Probability: from intuition to mathematics

Ting-Kam Leonard Wong

University of Southern California

EPYMT 2017

Probability has a right and a left hand. On the right is the rigorous foundational work... The left hand thinks probabilistically, reduces problems to gambling situations, coin-tossing, motions of a physical particle.

Leo Breiman (1968)

Warm up

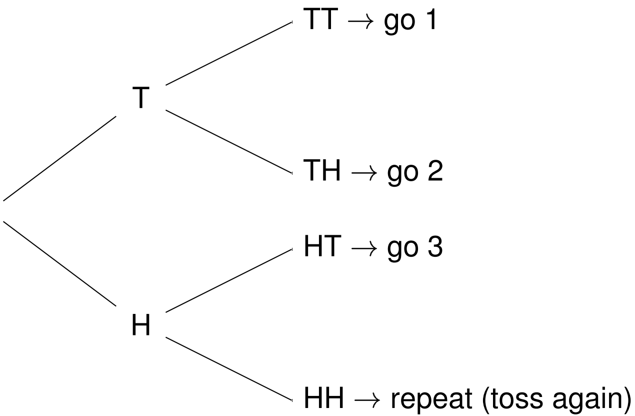
– You want to decide which of 3 restaurants to go.

– You have a fair coin, so that Pr(head) $=$ Pr(tail) $= \frac{1}{2}$.

– You want to choose randomly, so that each restaurant has probability $\frac{1}{3}$ of being chosen.

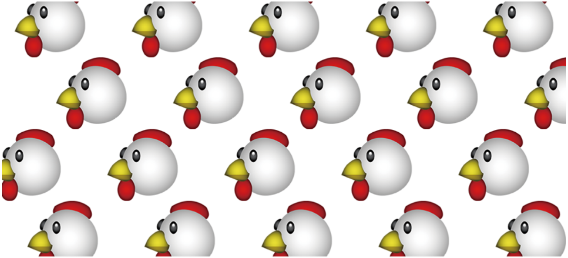
How should you use the coin?

Solution



- $\Pr(\text{HHHHHH}\ldots) = 0$.
- No solution can be constructed by using a fixed and finite number of tosses. Why?

One more



農場裡有 100 隻雞，安靜地圍成一個圈。忽然，每隻雞都把自己身旁的一隻雞啄了一下。

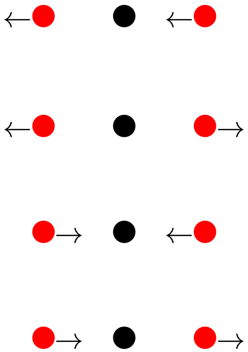
如果每隻雞啄左邊或右邊是隨機的，會有多少隻雞沒有被啄？

美國 Mathcounts 全國初中數學競賽總決賽決勝題

/ 曾鈺成

Solution

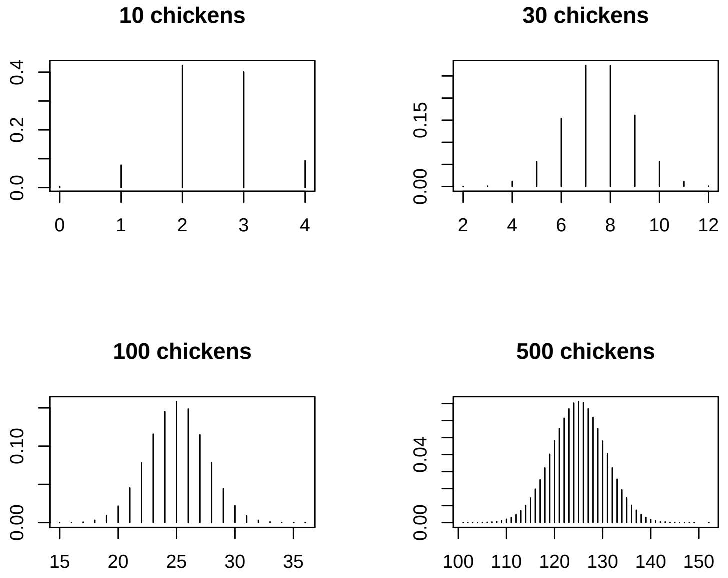
Consider the situation of a fixed chicken:



On average, $100 \times 1/4 = 25$ chickens are not pecked.

Expected *total* number does NOT depend on interdependence.

Probability distributions



What is probability?

- Frequency:

$$\text{probability} = \text{limiting proportion}$$

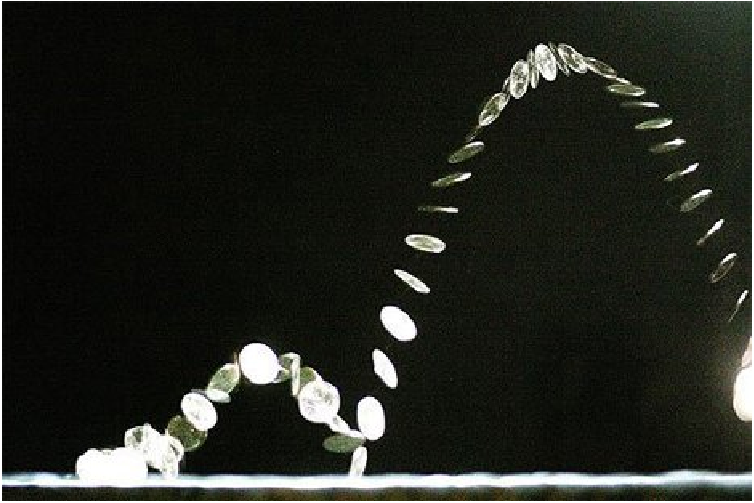
- Subjective (degree of belief):

"I am 90% sure I will get 5** in DSE Math."

Important:

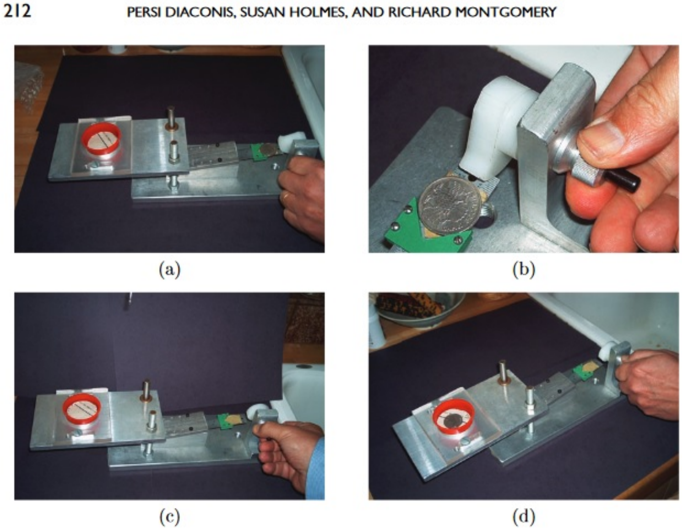
- Mathematics does not tell us what probability 'really' is.
- Axiomatic framework (Kolmogorov 1933).

Coin tossing



(Source: `https://uncommondescent.com/`)

Coin tossing machine



(Source: Diaconis, Holmes and Montgomery (2007))

Subjective probability

For each unknown X, give two numbers a and b such that you are 90% sure that $a \leq X \leq b$:

1. population of Vietnam in 2015
2. half-life of uranium-238 (a radioactive substance)
3. the year Universität Wien (in Austria) was founded
4. gestation (pregnancy) period of blue whale
5. average lifespan of *C. porosus* (saltwater crocodile)

Subjective probability

Answers:

1.	Vietnam:	about 91.7 million
2.	uranium-238:	about 4.5 billion years
3.	Universität Wien:	1365 AD
4.	blue whale:	about 11 months
5.	saltwater crocodile:	about 70 years

An interview question

- 50% of all people who receive a first interview receive a second interview.
- 95% of your friends that got a second interview felt they had a good first interview.
- 75% of your friends that DID NOT get a second interview felt they had a good first interview.

If you feel that you had a good first interview, what is the probability you will receive a second interview?

Source: `http://stats.stackexchange.com/questions/86015/amazon-interview-question-probability-of-2nd-interview`

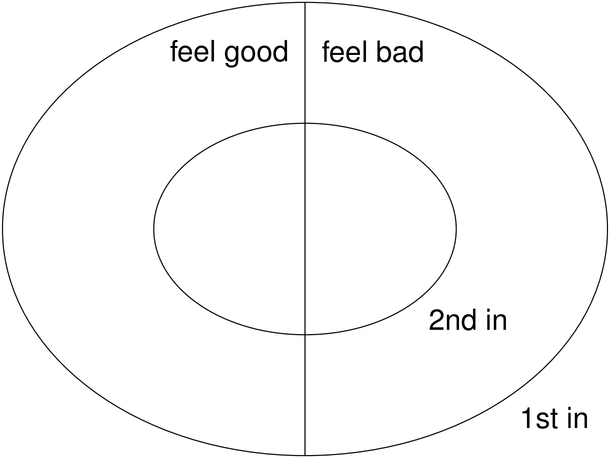
A picture

$$\Pr(\text{2nd in} | \text{feel good}) = \frac{\Pr(\text{feel good and 2nd in})}{\Pr(\text{feel good})}$$

A picture

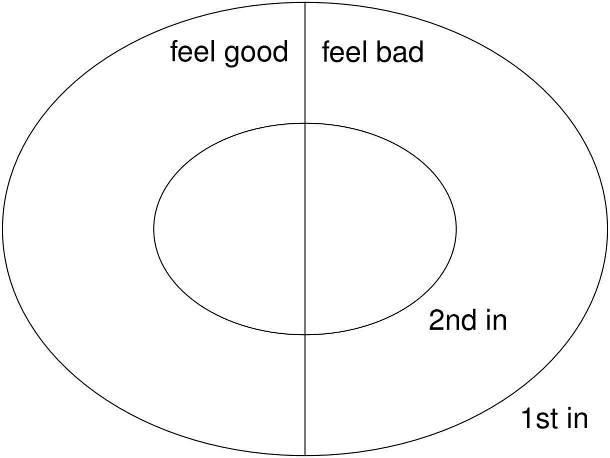
$$\Pr(\text{2nd in}|\text{feel good}) = \frac{\Pr(\text{feel good and 2nd in})}{\Pr(\text{feel good})}$$

A picture



$$\Pr(\text{2nd in} \mid \text{feel good}) = \frac{\Pr(\text{feel good and 2nd in})}{\Pr(\text{feel good})}$$

A picture



$$\Pr(\text{2nd in}|\text{feel good}) = \frac{\Pr(\text{feel good and 2nd in})}{\Pr(\text{feel good})}$$

Probability theory is nothing else but common sense reduced to calculations. (Laplace)

Product rule:

$$\begin{aligned}\Pr(\text{feel good and 2nd in}) &= \Pr(\text{2nd in})\Pr(\text{feel good}|\text{2nd in}) \\ &= 0.5 \times 0.95\end{aligned}$$

Similarly,

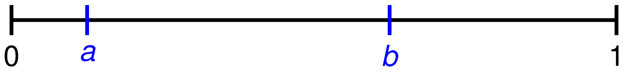
$$\Pr(\text{feel good}) = 0.5 \times 0.95 + 0.5 \times 0.75. \quad \textbf{(sum rule)}$$

So

$$\Pr(\text{2nd in}|\text{good}) = \frac{0.5 \times 0.95}{0.5 \times 0.95 + 0.5 \times 0.75} = 56\%.$$

A random point in [0, 1]

Imagine a point X is picked *uniformly at random* in the interval $[0, 1]$:



For any $a < b$, we should have $\Pr(a \leq X \leq b) = b - a$. Then

$$\Pr(X = x) = 0, \quad \text{for any } x,$$

but

$$\sum_{x \in [0,1]} \Pr(X = x) = 0 \neq 1 = \Pr(0 \leq X \leq 1).$$

A number is chosen, but the probability of getting a fixed number is zero!

Axiomatic framework

There is a set Ω (sample space) interpreted as the collection of all possible outcomes, e.g.,

$$\Omega = \{1, 2, 3, 4, 5, 6\} \quad \text{(throwing a die)}$$

$$\Omega = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\} \quad \text{(three coins)}$$

A probability *function* $\mathbb{P}$ assign numbers between 0 and 1 to *certain* subsets of Ω:

$$\begin{aligned}
\mathbb{P}(\Omega) &= 1 \quad \text{(by definition)} \\
\mathbb{P}(\{2, 4, 6\}) &= \frac{3}{6} \\
\mathbb{P}(H \text{ before } T) &= \mathbb{P}(\{HHH, HHT, HTH, HTT\}) = \frac{4}{8}
\end{aligned}$$

Axiomatic framework

The only general restriction is that $\mathbb{P}$ is *countably additive*:

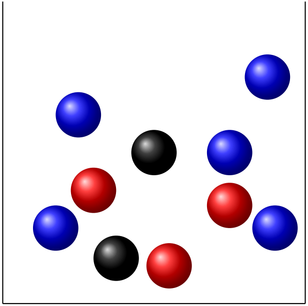
$$A_1, A_2, \ldots \text{ mutually exclusive}$$

$$\Rightarrow$$

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$$

Deep mathematical fact: In many cases (including choosing uniformly a random point in $[0, 1]$) it is **impossible** to assign probabilities to **all** subsets of the sample space such that countable additivity remains valid.

Case study: drawing balls



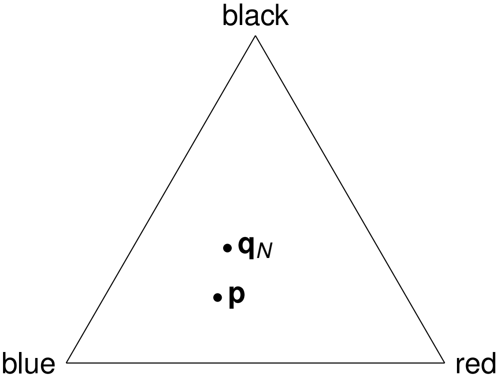
- 10 balls: 5 blue, 3 red, 2 black
- Sampling with replacement $\Rightarrow$ independence
- Simple model of realistic situations

Theoretical vs empirical distributions

Suppose N balls are drawn, get X_1 blue, X_2 red, X_3 black.

$$\mathbf{p} = \text{theoretical proportions} = \left(\frac{5}{10}, \frac{3}{10}, \frac{2}{10}\right)$$

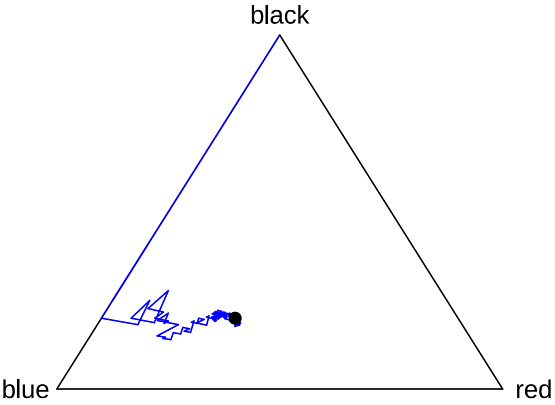
$$\mathbf{q}_N = \text{empirical proportions} = \left(\frac{X_1}{N}, \frac{X_2}{N}, \frac{X_3}{N}\right)$$



Law of large numbers

Theorem (Strong law of large numbers)

$\lim_{N\to\infty} \mathbf{q}_N = \mathbf{p}$ *with probability 1.*



Law of iterated logarithm

Theorem (Law of iterated logarithm)

With probability 1, we have

$$\limsup_{N\to\infty} \frac{\|\mathbf{q}_N - \mathbf{p}\|}{\sqrt{\frac{\log\log N}{N}}} = \textit{constant}.$$

Central limit theorem

Theorem (Central limit theorem)

The distribution of

$$\sqrt{N}\,(\mathbf{q}_N - \mathbf{p})$$

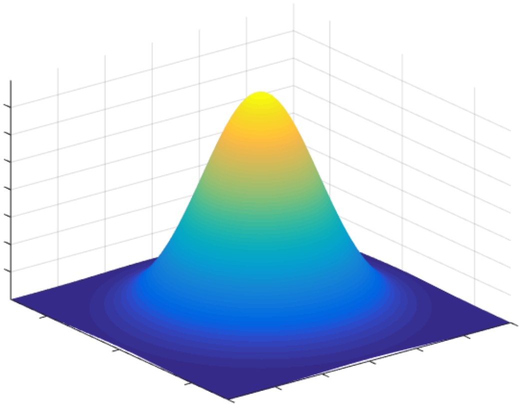
converges to a Gaussian distribution as $N \to \infty$.

Gaussian distributions

One dimension:

$$\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}}$$

Higher dimensions:



Large deviations: probabilities of improbable events

- We expect that $\mathbf{q}_N \approx \mathbf{p}$, at least when N is large.
- Thus, for any proportions $\mathbf{r} \neq \mathbf{p}$, we expect that

$$\mathbb{P}\left(\mathbf{q}_N \approx \mathbf{r}\right)$$

  is *very* small, when N is large.
- How small?
- The 'smallness' should depend on the 'distance' between $\mathbf{r}$ and $\mathbf{p}$.
- The appropriate 'distance' is NOT the Euclidean distance

$$\|\mathbf{r} - \mathbf{p}\| = \sqrt{(r_1 - p_1)^2 + (r_2 - p_2)^2 + (r_3 - p_3)^2}.$$

How small?

$$\mathbb{P}\left(\mathbf{q}_N = (\frac{x_1}{N}, \frac{x_2}{N}, \frac{x_3}{N})\right) = \binom{N}{x_1\ x_2\ x_3} p_1^{x_1} p_2^{x_2} p_3^{x_3}$$
$$= \frac{N!}{x_1!x_2!x_3!} p_1^{x_1} p_2^{x_2} p_3^{x_3}.$$

The key is the celebrated Stirling's formula:

$$n! \sim \sqrt{2\pi}\, n^{n+\frac{1}{2}} e^{-n}$$

Relative entropy

Writing $\mathbf{r} = (\frac{x_1}{N}, \frac{x_2}{N}, \frac{x_3}{N})$, we have

Theorem (Sanov's theorem)

$$\mathbb{P}\left(\mathbf{q}_N = \mathbf{r}\right) \approx e^{-N \cdot H(\mathbf{r}|\mathbf{p})},$$

where

$$H(\mathbf{r}|\mathbf{p}) = \sum_{i=1}^{3} r_i \log \frac{r_i}{p_i} \geq 0$$

is the relative entropy.

Important: Relative entropy is a fundamental distance-like quantity in probability theory.

Suggestions for further reading

- *Probability Theory: The Logic of Science*
  Jaynes and Bretthorst
  – Enlightening philosophical discussions
- *Thinking, Fast and Slow*
  Kahneman
  – How humans actually make decisions
- *An Introduction to Probability Theory and Its Applications*
  Feller
  – The *best* mathematical probability book ever written