# Sampling in forensic comparison problems

JM CURRAN, CM TRIGGS

*Department of Statistics, University of Auckland, Private Bag 90219, Auckland, New Zealand*

*e-mail: curran@stat.auckland.ac.nz or triggs@stat.auckland.ac.nz*

and

J BUCKLETON

*ESR, Private Bag 92-021, Auckland, New Zealand*

*e-mail: john.buckleton@esr.cri.nz*

*Science & Justice 1998; 38: 101–107*

*Received 9 June 1997; accepted 22 October 1997*

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. If a large number of fragments are recovered from the suspect, then it may be more efficient for the forensic scientist to examine a subset of these fragments. Such sampling incurs information loss. This paper will derive an expression that allows a partial quantification of this loss. The loss of such information due to the examination of a subset of recovered material arises with many forms of evidence.

Wenn Glas zerbrochen wird, können kleinste Glassplitter auf die betreffende Person übertragen werden. Geschieht dies in Zusammenhang mit einer Straftat, können die Glassplitter als Beweismittel dienen. Werden am Tatverdächtigen viele Teilchen gesichert, kann es ökonomischer sein nur eine bestimmte Gruppe von Teilchen auszuwerten. Diese Verfahrensweise beinhaltet aber Verlust von Information. Es wird eine Formel abgeleitet, mit der sich der Informationsverlust teilweise quantifizieren läßt. Die Frage des Informationsverlusts bei der Untersuchung nur bestimmter Gruppen aus der Gesamtheit eines Spurengutes ergibt sich auch bei vielen anderen Spurenarten.

Lorsque quelqu'un casse du verre, un certain nombre de petits fragments peuvent être transférés sur la personne. Si ce verre est cassé lors de la commission d'un délit, alors le verre peut être utilisé comme indice. Si un grand nombre de fragments sont retrouvés sur un suspect, il peut alors être plus efficace pour l'expert forensique d'examiner une partie seulement de ces fragments. Un tel échantillonnage provoque une perte d'information. La perte d'une telle information résultant de l'examen d'une partie seulement du matériel récolté est un probléme rencontré avec de nombreux types d'indices.

Cuando alguien rompe un vidrio, una cierta cantidad de pequeños fragmentos pueden ser transferidos a esa persona. Si el cristal se ha roto al cometer un delito, los fragmentos pueden ser utilizados como evidencia. Cuando se pueden obtener un gran número de fragmentos del sospechoso, puede ser más eficaz para el científico forense examinar un subconjunto de esos fragmentos. Este muestreo conlleva una pérdida de información. Esta trabajo obtiene una expresión que permite una cuantificación parcial de dicha pérdida. La pérdida de información, debida al análisis de un subconjunto del material obtenido, surge con muchas formas de evidencia.

*Key Words:* Forensic science; Statistics; Sampling; Comparison; Evidence; Glass.

## Introduction

As an example of a forensic comparison problem we consider evidence from fragments of glass although such problems arise with evidence derived from many other materials.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. A typical forensic treatment of glass evidence may involve taking a refractive index (RI) measurement on each recovered fragment of glass, or determining the elemental composition of each fragment of glass. Both of these methods are time consuming, thus if a large number of fragments are recovered from the suspect, it may be necessary for the forensic scientist to examine a subset of these fragments. However, in taking a sample from the recovered fragments the forensic scientist runs the risk that he or she is losing information on the number of 'matching' or 'non-matching' fragments. This information has considerable evidential value and may be included in a Bayesian analysis of the evidence. Therefore in taking a sample, the forensic scientist has incurred an information loss. Bates and Lambert [1] examined this problem and have answered only part of the question. This paper will explain the work of Bates and Lambert, and derive an expression that allows an extension of their answer. Clearly this problem is not restricted to glass analysis and can be applied to any membership type applications. This result has been extended for multi-category trace evidence (such as fibres) and will be the subject for another publication along with the considerations of the multi-category problem.

## Use of the hypergeometric distribution for sampling in forensic glass comparison

The comparison of a number of recovered items of the same generic type of material with a control sample of known origin is required in a wide range of the casework carried out by a forensic scientist [1]. In the course of this comparison it falls upon the forensic scientist to decide how many items should be compared to the control sample, and what information may be lost by leaving some of the recovered items out of the sample.

With respect to the statistical analysis of forensic glass evidence, there is considerable evidential value to be gained from information on the number and size of groups of glass on the suspect. This follows from the fact that it is uncommon to find a group of three or more glass fragments on the clothing of people unassociated with crime. Furthermore, even if some glass is found, it is uncommon that this comes from more than one source [2,3]. We consider a hypothetical case where the suspect has one set of glass fragments that match the glass from the crime scene on his clothing, as well as a number of sets of glass fragments from other sources. The fragments are classified as 'matching' or 'non-matching'. Assume that:

$N$ = the total number of fragments recovered from the suspect.

$M$ = the number of matching fragments in total. Therefore there are $N-M$ non-matching fragments. $M$ can only be determined by examination of all $N$ fragments. $M$ can be regarded as the *Evidence*.

$n$ = the number of fragments that are sampled without replacement from the $N$ recovered fragments.

$m$ = the number of fragments in the sub-sample of $n$ that match.

Then as Bates and Lambert [1] correctly point out, if we choose a sample of $n$ fragments from the total $N$ fragments, the probability that $m$ of these fragments match the control is given by the hypergeometric distribution, i.e.

$$\Pr(m|n,M,N) = \frac{\binom{M}{m}\binom{N-M}{n-m}}{\binom{N}{n}}$$

This, however, is not the probability we are interested in, nor do we have all the information to assess it. We wish to make a statement about the probability that $M$ fragments match in our recovered sample of $N$ fragments based on the fact that $m$ fragments matched in our sub-sample of $n$ fragments. The hypergeometric formula assumes that everything is known about the recovered fragments. In such a case there is nothing to be gained from taking a sample. In particular the number $M$ is known.

The real question of interest is 'what is the probability that there are $M$ matching recovered fragments out of the total $N$ given that I found $m$ matching fragments in my sample of $n$?' We wish to find $\Pr(M|m,n,N)$.

### Answering a more complete question

We wish to make inferential statements about the 'population', the entire set of recovered fragments, based on the information contained in our 'sample', a subset of those fragments. As previously noted this problem is not restricted to glass analysis and can be applied to any membership type applications. This result has been extended for multi-category trace evidence (such as fibres) and will be the subject for another publication along with the considerations of the multi-category problem. An expression for $\Pr(M|m,n,N)$ is derived using Bayes theorem. We know about $\Pr(m|n,M,N)$, and we wish to use this information to calculate $\Pr(M|m,n,N)$. Although one can calculate a formal expression for $\Pr(M|m,n,N)$ $\left(=\Pr(m|,n,M,N)\frac{\Pr(M|N)}{\Pr(m|n)}\right)$ relatively easily, the expression cannot be used without empirical knowledge.

We quantify uncertainty about $M$ in terms of a parameter $\theta$, the true but unknown proportion of matching fragments in any similar set of recovered fragments. In fact, the expected value of the ratio M/N is $\theta$, i.e. E[M/N] = $\theta$, where E[$X$] is the expected value of $X$.

In the analysis of glass evidence, and other trace evidence, $\theta$ could change under the competing hypotheses of 'The suspect *was* in contact with the crime scene' or 'The suspect *was not* in contact with the crime scene'. In order to model each situation it is necessary to select a distribution for $\theta$ which reflects the facts under each hypothesis. If nothing is known, we can put an uninformative prior distribution on $\theta$, i.e. $\theta$ is uniformly distributed between 0 and 1. In fact, this is a problem that was first considered by Reverend Bayes himself [4].

In a Bayesian analysis of forensic glass evidence, the scientist is asked to evaluate a likelihood ratio (LR), which considers the probability of the evidence under the two hypotheses:

$$\text{LR} = \frac{\Pr(Evidence \mid Contact)}{\Pr(Evidence \mid \overline{Contact})}$$

Therefore we must consider the distribution of $\theta$ under the hypotheses of *Contact* and $\overline{Contact}$. In the case of *Contact* one might take into account the distribution of the number of groups of glass found on people known to be associated with a crime. Under the assumption that the event *Contact* has taken place the prior probability of $\theta$ should correspond to this distribution, the assumption being that a higher proportion of fragments of glass would come from one source.

Under the assumption of $\overline{Contact}$, the prior probability of $\theta$ might correspond to the distribution of the number of groups of glass on people unassociated with crime. That is, if the suspect has glass on his clothing, then it is more likely to have come from a variety of sources rather than one concentrated source.

Choosing these prior distributions is modelled by selecting the parameters of a Beta distribution. The Beta distribution is a convenient choice for the prior distribution of $\theta$ for two reasons: (1) it can only take values between zero and one, and (2) the parameters of the distribution can be selected to model the hypotheses of *Contact* and $\overline{Contact}$.

The Beta distribution is often used to model the prior distribution of a probability or proportion. The reasons for this are threefold, and we shall explain each carefully.

The most desirable property of the Beta distribution, is that it only assigns probability to values between 0 and 1, i.e. it does not allow (quite sensibly) the proportion that is being modelled to take a value less than zero or greater than one.

The second property of the Beta distribution is that its parameters, $\alpha$ and $\beta$, control the mean of the distribution. That is, if a random variable has a Beta distribution with parameters $\alpha$ and $\beta$, then the mean of that random variable is

$$\frac{\alpha}{\alpha + \beta}$$

This fact is useful in that if you have a prior estimate on the proportion, say $p$, then by choosing $\alpha = p$ and $\beta = 1-p$, the Beta distribution will have a mean of $p$.

Finally, the parameters $\alpha$ and $\beta$ can be also be chosen to control the shape of the distribution so that a certain amount of probability is assigned to values less than some value $x$ is $p$. For example, a situation may demand that the probability assigned to all the values less than 0.5 is 0.25, i.e. there is a 25% chance of the proportion of matching fragments being 0.5 or less. In this situation a choice of $\alpha = 4.61$ and $\beta = 1$ would provide the desired properties. Given there are two unknown parameters there will never be a unique solution. However if $\beta$ is set to 1 and

$\alpha = \dfrac{\log(p)}{\log(x)}$ then it can be shown that $\Pr(X \le x \mid \alpha, \beta) = p$.

The parameters can be chosen so that the shape of the Beta distribution closely matches that of the empirical distributions of the proportion of matching fragments from other similar sources. For example, in glass we might use distributions of similar shape to those reported by Lambert et al. [2] or McQuillan and Edgar [3]. It must be noted that the choice of these parameters is always arbitrary and therefore subjective – there is no 'best' choice. It is hoped that in the future, methods will be developed to enable an objective selection method for the parameters.

Figure 1 shows the probability density function for a possible prior under the hypothesis of *Contact*. The distribution puts high probability (80%) on values of $\theta$ between 0.8 and 1, representing the fact that one would expect to find a higher concentration of glass from one source, i.e. the crime scene, if the suspect truly was at the crime scene.

Figure 2 shows the probability density function for a possible prior under the hypothesis of $\overline{Contact}$. The distribution more or less evenly weights all values of $\theta$ between 0 and 1, representing the fact that one would expect to find different groups or sources of glass on the suspect given that he was not at the crime scene.

**Result**

$$\Pr(M \mid m,n,N) = \frac{\binom{M}{m}\binom{N-M}{n-m}}{\binom{N}{n}} \frac{\binom{N}{M}}{\binom{n}{m}} \frac{\text{Beta}\,(M+\alpha,\, N-M+\beta)}{\text{Beta}\,(m+\alpha,\, n-m+\beta)},$$

$$M = m, \ldots, N-n+m$$

where Beta($a,b$) is the standard Beta function.

Proof: See Appendix A.

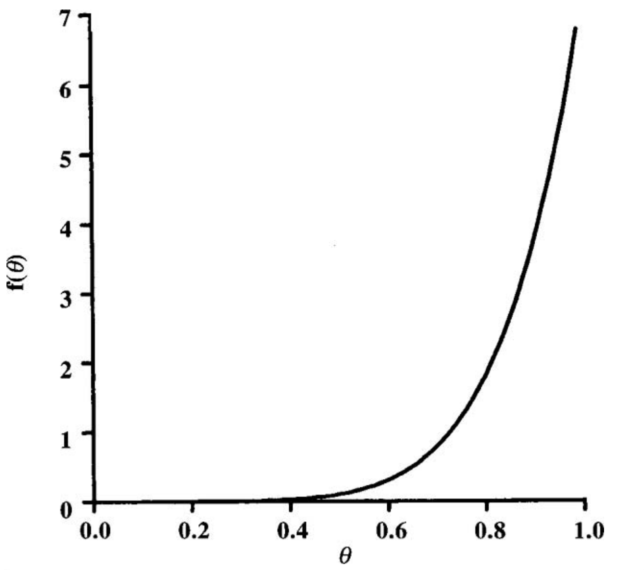


**FIGURE 1 Probability density function for a $\beta(7.21,1)$ distribution.**

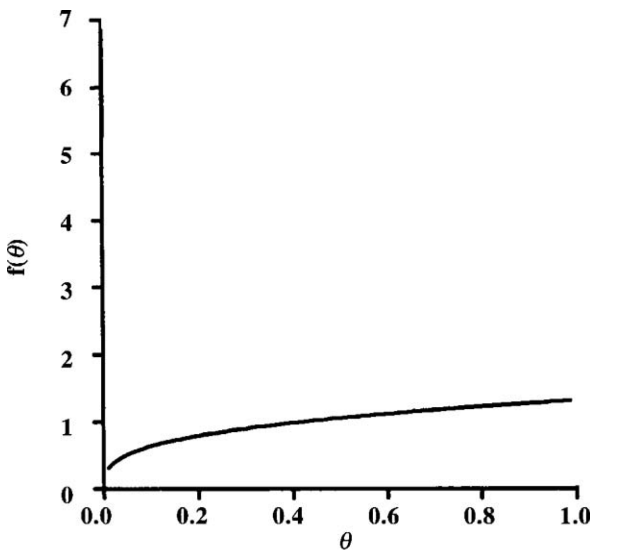


**FIGURE 2 Probability density function for a $\beta(1.32,1)$ distribution.**

## Examples

We give two examples to illustrate the 'common sense' behaviour of this probability function. A demonstration of the calculations required has been included in Appendix B.

The first example has 100 recovered fragments in total ($N = 100$), with 50 fragments examined ($n = 50$) and 25 of those fragments are deemed to 'match' the control sample ($m = 25$).

Figure 3 demonstrates what we would logically assume. Given that 50 fragments out of 100 have been examined, and 25 'match', we would expect to find close to 50 fragments that 'match' in the entire recovered sample. The distributions do not change much under each hypothesis because the evidence dominates the expression.

In the second example 50 fragments have been recovered from the suspect ($N = 50$). A smaller sub-sample of 10 fragments has been examined ($n = 10$), and 9 of those fragments 'match' the control sample.

In Figure 4 again we see what we would logically expect. Given that 90% of the examined fragments 'match' the control sample, we expect that around 90% of the entire recovered sample would 'match'. This time the prior distribution of $\theta$ does play some part in changing the conditional distribution, thus reflecting the low information contained in a sub-sample size of 10 from a sample of 50.

## Conclusions

We are now able to answer our question correctly, and quite sensibly.

$$\Pr(M|m,n,N) = \frac{\binom{M}{m}\binom{N-M}{n-m}}{\binom{N}{n}} \times \frac{\binom{N}{M}}{\binom{n}{m}} \frac{\text{Beta}\,(M+\alpha,\ N-M+\beta)}{\text{Beta}\,(m+\alpha,\ n-m+\beta)}$$

This function says that the probability that there are $M$ fragments that match the control sample out of the total $N$ fragments recovered is dependent on the probability of $m$ fragments matching in our sample of $n$ fragments **and** the likelihood of the data given that there are $M$ matching fragments and the hypothesis of $H = \textit{Contact}$ or $H = \overline{\textit{Contact}}$.

## APPENDIX A

### Proof of result

By the introduction of a nuisance parameter $\theta$, with some prior distribution assumptions, a moderately simple closed form solution can be found.

Assume that $\theta \sim \text{Beta}(\alpha,\beta)$

Given

$$\Pr(M|m,n,N,\theta) = \frac{\Pr(M,m,n,N,\theta)}{\Pr(m,n,N,\theta)} \quad (1)$$

and

$$\Pr(m|M,n,N,\theta) = \frac{\Pr(M,m,n,N,\theta)}{\Pr(M,n,N,\theta)} \quad (2)$$

then

$$\Pr(M|m,n,N,\theta) = \Pr(m|M,n,N,\theta)\frac{\Pr(M,m,n,N,\theta)}{\Pr(m,n,N,\theta)} \quad (3)$$

Now note that

$$\Pr(M,n,N,\theta) = \Pr(M|n,N,\theta)\Pr(n,N,\theta) \quad (4)$$

and

$$\Pr(m,n,N,\theta) = \Pr(m|n,N,\theta)\Pr(n,N,\theta) \quad (5)$$

so

$$\Pr(M|m,n,N,\theta) = \Pr(m|M,n,N,\theta)\frac{\Pr(M,m,n,N,\theta)}{\Pr(m,n,N,\theta)} \quad (6)$$

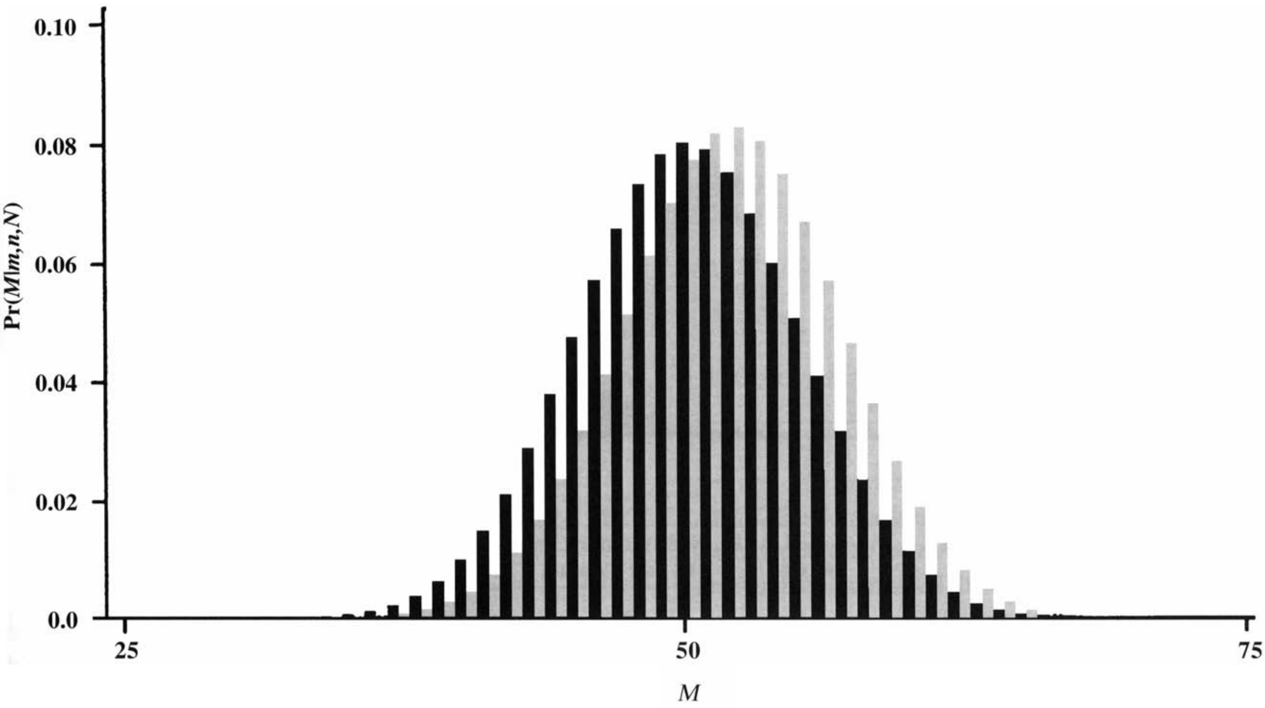


**FIGURE 3 Pr($M$|$m$=25, $n$=50, $N$=100) under *Contact* and $\overline{Contact}$.**
**■ Contact ■ not Contact**

$M$ does not depend on $n$ and $m$ is conditionally independent of $N$ given $n$, therefore (6) becomes

$$\Pr(M|m,n,N,\theta) = \Pr(m|M,n,N,\theta)\ \frac{\Pr(M|N,\theta)}{\Pr(m|n,\theta)} \qquad (7)$$

The ratio in (7) is simply a ratio of two binomial probabilities. If it is noted that

$$\Pr(\theta|M,m,n,N) = \frac{\Pr(M,m,n,N,\theta)}{\Pr(M,m,n,N)} \qquad (8)$$

then with the result from (2), equation (8) becomes

$$\Pr(m|M,n,N,\theta) = \Pr(\theta|M,n,N)\ \frac{\Pr(M,m,n,N)}{\Pr(M,n,N,\theta)} \qquad (9)$$

Furthermore, given $M$ and $N$, $\theta$ is independent of $m$ and $n$,

$$\Pr(M,m,n,N) = \Pr(m|M,n,N)\ \Pr(M,n,N) \qquad (10)$$

and

$$\Pr(\theta,M,n,N) = \Pr(\theta|M,n,N)\Pr(M,n,N) = \Pr(\theta,M,N)\Pr(M,n,N) \qquad (11)$$

so

$$\Pr(m|M,n,N,\theta)=\Pr(\theta|M,N)\ \frac{\Pr(m|M,n,N)}{\Pr(\theta|M,N)} = \Pr(m|M,n,N) \qquad (12)$$

Using these facts we have

$$\Pr(M|m,n,N,\theta)=\Pr(m|M,n,N)\ \frac{\Pr(M|N,\theta)}{\Pr(m|n,\theta)} \qquad (13)$$

If the joint density of $M$ and $\theta$ conditional on $m$, $n$, and $N$, is specified then we have

$$\Pr(M,\theta|m,n,N) = \frac{\Pr(M,m,n,N,\theta)}{\Pr(m,n,N)} \qquad (14)$$

Using this in conjunction with (1) gives

$$\Pr(M,\theta|m,n,N)=\Pr(M|m,n,N,\theta)\ \frac{\Pr(m,n,N,\theta)}{\Pr(m,n,N)} \qquad (15)$$

$$= \Pr(M|m,n,N,\theta)\Pr(\theta,m,n,N)$$

Therefore the joint density of $M$ and $\theta$ conditional on $m$, $n$, and $N$ is given by

$$\Pr(M,\theta|m,n,N)=\Pr(m|M,n,N)\ \frac{\Pr(M|N,\theta)}{\Pr(m|n,\theta)}\ \Pr(\theta|m,n) \qquad (16)$$

It now remains to remove the nuisance parameter $\theta$ to get the marginal distribution of $M$. This is done by integration of (16) with respect to $\theta$. It is convenient at this point to adopt the notation $f_{X,Y}(R,S,T)$ to represent the joint density of $X$ and $Y$ conditional on $R$, $S$ and $T$. Thus (16) can be rewritten as

$$f_{M,\theta}(m,n,N)=f_m(M,n,N)\ \frac{f_M(N,\theta)}{f_m(n,\theta)}\,f_\theta(m,n) \qquad (17)$$

Integration of (17) with respect to $\theta$ gives the marginal distribution of $M$, i.e.

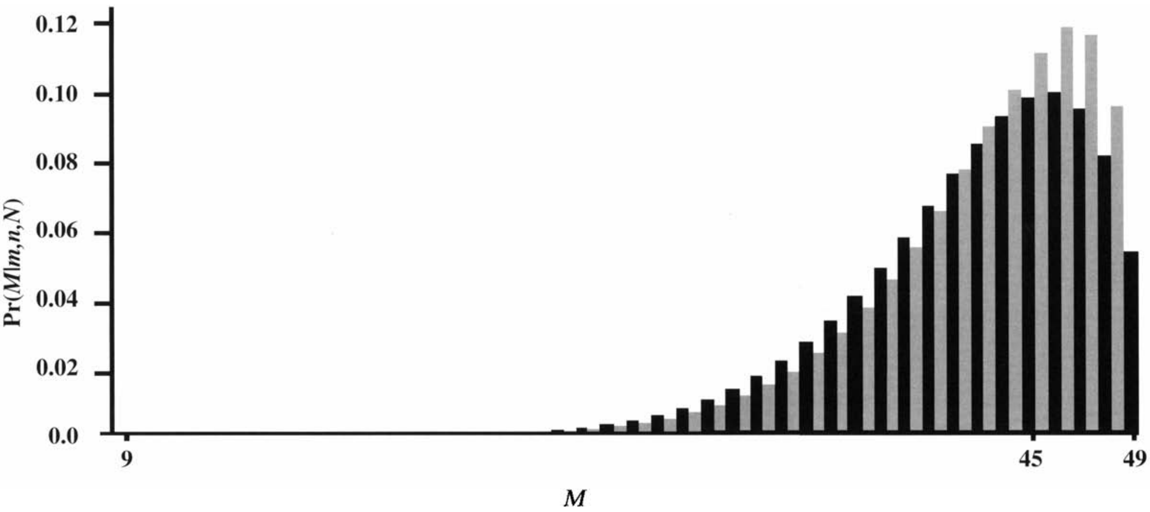


**FIGURE 4 Pr(*M*|*m*=9, n=10, N=50) under *Contact* and $\overline{\textit{Contact}}$.**
**■ Contact ■ not Contact**

$$\int_0^1 f_{M\theta}(m,n,N).d\theta = f_M(m,n,N)$$

$$= \int f_m(M,n,N)\frac{f_M(N,\theta)}{f_m(n,\theta)} f_\theta(m,n).d\theta$$

$$= f_m(M,n,N)\int \frac{f_M(N,\theta)}{f_m(n,\theta)} f_\theta(m,n).d\theta$$

$$= f_m(M,n,N)\int_0^1 \frac{\binom{N}{M}\theta^M(1-\theta)^{N-M}}{\binom{n}{m}\theta^m(1-\theta)^{n-m}} \frac{\theta^{M+\alpha-1}(1-\theta)^{N-M+\beta-1}}{\beta(m+\alpha,\ n-m+\beta)}.d\theta$$

$$= f_m(M,n,N)\frac{\binom{N}{M}}{\binom{n}{m}}\frac{1}{\beta(m+\alpha,\ n-m+\beta)}\int_0^1\theta^{M+\alpha-1}(1-\theta)^{N-M+\beta-1}.d\theta$$

$$= f_m(M,n,N)\frac{\binom{N}{M}}{\binom{n}{m}}\frac{\beta(M+\alpha,N-M+\beta)}{\beta(m+\alpha,\ n-m+\beta)}$$

where

$$f_m(M,n,N) = \frac{\binom{M}{m}\binom{N-M}{n-m}}{\binom{N}{n}}$$

## APPENDIX B

Take the first example: the forensic scientist recovered $N$ = 100 fragments. From these 100 fragments she took a subsample of $n$ = 50 fragments and determined the RI of each fragment using GRIM. She found that $m$ = 25 of these fragments 'matched' the control sample. We may now ask questions of the type: 'given this information, what is the probability that the total number of matching fragments, $M$, is $x$'. $x$ is a placeholder for any number of fragments from 25 to 75. Why? Well the scientist knows that there are 25 matching fragments already so she can't find any fewer than 25 matching fragments, and she also knows that there are 25 non-matching fragments, so there are at most 50 more fragments that match (25 + 50 = 75). We will calculate the probability for $M$ = 55,
i.e. we wish to find Pr($M$=55|$m$=25, $n$=50, $N$=100)

under the hypotheses of *Contact* and $\overline{\textit{Contact}}$.

Pr($M$=55|$m$=25,$n$=50,$N$=100) =

$$\frac{\binom{55}{25}\binom{45}{25}}{\binom{100}{50}} \times \frac{\binom{100}{55}}{\binom{50}{25}} \times \frac{\mathrm{Beta}(55+\alpha,45+\beta)}{\mathrm{Beta}(25+\alpha,25+\beta)}$$

where $\binom{x}{c} = \frac{x!}{(x-c)!c!} = \frac{x\times(x-1)\times(x-2)\times\ldots\times 1}{(x-c)\times(x-c-1)\times\ldots\times 1\times c\times(c-1)\times\ldots\times 1}$,

for example

$$\binom{50}{25} = \frac{50!}{25!25!} = \frac{50\times 49\times\ldots\times 1}{(25\times 24\times\ldots\times 1)^2} = 126{,}410{,}606{,}437{,}752.$$

Therefore

$$\frac{\binom{55}{25}\binom{45}{25}}{\binom{100}{50}} \times \frac{\binom{100}{55}}{\binom{50}{25}} = 47{,}129{,}212{,}243{,}960$$

Under the hypothesis of *Contact* we wish to assign 80% of the probability to the proportion of matching fragments being 0.8 or greater. Therefore, there is a 20% chance of getting a value less than 0.8. Using $p$ = 0.2 and $x$ = 0.8, we find that $\alpha$ = 7.21 if $\beta$ = 1. Hence, under the hypothesis of *Contact* Beta(55+$\alpha$,45+$\beta$) = 4.403287 × $10^{-33}$ and Beta(25+$\alpha$,25+$\beta$) = 2.77446 × $10^{-18}$, so that

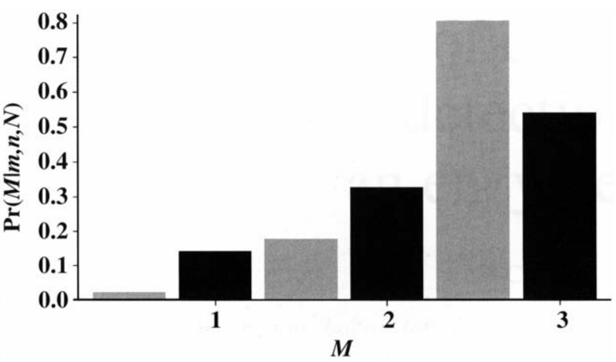


**FIGURE 5 Pr(M|m=1,n=1,N=3) under *Contact* and $\overline{Contact}$.**
**Contact ■ not Contact**

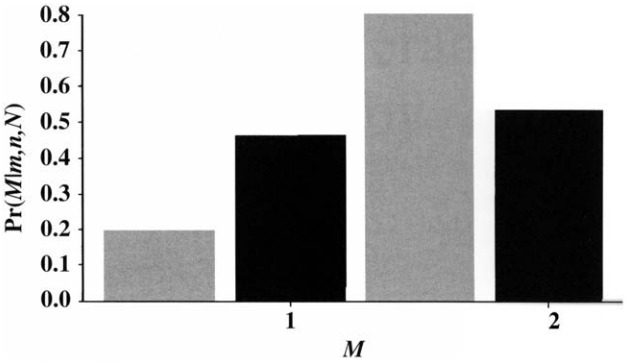


**FIGURE 6 Pr(M|m=1,n=1,N=3) under *Contact* and $\overline{Contact}$.**
**Contact ■ not Contact**

$\Pr(M{=}55|m{+}25,n{=}50,N{=}100,\overline{Contact})$

$= 47{,}129{,}212{,}243{,}960 \times 1.587079 \times 10^{-15}$

$= 0.075$ (3 decimal places).

Under the hypothesis *Contact*, we wish to assign a 60% chance to observing a proportion of matching fragments greater than 0.5. Therefore, there is a 40% chance of getting a value less than 0.5. Using $p = 0.4$ and $x = 0.5$, we find that $\alpha = 1.32$ if $\beta = 1$. Hence, under the hypothesis *Contact* = $\mathrm{Beta}(55{+}\alpha,45{+}\beta) = 1.328784 \times 10^{-31}$ and $\mathrm{Beta}(25{+}\alpha, 25{+}\beta) = 1.239952{\times}10^{-16}$, so that

$\Pr(M{=}55|m{=}25,n{=}50,N{=}100,Contact)$

$= 47{,}129{,}212{,}243{,}960 \times 1.071642 \times 10^{-15}$

$= 0.051$ (3 decimal places).

The Beta function can be programmed quite simply in Excel™ by using the function GAMMALN (the log of the Gamma function,) and the following identity.

$$\mathrm{Beta}(\alpha,\beta) = \exp(\log\Gamma(\alpha)+\log\Gamma-\log\Gamma(\alpha+\beta))$$

The log of the gamma function can also be used to work out $x!$, by using the fact that, $\Gamma(x+1)=x!$ for any positive integer $x$. In order to avoid overflow errors, it is a good idea to evaluate the log of the probability formula, and exponentiate the result.

## Example

For the sake of illustration we include a 'realistic' example, where three fragments have been recovered from a suspect, and ten fragments are taken as a control sample. The number of control fragments is superfluous to our example, as we are only making inferences about the recovered fragments. We demonstrate the behaviour of the formulae under both hypotheses for two situations. In the first situation the analyst has determined the refractive index of one fragment and found that it 'matches' the control sample mean. Figure 5 shows the probability that we might find 1, 2 or 3 ($M$=1,2,3) matching fragments, given that there are three fragments altogether ($N$=3) and from a sample of one fragment ($n$=1) one was found to match ($m$=1).

In the second situation the analyst has determined the refractive index of two fragments and found that one of these 'matches' the control sample mean. Figure 6 shows the probability that we might find 2 or 3 ($M$=2,3) matching fragments given that there are three fragments altogether ($N$=3) and from a sample of two fragments ($n$=2) one was found to match ($m$=1).

## References

1. Bates JW and Lambert JA. Use of the hypergeometric distribution in forensic glass comparison. Journal of the Forensic Science Society 1991; 314: 449–455.
2. Lambert JA, Satterthwaite MJ and Harrison PH. A survey of glass fragments recovered from clothing of persons suspected of involvement in crime. Science & Justice 1995; 35: 273–281.
3. McQuillan J and Edgar KA. Survey of the distribution of glass on clothing. Journal of the Forensic Science Society 1992; 32: 333–348.
4. Bernado JM and Smith AFM. Bayesian Inference. New York: John Wiley & Sons Inc, 1994.