# What the zeta!

James M. Curran[1]
Patrick Buzzini[2]
Tatiana Trejos[3]

30$^{th}$ November 2023

[1]Department of Statistics, University of Auckland
[2]College of Criminal Justice, Sam Houston State University, TX
[3]Department of Forensic and Investigative Science, West Virginia University, WV


I've long feared my sins
would come back to
haunt me and the cost
is more than I can bear.

# A case

- On 3 March, 1991, a float glass window was smashed in a pharmacy in Hamilton, New Zealand
- The offenders took drugs and prescription medicines worth thousands of dollars

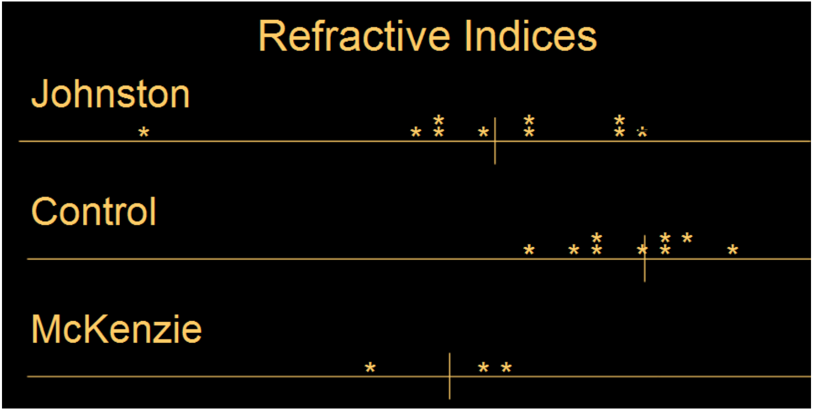
## The suspects

- Police apprehended two suspects, Michael Johnston and John MacKenzie, 90 minutes later
- Their clothing was taken but the drugs were not found

# The evidence

- Recovered from Johnston's clothing
  - small flakes of paint - indistinguishable from crime scene
  - 11 fragments of glass
- MacKenzie's clothing
  - 3 fragments of glass
- 3 fragments were original float surfaces
- 9 control fragments taken from scene window
- Evidence quantified using RI

# The evidence



# Three principles of interpretation

Evett and Weir (1998) proposed three basic principles of evidence interpretation

1. To evaluate the uncertainty of any given proposition it is necessary to consider at least one alternative proposition
2. Scientific interpretation is based on questions of the kind "What is the probability of the evidence given the proposition?"
3. Scientific interpretation is conditioned not only by the competing propositions, but also by the framework of circumstances within which they are to be evaluated

# A likelihood ratio approach to evidence interpretation

Many of my colleagues and I are proponents of what is called the "Bayesian," or "LR," or "logical" approach to evidence interpretation

This way of thinking encapsulates all of the ideas on the previous slide

We believe all forensic scientists should present evidence of in the form of a likelihood ratio

**Odds form of Bayes' Theorem**

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

$$\text{Posterior Odds} = \text{Likelihood Ratio} \times \text{Prior Odds}$$

# The LR under consideration

- We start with the denominator of the $LR$ because the explanation is slightly simpler

$$\Pr(Evidence|\overline{Contact})$$

- We want to know "The probability of the evidence given the suspect was **NOT** in contact with the crime scene"
- Possible reasons:
  - One group of glass there before
  - *and* it just happened (by random chance) to match the crime scene sample
- We write this as

$$\Pr(Evidence|\overline{Contact}) = P_1 S_L f$$

# The LR under consideration

- We start with the denominator of the *LR* because the explanation is slightly simpler

$$\Pr(Evidence|\overline{Contact})$$

- We want to know "The probability of the evidence given the suspect was **NOT** in contact with the crime scene"
- Possible explanations:
  - One group of glass there before
  - *and* it just happened (by random chance) to match the crime scene sample
- We write this as

$$\Pr(Evidence|\overline{Contact}) = P_1 S_L f$$

# The LR under consideration

- The numerator has at least two possible explanations

$$\Pr(Evidence|Contact)$$

- We want to know "The probability of the evidence given the suspect **WAS** in contact with the crime scene"
- Possible explanations:
  - No glass transferred , one group of glass was there before and it matched the crime scene sample by chance **or**
  - one group of fragments transferred from the scene
- We write this as

$$\Pr(Evidence|Contact) = T_0 P_1 S_L f + T_L P_0$$

# The LR under consideration

$$
\begin{aligned}
LR &= \frac{\Pr(E|C)}{\Pr(E|\overline{C})} \\
&= \frac{T_0 P_1 S_L f + T_L P_0}{P_1 S_L f} \\
&= T_0 + \frac{T_L P_0}{P_1 S_L} \times \frac{1}{f} \\
&\approx \frac{T_L P_0}{P_1 S_L} \times \frac{1}{f}
\end{aligned}
$$

# Interpretation of the LR

- Survey estimates give a likelihood ratio of 25 for Johnston and 10 for MacKenzie
- The statement I would give in court is "The **evidence** is 25(10) times more likely if Mr Johnston(MacKenzie) was in contact with the crime scene than if he wasn't"
- This method of interpretation gives a far more intuitive and usable result
- I might downgrade the last statement to "more logically consistent" as the judge, the jury, and the general public have many problems with this statement

# Where do the numbers for the P and S terms from?

➢ The glib and short answer is 'clothing surveys'

- Lau et al. (1997) surveyed the outer clothing and footwear of 213 high school students in Vancouver
- Petterd et al. (1998) who searched the upper outer garments of 2008 people at a shopping centre in Canberra, Australia
- Coulson et al. (2001) search the outer clothing and footwear of 122 people attending a university gymnasium and a private gymnasium.
- Roux et al. (2001) searched for glass fragments on the footwear of 776 people (students, friends, family)...

# Some data

Numbers of groups of glass found on a survey of pairs of shoes taken $N = 776$ people (Roux et al., 2001).

| $n$ | $r_n$ |
|---|---|
| 0 | 754 |
| 1 | 9 |
| 2 | 8 |
| 3 | 4 |
| 4 | 1 |

What do we do with this?

# Raw frequencies only take you so far...

| $n$ | $r_n$ | $\hat{P}_k$ |
|---|---|---|
| 0 | 754 | 0.972 |
| 1 | 9 | 0.006 |
| 2 | 8 | 0.010 |
| 3 | 4 | 0.005 |
| 4 | 1 | 0.001 |

How do I calculate $P_5$?

# A model

Coulson et al. (2001) proposed that the $P$ and $S$ terms could be modelling using a zeta distribution.

$$P_k = \frac{(k+1)^{-\alpha}}{\zeta(\alpha)}, k = 0, 1, 2, \ldots, \alpha > 1$$

and

$$S_n = \frac{n^{-\alpha}}{\zeta(\alpha)}, n = 1, 2, 3, \ldots, \alpha > 1,$$

where $\zeta(\alpha)$ is the Riemann zeta (or Euler-Riemann zeta) function.

# Problem solved. . . right?

SCIENTIFIC & TECHNICAL

## Glass on clothing and shoes of members of the general population and people suspected of breaking crimes

SA COULSON, JS BUCKLETON

*Institute of Environmental Science and Research Ltd., Private Bag 92-021, Auckland, New Zealand*

AB GUMMER

*40 Milton Street, Somerfield, Christchurch, New Zealand*

and

CM TRIGGS

*University of Auckland, Private Bag 92019, Auckland, New Zealand*

*Science & Justice 2001; 41: 39–48*

*Received 27 October 1999; accepted 1 November 2000*

# Problem solved... right?

should fit these criteria. The model used was the maximum likelihood estimation (MLE) of a power series.

The estimation statistic was:

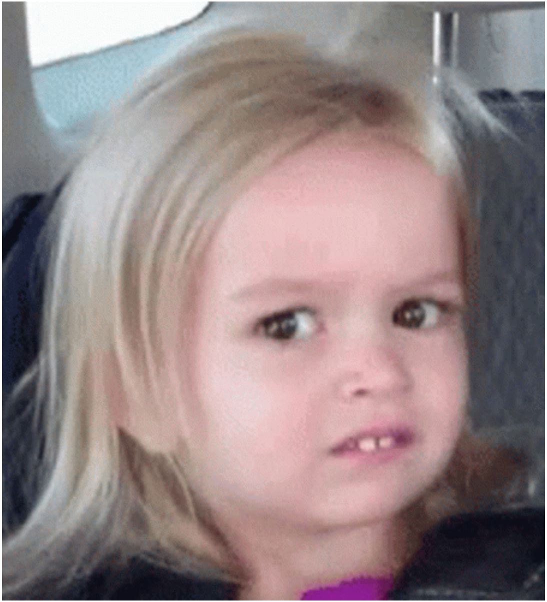
$$-\frac{\zeta(\hat{\alpha})'}{\zeta(\hat{\alpha})} = \frac{1}{N}\sum_{n=1}^{\infty} r_n \log(n+1)$$

This was calculated and the power $\alpha$ and normalising constant $\frac{1}{\zeta(\hat{\alpha})}$

were estimated by interpolation from a table of values of the estimation function $-\frac{\zeta(\hat{\alpha})}{\zeta(\hat{\alpha})}$

for $1 < \alpha < 6$.

# Yeah right



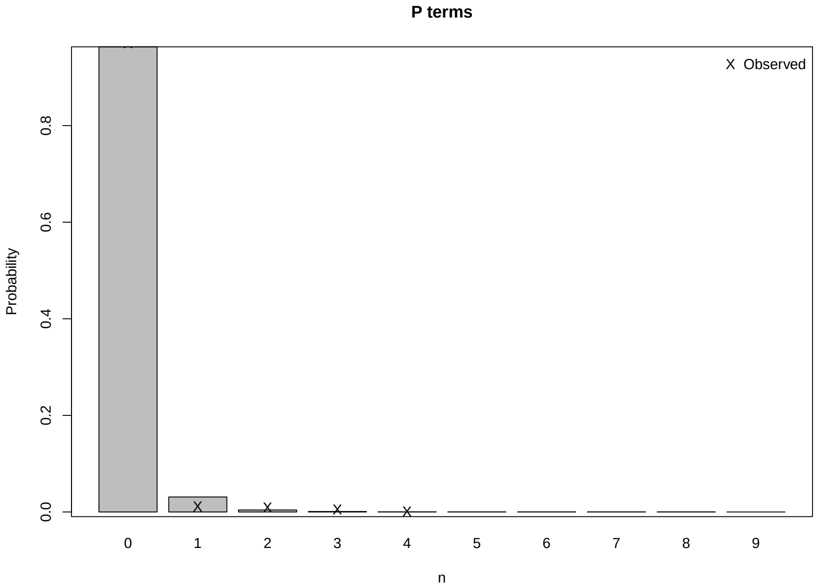
# Producing an R package

- It is relatively easy to implement this in R
- The `optim` function handles the constrained optimisation
- We do, however, need a stable implementation of the Reimann-Zeta function
- And Thomas Yee's `VGAM` package helpfully provides this
- The R package is called `fitPS` and is on CRAN

# Using fitPS with the Roux et al. data

```
> library(fitPS)
Loading required package: foreach

Attaching package: ‘fitPS’

The following object is masked from ‘package:stats’:

    var

> data("Psurveys")
> roux = Psurveys$roux
> fit.roux = fitDist(roux)
> fit.roux
The estimated shape parameter is 4.9544
The standard error of shape parameter is 0.2366
------
NOTE: The shape parameter is reported so that it is consistent with Coulson et al.
However, the value returned is actually s' = shape - 1 to be consistent with the
VGAM parameterisation, which is used for computation. This has flow on effects, for
example in confInt. This will be changed at some point.
------

The first  10 fitted values are:
          P0           P1           P2           P3           P4           P5           P6           P7           P8
9.631547e-01 3.106447e-02 4.167082e-03 1.001917e-03 3.316637e-04 1.344002e-04 6.262053e-05 3.231467e-05 1.802885e-05 1.0697
```

# Using fitPS with the Roux et al. data



# Some simple extensions

Adrian Baddeley said, at the UseR conference in Aalborg in 2015, that one of the interesting things about R is that the computation often inspires the statistics.

This inspired me to think about some simple extensions

# Some simple extensions

You have seen a couple of these already:

- It is very easy to compute a standard error for the estimate
- And it is simple to write a `plot` method for the fitted object

But there is more

- If you have a standard error then you can compute a Wald confidence interval
- And of course, given we are doing maximum likelihood it is not hard to compute a profile likelihood interval

# And I did promise Morris Dancing

- Many surveys yield very little glass
  1. Most observations for the number of groups is zero
  2. and when we do find glass there are only one or two fragments
- This means lots of zeros in the $P$ data, and abundance of ones in the $S$ data

```
> Psurveys$pettard


Number of Groups

  n     rn
---   -----
  0   2002
  1      6
Petterd C, McCallum I, Bradford L, Brinch K, Stewart S (1998). “Glass particles in the clothing of the general
population in Canberra-a survey.” In _Proceedings of the 14th International Symposium on the Forensic Sciences_.
> data("Ssurveys")
> Ssurveys$pettard


Group Size

  n   rn
---   ---
  1    6
Petterd C, McCallum I, Bradford L, Brinch K, Stewart S (1998). “Glass particles in the clothing of the general
population in Canberra-a survey.” In _Proceedings of the 14th International Symposium on the Forensic Sciences_.
```

## Behold its majesty!

The Zero/One Inflated Zeta

$$P_k^{ZIZ} = \begin{cases} \pi + \frac{1-\pi}{\zeta(\alpha)} & , k = 0, \\ \frac{(1-\pi)(k+1)^{-\alpha}}{\zeta(\alpha)} & , k = 1, 2, \ldots, \end{cases}$$

and

$$S_n^{ZIZ} = \begin{cases} \pi + \frac{1-\pi}{\zeta(\alpha)} & , n = 1, \\ \frac{(1-\pi)n^{-\alpha}}{\zeta(\alpha)} & , n = 1, 2, 3, \ldots, \end{cases}$$

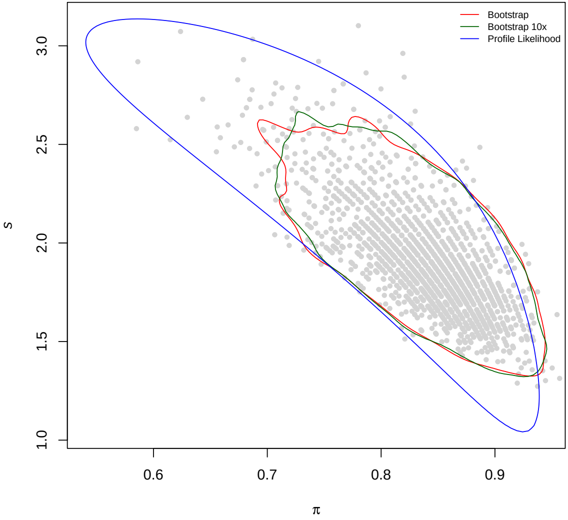
where $\pi \in (0, 1)$ and $\alpha > 1$

# Comparing to the Zeta

| $k$ | $\hat{P}_k^{\mathrm{zeta}}$ | $\hat{P}_k^{\mathrm{ZIZ}}$ |
|---|---|---|
| 0 | 0.9632 | 0.9716 |
| 1 | 0.0311 | 0.0169 |
| 2 | 0.0042 | 0.0053 |
| 3 | 0.0010 | 0.0023 |
| 4 | 0.0003 | 0.0012 |
| 5 | 0.0001 | 0.0007 |

Estimated probability that $k$ groups of glass would be found in shoes of a random member of the population based on the data of Roux et al. (2001) and the zeta and ZIZ models respectively.

# Confidence Intervals



# Acknowledgements


- Thanks of course go to Chris Triggs
- Thomas for VGAM - it would have been a lot harder without it
- And my forensic colleagues for persuading me to do this