FOUNDATIONS OF STATISTICAL QUALITY CONTROL

Richard E. Barlow, University of California at Berkeley

and

Telba Z. Irony[1], University of California at Berkeley

Abstract

The origins of statistical quality control are first reviewed relative to the concept of statistical control. A recent *Bayesian* approach developed at AT&T laboratories for replacing Shewart-type control charts is critiqued. Finally, a compound Kalman filter approach to an inventory problem, closely related to quality control and based on Bayesian decision analysis, is described and compared to other approaches.

Statistical Control

The control chart for industrial statistical quality control was invented by Dr. Walter A Shewhart in 1924 and was the foundation for his *Economic Control of Quality of Manufactured Product*–his 1931 book. (A highly recommended recent reference is Deming, 1986.) On the basis of Shewhart's industrial experience, he formulated several basic and important ideas. Recognizing that all production processes will show variation in product if measurements of quality are sufficiently precise, Shewhart described two sources of variation; namely

i) variation due to *chance causes* (called *common causes* by Deming, 1986);

ii) variation due to *assignable causes* (called *special causes* by Deming, 1986).

Chance causes are inherent in the system of production while assignable causes, if they exist, can be traced to a particular machine, a particular worker, a particular material, etc. According to both Shewart and Deming, if variation in product is only due to chance causes, then the process is said to be in *statistical control*. Nelson (1982) describes a process in statistical control as follows: "A process is said to have reached a state of *statistical control* when changes in

[1]Now at George Washington University, Washington, D.C. 20006.

This research was partially supported by the U.S. Air Force Office of Scientific Research (AFOSR-90-0087) to the University of California at Berkeley.

measures of variability *and* location from one sampling period to the next are no greater than statistical theory would predict. That is, assignable causes of variation have been detected, identified, and eliminated." Duncan (1974) describes chance variations: "If chance variations are ordered in time or possibly on some other basis, they will behave in a random manner. They will show no cycles or runs or any other defined pattern. *No specific variation to come can be predicted from knowledge of past variations.*" Duncan, in the last sentence, is implying statistical independence and not statistical control.

Neither Shewhart nor Duncan have given us a mathematical definition of what it means for a process to be in statistical control. The following example shows that statistical independence depends on the knowledge of the observer and, therefore, we think it should *not* be a part of the definition of statistical control.

Example

The idea of *chance causes* apparently comes from or can be associated with Monte Carlo experiments. Suppose I go to a computer and generate n *random* quantities normally distributed with mean 0 and variance 1. Since I know the distribution used to generate the observed quantities, I would use a $N(0,1)$ distribution to predict the $(n+1)^{st}$ quantity yet to be generated by the computer. For me, the process is *random* and the generated n random quantities provide no predictive information. However, suppose I show a plot of these n numbers to my friend and I tell her how the numbers were generated except that I neglect to tell her that the variance was 1. Then for her, x_{n+1} is not independent of the first n random quantities because she can use these n quantities to estimate the process variance and, therefore, better predict x_{n+1}.

What is interesting from this example is that for one of us the observations are from an independent process while for the other the observations are from a dependent process. But of course (objectively) the plot looks exactly the same to both of us. The probability distribution used depends on the state of knowledge of the analyst. I think we both would agree however that the process is in statistical control.

All authors seem to indicate that the concept of statistical control is somehow connected with probability theory although not with any specific probability model. We think de Finetti (1937, 1979) has given us the concept which provides the correct mathematical definition of statistical control.

Definition: Statistical control

We say that a product process is in *statistical control* with respect to some measurement variable, x, on units $1, 2, \ldots, n$ if and only if in our judgement

$$p(x_1, x_2, \ldots, x_n) = p(x_{i_1}, x_{i_2}, \ldots, x_{i_n})$$

for all permutations $\{i_1, i_2, \ldots, i_n\}$ of units $\{1, 2, \ldots, n\}$. That is, the units are *exchangeable* with respect to x in our opinion. This definition has two implications: namely that the order in which measurements are made is not important and, secondly, as a result, all marginal distributions are the same. It does not, however, imply that measurements are independent.

In addition, the process *remains* in statistical control if, in our judgement, future units are *exchangeable* with past units relative to our measurement variable.

The questions which concern all authors on quality control are:

(1) How can we determine if a production process is in statistical control?

and

(2) Once we have determined that a production process is in statistical control, how can we detect a departure from statistical control if it occurs?

The solution offered by most authors to both questions is to first plot the data. A plot of the measurements in time order is called a *run chart.* Run charts are also made of sample averages and sample ranges of equal sample sizes at successive time points. The grand mean is plotted and *control limits* are set on charts of sample averages and sample ranges. The process is judged to be in statistical control if

i) there are no obvious trends, cycles or runs below or above the grand mean;

ii) no sample average or sample range falls outside of *control limits.*

Samples at any particular time are considered to constitute a *rational sample* (i.e., in our terminology, to be exchangeable with units not sampled at that time). The only question is that of exchangeability of *rational samples* over time. In practice, *control limits* are based on a probability model for the rational samples and *all* observed sample averages and ranges over time.

The marginal probability model can, in certain cases, also be inferred from the judgement of exchangeability. If measurements are in terms of attributes; i.e., $x_i = 1$ (0) if the i^{th} unit is good (bad) and if the number of such measurements is conceptually unbounded, then it follows from de Finetti's representation theorem that

$$p(x_i = 1) = \int_0^1 p(x_i = 1|\theta)p(\theta)\,d\theta = \int_0^1 \theta p(\theta)\,d\theta$$

for some measure $p(\theta)d\theta$ and further, that x_1, $x_2,\ldots,x_n$ are conditionally independent given θ. In this case θ can be interpreted as the long run "chance" that a unit is good; i.e., $(\sum x_i)/n$ tends to θ with subjective probability one as n increases without limit. *Chance* in this case is considered a parameter – not a probability.

In general, however, exchangeability alone is too weak to determine a probability model and additional judgements are required to determine marginal probability distributions. Let $x_1, x_2, \ldots, x_n$ be exchangeable measurement errors. If, in addition, we judge measurement errors to be *spherically symmetric*; i.e., $p(x_1, x_2, \ldots, x_n)$ is invariant under rotations of the vector $(x_1, x_2, \ldots, x_n)$ and this for all n, then it follows that the joint probability function is a *mixture* of normal distributions and x_i given σ^2 is $N(0, \sigma^2)$ while $x_1, x_2, \ldots, x_n$ given σ^2 are conditionally independent. Also $(\sum x_i^2)/n$ tends to a limit, σ^2, with subjective probability one. For more details see Dawid (1986).

The problem of determining and justifying *control limits* remains. It was this problem which led Hoadley (1981) to develop his quality measurement plan critiqued in the next section. The usual method for computing *control limits* (e.g. Nelson, 1982) violates the likelihood principle. Basu (1988) has argued convincingly against such methods.

A Critique of the Quality Measurement Plan

A quality auditing method called the quality measurement plan or QMP was implemented throughout AT&T technologies in 1980 (see Hoadley, 1981). The QMP is a statistical method for analyzing discrete time series of quality audit data relative to the expected number of defects given standard quality. It contains three of the audit ingredients: defects assessment, quality rating and quality reporting.

A quality audit is a system of inspections done continually on a sampling basis. Sampled product is inspected and defects are assessed whenever the product fails to meet engineering requirements. The results are combined into a rating period and compared to a quality standard which is a target value of defects per unit. It reflects a trade-off between manufacturing cost, operating costs and customer need.

Suppose there are T rating periods: $t = 1, \ldots, T$ (current period). For period t, we have the following data from the audit:

$n_t =$ audit sample size;

$x_t =$ number of defects in the audit sample;

$s =$ standard number of defects per unit;

$e_t =$ expected number of defects in the sample when the quality standard is met; $e_t = sn_t$;

$I_t = \frac{x_t}{e_t}$ = quality index (measure of the defect rate).

I_t is the defect rate in units of standard defect rate. For instance, if $I_t = 2$, it means that twice as many defects as expected have been observed.

The statistical model used in QMP is a version of the Empirical Bayes model. The assumptions are the following:

1. x_t has a Poisson distribution with mean $n_t\lambda_t$, i.e. $(x_t|n_t\lambda_t) \sim Poi(n_t\lambda_t)$ where λ_t is the *true* defect rate per unit in time period *t*. If λ_t is reparametrized on a quality index scale, the result is:

$$\theta_t = \lambda_t/s = \text{true quality index.}$$

In other words, $\theta_t = 1$ is the standard value. Therefore, we can write:

$$(x_t|\theta_t) \sim Poi(e_t\theta_t).$$

2. For each rating period *t*, there is a true quality index θ_t. θ_t, $t = 1,\ldots,T$ is a random sample from a Gamma distribution with mean θ and variance γ^2. θ is called the *process average* and γ^2 is called the *process variance*. We can write $(\theta_t|\theta,\gamma^2) \sim Gamma(\theta^2/\gamma^2, \theta/\gamma^2)$. In this model, both θ and γ^2 are unknown.

3. θ and γ^2 have a joint prior distribution $\rho(\theta, \gamma^2)$.

The parameter of interest is θ_T given the past data, d_{T-1} and current data, x_T. Here $d_{t-1} = (x_1, x_2,\ldots,x_{T-1})$ and d_0 is a constant.

The model assumes that the process average, θ, although unknown, is fixed; i.e., the model assumes exchangeability. In reality θ may be changing. In order to handle this, the QMP procedures uses a moving window of six periods of data.

A suitable way to describe and to analyze the QMP model is via *probabilistic influence diagrams*. Probabilistic influence diagrams have been described by Shachter (1986) and Barlow and Pereira (1990).

A probabilistic influence diagram is a special kind of graph used to model uncertain quantities and the probabilistic dependence among them. It is a network with directed arcs and no directed cycles. Circular nodes (probabilistic nodes) represent random quantities and arcs into random quantities indicate probabilistic dependence. An influence diagram emphasizes the relationships among the random quantities involved in the problem and represents a complete probabilistic description of the model. The solution for the QMP model, i.e., the posterior distribution of θ_T given the past data, d_{T-1} and current data, x_T can be achieved through the use of influence diagrams operations, namely, node merging, node splitting, node elimination and arc reversal. These operations are described in Barlow and Pereira (1990). Figure 1 is an influence diagram representation corresponding to the QMP model.

The joint distribution for random quantities in the QMP model is completely defined by the influence diagram above. The absence of arrows into node (θ, γ^2) means that we start with the unconditional joint distribution of θ and γ^2. The arrows originating at note (θ, γ^2) and ending at nodes θ_t $(t = 1,\ldots,T)$ indicate that the distributions of θ_t are conditional on θ and γ^2. This means that the process is considered exchangeable, that is, the process average, θ, is constant over time. Finally, each node x_t is the sink of an arrow starting at node θ_t meaning that the distribution of the random quantity x_t is conditional on θ_t for each $t = 1,\ldots,T$.

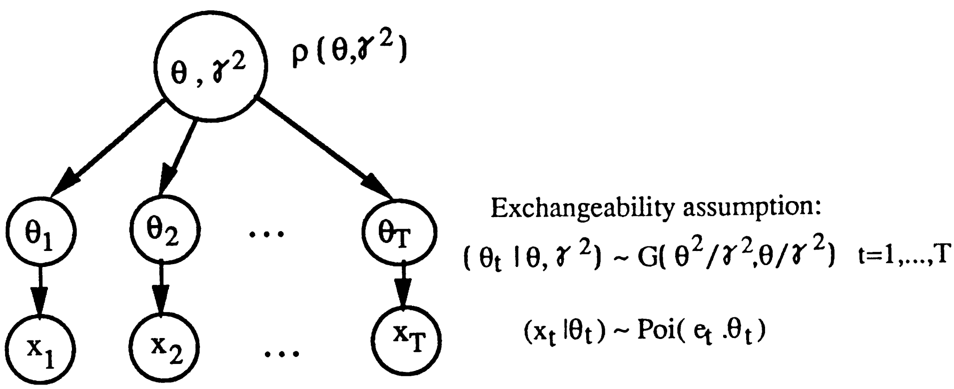


Figure 1

The QMP chart is a control chart for analyzing defect rates. Quality rating in QMP is based on posterior probabilities given the audit data. It provides statistical inference for the true quality process. Under QMP, a box and whisker plot (Figure 2) is plotted each period. The box plot is a graphical representation of the posterior distribution of θ_T given $d_{T-1} = (x_1,\ldots,x_{T-1})$ and x_T. The standard quality on the quality index scale is one. Two means twice as many defects as expected under the standard. Hence, the larger the quality index, the worse the process.

The posterior probability that the true quality index is less than the top whisker ($I_{99\%}$) is 99%. The top of the box ($I_{95\%}$), the bottom of the box ($I_{5\%}$) and the bottom whisker ($I_{1\%}$) correspond to probabilities of 95%, 5% and 1%, respectively.

The x is the observed value in the current sample, the heavy dot is the Bayes estimate of θ and the dash is the Bayes estimate of the current quality index (θ_T), a weighted average between x_T and θ.

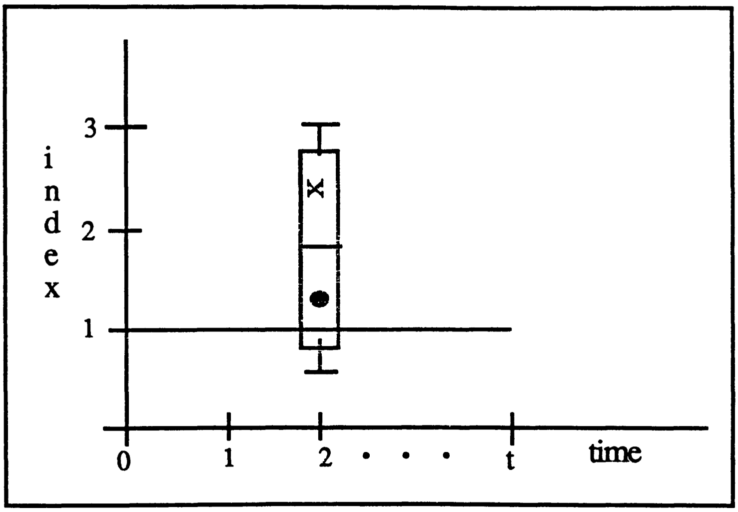


Figure 2

In a complete QMP chart (with all boxes), the dots are joined to show trends, i.e., it is assumed implicitly that the quality index θ_t may be changing from period to period.

1. Exception reporting

The objective of quality rating is to give a specific rule that defines quality exceptions and a measure (e.g., probability) associated with an exception. For QMP there are two kinds of exceptions:

a. A rating class is Below Normal (BN) if $I_{99\%} \geq 1$, i.e. if $P(\theta_T > 1) \geq 99\%$.

b. A rating class is on Alert if $I_{99\%} \leq 1 \leq I_{95\%}$, i.e., if $95\% \leq P(\theta_T > 1) \leq 99\%$. (See Figure 3.)

Products that meet these conditions are highlighted in an exception report.

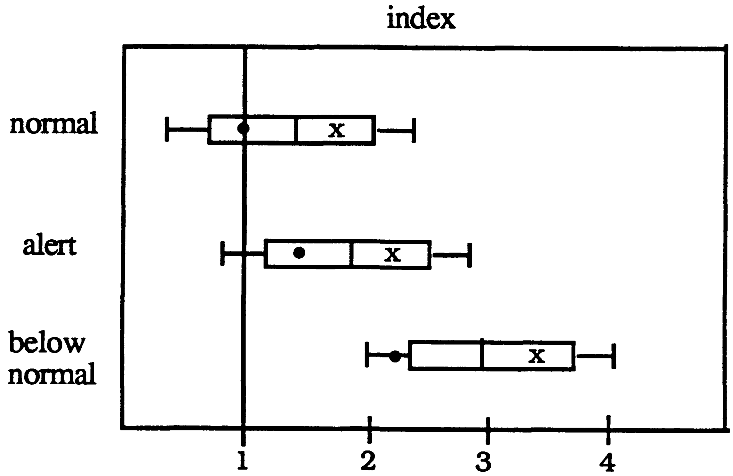
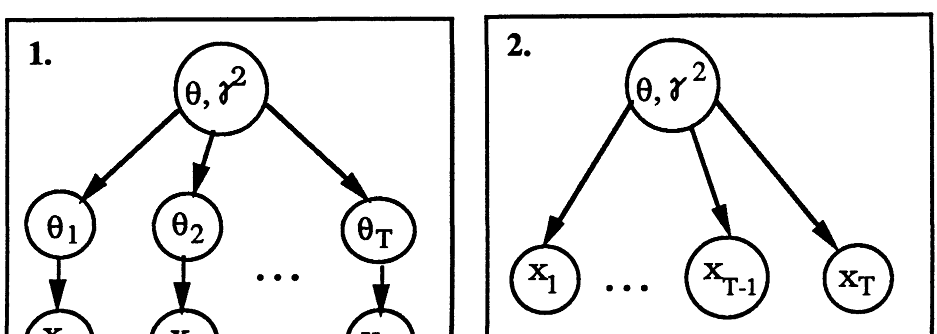


Figure 3

2. Posterior distribution of current quality

In order to get the exact solution for QMP, we have to compute the posterior distribution of θ_T given $d_{T-1} = (x_1,..,x_{T-1})$ and x_T. Hoadley (1981) describes a complicated mathematical "solution" to this model. It can be best understood through the following sequence of influence diagrams:



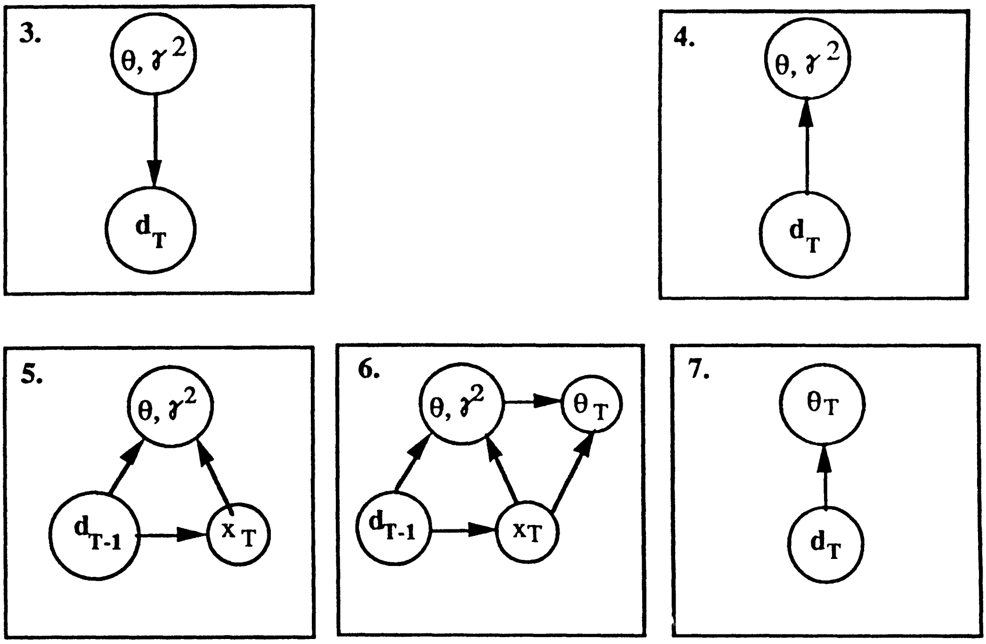


Figure 4

Diagram 1: Starting model: $(\theta,\ \gamma^2)\ \sim\ \rho(\theta,\ \gamma^2)$.

$(\theta_t|\theta,\gamma^2)\ \sim\ Gamma\left(\frac{\theta^2}{\gamma^2}, \frac{\theta}{\gamma^2}\right)$ for $t = 1,\ldots,T$. $(x_t|\theta_t)\ \sim\ Poi(e_t\theta_t)$.

Diagram 2: Nodes $\theta_1,\ldots,\theta_T$ are eliminated through integration.

$(x_t|\theta,\gamma^2)\ \sim$ Negative Binomial (Aitchison and Dunsmore, 1975).

Diagram 3: Nodes $x_1,\ x_2,\ldots,x_T$ are merged, i.e., the joint distribution of $d_T = (x_1,\ x_2,\ldots,x_T)$ is computed.

$(d_T|\theta,\gamma^2)\ \sim\ \prod_{t=1}^{T}$ Negative Binomials.

Diagram 4: The arc that goes from node (θ,γ^2) to node d_T is reversed, i.e., Bayes theorem is used to compute the posterior distribution of θ and γ^2 given d_T. $\rho(\theta,\gamma^2|d_T)$: posterior for θ and γ^2 given d_T.

Diagram 5: Node d_T is split. The joint distribution of $(x_1,\ x_2,\ldots,x_T)$ is written as the distribution of $d_{T-1} = (x_1,\ x_2,\ldots,x_{T-1})$ and the conditional distribution of x_T given d_{T-1}.

Diagram 6: Node θ_T is added again into the model. The distribution of θ_T given (θ,γ^2) and x_T is determined.

Diagram 7: Node (θ,γ^2) is eliminated through integration.

As we can see from the diagrams, the quality indexes $\theta_1,\ldots,\theta_T$ are eliminated in order to compute the distribution of the data, d_T, given θ and γ^2 and then, to compute the posterior distribution of θ and γ^2 given the data, d_T. Nevertheless, the parameter of interest is the current quality index, θ_T, which has to be re-introduced into the influence diagrams. This procedure is not correct. According to this, x_T is influencing θ_T twice in diagram 6. On one hand, directly (there is an arrow from x_T to θ_T), and on the other hand, through the posterior distribution of θ and γ^2 given d_T. In other words, node θ_T is eliminated (in influence diagram 2) and is added again (in influence diagram 6) and this is not the way one should solve an inference problem.

Even if this procedure were correct, the posterior distribution of θ_T would be a complex triple integral depending on the prior distribution assessed for θ and γ^2. This integral would have to be inverted in order to compute the QMP box chart. In other words, the exact solution is mathematically intractable, especially when many rating classes have to be analyzed each period. The result is a complicated algorithm (Hoadley, 1981) that computes all the parameters that are needed in order to construct the Gamma distribution for $\theta_T|d_T$. Hoadley's model assumes exchangeability, i.e., statistical control. Hence it does not provide an alternative to statistical control which can be used to decide whether or not the process is still in statistical control at the current time period. In the absence of an alternative model to exchangeability a better solution would have been to simply plot the standardized likelihoods (gamma densities) for θ_t at each time period based on the Poisson model. This would implicitly assume the θ_t's independent a priori.

A Kalman Filter Model for Inventory Control

As we have seen, the problem of quality control is to determine if and when a process has gone out of *statistical control*. The main difficulty with classical quality control procedures and also with the QMP model is that the models used assume the process is in statistical control and consider no alternative models to this situation. For coherent decision making, it is necessary to determine logical alternative models corresponding to a process out of statistical control.

In a paper dealing with inventory control (Barlow, Durst and Smiriga, 1984), a Kalman filter model was discussed from a decision theory point of view which could also be used for quality control problems. The paper describes an integrated decision procedure for deciding whether a diversion of Special Nuclear Material (SNM) has occurred. The problem is especially relevant for statistical analysis because it concerns (a priori) low probability events which would have high consequence if any occur. Two possible types of diversion are considered: a block loss during a single time period and a cumulative trickle loss over several time periods. The methodology used is based on a compound Kalman filter model.

Perhaps the simplest Kalman filter model is

$$\begin{aligned} y(t) &= \theta(t) + v(t) \\ \theta(t) &= \theta(t-1) + w_\theta(t), \end{aligned} \tag{1}$$

where $y(t)$ is the measured inventory at time period t and $\theta(t)$ is the actual inventory level. Our uncertainty with respect to measuring error is modeled by $v(t)$ while $w_\theta(t)$ models our uncertainty about the difference in the actual amounts processed between time period t–1 and t.

The $y(t)$ process will be in statistical control in the sense of the first section if and only if $w_\theta(t) \equiv 0$ for all t. For the inventory problem it seems reasonable to use (1) to model the process in the absence of any diversions. Later we will extend this model to account for possible diversions.

The compound Kalman filter model allows a decision maker to decide at each time period whether the data indicate a diversion. A block loss, by definition, will be a substantial amount which, it is hoped, will be detected at the end of the period in which it occurs. A trickle loss, on the other hand, is a smaller amount which is not expected to be detected in a single occurrence. A trickle loss may consist of a diversion or process holdup (or both), while a block loss is always a diversion. Two models are given for the process during each time period; in one, a block loss is assumed to have occurred, while in the other, only the usual trickle loss takes place. Since there are two models at each time period, a fully Bayesian analysis would required 2^n models at the end of n time periods, which is computationally untenable. A simple approximation is made which rests on the assumption that a block loss is a low-probability event. With this approximation only two models need be considered at each period, with all inference conditional on the assumption of no block loss in past periods (which has probability virtually equal to 1 as long as we have never come close to deciding that a block loss has occurred). By comparing these two models, we decide whether a block loss has occurred, and if we decide that it has an investigation is initiated. Since trickle loss, at least in the form of process holdup, is always assumed to occur, we will never decide that no trickle loss has occurred. We will either decide that a trickle diversion has occurred over several past periods, or we will decide that we as yet are unconvinced that a trickle loss beyond the normal holdup has occurred.

In Figure 5, $\beta(1)$, $\beta(2)$,..., etc., denote the amount of possible but unknown block losses during their respective time periods. The amount of possible but unknown trickle losses are denoted by $\tau(1)$, $\tau(2)$,..., etc. In our approach, we shall have two models: one model for block loss, say M_B, and one model for trickle loss, say M_T. We believe that model M_B holds with probability $p(M_B)$ and model M_T with probability $1 - p(M_B)$. Given data D, $p(M_B|D)$ is our updated probability for the block loss model M_B. If our updated probability for the block loss model is too high, then we will decide to investigate the possibility of a block loss. A decision regarding possible trickle loss, on the other hand, is based on the probability that loss beyond the normally expected holdup has occurred over several time periods; i.e.

$$P\{\tau(1) + \ldots + \tau(t) > c \mid D\}$$

where c is the normally expected holdup over t time periods. Thus, as indicated in Figure 5, our decision sequence is the customary one; at each time period we either decide that a

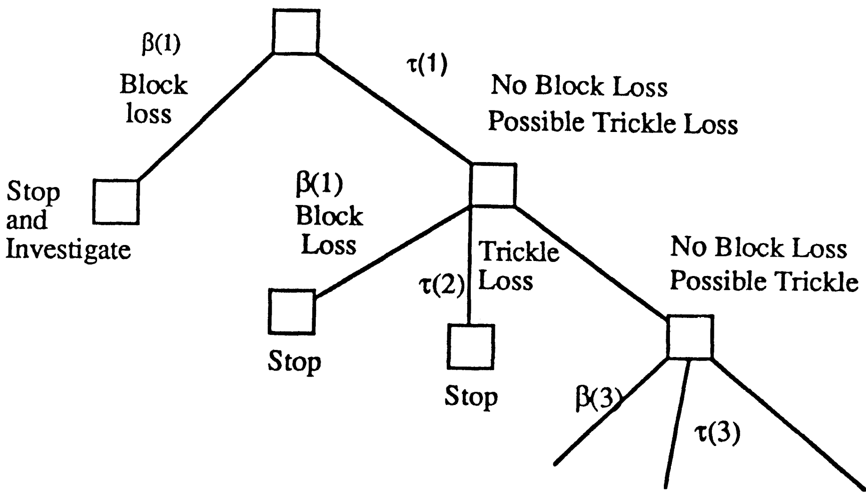


Figure 5 Diagram of possible decision sequences relative to diversion of special nuclear material

substantial block loss has occurred in the most recent period, that an unusually large trickle loss has been occurring in the past few periods, or that no block loss is likely to have occurred and that trickle loss is within acceptable limits. Our decision procedure does not formally permit the conclusion that a block loss has occurred other than within the most recent period, but it is shown that certain trickle alarms indicate the presence of an undetected block loss in some past period.

In order to clearly illustrate the salient features of these models, consider the simplified model (1) with only one measurement each period. At time t, $\theta(t)$

is the quantity of interest, but we can only observe $y(t)$. We assume that all variables in (1) are normally distributed.

The simplified trickle model M_T is:

$$y(t) = \theta(t) + v(t),$$

$$\theta(t) = \theta(t\text{-}1) - \tau(t) + w_\theta(t), \tag{2}$$

$$\tau(t) = \tau(t\text{-}1) + w_\theta(t).$$

The simplified Kalman filter block model M_B is:

$$y(t) = \theta(t) - \beta(t) + v(t),$$

$$\theta(t) = \theta(t\text{-}1) - \tau(t) + w_\theta(t), \tag{3}$$

$$\tau(t) = \tau(t\text{-}1) + w_\theta(t).$$

For the MB model, assume that $\beta(0)$ is also normally distributed.

The values of distribution parameters, even in our simplest model, must be carefully set. Too little initial uncertainty about possible trickle loss may make the model surprisingly unresponsive to large unexpected losses. A set of distribution parameters can be entirely self-consistent, seem on casual inspection quite sensible, and still produce undesirable behavior of the detection procedure. Thus distribution parameters should not be set arbitrarily or casually, but only after a careful assessment of process and loss uncertainties which takes into account the effect of the parameters on the resulting decision procedure.

The compound Kalman filter model provides a detection process which can compete with currently popular methods. Large block losses are detected handily, while somewhat smaller block losses are often detected later by the trickle model. Trickle losses consistently in excess of the expected holdup are detected rapidly, and smaller trickle losses are detected as the total amount of trickle loss becomes large.

With standard quality control methods, decisions must be made with a test of fixed significance level; otherwise, the frequentist interpretation of the test does not hold. Since we are dealing with probability distributions, we are not limited to setting a critical threshold and a critical probability. In fact, simulations indicate that it is best to take into account all the information given by the posterior probabilities. The results of a single hypothesis test, although a convenient summary, may be misleading. The user of these methods is encouraged to examine the probabilities of multiple critical regions, something which is not possible with standard quality control methods.

References

Aitchison, J. and Dunsmore, I. R. (1975): *Statistical Prediction Analysis*, Cambridge University Press.

Barlow, R. E., M. J. Durst, and N. G. Smiriga (1984): A Kalman filter model for determining block and trickle SNM losses, in *Low-Probability High-Consequence Risk Analysis*, eds. R. A. Waller and V. T. Covello, Plenum Press, New York, pp. 161-179.

Barlow, R.E. and Pereira, C. A. (1990): Conditional independence and probabilistic influence diagrams, in *Topics in Statistical Dependence, IMS Lecture Notes Series*, eds. H. W. Block, A. R. Sampson, and T. H. Savits, Institute of Mathematical Statistics, Hayward, Calif., pp. 1-14.

Basu, D. (1988): Statistical Information and Likelihood–A Collection of Critical Essays, *Lecture Notes in Statistics*, Number 45, Springer Verlag.

Dawid, A. P. (1986): A Bayesian view of statistical modelling, in *Bayesian Inference and Decision Techniques*, eds. P. Goel and A. Zellner, Elsevier Science Publishers, B. V., pp. 391-404.

Deming, W. J. (1986): *Out of the Crisis*, MIT Press, Cambridge, MA.

de Finetti, B. (1937): La Prevision: ses lois logiques, ses sources subjectives, *Annales de l'Institut Henri Poincare* 7, 1-68. Translated by Kyburg Jr. and H. E. Smokler (1980): *Studies in Subjective Probability*, 2nd ed., Robert E. Krieger Pub. Co., Huntington, New York.

de Finetti, B. (1979): Probability and exchangeability from a subjective point of view, *International Statistical Review* 47, 129-135.

Duncan, A. J. (1974): *Quality Control and Industrial Statistics*, 4th ed., Richard D. Irwin, Homewood, IL.

Hoadley, B. (1981): The quality measurement plan (QMP), *The Bell System Technical Journal* 60(2), 215-273.

Nelson, L. S. (1982): Control charts, *Encyclopedia of Statistical Sciences*, eds. Kotz and Johnson, J. Wiley Interscience, pp. 176-182.

Shachter, R. (1986): Evaluating influence diagrams, *Operations Research* 34, 871-882.

Shewart, W. A. (1931): *Economic Control of Quality Manufactured Product*, D. Van Nostrand, New York.