POLITECNICO DI MILANO

Basic notions of probability theory

(Part 2)

Contents

- Basic Definitions
- Boolean Logic
- Definitions of probability
- Probability laws
- Random variables
- Probability Distributions

Random variables

Random variables

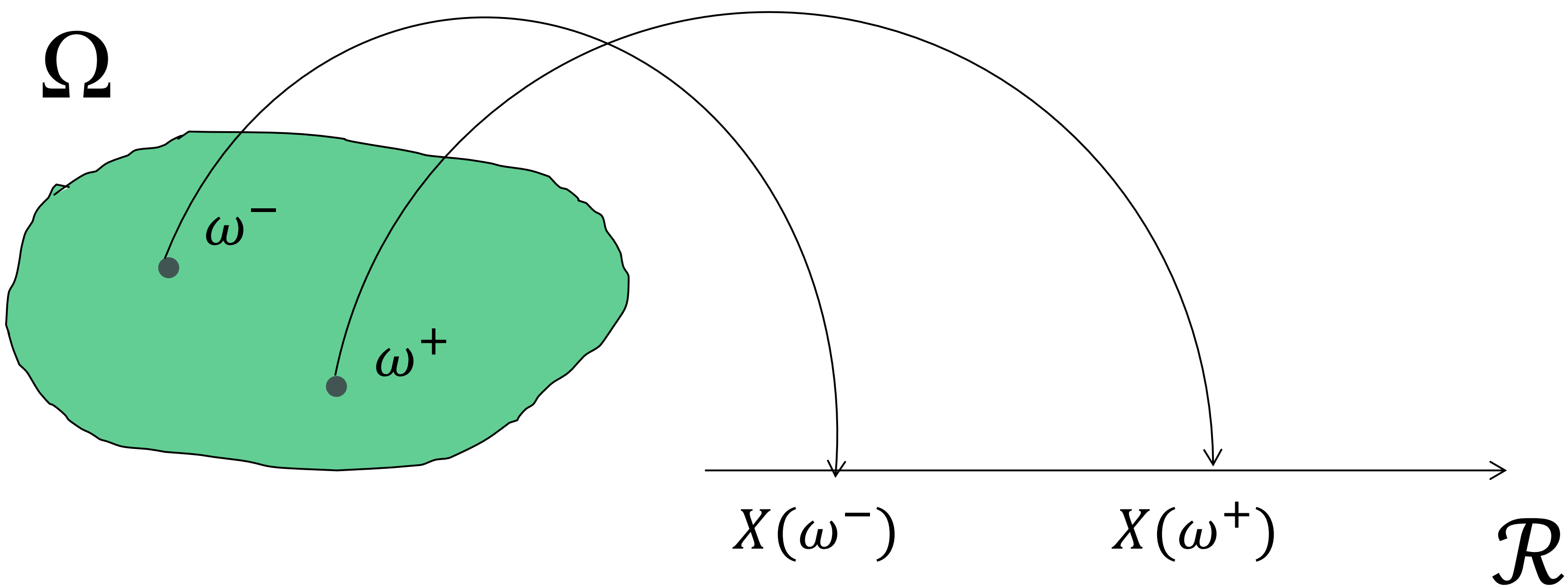
Experiment: ε

Sample space: Ω

Generic outcome: ω

→ $X(\omega)$ random variable in $\mathcal{R}$

Univocal mapping



Random variable - Example

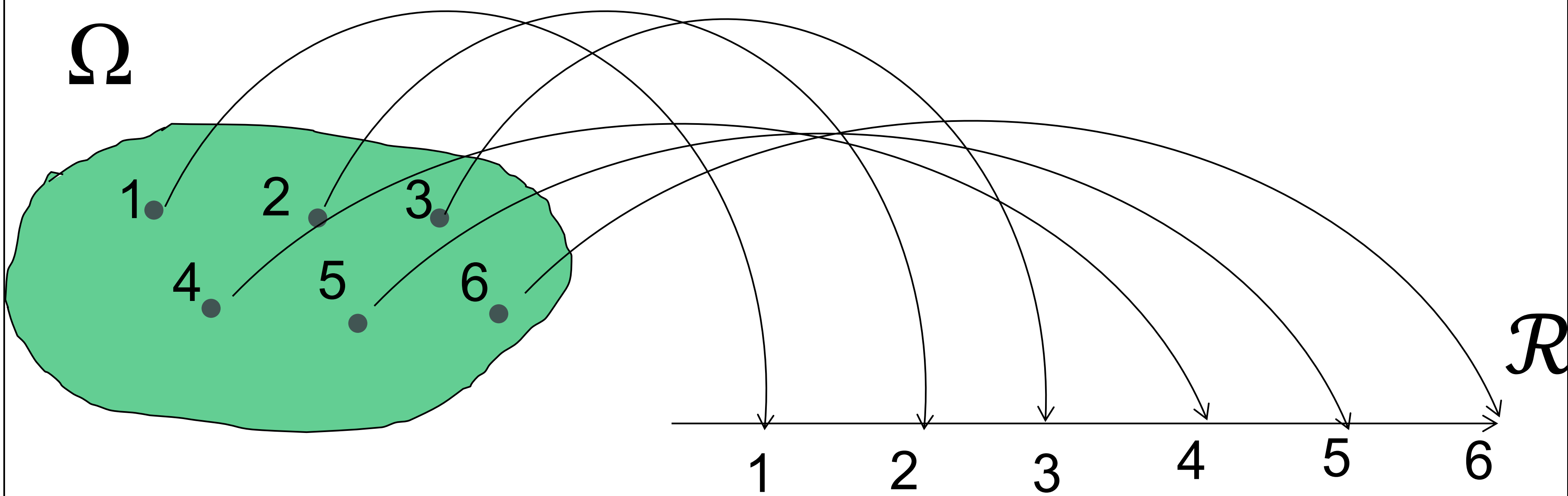
Experiment: ε = {tossing a dice}

Sample space: Ω={1,2,3,4,5,6}

Outcome: ω

$X(\omega)$ in $\Re$

Univocal mapping



Random variable - Event

Experiment: $\varepsilon =$ {tossing a dice}

Sample space: $\Omega=\{1,2,3,4,5,6\}$

Event: $E_1 = \{1,2,3,4\}$

$E_2 = \emptyset$

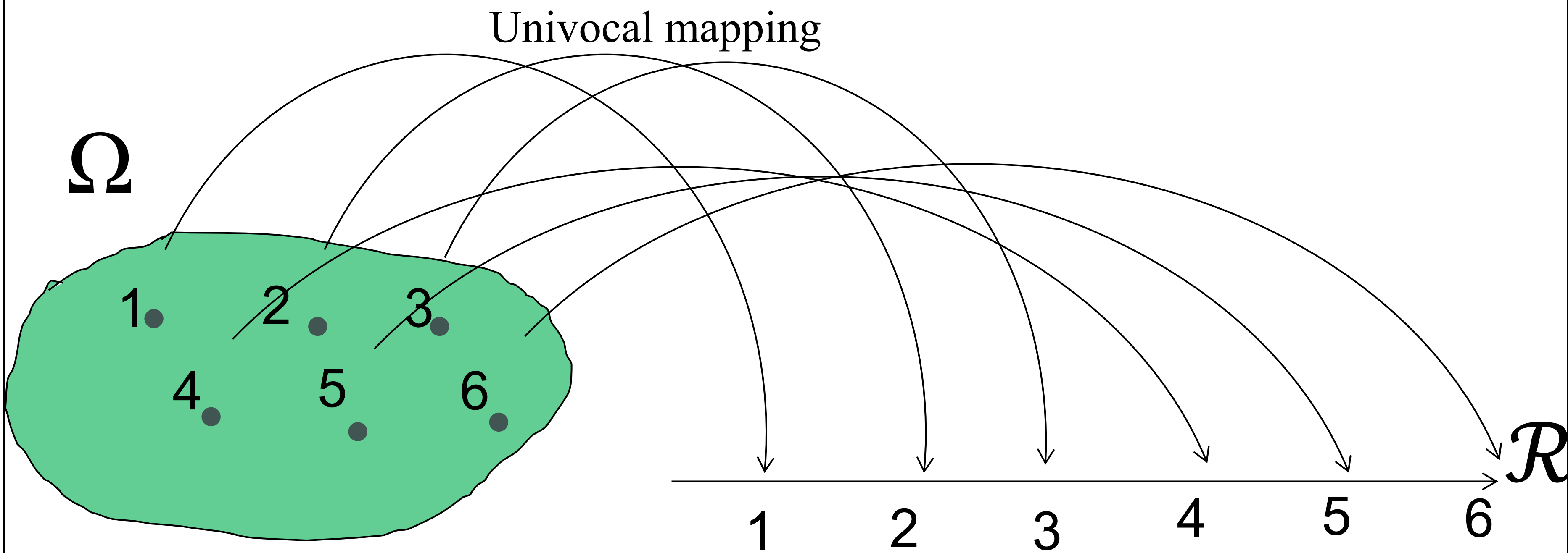
$E_3 = \Omega$

X(ω)

$E_1=\{X<4.236\}$

$E_2=\{X<0\}$

$E_3=\{X<+\infty\}$



Random variables

Experiment: ε

Sample space: Ω

Generic outcome: ω

→ $X(\omega)$ random variable in $\mathcal{R}$

General mathematical models of random behaviours

They apply to different physical phenomena which behave similarly

Probability distributions for reliability, safety and risk analysis

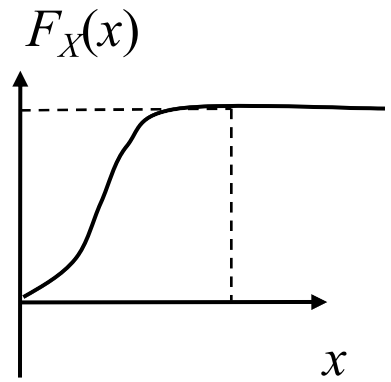
Probability functions (I)

- **Cumulative Distribution Function (cdf)**

  - $F_X(x)$ gives the probability of the event $\{X \leq x\}$ for any numerical value x.
  - Properties:

$$\lim_{x \to -\infty} F_X(x) = 0$$

$$\lim_{x \to +\infty} F_X(x) = 1$$



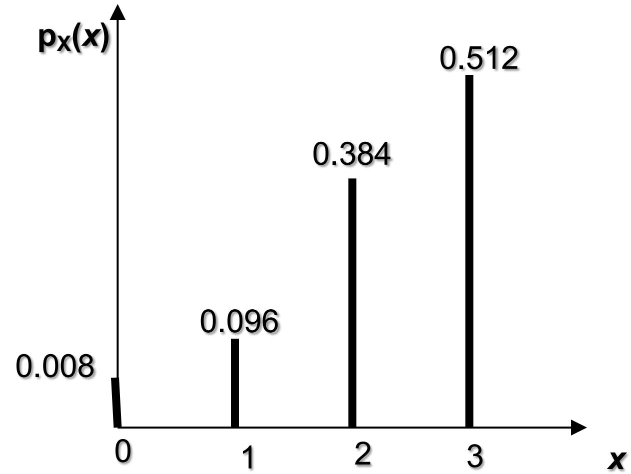
    - $F_X(x)$ is a non-decreasing function of x
    - The probability that X takes on a value in the interval $[a,b]$ is:

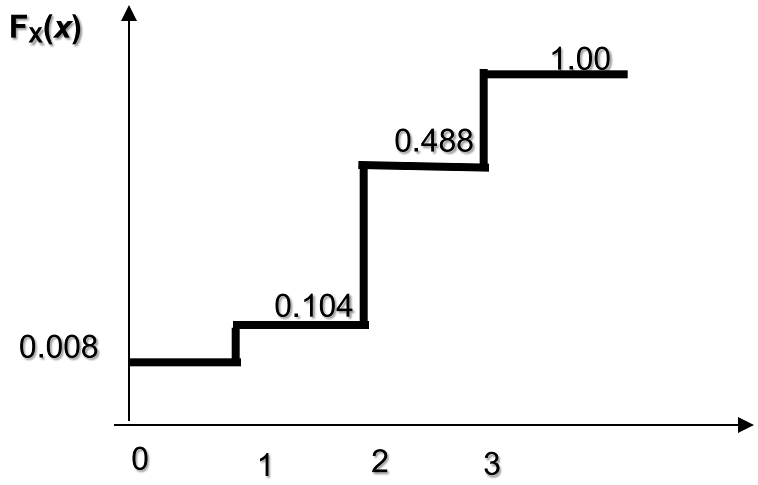
$$P\{a < X \leq b\} = F_X(b) - F_X(a)$$

Probability functions (II, discrete random variables)

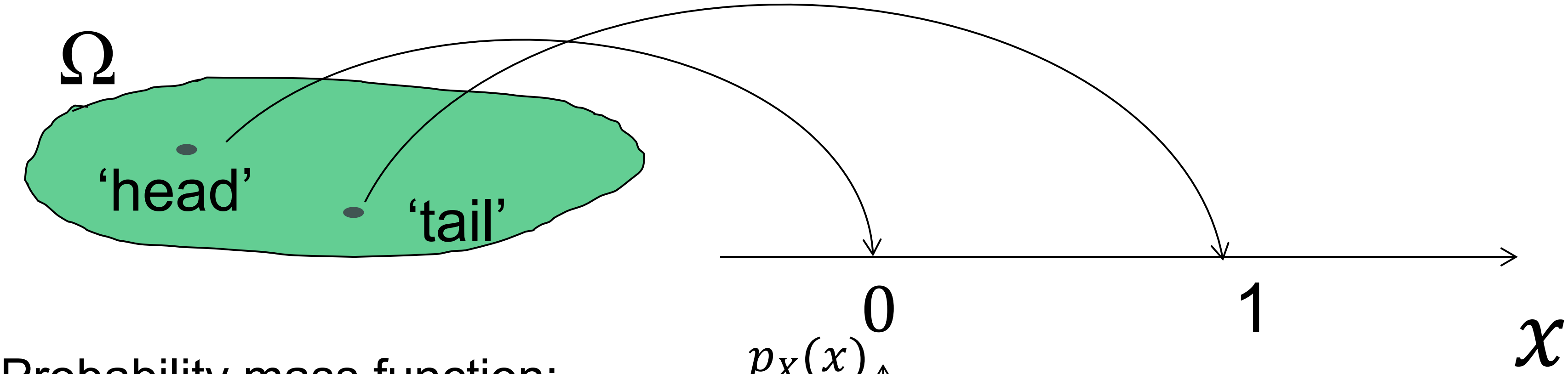
- **Probability Mass Function (pmf)**

  - X – random variable takes discrete values x_i, $i=1,2,\ldots n$:
    - $p_i = P\{X = x_i\}$
    - $F_X(x) = \sum_{x_i \le X} p_i$



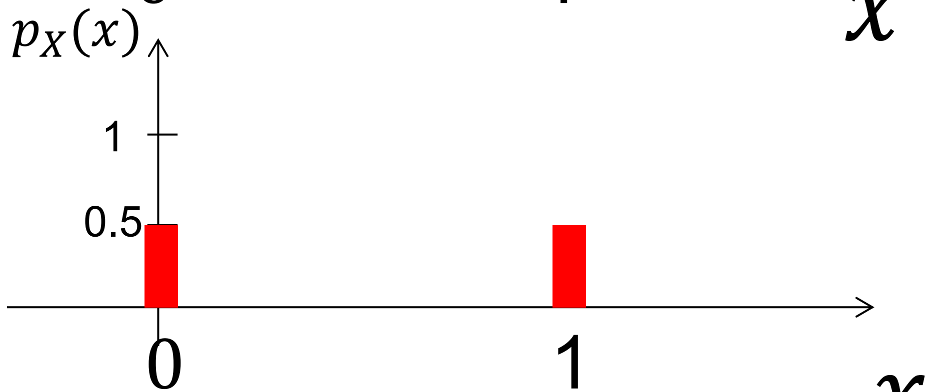


Probabilty Mass Function - Example

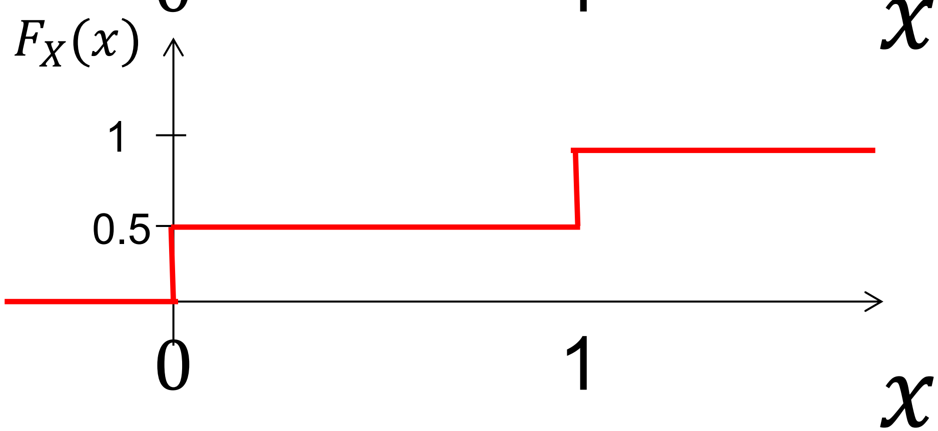


Probability mass function:

$$p_0 = P\{X = 0\} = 0.5$$
$$p_1 = P\{X = 1\} = 0.5$$



Cumulative distribution



Probability functions (II, continuous random variables)

- X – random variable takes continuous values in R
- Definition :

probability density function

$$P\{x \le X < x + dx\} = F_X(x+dx) - F_X(x) = f_X(x)dx$$

- $dx \to 0$:

$$f_X(x) = \lim_{dx \to 0} \frac{F_X(x+dx) - F_X(x)}{dx} = \frac{dF_X}{dx}$$

- $f_X(x)$ is not a probability but a probability per unit of x (probability density)

Summary measures: *percentiles, median, mean, variance*

- Distribution Percentiles (x_α):

$$F_X(x_\alpha) = \frac{\alpha}{100}$$

- Median of the distribution (x_{50}):

$$F_X(x_{50}) = 0.5$$

- Mean Value (Expected Value):

$$\mu_X = E[X] = < X > = \sum_{i=1}^{n} x_i p_i \qquad (\textit{discrete random variables})$$

$$= \int_{-\infty}^{\infty} x f_X(x) dx \qquad (\textit{continuous random variables})$$

- Variance (var[X]):

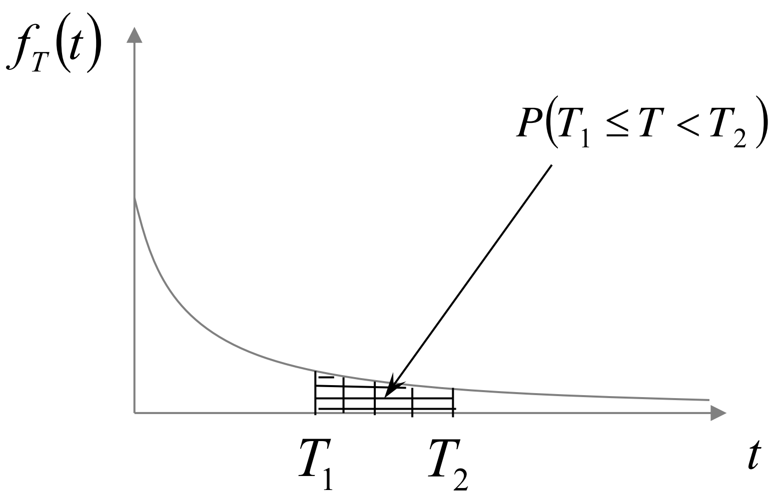
$$\sigma_X^2 = \sum_i (x_i - \mu_X)^2 p_i \qquad (\textit{discrete random variables})$$

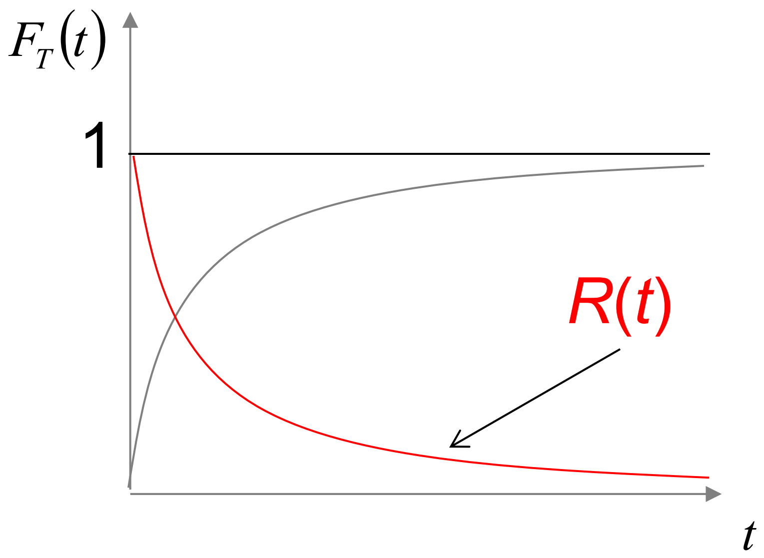
$$= \int_{-\infty}^{\infty} (x - \mu_X)^2 f_X(x) dx \qquad (\textit{continuous random variables})$$

Reliability

- T = Time to failure of a component (random variable)
  - Probability density function (pdf) at time t: $f_T(t)$
  - Cumulative distribution function (cdf) at time t = probability of having a failure before t : $F_T(t) = P(T<t)$
  - Reliability at time t = Probability that the component does not fail up to t:

$$R(t)=1-F_T(t)$$





The Hazard Function – *failure rate* (I)

$$h_T(t)dt = P(t < T \le t + dt \mid T > t) = \frac{P(t < T \le t + dt)}{P(T > t)} = \frac{f_T(t)dt}{R(t)}$$

Hazard Function and Reliability

$$h_T(t)dt = P(t < T \le t + dt \mid T > t) = \frac{P(t < T \le t + dt)}{P(T > t)} = \frac{f_T(t)dt}{R(t)}$$

$$h_T(t)dt = -\frac{dR(t)}{R(t)}$$

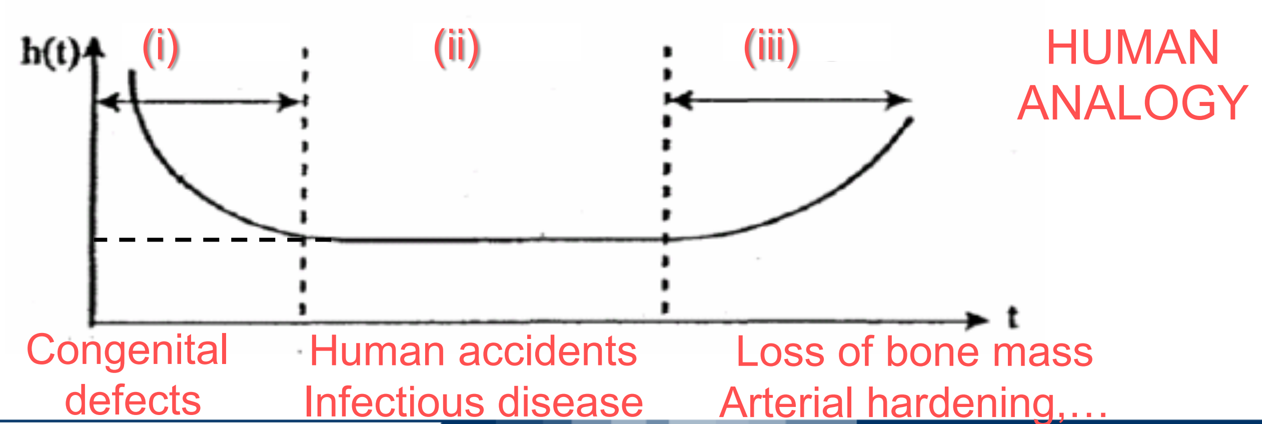
$$\int_0^t h_T(t')dt' = -\ln R(t)$$

$$R(t) = e^{-\int_0^t h_T(t')dt'}$$

$$f(t) = h(t)R(t) = h(t)e^{-\int_0^t h_T(t')dt'}$$

Hazard Function: the Bath-Tub Curve

- Usually, the hazard function shows three distinct phases:
  i. Decreasing - *infant mortality* or *burn in period:*
     - Failures due to defective pieces of equipment not manufactured or constructed properly (missing parts, substandard material batches, damage in shipping, ...)
  ii. Constant - *useful life*
     - Random failures due to unavoidable loads coming from without (earthquakes, power surges, vibration, temperature fluctuations,...)
  iii. Increasing – *ageing*
     - Aging failures due to cumulative effects such as corrosion, embrittlement, fatigue, cracking, …



Univariate probability distributions

Univariate Discrete Distributions: Binomial Distribution (I)

Y = discrete random variable with only two possible outcomes:

- Y=1 (success) with $P\{Y=1\}=p$
- Y=1 (failure) with $P\{Y=0\}=1-p$

Bernoulli process

We perform n different trials of the experiment, $Y_1, \dots, Y_n$

X= discrete random variable describing the number of success out of the n trials, independently from the sequence with which success appear:

$$X = \sum_{i=1}^{n} Y_i \qquad \Omega = \{1,2,\dots,n\}$$

The probability mass function:

$$b(k;n,p) = \binom{n}{k} p^k (1-p)^{n-k}$$

$k=1,\ 2,\dots n$

Univariate Discrete Distributions: Binomial Distribution (II)

$$b(k;n,p) = \binom{n}{k} p^k (1-p)^{n-k}$$

$k=1, 2, \ldots n$

$$E[X] = np$$
$$Var[X] = np(1-p)$$

T= trial of the first success (or number of trials between two successive occurrences of success)

The probability mass function:

$$g(t;p) = (1-p)^{t-1}\,p \qquad t=1,\ 2,\ldots$$

Expected value (return period):

$$E[T] = \sum_{t=1}^{\infty} t(1-p)^{t-1}\,p = p[1+2(1-p)+3(1-p)^2+\ldots] = \frac{p}{[1-(1-p)]^2} = \frac{1}{p}$$

Univariate Discrete Distributions, Poisson Distribution

Stochastic events that occur in a continuum period:

- Independent events
- Rate of occurrence, λ, is constant

↓

- Discrete Random Variable, K = number of events in the period of observation $(0,t)$

↓

- Probability mass function: $$p(k;(0,t),\lambda) = \frac{(\lambda t)^k}{k!} e^{-\lambda t}$$ *k=1, 2,…*

↓

$$E[K] = \lambda t$$
$$Var[K] = \lambda t$$

Continuous Distributions: Exponential Distribution

- $h_T(t)=\lambda$ constant
- T=failure time

$$P\{T>t\}=P\{\text{no failure in } (0,t)\}=$$

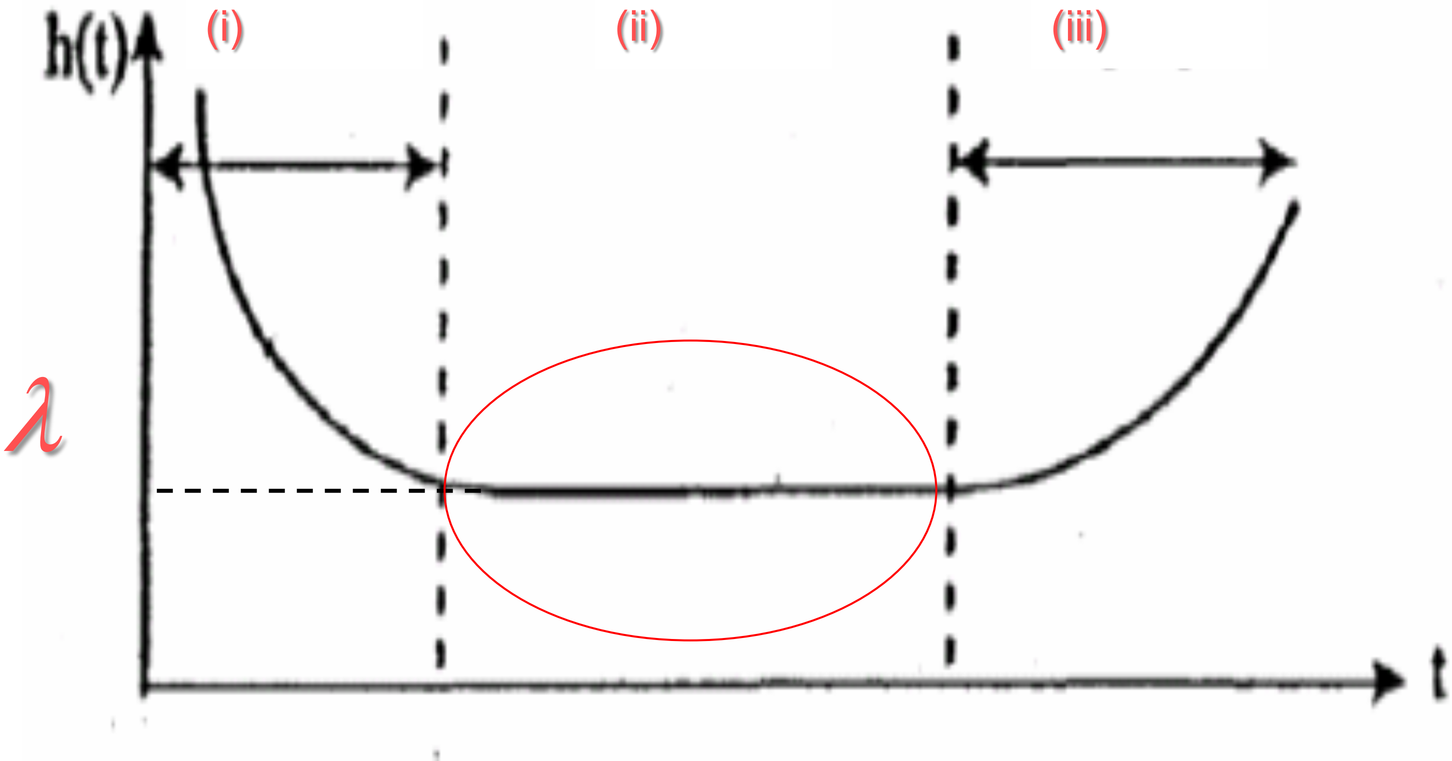
$$=\text{Poisson}(k=0,(0,t),\lambda)=\frac{(\lambda t)^0}{0!}e^{-\lambda t}=e^{-\lambda t}$$

$$F_T(t)=1-P\{T>t\}=1-e^{-\lambda t}$$

$$f_T(t)=\lambda e^{-\lambda t}$$

- It is the only distribution characterized by a constant failure rate

Exponential Distribution and bath tub curve



Exponential Distribution moments

$$E[T] = \int_0^{+\infty} t f(t) dt = \int_0^{+\infty} t \lambda e^{-\lambda t} dt = \frac{1}{\lambda}$$

$$\text{Var}[T] = \frac{1}{\lambda^2}$$

Exponential distribution: memorylessness

- A component with constant failure rate, λ, is found still operational at a given time t_1.

$$P\{T \leq t_2 | T > t_1\} = \frac{P(t_1 < T \leq t_2)}{P(T > t_1)} = \frac{F(t_2) - F(t_1)}{R(t_1)} =$$

$$\frac{(1 - e^{-\lambda t_2}) - (1 - e^{-\lambda t_1})}{e^{-\lambda t_1}} = \frac{(e^{-\lambda t_1} - e^{-\lambda t_2})}{e^{-\lambda t_1}} = 1 - e^{-\lambda (t_2 - t_1)}$$

- Still exponential with failure rate λ!
- The probability that it will fail in some period of time of lengths $\tau = t_2 - t_1$ does not depend from the component age t_1

Continuous Distributions : the Weibull Distribution

- In practice, the age of a component influences its failure process so that the hazard rate does not remain constant throughout the lifetime

$$F_T(t) = P(T \le t) = 1 - e^{-\lambda t^{\alpha}}$$

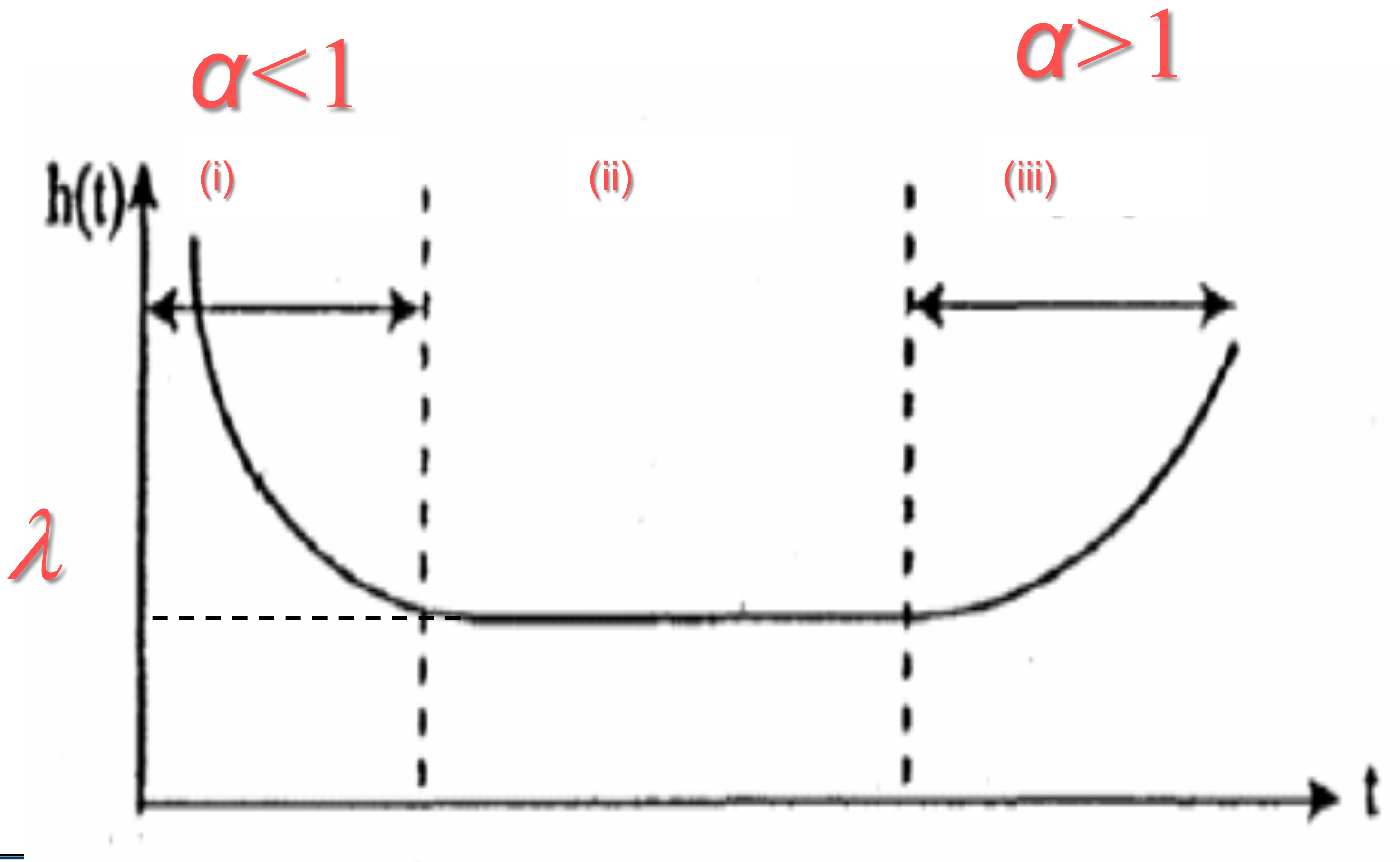
$$\begin{cases} f_T(t) = \lambda \alpha t^{\alpha-1} e^{-\lambda t^{\alpha}} & t \ge 0 \\ \quad = 0 & t < 0 \end{cases}$$

$$E[T] = \frac{1}{\lambda}\Gamma\left(\frac{1}{\alpha}+1\right) \quad ; \quad Var[T] = \frac{1}{\lambda^2}\left(\Gamma\left(\frac{2}{\alpha}+1\right) - \Gamma\left(\frac{1}{\alpha}+1\right)\right)^2$$

$$\Gamma(k) = \int_0^{\infty} x^{k-1} e^{-x} dx \qquad k > 0$$

Weibull Distribution and bath tub curve

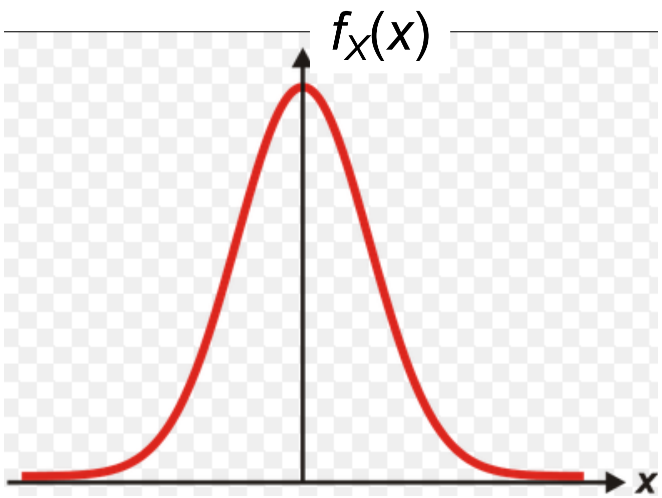
$$h(t) = \frac{f(t)}{1 - F(t)} = \lambda \alpha t^{\alpha - 1}$$



Continuous Distributions: Normal or Gaussian Distribution

Probability density function: $X \sim N(\mu_X, \sigma_X)$

$$f_X(x; \mu_X, \sigma_X) = \frac{1}{\sqrt{2\pi}\sigma_X} e^{-\frac{1}{2}\left(\frac{x-\mu_X}{\sigma_X}\right)^2} \qquad -\infty < x, \mu_X < \infty; \sigma_X > 0$$



It is the only distribution with a symmetric bell shape!

Expected value and variance:

$$E[X] = \mu_X$$
$$Var[X] = \sigma_X^2$$

Central limit theorem

- For any sequence of *n* independent random variable X_i, their sum $X = \sum_{i=1}^{n} X_i$ is a random variable which, for large *n*, tends to be distributed as a normal distribution

If X_i are independent, identically distributed random variables with mean μ and finite variance given by σ^2

$$S_n = \frac{\sum_{i=0}^{n} X_i}{n} \rightarrow N\left(\mu, \frac{\sigma^2}{n}\right)$$

Standard Normal Variable

$$P(a < X < b) = \frac{1}{\sigma\sqrt{2\pi}} \int_a^b e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$$

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

$$S \sim N(0,1)$$

$$\mathrm{P(a < X < b)} = \frac{1}{\sigma\sqrt{2\pi}} \int_{\frac{a-\mu}{\sigma}}^{\frac{b-\mu}{\sigma}} \mathrm{e}^{-\frac{1}{2}s^2} \sigma\, ds =$$

$$\Rightarrow \mathrm{P(a < X < b)} = \frac{1}{\sqrt{2\pi}} \int_{\frac{a-\mu}{\sigma}}^{\frac{b-\mu}{\sigma}} \mathrm{e}^{-\frac{1}{2}s^2} ds = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$$

Table of Standard Normal Probability

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{1}{2}\xi^2} d\xi$$

x	$\Phi(x)$	x	$\Phi(x)$		
0.00	0.500000	0.50	0.691463		
0.01	0.503989	0.51	0.694975		
0.02	0.507978	0.52	0.698468		
0.03	0.511966	0.53	0.701944		
0.04	0.515954	0.54	0.705401	1.04	0.850830
0.05	0.519939	0.55	0.708840	1.05	0.853141
0.06	0.523922	0.56	0.712260	1.06	0.855428
0.07	0.527904	0.57	0.715661	1.07	0.857690
0.08	0.531882	0.58	0.719043	1.08	0.859929
0.09	0.535857	0.59	0.722405	1.09	0.862143
0.10	0.539828	0.60	0.725747	1.10	0.864334
0.11	0.543796	0.61	0.729069	1.11	0.866500
0.12	0.547759	0.62	0.732371	1.12	0.868643
0.13	0.551717	0.63	0.735653	1.13	0.870762
0.14	0.555671	0.64	0.738914	1.14	0.872857
0.15	0.559618	0.65	0.742154	1.15	0.874928
0.16	0.563500	0.66	0.745374	1.16	0.876976
0.17	0.567494	0.67	0.748572	1.17	0.878999
0.18	0.571423	0.68	0.751748	1.18	0.881000
0.19	0.575345	0.69	0.754903	1.19	0.882977
0.20	0.579260	0.70	0.758036	1.20	0.884930
0.21	0.583166	0.71	0.761148	1.21	0.886860
0.22	0.587064	0.72	0.764238	1.22	0.888767
0.23	0.590954	0.73	0.767305	1.23	0.890651
0.24	0.549835	0.74	0.770350	1.24	0.892512
0.25	0.598706	0.75	0.773373	1.25	0.894350
0.26	0.602568	0.76	0.776373	1.26	0.896165
0.27	0.606420	0.77	0.779350	1.27	0.897958
0.28	0.610262	0.78	0.782305	1.28	0.899727
0.29	0.614092	0.79	0.785236	1.29	0.901475
0.30	0.617912	0.80	0.788145	1.30	0.903199
0.31	0.621720	0.81	0.791030	1.31	0.904902
0.32	0.623517	0.82	0.793892	1.32	0.906583
0.33	0.629301	0.83	0.796731	1.33	0.908241
0.34	0.633072	0.84	0.799546	1.34	0.909877
0.35	0.636831	0.85	0.802337	1.35	0.911492
0.36	0.640576	0.86	0.805105	1.36	0.913085
0.37	0.644309	0.87	0.807850	1.37	0.914656
0.38	0.648027	0.88	0.810570	1.38	0.916207
0.39	0.651732	0.89	0.813267	1.39	0.917735
0.40	0.655422	0.90	0.815940	1.40	0.919243
0.41	0.659097	0.91	0.818589	1.41	0.920730
0.42	0.662757	0.92	0.821214	1.42	0.922196
0.43	0.666402	0.93	0.823815	1.43	0.923641
0.44	0.670032	0.94	0.826391	1.44	0.925066
0.45	0.673645	0.95	0.828944	1.45	0.926471
0.46	0.677242	0.96	0.831473	1.46	0.927855
0.47	0.680823	0.97	0.833977	1.47	0.929219
0.48	0.684387	0.98	0.836457	1.48	0.930563
0.49	0.687933	0.99	0.838913	1.49	0.931888

Table of Standard Normal Probability

x	$\Phi(x)$
1.50	0.933193
1.51	0.934478
1.52	0.935744
1.53	0.936992
1.54	0.938220
1.55	0.939429
1.56	0.940620
1.57	0.941792
1.58	0.942947
1.59	0.944083
1.60	0.945201
1.61	0.946301
1.62	0.947384
1.63	0.948449
1.64	0.949497
1.65	0.950529
1.66	0.951543
1.67	0.952540
1.68	0.953521
1.69	0.954486
1.70	0.955435
1.71	0.956367
1.72	0.957284
1.73	0.958185
1.74	0.959071
1.75	0.959941
1.76	0.960796
1.77	0.961636
1.78	0.962426
1.79	0.963273
1.80	0.964070
1.81	0.964852
1.82	0.965621
1.83	0.966375
1.84	0.967116
1.85	0.967843
1.86	0.968557
1.87	0.969258
1.88	0.969946
1.89	0.970621
1.90	0.971284
1.91	0.971933
1.92	0.972571
1.93	0.973197
1.94	0.973810
1.95	0.974412
1.96	0.975002
1.97	0.975581
1.98	0.976148
1.99	0.976705

x	$\Phi(x)$
2.00	0.977250
2.01	0.977784
2.02	0.978308
2.03	0.978822
2.04	0.979325
2.05	0.979818
2.06	0.980301
2.07	0.980774
2.08	0.981237
2.09	0.981691
2.10	0.982136
2.11	0.982571
2.12	0.982997
2.13	0.983414
2.14	0.983823
2.15	0.984223
2.16	0.984614
2.17	0.984997
2.18	0.985371
2.19	0.985738
2.20	0.986097
2.21	0.986447
2.22	0.986791
2.23	0.987126
2.24	0.987455
2.25	0.987776
2.26	0.988089
2.27	0.988396
2.28	0.988696
2.29	0.988989
2.30	0.989276
2.31	0.989556
2.32	0.989830
2.33	0.990097
2.34	0.990358
2.35	0.990613
2.36	0.990863
2.37	0.991106
2.38	0.991344
2.39	0.991576
2.40	0.991802
2.41	0.992024
2.42	0.992240
2.43	0.992451
2.44	0.992656
2.45	0.992857
2.46	0.993053
2.47	0.993244
2.48	0.993431
2.49	0.993613

x	$\Phi(x)$
2.50	0.993790
2.51	0.993963
2.52	0.994132
2.53	0.994267
2.54	0.994457
2.55	0.994614
2.56	0.994766
2.57	0.994915
2.58	0.995060
2.59	0.995201
2.60	0.995339
2.61	0.995473
2.62	0.995604
2.63	0.995731
2.64	0.995855
2.65	0.995975
2.66	0.996093
2.67	0.996207
2.68	0.996319
2.69	0.996427
2.70	0.996533
2.71	0.996636
2.72	0.996736
2.73	0.996833
2.74	0.996928
2.75	0.997020
2.76	0.997110
2.77	0.997197
2.78	0.997282
2.79	0.997365
2.80	0.997445
2.81	0.997523
2.82	0.997599
2.83	0.997673
2.84	0.997744
2.85	0.997814
2.86	0.997882
2.87	0.997948
2.88	0.998012
2.89	0.998074
2.90	0.998134
2.91	0.998193
2.92	0.998250
2.93	0.998305
2.94	0.998359
2.95	0.998411
2.96	0.998462
2.97	0.998511
2.98	0.998559
2.99	0.998605

Table of Standard Normal Probability

x	Φ(x)	x	Φ(x)	x	1−Φ(x)
3.00	0.998630	3.50	0.999767	4.00	0.316712E-04
3.01	0.998694	3.51	0.999776	4.05	0.256088E-04
3.02	0.998736	3.52	0.999784	4.10	0.206575E-04
3.03	0.998777	3.53	0.999792	4.15	0.166238E-04
3.04	0.998817	3.54	0.999800	4.20	0.133458E-04
3.05	0.998856	3.55	0.999807	4.25	0.106883E-04
3.06	0.998893	3.56	0.999815	4.30	0.853906E-05
3.07	0.998930	3.57	0.999821	4.35	0.680688E-05
3.08	0.998965	3.58	0.999828	4.40	0.541234E-05
3.09	0.998999	3.59	0.999835	4.45	0.429351E-05
3.10	0.999032	3.60	0.999841	4.50	0.339767E-05
3.11	0.999065	3.61	0.999847	4.55	0.268230E-05
3.12	0.999096	3.62	0.999853	4.60	0.211245E-05
3.13	0.999126	3.63	0.999858	4.65	0.165968E-05
3.14	0.999155	3.64	0.999864	4.70	0.130081E-05
3.15	0.992184	3.65	0.999869	4.75	0.101708E-05
3.16	0.999119	3.66	0.999874	4.80	0.793328E-06
3.17	0.999238	3.67	0.999879	4.85	0.617307E-06
3.18	0.999264	3.68	0.999883	4.90	0.479183E-06
3.19	0.999289	3.69	0.999888	4.95	0.371067E-06
3.20	0.999313	3.70	0.999892	5.00	0.286652E-06
3.21	0.999336	3.71	0.999806	5.10	0.169827E-06
3.22	0.999359	3.72	0.999900	5.20	0.996443E-07
3.23	0.999381	3.73	0.999904	5.30	0.579013E-07
3.24	0.999402	3.74	0.999908	5.40	0.333204E-07
3.25	0.999423	3.75	0.999912	5.50	0.189896E-07
3.26	0.999443	3.76	0.999915	5.60	0.107176E-07
3.27	0.999462	3.77	0.999918	5.70	0.599037E-08
3.28	0.999481	3.78	0.999922	5.80	0.331575E-08
3.29	0.999499	3.79	0.999925	5.90	0.181751E-08
3.30	0.999516	3.80	0.999928	6.00	0.986588E-09
3.31	0.999533	3.81	0.999931	6.10	0.530343E-09
3.32	0.999550	3.82	0.999933	6.20	0.282316E-09
3.33	0.999566	3.83	0.999936	6.30	0.148823E-09
3.34	0.999581	3.84	0.999938	6.40	0.77688 E-10
3.35	0.999596	3.85	0.999941	6.50	0.40160 E-10
3.36	0.999610	3.86	0.999943	6.60	0.20558 E-10
3.37	0.999624	3.87	0.999946	6.70	0.10421 E-10
3.38	0.999637	3.88	0.999948	6.80	0.5231 E-11
3.39	0.999650	3.89	0.999950	6.90	0.260 E-11
3.40	0.999663	3.90	0.999952	7.00	0.128 E-11
3.41	0.999675	3.91	0.999954	7.10	0.624 E-12
3.42	0.999687	3.92	0.999956	7.20	0.301 E-12
3.43	0.999698	3.93	0.999958	7.30	0.144 E-12
3.44	0.999709	3.94	0.999959	7.40	0.68 E-13
3.45	0.999720	3.95	0.999961	7.50	0.32 E-13
3.46	0.999730	3.96	0.999963	7.60	0.15 E-13
3.47	0.999740	3.97	0.999964	7.70	0.70 E-14
3.48	0.999749	3.98	0.999966	7.80	0.30 E-14
3.49	0.999758	3.99	0.999967	7.90	0.15 E-14

Probability distributions:

Univariate Continuous Distributions Log-normal Distribution

Probability density function:

$$g_X(x;\mu_Z,\sigma_Z) = \frac{1}{\sqrt{2\pi}\sigma_Z}\frac{1}{x}e^{-\frac{1}{2}\left(\frac{\ln x-\mu_Z}{\sigma_Z}\right)^2} \qquad x,\sigma_Z > 0$$

Note: if

$$X \sim Log-normal(\mu_Z,\sigma_Z) \Rightarrow Z = \ln X \sim N(\mu_Z,\sigma_Z)$$

Expected value and variance:

$$E[X] = e^{\mu_Z+\frac{\sigma_Z^2}{2}}$$

$$Var[X] = e^{2\mu_Z+\sigma_Z^2}(e^{\sigma_Z^2}-1)$$

Bivariate probability distributions

Probability distributions:

Joint Probability Distributions, Bivariate Distribution

Joint cumulative distribution function:

X, Y - random variables

$$P[X \le x, Y \le y] = F_{XY}(x, y)$$

constraints:

a) $F_{XY}(-\infty, -\infty) = 0;$ $\quad F_{XY}(\infty, \infty) = 1$

b) $F_{XY}(-\infty, y) = 0;$ $\quad F_{XY}(\infty, y) = F_Y(\mathrm{y})$

$F_{XY}(x, -\infty) = 0;$ $\quad F_{XY}(x, \infty) = F_X(\mathrm{x})$

(the two right-hand relations of b): marginal CDFs)

c) $F_{XY}(x, y)$ is nonegative, and nondecreasing function of x and y

Joint probability density function:

$$P[x < X \le x + dx, y < Y \le y + dy] = f_{XY}(x, y)dxdy$$

$$\Rightarrow F_{XY}(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{XY}(x, y)dxdy$$

Probability distributions:

Joint Probability Distributions, Bivariate Distribution

Conditional probability density functions:

$$P[X \le x | Y \le y] = F_{X|Y}(x|y) = \frac{F_{XY}(x,y)}{F_Y(y)}, \text{ if } F_Y(y) \ne 0$$

$$P[Y \le y | X \le x] = F_{Y|X}(y|x) = \frac{F_{XY}(x,y)}{F_X(x)}, \text{ if } F_X(x) \ne 0$$

The marginal PDFs can be obtained from the joint PDF by applying the theorem of total probability:

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x,y)dy \qquad f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x,y)dx$$

X, Y - statistically independent

Joint probability density function:

$$f_{XY}(x,y) = f_X(x) \cdot f_Y(y)$$

Probability distributions:

Joint Probability Distributions, Bivariate Distribution

Covariance:

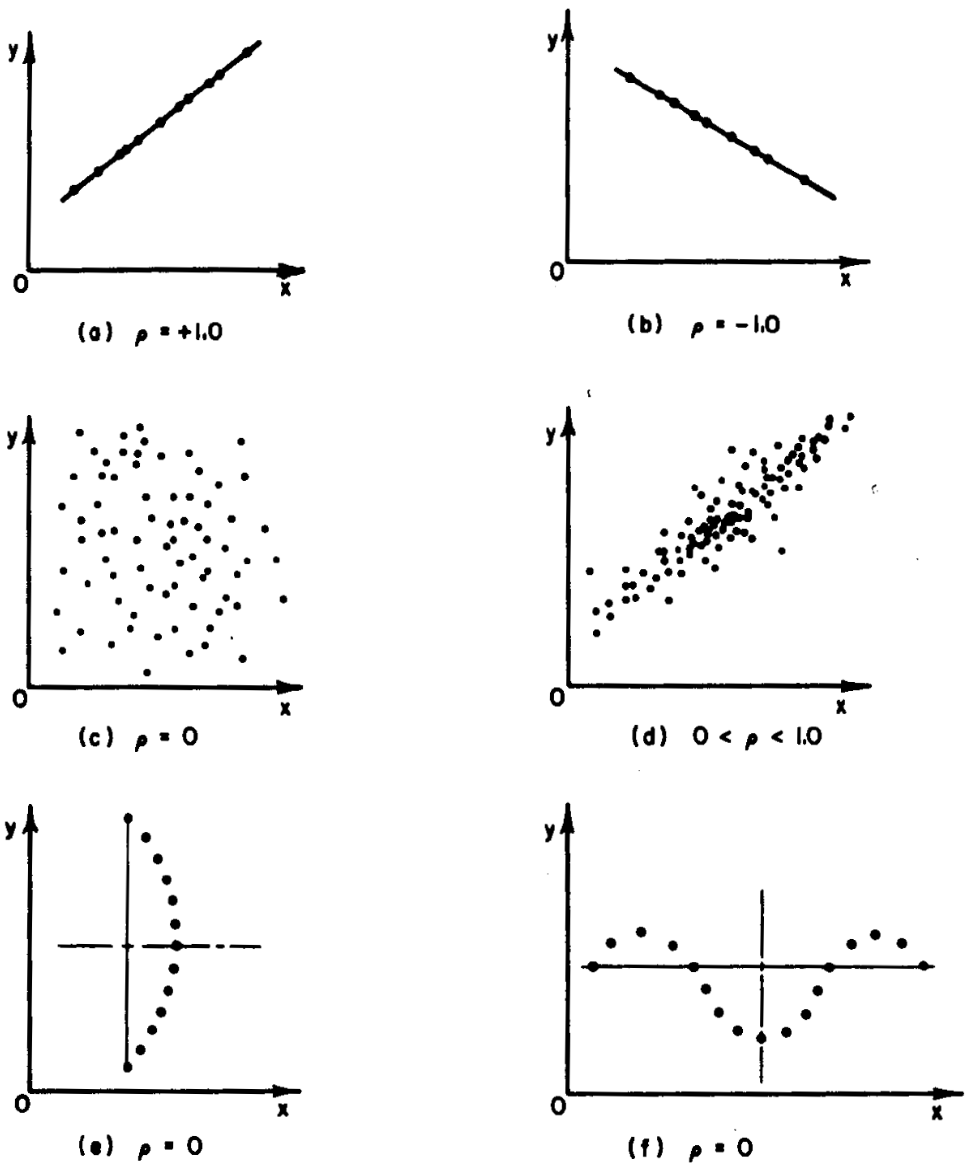
$$\mathrm{cov}[X,Y] = E[(X-\mu_X)\cdot(Y-\mu_Y)] = E[XY] - E[X]\cdot E[Y]$$

Correlation coefficient (normalized covariance):

$$\rho_{XY} = \frac{\mathrm{cov}[X,Y]}{\sigma_X \cdot \sigma_Y}, \qquad -1 \leq \rho_{XY} \leq 1$$

Probability distributions:

Joint Probability Distributions, Bivariate Distribution



- ρ just gives a 'descriptive' relationship: $\rho \neq 0$ does not mean that there is a casual relationship; the variables remain random

- ρ is a measure of linear relationship; you might have random variables that are completely dependent but have $\rho=0$

- $\rho=0 \not\Rightarrow$ *X* and *Y* are independent

if *X* and *Y* are independent $\Rightarrow \rho=0$

Functions of random variables

Functions of Random Variables: Derived Probability Distributions

$X, Y = g(X)$

Cumulative distribution function:

$$F_Y(y) = P[Y \le y] = P[g(X) \le y] = \\ = P[X \le g^{-1}(y)] = F_X[g^{-1}(y)]$$

assume that $g(x)$ is a monotonically increasing function

Probability density function:

$$f_Y(y)dy = f_X(x)dx \qquad , Y = g(X)$$

$$x_i = g^{-1}(y) \qquad x_i \text{ is the } i\text{-th root of } g^{-1}(y)$$

$$g'(x) = \frac{dy}{dx}$$

- derived density function:

$$\Rightarrow f_Y(y) = \sum_{i=1}^{k} \frac{f_X(x_i)}{|g'(x_i)|}$$

Functions of Random Variables:

Mean and Variance of a General Function

$X, Y = g(X)$

Mean:

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

Variance:

$$Var[Y] = \int_{-\infty}^{\infty} \left[g(x) - E[Y] \right]^2 f_X(x) dx$$

Expanding $g(X)$ in a Taylor series about the mean value $\bar{X} = \mu_x$

$$Y = g(\bar{X}) + (X - \bar{X}) \left. \frac{dg}{dX} \right|_{\bar{X}} + \frac{1}{2} (X - \bar{X})^2 \left. \frac{d^2 g}{dX^2} \right|_{\bar{X}} + \dots$$

Functions of Random Variables:

Approximate Mean and Variance of a General Function

First-order approximate mean:

$$E[Y] \approx g(\bar{X})$$

First-order approximate variance:

$$Var[Y] \approx Var[X]\left(\left.\frac{dg}{dX}\right|_{\bar{X}}\right)^2$$

Second-order approximate mean:

$$E[Y] \approx g(\bar{X}) + \frac{1}{2} Var[X] \left.\frac{d^2 g}{dX^2}\right|_{\bar{X}}$$

Second-order approximate variance:

$$Var[Y] \approx Var[X]\left(\left.\frac{dg}{dX}\right|_{\bar{X}}\right)^2 - \frac{1}{4} Var^2[X]\left(\left.\frac{d^2 g}{dX^2}\right|_{\bar{X}}\right)^2 + E[X-\bar{X}]^3 \left.\frac{dg}{dX}\right|_{\bar{X}} \left.\frac{d^2 g}{dX^2}\right|_{\bar{X}} + \frac{1}{4} E[X-\bar{X}]^4 \left(\left.\frac{d^2 g}{dX^2}\right|_{\bar{X}}\right)^2$$

Example of functions of multiple random variables

1.

$$Z = \sum_{i=1}^{n} X_i$$

a) $X_i \sim \text{Poisson}(\lambda_i)$,

b) X_i-independent

$$\Rightarrow Z \sim \text{Poisson}\left(\lambda = \sum_{i=1}^{n} \lambda_i\right)$$

2.

$$Z = X + Y$$

$$X \sim N(\mu_X, \sigma_X)$$

$$Y \sim N(\mu_Y, \sigma_Y)$$

$$\Rightarrow Z \sim \mathrm{N}\left(\mu_X + \mu_Y, \sqrt{\sigma_X^2 + \sigma_Y^2}\right)$$

$$Z = aX + bY + c$$

$$\Rightarrow Z \sim \mathrm{N}\left(a\mu_X + b\mu_Y + c, \sqrt{a^2\sigma_X^2 + b^2\sigma_Y^2}\right)$$

$$Z = X - Y$$

$$\Rightarrow Z \sim \mathrm{N}\left(\mu_X - \mu_Y, \sqrt{\sigma_X^2 + \sigma_Y^2}\right)$$

Any linear function of normal variates is also a normal variate

Example of functions of multiple random variables

3.

$$Y = aX^b$$
$$X \sim \text{log-normal}(\lambda, \xi)$$

$$\Rightarrow \ln Y = \ln a + b \ln X$$
$$\ln Y \sim N\left(\ln a + b\lambda, \sqrt{b^2 \xi^2}\right)$$

$$\Rightarrow Y \sim \text{log-normal}\left(\ln a + b\lambda, \sqrt{b^2 \xi^2}\right)$$

4.

$$Z = XY$$
$$X \sim \text{log-normal}(\lambda_X, \xi_X)$$
$$Y \sim \text{log-normal}(\lambda_Y, \xi_Y)$$

$$\Rightarrow \ln X = \ln X + \ln Y$$
$$\ln Z \sim N\left(\lambda_X + \lambda_Y, \sqrt{\xi_X^2 + \xi_Y^2}\right)$$

$$\Rightarrow Z \sim \text{log-normal}\left(\lambda_X + \lambda_Y, \sqrt{\xi_X^2 + \xi_Y^2}\right)$$

THE END (probably)

Thank you for the attention!!!!