Appendix: Bayesian control chart construction details

At each time point k, once the data point x_k becomes available, we obtain the posterior distribution $p\left(\theta \middle| x_1,...,x_k\right)$ and then the predictive distribution of the future observable x_{k+1} i.e. $f\left(x_{k+1} \middle| x_1,...,x_k\right)$

In the Normal prior – Normal likelihood scenario adopted here, both posterior and predictive distribution will be available in closed forms and can be obtained recursively as the following theorem shows:

Theorem: If the initial prior distribution of the unknown parameter is:

$$\theta \sim N\left(\mu,\sigma^2\right)$$

and the data constitute a random sample with likelihood:

$$X_k \mid \theta \overset{iid}{\sim} N\left(\theta,\tau^2\right)$$

then the posterior distribution at time $k=1,2,...$ will be given by:

$$\theta \mid X_1,X_2,...,X_k \sim N\left(\hat{\theta}_k,\hat{\sigma}_k^2\right)$$

and the predictive distribution will be:

$$X_{k+1} \mid X_1,X_2,...,X_k \sim N\left(\hat{\theta}_k,\hat{\sigma}_k^2+\tau^2\right)$$

where:

$$\hat{\theta}_k = \frac{\hat{\sigma}_{k-1}^2 x_k + \tau^2 \hat{\theta}_{k-1}}{\hat{\sigma}_{k-1}^2 + \tau^2} = w_k x_k + \left(1-w_k\right)\hat{\theta}_{k-1}$$

$$\hat{\sigma}_k^2 = \frac{\hat{\sigma}_{k-1}^2 \tau^2}{\hat{\sigma}_{k-1}^2 + \tau^2} = w_k \tau^2 = \left(1 - w_k\right) \hat{\sigma}_{k-1}^2$$

$$w_k = \frac{\hat{\sigma}_{k-1}^2}{\hat{\sigma}_{k-1}^2 + \tau^2}, \qquad \hat{\sigma}_0^2 = \sigma^2 \qquad \text{and} \qquad \hat{\theta}_0 = \mu$$

The predictive mean combines information from the prior setting and the incoming data; as k increases, the effect of the prior distribution decays. Similarly, the variance of the predictive distribution will decrease as k increases, converging on τ^2.

At each time k $\left(k \geq 1\right)$, we know the predictive distribution $X_{k+1} \mid X_1, ..., X_k$, for the future data point X_{k+1}. To construct a control chart, this predictive distribution can be summarized using an interval. The endpoints of the predictive interval (which can be obtained by taking $\alpha/2$ probability out of each tail) will provide the control limits for the upcoming observation x_{k+1}. Note that the control limits obtained from the predictive distribution are probabilistic (i.e., the next observable has a probability of $1-\alpha$ of being within these limits). This is radically different from traditional Shewhart-type chart control limits (e.g., 1_{2s}) that can be used for testing a point null hypothesis: i.e., we are able to reject or not the null hypothesis that the unknown parameter θ equals some constant value.

In the BCC we initially standardize the incoming observation according to the predictive distribution $N\left(\hat{\theta}_k, \hat{\sigma}_k^2 + \tau^2\right)$ by:

$$z_{k+1} = \frac{x_{k+1} - \hat{\theta}_k}{\sqrt{\hat{\sigma}_k^2 + \tau^2}}$$

and draw the control limits at $\pm z_{\alpha/2}$. Thus the

Thus the steps in constructing the Bayesian control chart are as follows:

1. Decide the appropriate α value and plot the control limits at $\pm z_{\alpha/2}$ (center line at zero).
2. Once a data point x_k $(k \geq 1)$ becomes available, calculate the predictive mean and variance of the future observable $X_{k+1} | X_1, X_2, ..., X_k$ based on the theorem.
3. Once x_{k+1} becomes available, standardize it to obtain z_{k+1} and plot it in the control chart.
4. If z_{k+1} falls
   a. within the limits, then the data conform to the in control scenario and the process will continue to operate: i.e., we replace k by $k+1$ as subscript and move back to step 2;
   b. outside the predictive limits, it is an outlier, providing an alarm that the process has moved away from the in control state.

The predictive distribution is first available right after x_1 becomes available; thus charting can start from the second data point on. In the first step of BCC construction, we need to specify the appropriate α value. This value will determine the performance of the control chart, since its choice is a compromise between detection power and false alarm rate.

We base our decision regarding α on the False Alarm Probability (FAP) performance metric, which is defined as the probability of getting at least one false alarm in the m observations of the preliminary phase. The value of α as a function of FAP and the number of data points m will be:

$$\alpha = 1 - (1 - FAP)^{1/(m-1)}$$

If type I error is quite expensive compared to type II, we will select a small FAP (α) value, moving the control limits away from the center line. On the other hand, if the converse is true (i.e., type II is more important), we use a large FAP (α) value, moving the limits closer to the center line.