Fooling juries with statistics

Peter Haggstrom
www.gotohaggstrom.com
mathsatbondibeach@gmail.com

April 20, 2015

1 Introduction

In his TED (www.ted.com) lecture titled "How stats fool juries" [1], Australian mathematician/statistician, Peter Donnelly who is Professor of Statistical Sciences at Oxford University, demonstrates just how poorly the judicial system is in handling probabilistic and statistical arguments. He does this in two ways. First he takes a really simple coin flipping experiment and then asks the audience for its opinion on the answer. The problem is this: you toss an unbiased coin many times and count the number of tosses until you get the pattern "HTH". You then calculate the average. You repeat the experiment but this time you look for the pattern "HTT". The question is: "Is the average number of trials required to get "HTH" greater than, less than or equal to the average number of trials required to get "HTT"? " The audience not surprisingly said the number of trials were equal. I say "not surprisingly" because the probability of "HTH" is $\frac{1}{8}$ which is the same as the probability of "HTT" so in that sense naive intuition might suggest the expected tosses are equal. In fact they are not: the expected number of trials to get "HTH" is 10 while the expected number of trials to get "HTT" is 8.

As Donnelly points out, even experienced professional matematicians get this wrong. I am unable to do the required calculations in my head (possibly Johnny von Neumann could do if he were alive today, given the documented instances of his genius for complex mental calculations) since they are inherently complicated but there are two ways to convince oneself that the naive answer is wrong. First, write a computer program to generate random strings of "H" and "T" and count how many tosses are required to get the relevant string and then repeat the experiment many times and then take the average. I did this in Mathematica with a short program which generated "0" and "1" for head and tail respectively. The second method, which is far more difficult, is to derive an analytic expression for the relevant expectations (ie average). This requires a knowledge of generating functions and recurrence/renewal theory. I will come to the detail of the calculations shortly. However, Donnelly's point is that even with such a simply stated problem we as "reasonable" people will invariably get it wrong. Once one translates those cognitive defects to the judicial system the problems can involve serious consequences.

With this background Donnelly gives an overview of Bayes' Theorem and its applications before he leads into the infamous Sally Clark case in the UK. Briefly, Sally Clark's first child died of sudden infant death syndrome (SIDS). Unfortunately her second child also died and this led to murder charges being laid against her. An important piece of "evidence" in her conviction was the opinion of a pediatrician who gave evidence that for a family such as Sally Clark's socio-economic status, the probability of one SIDS death is $\frac{1}{8500}$ and hence the probability of two SIDS deaths is $\frac{1}{8500} \times \frac{1}{8500} = \frac{1}{72250000}$. A detailed commentary on the case by lawyer Vincent Scheurer (who has a long standing interest in probability theory and statistics) can be accessed here: `http://understandinguncertainty.org/node/545#notes`. His article also includes links to the relevant judicial decisions.

The details of the Sally Clark legal decision are quite simply outrageous. Certain exculpatory evidence was withheld from the defence. That evidence suggested that the first child died of a bacterial infection. Professor Sir Roy Meadows is the pediatrician who came up with the faulty statistical "analysis" that played a role in the conviction. He had previously done work on the factors associated with SIDs and that research cautioned about assuming the very independence that formed the basis for Meadows' faulty calculation. The President of the Royal Statistical Society said the following to the Lord Chancellor in a letter in 2002:

> "The calculation leading to 1 in 73 million is invalid. It would only be valid if SIDS cases arose independently within families, an assumption that would need to be justified empirically. Not only was no such empirical justification provided in the case, but there are very strong reasons for supposing that the assumption is false. There may well be unknown genetic or environmental factors that predispose families to SIDS, so that a second case within the family becomes much more likely than would be a case in another, apparently similar, family.
>
> A separate concern is that the characteristics used to classify the Clark family were chosen on the basis of the same data as was used to evaluate the frequency for that classification. This double use of data is well recognised by statisticians as perilous, since it can lead to subtle yet important biases.
>
> For these reasons, the 1 in 73 million figure cannot be regarded as statistically valid. The Court of Appeal recognised flaws in its calculation, but seemed to accept it as establishing "... a very broad point, namely the rarity of double SIDS" [AC judgment, para 138]. However, not only is the error in the 1 in 73 million figure likely to be very large, it is almost certainly in one particular direction - against the defendant. Moreover, following from the 1 in 73 million figure at the original trial, the expert used a figure of about

> 700,000 UK births per year to conclude that "... by chance that happening will occur every 100 years". This conclusion is fallacious, not only because of the invalidity of the 1 in 73 million figure, but also because the 1 in 73 million figure relates only to families having some characteristics matching that of the defendant. This error seems not to have been recognised by the Appeal Court, who cited it without critical comment [AC judgment para 115]. Leaving aside the matter of validity, figures such as the 1 in 73 million are very easily misinterpreted. Some press reports at the time stated that this was the chance that the deaths of Sally Clark's two children were accidental. This (mis-)interpretation is a serious error of logic known as the Prosecutor's Fallacy1 . The jury needs to weigh up two competing explanations for the babies' deaths: SIDS or murder. The fact that two deaths by SIDS is quite unlikely is, taken alone, of little value. Two deaths by murder may well be even more unlikely. What matters is the relative likelihood of the deaths under each explanation, not just how unlikely they are under one explanation.
>
> The case of R v. Sally Clark is one example of a medical expert witness making a serious statistical error. Although the Court of Appeal judgment implied a view that the error was unlikely to have had a profound effect on the outcome of the case, it would be better that the error had not occurred at all. Although many scientists have some familiarity with statistical methods, statistics remains a specialised area. The Society urges you to take steps to ensure that statistical evidence is presented only by appropriately qualified statistical experts, as would be the case for any other form of expert evidence." [2]

The fact that this "evidence" was clearly not even expert statistical evidence (Meadows was not a statistician) but still found its way into the system suggests that not only was the defence counsel comatose but how little care the judge had in letting such dangerous nonsense being put to the jury. The intellectual gymnastics and dishonesty of the two appeal cases merely reinforce the view that the judicial system is more concerned with judicial and administrative convenience than anything approximating justice. Mrs Clark died in 2007 of alcohol poisoning. I wonder what the probability is that she drank herself to death. Maybe Professor Meadows knows. Her husband tried to have Meadows struck off for serious professional misconduct and was initially successful but Meadows appealed and the appeal court held that what he did was merely "misconduct", not "serious misconduct".

The judicial arrogance and stupidity do not stop here. A 2013 English Appeal Court case [3] has had the effect of banning Bayesian reasoning in court cases. The judges made the following comments:

> "The chances of something happening in the future may be expressed in terms of percentage. Epidemiological evidence may enable doctors to say

> that on average smokers increase their risk of lung cancer by X%. But you cannot properly say that there is a 25 per cent chance that something has happened: Hotson v East Berkshire Health Authority [1987] AC 750. Either it has or it has not. ″

Cambridge statistician David Spiegelhalter explains the substance of the case in his blog [4]. His exasperation at trying to explain the concepts to these serial dopes is clearly evident. They just don't get it at any level. It is like trying to explain quantum mechanics to Neanderthal man. Peter Donnelly's explanation of disease testing in his TED talk would be a good starting place for these judges as it explains the interaction between likely and unlikely events in a practical way that even they might be able to comprehend, although I suspect that the judicial world is too enamoured of word games and false logic to ever really understand the concepts at issue.

With this depressing background let's move to a mathematical analysis of the ″simple″ problem Peter Donnelly presented to his audience.

2 A simple program to verify the expected number of tosses

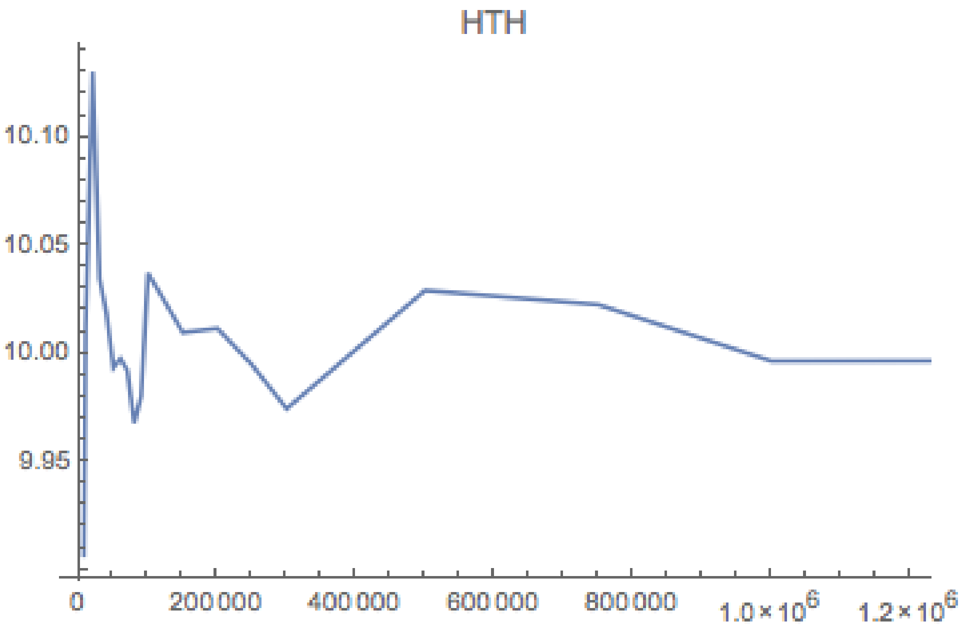
Here is a program in Mathematica 10.1.0.0 which calculates the average number of trials needed to generate ″HTH″. In this case 10,000 trials were run.

```
r = 10;
u[1] = 0;
u[2] = 1;
u[3] = 0;
lntot = 0;
trials = 1000 r;
Do[
  t = Table[RandomInteger[], {i, 1, 1000}];
  ln = 3;
  j = 1;
  While[
   ¬ ((t[[j]] == u[1]) ∧ (t[[j + 1]] == u[2]) ∧ (t[[j + 2]] == u[3])),
   ln = ln + 1; j ++];
  lntot = lntot + ln,
  {k, 1, trials}];
v[r] = N[lntot / trials]
```

```
10.1196
```

The average number of tosses obtained is 10.1196. This program works by generating a random string of 0*s* (ie "H") and 1*s* (ie "T") (1000 are generated so that one can be sure that a run of "HTH" will actually be generated) and the string "HTH" is slid along the complete random string until the first match occurs.

When the program is run for trials ranging from 10^3 to 20×10^6 the graph of the results is as follows:



The raw data is as follows (note that since the random number generator was restarted the average for 10,000 runs is different from that given above):

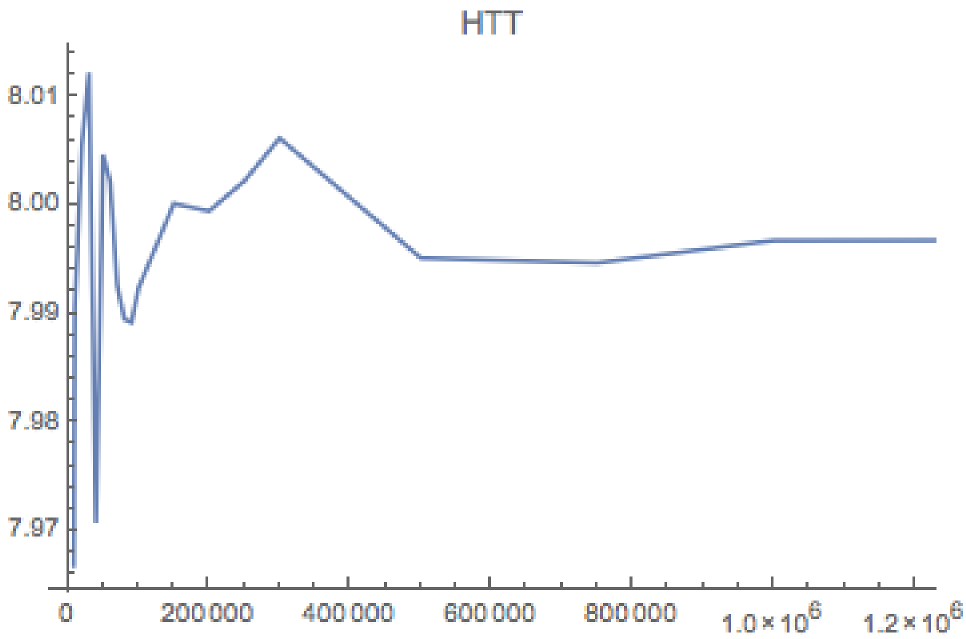
Number of trials	Average number of tosses to get "HTH"
1000	9.66
10000	10.0133
20000	10.1303
30000	10.0346
40000	10.0185
50000	9.99354
60000	9.997572
70000	9.9923
80000	9.96775
90000	9.98018
100000	10.0372
150000	10.0097
200000	10.0116
250000	9.99474
300000	9.97449
500000	10.0292
750000	10.0226
1000000	9.99648
20000000	10.0033

When the program is run 10,000 times for "HTT" the average is 8.0507:

```
r = 10;
u[1] = 0;
u[2] = 1;
u[3] = 1;
lntot = 0;
trials = 1000 r;
Do[
  t = Table[RandomInteger[], {i, 1, 1000}];
  ln = 3;
  j = 1;
  While[
   ¬ ((t[[j]] == u[1]) ∧ (t[[j + 1]] == u[2]) ∧ (t[[j + 2]] == u[3])),
   ln = ln + 1; j ++];
  lntot = lntot + ln,
  {k, 1, trials}];
v[r] = N[lntot / trials]
```

```
8.0507
```

The corresponding graph for "HTT" for trials ranging from 10^3 to 20×10^6 is as follows:



The raw data is as follows (note that since the random number generator was restarted the average for 10,000 runs is different from that given above):

Number of trials	Average number of tosses to get "HTT"
1000	7.867
10000	7.9906
20000	8.0056
30000	8.01223
40000	7.97073
50000	8.00466
60000	8.00217
70000	7.9926
80000	7.98951
90000	7.98921
100000	7.99238
150000	8.00015
200000	7.99946
250000	8.00224
300000	8.00619
500000	7.99514
750000	7.9947
1000000	7.99677
20000000	7.99888

3 An analytic expression for the expected number of tosses

To understand what follows it is necessary to read Chapters 11 and 13 of [5] where the theory of generating functions is developed. Feller gives the following concrete coin tossing example [5, page 278]. Let q_n be the probability that in n tosses of an ideal coin no run of three consecutive heads appears. He notes that q_n is not a probability distribution: if p_n is the probability that the first run of three consecutive heads end at the n^{th} trial, then p_n is a probability distrbution, and q_n represents its "tails", $q_n = p_{n+1} + p_{n+2} + \dots$). He then goes on to state the following recurrence relation:

$$q_n = \frac{1}{2}q_{n-1} + \frac{1}{4}q_{n-2} + \frac{1}{8}q_{n-3} \tag{1}$$

He explains (1) as follows:

"In fact, the event that n trials produce no sequence "HHH" can occur only when the trials begin with "T", "HT", or "HHT". The probabilities that the following trials lead to no run "HHH" are q_{n-1}, q_{n-2} and q_{n-3} respectively and the right hand side of (1) therefore contains the probabilities of the three mutually exclusive ways in which the event "no run HHH" can occur."

Clearly $q_0 = q_1 = q_2 = 1$ and q_n can be found recursively from (1). For instance, $q_9 = 0.535156$. To derive the generating function relating to (1) we define:

$$Q(s) = \sum_{n=0}^{\infty} q_n s^n \tag{2}$$

We multiply (1) by s^n and sum for $n \geq 3$:

$$\begin{aligned}
\sum_{n=3}^{\infty} q_n s^n =& \frac{1}{2}\sum_{n=3}^{\infty} q_{n-1}\, s^n + \frac{1}{4}\sum_{n=3}^{\infty} q_{n-2}\, s^n + \frac{1}{8}\sum_{n=3}^{\infty} q_{n-3}\, s^n \\
Q(s) - 1 - s - s^2 =& \frac{1}{2}(q_2 s^3 + q_3 s^4 + \dots) + \frac{1}{4}(q_1 s^3 + q_2 s^4 + \dots) + \frac{1}{8}(q_0 s^3 + q_1 s^4 + \dots) \\
=& \frac{s}{2}(q_2 s^2 + q_3 s^3 + \dots) + \frac{s^2}{4}(q_1 s + q_2 s^2 + \dots) + \frac{s^3}{8}(q_0 + q_1 s + \dots) \\
=& \frac{s}{2}(Q(s) - 1 - s) + \frac{s^2}{4}(Q(s) - 1) + \frac{s^3}{8}Q(s) \\
\therefore Q(s) =& \frac{2s^2 + 12s + 8}{8 - 4s - 2s^2 - s^3}
\end{aligned} \tag{3}$$

Feller shows [5, p.278] that:

$$q_n \approx \frac{1.236840}{1.0873778^{n+1}} \tag{4}$$

Using (4) we find that $q_9 = 0.535191$ which closely approximates q_9 by the recursive method.

The connection between a generating function and probabilities is explained by Feller in [5, pages 264-266]. For instance, if X is the number scored in a throw of a perfect die the probability distribution of X has the generating function $\frac{s+s^2+s^3+s^4+s^5+s^6}{6}$.

More generally, if X is a random variable assuming values $0, 1, 2, 3, \ldots$ we define:

$$P[X = j] = p_j \ , \ P[X > j] = q_j \tag{5}$$

Given (5) we have that:

$$q_k = p_{k+1} + p_{k+2} + \ldots \tag{6}$$

The generating functions of the sequences p_j and q_j are then:

$$P(s) = p_0 + p_1 s + p_2 s^2 + p_3 s^3 + \ldots \tag{7}$$

$$Q(s) = q_0 + q_1 s + q_2 s^2 + q_3 s^3 + \ldots \tag{8}$$

Note that $P(1) = 1$ so that $P(s)$ converges absolutely for $-1 \leq s \leq 1$ and because the coefficients of $Q(s)$ are all less than 1 the series for $Q(s)$ will converge for $-1 < s < 1$. Note that $Q(s)$ is not a probability generating function since the set of probabilites q_j is not a discrete probability function (see (6)).

This leads to the following theorem:

For $-1 < s < 1$:

$$Q(s) = \frac{1 - P(s)}{1 - s} \tag{9}$$

This can be proved be noting that the coefficient of s^n in $(1-s).Q(s)$ is $q_n - q_{n-1} = -p_n$ for $n \geq 1$. When $n = 0$ we have $q_0 = p_1 + p_2 + \cdots = 1 - p_0$. Hence, $(1-s).Q(s) = 1 - P(s)$.

Because we are interested ultimately in the expected number of tosses we need to get a general expression for the expectation of a particular random variable.

Differentiating (7) we get:

$$P^{'}(s) = \sum_{k=1}^{\infty} k\, p_k\, s^{k-1} \tag{10}$$

(10) certainly converges for $-1 < s < 1$ and for $s = 1$ we have formally:

$$E[X] = \sum_{k=1}^{\infty} k\, p_k \tag{11}$$

If the series in (10) exists, the expectation exists and, since $|P^{'}(s)| \leq \sum_{k=1}^{\infty} k\, p_k\, |s^{k-1}| \leq \sum_{k=1}^{\infty} k\, p_k$ for $-1 \leq s \leq 1$, the derivative is continuous on $[-1, 1]$. When $\sum_{k=1}^{\infty} k\, p_k$ diverges the expectation is said to be infinite (ie $P^{'}(1) = E[X] = \infty$).

By the Mean Value Theorem there exists an ξ between 1 and s such that:

$$P^{'}(\xi) = \frac{1 - P(s)}{1 - s} = Q(s) \tag{12}$$

Because both $P(s)$ and $Q(s)$ are monotone, $P^{'}(s)$ and $Q(s)$ will have the same limit (finite or infinite). This means that:

$$E[X] = \sum_{j=1}^{\infty} j\, p_j = \sum_{k=1}^{\infty} q_k \tag{13}$$

or, in terms of generating functions:

$$E[X] = P^{'}(1) = Q(1) \tag{14}$$

Although we don't need it for the purposes of this article the variance is:

$$Var[X] = P^{''}(1) + P^{'}(1) - (P^{'}(1))^2 = 2Q^{'}(1) + Q(1) - (Q(1))^2 \tag{15}$$

Consistent with Feller's terminology [5, pages 307-9] an attribute $\mathcal{E}$ defines a recurrent event if:

(a) In order that $\mathcal{E}$ occurs at the n^{th} and the $(n+m)^{th}$ place of the sequence $(\mathcal{E}_{j_1}, \mathcal{E}_{j_2}, \ldots \mathcal{E}_{j_{n+m}})$ it is necessary and sufficient that $\mathcal{E}$ occurs at the last place of each of the two subsequences $(\mathcal{E}_{j_1}, \mathcal{E}_{j_2}, \ldots \mathcal{E}_{j_n})$ and $(\mathcal{E}_{j_{n+1}}, \mathcal{E}_{j_{n+2}}, \ldots \mathcal{E}_{j_{n+m}})$;

(b) If ($\mathcal{E}$ occurs at the end of the n^{th} place then identically:

$$P[\mathcal{E}_{j_1}, \mathcal{E}_{j_2}, \ldots \mathcal{E}_{j_{n+m}}] = P[\mathcal{E}_{j_1}, \mathcal{E}_{j_2}, \ldots \mathcal{E}_{j_n}]\, P[\mathcal{E}_{j_{n+1}}, \mathcal{E}_{j_{n+2}}, \ldots \mathcal{E}_{j_{n+m}}] \tag{16}$$

With each recurrent event $\mathcal{E}$ (which could be something like "HTH" or "HTT") there are associated two sequences defined for $n = 1, 2, \ldots$:

$$u_n = P[\mathcal{E} \text{ occurs on the } n^{th} \text{ trial}] \tag{17}$$

$$f_n = P[\mathcal{E} \text{ occurs for the first time on the } n^{th} \text{ trial}] \tag{18}$$

The generating functions associated with u_n and f_n are:

$$U(s) = \sum_{k=0}^{\infty} u_k\, s^k \tag{19}$$

$$F(s) = \sum_{k=0}^{\infty} f_k\, s^k \tag{20}$$

We define $u_0 = 1$ and $f_0 = 0$.

The u_k do not form a probability distribution and for representative cases $\sum u_k = \infty$. However, because the events "$\mathcal{E}$ occurs for the first time at the n^{th} trial"are mutually exclusive we will have:

$$f = F(1) = \sum_{n=1}^{\infty} f_n \leq 1 \tag{21}$$

This means that $1 - f$ can be interpreted as the probability that $\mathcal{E}$ does not occur in an indefinitely prolonged sequence of trials.

The probability that $\mathcal{E}$ occurs for the first time on trial ν then again on a later trial $n > \nu$ is:

$$f_\nu\, u_{n-\nu} \tag{22}$$

This is purely a result of the definitions of u_n and f_n. The probability that $\mathcal{E}$ occurs at the n^{th} trial for the first time is:

$$f_n = f_n\, u_0 \tag{23}$$

Note the need for $u_0 = 1$ in (23). Because of mutual exclusivity we have:

$$u_n = f_1\, u_{n-1} + f_2\, u_{n-2} + \cdots + f_n\, u_0 \quad \text{for } n \geq 1 \tag{24}$$

At this point we need the concept of a convolution (see [5] pages 266-8). Suppose X and Y are independent integral valued random variables with probability distributions $P[X = j] = a_j$ and $P[Y = j] = b_j$. Then the event $(X = j, Y = k)$ has probability $a_j\, b_k$. If $S = X + Y$ then the event $S = r$ is the sum of the mutually exclusive events:

$(X = 0, Y = r),\ (X = 1, Y = r - 1), \ldots, (X = r, Y = 0)$

If $c_r = P[S = r]$ then the probability distribution of S is given by:

$$c_r = a_0 b_r + a_1 b_{r-1} + a_2 b_{r-2} + \cdots + a_{r-1} b_1 + a_r b_0 \tag{25}$$

Equation (25) is usually written a follows using the convolution symbol:

$$\{c_k\} = \{a_k\} * \{b_k\} \tag{26}$$

The sequences $\{a_k\}$ and $\{b_k\}$ have generating functions $A(s) = \sum a_k\, s^k$ and $B(s) = \sum b_k\, s^k$ respectively. When we multiply the two generating functions termwise and collect like powers of s we get:

$$\begin{split} a_0 b_0 + (a_0 b_1 + a_1 b_0)s + (a_0 b_2 + a_1 b_1 + a_2 b_0)s^2 + (a_0 b_3 + a_1 b_2 + a_2 b_1 + a_3 b_0)s^3 + \\ \cdots + (a_0 b_r + a_1 b_{r-1} + a_2 b_{r-2} + \cdots + a_{r-1} b_1 + a_r b_0)s^r + \ldots \end{split} \tag{27}$$

Thus we have:

$$C(s) = A(s)\, B(s) \tag{28}$$

Going back to (24) we see that the right hand side is the convolution $\{f_k\} * \{u_k\}$ which by (28) has generating function $F(s)\, U(s)$. The left hand side of (24) is $U(s) - u_0 = U(s) - 1$, hence $U(s) - 1 = F(s)\, U(s)$. Thus the basic relationship between the generating functions of $\{f_k\}$ and $\{u_k\}$ is:

$$F(s) = \frac{U(s) - 1}{U(s)} \tag{29}$$

Given the theory developed above for the expectation it is convenient to recast (29) as follows:

$$F(s) = \frac{1}{1 + \frac{1}{U(s) - 1}} = \frac{1}{1 + (1 - s)Q(s)} \tag{30}$$

Then:

$$F'(s) = \frac{Q(s) - (1-s)Q'(s)}{(1 + (1-s)Q(s))^2} \tag{31}$$

But:

$$\mu = F'(1) = Q(1) \tag{32}$$

With this background we can now perform the required expectations for "HTH" and "HTT".

3.1 Expectation calculation for "HTH"

To develop the recurrence relationship for the probability of the first occurence of "HTH" consider the following diagram:

There are matches at positions 1 and 3. In what follows we denoted the probability of a head by p and the probability of a tail by $q = 1 - p$. Later we will of course apply $p = q = \frac{1}{2}$ but the general approach is preferable. The probability of "HTH" is p^2q. The total probability of "HTH" in the last 3 trials is:

$$p^2q = u_n + pq\, u_{n-2} \tag{33}$$

Note that $u_0 = 1$ and $u_1 = u_2 = 0$.

Multiplying (33) by s^n and summing over k from 3 to ∞ (the equation is only valid for that range) we have:

$$
\begin{aligned}
\sum_{k=3}^{\infty} p^2 q\, s^k &= \sum_{k=3}^{\infty} u_k\, s^k + \sum_{k=3}^{\infty} pq\, u_{k-2}\, s^k \\
p^2 q \Big(\frac{1}{1-s} - (1+s+s^2)\Big) &= U(s) - (u_0 + u_1\, s + u_2\, s^2) + pq\,(u_1 s^3 + u_2 s^4 + \dots) \\
p^2 q \Big(\frac{1}{1-s} - \frac{(1-s^3)}{1-s}\Big) &= U(s) - 1 + pq\, s^2\,(u_1 s + u_2 s^2 + \dots) \\
\frac{p^2 q\, s^3}{1-s} &= U(s) - 1 + pq\, s^2 (U(s) - 1) \\
&= (U(s) - 1)\,(1 + pqs^2) \\
\therefore\ \frac{1}{U(s)-1} &= \frac{(1-s)(1+pq\, s^2)}{p^2 q\, s^3}
\end{aligned}
\tag{34}
$$

Recalling (30) we have:

$$
F(s) = \frac{1}{1 + (1-s)\frac{(1+pq\, s^2)}{p^2 q\, s^3}} \tag{35}
$$

where $Q(s) = \frac{1+pq\, s^2}{p^2 q\, s^3}$

Using (32) and the fact that $p = q = \frac{1}{2}$ we have that the mean number of tosses to obtain "HTH" is:

$$
\mu = Q(1) = \frac{1 + \frac{1}{4}}{\frac{1}{8}} = 10 \tag{36}
$$

3.2 Expectation calculation for "HTT"

We apply the same logic as before as illustrated by this diagram:

The required recurrence relation is:

$$pq^2 = u_n \quad \text{for } n \geq 1 \tag{37}$$

Note $u_0 = 1$.

As before, multiplying by s^k and summing we have:

$$\begin{aligned} \sum_{k=1}^{\infty} pq^2 s^k &= \sum_{k=1}^{\infty} u_k s^k \\ pq^2\Big(\frac{1}{1-s} - 1\Big) &= U(s) - 1 \\ \therefore \frac{1}{U(s)-1} &= \frac{1-s}{pq^2 s} \end{aligned} \tag{38}$$

In this case:

$$Q(s) = \frac{1}{pq^2 s} \tag{39}$$

With $p = q = \frac{1}{2}$ the mean number of tosses for "HTT" is:

$$\mu = Q(1) = \frac{1}{\frac{1}{8}} = 8 \tag{40}$$

4 References

1. http://www.ted.com/talks/peter_donnelly_shows_how_stats_fool_juries?language=en#

2.http://www.rss.org.uk/Images/PDF/influencing-change/rss-use-statistical-evidence-court-
pdf

3. http://www.bailii.org/ew/cases/EWCA/Civ/2013/15.html

4. http://understandinguncertainty.org/court-appeal-bans-bayesian-probability-and-sherloc

5. William Feller, "An Introduction to Probability Theory and its Applications", Third Edition. Volume 1, Wiley.

5 History

Created 19/04/2015

Corrected some typos 20/4/2015