Univariate Temporal Methods

“At the present time, even a good surveillance system is perhaps best viewed as a ‘smoke detector.’ They may pick up important public health diagnoses or syndromes, but only labor-intensive follow-up investigations will distinguish genuine fires from backyard barbecues.”

Michael Grey & Kenneth Spaeth (2006, P. 76)

This chapter describes univariate temporal methods useful for early event detection (EED) and how to appropriately apply them to the biosurveillance problem. Biosurveillance systems often apply variants of standard univariate statistical process control (SPC) methods – the Shewhart, cumulative sum (CUSUM), and exponentially weighted moving average (EWMA) methods – for early event detection. A challenge in applying these methods to biosurveillance is the data often violates classical SPC assumptions, particularly the assumptions of normality and independent and identically distributed observations.

When applied to biosurveillance, the Shewhart, CUSUM, and EWMA methods should be tailored to the biosurveillance problem and data. For example, biosurveillance is generally only focused on detecting increases in disease incidence and thus the methods should be designed only to signal for increases. (In SPC parlance, these are “one-sided control chart” problems.) In addition, because biosurveillance data frequently contains various systematic effects, the methods should be applied to the residuals from a model designed to remove such effects, not the raw data itself. By making these and other adjustments, the classical SPC methods can often be appropriately applied to biosurveillance and further, in so doing, EED performance can often be improved.

This chapter begins by describing the Historical Limits method, a method commonly used in public health surveillance that is a particular variant of the Shewhart method. It then proceeds to describe and develop the Shewhart, CUSUM, and EWMA methods, both in an SPC context and then in a biosurveillance context. In the SPC context, these methods are referred to as control charts while in the biosurveillance context they are referred to as EED methods. For each method the control chart is first introduced, followed by a
discussion of how to modify and implement the method for biosurveillance. Following this, the chapter then discusses other methods that are commonly used in health surveillance, including historical limits and the Early Aberra-
tion Reporting System (EARS) EED methods – both of which are actually variants of the Shewhart method, as well as scan statistics.

The discussions and examples in this chapter focus on the application of the EED methods to data that follow specific probability distributions. While somewhat abstract, in so doing the discussions make clear how to appropriately tailor the EED methods particular types of data. Chapter 9 then demonstrates how to apply the methods to actual biosurveillance data.

CHAPTER OBJECTIVES

Upon completion of this chapter, you should be able to:

• Discuss Shewhart, CUSUM, and EWMA univariate temporal methods, including
  – specifying their traditional SPC formulations;
  – describing how they are modified for biosurveillance applications;
  – determining how to appropriately parameterize them for implementation; and,
  – applying them to achieve desired ATFS, CED, and PSD performance.

• Describe the historical limits, temporal scan statistics, and EARS methods, including
  – discussing how the methods are or can be used for surveillance and biosurveillance;
  – describing how the performance of these methods compare to the Shewhart, CUSUM, and EWMA; and,
  – explaining why the EARS’ C1 and C2 methods are not CUSUMs.
MATHEMATICAL NOTATION

$C$ CUSUM statistic  
$D$ Duration of outbreak  
$E$ EWMA statistic  
$\mathbb{E}(Y)$ Expected value of $Y$  
$f_0, f_1$ Probability density functions (pdfs)  
$F_0, F_1$ Cumulative distribution functions (cdfs)  
$F_t$ Outbreak cdf at time $t$  
$F_t^{-1}$ Inverse cdf  
$h$ Threshold  
$k$ Reference value parameter of CUSUM  
$L$ Parameter for setting EWMA threshold  
$L_t$ Log likelihood ratio at time $t$  
$\lambda$ Smoothing parameter of EWMA  
$\lambda_0, \lambda_1$ Poisson distribution parameters under $F_0$ and $F_1$, respectively  
$M$ Magnitude of outbreak  
$\mu_i$ Expected value of random variable with distribution $F_i$  
$\hat{\mu}_i$ Estimated mean of random variable with distribution $F_i$  
$\hat{\mu}_{i,j,k}$ Historical limits sample mean for reportable disease $i$, in week $j$ and year $k$  
$\text{NBin}(r, l)$ Negative binomial distribution with parameters $r$ and $l$  
$N(0, 1)$ Normal distribution with mean 0 and variance 1  
$\text{Pois}(\lambda)$ Poisson distribution with parameter $\lambda$  
$S$ Shewhart statistic, scan statistic  
$\sigma_0$ Standard deviation of random variable with distribution $F_0$  
$\hat{\sigma}_0$ Estimated standard deviation of random variable with distribution $F_0$  
$\hat{\sigma}_Y$ Estimated standard error of $Y$  
$\hat{\sigma}_{i,j,k}^2$ Historical limits sample variance for reportable disease $i$, in week $j$ and year $k$  
$T_{i,j,k}$ Total number of reportable disease $i$, in week $j$ and year $k$  
$\tau$ (Unknown) time of change from $F_0$ to $F_1$  
$\tau_l$ Last time period of outbreak  
$\tau_s$ Start time period of outbreak  
$Y$ Observation  
$\bar{Y}$ Sample mean of $Y_1, \ldots, Y_m$  
$\hat{Y}_t$ Predicted value of $Y$ for day $t$  
$Z$ Standardized observation
7.1 Historical Limits Detection Method

The \textit{historical limits} method is commonly used by public health practitioners to compare data from a current time period to data from an equivalent historical period or periods. The idea is to assess whether the mean of the current data is significantly larger or smaller than the mean observed in the historical data after accounting for the natural variation inherent in the data.

An example of a system that uses historical limits is the CDC’s National Notifiable Diseases Surveillance System (NNDSS). NNDSS aggregates and summarizes data on specific diseases that health care providers are required by state law to report to public health departments. Reportable diseases include anthrax, botulism, plague, and tularemia.\footnote{See www.cdc.gov/ncphi/diss/nndss/phs/infdis.htm for a complete list of reportable diseases.} Each week the states report counts of cases for each of the reportable diseases to the CDC.

A simple use of comparisons to historical data is Table I of the “Notifiable Diseases/Deaths in Selected Cities Weekly Information” report published on-line each week in the CDC’s \textit{Morbidity and Mortality Weekly Report} (MMWR). For each reportable disease, Table I provides the current week’s national count total and a 5-year weekly average for an equivalent period. A significant deviation of the current count from the 5-year average is an indication of unusual activity in a particular week for any of the diseases. Table I also provides other useful statistics such as the cumulative number of cases reported for the year and the total number of cases reported for each of the past five years.

Figure I in the MMWR’s “Notifiable Diseases/Deaths in Selected Cities Weekly Information” employs a more formal use of historical data by incorporating a measure of variation into limits. Specifically, the most recent 4-week totals for each of the notifiable diseases, $T_{i,j,k}$, for reportable disease $i$, in week $j$ and year $k$, are compared to the mean number of cases reported for the same 4-week period, the preceding 4-week period, and the succeeding 4-week period for the previous 5 years as follows.

Calculating the mean of the 15 historical periods as

$$\hat{\mu}_{i,j,k} = \frac{1}{15} \sum_{s=1}^{5} \sum_{r=-1}^{1} T_{i,j-r,k-s},$$

and the variance as

$$\hat{\sigma}_{i,j,k}^2 = \frac{1}{14} \sum_{s=1}^{5} \sum_{r=-1}^{1} (T_{i,j-r,k-s} - \hat{\mu}_{i,j,k})^2,$$

the historical limits for the ratio $T_{i,j,k}/\hat{\mu}_{i,j,k}$ are defined to be
7.1 Historical Limits Detection Method

\[ 1 \pm \frac{2\hat{\sigma}_{i,j,k}}{\hat{\mu}_{i,j,k}}. \]  

(7.1)

Fig. 7.1. An example of Figure I from “Notifiable Diseases/Deaths in Selected Cities Weekly Information” for the week 47 of 2009 (CDC, 2009). For this week, the mumps count exceeded its historical limits.

Figure 7.1 is an example of Figure I from “Notifiable Diseases/Deaths in Selected Cities Weekly Information” for week 47 of 2009 (CDC, 2009), where for this week the mumps count exceeded its historical limits as shown by the crosshatched top of the bar. See Stroup et al. (1989) and Stroup et al. (1993) for additional discussion, including further discussion of the historical limits methodology used in the CDC’s National Notifiable Diseases Surveillance System and alternatives to that methodology.

Returning to Equation 7.1, multiplying through the equation by \( \hat{\mu}_{i,j,k} \) shows that the idea of the historical method is to signal whenever the mean of the current data is larger than the mean of observed in the historical data plus two standard deviations, or it is smaller than the mean minus two standard deviations. That is, a signal is produced if

\[ T_{i,j,k} \geq \hat{\mu}_{i,j,k} + 2\hat{\sigma}_{i,j,k} \]  

(7.2)

or

\[ T_{i,j,k} \leq \hat{\mu}_{i,j,k} - 2\hat{\sigma}_{i,j,k}. \]  

(7.3)

As the next section will show, this is just a specific form of the Shewhart method with two standard deviation signal limits.
7.2 Shewhart Detection Method

The Shewhart control chart (Shewhart, 1931) is perhaps the best known of all univariate statistical process control methods and is widely applied in industry. One variant compares a test statistic $S$, calculated as the absolute value of a standardized sample mean at time $t$,

$$S_t = \left| \frac{\bar{Y}_t - \hat{\mu}_0}{\hat{\sigma}_Y} \right|,$$

(7.4)

to a threshold $h > 0$, where at time $t$ a sample of size $m$ is drawn and the sample mean is calculated as

$$\bar{Y}_t = \frac{1}{m} \sum_{j=1}^{m} Y_{tj}.$$

The estimated mean, $\hat{\mu}_0$, and estimated standard error, $\hat{\sigma}_Y$, are calculated using historical data from what is referred to as the in-control distribution, $F_0$, which is the desired state of the process. In industrial quality control, it is often assumed that the process mean is stationary, meaning that as long as the process is “in control” $F_0$ does not change over time. Under these conditions, the mean and standard deviation need only be estimated once from some historical data and then the same values for $\hat{\mu}_0$ and $\hat{\sigma}_Y$ can be used in Equation 7.4 at each time period.

If $S_t$ exceeds the threshold then a signal is generated. If it does not, then another sample is drawn and the process is repeated.

The idea of the Shewhart control chart is that, if at some point in the future the observations have a new new distribution $F_1$ with mean $\mu_1$ (where $\mu_1 \neq \mu_0$), the resulting statistic $S$ will be more likely to exceed $h$. The Shewhart chart is really just a repeated two-sided $z$- or $t$-test, where samples continue to be drawn and tested until the test statistic falls far enough out in one of the tails of the sampling distribution that a signal is generated.

For a reasonably large iid sample of size $m$, the sample mean has an approximately normal distribution via the Central Limit Theorem (CLT). This simplifies control chart implementation by eliminating the need to first determine $F_0$. Rather, for known $\sigma_Y$ or for a reasonably large historical sample size it follows that $S$ is approximately distributed according to a standard normal distribution: $N(0, 1)$. This makes choosing thresholds relatively straightforward.

Note the correspondence between the Shewhart method in Equation 7.4 and the historical limits method. In particular, if $h = 2$ then the Shewhart method will signal if

$$\left| \frac{\bar{Y}_t - \hat{\mu}_0}{\hat{\sigma}_Y} \right| \geq 2.$$

(7.5)
7.2 Shewhart Detection Method

Rearranging Equation 7.5 results in equations equivalent to Equations 7.2 and 7.3, the main difference being that the historical limits method is monitoring totals while the Shewhart control chart usually monitors the process mean. In industrial practice, two Shewhart control charts are often run simultaneously, one to monitor the process mean and the other the process variation. More complicated versions of the Shewhart chart also exist that are designed to be more sensitive to small shifts in the mean using what are called “runs rules.”

Shewhart *individuals charts* can also be created for monitoring using individual observations, though the distribution of those individual observations must either be known or well-estimated in order to appropriately set the thresholds. See Montgomery (2004) for additional details about the design and application of Shewhart control charts in the industrial SPC setting.

### 7.2.1 Modified for Biosurveillance

The fundamental idea of the Shewhart chart is to monitor data, comparing it to a distribution that reflects the “normal” state ($F_0$) until an observation is observed that is sufficiently rare under $F_0$ that it is rejected as truth. As with classical hypothesis testing, “rare” is defined by the alternative to be detected.

In biosurveillance, monitoring the data for unusual increases is of most interest. In addition, because the goal is often to detect an outbreak as quickly as possible, it is preferable to monitor the individual observations rather than waiting for enough data to accumulate in order to calculate sample averages. These conditions imply that the standard SPC implementation should be modified as follows.

Given only increases are relevant to detect, individual observations are used, and assuming $F_0$ is stationary, then the *Shewhart early event detection method* monitors

$$S_t = \frac{Y_t - \bar{Y}}{\hat{\sigma}_Y}$$

The sample mean and standard deviation are calculated from historical data distributed as $F_0$. That is, $\bar{Y}$ is an estimate of $\mu_0$ and $\hat{\sigma}_Y$ is an estimate of the standard deviation of $Y$, $\sigma_0$.

If the distribution of the data is not stationary, which is the usual case in biosurveillance, apply the Shewhart method in two steps. First, model the systematic effects in the data, perhaps using one of the methods described in Chapter 5, and calculate the standardized residuals

$$Z_t = \frac{Y_t - \bar{Y}_t}{\hat{\sigma}_{Y - \bar{Y}}}$$

where $Y_t - \bar{Y}_t$ is the residual at time $t$ and $\hat{\sigma}_{Y - \bar{Y}}$ is the estimated standard deviation of the residuals. Then monitor the residuals $Z_t$, where ideally, but not
necessarily, $Z_t \sim N(0, 1)$. Regardless, the purpose of this first step is to remove the systematic effects from the data with the goal of achieving stationarity or near-stationarity of the residuals. If successful, then the Shewhart method is applied to the standardized residuals, where $Z_t \sim F_0$ for some distribution $F_0$.

The Shewhart method is applied to biosurveillance using the following “one-sided” scheme for monitoring the mean incidence: Choose a threshold $h$ and observe a sequence of residuals over time. As long as $Z_t < h$, assume the $F_0$ is true and thus there is no evidence of an outbreak. However, if $Z_t \geq h$, signal that an outbreak may be occurring.

**Example 7.1.** To illustrate the Shewhart detection method, consider the following ILI daily syndrome count data, along with forecasts from a regression-based model (see Chapter 5) and the associated estimated standard deviations of the residuals, for days $t = 1, \ldots, 15$.

\[
\begin{array}{cccccccc}
\text{day} & 1 & 2 & 3 & 4 & 5 & 6 & 7 \\
\hline
y_t & 20 & 18 & 21 & 24 & 17 & 22 & 23 \\
\hat{\sigma}_{y-\hat{y}} & 0.959 & 0.984 & 0.976 & 0.991 & 1.016 & 0.983 & 1.038 \\
\hline
\end{array}
\]

| \cdots & 8 & 9 & 10 & 11 & 12 & 13 & 14 & 15 |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| \cdots & 20 & 21 & 19 & 32 & 30 & 27 & 32 & 28 |
| \cdots & 1.055 & 1.023 & 1.019 & 1.154 & 1.126 & 1.144 & 1.170 & 1.211 |

Calculate the standardized residuals, $z_1, \ldots, z_{15}$, plot them, and then for $h = 1.9$ determine whether an outbreak may have occurred.

**Solution:** The standardized residuals are:

\[
\begin{array}{cccccccc}
\text{day} & 1 & 2 & 3 & 4 & 5 & 6 & 7 \\
\hline
z_t = \frac{y_t - \hat{y}_t}{\hat{\sigma}_{y-\hat{y}}} & -0.231 & 0.738 & -0.147 & -1.731 & 0.067 & 1.304 & -0.204 \\
\hline
\end{array}
\]

| \cdots & 8 & 9 & 10 & 11 & 12 & 13 & 14 & 15 |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| \cdots & -0.963 & 0.523 & -1.079 & 1.263 & 2.799 & 1.424 & 2.166 & 1.286 |

Figure 7.2 is a time series plot of the residuals. It shows that, for a threshold of $h = 1.9$, the Shewhart method signals at times 12 and 14 indicating a possible ILI outbreak. The signals are the result of large residuals which occur because the observed values on days 12 and 14 significantly exceeded the forecast values.
For the (simulated) data in Example 7.1, the outbreak actually started on day 11. Thus, the Shewhart method with a threshold of $h = 1.9$ had a delay of one day between the start of the actual outbreak and when it first signaled. As is evident in Figure 7.2, the speed of detection can be improved by using a lower threshold. Indeed, if the threshold had been set below the residual value on day 11 (1.263), the Shewhart method would have signaled for days 11-15. However, note that this improved sensitivity comes at the cost of increased false positive signals. For example, setting $h = 1.26$ would also have resulted in a false positive signal on day 6.

### 7.2.2 Implementation

To implement the Shewhart method, one must decide: (1) Which forecast model is most appropriate to use for a given set of data; (2) how to estimate the standard deviation of the residuals, and (3) how to set the threshold $h$. For forecast model alternatives, see Chapter 5. When estimating the standard deviation, the main question is how frequently it should be re-estimated and
how much historical data should be used in the calculation. Setting these two issues aside, which must be decided in the context of the actual data in a specific situation, a critical issue is how to choose \( h \).

The threshold \( h \) is chosen to achieve a desired level of performance as measured by the metrics described in Chapter 6. In particular, \( h \) is chosen to make the average time between false signals (ATFS) appropriately large, so that the rate of false positive signals is low enough that the organization has the resources to investigate all resulting signals.

Assuming the residuals are independent and identically distributed according to \( F_0 \), then \( \mathbb{P}(Z_t \geq h) = p \) for \( t = 1, 2, 3, \ldots \). Under these conditions, the time until first signal has a geometric distribution with probability \( p \) and thus the expected time to signal is \( 1/p \). Therefore, set \( h = F_0^{-1}(1-p) \) to achieve an ATFS = \( 1/p \).

CED performance and PSD performance are functions of the outbreak distribution. In an industrial quality control context where \( Z_t \sim F_1 \) for \( i = \tau, \tau + 1, \tau + 2, \ldots \), the average time until a true signal from the start of the outbreak is \( 1/p_1 \), where \( p_1 = 1 - F_1(h) \). In this situation, the PSD metric is unnecessary since, as long as \( p_1 > 0 \), the CED is finite and thus the probability of eventual detection is 1. That is, in the situation where the outbreak continues indefinitely, the Shewhart method will eventually signal during the outbreak and so PSD = 1.

These results do not apply to an outbreak of finite duration, say of \( D \) periods. In this, the typical biosurveillance situation, assuming \( F_1 \) is stationary over the outbreak period, the CED is calculated as

\[
\text{CED} = \sum_{i=1}^{D} i (1 - F_1(h)) F_1(h)^{i-1}. \tag{7.6}
\]

The probability of detecting an outbreak of duration \( D \), again assuming \( F_1 \) is constant over the outbreak period, is

\[
\text{PSD} = \sum_{i=1}^{D} (1 - F_1(h)) F_1(h)^{i-1} = 1 - F_1(h)^D. \tag{7.7}
\]

Now, if the outbreak distribution changes over time, \( Z_{\tau_s} \sim F_{\tau_s}, Z_{\tau_s+1} \sim F_{\tau_s+1}, \ldots, Z_{\tau_s+D-1} \sim F_{\tau_s+D-1} \), then the CED is calculated as

\[
\text{CED} = \left(1 - F_{\tau_s}(h)\right) + \sum_{i=2}^{D} \left[i (1 - F_{\tau_s+i-1}(h)) \prod_{j=1}^{i-1} F_{\tau_s+j-1}(h)\right] \tag{7.8}
\]

and the probability of detecting an outbreak of duration \( D \) is

\[
\text{PSD} = 1 - \prod_{i=1}^{D} F_{\tau_s+i-1}(h). \tag{7.9}
\]
The following examples, starting with one using a discrete distribution, should make these definitions and ideas more concrete.

Example 7.2. Let the daily count for chief complaints corresponding to the gastrointestinal syndrome for a large metropolitan hospital follow a Poisson distribution with parameter $\lambda_0 = 5$. That is, $F_0 = \text{Pois}(5)$. Figure 7.3 shows what 100 observations drawn randomly according to $F_0$ look like.

![Simulated gastrointestinal chief complaint data for 100 days](image)

**Fig. 7.3.** Simulated gastrointestinal chief complaint data for 100 days, where the data were drawn randomly from a Poisson distribution with $\lambda_0 = 5$.

**Question #1:** What threshold $h$ will achieve an ATFS of at least 100 days?

**Solution:** Figure 7.4 is the probability mass function for $F_0$. The goal is to determine the smallest $h$ such that $\Pr(Y \geq h) \leq 0.01$. Looking up the probabilities in a table or software package gives $\Pr(Y \geq 11) = 0.0137$ and $\Pr(Y \geq 12) = 0.0055$. Thus, the appropriate choice of threshold is $h = 12$ which achieves an ATFS of about 182 days $(1/0.0055 = 181.8)$, or roughly every six months.

**Question #2:** Given the threshold $h = 12$, what is the average time between signal events (ATBSE)?
Solution: Because Shewhart’s detection method “re-sets” after every signal, ATBSE=ATFS=182 days.

Question #3: Given the threshold $h = 12$, what is the expected time to detect an increase in the rate of gastrointestinal chief complaints to $F_1 = \text{Pois}(10)$ assuming the increase is constant and sustained?

Solution: Again using a table or statistical software, calculate that for $Y \sim \text{Pois}(10)$, $\Pr(Y \geq 12) = 0.303$. Therefore, the expected time until a signal for a sustained outbreak is $1/0.303 = 3.3$ or a little over three days.

Question #4: Given the outbreak only lasts 5 days, what is the CED?

Solution: The CED is the average time it takes to detect an outbreak, starting from the first day of the outbreak, given an outbreak is occurring. For this problem
CED = ∑

_{i=1}^{D} i \left(1 - F_1(h)\right) F_1(h)^{i-1}

= ∑

_{i=1}^{5} i \times (1 - 0.697) \times (0.697)^{i-1}

= 1.9 \text{ days.}

Note that the CED is less than the expected time to detect in Question #2 since (true) detection times greater than 5 days cannot occur.

**Question #5:** Finally, again given a threshold of \( h = 12 \), what is the probability of detecting an increase in the rate of gastrointestinal chief complaints to \( F_1 = \text{Pois}(10) \) over the course of the 5-day outbreak (after which the distribution reverts back to \( F_0 = \text{Pois}(5) \))?

**Solution:** From Question #2, the probability of a signal on any one day with \( F_1 = \text{Pois}(10) \) is 0.303. The probability of detecting the outbreak is one minus the probability that there are no signals on any of the outbreak days:

\[ \text{PSD} = 1 - (1 - 0.303)^5 = 0.8355, \]

so there is an 83.5% chance of detecting the outbreak.

**Alternate Solution:** To directly calculate the probability of one or more signals over the course of the five days, let \( X \) be the number of signals over the five days, \( X = 0, 1, \ldots, 5 \). Then \( X \) follows a binomial distribution with \( n = 5 \) trials and probability of success \( p = 0.303 \). Using a table or statistical software, calculate that \( \Pr(X \geq 1) = 0.8355 \), which matches the previous solution.

Note that ATFS, the CED, and the PSD all trade-off depending on the choice of threshold. High thresholds result in large ATFS values while lower thresholds result in smaller ATFS values. As described in Chapter 6, the ATFS should be used to set the threshold, and it should be based on the resources available for adjudicating signals. The more resources available the lower the threshold can be set and the more sensitive the system will be to detecting outbreaks.

The CED and PSD are then functions of the chosen threshold and the outbreak. They characterize the performance of the system in a particular situation where, obviously, smaller values for CED and larger values of the PSD are preferred. The lower the threshold the smaller the CED and the higher the PSD. However, raising the threshold does not necessarily increase the CED. Rather, the CED is a concave function that initially increases with increasing thresholds but then decreases as thresholds get very large. Specifically,
Univariate Temporal Methods

\[ \lim_{h \to -\infty} \text{CED} = 1, \]

since \( F(h) \to 0 \) as \( h \to -\infty \), and

\[ \lim_{h \to \infty} \text{CED} = 0, \]

since \( F(h) \to 1 \) as \( h \to \infty \). Therefore, when judging performance it is important to consider both the CED and PSD, since as \( h \to \infty \) the PSD \( \to 0 \).

Of course, these concepts apply equally to continuous distributions, as the next example illustrates.

**Example 7.3.** Let the standardized residuals from an adaptive regression model for the influenza-like illness (ILI) syndrome counts in a county follow a standard normal distribution: \( F_0 = N(0, 1) \). For \( h = 2.7775 \), assess the performance of Shewhart’s method at detecting an 8-day outbreak that manifests as \( F_{\tau_s+i} = N(\delta_i, 1) \), where \( \delta_0 = 0.5, \delta_1 = 1.0, \delta_2 = 1.5, \delta_3 = 2.0, \delta_4 = 2.0, \delta_5 = 1.5, \delta_6 = 1.0, \delta_7 = 0.5 \).

**Question #1:** Given \( h = 2.7775 \), what is the average time between false signals?

**Solution:** For \( Y \sim N(0, 1) \), the \( \mathbb{P}(Y \geq 2.7775) = 0.00274 \). Therefore, the average time between false signals is \( 1/0.00274 = 365 \) days, or about once a year.

**Question #2:** Given the outbreak described, what is the CED?

**Solution:** For this problem

\[
\text{CED} = (1 - F_{\tau_s}(h)) + \sum_{i=2}^{D} \left[ i (1 - F_{\tau_{s+i-1}}(h)) \prod_{j=1}^{i-1} F_{\tau_s+j-1}(h) \right]
\]

\[
= 1 \times 0.011 + \\
2 \times 0.038 \times 0.989 + \\
3 \times 0.101 \times 0.989 \times 0.962 + \\
4 \times 0.218 \times 0.989 \times 0.962 \times 0.899 + \\
5 \times 0.218 \times 0.989 \times 0.962 \times 0.899 \times 0.782 + \\
6 \times 0.101 \times 0.989 \times 0.962 \times 0.899 \times 0.782 \times 0.782 + \\
7 \times 0.038 \times 0.989 \times 0.962 \times 0.899 \times 0.782 \times 0.782 \times 0.782 + \\
8 \times 0.011 \times 0.989 \times 0.962 \times 0.899 \times 0.782 \times 0.782 \times 0.782 \times 0.782 \\
= 2.3 \text{ days}.
\]

**Question #3:** What is the probability of detecting the outbreak?
Solution: The probability of detecting the outbreak is

\[
\text{PSD} = 1 - \prod_{i=1}^{p} F_{\tau + i - 1}(h) \\
= 1 - (0.989 \times 0.962 \times 0.899 \times 0.782 \times \\
0.782 \times 0.899 \times 0.962 \times 0.989) = 0.55.
\]

Question #4: What do graphs of the ATFS, the CED, and the PSD for this outbreak as a function of the threshold, for $1.5 \leq h \leq 3$ show?

Solution: The plots are shown in Figure 7.5. The dotted lines show the values for each of the functions at $h = 2.7775$. Note how the ATFS is strictly increasing for increasing thresholds while the PSD is strictly decreasing. The CED, on the other hand, is concave. This occurs because, when the threshold is smaller, increases make it more difficult to detect and thus the CED increases. However, as the threshold continues to get larger it reaches a point where most outbreaks are simply not detected and that then causes the CED to decrease. As previously discussed, in the limit as $h \to \infty$, CED $\to 0$.

Note how the plots show that increasing the PSD comes at the cost of decreasing the ATFS. Decreasing the ATFS is equivalent to increasing the rate of false positive signals. For example, to achieve a PSD of 90 percent (up from 55 percent), the ATFS decreases to about 50 days (down from 365 days). Whether such a trade-off is desirable and achievable is a function of the resources available to adjudicate the signals. Also note how increasing the PSD also results in an increased CED – since more outbreaks are now being caught – and the increase is relatively modest.

7.3 CUSUM Detection Method

The cumulative sum (CUSUM) control chart of Page (1954) and Lorden (1971) is a well known statistical process control methodology that is also frequently applied in biosurveillance. Formally, the CUSUM is a sequential test for a change from a known distribution $F_0$ to a known alternative distribution $F_1$. The method monitors the statistic $C_t$, which satisfies the recursion

\[
C_t = \text{max}[0, C_{t-1} + L_t],
\]

where the increment $L_t$ is the log likelihood ratio

\[
L_t = \log \frac{f_1[Y_t]}{f_0[Y_t]}.
\]
Fig. 7.5. Graphs of ATFS, CED, and the PSD as a function of the threshold, $1.5 \leq h \leq 3$, for the outbreak defined in Example 7.3.
7.3 CUSUM Detection Method

The method is usually started at $C_0 = 0$; it stops and concludes that $Y \sim F_1$ at the first time when $C_t > h$, for some threshold $h$ that achieves a desired ATFS when $Y \sim F_0$ (i.e., when no outbreak is present).

If $f_0$ and $f_1$ are normal densities with common variance $\sigma^2$ and means $\mu_0$ and $\mu_1 = \mu_0 + \delta \sigma$ ($\delta > 0$), respectively, then Equation 7.10 becomes

$$C_t = \max \{0, C_{t-1} + (Y_t - \mu_0)/\sigma - k\}, \quad (7.12)$$

with $k = \frac{\mu_1 - \mu_0}{2\sigma} = \frac{\delta}{2}$, where $k$ is commonly referred to as the reference value. If the $Z$s are independent and identically distributed according to $f_0$ before some unknown change point and according to $f_1$ after the change point, then the CUSUM has certain optimality properties. See Moustakides (1986) and Ritov (1990).

Example 7.4. To illustrate the CUSUM, apply Equation 7.12 to the ILI residuals data of Example 7.1 (which can reasonably be assumed to have a standard normal distribution). In particular, calculate the CUSUM values, $C_1, \ldots, C_{15}$, plot them, and then for $h = 1.0$ determine whether an outbreak may have occurred. For the purposes of these calculations, use $k = 1$ and $C_0 = 0$.

Solution: If the residuals have a standard normal distribution, $F_0 = N(0,1)$, then $\mu_0 = 0$ and $\sigma = \sigma_0 = 1$. The CUSUM calculations are:

<table>
<thead>
<tr>
<th>$t$</th>
<th>$z_t$</th>
<th>$C_{t-1} + z_t - k$</th>
<th>$C_t$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-0.231</td>
<td>$0 - 0.231 - 1 = -1.231$</td>
<td>0.000</td>
</tr>
<tr>
<td>2</td>
<td>0.738</td>
<td>$0 + 0.738 - 1 = -0.262$</td>
<td>0.000</td>
</tr>
<tr>
<td>3</td>
<td>-0.147</td>
<td>$0 - 0.147 - 1 = -1.147$</td>
<td>0.000</td>
</tr>
<tr>
<td>4</td>
<td>-1.731</td>
<td>$0 - 1.731 - 1 = -2.731$</td>
<td>0.000</td>
</tr>
<tr>
<td>5</td>
<td>0.067</td>
<td>$0 + 0.067 - 1 = -0.933$</td>
<td>0.000</td>
</tr>
<tr>
<td>6</td>
<td>1.304</td>
<td>$0 + 1.304 - 1 = 0.304$</td>
<td>0.304</td>
</tr>
<tr>
<td>7</td>
<td>-0.204</td>
<td>$0.304 - 0.204 - 1 = -0.900$</td>
<td>0.000</td>
</tr>
<tr>
<td>8</td>
<td>-0.963</td>
<td>$0 - 0.963 - 1 = -1.963$</td>
<td>0.000</td>
</tr>
<tr>
<td>9</td>
<td>0.523</td>
<td>$0 + 0.523 - 1 = -0.477$</td>
<td>0.000</td>
</tr>
<tr>
<td>10</td>
<td>-1.079</td>
<td>$0 - 1.079 - 1 = -2.079$</td>
<td>0.000</td>
</tr>
<tr>
<td>11</td>
<td>1.263</td>
<td>$0 + 1.263 - 1 = 0.263$</td>
<td>0.263</td>
</tr>
<tr>
<td>12</td>
<td>2.799</td>
<td>$0.263 + 2.799 - 1 = 2.062$</td>
<td>2.062</td>
</tr>
<tr>
<td>13</td>
<td>1.424</td>
<td>$2.062 + 1.424 - 1 = 2.486$</td>
<td>2.486</td>
</tr>
<tr>
<td>14</td>
<td>2.166</td>
<td>$2.486 + 2.166 - 1 = 3.652$</td>
<td>3.652</td>
</tr>
<tr>
<td>15</td>
<td>1.286</td>
<td>$3.652 + 1.286 - 1 = 3.938$</td>
<td>3.938</td>
</tr>
</tbody>
</table>

Figure 7.6 is a time series plot of the CUSUM statistics. It shows that, for a threshold of $h = 1.0$, the CUSUM method signals on days 12 through 14.
indicating a possible ILI outbreak. The signals are the result of large positive residuals, which begin to accumulate in the CUSUM calculations once the residuals regularly start to exceed the reference value of \( k = 1 \), and which occur during the outbreak because the observed counts are significantly greater than the forecast values.

![CUSUM statistics](image)

**Fig. 7.6.** A plot of CUSUM statistics from Example 7.4. For a threshold of \( h = 1.0 \), the CUSUM method signals at for days 12 through 14 indicating a possible ILI outbreak.

To get some insight into how the CUSUM works, focus on the increment \( \frac{(Y_t - \mu_0)}{\sigma} - k = \frac{(Y_t - \mu_0)}{\sigma} - \delta/2 \), which is what gets added onto the cumulative sum at each iteration. In particular, note that when \( Y \sim F_0 \) that \( \mathbb{E}[\frac{(Y_t - \mu_0)}{\sigma} - \delta/2] = \frac{(\mu_0 - \mu_0)}{\sigma} - \delta/2 = -\delta/2 \). Hence, when the data are generated from \( F_0 \) the increment is negative (in an expected value sense) and so the CUSUM statistic will have a downward drift. Since the CUSUM is bounded below by 0, under these conditions the CUSUM will tend to be down around zero.

On the other hand, when \( Y \sim F_1 \) then \( \mathbb{E}[\frac{(Y_t - \mu_0)}{\sigma} - \delta/2] = \frac{(\mu_1 - \mu_0)}{\sigma} - \delta/2 = +\delta/2 \). Thus, when the data are generated from \( F_1 \) the increment
is positive and so the CUSUM statistic will have an upward drift. This positive drift will accumulate in the cumulative sum which will then tend to exceed the threshold fairly quickly.

Equation 7.12 is a one-sided CUSUM, meaning that it will only detect increases in the mean. In industrial quality control, where it is often important to detect both increases and decreases in the mean, a second CUSUM control chart is used to detect decreases. The CUSUM can also be used to monitor process variability. For example, to monitor an increase in process variability, following Hawkins & Olwell (1998, p. 67), use the CUSUM recursion

$$V_t = \max\{0, V_{t-1} + W_t - k\},$$

where

$$W_t = \frac{\sqrt{|Y_t|} - 0.822}{0.394}.$$

As recommended by Hawkins and Olwell, the same value for $k$ should be used in these CUSUMs for monitoring variability as in the CUSUMs for the mean.

Equation 7.12 is the CUSUM form routinely used, even when the underlying assumptions are only approximately met. However, for other $F_0$ and $F_1$ distributions, the appropriate log likelihood ratio (Equation 7.11) should be used and the CUSUM recursion adjusted.

For example, it may be reasonable to assume that the daily counts for some syndromes have a Poisson distribution. For $F_0 = \text{Pois}(\lambda_0)$ and $F_1 = \text{Pois}(\lambda_1)$ with $\lambda_1 > \lambda_0$, the CUSUM becomes

$$C_t = \max\{0, C_{t-1} + Y_t - k\},$$

where

$$k = \frac{\lambda_1 - \lambda_0}{\ln \lambda_1 - \ln \lambda_0}.$$  

For the Poisson distribution, $\mathbb{E}(Y) = \text{Var}(Y)$, so this CUSUM is simultaneously monitoring both the mean and variance of the distribution.

It may also be the case that other daily syndrome counts have a negative binomial distribution, where for given parameters $r$ and $l$, the probability mass function is

$$\mathbb{P}(Y = y) = \binom{y + r - 1}{r - 1} \left(\frac{r}{r + l}\right)^r \left(1 - \frac{r}{r + l}\right)^y.$$  

(7.14)

Then for $F_0 = \text{NBin}(r, l_0)$ and $F_1 = \text{NBin}(r, l_1)$ (i.e., assuming $r$ does not change) the CUSUM is

$$C_t = \max\{0, C_{t-1} + Y_t - k\},$$

where

$$k = \frac{-r \ln[(r + l_0)/(r + l_1)]}{\ln[(l_1(r + l_0))/(l_0(r + l_1))]}.$$
For this parametrization of the negative binomial, $E(Y) = l$ and $\text{Var}(Y) = l + l^2/r$ so this CUSUM also simultaneously monitors both the mean and variance.

### 7.3.1 Modified for Biosurveillance

In industrial settings, the CUSUM is applied directly to the observations because some control is exhibited over the process such that it is reasonable to assume $F_0$ is stationary. If in fact biosurveillance data can reasonably be assumed iid according to a Poisson or negative binomial distribution, then the recursions using the increments from Equations 7.13 or 7.15, respectively, can be used to monitor disease incidence using the data itself.

However, this is generally not the possible in biosurveillance as the data often have uncontrollable systematic trends, such as seasonal cycles and day-of-the-week effects. In this case, as with the Shewhart method, the CUSUM should be applied to the residuals of a model that accounts for and removes any systematic effects in the data. If the residuals are normally or nearly-normally distributed, then the CUSUM in Equation 7.12 should be used. If the residuals are not normally distributed, then the correct log likelihood ratio increment should be used in Equation 7.10 to derive the appropriate CUSUM.

And, of course, since biosurveillance is focused on detecting increases in disease incidence, only a single CUSUM is necessary.

### 7.3.2 Implementation

Implementing the CUSUM requires choosing a threshold $h$ and setting the reference value $k$. The choice of threshold should be based on the smallest ATFS that can be accommodated. That is, to achieve the greatest sensitivity for detecting outbreaks, the threshold should be set as low as possible, subject to the constraint that sufficient resources are available to investigate the resulting signals.

Now, in the CUSUM the ATFS is also a function of the reference value, which itself is a function of the outbreak to be detected (e.g., $\mu_1$ in Equation 7.12 or $\lambda_1$ in Equation 7.13 or $l_1$ in Equation 7.15). The choice of $\mu_1$ or $\lambda_1$ or $l_1$ is a subjective judgment based on the smallest outbreak which is important to detect quickly.

For the CUSUM based on the normal distribution, given $k$, Table 7.1 can be used to determine the choice of $h$ needed to achieve a desired ATFS. Alternatively, again for the CUSUM based a mean shift in a normal distribution, the following approximation for the expected time between false signals due to Siegmund (1985) can be used:

$$\text{ATFS} \approx \frac{\exp(2k(h + 1.166)) + 2k(h + 1.166) - 1}{2k^2}.$$
### Table 7.1.

<table>
<thead>
<tr>
<th>$h$</th>
<th>0.00</th>
<th>0.25</th>
<th>0.50</th>
<th>0.75</th>
<th>1.00</th>
<th>1.25</th>
<th>1.50</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.000</td>
<td>4.75</td>
<td>7.0</td>
<td>11.2</td>
<td>19.2</td>
<td>35.3</td>
<td>68.9</td>
<td>142.2</td>
</tr>
<tr>
<td>1.125</td>
<td>5.27</td>
<td>8.0</td>
<td>13.2</td>
<td>23.4</td>
<td>44.8</td>
<td>91.4</td>
<td>196.8</td>
</tr>
<tr>
<td>1.250</td>
<td>5.84</td>
<td>9.1</td>
<td>15.4</td>
<td>28.6</td>
<td>57.2</td>
<td>122.1</td>
<td>274.9</td>
</tr>
<tr>
<td>1.375</td>
<td>6.44</td>
<td>10.3</td>
<td>18.0</td>
<td>34.9</td>
<td>73.1</td>
<td>164.0</td>
<td>387.2</td>
</tr>
<tr>
<td>1.500</td>
<td>7.09</td>
<td>11.6</td>
<td>21.1</td>
<td>42.6</td>
<td>93.8</td>
<td>221.5</td>
<td>549.7</td>
</tr>
<tr>
<td>1.625</td>
<td>7.76</td>
<td>13.0</td>
<td>24.6</td>
<td>52.0</td>
<td>120.7</td>
<td>300.5</td>
<td>786.0</td>
</tr>
<tr>
<td>1.750</td>
<td>8.48</td>
<td>14.6</td>
<td>28.6</td>
<td>63.5</td>
<td>155.5</td>
<td>409.4</td>
<td>1,130.8</td>
</tr>
<tr>
<td>1.875</td>
<td>9.22</td>
<td>16.3</td>
<td>33.3</td>
<td>77.4</td>
<td>200.5</td>
<td>559.4</td>
<td></td>
</tr>
<tr>
<td>2.000</td>
<td>10.00</td>
<td>18.2</td>
<td>38.5</td>
<td>94.3</td>
<td>258.7</td>
<td>766.2</td>
<td></td>
</tr>
<tr>
<td>2.125</td>
<td>10.81</td>
<td>20.2</td>
<td>44.6</td>
<td>114.9</td>
<td>333.8</td>
<td>1,051.0</td>
<td></td>
</tr>
<tr>
<td>2.250</td>
<td>11.66</td>
<td>22.4</td>
<td>51.5</td>
<td>139.7</td>
<td>430.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.375</td>
<td>12.53</td>
<td>24.7</td>
<td>59.3</td>
<td>169.7</td>
<td>555.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.500</td>
<td>13.43</td>
<td>27.3</td>
<td>68.2</td>
<td>206.0</td>
<td>716.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.625</td>
<td>14.37</td>
<td>30.0</td>
<td>78.3</td>
<td>249.7</td>
<td>922.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.750</td>
<td>15.33</td>
<td>32.9</td>
<td>89.8</td>
<td>302.5</td>
<td>1,187.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.875</td>
<td>16.32</td>
<td>36.1</td>
<td>102.8</td>
<td>366.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.000</td>
<td>17.35</td>
<td>39.5</td>
<td>117.6</td>
<td>442.8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.125</td>
<td>18.41</td>
<td>43.1</td>
<td>134.4</td>
<td>535.3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.250</td>
<td>19.50</td>
<td>47.0</td>
<td>153.4</td>
<td>646.9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.375</td>
<td>20.61</td>
<td>51.2</td>
<td>175.0</td>
<td>781.4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.500</td>
<td>21.76</td>
<td>55.7</td>
<td>199.6</td>
<td>943.7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.625</td>
<td>22.95</td>
<td>60.5</td>
<td>227.4</td>
<td>1,139.4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.750</td>
<td>24.16</td>
<td>65.7</td>
<td>259.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.875</td>
<td>25.40</td>
<td>71.2</td>
<td>294.8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.000</td>
<td>26.68</td>
<td>77.1</td>
<td>335.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.125</td>
<td>27.99</td>
<td>83.4</td>
<td>381.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.250</td>
<td>29.32</td>
<td>90.2</td>
<td>433.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.375</td>
<td>30.69</td>
<td>97.4</td>
<td>492.8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.500</td>
<td>32.09</td>
<td>105.1</td>
<td>559.9</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.625</td>
<td>33.53</td>
<td>113.4</td>
<td>636.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.750</td>
<td>34.99</td>
<td>122.2</td>
<td>722.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.875</td>
<td>36.48</td>
<td>131.6</td>
<td>820.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.000</td>
<td>38.01</td>
<td>141.7</td>
<td>930.9</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.125</td>
<td>39.57</td>
<td>152.4</td>
<td>1,056.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.250</td>
<td>41.15</td>
<td>163.9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.375</td>
<td>42.77</td>
<td>176.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.500</td>
<td>44.42</td>
<td>189.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.625</td>
<td>46.11</td>
<td>203.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.750</td>
<td>47.82</td>
<td>218.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.875</td>
<td>49.56</td>
<td>233.9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6.000</td>
<td>51.34</td>
<td>250.8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 7.1. ATFS for a CUSUM for the mean shift of a standard normal distribution as a function of $k$ and $h$. Adapted from Hawkins & Olwell (1998, Table 3.1, p. 48).
If the observations are not normally distributed, but the probability distribution of the observations or residuals can be specified, simulation can be used to determine the threshold to achieve a specific ATFS. For a given $h$ and $k$, the following pseudo-code calculates the estimated ATFS and its standard error:

```
set $h$ and $k$ to desired values
dimension $tfs$ vector to size $n$
loop from $i = 1$ to $n$
    { 
        set $CUSUM = 0$
        set $tfs[i] = 0$
        while ($CUSUM < h$)
            { 
                $CUSUM = \max(0, \text{appropriate CUSUM recursion expression})$
                $tfs[i] = tfs[i] + 1$
            }
    }
print ATFS=average($tfs$) and s.e.(ATFS) = s.d.($tfs$)/$\sqrt{n}$
```

This simulation is iteratively run, starting with a relatively small $n$, to find the value of $h$ that achieves a desired ATFS. As the approximate value of $h$ is determined, $n$ can be increased to obtain as precise an estimate as desired (and achievable within the computing resources available). The above pseudo-code can also be enclosed within additional code to automate the search for $h$.

Another alternative for calculating the expected time between false signals is the Markov chain approach of Brook & Evans (1972).

The following example illustrates how to apply and calculate the CUSUM in a specific biosurveillance scenario.

**Example 7.5.** The following data are the standardized residuals from an adaptive regression fit to daily ILI counts: 0.819, 0.557, 1.733, 0.415, 0.155, -3.173, -0.549, -0.756, 1.060, 0.137, 0.744, 0.509, 1.206, 0.512, -1.347, 3.772, 1.142, 3.132, 2.846, 2.117.

Under non-outbreak conditions, the standardized residuals are well-characterized by the standard normal distribution.

**Question #1:** What reference value ($k$) should be used to detect a 2-standard deviation increase in the mean?

**Solution:** $k = \frac{\mu_1 - \mu_0}{2\sigma} = \frac{\mu_0 + 2\sigma - \mu_0}{2\sigma} = 1.$
Question #2: Given the $k$ from the first question, what threshold should be used to achieve an average time between false signals of approximately 45 days?

Solution: Using Table 7.1, for $k = 1$ and $h = 1.125$ gives an ATFS of 44.8 days.

Question #3: Given the threshold $h = 1.125$ and $k = 1$, what is the average time between signal events (ATBSE)?

Solution: Using simulation (based on the pseudo-code below), the average time between signal events is estimated as ATBSE=44.7 (s.e. 0.14). In this case, the ATBSE is close to the ATFS, but that will not necessarily be case in general.

```
initialize $h$ and $k$ to desired values
initialize $counter1$ to 0
initialize $counter2$ to 1
initialize $CUSUM$ to 0

dimension $tbse$ vector to size $n$ and set to a vector of 1s
loop from $i = 1$ to $n$
{
    set $lastCUSUM = CUSUM$
    $CUSUM = \max(0, CUSUM + random\ standard\ normal\ deviate - k)$
    $counter1 = counter1 + 1$
    if($CUSUM \geq h \& lastCUSUM < h$)
    {
        $TBSE[counter2] = counter1$
        $counter2 = counter2 + 1$
    }
    if($CUSUM < h \& lastCUSUM \geq h$)
    {
        $counter1 = 1$
    }
}
print $ATBSE=\text{average}(tbse[2 \text{ to } n])$
print s.e.($ATBSE) = s.d.(tbse[2 \text{ to } n])/\sqrt{n-1}$
```

Question #4: Calculate the CUSUM values over time, both when the CUSUM is re-set after each signal and when it is not re-set, and plot them on a chart showing the threshold. Is there any evidence of an outbreak?
Solution: Table 7.2 shows the CUSUM calculations and Figure 7.7 plots the CUSUM statistics over time. Both show that the first signal occurs at time $t = 16$ when the CUSUM statistic first exceeds $h$.

After time 16, both versions of the CUSUM first signal. After that time, the CUSUM that is re-set signals intermittently, while the CUSUM that is not re-set signals continuously. To understand whether and when it is appropriate to re-set the CUSUM, see the discussions in Chapter 6 on pages 183 and 187.

<table>
<thead>
<tr>
<th>$t$</th>
<th>$Z_t$</th>
<th>With re-setting:</th>
<th>Without re-setting:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$C_{t-1}$</td>
<td>$C_t$</td>
</tr>
<tr>
<td>1</td>
<td>0.819</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>2</td>
<td>0.557</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>3</td>
<td>1.733</td>
<td>0.000</td>
<td>0.733</td>
</tr>
<tr>
<td>4</td>
<td>0.415</td>
<td>0.733</td>
<td>0.148</td>
</tr>
<tr>
<td>5</td>
<td>0.155</td>
<td>0.148</td>
<td>0.000</td>
</tr>
<tr>
<td>6</td>
<td>-3.173</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>7</td>
<td>-0.549</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>8</td>
<td>-0.756</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>9</td>
<td>1.060</td>
<td>0.000</td>
<td>0.060</td>
</tr>
<tr>
<td>10</td>
<td>0.137</td>
<td>0.060</td>
<td>0.000</td>
</tr>
<tr>
<td>11</td>
<td>0.744</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>12</td>
<td>0.509</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>13</td>
<td>1.206</td>
<td>0.000</td>
<td>0.206</td>
</tr>
<tr>
<td>14</td>
<td>0.512</td>
<td>0.206</td>
<td>0.000</td>
</tr>
<tr>
<td>15</td>
<td>-1.347</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>16</td>
<td>3.772</td>
<td>0.000</td>
<td><strong>2.772</strong></td>
</tr>
<tr>
<td>17</td>
<td>1.142</td>
<td>0.000</td>
<td>0.142</td>
</tr>
<tr>
<td>18</td>
<td>3.132</td>
<td>0.142</td>
<td><strong>2.274</strong></td>
</tr>
<tr>
<td>19</td>
<td>2.846</td>
<td>0.000</td>
<td><strong>1.846</strong></td>
</tr>
<tr>
<td>20</td>
<td>2.117</td>
<td>0.000</td>
<td>1.117</td>
</tr>
</tbody>
</table>

Table 7.2: CUSUM values for Example 7.5, where $C_t = \max(0, C_{t-1} + Z_t - 1)$, both with and without re-setting after a signal. The bold $C_t$ values are signal times, where $C_t > h = 1.125$.

This next example illustrates how to set up the CUSUM and then assess its performance for various types of outbreaks.
Example 7.6. Consider a biosurveillance system which, when there is no outbreak, the standardized residuals from a model of the daily counts has a standard normal distribution.

Question #1: What values of \( k \) and \( h \) should be used to detect a 1-standard deviation increase in the mean with an ATFS of roughly once every 40 days?

Solution: For standardized residuals \( \mu_0 = 0 \) and a 1-standard deviation increase means \( \mu_1 = 1 \), so that \( k = 0.5 \). Then, from Table 7.1 if follows that \( h = 2 \). Then results in an average of one false signal every 38.5 days.

Question #2: Using simulation, what is the zero-state CED and PSD of an outbreak characterized by a triangle shaped mean shift with a duration (\( D \)) of 9 days and a maximum increase (\( M \)) of two standard deviations?

Solution: Using a simulation based on the pseudo-code below, the zero-state CED is 4.34 days (s.e. 0.01) with a PSD of 0.973. The pseudo-code:

```plaintext
set o = {0.0, 0.5, 1.0, 1.5, 2.0, 1.5, 1.0, 0.5, 0.0}
dimension tfs vector to size n
loop from i = 1 to n
  
\footnote{See Equation B.5 in Appendix B, with parameters \( M = 2 \) and \( D = 9 \).}
set CUSUM = 0
set tfs[i] = 0
loop from j = 1 to 9
{
    CUSUM = max(0, CUSUM + random normal variate + o[j] - 0.5)
    if(CUSUM ≥ 2 & tfs[i] = 0) then tfs[i] = j
}
Assign to good.tfs vector only those tfs values greater than 0
print CED=average(good.tfs), se(good.tfs), PSD=length(good.tfs)/n

Question #3: Using simulation, plot the zero-state CED and PSD for 3 ≤ D ≤ 16 and M = 2?
Solution: Figure 7.8 shows the plots. CED varies from 2.1 days for an outbreak of 3 days of duration to a CED of 5.2 days for an outbreak of 16 days duration. The PSD varies from 0.72 for an outbreak of 3 days to 0.999 for an outbreak of duration 16 days.

Fig. 7.8. Plots of CED and PSD for Example 7.6 for 3 ≤ D ≤ 16 and M = 2.

The following example illustrates the application of the CUSUM to data that follow a Poisson distribution. In such cases, the CUSUM based on Equa-
tion 7.12 would not be appropriate, particularly for Poisson distributions with very small means.

Example 7.7. In a biosurveillance system, in the absence of a flu outbreak the daily counts of GI syndrome follow a Poisson distribution with $\lambda_0 = 5$. It is important to quickly detect a flu outbreak that increases the mean of the daily counts to $\lambda_1 = 10$.

Question #1: What is the value of $k$ for a CUSUM monitoring this shift in the mean of a Poisson distribution?

Solution: For this CUSUM, $k = \frac{\lambda_1 - \lambda_0}{\ln \lambda_1 - \ln \lambda_0} = \frac{10 - 5}{\ln(10) - \ln(5)} = 7.213$.

Question #2: Using simulation, what is the threshold for this CUSUM that achieves an ATFS of approximately 30 days?

Solution: Via simulation, $h = 2.7$ gives an ATFS of 29.5 days (s.e. = 0.1).

Question #3: Using simulation, what is the (estimated) CED and PSD for outbreaks with a duration of 8 days and with means from 6 to 15 (i.e., $6 \leq \lambda_1 \leq 15$)?

Solution: Figure 7.9 shows the simulation results in which CED varies from 4.1 days for an outbreak with $\lambda_1 = 6$ to a CED of 1.1 days for an outbreak with $\lambda_1 = 15$. The PSD varies from 0.54 to 1.0 over this range of $\lambda_1$ values.

![Fig. 7.9. Plots of CED and PSD for Example 7.7 for $6 \leq \lambda_1 \leq 15$.](image)
7.4 EWMA Detection Method

The exponentially weighted moving average (EWMA) control chart of Roberts (1959) is another popular statistical process control method. It calculates

\[ E_t = \lambda \bar{Y}_t + (1 - \lambda)E_{t-1}, \]  

(7.16)

where \( \bar{Y}_t \) is the sample mean of \( m \) observations taken at time \( t \) and \( 0 < \lambda \leq 1 \) is the smoothing parameter. Note the similarities to the exponentially weighted moving average models in section 5.2.2 of Chapter 5.

In Equation 7.16, \( E_t \) is a weighted average of past observations; smaller values of \( \lambda \) put more emphasis on past observations, while large values put more emphasis on recent observations. For \( \lambda = 1 \) the EWMA reduces to the Shewhart. The method starts at \( E_0 = \mu_0 \), where \( \mu_0 \) is the expected value of the process under \( F_0 \), and a signal is generated at time \( t \) if \( E_t \geq h \) or \( E_t \leq -h \).

The threshold is set as

\[ h = \mu_0 + L \frac{\hat{\sigma}}{\sqrt{m}} \sqrt{\frac{\lambda}{2 - \lambda} [1 - (1 - \lambda)^2 t]} \]

for some value of \( L \). Note that the threshold changes over time, particularly with small values of \( t \). Letting \( t \to \infty \), the asymptotic or steady-state threshold is

\[ h = \mu_0 + L \frac{\hat{\sigma}}{\sqrt{m}} \sqrt{\frac{\lambda}{2 - \lambda}}. \]

For individual observations, Equation 7.16 becomes

\[ E_t = \lambda Y_t + (1 - \lambda)E_{t-1}, \]

(7.17)

where as before the method starts at \( E_0 = \mu_0 \). For \( m = 1 \),

\[ h = \mu_0 + L \hat{\sigma} \sqrt{\frac{\lambda}{2 - \lambda} [1 - (1 - \lambda)^2 t]} \]

for some value of \( L \), and the steady-state threshold is

\[ h = \mu_0 + L \hat{\sigma} \sqrt{\frac{\lambda}{2 - \lambda}}. \]

A nice property of the EWMA is that it can be robust to non-normality of the observations, a good property to have when working with individual observations. In particular, Borror et al. (1999) found the EWMA is robust for \( 0.05 \leq \lambda \leq 0.1 \) and Borror et al. (1998) recommend using the EWMA form in Equation 7.17 for \( Y \sim \text{Pois}(\lambda_0) \), setting \( E_0 = \mu_0 = \lambda_0 \).

\[ ^3 \text{Potential notation confusion: } \lambda_0 \text{ and } \lambda_1 \text{ are parameters of the Poisson distribution while } \lambda \text{ is the EWMA smoothing parameter.} \]
7.4 EWMA Detection Method

7.4.1 Modified for Biosurveillance

The EWMA is a two-sided method, meaning it is designed to detect both increases and decreases in the mean of a distribution. Since biosurveillance is only concerned with increases, the EWMA is “reflected” (in the spirit of the CUSUM) to improve its performance in detecting increases. In addition, in biosurveillance the EWMA is either run on individual observations or residuals from a model of the systematic effects, so Equation 7.16 is modified to:

\[ E_t = \max[\mu_0, \lambda Z_t + (1 - \lambda)E_{t-1}] \].

(7.18)

The method starts at \( E_0 = \mu_0 \) and a signal is generated only when \( E_t \geq h \), where for \( \mu_0 = \mathbb{E}(Z) \) the threshold is

\[ h = \mu_0 + \hat{\sigma} \sqrt{\frac{\lambda}{2 - \lambda}} [1 - (1 - \lambda)^2t], \]

again for some value of \( L \). When run on standardized residuals, so that \( \mu_0 = 0 \) and \( \sigma = 1 \), the threshold is

\[ h = L \sqrt{\frac{\lambda}{2 - \lambda}} [1 - (1 - \lambda)^2t], \]

and asymptotically it is

\[ h = L \sqrt{\frac{\lambda}{2 - \lambda}}. \]

Then Equation 7.18 becomes

\[ E_t = \max[0, \lambda Z_t + (1 - \lambda)E_{t-1}] \]

with \( E_0 = 0 \).

Example 7.8. To illustrate the EWMA, apply Equation 7.18 to the ILI residuals data of Examples 7.1 and 7.8. Calculate the EWMA values, \( E_1, \ldots, E_{15} \), plot them, and then for \( h = 0.6 \) determine whether an outbreak may have occurred. For the purposes of these calculations, use \( \lambda = 0.2 \) and \( E_0 = 0 \).

Solution: The EWMA calculations are:
7.4.2 Implementation

Implementing the EWMA requires choosing a threshold $h$ and the smoothing parameter $\lambda$. As with the Shewhart and CUSUM methods, the choice of threshold should be based on smallest ATFS that can be accommodated.

For the EWMA, the ATFS is also a function of the smoothing parameter, the choice of which depends on the type of outbreak to be detected. Specifically, for larger and more immediate outbreaks, larger values of $\lambda$ will result in an EWMA that more quickly detects the outbreak. As previously mentioned, with $\lambda = 1$ the EWMA reverts to the Shewhart which is capable of detecting large outbreaks in one observation. Conversely, smaller outbreaks (of sufficient duration) are better detected with smaller values of $\lambda$.

For the EWMA based on standardized observations or residuals that are normally distributed, given a choice of $\lambda$, Table 7.3 can be used to determine the choice of $h$ needed to achieve a desired ATFS.

If the observations are not normally distributed, but the probability distribution of the observations or residuals can be specified, simulation can be used to determine the threshold to achieve a specific ATFS.
The following example illustrates how to apply and calculate the EWMA in the biosurveillance scenario first posed in Example 7.5.

Example 7.9. Using the data from an adaptive regression fit to daily ILI counts in Example 7.5 (0.819, 0.557, 1.733, 0.415, 0.155, -3.173, -0.549, -0.756, 1.060, 0.137, 0.744, 0.509, 1.206, 0.512, -1.347, 3.772, 1.142, 3.132, 2.846, 2.117), set-up and repeat the analysis using the EWMA.

Question #1: What value of the smoothing parameter ($\lambda$) should be used to detect an outbreak one standard deviation in magnitude and 10 days in duration?

Solution: In the SPC literature $0.1 \leq \lambda \leq 0.3$ is frequently recommended for detecting small changes in the mean. However, in biosurveillance terms, this recommendation is based on the assumption of an outbreak of infinite duration. When the duration is finite, as it will be in biosurveillance, the appropriate choice of $\lambda$ likely depends both on the magnitude and duration of the outbreak.
Table 7.3. ATFS for an EWMA for the mean shift of a standard normal distribution as a function of $\lambda$ and $L$ determined via simulation. Standard errors of the estimates less than 0.1.

<table>
<thead>
<tr>
<th>$\lambda$</th>
<th>0.1</th>
<th>0.2</th>
<th>0.3</th>
<th>0.4</th>
<th>0.5</th>
<th>0.6</th>
<th>0.7</th>
<th>0.8</th>
<th>0.9</th>
</tr>
</thead>
<tbody>
<tr>
<td>$L$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.000</td>
<td>10.4</td>
<td>8.2</td>
<td>7.2</td>
<td>6.8</td>
<td>6.5</td>
<td>6.2</td>
<td>6.1</td>
<td>6.2</td>
<td>6.2</td>
</tr>
<tr>
<td>1.125</td>
<td>13.5</td>
<td>10.3</td>
<td>9.0</td>
<td>8.2</td>
<td>7.8</td>
<td>7.5</td>
<td>7.4</td>
<td>7.4</td>
<td>7.4</td>
</tr>
<tr>
<td>1.250</td>
<td>17.5</td>
<td>12.9</td>
<td>11.1</td>
<td>10.0</td>
<td>9.5</td>
<td>9.2</td>
<td>9.0</td>
<td>9.1</td>
<td>9.2</td>
</tr>
<tr>
<td>1.375</td>
<td>22.6</td>
<td>16.4</td>
<td>13.8</td>
<td>12.5</td>
<td>11.7</td>
<td>11.4</td>
<td>11.2</td>
<td>11.2</td>
<td>11.4</td>
</tr>
<tr>
<td>1.500</td>
<td>29.3</td>
<td>20.7</td>
<td>17.4</td>
<td>15.7</td>
<td>14.6</td>
<td>14.1</td>
<td>13.9</td>
<td>14.0</td>
<td>14.3</td>
</tr>
<tr>
<td>1.625</td>
<td>37.7</td>
<td>26.4</td>
<td>22.0</td>
<td>19.7</td>
<td>18.5</td>
<td>17.8</td>
<td>17.6</td>
<td>17.8</td>
<td>18.2</td>
</tr>
<tr>
<td>1.750</td>
<td>49.2</td>
<td>34.0</td>
<td>28.2</td>
<td>25.3</td>
<td>23.7</td>
<td>22.8</td>
<td>22.5</td>
<td>22.8</td>
<td>23.6</td>
</tr>
<tr>
<td>1.875</td>
<td>63.9</td>
<td>43.4</td>
<td>36.3</td>
<td>32.6</td>
<td>30.6</td>
<td>29.6</td>
<td>29.3</td>
<td>29.8</td>
<td>30.9</td>
</tr>
<tr>
<td>2.000</td>
<td>83.6</td>
<td>57.0</td>
<td>47.5</td>
<td>42.6</td>
<td>40.1</td>
<td>38.8</td>
<td>38.6</td>
<td>39.4</td>
<td>41.1</td>
</tr>
<tr>
<td>2.125</td>
<td>100.0</td>
<td>75.0</td>
<td>62.5</td>
<td>56.5</td>
<td>53.3</td>
<td>51.8</td>
<td>51.7</td>
<td>52.9</td>
<td>55.2</td>
</tr>
<tr>
<td>2.250</td>
<td>145.8</td>
<td>99.5</td>
<td>83.6</td>
<td>75.9</td>
<td>71.9</td>
<td>70.1</td>
<td>70.2</td>
<td>71.8</td>
<td>75.5</td>
</tr>
<tr>
<td>2.375</td>
<td>195.4</td>
<td>133.8</td>
<td>113.0</td>
<td>103.2</td>
<td>98.3</td>
<td>96.3</td>
<td>96.6</td>
<td>99.2</td>
<td>104.9</td>
</tr>
<tr>
<td>2.500</td>
<td>264.6</td>
<td>182.4</td>
<td>155.0</td>
<td>142.7</td>
<td>136.5</td>
<td>134.2</td>
<td>135.1</td>
<td>139.2</td>
<td>147.4</td>
</tr>
<tr>
<td>2.625</td>
<td>362.3</td>
<td>252.0</td>
<td>215.9</td>
<td>200.1</td>
<td>192.8</td>
<td>190.2</td>
<td>192.0</td>
<td>198.2</td>
<td>210.5</td>
</tr>
<tr>
<td>2.750</td>
<td>503.3</td>
<td>353.3</td>
<td>305.2</td>
<td>285.1</td>
<td>276.2</td>
<td>274.0</td>
<td>276.9</td>
<td>286.4</td>
<td>304.5</td>
</tr>
<tr>
<td>2.875</td>
<td>708.5</td>
<td>502.8</td>
<td>438.3</td>
<td>412.3</td>
<td>402.5</td>
<td>400.4</td>
<td>405.9</td>
<td>420.1</td>
<td>447.5</td>
</tr>
<tr>
<td>3.000</td>
<td>1,014.9</td>
<td>727.2</td>
<td>639.6</td>
<td>606.4</td>
<td>595.1</td>
<td>594.7</td>
<td>604.0</td>
<td>625.6</td>
<td>667.3</td>
</tr>
<tr>
<td>3.125</td>
<td>1,470.7</td>
<td>1,068.7</td>
<td>948.6</td>
<td>905.9</td>
<td>894.4</td>
<td>897.7</td>
<td>912.8</td>
<td>946.1</td>
<td>1,009.6</td>
</tr>
</tbody>
</table>

Since there have not been any studies in the biosurveillance literature to guide the choice of $\lambda$, Figure 7.11 shows the results of a small simulation conducted to get some insight into the EWMA’s behavior for various $\lambda$, in terms of CED and PSD, for this type of outbreak.

The simulation used the asymptotic form of the threshold equation with the parameter $L$ set to achieve 30 days between false signals (by interpolating from Table 7.3). Then the zero-state CED and PSD were estimated using simulated random observations drawn from a $N(1,1)$ distribution. The results are shown in Figure 7.11 where $\lambda = 0.3$ looks to achieve nearly the lowest CED (3.9 days out of a range of 3.8 to 4.5) for nearly the highest PSD (0.969 out of a range of 0.973 to 0.909).

**Question #2:** Given the choice of $\lambda = 0.3$ from the first question, what value of $L$ should be used in the threshold to achieve a average time between false signals of approximately 30 days? What is the asymptotic threshold for that $L$?

**Solution:** Using Table 7.3, $L = 1.75$ results in 28.2 while $L = 1.875$ results in 36.3. Interpolating gives $L = 1.778$. Thus, the asymptotic threshold
is \( h = \hat{L}\sigma \sqrt{\frac{\lambda}{2 - \lambda}} = (1.778)(1)\sqrt{\frac{0.3}{2 - 0.3}} = 0.75 \).

**Question #3:** Given the threshold \( h = 0.75 \) and \( \lambda = 0.3 \), what is the average time between signal events (ATBSE) for the EWMA?

**Solution:** Using simulation (based on pseudo-code similar to that from Example 7.5, modified for the EWMA), the average time between signal events is estimated as ATBSE=28.6 (s.e. 0.09). In this case, the ATBSE is slightly shorter than the ATFS.

**Question #4:** Calculate the EWMA values over time, both when the EWMA is re-set after each signal and when it is not re-set, and plot them on a chart showing the threshold. Is there any evidence of an outbreak?

**Solution:** Table 7.4 shows the EWMA calculations and Figure 7.12 plots the EWMA statistics over time. Both show that signals occur at times \( t = 3 \) and \( t = 16 \). The former is a false positive, while the latter is a true signal since day 16 was the first day of the outbreak. The false positive is a result of setting the threshold fairly aggressively to achieve an ATFS of 30 days: with 20 days of observations the occurrence of a false positive is quite likely. This can be mitigated by choosing a larger ATFS which will result in a higher threshold.

After time 16, both versions of the EWMA continue to signal, continuously for the EWMA that is not re-set and intermittently for the EWMA that is re-set. Again, see the discussions in Chapter 6 on pages 183 and 187.
Question #4: Using the results from this example and Example 7.5, how does the performance of the EWMA compare to the CUSUM?

Solution: Tables 7.2 and 7.4 (and the plots in Figures 7.7 and 7.12) show that both methods, whether re-set or not, signal on day 16, the first day of the outbreak. As discussed in the solution to the previous question, the EWMA also has a false signal on day 3, and both methods give similar patterns of signals after day 16.

Table 7.4. EWMA values for Example 7.9, where $E_t = \max(0, 0.3 \times Z_t + 0.7 \times E_{t-1})$, both with and without re-setting after a signal. The bold $E_t$ values are signal times, where $E_t > h = 0.75$.

<table>
<thead>
<tr>
<th>$t$</th>
<th>$Z_t$</th>
<th>With re-setting:</th>
<th>Without re-setting:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$E_{t-1}$</td>
<td>$E_t$</td>
</tr>
<tr>
<td>1</td>
<td>0.819</td>
<td>0.000</td>
<td>0.246</td>
</tr>
<tr>
<td>2</td>
<td>0.557</td>
<td>0.246</td>
<td>0.339</td>
</tr>
<tr>
<td>3</td>
<td>1.733</td>
<td>0.339</td>
<td><strong>0.757</strong></td>
</tr>
<tr>
<td>4</td>
<td>0.415</td>
<td>0.000</td>
<td>0.125</td>
</tr>
<tr>
<td>5</td>
<td>0.155</td>
<td>0.125</td>
<td>0.134</td>
</tr>
<tr>
<td>6</td>
<td>-3.173</td>
<td>0.134</td>
<td>0.000</td>
</tr>
<tr>
<td>7</td>
<td>-0.549</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>8</td>
<td>-0.756</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>9</td>
<td>1.060</td>
<td>0.000</td>
<td>0.318</td>
</tr>
<tr>
<td>10</td>
<td>0.137</td>
<td>0.318</td>
<td>0.264</td>
</tr>
<tr>
<td>11</td>
<td>0.744</td>
<td>0.264</td>
<td>0.408</td>
</tr>
<tr>
<td>12</td>
<td>0.509</td>
<td>0.408</td>
<td>0.438</td>
</tr>
<tr>
<td>13</td>
<td>1.206</td>
<td>0.438</td>
<td>0.669</td>
</tr>
<tr>
<td>14</td>
<td>0.512</td>
<td>0.669</td>
<td>0.622</td>
</tr>
<tr>
<td>15</td>
<td>-1.347</td>
<td>0.622</td>
<td>0.031</td>
</tr>
<tr>
<td>16</td>
<td>3.772</td>
<td>0.031</td>
<td><strong>1.153</strong></td>
</tr>
<tr>
<td>17</td>
<td>1.142</td>
<td>0.000</td>
<td>0.343</td>
</tr>
<tr>
<td>18</td>
<td>3.132</td>
<td>0.343</td>
<td><strong>1.180</strong></td>
</tr>
<tr>
<td>19</td>
<td>2.846</td>
<td>0.000</td>
<td><strong>0.854</strong></td>
</tr>
<tr>
<td>20</td>
<td>2.117</td>
<td>0.000</td>
<td>0.635</td>
</tr>
</tbody>
</table>

Returning to the scenario in Example 7.6, this next example illustrates how to set up the EWMA and then assess its performance for various types of outbreaks.
Fig. 7.12. Plot of EWMA statistics for Example 7.9. The solid line is the EWMA with re-setting and the dotted line is the EWMA without re-setting. The horizontal dotted line is the threshold $h = 0.75$.

Example 7.10. Consider a biosurveillance system which, when there is no outbreak, the standardized residuals from a model of the daily counts has a standard normal distribution.

Question #1: Assuming $\lambda = 0.3$, what value of $L$ should be used for an ATFS of once per quarter?

Solution: Defining a quarter as 91 days and interpolating from Table 7.3 with $L = 2.25$ for 83.6 days and $L = 2.375$ for 113 days gives $L = 2.28$.

Question #2: Using simulation, with $L = 2.28$, what is the zero-state CED and PSD of an outbreak characterized by a triangle shaped mean shift with a duration ($D$) of 9 days and a maximum increase ($M$) of two standard deviations?

Solution: Via simulation, the zero-state CED is estimated to be 5.87 days (s.e. < 0.01) with an estimated PSD of 0.998.

Question #3: Using simulation, plot the zero-state CED and PSD for $3 \leq D \leq 16$ and $M = 2$?

Solution: Figure 7.13 shows the plots. CED varies from 2.3 days for an outbreak of 3 days of duration to a CED of 6.1 days for an outbreak of 16 days duration. The PSD varies from 0.522 for an outbreak of 3 days to 0.997 for an outbreak of duration 16 days. Note the similarity of these results to
Example 7.6 with the CUSUM.

*Question #4:* Using the results from this example and Example 7.6, how does the performance of the EWMA compare to the CUSUM?

*Solution:* Comparing the results for the CUSUM in Figure 7.8 to the results for the EWMA in Figure 7.13, the EWMA seems to have a slightly lower PSD for similar CED values for the shorter duration outbreaks and it has slightly higher CED values for similar PSD values for the longer duration outbreaks. This suggests the CUSUM is to be preferred for these types of outbreaks compared to the EWMA with smoothing parameter $\lambda = 0.3$. It’s left as an exercise to see if another choice of $\lambda$ would result in an EWMA with equivalent or better performance properties compared to the CUSUM.

---

**Fig. 7.13.** Plots of CED and PSD for Example 7.10 for $3 \leq D \leq 16$ and $M = 2$.

The following example, based on the scenario of Example 7.7, illustrates the application of the EWMA to data that follow a Poisson distribution.

*Example 7.11.* In a biosurveillance system, in the absence of a flu outbreak the daily counts of GI syndrome follow a Poisson distribution with $\lambda_0 = 5$. It is important to quickly detect a flu outbreak that increases the mean of the daily counts to $\lambda_1 = 10$. 
Question #1: What value of $L$ should be used to achieve an ATFS of 30 days for an EWMA with $\lambda = 0.3$ monitoring the shift in the mean of a Poisson distribution?

Solution: Using simulation on the standardized observations $(Z_t = (Y_t - 5)/\sqrt{5})$, $L = 1.88$ results in an ATFS of 30.1 days (s.e. 0.1).

Question #2: Using simulation, what is the estimated CED and PSD for outbreaks with $2 \leq \lambda_1 \leq 15$ and a duration of 8 days?

Solution: Figure 7.14 shows the simulation results in which CED varies from 4.1 days with $\lambda_1 = 6$ to a CED of 1.1 days for an outbreak with $\lambda_1 = 15$. The PSD varies from 0.59 to 1.0 over this range of $\lambda_1$s. Again, note the similarity to the CUSUM results in Figure 7.9 of Example 7.7.

Question #3: Using the results from this example and Example 7.7, how does the performance of the EWMA compare to the CUSUM?

Solution: Comparing the results for the CUSUM in Figure 7.9 to the results for the EWMA in Figure 7.14, the two methods perform very similarly for these outbreak scenarios.

![Fig. 7.14. Plots of CED and PSD for Example 7.11 for $6 \leq \lambda_1 \leq 15$.](image)
The Shewhart, CUSUM, and EWMA methods are certainly not the only methods used for EED. In fact, the CUSUM and EWMA methods are rarely applied to biosurveillance in spite of the fact that they are popular in industrial quality control. Other methods that are used are scan statistics and methods first implemented in the Early Aberration Reporting System where, interestingly, two of the original EARS methods are variants of the Shewhart method (but are often mistakenly called CUSUM methods).

7.5.1 Temporal Scan Statistics

The idea of the temporal scan statistic is to literally scan through a set of longitudinal data, looking at a fixed number of adjacent observations (a “window” of observations), typically with the goal of identifying the set of observations for which the total is largest.

Description

Mathematically, the temporal scan statistic for day $t$ is defined as

$$ S_t = \max_{1 \leq i \leq t} \sum_{j=\max(1,i-m+1)}^{i} Y_j $$

where $m$ is the fixed window size. Literally the sequence of data, $\{Y_1, Y_2, \ldots, Y_t\}$, is scanned and $S_t$ takes on the largest sum of $m$ sequential observations. A signal is generated for time period $t$ if $S_t \geq h$, for some threshold $h$.

For prospective monitoring this approach is inefficient since at time $t$

$$ S_t = \max\{S_{t-1}, T_t\} $$

where

$$ T_t = \sum_{j=\max(1,t-m+1)}^{t} Y_j. $$

That is, at time $t$ it is not necessary to re-scan through $\{Y_1, Y_2, \ldots, Y_{t-1}\}$ since it has already been scanned at time $t-1$. This also means that, for prospective monitoring, it is only necessary to monitor $T_t$, reducing the temporal scan statistic to monitoring

$$ T_t = \sum_{j=\max(1,t-m+1)}^{t} Y_j $$

and signaling at the first time $T_t \geq h$. 
The temporal scan statistic is essentially SPC’s moving average control chart, differing only in that the total of the window of observations are monitored rather than the average. It is also similar to the CUSUM and EWMA in the sense that it is monitoring the sum of the current observation and some historical data. However, it differs in the way the sum is calculated.

Performance Comparisons

Using standard SPC metrics, Han et al. (2009) found that the CUSUM and EWMA outperform the scan statistic for detecting increases in Poisson rates. Similarly, Joner et al. (2008a) found the CUSUM performs better compared to the temporal scan statistic when monitoring for increases in incidence rates under the assumption of independent Bernoulli observations. However, they also concluded that the scan method may be preferred in some applications because of its simplicity at a cost of relatively little loss in performance efficiency. On the other hand, Woodall et al. (2008) showed that the temporal scan methods can be slightly more effective than the CUSUM in the case where the duration of the outbreak is close to the width of the scan window.

7.5.2 EARS Methods and Associated Variants

The Early Aberration Reporting System (EARS) was designed to be a drop in biosurveillance system, meaning it is intended to provide enhanced surveillance for a short duration around a discrete event (e.g., the Olympic Games or a national political convention), generally for which little or no prior data exists (Henning, 2004). Since September 11, 2001, the EARS system has also been increasingly used as a standard surveillance system (CDC, 2007) by state and local health departments. EARS uses three methods entitled “C1-MILD,” “C2-MEDIUM,” and “C3-ULTRA.” The C1, C3, and a modified form of the C2 methods were also implemented in BioSense 1.0. BioSense 2.0 currently only uses a modified form of the C2 method.

Description

The C1, C2, and C3 methods are often described as CUSUM methods (Hutwagner et al., 2003a; Zhu et al., 2005). The EARS V4.5 User’s Guide states that the EARS methods “were developed based on a one-sided positive CUSUM” (CDC, 2006e, p. 4). Even the latest BioSense 2.0 guidance incorrectly states that the EARS C2 method is “based on [a] cumulative sum control chart (CUSUM) methodology” (CDC, 2012b, p. 1).

However, as implemented in the EARS SAS code, the C1 and C2 are actually Shewhart variants that use a moving sample average and sample standard deviation to standardize daily counts. The C1 uses the seven days
prior to the current observation to calculate the sample average and sample standard deviation. The C2 is similar to the C1 but uses the seven days prior to a two-day lag. The C3 is an *ad hoc* method that combines information from the current observation and previous two periods as described below.

Let $Y_t$ be the observed count for period $t$ representing, for example, the number of individuals arriving at a particular hospital emergency room with a specific syndrome on day $t$. The C1 calculates the statistic $C_{1t}$ as

$$C_{1t} = \frac{Y_t - \bar{Y}_t}{S_t}$$

(7.19)

where $\bar{Y}_t$ and $S_t$ are the moving sample mean and standard deviation, respectively, based on the previous seven day’s observations:

$$\bar{Y}_t = \frac{1}{7} \sum_{i=t-7}^{t-1} Y_i \quad \text{and} \quad S_t = \sqrt{\frac{1}{6} \sum_{i=t-1}^{t-7} [Y_i - \bar{Y}_t]^2}.$$  

If $S_t = 0$ then EARS sets it to a small positive number. As implemented in the EARS system, the C1 signals on day $t$ when the C1 statistic exceeds a threshold $h$, which is fixed at three sample standard deviations above the sample mean: $C_{1t} > 3$.

The C2 is similar to the C1, but incorporates a two-day lag in the mean and standard deviation calculations. Specifically, it calculates

$$C_{2t} = \frac{Y_t - \bar{Y}_t^*}{S_t^*}$$

(7.20)

where

$$\bar{Y}_t^* = \frac{1}{7} \sum_{i=t-3}^{t-9} Y_i \quad \text{and} \quad S_t^* = \sqrt{\frac{1}{6} \sum_{i=t-3}^{t-9} [Y_i - \bar{Y}_t^*]^2}.$$  

As with the C1 method, if $S_t^* = 0$ then EARS sets it to a small positive number and the C2 method signals on day $t$ when $C_{2t} > 3$.

The C3 combines current and historical data from day $t$ and the previous two days, calculating the statistic $C_{3t}$ as

$$C_{3t} = \sum_{i=t-2}^{t-1} \max (0, C_{2i} - 1).$$

(7.21)

The C3 method signals on day $t$ when $C_{3t} > 2$.

**BioSense Variants**

BioSense 1.0 initially incorporated the original C1, C2, and C3 methods. It then replaced the C2 method with two modified forms it called the “W2 count” (W2c) and “W2 rate” (W2r) methods.
The W2c method calculated the mean and standard deviation separately for weekdays and weekends using the relevant prior seven days of data with a two-day lag. That is, the sample mean for weekdays was the average of seven weekdays occurring prior to a two day lag, and the sample mean for weekends was the average of the seven weekend days prior to a two day lag. The specific days included in the averages varied by day of the week.

The W2r method calculated a standardized statistic based on the proportion of visits for a particular syndrome out of the total number of visits to a health care facility on a given day. BioSense 1.0 allowed users to set the W2c and W2r method thresholds using a recurrence interval methodology. See Szarka et al. (2011) for additional details about both the W2c and W2r methods.

BioSense 2.0 introduced a new modified version of the EARS C2 method. This modification, which BioSense 2.0 calls the “C2 - Proportion Method,” is similar to the W2r method in the sense that it the observations (the $Y_t$ in Equation 7.20) are the proportion of visits for a particular syndrome out of the total number of visits to a health care facility. Like the EARS C2 method, it uses a two-day lag in the calculation of the average and standard deviation of the proportions, and it allows users to set the baseline period over which these statistics are calculated as either 7, 14, or 28 days. The threshold default is 3.9 standard deviations, but the user can override the default with any value from 0.5 to 5.0. See CDC (2012b) for additional details.

**Performance Comparisons**

Fricker et al. (2008a) compared the CUSUM to the EARS C1, C2, and (original) C3 methods in the univariate and multivariate cases. They found that the CUSUM applied to residuals from an adaptive regression model with an 8-week sliding baseline outperformed the EARS methods. Szarka et al. (2011) compared the performance of the W2c and W2r methods to Shewhart and EWMA methods using adaptive thresholds, concluding that EWMA performed better than the Shewhart and both the Shewhart and EWMA performed better than the W2r and W2c methods.

**7.6 Discussion & Summary**

The univariate temporal methods presented in this chapter are commonly used in biosurveillance, particularly the Shewhart variants in EARS and the CUSUM. The EWMA, while popular in industrial SPC, is less commonly used in biosurveillance.

The term “CUSUM” is sometimes used in the biosurveillance literature to incorrectly describe other methods such as the EARS C1, C2, and C3 methods. Referring back to section 7.3, it should now be absolutely clear: The
EARS methods and their derivatives are not CUSUMs. Indeed, the EARS C1 and C2 methods and associated variants are really just the Shewhart method applied to monitoring the standardized residuals of a moving average model (such as that described in section 5.2.1 of Chapter 5).

Industrial SPC research has demonstrated that the CUSUM and EWMA, with appropriate choices of \(k\), \(\lambda\), and thresholds, have very similar detection capabilities. It has also demonstrated that the CUSUM and EWMA can detect smaller sustained mean shifts quicker than the Shewhart, while the Shewhart can more quickly detect larger mean shifts. Presumably these broad results carry over to biosurveillance, though no research has been published that confirms this conjecture. In addition, more research is required to better understand how these methods behave in the biosurveillance context of transient outbreaks and, in particular, under what conditions each is to be preferred.

In biosurveillance, these industrial methods are sometimes incorrectly applied implicitly (and perhaps unknowingly) assuming the data is normally distributed. For example, the normal variant of the CUSUM (Equation 7.12) is often used with count data when, in fact, the Poisson or negative binomial variants (Equations 7.13 and 7.15) would likely be more appropriate. On the other hand, the normal CUSUM variant may be perfectly appropriate when monitoring the residuals of a model used to account for and remove systematic effects present in biosurveillance data.

A key point is that the choice of EED method should be based on the underlying probabilistic behavior of the data to which it will be applied and the type of shift to be detected. Current biosurveillance EED practice tends to focus only on a few of the most commonly-known methods and applies them without much regard for the distribution of the data or that the methods are mainly designed to detect changes in distributional means. Similarly, the idea that the CUSUM or EWMA should be used if one is looking for smaller shifts or the Shewhart if one is looking for larger shifts does not seem to be well-known.

Given that the industrial SPC literature is replete with many different types of control charts designed for many different situations and types of data, deeper mining of that literature would likely benefit biosurveillance. For example, the industrial literature has developed the notion of the fast initial response control chart which is designed to more quickly detect changes that occur immediately after the chart is started. This idea is directly relevant to drop-in surveillance biosurveillance systems such as EARS.

**Additional Reading**

For those who would like to delve more deeply into univariate temporal methods, consider the following.
• “The Use of Control Charts in Health-Care and Public-Health Surveillance” (Woodall, 2006) is a comprehensive overview of medical and public health applications of control charts.

• Examples of the application of Shewhart, CUSUM, and EWMA detection methods in traditional medical settings and to public health surveillance include:
  – “Performance of Risk-adjusted Control Charts to Monitor In-hospital Mortality of Intensive Care Unit Patients: A Simulation Study” (Koetsier et al., 2012).
  – “A Novel Experience in the Use of Control Charts for the Detection of Nosocomial Infection Outbreaks” (Gomes et al., 2011).
  – “Assessing the Early Aberration Reporting System’s Ability to Locally Detect the 2009 Influenza Pandemic” (Hagen et al., 2011).
  – “A One-Sided MEWMA Chart for Health Surveillance (Joner et al., 2008b).
  – “Methods for Monitoring Influenza Surveillance Data” (Cowling et al., 2006).
  – “Approaches to Syndromic Surveillance When Data Consist of Small Regional Counts” (Rogerson & Yamada, 2004a).

• Introduction to Statistical Quality Control (Montgomery, 2004) is an authoritative yet accessible introduction to statistical process control methods from an industrial quality control perspective.

• Useful control chart references:
  – CUSUM: Cumulative Sum Charts and Charting for Quality Improvement (Hawkins & Olwell, 1998) is a comprehensive treatment of the CUSUM, mainly from an industrial quality control perspective. See also Ewan (1963), Gan (1991), and Woodall & Adams (1993).