Lecture 1

David Aldous

26 August 2015

- ''Old school'' course – lectures with chalk, discussion section, homework, midterms, final, textbooks.
- Info on web page – Google "Aldous STAT 150".
- Pencil and paper mathematical probability theory – not data/computing.
- Need to understand STAT 134 (seriously).
- First few lectures are review of STAT 134 and extra material involving conditioning; mostly by working questions. Questions presented on slides, work on blackboard, math summarized on slides, brief mentions of conceptual background.
- First homework due Wednesday 9 September.
- In Spring I will teach a fun "Probability in the Real World" course (STAT 157) limited to 36 students; admission by pre-quiz.

Outcome	PredictWise	Derived Betfair Price	Betfair Back	Betfair Lay
Seattle Seahawks	18 %	$ 0.171	5.80	5.90
Green Bay Packers	14 %	$ 0.137	7.20	7.40
Indianapolis Colts	9 %	$ 0.089	11.00	11.50
New England Patriots	8 %	$ 0.077	12.50	13.50
Denver Broncos	6 %	$ 0.063	15.50	16.50
Dallas Cowboys	5 %	$ 0.047	21.00	22.00
Baltimore Ravens	4 %	$ 0.038	26.00	27.00
Philadelphia Eagles	3 %	$ 0.036	26.00	30.00
Pittsburgh Steelers	3 %	$ 0.035	28.00	30.00
Miami Dolphins	3 %	$ 0.033	29.00	32.00
Arizona Cardinals	3 %	$ 0.026	36.00	40.00
Cincinnati Bengals	2 %	$ 0.024	40.00	42.00
Kansas City Chiefs	2 %	$ 0.022	44.00	46.00
Carolina Panthers	2 %	$ 0.021	44.00	55.00
Atlanta Falcons	2 %	$ 0.020	48.00	50.00
New Orleans Saints	2 %	$ 0.019	50.00	55.00
Buffalo Bills	2 %	$ 0.018	50.00	65.00
Detroit Lions	2 %	$ 0.017	55.00	60.00
New York Giants	2 %	$ 0.017	55.00	60.00
Minnesota Vikings	2 %	$ 0.017	55.00	65.00
San Diego Chargers	1 %	$ 0.016	60.00	65.00
St Louis Rams	1 %	$ 0.014	65.00	75.00
Houston Texans	1 %	$ 0.013	75.00	80.00
San Francisco 49ers	1 %	$ 0.010	90.00	110.00
New York Jets	1 %	$ 0.010	100.00	110.00

Just for fun, here are probabilities (as perceived by gamblers) for next Superbowl winner.

No simple math theory for the actual numbers here (can build complex statistical models based on players statistics for past performance) but there is math theory for how these probabilities will fluctuate as the season progresses:

there will be (on average) 5 teams whose perceived probability will sometime exceed 20%.

I'll explain this in the "martingales" section of the course.

1. A famous hypothetical example.

A taxi was involved in a hit and run accident at night. Two taxi companies, the Green and the Blue, operate in the city. You are given the following data:

- 85% of the taxis in the city are Green and 15% are Blue.
- A witness identified the taxi as Blue. The court tested the reliability of the witness under the same circumstances that existed on the night of the accident and concluded that the witness correctly identified each one of the two colors 80% of the time and failed 20% of the time.

What is the probability that the taxi involved in the accident was Blue rather than Green?

This is **Bayes rule**, of course. Rather than memorize the formula I find it easier to do the argument directly, as follows.

$$\mathbb{P}(\text{is Blue; and identified as Blue}) = 15\% \times 80\% = 12\%$$

$$\mathbb{P}(\text{is Green; and identified as Blue}) = 85\% \times 20\% = 17\%$$

$$\mathbb{P}(\text{ identified as Blue}) = 12\% + 17\% = 29\%$$

$$\mathbb{P}(\text{is Blue } \mid \text{identified as Blue}) = 12\%/29\% \approx 41\%$$

So the witness is probably wrong!

Extensive work in Psychology and Behavioral Economics on how "ordinary people" think about questions involving probability and risks – "decisions under uncertainty" – see popular book *Thinking, Fast and Slow* by Kahneman.

2. When should you buy every Lotto combination?

Simplifying some real-world issues, assume

- Ticket cost 1 dollar.
- N combinations of numbers
- "hold over" A dollars from previous drawing (no winner)
- M other tickets sold for this drawing
- 50% of ticket price goes to "pool" of prize money.

We want a formula for $\mathbb{E}(\text{gain})$ – when does this work out to be > 0?

If you buy all N combinations

$$\text{pool of prize money} = A + (M+N)/2.$$

You have a winning ticket but so do some random number X of other people; so your winnings are $\frac{1}{1+X} \times [A + (M+N)/2]$ and your

$$\text{gain} \ = \frac{1}{1+X} \times [A + (M+N)/2] \ - \ N.$$

A rough model is that X has Poisson($\lambda = M/N$) distribution. We need to calculate $\mathbb{E}\frac{1}{1+X}$; recall the general formula:

$$\mathbb{E}g(X) = \sum_x g(x)\mathbb{P}(X = x).$$

$$\mathbb{E}\frac{1}{1+X} = \sum_{x=0}^{\infty} \frac{1}{1+x} e^{-\lambda}\lambda^x/x! = \lambda^{-1}e^{-\lambda}\sum_{x=1}^{\infty}\lambda^x/x!$$

$$= \lambda^{-1}e^{-\lambda}(e^{\lambda}-1) = \lambda^{-1}(1-e^{-\lambda}).$$

This is a bit complicated but we end with explicit formula

$$\mathbb{E}(\text{gain}) = \frac{1-e^{-M/N}}{M/N}[A + (M+N)/2] - N.$$

Bottom line: need "hold over" amount A to be large, and the number of other tickets M to be small; but this doesn't happen because a large "hold over" encourages other people to buy tickets.

Recall terminology about gambling – also apply to stock market investment, insurance etc.

- You end with a random "gain" , taking your costs into account.
- A loss is a negative gain.

A bet is called

- **fair** if $\mathbb{E}(\text{gain}) = 0$
- **favorable** if $\mathbb{E}(\text{gain}) > 0$
- **unfavorable** if $\mathbb{E}(\text{gain}) < 0$

Using expectation as a criterion is not always appropriate. If potential gains/losses are large then issues of **utility** arise. Here is another case – imagining stock market investment as a favorable game.

3. Gambling on a favorable game.

Hypothetical setting: events $A_1, A_2, \ldots$ independent, $\mathbb{P}(A_i) = p > 1/2$. Imagine $p = 0.51$. But you can make bets at even odds (as if p were $1/2$). What is the best way to exploit this favorable game, if you have an initial "fortune" x_0 and can only work with your winnings, not with extra money?

Common sense suggests a strategy: choose $0 < \alpha < 1$ and, at each step, bet a proportion α of your current fortune. If you use this strategy, what happens in the long run?

Students often find it hard to get started on this question. It's helpful to remember the **indicator r.v.** $I(A)$. We want to study

$$X_n = \text{ your fortune after } n \text{ bets}$$

and we can relate X_n to X_{n-1} as follows.

$$X_n = V_n X_{n-1}, \quad \text{where}$$

$$V_n = 1 + \alpha \text{ if } A_n \text{ happens}, \ V_n = 1 - \alpha \text{ if not.}$$

Then

$$\log X_n = \log x_0 + \sum_{i=1}^{n} \log V_i .$$

By the law of large numbers we have

$$\frac{1}{n} \log X_n \to \mathbb{E} \log V_1 .$$

Now saying $\frac{1}{n} \log X_n \to b$ is roughly saying that X_n behaves as $\exp(bn)$, so our "long-run optimality" criterion is to maximize this long run growth rate $b = \mathbb{E} \log V_1$.

Rewrite

$$p = \tfrac{1}{2} + \delta \text{ for small } \delta > 0.$$

So the long run growth rate is

$$\begin{aligned} \mathbb{E} \log V_1 &= (\tfrac{1}{2} + \delta) \log(1+\alpha) + (\tfrac{1}{2} - \delta) \log(1-\alpha) \\ &\approx (\tfrac{1}{2} + \delta)(\alpha - \alpha^2/2) + (\tfrac{1}{2} - \delta)(-\alpha - \alpha^2/2) \\ &= 2\alpha\delta - \alpha^2/2. \end{aligned}$$

In this question, δ is given, but we can choose α to maximize this expression. So we choose $\alpha = 2\delta$ and get a long run growth rate $\approx 2\delta^2$.

This is called the **Kelly strategy –** see popular books such as *Fortune's Formula* (William Poundstone) or *Red-Blooded Risk: The Secret History of Wall Street* (Aaron Brown), who describes it as "getting rich exponentially slowly". A conceptual point is that this strategy is quite different from "maximize $\mathbb{E}(\text{gain})$ on each step", which would make you bet all your fortune on each bet, and therefore likely lose all your money quickly.