DISTRIBUTIONS OF CONTINUOUS RVS

Dr. Aleksandra Kaszubowska-Anandarajah

anandara@tcd.ie

Some material has been adopted from
the Pennsylvania State University online course STAT 414

Previous lecture

- ➢ Moments of RV
- ➢ Discrete RV:
  - ✓ Binomial
  - ✓ Bernoulli
  - ✓ Geometric
  - ✓ Poisson

Today’s lecture

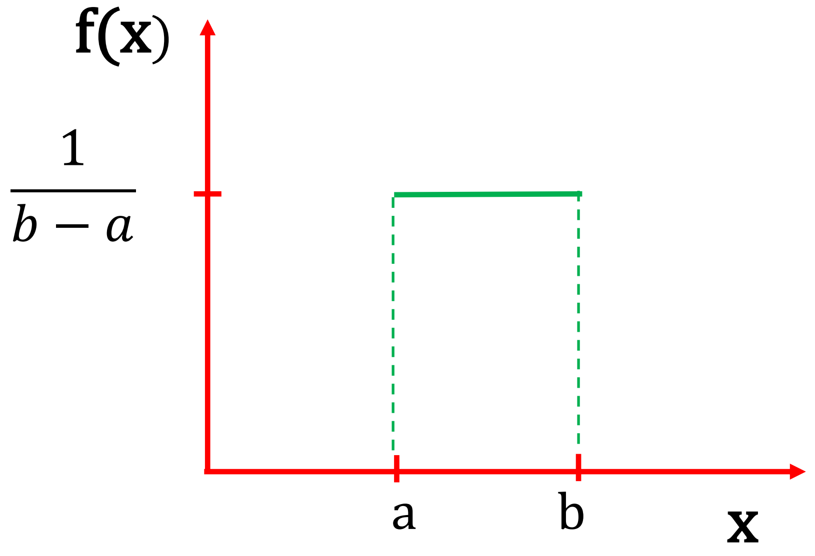
- Continuous distributions
  - Uniform
  - Exponential
  - Normal
- Gaussian noise

Continuous RV: Uniform

PDF of Uniform RV is

$$f(X) = \begin{cases} \dfrac{1}{b-a} & when\ a < x < b \\ 0 & otherwise \end{cases}$$

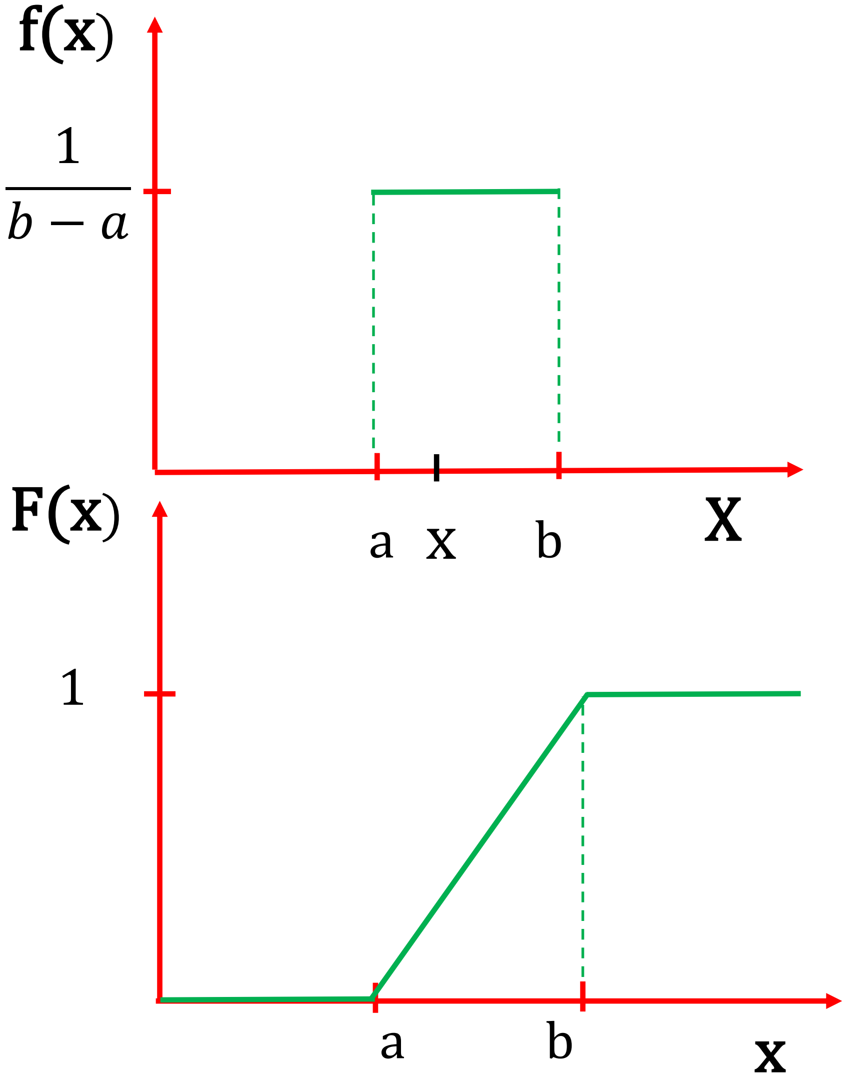
- A uniform distribution is denoted as U(a, b)



CDF of a Uniform RV

CDF of Uniform RV is

$$F(X) = \begin{cases} 0 & \text{for } x < a \\ \dfrac{x-a}{b-a} & \text{for } a < x < b \\ 1 & \text{for } x > b \end{cases}$$



Moments of a Uniform RV

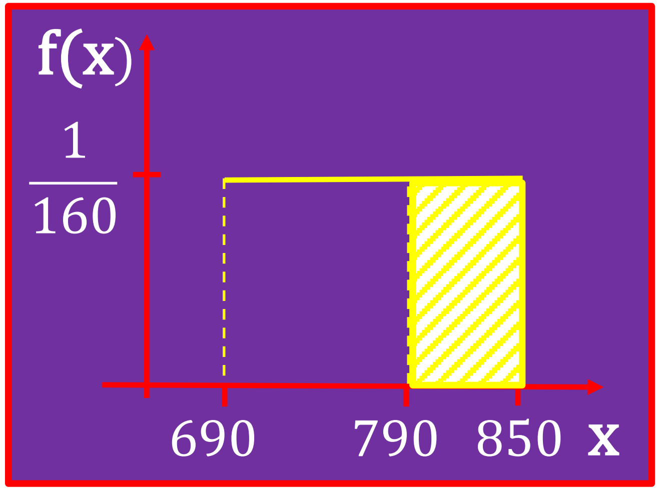
Mean and variance of Uniform RV

$$\mu_X = \frac{b-a}{2}$$

$$\sigma_X^2 = \frac{(b-a)^2}{12}$$

Uniform Distribution - Example

A random number N is taken from 690 to 850 in uniform distribution.
Find the probability number N is greater than the 790?



Solution:

Interval of numbers in probability distribution = [690, 850]

$$f(x) = \frac{1}{b-a} = \frac{1}{850-690} = \frac{1}{160}$$

We are looking for probability $P(790 < x < 850)$

Interval of probability distribution of successful event = [790, 850] = 60

The probability ratio $\frac{60}{160} = 0.375$

Hence the probability of N number greater than 790 = 37.5%

Continuous RV: Exponential

- The probability distribution of the time between successive random events for the same conditions as apply to the Poisson distribution.
- Used for studies of reliability and of queuing theory.
  - ✓ Queuing theory gives probability as a function of waiting time in a queue for service.
  - ✓ For example: what is the probability that the time between arrival of 2 customers at a service counter will be more than a stated time, e.g. three minutes?

Continuous RV: Exponential

- If X is an RV Po(λ), denoting the number of times an event occurs in an intervall,
- Then RV Y, expressing the time before the 1st success occurs, has an exponential distribution with following:

PDF of Exponential RV

$$f(Y) = \lambda e^{-\lambda y}, \quad \text{for } \lambda > 0 \;\; \& \;\; y \geq 0$$

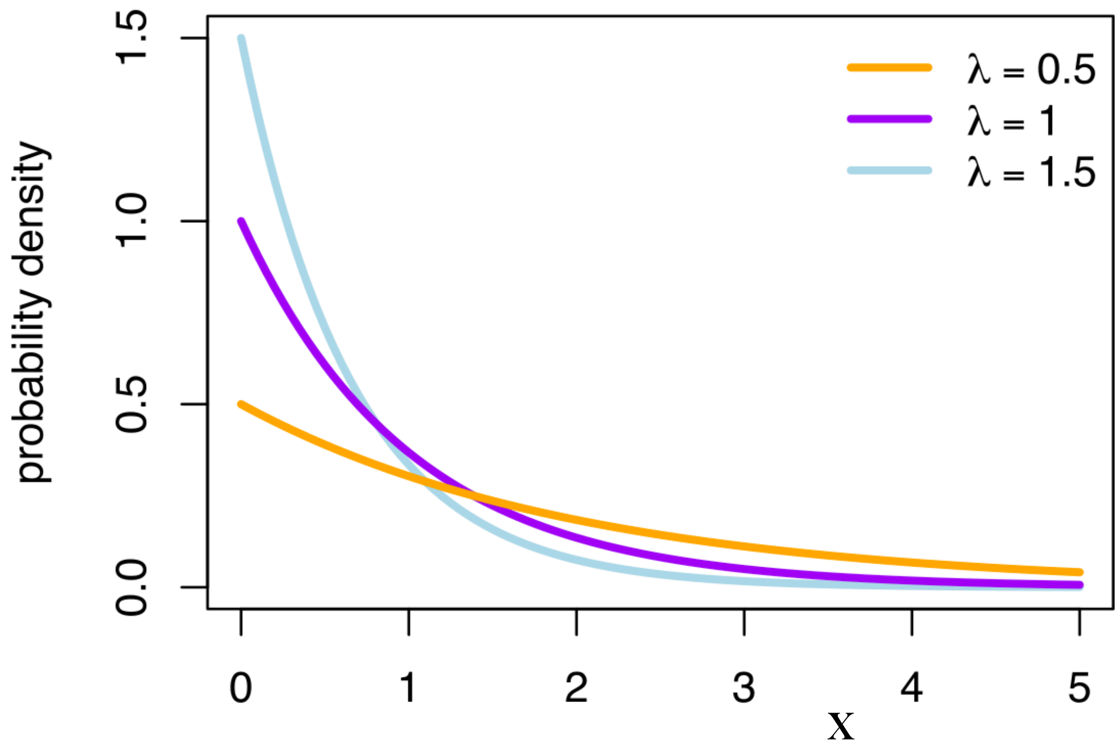
$$\mathrm{F(Y)} = 1 - \mathrm{e}^{-\lambda y} \qquad \text{for y} \geq 0$$

$$\mu_{\mathrm{Y}} = \frac{1}{\lambda} \quad \text{and} \quad \sigma_{\mathrm{Y}}^2 = \left(\frac{1}{\lambda}\right)^2$$

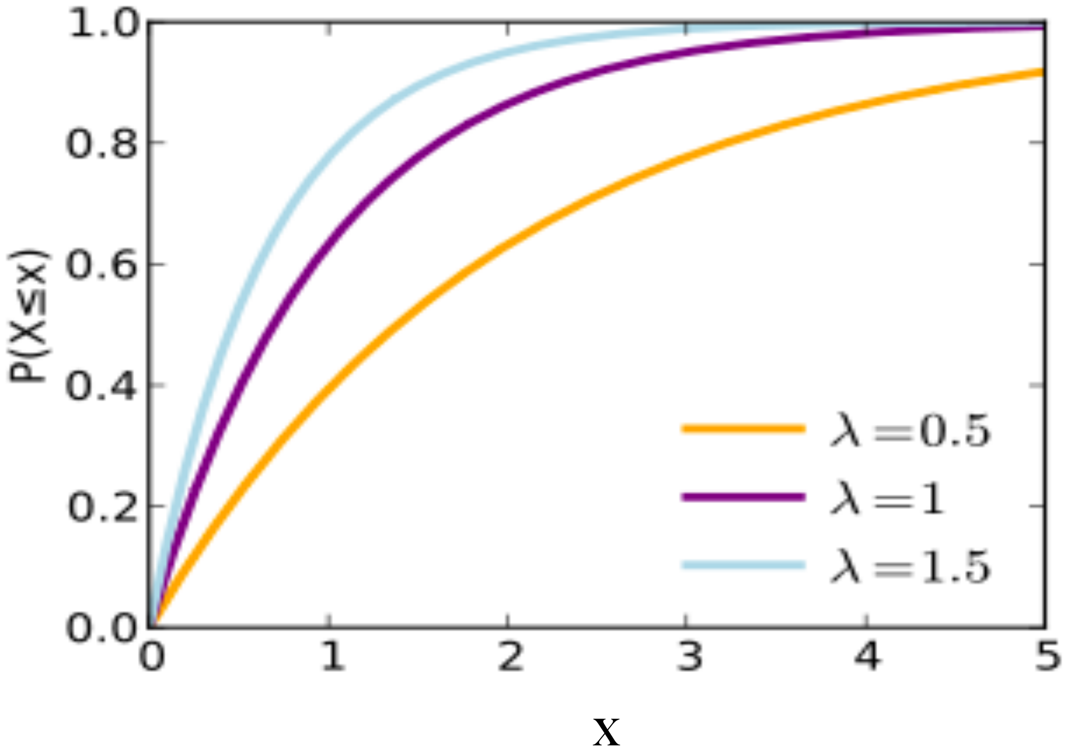
$$PMF\ Po(\lambda):\ f(X) = \frac{\lambda^x e^{-\lambda}}{x!}$$

Continuous RV: Exponential

PDF



CDF



Memoryless Property of Exponential RV

Example:

The number of miles that a car can run, before its battery wears out, is exponentially distributed with $\mu_Y = 10{,}000$ km. What is the probability that the driver will be able to complete the 5000 km trip without having to replace the car battery?

Solution:

Let set X= no. of km before battery has to be changed

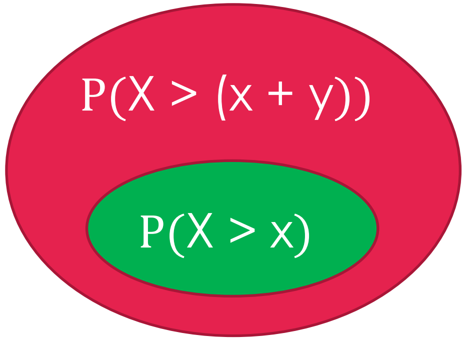
x - no. of km the car has driven

y – 5000 km

Find: $P\{(X > (x + y))|(X > x)\}$ i.e., the probability that the no. of km after the journey $(x + y)$ is less than the no. of km, after which the battery must be changed.

Memoryless property of Exponential RV

From conditional probability: $P(B|A) = \frac{P(AB)}{P(A)}$



$$P\{(X > (x + y)\ )\,|\ (X > x)\} = \frac{P\{(X > (x + y)) \cap (X > x)\}}{P(X > x)}$$

If X > (x + y), then X > x is redundant (in our case the car is running so X is >x)

$$P\{(X > (x + y))|(X > x)\} = \frac{P(X > (x + y))}{P(X > x)} = \frac{1 - P(X \ \leq \ (x + y))}{1 - P(X \ \leq \ x)} \leftarrow \text{CDF}$$

$$= \frac{1 - (1 - e^{-\lambda(x+y)})}{1 - (1 - e^{-\lambda x})} = \frac{e^{-\lambda x} e^{-\lambda y}}{e^{-\lambda x}} = e^{-\lambda y} = P(X > y)$$

Memoryless Property of Exponential RV

- The conditional probability does not depend on x!
- The probability of an exponential RV exceeding the value x+y, given x, equals the probability of the RV originally exceeding that value y, regardless of x
- The exponential distribution is memoryless because the past has no bearing on its future behaviour.
- Every instant is like the beginning of a new random period, which has the same distribution regardless of how much time has already elapsed.
- The exponential is the only memoryless continuous random variable

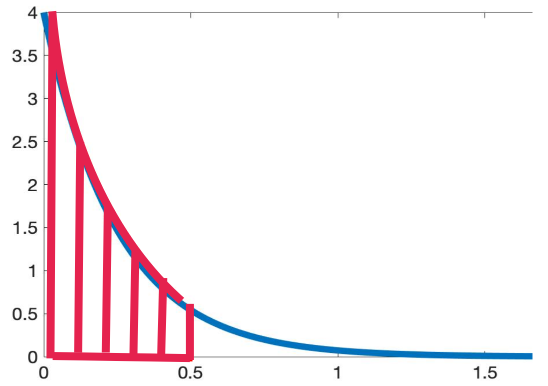
Exponential RV - Example

1. If jobs arrive every 15 seconds on average (λ = 4 per minute), what is the probability of waiting less than or equal to 30 seconds, i.e. 0.5 min?

$\mu_Y = \frac{1}{\lambda}$

Solution:

$$P(X \le 0.5) = \int_0^{0.5} 4e^{-4t}\, dt = -e^{-4t}\Big|_0^{0.5} = 1 - e^{-2} = 0.86$$



2. What is the max. waiting time between 2 jobs submissions with 95% confidence?

To find k so that $P(X \le k) = 0.95$, use quantile function, which specifies the value of RV, for which the probability $P(X \le x)$ is equal to a given value.

For exponential distribution:

$$Q(p, \lambda) = \frac{-\ln(1-p)}{\lambda} = \frac{-\ln(1-0.95)}{4} = 0.7489 \text{ min}$$

Continuous RV: Normal

PDF of Normal RV

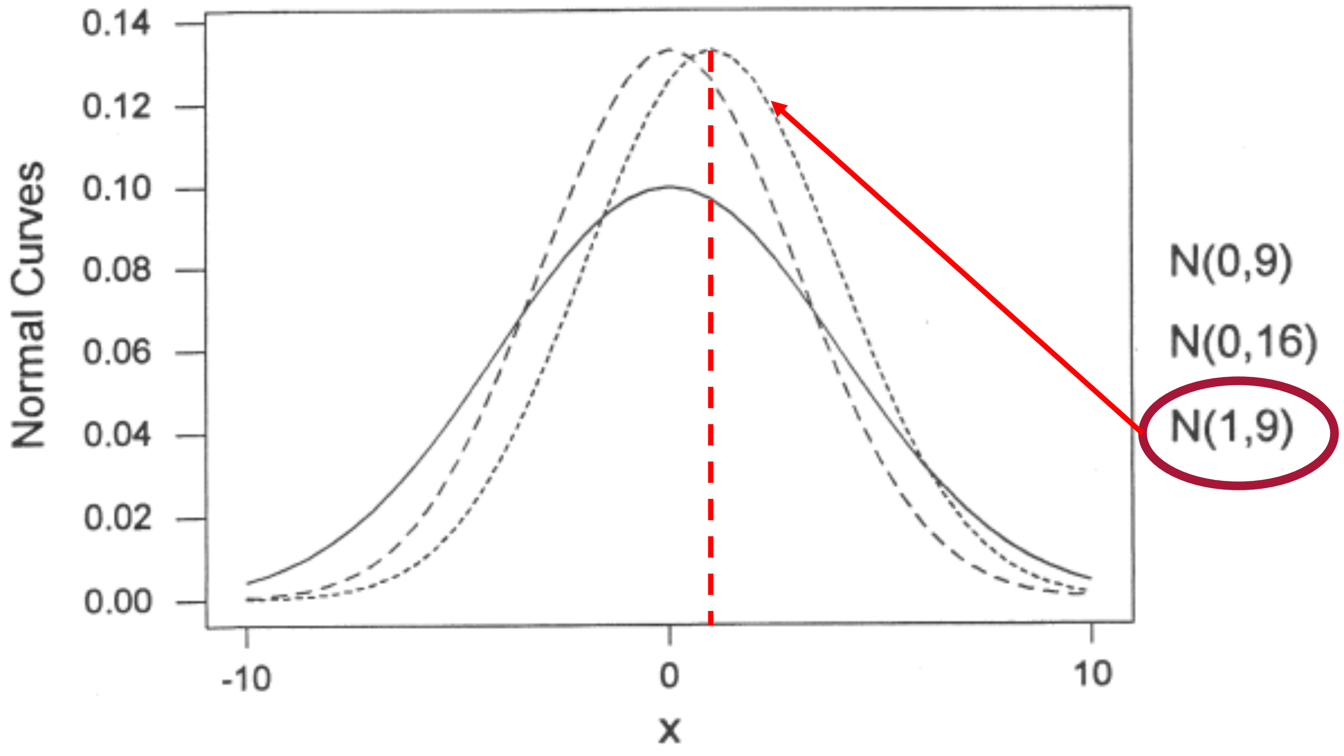
$$f(X) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

for $-\infty < x < \infty$,
$-\infty < \mu < \infty$
$0 < \sigma < \infty$

- All normal curves are **bell-shaped** with points of inflection at $\mu \pm \sigma$.
- All normal curves are **symmetric about the mean μ** with max. height at **x** = **μ**
- $\lim_{x \to -\infty} f(x) = 0$ and $\lim_{x \to \infty} f(x) = 0$
- As $\int_{-\infty}^{\infty} e^{-\left(\frac{t}{\alpha}\right)^2} dt = \sqrt{\alpha\pi}$ we need to normalise f(x) to obtain $\int_{-\infty}^{\infty} f(x)dx = 1$
- $\alpha = \sqrt{2}\sigma$ so we normalise f(x) by $\frac{1}{\sigma\sqrt{2\pi}}$

Continuous RV: Normal

- The shape of normal curve depends on its mean μ and standard deviation σ.

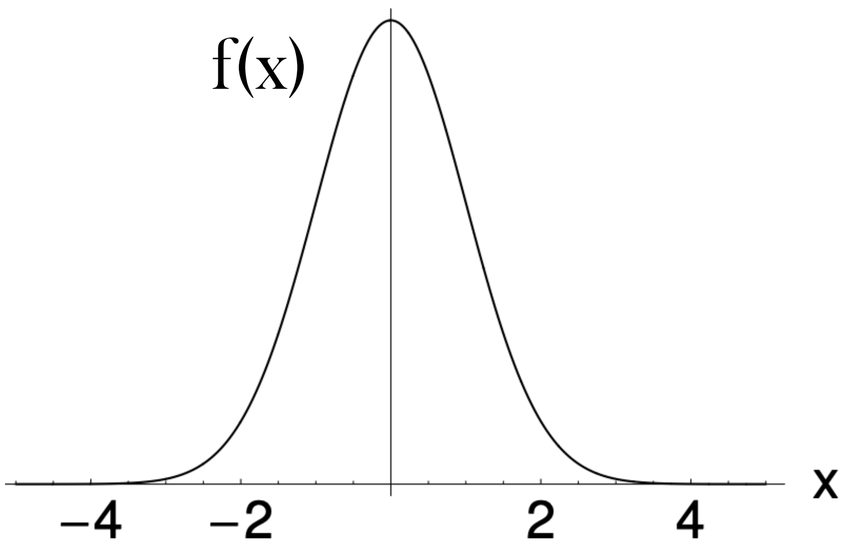


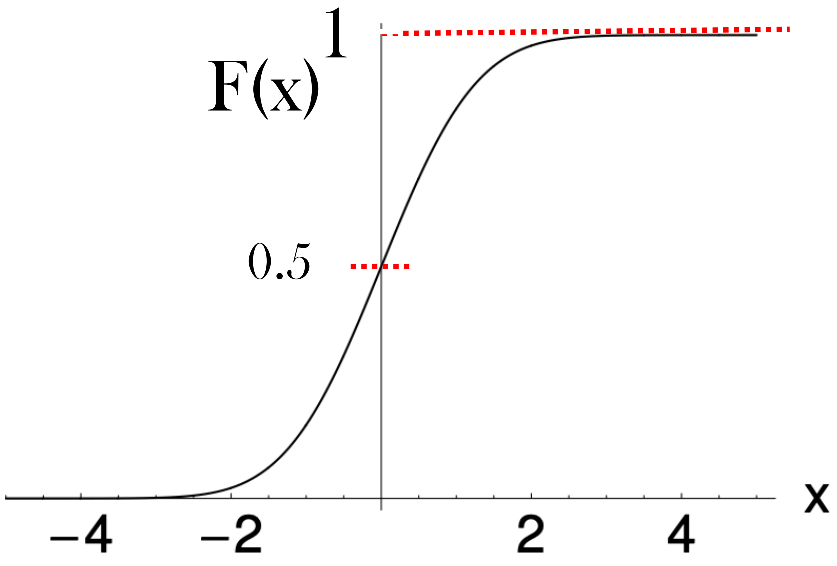
- Normal distribution is also called Gaussian distribution

CDF of Normal Distribution

- For continuous RV the probability is the area under the f(x)

$$F(X) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx^*$$





* https://www.youtube.com/watch?v=3K6UUAvFCjg

CDF of Normal Distribution

- Difficult to integrate normal PDF – we use tables
- Since there are infinite no. of normal PDFs, we use the standardised normal distribution with $\mu = 0$ and σ^2=1

If X ~ N(μ , σ^2) then

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

follows the standard normal distribution N(0,1)

- Z – # of standard deviations between any point and μ

Calculating probabilities using tables

Heights of students are normally distributed with mean=150 cm and stand. dev. of 20 cm. What proportion of the student's population will be shorter than 161.4 cm?

Solution:

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

Table A (*Continued*)

z	.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

Calculating probabilities using tables

$$Z = \frac{X - \mu}{\sigma} = \frac{161.4 - 150}{20} = \frac{11.4}{20} = 0.57$$

Z=0.57 means that 161.4 cm is 0.57 σ above the mean

$$P(X < 161.4) = 0.7157 = 72\%$$

Table A (*Continued*)

z	.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

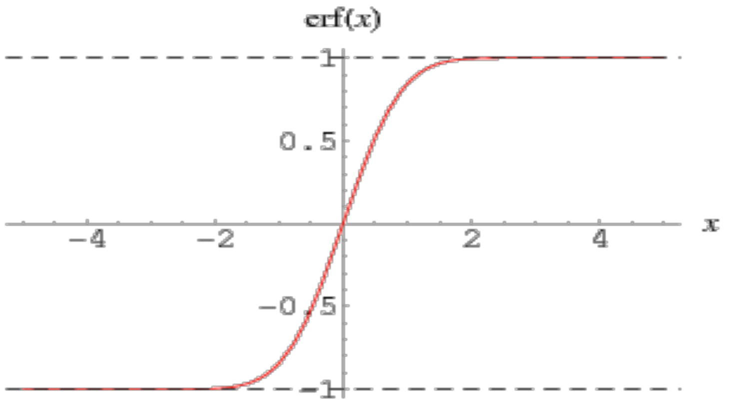
Error Function

- Used to calculate the probabilities for N. distribution P(-x<X<x)

Error function

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt$$

- Uses the symmetry of N distribution about y-axis (i.e. 2x integral form 0 to x)
- Calculation done using Maclaurin series:
- $\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt = \frac{2}{\sqrt{\pi}} \left[x - \frac{x^3}{3} - \frac{x^5}{10} - \cdots . \right]$
- erf(-x)=-erf(x) – odd function



Finding probabilities for Normal RV

- CDF of normal distribution can be expressed using Erf(x)
- Since $\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt$ we can write it as CDF(x)-CDF(0)
- $F(X) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{z^2}{2}} dz$
- $\operatorname{erf}(x) = 2\left(\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x\sqrt{2}} e^{-\frac{z^2}{2}} dz - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{0} e^{-\frac{z^2}{2}} dz \right)$ (change variable from t to $\frac{z}{2}$)
- Since F(∞)=1 and F(0)=0.5 we get

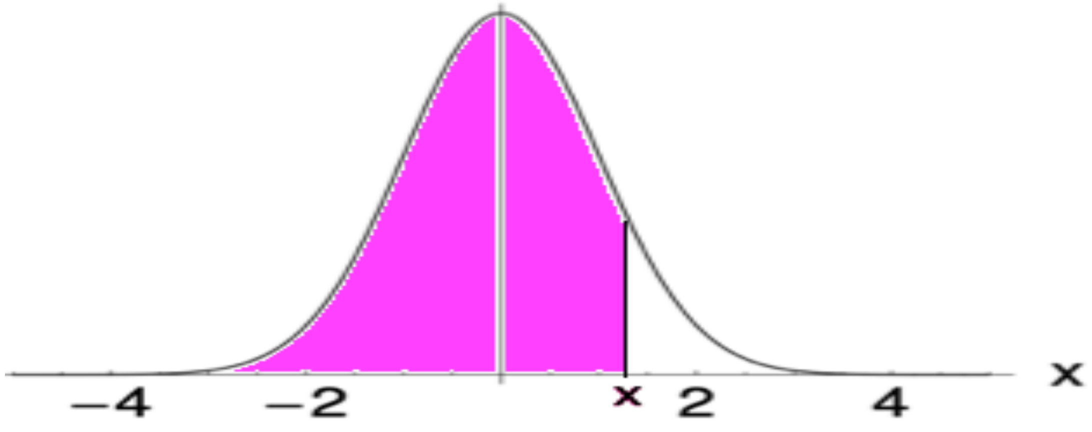
$$\operatorname{erf}(x) = 2\left(F(x\sqrt{2}) - \frac{1}{2}\right) \text{ and}$$

$$F(x\sqrt{2}) = \frac{1}{2}[1 + erf(x)]$$

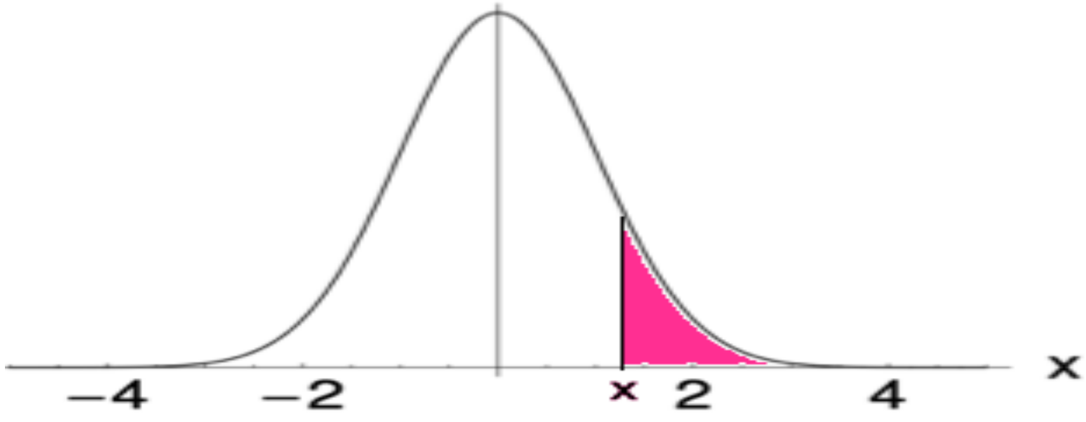
Finding probabilities for Normal RV

CDF

$$CDF = P(X \le x) = \frac{1}{2}\left[1 + erf\left(\frac{x-\mu}{\sigma\sqrt{2}}\right)\right]$$

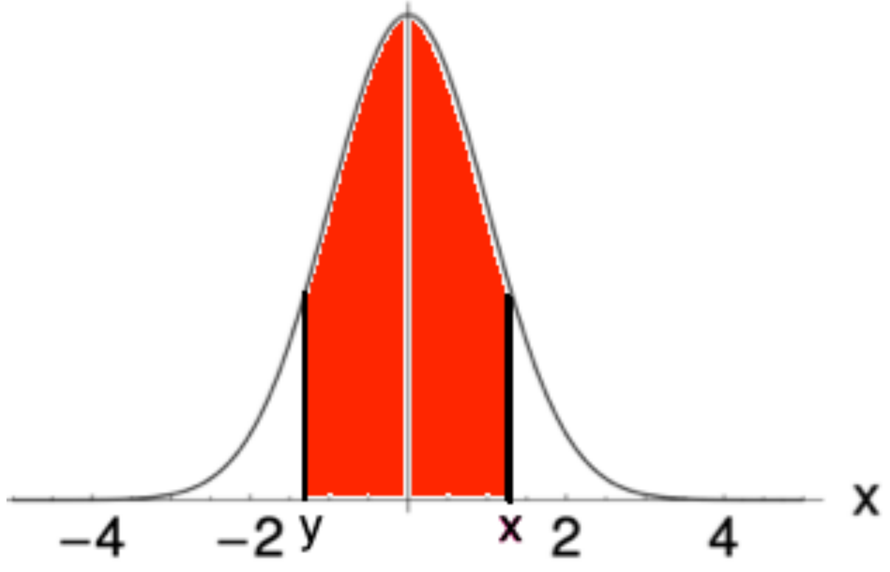


$P(X \le x)$



$P(X > x) = 1 - P(X \le x)$

Finding probabilities for Normal RV



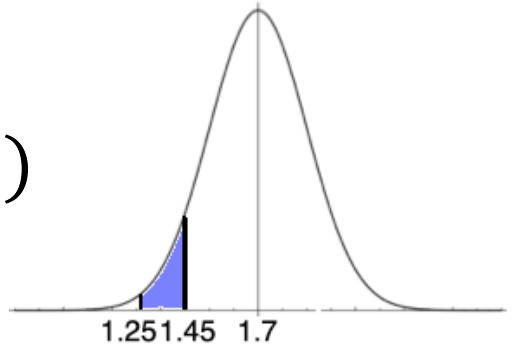
$$P(y < X < x)$$
$$= P(X < x) - P(X < y)$$

$$P(X < y) \cup P(X > x)$$
$$= 1 - P(y < X < x)$$

https://newonlinecourses.science.psu.edu/stat414/node/150/

Calculating probabilities using erf(x)

In a class with 100 students the mean height of a student is 1.7 m and stand. dev. is 0.2 m. Calculate the probability that a student will be taller than 1.25m and shorter than 1.45 m.



Solution: $P(1.25 < X < 1.45) = P(X < 1.45) - P(X < 1.25)$

$$= \frac{1}{2}\left[1 + erf\left(\frac{1.45-1.7}{0.2\sqrt{2}}\right)\right] - \frac{1}{2}\left[1 + erf\left(\frac{1.25-1.7}{0.2\sqrt{2}}\right)\right]$$

$$= \frac{1}{2}\left[erf\left(\frac{-0.25}{0.2\sqrt{2}}\right) - erf\left(\frac{-0.45}{0.2\sqrt{2}}\right)\right] =$$

-0.88 -1.59

erf(-x)=-erf(x)

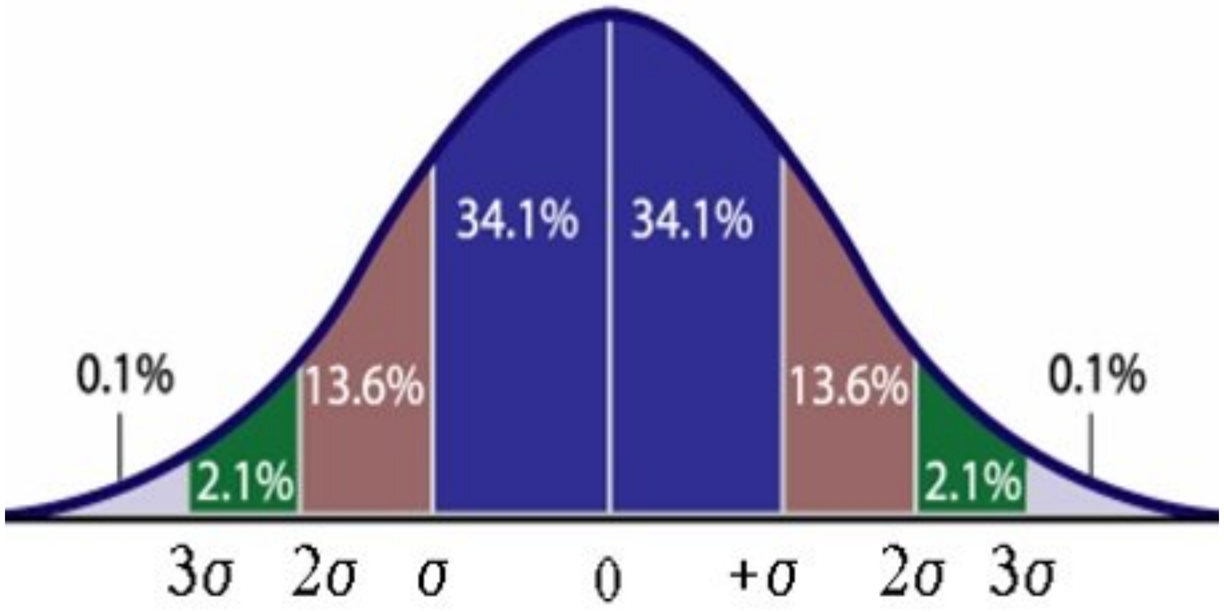
$$= \frac{1}{2}[-0.7866 - (-0.9754)] = 0.093 \text{ or } 9 \text{ students}$$

TABLE 5.1 Tabulation of Error Function Values

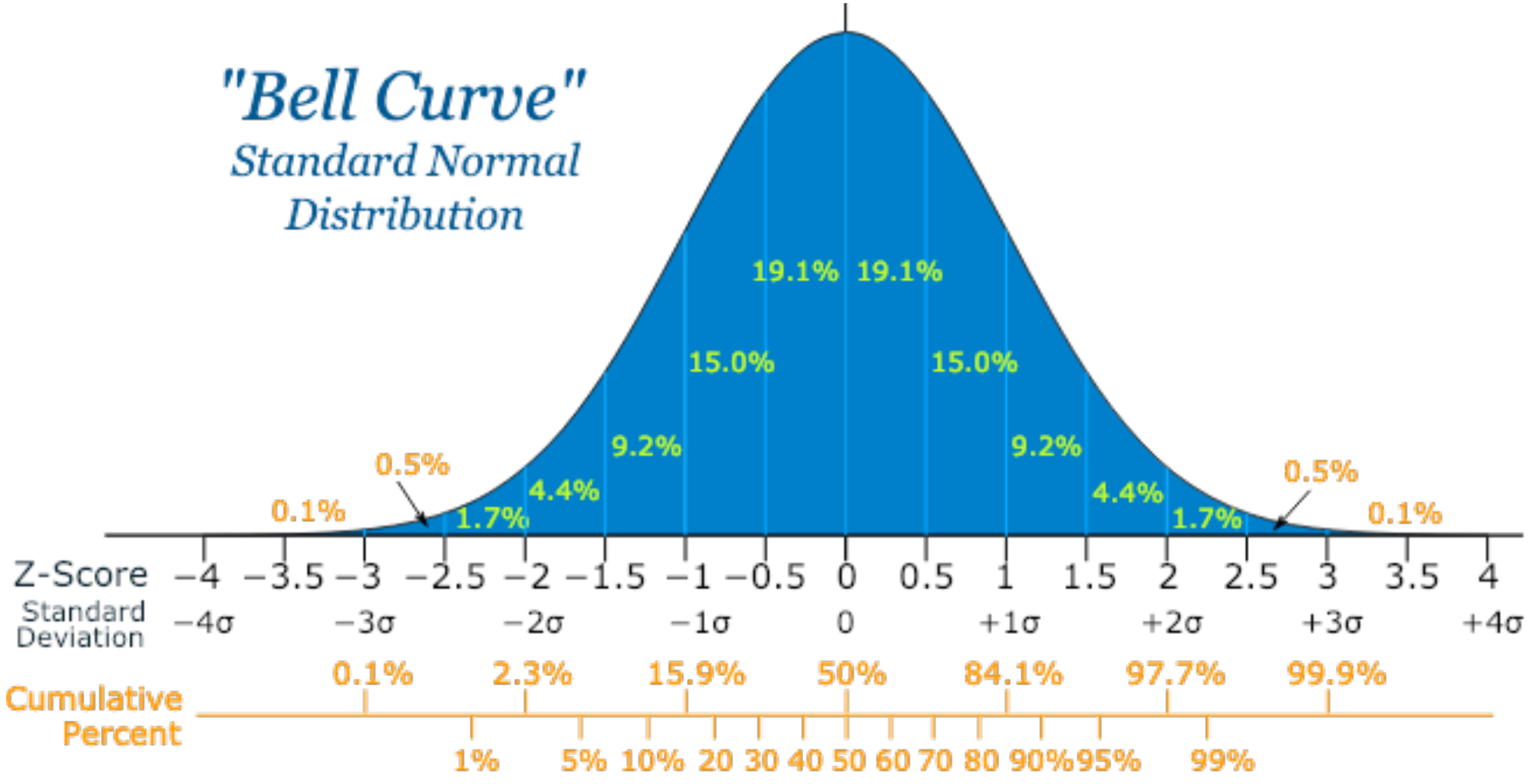
z	$erf(z)$	z	$erf(z)$	z	$erf(z)$
0	0	0.55	0.5633	1.3	0.9340
0.025	0.0282	0.60	0.6039	1.4	0.9523
0.05	0.0564	0.65	0.6420	1.5	0.9661
0.10	0.1125	0.70	0.6778	1.6	0.9763
0.15	0.1680	0.75	0.7112	1.7	0.9838
0.20	0.2227	0.80	0.7421	1.8	0.9891
0.25	0.2763	0.85	0.7707	1.9	0.9928
0.30	0.3286	0.90	0.7970	2.0	0.9953
0.35	0.3794	0.95	0.8209	2.2	0.9981
0.40	0.4284	1.0	0.8427	2.4	0.9993
0.45	0.4755	1.1	0.8802	2.6	0.9998
0.50	0.5205	1.2	0.9103	2.8	0.9999

Empirical rule of normal RV

- Approximately 68% of the data fall within 1 standard deviation of the mean
- Approximately 95% of the data fall within 2 standard deviations of the mean
- Approximately 99.7% of the data fall within 3 standard deviations of the mean



Empirical rule of Normal RV



Empirical rule - Example

A machine produces electrical components. 99.7% of the components have lengths between 1.176 cm and 1.224 cm. Assuming this data is normally distributed, what are the mean and standard deviation?

Solution:

The mean is halfway between 1.176 cm and 1.224 cm:

Mean = (1.176 cm + 1.224 cm)/2 = 1.200 cm

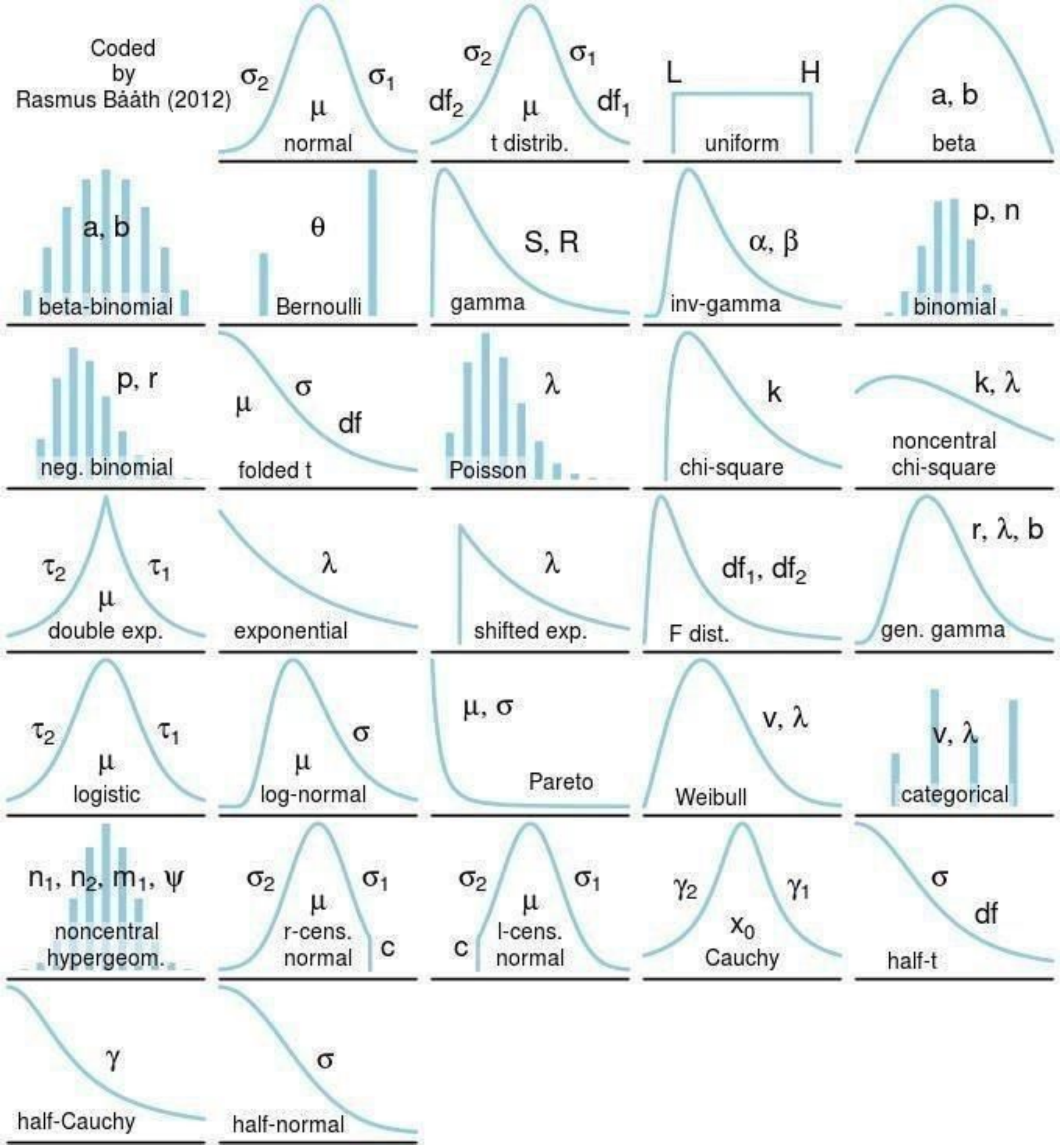
99.7% is 3 standard deviations either side of the mean (a total of 6 standard deviations) so:

1 standard deviation = (1.224 cm - 1.176 cm)/6 = 0.048 cm/6 = 0.008 cm

Gaussian noise

- Basic noise model used to mimic the effect of many random processes that occur in nature is additive white Gaussian noise (AWGN):
  - ✓ Additive - it is added to any noise that might be intrinsic to the information system.
  - ✓ White refers to the idea that it has uniform power across the frequency band for the information system. It is an analogy to the colour white, which has uniform emissions at all frequencies in the visible spectrum.
  - ✓ Gaussian because it has a **normal distribution** in the time domain with an average time domain value of zero

Some other distributions



What have we learnt?

- Continuous distributions
  - ✓ Uniform
  - ✓ Exponential
  - ✓ Normal
- Gaussian noise