

Statistical Process Control

MSc: Statistics and Actuarial-Financial Mathematics

Konstantinos Bourazas
kbourazas@aegean.gr

Department of Statistics and Actuarial-Financial Mathematics

Course #9

April 19, 2026

Why we need multivariate control charts

- Product quality almost always depends on **several characteristics** measured simultaneously
- Shifts that are **invisible to univariate charts** can become detectable in the multivariate setting
 - ▶ Each component may shift by a tiny amount that stays well within its individual control limits, yet the **joint shift** is statistically significant
 - ▶ The **correlation structure** between variables can change even when the marginal means and variances stay the same
- Running p separate univariate charts is tempting, but the design is complicated ($2p$ parameters to set when using CUSUMs) and the resulting scheme is generally **less powerful** than a genuinely multivariate procedure
- Multivariate charts treat each observation as a **vector** and exploit correlations, giving better detection power

The multivariate normal distribution and the Mahalanobis distance

- A p -dimensional vector $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ has pdf

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{p/2} |\boldsymbol{\Sigma}|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})\right]$$

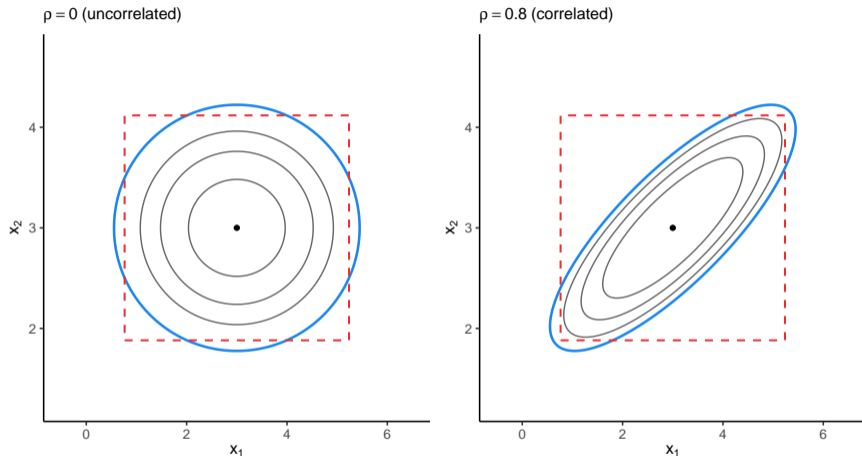
- The quadratic form inside the exponent is the **squared Mahalanobis distance**¹ from \mathbf{x} to $\boldsymbol{\mu}$

$$d_M^2(\mathbf{X}, \boldsymbol{\mu}) = (\mathbf{X} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{X} - \boldsymbol{\mu}) \sim \chi_p^2$$

- The Mahalanobis distance accounts for correlations and different scales among the p variables. It is the foundation of all multivariate Shewhart charts
- Under normality, \mathbf{X} has a $1 - \alpha$ chance to fall inside the ellipsoid $\{\mathbf{x} : d_M^2(\mathbf{x}, \boldsymbol{\mu}) \leq \chi_{1-\alpha, p}^2\}$

¹Mahalanobis, P. C. (1936). On the generalised distance in statistics. *Proceedings of the National Institute of Sciences of India*, 2, 49–55.

Bivariate normal contours and the probability ellipse



Left: uncorrelated ($\rho = 0$), contour ellipses aligned with axes. Right: correlated ($\rho = 0.8$), ellipses tilted. The solid blue curve is the 95% probability ellipse under the bivariate normal model. The dashed red rectangle is based on separate marginal intervals and ignores correlation.

Estimation and Hotelling's T^2 statistic

- Given a sample X_1, \dots, X_n from $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, the unbiased estimators are

$$\hat{\boldsymbol{\mu}} = \bar{X}, \quad S = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})(X_i - \bar{X})'$$

- Key sampling properties: $\bar{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}/n)$, $(n-1)S \sim W_{n-1, \boldsymbol{\Sigma}}$, and \bar{X} and S are independent
- Hotelling² (1931) generalised the t -test: $T^2 = n(\bar{X} - \boldsymbol{\mu})'S^{-1}(\bar{X} - \boldsymbol{\mu})$ is the estimated squared Mahalanobis distance of the sample mean from the target, scaled by n

²Hotelling, H. (1931). The generalization of Student's ratio. *Ann. Math. Statist.*, 2, 360–378.

The Hotelling's T^2 control chart

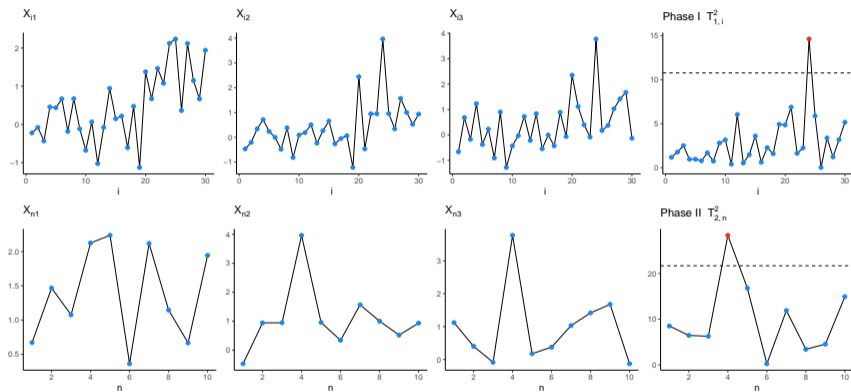
- **Known parameters.** At time i , compute the squared Mahalanobis distance $T_{0,i}^2 = (\mathbf{X}_i - \boldsymbol{\mu}_0)' \boldsymbol{\Sigma}_0^{-1} (\mathbf{X}_i - \boldsymbol{\mu}_0)$. Under IC, $T_{0,i}^2 \sim \chi_p^2$, so signal when $T_{0,i}^2 > \chi_{1-\alpha,p}^2$
- **Phase I, estimated parameters.** Use $T_{1,i}^2 = (\mathbf{X}_i - \bar{\mathbf{X}})' \mathbf{S}^{-1} (\mathbf{X}_i - \bar{\mathbf{X}})$. Tracy et al.³ (1992) showed $\frac{M}{(M-1)^2} T_{1,i}^2 \sim \text{Beta}(\frac{p}{2}, \frac{M-p-1}{2})$
- **Phase II, estimated parameters.** The charting statistic $T_{2,n}^2 = (\mathbf{X}_n - \hat{\boldsymbol{\mu}}_0)' \hat{\boldsymbol{\Sigma}}_0^{-1} (\mathbf{X}_n - \hat{\boldsymbol{\mu}}_0)$ has a scaled F distribution: $\frac{(M-p)M}{p(M-1)(M+1)} T_{2,n}^2 \sim F_{p,M-p}$
- In all cases the chart signals when the Mahalanobis distance of the current observation from the target is too large

³Tracy, N. D., Young, J. C., and Mason, R. L. (1992). Multivariate control charts for individual observations. *J. Qual. Technol.*, 24, 88–95.

Example 7.1 Hotelling's T^2 control for a 3-dimensional process

- 30 observations from a 3-dimensional process (Table 7.1 in Qiu), $\alpha = 0.005$ ($ARL_0 = 200$)
- **Phase I:** parameters estimated from all 30 observations. Beta-based control limit = 10.773. Signal at the **24th** observation
- **Phase II:** first 20 observations used as IC data, remaining 10 monitored. F -based control limit = 21.671. Signal at the **4th** Phase II observation
- The Phase II chart detects the shift much sooner because it compares against a fixed IC baseline, while the Phase I chart includes the shifted observations in its own parameter estimates

Example 7.1 Phase I and Phase II Hotelling T^2 control charts



Multivariate CUSUM via regression adjustment (Hawkins, 1991)⁴

- We assume $\mathbf{X} \sim N_p(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$, under the IC state. The conditional distribution of X_j given the remaining components is again normal

$$X_j | \mathbf{X}_{-j} \sim N\left(\mu_{0j} + \sum_{\ell \neq j} \beta_{\ell j} (X_\ell - \mu_{0\ell}), \tau_{jj}\right)$$

which yields the regression representation $X_j - \mu_{0j} = \sum_{\ell \neq j} \beta_{\ell j} (X_\ell - \mu_{0\ell}) + \varepsilon_j$, $\varepsilon_j \sim N(0, \tau_{jj})$, with $\beta_{\ell j} = -\sigma_0^{\ell j} / \sigma_0^{jj}$ and $\tau_{jj} = 1 / \sigma_0^{jj}$ from the elements of $\boldsymbol{\Sigma}_0^{-1}$

- The **standardised residual** $Z_j = \varepsilon_j / \sqrt{\tau_{jj}}$ removes the linear effect of the other components. In matrix form

$$\mathbf{Z} = [\text{diag}(\boldsymbol{\Sigma}_0^{-1})]^{-1/2} \boldsymbol{\Sigma}_0^{-1} (\mathbf{X} - \boldsymbol{\mu}_0)$$

Under IC, each $Z_j \sim N(0, 1)$

- Apply a standard two-sided univariate CUSUM to each Z_j

$$C_{n,j}^+ = \max(0, C_{n-1,j}^+ + Z_{nj} - k), \quad C_{n,j}^- = \min(0, C_{n-1,j}^- + Z_{nj} + k)$$

and signal when $C_n = \max_j [\max(C_{n,j}^+, -C_{n,j}^-)] > h$

⁴Hawkins, D. M. (1991). Multivariate quality control based on regression-adjusted variables. *Technometrics*, 33, 61–75.

Single-statistic multivariate CUSUMs (Crosier, 1988)⁵

- **Crosier's MCUSUM.** Define

$$Y_n = \left[(U_{n-1} + X_n - \boldsymbol{\mu}_0)' \boldsymbol{\Sigma}_0^{-1} (U_{n-1} + X_n - \boldsymbol{\mu}_0) \right]^{1/2}$$

and update

$$U_n = \begin{cases} 0, & Y_n \leq k \\ (U_{n-1} + X_n - \boldsymbol{\mu}_0)(1 - k/Y_n), & Y_n > k \end{cases}$$

Signal when $C_n = (U_n' \boldsymbol{\Sigma}_0^{-1} U_n)^{1/2} > h$

- **CUSUM of T (COT).** First reduce each observation to its Mahalanobis distance

$$T_n = \left[(X_n - \boldsymbol{\mu}_0)' \boldsymbol{\Sigma}_0^{-1} (X_n - \boldsymbol{\mu}_0) \right]^{1/2},$$

then apply a univariate CUSUM:

$$\tilde{C}_n = \max(0, \tilde{C}_{n-1} + T_n - k)$$

- **MCUSUM keeps directional information; COT keeps only the distance from target**

⁵Crosier, R. B. (1988). Multivariate generalizations of cumulative sum quality-control schemes. *Technometrics*, 30, 291–303.

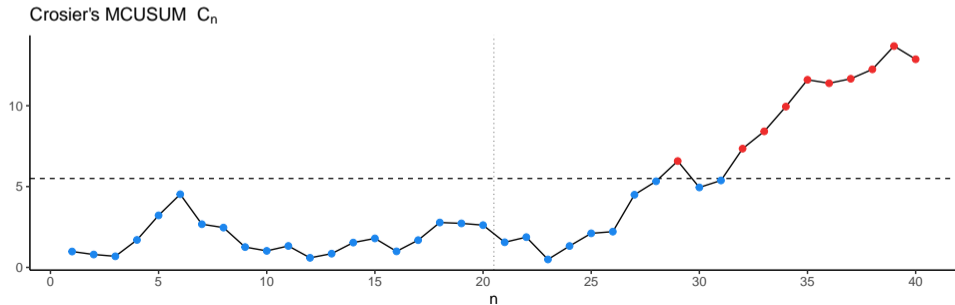
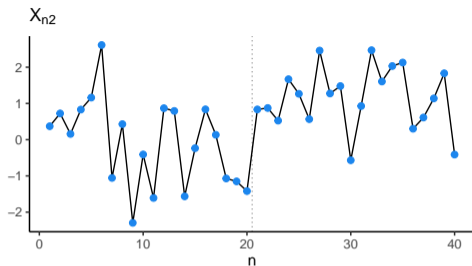
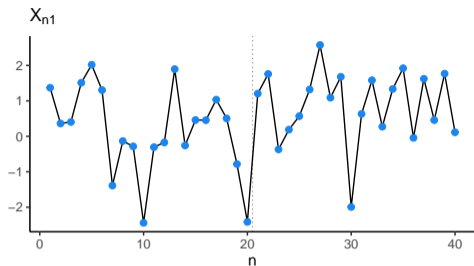
Example Crosier's MCUSUM in action

- Simulated bivariate process ($p = 2$) with $\boldsymbol{\mu}_0 = 0$ and IC covariance

$$\boldsymbol{\Sigma}_0 = \begin{pmatrix} 1.0 & 0.6 \\ 0.6 & 1.0 \end{pmatrix}$$

- 40 observations are generated and a mean shift from $\boldsymbol{\mu}_0$ to $\boldsymbol{\mu}_1 = (1, 1)'$ starts at the 21st observation. The Mahalanobis length of the shift is $[\boldsymbol{\delta}'\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\delta}]^{1/2} \approx 1.12$
- Crosier's MCUSUM with $k = 0.5$ and $h = 5.5$ ($ARL_0 \approx 200$ for $p = 2$)
- The chart signals at the $n = 29$ th observation, well after the shift has accumulated enough evidence through the vector recursion U_n
- The raw series X_1 and X_2 both drift upward after $n = 21$, but the multivariate statistic C_n captures the joint departure from $\boldsymbol{\mu}_0$ more efficiently than either component alone

Example Bivariate data and Crosier's MCUSUM chart



MCUSUM for monitoring the covariance matrix (Healy, 1987)⁷

- For a proportional covariance shift $\Sigma_1 = c\Sigma_0$, Healy applied a sequential likelihood-ratio argument: the log-LR for $N_p(\mu_0, c\Sigma_0)$ vs $N_p(\mu_0, \Sigma_0)$ is linear in T_n^2 , giving the optimal CUSUM

$$C_n = \max[0, C_{n-1} + T_n^2 - k], \quad k = \frac{pc \log c}{c - 1}$$

Signal when $C_n > h$

- The only difference from the COT chart is T_n^2 vs. $\sqrt{T_n^2}$. Both detect mean and covariance shifts, but COT tends to be better for mean shifts, while Healy's chart is optimal for the proportional covariance shift it was designed for
- For individual component variance shifts, Hawkins'⁶ regression-adjusted approach can be adapted using $W_{nj} = (|Z_{nj}|^{1/2} - 0.822)/0.349$

⁶Hawkins, D. M. (1993). Regression adjustment for variables in multivariate quality control. *J. Qual. Technol.*, 25, 170–182.

⁷Healy, J. D. (1987). A note on multivariate CUSUM procedures. *Technometrics*, 29, 409–412.

The multivariate EWMA chart (Lowry et al., 1992)⁸

- The MEWMA extends the univariate EWMA recursion to vectors, with the same smoothing parameter $\lambda \in (0, 1]$ applied componentwise

$$E_n = \lambda(X_n - \mu_0) + (1 - \lambda)E_{n-1}, \quad E_0 = 0$$

- Signal a mean shift when the Mahalanobis distance of E_n exceeds the limit

$$V_n^2 = E_n' \Sigma_{E_n}^{-1} E_n > h$$

where $\Sigma_{E_n} = \frac{\lambda}{2-\lambda} [1 - (1 - \lambda)^{2n}] \Sigma_0$ comes from the weighted sum of independent $N_p(0, \Sigma_0)$ increments

- The IC distribution of V_n^2 is χ_p^2 , but because consecutive V_n^2 are correlated, h must be found by **Monte Carlo simulation**
- The OOC performance depends on the shift only through its **Mahalanobis length** $[\delta' \Sigma_0^{-1} \delta]^{1/2}$, regardless of direction

⁸Lowry, C. A., Woodall, W. H., Champ, C. W., and Rigdon, S. E. (1992). A multivariate exponentially weighted moving average control chart. *Technometrics*, 34, 46–53.

MEWMA for monitoring the covariance matrix (Hawkins and Maboudou-Tchao, 2008)⁹

- Transform via Cholesky decomposition: $Y_n = L(X_n - \mu_0)$ where $\Sigma_0 = UU'$, $L = U^{-1}$. Under IC, $Y_n \sim N_p(0, I)$, so $E[Y_n Y_n'] = I$
- Replace the vector EWMA on Y_n with a **matrix EWMA** on the outer product $Y_n Y_n'$, which smooths the sample covariance itself. The charting statistic is the log-likelihood ratio for $H_0 : \Sigma = I$ vs $H_1 : \Sigma = E_n$:

$$E_n = \lambda Y_n Y_n' + (1 - \lambda)E_{n-1}, \quad E_0 = I \quad \implies \quad C_n = \text{Tr}(E_n) - \log |E_n| - p > h$$

- Under IC, E_n stays close to I and C_n near zero. Writing $C_n = \sum_i (\lambda_i - \log \lambda_i - 1)$ in terms of the eigenvalues of E_n shows any departure from $\lambda_i = 1$ inflates C_n

⁹Hawkins, D. M. and Maboudou-Tchao, E. M. (2008). Multivariate exponentially weighted moving covariance matrix. *Technometrics*, 50, 155–166.

Multivariate change-point detection (Zamba and Hawkins, 2006)¹⁰

- CPD charts need **less prior information** about the OOC distribution and provide an **estimate of the change point** when a signal occurs
- At each new observation n , perform a retrospective scan: for every candidate split $r \in \{1, \dots, n-1\}$, compute the two-sample Hotelling T^2 comparing the first r observations to the last $n-r$, and take the maximum

$$T_{\max,n}^2 = \max_{1 \leq r \leq n-1} T_r^2, \quad T_r^2 = \frac{r(n-r)}{n} (\bar{X}_{0,r} - \bar{X}_{r,n})' (\tilde{S}_r^2)^{-1} (\bar{X}_{0,r} - \bar{X}_{r,n})$$

where \tilde{S}_r^2 is the **pooled covariance estimator** built from both segments

- Signal when $T_{\max,n}^2 > h_n$, and estimate $\hat{\tau} = \hat{r} + 1$. The time-varying control limit is

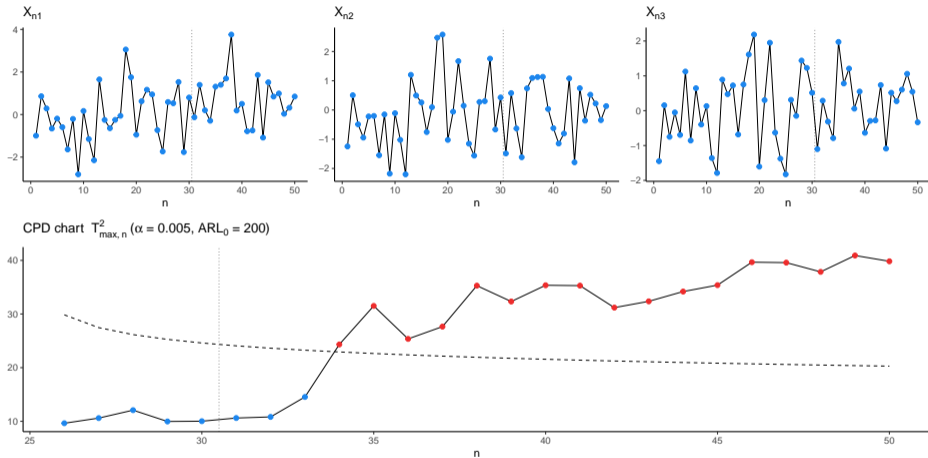
$$\log(h_n) \approx 2.706 + 0.230p - \frac{p+3}{50} \log(n-25) \quad (\alpha = 0.005)$$

¹⁰Zamba, K. D. and Hawkins, D. M. (2006). A multivariate change-point model for statistical process control. *Technometrics*, 48, 539–549.

Example 7.8 CPD chart in action

- 50 observations from a 3-dimensional process; IC parameters **unknown**. First 25 IC; monitoring begins at $n = 26$
- A mean shift occurs at the 31st observation
- With $\alpha = 0.005$ ($ARL_0 = 200$), the CPD chart signals at $n = 36$
- The estimated change point $\hat{r} = 30$ gives $\hat{\tau} = 31$, which is exactly correct
- The IC and OOC mean vectors can then be estimated from the two segments of the data

Example 7.8 CPD chart for a 3-dimensional process



Top row: the three components of X_n . Bottom: the CPD charting statistic $T_{\max, n}^2$ with the time-varying control limit h_n (dashed curve).