Chapter 4. Continuous Random Variables and Probability Distributions

Weiqi Luo (骆伟祺)
School of Data & Computer Science
Sun Yat-Sen University
Email：luoweiqi@mail.sysu.edu.cn

Chapter four:
Continuous Random Variables and Probability Distributions

- 4.1 Probability Density Functions
- 4.2 Cumulative Distribution Functions and Expected Values
- 4.3 The Normal Distribution
- 4.4 The Exponential and Gamma Distributions
- 4.5 Other Continuous Distributions
- 4.6 Probability Plots

- Continuous Random Variables

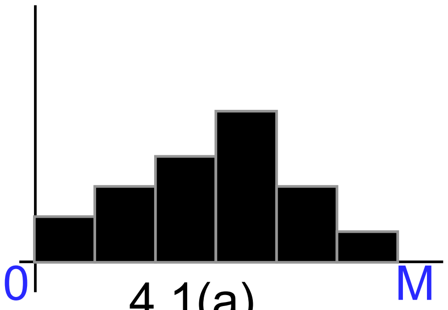
  A random variable X is said to be continuous if its set of possible values is an entire interval of numbers – that is , if for some A<B, any number x between A and B is possible

- Example 4.2

  If a chemical compound is randomly selected and its PH X is determined, then X is a continuous rv because any PH value between 0 and 14 is possible. If more is know about the compound selected for analysis, then the set of possible values might be a subinterval of [0, 14], such as $5.5 \leq x \leq 6.5$, but X would still be continuous.

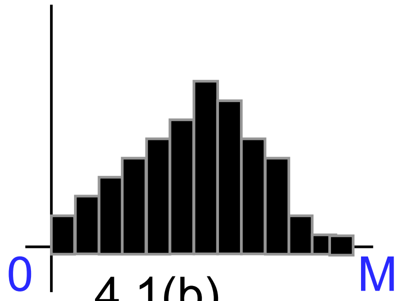
- Probability Distribution for Continuous Variables

Suppose the variable X of interest is the depth of a lake at a randomly chosen point on the surface. Let M be the maximum depth, so that any number in the interval [0,M] is a possible value of X.



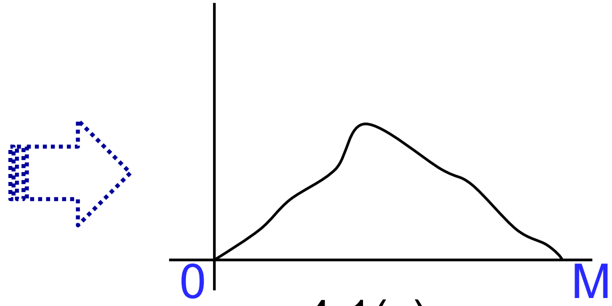
4.1(a)

Measured by meter



4.1(b)

Measured by centimeter



4.1(c)

A limit of a sequence of discrete histogram

Discrete Cases

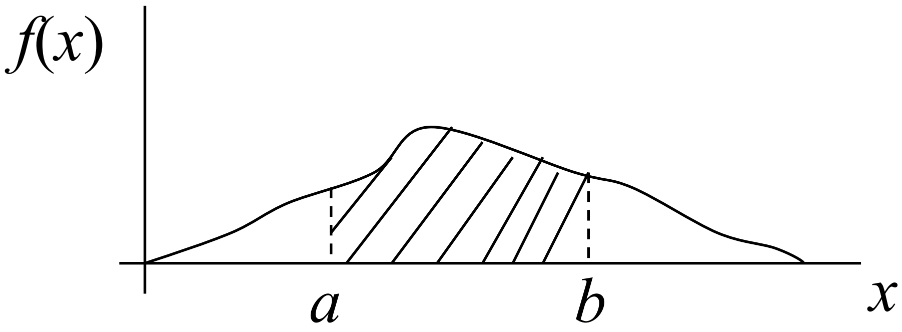
Continuous Case

Probability Distribution

Let X be a continuous rv. Then a probability distribution or probability density function (pdf) of X is f(x) such that for any two numbers a and b with a ≤ b

$$P(a \leq X \leq b) = \int_a^b f(x)\,dx$$

The probability that *X* takes on a value in the interval [*a*,*b*] is the area under the graph of the density function as follows.

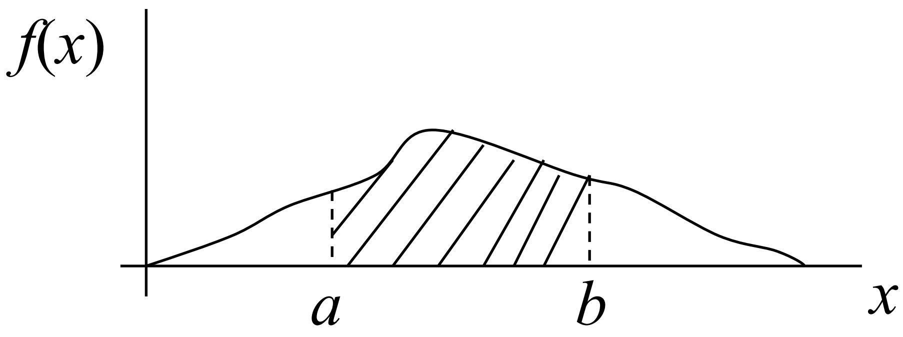


- A legitimate pdf should satisfy

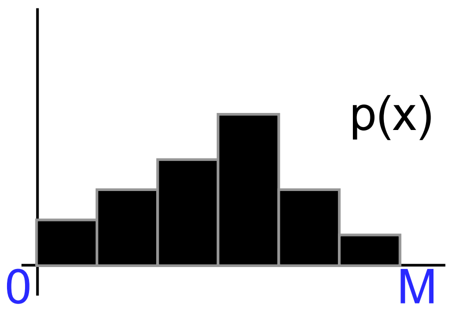
1. $f(x) \geq 0$ for all x

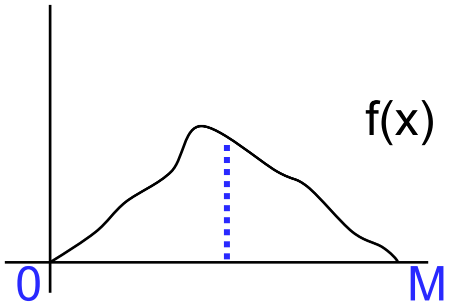
2. $\int_{-\infty}^{\infty} f(x)\mathrm{d}x =$ area under the entire graph of $f(x)$

$= 1$



- pmf (Discrete) vs. pdf (Continuous)





P(X=c) = p(c)

P(X=c) = f(c) ?

$$P(X = c) = \int_c^c f(x)dx = 0$$

- Proposition

  If X is a continuous rv, then for any number c, $P(X=c)=0$. Furthermore, for any two numbers a and b with $a<b$,

$$
\begin{aligned}
P(a \le X \le b) &= P(a < X \le b) \\
&= P(a \le X < b) \\
&= P(a < X < b)
\end{aligned}
$$

Impossible event :the event contain no simple element

P(A)=0 ➔ A is an impossible event ?

- Uniform Distribution

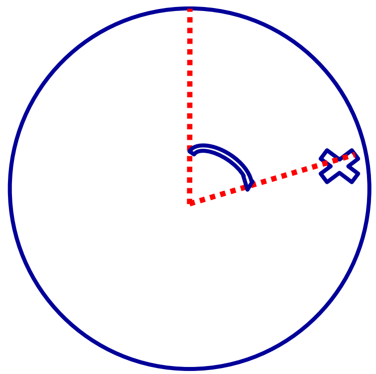
A continuous rv X is said to have a uniform distribution on the interval [A, B] if the pdf of X is

$$f(x; A, B) = \begin{cases} \dfrac{1}{B - A} & A \le x \le B \\ 0 & \text{otherwise} \end{cases}$$

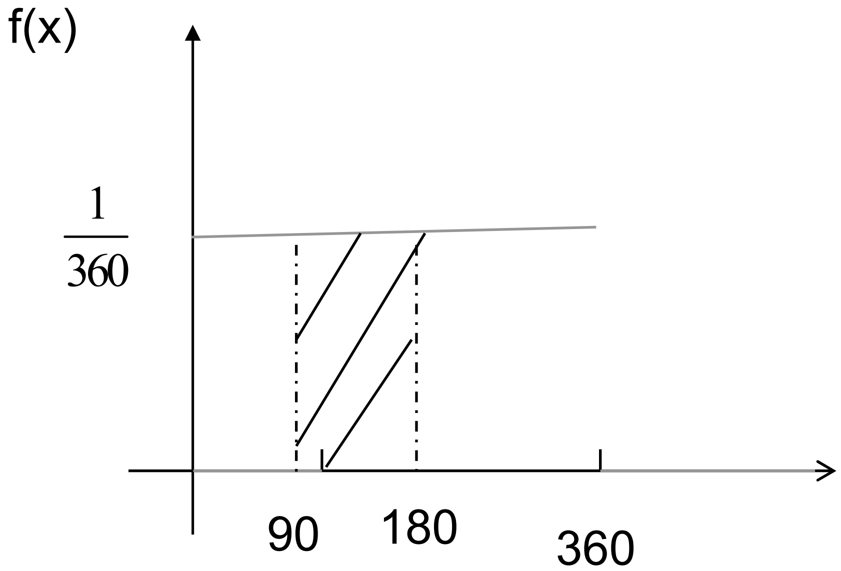
- Example 4.4

The direction of an imperfection with respect to a reference line on a circular object such as a tire, brake rotor, or flywheel is, in general, subject to uncertainty. Consider the reference line connecting the valve stem on a tire to the center point, and let X be the angle measured clockwise to the location of an imperfection, One possible pdf for X is

$$f(x) = \begin{cases} \dfrac{1}{360} & 0 \le x \le 360 \\ 0 & \text{otherwise} \end{cases}$$



- Example 4.3 (Cont')



$$P(90 \le X \le 180)$$
$$= \int_{90}^{180} \frac{1}{360} dx = 0.25$$

- Example 4.5

"Time headway" in traffic flow is the elapsed time between the time that one car finishes passing a fixed point and the instant that the next car begins to pass that point. Let X = the time headway for two randomly chosen consecutive cars on a freeway during a period of heavy flow. The following pdf of X is essentially the one suggested in "The Statistical Properties of Freeway Traffic".

$$f(x) = \begin{cases} 0.15e^{-0.15(x-0.5)} & x \geq 0.5 \\ 0 & \text{otherwise} \end{cases}$$

- Example 4.4 (Cont')

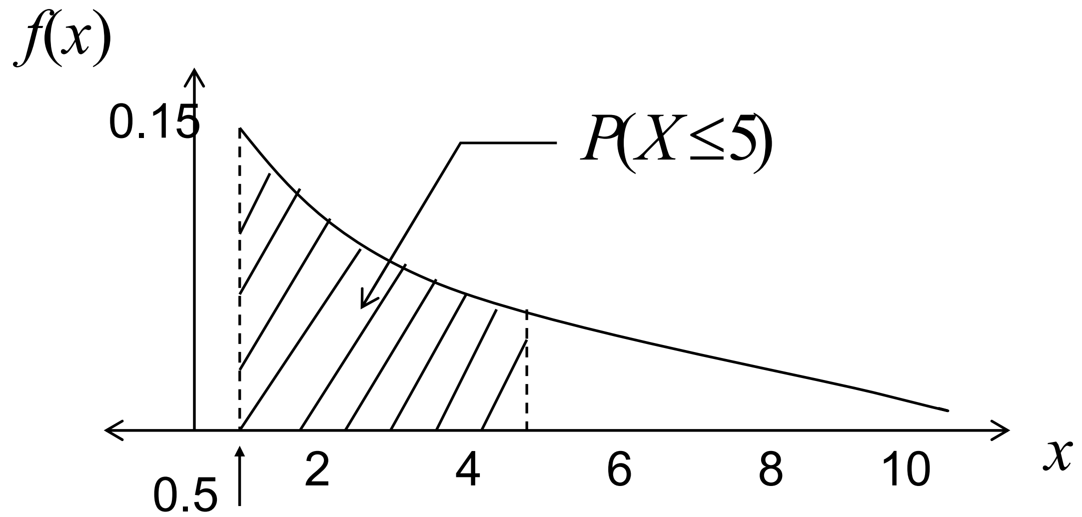
$$f(x) = \begin{cases} 0.15\mathrm{e}^{-0.15(x-0.5)} & x \geq 0.5 \\ 0 & \text{otherwise} \end{cases}$$

1. $f(x) \geq 0$;
2. to show $\int_{-\infty}^{\infty} f(x) \geq 0 dx = 1$, we use the result $\int_a^{\infty} e^{-kx} dx = \frac{1}{k} e^{-ka}$

$$\int_{-\infty}^{\infty} f(x) dx = \int_{0.5}^{\infty} 0.15 e^{-0.15(x-5)} dx = 0.15 e^{0.075} \int_{0.5}^{\infty} e^{-0.15x} dx$$

$$= 0.15 e^{0.075} \cdot \frac{1}{0.15} e^{-(0.15)(0.5)} = 1$$

- Example 4.4 (Cont')



$$P(X \le 5) = \int_{-\infty}^{5} f(x)dx = \int_{0.5}^{5} .15e^{-0.15(x-5)}dx = 0.15e^{0.075}\int_{0.5}^{5} e^{-0.15x}dx$$

$$= 0.15e^{0.075} \cdot \left(-\frac{1}{0.15} e^{-0.15x} \right)\Bigg|_{0.5}^{5} = 0.491 = P(X < 5)$$

- Homework

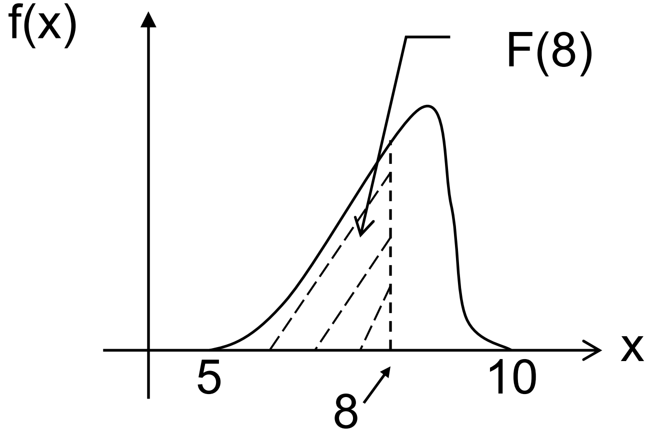
Ex. 2, Ex. 5, Ex. 8

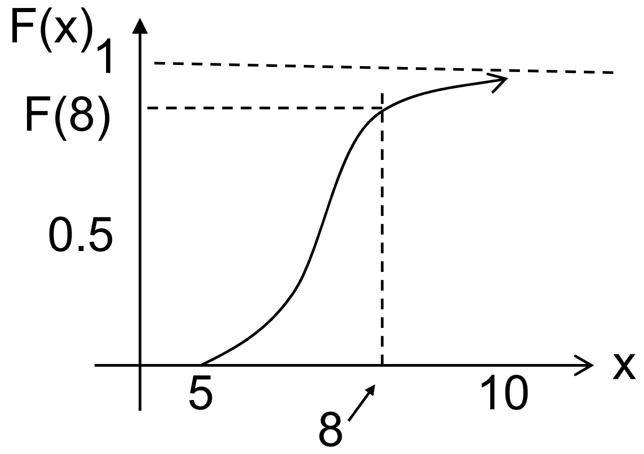
Cumulative Distribution Function

The cumulative distribution function F(x) for a continuous rv X is defined for every number x by

$$F(x) = P(X \le x) = \int_{-\infty}^{x} f(y)\,dy$$

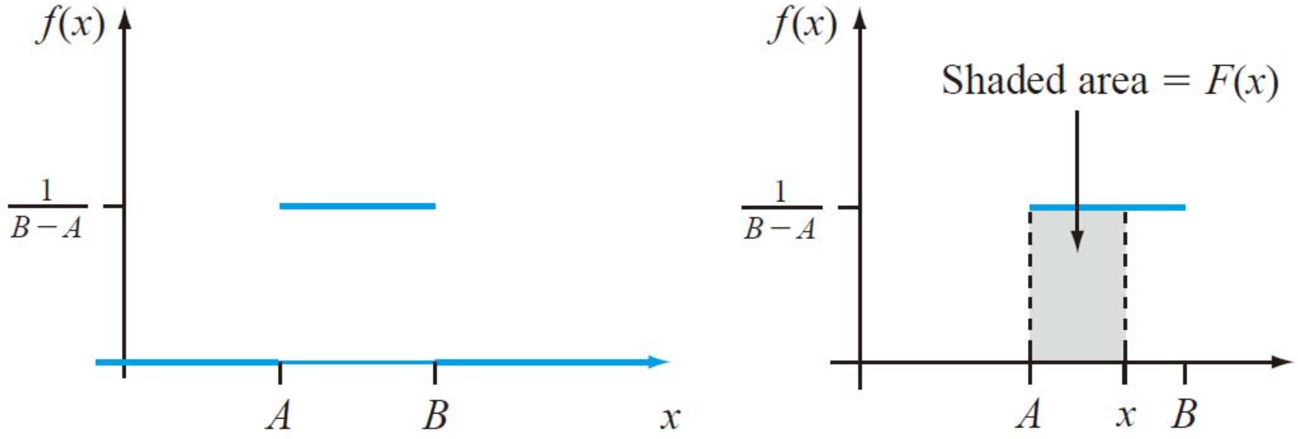
For each x, F(x) is the area under the density curve to the left of x as follows





- Example 4.6

  Let X, the thickness of a certain metal sheet, have a uniform distribution on $[A, B]$. The density function is shown as follows.



For $x < A$, $F(x) = 0$, since there is no area under the graph of the density function to the left of such an x.

For $x \geq B$, $F(x) = 1$, since all the area is accumulated to the left of such an x.

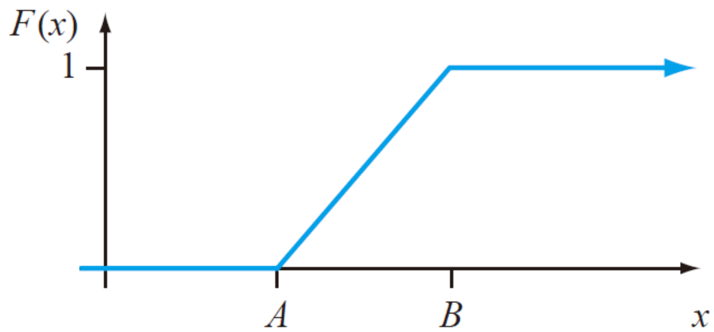
- Example 4.6 (Cont')

For A ≤X≤ B

$$F(x)=\int_{-\infty}^{x} f(y)dy=\int_{A}^{x}\frac{1}{B-A}dy=\frac{1}{B-A}\cdot y\Big|_{y=A}^{y=x}=\frac{x-A}{B-A}$$

Therefore, the entire cdf is

$$F(x)=\begin{cases} 0 & x<A \\ \dfrac{x-A}{B-A} & A\le x<B \\ 1 & x\ge B \end{cases}$$



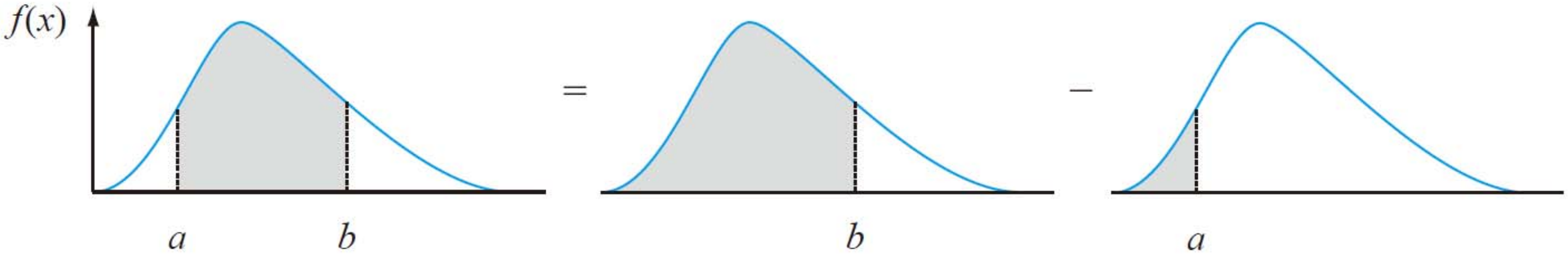
- Using F(x) to compute probabilities

Let X be a continuous rv with pdf f(x) and cdf F(x). Then for any number a

$$P(X > a) = 1 - F(a)$$

and for any two numbers a and b with a<b

$$P(a \le X \le b) = F(b) - F(a)$$



- Example 4.7

Suppose the pdf of the magnitude X of a dynamic load on a bridge is given by

$$f(x)=\begin{cases}\frac{1}{8}+\frac{3}{8}x, & 0\le x\le 2\\ 0, & otherwise\end{cases}$$

For any number x between 0 and 2,

$$F(x)=\int_{-\infty}^{x} f(y)dy=\int_{0}^{x}(\frac{1}{8}+\frac{3}{8}x)dy=\frac{x}{8}+\frac{3}{16}x^2$$

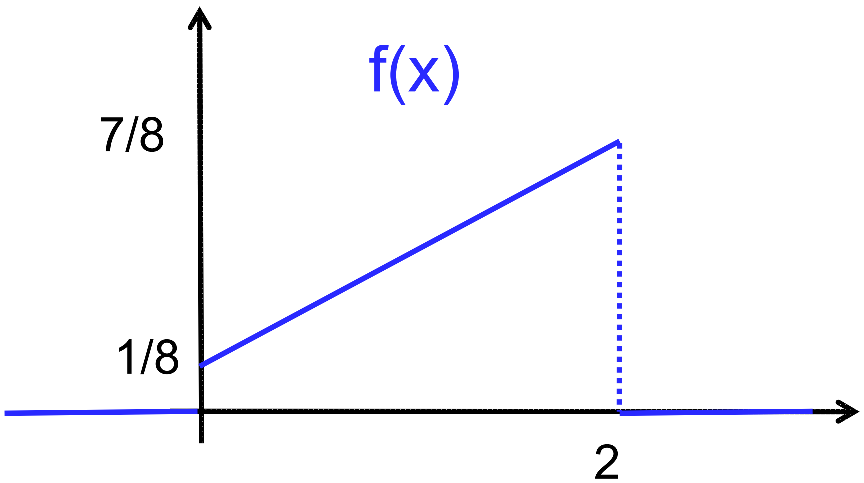
thus

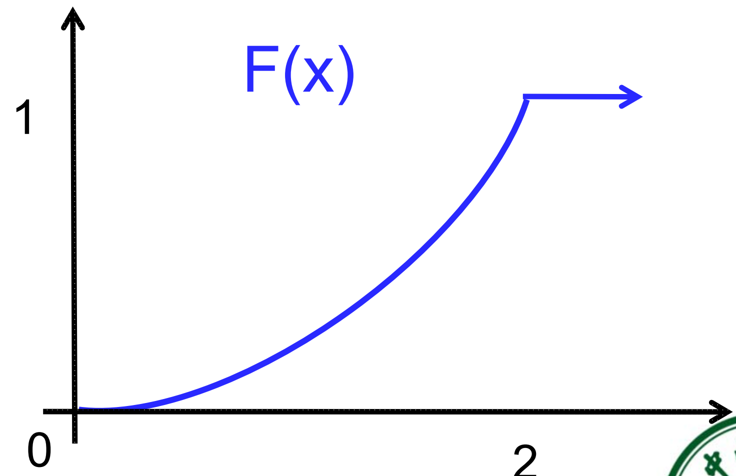
$$F(x)=\begin{cases}0, x<0\\ \int_{-\infty}^{x} f(y)dy=\int_{0}^{x}(\frac{1}{8}+\frac{3}{8}x)dy=\frac{x}{8}+\frac{3}{16}x^2, x\in[0,2]\\ 1, x>2\end{cases}$$

- Example 4.7 (Cont')

$$P(1 \le X \le 1.5) = F(1.5) - F(1) = 0.297$$

$$P(X > 1) = 1 - F(X = 1) = 0.688$$





4.2 Cumulative Distribution Functions and Expected Values

- Obtaining f(x) form F(x)

  If X is a continuous rv with pdf f(x) and cdf F(x), then at every x at which the derivative F'(x) exists, F'(x)=f(x)

$$f(x) \Longrightarrow F(x) \qquad F(x) = P(X \le x) = \int_{-\infty}^{x} f(y)\,dy$$

$$F(x) \Longrightarrow f(x) \qquad f(x) = F'(x) = \left(\int_{-\infty}^{x} f(y)\,dy\right)'$$

- Example 4.8 (Ex. 4.6 Cont’)

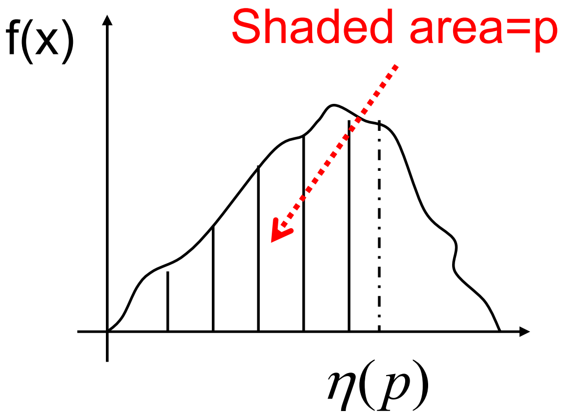
  When X has a uniform distribution, F(x) is differentiable except at x=A and x=B, where the graph of F(x) has sharp corners. Since F(x)=0 for x<A and F(x)=1 for x>B, F’(x)=0=f(x) for such x. For A<x<B

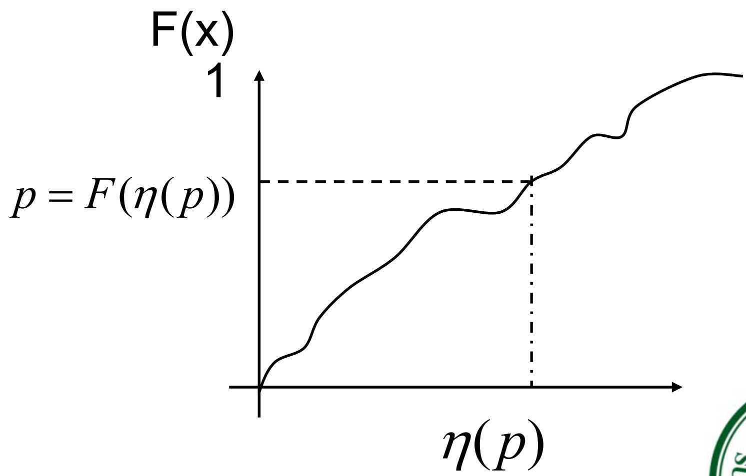
$$F'(x) = \frac{d}{dx}\left(\frac{x-A}{B-A}\right) = \frac{1}{B-A} = f(x)$$

- Percentiles of a Continuous Distribution

Let p be a number between 0 and 1. The (100p)th percentile of the distribution of a continuous rv X , denoted by η(p), is defined by

$$p = F(\eta(p)) = \int_{-\infty}^{\eta(p)} f(y)dy$$





- Example 4.9

The distribution of the amount of gravel (in tons) sold by a particular construction supply company in a given week is a continuous rv X with pdf

$$f(x) = \begin{cases} \frac{3}{2}(1 - x^2) & 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

The cdf of sales for any x between 0 and 1 is

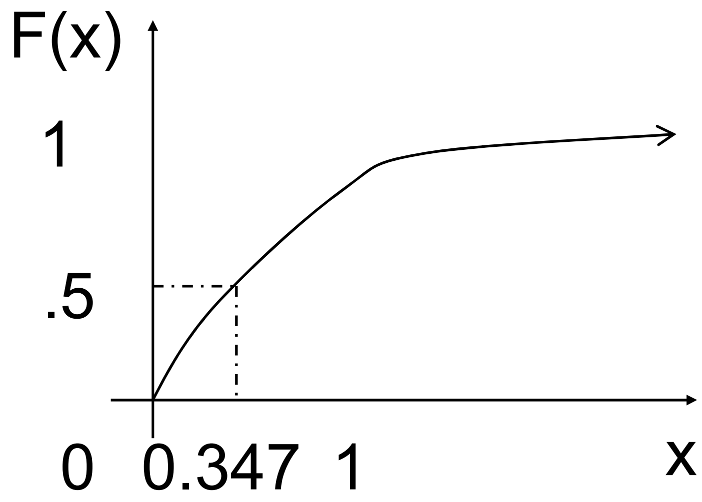
$$F(x) = \int_0^x \frac{3}{2}(1 - y^2)dy = \frac{3}{2}(y - \frac{y^3}{3})\Big|_{y=0}^{y=x} = \frac{3}{2}(x - \frac{x^3}{3})$$

- Example 4.9 (Cont')

$$p = F(\eta(p)) = \frac{3}{2}\left[\eta(p) - \frac{(\eta(p))^3}{3}\right]$$

$$(\eta(p))^3 - 3\eta(p) + 2p = 0$$

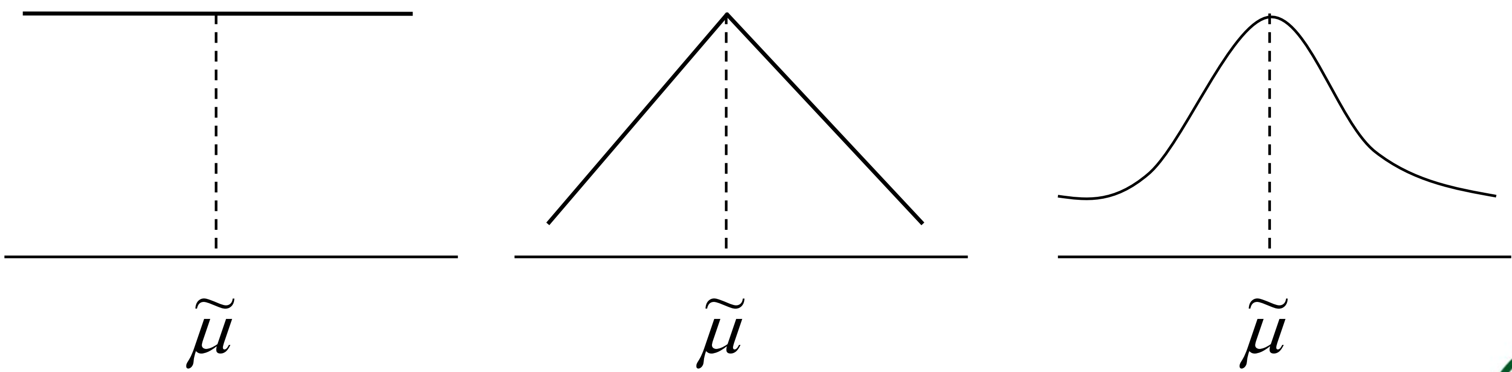
$$If \;\; p = 0.5, \eta(p) = 0.347$$



- The median

  The median of a continuous distribution, denoted by $\tilde{\mu}$, is the 50th percentile, so satisfies $0.5 = F(\tilde{\mu})$, that is, half the area under the density curve is to the left of $\tilde{\mu}$ and half is to the right of $\tilde{\mu}$

Symmetric Distribution



- Expected/Mean Value

The expected/mean value of a continuous rv X with pdf f(x) is

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

$$\mu_X = E(X) = \sum_{x \in D} x \cdot p(x)$$

Discrete Case

- Example 4.10 (Ex. 4.9 Cont')

The pdf of weekly gravel sales X was

$$f(x) = \begin{cases} \frac{3}{2}(1-x^2) & 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

So

$$E(x) = \int_{-\infty}^{+\infty} xf(x)dx = \int_0^1 x\frac{3}{2}(1-x^2)dx = \frac{3}{2}\left(\frac{x^2}{2} - \frac{x^4}{4}\right)\Big|_{x=0}^{x=1} = \frac{3}{8}$$

- Expected value of a function

  If X is a continuous rv with pdf $f(x)$ and $h(X)$ is any function of X, then

$$E\left[h(X)\right] = \mu_{h(X)} = \int_{-\infty}^{\infty} h(x) f\left(x\right) dx$$

$$\mu_{h(X)} = E(h(X)) = \sum_{x \in D} h(x) \cdot p(x)$$

Discrete Case

- Example 4.11

Two species are competing in a region for control of a limited amount of a certain resource. Let X = the proportion of the resource controlled by species 1 and suppose X has pdf

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

which is a uniform distribution on [0,1]. Then the species that controls the majority of this resource controls the amount

$$h(X) = \max(X, 1-X) = \begin{cases} 1-X & \text{if } 0 \le x < \frac{1}{2} \\ X & \text{if } \frac{1}{2} \le X \le 1 \end{cases}$$

The expected amount controlled by the species having majority control is then

$$E[h(X)] = \int_{-\infty}^{\infty} \max(x, x-1) \cdot f(x)dx = \int_0^1 \max(x, 1-x) \cdot 1dx$$

$$= \int_0^{1/2} (1-x) \cdot 1dx + \int_{1/2}^{1} x \cdot 1dx = \frac{3}{4}$$

- The Variance

The variance of a continuous random variable X with pdf $f(x)$ and mean value μ is

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

The standard deviation (SD) of X is

$$\sigma_X = \sqrt{V(X)}$$

- Proposition

$$E(aX+b)=aE(X)+b$$

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

The Same Properties as Discrete Cases

- Homework

  Ex. 13, Ex. 18, Ex. 22, Ex. 24

4.3 The Normal Distribution

- Normal (Gaussian) Distribution

A continuous rv X is said to have a normal distribution with parameters μ and σ (or μ and σ^2), where $-\infty < \mu < +\infty$ and $0 < \sigma$, if the pdf of X is

$$f(x;\mu,\sigma)=\frac{1}{\sqrt{2\pi}\sigma}e^{-(x-\mu)^2/(2\sigma^2)} \qquad -\infty < x < \infty$$

Note:

1. The normal distribution is the most important one in all of probability and statistics. Many numerical populations have distributions that can be fit very closely by an appropriate normal curve.
2. Even when the underlying distribution is discrete, the normal curve often gives an excellent approximation.
3. Central Limit Theorem (see next Chapter)

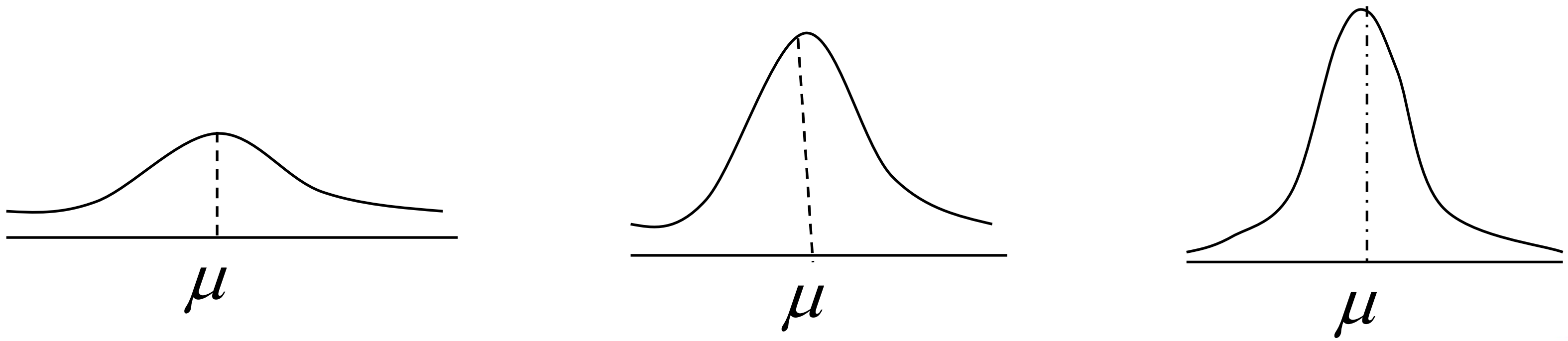
4.3 The Normal Distribution

- Properties of f(x;μ,σ)

$$f(x;\mu,\sigma)\geq 0, \int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}\sigma}e^{-(x-\mu)^2/(2\sigma^2)}dx=1$$ **Proof?**

E(X) = μ & V(X) = σ^2 , X~ N(μ, σ^2)

$\boldsymbol{\sigma}$ is large $\boldsymbol{\sigma}$ is medium $\boldsymbol{\sigma}$ is small



Symmetry Shape

4.3 The Normal Distribution

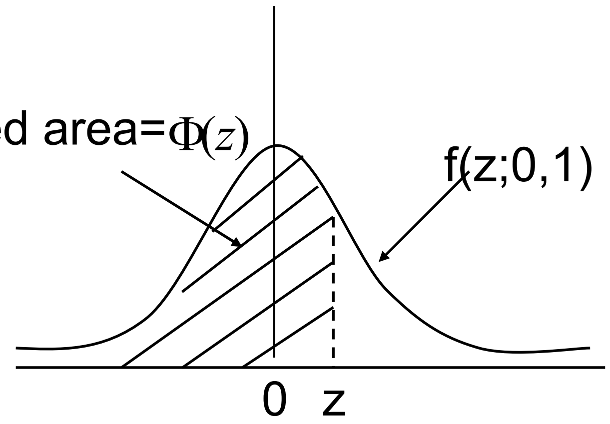
- Standard Normal Distribution

The normal distribution with parameter values μ=0 and σ=1 is called the standard normal distribution. A random variable that has a standard normal distribution is called a standard normal random variable and will be denoted by Z. The pdf of Z is

$$f(z;0,1)=\frac{1}{\sqrt{2\pi}}e^{-z^2/2},\ -\infty<z<\infty$$

The cdf of Z is

$$\Phi(z)=\int_{-\infty}^{z}f(t)dt=\int_{-\infty}^{z}\frac{1}{\sqrt{2\pi}}e^{-t^2/2}dt$$



Refer to Appendix Table A.3

4.3 The Normal Distribution

- Properties of Φ(z)

$$\Phi(-z) = 1 - \Phi(z)$$

$$\Phi(0) = 0.5$$

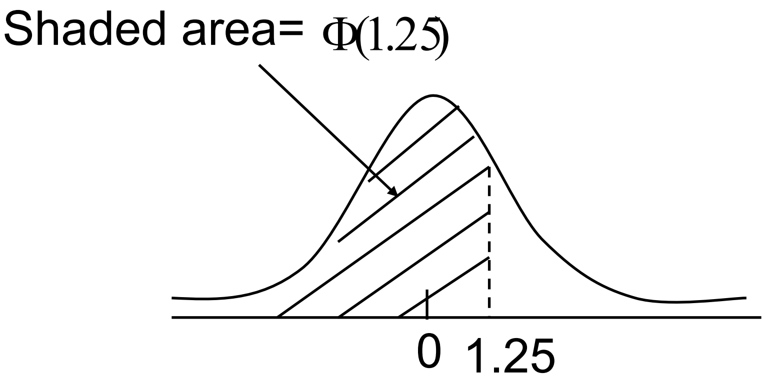
$$P(|X| \le z) = 2\Phi(z) - 1$$

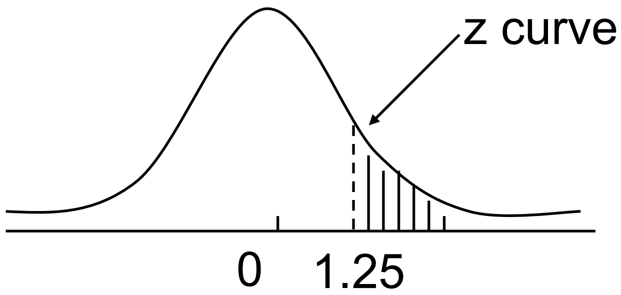
$$P(|X| \ge z) = 2[1 - \Phi(z)]$$

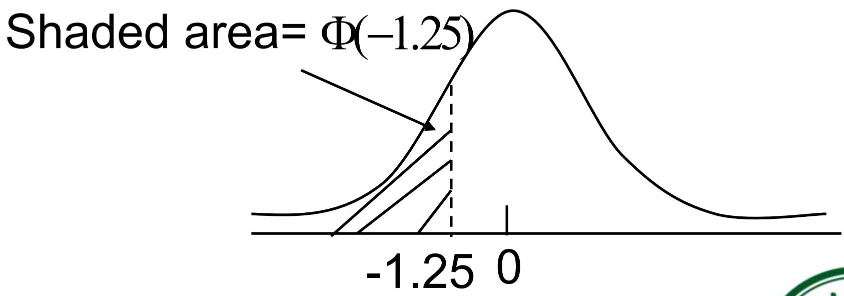
4.3 The Normal Distribution

- Example 4.13

  (a) P(Z≤1.25) (b) P(Z>1.25) (c) P(Z ≤-1.25)



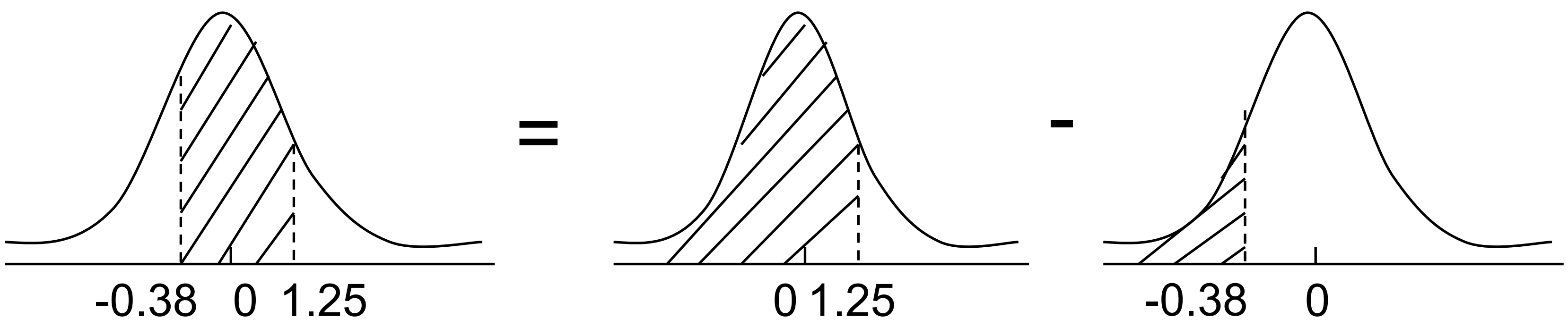




4.3 The Normal Distribution

- Example 4.13 (Cont')

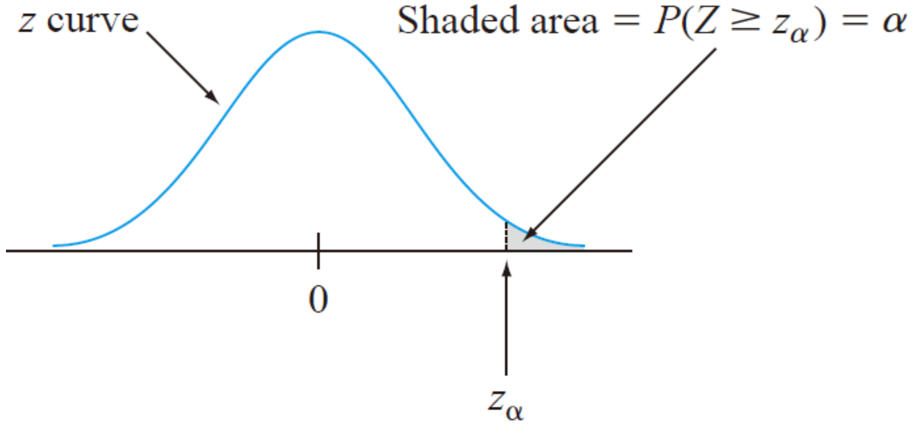
(d) $P(-0.38 \le Z \le 1.25)$



4.3 The Normal Distribution

- z_α notation

  z_α will denote the values on the measurement axis for which α of the area under the z curve lies to the right of z_α



Note: Z_α is the $100(1-\alpha)$th percentile of the standard normal distribution

Percentile	90	95	97.5	99	99.5	99.9	99.95
α (tail area)	0.1	0.05	0.025	0.01	0.005	0.001	0.0005
$z_\alpha = 100(1-\alpha)th$ percentile	1.28	1.645	1.96	2.33	2.58	3.08	3.27

4.3 The Normal Distribution

- Nonstandard Normal Distribution

If X has the normal distribution with mean μ and standard deviation σ, then

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

has a standard normal distribution (why?). Thus

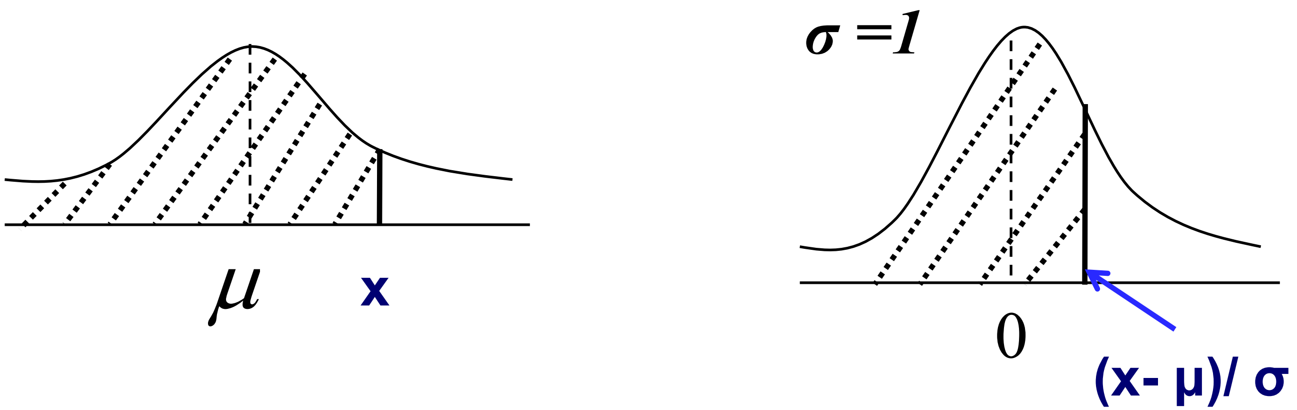
$$P(a \le X \le b) = P\left(\frac{a-\mu}{\sigma} \le Z \le \frac{b-\mu}{\sigma}\right) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$$

$$P(X \le a) = \Phi\left(\frac{a-\mu}{\sigma}\right) \qquad P(X \ge b) = 1 - \Phi\left(\frac{b-\mu}{\sigma}\right)$$

4.3 The Normal Distribution

- Equality of nonstandard and standard normal curve area

$$P(Z \le z) = P\left(X \le \sigma z + \mu\right) = \int_{-\infty}^{\sigma z+\mu} f(x;\mu,\sigma)dx$$



Percentiles of an Arbitrary Normal Distribution

$$\begin{aligned}&(100p)\text{th percentile for normal } (\mu,\sigma)\\&= \mu + \left[(100p)\text{th for standard normal}\right]\cdot\sigma\end{aligned}$$

4.3 The Normal Distribution

- Example 4.16

The time that it takes a driver to react to the brake lights on a decelerating vehicle is critical in helping to avoid rear-end collisions . Reaction time for an in-traffic response to a brake signal from standard brake lights can be modeled with a normal distribution having mean value 1.25 sec and standard deviation of .46 sec . What is the probability that reaction time is between 1.00 sec and 1.75 sec?

$$P(1.00 \le X \le 1.75) = P\left(\frac{1.00-1.25}{0.46} \le Z \le \frac{1.75-1.25}{0.46}\right)$$

$$= \Phi(1.09) - \Phi(-0.54)$$

$$= 0.8621 - 2.946 = 0.5675$$

4.3 The Normal Distribution

- Example 4.17

  The breakdown voltage of a randomly chosen diode of a particular type is known to be normally distributed. What is the probability that a diode's breakdown voltage is within 1 standard deviation of its mean value?

$$P(X \text{ is within 1 standard deviation of its mean })$$

$$= P(\mu - \sigma \le X \le \mu + \sigma) \quad = P\left(\frac{\mu - \sigma - \mu}{\sigma} \le Z \le \frac{\mu + \sigma - \mu}{\sigma}\right)$$

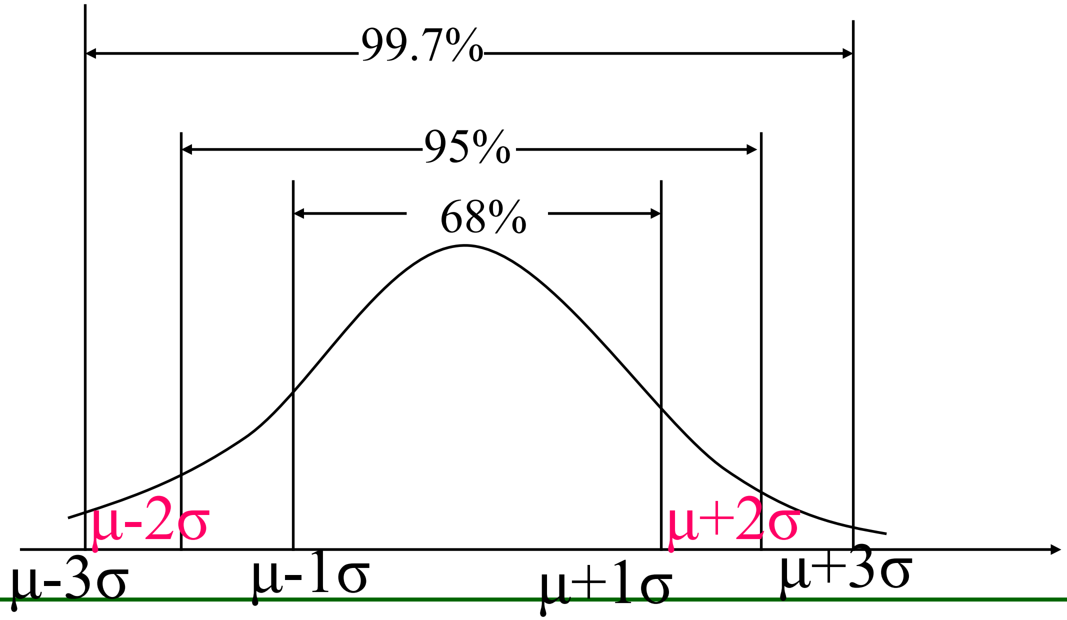
$$= P(-1.00 \le Z \le 1.00) \quad = \Phi(1.00) - \Phi(-1.00) = 0.6826$$

Note: This question can be answered without knowing either μ or σ, as long as the distribution is known to be normal; in other words , the answer is the same for any normal distribution:

4.3 The Normal Distribution

If the population distribution of a variable is (approximately) normal, then

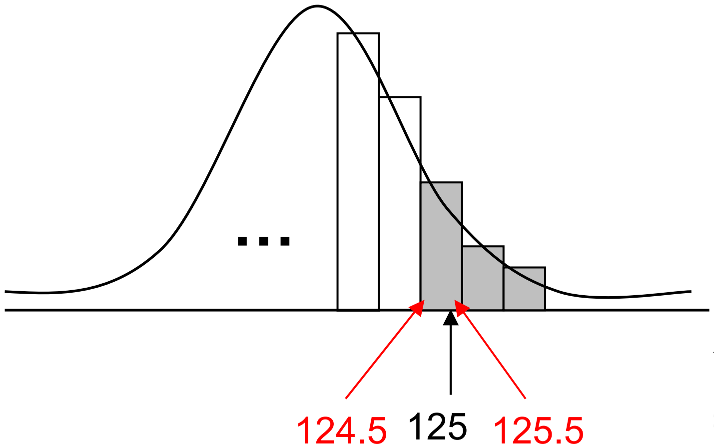
1. Roughly **68%** of the values are within **1 SD** of the mean.
2. Roughly **95%** of the values are within **2 SDs** of the mean
3. Roughly **99.7%** of the values are within **3 SDs** of the mean



4.3 The Normal Distribution

- The Normal Distribution and Discrete Populations

Ex. 4.19: IQ in a particular population is known to be approximately normally distributed with $\mu = 100$ and $\sigma = 15$. What is the probability that a randomly selected individual has an IQ of at least 125? Letting X = the IQ of a randomly chosen person, we wish $P(X \geq 125)$. The temptation here is to standardize $X \geq 125$ immediately as in previous example. However, the IQ population is actually discrete, since IQs are integer-valued, so the normal curve is an approximation to a discrete probability histogram,



continuity correction

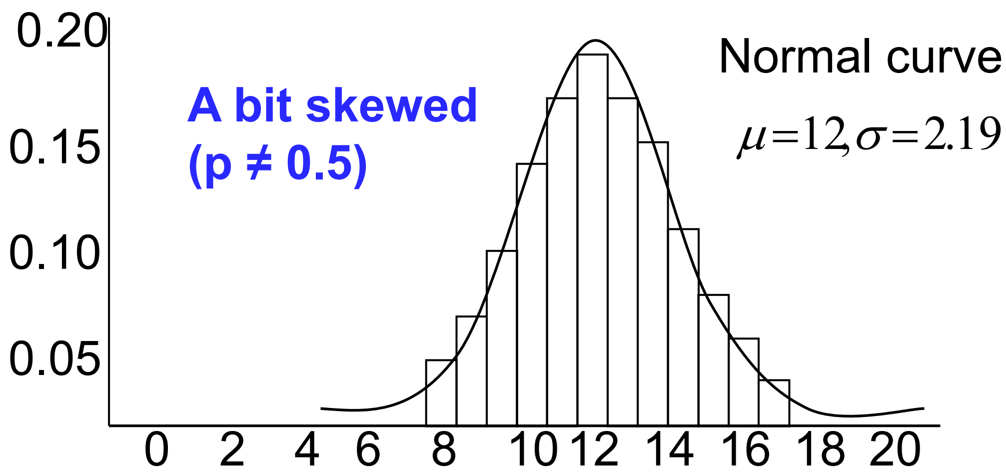
$$P(X \geq 125) = P(Z \geq [(125 - 0.5) - 100)] / 15)$$
$$= P(Z \geq 1.63) = 0.0516$$

$$P(X = 125)$$
$$= P([(125 - 0.5) - 100] / 15 \leq Z \leq [(125 + 0.5) - 100] / 15)$$
$$= P(1.63 \leq Z \leq 1.7) \neq 0$$

4.3 The Normal Distribution

- The Normal Approximation to the Binomial Distribution

Recall that the mean value and standard deviation of a binomial random variable X are μ_X = np and σ_X=(npq)$^{1/2}$. Consider the binomial probability histogram with n = 20, p = 0.6. It can be approximated by the normal curve with μ = 12 and σ = 2.19 as follows.



4.3 The Normal Distribution

- Proposition

Let X be a binominal rv based on n trials with success probability p. Then if the binomial probability histogram is not too skewed, X has approximately a normal distribution with $\mu = np$ and σ_X=(npq)$^{1/2}$. In particular, for x = a possible value of X ,

$$p(X \le x) = B(x; n, p)$$

$$\approx (\text{area under the normal curve to the left of } x + 0.5)$$

$$= \Phi\left(\frac{x + 0.5 - np}{\sqrt{npq}}\right)$$

Rule: In practice, the approximation is adequate provided that both np≥10 and nq ≥10. (where q=1-p)

- Example 4.20

Suppose that 25% of all licensed drivers in a particular state do not have insurance. Let X be the number of uninsured drivers in a random sample of size 50, so that p=0.25. Since np=50(0.25)=12.5≥10 and nq=37.5 ≥ 10, the approximation can safely be applied. Then μ = 12.5 and σ = 3.06.

$$P(X \le 10) = B(10;50,0.25) \approx \Phi\left(\frac{10+0.5-12.5}{3.06}\right)$$

$$= \Phi(-0.65) = 0.2578$$

Similarly , the probability that between 5 and 15 (inclusive) of the selected drivers are uninsured is

$$P(5 \le X \le 15) = B(15;50,0.25) - B(4;50,0.25)$$

$$\approx \Phi\left(\frac{15.5-12.5}{3.06}\right) - \Phi\left(\frac{4.5-12.5}{3.06}\right) = 0.8320$$

4.3 The Normal Distribution

- Homework

Ex. 29, Ex. 32, Ex. 44, Ex. 48, Ex. 52

4.4 The Exponential and Gamma Distributions

- Gamma Function

For $\alpha>0$, the gamma function $\Gamma(\alpha)$ is defined by

$$\Gamma(\alpha) = \int_0^{\infty} x^{a-1} e^{-x} dx$$

The most important properties of the gamma function are the following:

1. For any $\alpha>1$, $\Gamma(\alpha) = (\alpha-1)\,\Gamma(\alpha-1)$;
2. For any positive integer n, $\Gamma(n)=(n-1)!$
3. $\Gamma(1/2) = \Pi^{1/2}$

4.4 The Exponential and Gamma Distributions

- Standard Gamma Distribution

$$f(x;\alpha) = \begin{cases} \dfrac{x^{\alpha-1}e^{-x}}{\Gamma(\alpha)} & x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

Satisfying the two Basic Properties of a pdf:

$$1: f(x;a) \geq 0$$

$$2: \int_0^\infty f(x;a)dx = \frac{\int_0^\infty x^{\alpha-1}e^{-x}dx}{\Gamma(\alpha)} = \frac{\Gamma(\alpha)}{\Gamma(\alpha)} = 1$$

4.4 The Exponential and Gamma Distributions

- Example 4.23

 Suppose the reaction time X of a randomly selected individual to a certain stimulus has a standard gamma distribution with α=2 sec. Then

$$P(3 \le X \le 5) = F(5;2) - F(3;2)$$

$$= 0.960 - 0.801 \ = 0.159$$

$$P(\ X>4) = 1 - P(\ X \le 4) = 1 - F(4;2) = 1 - 0.908 = 0.902$$

4.4 The Exponential and Gamma Distributions

- The Family of Gamma Distributions

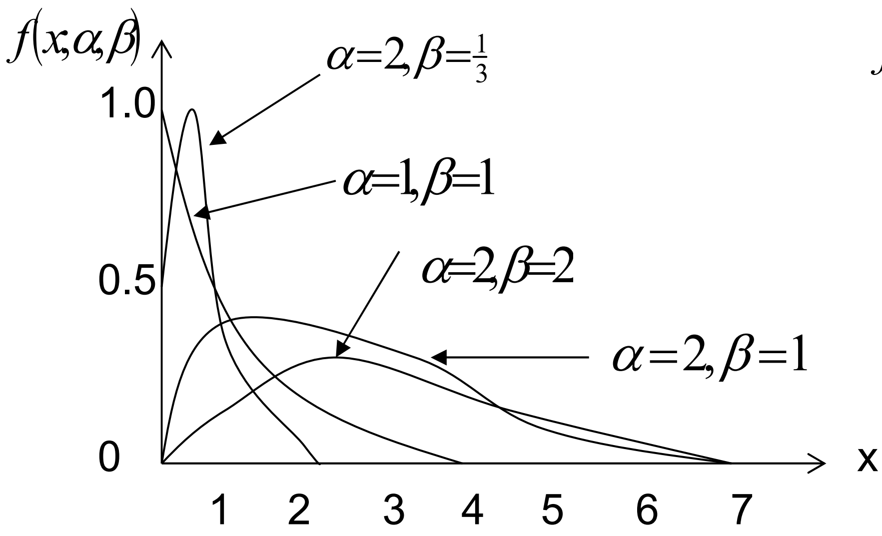
A continuous random variable X is said to have a gamma distribution if the pdf of X is

$$f(x;\alpha,\beta)=\begin{cases}\dfrac{1}{\beta^{\alpha}\Gamma(\alpha)}x^{\alpha-1}e^{-x/\beta} & x\geq 0\\ 0 & \text{otherwise}\end{cases}$$

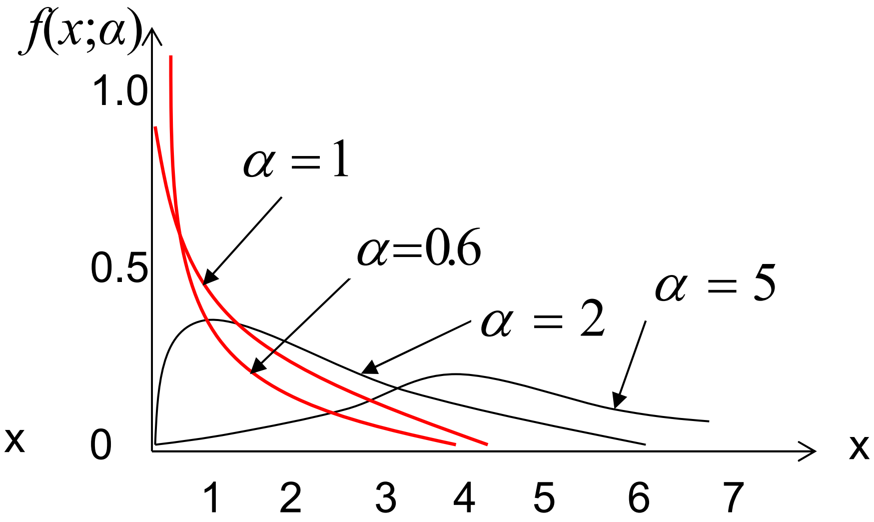
where the parameters α and β satisfy $\alpha>0$, $\beta>0$. The standard gamma distribution has $\beta=1$.

4.4 The Exponential and Gamma Distributions

- Illustrations of the Gamma pdfs



(a) Gamma density curves



(b) Standard gamma density curves

4.4 The Exponential and Gamma Distributions

- Mean and Variance

The mean and variance of a random variable X having the gamma distribution $f(x;\alpha,\beta)$ are

$$E(X) = \mu = \alpha\beta$$

$$V(X) = \delta^2 = \alpha\beta^2$$

- The cdf of a standard gamma distribution

$$F(x;\alpha) = \int_0^x \frac{y^{\alpha-1}e^{-y}}{\Gamma(\alpha)}dy \qquad x > 0$$

Incomplete gamma function (or without the denominator $\Gamma(\alpha)$ sometimes)

4.4 The Exponential and Gamma Distributions

- Proposition

  Let X have a gamma distribution with parameters α and β. Then for any $x > 0$, the cdf of X is given by

$$P(X \leq x) = F(x; \alpha, \beta) = F\left(\frac{x}{\beta}; \alpha\right)$$

  where F(•; α) is the incomplete gamma function.

- Example 4.24

Suppose the survival time X in weeks of a randomly selected male mouse exposed to 240 rads of gamma radiation has a gamma distribution with α=8 and β=15, then the probability that a mouse survives between 60 and 120 weeks is

$$
\begin{aligned}
P(60 \le X \le 120) &= P(X \le 120) - P(X \le 60) \\
&= F(120/15;8) - F(60/15;8) \\
&= F(8;8) - F(4;8) = 0.496
\end{aligned}
$$

the probability that a mouse survives at least 30 weeks is

$$
\begin{aligned}
P(X \ge 30) &= 1 - P(X < 30) = 1 - P(X \le 30) \\
&= 1 - F(30/15;8) = 0.999
\end{aligned}
$$

4.4 The Exponential and Gamma Distributions

- The Exponential Distribution

X is said to have an exponential distribution with parameter λ (λ>0) if the pdf of X is

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

Just a special case of the general gamma pdf

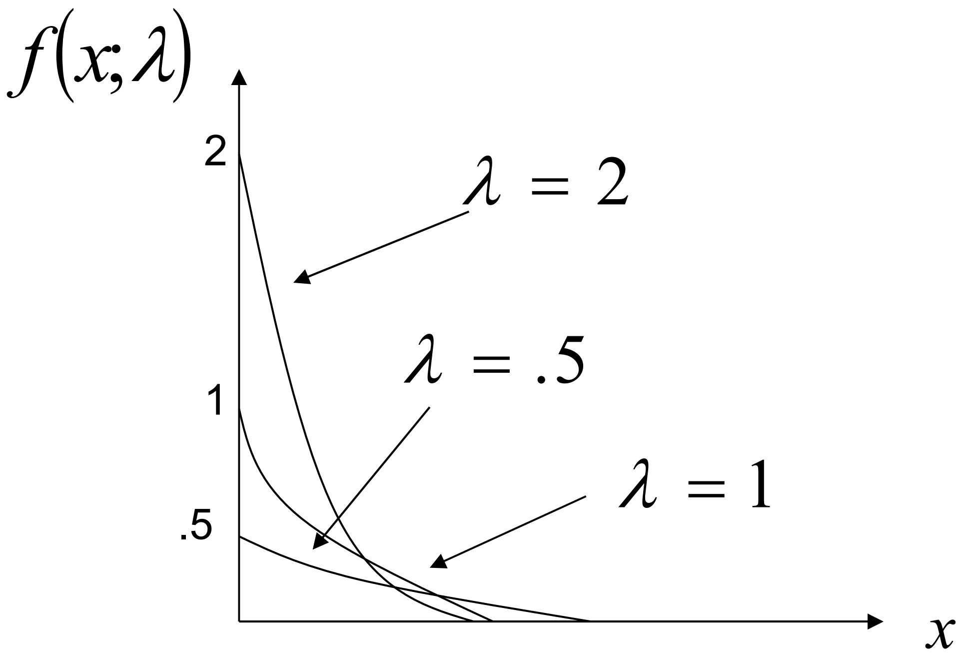
$\alpha=1$ and $\beta = 1/\lambda$

therefore, we have

$E(X) = \alpha\beta = 1/\lambda$; $V(X) = \alpha\beta^2 = 1/\lambda^2$

4.4 The Exponential and Gamma Distributions

- Illustrations of the Exponential pdfs



4.4 The Exponential and Gamma Distributions

- The cdf of Exponential Distribution

  Unlike the general gamma pdf, the exponential pdf can be easily integrated.

$$F(x;\lambda)=\begin{cases} 0 & x<0 \\ 1-e^{-\lambda x} & x\geq 0 \end{cases}$$

4.4 The Exponential and Gamma Distributions

- Example

Suppose the response time X at a certain on-line computer terminal (the elapsed time between the end of a user's inquiry and the beginning of the system's response to inquiry) has an exponential distribution with expected response time equal to 5 sec. then E(X) =1/ λ =5, so λ=0.2. the probability that the response tine is at most 10 sec is

$$P(X \le 10) = F(10;0.2) = 1 - e^{-(0.2)(10)} = 0.865$$

The probability that response time is between 5 and 10 sec is

$$P(5 \le X \le 10) = F(10;0.2) - F(5;0.2) = 0.233$$

4.4 The Exponential and Gamma Distributions

- The Chi-Squared Distribution

Let v be a positive integer. Then a random variable X is said to have a chi-squared distribution with parameter v if the pdf of X is the gamma density with $\alpha = v/2$ and $\beta = 2$. The pdf of a chi-squared rv is thus

$$f(x, v) = \begin{cases} \dfrac{1}{2^{v/2}\Gamma(v/2)} x^{(v/2)-1} e^{-x/2} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

The parameter v is called the number of degrees of freedom of X. The symbol χ^2 is often used in place of "chi-squared."

4.4 The Exponential and Gamma Distributions

- Homework

  Ex. 64, Ex. 66, Ex. 70

4.5 Other Continuous Distributions

- The Weibull Distribution

  A random variable X is said to have a Weibull distribution with parameters α and β ($\alpha > 0$, $\beta > 0$) if the cdf of X is

$$f(x;\alpha,\beta)=\begin{cases}\dfrac{\alpha}{\beta^{\alpha}}x^{\alpha-1}e^{-(x/\beta)^{\alpha}} & x\geq 0\\ 0 & x<0\end{cases}$$

When α =1, the pdf reduces to the exponential distribution (with λ =1/ β), so the exponential Distribution is a special case of both the Gamma and Wellbull distributions.

4.5 Other Continuous Distributions

- Mean and Variance

$$\mu = \beta\Gamma\left(1+\frac{1}{\alpha}\right); \quad \sigma^2 = \beta^2\left\{\Gamma\left(1+\frac{2}{\alpha}\right)-\left[\Gamma\left(1+\frac{1}{\alpha}\right)\right]^2\right\}$$

- The cdf of a Weibull Distribution

$$F(x;\alpha,\beta)=\begin{cases} 0 & x<0 \\ 1-e^{-(x/\beta)^{\alpha}} & x\geq 0 \end{cases}$$

4.5 Other Continuous Distributions

- The Lognormal Distribution

A nonnegative rv X is said to have a lognormal distribution if the rv $Y = \ln(X)$ has a normal distribution . The resulting pdf of a lognormal rv when $\ln(X)$ is normally distributed with parameters μ and σ is

$$f(x;\mu,\sigma) = \begin{cases} \dfrac{1}{\sqrt{2\pi}\sigma x} e^{-(\ln(x)-\mu)^2/(2\sigma^2)} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

4.5 Other Continuous Distributions

- Mean and Variance

$$E(X) = e^{\mu+\sigma^2/2} \quad ; V(X) = e^{2\mu+\sigma^2} \cdot \left(e^{\sigma^2} - 1\right)$$

- The cdf of Lognormal Distribution

$$F(x;\mu,\sigma) = P(X \le x) = P[\ln(X) \le \ln(x)]$$
$$= P\left(Z \le \frac{\ln(x)-\mu}{\sigma}\right) = \Phi\left(\frac{\ln(x)-\mu}{\sigma}\right)$$

4.5 Other Continuous Distributions

- The Beta Distribution

A random variable X is said to have a beta distribution with parameters α, β, A, and B if the pdf of X is

$$f(x;\alpha,\beta,A,B)=\begin{cases}\frac{1}{B-A}\cdot\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\cdot\Gamma(\beta)}\left(\frac{x-A}{B-A}\right)^{\alpha-1}\left(\frac{B-x}{B-A}\right)^{\beta-1}, & A\le x\le B\\ 0 & \text{otherwise}\end{cases}$$

The case A = 0, B =1 gives the standard beta distribution. And the mean and variance are

$$\mu=A+(B-A)\cdot\frac{\alpha}{\alpha+\beta};\quad \sigma^2=\frac{(B-A)^2\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

4.5 Other Continuous Distributions

- Homework

  Ex. 72, Ex. 81

4.6 Probability Plots

- Probability Plot

An investigator obtained a numerical sample $x_1,x_2,\ldots,x_n$ and wish to know whether it is plausible that it came from a population distribution of some particular type (and/or the corresponding parameters).

An effective way to check a distributional assumption is to construct the so-called Probability plot.

4.6 Probability Plots

- Sample Percentiles

Order the n sample observations from the smallest to the largest. Then the ith smallest observation in the list is taken to be the $[100(i\text{-}.5)/n]$th sample percentile. Considering the following pairs (as a point on a 2-D coordinate system) in a figure

$$\begin{pmatrix} [100(i-0.5)/n]\text{th percentile,} & i\text{th smallest sample} \\ \text{of the distribution} & \text{observation} \end{pmatrix}$$

Note: If the sample percentiles are close to the corresponding population distribution percentiles, then all points will fall close to a 45^{o} line.

4.6 Probability Plots

- Normal Probability Plot

  Just a special case of the probability plot

$$\begin{pmatrix} [100(i-0.5)/n]\text{th percentile,} & i\text{th smallest sample} \\ \text{of the distribution} & \text{observation} \end{pmatrix}$$

$$\Downarrow$$

$$\begin{pmatrix} [100(i-0.5)/n]\text{th z percentile,} & \begin{matrix} i\text{th smallest sample} \\ \text{observation} \end{matrix} \end{pmatrix}$$

Used to check the Normality of the sample data

4.6 Probability Plots

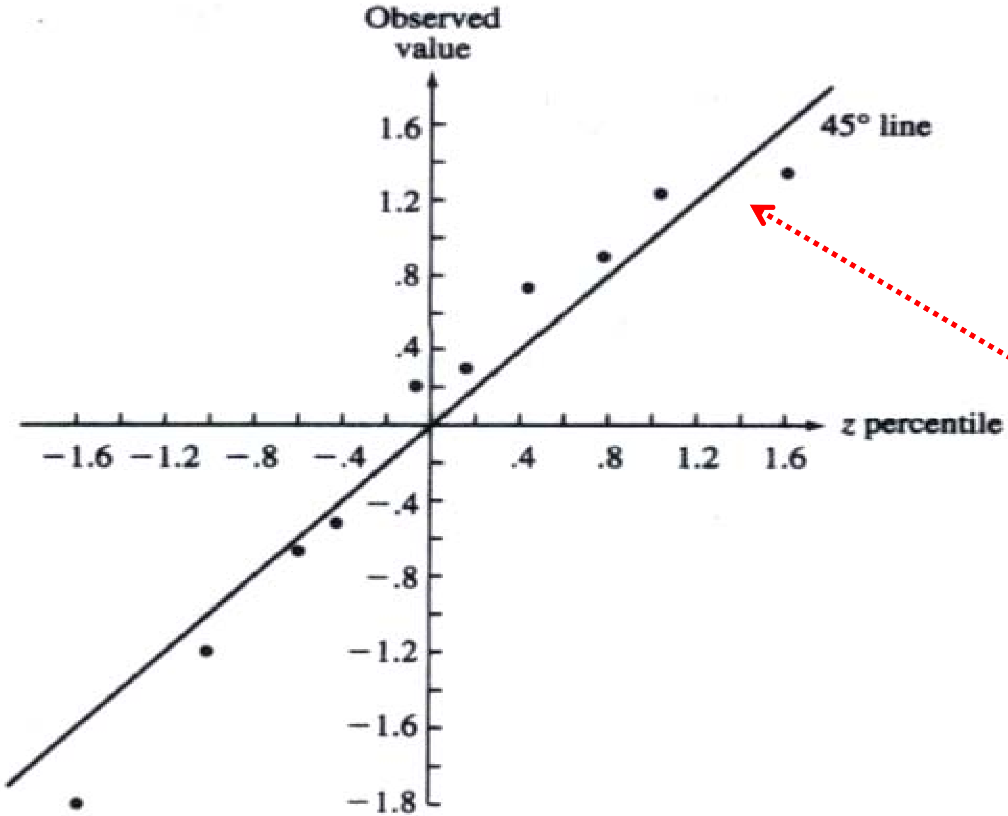
- Example 4.29

The value of a certain physical constant is known to an experimenter. The experimenter makes $n = 10$ independent measurements of this value using a particular measurement device and records the resulting measurement errors (error = observed value - true value). These observations appear in the accompanying table.

Percentage	5	15	25	35	45
z percentile	−1.645	−1.037	−.675	−.385	−.126
Sample observation	−1.91	−1.25	−.75	−.53	.20
Percentage	55	65	75	85	95
z percentile	.126	.385	.675	1.037	1.645
Sample observation	.35	.72	.87	1.40	1.56

4.6 Probability Plots

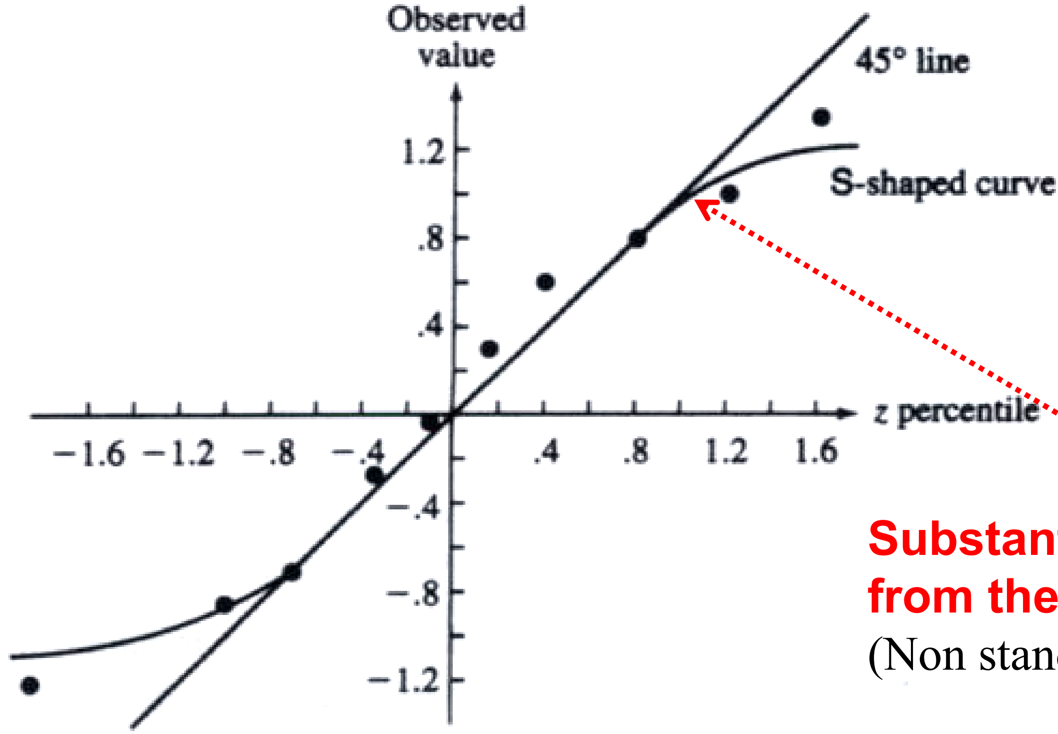
- Example 4.29 (Cont')



Close the 45^0 line
(Approximately standard normal distribution)

4.6 Probability Plots

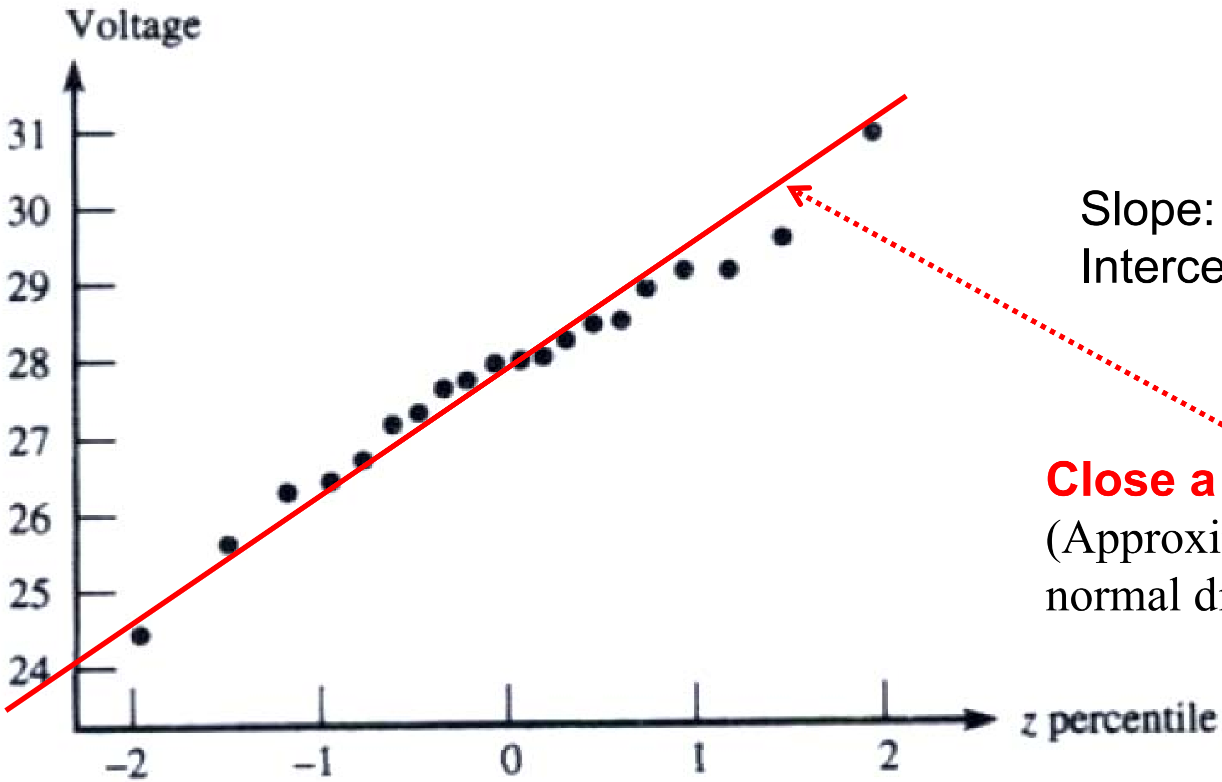
- Example 4.29 (Cont')



Substantial deviations from the 45^0 line
(Non standard normal distribution)

4.6 Probability Plots

- Example 4.29



Slope: δ
Intercept : μ

Close a straight line
(Approximately normal distribution)

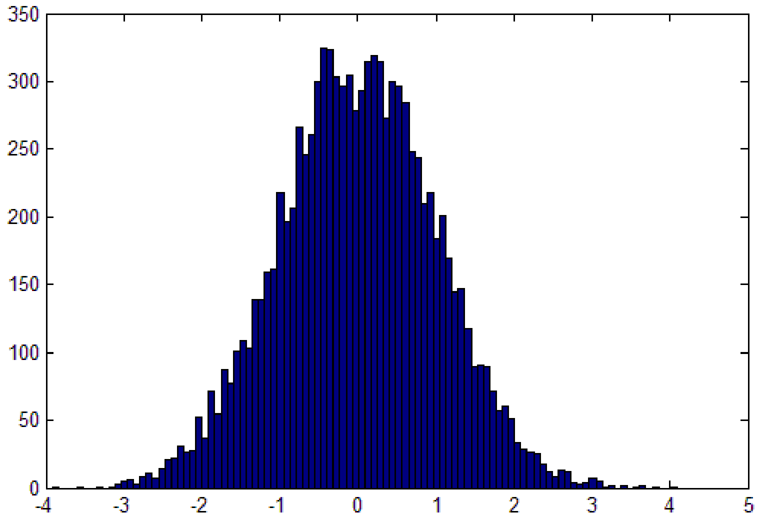
4.6 Probability Plots

- Categories of a non-normal population distribution

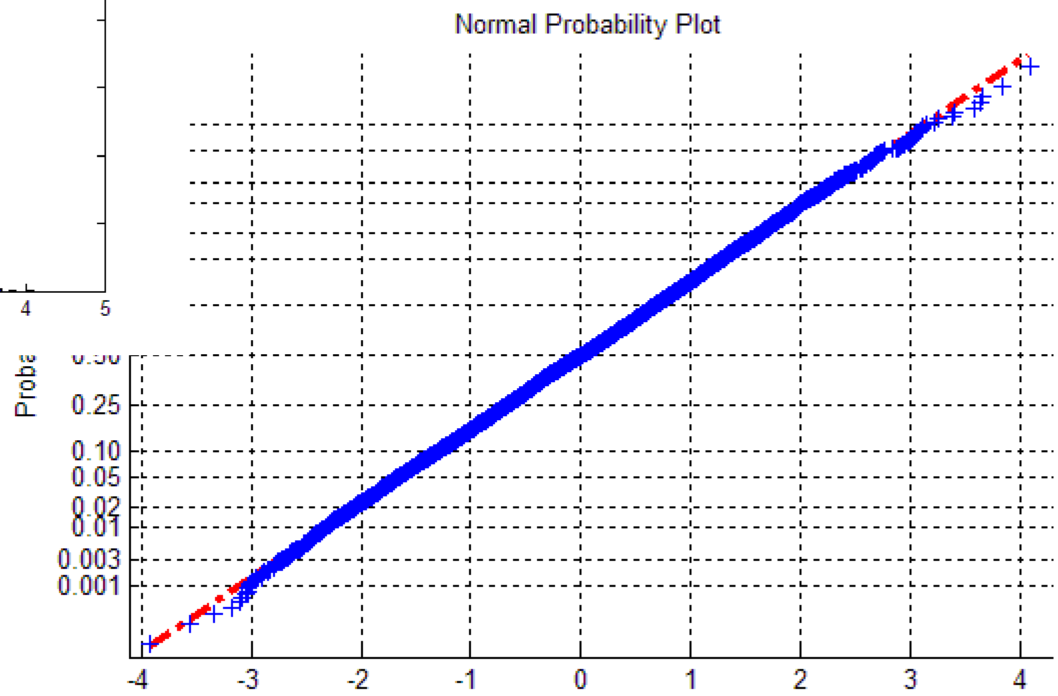
1. It is symmetric and has “lighter tails” than does a normal distribution; that is, the density curve declines more rapidly out in the tails than does a normal curve.
2. It is symmetric and heavy-tailed compared to normal distribution.
3. It is skewed.

4.6 Probability Plots

- Normal Probability plot of the normal distribution

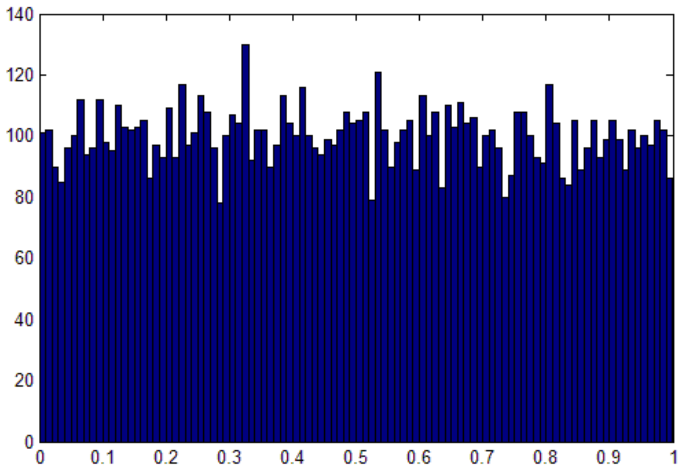


Simulation Data

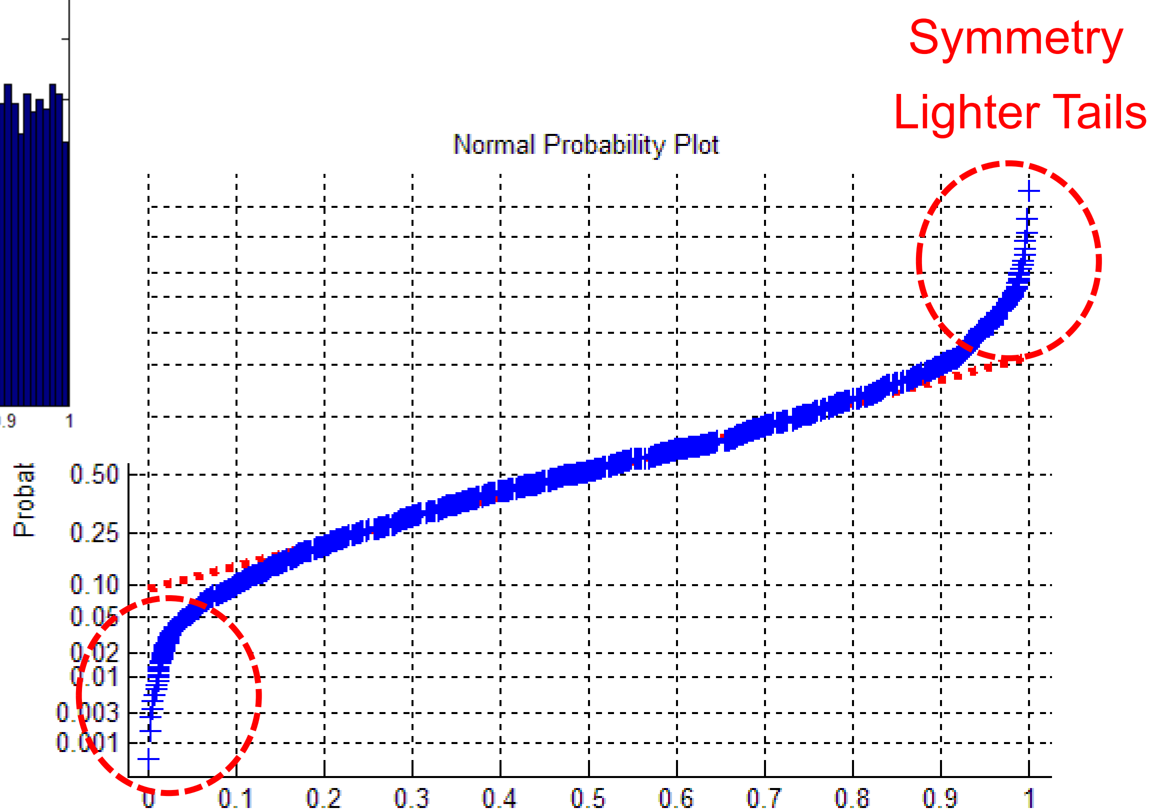


4.6 Probability Plots

- Normal Probability plot of the uniform distribution

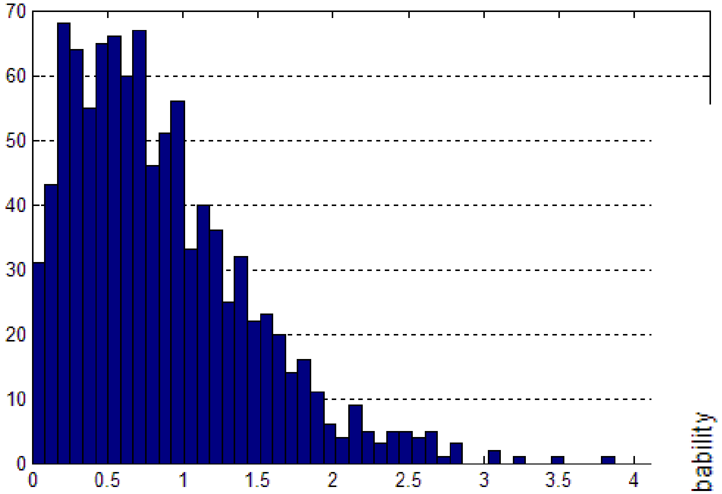


Simulation Data

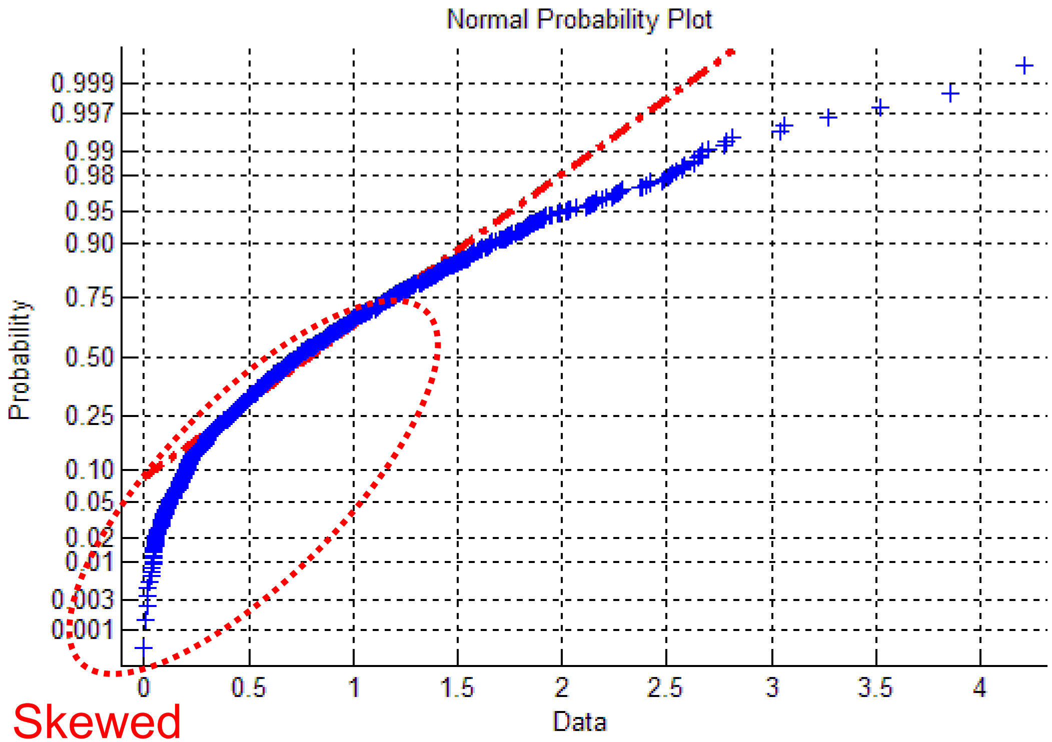


4.6 Probability Plots

- Normal Probability plot of the Weibull distribution



Simulation Data



Positively Skewed

4.6 Probability Plots

- Homework

  Ex. 88