ELSEVIER

Forensic Science International
104 (1999) 91–103

Forensic Science International

www.elsevier.com/locate/forsciint

# The robustness of a continuous likelihood approach to bayesian analysis of forensic glass evidence

J.M. Curran[a,*], J. Buckleton[b], C.M. Triggs[c]

[a]*Program in Statistical Genetics, Department of Statistics, North Carolina State University, Box 8203, Raleigh, NC 27606-8203, USA*
[b]*ESR, Private Bag 92021, Auckland, New Zealand*
[c]*Department of Statistics, Private Bag 92019, Auckland, New Zealand*

Received 26 January 1999; received in revised form 25 June 1999; accepted 12 July 1999

**Abstract**

In previous work [1–3] the authors [K.A.J. Walsh, J.S. Buckleton, C.M. Triggs, A practical example of glass interpretation, Sci. Justice 36 (1996) 213–218; J.M. Curran, Forensic application of Bayesian interference to glass evidence, Ph.D. Thesis, Department of Statistics, University of Auckland, 1997; J.M. Curran, C.M. Triggs, J.S. Buckleton, S. Coulson, Combining a continuous Bayesian approach with grouping information, Forensic Sci. Int. 91 (1998) 181–196] have presented various aspects of a Bayesian interpretation of forensic glass evidence. Such an interpretation relies on assumptions that may not hold. This paper demonstrates the robustness of the Bayesian approach to deviations from the statistically convenient notion of normality of the measurements. 

*Keywords:* Glass analysis; Statistics; Bayesian; Robustness; Sensitivity

## 1. Introduction

When someone breaks glass a number of tiny fragments may be transferred to that person. If the glass is broken in the commission of a crime then these fragments may be used as evidence. Interpretation of this evidence relies on, among other things, the

*Corresponding author.
*E-mail address:* curran@stat.auckland.ac.nz (J.M. Curran)

0379-0738/99/$ – see front matter 
PII: S0379-0738(99)00097-3

difference in the mean refractive indices of the glass recovered from the suspect and a control sample taken from the crime scene.

In previous publications [1–3] the authors note that a true Bayesian analysis would include estimation of the difference in the distribution of the refractive indices of the recovered fragments and the control fragments. Because this is considered difficult, if not completely impossible to do, it is useful to note that without loss of generality [1] the difference in the means can be examined instead. However, just as the true distribution of the refractive indices are not known, neither is the true distribution of the difference in the means. Walsh et al. [1] found an approximate solution by modelling the distribution by using a *t*-distribution with Welch's modification [4]. This approximation is valid if the assumption of localised normality is valid. We have become concerned that this assumption, made by ourselves and many other authors in the field of glass evidence, does not have experimental validation. There are theoretical reasons to believe that the sample recovered from the clothing may have 'heavy tails' due to a preponderance of surface or near surface fragments in the recovered samples. Experimental work (pers. comm. J. Buckleton and A. Newton) has suggested that a sample of recovered glass may contain 1/3 to 1/2 of its fragments originating from a surface or near to one. It has also been shown that typically one surface has a higher refractive index than the mean of the bulk glass (and the other has a lower refractive index). Recovered surface fragments are understood to come from both the surface facing the breaker and the opposing one. Such considerations give plausibility to the suggestion that the distribution of the refractive indices in the recovered sample (at least) may not have localised normality and may in fact have 'heavy tails'.

Statistical folklore has it that even if the assumption of localised normality is not true the *t*-distribution will still work reasonably well, although violations will engender a certain amount of information loss. This paper will examine the effect of departures from normality on the likelihood ratio. The ultimate goal is to be able to quantify the distribution of the difference between the control and recovered samples. However in the interim we must work towards a system whereby the information loss that occurs by departures from model assumptions is quantified.

## 2. Methodology

As in previous work [1–3], we take simple examples to demonstrate our findings. Consider the following hypothetical crime:

A window is broken during a burglary. Some time later a suspect is apprehended and his clothes, shoes and hair combings are taken as evidence. A forensic scientist finds a number of fragments of recovered glass upon the suspect's clothing, but finds no glass on the suspect's shoes or from in the suspect's hair combings. The localisation of the recovered glass to the suspect's clothing increases the chance that all the recovered fragments are from one source. The police provide the forensic scientist with a sample of fragments from the window. Each fragment has its refractive index (RI) determined. This information must be assessed.

The likelihood ratio provides a means of examining the plausibility of the evidence under two competing hypotheses. We shall call these hypotheses the prosecution hypothesis, $H_p$ and the defence hypothesis, $H_d$. In this particular example these hypotheses are

$H_p$: The suspect broke the window

$H_d$: The suspect did not break the window

The likelihood ratio is then

$$LR = \frac{\Pr(Evidence|H_p)}{\Pr(Evidence|H_d)}$$

Using the notation introduced by Evett and Buckleton [5–7] we examine each of these terms in turn.

The denominator must answer the following question:

*If we examine the clothing of a man who has come to police notice on suspicion of a glass breaking offence, yet he is unconnected with the offence, what is the probability that we will find one group of glass fragments of the observed size and properties*

Let

$P_1$ be the probability that the suspect's clothing would have one group of glass fragments upon his clothing beforehand.

$S_{n_r} g(\bar{Y})$ be the likelihood that a group of glass fragments on clothing will have $n_r$ fragments and the observed RI, $\bar{Y}$ has density $g(\bar{Y})$.

$T_k$ be the conditional probability of recovering $k$ fragments of glass from the suspect given that an unknown number of fragments were transferred to the suspect from the crime scene, the retention properties of the suspect's clothing, and the time between the commission of the crime and the arrest.

Using these terms, and the implicit assumptions of independence, the denominator can be evaluated as

$$\Pr(Evidence|H_d) = P_1 S_{n_r} g(\bar{Y})$$

We follow [8] and allow for at least two possible explanations in the numerator[1]:

(a): no glass was transferred from the window at the crime scene, but the suspect already had one group of glass on his clothing beforehand, or

(b): one group of glass was transferred from the crime scene, thus

$$\Pr(Evidence|H_p) = T_0 P_1 S_{n_r} f + T_{n_r} P_0 f(\bar{X} - \bar{Y}|s_X, s_Y)$$

where $f(\bar{X} - \bar{Y}|s_X, s_Y)$ is the rescaled value of the density function for Student's $t$ density at the difference between the control and recovered sample means. This function measures the strength of the match between the mean of the control fragments and the mean of all the recovered fragments. $\bar{X}$, $\bar{Y}$, $s_X$ and $s_Y$ are the sample means and sample standard deviations of the control and recovered samples respectively. If $\{x_1, x_2, \ldots,$

[1] There are in fact $n+1$ possible explanations: $r$ fragments were transferred, $n-r$ being there beforehand; where $r=0, 1, 2, \ldots, n$. But it is possible to show that most of the terms associated with these alternatives are likely to be small and leaving them out is, in any case conservative, in the sense that the numerator will be smaller than it should be [7].

$x_{n_c}\}$ and $\{y_1, \ldots, y_{n_r}\}$ represent refractive index measurements on $n_c$ control fragments and $n_r$ recovered fragments, then the summary statistics are given by:

$$\bar{X} = \frac{1}{n_c}\sum_{i=1}^{n_c} x_i \quad \text{and} \quad \bar{Y} = \frac{1}{n_r}\sum_{j=1}^{n_r} y_j$$

and

$$s_X = \sqrt{\frac{1}{n_c - 1}\sum_{i=1}^{n_c}(x_i - \bar{X})^2} \quad \text{and} \quad s_Y = \sqrt{\frac{1}{n_r - 1}\sum_{j=1}^{n_r}(y_j - \bar{Y})^2}$$

The likelihood ratio in this case is:

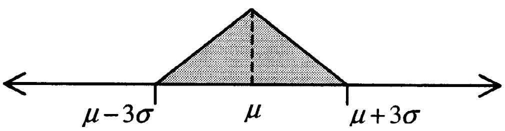
$$\mathrm{LR} = T_0 + \frac{T_n P_0}{P_1 S_n} \cdot \frac{f(\bar{X} - \bar{Y} | s_X, s_Y)}{g(\bar{Y})}$$

In this paper we consider four experimental situations:

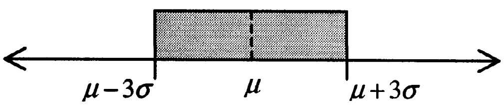
A1: The recovered fragments are distributed normally with mean $\mu$ and standard deviation $\sigma$

A2: The recovered fragments come from a truncated $t$-distribution on 3 degrees of freedom. By truncated we mean we only accept random variates between $\mu - 3\sigma$ and $\mu + 3\sigma$. This will result in a t-shaped curve with more density mass around the central values. This distribution should behave in a similar fashion to a standard $t$-distribution but have heavier tails, than the normal distribution between $(\mu - 3\sigma, \mu - \sigma)$ and $(\mu + \sigma, \mu + 3\sigma)$.

A3: The recovered fragments come from a triangular distribution on the interval $[\mu - 3\sigma, \mu + 3\sigma]$ and mode at $\mu$. That is




A4: The recovered fragments come from a uniform distribution on the interval $[\mu - 3\sigma, \mu + 3\sigma]$ and mode at $\mu$. That is




Each set of control fragments is assumed to be normally distributed with mean $\mu + \delta$ and standard deviation $\delta$, where $\delta = 0, \ldots, 3$. Therefore the standardised distance between the means of the control and recovered samples will be approximately $\delta$. In the very worst situation (the uniform case), the standard deviation is

$$X \sim U[a, b], sd[X] = \sqrt{\mathrm{Var}[X]} = \frac{(b - a)}{\sqrt{12}}$$

so $\mathrm{sd}[X] = (6\sigma/\sqrt{12}) \approx 1.73\sigma$, which is slightly larger than desired, but still acceptable.

## 3. Results

Simulation was used to generate 10,000 samples of each combination of sample sizes $n_r = \{2, 3, 4, 5, 6, 10\}$ and $n_c = \{2, 3, 4, 5, 6, 10\}$ respectively and for each value of $\delta = 0, \ldots, 3$ where $\sigma = 4 \times 10^{-5}$, and $\delta$ is the approximate standardised distance between the recovered sample mean and the control sample.

To evaluate the likelihood ratios the $P_n$, and $S_{n_i}$ terms were estimated using information obtained from Lambert et al. [9]. The $T_k$ probabilities were estimated using the graphical model of Curran et al. [3]. The control sample (the crime scene sample) was assumed to have a true mean refractive index of 1.51900. The frequencies of the sample mean refractive indices were calculated from a density estimate of refractive index based on observed New Zealand case work[2] (Fig. 1).

For each sample, each experiment provides one likelihood ratio (LR), so there are 10,000 LR's calculated for each combination of $\delta$, $n_c$ and $n_r$. The results are first presented in terms of the mean likelihood ratio of the 10,000 generated at each combination of the parameters.

In Fig. 2 we examine the results when the number of control fragments, $n_c = 5$, and the number of recovered fragments, $n_r = \{2, 3, 4, 5, 6, 10\}$ and there is no difference between the means of the distributions ($\delta = 0$). Fig. 2 clearly demonstrates what statistical folklore has suggested. The results from the normally distributed (best case) samples have the highest mean LR and the results from the uniformly distributed (worst

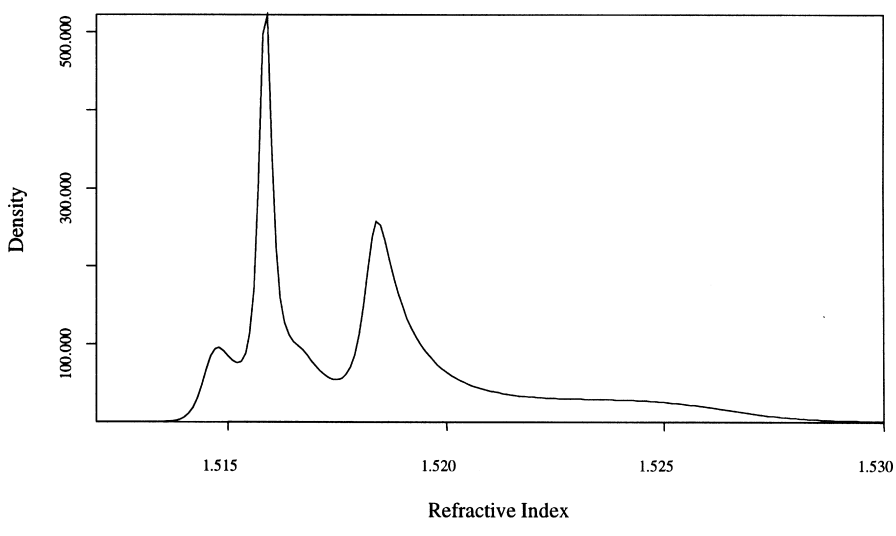



Fig. 1. Density estimate of the refractive index of glass samples based on NZ casework data.

[2] See Appendix A for details of this data and the probabilities.

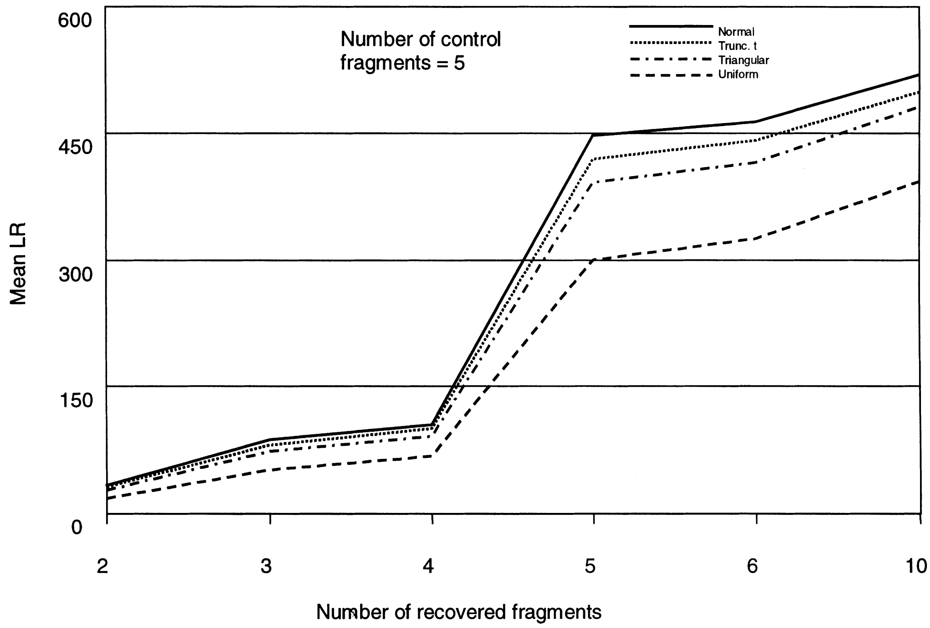



Fig. 2. Mean likelihood ratio when there is no difference between the means of the distributions ($\delta = 0$).

case) samples have the lowest LR. As the sample sizes get larger this disparity seems to increase. It is worth noting however that the results are comparable. That is, even though the true situation may be as 'bad' as uniform, the *t*-approximation still gives a comparable result to the best case situation.

The abrupt jump in the value of the likelihood ratio between $n_r = 4$ and $n_r = 5$ is due to the sharp reduction in the values of $S_4$ and $S_5$. Special attention should be paid to the fact that the horizontal scale is not linear, in that it jumps from 6 to 10.

The situation is slightly different however when $\delta = 1.0$. Although it is difficult to see from Fig. 3, the ordering has been partially destroyed. However, once again, all the results are comparable.

When $\delta = 3.0$ (Fig. 4) the ordering of results seem to have completely reversed. However, we must consider the situation here. The data appear to have come from two completely different distributions. Therefore an optimal result is the one that provides the lowest LR. The results from the uniformly distributed data have the highest LR, whereas the results from the normally distributed data have the lowest LR. However, it must be borne in mind that this is the situation where the difference in means is the greatest, and thus we expect the normally distributed data to provide the best discrimination as demonstrated. Whilst the deviation of the uniformly distributed results from the optimum seems large, it is still relatively minor in the scale of things. That is, no one of these results would lead to a different/contradictory conclusion. The general results for all combinations can be seen in Figs. 7–10.

It is important to realise that the mean alone is not an adequate summary of this data,

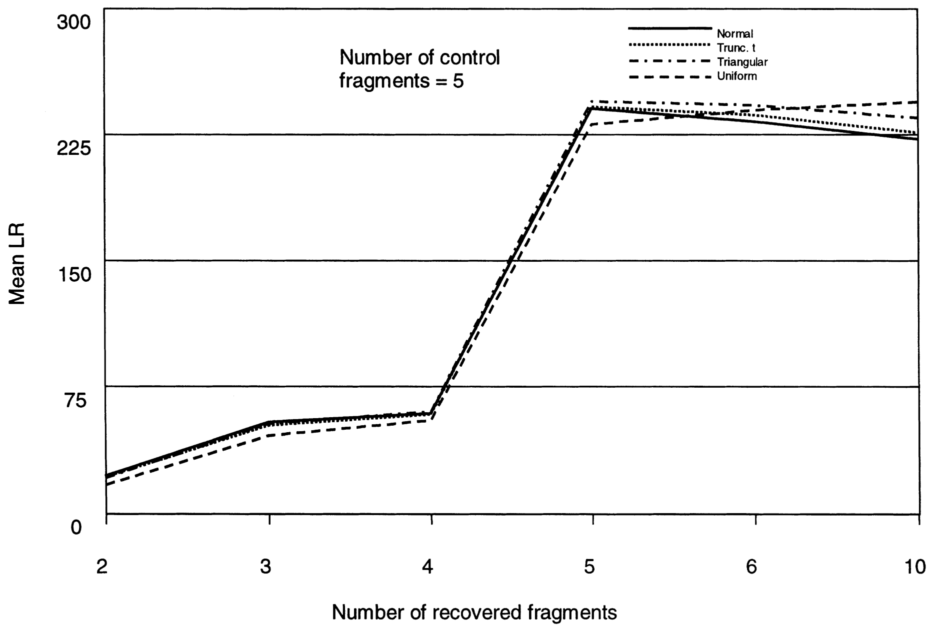



Fig. 3. Mean likelihood ratio when the means of the two distributions are slightly separated ($\delta = 1.0$).

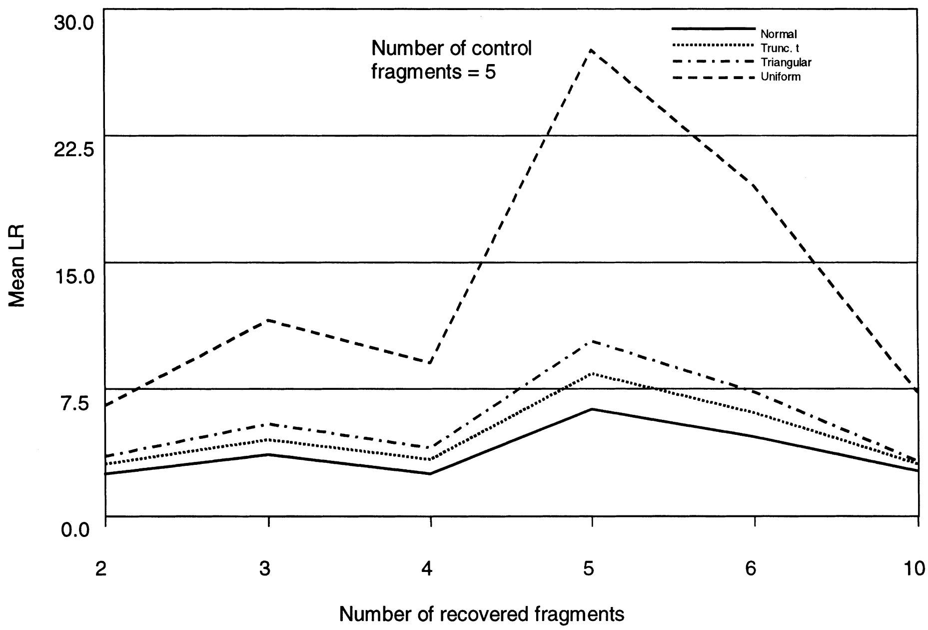



Fig. 4. Mean likelihood ratio when the means of the two distributions are strongly separated for $\delta = 3.0$.

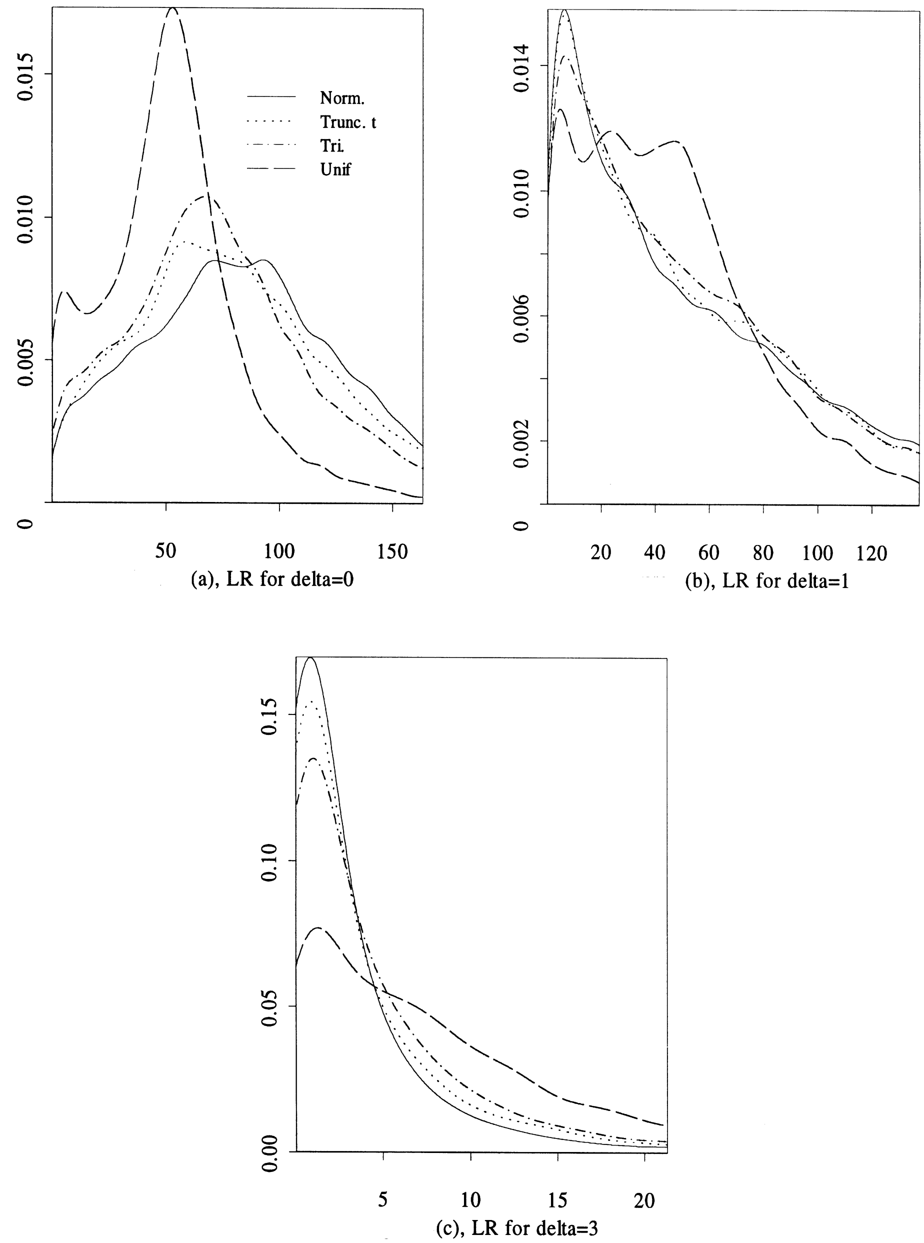



Fig. 5. LR kernel density estimate when $n_c=6$, $n_r=3$ and $\delta=\{0, 1, 3\}$.

although it provides a concise summary. This is especially true for the likelihood ratio as we have good reason to expect the distribution of the likelihood ratio to be positively skewed. For that reason we show density estimates of the simulation results for

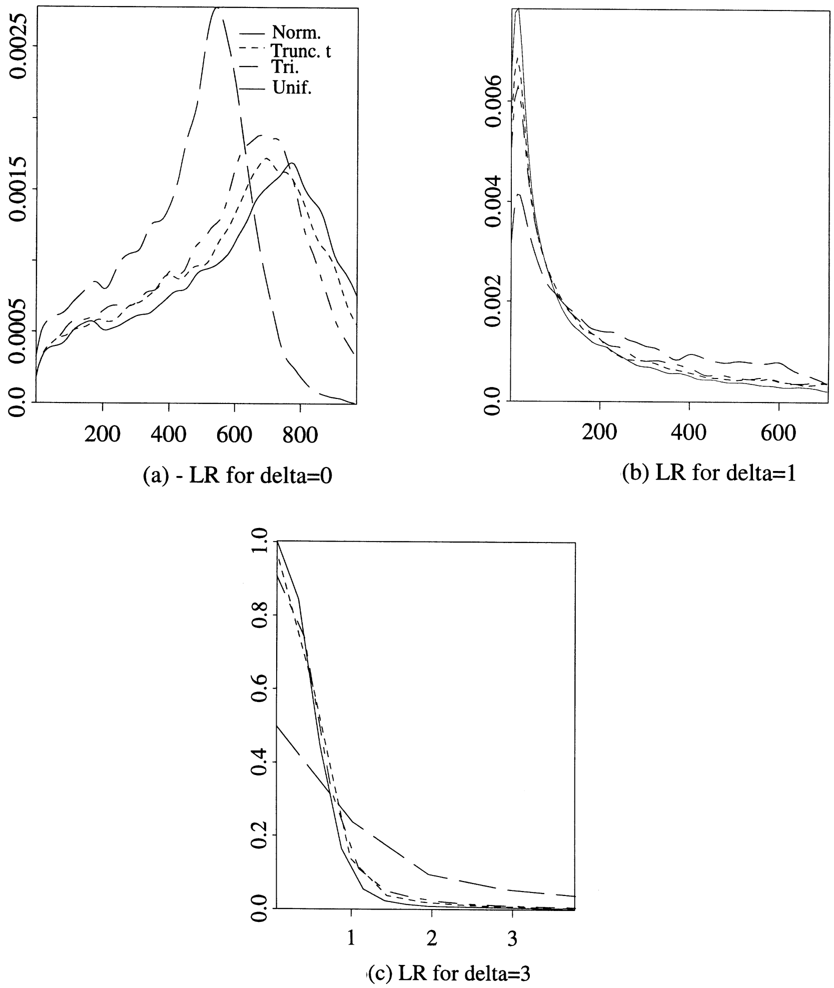



Fig. 6. LR kernel density estimate when $n_c = 10$, $n_r = 10$ and $\delta = \{0, 1, 3\}$.

specifically chosen cases. These cases were selected in order to display the features most commonly exhibited by all combinations of the parameters.

We consider the two sample size combinations: a 'realistic' combination of 6 control fragments ($n_c = 6$) and 3 recovered fragments ($n_r = 3$), and an 'optimal' combination of 10 control fragments and 10 recovered fragments($n_c = n_r = 10$). We simulate data under 3 situations: 'the samples match ($\delta = 0$)', 'the samples match poorly' ($\delta = 1$), and 'the samples don't match ($\delta = 3$)'.

The most interesting feature to note from all of these graphs is that the density estimate for the situation where the recovered sample is uniformly distributed is vastly different to the other three situations. In Figs. 5(a) and 6(a) the bulk of the density mass for the uniform results lies to the left of the other density estimates. This implies that the

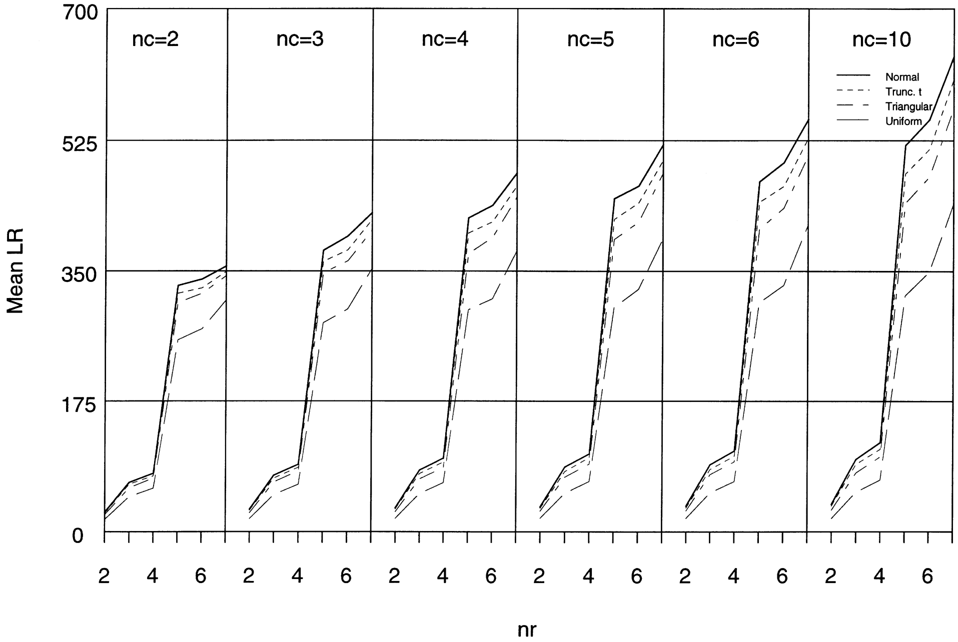



Fig. 7. Mean LR for $\delta = 0$.

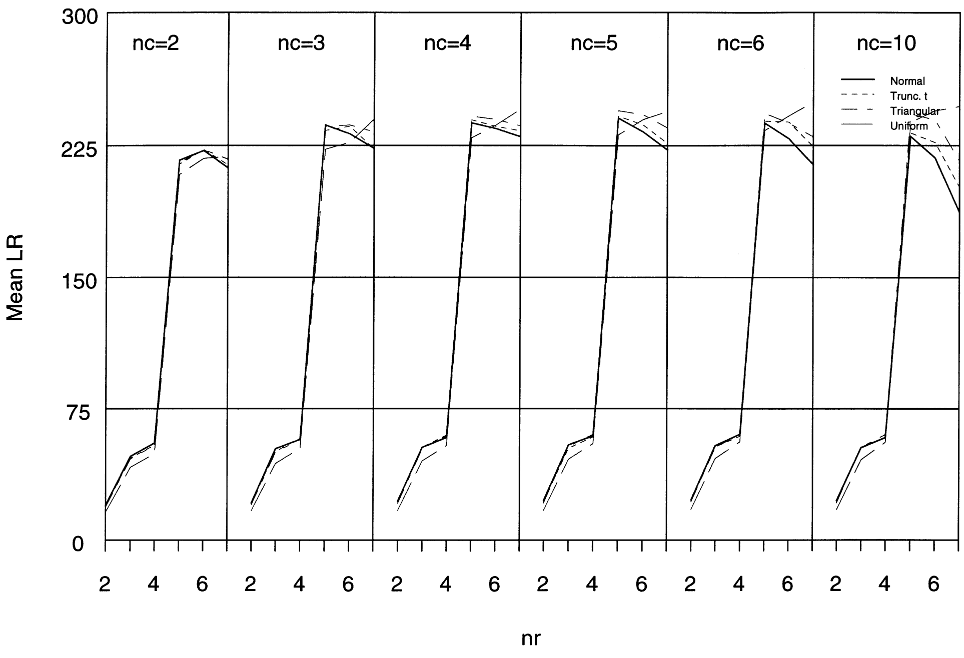



Fig. 8. Mean LR for $\delta = 1.0$.

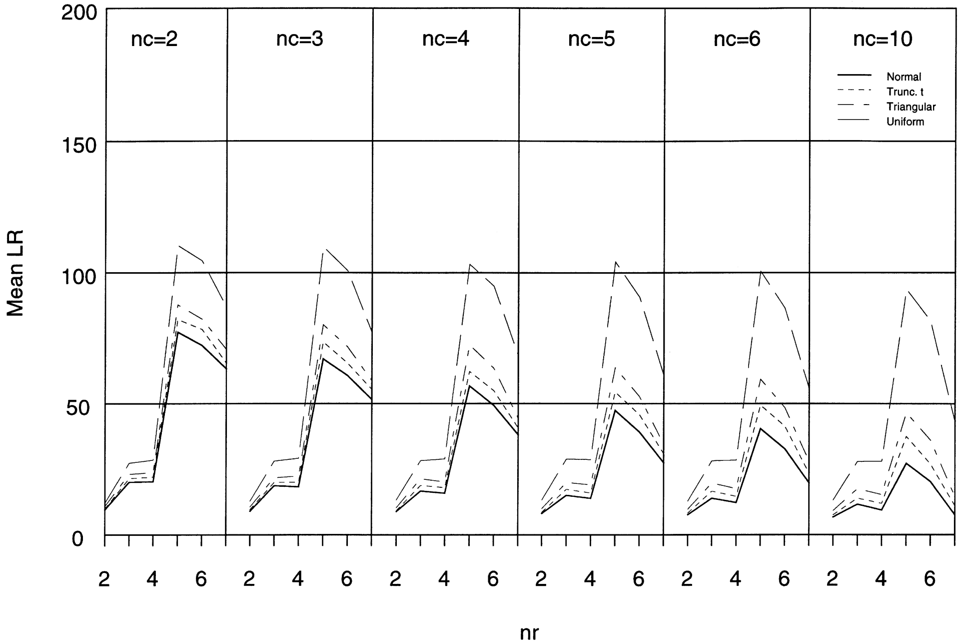



Fig. 9. Mean LR for $\delta = 2.0$.

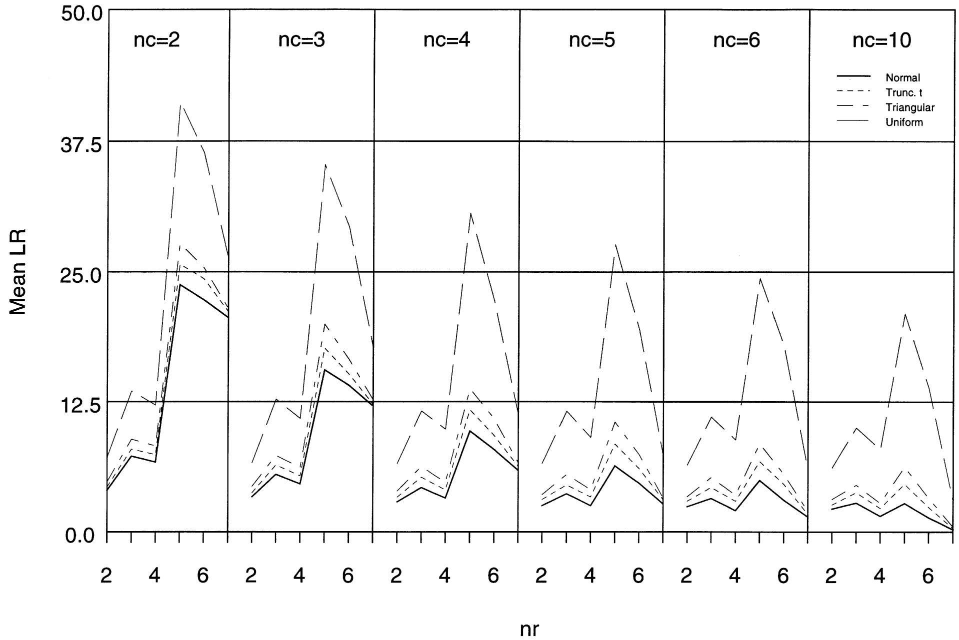



Fig. 10. Mean LR for $\delta = 3.0$.

'worst case' scenario does result in a departure from the optimal, but that departure is conservative, in that it returns a lower LR than expected. However, the LR is not so much lower in the uniform case that the conclusions would be altered. Figs. 5(b), 5(c), 6(b) and 6(c) show the gradual convergence of the likelihood ratio to values less than 1. It is interesting to note that the uniform density is more dispersed, implying a slower rate of convergence. This behaviour accounts for the higher LR's when $\delta$ is large for the uniform case. Finally, the order of performance that we would expect from each situation is preserved in every picture.

## 4. Conclusions

We have examined the behaviour of the continuous likelihood ratio with respect to departures from normality in the recovered sample. These departures ranged from minor (truncated *t*) to major (triangular, uniform). In each and every case the approximation provided by the *t*-distribution for the true distribution of the difference in the recovered and control means provided a result that was comparable in magnitude and conclusion to the situation where the data were true normally distributed. We conclude from these experiments that using the *t*-distribution (with Welch's modification) will provide a robust approximation to the true difference in the means of the control and recovered samples. It is also worth noting that the results from the uniform situation give us some sort of lower bound on the performance of the LR. These results are undoubtedly sensitive to the choice of the $P_n$, and $S_{n_i}$ and $T_k$ probabilities, however, it is expected that the order of the results would remain be preserved for different choices.

We hope that in presenting the results of these experiments, we have provided forensic scientists with a valuable resource for defending the validity of the assumptions associated with the use of the likelihood ratio.

## Acknowledgements

This work was supported in part by a postdoctoral fellowship from the New Zealand Foundation for Research in Science and Technology (FORST) to JMC.

## Appendix A

The $P_n$ and $S_{n_i}$ terms were estimated using information obtained from Lambert et al. [9]. These terms were interpolated from graphs as

| $P_0$ | $P_1$ | $P_2$ | $P_3$ | $P_4$ | $P_5$ | $P_6$ | $P_{10}$ |
|---|---|---|---|---|---|---|---|
| 0.42 | 0.28 | 0.08 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 |

| $S_1$ | $S_2$ | $S_3$ | $S_4$ | $S_5$ | $S_6$ | $S_{10}$ |
|---|---|---|---|---|---|---|
| 0.70 | 0.12 | 0.05 | 0.04 | 0.01 | 0.01 | 0.01 |

The $T_k$ probabilities were estimated using the program detailed in Curran et al. [8], assuming that the breaker was around 70 cm from the window when he broke it and was apprehended around $1\frac{1}{2}$ h later.

| $T_0$ | $T_1$ | $T_2$ | $T_3$ | $T_4$ | $T_5$ | $T_6$ | $T_{10}$ |
|---|---|---|---|---|---|---|---|
| 0.083 | 0.0524 | 0.0497 | 0.0462 | 0.0406 | 0.0407 | 0.0390 | 0.01 |

The New Zealand case work database consists of some 600 cases from the Physical Evidence unit at ESR in Auckland. The density estimate is constructed from around 200 of these cases.

## References

[1] K.A.J. Walsh, J.S. Buckleton, C.M. Triggs, A practical example of glass interpretation, Sci. Justice 36 (1996) 213–218.
[2] J.M. Curran, Forensic applications of Bayesian inference to glass evidence. Ph.D. Thesis. Department of Statistics, University of Auckland, 1997.
[3] J.M. Curran, C.M. Triggs, J.S. Buckleton, S. Coulson, Combining a continuous Bayesian approach with grouping information, Forensic Sci. Int. 91 (1998) 181–196.
[4] B.L. Welch, The significance of the difference between two means when the population variances are unequal, Biometrika 29 (1937) 350–362.
[5] I.W. Evett, A Bayesian approach to the problem of interpreting glass evidence in forensic science casework, J. Forensic Sci. Soc. 26 (1986) 3–18.
[6] I.W. Evett, J.S. Buckleton, Some aspects of the Bayesian approach to evidence evaluation, J. Forensic Sci. Soc. 29 (1989) 317–324.
[7] I.W. Evett, J.S. Buckleton, The interpretation of glass evidence: a practical approach, J. Forensic Sci. Soc. 30 (1990) 215–223.
[8] J.M. Curran, C.M. Triggs, T. Hicks, J.S. Buckleton, K.A.J. Walsh, Assessing transfer probabilities in a Bayesian interpretation of forensic glass evidence, Sci. Justice 38 (1998) 15–22.
[9] J.A. Lambert, M.J. Satterthwaite, P.H. Harrison, A survey of glass fragments recovered from clothing of persons suspected of involvement in crime, Sci. Justice 35 (1995) 273–281.