Statistical Process Control

MSc: Statistics and Actuarial-Financial Mathematics

Konstantinos Bourazas
kbourazas@aegean.gr

Department of Statistics and Actuarial-Financial Mathematics

Course #4

March 12, 2026

Shewhart charts for categorical variables

- So far we have monitored **numerical quality characteristics** such as length, weight, or strength
- In many applications the quality characteristic is **categorical**
  - Each inspected item is classified as **conforming** or **non-conforming**
  - Or we count the number of **defects** found in an inspection unit
- The natural probability models change accordingly
  - Proportions of non-conforming items follow a **Binomial** model
  - Counts of defects follow a **Poisson** model
- We will study five charts designed for these situations
  - The p **chart** and mp **chart** for proportions and counts of non-conforming items
  - The c **chart**, u **chart**, and D **chart** for defect counts (in the next course)

The p chart: monitoring the fraction non-conforming

- At each time point we draw a random sample of m products and classify each one as conforming or non-conforming
- Let X be the number of non-conforming items in the sample and let π be the true non-conforming proportion when the process is IC
- Then $X \sim \text{Binomial}(m, \pi)$, which gives

$$\mu_X = m\pi, \qquad \sigma_X^2 = m\pi(1 - \pi)$$

- The sample proportion $p = X/m$ has

$$\mu_p = \pi, \qquad \sigma_p^2 = \frac{\pi(1 - \pi)}{m}$$

- When m is **large enough**, the CLT tells us that p is approximately normal, and we can build control limits exactly as in the $\bar{X}$ chart

Control limits when π is known

- If π were known, we would declare the process OOC at time i whenever p_i falls outside

$$\text{UCL} = \pi + Z_{1-\alpha/2}\sqrt{\frac{\pi(1-\pi)}{m}}, \qquad \text{LCL} = \pi - Z_{1-\alpha/2}\sqrt{\frac{\pi(1-\pi)}{m}}$$

- The structure is identical to the known-parameter $\bar{X}$ chart, with π replacing the process mean and $\sqrt{\pi(1-\pi)/m}$ replacing the standard error
- In practice π is unknown and must be estimated from Phase I data, just as μ_0 and σ were estimated from Phase I subgroups in the variables charts

The p chart with estimated limits

- Collect n samples of size m at n time points and compute the sample proportions $p_i = X_i/m$ for $i = 1, \ldots, n$
- Estimate π by the overall average

$$\bar{p} = \frac{1}{n}\sum_{i=1}^{n} p_i$$

- The estimated control limits of the p chart are

$$\text{UCL} = \bar{p} + Z_{1-\alpha/2}\sqrt{\frac{\bar{p}(1-\bar{p})}{m}}, \qquad \text{CL} = \bar{p}, \qquad \text{LCL} = \bar{p} - Z_{1-\alpha/2}\sqrt{\frac{\bar{p}(1-\bar{p})}{m}}$$

- As usual we take $Z_{1-\alpha/2} = 3$ by convention
- A negative LCL is set to 0, since proportions are non-negative

When is the Binomial normal approximation reliable

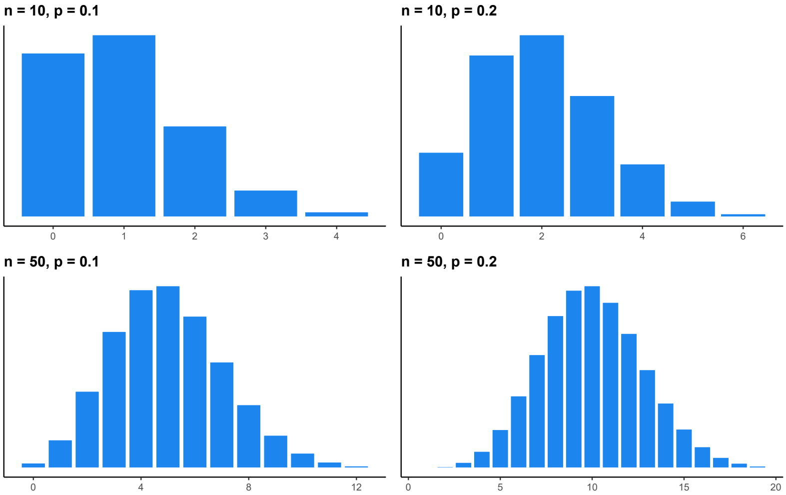
- The Binomial(m, π) distribution approaches $N\big(m\pi,\ m\pi(1-\pi)\big)$ as m increases, provided that π is not too close to 0 or 1
- For the p chart, this implies that the sample proportion $p = X/m$ is approximately $N\big(\pi,\ \pi(1-\pi)/m\big)$
- A common rule of thumb is

$$m\bar{p} \geq 5 \qquad \text{and} \qquad m(1-\bar{p}) \geq 5$$

- If the process has a very small (or very large) non-conforming rate, these conditions may fail even for moderate m
- When the conditions are not met, the actual Type I error probability $\tilde{\alpha}$ can differ substantially from the nominal α, which means the chart may give far more (or fewer) false alarms than expected

Binomial probabilities for different values of n and p

The panels below show exact **Binomial probabilities** for different values of n and p. Larger n and less extreme p values make the distribution more symmetric



Example 3.4 Non-conforming products in 20 samples

- We have $n = 20$ samples, each containing $m = 50$ products
- The numbers of non-conforming products in the 20 samples are

i	1	2	3	4	5	6	7	8	9	10
X_i	7	3	10	1	8	5	4	9	3	9
p_i	0.14	0.06	0.20	0.02	0.16	0.10	0.08	0.18	0.06	0.18

i	11	12	13	14	15	16	17	18	19	20
X_i	5	7	2	10	4	6	9	3	11	5
p_i	0.10	0.14	0.04	0.20	0.08	0.12	0.18	0.06	0.22	0.10

- We compute $\bar{p} = 0.121$, so $m\bar{p} = 6.05 > 5$ and $m(1 - \bar{p}) = 43.95 > 5$; the normal approximation is adequate

Example 3.4 Computing the p chart limits

- With $\bar{p} = 0.121$, $m = 50$, and $Z_{1-\alpha/2} = 3$ we get

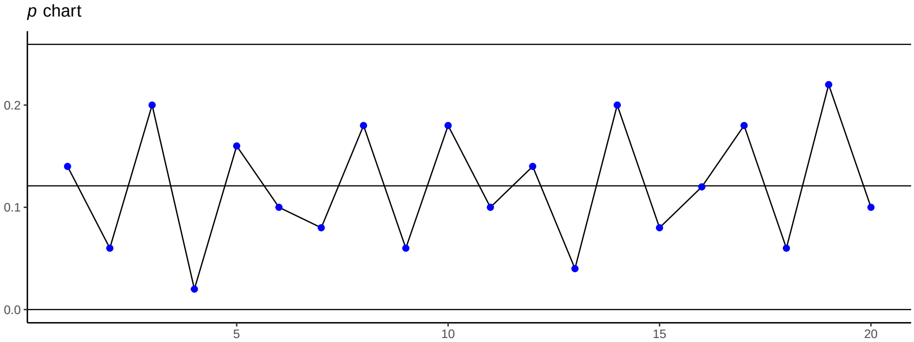
$$\text{UCL} = 0.121 + 3\sqrt{\frac{0.121 \times 0.879}{50}} = 0.259$$

$$\text{CL} = 0.121$$

$$\text{LCL} = 0.121 - 3\sqrt{\frac{0.121 \times 0.879}{50}} = -0.017 \longrightarrow 0$$

- Since proportions are non-negative, the LCL is set to 0
- All 20 sample proportions lie within the limits, so the process appears IC at all time points

Example 3.4 The *p* chart



Small samples and the actual Type I error

- When m is small or π is close to 0 or 1, the Binomial distribution is noticeably skewed
- In such cases, the actual false alarm probability $\tilde{\alpha}$ can be **much larger or smaller** than the nominal α
- For instance, if $\alpha = 0.01$ but $\tilde{\alpha} = 0.02$, the chart would stop the process for investigation **twice as often** as intended, wasting time and resources
- The discrepancy between $\tilde{\alpha}$ and α grows worse as m decreases or as π moves towards the boundaries

Two numerical examples

Example A $m = 10,\ \pi = 0.10,\ Z_{1-\alpha/2} = 3$

The normal-based UCL is $0.10 + 3\sqrt{0.10 \times 0.90\,/\,10} = 0.385$ and LCL $\to 0$. Since $p = X/m$, a signal occurs when $X/10 > 0.385$, that is, when $X > 3.85$. Because X is an integer, this means $X \geq 4$. Computing the exact Binomial tail

$$\tilde{\alpha} \ = \ P(X \geq 4 \mid X \sim \text{Bin}(10,\, 0.10)) \ = \ 1 - P(X \leq 3) \ \approx \ 0.013$$

The nominal α is 0.0027, so the **actual false alarm rate is almost 5 times larger** than intended

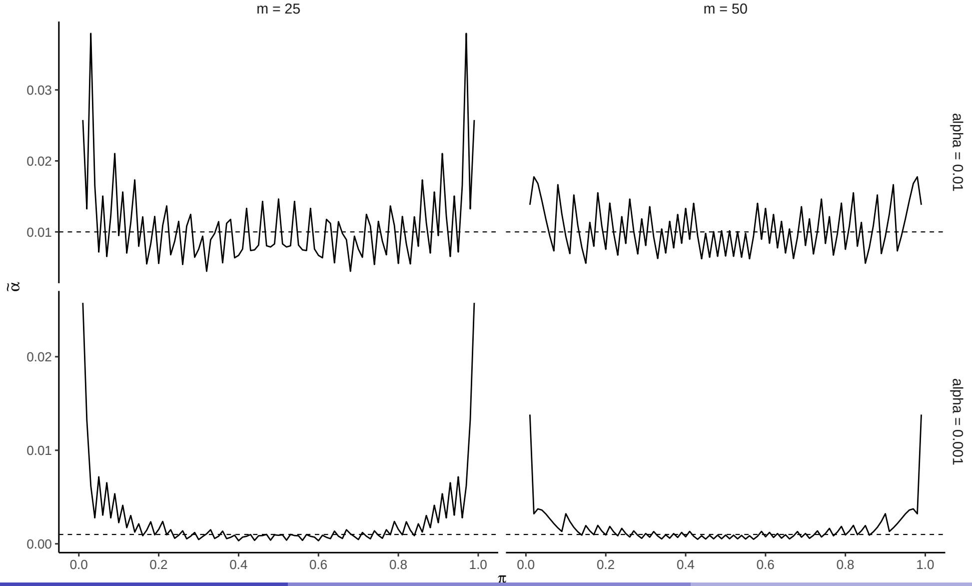
Example B $m = 15,\ \pi = 0.50,\ Z_{1-\alpha/2} = 3$

Now UCL $= 0.887$ and LCL $= 0.113$. Since $p = X/m$, a signal occurs when $X/15 > 0.887$ or $X/15 < 0.113$. Equivalently, $X > 13.305$ or $X < 1.695$. This translates to $X \geq 14$ or $X \leq 1$, thus

$$\tilde{\alpha} \ = \ 2\,P(X \leq 1) + \ P(X \geq 14) \ = \ \approx \ 0.001$$

Here $\tilde{\alpha}$ is roughly **one third of the nominal** $\alpha = 0.0027$ and the chart is too conservative and will detect shifts more slowly

Actual Type I error probability for the *p* chart



Exact control limits using the Binomial distribution

- When the normal approximation is unreliable we can use the **exact Binomial quantiles** directly
- If π is known and the nominal Type I error is α, define

$$L^* = \frac{\max\{a \,:\, P(X \leq a) \leq \alpha/2\}}{m}, \qquad U^* = \frac{\min\{a \,:\, P(X \geq a) \leq \alpha/2\}}{m}$$

  where $X \sim \text{Binomial}(m, \pi)$
- Because of the discreteness of the Binomial, the actual $\tilde{\alpha}$ may still not equal α exactly, but it can be computed in advance

$$\tilde{\alpha} = P(X < mL^*) + P(X > mU^*)$$

- When π is unknown we replace it with $\bar{p}$

The *mp* chart: monitoring counts instead of proportions

- Sometimes it is more natural to chart the **number** of non-conforming items X_i directly rather than the proportion p_i
- The two approaches are equivalent because $X_i = m\, p_i$
- The estimated control limits for the *mp* chart are

$$\text{UCL} = \bar{X} + Z_{1-\alpha/2}\sqrt{m\bar{p}(1-\bar{p})}, \qquad \text{CL} = \bar{X}, \qquad \text{LCL} = \bar{X} - Z_{1-\alpha/2}\sqrt{m\bar{p}(1-\bar{p})}$$

  where $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i = m\bar{p}$
- The large-sample conditions become

$$\bar{X} \geq 5 \qquad \text{and} \qquad m - \bar{X} \geq 5$$

- The same discussion about exact Binomial limits applies here through the relationship $X_i = m\, p_i$

Simulated example for the *mp* chart

- We generate $n = 25$ Phase I samples of size $m = 100$ from $\text{Binomial}(100, 0.08)$
- We then generate 5 Phase II samples
  - The first three from the same IC distribution $\text{Binomial}(100, 0.08)$
  - The last two from $\text{Binomial}(100, 0.18)$, simulating a **substantial increase in the non-conforming proportion**
- We build the *mp* chart using the Phase I data and the usual three-sigma limits with $\alpha = 0.0027$

$$\text{UCL} = \bar{X} + 30\sqrt{\bar{p}(1-\bar{p})}, \qquad \text{CL} = \bar{X}, \qquad \text{LCL} = \bar{X} - 30\sqrt{\bar{p}(1-\bar{p})}$$

- This example illustrates how the *mp* chart detects a **shift in the fraction non-conforming**, with the final two Phase II samples exceeding the UCL and producing an alarm

Simulated example *mp* chart