Topic 6 Continuous Random Variables

Reference: Chapter 5.1-5.3

Probability Density Function

The Uniform Distribution

The Normal Distribution
- "Standardizing" a Normal Distribution
- Using the Standard Normal Table
- Linear Transformations of Random Variables

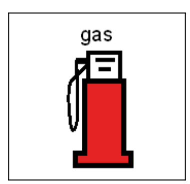
Further Examples of Continuous Distributions

Probability Density Function

If an experiment can result in an infinite, non-countable number of outcomes, then the random variable defined can be **"continuous"**.

Whenever the value of a random variable is measured rather than counted, a continuous random variable is defined.

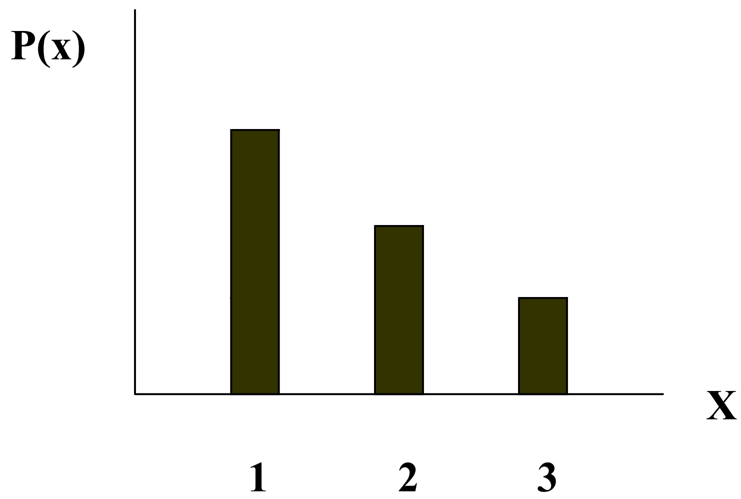
- Water level in the reservoir
- Distance between two points
- Amount of peanut butter in a jar
- Level of gasoline in fuel tank of a car



The values of the random variables in these examples can be any of an infinite number of values within a defined interval [a,b].

These random variables could be any number from **minus ∞ to plus b**.

Recall, that for the probability mass function, (p.m.f.), the value of P(X=x) was represented by the **height** of the spike at the point X=x.



One of the major differences between discrete and continuous probability distributions is that this representation **no** longer holds.

As you will see, a continuous p.d.f. is represented by the **area** between the x axis and the density function.

An important fundamental rule of the continuous random variable is when the random variable is continuous, the probability that any one specific value takes place is **zero**.

$$\mathbf{P(X=x)=}\left[\frac{1}{\infty}\right] = 0$$

•

Hence, we can determine probability values only for intervals, such as:

$P(a \leq x \leq b)$, (where a and b are some 2 points.)

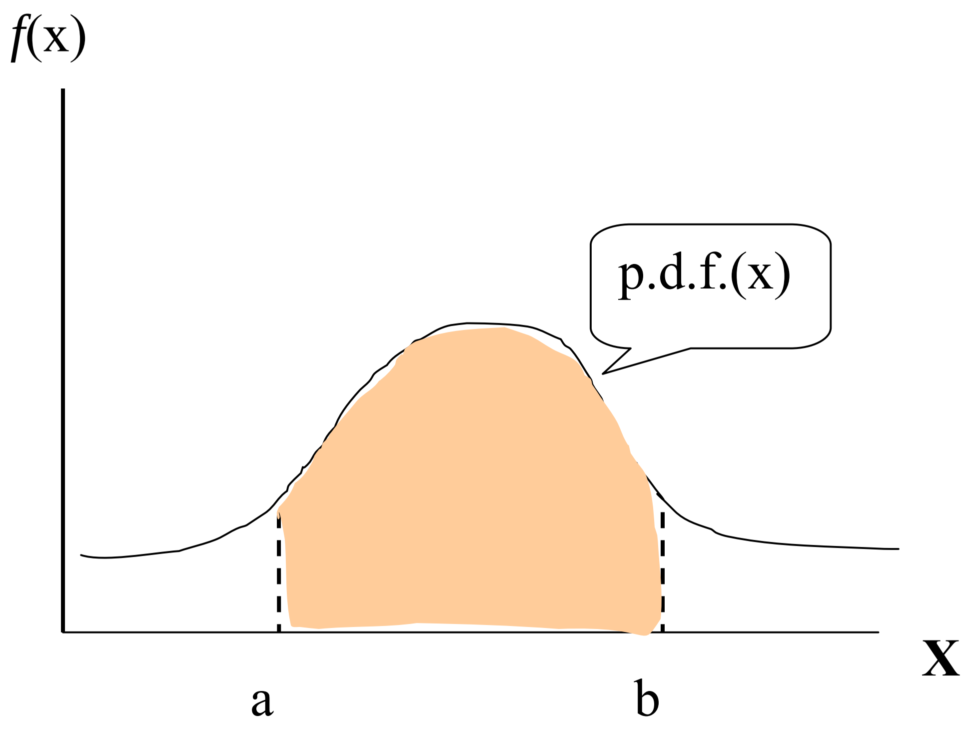
NOTE: f(x) does not represent probability!

→It represents how high or dense the function is at any specified value of x.

$P(X=x) = 0$ in the continuous case, so $f(x) \neq$ probability.

$P(a \leq x \leq b)$ = Area under $f(x)$ from a to b.

P(a <x < b) in the continuous case.



$$P(a < x < b) = \int_a^b f(x)dx$$

$$P(a < x < b) = \int_a^b f(x)dx$$

To determine the probability, one must **integrate** the function between two points.

Properties of All Probability Density Functions

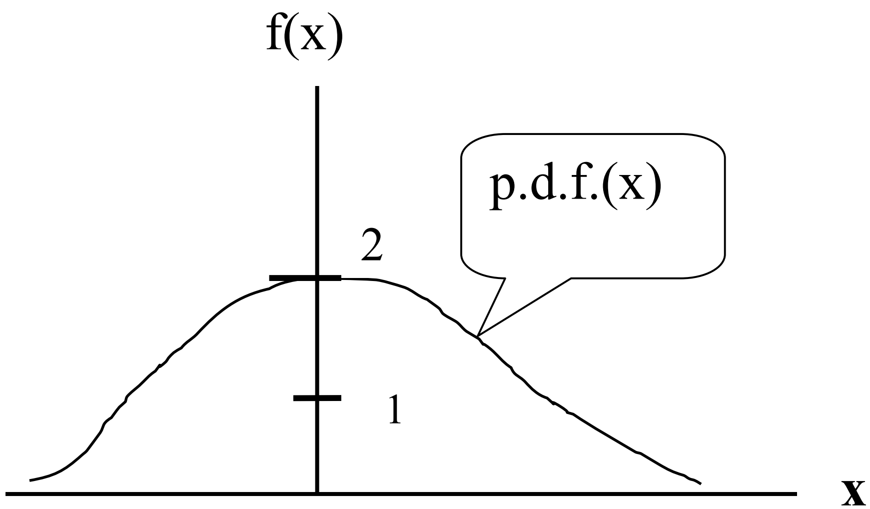
1) The total area under f(x), from -∞ to ∞ has to equal 1.

$$P(-\infty \le x \le \infty) = \int_{-\infty}^{\infty} f(x)dx = 1$$

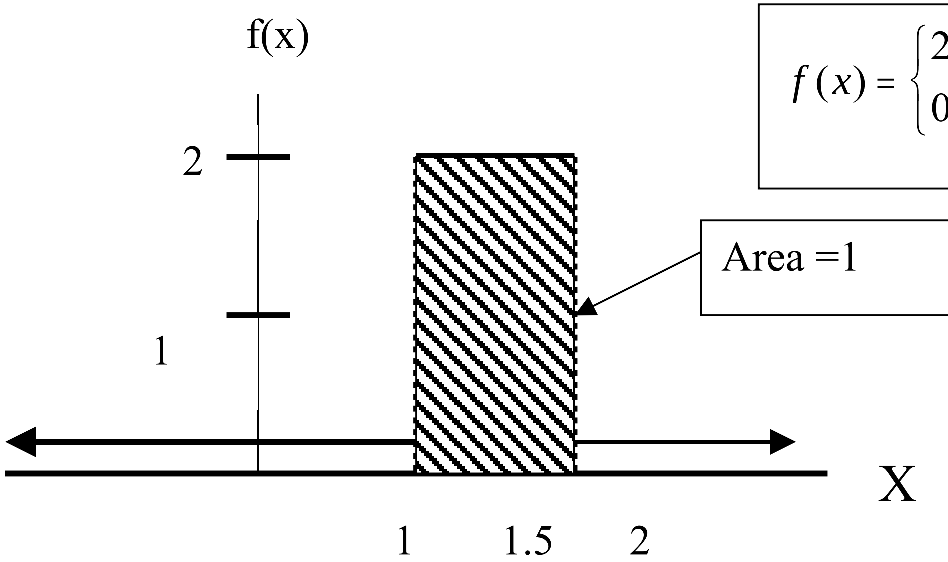
2) $f(x) \ge 0$ ⇐ The density function is never negative.

A major difference between the p.m.f. and p.d.f. is *f(x)* does not have to be $\le$ 1.

Also, for a continuous random variable, it does not matter whether the endpoints are included in the interval or not. The addition of the endpoint changes the probability only by the value of $\frac{1}{\infty} = 0,$ so it has no effect.



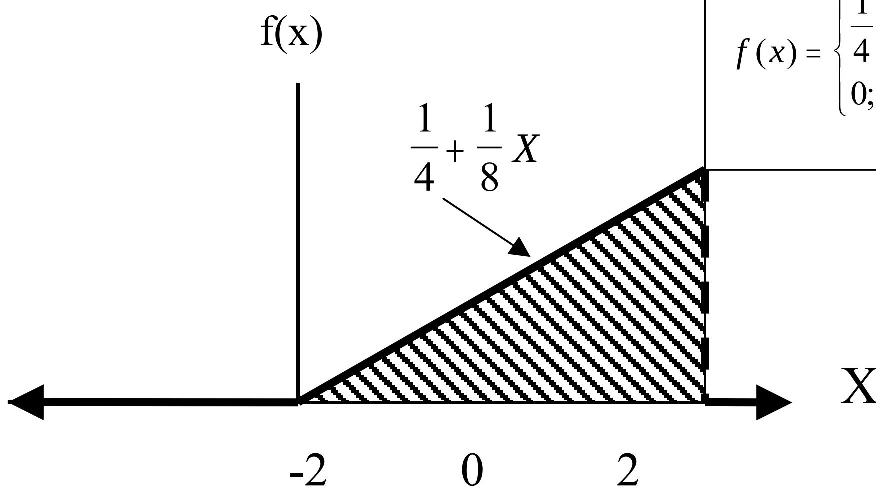
Example 1:



$$f(x) = \begin{cases} 2; & 1 \le x \le 1.5 \\ 0; & otherwise \end{cases}$$

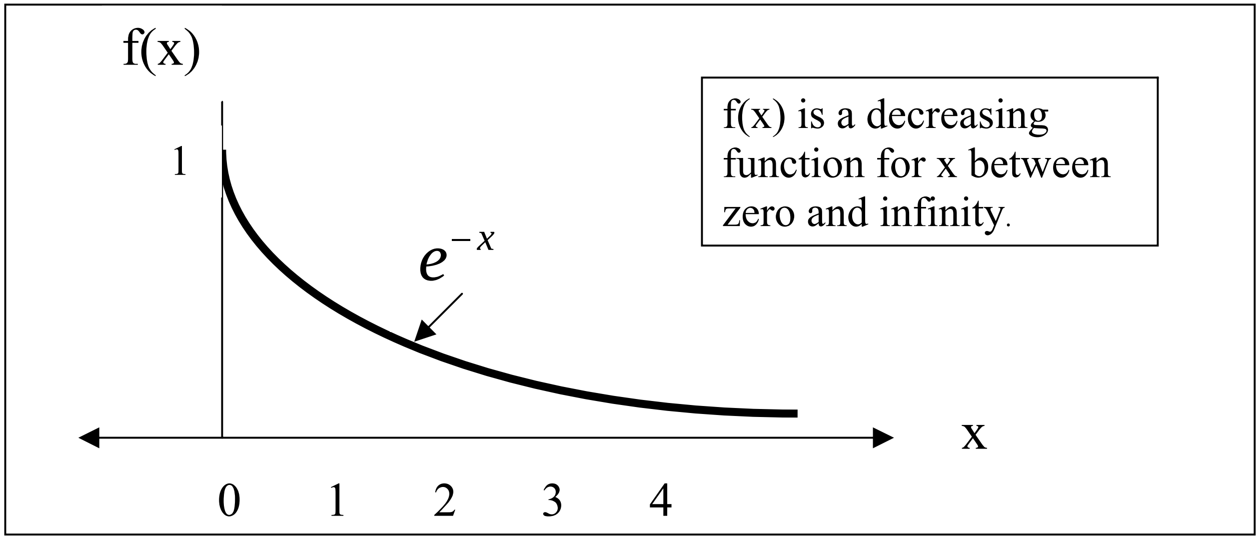
Example 2:

$$f(x) = \begin{cases} \frac{1}{4} + \frac{1}{8}X; & -2 \le x \le 2 \\ 0; & otherwise \end{cases}$$



f(x) is a straight line for values between x= -2 and x=2. (X does not have to be positive.)

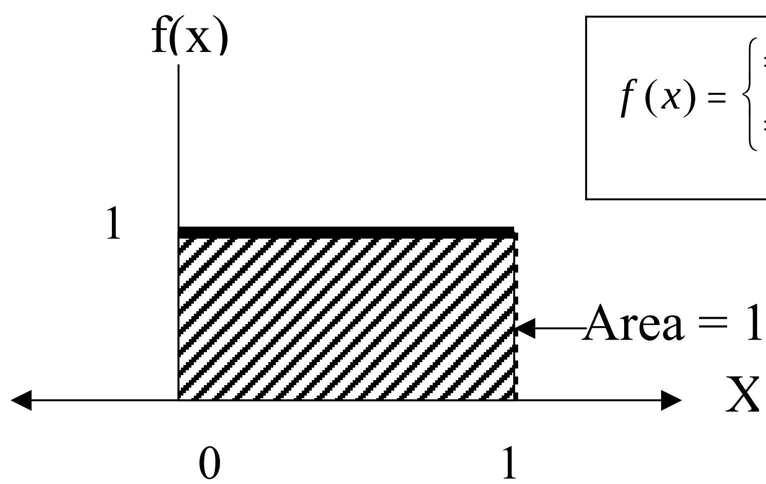
<u>Example 3</u>:


f(x)
1
e^{-x}
f(x) is a decreasing function for x between zero and infinity.
x
0 1 2 3 4

Uniform Distribution:

The uniform distribution is a rectangular distribution.

"Uniform" random variable.



$$f(x) = \begin{cases} = 1; & 0 < x < 1 \\ = 0; & otherwise \end{cases}$$

$$\int_{-\infty}^{\infty} f(x)dx = \int_{-\infty}^{0} f(x)dx + \int_{0}^{1} f(x)dx + \int_{1}^{\infty} f(x)dx$$

$$= \quad 0 \quad + \int_{0}^{1} f(x)dx \; + \; 0$$

$$= \int_{0}^{1} 1\text{dx} = x]_0^1 = 1$$

<u>Rule:</u> $\int dx = \int 1x^0 dx = \dfrac{1\times x^{0+1}}{0+1} = X]_0^1 = 1-0 = 1$

Example:

For the <u>uniform</u> distribution:

$$P(0 < x < \tfrac{1}{2}) = \int_{0}^{1/2} f(x)dx = x]_0^{1/2} = (\tfrac{1}{2} - 0) = \tfrac{1}{2}$$

Recall, in the discrete case, the cumulative distribution function:

$$F(X^*) = \sum_{x \le x^*} P(x)$$

and

$$E(X) = \sum_{x} xP(x)$$

$$V(x) = E(X^2) - [E(X)]^2;$$

In the **continuous case:**

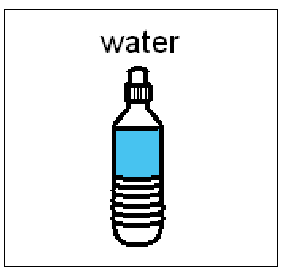
$$F(X^*) = \int_{-\infty}^{x^*} f(x)dx$$

and

$$E(X) = \int_{-\infty}^{\infty} xf(x)dx$$

$$V(x) = \int_{-\infty}^{\infty} x^2 f(x)dx - [E(x)]^2$$

$$= \left(\int_{-\infty}^{\infty} x^2 f(x)dx \right) - \left(\int_{-\infty}^{\infty} xf(x)dx \right)^2$$



From the last example: $f(x)$=1 for (0 < x < 1); 0 otherwise:

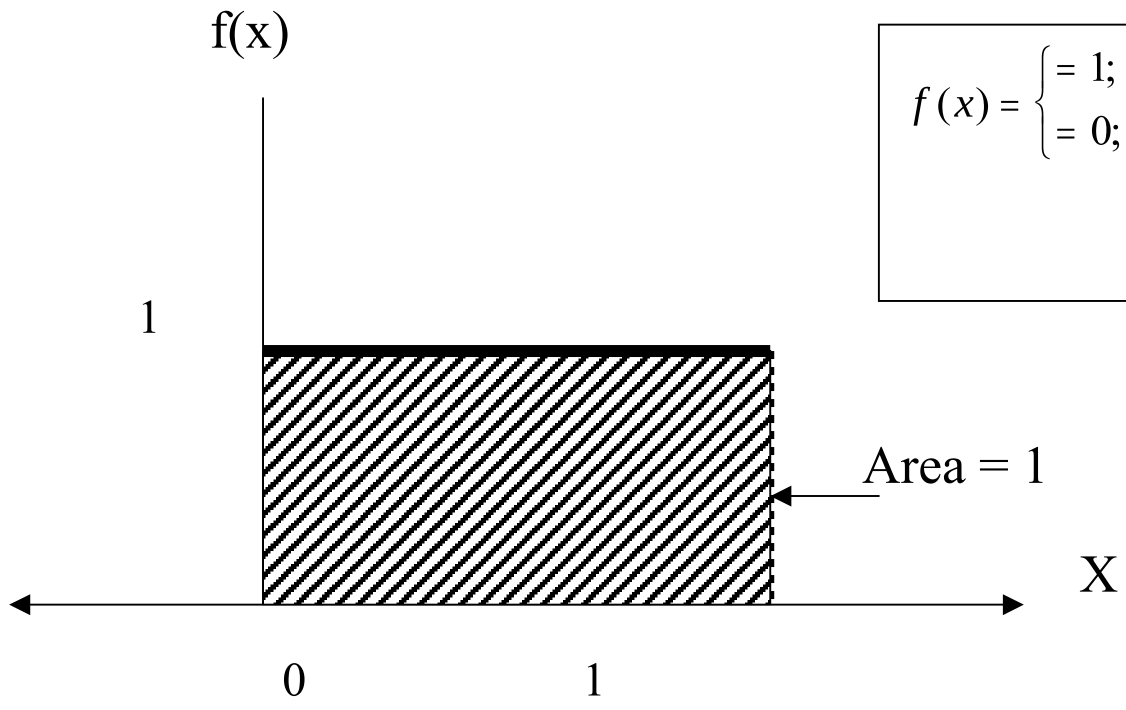
$$E(X) = \int_0^1 xf(x)dx = \int_0^1 1x dx = \tfrac{1}{2}x^2\Big]_0^1 = (\tfrac{1}{2} - 0) = \tfrac{1}{2}$$

$$V(x) = \int_0^1 (x-\mu)^2 f(x)dx = \int_0^1 (x-\tfrac{1}{2})^2 dx = \int_0^1 (x^2 - x + \tfrac{1}{4})dx$$

$$= \Big[\tfrac{1}{3}x^3 - \tfrac{1}{2}x^2 + \tfrac{1}{4}x\Big]_0^1 = \Big[\tfrac{1}{3} - \tfrac{1}{2} + \tfrac{1}{4}\Big] = \tfrac{1}{12}$$

$(Sheppard's\ Correction)$

“Uniform” random variable.



$$f(x) = \begin{cases} = 1; & 0 < x < 1 \\ = 0; & otherwise \end{cases}$$

The Cumulative Distribution Function:

Recall that for a discrete random variable, we obtained the cumulative probability distribution by **adding up** the probabilities accordingly:

For example:

x	p(x)	Σ p(x)
0	**0.3**	**0.3**
1	**0.5**	**0.8**
2	0.2	1

That is,

$$P(X \le 1) = 0.8$$

$$P(X \le 2) = 1.0; \quad etc.$$

Similarly, for a continuous random variable, we can work out:

F(b)=P(X≤b)=P(X<b)= $\int_{-\infty}^{b} f(x)dx.$

(F(b) = Area under the p.d.f. to the left of X=b.)

<u>Note:</u> ■Values of F(b) = probabilities.

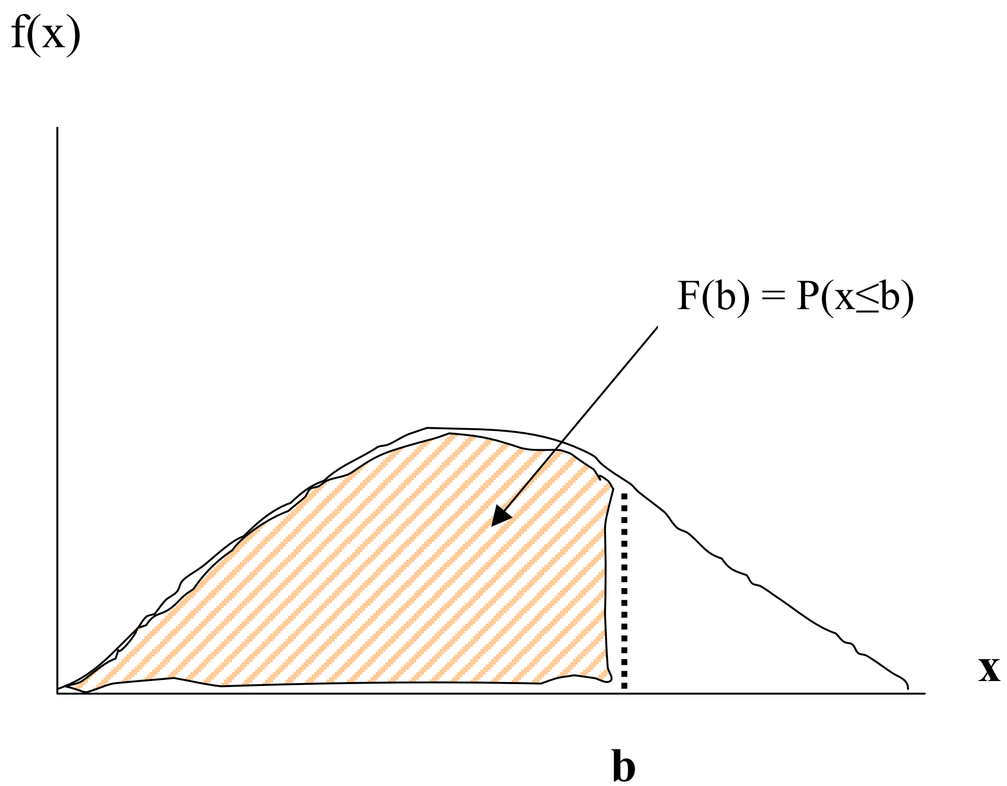
■Values of *f*(x)≠probabilities. (Density function)

For example: F(b) represents the probability that the random variable x assumes a value less than or equal to some specified value, say b.

To calculate F(b) in the continuous case, it is necessary to **integrate** *f*(x) over the relevant range, rather than sum discrete probabilities.

$F(b) = P(x \leq b)$ = all area under f(x) from $x \leq b$.

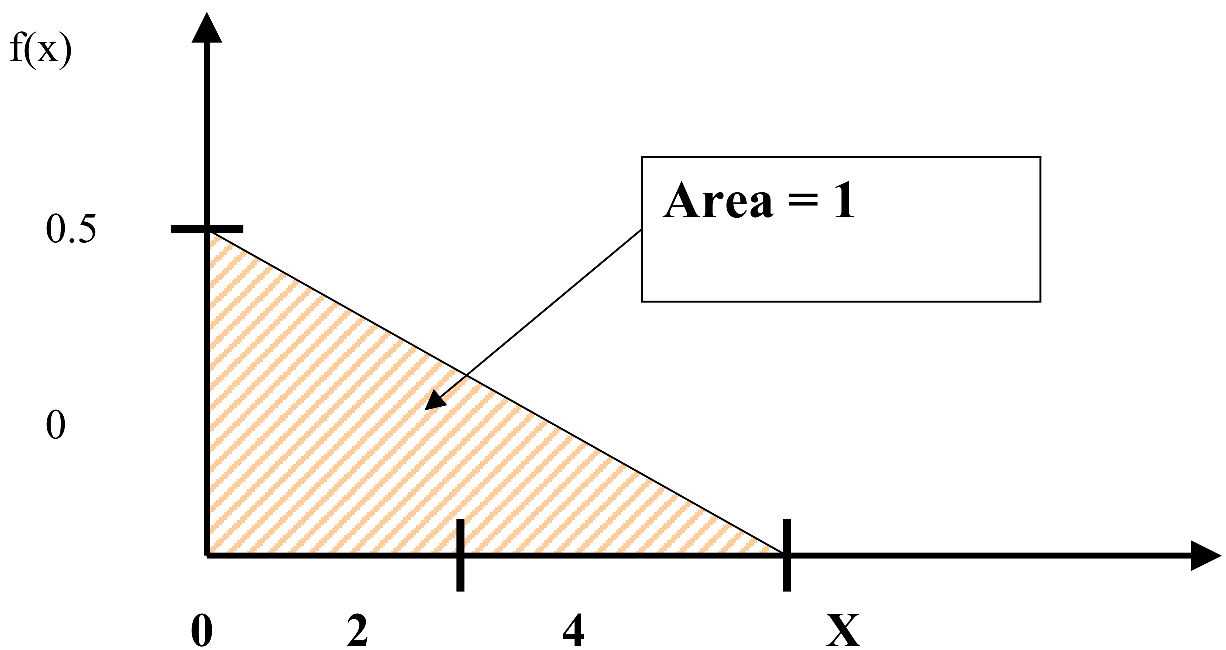
<u>Notation:</u> Values of F(x) represent probabilities, while values of *f(x)* do not.



Example: **Let X= time (months) taken to get a job.**

$$f(x) = \begin{cases} = \tfrac{1}{2} - \tfrac{1}{8}X; & 0 \le x \le 4 \\ = 0; & otherwise \end{cases}$$

First determine if this is a proper p.d.f..



(a)

$$\int_{-\infty}^{\infty} f(x)dx = \int_{0}^{4} f(x)dx = \int_{0}^{4} \left(\tfrac{1}{2} - \tfrac{X}{8}\right)dx$$

$$= \left(\tfrac{X}{2} - \tfrac{X^2}{16}\right]_0^4$$

$$= \left[\left(\tfrac{4}{2} - \tfrac{16}{16}\right) - \left(\tfrac{0}{2} - \tfrac{0}{16}\right)\right]$$

$$= \left[(2-1) - 0\right] = 1$$

So this is a “proper” p.d.f..

(b) What is the probability of waiting more than 2 months to get a job?

$$P(X > 2) = \int_2^4 f(x)dx = \int_2^4 \left(\tfrac{1}{2} - \tfrac{X}{8}\right)dx$$

$$= \left(\tfrac{X}{2} - \tfrac{X^2}{16}\right]_2^4$$

$$= \left[\left(\tfrac{4}{2} - \tfrac{16}{16}\right) - \left(\tfrac{2}{2} - \tfrac{4}{16}\right)\right]$$

$$= \left[(2 - 1) - (1 - \tfrac{1}{4})\right] = 0.25$$

or:

$$P(X > 2) = 1 - P(X \le 2) = 1 - F(2)$$

$$= 1 - \int_0^2 f(x)dx = 1 - \int_0^2 \left(\tfrac{1}{2} - \tfrac{x}{8}\right)dx$$

$$= 1 - \left(\tfrac{x}{2} - \tfrac{x^2}{16}\right]_0^2$$

$$= 1 - \left[\left(\tfrac{2}{2} - \tfrac{4}{16}\right) - \left(\tfrac{0}{2} - \tfrac{0}{16}\right)\right]$$

$$= 1 - \left[(1 - \tfrac{1}{4}) - 0\right] = 0.25$$

Mean and Variance:

Recall, for discrete random variables:

$$E(X) = \sum_x xP(x) = \mu$$

$$V(x) = E\left[(x-\mu)^2\right] = \sum_x (x-\mu)^2 P(x)$$

and

$$E(g(x)) = \sum_x g(x)P(x).$$

In the **continuous case**:

$$E(X) = \int_{-\infty}^{\infty} xf(x)dx = \mu$$

$$V(X) = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx$$

$$= \int_{-\infty}^{\infty} x^2 f(x)dx - \left[\int_{-\infty}^{\infty} xf(x)dx\right]^2$$

since

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

The summary measures of central location and dispersion are important in describing the density function.

The mean represents central location: the balancing point of p.d.f.:

$$\mu = E(x)$$

The variance and standard deviation of p.d.f.: dispersion

$$\sigma^2 = V(x) \quad and \quad \sigma = \sqrt{V(x)}$$

"Rule of Thumb" Interpretation of Standard Deviation:

$$\mu \pm \sigma \Rightarrow \text{ contains} \approx 68\% \text{ of probability (area)}$$

$$\mu \pm 2\sigma \Rightarrow \text{ contains} \approx 95\% \text{ of probability (area)}$$

For the continuous probability distribution, the weights are "areas" given by the f(x)dx (height times width).

Example: From the example on waiting time to get a job, what are the average and the standard deviation of waiting time?

$$f(x) = \begin{cases} = \tfrac{1}{2} - \tfrac{1}{8}X; & 0 \le x \le 4 \\ = 0; & otherwise \end{cases}$$

$$E(X) = \int_{-\infty}^{\infty} xf(x)dx = \mu$$

$$= \int_0^4 xf(x)dx = \int_0^4 x\left(\tfrac{1}{2} - \tfrac{X}{8}\right)dx = \int_0^4 \left(\tfrac{X}{2} - \tfrac{X^2}{8}\right)dx$$

$$= \left[\tfrac{X^2}{4} - \tfrac{X^3}{24}\right]_0^4 = \left[\left(\tfrac{16}{4} - \tfrac{64}{24}\right) - (0 - 0)\right] = \left(4 - \tfrac{8}{3}\right) = 1\tfrac{1}{3}\ months.$$

$$V(X) = E(x - E(x))^2 = E(x^2) - [E(x)]^2$$

$$= \int_0^4 (x)^2 f(x)dx - \left(\tfrac{4}{3}\right)^2 = \left[\int_0^4 (x)^2\left(\tfrac{1}{2} - \tfrac{X}{8}\right)dx\right] - \frac{16}{9}$$

$$= \int_0^4 \left(\tfrac{X^2}{2} - \tfrac{X^3}{8}\right)dx - \left(\tfrac{16}{9}\right) = \left[\tfrac{X^3}{6} - \tfrac{X^4}{32}\right]_0^4 - \left(\tfrac{16}{9}\right)$$

$$= \left[\left(\tfrac{64}{6} - \tfrac{256}{32}\right) - 0\right] - \left(\tfrac{16}{9}\right) = 0.889$$

$\Rightarrow$ standard deviation $= \sqrt{0.889} = 0.943\ months.$

The Median Value:

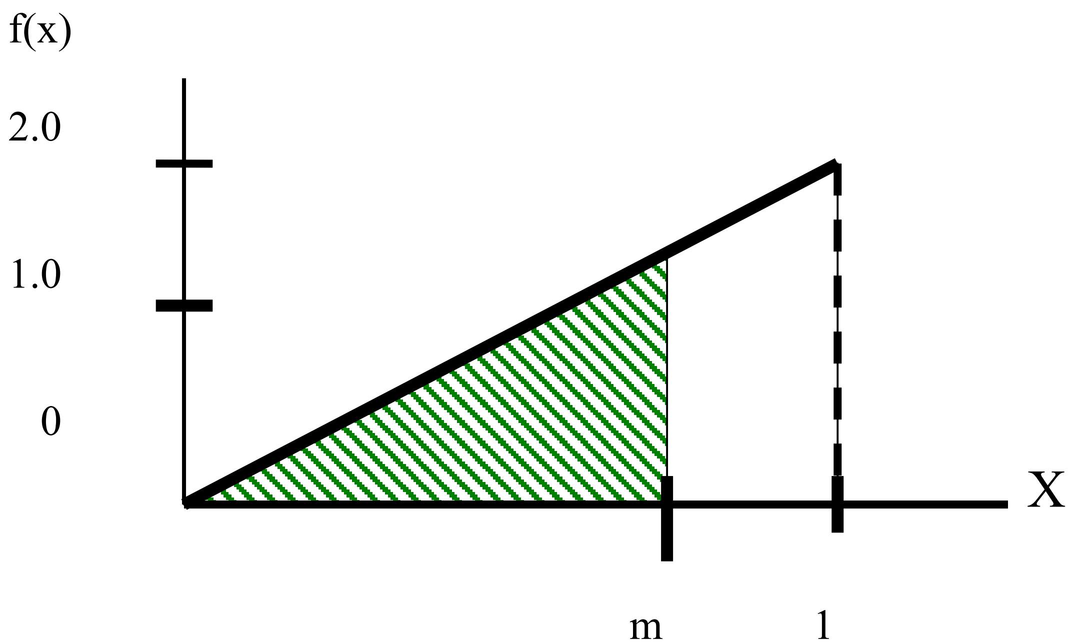
Recall that the median divides the (ranked) data into two equal parts.

For a continuous random variable we need to find the value of X, say "m", such that:

$$P(X \leq m)=0.5.$$

(That is the point where 50% of the area is on either side of x, or F(m) = 0.5, the cumulative area up to some point where area = 50%.)

<u>Example:</u> Suppose: $$f(x) = \begin{cases} = 2X; & 0 \leq x \leq 1 \\ = 0; & otherwise \end{cases}$$



Now find the median. Find "m" such that F(m) = 0.5

$$\int_0^m f(x)dx = 0.5$$

$$\int_0^m (2x)dx = \left[x^2\right]_0^m = 0.5$$

$$m^2 = 0.5$$

$$m = \sqrt{0.5} = 0.707$$

The Normal Distribution

The most important distribution in statistics is the **Normal** or **Gaussian** distribution:

(i) Relates to a ***continuous random variable*** that can take any real value.

(ii) Many natural phenomenons involve this distribution.

(iii) Even if a phenomenon of interest follows some other probability distribution, if we average enough I independent random effects, the result is a normal random variable.

A normal r.v. is characterized by two parameters:
→ Mean (μ)
→ Variance (σ^2)

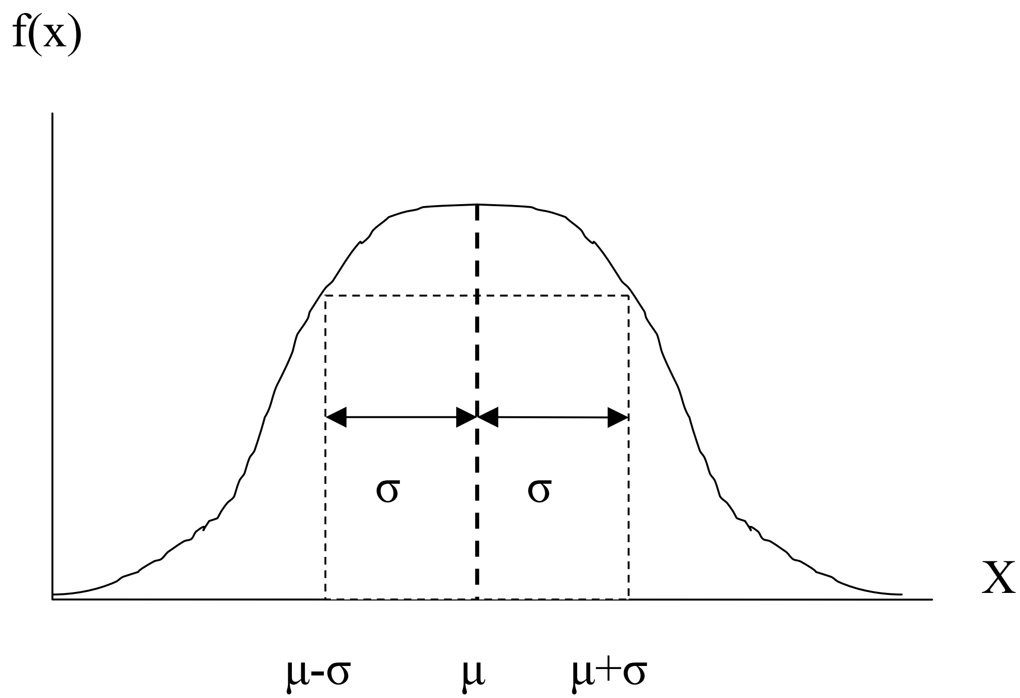
The formula for the p.d.f. when X is N(μ, σ^2) is:

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

$$where\ (-\infty < x < \infty).$$

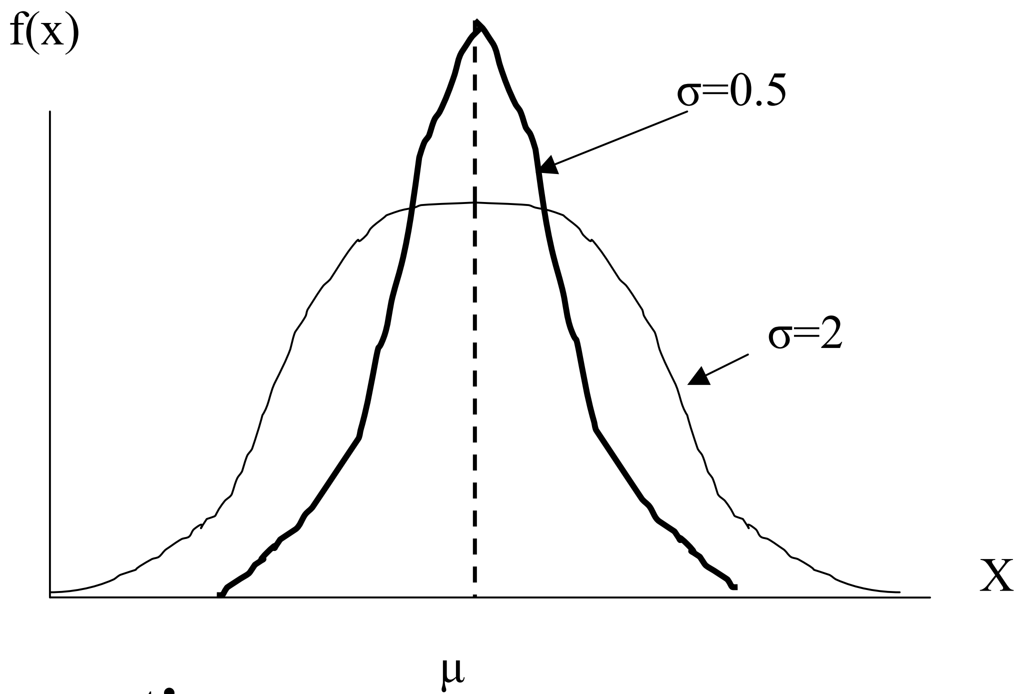
The normal distribution is a continuous distribution in which x can assume any value between minus infinity and plus infinity.

The normal density function is symmetrical bell-shaped probability density function.



Since Π and e are constants, if μ and σ are known, it is possible to evaluate areas under this function by using calculus. (i.e. integration.)

All normal distributions have the same bell shaped curve regardless of the μ (center) and σ (the spread or width) of the distribution.

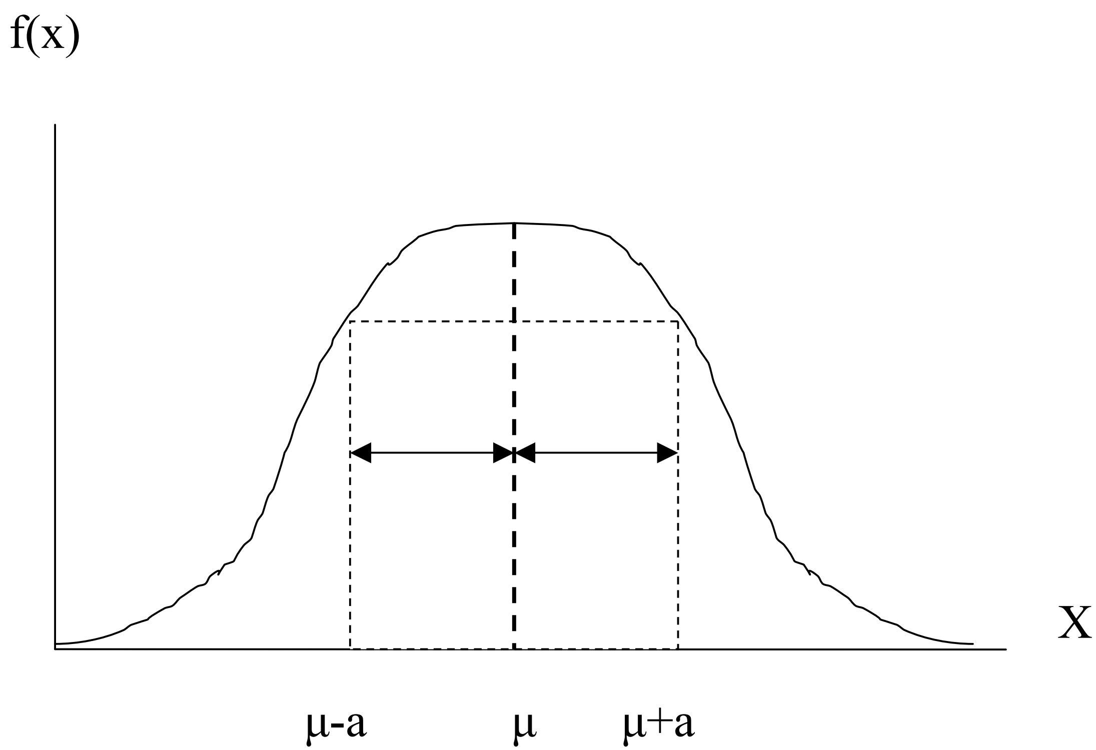


<u>Properties:</u>

A symmetric “bell shaped” p.d.f., centered at μ. So,

$$\boldsymbol{\mu} = \textbf{Mean} = \textbf{Mode} = \textbf{Median}$$

The "spread" or shape is determined by the variance.



Area under the curve = 1.

So, $P(x<\mu) = P(x>\mu) = 0.5$ by symmetry.
Similarly, $P(x > \mu+a)=P(x < \mu-a)$.

Example: Suppose X~N(μ=5,σ^2=4). What is the P(x≤4)? (F(4))

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

$$P(x \le 4) = \int_{-\infty}^{4} f(x)dx = \int_{-\infty}^{4} \frac{1}{2\sqrt{2\Pi}} e(-1/8)(x-5)^2$$

☺ ***This is going to be tedious***!

Fortunately, we can use an idea we met earlier to assist us.

We learned that we can **standardize** a random variable by subtracting its mean, and dividing by its standard deviation:

$$Z=(x-\mu) / \sigma.$$

The Two Properties of the Normal Standardized Random Variable:

1)E(Z) =0

Proof:

$$E(Z) = E\left(\frac{(x-\mu)}{\sigma}\right) = \left(\frac{1}{\sigma}\right)E(x-\mu)$$

$$= \left(\frac{1}{\sigma}\right)\left[E(x)-\mu\right]$$

$$= \left(\frac{1}{\sigma}\right)(\mu-\mu) = 0$$

Expected value of a standard normal variable is zero.

2)V(Z)=1

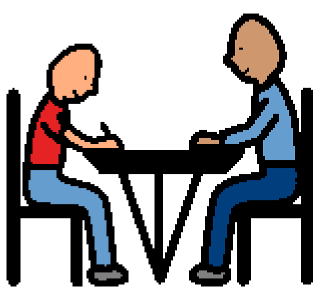
$$V(Z) = V\left(\frac{(x-\mu)}{\sigma}\right) = \left(\frac{1}{\sigma}\right)^2 V(x-\mu)$$

Proof:

$$= \left(\frac{1}{\sigma}\right)^2 \left[V(x)\right]$$

$$= \left(\frac{1}{\sigma^2}\right)(\sigma^2) = 1$$

Variance of a standard normal variable is one.



Standardized Normal

Values of x for the normal distribution usually are described in terms of how many standard deviations they are away from the mean.

Treating the values of x in a normal distribution in terms of standard deviations about the mean, has the advantage of permitting all normal distributions to be compared to one common or standard normal distribution.

It is easier to compare normal distributions having different values of μ and σ if these curves are transformed to one common form: **Standardized Normal Distribution.**

The standardized variable represents a "new" variable.

$$Z = \left[\frac{x - \mu}{\sigma} \right].$$

Using the Standardized Normal:

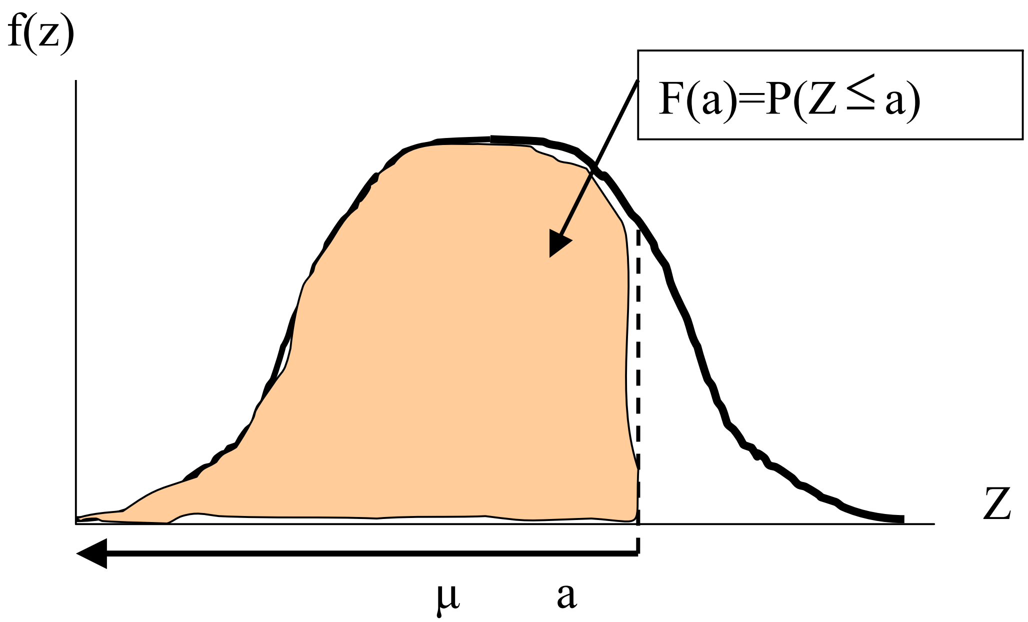
Since the normal distribution is completely symmetrical, tables of Z-values usually include only positive values of Z.

The lowest value in the text table is Z=0.00.

$F(0) = P(Z \leq 0) = 0.50$, the cumulative probability up to this point.

Four Basic Rules:

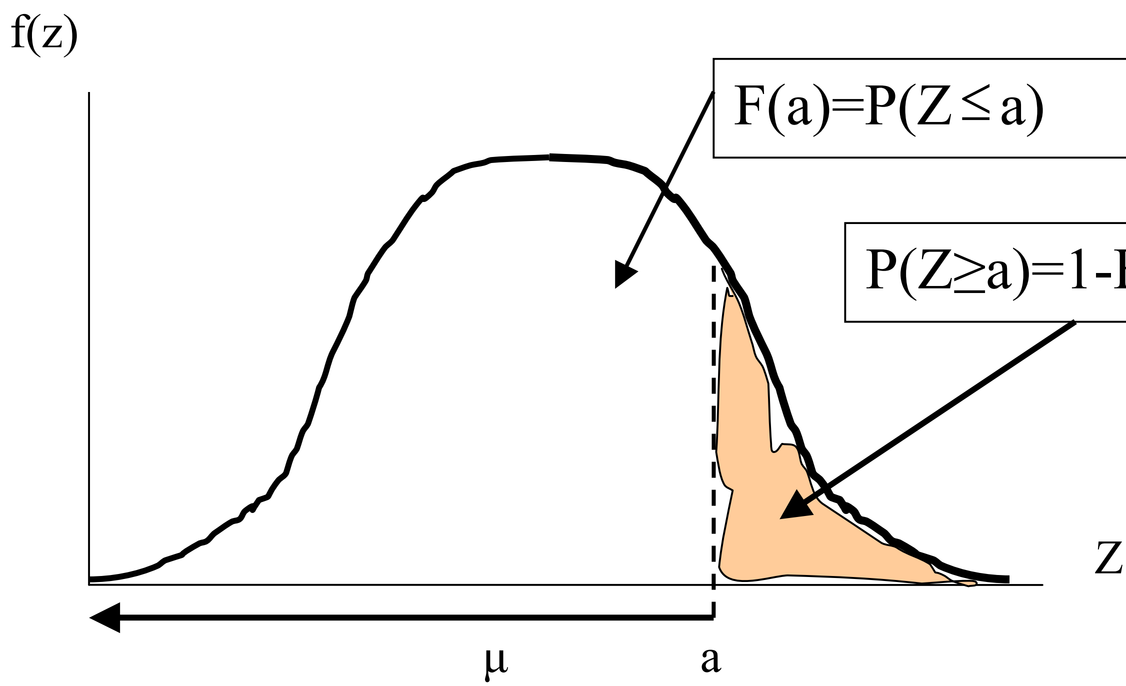
Rule 1: P(Z ≤ a) is given by F(a) where "a" is positive.



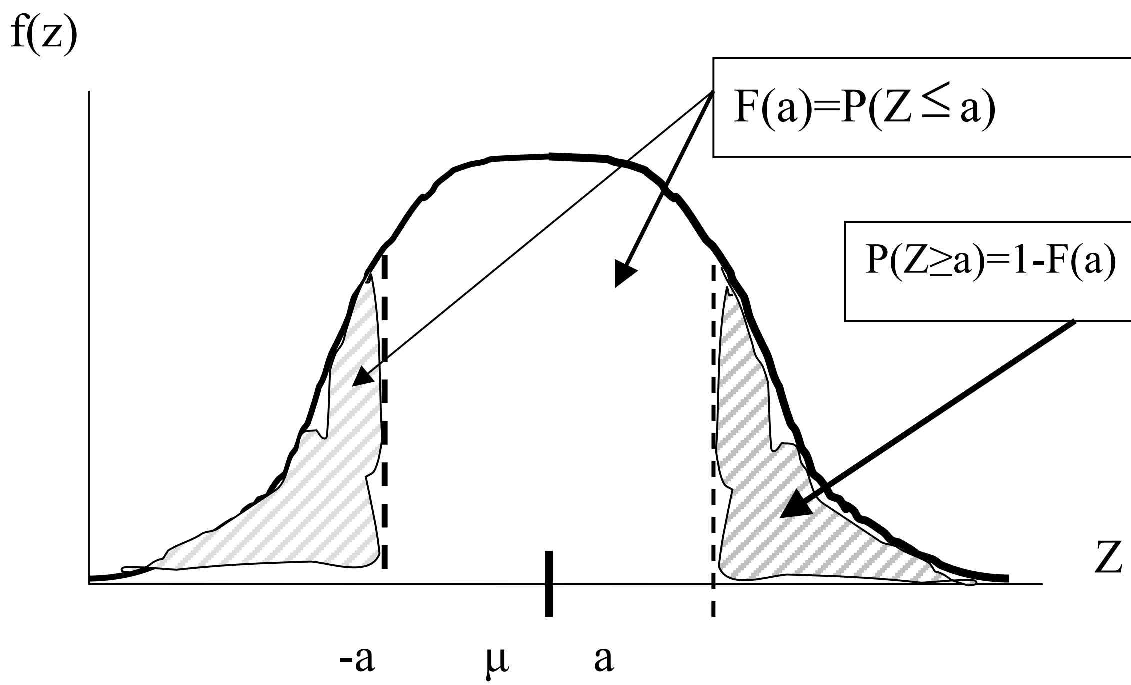
If a=2.25

P(Z ≤ 2.25)=0.9878

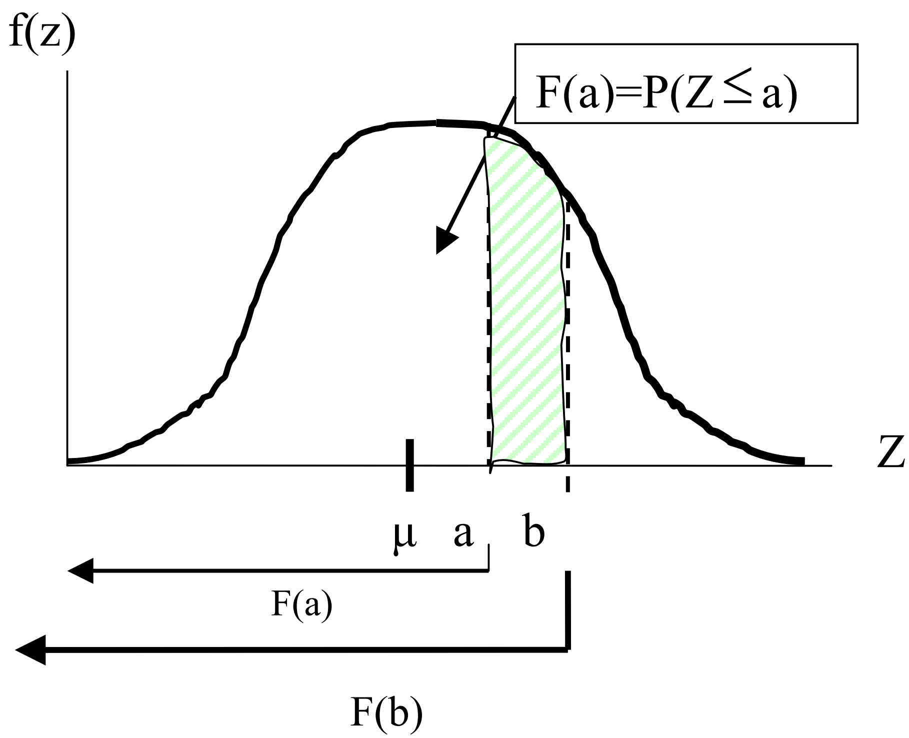
Rules 2: $P(Z \geq a)$ is given by the complement rule as 1-F(a).



Rule 3: $P(Z \leq -a)$, where (-a) is negative is given by (1-F(a)).



Rule 4: P(a≤Z≤b) is given by F(b) – F(a):



The area under the curve between two points a and b is found by subtracting the area to the left of a, F(a), from the area to the left of b, F(b).

If a and b are negative:

$$P(-a \leq z \leq -b) = P(b \leq Z \leq a) = F(a) - F(b).$$

Also, linear transformations of Normal random variables are also normal. So:

$$\mathbf{Z \sim N(0,1)}$$

Why is this transformation useful?

To solve probability questions!!

<u>Example:</u> $X \sim N(\mu = 5, \sigma^2 = 4)$

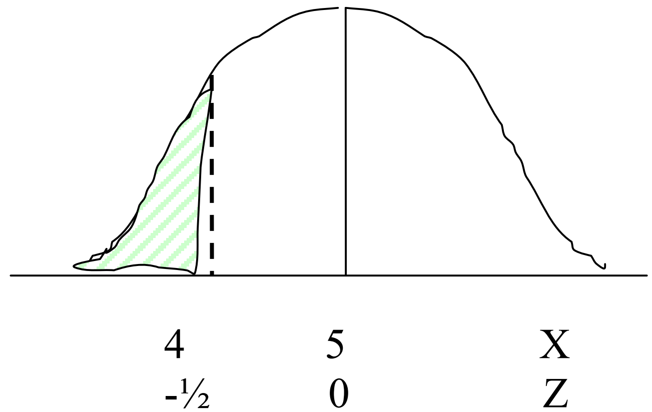
$P(x<4)$:

$$P(x<4)=P\left(\frac{x-\mu}{\sigma}<\frac{a-\mu}{\sigma}\right)$$

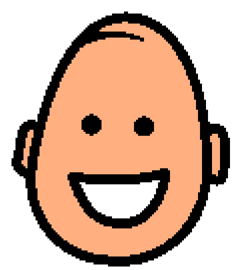
$$=P\left(Z<\frac{4-5}{2}\right)$$

$$=P(Z<-\tfrac{1}{2})$$

$$\int_{-\infty}^{\frac{1}{2}} f(z)dz=F(-\tfrac{1}{2})$$



If we have a table of such integrals already evaluated for Z, we can always transform ***any*** problem into the form where we can evaluate probability from just **one table**.



Using Table from the text:

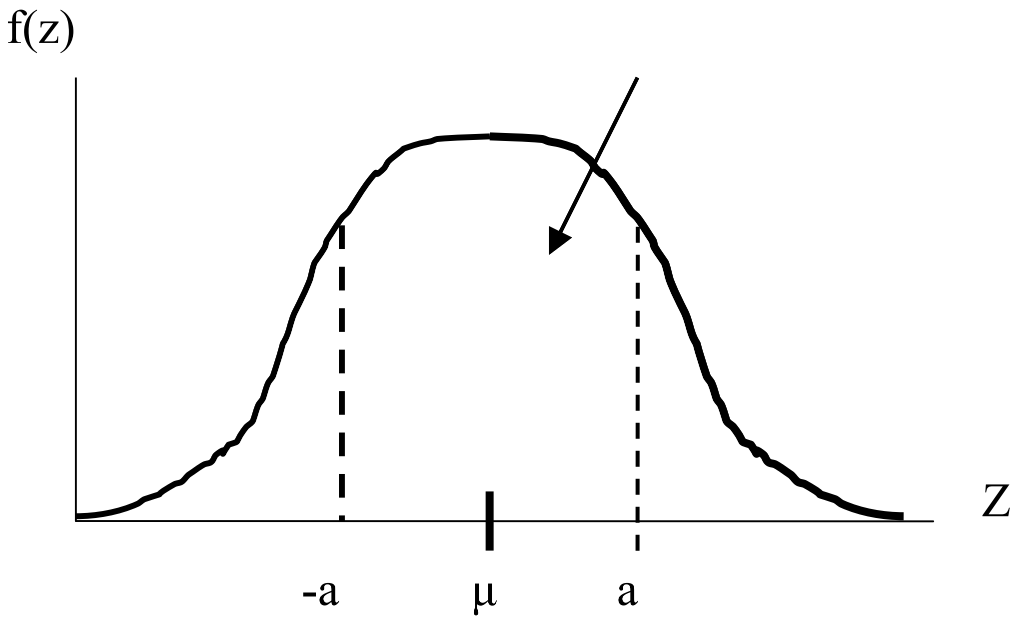


Table III: $$F(a) = \int_{-\infty}^{a} f(z)dz = P(Z \le a)$$

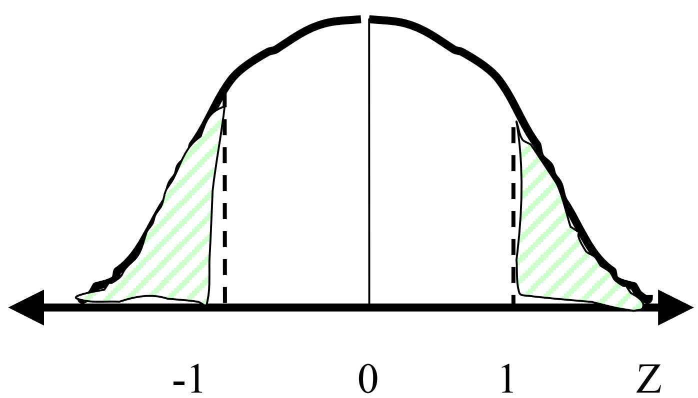
Examples:

(i) $P(Z < 0) = 0.5$

(ii) $P(Z < 1) = 0.8413$

(iii) $P(Z > 2) = 1 - P(Z \leq 2) = 1 - 0.9772 = 0.0228$

(iv) $P(Z < -1) = P(Z > 1) = 1 - P(Z \leq 1) = 1 - 0.8413 = 0.1587$



(v) $P(-1<Z<2) = F(2) - F(-1) = 0.9772 - 0.1587 = 0.8185$