Chapter 3

Estimation Effect in Control Charts

3.1 Introduction

A feature that may affect the performance of a control chart is the estimation effect. In this chapter we present the current status of research of this field and some new results. In Section 3.2, we present the case on the estimation effect issue in univariate and multivariate Shewhart charts. New results on the effect of estimation on the values of average run length (ARL) and standard deviation of the run length (SDRL) of the S chart with three sigma and probability limits in the case of subgroups are also presented. Corresponding results for the X chart for individual observations are also presented. In Section 3.3 we refer to the estimation effect in the EWMA chart.

3.2 Estimation Effect in Univariate and Multivariate Shewhart Charts

The estimation effect issue in Shewhart charts was investigated by many authors. Proschan and Savage (1960) considered the effect of the number of samples and sample size on the performance of the $\overline{X}$ chart in terms of the probability of the mean plotting outside the control limits if we are in-control when the average range or a pooled estimate of the variance is used as an estimate of the process variability. They provided values for the number of samples needed for keeping stable the probability of the mean plotting outside the control limits if we are in-control for given values of the sample size.

Table 3.1.Correlation for several values of m and n

	n			
m	5	10	20	50
5	0.46581	0.37055	0.30735	0.25370
10	0.30362	0.22741	0.18158	0.14528
20	0.17898	0.12829	0.09986	0.07833
30	0.12689	0.08935	0.06886	0.05362
50	0.08020	0.05560	0.04249	0.03288
100	0.04178	0.02859	0.02171	0.01671
200	0.02133	0.01450	0.01097	0.00843
500	0.00864	0.00585	0.00442	0.00339
1000	0.00434	0.00293	0.00221	0.00170

However, they did not take into account the dependence between the event that the sample mean of sample i exceeds UCL and the event that the sample mean of another sample j exceeds UCL. Therefore, there results are of limited use. Hillier (1969) dealt with the problem of estimated control limits in the case of $\overline{X}$ and R chart. He provided a method of evaluating the probability of the mean plotting outside the control limits in the case of the $\overline{X}$ with the range R used to compute the process variability. This method did not consider the dependence issue as the method of Proschan and Savage (1960), consequently we can not base the design of our chart on these results.

Ghosh et al. (1981) gave formulas for the computation of the run length distribution in the case of the $\overline{X}$ chart with unknown variance. Quesenberry (1993) examined the effect of estimation of the process mean and standard deviation on the control limits of the Shewhart chart for the mean for both rational subgroups and individual observations.

Table 3.2. ARL and SDRL values for the S (three sigma) chart when $n = 5$

	σ_1^2/σ_0^2									
	1		1.2		1.4		1.6		1.8	
m	ARL	$SDRL$	ARL	$SDRL$	ARL	$SDRL$	ARL	$SDRL$	ARL	$SDRL$
5	$4 \cdot 10^5$	$1 \cdot 10^5$	2223.5	$7 \cdot 10^4$	594.02	$2 \cdot 10^4$	105.38	1353.3	37.41	143.16
10	2200.1	$3 \cdot 10^4$	310.65	2288.7	86.99	330.39	39.29	104.72	21.06	45.34
20	551.16	1699.7	139.42	297.08	54.88	95.14	27.47	41.75	16.48	21.70
30	415.06	840.14	112.55	182.48	48.74	74.75	25.79	34.47	15.52	18.54
50	346.68	545.72	101.62	134.96	43.32	55.43	23.36	27.19	14.73	16.19
100	298.59	407.05	91.09	106.99	40.75	44.42	22.35	23.56	14.20	14.55
200	276.08	318.09	85.28	93.97	39.28	41.12	21.55	22.18	13.93	13.92
500	262.29	275.14	85.20	88.07	38.55	39.24	21.75	21.79	13.94	13.52
1000	253.76	258.84	84.37	87.67	37.32	37.06	20.97	20.66	13.59	13.14
∞	249.31	248.81	82.44	81.94	37.72	37.21	21.22	20.71	13.69	13.18

He proved that

$$Corr(\overline{X}_i - \widehat{UCL}, \overline{X}_j - \widehat{LCL}) = \frac{Var(\widehat{UCL})}{Var(\overline{X}_i - \widehat{UCL})} = \left[1 + m\left(1 + \frac{9\,(1 - c_4^2)}{c_4^2}\right)^{-1}\right]^{-1},$$

which means that there is a correlation between the events $\overline{X}_i - \widehat{UCL}$ and $\overline{X}_j - \widehat{LCL}$. He concluded that $\overline{X}$ chart requires about $400/(n-1)$ samples for estimating the parameters in order for the estimated control limits to behave as the theoretical ones, where n is the subgroup size. In the case of individual observations he showed that 300 observations are needed for the estimated control limits to behave as the theoretical ones. The control chart he used is the X chart with the variability estimated by the moving range.

Chen (1997) extended this work by using three different estimators of the standard deviation in the $\overline{X}$ chart case. Let $X_{ij}, i = 1, 2, ...m$ and $j = 1, 2, ..., n$ represent data from a period known to operate in-control and let $Y_{ij}, i = 1, 2, ...$ and $j = 1, 2, ..., n$ represent

Table 3.3. ARL and SDRL values for the S (three sigma) chart when $n = 10$

	σ_1^2/σ_0^2									
	1		1.2		1.4		1.6		1.8	
m	ARL	$SDRL$	ARL	$SDRL$	ARL	$SDRL$	ARL	$SDRL$	ARL	$SDRL$
5	606.61	1064.81	236.14	634.06	78.31	263.87	29.43	112.67	14.39	41.18
10	538.65	919.10	145.57	329.89	45.83	99.37	19.21	31.64	10.17	13.87
20	461.44	725.80	106.92	175.82	33.86	48.22	15.85	19.75	9.04	10.34
30	430.50	626.79	95.59	137.72	32.34	40.07	15.02	17.08	8.59	9.14
50	389.91	510.09	88.05	106.54	30.29	33.98	14.38	15.65	8.37	8.58
100	359.35	411.69	80.89	88.11	28.79	30.81	13.85	14.24	8.26	8.16
200	344.38	367.08	78.19	82.25	28.46	28.96	13.43	13.31	7.97	7.58
500	334.53	340.97	76.10	76.60	27.45	27.27	13.52	13.14	8.06	7.65
1000	334.56	337.96	75.88	75.93	27.31	27.01	13.50	13.04	7.98	7.48
∞	331.17	330.67	75.66	75.16	27.52	27.01	13.47	12.96	8.00	7.48

current or future data. Also, let $X_{ij} \backsim N(\mu, \sigma^2)$ and $Y_{ij} \backsim N(\mu + \alpha\sigma, b^2\sigma^2)$ with α, b constants. Since $\overline{\overline{X}} \backsim N(\mu, \sigma^2/(mn))$ and $\overline{Y}_i \backsim N(\mu + \alpha\sigma, b^2\sigma^2/n)$ for given $\overline{\overline{X}} = \overline{\overline{x}}$ and given $\widehat{\sigma}$ we have

$$P(\overline{Y}_i < \widehat{LCL} \text{ or } \overline{Y}_i > \widehat{UCL}|\overline{\overline{x}}, \widehat{\sigma}) = 1 - \Phi\left(\frac{z}{b\sqrt{m}} + \frac{3}{b}w - \frac{\alpha}{b}\sqrt{n}\right) + \Phi\left(\frac{z}{b\sqrt{m}} - \frac{3}{b}w - \frac{\alpha}{b}\sqrt{n}\right),$$

where $z = (\overline{\overline{x}} - \mu) / (\sigma/\sqrt{mn})$ and $w = \widehat{\sigma}/\sigma$. Then, the ARL is computed through the following relation

$$ARL = \int_{-\infty}^{+\infty} \int_{0}^{+\infty} \frac{1}{P(\overline{Y}_i < \widehat{LCL} \text{ or } \overline{Y}_i > \widehat{UCL}|\overline{\overline{x}}, \widehat{\sigma})} \frac{1}{\sqrt{2\pi}} \exp\left(-0.5z^2\right) f(w) dz dw,$$

where $f(w)$ is calculated for three different estimators of σ. For a detailed discussion on the different estimators of σ, see Vardeman (1999).

Table 3.3.(continued) ARL and SDRL values for the S (three sigma) chart when $n = 10$

	σ_1^2/σ_0^2							
	0.2		0.4		0.6		0.8	
m	ARL	$SDRL$	ARL	$SDRL$	ARL	$SDRL$	ARL	$SDRL$
5	24.01	37.43	306.28	520.37	1019.6	1274.5	1136.2	1433.6
10	21.04	25.88	254.28	377.34	1071.7	1253.2	1316.9	1514.7
20	19.77	22.62	230.60	275.74	1079.4	1207.5	1472.6	1603.4
30	19.15	20.62	223.33	249.71	1056.5	1155.1	1569.2	1656.9
50	18.47	19.15	218.20	229.50	1047.3	1106.2	1644.7	1686.9
100	18.21	18.10	210.57	215.40	1037.5	1061.3	1696.2	1729.9
200	17.95	17.93	205.32	205.49	1023.1	1027.8	1744.5	1746.7
500	18.20	17.79	205.59	203.14	1009.1	1022.0	1773.9	1785.0
1000	17.53	17.28	206.96	204.95	1006.7	1007.9	1768.3	1773.9
∞	17.90	17.39	206.06	205.56	1011.7	1011.2	1777.2	1776.7

Nedumaran and Pignatiello (2001) developed new control limits for the $\overline{X}$ chart taking into account the estimation effect. Specifically, let $\overline{X}_i$ be the average of a future subgroup, $\overline{V}$ be the average variance of the m initial in-control subgroups and $T_i = \frac{\overline{X}_i - \overline{\overline{X}}}{\sqrt{\frac{m+1}{mn}}\sqrt{\overline{V}}}$. Then, $(T_{m+1}, T_{m+2}, ..., T_{m+k})$ has a positively equicorrelated multivariate t distribution with correlation $1/(m+1)$, where k is a specified number of future subgroups. If $P\left(\widehat{LCL} \leq \overline{X}_i \leq \widehat{UCL}\right) = 1 - \gamma$, $i = m+1, m+2, ..., m+k$ then γ must be equal to the run length distribution percentile when we have true limits, for the estimated limits to have equivalent performance with the true ones. Then $\gamma = P[RL \leq k] = 1 - (1-\alpha)^k$ where α is the probability of a false alarm for a single subgroup. If $P\left[\max_{m+1 \leq i \leq m+k} |T_i| \leq h'_{\gamma,m,k,\nu}\right] = 1 - \gamma$ where $\nu = m(n-1)$, we have that $P\left[\overline{\overline{X}} - h'_{\gamma,m,k,\nu}\sqrt{\frac{m+1}{mn}}\sqrt{\overline{V}} \leq \overline{X}_i \leq \overline{\overline{X}} + h'_{\gamma,m,k,\nu}\sqrt{\frac{m+1}{mn}}\sqrt{\overline{V}}\right] = 1 - \gamma$. Consequently, the control limits are

$$\widehat{UCL} = \overline{\overline{X}} + h'_{\gamma,m,k,\nu}\sqrt{\frac{m+1}{mn}}\sqrt{\overline{V}}$$
$$\widehat{LCL} = \overline{\overline{X}} - h'_{\gamma,m,k,\nu}\sqrt{\frac{m+1}{mn}}\sqrt{\overline{V}}.$$

Table 3.4. ARL and $SDRL$ values for the S (three sigma) chart when $n = 20$

	σ_1^2/σ_0^2									
	1		1.2		1.4		1.6		1.8	
m	ARL	$SDRL$	ARL	$SDRL$	ARL	$SDRL$	ARL	$SDRL$	ARL	$SDRL$
5	332.72	444.02	121.96	244.63	32.29	78.86	11.46	27.03	5.54	10.33
10	362.96	457.01	92.99	166.56	23.32	45.79	8.71	11.78	4.62	5.39
20	371.24	439.25	75.00	115.39	19.63	25.36	7.87	8.77	4.32	4.20
30	372.32	430.13	68.53	86.90	18.23	21.84	7.67	8.17	4.29	4.12
50	362.66	403.51	63.80	76.74	17.52	18.73	7.49	7.60	4.24	3.97
100	364.01	393.80	60.17	65.57	17.01	17.34	7.36	7.15	4.11	3.58
200	359.00	374.30	59.56	60.31	16.61	16.45	7.15	6.82	4.11	3.59
500	355.18	358.14	59.11	59.61	16.36	16.15	7.13	6.70	4.09	3.56
1000	353.23	353.28	57.59	57.23	16.26	15.79	7.15	6.66	4.08	3.59
∞	356.50	356.00	57.37	56.87	16.39	15.88	7.15	6.63	4.07	3.53

In the case of the attributes charts p and c with estimated control limits Braun (1999) computed the run length distributions. If W is the run length until the next signal we have that

$$P\left(W \leq w\right) = 1 - \sum_x P\left(\sum_{j=1}^m X_j = x\right)\left(1 - P\left(F_1 | \sum_{j=1}^m X_j = x\right)\right)^w,$$

where F_i is the event that the ith new observation is outside the estimated control limits.

In the case of the c chart we have that

$$P\left(F_1|\sum_{j=1}^{m}X_j=x\right)=1-\sum_{j=\left[x/m-3\sqrt{x/m}\right]+1}^{\left[x/m+3\sqrt{x/m}\right]}\frac{e^{-bc}\left(bc\right)^j}{j!},x=0,1,2,...$$

and in the case of the p chart

$$P\left(F_1|\sum_{j=1}^{m}X_j=x\right)=1-\sum_{j=\left[nx/m-3\sqrt{nx/m(1-x/m)}\right]+1}^{\left[nx/m+3\sqrt{nx/m(1-x/m)}\right]}\binom{n}{j}\left(bp\right)^j\left(1-bp\right)^{n-j},$$

where $x=0,1/n,2/n,...,(mn-1)/n,mn/n$. In the case of the c chart $\sum_{j=1}^{m}X_j$ is distributed as a Poisson random variable with mean mc therefore $P\left(\sum_{j=1}^{m}X_j\right)=\frac{e^{-mc}(mc)^x}{x!}$, $x=0,1,2,...$ In the case of the p chart $n\sum_{j=1}^{m}X_j$ is distributed as a Binomial random variable with parameters mn and p that is $P\left(\sum_{j=1}^{m}X_j\right)=\binom{mn}{nx}p^{nx}\left(1-p\right)^{(m-x)n}$, $x=0,1/n,2/n,...,(mn-1)/n,mn/n$. Finally, the ARL is equal to

$$ARL=\sum_{x}P\left(\sum_{j=1}^{m}X_j=x\right)P\left(F_1|\sum_{j=1}^{m}X_j=x\right)^{-1}$$

Braun (1999) showed that, as for variables control charts, the estimation effect can be serious.

Yang et al. (2002) examined the case of the Geometric chart with estimated control limits. The run length distribution in this case is equal to

$$P\left(R\leq r;p,p_0\right)=\sum_{n=0}^{m}\left[1-\alpha\left(n\right)\right]^{r-1}\alpha\left(n\right)\binom{m}{n}p_0^n\left(1-p_0\right)^{m-n},$$

where $\alpha\left(n\right)=\left(1-p\right)^{\ln(\alpha/2)[\ln(1-n/m)]}-\left(1-p\right)^{\ln(1-\alpha/2)[\ln(1-n/m)]}+1$, p_0 is the fraction nonconforming, m is the sample size and n is the number of nonconforming items. The

ARL in this case is equal to

$$ARL = \sum_{n=0}^{m} \frac{1}{\alpha(n)} \binom{m}{n} p_0^n (1 - p_0)^{m-n}.$$

Yang et al. (2002) showed that the effect on the alarm probability is significant even when the sample size is very large e.g. 10000. Despite that fact, the ARL is not affected that seriously, unless we have a small sample size and a large process improvement.

Nedumaran and Pigniatiello (1999) investigated the estimation effect on the T^2 control charts. They proposed that the number of subgroups needed for the estimated control limits to behave as the theoretical ones must be between $800p/3(n-1)$ and $400p/(n-1)$, where p is the number of variables and n is the sample size. Moreover, they gave an exact procedure for the construction of the T^2 control charts when we estimate the parameters so as to perform similar to the ones with known parameters.

Table 3.4. (continued) ARL and $SDRL$ values for the S(three sigma) chart when $n = 20$

	σ_1^2/σ_0^2							
	0.2		0.4		0.6		0.8	
m	ARL	$SDRL$	ARL	$SDRL$	ARL	$SDRL$	ARL	$SDRL$
5	1.32	0.75	11.92	18.83	111.01	190.51	383.43	457.13
10	1.28	0.64	10.03	12.23	90.20	127.11	423.04	463.05
20	1.26	0.60	9.21	9.96	80.28	94.68	442.70	473.45
30	1.26	0.58	8.97	9.20	78.03	88.22	444.96	471.63
50	1.24	0.54	8.90	8.59	75.70	80.02	451.20	469.84
100	1.25	0.57	8.68	8.20	73.42	75.19	450.90	455.92
200	1.23	0.54	8.70	8.27	73.69	74.61	447.50	446.43
500	1.24	0.55	8.55	8.05	73.62	73.39	441.19	441.96
1000	1.24	0.56	8.54	8.14	72.08	71.77	445.81	448.49
∞	1.24	0.54	8.56	8.04	72.91	72.41	449.79	449.29

Woodall and Montgomery (1999) emphasized the need for much more research in this area since it is proved that more data than usually recommended is needed for the control charts to behave as expected from theory. In the same paper, Woodall and Montgomery state that much work has been done concerning the control of the process mean but not that much for the process dispersion.

Table 3.5. ARL and SDRL values for the S (three sigma) chart when $n = 50$

	σ_1^2/σ_0^2									
	1		1.2		1.4		1.6		1.8	
m	*ARL*	*SDRL*	*ARL*	*SDRL*	*ARL*	*SDRL*	*ARL*	*SDRL*	*ARL*	*SDRL*
5	263.03	325.32	59.79	125.84	9.49	18.56	3.23	3.96	1.86	1.57
10	304.52	359.74	44.11	76.18	7.69	9.98	2.89	2.87	1.73	1.23
20	328.25	365.28	36.59	49.56	6.91	7.54	2.83	2.52	1.69	1.15
30	340.23	369.51	33.55	39.88	6.65	6.77	2.76	2.37	1.68	1.11
50	345.02	369.81	32.36	35.89	6.64	6.61	2.72	2.24	1.67	1.09
100	355.17	366.97	30.64	31.98	6.37	6.11	2.7	2.2	1.67	1.08
200	357.85	364.35	30.75	30.97	6.39	6.06	2.67	2.09	1.67	1.06
500	362.32	358.59	30.32	30.28	6.38	5.87	2.65	2.1	1.67	1.06
1000	356.30	352.76	30.62	29.97	6.29	5.89	2.67	2.08	1.67	1.05
∞	365.96	365.46	30.23	29.72	6.35	5.83	2.67	2.11	1.66	1.04

Chen (1998) deals with the run length properties of the R, s and s^2 control charts in the case that σ is estimated. Let $X_{ij}, i = 1, 2, ...m$ and $j = 1, 2, ..., n$ denote historically in-control data and $Y_{ij}, i = 1, 2, ...$ and $j = 1, 2, ..., n$ represent current or future data. Let $X_{ij} \backsim f((x - \mu)/\sigma)/\sigma$ and $Y_{ij} \backsim f((y - \mu)/(b\sigma))/(b\sigma)$ with b constant, μ, σ the process mean and standard deviation respectively and $f(\cdot)$ the form of the known density function. Denote $U = \widehat{\sigma}/\sigma$, where $\widehat{\sigma}$ is an estimate of σ calculated from the historical data set and $U \backsim h(u; m, n)$. Let $T_i = \widehat{\sigma}_i/\sigma$, where $\widehat{\sigma}_i$ is an estimate of $b\sigma$ using Y_{ij}. Also, denote $G(t; b, n) = P(T_i \leq t)$. Then, if L_n and U_n are the constants multiplied with σ for the known lower and upper control limits case respectively, we have that

$$P(\widehat{\sigma}_i < L_n\widehat{\sigma} \text{ or } \widehat{\sigma}_i > U_n\widehat{\sigma}|\widehat{\sigma}) = G(L_n u/b; b, n) + 1 - G(U_n u/b; b, n) = l(u; b, n),$$

where $u = \widehat{\sigma}/\sigma$. Then, the ARL is computed through the following relation

$$ARL = \int_0^{+\infty} \frac{1}{l(u; b, n)} h(u; m, n) du.$$

Table 3.5. (continued) ARL and SDRL values for the S (three sigma) chart when $n = 50$

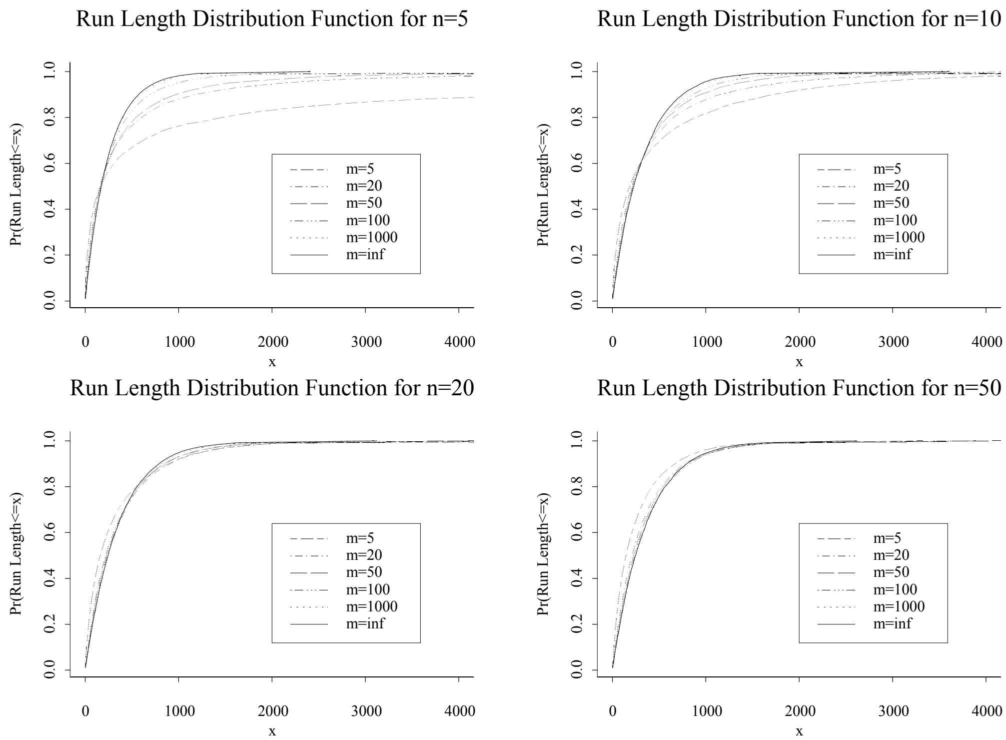
	σ_1^2/σ_0^2							
	0.2		0.4		0.6		0.8	
m	ARL	$SDRL$	ARL	$SDRL$	ARL	$SDRL$	ARL	$SDRL$
5	1	0	1.25	0.66	8.68	13.45	124.56	199.43
10	1	0	1.23	0.56	7.20	8.36	110.20	171.69
20	1	0	1.21	0.51	6.80	7.10	97.84	128.55
30	1	0	1.20	0.50	6.55	6.53	93.47	110.48
50	1	0	1.20	0.48	6.51	6.38	89.64	98.31
100	1	0	1.20	0.48	6.44	6.04	85.98	91.30
200	1	0	1.19	0.47	6.37	5.91	85.92	88.10
500	1	0	1.18	0.47	6.34	5.92	85.47	85.74
1000	1	0	1.18	0.47	6.26	5.82	85.82	85.86
∞	1	0	1.19	0.48	6.28	5.76	84.25	83.75

Maravelakis, Panaretos and Psarakis (2002) examine the effect of estimation of the process parameters on the control limits of charts for process dispersion by extending the results of Chen (1998) for both rational subgroups and individual observations. In sections 3.2.1-3.2.4 we present this work.

3.2.1 The S (Three Sigma) Control Chart

Assume that we have the control limits (2.6) and their estimated counterparts in (2.8). Let A_i denote the event that the ith sample standard deviation S_i exceeds UCL or is exceeded by LCL. Then, since S_i and S_j are independent for $i \neq j$, the sequence of trials A_i and A_j are independent meaning that they constitute a sequence of Bernoulli trials.

Figure 3.1. Empirical Run Length Distribution Functions for the 3 sigma chart



The mean and standard deviation of the run length distribution, ARL and $SDRL$ respectively, of this process is that of a geometric distribution given by the following formulas

$$\begin{aligned} ARL &= \frac{1}{1-\beta} \\ SDRL &= \frac{\sqrt{\beta}}{1-\beta} \end{aligned} \tag{3.1}$$

where $\beta = 1 - \Pr(A_i) = \Pr(LCL \leq S_i \leq UCL)$.

Assume now that we are in the case when the true value of the standard deviation is not known, which is the most usual case. Let B_i denote the event that the ith sample standard deviation S_i exceeds $\widehat{UCL}$ or is exceeded by $\widehat{LCL}$.

Table 3.6.Correlation for several values of m and n

	n			
m	5	10	20	50
5	0.51095	0.39568	0.32137	0.26032
10	0.34314	0.24663	0.19144	0.14964
20	0.20710	0.14066	0.10585	0.08087
30	0.14831	0.09839	0.07315	0.05541
50	0.09460	0.06145	0.04521	0.03400
100	0.04965	0.03170	0.02313	0.01729
200	0.02545	0.01611	0.01170	0.00872
500	0.01034	0.00650	0.00471	0.00351
1000	0.00520	0.00326	0.00236	0.00176

The formulas (3.1) for ARL and $SDRL$ are not valid any more because the events B_i and B_j are not independent for $i \neq j$. We can prove that $E(\widehat{UCL}) = UCL$ and $Var(\widehat{UCL}) = \left(1 + \frac{3}{c_4}\sqrt{1 - c_4^2}\right)^2 \sigma^2 \frac{(1-c_4^2)}{m}$ and using these relations we prove after some calculations that

$$Cov(S_i - \widehat{UCL}, S_j - \widehat{LCL}) = Var(\widehat{UCL}) = \left(1 + \frac{3}{c_4}\sqrt{1 - c_4^2}\right)^2 \sigma^2 \frac{(1 - c_4^2)}{m}$$

and

$$Var(S_i - \widehat{UCL}) = \left[1 + \frac{\left(1 + \frac{3}{c_4}\sqrt{1 - c_4^2}\right)^2}{m}\right] \sigma^2(1 - c_4^2)$$

Therefore, the correlation between the random variables $S_i - \widehat{UCL}$ and $S_j - \widehat{LCL}$ is

$$Corr(S_i - \widehat{UCL}, S_j - \widehat{LCL}) = \frac{Var(\widehat{UCL})}{Var(S_i - \widehat{UCL})} = \frac{\left(1 + \frac{3}{c_4}\sqrt{1 - c_4^2}\right)^2}{m + \left(1 + \frac{3}{c_4}\sqrt{1 - c_4^2}\right)^2}$$

Table 3.7. ARL and $SDRL$ values for the S (probability limits) chart when $n = 5$

	σ_1^2/σ_0^2									
	1		1.2		1.4		1.6		1.8	
m	ARL	$SDRL$	ARL	$SDRL$	ARL	$SDRL$	ARL	$SDRL$	ARL	$SDRL$
5	359.97	463.12	267.35	405.54	173.40	312.06	111.11	231.57	71.17	173.0
10	401.46	491.51	268.52	395.19	154.68	263.77	83.88	161.01	47.93	102.77
20	441.09	495.15	254.39	350.22	127.40	199.04	64.92	106.92	36.69	58.40
30	462.04	509.78	247.68	320.05	115.34	164.84	58.02	84.90	33.35	49.45
50	472.24	504.56	239.29	295.97	108.19	137.80	52.48	65.23	30.29	35.47
100	489.90	512.64	229.28	262.50	99.08	115.37	49.79	54.68	28.81	31.03
200	498.35	505.20	221.61	240.21	94.66	102.45	48.20	50.97	27.67	29.10
500	500.93	505.59	216.74	223.58	93.45	95.24	46.06	46.00	28.00	27.60
1000	497.73	503.09	213.01	217.36	92.29	94.50	47.12	47.08	27.31	26.70
∞	500.02	499.52	214.74	214.24	91.78	91.28	46.51	46.01	27.33	26.82

It is obvious that the correlation is a function of m and n only. In Table 3.1 we present values of the correlation for combinations of m and n. From this Table we see that as the sample size and the number of samples increases the correlation decreases. For small or moderate sample size ($n \leq 20$) we need 200 samples for the correlation to be negligible. However, for larger sample size the value $m = 50$ is suitable.

In order to examine the values of the first two moments of the run length distribution, we performed a simulation study based on various numbers of samples and various sample sizes. In particular the number of samples and samples sizes considered were $m = 5$, 10, 20, 30, 50, 100, 200, 500, 1000 and $n = 5$, 10, 20, 50. For every combination of m and

n we simulated m samples of size n from a $N(\mu, \sigma_0^2)$ distribution and computed $\widehat{UCL}$ and $\widehat{LCL}$. Then, we simulated samples from a $N(\mu, \sigma_1^2)$ distribution until we obtained a value above $\widehat{UCL}$ or below $\widehat{LCL}$. The number of samples simulated up to the one that lead to a value outside the control limits constitutes one observation of the run length distribution. This procedure was repeated 10000 times in order to get estimates of the values of ARL and $SDRL$. The results are presented in Tables $3.2 - 3.5$.

Table 3.7.(continued) ARL and $SDRL$ values for the S (probability limits) chart when $n = 5$

	σ_1^2/σ_0^2							
	0.2		0.4		0.6		0.8	
m	ARL	$SDRL$	ARL	$SDRL$	ARL	$SDRL$	ARL	$SDRL$
5	59.28	92.90	207.50	279.98	367.15	426.84	423.91	494.16
10	51.33	60.99	188.80	229.67	383.62	419.59	478.62	506.71
20	49.25	53.77	178.80	195.31	381.22	406.88	535.06	558.84
30	47.36	50.56	174.26	182.63	378.10	395.01	551.71	561.82
50	47.06	47.90	172.37	175.45	374.69	387.19	572.90	579.90
100	46.45	46.53	170.21	172.41	369.25	373.72	588.21	585.37
200	44.99	44.60	169.92	169.76	369.89	371.74	595.42	594.84
500	45.64	45.22	168.09	168.67	364.29	363.97	604.03	601.67
1000	45.64	44.65	165.86	166.04	364.05	369.30	598.0	601.84
∞	45.09	44.59	167.40	166.90	366.87	366.37	597.91	597.41

From Tables 3.2 through 3.5 certain conclusions are drawn. We see that we have results for both upward and downward shifts when $n > 5$ but only for upward when $n = 5$. This happens because for $n \leq 5$ the lower control limit is set to zero. Therefore, it can never be crossed. For upward shifts as m increases the ARL and $SDRL$ values decrease and approach their theoretical values. For downward shifts as m increases the same thing happens for $n = 50$. For $n = 10, 20$ the ARL and $SDRL$ values do not follow a specific trend. In the in-control state we also do not have a clear pattern for either ARL or $SDRL$ values. What we can say in every case is that ARL and $SDRL$ values behave in the same way.

Table 3.8. ARL and $SDRL$ values for the S (probability limits) chart when $n = 10$

	σ_1^2/σ_0^2									
	1		1.2		1.4		1.6		1.8	
m	ARL	$SDRL$	ARL	$SDRL$	ARL	$SDRL$	ARL	$SDRL$	ARL	$SDRL$
5	341.44	422.99	217.14	329.46	110.54	218.04	52.43	130.38	25.60	62.05
10	391.03	456.05	208.08	307.21	86.61	155.83	36.61	67.22	17.72	28.51
20	428.95	469.37	194.55	257.65	70.14	106.07	28.50	39.48	14.96	18.23
30	448.41	480.41	187.90	234.75	65.03	88.79	27.33	33.58	14.20	16.15
50	464.28	481.37	178.40	209.85	60.27	72.62	25.81	28.64	13.61	14.78
100	479.05	488.20	169.77	184.28	56.35	61.10	24.52	25.62	13.16	13.69
200	484.86	493.03	166.70	176.09	54.73	56.70	24.26	24.50	12.81	12.66
500	490.54	489.97	161.32	164.74	52.91	53.41	24.02	24.08	13.11	12.82
1000	492.16	480.65	161.60	161.87	53.81	53.11	23.60	23.23	12.70	12.38
∞	500.05	499.55	161.99	161.48	53.44	52.94	23.46	22.95	12.74	12.23

As m increases the ARL is getting closer to the theoretical value faster than the $SDRL$. Moreover, as n increases the theoretical values, in the in-control state, approach the ones from a normal distribution, which are $ARL = 370.4$ and $SDRL = 369.9$. The same of course happens and for the out-of-control states.

If we use this type of chart for identifying shifts in process dispersion we have to use samples of size n at least 20, for minimizing the effect of estimating S. If n is less than this value the practitioner will face an increased number of false alarms. The effect of estimation is also severe for $m \leq 20$, especially in the in-control state and for small shifts.

For values $30 \leq m \leq 50$ the effect is moderate and for values of 100 or larger the effect is small enough. A last point we have to make is that when we have small downward shifts for $n \leq 20$ the ARL and $SDRL$ values are larger than the corresponding in-control values. This result is also confirmed by Klein (2000). Consequently, in such cases special care must be given and it is better to use control charts for small shifts like CUSUM and EWMA.

Table 3.8. (continued) ARL and $SDRL$ values for the S (probability limits) chart when $n = 10$

	σ_1^2/σ_0^2							
	0.2		0.4		0.6		0.8	
m	ARL	$SDRL$	ARL	$SDRL$	ARL	$SDRL$	ARL	$SDRL$
5	5.34	7.41	46.09	81.74	182.01	262.85	339.52	406.60
10	4.63	4.92	38.28	48.93	162.80	206.09	378.42	422.41
20	4.40	4.27	35.18	40.33	154.72	181.67	396.79	417.98
30	4.36	4.10	34.06	37.31	147.37	159.98	400.78	424.06
50	4.20	3.76	33.36	35.08	144.77	156.54	401.66	421.77
100	4.27	3.81	32.66	34.24	139.43	140.95	402.89	413.00
200	4.26	3.73	32.78	33.25	137.34	137.24	400.48	403.36
500	4.17	3.64	32.44	31.76	136.91	133.59	405.11	405.09
1000	4.21	3.63	31.95	31.10	133.69	132.26	398.98	400.52
∞	4.23	3.70	32.13	31.62	136.47	135.97	400.85	400.35

In Figure 3.1 we present the empirical run length distribution functions (ERL) for $n = 5, 10, 20, 50$. In each Figure we plot six different lines representing the ERL functions for $m = 5, 20, 50, 100, 1000$ and the theoretical run length distribution (inf). It is obvious that as m increases the ERL approaches the theoretical run length distribution. Moreover, as n increases the ERL's for the m values approach the theoretical run length distribution faster.

3.2.2 The S (Probability Limits) Control Chart

Consider the control limits 2.10 and 2.11. In the same way of thinking as in the case of three sigma limits we can prove that $Var(\widehat{UCL}) = [\sigma^2(1-c_4^2)\chi^2_{0.999}]/[(n-1)c_4^2 m]$ and consequently

$$Cov(S_i - \widehat{UCL}, S_j - \widehat{LCL}) = Var(\widehat{UCL}) = \frac{\sigma^2(1-c_4^2)\chi^2_{0.999}}{(n-1)c_4^2 m}.$$

Moreover,

$$Var(S_i - \widehat{UCL}) = \sigma^2(1-c_4^2)\left[1 + \frac{\chi^2_{0.999}}{(n-1)c_4^2 m}\right]$$

Table 3.9. ARL and $SDRL$ values for the S (probability limits) chart when $n = 20$

	σ_1^2/σ_0^2									
	1		1.2		1.4		1.6		1.8	
m	ARL	$SDRL$	ARL	$SDRL$	ARL	$SDRL$	ARL	$SDRL$	ARL	$SDRL$
5	327.65	381.36	170.43	279.59	56.40	125.43	18.47	47.23	8.07	17.04
10	379.94	415.88	154.94	241.56	40.09	71.48	13.24	19.53	6.45	8.01
20	421.10	434.89	135.60	194.87	33.08	46.55	11.89	14.66	5.93	6.32
30	442.13	451.34	126.95	170.03	30.45	37.64	11.46	12.67	5.81	6.0
50	461.32	467.99	117.98	139.49	29.17	32.83	11.09	11.54	5.69	5.50
100	476.40	478.77	113.42	126.66	27.69	28.82	10.97	10.94	5.57	5.26
200	486.38	486.97	109.86	115.12	27.35	27.82	10.50	10.20	5.50	5.17
500	485.13	488.22	108.31	108.90	26.81	26.19	10.43	10.12	5.41	4.89
1000	494.29	488.10	106.57	108.41	26.63	25.79	10.30	9.78	5.48	4.91
∞	500.01	499.51	106.64	106.14	26.67	26.17	10.42	9.91	5.46	4.93

and finally

$$Corr(S_i - \widehat{UCL}, S_j - \widehat{LCL}) = \frac{Var(\widehat{UCL})}{Var(S_i - \widehat{UCL})} = \frac{\chi^2_{0.999}}{\chi^2_{0.999} + (n-1)c_4^2 m}.$$

Table 3.9.(continued) ARL and $SDRL$ values for the S (probability limits) chart when $n = 20$

	σ_1^2/σ_0^2							
	0.2		0.4		0.6		0.8	
m	ARL	$SDRL$	ARL	$SDRL$	ARL	$SDRL$	ARL	$SDRL$
5	1.17	.51	7.0	10.62	56.41	103.25	245.11	324.91
10	1.14	.43	6.11	7.0	45.91	69.83	247.16	304.65
20	1.13	.40	5.71	5.85	40.82	47.93	233.76	273.54
30	1.13	.39	5.44	5.32	40.07	44.78	228.37	256.46
50	1.12	.36	5.49	5.21	38.59	41.06	223.43	242.44
100	1.12	.37	5.36	4.95	37.85	39.05	219.40	225.87
200	1.11	.36	5.40	4.92	37.99	38.81	217.81	217.89
500	1.11	.36	5.36	4.87	37.41	36.61	213.94	209.40
1000	1.12	.37	5.29	4.80	37.69	37.00	213.70	213.52
∞	1.12	.36	5.29	4.77	37.44	36.94	215.93	215.43

As in the case of three sigma limits this correlation depends only on m and n. In Table 3.6 we calculated the correlation for various combinations of m and n. From this Table we conclude again that as the sample size and the number of samples increases the correlation decreases. The recommendation for sample sizes and number of samples is the same as in the case of three sigma limits.

We computed the ARL and $SDRL$ values for several values of m and n via simulation along the same lines as in the three sigma limits. The number of samples and sample sizes considered were $m = 5, 10, 20, 30, 50, 100, 200, 500, 1000$ and $n = 5, 10, 20, 50$. The results are presented on Tables 3.7 − 3.10. From Tables 3.7 through 3.10 we deduce the following points. For upward shifts as m increases the ARL and $SDRL$ values generally decrease and approach their theoretical values. For downward shifts as m increases the same thing happens for $n = 20$, 50. For $n = 5$, 10 the ARL and $SDRL$ values do not follow a specific pattern. In the in-control state the ARL and $SDRL$ values increase until they get close to their theoretical values, which is in accordance with the results of Chen (1998). As an overall conclusion we can say that the ARL and $SDRL$ values behave in the same way except that as m increases the ARL is getting closer to the theoretical value faster than the $SDRL$.

Table 3.10. ARL and $SDRL$ values for the S (probability limits) chart when $n = 50$

	σ_1^2/σ_0^2									
	1		1.2		1.4		1.6		1.8	
m	ARL	$SDRL$	ARL	$SDRL$	ARL	$SDRL$	ARL	$SDRL$	ARL	$SDRL$
5	320.32	380.78	93.28	184.90	13.83	29.14	4.02	5.49	2.10	1.90
10	369.19	410.82	70.91	122.34	10.73	15.32	3.61	4.02	1.93	1.49
20	411.62	433.18	58.04	85.60	9.50	10.86	3.37	3.19	1.93	1.42
30	431.22	447.76	53.56	68.63	8.96	9.54	3.42	3.10	1.90	1.36
50	452.27	459.10	50.77	58.78	8.96	8.95	3.28	2.85	1.89	1.30
100	472.90	472.99	48.14	50.64	8.62	8.59	3.25	2.79	1.88	1.32
200	482.50	481.24	47.71	48.24	8.51	8.24	3.25	2.75	1.86	1.29
500	493.58	498.61	47.47	48.19	8.60	8.11	3.24	2.69	1.85	1.23
1000	490.32	499.04	47.59	47.66	8.56	8.10	3.23	2.66	1.86	1.27
∞	500.01	499.51	47.23	46.73	8.52	8.01	3.22	2.67	1.86	1.27

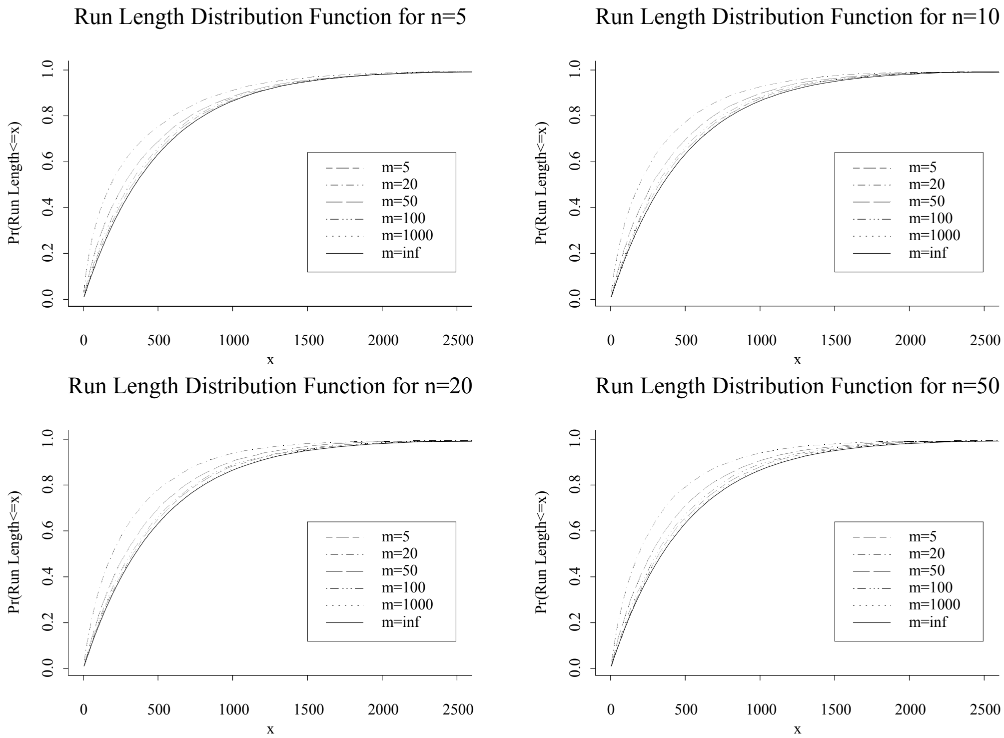
When we are in-control we need at least $m = 200$, otherwise the practitioner will face many false alarms whereas the value of n is not equally important. In the out-of-control situations the value of n is important for minimizing the effect of estimating S. Specifically, when $\sigma_1^2/\sigma_0^2 = 1.2$ the ARL values for $n = 5, 10, 20, 50$ are $239.29, 178.40, 117.98$ and 50.77 respectively. Therefore, we observe a dramatic reduction as n becomes larger. A similar situation occurs for downward shifts. Consequently, large values of n, larger than 20, are recommended. The effect of estimation is severe for $m \leq 20$, especially for small shifts. For values $30 \leq m \leq 50$ the effect is moderate and for values of 100 or larger the effect is small enough. When we have small downward shifts for $n = 5$, and for $n = 10$ when $m \leq 10$, the ARL and $SDRL$ values are larger than the corresponding in-control values. In such a situation it is better to use control charts for detecting small shifts like CUSUM and EWMA charts.

Table 3.10. (continued) ARL and $SDRL$ values for the S (probability limits) chart when $n = 50$

	σ_1^2/σ_0^2							
	0.2		0.4		0.6		0.8	
m	ARL	$SDRL$	ARL	$SDRL$	ARL	$SDRL$	ARL	$SDRL$
5	1	0	1.21	.62	7.47	11.58	107.80	188.75
10	1	0	1.19	.50	6.19	6.96	92.22	141.22
20	1	0	1.18	.47	5.86	5.90	79.17	102.14
30	1	0	1.16	.45	5.69	5.50	74.86	87.60
50	1	0	1.16	.45	5.65	5.46	71.97	78.71
100	1	0	1.15	.42	5.64	5.22	69.87	73.40
200	1	0	1.16	.43	5.62	5.18	69.61	69.69
500	1	0	1.15	.42	5.49	4.95	68.90	70.23
1000	1	0	1.15	.42	5.51	5.06	69.27	70.27
∞	1	0	1.16	.43	5.48	4.96	68.04	67.54

In Figure 3.2 we present the empirical run length distribution functions (ERL) for $n = 5, 10, 20, 50$. In each Figure we plot six different lines representing the ERL functions for $m = 5, 20, 50, 100, 1000$ and the theoretical run length distribution (inf). We see that as m increases the ERL approaches the theoretical run length distribution. Also, an increasing n value causes the ERL's for the m values to approach the theoretical run length distribution faster.

Figure 3.2. Empirical Run Length Distribution Functions for the probability limits chart



3.2.3 The X Chart for Monitoring Process Dispersion

Consider the control limits of Section 2.3.3. In order to assess the effect of the number of observations on the control limits of the X chart we performed a simulation study. The results are presented in Table 3.11. For each value in the Table, we simulated N values from a $N(\mu, \sigma_0^2)$ distribution, we computed the $\widehat{UCL}$ and $\widehat{LCL}$ and subsequently we generated values from a $N(\mu, \sigma_1^2)$ distribution until we obtained a value above $\widehat{UCL}$ or below $\widehat{LCL}$. The number of samples simulated up to the one that was outside the control limits constitutes one observation on the run length. This procedure was repeated 32000 times in order to get estimates of the values of ARL and $SDRL$.

Table 3.11. ARL and $SDRL$ values for the X control chart

	σ_1^2/σ_0^2									
	1		1.2		1.4		1.6		1.8	
N	ARL	$SDRL$	ARL	$SDRL$	ARL	$SDRL$	ARL	$SDRL$	ARL	$SDRL$
30	986.31	5024.83	315.36	1058.44	147.93	439.79	84.36	187.50	53.74	98.54
50	614.94	1565.0	229.95	476.60	116.69	200.50	69.61	107.23	47.23	66.81
75	503.75	948.78	202.02	318.54	105.18	150.77	64.51	84.15	43.99	54.87
100	467.07	770.60	190.53	274.54	100.73	131.39	61.98	75.26	42.78	50.48
200	413.88	518.65	173.68	205.96	93.86	105.77	58.63	63.56	40.67	42.81
300	398.94	476.34	167.79	187.69	92.76	100.47	57.93	61.37	41.26	42.29
500	387.38	429.45	167.90	179.39	90.34	93.58	56.80	58.96	39.69	40.54
1000	379.32	401.55	162.96	168.50	89.12	91.10	57.03	57.78	39.90	39.85
2000	372.64	383.71	162.70	166.87	89.45	89.41	56.35	55.82	39.62	39.17
∞	370.40	369.90	162.08	161.58	89.05	88.55	56.48	55.98	39.45	38.95

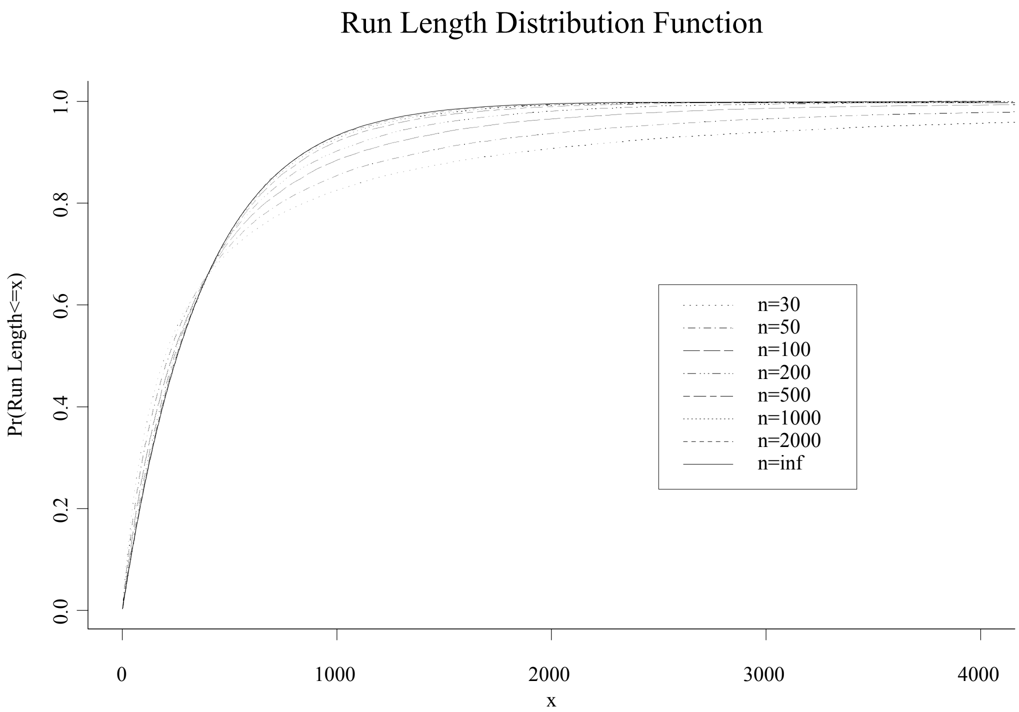
From Table 3.11 we see that we do not have results for downward shifts. This happens because a decreasing standard deviation will never cause a value below the lower control limit. The simulation reveals that the ARL and $SDRL$ values decrease until they approach their theoretical values. We need at least 300 observations to minimize the effect of estimation in the control limits of the X chart.

In Figure 3.3 we present the empirical run length distribution function (ERL) for $n = 30, 50, 100, 200, 500, 1000, 2000$ and the theoretical run length distribution (inf). The result is that as n increases the ERL approaches the theoretical run length distribution.

3.2.4 Discussion

In the rational subgroups case we propose larger n values than usual and someone may report that this is a problem. However, Woodall and Montgomery (1999) remarked that in industry now there are large data sets available in contrast to the past. Therefore, such values for the sample size should not be a problem, generally. On the other hand, if for some special applications this still remains a problem, the practitioner should keep in mind the great influence on the estimated control chart performance displayed on the tables of this work.

Figure 3.3. Empirical Run Length Distribution Functions for the X chart



3.3 Estimation Effect in the EWMA Chart

Jones et al. (2001), considered the problem of estimating the parameters of the EWMA chart in the normal case. They proved that if the random variable T is the run length of the EWMA chart, then the ARL of such a chart is given by

$$ARL = E\left[T|\gamma, \delta, u\right] = \int_{-\infty}^{\infty} \int_{0}^{\infty} M\left(w, z_0, \gamma, \delta, u\right) f_w\left(w\right) \phi\left(z_0\right) dw dz_0,$$

where $M\left(w, z_0, \gamma, \delta, u\right) = 1 + \frac{w}{r\gamma} \int_{-h}^{h} M\left(w, z_0, \gamma, \delta, v\right) \phi\left(\frac{w}{r\gamma}\left[v - \left(1-r\right)u\right] - \frac{\delta}{\gamma} + \frac{z_0}{\gamma\sqrt{m}}\right) dv$, $\gamma = \sigma/\sigma_0$, $\delta = \left(\mu - \mu_0\right)/\left(\sigma_0/\sqrt{n}\right)$, w, z_0 are specific values of the random variables $W = \widehat{\sigma}_0/\sigma_0$, $Z_0 = \sqrt{m}\frac{\left(\widehat{\mu}_0 - \mu_0\right)}{\left(\sigma_0/\sqrt{n}\right)}$ and u is the starting value of the EWMA. Also, μ_0, σ_0 are the in-control mean and standard deviation, $\widehat{\mu}_0$, $\widehat{\sigma}_0$ are their estimates respectively and μ, σ are the mean and standard deviation at time t. Additionally, the second moment of T is given by

$$E\left[T^2|\gamma, \delta, u\right] = \int_{-\infty}^{\infty} \int_{0}^{\infty} M_2\left(w, z_0, \gamma, \delta, u\right) f_w\left(w\right) \phi\left(z_0\right) dw dz_0,$$

where $M_2\left(w, z_0, \gamma, \delta, u\right) = 1 + \frac{2w}{r\gamma} \int_{-h}^{h} M\left(w, z_0, \gamma, \delta, v\right) \phi\left(\frac{w}{r\gamma}\left[v - \left(1-r\right)u\right] - \frac{\delta}{\gamma} + \frac{z_0}{\gamma\sqrt{m}}\right) dv + \frac{w}{r\gamma} \int_{-h}^{h} M_2\left(w, z_0, \gamma, \delta, v\right) \phi\left(\frac{w}{r\gamma}\left[v - \left(1-r\right)u\right] - \frac{\delta}{\gamma} + \frac{z_0}{\gamma\sqrt{m}}\right) dv$. The SDRL can be computed by $SDRL = \sqrt{E\left[T^2\right] - \left(E\left[T\right]\right)^2}$. Jones et al. (2001) concluded that in both in-control and out-of-control cases the process's run length performance is affected. In particular, the estimation effect results in more false alarms and generally leads to a reduction of the ability of the chart to detect process shifts.

Additionally, Jones (2002) developed a procedure for designing an EWMA chart with estimated parameters. Using this procedure a practitioner is able to design an EWMA chart to have the desirable performance. The steps of this method are

Step 1. Identify the desired in-control ARL of the chart

Step 2. Determine the subgroup size n and number of subgroups m that will be used to estimate the parameters of the in-control process. Obtain a reference sample of m

subgroups, of n observations each

Step 3. Ensure that the reference sample is representative of the in-control state of the process. Estimate the parameters according to $\widehat{\mu}_0 = \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n} X_{ij}$ and $\widehat{\sigma}_0 = \frac{S_p}{c_{4,m}}$ where $S_p = \sqrt{\frac{\sum_{i=1}^{m}\sum_{j=1}^{n}\left(X_{ij}-\overline{X}_{i.}\right)^2}{m(n-1)}}$ and $c_{4,m} = \frac{\sqrt{2}\Gamma\left(\frac{m(n-1)+1}{2}\right)}{\sqrt{m(n-1)}\Gamma\left(\frac{m(n-1)}{2}\right)}$.

Step 4. Select the smoothing constant λ.

Step 5. Using λ from Step 4, identify the constant L that produces an EWMA chart with the desired in-control ARL.