PPQ Power Assessment Theoretical Results

Yalin Zhu (yalin.zhu@merck.com)

2018-02-18

Preliminaries

Without loss of generality, suppose n outcomes of the Critical Quality Attribute (CQA) are normally distributed, which is denoted by $X_i \overset{\text{i.i.d}}{\sim} \mathcal{N}(\mu, \sigma^2)$, where $i = 1, \ldots, n$, then the distributions of sample mean and standard deviation are as known:

$$\bar{X} \sim \mathcal{N}(\mu, \frac{\sigma^2}{n}) \tag{1}$$

and

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1). \tag{2}$$

Moreover, sample mean and sample standard deviation are independent under normal distribution assumption.

Denote the lower and upper specification limits as L and U, respectively. The prediction or tolerance interval can be expressed by

$$[Y_1, Y_2] = \left[\bar{X} - kS,\ \bar{X} + kS\right], \tag{3}$$

where k is a speicific multiplier for the interval. For example, for prediction interval, $k = t_{1-\alpha/2, n-1}\sqrt{1 + \frac{1}{n}}$.

Specification test for one release batch

The outcome at release can be any one of the sample, so $X_{rl} \sim \mathcal{N}(\mu, \sigma^2)$, then the probability of passing PPQ at release should be

$$\begin{aligned} \Pr(\text{Passing Specification for Release}) &= \Pr(L \leq X_{rl} \leq U) \\ &= \Phi(U) - \Phi(L) \end{aligned} \tag{4}$$

This probability is very easy to calculate using software, such as `pnorm()` in R.

Test for PPQ Batches

$$\begin{aligned} \Pr(\text{Passing a Single PPQ Batch}) &= \Pr(L \leq Y_1 \leq Y_2 \leq U) \\ &= \int_L^U \int_L^{y_2} f_{Y_1,Y_2}(y_1, y_2) dy_1 dy_2 \end{aligned} \tag{5}$$

Now it is essential to obtain the bivariate joint distribution of the lower and upper prediction/tolerance interval, that is, find joint probability density function (PDF) $f_{Y_1,Y_2}(y_1, y_2)$.

Since $Y_1 = \bar{X} - kS$ and $Y_2 = \bar{X} + kS$, we can use another bivariate PDF $f_{\bar{X},S}(x, s)$ to calculate $f_{Y_1,Y_2}(y_1, y_2)$ by using Jacobian transformation.

Solve $\bar{X}$ and S as $x = \dfrac{y_1 + y_2}{2}$ and $s = \dfrac{y_2 - y_1}{2k}$, then Jacobian of the transformation is

$$|J| = \begin{vmatrix} \dfrac{\partial x}{\partial y_1} & \dfrac{\partial x}{\partial y_2} \\ \dfrac{\partial s}{\partial y_1} & \dfrac{\partial s}{\partial y_2} \end{vmatrix} = \begin{vmatrix} \dfrac{1}{2} & \dfrac{1}{2} \\ -\dfrac{1}{2k} & \dfrac{1}{2k} \end{vmatrix} = \frac{1}{2k}. \tag{6}$$

Thus, (5) can be calulated as

$$\begin{aligned} \int_L^U \int_L^{y_2} f_{Y_1,Y_2}(y_1, y_2) dy_1 dy_2 &= \int_L^U \int_L^{y_2} f_{\bar{X},S}\left(\frac{y_1 + y_2}{2}, \frac{y_2 - y_1}{2k}\right) |J| dy_1 dy_2 \\ &= \frac{1}{2k} \int_L^U \int_L^{y_2} f_{\bar{X}}\left(\frac{y_1 + y_2}{2}\right) f_S\left(\frac{y_2 - y_1}{2k}\right) dy_1 dy_2. \end{aligned} \tag{7}$$

The second equation follows from normal sample mean and standard deviation being independent.

Similarly, we can obtain the PDF of sample standard deviation $f_S(s)$. By (2), let $v = \dfrac{(n-1)s^2}{\sigma^2}$, then Jacobian of the transformation is

$$|J| = \left|\frac{dv}{ds}\right| = \frac{2(n-1)s}{\sigma^2}. \tag{8}$$

Thus,

$$\begin{aligned} f_S(s) &= f_V\left(\frac{(n-1)s^2}{\sigma^2}\right) |J| \\ &= \frac{2(n-1)s}{\sigma^2} f_V\left(\frac{(n-1)s^2}{\sigma^2}\right) \end{aligned} \tag{9}$$

Plug (9) in (7), we can get the final results.

$$\begin{aligned} &\Pr(\text{Passing a Single PPQ Batch}) \\ =&\frac{1}{2k} \int_L^U \int_L^{y_2} f_{\bar{X}}\left(\frac{y_1 + y_2}{2}\right) \frac{2(n-1)\frac{y_2 - y_1}{2k}}{\sigma^2} f_V\left\{\frac{(n-1)\left[\frac{y_2 - y_1}{2k}\right]^2}{\sigma^2}\right\} dy_1 dy_2 \\ =&\frac{n-1}{2k^2\sigma^2} \int_L^U \int_L^{y_2} f_{\bar{X}}\left(\frac{y_1 + y_2}{2}\right) f_V\left\{\frac{(n-1)(y_2 - y_1)^2}{4k^2\sigma^2}\right\} (y_2 - y_1) dy_1 dy_2, \end{aligned} \tag{10}$$

where $\bar{X} \sim \mathcal{N}(\mu, \frac{\sigma^2}{n})$ and $V \sim \chi^2(n-1)$. Then this quantity can be easily calculated by software, such as functions `dnorm()`, `dchisq()` and `integrate()` in R.

We can also calculate the probability of passing m PPQ batches, then under the assumption of independence and similar expected performance across batches, the probability will be

$$\Pr(\text{Passing } m \text{ batches }) = \{\Pr(\text{Passing a Single PPQ Batch})\}^m$$