A First Course in Quality Engineering – 2nd edition

K.S. Krishnamoorthi
With V. Ram Krishnamoorthi

Chapter 2: Statistics for Quality – Part 2

Random Variable

A random variable is a variable, which takes for it's the outcomes of a random experiment.

A random variable assumes only real numbers.

The set of all possible values of a random variable X is called its range space, denoted by Rx.

Events can be defined in the range space and their probabilities can be found.

Random Variable (contd.)

Examples of RVs.

1. The number that shows when a die is thrown
2. Number of heads obtained when a coin is tossed three times
3. Number of bad widgets in a sample of 20
4. Number of coin tosses needed to get 5 heads in a row
5. Weight of sugar in a one-pound sugar bag
6. Amount of snow fall in January in Peoria, Illinois
7. Length of life of a car battery in hours

Two Types of random variables

Random variables are of two kinds:

1. Discrete random variables
2. Continuous random variables.

A discrete random variable takes a finite (or countably infinite) number of possible values.

A continuous random variable takes infinite number of possible values. It takes values in an interval.

Probability Distributions

Probability distributions are models used to describe the behavior of random variables.

The two types of random variables have two types of distributions, because they are mathematically different

– one is continuous and the other non-continuous.

Probability mass function of a discrete r.v.

If X is a discrete random variable, a function *p(x)* is defined as the probability mass function (p.m.f.) of the random variable with the following properties

1) $\boxed{p(x) \geq 0}$, for all x

2) $\boxed{\sum_x p(x) = 1}$

3) $p(x) = P(X = x)$

Probab. mass func. of a discrete r.v.(contd.)

Example of a p.m.f.

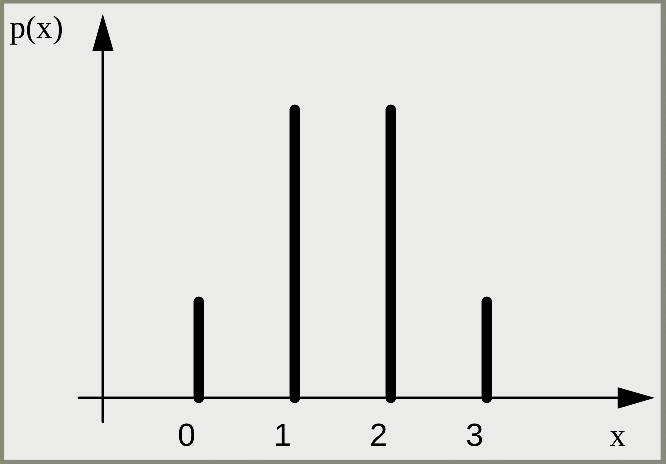
In a table

x	0	1	2	3
p(x)	1/8	3/8	3/8	1/8

In a closed form:

$$p(x) = \frac{\binom{3}{x}}{8}, x = 0,1,2,3$$

Or, as a graph:



Probability density function of a continuous r.v.

If X is a continuous random variable, a function $f(x)$ is defined with the following properties and is called the probability density function (p.d.f) of X.

1. $f(x) \geq 0$, for all values of X

2. $\int_{x} f(x)dx = 1$

3. $P(a \leq X \leq b) = \int_{a}^{b} f(x)dx$

The pdf of a cont. r.v. (contd.)

Example of a p.d.f.

$$f(x) = \begin{cases} 0.01x, & 0 \le x \le 10 \\ 0.01(20 - x), & 10 \le x \le 20 \\ 0, & otherwise \end{cases}$$

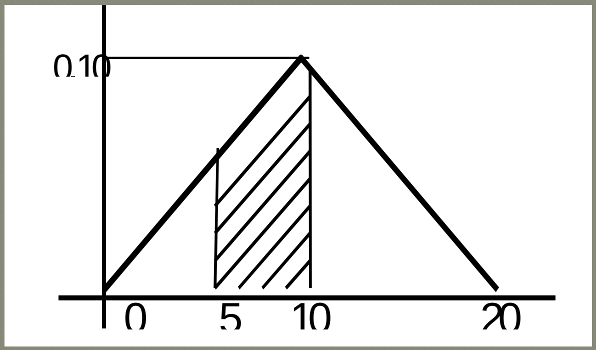
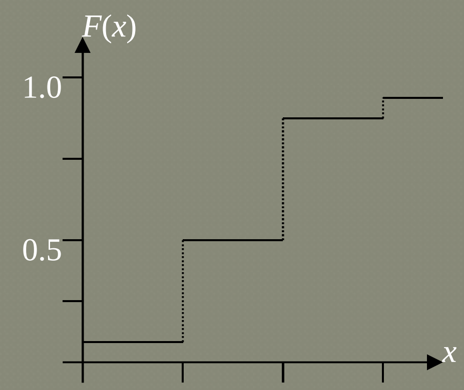


Figure 2.10 Graph of a pdf

Cumulative Distribution Function (CDF)

X is a discrete with p.m.f. *p(x)*

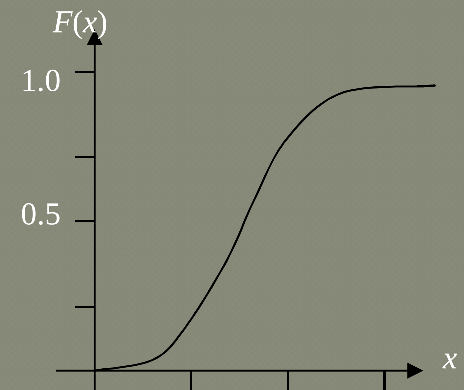
CDF: $F(x)=P(X\leq x)=\sum_{t\leq x} p(t)$



CDF of a discrete random variable

If X is continuous with pdf *f(x)*,

$F(x)=P(X\leq x)=\int_{t\leq x} f(t)\,dt$



CDF of a continuous random variable

Mean and Variance

If discrete

If continuous

Mean: $$\mu_x = \sum_x x p(x)$$

$$\mu_x = \int_x x f(x) dx$$

Variance: $$\sigma_x^2 = \sum_x (x - \mu_x)^2 p(x)$$

$$\sigma_x^2 = \int_x (x - \mu_x)^2 f(x) dx$$

Some Important Probability Distributions

1. Binomial Distribution –discrete

2. Poisson Distribution – discrete

3. Normal Distribution – continuous

(Will see later in other chapters: Exponential, Weibull, *t*-dist., Chi-squared dist., and *F*-dist..)

The Binomial Distribution

The Binomial Distribution

A random variable *X* is said to have the binomial distribution with parameters *n* and *p* if its probability distribution is given by:

$$p(x) = \binom{n}{x} p^x (1-p)^{n-x}, \; x = 0, 1, \ldots, n$$

We write $X \sim Bi(n, p)$ to indicate *X* has the binomial dist.

X represents the number of "successes" of of *n* <u>independent</u> Bernoulli trials where *p* is the probability of success and (1 - *p*) is the probability of "failure" in one trial.

Examples of Binomial R.Vs.

1. X: The number of heads when a fair coin is tossed 10 times:

$$X \sim B_i(10,\ 1/2)$$

2. Y: The number of baskets a ball player makes in 12 free throws, if her average is 0.4:

$$Y \sim B_i(12,\ 0.4)$$

3. W: The number of defectives in a sample of 20 taken from a large (??) lot having 2% defectives:

$$W \sim B_i(20,\ 0.02)$$

Calculations with Binomial Distribution

Example2.26
A sample of 12 bolts is picked from a production line and inspected. If the process produces 2% defectives, what is the probability the sample will have exactly 1 defective?

Let *X* be the number of defectives out of 12. Then:

$$X \sim \mathrm{Bi}(12,\ 0.02)$$

$$p(x) = \binom{12}{x}(.02)^{x}(.98)^{12-x},\ for\ x = 0, 1, \ldots, 12.$$

$$p(1) = \binom{12}{1}(.02)^{1}(.98)^{11} = 0.192$$

Calculations with Binomial Distribution

What is the probability there will be no more than 1 defective?

$$P(X \le 1) = p(0) + p(1)$$

P (no more than 1 def.) =

$$= \binom{12}{0}(0.02)^0(.98)^{12} + \binom{12}{1}(.02)^1(.98)^{11}$$

= 0.784 + 0.192 = 0.976

The Mean and Variance of a Binomial Variable

If $X \sim Bi(n,p)$,

it can be shown, using the definition for mean and variance that:

$$\mu_x = \sum_{x=0}^{n} x \binom{n}{x} p^x (1\text{-}p)^{n\text{-}x} = np$$

$$\sigma^2{}_x = \sum_{x=0}^{n} (x - \mu_x)^2 \binom{n}{x} p^x (1\text{-}p)^{n\text{-}x} = np(1 - p)$$

μ and σ represent the long-run average and standard deviation respectively of the binomial random variable.

The Poisson Distribution

A random variable X is said to have the Poisson distribution if its p.m.f. is given by:

$$p(x) = \frac{e^{-\lambda}\lambda^{x}}{x!} \quad , x = 0, 1, 2, \ldots$$

Examples of Poisson random variables:

1. Number of knots per sheet of plywood
2. Number of blemishes per shirt
3. Number of pinholes per sq. ft. of galvanized sheet
4. Number of accidents per month in a factory etc.

Poisson Distribution (contd.)

Mean and variance of a Poisson variable:

$$\mu_x = \sum_{x=0}^{\infty} x \frac{e^{-\lambda}\lambda^x}{x!} = \lambda$$

$$\sigma_x^2 = \sum_{x=0}^{\infty} (x-\mu_x)^2 \frac{e^{-\lambda}\lambda^x}{x!} = \lambda$$

Note that the mean and variance are equal to the value of the parameter of the distribution.

This is an unique property of the Poisson distribution.

Calculating Poisson Probabilities

Example 2.28

A typist makes on the average 3 mistakes per page. What is the probability that the page he types for a typing test will have no more than one mistake?

Let X be the number of mistakes per page.

$X \sim \text{Po}(3)$

=> $$p(x) = \frac{e^{-3} 3^{x}}{x!}$$

P (no more than one defect) = $P(X \leq 1)$ = p(0) + p(1)

$$= e^{-3}\left[\frac{3^0}{0!} + \frac{3^1}{1!}\right] = e^{-3}[4] = 0.199$$

The Normal Distribution

A random variable *X* is said to have the normal distribution with parameters μ and σ^2, if its p.d.f. is given by:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, -\infty < x < \infty, \sigma > 0$$

The graph of the density function looks as shown in Figure 2.12.

We use the notation: $\mathrm{X} \sim \mathrm{N}\left(\mu,\ \sigma^2\right)$

The Normal Distribution

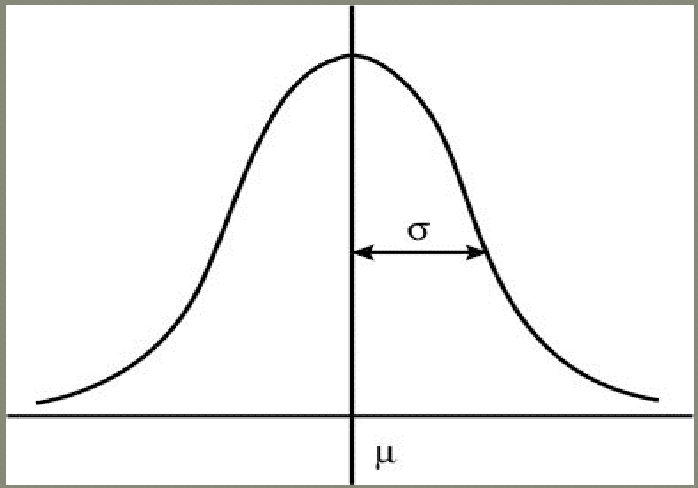


Figure 2.12a The graph of the normal p.d.f.

1. It is asymptotic with respect to the *x*-axis
2. It is symmetric with respect to a vertical line at $x = \mu$
3. The maximum value of $f(x)$ occurs at $x = \mu$
4. The two points of inflexion occur at σ distances on each side of μ

Parameters of the Normal Distribution

It can be shown:

$$\int_x f(x)dx = 1$$

[Area under the curve = 1]

$$\int_x x\, f(x)\, dx = \mu$$

[Mean of the distribution = μ]

$$\int_x (x-\mu)^2 f(x)\, dx = \sigma^2$$

[Variance of the distribution = σ^2
]

μ and σ^2 are the two parameters of the normal distribution. One of them equals its mean and the other its variance.

Finding Areas Under the Normal Curve

The Standard Normal Distribution:

A random variable that is normally distributed with mean = 0 and variance = 1 is called a Standard Normal Variable, denoted by Z. So, $Z \sim N(0,1)$.

And its pdf is given by: .

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}, -\infty < z < \infty$$

And its CDF is given by:

$$\Phi(z) = \int_{-\infty}^{z} \phi(t)dt$$

The table of CDF for various values of z is called the Normal Table.

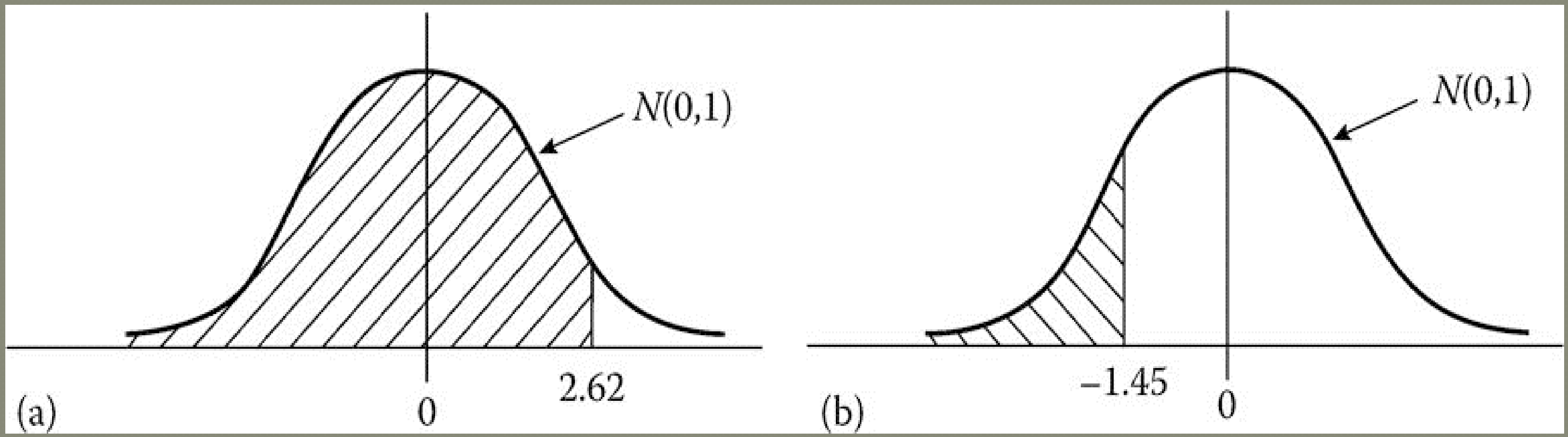
Using Standard Normal Table

Example 2.30 on how to use (Standard) Normal table.

If $Z \sim N(0,1)$

a) Find $P(Z \leq 2.62)$
From Table $P(Z \leq 2.62) = 0.9956$

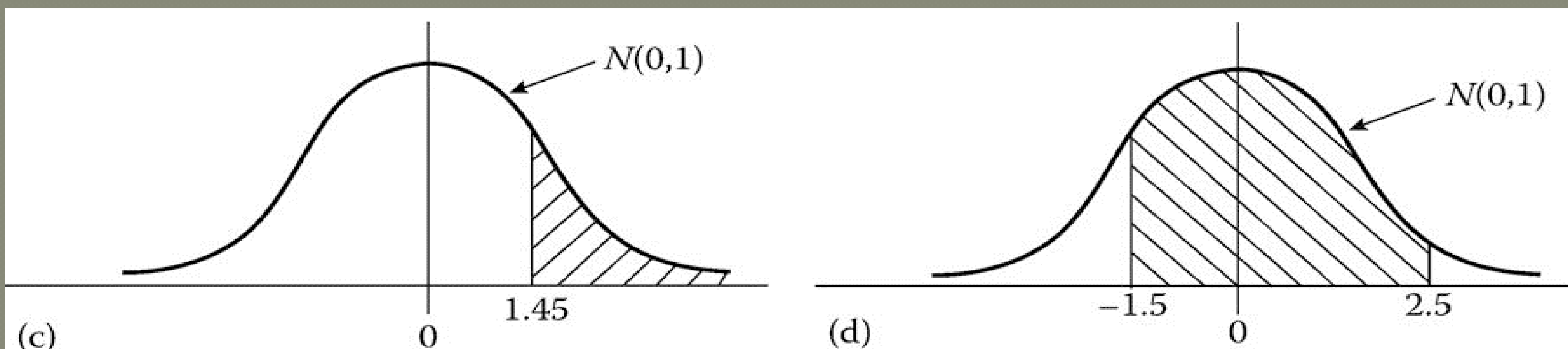
b) Find $P(Z \leq -1.45)$
From Table, $P(Z \leq -1.45) = 0.0735$



Using Standard Normal Table (contd.)

c) Find $P(Z > 1.45)$

d) Find $P(-1.5 \leq Z \leq 2.5)$



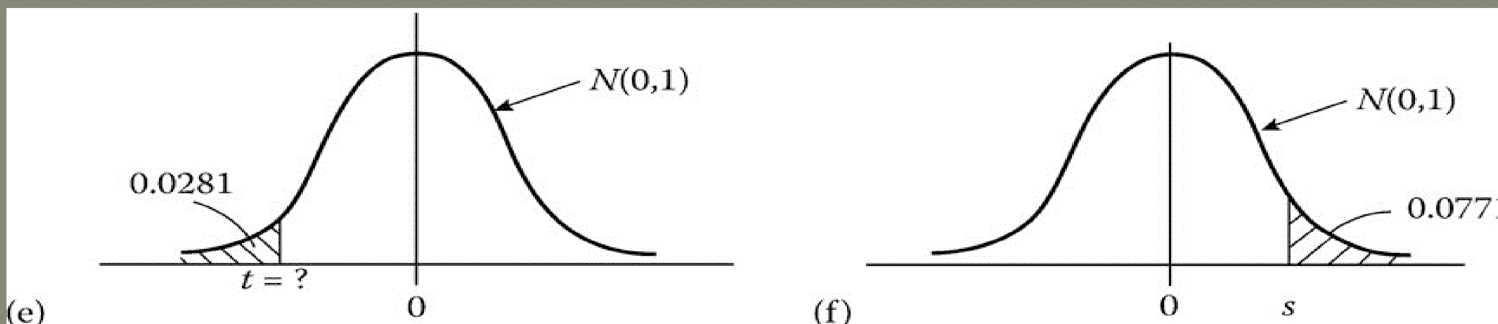
$P(Z > 1.45) = 1 – P(Z \leq 1.45) =$
$1 – 0.9265 = 0.0735$

$P(-1.5 \leq Z \leq 2.5) = P(Z \leq 2.5) – P(Z \leq -1.5)$
$= 0.9938 – 0.0668 = 0.9270$

Using Standard Normal Table (contd.)

e) Find t such that $P(Z < t) = 0.0281$

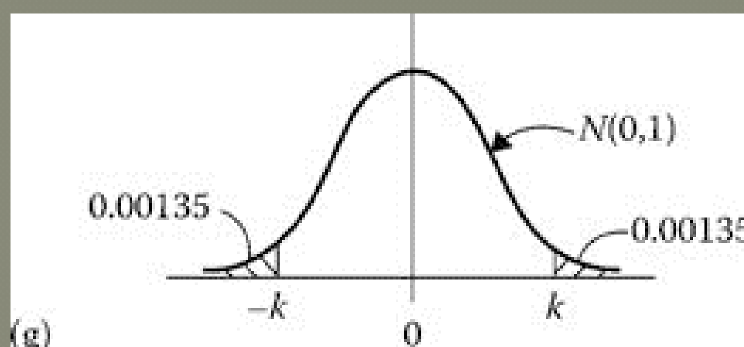
f) Find s such that $P(Z > s) = 0.0771$



From the table, $t = -1.91$

$s = 1.425$

g) Find k such that $P(-k < Z < k) = 0.9973$



Because of symmetry $P(Z \leq -k) = P(Z > k) = 0.00135$, $k = 3$

Areas Under any Normal Distribution

Theorem: $$\text{If } X \sim N(\mu,\sigma^2), \text{ then } \frac{X-\mu}{\sigma} \sim N(0,1)$$

Example 2.31

A random variable $X \sim N(2.0,\ 0.0025)$.

(a) $$\mathrm{P}(X \le 1.87) = \mathrm{P}\left(\frac{X-2.0}{0.05} \le \frac{1.87-2.0}{0.05}\right) = \mathrm{P}\left(Z \le -\frac{0.13}{0.05}\right) = \mathrm{P}(Z \le -2.6) = 0.0047$$

(b) $$P(X > 2.2) = P\left(Z > \frac{2.2-2.0}{0.05}\right) = P(Z > 4.0) = 0.0$$

Finding Areas Under Normal Distribution (contd.)

c)

$$\mathrm{P}(1.9 \le X \le 2.1) = \mathrm{P}\left(\frac{1.9-2.0}{0.05} \le Z \le \frac{2.1-2.0}{0.05}\right)$$
$$= 0.972 - 0.0228 = 0.9544$$

d) Find t such that P(X > t) = 0.05

$$P\left(Z > \frac{t-2}{0.05}\right) = 0.05 \Rightarrow \frac{t-2}{0.05} = 1.645 \Rightarrow t = 2.08225$$

Normal Distribution Problems (contd.)

Example 2.31

(e) Find *k* such that $\mathrm{P}(\mu - k\sigma \le X \le \mu + k\sigma) = 0.9973$

$$\Rightarrow \quad \mathrm{P}\left(\frac{\mu - \mathrm{k}\sigma - \mu}{\sigma} \le Z \le \frac{\mu + \mathrm{k}\sigma - \mu}{\sigma}\right) = 0.9973$$

$$P(\text{-}k \le Z \le k) = 0.9973$$

$$k = 3$$

(g) Find k such that $\mathrm{P}(\mu - k\sigma \le X \le \mu + k\sigma) = 0.6826$

$$\Rightarrow \mathrm{P}\left(\frac{\mu - \mathrm{k}\sigma - \mu}{\sigma} \le Z \le \frac{\mu + \mathrm{k}\sigma - \mu}{\sigma}\right) = 0.6826$$

$$\mathrm{P}(-k \le Z \le k) = 0.6826$$

$$k = 1.0$$

Application of Normal Distribution

Example 2.32

D is the diameter of bolts, and, $D \sim N(0.25, 0.01^2)$

Bolt specs call for 0.24 ± 0.02”.

a) What proportion of the bolts are outside spec.

We need: $$\mathrm{P}\left(D < 0.22\right) + \mathrm{P}\left(D > 0.26\right)$$

$$= \mathrm{P}\left(Z < \frac{0.22 - 0.25}{0.01}\right) + \mathrm{P}\left(Z > \frac{0.26 - 0.25}{0.01}\right)$$
$$= P(Z < -3) + P(Z > 1)$$
$$= 0.00135 + 0.1587 = 0.16$$

i.e., 16% of the bolts are outside specification.

Centering Improves a Process

b) If the process mean is moved to coincide with the center of spec, what proportion will be defective?

When process mean coincides with spec center:
P(D < 0.22) + P(D > 0.26)

$$
\begin{aligned}
&= \mathrm{P}\left(Z < \frac{0.22 - 0.24}{0.01} \right) + \mathrm{P}\left(Z > \frac{0.26 - 0.24}{0.01} \right) \\
&= 2 \times 0.0228 = 0.0456
\end{aligned}
$$

i.e., 4.56% will be outside specification

Centering a process will improve process conditions; further improvement has to come from reducing variability.

Reducing Variability

Example 2.33

X is the random variable that denotes the life of the batteries in years.

$$X \sim N(5,\, 0.5^2)$$

a) If any battery failing before 4 years is replaced under warranty what proportion of batteries need replacement?

$$\mathrm{P}\left(\mathrm{X} < 4\right) = \mathrm{P}\left(Z < \frac{4-5}{0.5}\right) = \mathrm{P}\left(Z < -2.0\right) = 0.0228$$

i.e., 2.28% will have to be replaced during warranty.

Reducing Variability

b) What should be the std. deviation if no more than 0.02% should require replacement?

Let σ' be the new standard deviation.
Find σ' such that $P(X<4) = 0.005$

$$\Rightarrow \mathrm{P}\left(Z < \frac{4-5}{\sigma'}\right) = 0.005 \Rightarrow \frac{4-5}{\sigma'} = -2.575$$

$$\Rightarrow \sigma' = \frac{1}{2.575} = 0.3883$$

The standard deviation must be reduced from 0.5 to 0.39; i.e., a 23% reduction in variability is needed.

Distribution of the Sample Average

Theorem:

If , $X \sim N(\mu, \sigma^2)$ then, $\bar{X}_n \sim N\left(\mu, \frac{\sigma^2}{n}\right)$

The Central Limit Theorem (CLT)

Let a population have any distribution with a finite mean $= \mu$ and finite variance $= \sigma^2$.

Then, $\bar{X}_n \xrightarrow[n \to \infty]{} N(\mu, \frac{\sigma^2}{n})$

Problem relating to X-bar

Example 3.34

$$X \sim N\left(2.0, 0.15^2\right)$$

$$\bar{X}_9 \sim N\left(2.0, 0.15^2 / 4\right)$$

$$\mu_{\bar{X}} = 2.0$$

$$\sigma_{\bar{X}} = \sqrt{\left(0.0.15^2 / 9\right)} = (0.15 / 3) = 0.05$$

We need :

$$P(X \leq 2.1) = P\left(Z \leq \frac{2.1 - 2.0}{0.05}\right) = P(Z \leq 2) = 0.9772$$

That is, 97.72% of the averages will be less than 2.1 in.

Problem relating to X-bar

Example 2.35:
Setting limits for *X*-bar (n = 4) to include 99.73%:

$$X \sim N\left(2.0, 0.0225\right) \Rightarrow \bar{X} \sim N\left(2.0, 0.0225/4\right)$$

$$\mu_{\bar{X}} = 2.0 \qquad \sigma_{\bar{X}} = \sqrt{\left(0.0225/4\right)} = 0.15/2 = 0.075$$

In order to include 99.73% of $\bar{X}$s, the limits must be located at 3-sigma distance from the .

Therefore the limits must be at 2.0 $\pm$ 3(0.075),
i.e., at l.775 and 2.225.

EXERCISE IN DISTRIBUTIONS