Probability 1

Ian Crawford

Department of Economics

Sources:

Kendall's Advance Theory of Statistics, Vol 1, by A. Stuart and K. Ord, Arnold, 2000

Introduction to Probability Models, by S. M. Ross, Academic Press, 2003

Theoretical Statistics by D.R. Cox and D.V. Hinkley, Chapman & Hall, 1974

Probability, manuscript by Alison Etheridge, 2008

SticiGUI, by Philip B. Stark, Dept of Statistics, UCBerkeley

Introduction to Probability by C M. Grinstead and J. L. Snell, American Mathematical Society, 2001

n*Introduction to Econometric Analysis* by D. Hendry and B. Nielsen. Princeton University Press, 2007

Principles of Statistical Inference, D.R. Cox, Cambridge, 2006

Preliminary to (and formally part of) the core QE course which will run in TT.

Aim: to familiarise you with the main concepts behind elementary probability and statistics in preparation for

1. Microeconomics (e.g. choice under uncertainty and game theory)

2. QE (e.g. statistics and econometrics)

Probability

Probability is the study of randomness. It has a mathematical and a philosophical aspect.

The mathematical aspect concerns the question of **what probability is**

The philosophical aspect connects the mathematical theory with the world – it's concerned with **what probability means.**

Interpretations of Probability

Theories of Probability assign meaning/interpretations to probability statements about the world.

1. *The Frequency Theory* says that the probability of an event is the limit of the relative frequency with which the event occurs in repeated trials under essentially identical conditions.

2. *The Subjective Theory* says that probability is a measure of strength of belief on a scale of 0 to 1.

What is Probability?

For our purposes *probability is what probability does* - it's a measurement system which satisfies certain axioms.

This leaves open the question of how to interpret probability.

Each interpretation is valuable in certain situations, but may not make sense in others.

The issue of interpretation is not one that we'll resolve in this course.

I'll refer you to your Philosophy tutors - good luck!

A Brief History of Probability

Probability is a young branch of mathematics. It began in 16th/17th centuries when the mathematicians Gerolamo Cardano (in Italy), Blaise Pascal and Pierre de Fermat (in France), started to analyse problems arising from games of chance.

This eventually culminated in modern probability theory, the foundations of which were laid by Andrei Nikolaevich Kolmogorov.

Modern probability is built around *set theory*.

A Brief History of Probability

Kolmogorov combined the notion of sample space, introduced by Richard von Mises, and measure theory and presented his axiomatic system for probability theory in 1933.

This axiomatic approach asks *"What do you want your probability measure to do? What properties would you like it to have?"*.

These requirements are enshrined as *axioms* and measures which satisfy these axioms are *Probabilities.*

The Mathematics of Probability Theory

To begin with probability theory mainly considered *discrete* events (with countable* outcomes)

Its methods mainly involved counting (number of positive outcomes and number of possible outcomes) and combinatorics.

Eventually, the theory moved on to events with *continuous* outcomes.

The method of counting outcomes breaks down with continuous outcomes.

*Could be infinite, just keep counting ...

A good way to *begin* to think about probability is with discrete events and equiprobable alternatives e.g. dice. This leads naturally to what is called the classical definition of probability:

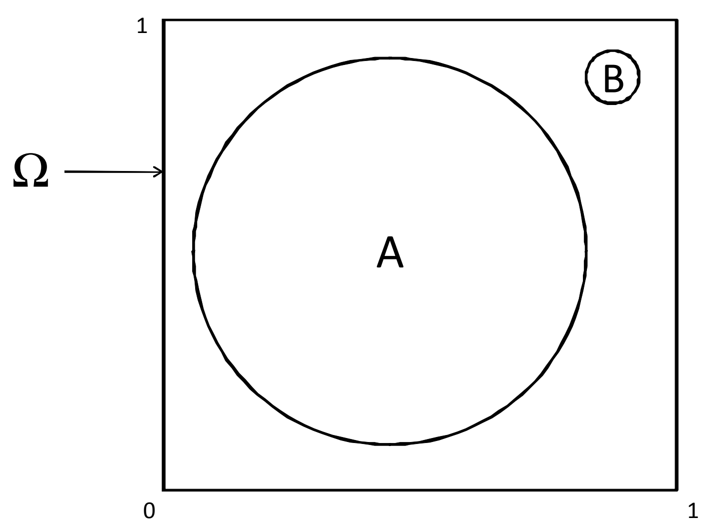
$$\mathsf{Pr}\,(\text{event}) = \frac{\text{number of outcomes which correspond to the event}}{\text{number of all possible outcomes}}$$

But probabilities are also defined for continuous events and alternatives, and the classical definition runs into trouble:

$$\mathsf{Pr}\,(\text{event}) = \frac{\text{number of outcomes which correspond to the event}}{\text{number of all possible outcomes}} = \frac{\infty}{\infty} = 1$$

So perhaps a better way to think about probability (which works when the discrete/dice rolling intuition is inappropriate) is as being analogous to **area** or **volume** or **mass**.

Consider the unit square (which we will call Ω) and regions inside the square (subsets of Ω) like A and B. We can think of their probabilities being related to their **areas**.



Basic Set Theory

The following are a list of definitions and operations which will be useful.

Set. A set is a finite or infinite collection of objects in which order has no significance and which is considered as an object in its own right.

Empty Set. The empty set is the set containing no elements, commonly denoted $\varnothing$.

Element. Members of a set are often referred to as elements and the notation $a \in A$ is used to denote that a is an element of a set A.

Union. The union operation $A \cup B$ produces a new set containing all elements that are in either set.

Intersection. The intersection $A \cap B$ operation produces the set consisting of all elements that are in both of the original sets. Sometimes denotes AB

Subset. A subset is a portion of a set. B is a subset of A (written $B \subseteq A$) iff every member of B is a member of A.

Complementation: Given a set U with a subset A, the complement of A (denoted A^c or A' or $\overline{A}$) is the set of elements of U that are not in A. The set of elements of U not in A is also called the set difference, denoted $U \backslash A$.

[Those of you who have studied logic will gather that there is a close connection between logical operations and set operations. Every logical operation can be represented as an operation on sets by thinking of propositions as subsets of some super set. Logical NOT becomes the set complement, logical AND becomes the set intersection, logical OR becomes the set union, etc.]

The Elements of Probability Theory

Sample space. The set Ω, is called the sample space, if it contains all possible (primitive) outcomes that we are considering.

Event. An event is a subset of Ω (including Ω itself)

Event Space. An event space $\mathcal{F}$ is a set of subsets of Ω which must satisfy certain properties [which need not overly concern us].

Examples:

Consider the roll of a fair dice.

The sample space, is $\Omega = \{1, 2, 3, 4, 5, 6\}$.

The event set $\mathcal{F}$ is a big set composed of all of the subsets of this (there are 2^6 of these so I won't list them!)

For example $A = \{6\}$ is the event 'the result is a 6' and $B = \{2, 4, 6\}$ is the event 'the result is even'. A and B are subsets of Ω so they are in $\mathcal{F}$

Probability axioms based on the triple $(\Omega, \mathcal{F}, \text{Pr})$

For a sample space Ω a probability function is a function Pr that satisfies, for all possible events $\mathcal{F}$ which are subsets of Ω,

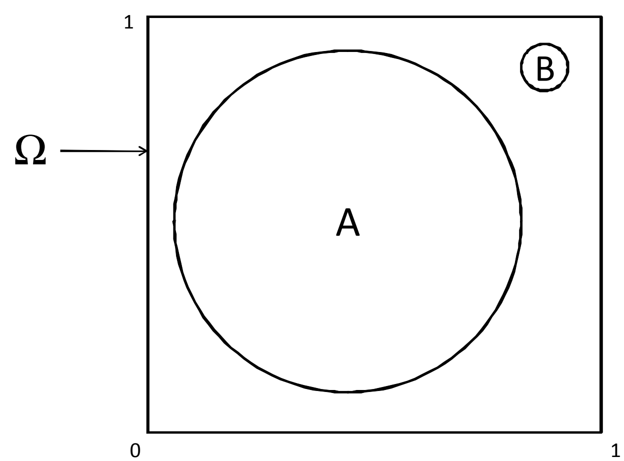
1. $\text{Pr}(A) \geq 0$, for all $A \in \mathcal{F}$

2. $\text{Pr}(\Omega) = 1$

3. If $A_1, A_2, ... \in \mathcal{F}$ are disjoint then $\text{Pr}(A_1 \cup A_2 \cup ....) = \text{Pr}(A_1) + \text{Pr}(A_2) + ...$

Pardon?

Go back to our area/volume intuition for probability. Consider regions like A and B.



The area of any such region is at least zero; the area of Ω is 1; and the area of the union of two regions is the sum of their areas, if they do not overlap (i.e., if they are disjoint). These facts are direct analogues of the axioms of probability.

To Summarise

Probabilities are function of sets.

The event space represents all subsets of the event space and so provides the basis on which valid events can be defined.

The Axioms give us a basis on which to proceed.

The Consequence: Rules

We have seen how to formalise the notion of probability - it's a measure on $[0, 1]$ which satisfies some properties.

The probability is a function Pr which assigns a number from $[0, 1]$ to each element of $\mathcal{F}$.

The axioms allow us to deduce the basic **rules of probability.**

Everything that is mathematically true of probability (including these rules) is a consequence of the Axioms of Probability.

The Consequence: Rules

I'm going to walk you through how some of rules are implied by the Axioms.

It's good for you to know the basic rules of probablity.

But mainly I'm concerned that you see the connection between the axioms and these rules.

These rules are *not* just a set of plausible-sounding laws which appear out of nowhere and which must be learned by rote.

1. The Complement Rule

Rule: The probability that an event occurs is always equal to 1 minus the probability that the event does not occur: $\text{Pr}(A^c) = 1 - \text{Pr}(A)$.

The union of A and A^c is Ω (either A happens or it does not, and there is no other possibility), so $\text{Pr}(A \cup A^c) = \text{Pr}(\Omega) = 1$ by axiom 2.

The event A and A^c disjoint (if "A does not happen" happens, A does not happen; if A happens, "A does not happen" does not happen), so $\text{Pr}(A \cup A^c) = \text{Pr}(A) + \text{Pr}(A^c)$ by axiom 3.

Putting these together, we get $\text{Pr}(A) + \text{Pr}(A^c) = 1$ (which is the rule we were after).

A special case of the Complement Rule is that the probability of the empty set is always zero.

2. The Probability of the Union of Two Events

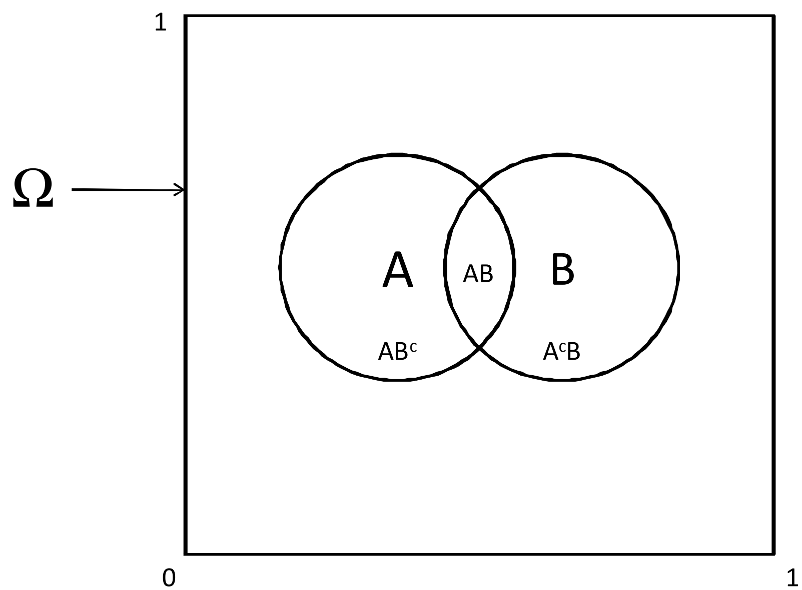
Rule: The probability of A and B is: $P(A \cup B) = P(A) + P(B) - P(AB)$

Axiom 3 tells us how to find the probability of a union of disjoint events in terms of their individual probabilities.

The Axioms can be used together to find a formula for the probability of a union of two events that are not necessarily disjoint in terms of the probability of each of the events and the probability of their intersection.

The union of two events, $A \cup B$, can be partitioned into three *disjoint* sets:

Elements of A that are not in B (denoted AB^c)
Elements of B that are not in A (denote A^cB)
Elements of both A and B (denoted AB)



Together, these three *disjoint* events are an exhaustive partition of $A \cup B$ so

$$A \cup B = AB^c \cup A^cB \cup AB$$

By axiom 3

$$P(A \cup B) = P(AB^c) + P(A^cB) + P(AB)$$

On the other hand,

$$\begin{aligned} P(A) &= P(AB^c \cup AB) = P(AB^c) + P(AB) \\ P(B) &= P(A^cB \cup AB) = P(A^cB) + P(AB) \end{aligned}$$

because AB^c and AB, and A^cB and AB are disjoint.

Add these up and you will find that

$$P(A) + P(B) = P(AB^c) + P(A^cB) + 2P(AB).$$

We already worked out that

$$P(A \cup B) = P(AB^c) + P(A^cB) + P(AB)$$

So the rhs of these are the same but $P(AB)$ is counted twice in the first expression. It follows that in general

$$P(A \cup B) = P(A) + P(B) - P(AB)$$

3. Bounds on Probabilities

Rule: The probability of A and B is weakly less than the sum of their probabilities: $P(A \cup B) \leq P(A) + P(B)$

It follows the previous rule and from the facts that Axiom 1 guarantees that $P(AB) \geq 0$ and that $P(A \cup B) = P(A) + P(B) - P(AB)$

Furthermore, taking a union cannot exclude outcomes, so $P(A \cup B) \geq P(A)$, and taking an intersection cannot include additional outcomes, so $P(AB) \leq P(A)$. Thus

$$0 \leq P(AB) \leq P(A) \leq P(A \cup B) \leq P(A) + P(B)$$

This result extents in the obvious way to families of events.

Conditional Probability.

Our assessment of probability often changes as we acquire more information and our next task is to formalise that idea.

On an abstract level the issue is that as we acquire more information **the event space $\mathcal{F}$ changes**.

Of course that's just a load of abstract nonsense (and you've seen enough of that already)

First, to get a feel for what I mean let's look at a couple of simple examples.

Example 1. You visit the home of who you know has two children. From the shoes in the hall you guess that at least one is a boy. What is the probability that your acquaintance has two boys?

Beforehand, the sample space was (in an obvious notation)

$$\{(b,b),(b,g),(g,b),(g,g)\}$$

but if there is at least one boy, we know that in fact the only possibilities are

$$\{(b,b),(b,g),(g,b)\}$$

All of these are about equally likely, so the answer to our question is about one third.

We write $\Pr[A|B]$ to indicate the probability of an event A given we known that event B has happened. This is called a **conditional probability**.

Example 2. Suppose that in a single roll of a fair die we know that the outcome is an even number. What is the probability that it is in fact a six?

Let $B = \{\text{result is even}\} = \{2, 4, 6\}$ and $C = \{\text{result is a six}\} = \{6\}$. Then $\Pr(B) = 1/2$ and $\Pr(C) = 1/6$

But *if I know* that B has happened, then $\Pr(C|B)$ is $1/3$ because given that B happened, we know the outcome was one of $\{2, 4, 6\}$ and **not** one out of $\{1, ..., 6\}$. Since the die is fair, in the absence of any other information, we assume each of these is equally likely.

Now let $A = \{\text{result is divisible by 3}\} = \{3, 6\}$.

If we *know* that B happened, then the only way that A can also happen is if the outcome is in $A \cap B$ i.e. if the outcome is $\{6\}$. The $\Pr(A \cap B) = 1/6$ and so $\Pr(A|B) = 1/3$ again which is $P(A \cap B) / P(B)$.

Bayes' Rule

$$\Pr(A|B) = \frac{\Pr(A \cap B)}{\Pr(B)}$$

We need B not to be impossible for this to make sense, i.e. $\Pr(B) > 0$

The intuition is as follows: since we know that B happens, the possible outcomes of our experiment are just the elements of B. i.e. **we change our sample space.**

Bayes' Rule is symmetric

$$\Pr(A|B) = \frac{\Pr(A \cap B)}{\Pr(B)} \text{ and } \Pr(B|A) = \frac{\Pr(B \cap A)}{\Pr(A)}$$

Since $\Pr(A \cap B) = \Pr(B \cap A)$ the probabilities that A happens given B happens and B happens given A happens are related:

$$\Pr(A|B)\Pr(B) = \Pr(B|A)\Pr(A)$$

$$\Pr(A|B) = \frac{\Pr(B|A)\Pr(A)}{\Pr(B)}$$

Example - die again. The events are:

$$\begin{aligned} A &: N < 3 \Rightarrow P(A) = 1/3 \\ B &: N \text{ is even} \Rightarrow P(B) = 1/2 \end{aligned}$$

$$\begin{aligned} \Pr(A|B) &= \frac{\Pr(A \cap B)}{\Pr(B)} = \frac{1/6}{1/2} = \frac{1}{3} \\ \Pr(B|A) &= \frac{\Pr(B \cap A)}{\Pr(A)} = \frac{1/6}{1/3} = \frac{1}{2} \end{aligned}$$

$$\begin{aligned} \Pr(A|B)\Pr(B) &= \frac{1}{3} \times \frac{1}{2} = \frac{1}{6} \\ \Pr(B|A)\Pr(A) &= \frac{1}{2} \times \frac{1}{3} = \frac{1}{6} \end{aligned}$$

Independence

Conditional probability gives us an important new concept - that of **independence.** Heuristically, two events are independent if knowing about one tells you **nothing** about the other. That is

$$\mathsf{Pr}(A|B) = \mathsf{Pr}(A)$$

Substituting in Bayes' rule it turns out that this is just the same as saying

$$\mathsf{Pr}\,(A \cap B) = \mathsf{Pr}(A)\,\mathsf{Pr}\,(B)$$

NB this is symmetric in A and B, so A is independent of B iff B is independent of A.

This notion of independence extends in an obvious way to families of events.

So we have defined a new kind of thing - a *conditional probability*.

Is it still a probability in the sense that it satisfies the Axioms?

"It turns out" that it is.

The key thing to remember about them (other than perhaps the formula for Bayes rule) is that they turn on **the change in the event space** which results from the new information.

This give an enormously important application of Bayes' Rule: a model of *learning*.

Bayesian Learning

Bayes' Rule shows how to update probabilities given evidence. If H denotes a hypothesis and E denotes some evidence then by Bayes' Rule:

$$\Pr(H|E) = \frac{\Pr(E|H)\Pr(H)}{\Pr(E)}$$

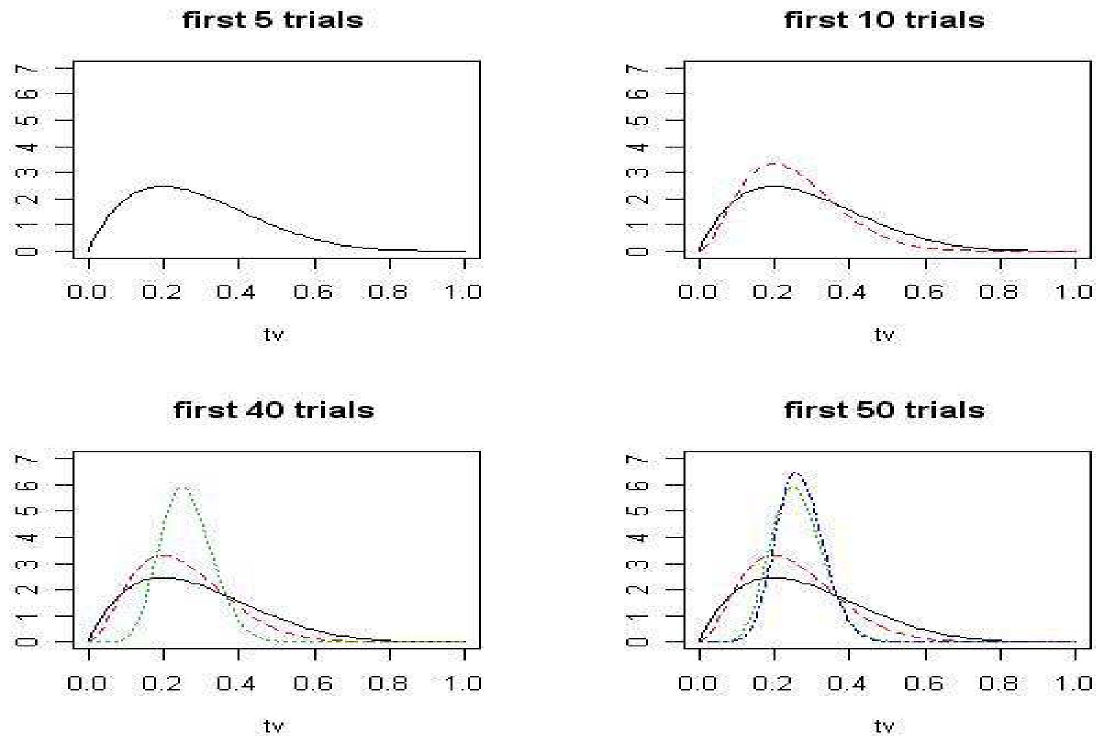
$P(H)$ is the *prior probability* of H that was formed before E, became available.

$P(E|H)$ is the *likelihood*. It's the conditional probability of seeing E if H is to be true.

$P(E)$ is the marginal probability of E: the unconditional probability of witnessing the new evidence.

$P(H|E)$ is called the *posterior probability* of H given E.

Example (from Tony Lancaster): The data from 50 coin-flips: T T H T T H T T T T T H T H H H T T T T H T H T T T H T T T T T T T T T T T H H T T T T H T H T T. Prior probability uniform on [0,1]



The Partition Rule

Since conditional probabilities are just like normal probabilities and satisfy the same axioms, we can use them to deduce some rules of conditional probability. The main one of interest is the partition rule also know as the **Law of Total Probability**.

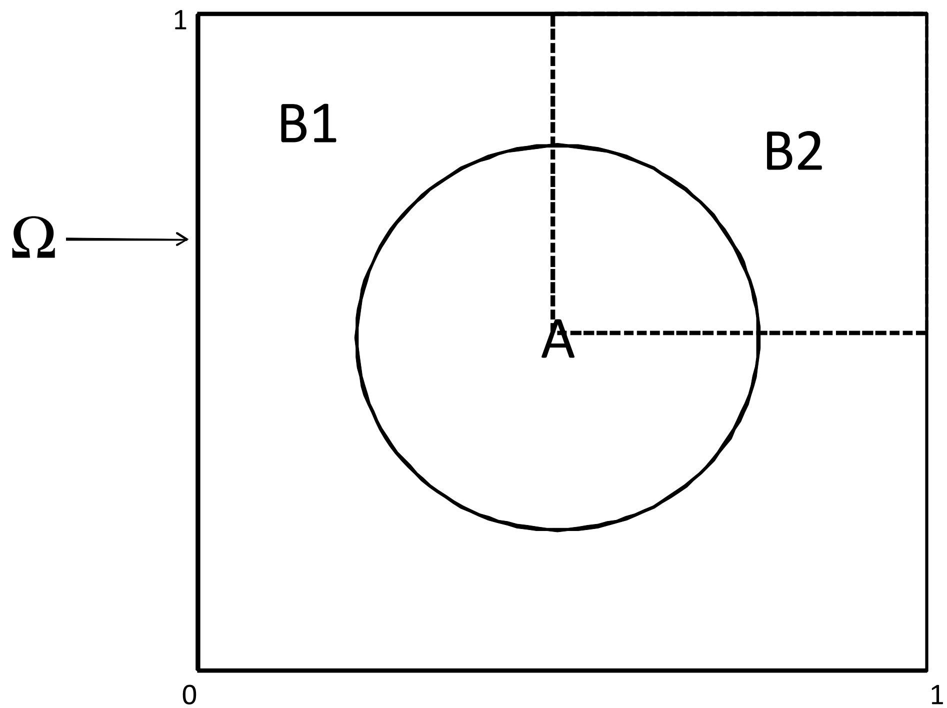
This says that given a partition $\{B_i\}_{i \in I}$ of Ω of events from $\mathcal{F}$, then for any event A

$$\mathsf{Pr}(A) = \sum_{i \in I} \mathsf{Pr}(A|B_i)\,\mathsf{Pr}(B_i)$$

E.g. for a two-partition of Ω into B_1 and B_2:

$$\begin{aligned} \mathsf{Pr}(A) &= \mathsf{Pr}(A \cap B_1) + \mathsf{Pr}(A \cap B_2) \\ &= \mathsf{Pr}(A|B_1)\,P(B_1) + \mathsf{Pr}(A|B_2)\,P(B_2) \end{aligned}$$

The Partition Rule



Bayes Theorem

If you combine the Partition Rule with *Bayes' Rule* you get *Bayes' Theorem* or the *Generalised Bayes' Rule* (you'll see by now the kind of beautiful/hideous* edifice which can quickly pile up on top of a few fairly innocuous axioms)

$$\mathsf{Pr}\left(B_i|A\right) = \frac{\mathsf{Pr}\left(A|B_i\right)\mathsf{Pr}\left(B_i\right)}{\sum_{j\in I}\mathsf{Pr}\left(A|B_j\right)\mathsf{Pr}\left(B_j\right)}$$

[* delete according to taste]

Example. Suppose that 15% of professional road cyclists take EPO. The blood test for EPO is imperfect so

$$\Pr(\text{Fails blood test} \mid \text{Dopes}) = 0.9$$

$$\Pr(\text{Fails blood test} \mid \text{Clean}) = 0.12$$

What is the probability that a cyclist who tests positive for EPO actually dopes?

$$\begin{aligned}
\Pr(\text{Dopes} \mid \text{Fails}) &= \frac{\Pr(\text{Fails} \mid \text{Dopes})\Pr(\text{Dopes})}{\Pr(\text{Fails} \mid \text{Dopes})\Pr(\text{Dopes}) + \Pr(\text{Fails} \mid \text{Clean})\Pr(\text{Clean})} \\
&= \frac{0.9 \times 0.15}{0.9 \times 0.15 + 0.12 \times 0.85} \approx 0.57
\end{aligned}$$

Maybe they are all innocent!