Engineering Statistics
IES 302

Dr. Prapun Suksompong

prapun@siit.tu.ac.th

5 Foundation of Probability Theory

Axioms of probability theory

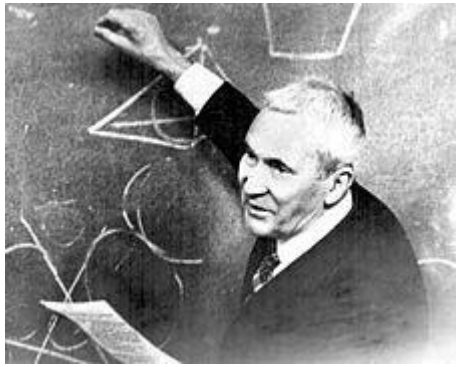
- Abstractly, a **probability measure** is a function that assigns numbers to events, which satisfies the following assumptions:

1. Nonnegativity: For any event A, $P(A) \geq 0$
2. Unit normalization: $P(\Omega) = 1$
3. If $A_1, A_2, \ldots,$ is an infinite sequence of (pairwise) **disjoint** events, then

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

Kolmogorov

- Andrey Nikolaevich Kolmogorov
- Soviet Russian mathematician
- Advanced various scientific fields
  - probability theory
  - topology
  - classical mechanics
  - computational complexity.



- 1922: Constructed a Fourier series that diverges almost everywhere, gaining international recognition.
- 1933: Published the book, **Foundations of the Theory of Probability**, laying the modern axiomatic foundations of probability theory and establishing his reputation as the world's leading living expert in this field.

I learn probability theory from

Eugene Dynkin

Philip Protter

Gennady Samorodnitsky

Terrence Fine

Xing Guo

Toby Berger

Rick Durrett

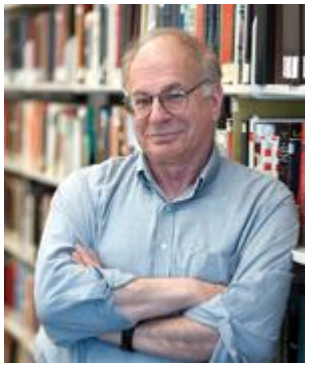
Not too far from Kolmogorov

You can be

the 4th-generation

probability theorists

Daniel Kahneman



- Daniel Kahneman
- Israeli-American **psychologist**
- 2002 **Nobel** laureate
  - In **Economics**
- Hebrew University, Jerusalem, Israel.
- Professor emeritus of psychology and public affairs at **Princeton** University's Woodrow Wilson School.
- With Amos **Tversky**, Kahneman studied and clarified the kinds of misperceptions of randomness that fuel many of the common fallacies.

[outspoken = given to expressing yourself freely or insistently]

K&T: Q1

Imagine a **woman** named **Linda**, **31** years old, **single**, **outspoken**, and very **bright**. In college she majored in **philosophy**. While a student she was deeply concerned with **discrimination** and **social justice** and participated in **antinuclear demonstrations**.

- K&T presented this description to a group of 88 subjects and asked them to **rank** the eight statements (shown on the next slide) on a scale of 1 to 8 according to their probability, with 1 representing the most probable and 8 the least.

[Daniel Kahneman, Paul Slovic, and Amos Tversky, eds., Judgment under Uncertainty: Heuristics and Biases (Cambridge: Cambridge University Press, 1982), pp. 90–98.]

[feminist = of or relating to or advocating equal rights for women]

K&T: Q1 - Results

- Here are the results - from most to least probable

Statement	*Average Probability Rank*
Linda is active in the feminist movement.	2.1
Linda is a psychiatric social worker.	3.1
Linda works in a bookstore and takes yoga classes.	3.3
Linda is a bank teller and is active in the feminist movement.	4.1
Linda is a teacher in an elementary school.	5.2
Linda is a member of the League of Women Voters.	5.4
Linda is a bank teller.	6.2
Linda is an insurance salesperson.	6.4

K&T: Q1 – Results (2)

- At first glance there may appear to be nothing unusual in these results: the description was in fact designed to be
  - representative of an active feminist and
  - unrepresentative of a bank teller or an insurance salesperson.

Most probable

Least likely

Statement	*Average Probability Rank*	
Linda is active in the feminist movement.	2.1	←
Linda is a psychiatric social worker.	3.1	
Linda works in a bookstore and takes yoga classes.	3.3	
Linda is a bank teller and is active in the feminist movement.	4.1	←
Linda is a teacher in an elementary school.	5.2	
Linda is a member of the League of Women Voters.	5.4	
Linda is a bank teller.	6.2	←
Linda is an insurance salesperson.	6.4	

K&T: Q1 – Results (3)

- Let's focus on just three of the possibilities and their average ranks.
- This is the order in which **85 percent** of the respondents ranked the three possibilities:

Statement	*Average Probability Rank*
Linda is active in the feminist movement.	2.1
Linda is a bank teller and is active in the feminist movement.	4.1
Linda is a bank teller.	6.2

- If nothing about this looks strange, then K&T have fooled you

K&T: Q1 - Contradiction

The probability that two events will both occur can never be greater than the probability that each will occur individually!

Statement	*Average Probability Rank*
Linda is active in the feminist movement.	2.1
Linda is a bank teller and is active in the feminist movement.	4.1
Linda is a bank teller.	6.2

K&T: Q2

- K&T were not surprised by the result because they had given their subjects a large number of possibilities, and the connections among the three scenarios could easily have gotten lost in the shuffle.
- So they presented the description of Linda to another group, but this time they presented only three possibilities:
  - Linda is active in the feminist movement.
  - Linda is a bank teller and is active in the feminist movement.
  - Linda is a bank teller.

K&T: Q2 - Results

- To their surprise, **87 percent** of the subjects in this trial also **incorrectly** ranked the probability that "Linda is a bank teller and is active in the feminist movement" higher than the probability that "Linda is a bank teller".
- If the **detail**s we are given **fit our mental picture** of something, then the more details in a scenario, the more real it seems and hence the **more probable** we consider it to be
  - even though any act of adding less-than-certain details to a conjecture makes the conjecture less probable.
- Even **highly trained doctors** make this error when analyzing symptoms.
  - 91 percent of the doctors fall prey to the same bias.

[Amos Tversky and Daniel Kahneman, "Extensional versus Intuitive Reasoning: The Conjunction Fallacy in Probability Judgment," Psychological Review 90, no. 4 (October 1983): 293–315.]

Misuse of probability in law

- It is not uncommon for experts in **DNA analysis** to testify at a criminal trial that a DNA sample taken from a crime scene matches that taken from a suspect.
- How certain are such matches?
- When DNA evidence was first introduced, a number of experts testified that false positives are impossible in DNA testing.
- Today DNA experts regularly testify that the odds of a random person's matching the crime sample are less than **1 in 1 million** or **1 in 1 billion**.
- In Oklahoma a court sentenced a man named Timothy Durham to prison even though **eleven witnesses** had placed him in another state at the time of the crime.

[Mlodinow, 2008, p 37]

Lab/Human Error

- There is **another stat**istic that is often **not presented** to the jury, one having to do with the fact that **labs make errors**, for instance, in collecting or handling a sample, by accidentally mixing or swapping samples, or by misinterpreting or incorrectly reporting results.
- Each of these errors is rare but not nearly as rare as a random match.
- The Philadelphia City Crime Laboratory admitted that it had swapped the reference sample of the defendant and the victim in a rape case
- A testing firm called Cellmark Diagnostics admitted a similar error.

Timothy Durham's case

- It turned out that in the initial analysis the lab had failed to completely separate the DNA of the rapist and that of the victim in the fluid they tested, and the combination of the victim's and the rapist's DNA produced a positive result when compared with Durham's.
- A later retest turned up the error, and Durham was released after spending nearly **four years** in prison.

DNA-Match Error + Lab Error

- Estimates of the error rate due to human causes vary, but many experts put it at around 1 percent.
- Most jurors assume that given the two types of error—the **1 in 1 billion** accidental match and the **1 in 100 lab-error match**—the overall error rate must be somewhere in between, say 1 in 500 million, which is still for most jurors **beyond a reasonable doubt**.

Wait!...

- Even if the DNA match error was extremely accurate + Lab error is very small,
- there is also another probability concept that should be taken into account.
- More about this later.
- Right now, back to notes for more properties of probability measure.