Analysing Uncertainty

David Spiegelhalter

Statistical Laboratory and
MRC Biostatistics Unit, Cambridge

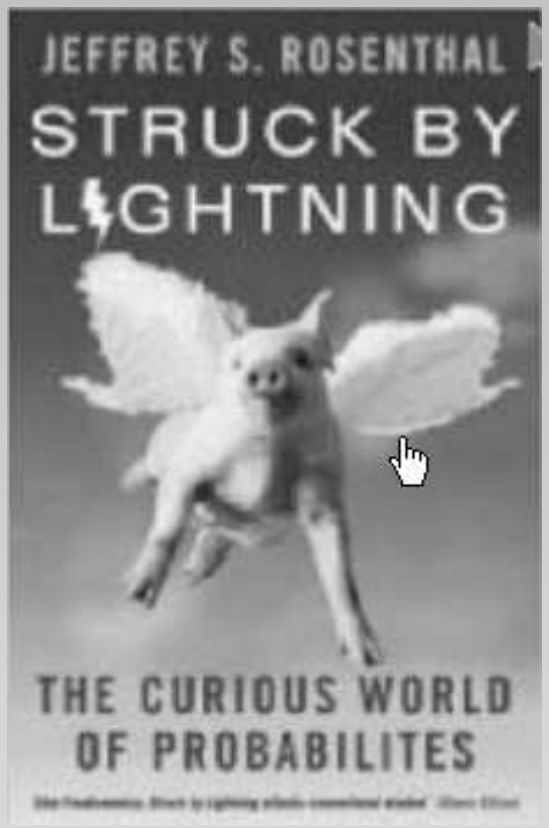
Fields Institute, May 27th 2009

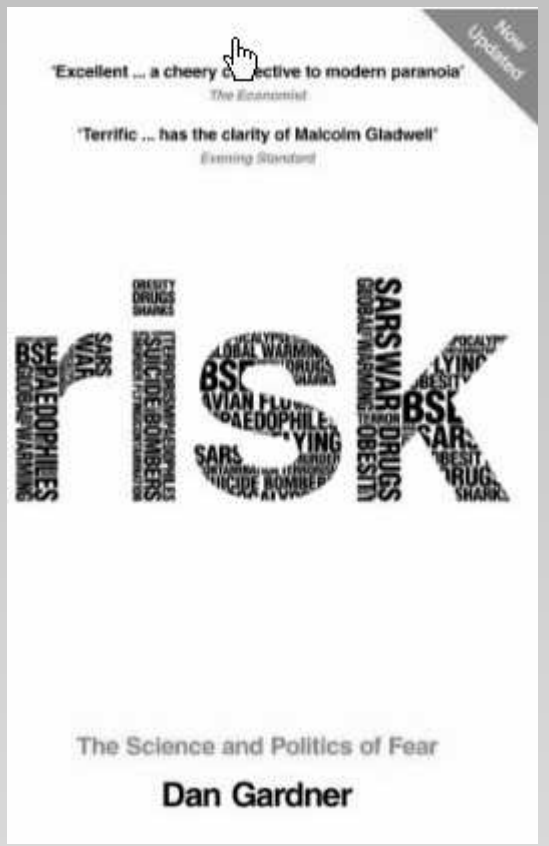
With thanks to Mike Pearson, Ian Short , Yin-Lam Ng, Hauke Riesch, Owen Smith, Arciris Garay, etc etc

What is the Winton programme trying to do?

Improve the public handling of quantitative aspects of risk and uncertainty, through

- Educational lectures, workshops
- The 'Risk Roadshow'
- Website
- Engagement with media
- Working with people who want to communicate risk
- Inter-disciplinary research


JEFFREY S. ROSENTHAL
STRUCK BY LIGHTNING
THE CURIOUS WORLD OF PROBABILITES


Now Updated
'Excellent ... a cheery ... ective to modern paranoia'
The Economist
'Terrific ... has the clarity of Malcolm Gladwell'
Evening Standard
RISK
The Science and Politics of Fear
Dan Gardner

www.understandingunceratinty.org

Search

Understanding Uncertainty

Home | Blog | For Educators | Fun Stuff | About Us

Contact us Log out

Home :: Featured Content

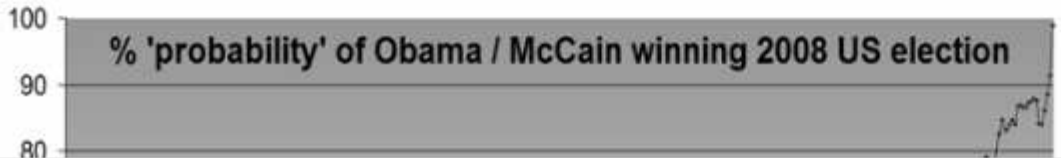
What was the probability that Barack Obama would win the US election?

View | Edit | Revisions | Workflow | Clone

Posted December 1st, 2008 by David Spiegelhalter in level 1, probability betting

On the face of it this seems an odd question. After all, he won. But before the election it was uncertain whether Obama would win, and probability is the way that uncertainty is quantified, so maybe it is reasonable to ask what that probability was.

We know that there were betting odds – a betting exchange such as Intrade allows people both to accept or make bets and so converges, at any point in time, to a certain set of odds at which people are willing to be either the better or the bookmaker. This prediction market provides a 'probability' on Obama winning that kept changing for the year before the election – this is shown in Figure 1 with some of the main events of the year marked in.



Featured Content

- One game to play!
- 2845 ways to spin the Risk
- A predictable pattern of murder?
- Nightingale's 'Coxcombs'
- What was the probability that Barack Obama would win the US election?
  - Laplace's law of succession
- Coincidences
- National Lottery
- Premier League
- What is Probability?
- Risk in the media
- How long are you going to live?

Pending Content

- Bayes' Theorem

Classify into uncertainty about :

1. Specific future events
2. Quantities in a model
3. The structure of the ‘best’ model
4. Inadequacies of our ‘best’ model

Classify into uncertainty about :

1. Specific future events

Let's start with a nice, 'known-risk' situation!

“6/49” lottery


LOTTO 649
27-MAY/MAI-2009
$2.00
QUICK PICK/MISE-ÉCLAIR
10 12 17 25 30 42
ENCORE ENCORE ENCORE
NOT ENTERED/ NON-INSCRIT
Ticket No. Billet
11153-1954-2022-0618-56406
008901
400100
SIGN/SIGNEZ
SIGNATURE
WEDNESDAY´S LOTTO 6/49
OLG
OLG
OLG
OLG


City CONVENIENCE

Statistics : Lotto 6/49

LOTTO 649

Numeric Order

Appearance of Winning Numbers by Frequency Order

Past six months: **52 draws** From 26/11/2008 to 23/05/2009			Since the start: **2,644 draws** From 12/06/1982 to 23/05/2009		
Numbers	**Total**	**Bonus**	**Numbers**	**Total**	**Bonus**
34	16	2	34	426	61
11	14	3	31	421	55
48	14	2	43	419	58
20	13	1	47	415	59
39	12	1	45	408	55
4	11	2	46	403	55
8	11	1	40	400	46
37	11	0	27	399	57

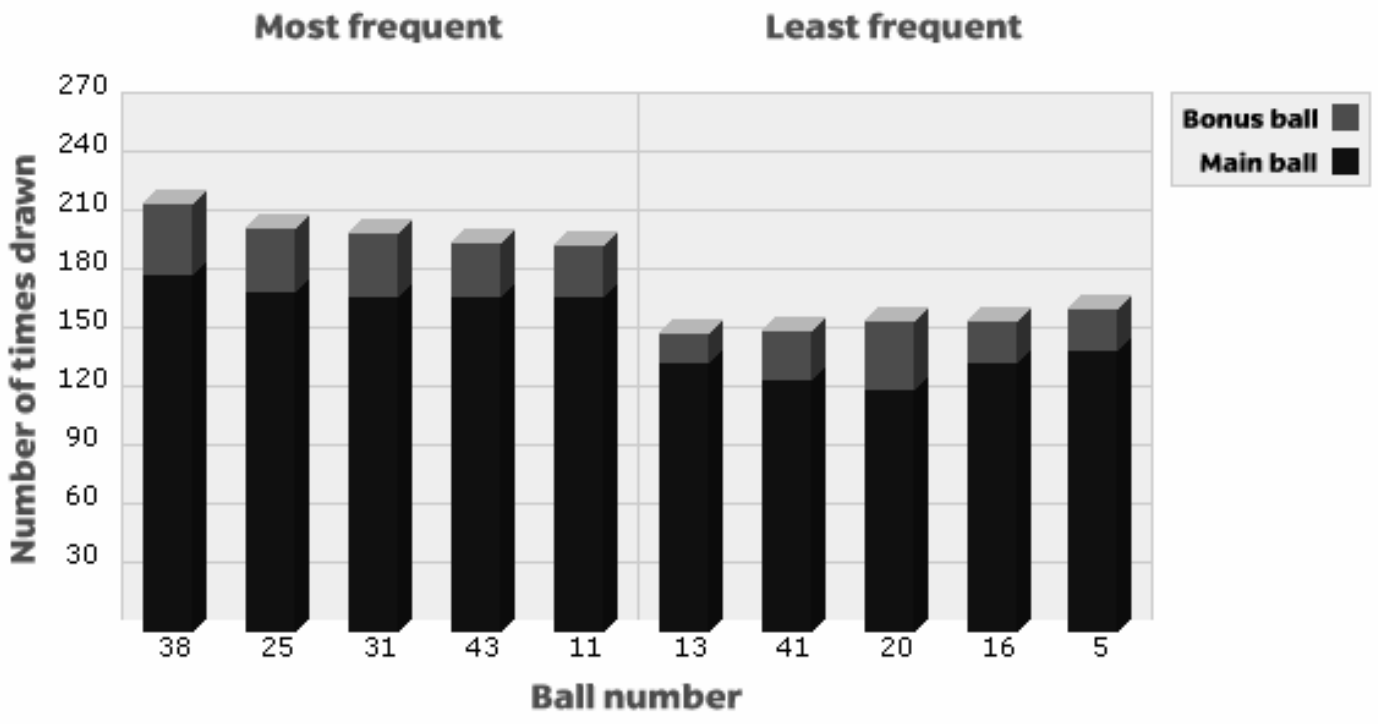
Find out which numbers have been drawn the most frequently, and which have been drawn the least. Despite the draws being totally random, some numbers have a habit of cropping up more than others, while others hardly appear at all! Please note, these results include the Lotto Bonus Draws held on 18th May 2002, 1st June 2002, 6th November 2004, and the £5 million jackpot-only draw held on 29th April 2006.

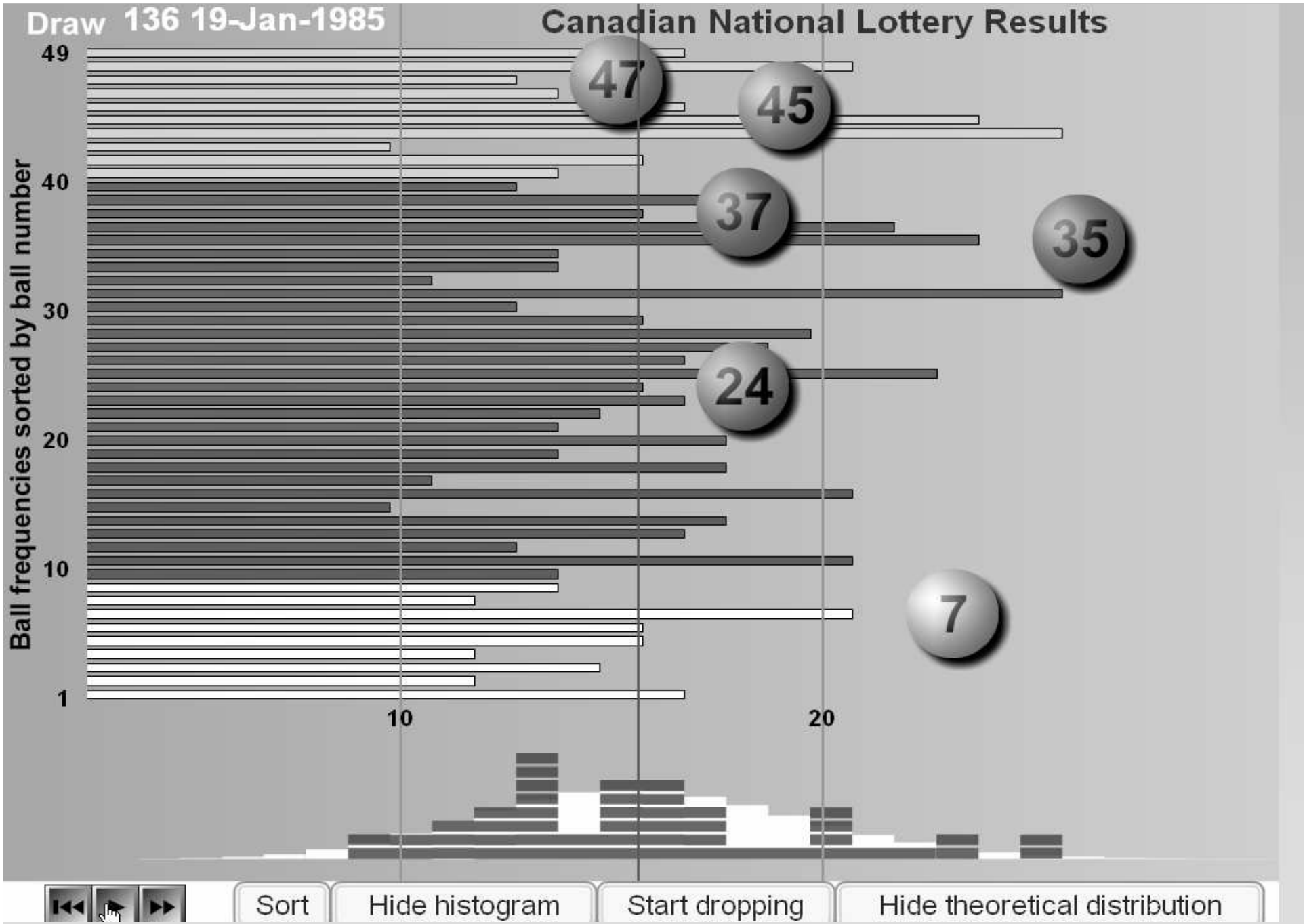
All numbers | **Individual number**

Game Lotto **Time period** To Date GO

lotto

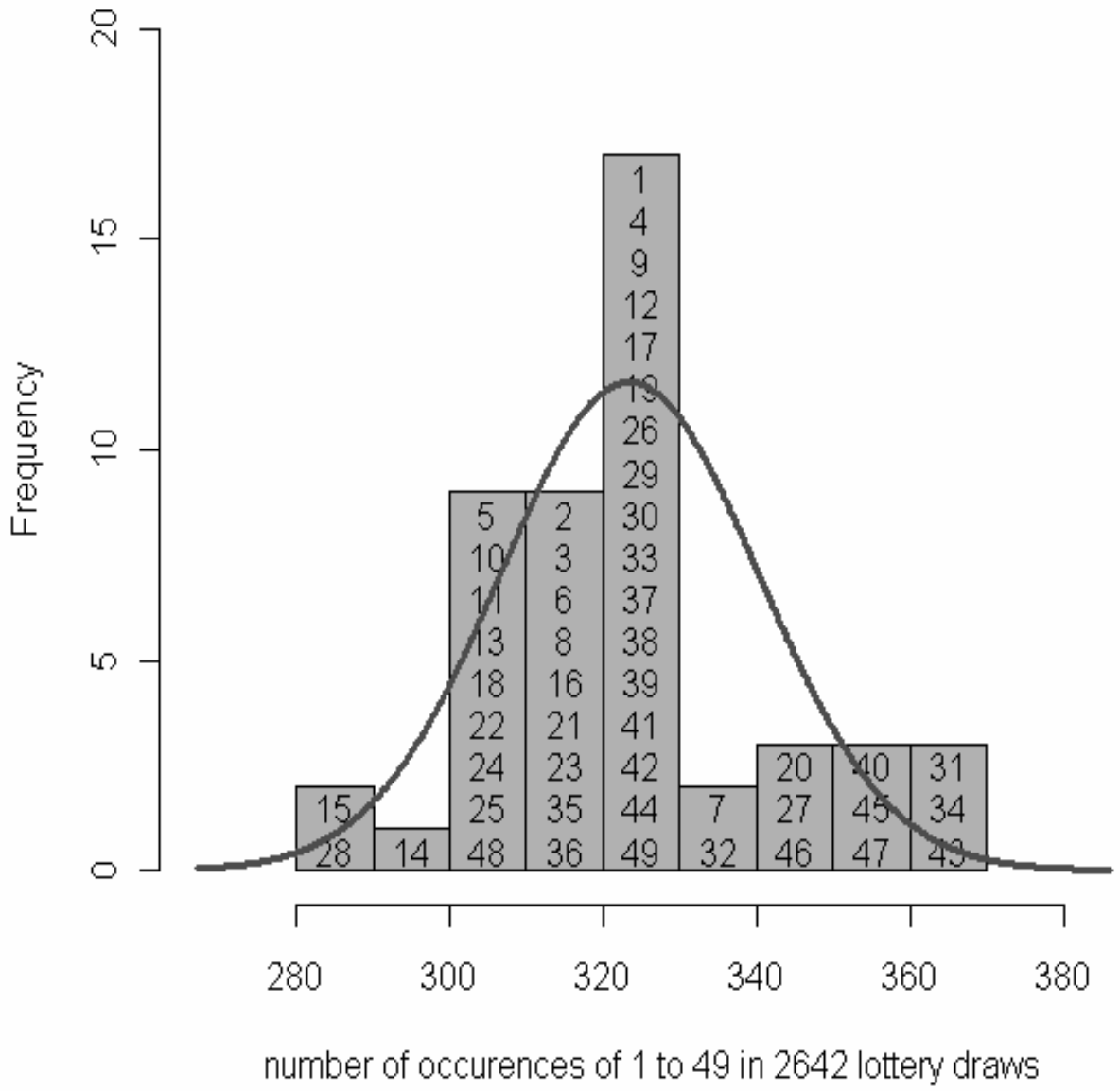
Five most and least frequently drawn balls



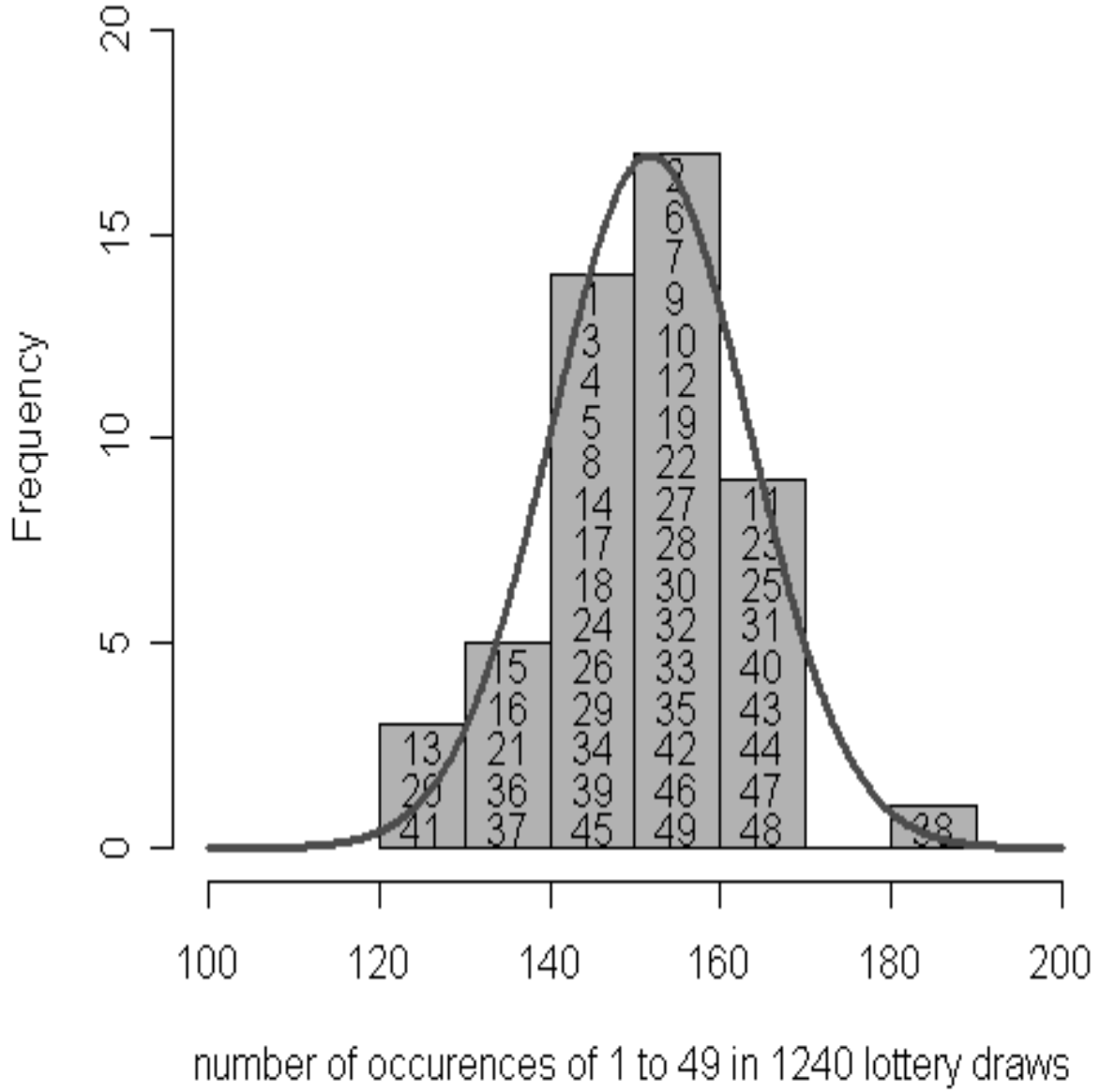


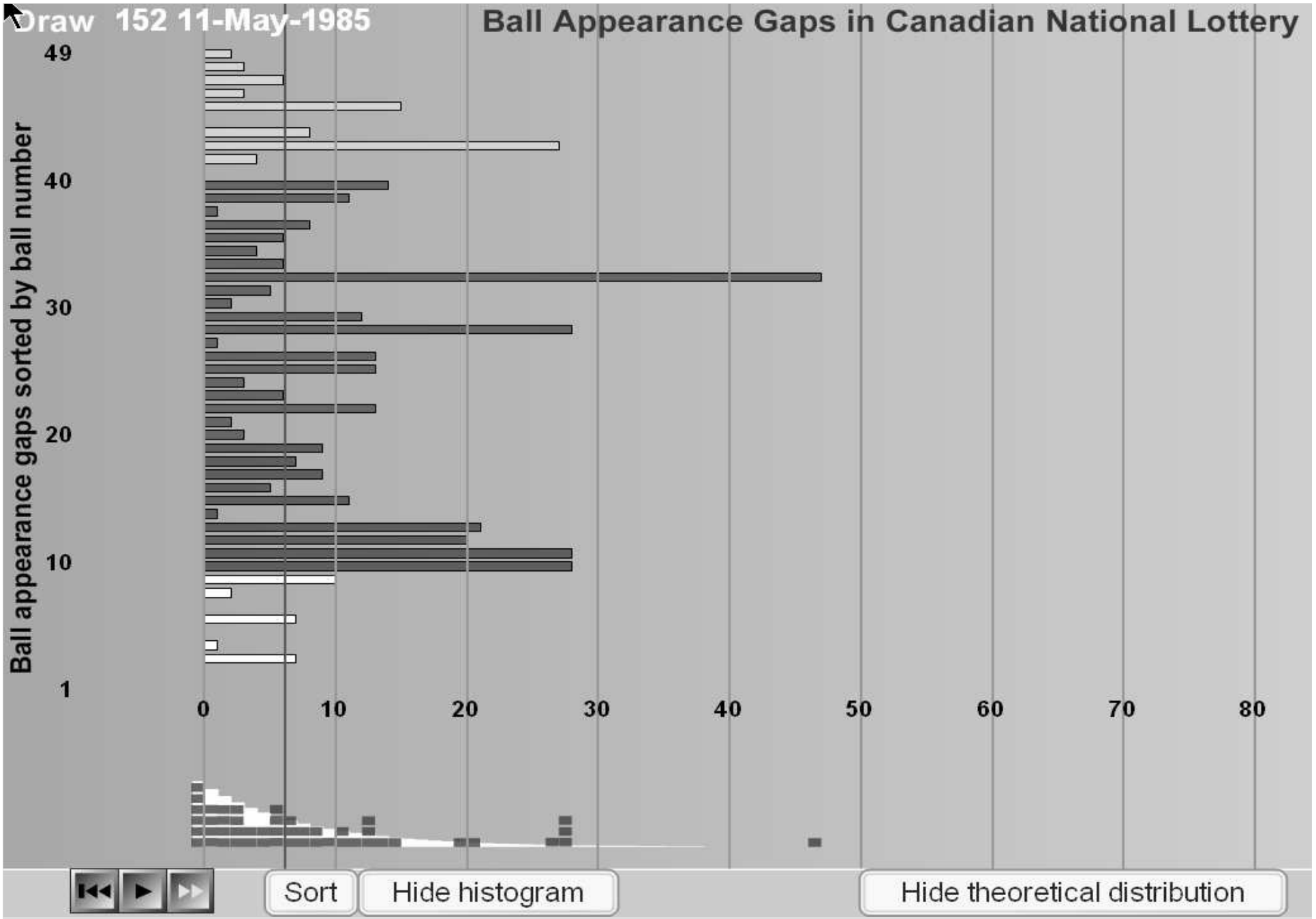
Lottery animation: www.understandinguncertainty.org/node/39

Counts in Canada 649, 1982-2009



Counts (in the UK!) obey the rules of probability

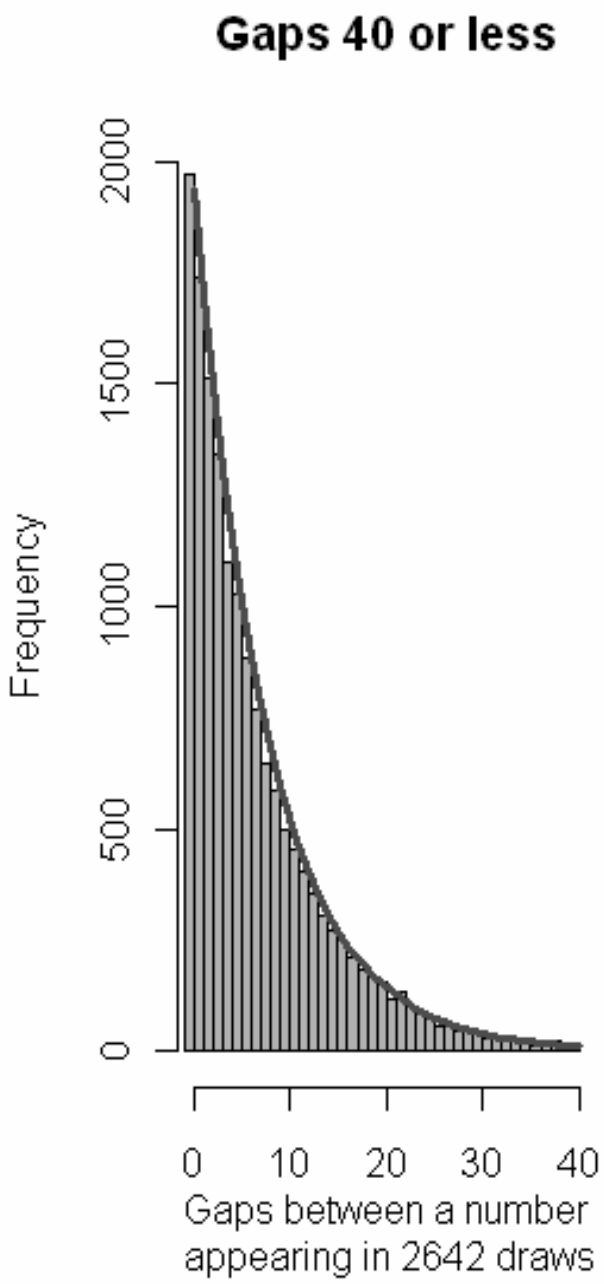


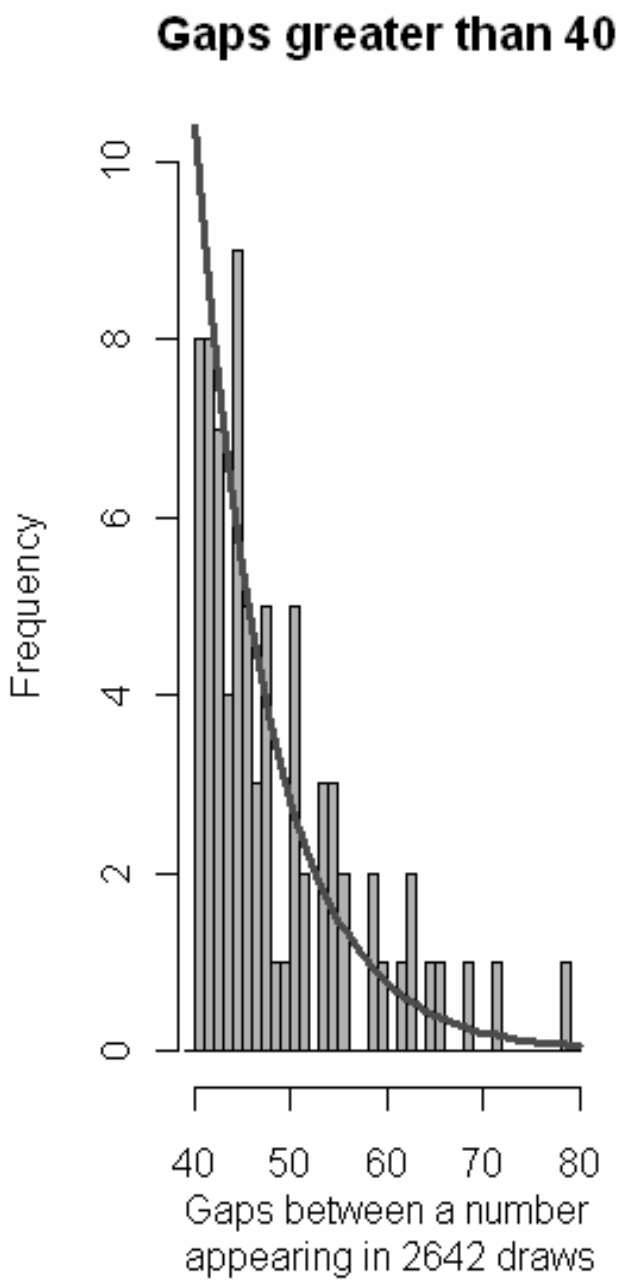


Lottery animation: www.understandinguncertainty.org/node/39

As expected, a geometric distribution of gaps

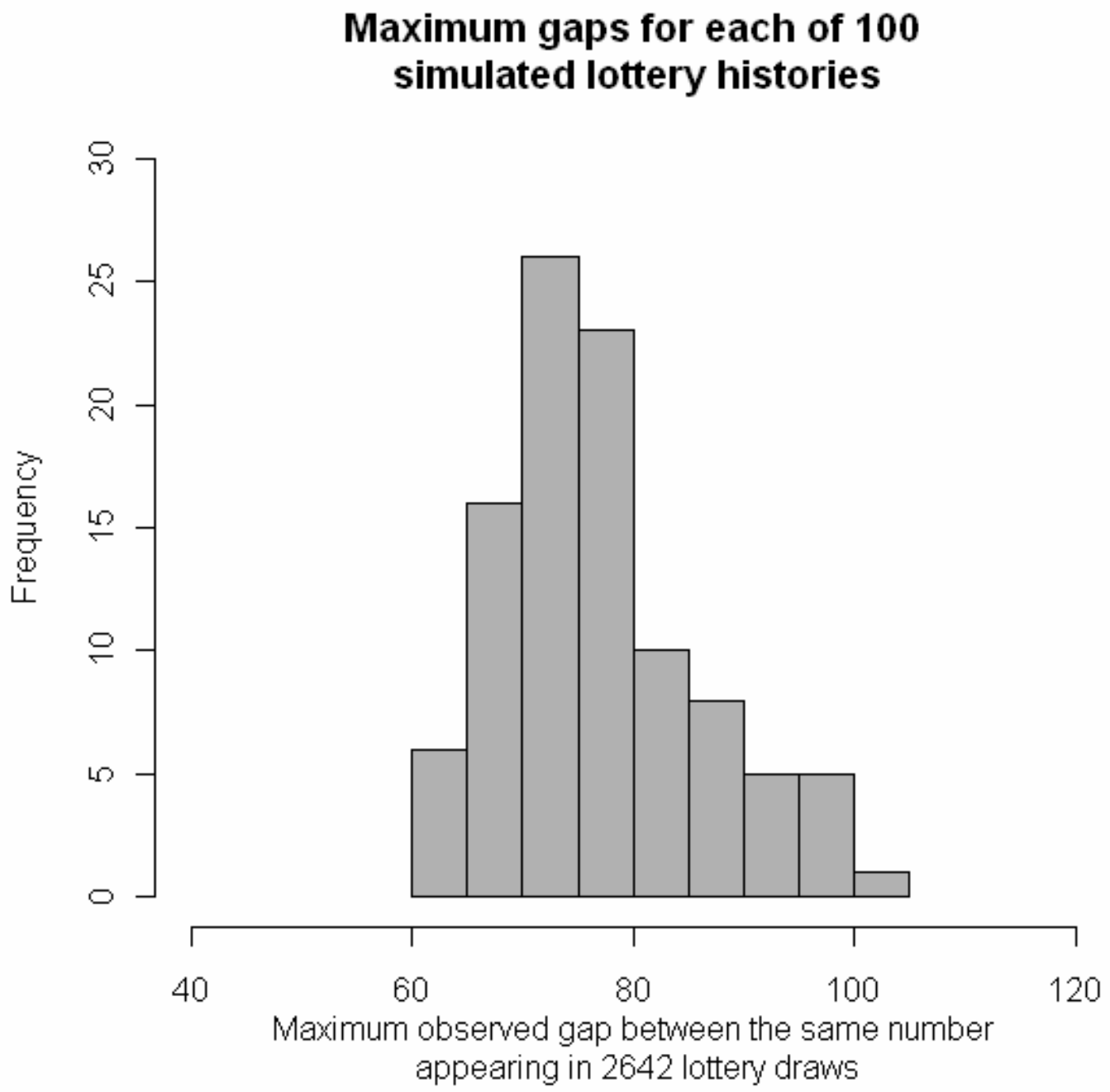
But is a maximum gap length of 79 surprising?





Simulate 100 full lottery histories

79 is almost exactly the expected maximum gap



Classify into uncertainty about:

1. Specific future events
2. **Quantities in a model**

- Missing data
- Fixed or random effects
- Parameters (including systematic biases)

Another classification of 'uncertainty'

Type A evaluation
method of evaluation of uncertainty by the **statistical analysis** of series of observations,

Type B evaluation
method of evaluation of uncertainty by means **other than the statistical analysis** of series of observations.

The NIST Reference on Constants, Units, and Uncertainty

Table D.2 – Uncertainty Budget: End-Gauge Calibration

Source of uncertainty	Standard uncertainties from random effects in the current measurement process (nm)		Standard uncertainties from systematic effects in the current measurement process (nm)	
	Type A evaluation	Type B evaluation	Type A evaluation	Type B evaluation
Calibration of standard end gauge				25
Measured difference between end gauges:				
repeated observations	5.8			
random effects of comparator			3.9	
systematic effects of comparator				6.7
Thermal expansion of standard end gauge				1.7
Temperature of test bed:				
mean temperature of bed	5.8			
cyclic variation of temperature of room				10.2
Difference in expansion coefficients of end gauges				2.9
Difference in temperatures of end gauges		16.6		
Combined standard uncertainty: $u_c(l)$ = 34 nm				

Using judgement to adjust for internal and external biases when doing meta-analysis for technology appraisals (Turner *et al*, *JRSSA*, 2009)

Internal biases

Selection bias
(eg. lack of randomisation)

Performance bias
(eg. lack of blinding)

Attrition bias
(e.g. loss to follow-up)

Detection bias
(eg. differences in outcome)

Other bias suspected
(eg. inconsistent results)

External biases

Population bias
(eg. differences in participants)

Control bias
(eg. different control strategies)

Intervention bias
(eg. differences in dose / timing)

Outcome bias
(eg. different outcome definitions)

Elicitation of ranges for internal biases

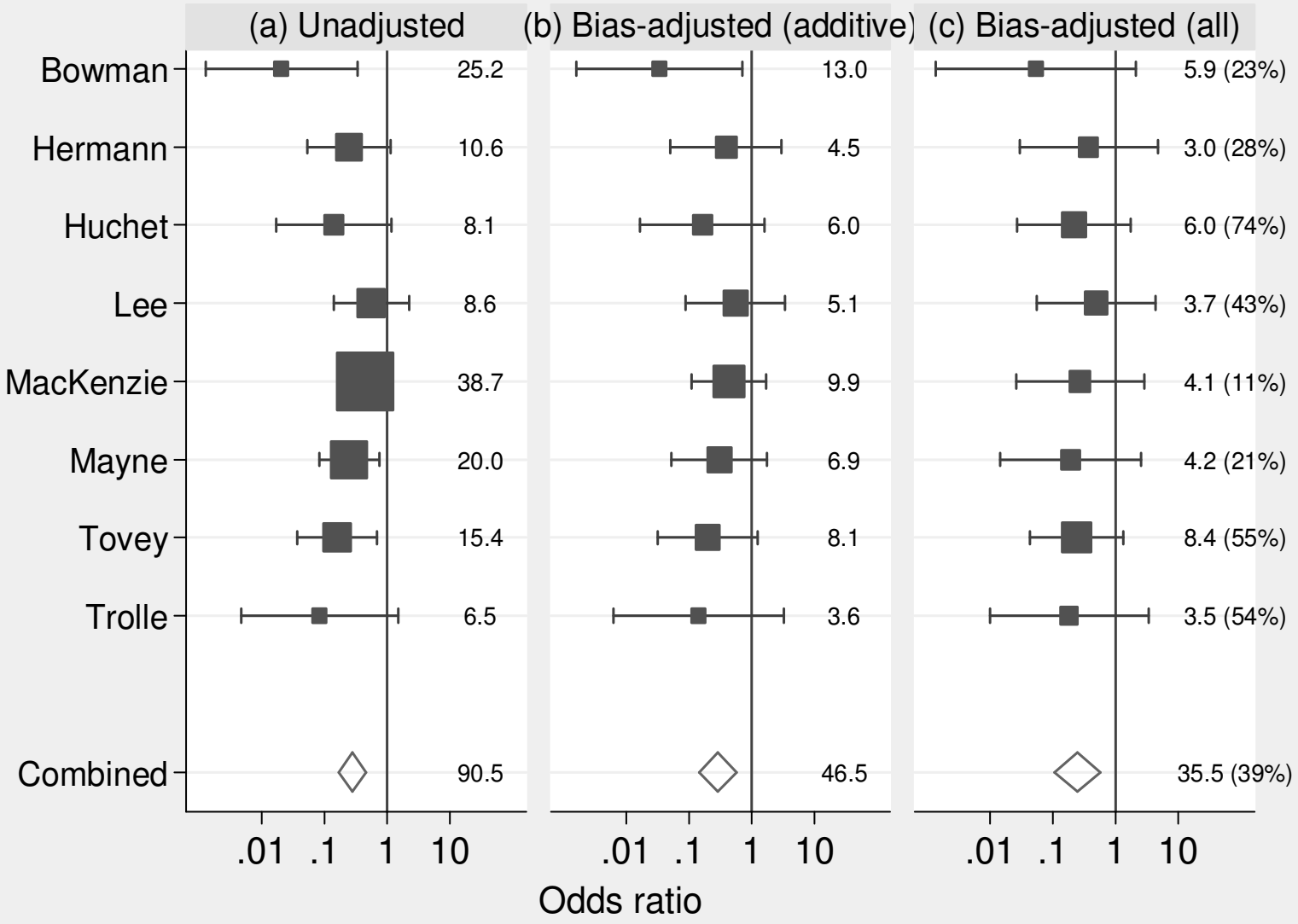
Question: Even if there were no intervention effect in this study, what apparent effect (ignoring sampling variation) might be induced by this bias?

- Each assessor considers the question for each bias
- Begins by making an informal judgement of the severity of the bias in each direction, as Low / Medium / High
- Next marks a 67% range on the (symmetric) relative risk scale

Risk lower in intervention group (or higher in control group)

Risk lower in control group (or higher in intervention group)

Adjusting for additive and proportional biases in all 8 studies: odds ratios, 95% CIs and 'effective number of events'



- How much of the English Premier football league is due to chance?

- 22%

TIMESONLINE

NEWS | COMMENT | BUSINESS | MONEY | SPORT | LIFE & STYLE | TRAVEL | DRIVING | AR

UK NEWS | WORLD NEWS | POLITICS | ENVIRONMENT | WEATHER | TECH & WEB | VIDEO | PHOTO GA

Where am I? › Home › Sport › Football › Premier League › **Newcastle**

From The Times

May 23, 2009

Alan Shearer believes Newcastle will stay up

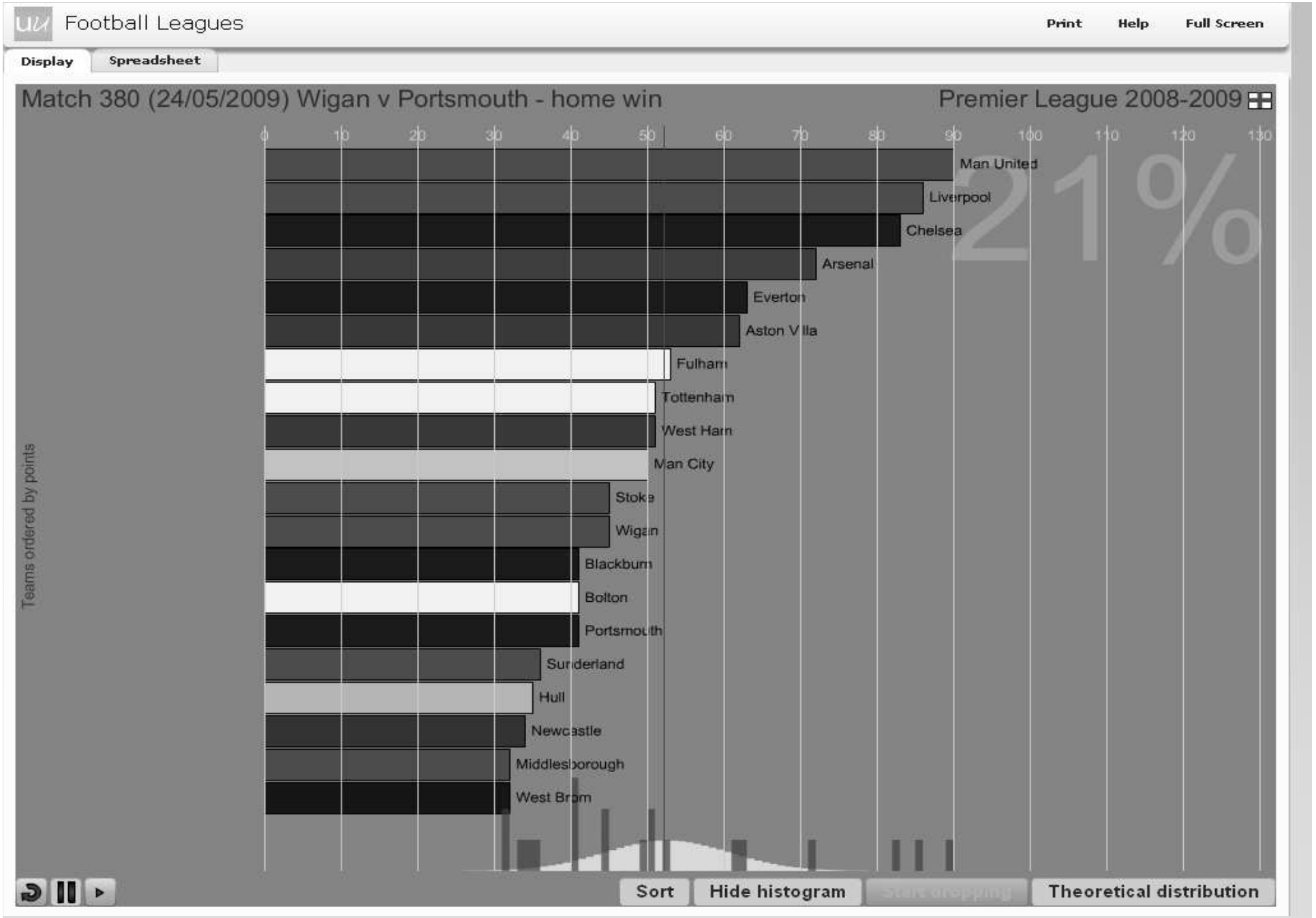
Manager expresses confidence that his club will escape relegation but warns of the implications of failure on Sunday

George Caulkin

Alan Shearer employed words such as "disaster" and "devastation" yesterday to describe the prospect of relegation, but while he spoke about the possibility of job losses at St James' Park, his is increasingly unlikely to be

CLUB DETAILS

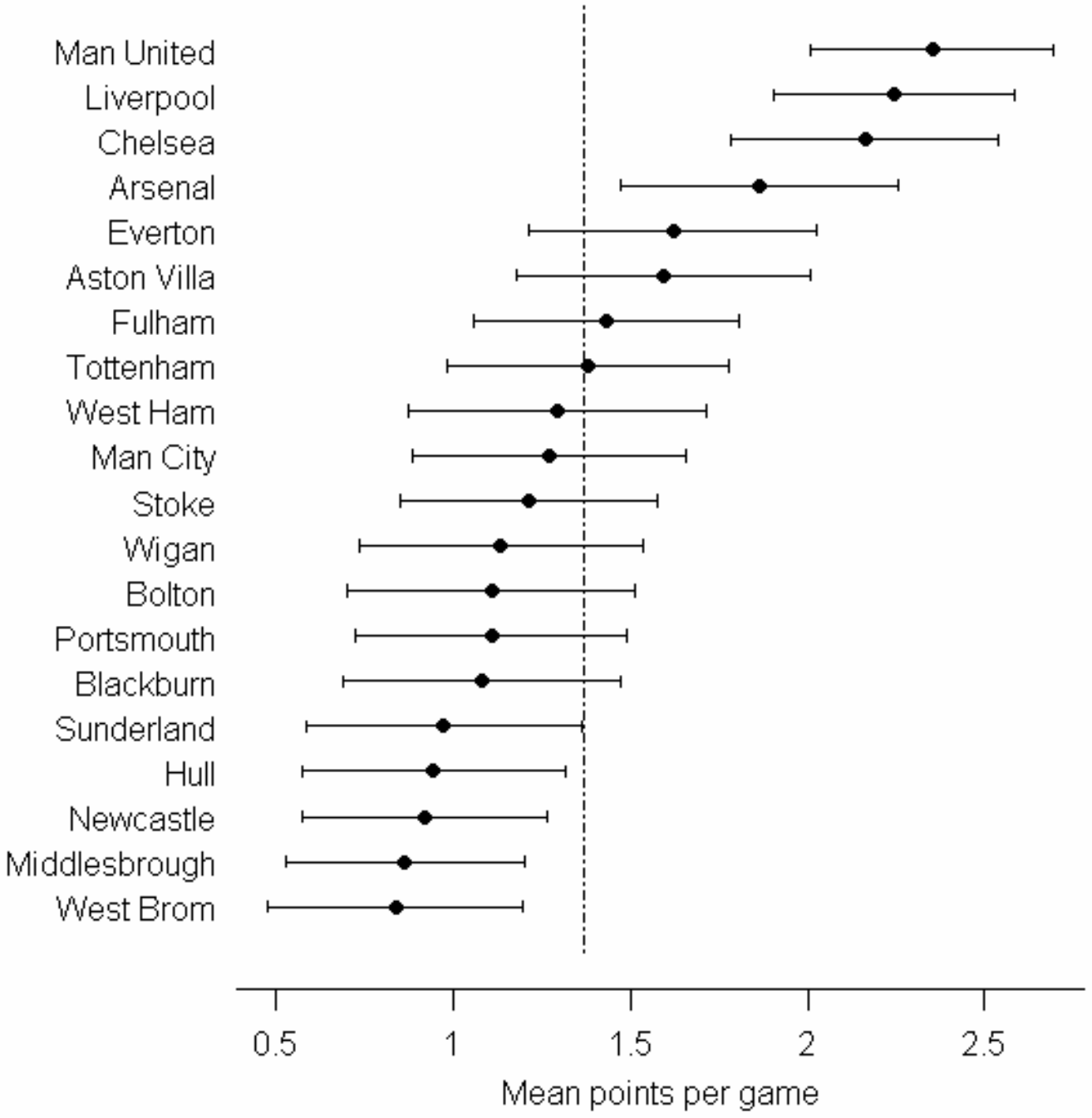
Ground:	St James' Park
Capacity:	52,387



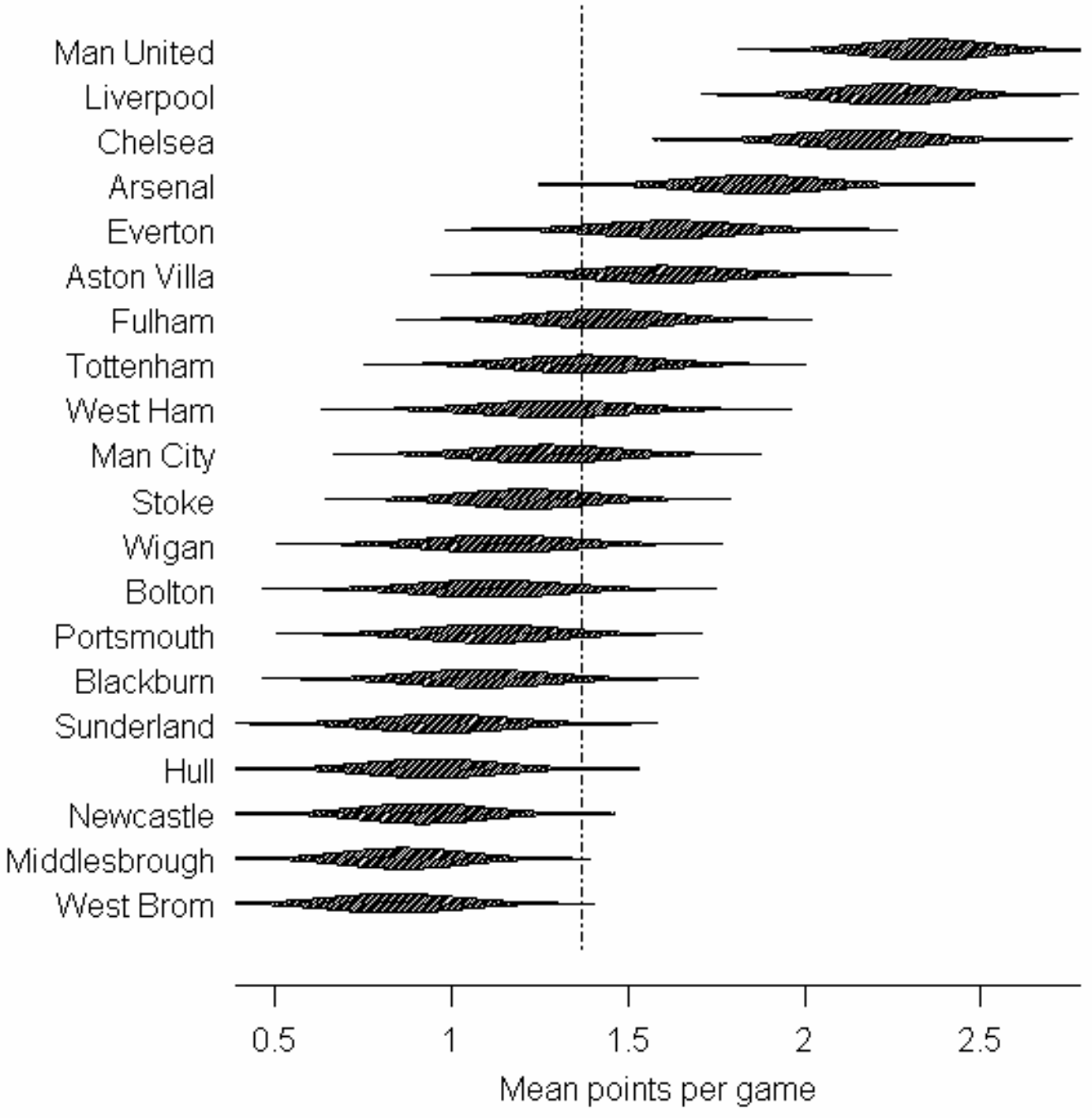
Football animation: www.understandinguncertainty.org/node/228

League with one match left to play (last Saturday)

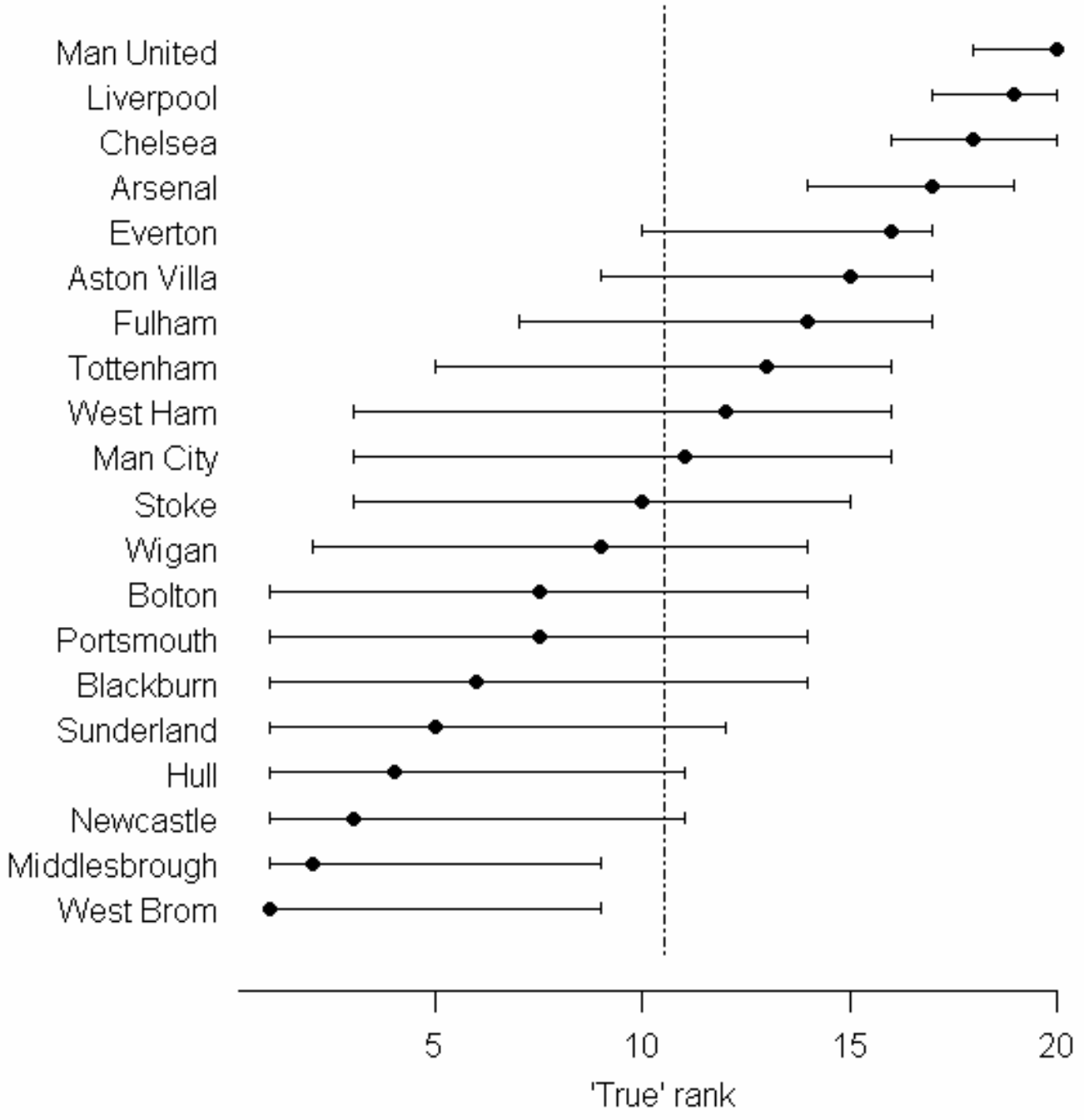
- Intervals based on standard errors derived from sample variance of points for each team, stratified by home/away



Avoiding 95% intervals



Simulate ‘true’ mean point score and rank



Only 56% chance that Man U really the ‘best’ team

Team	Points	Goals for	'Attack strength'	Goals against	'Defence weakness'
Man United	87	67	1.46	24	0.52
Liverpool	83	74	1.61	26	0.57
Chelsea	80	65	1.41	22	0.48
Arsenal	69	64	1.39	36	0.78
Everton	60	53	1.15	37	0.80
Aston Villa	59	53	1.15	48	1.04
Fulham	53	39	0.85	32	0.70
Tottenham	51	44	0.96	42	0.91
West Ham	48	40	0.87	44	0.96
Man City	47	57	1.24	50	1.09
Stoke	45	37	0.80	51	1.11
Wigan	42	33	0.72	45	0.98
Bolton	41	41	0.89	52	1.13
Portsmouth	41	38	0.83	56	1.22
Blackburn	40	40	0.87	60	1.30
Sunderland	36	32	0.70	51	1.11
Hull	35	39	0.85	63	1.37
Newcastle	34	40	0.87	58	1.26
Middlesbrough	32	27	0.59	55	1.20
West Brom	31	36	0.78	67	1.46

Predicting results using simple independent Poisson model

Hull City vs Manchester United: expected goals

Hull: = home-average x attack strength x defence weakness of opposition

=1.06 x 0.85 x 0.52 = 0.60

Man U: = 1.36 x 1.46 x 1.37 = 2.12

Team	Expected goals	0	1	2	3	4	5
Hull City	0.60	55	33	10	2	0	0
Man U	2.12	12	25	27	19	10	4

Assume independent Poisson distributions to give probability of any result

Add to give win/draw/lose probabilities

Home	Away	Most likely	2nd most likely	3rd most likely	4th most likely
Arsenal	Stoke	2-0 (14%)	1-0 (13%)	2-1 (9%)	3-0 (9%)
Aston Villa	Newcastle	1-0 (10%)	2-0 (10%)	2-1 (10%)	1-1 (10%)
Blackburn	West Brom	1-1 (10%)	2-0 (10%)	2-1 (10%)	1-1 (10%)
Fulham	Everton	0-0 (19%)	1-0 (16%)	0-1 (14%)	1-1 (13%)
Hull	Man United	0-2 (14%)	0-1 (14%)	1-2 (9%)	1-1 (8%)
Liverpool	Tottenham	1-0 (16%)	2-0 (15%)	3-0 (10%)	2-1 (9%)
Man City	Bolton	2-1 (10%)	1-1 (10%)	1-0 (10%)	2-0 (10%)
Sunderland	Chelsea	0-1 (20%)	0-2 (15%)	0-0 (13%)	1-2 (8%)
West Ham	Middlesbrough	1-0 (19%)	0-0 (14%)	2-0 (13%)	1-1 (11%)
Wigan	Portsmouth	1-0 (17%)	2-0 (14%)	0-0 (11%)	1-1 (10%)

- Actual model used is Bivariate Poisson, allowing correlations (R function lm.bp)
- Found to best fit European league results
- Many more sophisticated models used

Page last updated at 14:11 GMT, Friday, 22 May 2009 15:11 UK

The professor's Premiership probabilities

By Professor David Spiegelhalter
Professor for the Understanding of Risk, Cambridge University

Why should anyone take any notice of what a Cambridge professor of statistics, who knows little about football and does not even support [illegible] about this weekend's

Professor David Spiegelhalter analyses the football table

in the increasingly mathematical models used by sports betting t odds and identify bets.

So I would not recommend anyone using these odds for betting.

You have been warned.

Understanding Uncertainty: Animated Premier League Statistics

PREMIER LEAGUE PROBABILITIES

Read how the professor did

ARSENAL V STOKE

Home win: 72%

Draw: 19%

Away win: 10%

Verdict: 2-0 (14%)

ASTON VILLA V NEWCASTLE

Home win: 62%

Draw: 21%

Away win: 17%

Verdict: 1-0 (10%)

Two types of uncertainty

Aleatory

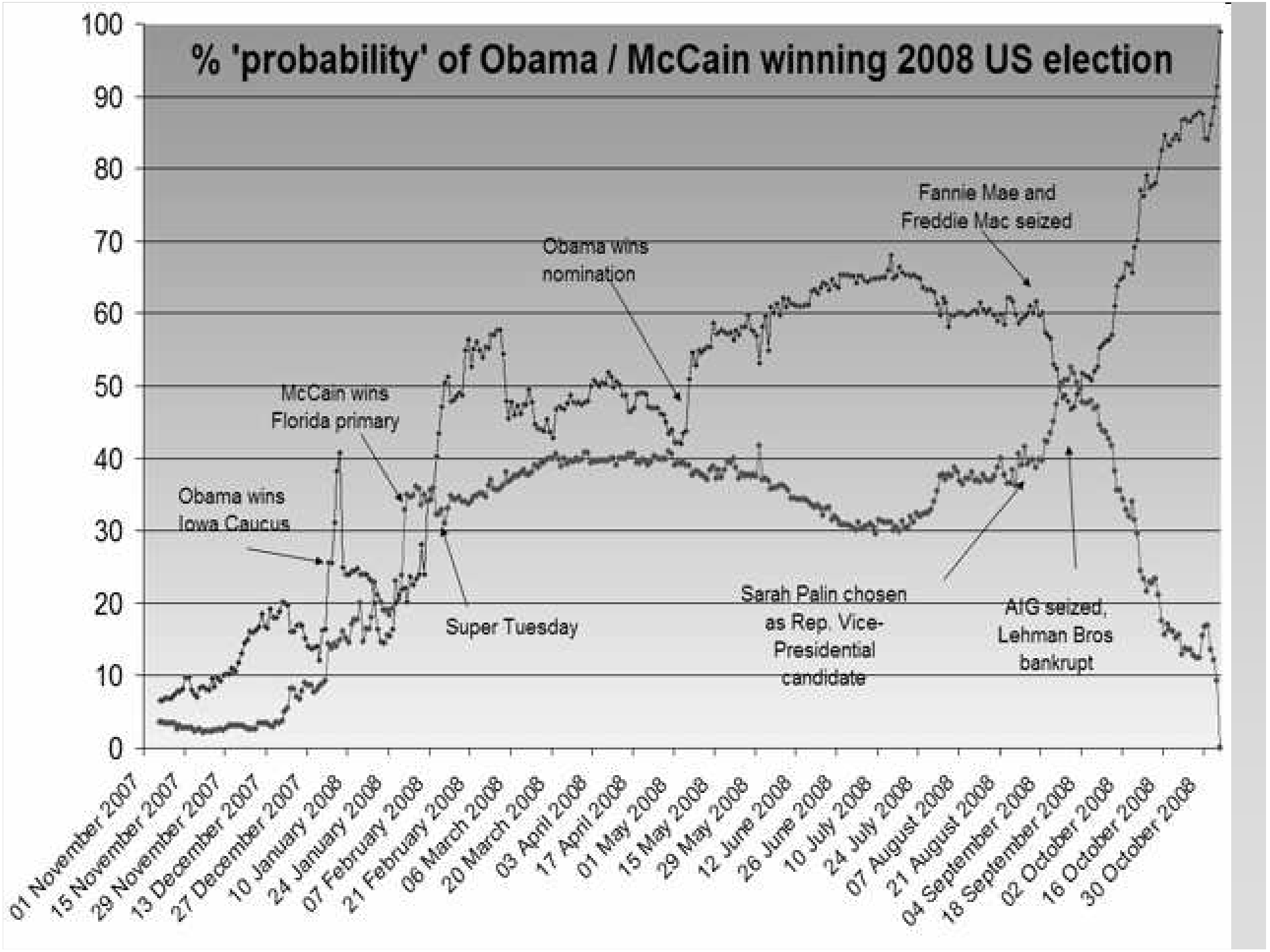
– chance, unpredictable

Epistemic

- lack of knowledge

‘Personalist’ approach to risk and uncertainty

- ‘*Probability does not exist*’ (objectively)
- Constructed based on current knowledge
- Is always contingent
- Assessed, not measured
- Best thought of as betting odds
- Allowed to use probability for epistemic uncertainty
- Calibrate by thinking of ‘chance’ situation that would be equally preferable to bet on


% 'probability' of Obama / McCain winning 2008 US election
100
90
80
70
60
50
40
30
20
10
0
Obama wins Iowa Caucus
McCain wins Florida primary
Super Tuesday
Obama wins nomination
Sarah Palin chosen as Rep. Vice-Presidential candidate
Fannie Mae and Freddie Mac seized
AIG seized, Lehman Bros bankrupt
01 November 2007
15 November 2007
29 November 2007
13 December 2007
27 December 2007
10 January 2008
24 January 2008
07 February 2008
21 February 2008
06 March 2008
20 March 2008
03 April 2008
17 April 2008
01 May 2008
15 May 2008
29 May 2008
12 June 2008
26 June 2008
10 July 2008
24 July 2008
07 August 2008
21 August 2008
04 September 2008
18 September 2008
02 October 2008
16 October 2008
30 October 2008

Classify into uncertainty about :

1. Specific future events
2. Quantities in a model
3. **The structure of the 'best' model**

- Which model selection criterion: AIC, BIC, DIC etc etc?
- Can we put weights on models?
- Does it make sense to talk of probabilities of models?

Bayes factors

$$\frac{p(M_0|y)}{p(M_1|y)} = \frac{p(M_0)}{p(M_1)} \frac{p(y|M_0)}{p(y|M_1)}$$

where

$$\frac{p(y|M_0)}{p(y|M_1)} = \frac{\int p(y|\theta_0)p(\theta_0)d\theta_0}{\int p(y|\theta_1)p(\theta_1)d\theta_0} = B_{01}$$

- Seems 'correct' thing to do
- Consistent – asymptotically finds 'true' model
- Can convert into model probabilities (if assume 'closed' world)
- Becoming popular in cosmology, psychology etc

BUT

- Very dependent on priors within models
- Do we believe there is a true model to find?

An alternative perspective

- Acknowledge that all models are wrong
- Admit interest is in predictions
- Led to Akaike-like procedures

Deviance Information Criterion:

$$p_D = E_{\theta|y}[-2\log p(y|\theta)] + 2\log p(y|\tilde{\theta}(y)).$$

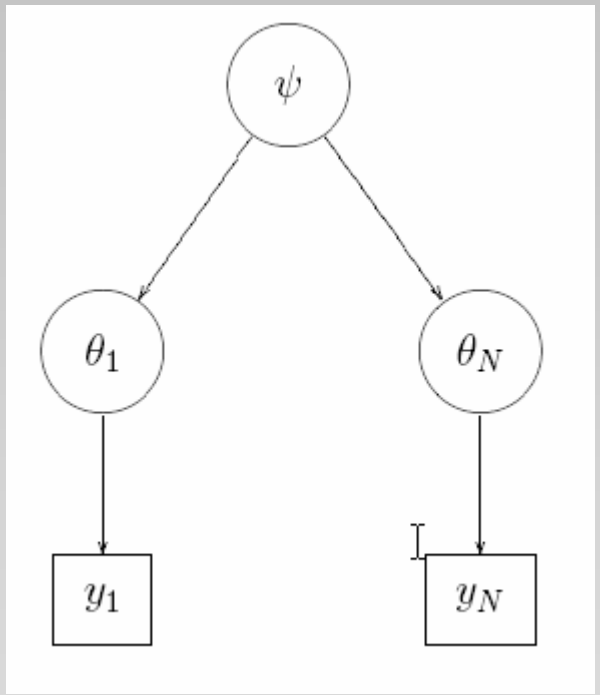
If we take $\tilde{\theta} = E[\theta|y]$, then

$p_D =$ "posterior mean deviance - deviance of posterior means"

Let $D(\theta) = -2\log p(y|\theta)$.

$$\begin{aligned} \text{DIC} &= D(\overline{\theta}) + 2p_D \\ &= \overline{D} + p_D \end{aligned}$$

Model choice in hierarchical models



- **DIC**

$$\text{DIC} = D(\overline{\theta}) + 2p_D$$

- **AIC**

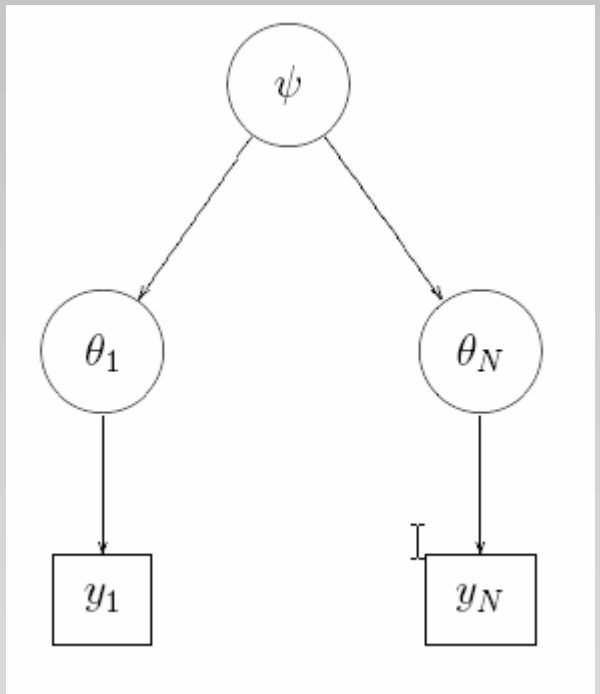
$$\text{AIC} = -2\log p(y|\widehat{\psi}) + 2p_\psi$$

where p_ψ is the number of hyperparameters

- **BIC**

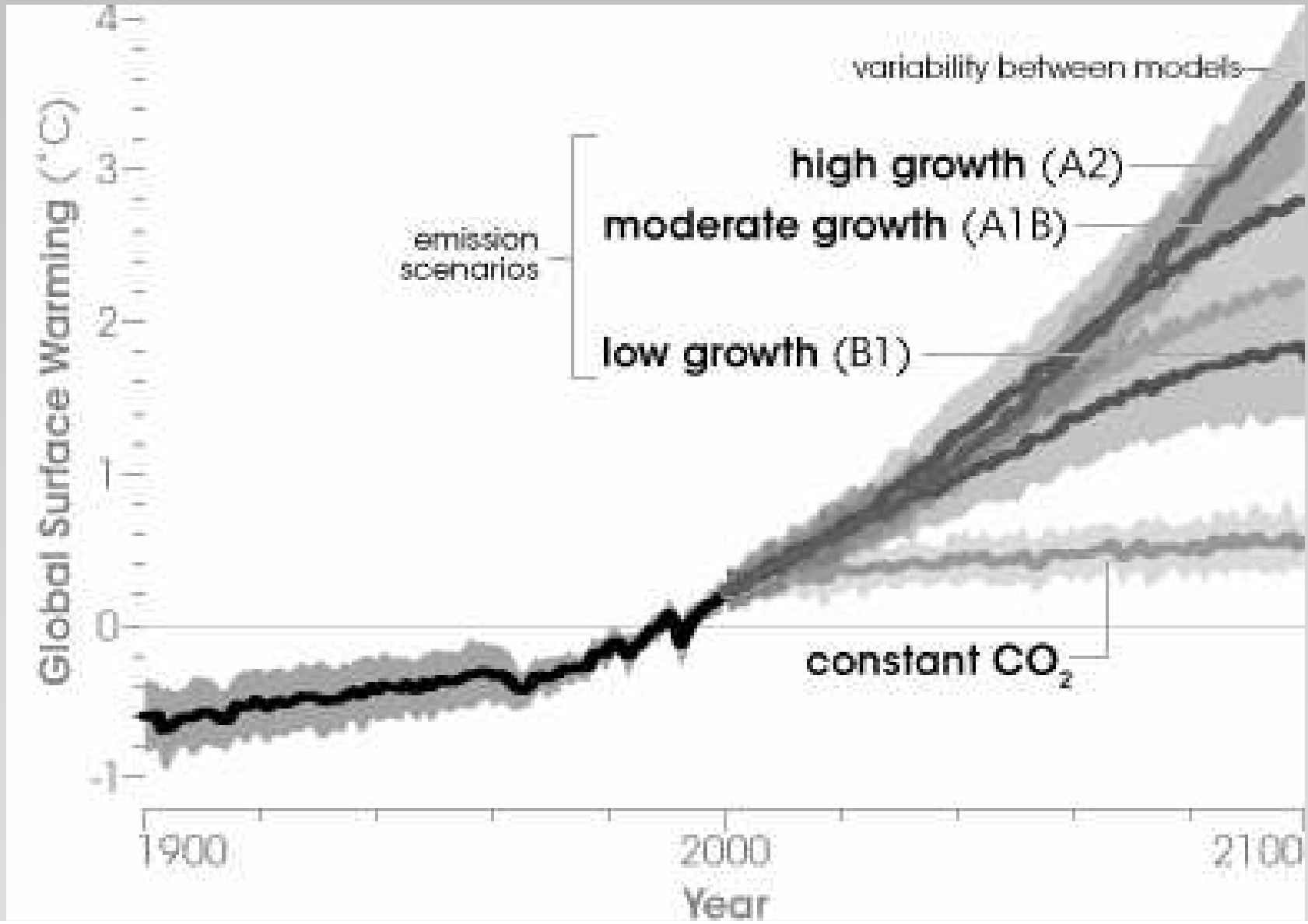
$$\text{BIC} = -2\log p(y|\widehat{\psi}) + p_\psi \log n$$

Criterion depends on prediction problem of interest



- Interested in predicting Y^{rep} *with* θ*'s fixed?*
  - **DIC:**
  - $\log p(Y^{\text{rep}}|\widehat{\theta})$ estimated by $\log p(y|\widehat{\theta})$,
  - penalised by $2p_D$
- Interested in predicting Y^{rep} *with* ψ *fixed?*
  - **AIC:**
  - integrate out θ's
  - $\log p(Y^{\text{rep}}|\widehat{\psi})$ estimated by $\log p(y|\widehat{\psi})$,
  - penalised by $2k$
- Interested in predicting Y marginally
  - **Bayes Factors:**
  - integrate out θ's and ψ
  - $\log p(Y^{\text{rep}})$ estimated by $\log p(y)$,
  - no penalty

IPCC projections



Classify into uncertainty about :

1. Specific future events
2. Quantities in a model
3. The structure of the ‘best’ model
4. **Inadequacies of our ‘best’ model**

- Recognised inadequacies
- Unrecognised inadequacies

‘Deep’ uncertainty

- Can we make any quantitative analysis of such uncertainties?
- Do we just have to admit ‘we don’t know’?
- Then make decisions that are resilient to mistakes
- ‘Precautionary principle’

Frank Knight (1885-1972)

Risk, Uncertainty, and Profit (1921)

- *The essential fact is that 'risk' means in some cases a quantity susceptible of measurement, while at other times it is something distinctly not of this character; and there are far-reaching and crucial differences in the bearings of the phenomena depending on which of the two is really present and operating....*
- *It will appear that a measurable uncertainty, or 'risk' proper, as we shall use the term, is so far different from an unmeasurable one that it is not in effect an uncertainty at all*

Memorable quote #325

"But there are also unknown unknowns. There are things we do not know we don't know"

"Cromwell's Law"

OLIVER CROMWELL'S

LETTERS AND SPEECHES:

WITH ELUCIDATIONS.

BY

THOMAS CARLYLE.

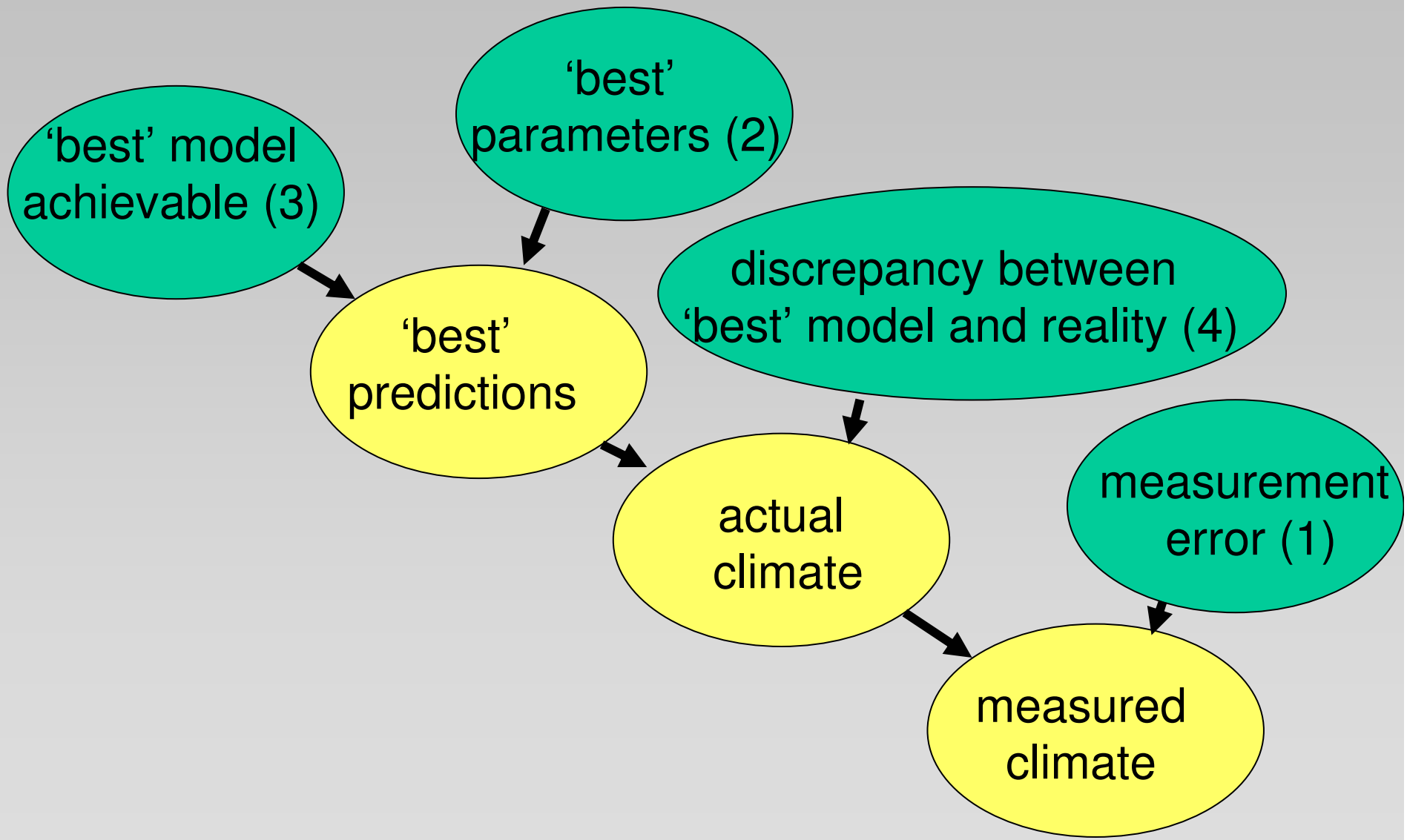
To the General Assembly of the Kirk of Scotland; or, in case of their not sitting, To the Commissioners of the Kirk of Scotland: These.

SIRS, Musselburgh, 3d August 1650.

Is it therefore infallibly agreeable to the Word of God, all that *you* say? I beseech you, in the bowels of Christ, think it possible you may be mistaken. Precept may be upon

picking our horses' beans, eating our soldiers' leavings: 'they are much enslaved to their Lords,' poor creatures; almost destitute of private capital,—and ignorant of soap to a terrible extent![25] Cromwell distributes among them 'pease and wheat

Structure of Goldstein+Rougier



UK Climate Impact Programme approach to uncertainty:

Producing pdfs for future climate, and attempting to classify uncertainty into

- Inevitable – chaotic models
- Parameter
- Structural
- (a) Known inadequacies due to scaling etc
  (b) Unknown inadequacies (use model disagreement as proxy)

UK Climate Impact Programme website

*

explicitly acknowledges role of judgment

What are probabilistic climate projections?

Before using the UKCIP08 probabilistic climate projections, it is important to understand what they are and what they are not. A probabilistic climate projection:

- IS NOT an objective probability, where a situation is well understood, where all outcomes can be accounted for or where probabilities can be revised based on observed outcomes (such as tossing a coin or rolling a dice);
- IS rather a subjective probability, providing an estimate based on the available information and strength of evidence (similar to horse-racing odds);
- encapsulates some, but not all, of the uncertainty associated with projecting future climate;
- is dependent on the method used, including assumptions and choices made, meaning that a different method would produce different results;
- is based on the current evidence (i.e. models and observations), and new evidence in the future may lead to the results being modified;

HOW DID DR SPIEGELHALTER DO?

(All fixtures were played on Sunday 24th May at 1600 BST)

SUNDAY

Arsenal 4-1 Stoke

The professor's probable score: 2-0 (14%)

Match report

Aston Villa 1-0 Newcastle

The professor's probable score: 1-0 (10%)

Match report

Blackburn 0-0 West Brom

The professor's probable score: 1-1 (10%)

Match report

Statistics:

9/10 win/draw/lose, 2 exact scores

BBC expert Mark Lawrenson:

7/10 win/draw/lose, 1 exact

Vague ‘conclusions’

- Statisticians tend to have (or at least are taught) a rather narrow view of uncertainty
- There are many communities out there with opinions about risk and uncertainty, and they are not all stupid
- Different groups approach the hierarchy of uncertainty from opposite ends
- But robust use of quantitative methods, with due humility, is of huge value

Structural uncertainty

The Treatment of Uncertainties in the Fourth IPCC Assessment Report

Structural uncertainty	Inadequate models, incomplete or competing conceptual frameworks, lack of agreement on model structure, ambiguous system boundaries or definitions, significant processes or relationships wrongly specified or not considered	Specify assumptions and system definitions clearly, compare models with observations for a range of conditions, assess maturity of the underlying science and degree to which understanding is based on fundamental concepts tested in other areas

Table 3 Quantitatively calibrated levels of confidence

Terminology	Degree of confidence in being correct
Very High confidence	At least 9 out of 10 chance of being correct
High confidence	About 8 out of 10 chance
Medium confidence	About 5 out of 10 chance
Low confidence	About 2 out of 10 chance
Very low confidence	Less than 1 out of 10 chance