PRO 5971

Statistical Process Monitoring - monitoring count data

Linda Lee Ho

April 11, 2023

Department of Production Engineering
University of São Paulo

Outline

- Sometimes count data are available and we wish to monitor if its average rate λ is stable or not
- Procedure: at regular sampling interval, a random sample of size n is collected $X_1, X_2, \ldots, X_n$
- Usually its average is calculated and the decision (if the process is stable or not) based on this value
- Assumption: X follows a Poisson distribution with the probability density function

$$P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}$$

x=0, 1, 2, 3, ∞

u chart: Control Limits

- Monitored statistic: $\overline{X}$ is an unbiased estimator of λ as $E(\overline{X}) = \lambda$; and consistent since $Var(\overline{X}) = \dfrac{\lambda}{n} \to 0$ as n increases
- To determine the control limits, the distribution of $\overline{X}$ or $\sum_{i=1}^{n} X_i$ must be known.
- If the process is stable $X_i \sim Poisson(\lambda_0)$. According to the statistics theory, the exact distribution of $Y = \sum_{i=1}^{n} X_i \sim Poisson(n\lambda_0)$
- Fixed α, the integers a and b can be determined such that satisfy $\alpha/2 = P(Y > a) = P(Y < b)$ a and b are the quantiles of Y such that $P(Y < a) = 1 - \alpha/2$ and $P(Y < b) = \alpha/2$
- Then the probability control limits are: $UCL_u = \dfrac{a}{n}$ and $LCL_u = \dfrac{b}{n}$
- As Y assumes only integer values, the value α may not be reached

u chart: normal approximation

- By the central limit theorem, $\sum_{i=1}^{n} X_i$ asymptotically follow $\sim N(n\lambda, n\lambda)$
- Thus $\overline{X}$ asymptotically $\sim N(\lambda, \lambda/n)$
- Fixed α, $\alpha/2 = P(\overline{X} > UCL_u) = P(\overline{X} < LCL_u)$
- Central line: λ_0; asymptotic control limits: $UCL_u = \lambda_0 + z_{\alpha/2}\sqrt{\frac{\lambda_0}{n}}$; $LCL_u = \lambda_0 - z_{\alpha/2}\sqrt{\frac{\lambda_0}{n}}$
- When λ is unavailable:
  - take m samples each one with n units; replace λ_0 by $\overline{\overline{X}} = \sum_{i=1}^{m} \frac{\overline{X}_i}{m}$;
  - Central line: $\overline{\overline{X}}$; Trial Control limits: $\overline{\overline{X}} \pm z_{\alpha/2}\sqrt{\frac{\overline{\overline{X}}}{n}}$
  - In case of variable sample sizes :
    - Replace n by $\overline{n} = \sum_{i=1}^{m} \frac{n_i}{m}$, m= number of sample
    - Preferable: Use as monitored statistic $Z_i = \frac{\overline{X}_i - \overline{\overline{X}}}{\sqrt{\frac{\overline{\overline{X}}}{n_i}}}$; Control limits: $\pm z_{\alpha/2}$

Find the control limits of u chart considering $n = 4, 25, 100$; $\alpha = 0.1; 0.05; 0.01.0.0027$ and $\lambda_0 = 0.5; 10$ for two sided and one sided (increases). Use the exact and approximated distribution of $\overline{X}$

Find the power of this chart when the parameter λ_0 shifts to $\lambda_1 = \delta\lambda_0$, $\delta = 1.2; 1.5, 3$ using the exact and approximated distribution of $\overline{X}$

Make comparison, discuss the results.

c chart: Control Limits

- A single unit is taken
- Monitored statistic: X; $E(X) = \lambda$; $Var(X) = \lambda$
- When λ is available:
  - Central line: λ; Control Limits: $\lambda \pm z_{\alpha/2}\sqrt{\lambda}$
- When λ not available:
  - Take m preliminary samples; let X_i be the number of nonconformities at i-th sample, replace λ by $\overline{X}$
  - Central line: $\overline{X}$; Trial Control limits: $\overline{X} \pm z_{\alpha/2}\sqrt{\overline{X}}$

Exercise

- A process yields averagely 0.8 defects per unit when the process is in-control and 1.4 when the process is out-of-control. Which of the two control provide a higher power: a) a *np* control chart with $n = 100$ (an unit is classified as non-conforming if it shows more than 3 defects) or b) C chart with $n = 5$
- From historical data, averagely 25 defects are observed per 100 produced units. Design a control chart to monitor the number of defect per unit when the average increases to 2 defects per unit considering $\alpha = 0.1\%$ and $\beta = 0.30$
- In aeronautics industry, the average number misplaced rivets in a wing of an airplane is 2.13 rivets per 10 m^2. The manager decides to monitor taking sampling of 50 m^2. a) What kind control chart do you suggest? b) Provide its control limits by $\alpha = 0.05$

Dealing with Low Count rates

- c and u charts - ineffective to monitor process with low count rates
- Many samples with zero defects; charts plotting at zero - no informative
- Long periods of time between the occurrence of a nonconforming unit
- Relationship: Counts - **Poisson** Distribution $\rightarrow$ Time between events - **Exponential** distributions
- Alternative: a chart using the transformation of **the times between successive occurrences**
  - Proposed by Nelson(1994): $X = Y^{(1/3.6)} = Y^{0.2777}$; Y, the time between the events
  - X, approximately normal distribution
  - Build a chart for individual observations using X

Highlights:

- c and u control charts: assume that Poisson distribution is the correct probabilistic model
- Poisson distribution has a strong restriction: the mean and the variance are equal (equidispersion) not able to deal with under and overdispersion.
- Poisson is not the only distribution to model the counts
- Other count distributions have been used to design control charts for count data as Negative Binomial, Conway-Maxwell, Berg, Touchard distributions, among others

Transformations

X$\sim$ Poisson (λ), some transformations are used to build control charts:

$$Z_1 = \frac{X - \lambda}{\sqrt{\lambda}}$$

$$Z_2 = 2\left(\sqrt{X} - \sqrt{\lambda}\right)$$

Both transformations Z_1 and Z_2 are asymptotically normally distributed so their control limits are ± 3

Rossi et al. (1999) proposed

$$Z_3 = 0.5Z_1 + 0.5Z_2$$

as the monitored statistic also with control limits are ± 3

g and h Control Charts

- Geometric Distribution can also used to build control chart to monitor counts or events
- This chart was proposed by Kaminsky et al (1972) and they used the following probabilistic model

$$P(X = x) = p(1 - p)^{(x-a)} \text{ for x=a, a+1, \ldots}$$

- a is the known minimum possible number of events, with

$$E(X) = \frac{1-p}{p} + a$$

$$Var(X) = \frac{1-p}{p^2}$$

g and h charts

- For a sample of n values of X, say $x_1, \ldots x_n$, two statistics can be used to build the control chart

- g chart

$$\overline{x} = \frac{x_1 + \ldots + x_n}{n}$$

- Center line:

$$\frac{1-p}{p} + a$$

- Control Limits:

$$\left(\frac{1-p}{p} + a\right) \pm L\sqrt{\frac{1-p}{np^2}}$$

- h chart

$$t = \sum_{1}^{n} x_1 + \ldots + x_n$$

- Center line:

$$n\left(\frac{1-p}{p} + a\right)$$

- Control Limits:

$$n\left(\frac{1-p}{p} + a\right) \pm L\sqrt{\frac{n(1-p)}{p^2}}$$

g and h charts

- In many practical situation, the value of p is unknown. An estimator can be obtained as

$$\hat{p} = \frac{1}{\overline{\overline{x}} - a + 1}$$

- Consider m rational groups data available, thus

- g chart
- Center line:

$$\overline{\overline{x}} = \frac{\overline{t}}{n} = \frac{t_1 + \ldots + t_m}{nm}$$

- Control Limits:

$$\frac{\overline{t}}{n} \pm \frac{L}{\sqrt{n}} \sqrt{\left(\frac{\overline{t}}{n} - a\right)\left(\frac{\overline{t}}{n} - a + 1\right)}$$

- h chart
- Center line: $\overline{t}$
- Control Limits:

$$\overline{t} \pm L \sqrt{n \left(\frac{\overline{t}}{n} - a\right)\left(\frac{\overline{t}}{n} - a + 1\right)}$$

References

Rossi, G., Lampugnani, L. & Marchi, M. (1999), 'An approximate cusum procedure for surveillance of health events', *Statistics in medicine* **18**(16), 2111–2122.