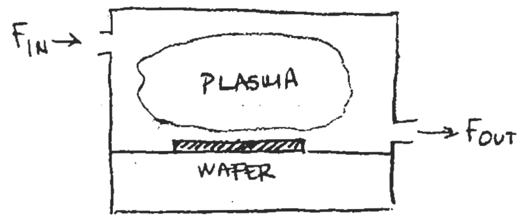
6.780 Semiconductor Manufacturing
Spring Term, 2003

D. Boning

SENSORS AND SIGNALS IN SEMICONDUCTOR MANUFACTURING

Available data for monitoring, control, and diagnosis can be categorized in several ways

- STATE SENSED



① Wafer State
e.g. film thickness, resistivity

② Process/Environment State
e.g. gas concentrations, ion density in plasma

③ Equipment State
e.g. machine settings, flows tool wear

- Often only indirect measures of state parameters are available
- MODELS invariably needed to convert sensor signal to state information

- DATA AVAILABILITY & TIMELINESS

① Real time – information available during process step
vs. Run to Run – information available from one run (wafer, lot, batch) to the next
vs. Many Timescales!

② In-situ – available with wafer in process (i.e. during processing)
On-line – usually means the measurement can be automatically taken without additional operator intervention (e.g. on the process tool)
In-line – usually a prescribed measurement as part of process flow
Ex-situ – requires additional/separate measurement
End-of-Line – available at end of process rather than right after a particular process step

- DATA STREAMS – Multivariate & coupled
  – Time-Series
  – Spatial

MULTIPLE CONTROL CHARTS – INDEPENDENT VARIABLES

- If we are simultaneously monitoring multiple quality characteristics, the univariate assumptions on type I (false alarm) rate are inaccurate...

Consider p statistically independent characteristics, where each chart has been set with $Pr\{\text{false alarm}\} = \alpha$;

⇒ What is overall probability of a false alarm? (α')

$$\alpha' = 1 - (1-\alpha)^p \approx p \cdot \alpha \quad \text{for } \alpha \text{ small}$$

E.g. for 3σ control limits, $\alpha = 0.0027$

With 2 charts	$\alpha' = 0.0054$	Many more false alarms!!
5 charts	$\alpha' = 0.0134$	
10 charts	$\alpha' = 0.0267$	
100 charts	$\alpha' = 0.2369$	

Or to look at it another way, the

$$Pr\{\text{all plot inside limit} \mid \text{in control}\} = (1-\alpha)^p \approx 1 - p\alpha$$

- This is a common problem: a complicated tool may have 30 charts! with ARL = 370 for false alarm on univariate chart design ⇒ false alarm every 12 runs!

- Approach: ① USE CORRECT α based on OVERALL α' RISK!

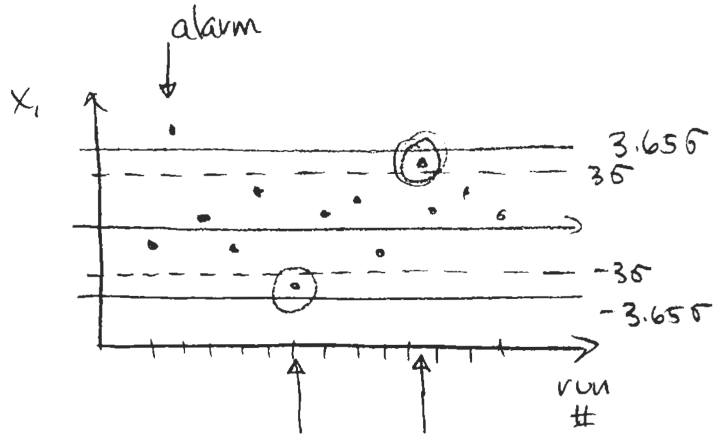
$$\alpha = 1 - (1-\alpha')^{1/p} \qquad (\text{approximately } \alpha = \alpha'/p)$$

Example: 10 control charts

$\alpha' = 0.0027$ gives $z_{\alpha'/2} = 3$ (3σ rule)

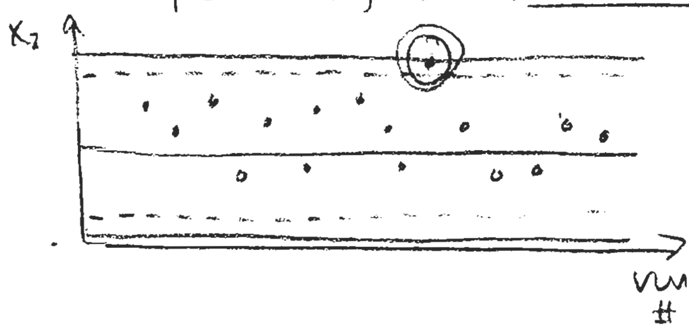
multivariate limits ⇒ $\alpha = 0.00027$ $\quad z_{\alpha/2} = \underline{\underline{3.65}}$

Appropriately Expanded Control Limit

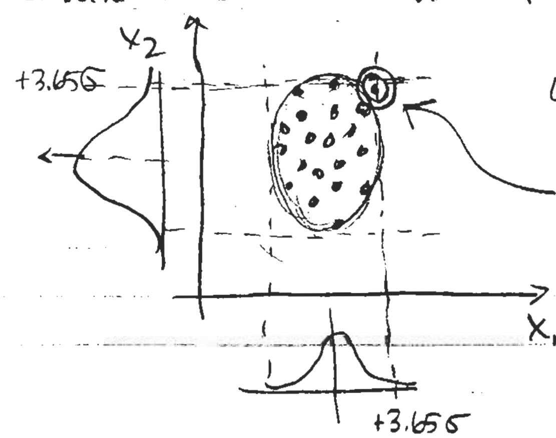


Would have been univariate alarms,
but we ignore. Why?
$\Rightarrow$ build in extra "deadzone" for false alarms

② We could also include additional rules based on probability of simultaneous excursions ..



- Indeed, we can formalize this "joint distance" notion. Suppose, as we have assumed, that x_1 & x_2 are bivariate independent normal distributions:



where $\sigma_2 > \sigma_1$

e.g. Point inside individual charts — but if "high" in multiple charts this is very unlikely & thus should be an alarm.

What we're looking for is some "generalized distance" measure that
(1) captures total deviation from mean (centroid)
<u>and</u> (2) weights distance in each direction by likelihood of points in that direction (σ_{x_i})

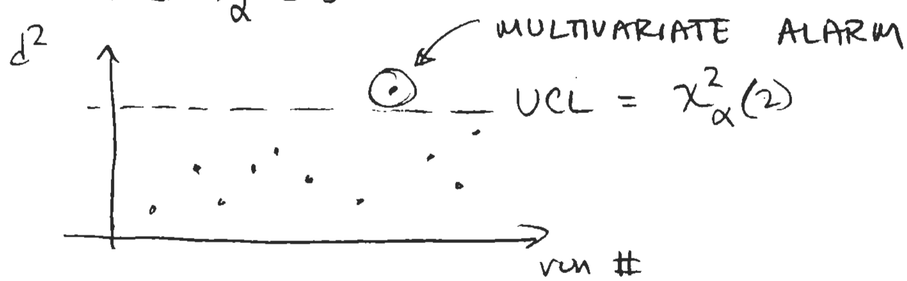
$$d = \sqrt{\left(\frac{x_1 - \mu_1}{\sigma_1}\right)^2 + \left(\frac{x_2 - \mu_2}{\sigma_2}\right)^2}$$

$$d^2 = \left(\frac{x_1 - \mu_1}{\sigma_1}\right)^2 + \left(\frac{x_2 - \mu_2}{\sigma_2}\right)^2 = \frac{\sigma_2^2 (x_1 - \mu_1)^2 + \sigma_1^2 (x_2 - \mu_2)^2}{\sigma_1^2 \sigma_2^2}$$

Where $d^2 \sim \chi^2_{(2)}$ Chi-Square with 2 degrees of freedom (by definition of Chi-Square!)

- Chi-Square Control Chart

We can plot for each run the d^2 statistic, and set control limits based on OVERALL α-risk in the χ^2_α distribution:



For UNCORRELATED set of data, we can also extend to p dimensions: the distance measure for any single point at $(x_1, x_2 \ldots x_p)$ is

$$d^2 = \begin{bmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \\ \vdots \\ x_p - \mu_p \end{bmatrix}^T \begin{bmatrix} \sigma_1^2 & 0 \cdots & 0 \\ 0 & \sigma_2^2 & \\ \vdots & & \ddots \\ 0 & 0 & \sigma_p^2 \end{bmatrix}^{-1} \begin{bmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \\ \vdots \\ x_p - \mu_p \end{bmatrix}$$

$$= \begin{bmatrix} x_1 - \mu_1 & \ldots & x_p - \mu_p \end{bmatrix} \begin{bmatrix} 1/\sigma_1^2 & & & 0 \\ & \frac{1}{\sigma_2^2} & & \\ & & \ddots & \\ 0 & & & \frac{1}{\sigma_p^2} \end{bmatrix} \begin{bmatrix} x_1 - \mu_1 \\ \vdots \\ x_p - \mu_p \end{bmatrix}$$

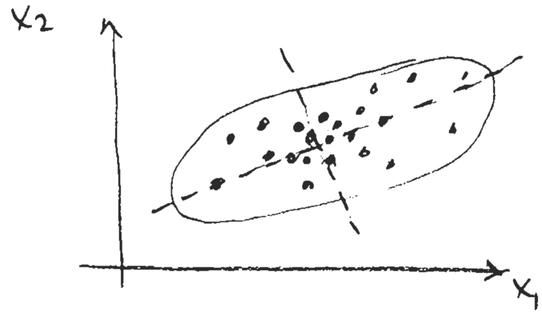
$$= \begin{bmatrix} \frac{x_1 - \mu_1}{\sigma_1^2} & & & 0 \\ & \frac{x_2 - \mu_2}{\sigma_2^2} & & \\ & & \ddots & \\ 0 & & & \frac{x_p - \mu_p}{\sigma_p^2} \end{bmatrix} \begin{bmatrix} x_1 - \mu_1 \\ \vdots \\ x_p - \mu_p \end{bmatrix}$$

$$= \frac{(x_1 - \mu_1)^2}{\sigma_1^2} + \ldots \frac{(x_p - \mu_p)^2}{\sigma_p^2}$$

CHI-SQUARE CONTROL CHARTS – CORRELATED VARIABLES

- If we are monitoring quality characteristics on the same product, or on the same process, it is quite likely that these characteristics are indeed CORRELATED.

How handle such cases?



Statistically dependent
⇒ knowing value of x_1 changes the probability of some x_2 value

(e.g. a "high" x_1 means I'm also more likely to have a "high" x_2).

∴ Our independent case "distance" metric will <u>overstate</u> the true probability distance; we should discount for correlation:

$$d^2 = \frac{1}{\sigma_1^2\sigma_2^2 - \rho_{12}^2}\left[\sigma_2^2(x_1-\mu_1)^2 + \sigma_1^2(x_2-\mu_2)^2 - 2\rho_{12}(x_1-\mu_1)(x_2-\mu_2)\right]$$

Or, in matrix notation for p output variables:

$$d^2 = (\vec{x}-\vec{\mu})^T \vec{\Sigma}^{-1} (\vec{x}-\vec{\mu})$$

where $\vec{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix}$ mean vector

$$\vec{\Sigma} = \text{covariance matrix} = \begin{bmatrix} \sigma_1^2 & \rho_{12} & \rho_{13} & \cdots \\ \rho_{12} & \sigma_2^2 & & \\ \rho_{13} & \rho_{23} & \sigma_3^2 & \\ \vdots & & & \ddots & \\ & & & & \sigma_p^2 \end{bmatrix}$$

$d^2 \sim \chi^2(p)$ Chi-square with p d.o.f.

Variance & Covariance - Linear Algebraic Picture

① Sample variance

$$\vec{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_r \end{bmatrix} \quad r \text{ samples}$$

$$\vec{m} = \vec{x} - \vec{\bar{x}} = \begin{bmatrix} x_1 - \bar{x} \\ \vdots \\ x_r - \bar{x} \end{bmatrix} \quad \text{mean centered}$$

$$s^2 = \frac{m^T m}{r-1} = \begin{bmatrix} x_1 - \bar{x} \\ \vdots \\ x_r - \bar{x} \end{bmatrix}^T \begin{bmatrix} x_1 - \bar{x} \\ \vdots \\ x_r - \bar{x} \end{bmatrix} \frac{1}{r-1} = \frac{\|\vec{m}\|^2}{r-1} \quad \text{"length" of vector}$$

$$= \begin{bmatrix} x_1 - \bar{x} & \dots & x_r - \bar{x} \end{bmatrix} \begin{bmatrix} x_1 - \bar{x} \\ \vdots \\ x_r - \bar{x} \end{bmatrix} \frac{1}{r-1}.$$

$$\hat{\sigma}_x^2 = \hat{\sigma}_{xx}^2 = s^2 = \frac{1}{r-1} \sum_{i=1}^{r} (x_i - \bar{x})^2 \quad \text{as before}$$

② Covariance between two variables ...

$$\hat{\sigma}_{xy}^2 = S_{xy}^2 = \frac{1}{r-1} (\vec{x} - \vec{\bar{x}})^T (\vec{y} - \vec{\bar{y}})$$

$$= \frac{1}{r-1} \begin{bmatrix} x_1 - \bar{x} & \dots & x_r - \bar{x} \end{bmatrix} \begin{bmatrix} y_1 - \bar{y} \\ \vdots \\ y_r - \bar{y} \end{bmatrix}$$

$$\hat{\sigma}_{x_1, x_2}^2 = S_{12}^2 = \frac{1}{r-1} \left[\vec{x}_1 - \vec{\bar{x}}_1 \right]^T \left[\vec{x}_2 - \vec{\bar{x}}_2 \right] \quad \text{where } \vec{x}_2 = \begin{bmatrix} x_{21} \\ x_{22} \\ \vdots \\ x_{2r} \end{bmatrix} \downarrow r \text{ rows}$$

2^{nd} variable

(3) Extend to n variables

$$\left[\vec{x}_1 - \bar{\vec{x}}_1 \quad \vec{x}_2 - \bar{\vec{x}}_2 \quad \cdots \quad \vec{x}_n - \bar{\vec{x}}_n \right] = \vec{M} = \left[\vec{X} - \bar{\vec{X}} \right]$$

So

$$C = \underset{\text{covariance}}{\overset{\text{sample}}{}} = \frac{1}{r} M^T M$$

T^2 CONTROL CHART - Estimated means & covariance

- In direct analog to the univariate case, we often do not know the process mean and covariance, and must estimate them:

$$\bar{\vec{x}} = \hat{\vec{\mu}} \qquad \text{That is,} \quad \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \vdots \\ \bar{x}_p \end{bmatrix} = \begin{bmatrix} \hat{\mu}_1 \\ \vdots \\ \hat{\mu}_p \end{bmatrix}$$

$$\vec{\vec{S}} = \hat{\vec{\vec{\Sigma}}}$$

That is, the sample covariance estimates the process covariance.

Then a statistic, the HOTELLING T^2, is

$$T^2 = (\vec{x} - \bar{\vec{x}})^T \vec{\vec{S}}^{-1} (\vec{x} - \bar{\vec{x}}) \sim F_{p,\, m-p}$$

gives us the same "weighted" generalized distance measure, but now using the estimated covariance.

- In general, we may use samples with n individuals, and base our estimates $\vec{\vec{S}}$ on m previous or preliminary runs:

$$T^2 = n\,(\bar{x} - \bar{\bar{x}})^T S^{-1} (\bar{x} - \bar{\bar{x}})$$

(dropped vector, use $\bar{\bar{x}} = \hat{\mu}$)

and set $UCL = \dfrac{p(m+1)(n-1)}{mn - m - p + 1} F_{\alpha,\, p,\, mn-m-p+1}$

- Example: two variable uncorrelated case

$$T^2 = \frac{(\bar{x}_1 - \bar{\bar{x}})^2}{s_1^2} + \frac{(\bar{x}_2 - \bar{\bar{x}})^2}{s_2^2} \sim F_{2,\, m-2}$$

- When m is large ($m > 20-50$), usually take $\vec{\vec{S}}$ estimate to be very good, and simply use

$$UCL = \chi^2_{\alpha,\, p}$$

Estimating Covariance

Samples of size n, $i = 1, 2 \ldots n$
Output variables p, $j = 1, 2, \ldots p$
Preliminary runs m, $k = 1, 2, \ldots m$

$$\bar{x}_{jk} = \frac{1}{n} \sum_{i=1}^{n} x_{ijk}$$ … mean for each sample

$$S^2_{jk} = \frac{1}{n-1} \sum_{i=1}^{n} (x_{ijk} - \bar{x}_{jk})^2$$ … variance for each sample

as usual

Covariance between characteristic j and h in k^{th} run:

$$S_{jh,k} = \frac{1}{n-1} \sum_{i=1}^{n} (x_{ijk} - \bar{x}_{jk})(x_{ihk} - \bar{x}_{hk}) \quad , \quad k = 1, 2, \ldots m, \quad j \neq h$$

Now we average over the m runs:

$$\bar{\bar{x}}_j = \frac{1}{m} \sum_{k=1}^{m} \bar{x}_{jk} \qquad j = 1, 2, \ldots, p$$

$$\bar{S}_j^2 = \frac{1}{m} \sum_{k=1}^{m} S^2_{jk} \qquad j = 1, 2, \ldots, p$$ – variance parts

and

$$\bar{S}_{jh} = \frac{1}{m} \sum_{k=1}^{m} S_{jhk} \qquad j \neq h$$

So

$$\vec{S} = \begin{bmatrix} \bar{S}_1 & \bar{S}_{12} & \bar{S}_{13} & \cdots & \bar{S}_{1p} \\ \vdots & \bar{S}_2 & & & \\ & & \ddots & & \\ \bar{S}_{p1} & & & & \bar{S}_p \end{bmatrix}$$

- EXAMPLE: tensile strength & diameter

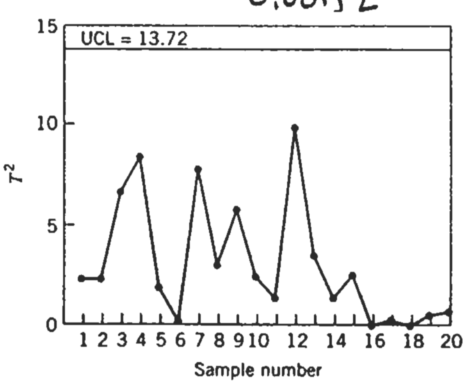
  $n = 10$ sample size
  $m = 20$ preliminary samples

Sample Number k	(a) Sample Means: Tensile Strength ($\bar{x}_{1k}$)	Diameter ($\bar{x}_{2k}$)	(b) Variances and Covariances: S_{1k}^2	S_{2k}^2	S_{12k}	(c) Control Chart Statistics: T_k^2	$\lvert S_k \rvert$
1	115.25	1.04	1.25	0.87	0.80	2.16	0.45
2	115.91	1.06	1.26	0.85	0.81	2.14	0.41
3	115.05	1.09	1.30	0.90	0.82	6.77	0.50
4	116.21	1.05	1.02	0.85	0.81	8.29	0.21
5	115.90	1.07	1.16	0.73	0.80	1.89	0.21
6	115.55	1.06	1.01	0.80	0.76	0.03	0.23
7	114.98	1.05	1.25	0.78	0.75	7.54	0.41
8	115.25	1.10	1.40	0.83	0.80	3.01	0.52
9	116.15	1.09	1.19	0.87	0.83	5.92	0.35
10	115.92	1.05	1.17	0.86	0.95	2.41	0.10
11	115.75	0.99	1.45	0.79	0.78	1.13	0.54
12	114.90	1.06	1.24	0.82	0.81	9.96	0.36
13	116.01	1.05	1.26	0.55	0.72	3.86	0.17
14	115.83	1.07	1.17	0.76	0.75	1.11	0.33
15	115.29	1.11	1.23	0.89	0.82	2.56	0.42
16	115.63	1.04	1.24	0.91	0.83	0.70	0.19
17	115.47	1.03	1.20	0.95	0.70	0.19	0.65
18	115.58	1.05	1.18	0.83	0.79	0.00	0.36
19	115.72	1.06	1.31	0.89	0.76	0.35	0.59
20	115.40	1.04	1.29	0.85	0.68	0.62	0.63
Averages	$\bar{\bar{x}}_1 = 115.59$	$\bar{\bar{x}}_2 = 1.06$	$\bar{S}_1^2 = 1.23$	$\bar{S}_2^2 = 0.83$	$\bar{S}_{12} = 0.79$		

For control, will calculate

$$T^2 = \frac{10}{(1.23)(0.83) - (0.79)^2}\Big[0.83(\bar{x}_1 - 115.89)^2 + 1.23(\bar{x}_2 - 1.06)^2 - 2(0.79)(\bar{x}_1 - 115.59)(\bar{x}_2 - 1.06)\Big]$$

Use $UCL \simeq \chi^2_{0.001,2} = 13.816$



No evidence of poor control during preliminary runs $\Rightarrow$ accept chart

- INTERPRETING ALARMS

  - A disadvantage of multivariate charts is determing what is responsible for signal.

Approaches: (1) standard practice: plot univariate $\bar{x}_i$

$\rightsquigarrow$ may not succeed!

(2) Decompose T^2 into components reflecting contribution of ith variable:

$$d_i = T^2 - T^2_{(i)}$$

where $T^2_{(i)}$ = statistic for all process variable <u>except</u> ith

$\Rightarrow$ look for d_i large...

PRINCIPAL COMPONENT ANALYSIS

- Raw Data L, $r \times n$, $r = \#$ experimental runs
  $n = \#$ measurements or variables per run

  L_{ij} = i^{th} run, j^{th} data variable

- Mean Centering: Average across # runs (that is, average column)

  $$M_{ij} = L_{ij} - \bar{L}_j$$ so $\bar{L}_j$ is sample mean of the j^{th} column across r runs

- Sample Covariance

  $$C = \frac{1}{r-1} M^T M$$, C is $n \times n$ capturing relationships between variables

- Eigenvector - Right Eigenvectors

  $$CV = V\Lambda$$ where V is $n \times n$ eigenvector matrix
  Λ is $n \times n$ diagonal

  -Can efficiently compute using SVD

  $$\Lambda = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & & \vdots \\ 0 & 0 & \cdots & \lambda_2 \end{bmatrix}$$

- Projection onto <u>some</u> principal direction (eigenvector):

  $$M v_j = t_j$$ where t_j are the "score" or loadings on the principal directions

  or, as a whole

  $$MV = T$$

  Note that $\Lambda = \frac{1}{r-1} T^T T$; that is to say, the T vectors are orthogonal and capture all the covariance in the original data.

  Since $V^T V = I$

- Dimensionality Reduction

A key use of PCA is to select only those components that matter – a reduced # of effective dimensions in the data.

– Variation captured in each P.C. (eigenvector):

order $\Lambda = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix}$ in decreasing order

$$\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \dots \lambda_n$$

% variance in eigenvector v_j $= \dfrac{\lambda_j}{\sum_j \lambda_j} \cdot 100 = \dfrac{\lambda_j}{\mathrm{Trace}(C)} \cdot 100$

So, if want say 97% or 99% of variation to be captured in the model, pick top λ_j, v_j that capture as desired:

e.g. $\hat{T} = [t_1 \; t_2 \; t_3 \; t_4]$

$$M \hat{V} = \hat{T}$$

just project onto desired "top" directions