Multivariate Control Charts

Stat 3570

28 Feb, 2013
Multivariate Control Charts

- In all control charts procedures we discussed so far aims at monitoring one quality characteristic at a time.
- But in many situations, we are interested to control more than two quality characteristics at the same time. One idea is to monitor separate control charts. But in many cases, both quality characteristics may be correlated.
- Another option is to use multivariate techniques to construct a single chart to monitor all quality characteristics together.
- Hotelling’s $T^2$ control chart is the most commonly used multivariate control charts, when all the quality characteristics are normally distributed.
Hotelling’s $T^2$ Control Chart

- Let $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_p$ are the $p$ quality characteristics we are interested to monitor and we assume that all are normally distributed with multivariate mean, $\mu$ and covariance matrix $\Sigma$.

- We collect samples of size $n$ for each subgroup and repeat it for $m$ subgroups.

- For the $i$th subgroup, we have
  
  $$(x_{i11}, x_{i12}, \ldots, x_{ip}), (x_{i21}, x_{i22}, \ldots, x_{i2p}), \ldots$$
  
  $$(x_{in1}, x_{in2}, \ldots, x_{inp})$$

  For the $i$th subgroup,

  $$\bar{\mathbf{x}}_i = (\bar{x}_{i1}, \bar{x}_{i2}, \ldots, \bar{x}_{ip})$$

  sample covariance matrix is $S_i$ where the diagonal elements are

  $$s_{jj}^2 = \frac{1}{n-1} \sum_{l=1}^{n} (x_{lj} - \bar{x}_j)^2$$

  and the sample covariances are

  $$s_{jk} = \frac{1}{n-1} \sum_{l=1}^{n} (x_{lj} - \bar{x}_j)(x_{lk} - \bar{x}_k)$$

  Note: we may need on more subscript notation, but ignored to avoid to simplify the notation.
Hotelling’s $T^2$ Control Charts

- We can show that sample mean and sample covariance matrix are unbiased estimators of $\mu$ and $\Sigma$.
- We estimate $\mu$ and $\Sigma$ by averaging over the all $m$ subgroups, as

$$\bar{x} = \frac{1}{m} \sum_{i=1}^{m} \bar{x}_j$$

$$S = \frac{1}{m} \sum_{i=1}^{m} S_j$$
Hotelling’s $T^2$ Control Charts

- Hotelling’s $T^2$ statistics for $i$th subgroup is defined as
  
  $$T^2_i = n(\bar{x}_i - \bar{x})'S^{-1}(\bar{x}_i - \bar{x})$$

- Phase I control limits for $T^2$ charts are
  
  $$UCL = \frac{p(m-1)(n-1)}{mn - m - p + 1} F_{\alpha, p, mn-m-p+1}; \quad LCL = 0$$

- Phase II control limits for $T^2$ charts are
  
  $$UCL = \frac{p(m+1)(n-1)}{mn - m - p + 1} F_{\alpha, p, mn-m-p+1}; \quad LCL = 0$$
Interpretation of Control Charts

- It is not easy to find which of the p variable is responsible for an out of control signal. The standard practice is to plot individual $\bar{x}$ charts on individual variables, which may not be successful due to correlation among variables.

- Alt (1985) suggest to plot $\bar{x}$ chart with Bonferroni correction.

- Another very useful approach to diagnosis of an out of control signal is to decompose the $T^2$ into components that reflect the contribution of each individual variables. If $T^2$ is the current value of the statistic, and $T^2_{(i)}$ is the value of the statistic for all process variable except the ith one, then Runger et al (1996) show that $d_i = T^2 - T^2_{(i)}$, $i = 1, \ldots, p$ is an indicator of the relative contribution of the ith variable to the overall statistic.

- When an out of control signal is generated, compute $d_i$ for all the variable and focus the attention of the variable which has high $d_i$. 

Control Charts for Individual Observations

- In practice, subgroup size of one is mostly preferred due to time and cost. So control chart based on individual observation is of great interest.

- In this case, for each multivariate data point $x_i)$, we compute

$$T^2(j) = (x_i - \bar{x})S^{-1}(x_i - \bar{x})$$

- Tracey et al. (1992) constructed the control charts based on individual observations and the limits for Phase I charts are

$$UCL = \frac{(m - 1)^2}{m} \beta_{\alpha, p/2, (m - p - 1)/2}; \quad LCL = 0$$

- Control limits for Phase II charts are

$$UCL = \frac{p(m + 1)(m - 1)}{m^2 - mp} F_{\alpha, p, m - p}; \quad LCL = 0$$
Sample mean and sample covariance are sensitive to outliers, so robust Estimation methods preferred.

Robustness is often measured in terms of Breakdown Point.

Breakdown Point: The breakdown point of an estimator is the proportion of incorrect observations (i.e. arbitrarily large observations) an estimator can handle before giving an arbitrarily large result. i.e.. The smallest proportion of observations which can render an estimator meaningless.

For example, mean has breakdown point of $1/n$ and median has $(n-1)/2n$

Estimates having large breakdown point is preferred.
Robust Multivariate Control Charts for Mean

- $T^2$ chart with sample covariance estimated using successive differences - Sullivan and Woodall (1996)
- $T^2$ chart using MVE and MCD estimators of mean and covariance - Vergas (2003), Jensen et al. (2007)
  - Minimum Volume Ellipsoid - Estimates of mean and covariance based on the smallest ellipsoid containing a subset (at least half) of the observations
    Asymptotic breakdown point - $(n-p+1)/2n$
  - Minimum Covariance Determinant - Estimates of mean and covariance based on the subset of observations having covariance matrix with lowest determinant.
    Asymptotic breakdown point - $1/2$
Robust Multivariate Control Charts

- Using MVE and MCD estimates, modified $T^2$ charts are proposed. Control limits arrived empirically
- Studies showed that Robust Control Charts works well in detecting outliers
- Performance assessed based on Phase I outlier detection
- For monitoring phase II, use ordinary $T^2$ charts with estimates from phase I after removing outliers.
- Studies by Jensen et al. (2007) indicated that performance of MCD and MVE charts depends on the dimensionality, sample size and proportion of outliers of phase I samples.
- Re-weighted MCD and MVE estimators are more efficient than MCD and MVE estimates.
Re-Weighted MCD

\[ \bar{X}_{RMCD} = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i} \]

\[ S_{RMCD} = c_{\alpha, p} \frac{\sum_{i=1}^{n} w_i (x_i - \bar{X}_{RMCD})(x_i - \bar{X}_{RMCD})'}{\sum_{i=1}^{n} w_i} \]

Weight, \( w \) is defined as
\[ w_i = 1, \text{ if } D(X_i) \leq q_\alpha, 0 \text{ otherwise} \]

\[ D(X_i) = \sqrt{(x_i - \bar{X}_{MCD})' S_{MCD}^{-1} (x_i - \bar{X}_{MCD})} \]

\[ c_{\alpha, p} = \frac{\alpha}{P(\chi^2_{(p+2)} \leq q_\alpha)} \]

\[ d_{\gamma, \alpha}^{n, p} \text{ a finite sample correction} \]

\( q_\alpha, \alpha\)-th quantile of the chi-square distribution with \( p \) degrees of freedom
Re-Weighted MVE

\[ \bar{X}_{RMVE} = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i} \]

\[ S_{RMVE} = c_{\alpha, p} \left( \sum_{i=1}^{n} \frac{w_i (x_i - \bar{X}_{RMVE})(x_i - \bar{X}_{RMVE})'}{\sum_{i=1}^{n} w_i} \right) \]

Weight, \( w \) is defined as
\[ w_i = 1, \text{if } D(X_i) \leq q_{\alpha}, 0 \text{ otherwise} \]

\[ D(X_i) = \sqrt{(x_i - \bar{X}_{MVE})'S_{MVE}^{-1}(x_i - \bar{X}_{MVE})} \]

\[ c_{\alpha, p} = \frac{\alpha}{P(\chi^2_{(p+2)} \leq q_{\alpha})} \]

\[ d_{\gamma, \alpha}^{n, p} \text{ a finite sample correction} \]

\( q_{\alpha}, \alpha \text{-th quantile of the chi-square distribution with } p \text{ degrees of freedom} \)
Re-weighted MCD & Re-MVE provide efficient estimates of mean and covariance.

$T^2$ based on RMVE and RMCD are

$$T^2_{RMCD}(j) = (x_i - \bar{x}_{RMCD})S^{-1}_{RMCD}(x_i - \bar{x}_{RMCD})$$

$$T^2_{RMVE}(j) = (x_i - \bar{x}_{RMVE})S^{-1}_{RMVE}(x_i - \bar{x}_{RMVE})$$

$T^2$ control chart based on Re-MCD and Re-MVE estimates as has better performance in Phase I as well as Phase II

Software (implemented in R - library rrcov)