Three systems of vocabulary

	1	2	3
Position on chromosome	locus	locus	locus
Protein-coding locus	gene	gene	gene
Physical copy of DNA at locus	gene	allele	gene copy
One of several variants at a locus	allele	allele	allele

1 is classical usage, 2 is Gillespie's, and we try to keep to 3.

> *Consistency is the hobgoblin of little minds*
> —*Ralph Waldo Emerson*

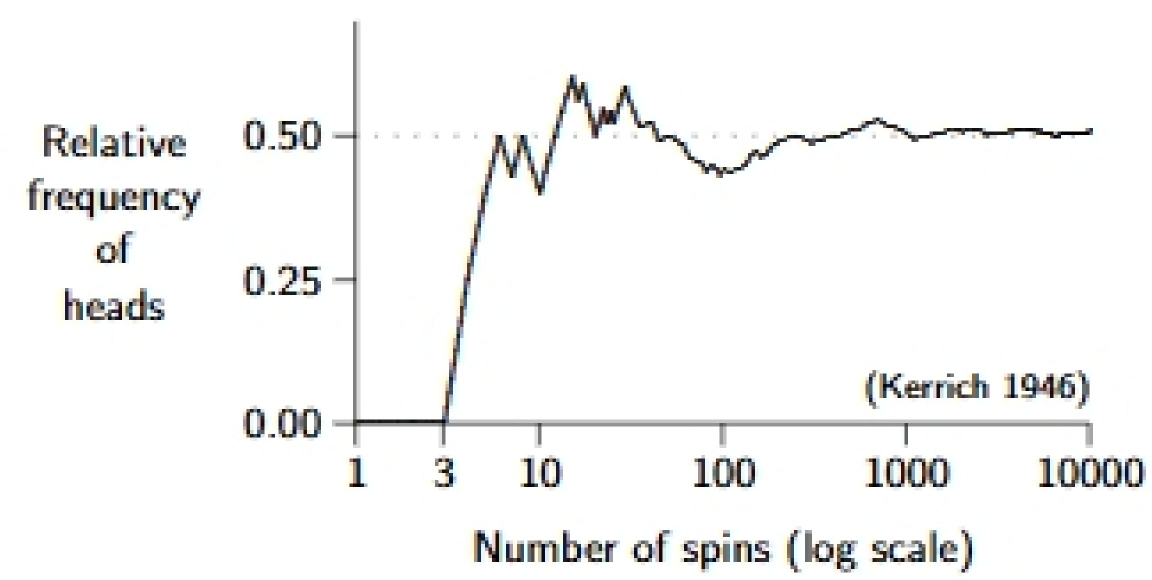
Illustration of classical usage

> Those organisms (homozygotes) which received like genes, in any pair of corresponding loci, from their two parents, would necessarily hand on genes of this kind to all of their offspring alike; whereas those (heterozygotes) which received from their two parents genes of different kinds... (Fisher, 1930, p. 8)

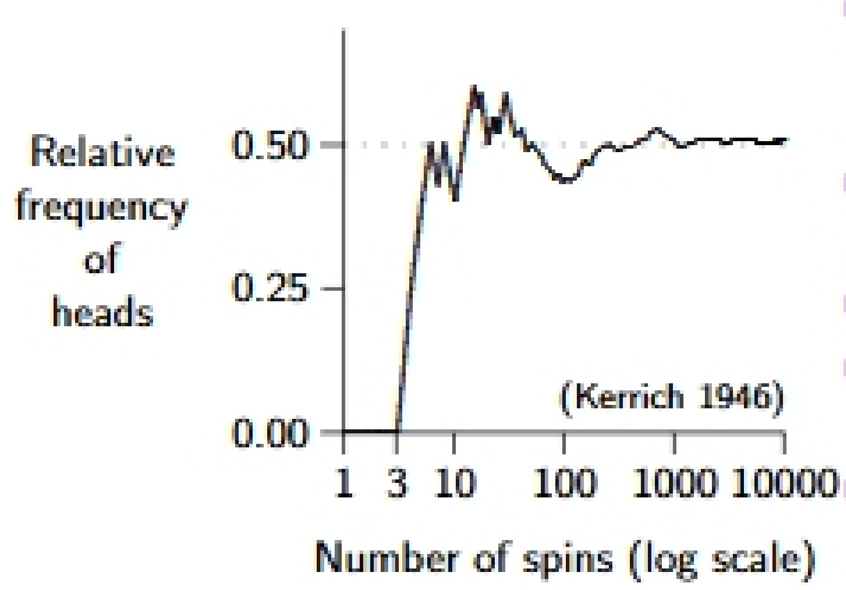
Probability

Alan R. Rogers

September 10, 2014



Probability and relative frequency in repeated trials



- rel. freq. of heads gradually approaches limiting value.
- Limiting value is the *probability* of heads
- Need not equal 1/2.
- We estimate probabilities from relative frequencies.
- We never know them exactly.

Kerrich's "urn" experiment

(○ ○ ● ●)

- Urn contains 4 balls: 2 black and 2 white
- Mix them up.
- Draw one at random
- Mix them up and draw a second *without* replacing first.
- Repeat 5000 times.

Results from Kerrich's urn experiment

First ball	Second ball Black	White	sum
Black	756	1689	2445
White	1688	867	2555
sum	2444	2556	5000

- If 1st ball is B, 2nd is likely to be W
- And vice versa

Model of Kerrich's urn experiment

Assumption: we are equally likely to draw any ball in urn.

1st Ball

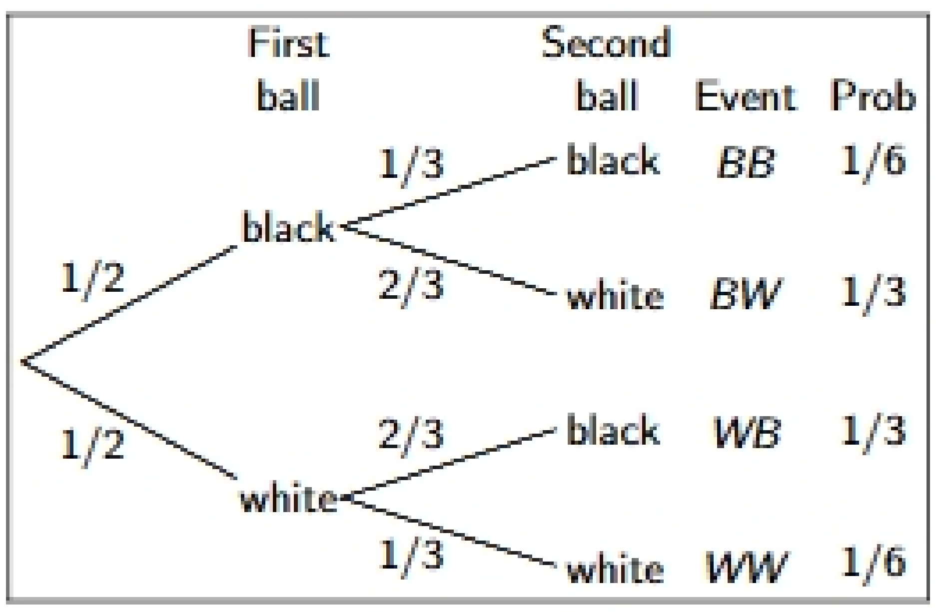
(∘ ∘ • •)

We are equally likely to draw black or white

2nd Ball

First ball	Remaining balls	Prob. of black
•	(∘ ∘ •)	1/3
∘	(∘ • •)	2/3

2nd ball usually black if 1st was white, and vice versa.



Tree diagram for urn model

Kerrich's urn experiment: model versus data

Event	Theoretical probability	Observed relative frequency
BB	0.167	0.151
BW	0.333	0.338
WB	0.333	0.338
WW	0.167	0.173

Theory and observation are not identical, but they are close.

Why do we multiply along branches?

Conditional probability

- What is the conditional probability that the 2nd ball is white given that the first was black?
- 2/3.
- Called a *conditional probability* and written

$$\Pr[\text{2nd ball white}|\text{1st one black}].$$

- "|" is pronounced "given."

Conditional relative frequencies

First ball	Second ball Black	White	sum
Black	**756**	**1689**	**2445**
White	1688	867	2555
sum	2444	2556	5000

- On trials where the 1st ball was black, how often was the 2nd white?
- A fraction 1689/2445 of the time, or ≈ 0.69.

This is a conditional relative frequency. If the number of trials is large, this approximates a conditional probability.

The results of 20,000 throws with two dice (Wolf 1850, cited in Bulmer 1967)

Black	White 1	2	3	4	5	6	$\sum$	f
1	547	587	500	462	621	690	3407	.170
2	609	655	497	535	651	**684**	**3631**	.182
3	514	540	468	438	587	629	3176	.159
4	462	507	414	413	509	611	2916	.146
5	551	562	499	506	658	672	3448	.172
6	563	598	519	487	609	646	3422	.171
$\sum$:	3246	3449	2897	2841	3635	3932	20000	1.000
f:	.162	.172	.145	.142	.182	.197	1.000	

- What is the conditional frequency of $W6$ given $B2$?
- $684/3631 \approx 0.188$

Product rule for relative frequencies

How often did Kerrich get $B1$ and $W2$?

First ball	Second ball Black	White	sum
Black	756	**1689**	2445
White	1688	867	2555
sum	2444	2556	**5000**

A fraction 1689/5000 of the time.

$$\frac{1689}{5000} = \frac{1689}{2445} \times \frac{2445}{5000}$$

$$\overbrace{\frac{1689}{5000}}^{f(B1\ \&\ W2)} = \overbrace{\frac{1689}{2545}}^{f(W2|B1)} \times \overbrace{\frac{2445}{5000}}^{f(B1)}$$

As N increases, relative frequencies (f) become probabilities.

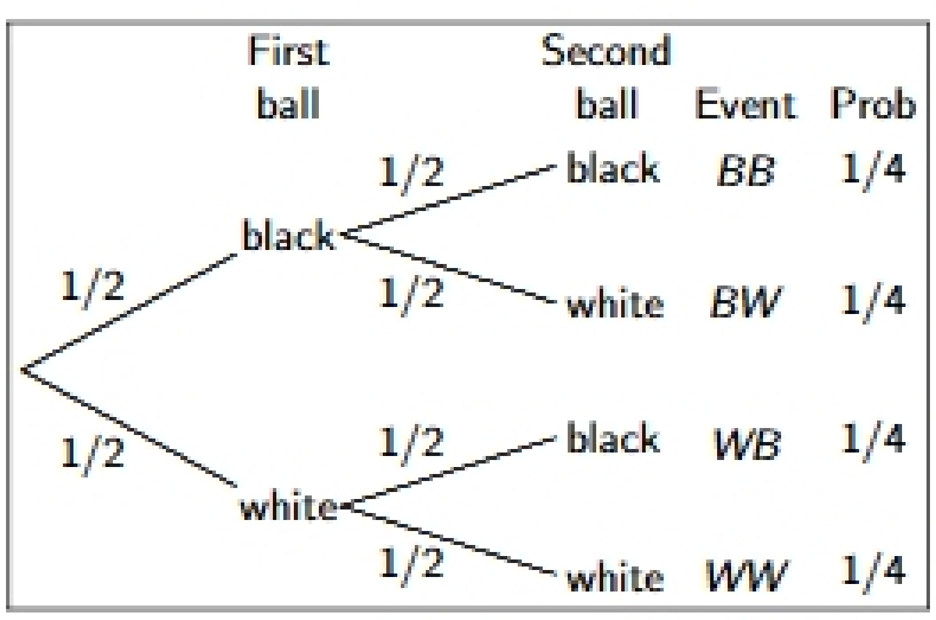
Product rule

The probability of A and B is

$$\Pr[A\ \&\ B] = Pr[B|A]\Pr[A]$$

This is why we multiply along the branches of a tree diagram.

Statistical independence: sampling w/ replacement



$$\Pr[W_2|B_1] = \Pr[W_2|W_1] = \Pr[W_2] = 1/2$$

Sampling with replacement: model versus data

Event	Theoretical probability	Observed relative frequency
BB	0.25	0.254
BW	0.25	0.255
WB	0.25	0.252
WW	0.25	0.239

Data from computer simulation

Theory and observation are not identical, but they are very close.