Spring 1999 EE290H Tentative Weekly Schedule

1. Functional Yield of ICs and DFM. 2. Parametric Yield of ICs. 3. Yield Learning and Equipment Utilization.	IC Yield & Performance
4. Statistical Estimation and Hypothesis Testing. 5. Analysis of Variance. 6. Two-level factorials and Fractional factorial Experiments. 7. System Identification. 8. Parameter Estimation.	Process Modeling
9. Statistical Process Control. ---> Distribution of projects. (week 9) 10. Break 11. Run-to-run control. 12. Real-time control. ---> *Quiz on Yield, Modeling and Control (week 12)*	Process Control
13. Off-line metrology - CD-SEM, Ellipsometry, Scatterometry 14. In-situ metrology - temperature, reflectometry, spectroscopy	Metrology
15. The Computer-Integrated Manufacturing Infrastructure ---> *Presentations of project results. (week 17)*	Manufacturing Enterprise

A Brief Statistical Primer

Basic distributions. The central limit theorem. Sampling, estimation and hypothesis testing.

IC Production Requirements

- IC design must be modified for manufacturability.
- Incoming material and utilities must be qualified.
- Equipment must be qualified for volume production.
- Process must be debugged and transferred to high volume fab line.
- Yield detractors must be identified and eliminated.
- Efficient test procedures must be established.

Issues that SPC & DOE deal with

SPC helps remove the causes of problems and makes the process more stable over time.

DOE allows the exploration of the process settings and helps find the optimum operating conditions.

Quality is fitness for use.

There are two kinds of Quality:

a. Quality of design.

b. Quality of conformance.

DOE and SQC deal with quality of conformance.

SPC is one of the tools used to implement SQC.

Statistics is the art of making inferences about the whole by observing a sample.

The Evolution of Manufacturing Science

1. Invention of machine tools. English system (1800).
   mechanical - accuracy
2. Interchangeable components. American system (1850).
   manufacturing - repeatability
3. Scientific management. Taylor system (1900).
   industrial - reproducibility
4. Statistical Process Control (1930).
   quality - stability
5. Information Processing and Numerical Control (1970).
   system - adaptability
6. Intelligent Systems and CIM (1980).
   knowledge - versatility

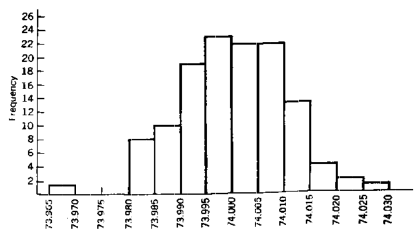
Outline

- Discrete Distributions
- Continuous Distributions (Normal)
- Central Limit Theorem

- Sampling
- Estimation
- Hypothesis Testing

Presenting and Summarizing Data

The histogram



The sample average

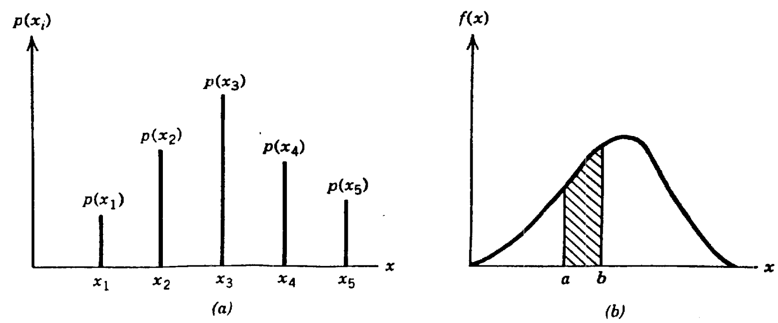
$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n}$$

The sample variance

$$s^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}$$

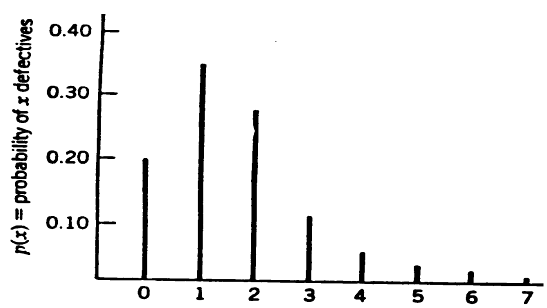
The Distribution of x

Discrete Distributions   Continuous Distributions



Probability distributions. (*a*) Discrete case. (*b*) Continuous case.

A discrete distribution: the binomial



Binomial Distribution with p = 0.10 and n=15.

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

$$\text{mean} = np, \quad \text{variance} = np(1-p)$$

$$\text{where } \binom{n}{x} \equiv \frac{n!}{x!(n-x)!}, \quad n! \equiv 1 \cdot 2 \cdot \ldots \cdot (n-1) \cdot n$$

$$\text{the sample fraction } \hat{p} = \frac{D}{n} \qquad \text{mean} = p, \quad \text{variance} = \frac{p(1-p)}{n}$$

Example: Using the Binomial Distribution

The probability of a broken wafer in a loading operation is 0.1% After loading 100 wafers, what is the probability that:

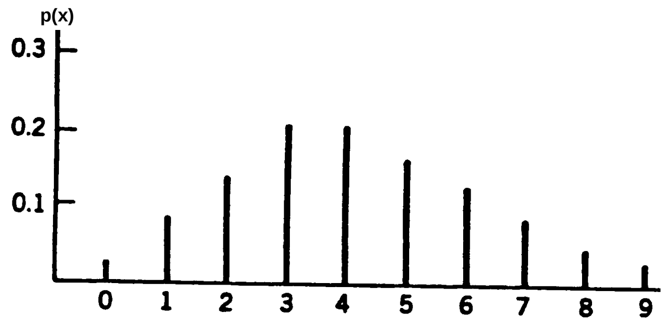
All wafers are OK?

1 wafer broke?

5 wafers broke?

More than 5 wafers broke?

Another discrete distribution: the Poisson



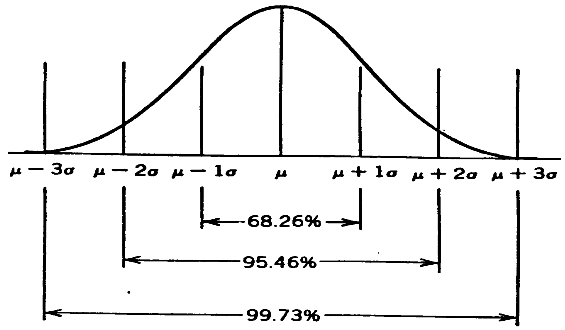
The Poisson Distribution with $\lambda = 4$.

$$p(x) = \frac{e^{-\lambda}\, \lambda^x}{x!} \qquad x = 0, 1, \qquad \text{where } x! \equiv 1 \cdot 2 \cdot 3 \cdot \ldots \cdot (x-1) \cdot x$$

mean: $\mu = \lambda$

variance: $\sigma^2 = \lambda$

A continuous distribution: the Normal



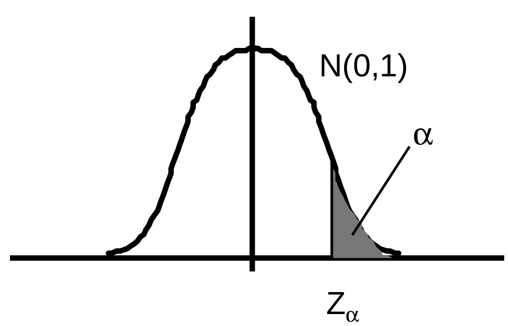
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \qquad -\infty < x < \infty \qquad x \sim N(\mu, \sigma^2)$$

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

$$z \equiv \frac{x-\mu}{\sigma}, \quad P\{x < a\} = P\left\{z < \frac{a-\mu}{\sigma}\right\} \equiv \Phi\left(\frac{a-\mu}{\sigma}\right)$$

Useful Definitions on the Normal Distribution

$z \sim N(0, 1)$ This is the Standard Normal Distribution



for any Normal distribution, z signifies the "number of sigmas away from the mean".
z_α is the number that "cuts" an upper tail with area α

Notice that $z_\alpha = -z_{1-\alpha}$

The Normal Distribution

THE
NORMAL
LAW OF ERROR
STANDS OUT IN THE
EXPERIENCE OF MANKIND
AS ONE OF THE BROADEST
GENERALIZATIONS OF NATURAL
PHILOSOPHY ♦ IT SERVES AS THE
GUIDING INSTRUMENT IN RESEARCHES
IN THE PHYSICAL AND SOCIAL SCIENCES AND
IN MEDICINE AGRICULTURE AND ENGINEERING ♦
IT IS AN INDISPENSABLE TOOL FOR THE ANALYSIS AND THE
INTERPRETATION OF THE BASIC DATA OBTAINED BY OBSERVATION AND EXPERIMENT

Example: Table Lookup for Normal Distribution

The wafer to wafer thickness of a poly layer is distributed normally around 500nm with a σ of 20nm:

$$Pth \sim N(500nm, 400nm^2)$$

What is the probability that a given wafer will have polysilicon thicker than 510nm?

... thinner than 480nm?

... between 490 and 515nm?

The Additivity of Variance

If $y = a_1x_1+a_2x_2+...+a_nx_n$ then $\mu_y = a_1\mu_1+...+a_n\mu_n$

and

$$\sigma_y^2 = a_1^2\sigma_1^2+...+a_n^2\sigma_n^2$$

This applies under the assumption that the parameters x are *independent*..

For example, consider the thickness variance of a layer defined by two consecutive growths:

$$\mu_t = \mu_{g1}+\mu_{g2}$$
$$\sigma_t^2 = \sigma_{g1}^2+ \sigma_{g2}^2$$

or by a growth step followed by an etch step:

$$\mu_t = \mu_g-\mu_e$$
$$\sigma_t^2 = \sigma_g^2+ \sigma_e^2$$

Example: How to combine consecutive steps

The thickness of a SiO_2 layer is distributed normally around 600nm with a σ of 20nm:

GOx~N (600nm, 400nm^2)

During a polysilicon removal step with limited selectivity, some of the oxide is removed. The removed oxide is:

ROx~N (50nm, 25nm^2)

What is the probability that the final oxide thickness is between 540 and 560nm?

The Correlation Coefficient

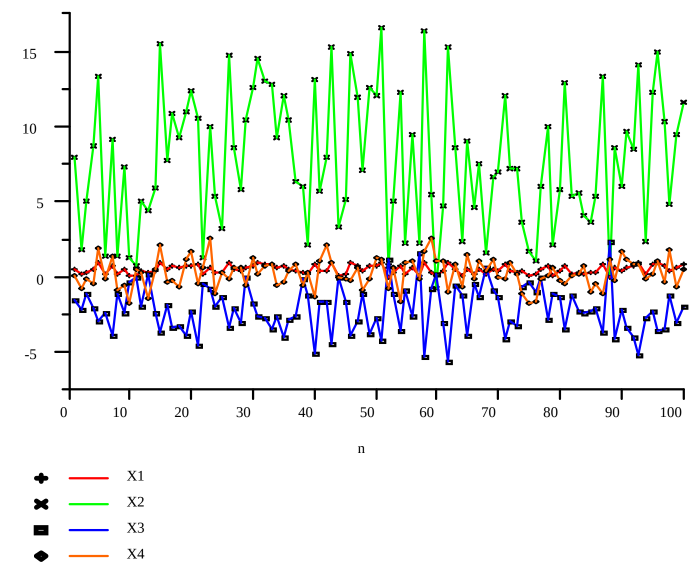
The correlation coefficient ρ, is a statistical moment that gives a measure of *linear dependence* between two random variables. It is estimated by:

$$r = \frac{s_{xy}}{s_x s_y}$$

where s_x and s_y are the square roots of the estimates of the variance of x and y, while s_{xy} is an estimate of the *covariance* of the two variables and is estimated by:

$$s_{xy} = \sum_{i=1}^{n} \frac{(x_i - \bar{x})(y_i - \bar{y})}{n-1}$$

Correlation Example

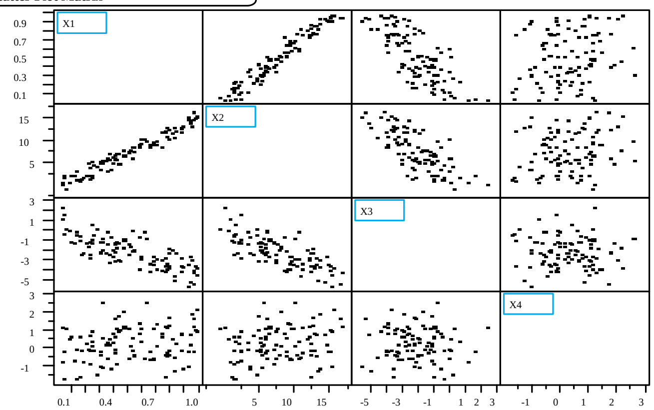


Correlation Example

Correlations

Variable	X1	X2	X3	X4
X1	1.0000	0.9719	-0.7728	0.2089
X2	0.9719	1.0000	-0.7518	0.2061
X3	-0.7728	-0.7518	1.0000	-0.0753
X4	0.2089	0.2061	-0.0753	1.0000

Scatter Plot Matrix



Using the Correlation Coefficient

In fact, the only way to stop variances from accumulating is by making sure that the parameters have the appropriate *correlation* ρ.

$$\mu_z = a\mu_x + b\mu_y$$

$$\sigma_z^2 = a^2\sigma_x^2 + b^2\sigma_y^2 + 2ab\rho\,\sigma_x\sigma_y$$

For example, two consecutive growths can be controlled if we set $\rho = -1$

$$\mu_t = \mu_{g1} + \mu_{g2}$$

$$\sigma_t^2 = \sigma_{g1}^2 + \sigma_{g2}^2 - 2\sigma_{g1}\sigma_{g2}$$

and a growth followed by an etch can be controlled by setting $\rho = 1$

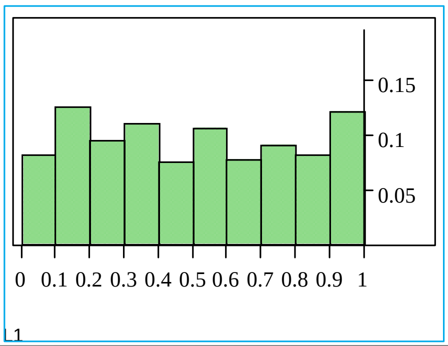
$$\mu_t = \mu_g - \mu_e$$

$$\sigma_t^2 = \sigma_g^2 + \sigma_e^2 - 2\sigma_g\sigma_e$$

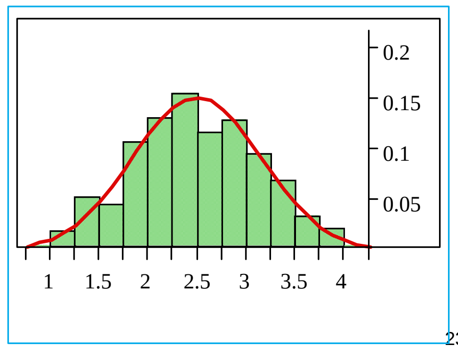
The Central Limit Theorem:

The sum of <u>independent</u> random variables <u>tends</u> to have a normal distribution.

Uniformly distributed number



Sum of 5 unif. distr. numbers:



Statistical Moments and their Estimators

The mean (μ), the sigma (σ), the correlation (ρ), etc. are symbolized with Greek characters and they refer to "true", but hidden values *which we cannot know exactly*.

These are the *moments* of a population.

The average, the standard deviation, the correlation coefficient, etc. are represented by Latin characters and they refer to values, *which we can calculate from the data.*

These values are used to estimate the true (but impossible to know exactly) moments. They are *estimators*.

Sampling and Estimation

Sampling: the act of making inferences about populations.

Random sampling: when each observation is identically and independently distributed.

Statistic: a function of sample data containing no unknowns. (e.g. average, median, standard deviation, etc.)

A statistic is a random variable. Its distribution is a sampling distribution.

Example: Estimating the mean of a normal dist.

The thickness of a poly layer is distributed normally around 500nm with a σ of 10nm:

$$Pth \sim N\,(500nm,\ 100m^2)$$

We randomly select 50 wafers, measure the poly thickness and calculate the average of the fifty readings:

$$\overline{Pth} = \frac{1}{50}\sum_{i=1}^{50} Pth_i$$

What is the distribution of $\overline{Pth}$?

What is the probability that $\overline{Pth}$ will be between 495 and 505nm?

Example: The Mean of a Normal Dist. (cont)

Now that we understand how the average is distributed, lets use it to estimate the mean (assuming that we know that σ is 10nm).

If the 50 measurements yielded an average of 503nm, what can we say about the unknown mean?

What is the estimated value of the unknown mean?

What is the probability that the unknown mean is between 500 and 506nm?

The chi-square distribution

The *chi-square* distribution with n degrees of freedom is defined as:

if $x_i \sim N(0,1)$ for i = 1,2,...,n then:

$$x_1^2 + x_2^2 + \dots + x_n^2 \sim \chi^2_n$$

Usage: To find the distribution of the standard deviation when the mean is known. Also to define the student-t distribution (more on this later).