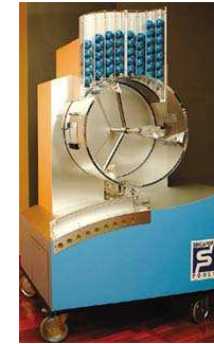
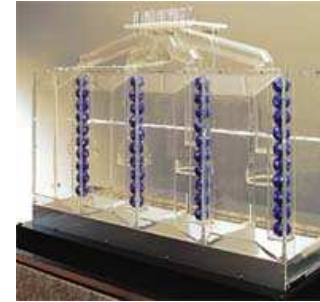


Having Fun with Lottery Data

Berwin A Turlach
 berwin@maths.uwa.edu.au
 School of Mathematics and Statistics (M019)
 The University of Western Australia



Hanley, J.A. (1984). Lotteries and Probabilities: Three Case Reports, *Teaching Statistics* **6**(3): 88–92.
 Hanley, J.A. (1992). Jumping to Coincidences: Defying Odds in the Realm of the Preposterous, *The American Statistician* **46**(3): 197–202.
 Paulson, R.A. (1992). Using Lottery Games to Illustrate Statistical Concepts and Abuses, *The American Statistician* **46**(3): 202–204.

Stories from the UK National Lottery

- 6/49 Lottery with bonus number.
- 45p for every £1 spend is returned on average.
- Match 3 winners receive a fixed amount of £ 10; the remaining price money is divided among the other winners (Match 4, Match 5, Bonus and Match 6).
- In Draw 366, 46 tickets shared the jackpot pool and only 13 shared the Bonus pool; the prize for a Bonus-winning ticket did exceed that for the jackpot.
- Sole jackpot winner in Draw 4 scooped the entire £ 17 million, while in Draw 9 the jackpot pool of about £ 16 million was shared by 133 winners; each receiving a mere £ 122 510.
- Tipping 1, 2, 3, 4, 5 and 6 is extremely popular (10 000+ per week).
- The biggest potential win was a jackpot of £ 42 million (shared by 3, but in that week 30 000 entries tipped the numbers above).

Haigh, J. (2003). *Taking Chances: Winning with Probability (2nd ed)*, Oxford University Press.

Astonishing events

- During a consecutive sequence of 37 draws, some *triple* had appeared in the winning combination eight times.
 - The number 44 was drawn eight times in the ten draws from Draw 86 to Draw 95.
- Haigh, J. (2003). *Taking Chances: Winning with Probability (2nd ed)*, Oxford University Press.
- The sets of five numbers picked North Carolina Cash-5 Lottery game were identical on July 9th and 11th, 2007.
- Stefanski, L.A. (2008). The North Carolina Lottery Coincidence, *The American Statistician* **62**(2): 130–134.
- In Draw 3016 (Wednesday, June 21st, 1995) of the German lotto (6/49 Lotto) the winning numbers were identical to those from the draw on Saturday, December 20th, 1986.
- Henze, N. (1997) *Stochastik für Einsteiger*, Vieweg Verlag.

Astonishing events: multiple winner

- Evelyn Adams won in the mid-1980s the New Jersey State Lottery twice within four months (raking in a total of \$5.4 million).
- Maureen Wilcox bought tickets in June 1980 for both the Massachusetts Lottery and the Rhode Island Lottery. She managed to pick the winning numbers for both lotteries but did not win a dime—her Massachusetts numbers won the Rhode Island Lottery, and her Rhode Island numbers won the Massachusetts Lottery.

Olofsson, P. (2007). *Probabilities: The Little Numbers That Rule Our Lives*, Wiley & Sons.

SSAI WA Branch Meeting, 18 August 2009

5 / 27

Playing Lotto is an expected loss?

In 1992 some investors in Melbourne, Australia, noticed:

- The Virginia Lottery, being a 6/44 lottery, has a chance of 1 in 7 059 052 of winning.
- The jackpot had grown to \$27 million; with second, third and fourth prizes included an available pot of \$27 918 561.
- Buy a ticket for each of the combinations and the value of those tickets would equal the value of the pot; about \$3.95 per ticket.
- But the tickets were sold for \$1!
- They quickly found 2 500 small investors willing to put up an average of \$3 000 each; if the scheme worked, the yield on that investment would be about \$10 800.

Mlodinow, L. (2009). *The Drunkard's Walk: How Randomness Rules Our Lives*, Vintage.

SSAI WA Branch Meeting, 18 August 2009

6 / 27

Playing Lotto is an expected loss? (cont.)

- They filled out 1.4 million slips by hand.
- Placed groups of buyers at 125 retail outlets and obtained cooperation from grocery stores.
- The scheme got going just 72 hours before the deadline.
- Grocery-store employees worked in shifts to sell as many tickets as possible.
- One store sold 75 000 tickets in the last 48 hours.
- A chain store accepted bank cheques for 2.4 million tickets, assigned the work of printing the tickets among its stores and hired couriers to gather them.
- In the end, they had purchased just 5 million of the 7 059 052 tickets.
- It took them several days to find the winning ticket.

Mlodinow, L. (2009). *The Drunkard's Walk: How Randomness Rules Our Lives*, Vintage.

SSAI WA Branch Meeting, 18 August 2009

7 / 27

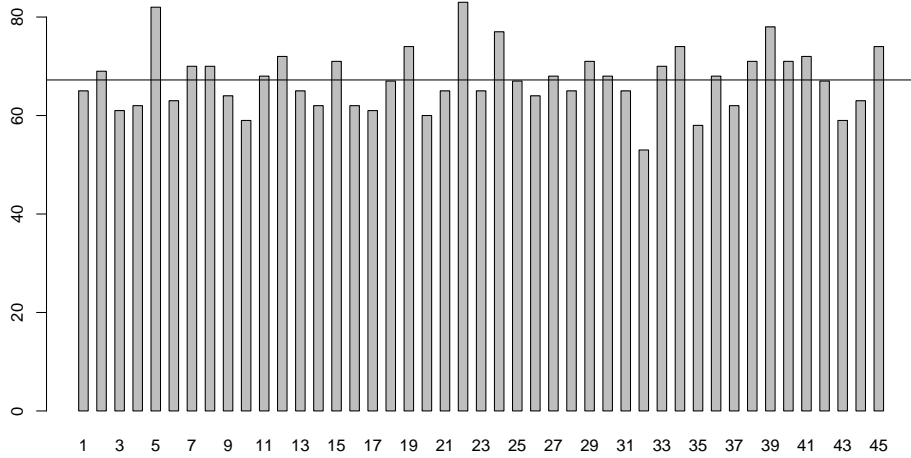
How It All Started

POWERBALL		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45
w #		65	69	61	62	82	63	70	70	64	59	68	72	65	62	71	62	61	67	74	60	65	83	65	77	67	64	68	65	71	68	65	53	70	74	58	68	62	71	78	71	72	67	59	63	74
22/12/07		12	3	1	9	7	9	8	11	18	5	1	13	12	3	6	1	11	0	16	1	28	0	8	7	1	21	5	0	10	3	8	22	0	7	72	2	0	6	3	8	5	13	33	2	6
		14	14	14	19	11	11	17	11	8	12	12	11	8	14	18	13	15	16	19	10	9	13	11	18	8	18	12	17	18	11	15	18	17	13	10	11	10	15	17	13	7	12	11	16	18
		6	3	2	38	24	170	34	5	68	77	65	114	13	0	52	41	45	186	37	11	10	15	9	29	80	44	56	50	4	54	58	36	39	53	8	16	1	46	7	25	86	28	30	20	62

SSAI WA Branch Meeting, 18 August 2009

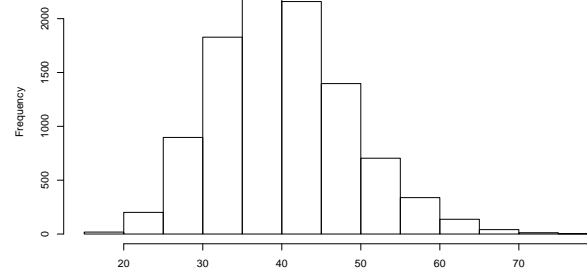
8 / 27

Barrel A: number of times drawn



What are the appropriate degrees of freedom?

Histogram of 10,000 simulated chi-squared statistics



Min:	15.20	Mean:	40.09
Q1:	33.94	Var:	72.81
Median:	39.52		
Q3:	45.36		
Max:	78.78		

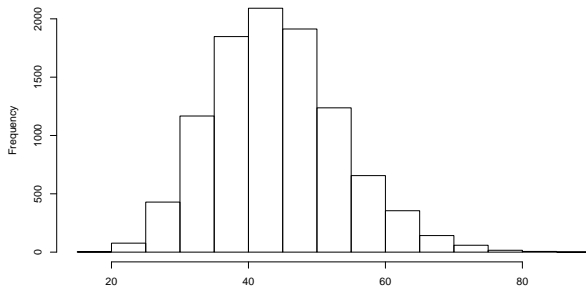
Does not look as if degrees of freedom is 44.

Powerball data I: χ^2 -test (cont.)

44 is correct degrees of freedom; however, for a k/N lottery use a scaled χ^2 -statistics, namely:

$$\frac{N-1}{N-k} \sum_{i=1}^N \frac{(O_i - E_i)^2}{E_i}$$

Histogram of 10,000 simulated scaled chi-squared statistics

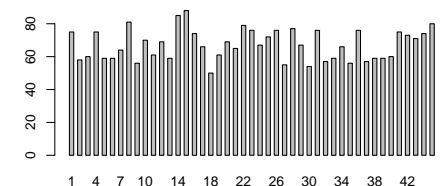
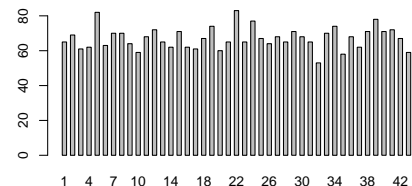
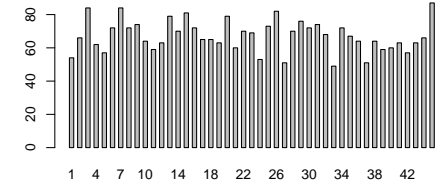
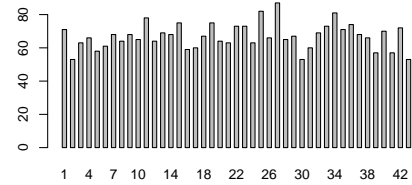


Min:	16.72	Mean:	44.09
Q1:	37.34	Var:	88.11
Median:	43.47		
Q3:	49.91		
Max:	86.65		

Joe, H. (1993). Tests of uniformity for sets of lotto numbers, *Statistics & Probability Letters* **16**(3): 181–188.

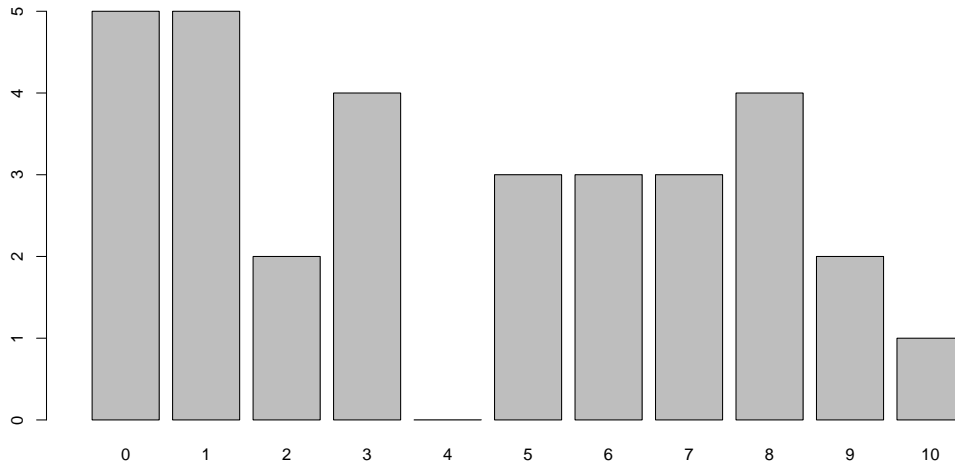
Haigh, J. (1997). The Statistics of the National Lottery, *Journal of the Royal Statistical Society, Series A* **160**(2): 187–206.

Powerball data I: Visual check

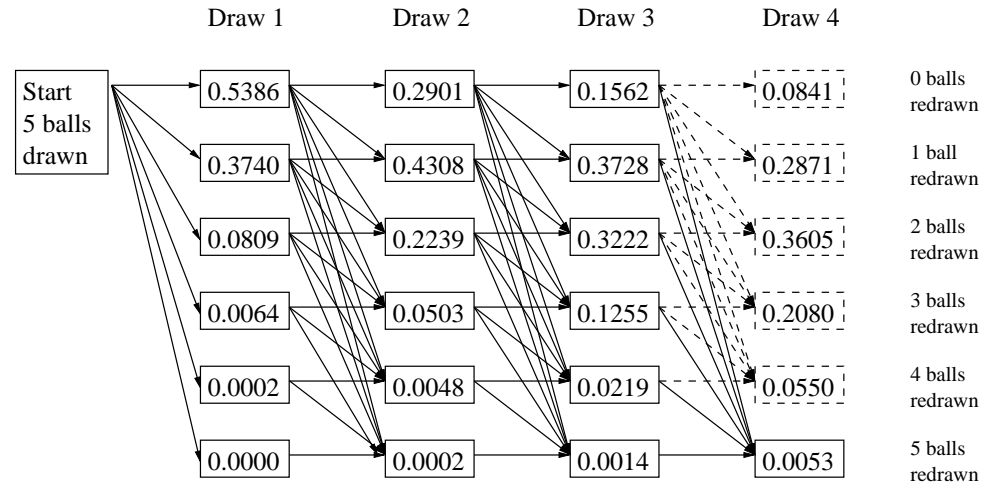


Powerball data II

Barrell A: number of weeks since last drawn



Powerball data II (cont.)



Diaconis, P. and Mosteller, F. (1989). Methods for Studying Coincidences, *Journal of the American Statistical Association* **84**(408): 853–861.

Powerball data III

Are the waiting times for Barrel B suspiciously long?

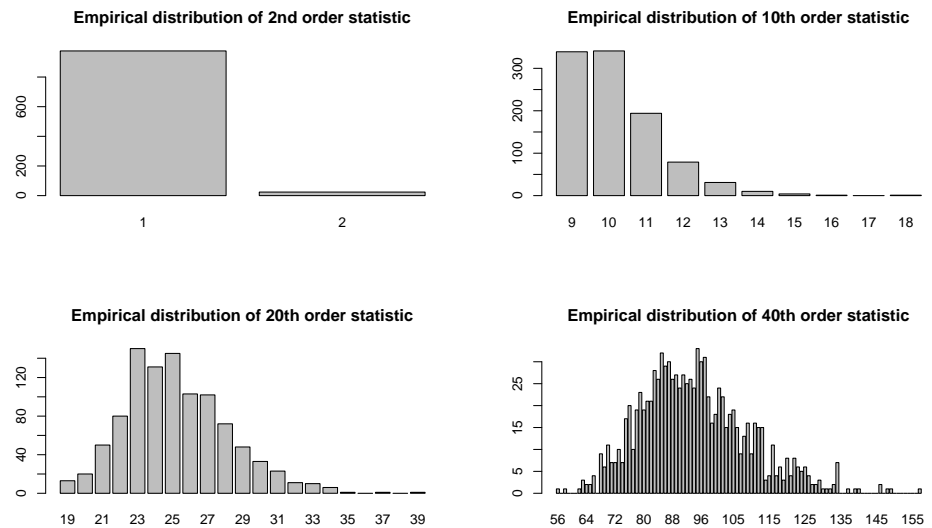
Idea: instead of studying the actual waiting times:

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
6	3	2	38	24	170	34	5	68	77	65	114	13	0	52
16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
41	45	186	37	11	10	15	9	29	80	44	56	50	4	54
31	32	33	34	35	36	37	38	39	40	41	42	43	44	45
58	36	39	53	8	16	1	46	7	25	86	28	30	20	62

study the order statistics:

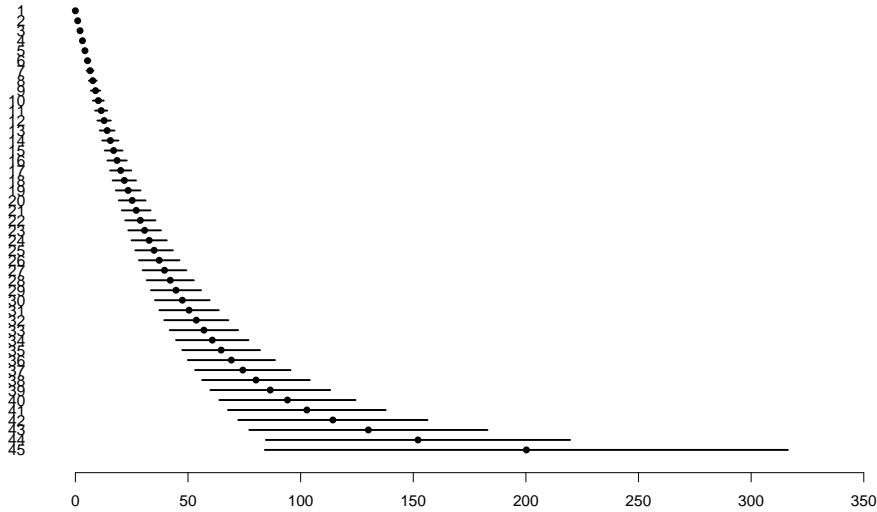
0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 13, 15, 16, 20, 24, 25, 28, 29, 30, 34, 36, 37, 38, 39, 41, 44, 45, 46, 50, 52, 53, 54, 56, 58, 62, 65, 68, 77, 80, 86, 114, 170, 186
(via simulation)

Powerball data III (cont.)



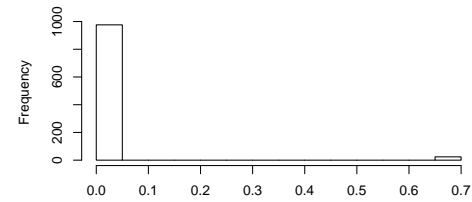
Powerball data III (cont.)

Empirical mean and StDev of order statistics

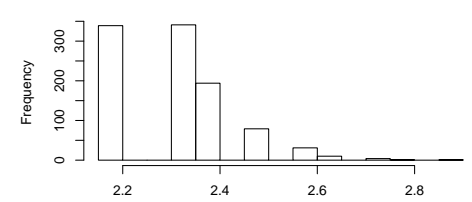


Powerball data III (cont.)

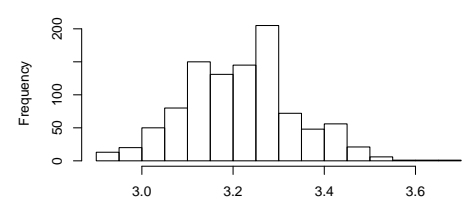
Empirical distribution of log(2nd order statistic)



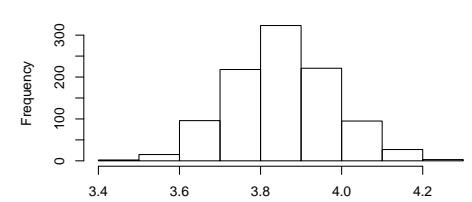
Empirical distribution of log(10th order statistic)



Empirical distribution of log(20th order statistic)

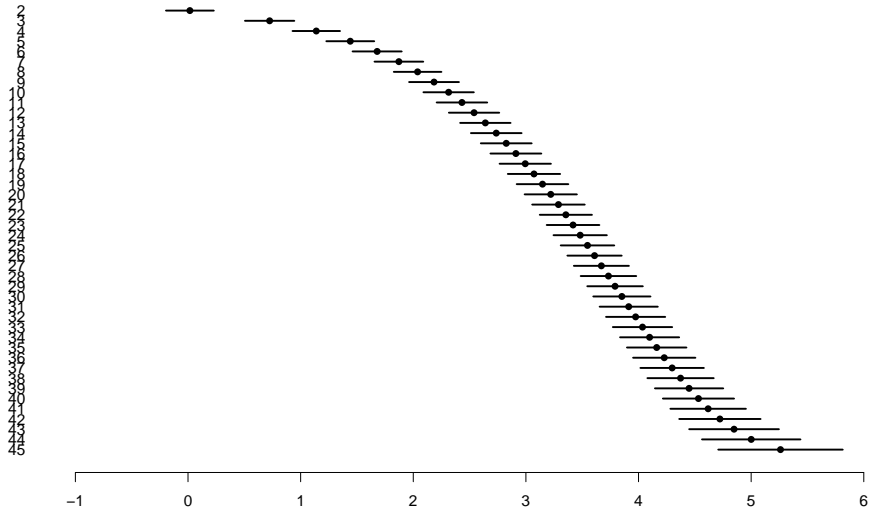


Empirical distribution of log(40th order statistic)



Powerball data III (cont.)

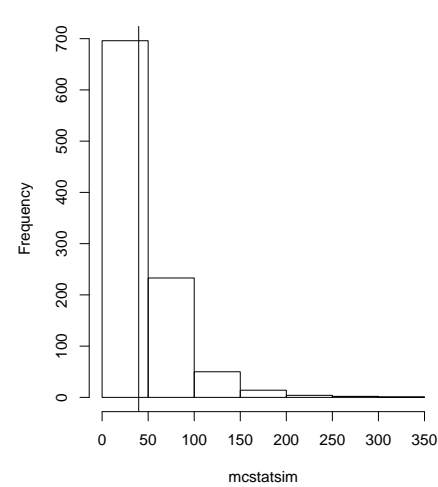
Empirical mean and StDev of log(order statistics)



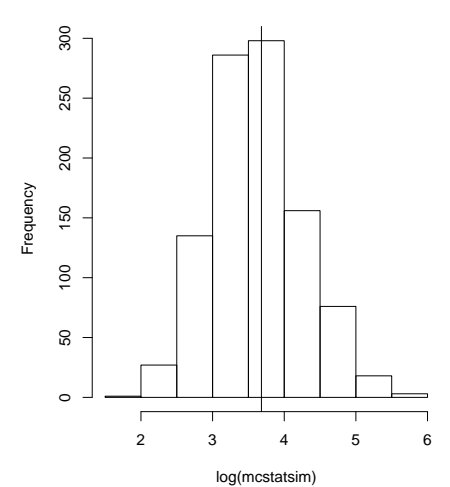
Powerball data III (cont.)

Using as test statistics a Pearson- χ^2 type statistic.

Histogram of simulated test statistic



Histogram of simulated log(test statistic)



Powerball data IV

How many draws (weeks) does it take, on average, until each ball in Barrel B was drawn at least once?

This is a variation of the coupon collector's problem.
The answer is

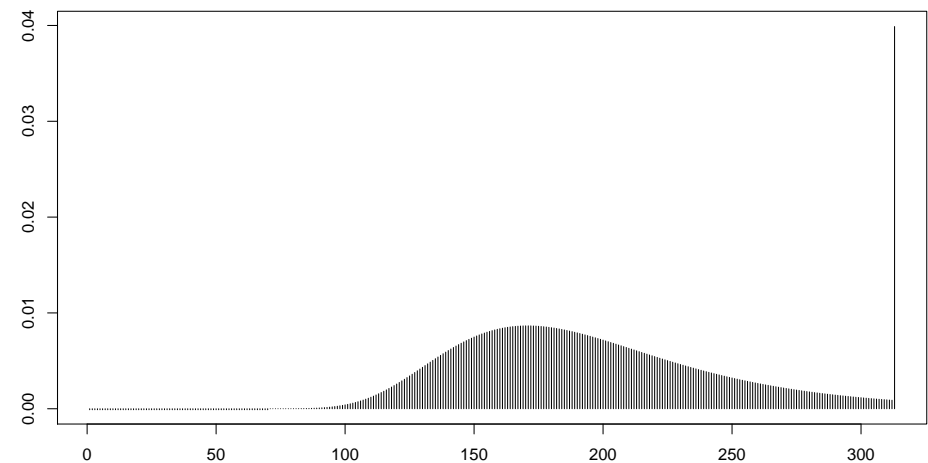
$$\sum_{i=1}^{45} \frac{45}{i} \approx 197.77$$

That is, a bit under four years.

After 6 years what is the most likely value for the 45th order statistic for “weeks since last drawn” from Barrel B? After 7 years? After 8 years? After 605 draws (roughly 12 years)?

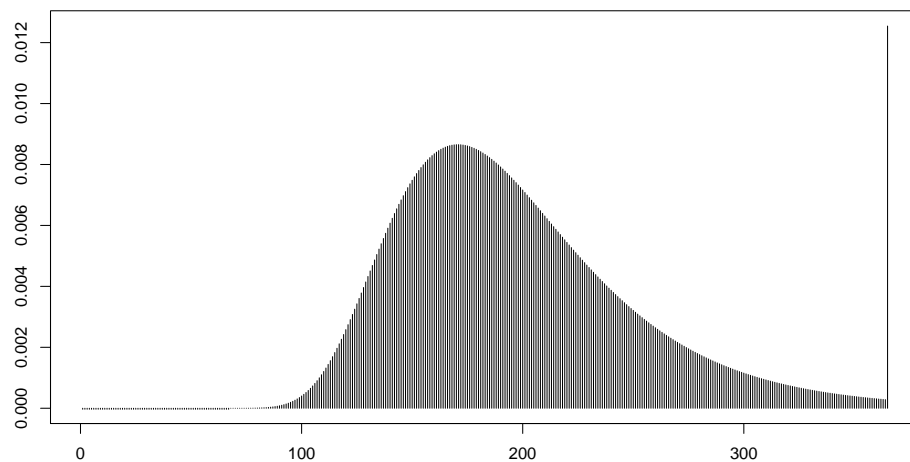
Powerball data IV (cont.)

312 draws, 1 from 45, distribution of max gap



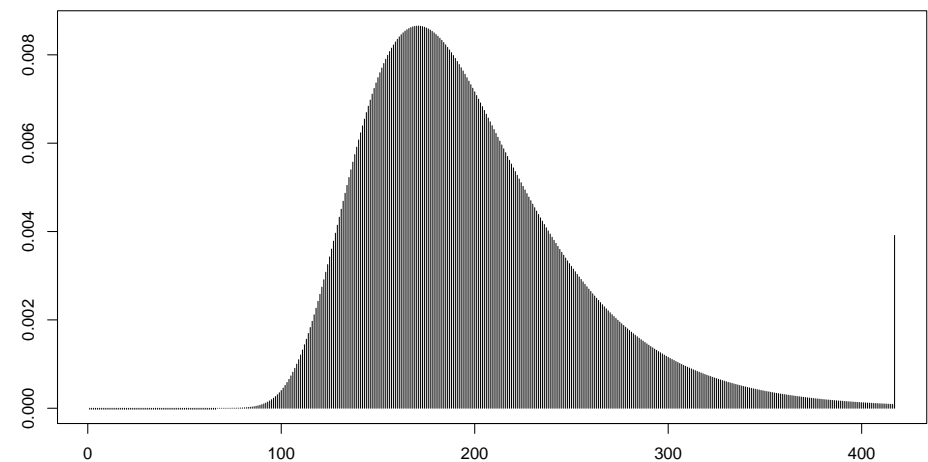
Powerball data IV (cont.)

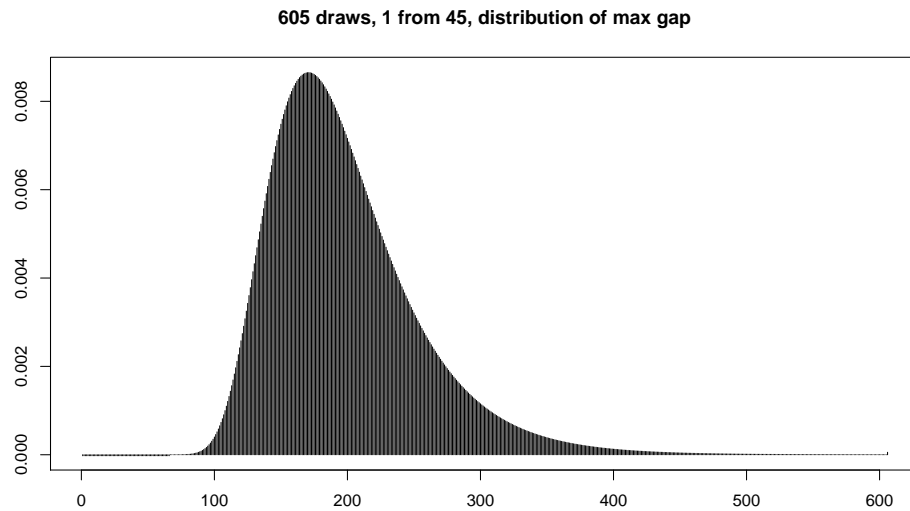
364 draws, 1 from 45, distribution of max gap



Powerball data IV (cont.)

416 draws, 1 from 45, distribution of max gap





<http://www.understandinguncertainty.org/>

Gigerenzer, G. (2002). *Calculated Risks: How to know when Numbers Deceive You*, Simon & Schuster.

Gigerenzer, G. (2007). *Gut feelings : The Intelligence of The Unconscious*, Viking.

Haigh, J. (2003). *Taking Chances: Winning with Probability (2nd ed)*, Oxford University Press.

Henze, N. and Riedwyl, H. (1998). *How to Win More: Strategies for Increasing a Lottery Win*, A K Peters.

Joe, H. (1990). A Winning Strategy for Lotto Games?, *Canadian Journal of Statistics* **18**(3): 233–244.

Literature of interest (cont.)

Mlodinow, L. (2009). *The Drunkard's Walk: How Randomness Rules Our Lives*, Vintage.

Nahin, P.J. (2008). *Digital Dice: Computational Solutions to Practical Probability Problems*, Princeton University Press.

Olofsson, P. (2007). *Probabilities: The Little Numbers That Rule Our Lives*, Wiley & Sons.

Rosenthal (2006). *Struck by Lightning: The Curious World of Probabilities*, Joseph Henry Press.

Senn, S. (2003). *Dicing with Death: Chance, Risk and Health*, Cambridge University Press.

Woolfson, M.M. (2008). *Everyday Probability and Statistics: Health, Elections, Gambling and War*, Imperial College Press.