its use. The computation of the limits of the "classical" control charts assumes that the distribution of the data must be normally distributed. If this is not the case, the type I error α really obtained will be different than the expected type I error $\alpha = 0.0027$ (3σ limits). But how large is the difference between the observed and the expected type I error when the kurtosis increases? In order to evaluate the impact of the kurtosis γ_2 of the data on the observed type I error we used the following approach:

- for a sample size $n \in \{5, 7, 9\}$.
- for a kurtosis $\gamma_2 \in [0, 10]$, mean $m_1 = 0$, and variance $\mu_2 = 1$.
- compute b and d using Eqs. (3) and (4).
- generate (by inverse simulation of Eq. (2)) a set of m samples of n symmetrical Johnson S_U random variables having parameters (b, d). The number m of sample has been chosen such that the total number of generated data is $m \times n = 3\,465\,000$.
- compute the mean, median, standard deviation, and range for each sample.
- compute the proportion of data outside the control limits for each of the four statistics. For a kurtosis $\gamma_2 = 0$, the estimated observed proportion of data outside the control limits (the type I error) must be close to the expected one $\alpha = 0.0027$, for all the control charts.

In Fig. 3 we have plotted the observed type I error versus the kurtosis for sample size $n = 5, 7, 9$. The conclusions of these simulations are:

- The mean and median charts seem to be very insensitive to the kurtosis of the data. As expected, this is particulary true for the median. The larger the sample size, the more insensitive the charts.
- In contrary, the standard deviation and range charts seem to be very sensitive to the kurtosis. The range chart is the most sensitive. The larger the sample size, the more sensitive the charts.

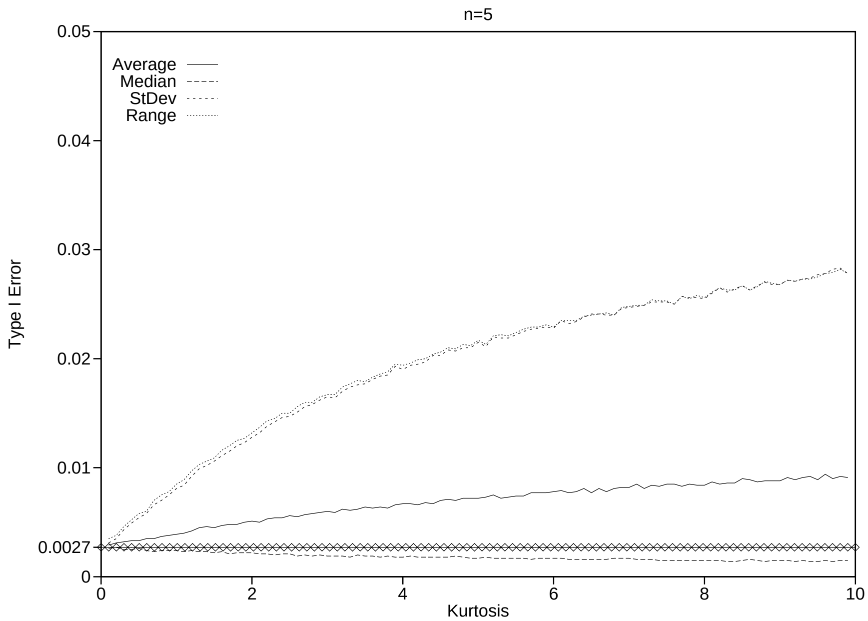
From this, we can conclude that the method proposed in this paper is mostly devoted to the computation of the limits of standard deviation or range charts (dispersion charts in general) of data having a positive kurtosis, even if it can also be applied to the other statistics.

6. OC and ARL Curves

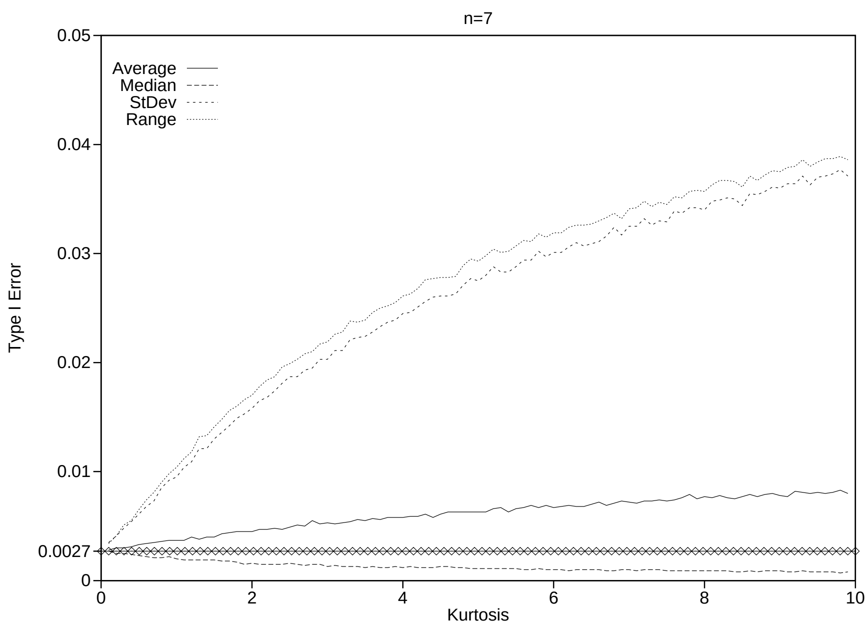
Let X be a symmetrical Johnson S_U with parameters (b, d). Without loss of generality, we will assume that $m_1 = 0$. By definition, the random variable Z defined by:

$$Z = b \sinh^{-1}\left(\frac{X}{d}\right) \tag{6}$$

is a normal $(0, 1)$ random variable. In order to compute the OC and ARL curves of a control chart using the method proposed in this paper, we have to first find the distributions of the random variables U and V defined as:

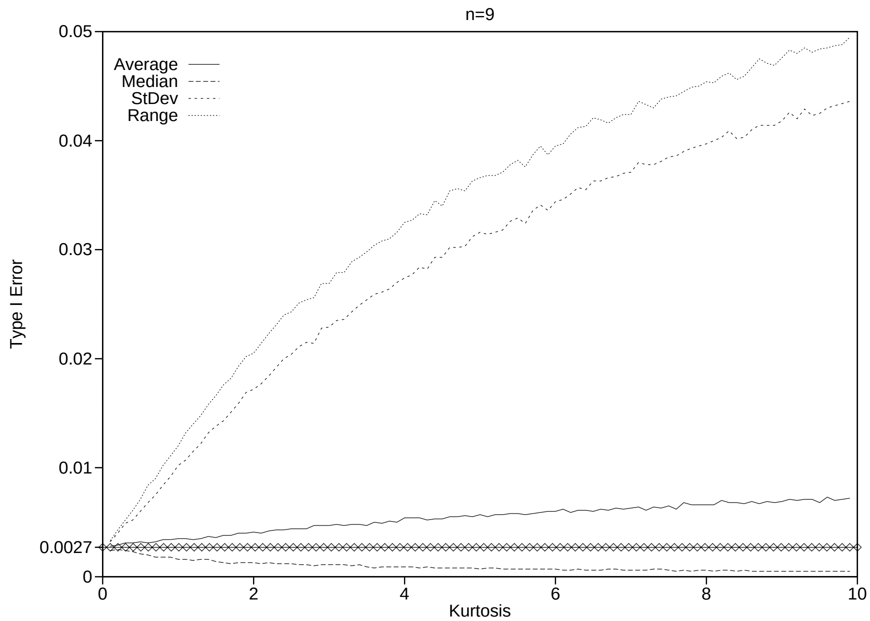


(a)



(b)

Fig. 3. Influence of the kurtosis on the observed type I error.



(c)

Fig. 3. (*Continued*).

$$U = b \sinh^{-1}\left(\frac{\epsilon + X}{d}\right) \tag{7}$$

$$V = b \sinh^{-1}\left(\frac{\tau X}{d}\right) \tag{8}$$

These random variables result from the transformation of the random variables $X + \epsilon$ and τX by the symmetrical Johnson S_U transformation with parameters (b, d), where ϵ and $\tau > 0$ are shifts corresponding respectively to the mean and standard deviation of the process. From Eqs. (6) and (7) we easily deduce:

$$\frac{X}{d} = \sinh\left(\frac{Z}{b}\right)$$

$$\frac{X + \epsilon}{d} = \sinh\left(\frac{U}{b}\right)$$

Substracting the first equation from the second yields $\sinh(U/b) - \sinh(Z/b) = \epsilon/d$. Then:

$$Z = b \sinh^{-1}\left[\sinh\left(\frac{U}{b}\right) - \frac{\epsilon}{d}\right]$$

Because Z is a normal $(0,1)$ random variable, we immediately have the cummulative distribution function of the random variable U

$$F_U(u) = \Phi\left\{b \sinh^{-1}\left[\sinh\left(\frac{u}{b}\right) - \frac{\epsilon}{d}\right]\right\} \tag{9}$$

and from the equation above we immediately deduce its inverse cummulative distribution function:

$$F_U^{-1}(\alpha) = b \sinh^{-1}\left\{\sinh\left[\frac{\Phi^{-1}(\alpha)}{b}\right] + \frac{\epsilon}{d}\right\} \tag{10}$$

Deriving Eq. (9) gives the density function $f_U(u)$ of U

$$f_U(u) = \frac{\cosh\left(\frac{u}{b}\right)}{\left\{1 + \left[\sinh\left(\frac{u}{b}\right) - \frac{\epsilon}{d}\right]^2\right\}^{1/2}} \phi\left\{b \sinh^{-1}\left[\sinh\left(\frac{u}{b}\right) - \frac{\epsilon}{d}\right]\right\} \tag{11}$$

Of course, if $\epsilon = 0$ (no shift in the process mean) then Eqs. (9)–(11) simply become $F_U(u) = \Phi(u)$, $F_U^{-1}(\alpha) = \Phi^{-1}(\alpha)$, and $f_U(u) = \phi(u)$. From Eqs. (6) and (8) we have:

$$\frac{X}{d} = \sinh\left(\frac{Z}{b}\right)$$

$$\frac{\tau X}{d} = \sinh\left(\frac{V}{b}\right)$$

Dividing the second equation by the first gives:

$$\tau = \frac{\sinh\left(\frac{V}{b}\right)}{\sinh\left(\frac{Z}{b}\right)}$$

and then:

$$Z = b \sinh^{-1}\left[\frac{1}{\tau} \sinh\left(\frac{V}{b}\right)\right]$$

The cummulative distribution function of the random variable V is then equal to:

$$F_V(v) = \Phi\left\{b \sinh^{-1}\left[\frac{1}{\tau} \sinh\left(\frac{v}{b}\right)\right]\right\} \tag{12}$$

and its inverse cummulative distribution function can be obtained from the previous equation:

$$F_V^{-1}(\alpha) = b \sinh^{-1}\left\{\tau \sinh\left[\frac{\Phi^{-1}(\alpha)}{b}\right]\right\} \tag{13}$$

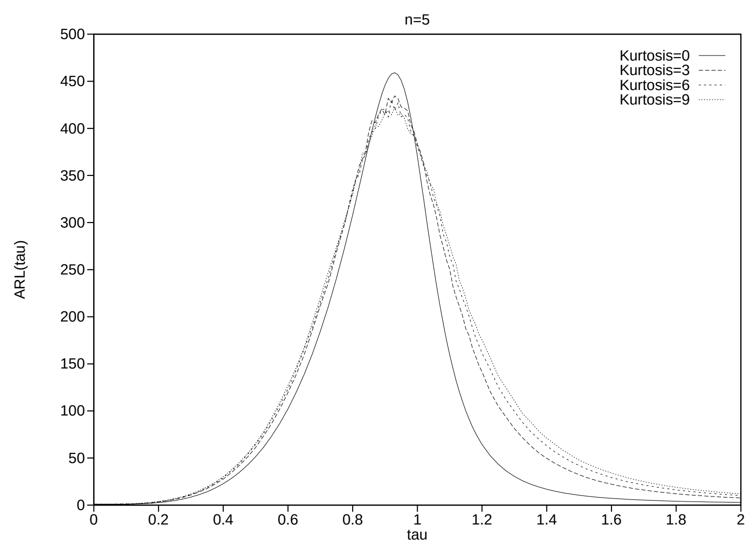
Deriving Eq. (12) gives the density function $f_V(v)$ of V

$$f_V(v) = \frac{\cosh\left(\frac{u}{b}\right)}{\left[\tau^2 + \sinh^2\left(\frac{u}{b}\right)\right]^{1/2}} \phi\left\{b \sinh^{-1}\left[\frac{1}{\tau} \sinh\left(\frac{v}{b}\right)\right]\right\} \tag{14}$$

As for the case $\epsilon = 0$, if $\tau = 1$ (no shift in the process standard deviation) then Eqs. (12)–(14) simply become $F_V(v) = \Phi(v)$, $F_V^{-1}(\alpha) = \Phi^{-1}(\alpha)$ and $f_V(v) = \phi(v)$. Unfortunately, because the random variables U and V are clearly non-normally distributed, it seems impossible to straightforwardly find the distribution of the sample mean or median (idem for the sample standard deviation or range). The only way to achieve the computation of the OC or ARL curves seems to be simulation. As an example, we have computed the ARL curve for the standard deviation using the following approach:

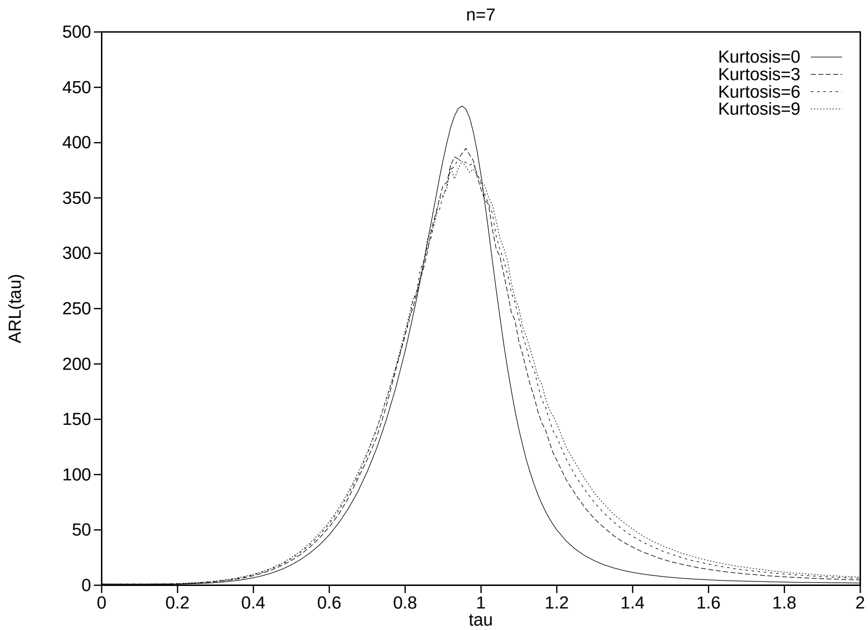
- for a sample size $n \in \{5, 7, 9\}$.
- for a kurtosis $\gamma_2 \in [0, 10]$, mean $m_1 = 0$, and variance $\mu_2 = 1$.
- for a shift $\tau \in [0, 2]$.
- compute b and d using Eqs. (3) and (4).
- generate (by inverse simulation of Eq. (13)) a set of m samples of n random variables V. The number m of sample has been chosen such that the total number of generated data is $m \times n = 3\,465\,000$.
- compute the standard deviation for each sample.
- compute the proportion of data inside the control limits for the standard deviation. This gives an estimate for OC(τ). Then, deduce an estimate for ARL(τ) = $1/(1 - \text{OC}(\tau))$.

The results of this simulation are plotted in Fig. 4. The main conclusion is that the larger the kurtosis, the larger the ARL of the standard deviation chart. Of course, for a specific kurtosis, when the sample size n increases, the ARL decreases.

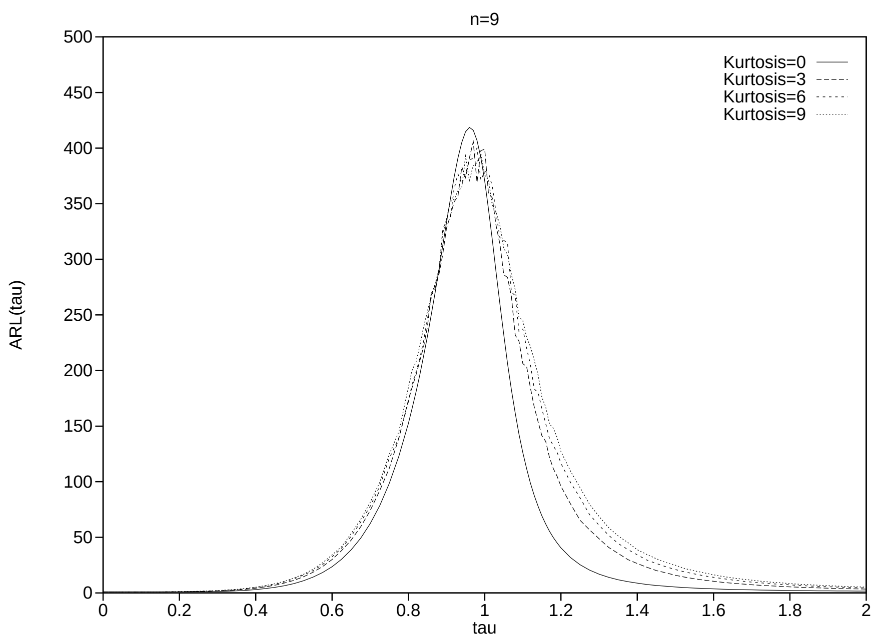


(a)

Fig. 4. Examples of ARL curves for the standard deviation chart.



(b)



(c)

Fig. 4. (*Continued*).