PRO 5971 - Statistical Process Monitoring

Control chart to monitor a proportion p

Linda Lee Ho

April 11, 2023

Department of Production Engineering
University of São Paulo

Outline

- Consider that the outcomes of event may be summarized by two possible results: success or failure (or conform and non-conforming) (approved or disapproved), etc.
- Such event is known as Bernoulli trial which can be described by a probabilistic model: Bernoulli
- Let X, a Bernoulli random variable of parameter p, being p $\rightarrow$ the probability to occur some event of interest.
- X assumes 2 possible values: $X = 1$ if the event of interest occurs with probability p and $X = 0$ if the event does not occur with probability $1 - p$ with E(X)=p and Var(x)=p(1-p).
- A sample of size n is collected and observations annotated as $x_1, x_2, \ldots, x_n$.
- Based on these results, one has to decided if the process is stable at p_0 (in-control) or not (out-of-control).
- Problem: how to create decision rules based on statistic criterion?

Monitored statistics

- Definition of Statistic - any function based on the observed values.
- Statistics are random variables.
- Two statistics used to monitor the proportion p: $\widehat{p} = \frac{\sum_{i=1}^{n} X_i}{n}$ (sample proportion and $\sum_{i=1} X_i$ (the # of success in n)
- Reasons for such choices: $\widehat{p} = \frac{\sum_{i=1}^{n} X_i}{n}$ is an unbiased estimator of p.

About the distribution of the monitored statistics

- Ingredients to build any control chart: 1-The probability distribution or the approximated distribution of the monitored statistics should be known; 2 - Choose the type I error α.
- In our case: what is the distribution of $\widehat{p} = \frac{\sum_{i=1}^{n} X_i}{n}$ or $Y = \sum_{i=1} X_i$?
- According to the statistic theory, the exact distribution of Y is a binomial distribution with parameter np.
- So $P(Y = y) = \left(\begin{array}{c} n \\ y \end{array} \right) p^y (1-p)^{n-y}$ with $E(Y) = np$ and $Var(Y) = np(1-p)$.
- According to the central limit theorem, an approximated distribution of Y is $Y \sim N[np, np(1-p)]$

Control limits of the *np* and *p* control charts - using the exact distribution of Y

- Let α be the type I error then
  $\alpha/2 = P(Y < LCL_{np}|p = p_0) = P(Y > UCL_{np}|p = p_0)$
- Probability Control limits of *np* chart are the quantiles of a binomial distribution Y: UCL_{np}, the 1-$\alpha/2$th quantile and LCL_{np}, the $\alpha/2$th quantile control chart *np*.
- As Y is a discrete random variable, the desired level α may not be reached.
- The control limits for *p* control chart are $UCL_p = \dfrac{UCL_{np}}{n}$ and $LCL_p = \dfrac{LCL_{np}}{n}$,
- These are the control limits for $H_1 : p \neq p_0$. for other alternative hypothesis, some adjustments are needed.
- If $H_1 : p < p_0$ then $\alpha = P(Y < LCL_{np}|p = p_0)$ and if $H_1 : p > p_0$ then $\alpha = P(Y > UCL_{np}|p = p_0)$

Control limits of the np and p control charts - using the approximate distribution of Y

- As $Y \sim N(np, np(1-p)$ then the asymptotic control limits of np chart are $np_0 \pm z_{1-\alpha/2}\sqrt{np_0(1-p_0)}$
- Similarly $p_0 \pm z_{1-\alpha/2}\sqrt{p_0(1-p_0)/n}$ are the asymptotic control limits for p chart
- If $H_1 : p < p_0$ then $LCL_{np} = np_0 - z_{1-\alpha}\sqrt{np_0(1-p_0)}$ or $LCL_p = p_0 - z_{1-\alpha}\sqrt{p_0(1-p_0)/n}$
- if $H_1 : p > p_0$ then $UCL_{np} = np_0 + z_{1-\alpha}\sqrt{np_0(1-p_0)}$ or $UCL_p = p_0 + z_{1-\alpha}\sqrt{p_0(1-p_0)/n}$

The power

- Power $= 1\text{-}\beta$, probability to detect a shift of p_0 to p_1
- β - the type II error - is the probability of no signal of the control chart when the process is out-of-control
- For $H_1 : p \neq p_0$, by the exact distribution:
  $1-\beta = P(Y > UCL_{np} | p = p_1) + P(Y < LCL_{np} | p = p_1)$
- If $H_1 : p < p_0$ then Power$=1-\beta = P(Y < LCL_{np} | p = p_1)$ and if $H_1 : p > p_0$ then Power$=1-\beta = P(Y > UCL_{np} | p = p_1)$
- Using the normal approximation $1-\beta = P(Y > np_0 + z_{\alpha/2}\sqrt{np_0(1-p_0)} | p = p_1) + P(Y < np_0 - z_{\alpha/2}\sqrt{np_0(1-p_0)} | p = p_1)$
- If $H_1 : p < p_0$ then $1-\beta = P(Y < np_0 - z_{1-\alpha}\sqrt{np_0(1-p_0)} | p = p_1)$
- if $H_1 : p > p_0$ then $1-\beta = P(Y > np_0 + z_{1-\alpha}\sqrt{np_0(1-p_0)} | p = p_1)$

Comparing the performance - Discussion

- Example: Monitoring the quality of a service
- p=probability of customer be unsatisfied in a normal condition, p_0=0.01
- Consider 3 plans, all with LCL=0:

	n=100		n=200
p	UCL=3	UCL=4	UCL=6
0.01	52.63($\alpha = 0.019$)	250.00($\alpha = 0.004$)	200.00($\alpha = 0.005$)
0.02	6.99	18.87	9.01
0.03	2.83	5.41	2.54
0.05	1.36	1.79	1.15
0.10	1.01	1.03	1.00

Table 1: Comparing the performance

Which plan do you prefer?

Exercises

1. To solve the exercises, use the functions of software R as dbinom; pbinom; qbinom; rbinom if necessary
2. For $n = 10$ and $p_0 = 0.1$, obtain the control limits for $\alpha = 0.1; 0.5; 0.0027$ using the exact distribution of Y and the normal approximation considering $H_1 : p \neq p_0$
3. Repeat item 1 for n=50
4. Repeat items 1 and 2 considering $H_1 : p > p_0$
5. Discuss about the control limits and the type I error level between items 1 and 2.
6. Obtain the power of item 1, 2 and 3 if the p shifts to $p_1 = 0.2, 0.3$ and discuss the results

Exercise

- When the process is in-control, one item is non-conforming at every 80 produced ones. A p chart is used to monitor the non-conforming fraction. What is the probability of control chart signals an increase of 100% at the first sample of size equal to 200 items?
- Consider an np control chart. Which is the best option to detect an increase of non-conforming fraction: $p_0 = 0.015$ to $p_1 = 0.04$: a) take a sample of n=200 at every half hour? b) or a sample of n=400 at every hour? Compare these two options under an equal condition (similar occurrence of false alarms)
- Design an np control chart with LCL=0 to detect a shift from $p_0 = 0.015$ to $p_1 = 0.04$, with $\alpha < 0.003$ and $\beta < 0.50$
- A quality characteristic X has as mean 20 and standard deviation 2.5 when the process is in-control. One item is considered as non-conforming if it does not satisfy the specification limits 20 ± 7. Usually when the process is out-of-control, its mean shifts to 23.75 keeping standard deviation unaltered. Fixed $\alpha = 0.01$ Which option is the better and why: a) an np control with $n = 120$ or $\overline{X}$ chart with $n = 4$?

p chart: when p_0 is not available

- Use the control limits as **trial control limits**
  - Central line: $\overline{p}$; Control limits: $\overline{p} \pm z_{\alpha/2}\sqrt{\dfrac{\overline{p}(1-\overline{p})}{n}}$
  - $\overline{p} = \sum_{i=1}^{m} \dfrac{\hat{p}_i}{m}$, $\hat{p}_i = \dfrac{Y_i}{n}$, Y_i, # of non-conforming in n units taken from in-control process
  - In case of variable sample size, use $\overline{p} = \sum_{i=1}^{m} \dfrac{Y_i}{\sum_{i=1}^{m} n_i}$; $\overline{n} = \sum_{i=1}^{m} \frac{\sum_{i=1}^{m} n_i}{m}$ and control limits $\overline{p} \pm z_{\alpha/2}\sqrt{\dfrac{\overline{p}(1-\overline{p})}{\overline{n}}}$

np chart: when p_0 is not availaable

- Use as the control limits
  - Central line: $n\overline{p}$; Control limits: $n\overline{p} \pm z_{\alpha/2}\sqrt{n\overline{p}(1-\overline{p})}$
  - $\overline{p} = \sum_{i=1}^{m} \frac{\hat{p}_i}{m}$, $\hat{p}_i = \frac{Y_i}{n}$, Y_i, # of non-conforming in n units taken from in-control process

Exercises

Samples of 100 units are taken and the # of non-conforming in each is computed

Table 2: Data

sample	# of non-conforming	sample	# of non-conforming
1	7	11	6
2	4	12	15
3	1	13	2
4	3	14	9
5	6	15	5
6	8	16	1
7	10	17	4
8	5	18	5
9	2	19	7
10	7	20	12

Use data from Table 2 to provide estimates for the non-conforming fraction. Which should be the control limits for a p chart?

References