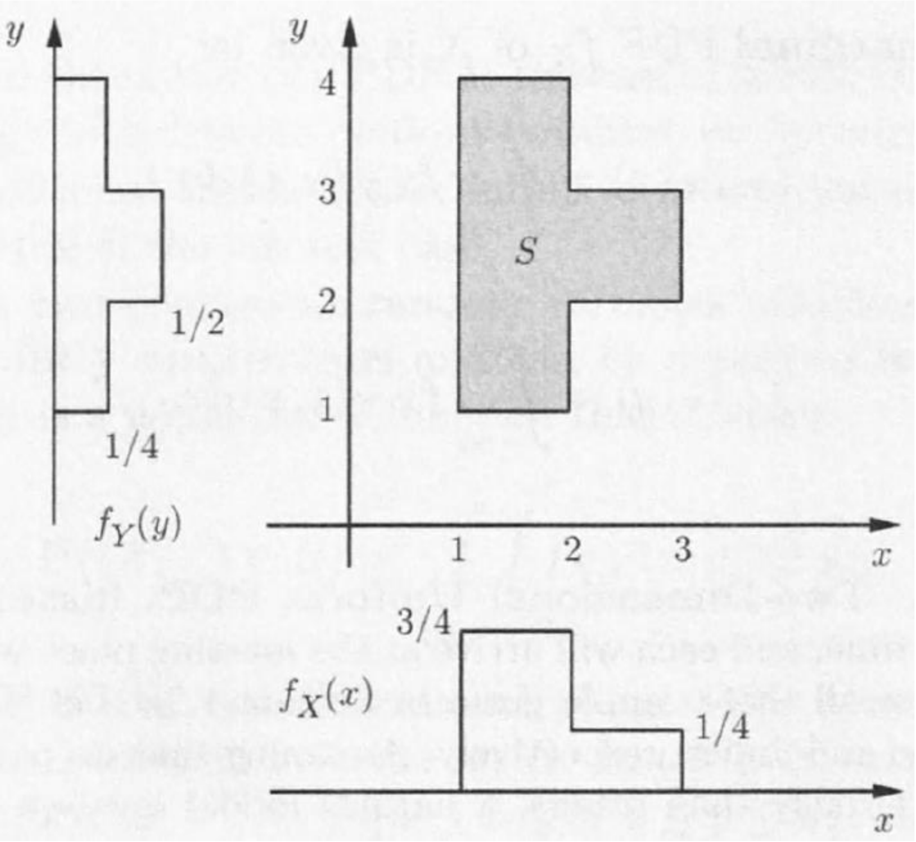
- Example 1.

Example 3.10. We are told that the joint PDF of the random variables X and Y is a constant c on the set S shown in Fig. 3.12 and is zero outside. We wish to determine the value of c and the marginal PDFs of X and Y.

The area of the set S is equal to 4 and, therefore, $f_{X,Y}(x, y) = c = 1/4$, for $(x, y) \in S$. To find the marginal PDF $f_X(x)$ for some particular x, we integrate (with respect to y) the joint PDF over the vertical line corresponding to that x. The resulting PDF is shown in the figure. We can compute f_Y similarly.

Marginals

- Example 1.



The joint PDF in Example 3.10 and the resulting marginal PDFs.

Joint CDFs

- If X and Y are two random variables associated with the same experiment, their **joint CDF** is defined by:

$$F_{X,Y}(x, y) = \mathbf{P}(X \leq x, Y \leq y)$$

- If X and Y are described by a joint PDF $f_{X,Y}$, then:

$$F_{X,Y}(x, y) = \mathbf{P}(X \leq x, Y \leq y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(s, t)\, dt\, ds$$

Joint PDF of Multiple Random Variables

- Conversely, if X and Y are continuous with joint CDF $F_{X,Y}$ their **joint PDF** is the derivative of the joint CDF with respect to *x* and *y*:

$$f_{X,Y}(x,y) = \frac{\partial^2}{\partial x \partial y} F_{X,Y}(x,y)$$

Joint CDF of Multiple Random Variables

- Let X and Y be described by a uniform PDF on the unit square. The joint CDF is given by:

$$F_{X,Y}(x,y) = \mathbf{P}(X \leq x, Y \leq y) = xy, \qquad \text{for } 0 \leq x, y \leq 1$$

- It can be verified that:

$$\frac{\partial^2 F_{X,Y}}{\partial x \partial y}(x,y) = \frac{\partial^2 (xy)}{\partial x \partial y}(x,y) = 1 = f_{X,Y}(x,y)$$

for al (x, y) in the unit square.

Expectation

- If X and Y are jointly continuous random variables and *g* is some function, then Z = *g* (X, Y) is also a random variable. Thus the expected value rule applies:

$$\mathbf{E}[g(X,Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y) f_{X,Y}(x,y)\,dx\,dy$$

- As an important special case, for any scalars *a*, *b*, and *c*, we have:

$$\mathbf{E}[aX + bY + c] = a\mathbf{E}[X] + b\mathbf{E}[Y] + c$$

More than Two Random Variables

- The joint PDF of three random variables X, Y, and Z is defined in analogy with the case of two random variables. For example:

$$\mathbf{P}\big((X,Y,Z)\in B\big) = \iiint_{(x,y,z)\in B} f_{X,Y,Z}(x,y,z)\,dx\,dy\,dz$$

- For any set B. We have the relations such as:

$$f_{X,Y}(x,y) = \int_{-\infty}^{\infty} f_{X,Y,Z}(x,y,z)\,dz$$

$$f_X(x) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y,Z}(x,y,z)\,dy\,dz$$

More than Two Random Variables

- The expected value rule takes the form:

$$\mathbf{E}[g(X,Y,Z)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y,z) f_{X,Y,Z}(x,y,z)\,dx\,dy\,dz$$

- If *g* is linear, of the form *a*X +*b*Y + *c*Z, then:

$$\mathbf{E}[aX + bY + cZ] = a\mathbf{E}[X] + b\mathbf{E}[Y] + c\mathbf{E}[Z]$$

Furthermore, there are obvious generalizations of the above to the case of more than three random variables. For example, for any random variables $X_1, X_2, \ldots,$ X_n and any scalars $a_1, a_2, \ldots, a_n$, we have

$$\mathbf{E}[a_1X_1 + a_2X_2 + \cdots + a_nX_n] = a_1\mathbf{E}[X_1] + a_2\mathbf{E}[X_2] + \cdots + a_n\mathbf{E}[X_n].$$