INSE6300 Jia Yuan Yu

Assignment 1

Concordia February 9, 2016

Assignment answers should be submitted on Moodle in PDF format.

The dataset that we will use for this assignment comes from CapBeast (cf. Moodle). We will use the column "Production time" as the measurements of interest $(X_1, X_2, \ldots, X_{1498})$. The data is in CSV format: you should import it into your favorite scientific computing software (e.g., R, Python, Matlab, etc.).

1 Distribution estimation

1. (10 points) Estimate the probability density function using histograms (bins) of width 20. Plot it.

2. (10 points) Using the density, estimate the cumulative distribution function. Plot it.

3. (10 points) Suppose that $X_1, X_2, \ldots$ are i.i.d., what kind of guarantee does the DKW Inequality give us about the estimated distribution and the true unknown distribution?

4. (10 points) Pick a value of δ in the interval $(0, 1)$. Using the above estimated distribution, find a value $z(\delta)$ such that you can give a guarantee of the form

$$\mathbb{P}(X_{1500} \leq z(\delta)) \geq 1 - \delta.$$

2 Control charts

1. (10 points) Draw the $\bar{X}$ control chart diagram on the production time data using chunks of $n = 10$ data points. First estimate the mean via the sample mean $\hat{m}u$ and standard deviation via the sample variance $\hat{\sigma}^2$ using the first chunk of data points.

2. (10 points) Repeat the above using $n = 20$.

3. (10 points) Draw the R control chart. Estimate $\bar{R}$ using the expression $\bar{R}_m = \frac{1}{m}\sum_{i=1}^{m} R^i$, with $m = 5$ and chunks of $n = 10$ data points.

3 CUSUM

(20 points) Explain how to use the CUSUM chart to detect changes in distribution in the production time data.