Statistical Process Control

MSc: Statistics and Actuarial-Financial Mathematics

Konstantinos Bourazas
kbourazas@aegean.gr

Department of Statistics and Actuarial-Financial Mathematics

Course #5

March 19, 2026

From non-conforming items to defect counts

- The p and mp charts classify each item as either conforming or non-conforming, a **binary** decision
- In many applications a product can contain several **defects** and still be acceptable
  - A new car may have minor scratches but still meet all functional requirements
  - A roll of fabric may have a few blemishes but remain sellable
- Manufacturers want to monitor the **occurrence of defects over time** even when individual items are not rejected
- The natural probability model for the number of defects in a fixed inspection unit is the **Poisson distribution**
- To compare defect counts across inspection units, the **size of the inspection unit** must be consistent

The Poisson distribution as a limit of the Binomial

- Recall that if $X \sim \text{Binomial}(n, p)$ with $n \to \infty$ and $p \to 0$ in such a way that $\lambda = np$ remains fixed, then X converges in distribution to $\text{Poisson}(\lambda)$
- This makes the Poisson model natural whenever we count **rare events in a large population**
  - The "population size" n may be unknown or even infinite, such as the number of opportunities for a scratch to appear on a car surface
  - Even when n is known but very large and p very small, working with $\text{Poisson}(\lambda)$ is simpler than with $\text{Binomial}(n, p)$
- Typical examples include the number of defects per unit, customer arrivals per hour, traffic accidents per day, or disease cases per region in a given period
- The Poisson distribution is also connected to the **Homogeneous Poisson Process**, where events occur independently at a constant rate λ per unit time
  - The number of events in any interval of fixed length follows a Poisson distribution
  - The waiting time between two consecutive events follows an $\text{Exponential}(\lambda)$ distribution with mean $1/\lambda$

The c chart: monitoring defect counts

- Let c be the number of defects found in one inspection unit
- If the process is IC, we model $c \sim \text{Poisson}(\lambda)$, which gives

$$\mu_c = \lambda, \qquad \sigma_c^2 = \lambda$$

- A key property of the Poisson distribution is that its **mean equals its variance**
- For the c chart, n inspection units of the **same size** are selected at n time points, yielding defect counts $c_1, c_2, \ldots, c_n$
- We estimate λ by $\bar{c} = \frac{1}{n}\sum_{i=1}^{n} c_i$ and use the normal approximation to the Poisson, giving the control limits

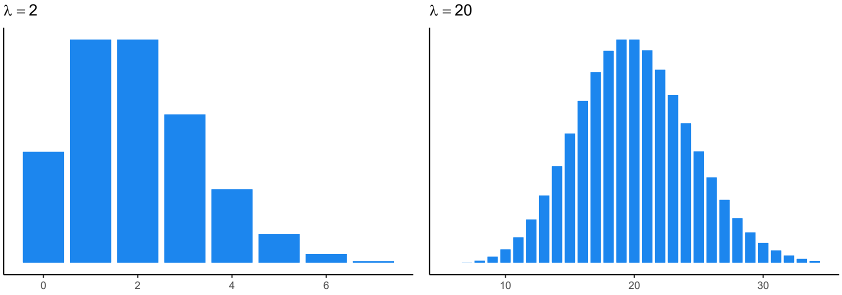
$$\text{UCL} = \bar{c} + Z_{1-\alpha/2}\sqrt{\bar{c}}, \qquad \text{CL} = \bar{c}, \qquad \text{LCL} = \bar{c} - Z_{1-\alpha/2}\sqrt{\bar{c}}$$

When is the Poisson normal approximation reliable

- The Poisson(λ) distribution approaches $N(\lambda, \lambda)$ as λ increases
- A common guideline is that the approximation is reasonable when $\lambda \geq 10$
- For small λ the Poisson distribution is heavily **right-skewed**, and the normal-based limits may produce an actual Type I error rate that differs from the nominal one
- In such cases we can use the exact Poisson quantiles as control limits
  - Set the LCL and UCL as the $(\alpha/2)$-th and $(1 - \alpha/2)$-th quantiles of the Poisson($\bar{c}$) distribution
- A negative LCL is always replaced by 0 since defect counts are non-negative

Poisson probabilities for small and large λ

The panels below show exact **Poisson probabilities** for $\lambda = 2$ and $\lambda = 20$. As λ increases the distribution becomes more symmetric and closer to normal



Example 3.6 Surface defects on steel plates

- Surface defects have been counted on $n = 15$ rectangular steel plates
- The observed defect counts are

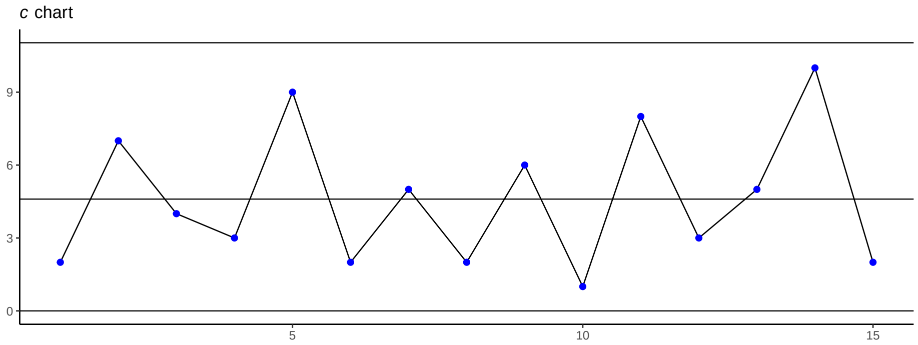
$$2,\ 7,\ 4,\ 3,\ 9,\ 2,\ 5,\ 2,\ 6,\ 1,\ 8,\ 3,\ 5,\ 10,\ 2$$

- We compute $\bar{c} = 4.6$. With $Z_{1-\alpha/2} = 3$ the limits are

$$\begin{aligned} \text{UCL} &= 4.6 + 3\sqrt{4.6} = 11.034 \\ \text{CL} &= 4.6 \\ \text{LCL} &= 4.6 - 3\sqrt{4.6} = -1.834 \ \longrightarrow \ 0 \end{aligned}$$

- All 15 counts lie within the limits, so the process producing the steel plates appears IC
- Note that $\bar{c} = 4.6 < 10$, so the normal approximation should be used with some caution here

Example 3.6 The c chart



The u chart: inspection units of different sizes

- The c chart assumes all inspection units have the **same size**
- In practice, different inspection units may differ in size
  - Different shipments may contain different numbers of products
  - Different panels of fabric may have different areas
- Raw defect counts are then **not comparable** across time points, because a larger unit naturally contains more defects
- The solution is to work with the **defect rate per size unit**

$$u_i = \frac{c_i}{m_i}, \qquad i = 1, 2, \ldots, n$$

where m_i is the size of the i-th inspection unit and c_i is its defect count

Control limits for the u chart

- Assume the number of defects per size unit follows Poisson$(\tilde{\lambda})$
- Then $c_i \sim$ Poisson$(m_i\tilde{\lambda})$, which gives $\mu_{u_i} = \tilde{\lambda}$ and $\sigma^2_{u_i} = \tilde{\lambda}/m_i$
- We estimate $\tilde{\lambda}$ by

$$\bar{u} = \frac{\sum_{i=1}^{n} c_i}{\sum_{i=1}^{n} m_i}$$

- The control limits for the u chart are

$$\text{UCL} = \bar{u} + Z_{1-\alpha/2}\sqrt{\frac{\bar{u}}{m_i}}, \qquad \text{CL} = \bar{u}, \qquad \text{LCL} = \bar{u} - Z_{1-\alpha/2}\sqrt{\frac{\bar{u}}{m_i}}$$

- Because the limits depend on m_i, they **vary from one time point to another** whenever the inspection unit sizes differ
- The same caveats about the normal approximation apply here

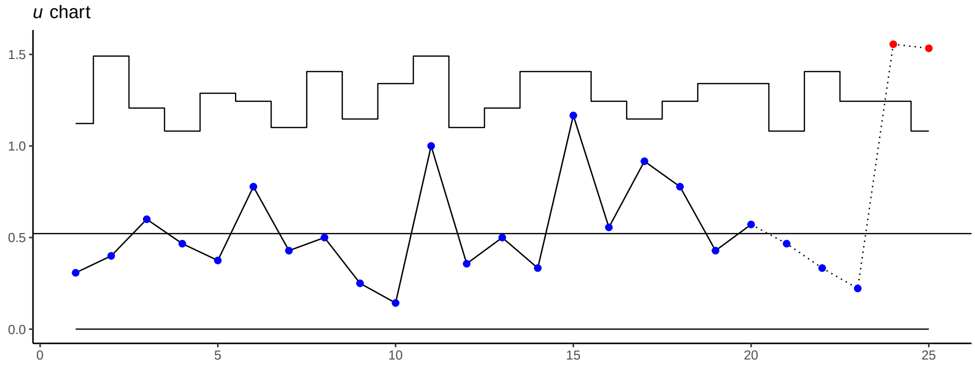
Simulated example for the u chart

- We generate $n = 20$ Phase I inspection units with sizes m_i drawn uniformly from $\{5, 6, \ldots, 15\}$
- Defect counts are generated from Poisson$(m_i\tilde{\lambda})$ with $\tilde{\lambda} = 0.5$ defects per size unit
- We then generate 5 Phase II inspection units
  - The first three from the same IC rate $\tilde{\lambda} = 0.5$
  - The last two from $\tilde{\lambda} = 1.5$, simulating a **tripling of the defect rate**
- We build the u chart using the Phase I data and $Z_{1-\alpha/2} = 3$

$$\text{UCL}_i = \bar{u} + 3\sqrt{\bar{u} \,/\, m_i}, \qquad \text{CL} = \bar{u}, \qquad \text{LCL}_i = \bar{u} - 3\sqrt{\bar{u} \,/\, m_i}$$

- Because the inspection unit sizes vary, the control limits **move from one time point to another**

Simulated example *u* chart



The D chart: weighting different types of defects

- In many products, not all defects are equally serious
  - A defect in the engine of a car is far more critical than a minor scratch on the surface
- Suppose defects are classified into k groups with counts $c_1^*, c_2^*, \ldots, c_k^*$, each assumed Poisson$(\lambda_j^*)$ and independent
- Assign pre-specified weights $w_1, w_2, \ldots, w_k \geq 0$ reflecting their relative importance and define

$$D = \sum_{j=1}^{k} w_j \, c_j^*$$

- Then

$$\mu_D = \sum_{j=1}^{k} w_j \lambda_j^*, \qquad \sigma_D^2 = \sum_{j=1}^{k} w_j^2 \, \lambda_j^*$$

- Note that the variance involves w_j^2 while the mean involves w_j, because the variance of $w_j c_j^*$ is $w_j^2 \operatorname{Var}(c_j^*)$

Control limits for the D chart

- From n inspection units of the same size we observe the grouped defect counts c_{ij}^* for $i = 1, \ldots, n$ and $j = 1, \ldots, k$
- The weighted defect count for unit i is $D_i = \sum_{j=1}^k w_j \, c_{ij}^*$
- We estimate λ_j^* by $\bar{c}_j^* = \frac{1}{n} \sum_{i=1}^n c_{ij}^*$, and then

$$\bar{D} = \sum_{j=1}^k w_j \, \bar{c}_j^*, \qquad \widehat{\sigma_D^2} = \sum_{j=1}^k w_j^2 \, \bar{c}_j^*$$

- The control limits of the D chart are

$$\text{UCL} = \bar{D} + Z_{1-\alpha/2} \sqrt{\sum_{j=1}^k w_j^2 \, \bar{c}_j^*}, \quad \text{CL} = \bar{D}, \quad \text{LCL} = \bar{D} - Z_{1-\alpha/2} \sqrt{\sum_{j=1}^k w_j^2 \, \bar{c}_j^*}$$

- These limits are appropriate when $\{\lambda_j^*\}$ are all reasonably large and the inspection units have the same size

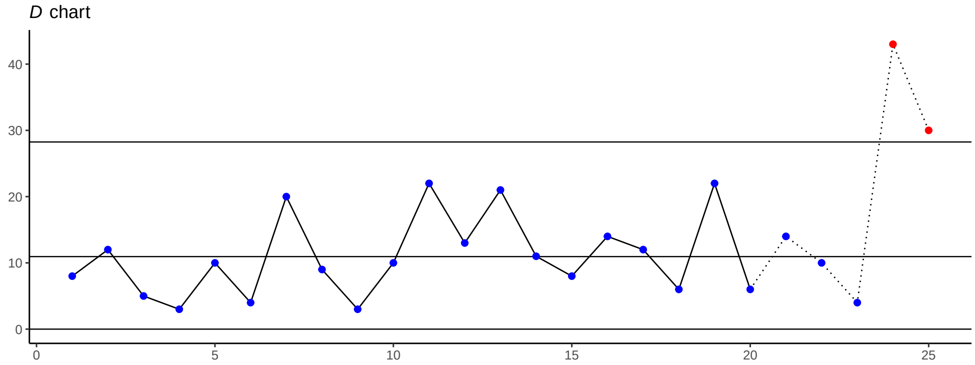
Simulated example for the D chart

- A product has $k = 3$ types of defects with weights $w_1 = 1$ (cosmetic), $w_2 = 3$ (functional), $w_3 = 5$ (safety)
- We generate $n = 20$ Phase I inspection units with IC rates $\lambda_1^* = 4,\ \lambda_2^* = 2,\ \lambda_3^* = 0.5$
- We then generate 5 Phase II inspection units
  - The first three from the same IC rates
  - The last two with the safety defect rate increased tenfold to $\lambda_3^* = 5$, while the other rates remain unchanged
- We build the D chart using the Phase I data and $Z_{1-\alpha/2} = 3$

$$\text{UCL} = \bar{D} + 3\sqrt{\sum_{j=1}^{3} w_j^2\ \bar{c}_j^*}, \qquad \text{CL} = \bar{D}, \qquad \text{LCL} = \bar{D} - 3\sqrt{\sum_{j=1}^{3} w_j^2\ \bar{c}_j^*}$$

- Even a modest increase in the **most heavily weighted** defect type can push D_i above the UCL, because the weight amplifies the shift

Simulated example *D* chart



Summary of charts for categorical variables

Chart	What we monitor	Model	Equal sizes?
p	proportion non-conforming p_i	Binomial	Yes
mp	count non-conforming X_i	Binomial	Yes
c	defect count c_i	Poisson	Yes
u	defect rate $u_i = c_i/m_i$	Poisson	No
D	weighted defect count D_i	Poisson	Yes

- All charts share the **same Shewhart structure**: CL $\pm\, Z_{1-\alpha/2} \times$ estimated SD
- All rely on a **normal approximation** whose quality depends on the sample size and the parameter values
- When the approximation is poor, exact distributional limits or alternative approximations should be used

A note on discreteness and the ARL

- Unlike continuous charting statistics, the statistics used in attribute charts take **discrete values**
- This means the chart can only achieve a **limited set of ARL_0 values** for a given set of parameters
- For example, with a specific m and $\bar{p}$, there may be no choice of control limits that yields exactly $\text{ARL}_0 = 370$
- When designing an attribute chart, we should pick a control limit that achieves an ARL_0 as close as possible to the desired level, and report the actual ARL_0 value alongside the chart
- This phenomenon is common to all charts based on discrete distributions and will appear again when we discuss CUSUM charts for attribute data