Probability and Statistics

Chapter 4, Figliola & Beasley

Suppose we manufacture a large numbe of round bearings.

When we measure the diameter of the bearings, we find their diameters are not the same.

Random Sample

It costs too much to measure each bearing, so we draw a random sample from the total population.

Important Considerations

1. How many bearings are needed in the sample? $N =$

2. How do we ensure that the sample is random and representative of the population?

Measurement Model

$X_i = x_t + b + E_i \qquad i = 1, 2 \cdots N$

X, E - random variables (R.V.)

x_t, b - deterministic (d)

x_t - true value of diameter (d)

b - measurement bias (d)

E_i - random measurement error (R.V.)

X_i - measured value of diameter (R.V.)

Data Set

Sample values of R.V. X_i; $i = 1, \cdots 20$

Table 4.1 Sample of Random Variable x

i	x_i	i	x_i
1	0.98	11	1.02
2	1.07	12	1.26
3	0.86	13	1.08
4	1.16	14	1.02
5	0.96	15	0.94
6	0.68	16	1.11
7	1.34	17	0.99
8	1.04	18	0.78
9	1.21	19	1.06
10	0.86	20	0.96

Why is X_1 called a R.V. when it has a determistic value, 0.98, assigned?

If we drew another random sample, 20 bearings, X_1 would have a different value. Every time we draw a random sample X_1 tak

on a different value.

$X_1 = 0.98$ is called a realization of the R.

Plot the data set on the x-axis

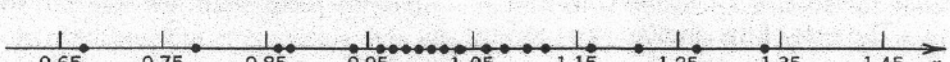


Figure 4.1 Concept of density in reference to a measured variable (from Example 4.1).

Histogram

Divide the x axis into K equally sized bins starting at x_{min} and ending at x_{max}. Number of bins

$$K = 1.87(N-1)^{0.40} + 1$$

$$= 1.87(20-1)^{0.40} + 1 = 7.072$$

Use $K = 7$.

Divide the x axis into 7 bins startin at 0.65 and ending at 1.35. The following Table shows the bins, counts and relative frequency.

j	Interval	n_j	$f_j = n_j/N$
1	$0.65 \le x_i < 0.75$	1	0.05
2	$0.75 \le x_i < 0.85$	1	0.05
3	$0.85 \le x_i < 0.95$	3	0.15
4	$0.95 \le x_i < 1.05$	7	0.35
5	$1.05 \le x_i < 1.15$	4	0.20
6	$1.15 \le x_i < 1.25$	2	0.10
7	$1.25 \le x_i < 1.35$	2	0.10

Plot the number of counts in each bin and the relative frequency

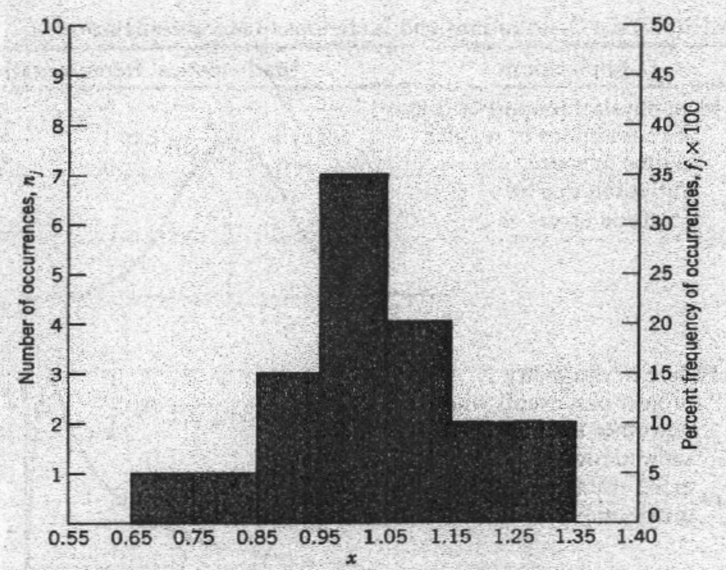


Figure 4.2 Histogram and frequency distribution for data in Table 4.1.

Continuous Random Variables

If the realizations of a random variable are any points on the x-axis, i.e. real numbers, they are continuous R.V.'s. The bearing diameter measurements, X_i, are continuous R.V.'s

Continuous R.V.'s have continuous probability density functions. The histogra suggests that the best model for the RV X_i representing bearing diameter is

$$p_X(x) = \frac{1}{\sqrt{2\pi}\,\sigma_X} e^{-\frac{1}{2}\frac{(x-\mu_X)^2}{\sigma_X^2}}$$

μ_X is the mean

σ_X is the standard deviation

There are statistical test called Goodness of Fit Criterion that indicate whether the data set can be modeled by a normal density function. However, it tak a lot more data, say $N = 1000$, to satisfy these tests. This analysis is beyond the scope of this course.

Discrete Random Variable

If the realizations of a RV are a set of integers, the R.V. is discrete

Ex: Toss a die. Let the R.V. Y represent the number that comes up

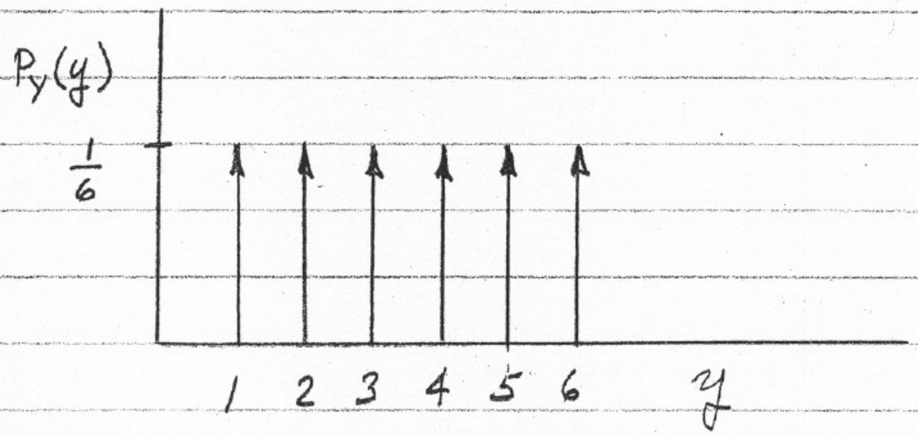
$$Y = \{1, 2, 3, 4, 5, 6\}$$

$$P(Y=1) = \frac{1}{6}$$

$$\vdots$$

$$P(Y=6) = \frac{1}{6}$$

Density function



Continuous R.V.'s

Expected-value

$$E\{g(x)\} = \int_{-\infty}^{\infty} g(x)\, p_X(x)\, dx$$

Mean μ_X

Let $g(x) = x$

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

Variance σ_X^2

Let $g(x) = (x - \mu_X)^2$

$$\sigma_X^2 = E\{(X-\mu_X)^2\} = \int_{-\infty}^{\infty} (x-\mu_X)^2 p_X(x)\, dx$$

Standard Deviation

$$\sigma_X = \sqrt{\sigma_X^2}$$

standarized Normal R.V. Z

$$Z = \frac{X - \mu_X}{\sigma_X}$$

$$\mu_Z = 0$$

$$\sigma_Z = 1$$

Probability

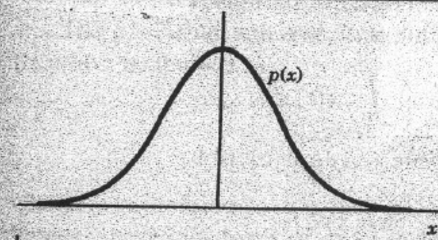
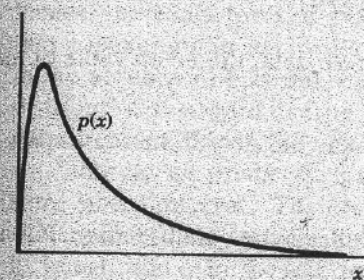
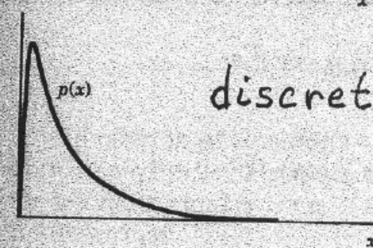
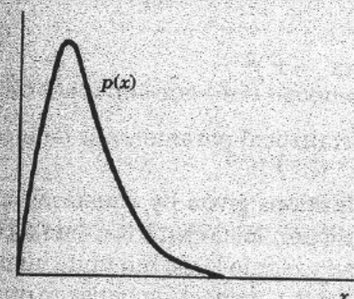
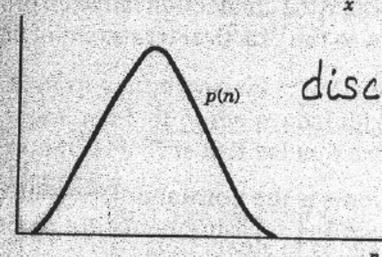
$$P\{Z < z_1\}$$

$$= \frac{1}{\sqrt{2\pi}\,1} \int_{-\infty}^{z_1} e^{-z^2/2}\, dz$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{0} e^{-z^2/2}\, dz + \frac{1}{\sqrt{2\pi}} \int_{0}^{z_1} e^{-z^2/2}\, dz$$

$$= \frac{1}{2} + \text{Table 4.3}$$

Table 4.2 Standard Statistical Distributions and Relations to Measurements

Distribution	Applications	Mathematical Representation	Shape
Normal	Most physical properties that are continuous or regular in time or space. Variations due to precision error. Measurement error	$p(x) = \frac{1}{\sigma(2\pi)^{1/2}} \exp\left[-\frac{1}{2}\frac{(x-x')^2}{\sigma^2}\right]$	
Log normal	Failure or durability projections; events whose outcomes tend to be skewed toward the extremity of the distribution. Impurities in gases	$p(x) = \frac{1}{x\sigma(2\pi)^{1/2}} \exp\left[-\frac{1}{2}\ln\frac{(x-x')^2}{\sigma^2}\right]$	
Poisson	Events randomly occurring in time; $p(x)$ refers to probability of observing x events in time t. Here λ refers to x'. Particle counts in gases	$p(x) = \frac{e^{-\lambda}\lambda^x}{x!}$	 discrete
Weibull	Fatigue tests; similar to log normal applications. strength of ceramics	See [4]	
Binomial	Situations describing the number of occurrences, n, of a particular outcome during N independent tests where the probability of any outcome, P, is the same. Pass/fail	$p(n) = \left[\frac{N!}{(N-n)!n!}\right] P^n(1-P)^{N-n}$	 discrete

Confidence Intervals

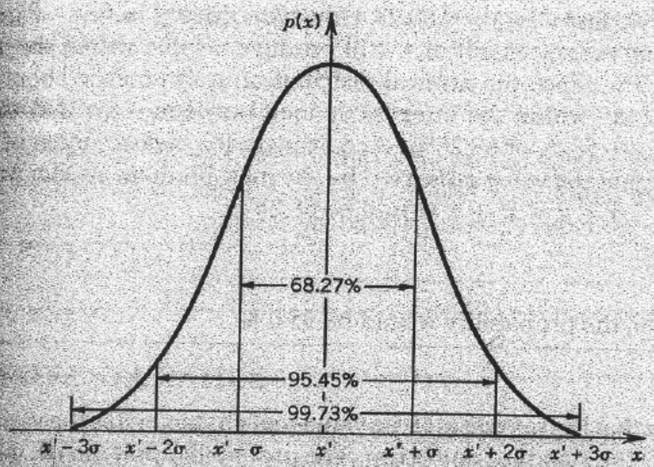


Figure 4.4 Relationship between the probability density function and its statistical parameters x' and σ for a normal distribution.

Assume a normal density with mean μ_x and variance σ_x^2

$$P\{\mu_x - \sigma_x < X \leq \mu_x + \sigma_x\} = 68.27\%$$

$$P\{\mu_x - 2\sigma_x < X \leq \mu_x + 2\sigma_x\} = 95.45\%$$

$$P\{\mu_x - 3\sigma_x < X \leq \mu_x + 3\sigma_x\} = 99.73$$

Also referred to as Sigma Bounds.