



Research Report
Department of Statistics
Göteborg University
Sweden

**Evaluations of
likelihood ratio methods
for surveillance**

**Marianne Frisé
Peter Wessman**

**Research Report 1996:3
ISSN 0349-8034**

| | | | |
|--------------------------|-----------------------|-----------------------|---|
| Mailing address: | Fax | Phone | Home Page: |
| Department of Statistics | Nat: 031-773 12 74 | Nat: 031-773 10 00 | http://www.stat.gu.se |
| Göteborg University | Int: +46 31 773 12 74 | Int: +46 31 773 10 00 | |
| S-411 80 Göteborg | | | |
| Sweden | | | |

EVALUATIONS OF LIKELIHOOD RATIO METHODS FOR SURVEILLANCE.

By Marianne Frisé and Peter Wessman
Göteborg University

Methods based on likelihood ratios are known to have several optimality properties. When control charts are used in practice, knowledge about several characteristics of the method is important for the judgement of which action is appropriate at an alarm. The probability of a false alarm, the delay of an alarm and the predictive value of an alarm are qualities (besides the usual ARL) which are described by a simulation study for the evaluations. Since the methods also have interesting optimality properties, the results also enlighten different criteria of optimality. Evaluations are made of the "The Likelihood Ratio Method" which utilizes an assumption on the intensity and has the Shiryaev optimality. Also, the Roberts and the CUSUM method are evaluated. These two methods combine the likelihood ratios in other ways. A comparison is also made with the Shewhart method, which is a commonly used method.

This work was supported by the Swedish Council for Research in the Humanities and Social Sciences.

AMS subject classifications. Primary 62L15 ; secondary 62N10.

Key words and phrases. Quality control, Warning system, Control chart, Predictive value, Performance, Shiryaev, Roberts, Shewhart, CUSUM.

Running title: Evaluations of likelihood ratio based surveillance

In many areas there is a need for continual observation of time series, with the goal of detecting an important change in the underlying process as soon as possible after it has occurred. In recent years there has been a growing number of papers in economics, medicine, environmental control and other areas dealing with the need of methods for surveillance. Examples are given in Frisé (1992) and Frisé (1994a). The timeliness of decisions is taken into account in the vast literature on quality control charts where simplicity is often a major concern. Also, the literature on stopping rules is relevant. For an overview, see the textbook by Wetherill and Brown (1990) the survey by Zacks (1983) and the bibliography by Frisé (1994 b).

Methods based on likelihood ratios are known to have several optimality properties. Evaluations are made of the LR method that is based on likelihood ratios and which in the case studied here has the Shiryaev optimality. Also, the Roberts method and the CUSUM method are evaluated. These two methods combine the likelihood ratios in other ways. A comparison is also made with the Shewhart method which is a commonly used method. When control charts are used in practice, it is necessary to know several characteristics of the method. Asymptotic properties have been studied by e.g. Srivastava and Wu (1993). Here properties for finite time of change are studied. The probability of a false alarm, the expected delay, the probability of successful detection and the predictive value are measures (besides the usual ARL) used for evaluations. Since the methods have interesting optimality properties, the results also enlighten different criteria of optimality.

In Section 1 some notations are given and there is a specification of the situation that is studied. In Section 2 the methods are described and their relations to each other and to different optimality criteria are discussed. In Section 3 comparisons based on a simulation study are reported. In Section 4 some concluding remarks are given.

1. NOTATIONS AND SPECIFICATIONS

The variable under surveillance is $X = \{X(t): t = 1, 2, \dots\}$, where the observation at time t is $X(t)$. It may be an average or some other derived statistic. The random process which determines the state of the system is denoted $\mu = \{\mu_t: t = 1, 2, \dots\}$.

The critical event of interest at decision time s is denoted $C(s)$. As in most literature on quality control, the case of shift in the mean of Gaussian random variables from an acceptable value μ^0 (say zero) to an unacceptable value μ^1 is considered. Only one-sided procedures are considered here. It is assumed that if a change in the process occurs, the level suddenly moves to another constant level, $\mu^1 > \mu^0$, and remains on this new level. That is $\mu_t = \mu^0$ for $t = 1, \dots, \tau-1$ and $\mu_t = \mu^1$ for $t = \tau, \tau+1, \dots$. We want to discriminate between

$$C(s) = \{\tau \leq s\} = \{\mu(s) = \mu^1\} \text{ and } D(s) = \{\tau > s\} = \{\mu(s) = \mu^0\}.$$

Here μ^0 and μ^1 are regarded as known values and the time τ where the critical event occurs is regarded as a random variable with the density

$$\pi_t = \Pr(\tau = t)$$

and $\sum \pi_t = 1 - \pi_\infty$. The intensity v_t of a change is

$$v_t = \Pr(\tau = t | \tau \geq t)$$

The aim is to discriminate between the states of the system at each decision time $s, s=1, 2, \dots$ by the observation $X_s = \{X(t): t \leq s\}$ under the assumption that $X(1) - \mu_1, X(2) - \mu_2, \dots$ are independent normally distributed random variables with mean zero and with the same known standard deviation (say $\sigma=1$). In some calculations below, where no confusion is possible, μ^1 is denoted μ and $\mu^0=0$ and $\sigma=1$ for typographical clarity.

Active surveillance (Frisén and de Maré 1991) is assumed. That means that the surveillance is stopped at an alarm. Thus, only one alarm is possible. A study of the properties of the first alarm at passive surveillance gives the same results.

2. METHODS

2.1 The Likelihood Ratio method

The problem of finding the method which maximizes the detection probability for a fixed false alarm probability and a fixed decision time was treated by de Maré (1980) and Frisé and de Maré (1991). The likelihood ratio (LR) method, discussed below, is the solution to this criterion. Different kinds of utility functions were discussed by Frisé and de Maré (1991). An important specification of utility is that of Girshick and Rubin (1952) and Shiryaev (1963). They treat the case of constant intensity where the gain of an alarm is a linear function of the difference $t_A - \tau$ between the time of the alarm and the time of the change. The loss of a false alarm is a function of the same difference. The utility can be expressed as $U = E\{u(\tau, t_A)\}$, where

$$u(\tau, t_A) \begin{cases} = l(\tau - t_A) & \text{if } \tau > t_A \\ = a_1(\tau - t_A) + a_2 & \text{else} \end{cases}$$

Their solution to the maximization of the expected utility is identical to the LR method for the situation specified in Section 1 (Frisén 1996).

The general method uses combinations of likelihood ratios. Even though methods based on likelihood ratios have been suggested earlier, for other reasons, the use in practice is (yet) rare. Here, the method is applied to the

shift case specified in Section 1. The "catastrophe" to be detected at decision time s is $C = \{ \tau \leq s \}$ and the alternative is $D = \{ \tau > s \}$.

The LR method has an alarm set consisting of those X_s for which the likelihood ratio exceeds a limit:

$$\begin{aligned} f_{X_s}(x_s | C) / f_{X_s}(x_s | D) &= p(x_s) = \\ &= \sum_{t=1}^{\infty} w(t)L(t) > K(s) \end{aligned}$$

where $w(t) = \Pr(\tau=t)/\Pr(\tau \leq s)$ and $L(t)$ is the likelihood ratio when $\tau=t$.

For the case of normal distribution, $C(s)=\{\tau \leq s\}$ and $D(s)=\{\tau > s\}$ specified in Section 1 we have

$$p(x_s) = h(s)p_s(x_s)$$

where

$$h(s) = \frac{\exp(-(s+1)(\mu^1)^2/2)}{\Pr(\tau \leq s)}$$

and

$$p_s(x_s) = \sum_{k=1}^s \pi_k \exp\left\{\frac{1}{2}k(\mu^1)^2\right\} \exp\left\{\mu^1 \sum_{u=k}^s x(u)\right\}$$

which is a nonlinear function of the observations.

In order to achieve the optimal error probabilities discussed by Frisén and de Maré (1991) an alarm should be given as soon as $p(x_s) > K(s)$.

In order also to achieve maximization of the utility of Shiryaev it is required that

$$\Pr(\tau \leq s | X_s = x_s) > K \Rightarrow p(x_s) > \frac{\Pr(\tau > s)}{\Pr(\tau \leq s)} \frac{K}{1-K}$$

where K is a constant. Now we must also consider the function $h(s)$.

2.2 Roberts method

Roberts (1966) suggested that an alarm is triggered at the first time s , for which

$$\sum_{t=1}^s L(t) > K$$

where K is a constant.

This is the limit of the LR method

$$\sum_{t=1}^s w(t)L(t) > K(s)$$

when ν tends to zero since both the weights $w(t)$ and the limit $K(s)$ tend to constants. Thus the Roberts method approximately satisfies the optimality criterion of Shiryaev (1963) for small values of the intensity ν . Roberts (1966) motivated the method by the conjecture that the intensity parameter ν has very little influence on the LR method, when ν is in the interval zero to 0.2 and thus the weights which depend on ν can be omitted. The Roberts method is sometimes called the Shiryaev-Roberts method. Pollak (1985) demonstrated that the Roberts procedure is asymptotically (as $K \rightarrow \infty$) minimax. Mevorach and Pollak (1991) examined the expected delay (see below) by simulations in a small sample setting and concluded that the difference between this method and the CUSUM method is small.

2.3 CUSUM

The cumulative sum

$$C_t = \sum_{i=1}^t (X(i) - \mu^0)$$

is used in several CUSUM- variants. The most commonly advocated variant gives an alarm for the first t for which $C_t - C_{t-i} > h + ki$ (for some $i=1,2,\dots, t$), where $C_0=0$ and h and k are chosen constants (Page 1954).

Sometimes the CUSUM test is defined in a more general way by likelihood ratios (Siegmund 1985 and Park and Kim 1990). This reduces to the method above in the case specified in Section 1.

In the simulation study below the slope $(\mu^0 + \mu^1)/2$ is used. The values $\mu^0=0$ and $\mu^1=1$ gives the slope $k=1/2$. The requirement of $ARL \underline{=} 11$ gives the parameter $h=0.985$. The short ARL^0 was chosen to make the computer time, necessary for the study, reasonable.

2.4 The Shewhart method

Shewhart (1931) suggested that an alarm is triggered as soon as a value which deviates too much from the target is observed. Frisén and de Maré (1991) demonstrated that this is the same as the LR method with $C = \{\tau = s\}$. That is $p(x_s) = LR(s)$. For the normal case described in Section 1 this reduces to $p(x_s) = x(s)$. The limit G for an alarm is calculated by the relation: $\Pr(X(s) > G | \mu = \mu^0) = 1/ARL^0$. $ARL^0 = 11$ gives $G = 1.3353$.

3. RESULTS

For the Shewhart method exact calculations were made. For the other methods simulations of 10 000 000 replicates were made for each point in the diagrams.

3.1 Probability of a false alarm

In all simulations the parameters of the different methods were chosen so that ARL^0 was equal for all methods (equal to 11) to make them comparable. To fix the value of ARL^0 is not the only way to do this, but it is in accordance with most comparative studies in this area.

Equal values of ARL^0 do not imply that the run length distributions are identical when $\mu \equiv \mu^0$. The distributions can have different shapes. This is demonstrated in Figures 1 and 2. In Figure 1 the probabilities of an alarm at a specific time point, when no change has occurred, are given for some different methods. The skewness of the distributions is less pronounced for the LR method with a great intensity parameter. For the chosen parameters, the distributions for the Shewhart and the CUSUM methods are very similar except at the first point. In Figure 2, the probability of an alarm no later than at t given that no change has occurred, $\alpha_t = \Pr(t_A \leq t \mid \mu_t = \mu^0)$, is illustrated. This is also the cumulative distribution function of the run length when the process is in control. The results of the LR method with parameters $v=0.001$ and $v=0.01$, cannot be distinguished from those of the Roberts method in the scale of the figures and are therefore not included. The results for $v=0.1$ are also very close to those of Roberts method and are included only in Figure 1 where more details can be seen.

A summarizing measure of the false alarm distribution is the probability of a false alarm, when the probability of a change has a geometric distribution with the intensity v . $\Pr(t_A < \tau)$ is illustrated as a function of v in Figure 4.

$$\Pr(t_A < \tau) = \sum_{t=1}^{\infty} \Pr(\tau=t) \Pr(t_A < t \mid \tau=t)$$

The first factor in the sum does not depend on the method but only on the true intensity v . The second factor depends only on the run length distribution when $\mu \equiv \mu^0$. Since the values of the ARL^0 are equal for all methods only the different shapes and not their locations will influence the false alarm probability. Thus only the very modest differences as are seen in Figure 4 can be expected.

3.2 Delay of an alarm

As was seen above, the correspondence to the level of significance in an ordinary test is not a value but a distribution. For the power the correspondence is still more complicated. To describe the ability of detecting a change we need a set of run length distributions. Some kind of summarizing measure is of value.

The distribution of t_A , the time of an alarm, when the change occurred before the surveillance started ($\tau=1$), that is $\mu=\mu^0$, is illustrated in Figure 3 for some methods.

The average run length under the alternative hypothesis, ARL^1 , is the mean number of decisions that must be taken to detect a true level change (that occurred at the same time as the inspection started). The part of the definition in the parenthesis is seldom spelled out but seems to be generally used in the literature on quality control. The values of ARL^1 for the methods and situations examined are given in Table 1.

Since the case $\tau=1$ of is not the only case of interest the expected delay is calculated also for other values $\tau=t$.

$$E(t_A - \tau | t_A \geq \tau, \tau=t)$$

is given in Figure 5. For $\tau=1$ the values of this function equal the values of $ARL^1 - 1$. The differences in shapes of these curves demonstrate the need for other measures than the conventional ARL. Although the Roberts method has worse ARL^1 , and thus worse delay for a change at $\tau=1$, than the Shewhart method, it is better for all other times of change. The CUSUM and the Shewhart methods are very much alike for $\tau=1$ but the CUSUM is here much better for other change points.

In some applications the loss of a delay is directly proportional to the expected value of the delay. The expected delay also with the respect to the distribution of τ is:

$$E(t_A - \tau | t_A \geq \tau)$$

It is given as a function of v in Figure 6, for the case when the distribution of τ is geometrical with the intensity v . Since the computer time necessary for reliable values for small values of v is too great, these values were not given in the figure. The shapes of the curves for small values are similar to the curve (exactly calculated) for the Shewhart method. When v tends to one, the expected delay tends to ARL^{-1} .

In some applications there is a limited time available for rescuing actions. Then, the expected value of the difference $\tau - t_A$ is not of main interest. Instead of using the expected value, the probability that the difference does not exceed a fixed limit is used. The fixed limit, say d , is the time available for successful detection.

The probability of successful detection,

$$PSD(\tau, d) = \Pr(t_A - \tau \leq d | t_A \geq \tau).$$

was suggested by Frisén (1992) as a measure of the performance. It is illustrated in Figure 7. The PSD is better for the CUSUM than for the Shewhart method for $d=3$. The shape of the curve for CUSUM with the present parameters is very similar to the constant curve for Shewhart. The Roberts and the LR methods have a worse probability of detection of a change which happens early but better for late changes.

3.3 Predicted value

The predictive value $PV(t) = \Pr(\tau \leq t | t_A = t)$ has been used as a criterion of evaluation by Frisén (1992), Frisén and Åkermo (1993) and Frisén and Cassel (1994). It is illustrated in Figure 8. The price for the high probability of detection of a change in the beginning of the surveillance (as was demonstrated in Figure 7) for the CUSUM and the Shewhart method is that the early alarms are not reliable.

4. CONCLUDING REMARKS

At large, the properties differ between the LR methods on one hand and the Shewhart and the CUSUM methods on the other, in the simulations. In comparison with this, the choice of the intensity parameter of the LR method has very little influence on the performance. The results here confirm the conjecture by Roberts (1966) about the robustness of the LR method.

That the Roberts method is the limit of the LR method when the intensity v tends to zero is also seen by the simulated results. The smaller the intensity parameter, the smaller the difference between the LR method and the Roberts method. Simulations were made also for $v=0.001$ and $v=0.01$ but these results were not included in the figures since the differences to those of the Roberts method are less than the line width. The results for $v=0.1$ were included only in the first figure where more details are given. The Roberts method is a good approximation of the optimal LR method and thus approximately optimal, for small values of v .

Sometimes the LR method is considered to be a Bayesian method while the Roberts method is considered a frequentistic one. Here, however all evaluations are made in the frequentistic framework. No Bayesian assumptions are necessary for the LR method. The properties of the method will be better if the intensity parameter is not far from the actual intensity. However, since the LR method is very robust for misspecification of the value of the intensity parameter v the gain with a precise search for the best value of the intensity parameter might not be worthwhile for a specific application.

The optimal method when the intensity is great should intuitively increase the probability of early alarms. Some of the present results might seem surprising in this light and will now be discussed.

The shapes (see Figures 1-3) of the distributions of the alarm time t_A might seem surprising. The probability of a very early alarm by the LR method

with a low intensity parameter is greater than for a large value of the parameter. However, for a low intensity the probability of a late change is great and thus a thick tail of the distribution of t_A is appropriate. As the expected value ARL^0 is fixed the only possibility is a high probability in the beginning. This also causes the differences in false alarm probabilities in Figure 4.

The greater expected delay (see Figures 5 and 6) for the LR method with a great intensity parameter v might seem surprising. For a fixed false alarm loss l the delay will be less for greater values of the intensity parameter v . However, now we have a fixed ARL^0 , and thus a greater value of l , which implies that the delay is increased to make the locations of the distributions of t_A equal. The only difference between the distributions is the shape and this difference causes the expected delay to be greater for the greater values of the intensity parameter.

The result by Mevorach and Pollak (1991), that the expected delay is similar for the Robert and CUSUM methods, was not confirmed. They studied an artificial quasi-stationary situation where the time of change, τ , does not have any influence. For a more realistic situation and the low ARL^0 used here the CUSUM in many aspects is very similar to the Shewhart method and not to the Roberts method.

Figure 7 demonstrates that the Roberts method has higher probability of a quick (within three time units) detection than the Shewhart method, unless the change occurs very early. For the LR(0.5) method this difference to the Shewhart method is still more pronounced.

In Figure 8 the predicted value of an alarm is given. This reflects the trust you should have in an alarm. The Roberts method has a relatively constant predicted value. This means that the same kind of action is appropriate both for early and late alarms.

REFERENCES

Frisén, M. (1992), Evaluations of methods for statistical surveillance, *Statistics in Medicine*, 11, 1489-1502.

Frisén M. (1994a) Statistical Surveillance of Business Cycles. *Research report*, 1994:1, Department of Statistics, Göteborg University.

Frisén M. (1994b) A classified bibliography on statistical surveillance. *Research report*, 1994, Department of Statistics, Göteborg University.

Frisén M. (1996) Characterization of methods for surveillance by optimality. Manuscript.

Frisén M. and Cassel C. (1994) Visual evaluations of statistical surveillance. *Research report*, 1994:3, Department of Statistics, Göteborg University.

Frisén, M. and de Maré, J. (1991), Optimal surveillance, *Biometrika*, 78, 271-280.

Frisén, M. and Åkermo, G. (1993), Comparison between two methods of surveillance: exponentially weighted moving average vs CUSUM, *Research report*, 1993:1, Department of Statistics, Göteborg University.

Girshick, M. A. and Rubin, H. (1952), A Bayes approach to a quality control model, *Ann. Math. Statist.*, 23, 114-125.

Maré, J. de (1980), Optimal prediction of catastrophes with application to Gaussian processes, *Ann. Prob.* 8, 841-850.

Mevorach, Y. and Pollak, M. (1991), A small sample size comparison of the CUSUM and Shiryaev-Roberts approaches to changepoint detection, *American J. of Mathematical and Management Sciences*, 11, 277-298.

Page, E. S. (1954), Continuous inspection schemes, *Biometrika*, 41, 100-114.

Park, C. S. and Kim, B. C. (1990) A CUSUM chart based on log probability ratio statistic, *J. Korean Statistical Society*, 19, 160-170.

Pollak, M. (1985) Optimal stopping times for detecting changes in distributions. *Ann. Statist.* 13, 206-227.

Roberts S. W. (1966) A comparison of some control chart procedures. *Technometrics*, 8, 411-30.

Shewhart, W. A. (1931) *Economic Control of Quality Control*. Reinhold Company, Princeton N.J.

Shiryaev, A. N. (1963), On optimum methods in quickest detection problems, *Theory Probab. Appl.*, 8, 22-46.

Siegmund, D. (1985), *Sequential analysis. Tests and confidence intervals*, Springer.

Srivastava, M.S. and Wu, Y (1993) Comparison of EWMA, CUSUM and Shiryaev-Roberts procedures for detecting a shift in the mean. *Ann. Statist.*, 21, 645-670.

Wetherill, G.B. and Brown, D.W. (1990), *Statistical process control*, London: Chapman and Hall.

Wessman, P. (1996) Some principles for surveillance adopted for multivariate processes with a common change point. *Research report*, 1996:2, Department of Statistics, Göteborg University.

Zacks, S. (1983) Survey of classical and Bayesian approaches to the change-point problem: Fixed sample and sequential procedures of testing and estimation. in *Recent advances in statistics*, 245-269.

DEPARTMENT OF STATISTICS
GÖTEBORG UNIVERSITY
S-41180 GÖTEBORG
SWEDEN

| Method | ARL¹ |
|---------------|------------------------|
| CUSUM | 2.61 |
| Shewhart | 2.71 |
| Roberts | 3.00 |
| LR(0.001) | 3.00 |
| LR(0.01) | 3.01 |
| LR(0.1) | 3.07 |
| LR(0.5) | 3.85 |

Table 1. The expected value of the alarm time t_A , when the change occurs immediately ($\tau=1$), and the parameters are set to make the expected value, when no change occurs, the same for all methods ($ARL^0=11$).

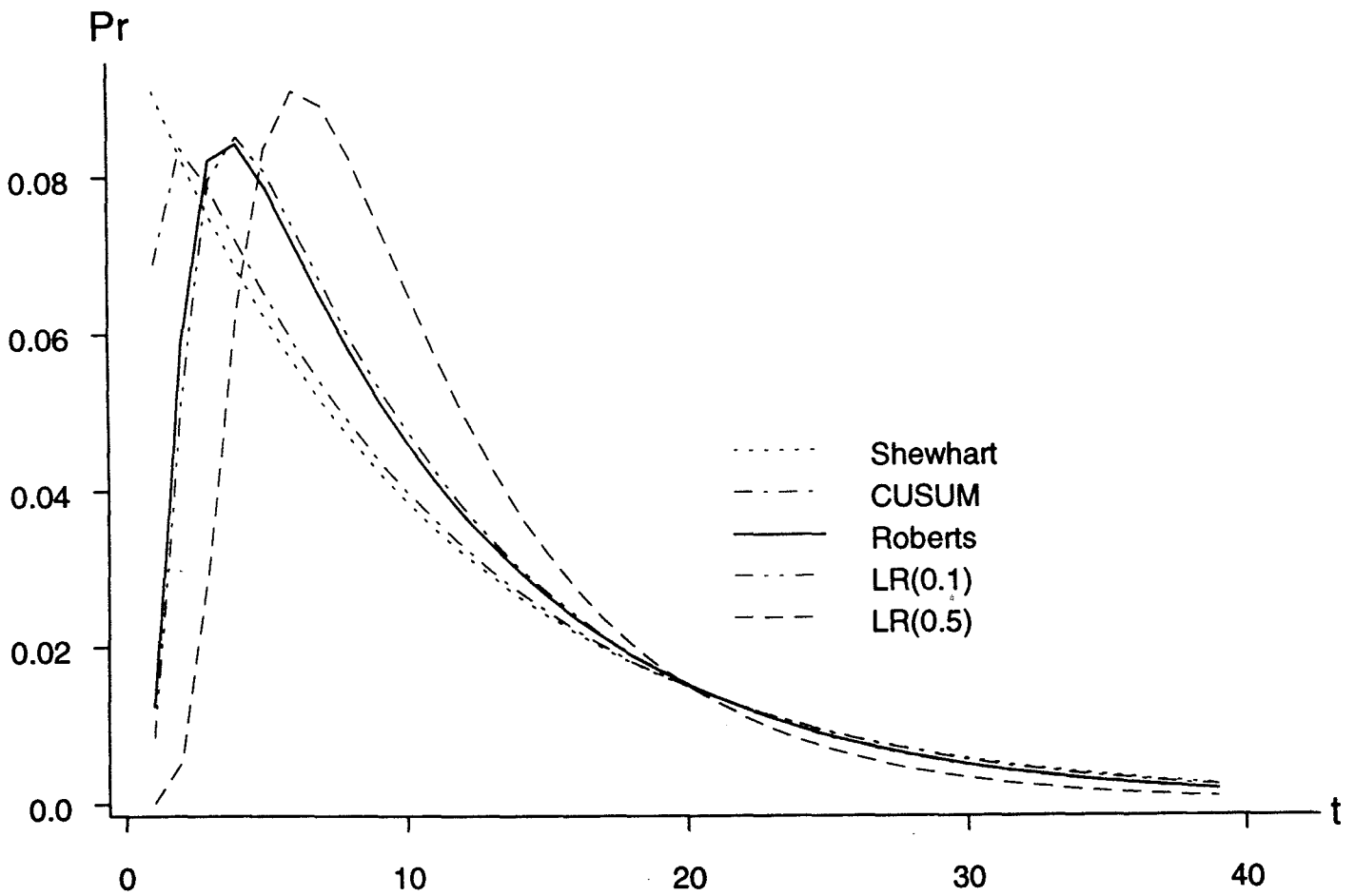


Figure 1. The density of the time of alarm, $\Pr(t_A=t|\mu_i=\mu^0)$. This is the probability of an alarm at time t , when no change has occurred. The expectations, ARL^0 , of the distributions are all equal to 11.

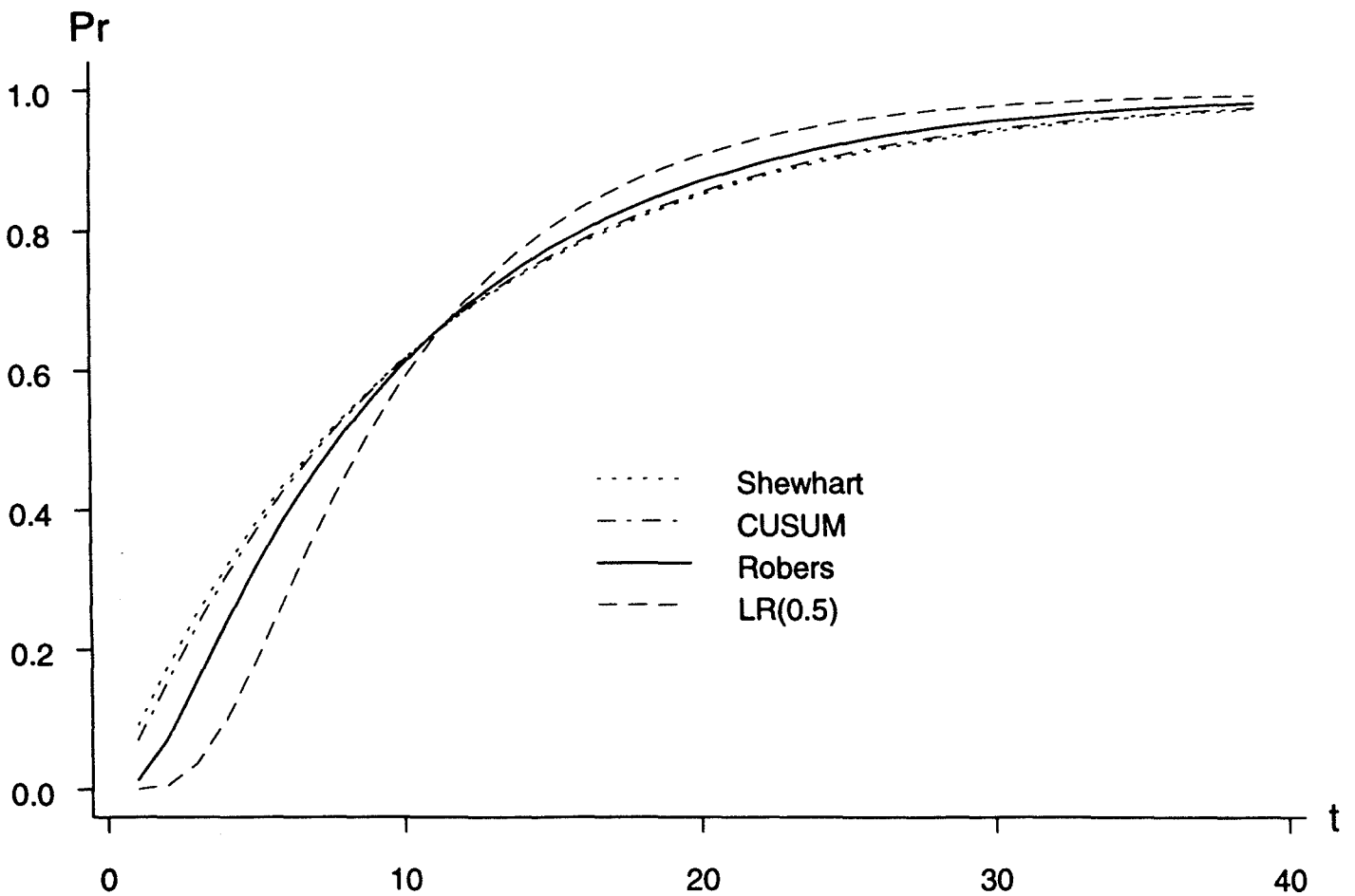


Figure 2. The distribution of the time of an alarm, when no change has occurred, $\alpha_i = \Pr(t_A \leq t | \mu_i = \mu^0)$. The expectations, ARL^0 , of the distributions are all equal to 11.

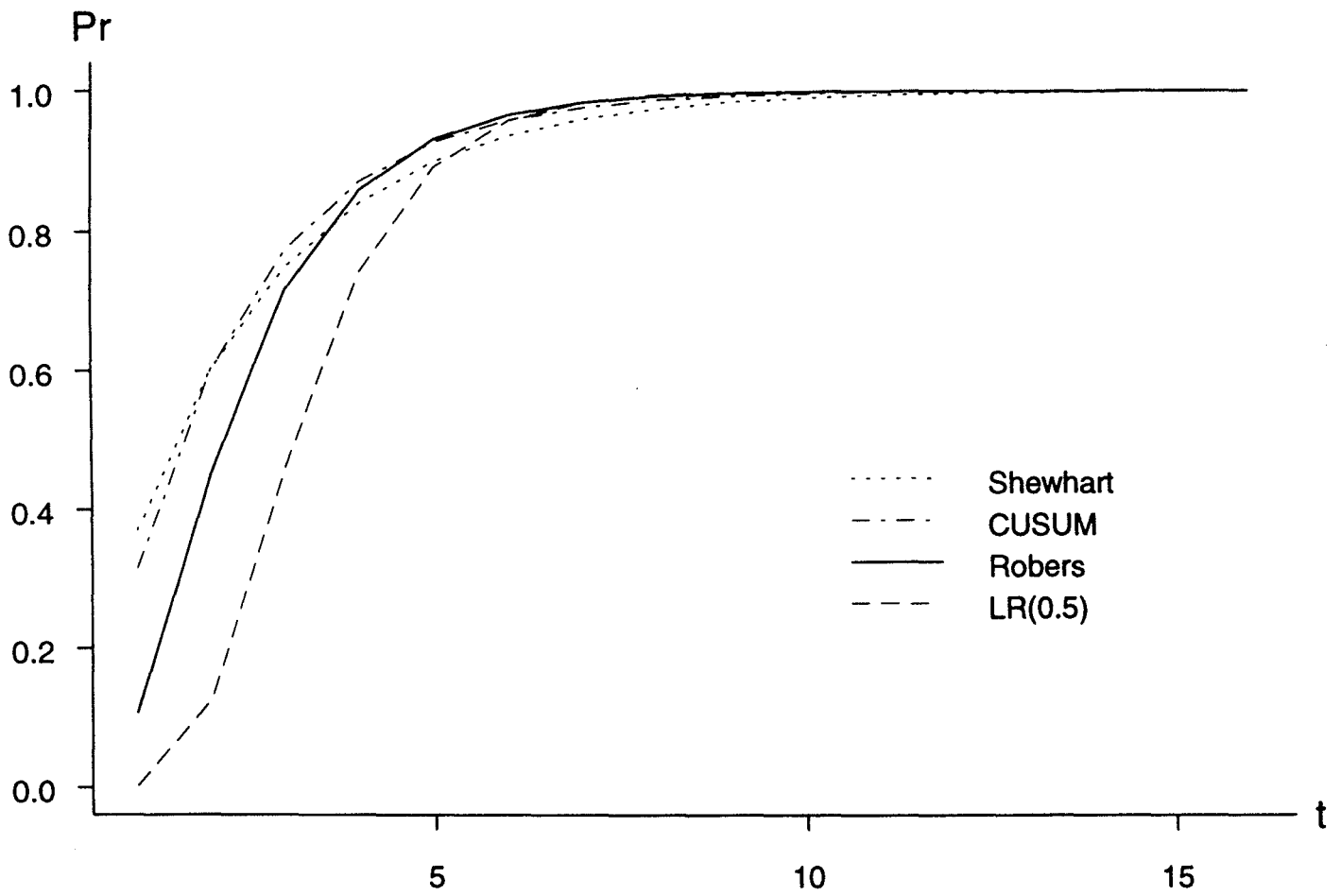


Figure 3. The distribution of the time of an alarm, when the change occurred before the surveillance started, $Pr(t_A \leq t | \mu_i = \mu^1)$. The expectations, ARL^0 , of the distributions are all equal to 11.

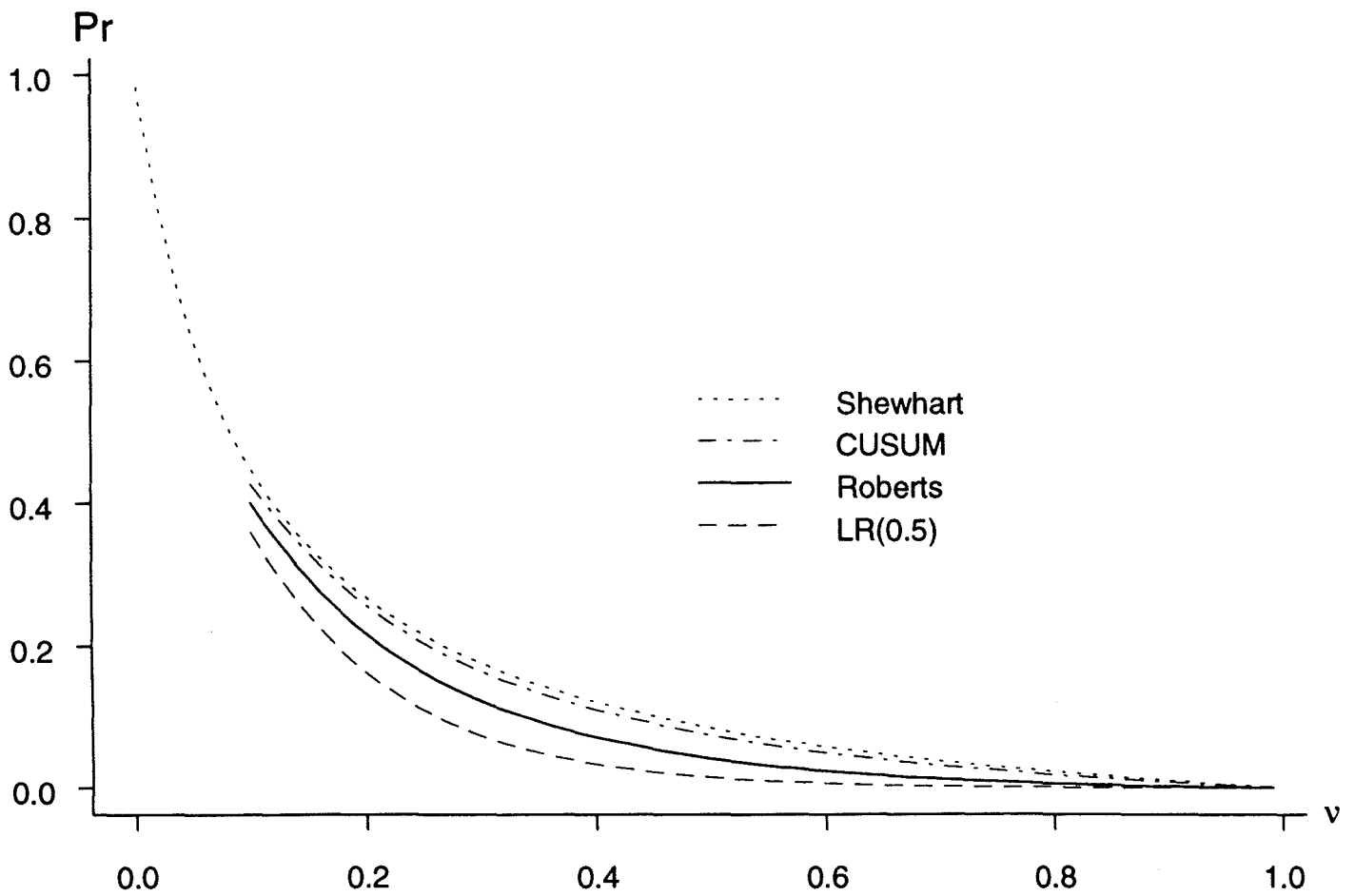


Figure 4. The expected probability of a false alarm $Pr(\tau > t_A)$ when the distribution of τ is geometrical with intensity ν .

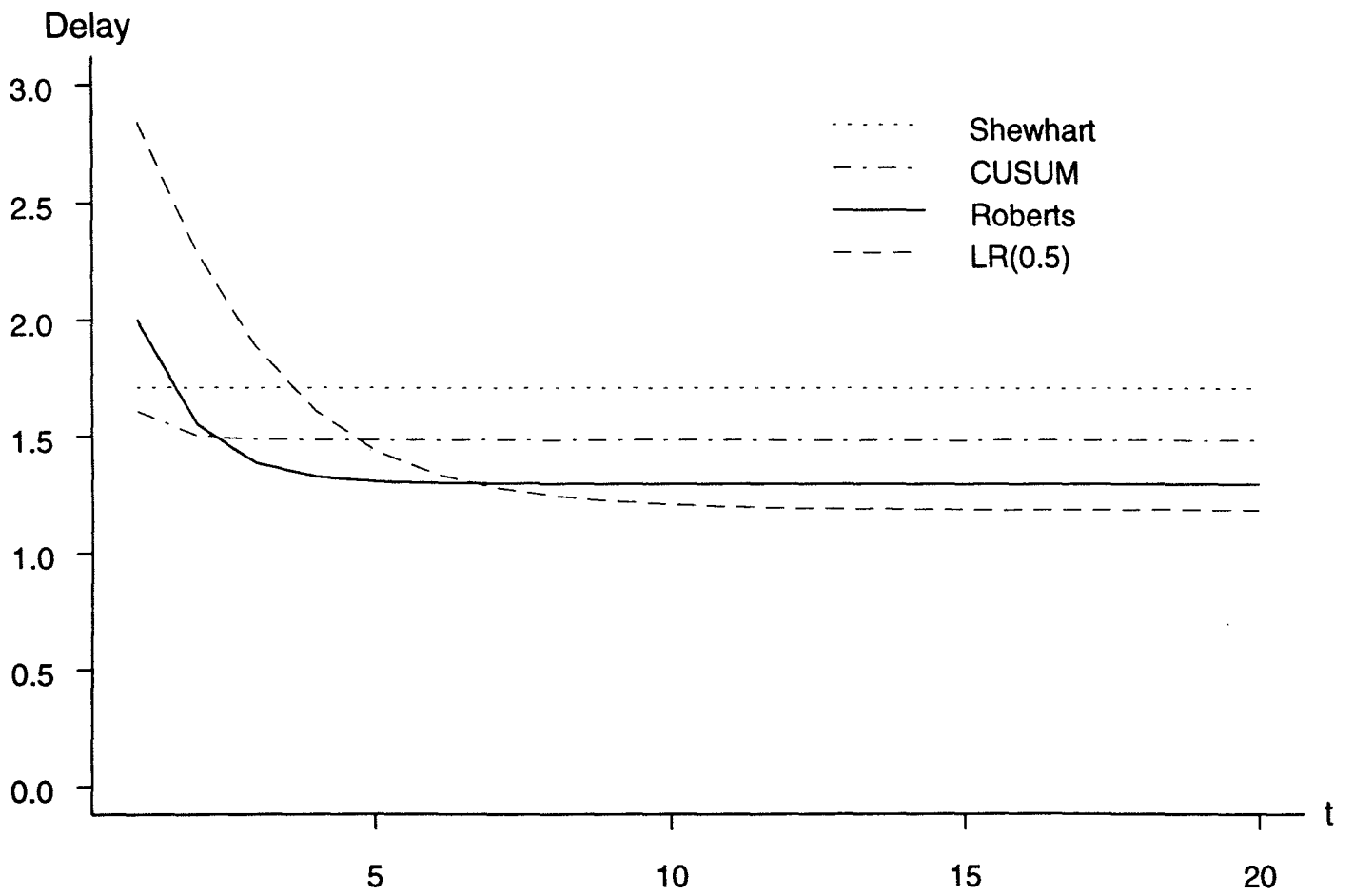


Figure 5. The expected delay after a change at time $\tau=t$. That is:
 $E(t_A - \tau | t_A \geq \tau, \tau=t)$.

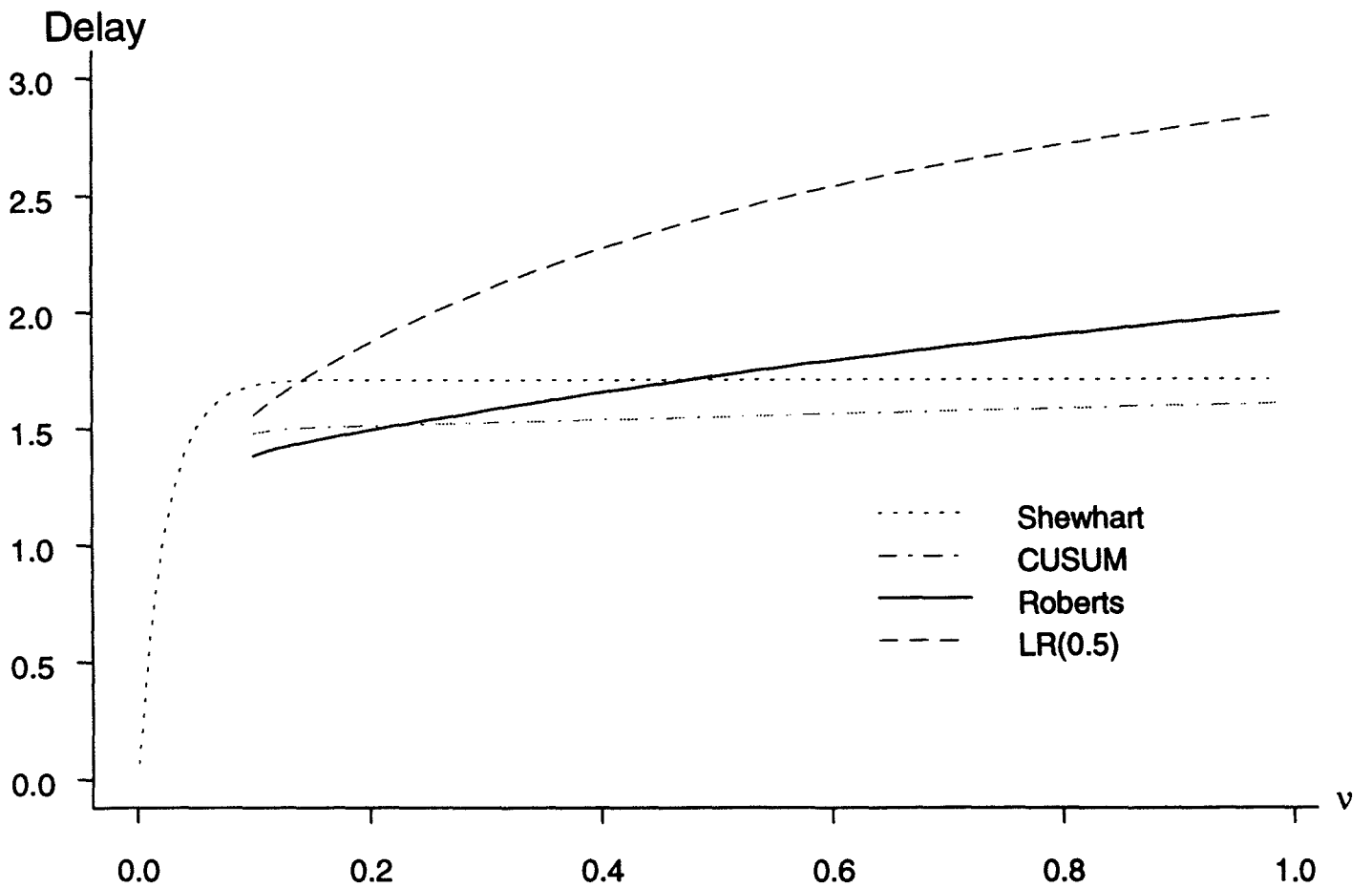


Figure 6. The expected delay when τ has a geometric distribution with intensity v . That is:
 $E(t_A - \tau | \tau \geq t_A)$

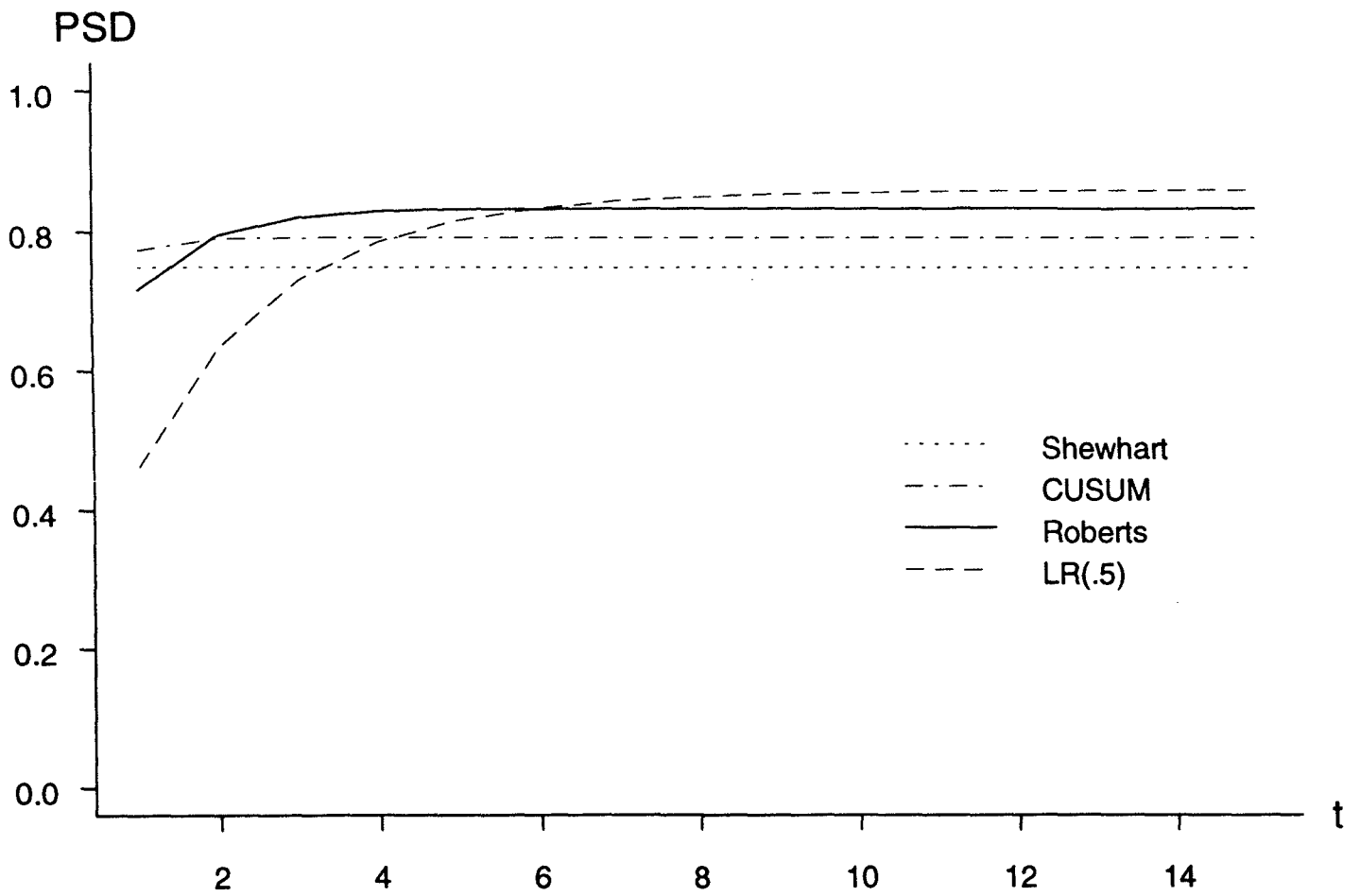


Figure 7. The probability of successful detection (for $d=3$)

$$PSD(\tau, d) = \Pr(t_A - \tau \leq d | t_A \geq \tau, \tau = t).$$

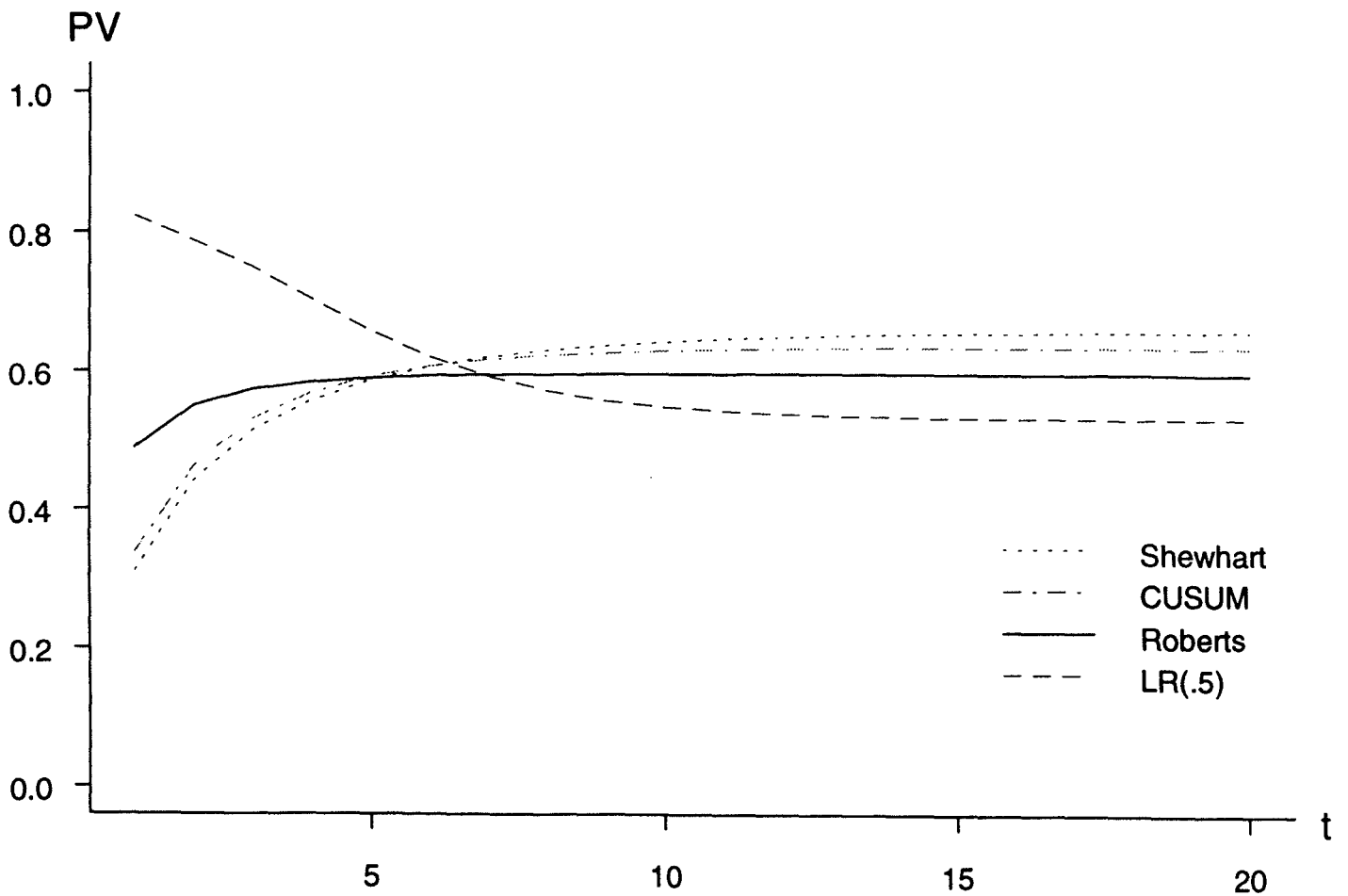


Figure 8. The predictive value $PV(t) = \Pr(\tau \leq t | t_A = t)$ is given for the case when τ has a geometric distribution with intensity 0.1.

Research Report

| | | |
|---------|-----------------------|--|
| 1994:1 | Frisén, M | Statistical surveillance of business cycles. |
| 1994:2 | Frisén, M | Characterization of methods for surveillance by optimality. |
| 1994:3 | Frisén, M & Cassel, C | Visual evaluations of statistical surveillance. |
| 1994:4 | Ekman, C | A comparison of two designs for estimating a second order surface with a known maximum. |
| 1994:5 | Palaszewski, B | Comparing power and multiple significance level for step up and step down multiple test procedures for correlated estimates. |
| 1994:6 | Åkermo, G | Constant predictive value of an alarm. |
| 1994:7 | Åkermo, G | On performance of methods for statistical surveillance. |
| 1994:8 | Palaszewski, B | An abstract bootstrap based step down test procedure for testing correlated coefficients in linear models. |
| 1994:9 | Ekman, C | Saturated designs for second order models. |
| 1994:10 | Ekman, C | A note on rotatability. |
| 1995:1 | Arnkelsdottir, H | Surveillance of rare events. On evaluations of the sets method. |
| 1995:2 | Sveréus, A | Detection of gradual changes. Statistical methods in post marketing surveillance. |
| 1995:3 | Ekman, C | On second order surfaces estimation and rotatability. |
| 1996:1 | Ekman, A | Sequential analysis of simple hypotheses when using play-the-winner allocation. |
| 1996:2 | Wessman, P | Some principles for surveillance adopted for multivariate processes with a common change point. |