Lecture Notes EMSE 4765:
DATA ANALYSIS - Probability Review

Chapter 5: Continuous Random Variables

Version: 01/14/2021

Text Book: A Modern Introduction to Probability and Statistics, Understanding Why and How

By: F.M. Dekking. C. Kraaikamp, H.P.Lopuhaä and L.E. Meester

5 Continuous Random Variables

5.1 Probability Density Functions and cumulative distribution function...

Continuous random variables as a refinement of discrete random variables.

Imagine that the outcome that a component fails in an interval $[a, b]$ equals p. Introducing the random variable function $X : \Omega \rightarrow [0, \infty)$ one defines:

$$X(\{\text{Component fails at time } x\}) = x, \text{ where } x \in [0, \infty).$$

Definition: **A random variable X is continuous** if for some function $f: \mathbb{R} \rightarrow \mathbb{R}$ and for any numbers a and b, such that $a \leq b$:

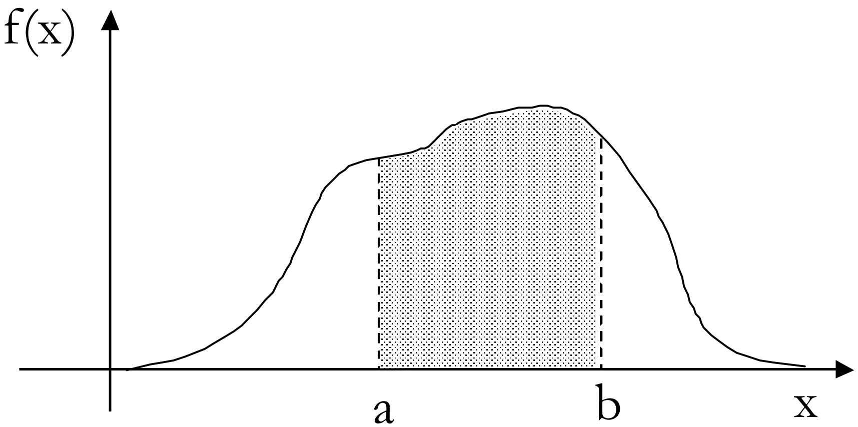
$$Pr(X \in [a, b]) = \int_a^b f(u)du, F(x) = Pr(X \leq x) = \int_{-\infty}^{x} f(u)du$$

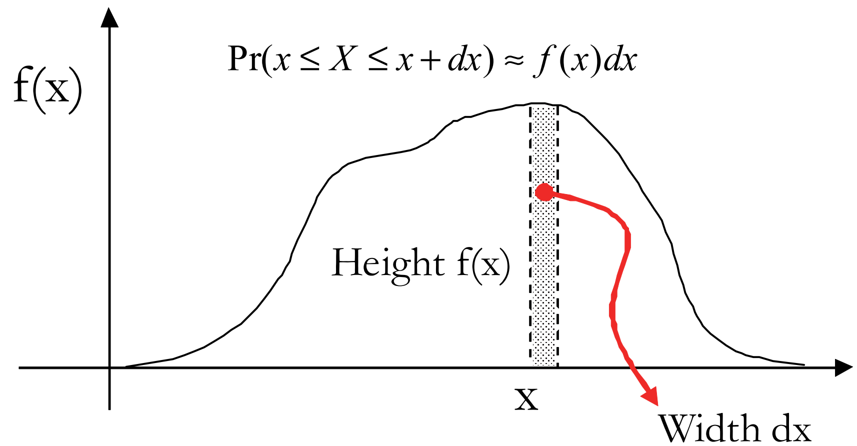
The function f has to satisfy $f(x) \geq 0$ for all x and $\int_{-\infty}^{\infty} f(x)dx = 1$. **We call $f(\cdot)$ the probability density function (pdf) of X. We call $F(x)$ the cumulative distribution function of X and $f(x) = \frac{d}{dx}F(x)$.**

5 Continuous Random Variables

5.1 Probability Density Functions. . .

As with discrete random variables one ommits **the wording "function"** and **the function argument ({Component fails at time x})** and one writes in short $X = x$, but with providing the definition "$X \equiv$ Failure time of the component".





Note that the probability that X lies in an interval $[a, b]$ is equal to **the area under the probability density function $f(\cdot)$ of X over the interval $[a, b]$**.

- **Interestingly for a continuous random variables *X*** we have:

$$Pr(a - \epsilon \leq X \leq a + \epsilon) \approx 2\epsilon \times f(a)$$

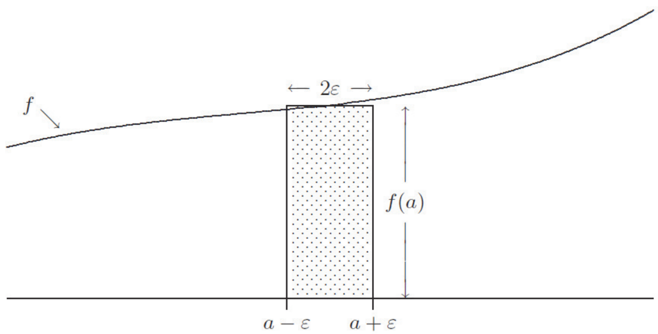


Fig. 5.2. Approximating the probability that X lies ε-close to a.

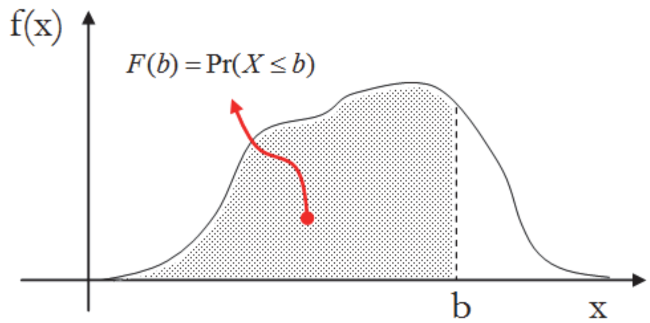
$$Pr(X = a) = \lim_{\epsilon \downarrow 0} \int_{a-\epsilon}^{a+\epsilon} f(x)dx = \lim_{\epsilon \downarrow 0} 2\epsilon \times f(a) = 0$$

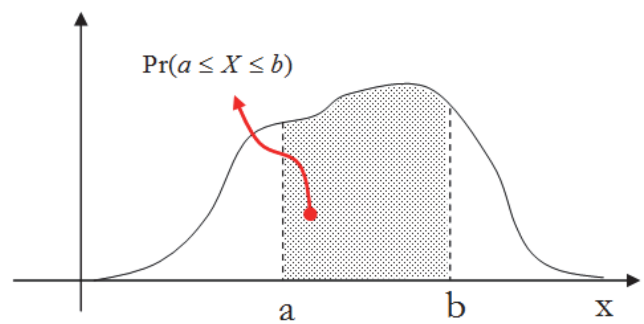
- **Thus for a continuous random variables *X*** one obtains:

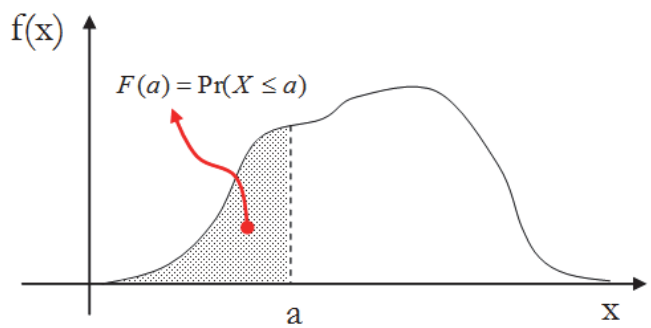
$$Pr(a < X < b) = Pr(a \leq X < b) = Pr(a < X \leq b) = Pr(a \leq X \leq b).$$

5 Continuous Random Variables

5.1 Cumulative Distribution Functions...







Conclusion:

$\Pr(a < X \leq b) = F(b) - F(a)$, where

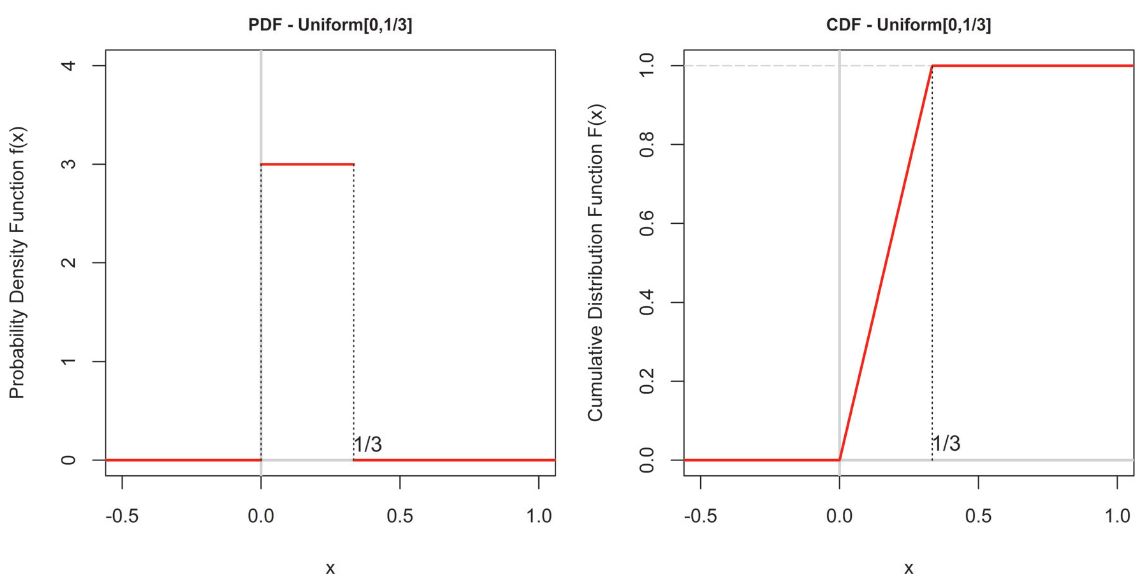
$F(b) = \int_{-\infty}^{b} f(x)dx$ and $\dfrac{dF(x)}{dx} = f(x)$

- For $a \leq b$ we have $\{X \leq b\} = \{X \leq a\} \cup \{a < X \leq b\} \Rightarrow$

$$Pr(X \leq b) = Pr(X \leq a) + Pr(a < X \leq b\} \Leftrightarrow$$
$$Pr(a < X \leq b\} = Pr(X \leq b) - Pr(X \leq a) = F(b) - F(a)$$

Definition: A continuous random variable X has **a uniform distribution** on the interval $[a, b]$ denoted $U(a, b)$ when :

$$f(x) = \begin{cases} 1/(b-a), & a \leq x \leq b, \\ 0, & \text{elsewhere.} \end{cases}$$



Definition: A continuous random variable X has **an exponential distribution** with parameter $\lambda > 0$ denoted $Exp(\lambda)$ when :

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0, \\ 0, & \text{elsewhere.} \end{cases}$$

The distribution function $F(\,\cdot\,)$ of an $Exp(\lambda)$ distribution is given by:

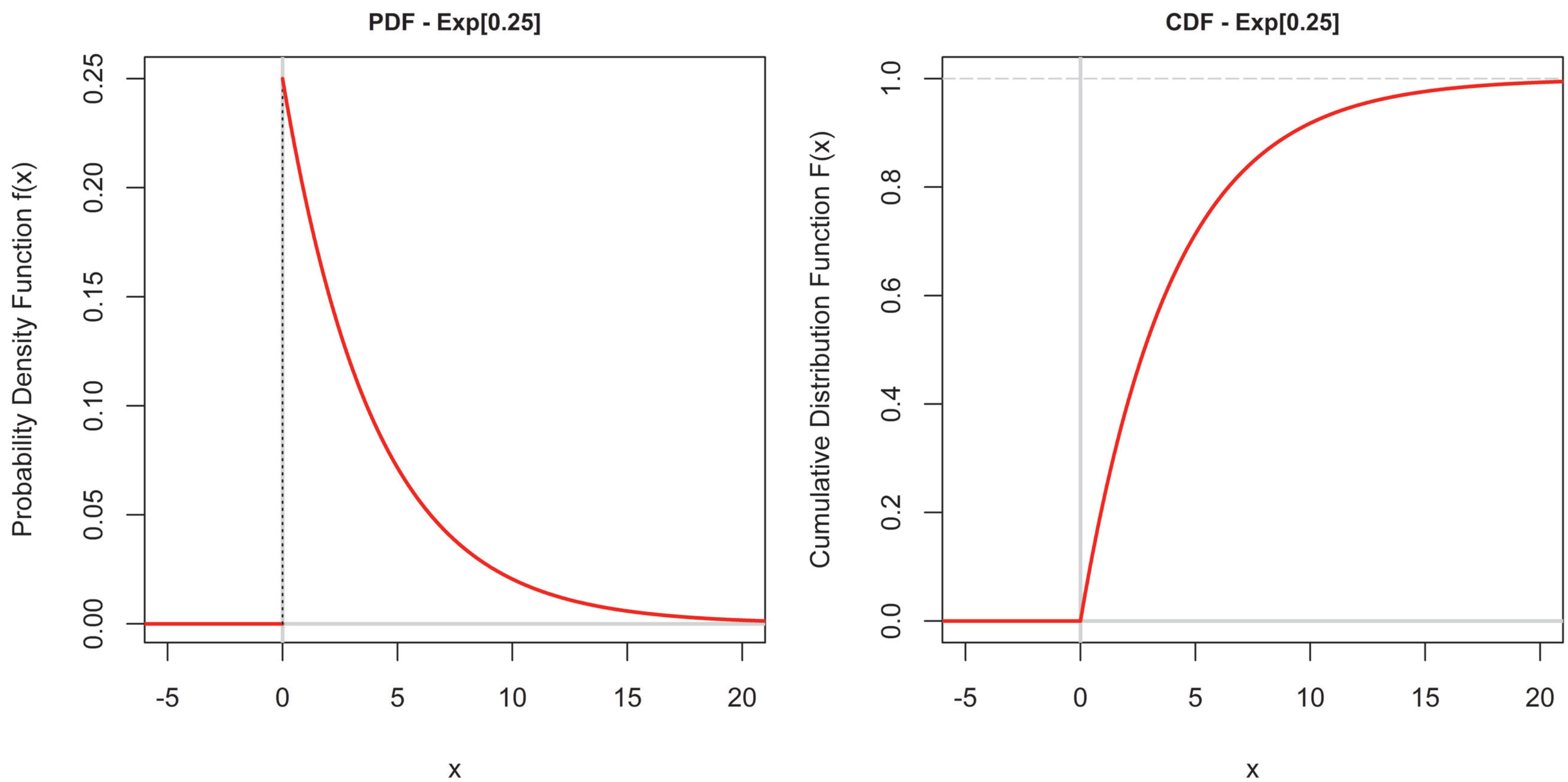
$$F(a) = Pr(X \leq a) = \int_0^a \lambda e^{-\lambda x} dx = \; - e^{-\lambda x}|_0^a = 1 - e^{-\lambda a}, \text{ for } a \geq 0.$$

The exponential distribution also **satisfies the memoryless property**, i.e.,

$$Pr(X > t + s | X > s) = \frac{Pr(X > t + s)}{Pr(X > s)} = \frac{e^{-\lambda(t+s)}}{e^{-\lambda s}} = e^{-\lambda t}, \text{ for } s, t > 0$$

5 Continuous Random Variables

5.3 The Exponential Distribution. . .



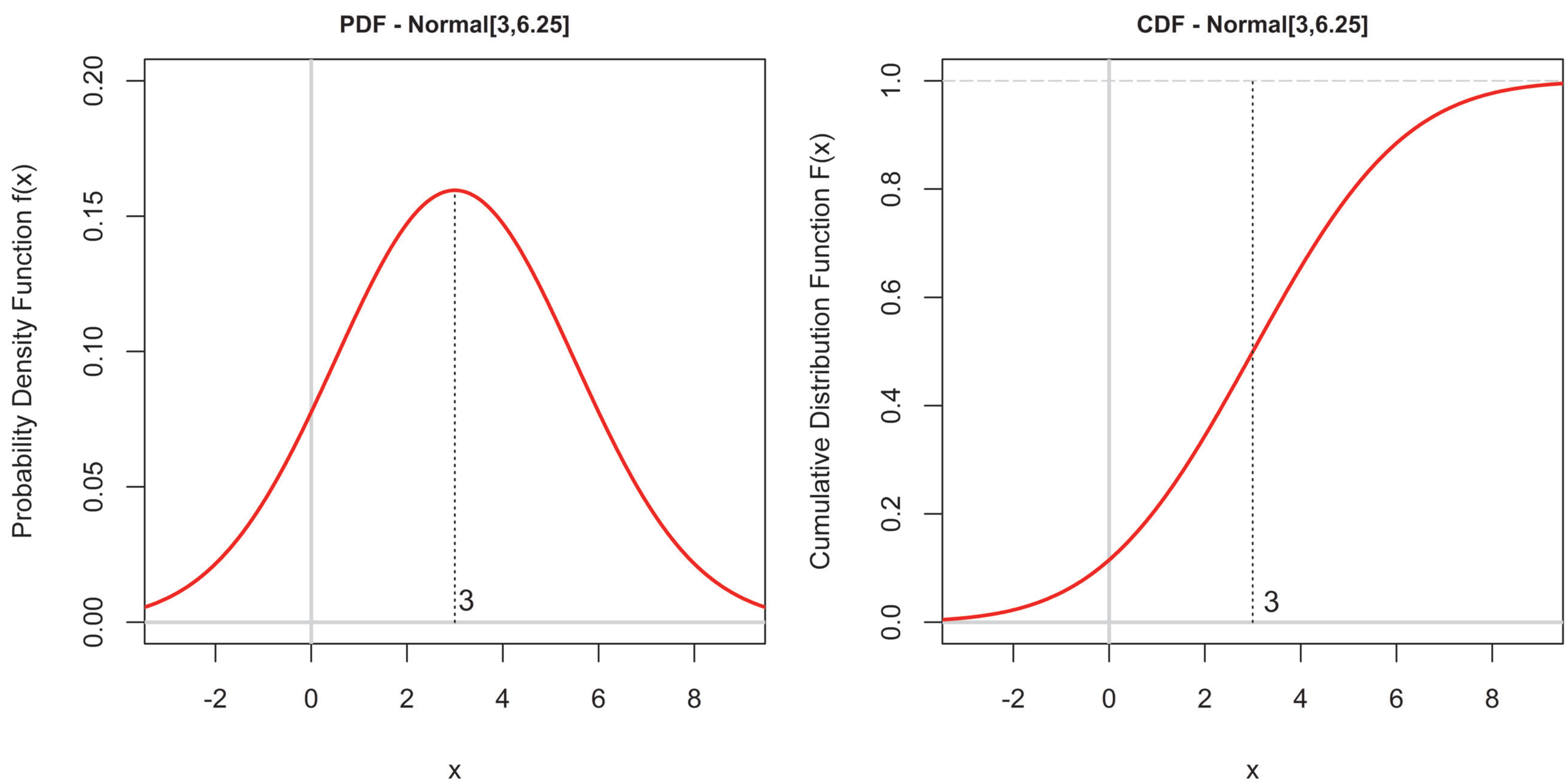
- **The normal distribution plays a central role in probability theory and statistics.** C.F. Gauss (1809) used it to model **observational errors in astronomy**; The normal distribution is an important tool **to approximate the probability distribution of the average of independent random variables**.

Definition: A continuous random variable X has **a normal distribution** with parameters μ and $\sigma^2 > 0$ denoted $N(\mu, \sigma^2)$ when :

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \text{ for } -\infty < x < \infty.$$

- **Cumulative distribution function (no closed form expression exists!):**

$$F(a) = Pr(X \leq a) = \int_{-\infty}^{a} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx \quad \text{for } -\infty < a < \infty$$



PDF - Normal[3,6.25]
Probability Density Function f(x)
0.00
0.05
0.10
0.15
0.20
3
-2
0
2
4
6
8
x
CDF - Normal[3,6.25]
Cumulative Distribution Function F(x)
0.0
0.2
0.4
0.6
0.8
1.0
3
-2
0
2
4
6
8
x

- **The Standard Normal Distribution:**

$$X \sim N(\mu, \sigma^2) \Leftrightarrow Z = \frac{X-\mu}{\sigma}, \; Z \sim N(0,1)$$

$$f_Z(z) = \phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} \text{ and } F_Z(a) = \Phi(a) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a} e^{-\frac{1}{2}z^2} dz$$

- **Symmetry of Standard Normal Distribution:**

$$\phi(z) = \phi(-z) \Rightarrow P(Z \leq -z) = P(Z \geq z) \Leftrightarrow \Phi(-z) = 1 - \Phi(z)$$

- **Table of Standard Normal Distribution** (Appendix B) contains the values of the right tail probabilities, i.e. $P(Z \geq z) = 1 - \Phi(z)$ for $z \geq 0$:

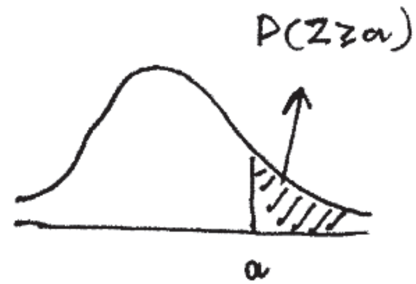
Example: $P(Z \leq 1) = 1 - P(Z \geq 1)$. In Appendix B we find that

$$P(Z \geq 1) = 0.1587 \Leftrightarrow P(Z \leq 1) = 1 - 0.1587 = 0.8413.$$

5.5 The Normal Distribution. . .

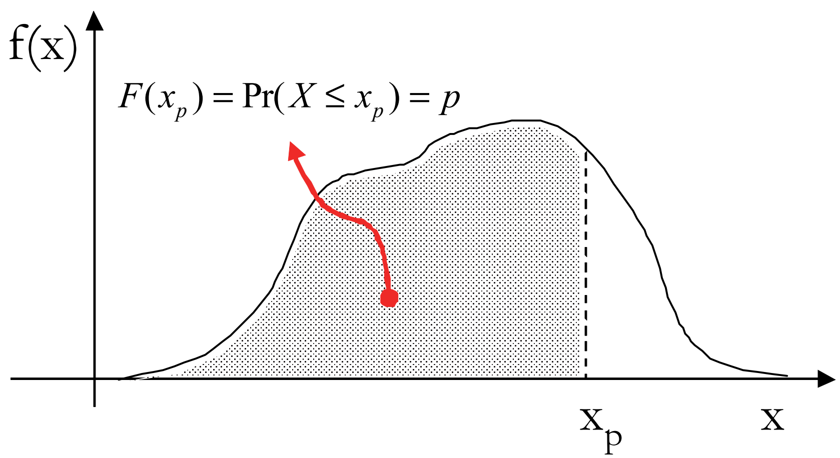
Table B.1. Right tail probabilities $1 - \Phi(a) = P(Z \geq a)$ for an $N(0, 1)$ distr
random variable Z.

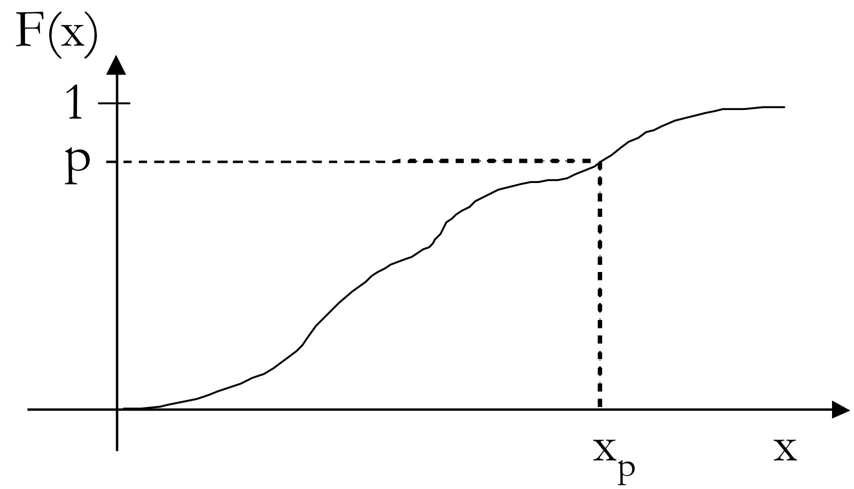
a	0	1	2	3	4	5	6	7	8	9
0.0	5000	4960	4920	4880	4840	4801	4761	4721	4681	4641
0.1	4602	4562	4522	4483	4443	4404	4364	4325	4286	4247
0.2	4207	4168	4129	4090	4052	4013	3974	3936	3897	3859
0.3	3821	3783	3745	3707	3669	3632	3594	3557	3520	3483
0.4	3446	3409	3372	3336	3300	3264	3228	3192	3156	3121
0.5	3085	3050	3015	2981	2946	2912	2877	2843	2810	2776
0.6	2743	2709	2676	2643	2611	2578	2546	2514	2483	2451
0.7	2420	2389	2358	2327	2296	2266	2236	2206	2177	2148
0.8	2119	2090	2061	2033	2005	1977	1949	1922	1894	1867
0.9	1841	1814	1788	1762	1736	1711	1685	1660	1635	1611
1.0	1587	1562	1539	1515	1492	1469	1446	1423	1401	1379
1.1	1357	1335	1314	1292	1271	1251	1230	1210	1190	1170
1.2	1151	1131	1112	1093	1075	1056	1038	1020	1003	0985
1.3	0968	0951	0934	0918	0901	0885	0869	0853	0838	0823
1.4	0808	0793	0778	0764	0749	0735	0721	0708	0694	0681
1.5	0668	0655	0643	0630	0618	0606	0594	0582	0571	0559
1.6	0548	0537	0526	0516	0505	0495	0485	0475	0465	0455
1.7	0446	0436	0427	0418	0409	0401	0392	0384	0375	0367
1.8	0359	0351	0344	0336	0329	0322	0314	0307	0301	0294
1.9	0287	0281	0274	0268	0262	0256	0250	0244	0239	0233



- Nowadays one evaluates $P(X \leq x|\mu, \sigma)$ using numerical functions such as $Norm.Dist(x, \mu, \sigma, 1)$ in Excel or $pnorm(x, \mu, \sigma)$ in R.

5.6 Quantiles...





Definition: Let X be a **continuous random variable** and let $p \in [0, 1]$. **The p-th quantile or $100p$-th percentile of the distribution of X** is the smallest number x_p such that :

$$F(x_p) = P(X \leq x_p) = p$$

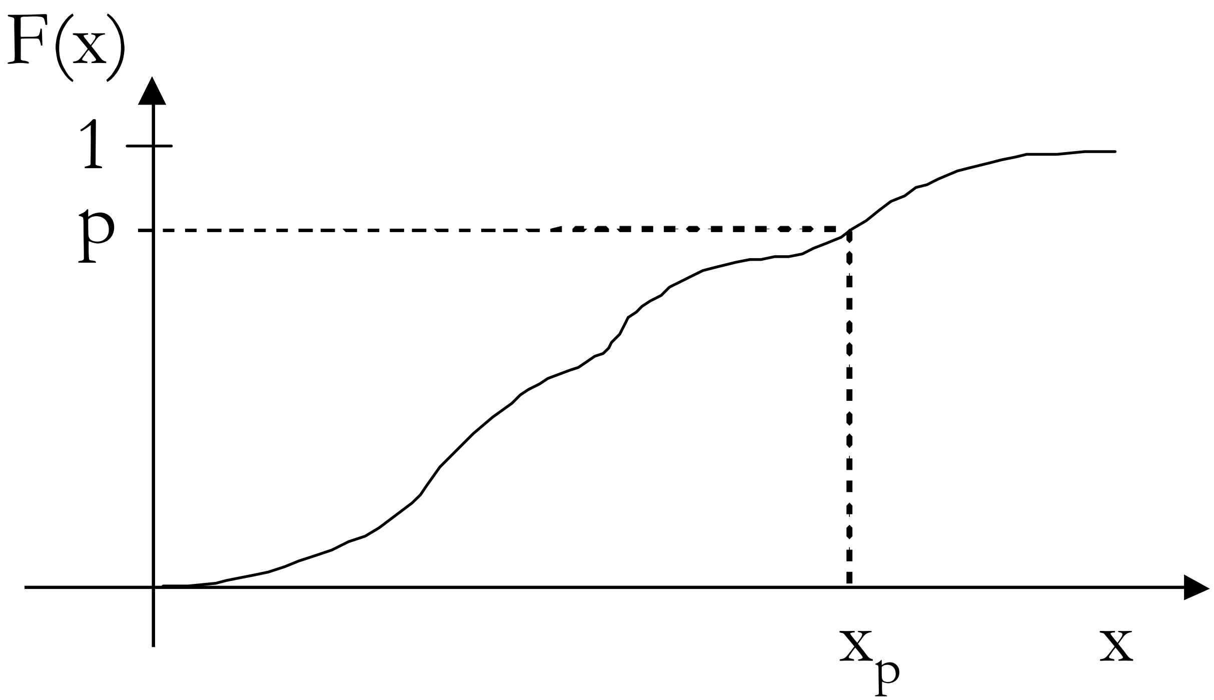
The median of a distribution is its 50-th percentile.

5 Continuous Random Variables

5.6 Quantile function is the inverse function of the cdf...

- **The quantile function** of $X \sim F(\cdot)$ is defined is defined as **the inverse function of the cumulative distributon function**: $F^{-1}:[0,1] \to \mathbb{R}$, where

$$F^{-1}(p) = x_p \Leftrightarrow F(F^{-1}(p)) = F(x_p) \Leftrightarrow p = F(x_p).$$



5 Continuous Random Variables

5.6 Quantile function is the inverse function of the cdf...

Example: Exponential distribution with parameter λ. That is: $X \sim Exp(\lambda)$

$$
\begin{aligned}
&F(x) = 1 - e^{-\lambda x}, F(x_p) = p \Leftrightarrow \\
&1 - e^{-\lambda x_p} = p \Leftrightarrow \\
&e^{-\lambda x_p} = 1 - p \Leftrightarrow \\
&-\lambda x_p = \ln(1-p) \Leftrightarrow \\
&x_p = -\frac{\ln(1-p)}{\lambda} \Leftrightarrow F^{-1}(x_p) = -\frac{\ln(1-p)}{\lambda}
\end{aligned}
$$

- **The quantile function** is not always available in a closed form, e.g. to calculate the 95-th quantile/percentile of $X \sim N(\mu, \sigma)$, where $\mu = 2, \sigma = 4$ one can use the MS Excel Function. $x_{0.95} = Norm.Inv(0.95, 2, 4, 1)$ or the R function, $x_{0.95} = qnorm(0.95, 2, 4)$.