A Pattern Test for Distinguishing Between Autoregressive and Mean-Shift Data

WAYNE A. TAYLOR
Baxter Healthcare Corporation, Round Lake, IL 60073

Statistical methods such as control charts and change-point analysis are commonly used to determine whether the mean has shifted. Such methods assume independent errors around a possibly changing mean. When such techniques are applied to autoregressive data, erroneous conclusions can result. However, shifts of the mean create autocorrelation between the observations making it difficult to distinguish mean-shift data from autoregressive data. A pattern test has been devised that can reliably distinguish between these two important cases.

Introduction

Look at Figures 1-3. Which two sets of data are most similar in structure?

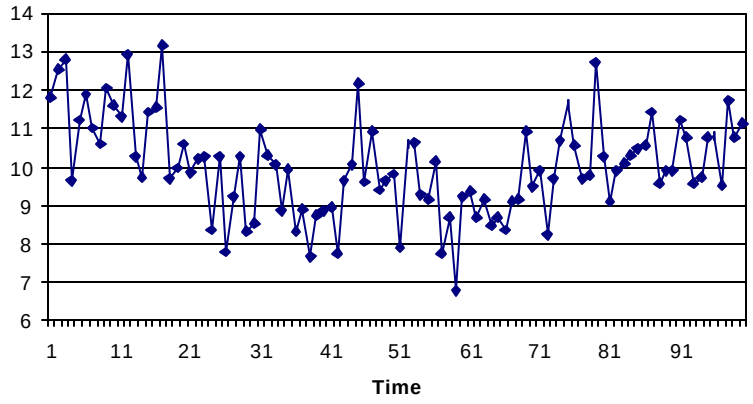

Figure 1: Mean-Shift Model

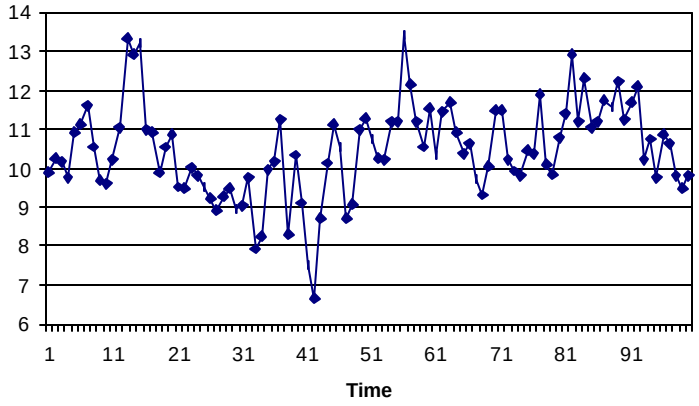

Figure 2: First Order Autoregressive Model - Positive Correlation

Dr. Taylor is Director Quality Technologies and Head of Baxter's Six Sigma Program. He is a Fellow of ASQ. His email address is wayne@variation.com.

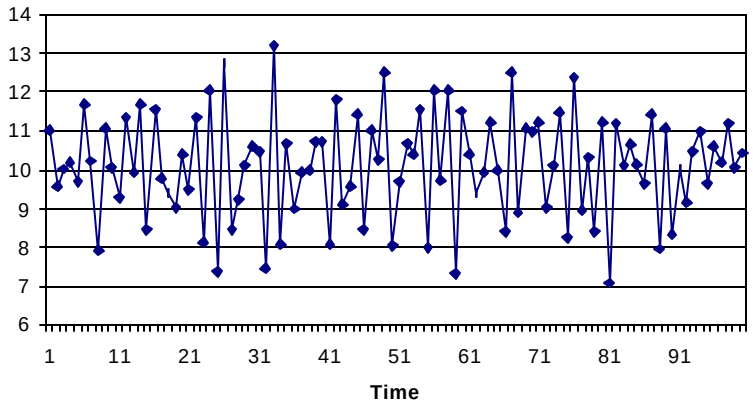

Figure 3: First Order Autoregressive Model - Negative Correlation

Would you be surprised to find out it is the plots in Figures 2 and 3? Both were generated using a first order autoregressive model. The plot in Figure 1 was generated using a different model, called the mean-shift model. When analyzing data collected over time, it is important to be able to distinguish between these two important cases. Visual inspection of such data is unreliable. A pattern test has been developed which can reliably distinguish between these two models.

The Mean-Shift Model

Statistical methods such as control charts and change-point analysis assume a series of independent observations collected over time. At one or more points in time the mean may shift. Let X_1, X_2, ... represent the data in time order. The mean-shift model can be written as

$$X_i = \mu_i + \varepsilon_i$$

where μ_i is the average at time i Generally $\mu_i = \mu_{i-1}$ except for a small number of values of i called the change-points. ε_i is the random error associated with the i-th value. It is assumed that the ε_i are independent and identically distributed with means of zero. Other assumptions including normality may also be made by some of these statistical methods but are not required for the proposed pattern test.

The data shown in Figure 1 was generated using the following model:

$\varepsilon_i \sim N(0,1)$ and independent
$\mu_1, \mu_{21}, \mu_{41}, \mu_{61}, \mu_{81} \sim N(10,1)$ and independent
For all other i, $\mu_i = \mu_{i-1}$

$N(\mu,\sigma)$ means normally distributed with mean μ and standard deviation σ. This model could result from a process where the mean shifts as a result of periodic material changes. It could also result from a process subject to both setup and within setup variation. In other cases, the mean-shifts could occur at random times. The proposed pattern test works for any of these situations.

The First Order Autoregressive Model

The data shown in Figures 2 and 3 were generated using the first order autoregressive model:

$\varepsilon_i \sim N(0,1)$ and independent
$r_i = \phi\, r_{i-1} + \varepsilon_i$
$r_0 = 0$
$X_i = 10 + r_i$

ϕ is a constant between -1 and 1. The above model results in a correlation between successive values of:

$$\text{Corr}\{X_i, X_{i-1}\} = \phi$$

Values of ϕ=0.7 and ϕ=-0.7 were used respectively in Figures 2 and 3. When ϕ=0, the autoregressive model reduces to what is called the white noise model where $X_i \sim N(10,1)$ and independent. This is also a special case of the mean-shift model with no shifts.

When checking for an autoregressive model, one frequently calculates the autocorrelations and displays them in the form of a correeogram. However, this is only useful for distinguishing between an autoregressive model and white noise. The mean-shift model also results in autocorrelations between the values. In Figure 1 the correlation between consecutive values is 0.43. Looking at the autocorrelations will not allow one to distinguish between these two models.

The Pattern Test

Figure 4 shows the six possible patterns that can result from plotting three consecutive points when there are no ties. Pattern 1 is called the double up pattern and Pattern 6 is called the double down pattern. The other 4 patterns will be referred to as reversal patterns. For the autoregressive model, the double up and double down patterns are most common when there is a positive autocorrelation as in Figure 2. The reversal patterns are most common when there is a negative correlation as in Figure 3.

When the means of the 3 points are the same, all six patterns are equally likely. In this case, the double up and double down patterns should occur 1/3 the time and the reversal patterns should occur 2/3 of the time. The pattern test involves counting the number of times the double up/down patterns occur. This count is slightly biased when the mean shifts or there is an outlier. However the bias is small and easily compensated for making this count useful for distinguishing between mean-shift and autoregressive data. If this count is significantly greater than a third the number of values, the data is autoregressive with positive correlation. If this count is significantly less than a third, the data is autoregressive with negative correlation. Otherwise the mean-shift model fits the observed data.

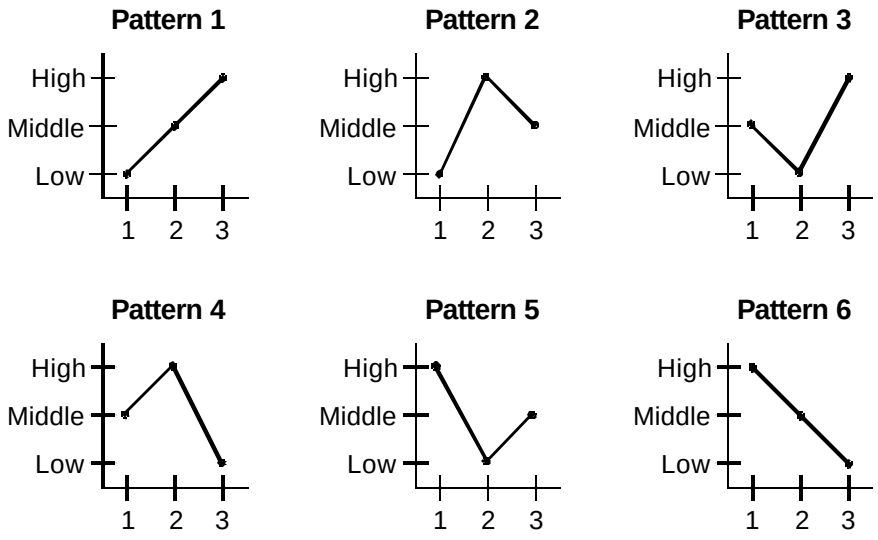

Figure 4: Six Patterns for Three Consecutive Points

Table 1 gives critical values for S for a 2-sided test with α=0.05 for n between 10 and 200. If $S \le s_{lower}$, the data is autocorrelated with negative correlation. If $S \ge s_{upper}$, the data is autocorrelated with positive correlation. Otherwise, the data is consistent with the mean-shift model. These critical values and the approximations given below are all based on the assumption that the number of shifts and outliers is less than 1 per 20 data points. This assumption should rarely restrict the use of this procedure.

Formulas 1 and 2 can also be used to calculate significance levels. If $\alpha_{lower} \le 0.025$, the data is autocorrelated with negative correlation. If $\alpha_{upper} \le 0.025$, the data is autocorrelated with positive correlation. Otherwise, any correlation in the data is the result of mean shifts.

$$\alpha_{lower} \approx 1 - I_{p_{lower}}(a_{lower}, b_{lower}) \tag{1}$$

$$\text{where} \quad p_{lower} = \frac{14n-31}{30n-60}, \quad a_{lower} = S+1 \quad \text{and} \quad b_{lower} = \frac{n-2}{3p_{lower}} - S$$

$$\alpha_{upper} \approx I_{p_{upper}}(a_{upper}, b_{upper}) \tag{2}$$

$$\text{where} \quad p_{upper} = \frac{147n-310}{315n-600}, \quad a_{lower} = S \quad \text{and} \quad b_{lower} = \frac{21n-40}{60p_{upper}} - S+1$$

$I_p(a,b)$ is the incomplete beta function. The derivation of these formulas is given in Appendix A. They are within 2% of the true value for $0.01 \le \alpha \le 0.1$ and $n \ge 10$. Formulas 3 and 4 give a second less accurate approximation that can be used when $n \ge 100$.

Table 1: Two-Sided Critical Values for S = Number of Double Up/Down Patterns (α=0.05)

n	s_{lower}	s_{upper}
10	0	6
11	0	6
12	0	7
13	0	7
14	1	8
15	1	8
16	1	9
17	1	9
18	1	9
19	2	10
20	2	11
21	2	11
22	2	11
23	3	12
24	3	13
25	3	13
26	3	13
27	4	14
28	4	14
29	4	14
30	4	15
31	4	15
32	5	16
33	5	16
34	5	16
35	6	17
36	6	17
37	6	18
38	6	18
39	7	19
40	7	19
41	7	20
42	7	20
43	8	21
44	8	21
45	8	21
46	9	22
47	9	22
48	9	22
49	9	23
50	9	23
51	10	24
52	10	24
53	10	24
54	11	25
55	11	25
56	11	25
57	12	26

n	s_{lower}	s_{upper}
58	12	26
59	12	27
60	12	27
61	13	28
62	13	28
63	13	28
64	13	29
65	14	30
66	14	30
67	14	30
68	15	31
69	15	31
70	15	31
71	16	32
72	16	32
73	16	32
74	16	33
75	16	33
76	17	34
77	17	34
78	17	34
79	18	35
80	18	35
81	18	36
82	18	36
83	19	37
84	19	37
85	19	37
86	20	38
87	20	38
88	20	38
89	21	39
90	21	39
91	21	40
92	21	40
93	22	41
94	22	41
95	22	41
96	23	42
97	23	42
98	23	42
99	24	43
100	24	44
101	24	44
102	24	44
102	25	45
104	25	45
105	25	45

n	s_{lower}	s_{upper}
106	26	46
107	26	46
108	26	46
109	27	47
110	27	47
111	27	47
112	27	48
113	27	48
114	28	49
115	28	49
116	28	49
117	29	50
118	29	50
119	29	50
120	30	51
121	30	52
122	30	52
123	30	52
124	31	53
125	31	53
126	31	53
127	32	54
128	32	54
129	32	54
130	33	55
131	33	55
132	33	55
133	34	56
134	34	57
135	34	57
136	34	57
137	35	58
138	35	58
139	35	58
140	36	59
141	36	59
142	36	60
143	37	60
144	37	61
145	37	61
146	37	61
147	38	62
148	38	62
149	38	62
150	39	63
151	39	63
152	39	63
153	40	64

n	s_{lower}	s_{upper}
154	40	64
155	40	64
156	41	65
157	41	65
158	41	65
159	41	66
160	42	67
161	42	67
162	42	67
163	43	68
164	43	68
165	43	68
166	44	69
167	44	69
168	44	70
169	44	70
170	45	71
171	45	71
172	45	71
173	46	72
174	46	72
175	46	72
176	46	72
177	47	73
178	47	73
179	47	73
180	47	74
181	48	75
182	48	75
183	48	75
184	49	76
185	49	76
186	49	76
187	50	77
188	50	77
189	50	77
190	51	78
191	51	78
192	51	78
193	52	79
194	52	80
195	52	80
196	52	80
197	53	81
198	53	81
199	53	81
200	54	82

Note: n = sample size. If $S \leq s_{lower}$, the data is autocorrelated with negative correlation. If $S \geq s_{upper}$, the data is autocorrelated with positive correlation. Otherwise, the data is consistent with the mean-shift model.

$$\alpha_{lower} \approx \Phi\left(\frac{3S - n + 3.5}{\sqrt{1.6n - 2.9}}\right) \tag{3}$$

$$\alpha_{upper} \approx 1 - \Phi\left(\frac{3S - 1.05n + 0.5}{\sqrt{1.68n - 2.9}}\right) \tag{4}$$

Applications of the Pattern Test

Table 2 shows the results of applying the pattern test to the three sets of generated data in Figures 1-3 plus the three real sets of data shown in Figures 5-7. In Figures 1-3, n=100 resulting in critical values s_{lower}=24 and s_{upper}=44. For the mean-shift data in Figure 1, S=38 which falls between the two critical values. This is consistent with a mean-shift model. For the Figure 2 autoregressive data with positive correlation, S=46. This exceeds the upper critical value proving the data is not consistent with a mean-shift model. For the Figure 3 autoregressive data with negative correlation, S=19. This is below the lower critical value again proving the data is not consistent with a mean-shift model. The α values from Equations 1-4 support these same conclusions. Also shown are the true α values obtained through simulation. All four approximations are accurate to three digits when n=100.

Table 2: Analysis of Example Data Sets

Fig.	Model	n	S	s_{lower}	s_{upper}	α_{lower} true	α_{lower} (Eq. 1)	α_{lower} (Eq. 3)	α_{upper} true	α_{upper} (Eq. 2)	α_{upper} (Eq. 4)
1	Mean-Shift	100	38	24	44	0.9187	0.9185	0.9187	0.2300	0.2296	0.2298
2	Autoregressive - Positive	100	46	24	44	0.9995	0.9996	0.9995	0.0047	0.0045	0.0046
3	Autoregressive - Negative	100	19	24	44	0.0007	0.0007	0.0008	0.9999	0.9999	0.9999
5	Number Sunspots	50	38	9	23	1.0000	1.0000	1.0000	0.0000	0.0000	0.0000
6	Batch Yields	70	9	15	31	0.0001	0.0000	0.0000	1.0000	1.0000	1.0000
7	Part Strength	52	19	10	24	0.8294	0.8286	0.8286	0.3491	0.3499	0.3509

Figure 5 shows the number of sunspots for a 50 year period of time. This data is Series E from Box and Jenkins (1976). The number of double up/down patterns is S=38. This exceeds the upper critical value s_{upper}=23 indicating the data is autoregressive with positive correlation. The α values from Equations 1-4 support this same conclusion.

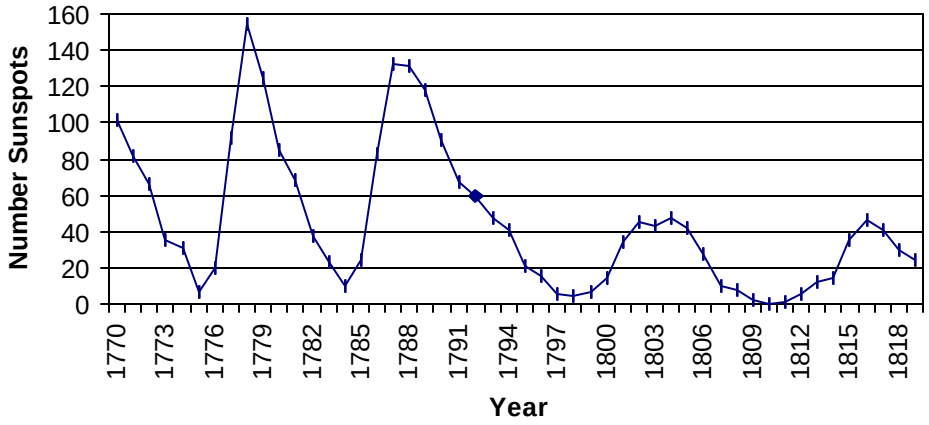

Figure 5: Wölfer Sunspot Data

Figure 6 shows the yields from 70 consecutive batches of a chemical process. This data is Series F from Box and Jenkins (1976). The number of double up/down patterns is S=9. This is below the lower critical value s_{lower}=15 indicating the data is autoregressive with negative correlation. The α values from Equations 1-4 support this same conclusion.

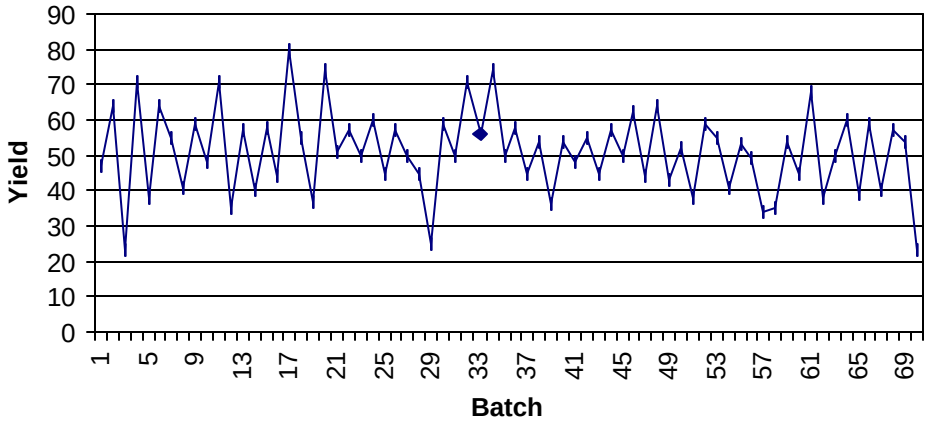

Figure 6: Batch Yields

Figure 7 shows part strength readings taken once an hour over 52 consecutive hours. The number of double up/down patterns is S=19. This is between the lower critical value s_{lower}=10 and the upper critical value s_{upperr}=24 indicating the data is consistent with the mean-shift model. The α values from Equations 1-4 support this same conclusion.

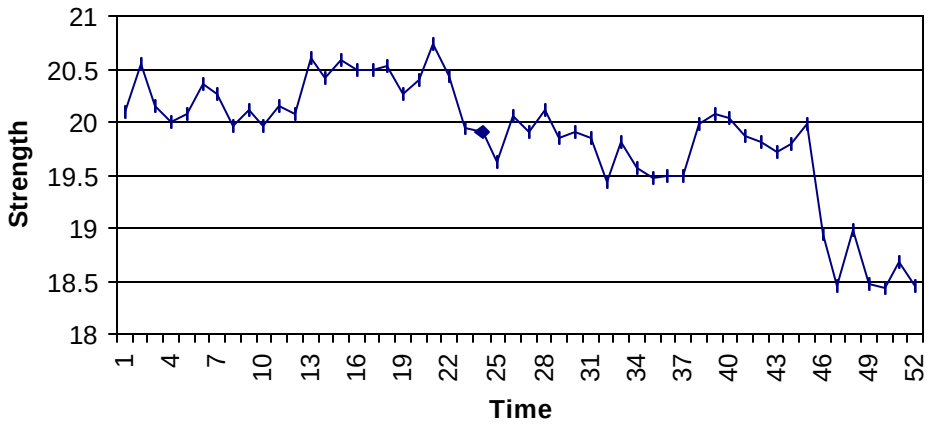

Figure 7: Part Strength

Handling Ties

When ties are possible, two new patterns can occur: the single tie and the double tie. In this case, let Pi be defined in terms of X_{i-2}, X_{i-1}, X_i as follows:

$$P_i = \begin{cases} 1 & \text{double up/down pattern} \\ \frac{1}{2} & \text{single tie pattern} \\ \frac{1}{3} & \text{double tie pattern} \\ 0 & \text{reversal pattern} \end{cases}$$

Further, let S be defined as:

$$S = \sum_{i=3}^{n} P_i$$

When X_{i-2}, X_{i-1}, X_i are identically distributed, $E\{P_i\} = 1/3$. Again a test for autoregression can be constructed based on S averaging above or below 1/3 the number of patterns. If the number of ties is small, Table 1 and Equations 1-4 may still be used. But if ties are more common, Table 1 and Equations 1-4 can no longer be used because the ties reduce the variation of S. Instead Equations 5-8 should be used:

$$\alpha_{lower} \approx 1 - I_{p_{lower}}(a_{lower}, b_{lower}) \qquad (5)$$

$$\text{where} \quad p_{lower} = 1 - \frac{3[(n-2)\mathrm{Var}\{P_i\} + 2(n-3)\mathrm{Cov}\{P_i, P_{i+1}\} + 2(n-4)\mathrm{Cov}\{P_i, P_{i+2}\}]}{n-2},$$

$$a_{lower} = S + 1 \quad \text{and} \quad b_{lower} = \frac{n-2}{3p_{lower}} - S$$

$$\alpha_{upper} \approx I_{p_{upper}}\left(a_{upper}, b_{upper}\right) \tag{6}$$

$$\text{where} \quad p_{upper} = 1 - \frac{60\left[(n-2)\mathrm{Var}\{P_i\} + 2(n-3)\mathrm{Cov}\{P_i, P_{i+1}\} + 2(n-4)\mathrm{Cov}\{P_i, P_{i+2}\}\right]}{21n-40},$$

$$a_{lower} = S \quad \text{and} \quad b_{lower} = \frac{21n-40}{60\,p_{upper}} - S + 1$$

$$\alpha_{lower} \approx \Phi\left(\frac{S - \frac{n}{3} + \frac{7}{6}}{\sqrt{(n-2)\mathrm{Var}\{P_i\} + 2(n-3)\mathrm{Cov}\{P_i, P_{i+1}\} + 2(n-4)\mathrm{Cov}\{P_i, P_{i+2}\}}} \right) \tag{7}$$

$$\alpha_{upper} \approx 1 - \Phi\left(\frac{S - \frac{7n}{20} + \frac{1}{6}}{\sqrt{(n-2)\mathrm{Var}\{P_i\} + 2(n-3)\mathrm{Cov}\{P_i, P_{i+1}\} + 2(n-4)\mathrm{Cov}\{P_i, P_{i+2}\}}} \right) \tag{8}$$

Estimates of $\mathrm{Var}\{P_i\}$, $\mathrm{Cov}\{P_i,P_{i+1}\}$ and $\mathrm{Cov}\{P_i,P_{i+2}\}$ can be obtained from the data. A special case with numerous ties is pass/fail data. In this case:

$$X_i = \begin{cases} 1 & \text{with probabilit y } p \\ 0 & \text{with probabilit y } (1-p) \end{cases}$$

Then:

$$P_i = \begin{cases} 1 & \text{with probabilit } y = 0 \\ ½ & \text{with probabilit } y = p(1-p)^2 + (1-p)^2 p + p^2(1-p) + (1-p)p^2 \\ ⅓ & \text{with probabilit } y = p^3 + (1-p)^3 \\ 0 & \text{with probabilit } y = p(1-p)p + (1-p)p(1-p) \end{cases}$$

This gives:

$$E\{P_i\} = ⅓\left[p^3 + (1-p)^3 + 3p^2(1-p) + 3(1-p)^2 p\right] = ⅓\left[p + (1-p)\right]^3 = ⅓$$

For pass/fail data, the variance and covariances of P_i are:

$$\mathrm{Var}\{P_i\} = \frac{1}{6}p(1-p) \tag{9}$$

$$\mathrm{Cov}\{P_i, P_{i+1}\} = -\frac{1}{9}p(1-p)\left[p^2 - 3p(1-p) + (1-p)^2\right] \tag{10}$$

$$\mathrm{Cov}\{P_i, P_{i+2}\} = \frac{1}{36}p(1-p)\left[p^3 - p^2(1-p) - p(1-p)^2 + (1-p)^3\right] \tag{11}$$

For pass/fail data, an estimate of p can be obtained from the data and substituted into Equations 9-11 to estimate $Var\{P_i\}$, $Cov\{P_i,P_{i+1}\}$ and $Cov\{P_i,P_{i+2}\}$. These estimates can then be plugged into Equations 5-8 to obtain approximate α levels.

Other Applications of P_i

An example of a data set with ties is shown in Figure 8. 197 chemical concentrations are shown. This data is Series A from Box and Jenkins (1976).

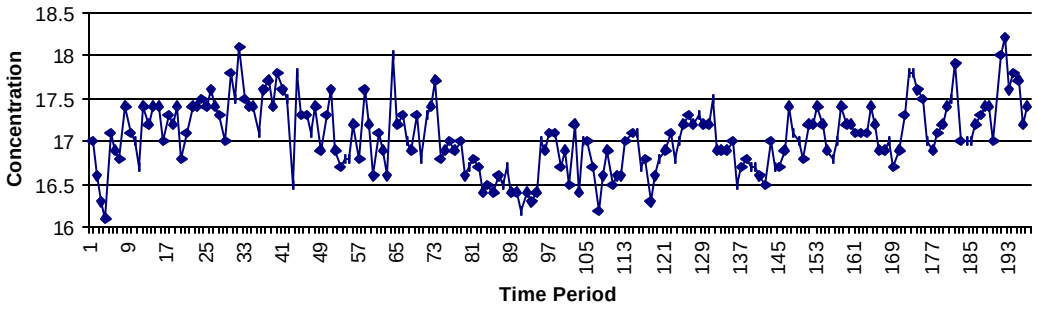

Figure 8: Chemical Concentration Data

From this data P_3, ..., P_{197} can be calculated. The P_i values are time ordered data that reacts to changes in the autoregressive behavior of the data. A CUSUM chart of the P_i values is shown in Figure 9. The sudden change in direction in the CUSUM chart indicates a sudden change in the autoregressive behavior of this data.

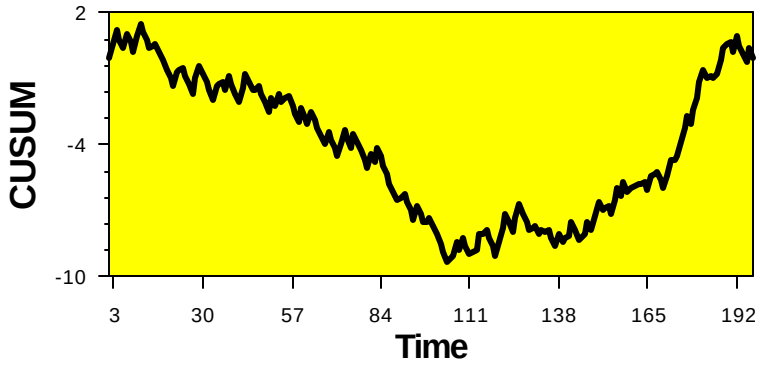

Figure 9: CUSUM Chart of P_i for Chemical Concentration Data

A change-point analysis was then performed on the P_i using Taylor (2000). This software performs a bootstrap analysis on the CUSUM chart to obtain confidence levels and confidence intervals for the change. The results of this analysis are shown in Figure 10. It verifies a change occurred with 98% confidence. The change is estimated to have occurred just prior to point 145. With 95% confidence it occurred between points 83 and 179.

Table of Significant Changes for Pi

Confidence Level = 90%, Confidence Interval = 95%, Bootstraps = 1000, Sampling Without Replacement

Row	Confidence Interval	Conf. Level	From	To	Level
145	(83, 179)	98%	0.32629	0.54088	1

Figure 10: Results of Change-Point Analysis of P_i for Chemical Concentration Data

The average P_i before the change is 0.326, which is close to 1/3, indicating a lack of autoregressive behavior. The average P_i following the change is 0.542 indicating autoregression with a positive correlation. Separate tests for autoregression were performed on points 1-144 and points 1405-197. The results are shown in Table 3. These tests confirm that following the change, the data is autoregressive with positive correlation, while before the change, the data is consistent with the mean-shift model.

Table 3: Pattern Test for Chemical Concentration Data

Points	n	S	α_{lower} (Eq. 5)	α_{lower} (Eq. 7)	α_{upper} (Eq. 6)	α_{upper} (Eq. 8)
1-144	144	47.33	0.4358	0.4442	0.8624	0.8631
145-197	53	26.67	1.0000	1.0000	0.0000	0.0000

Conclusion

The pattern test has proven to be useful for distinguishing between two very important models: the mean-shift model and the first order autoregressive model. The pattern test can be used to detect a violation of the assumption of independent errors when control charting data and performing a change-point analysis. The series P_i can also be used to detect changes in the autoregressive behavior of the data. It provides a useful new tool for helping to analyze complicated time series data.

Appendix A

The distribution of the test statistic S will be derived assuming no mean shifts or ties. Assume that a series of n data points X_1, X_2, ..., X_n has been collected in time order. Let P_i be an indicator function of whether the double up/down pattern occurred for points X_{i-2}, X_{i-1}, X_i. Further let:

$$S = \sum_{i=3}^{n} P_i$$

The average and variance of S are:

$$E\{S\} = \sum_{i=3}^{n} E\{P_i\} \tag{12}$$

$$Var\{S\} = \sum_{i=3}^{n} Var\{P_i\} + 2\sum_{i=3}^{n-1} Cov\{P_i, P_{i-1}\} + 2\sum_{i=3}^{n-2} Cov\{P_i, P_{i-2}\} \tag{13}$$

Assuming no ties or mean shifts, the P_i are identically distributed with:

$E\{P_i\} = 1/3$
$Var\{P_i\} = 2/9$
$Cov\{P_i,P_{i+1}\} = -1/36$
$Cov\{P_i,P_{i+2}\} = 1/180$

All other covariances are zero. The above moments were calculated by generating the 5!=120 possible patterns for 5 points. Substituting the moments of P_i into Equations 12 and 13 gives the following moments for S:

$$E\{S\} = \frac{n-2}{3} \tag{14}$$

$$\begin{aligned} Var\{S\} &= \left[(n-2)\frac{2}{9} + 2(n-3)\frac{-1}{36} + 2(n-4)\frac{1}{180}\right] \\ &= \frac{16n-29}{90} \end{aligned} \tag{15}$$

When the mean shifts between time i-1 and i, the following values change:

$E\{P_i\} = E\{P_{i+1}\} = 1/2$
$Var\{P_i\} = Var\{P_{i+1}\} = 1/4$
$Cov\{P_{i-1},P_i\} = 0$
$Cov\{P_i,P_{i+1}\} = 0$
$Cov\{P_{i+1},P_{i+2}\} = 0$
$Cov\{P_{i-2},P_i\} = 0$
$Cov\{P_{i-1},P_{i+1}\} = 0$
$Cov\{P_i,P_{i+2}\} = 0$
$Cov\{P_{i+1},P_{i+3}\} = 0$

All other values are as before. The above moments were calculated by generating the $(4!)^2$= 576 possible patterns for 8 points where the first 4 points are all less than the last four points. Let t be the number of shifts. When t shifts occur:

$$E\{S\} = (n-2-2t)\frac{1}{3} + (2t)\frac{1}{2} = \frac{n+t-2}{3} \tag{16}$$

$$\begin{aligned}\mathrm{Var}\{S\} &= \left[(n-2t-2)\frac{2}{9}+2t\frac{1}{4}+2(n-3t-3)\frac{-1}{36}+2(n-4t-4)\frac{1}{180}\right] \\ &= \frac{16n+16t-29}{90}\end{aligned} \tag{17}$$

Shifts increase both E{S} and Var{S}. To see what effect this has on the critical values, take E{S} ± 2 SD{S} as an approximate critical values. Both upper and lower critical values increase as t increases. Figure 11 shows the percentage increase in these approximate critical values as t ranges from 0% to 10% of n. When t is 5% of n, i.e. a change occurs once every 20 points, the critical values increase only 5%.

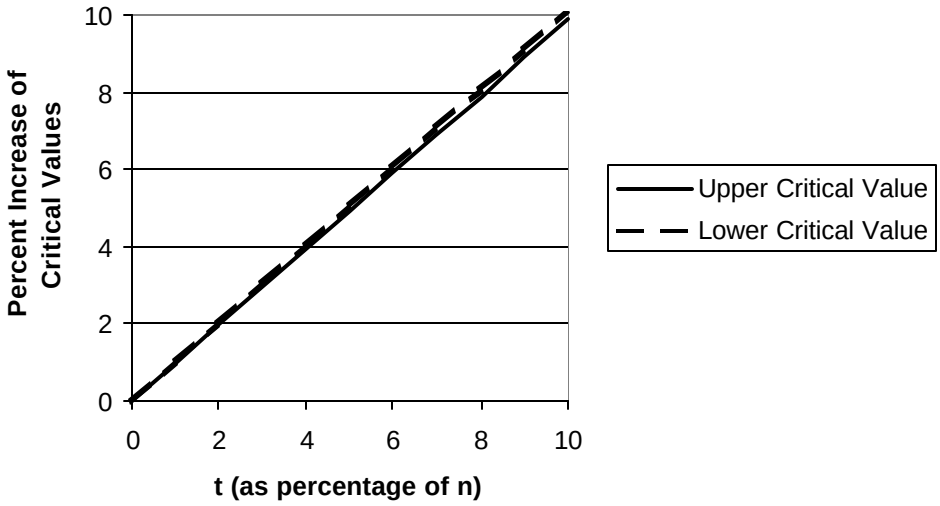

Figure 11: Approximate Percent Increase in Critical Values As t Increases

Since the number of changes is not known, one cannot exactly determine the distribution of S. However, by assuming an upper bound on the number of changes, one can bound its distribution. It would seem reasonable to expect no more than one change per twenty points ($t \le n/20$). A lower critical value is then calculated based on t=0 changes while the upper critical value is based on t=n/20 changes.

If the P_i where uncorrelated, S would follow the binomial distribution. Since the correlations are small, one would expect the binomial distribution to provide a close approximation. The binomial distribution $B(x|n_b,p_b)$ has parameters n_b and p_b. It has a mean of n_bp_b and variance $n_bp_b(1-p_b)$. Setting $E\{S\} = n_bp_b$ and $\mathrm{Var}\{S\} = n_bp_b(1-p_b)$ and solving for n_b and p_b gives:

$$p_b = 1-\frac{\mathrm{Var}\{S\}}{E\{S\}} = \frac{14n+14t-31}{30(n+t-2)} \tag{18}$$

$$n_b = \frac{E\{S\}}{p_b} = \frac{n+t-2}{3p_b} = \frac{10(n+t-2)^2}{14n+14t-31} \tag{19}$$

Since n_b may not be an integer as required by the binomial distribution, the more general incomplete Beta function, $I_p(a,b)$, will be used. Assuming t changes, the upper and lower significance levels for S can be approximated by:

$$\alpha_{lower} \approx B(S \mid n_b, p_b) = 1 - I_{p_b}(S+1, n_b - S) \tag{20}$$

$$\alpha_{upper} \approx 1 - B(S-1 \mid n_b, p_b) = I_{p_b}(S, n_b - S + 1) \tag{21}$$

Equation 1 was obtained from Equation 20 by substituting Equations 18 and 19 and setting t=0. Equation 2 was derived from Equation 21 by substituting Equations 18 and 19 and setting t=n/20. Equation 5 was obtained from Equation 20 by substituting Equations 13 and 16 and setting t=0. Equation 6 was derived from Equation 21 by substituting Equations 13 and 16 and setting t=n/20. Simulations indicate that Equations 20 and 21 are accurate to within 2% of the true value for $0.01 \le \alpha \le 0.1$ and $n \ge 10$.

A second less accurate estimate can be obtained by approximating the distribution of S using the normal distribution with continuity correction. This results in Equations 22 and 23. Equation 3 was derived from Equation 22 by substituting Equations 16 and 17 and setting t=0. Equation 4 was derived from Equation 23 by substituting Equations 16 and 17 and setting t=n/20. Equation 7 was derived from Equation 22 by substituting Equations 13 and 16 and setting t=0. Equation 8 was derived from Equation 23 by substituting Equations 13 and 16 and setting t=n/20. These approximations should only be used when $n \ge 100$.

$$\alpha_{lower} \approx \Phi\left(\frac{S + 0.5 - E\{S\}}{\sqrt{Var\{S\}}}\right) \tag{22}$$

$$\alpha_{upper} \approx 1 - \Phi\left(\frac{S - 0.5 - E\{S\}}{\sqrt{Var\{S\}}}\right) \tag{23}$$

References

Box, George E. P. and Jenkins, Gwilym (1976). Time Series Analysis: Forecasting and Control, Holden-Day, San Francisco, California.

Taylor, Wayne (2000). Change-Point Analyzer 2.0 software package, Taylor Enterprises, Libertyville, Illinois. WEB: http://www.variation.com/cpa

Key Words: Mean-Shift, Autoregression, Change-Point Analysis, Control Chart, Time Series