IEOR 165 – Lecture 21
Quality Control

The idea of statistical quality control is that we will periodically take measurements of the output of some service system or manufacturing process and then use these measurements to help identify when the process does not satisfy our engineering specifications. For example, we might be manufacturing screws and need the dimensions of the screws to fit within a tight tolerance. And so we would like a set of tools that will help us identify when the manufacturing process has encountered some change that causes the screws to deviate from the engineering specifications.

1 The $\overline{X}$ Control Chart

1.1 Known Variance

Suppose $X_1, X_2, \ldots$ are iid random variables with distribution $\mathcal{N}(\mu, \sigma^2)$, where μ, σ^2 are known values. One idea for quality control is to consider subgroups of n measurements and look for deviations of the sample average of these subgroups outside of suitably defined confidence bounds. In particular, define

$$\overline{X}^1 = \frac{1}{n}(X_1 + X_2 + \ldots + X_n)$$

$$\overline{X}^2 = \frac{1}{n}(X_{n+1} + X_{n+2} + \ldots + X_{2n})$$

$$\vdots$$

$$\overline{X}^{k+1} = \frac{1}{n}(X_{kn+1} + X_{kn+2} + \ldots + X_{(k+1)n})$$

Next, recall that the level $(1 - \alpha)$ confidence interval for *a single* sample average would be

$$\underline{\mu}^j = \overline{X}^j - \sigma \cdot z(1 - \alpha/2)/\sqrt{n}$$

$$\overline{\mu}^j = \overline{X}^j + \sigma \cdot z(1 - \alpha/2)/\sqrt{n}.$$

In the context of quality control, we will say that the process is out of control if $\mu \notin [\underline{\mu}^j, \overline{\mu}^j]$. Note that this is simply a hypothesis test (specifically a Z-test) with significance level α.

It is conventional to deal with an alternative representation of this hypothesis testing process. Define

$$LCL = \mu - \sigma \cdot z(1 - \alpha/2)/\sqrt{n}$$

$$UCL = \mu + \sigma \cdot z(1 - \alpha/2)/\sqrt{n}.$$

these values are called the *lower control limit* (LCL) and the *upper control limit* (UCL). With this alternative representation, we say that the process is out of control if $\overline{X}^k \notin [LCL, UCL]$.

1.2 Example

Q: Assume that items produced are supposed to be normally distributed with mean 35 and standard deviation 3. To monitor this process, subgroups of size 5 are sampled. If the following represents the averages of the first 20 subgroups, does it appear that the process was in control with significance level 0.05?

Subgroup No.	$\bar{X}$	Subgroup No.	$\bar{X}$
1	34.0	11	35.8
2	31.6	12	35.8
3	30.8	13	34.0
4	33.0	14	35.0
5	35.0	15	33.8
6	32.2	16	31.6
7	33.0	17	33.0
8	32.6	18	33.2
9	33.8	19	31.8
10	35.8	20	35.6

A: When in control, $X \sim N(35, 9)$. n = 5.

$$LCL = \mu - \sigma \cdot z_{(1-\alpha/2)}/\sqrt{n} = 35 - 3 \cdot 1.96/\sqrt{5} = 32.37$$

$$UCL = \mu + \sigma \cdot z_{(1-\alpha/2)}/\sqrt{n} = 35 + 3 \cdot 1.96/\sqrt{5} = 37.63$$

$\bar{X}^2 < LCL$, so the process is out of control starting with the 2nd subgroup.

1.3 Choosing Significance Level

The next question is how to pick the appropriate value of α. One convention is to pick $z(1 - \alpha/2) = 3$, since this corresponds to a level 0.0027 confidence interval. Another option is to pick the value of α such that we control the expected number of testing periods before concluding the process is out of control, under the assumption that the process never goes out of control. The reason that we will eventually conclude the process it out of control, even if the process were to never go out of control, is due to multiple testing. The choice of α only controls the probability of erroneously concluding the process is out of control for only one subgroup. However, we are repeatedly conducting tests and so we will suffer a familywise error rate.

Suppose we chosen a particular value α for the significance level. Assuming the process never goes out of control, the expected number of subgroups we analyze before concluding the process goes out of control can be abstracted using coin flips. In particular, consider a scenario where a coin is heads with probability $(1 - \alpha)$ and tails with probability α. What is the expected number of flips before we get a tails (including the last tails flip in the expected number)? To compute this value, we will define a difference equation and then solve this using standard approaches. Note that

$$\mathbb{E}(Flips) = \mathbb{E}[Flips|Flip_1 = Tail] \cdot \alpha + \mathbb{E}[Flips|Flip_1 = Head] \cdot (1 - \alpha)$$
$$= \alpha + (1 + \mathbb{E}(Flips)) \cdot (1 - \alpha).$$

Thus, we can conclude that $\mathbb{E}(Flips) = \frac{1}{\alpha}$. This equation gives us a way to choose the value of α given a desired stopping time.

For example, if we would like to have an expected number of analyzed subgroups before stopping to be 100, then we should choose $\alpha = 0.01$. The LCL and UCL can then be computed using a z-table to determine the value $z(1 - \alpha/2)$ and inserting this value into the equation for LCL and UCL.

2 The S-Control Chart

The idea of the $\overline{X}$-control chart is that we examined whether the sample mean deviated significantly from what was expected. However, in other cases we might be interested in detecting when the sample standard deviation deviates significantly from what was expected.

Suppose $X_1, X_2, \ldots$ are iid random variables with distribution $\mathcal{N}(\mu, \sigma^2)$, where σ^2 is known. Define

$$S_{k+1}^2 = \frac{1}{n-1} \sum_{i=kn+1}^{(k+1)n} (X_i - \overline{X}^{k+1})^2.$$

Then this value is distributed as $\sigma^2 \chi^2(n-1)/n-1$. Let $C_{n-1} \sim \chi^2(n-1)$ and define $c_{n-1}(1-\alpha)$ to be the value θ such that $\mathbb{P}(C_{n-1} \leq \theta) = 1 - \alpha$. We can compute the corresponding confidence intervals

$$\mathbb{P}\left(c_{n-1}(\alpha/2) \leq \frac{(n-1) \cdot S^2}{\sigma^2} \leq c_{n-1}(1 - \alpha/2) \right) = 1 - \alpha$$

$$\Rightarrow \underline{\sigma^2} = (n-1) \cdot S^2/c_{n-1}(1 - \alpha/2), \qquad \overline{\sigma^2} = (n-1) \cdot S^2/c_{n-1}(\alpha/2).$$

Consequently, we have that

$$LCL = (n-1) \cdot \sigma^2/c_{n-1}(1 - \alpha/2)$$
$$UCL = (n-1) \cdot \sigma^2/c_{n-1}(\alpha/2),$$

and we say the process is out of control if $S_{k+1}^2 \notin [LCL, UCL]$.

3 Nonparametric Control Charts

One of the challenges with control charts is that closed-form expressions may not be available for random variables with complicated distributions. Furthermore, in many cases, we may not even *a priori* know the distribution of the random variables. We can use a nonparametric approach to construct a control chart.

Suppose we have iid data X_i for $i = 1, \ldots, n_t$ from a process that is known to be in control. Let

$$\hat{\mu} = \frac{1}{n_t} \sum_{i=1}^{n_t} X_i$$

be the sample mean. Assuming n (i.e., the subgroup size is fairly large), we can use the following algorithm to construct nonparametric control charts.

1. For the j-th subgroup of data, use the percentile bootstrap to compute a confidence interval for the j-th subgroup of data. Let $[\underline{\mu}^j, \overline{\mu}^j]$ denote the range of the confidence interval.

2. If the sample mean of the process in control is not in the confidence interval for the j-th subgroup (i.e., $\hat{\mu} \notin [\underline{\mu}^j, \overline{\mu}^j]$), then declare the process out of control.