# Biased Coins

Or “*Why I am considering buying 840 blank dice*”

Michael Gibson

# Some “QI” Questions…

What is the total value of the coins in this bag?

(I am holding up a bag which appears to contain 6 10p coins.)

If a coin is tossed once and shows heads, what is the probability that it will show heads next time?

What if the coin was chosen from a collection with a uniform distribution of biases?

# Defining Bias…

We can define the bias of a coin by its *weighting,* $W$, i.e. the probability of heads on any toss.

$$P(H \mid W=w)=w,$$

$$P(T \mid W=w)=1-w$$

We consider an infinite collection of coins with a uniform distribution of weightings, i.e.

$$f_W(w)=1 \qquad \text{if } 0\leq w\leq 1,$$

$$f_W(w)=0 \qquad \text{otherwise.}$$

This will later be referred to as “our model”.

# Outcomes for 1 or 2 tosses

$$P(H)=\int_0^1 f(w)P(H\mid W=w)\,dw=\int_0^1 1\times w\,dw=1/2$$

$$P(T)=\int_0^1 f(w)P(T\mid W=w)\,dw=\int_0^1 1\times(1-w)\,dw=1/2$$

$$P(HH)=\int_0^1 f(w)P(HH\mid W=w)\,dw=\int_0^1 1\times w^2\,dw=1/3$$

$$P(HT)=\int_0^1 f(w)P(HT\mid W=w)\,dw=\int_0^1 1\times w(1-w)\,dw=1/6$$

$$P(TH)=\int_0^1 f(w)P(TH\mid W=w)\,dw=\int_0^1 1\times(1-w)w\,dw=1/6$$

$$P(TT)=\int_0^1 f(w)P(TT\mid W=w)\,dw=\int_0^1 1\times(1-w)^2\,dw=1/3$$

# Results for 1 or 2 tosses

Interestingly, we find

$$P(HH) = P(HT \text{ or } TH) = P(TT) = 1/3$$

To answer our original question, we can use Bayes' Theorem:

$$PHHH = PHHH\,P(HH)/P(H) = 1 \times 1/3 \;\; /1/2$$
$$= 2/3$$

# Extension to *r* heads from *n* tosses

We will use the shorthand "*r/n*" for this. Binomial theory gives:

$$Pr/n \;\; W = w = \binom{n}{r} w^{r} (1-w)^{n-r}.$$

For a randomly selected coin, we have

$$P(r/n) = \binom{n}{r} \int_0^1 w^{r} (1-w)^{n-r} \, dw = 1/n+1$$

What if we toss a coin *n* times, resulting in *r* heads? What is the probability that we get heads on the next toss? Using Bayes' Theorem again, it can be shown that

$$P((r+1)/(n+1) \mid r/n) = P(r/n \mid (r+1)/(n+1)) P((r+1)/(n+1)) / P(r/n) = r+1/n+2$$

Spend a moment thinking about the meaning of this result

# Replication With 2 Coins?

Suppose we have just 2 coins in the collection, but can choose their biases to be $1/2 \pm a$. We require

$$P(HH) = 1/2\,(1/2 + a)^2 + 1/2\,(1/2 - a)^2$$
$$= 1/4 + a^2 = 1/3$$

which gives $a = 1/6\,\sqrt{3}$.

It can be shown that this actually replicates our model perfectly for up to **three** tosses of the chosen coin. For example,

$$P(HHH) = 1/2\,(1/2 + 1/6\,\sqrt{3}\,)^3 + 1/2\,(1/2 - 1/6\,\sqrt{3}\,)^3$$

$$= 1/8 + 1/2 \times 1/36 \times 9 = 1/4$$

# Physical Replication With Actual Coins

Suppose we have $N$ coins, $a$ fair, $b$ double-headed and $b$ double-tailed. We require

$$P(HH)=1/N\ (a(1/2)^2+b.1^2+b.0^2)$$

$$=1/a+2b\ (1/4\ a+b)=1/3$$

Which leads to $a=4b$. So $a=4$, $b=1$ will work!

**if you have a collection of coins from which you repeatedly choose a coin at random, toss it up to 3 times, then return it to the collection, the results obtained from a collection consisting of 4 fair coins, 1 double-headed coin and 1 double-tailed coin are *statistically indistinguishable* from those that would be obtained from an infinite collection of coins with a continuous uniform distribution of biases.**

# Let's Keep Going!

OK so we've done all we realistically can with coins, but we can go further if we simulate coins using 6-sided **dice** with "H" or "T" written on their faces.

There are 7 different biases possible now. We will restrict ourselves to collections with a symmetric distribution, described by the vector $(a,b,c,d)$ as follows:

| Ratio H:T | Number of such dice |
|---|---|
| 0:6 | *d* |
| 1:5 | *c* |
| 2:4 | *b* |
| 3:3 | *a* |
| 4:2 | *b* |
| 5:1 | *c* |
| 6:0 | *d* |

# Let's Keep Going!

If we initially restrict ourselves to only 2 non-zero components, then we find the following vectors (and their multiples) work for up to **3 throws**:

(2, 0, 3, 0)

(4, 0, 0, 1)  * note this is equivalent to our coins earlier

(0, 1, 2, 0)

(0, 3, 0, 1)

Any (physically possible) linear combination of these will also work for at least 3 throws (but note that these 4 vectors can't all be linearly independent, for the same reason).

It turns out that adding the first 3 together to give (6, 1, 5, 1) actually works for **5 throws**, as does the vector (52, 27, 54, 13), or any linear combination these two.  The former seems optimal.

# Let’s Keep Going!

**If you have a collection of six sided dice, representing coins by having either heads or tails marked on each face, in the ratio H:T, from which you repeatedly choose one at random, roll it up to 5 times, then return it to the collection, the results obtained from a collection consisting of 20 dice as detailed below**

| Ratio H:T | Number of dice |
|---|---|
| 0:6 | 1 |
| 1:5 | 5 |
| 2:4 | 1 |
| 3:3 | 6 |
| 4:2 | 1 |
| 5:1 | 5 |
| 6:0 | 1 |

**are *statistically indistinguishable* from those that would be obtained from an infinite collection of coins with a continuous uniform distribution of biases.**

# Let’s Keep Going!

We still have two linearly independent vectors to combine, so this should give us hope of going further still.

There is only one vector (plus its multiples) which will work, and it turns out to be (272, 27, 216, 41).

This matches our model for up to **7 throws**.

# Let's Keep Going!

**If you have a collection of six sided dice, representing coins by having either heads or tails marked on each face, in the ratio H:T, from which you repeatedly choose one at random, roll it up to 7 times, then return it to the collection, the results obtained from a collection consisting of 840 dice as detailed below**

| Ratio H:T | Number of dice |
|---|---|
| 0:6 | 41 |
| 1:5 | 216 |
| 2:4 | 27 |
| 3:3 | 272 |
| 4:2 | 27 |
| 5:1 | 216 |
| 6:0 | 41 |

**are *statistically indistinguishable* from those that would be obtained from an infinite collection of coins with a continuous uniform distribution of biases.**

And that is why I am considering buying 840 blank dice.

Anyone fancy chipping in?