Notes: Practical Statistics for Physics & Astronomy

R.B. Metcalf

Alma Mater Studiorum - Universitá di Bologna

February 22, 2018

Contents

1 What is Probability?

1.1 Frequentist interpretation of probability

Imagine there is some event, instance or outcome of an experiment or observation called A. The probability of A is the fraction of times A occurs when the experiment or observation repeated in the same way or circumstances an *infinite* number of times.

$$P(A) = \lim_{N\to\infty} \frac{\text{number of trials where A is true}}{N(\text{total number of trials})} \tag{1.1}$$

This is the traditional definition of probability as formally stated by Laplace in 1774 and almost universally used for centuries despite no one ever having done anything *exactly* the same twice let along an *infinite* number of times.

Applying this definition to any physical phenomenon requires a partitioning of the world into things that are known and fixed on each repatriation of the observation and those things that are not known and change every repetition. If nature is deterministic and an experiment could be set up *exactly* the same way in all respects than the outcome would always be the same and probability would not apply. Of course even in classical physics it is not possible to know the state of every atom and photon that might possibly influence your measurement apparatus (or brain). It is these things that change when repeating the observation.

This partitioning between known and unknown factors seems reasonable when we talk about the positions and momenta of particles in a gas or flipping a coin, but in many other common situations where probability is used it seems less well defined. Say someone tells you that there is a 30% probability that candidate A will win an election tomorrow. Of course an identical election will never be run again and was never run in the past. There are many factors, known and unknown, that could affect an election. This statement was probably based on polling data. By the above definition of probability, this means that if the election were held an infinite number of times in which the polling data were exactly the same the candidate would win a 30% of them. This seems like a completely unverifiable claim. If scientific knowledge must be reproducible to be considered true then it would seem that any such argument should be considered unscientific. And yet probability through statistics is at the foundation of all quantitative measurements.

Lets be a bit more practical. Lets say we don't need an infinite number of trials, but just a very *large number* of them. Lets say we flip a coin a *very large number* of times. If we did it say one billion times we would not expect that *exactly* 500 million times it would be heads. We would expect that roughly half, but not exactly half of the times it would be heads even if the probability of getting heads in each flip is 1/2. We might try to quantify how close the number of heads should be to 500 million, but in doing so we would need to use a probabilistic argument that would use the very concept we are trying to define.

Many statisticians and philosophers have found this definition of probability problematic. Despite this it is the definition usually used by scientists when they are forced to addressing this subject.

1.2 classical interpretation of probability

The classical interpretation of probability relies on identifying events that are equally likely or probable. This is often the argumentation used in classical statistical mechanics where each microstate of the system is taken to be equally probable. If one then says that the probability of being in either of two mutually exclusive states is the sum of their probabilities and that the some of the probabilities of being in all possible states is one then you can find a numerical value for the probability of each state. A macro-state (one with temperature equal to some value or total energy equal to some value) corresponds to many micro-states so by adding up their probabilities you can find the probability of macro states which will not necessarily be equal.

The biggest criticism of this interpretation is that it doesn't really say what probability is, it just tells you how to calculate it in a restricted class of problems. What does it mean that two states are equally probable? What does the probability of a macro-state mean? Another problem is that not all events that we commonly apply probability to can reduced in this way to a collections of equally probable mutually exclusive events.

1.3 Subjective or Bayesian interpretation of probability

Thomas Bayes (1701 - 1761) (and Jacob Bernoulli 1655-1705) had a different conception of what probability is although the idea was not put on a firm theoretical foundation until the 1940's and 50's by G. Polya, R.T. Cox and E.T. Jaynes. It did not make its way into common use in science, in the form of Bayesian statistics, until relatively recently (80s and 90s).

In this school of thought, probability theory is an extension of formal logic to situations where the truth or falsehood of a proposition (e.g. "It will rain tomorrow." or "The mass of the Earth is between 5.972×10^{24} kg and 5.978×10^{24} kg.") cannot be deduced conclusively by deductive reasoning. A proposition has a probability function that depends on the evidence for and against its truth. When deductive reasoning can be applied conclusively this function is either zero (false) or one (true). In this way Boolean logic is a limiting case of probability theory. Surprisingly from just the following requirements (or *desiderata*) on the probability function of a proposition you can deduce the rules of probability and show that they are complete without ever mentioning randomness or repetition of experiments.

1. Degrees of plausibility are represented by real numbers.

2. The measure of plausibility must exhibit qualitative agreement with rationality. This means that as new information supporting the truth of a proposition is supplied, the number which represents the plausibility will increase continuously and monotonically. Also, to maintain rationality, the deductive limit must be obtained where appropriate.

3. Consistency

    (a) *Structured consistency* : If the conclusion can be reasoned out in more than one way, every possible way must lead to the same result.

    (b) *Propriety*: The theory must take account of all information that is relevant to the question.

    (c) *Jaynes consistency*: Equivalent states of knowledge must be represented by equivalent plausibility assignments. For example, if $A, B|C = B|C$, then the plausibility of $A, B|C$ must equal the plausibility of $B|C$

(taken from **?**).

These foundational proofs are very interesting, but outside the scope of this course (for those that are interested see chapter 2 of **?** or, more comprehensively, **?**). One thing that is of importance here is that this definition allows one to define the probability of something that would not usually be considered a *random variable* or a repeated even. It also establishes the accumulation of supporting evidence as central to the meaning of probability. Probability is a measure of knowledge, or ignorance, of an event and not a property of the event itself. These principles are central to the Bayesian method of parameter estimation and model selection that we will study later.

A	B	A, B	$\overline{A, B}$	$\overline{A} \cup \overline{B}$	$A \cup B$	$\overline{A \cup B}$	$\overline{A}, \overline{B}$
F	T	F	T	T	T	F	F
F	F	F	T	T	F	T	T
T	T	T	F	F	T	F	F
T	F	F	T	T	T	F	F

Table 1: The truth table for binary logical expressions.

1.4 Quantum mechanical probability

Probability in standard quantum mechanics is a fundamentally different thing than the probability that was in use before. In the frequentist interpretation of probability it is assumed that there are some "hidden variable" that are different every trial. In quantum mechanics it can be proven that such hidden variables do not exist or do not take on deterministic values for example with Bell's inequalities. When a measurement is made the square of the wave function gives the probability of an observation, but up to that point the outcome was impossible to determine, not just difficult to determine. This makes probability a property of physical systems and not solely a property of the observer. This seems to imply an intimate connection between physical laws and knowledge.

This is obviously a subject for a different course (or a *Star Trek* episode) so I will go no further.

1.5 the rules of probability

Suppose the A, B, ... are events that either occur or don't occur , that is they have values true or false (or 0 and 1 if you prefer). $P(A)$ is the probability of A occurring or being true. We can combine events in one of two ways. (A, B) means "A *and* B". It is true if both of them are true and false if both are false. $(A \cup B)$ means "A *or* B" it is true if either A or B is true. It is true if both are true. $\overline{A}$ means "not A". Note that $\overline{A \cup B} = \overline{A}, \overline{B}$ and $\overline{A, B} = \overline{A} \cup \overline{B}$ in the sense that there are no combinations of trues and falses for A and B that give different answers on either side of the equality. See table 1. In the language of Boolean algebra, they have the same truth table and are therefor equivalent statements. There probabilities must also be the same.

$P(A, B)$ is called the **joint probability** of events A and B. $P(A \cup B)$ is often called the **disjoint probability** of events A and B.

$P(A|B)$ is called a **conditional probability**. It means the probability of A *given* that B is true. You can imagine every probability being a conditional probability where it is "conditioned" on everything that you assume about the state of the Universe. Some of these things are assumed to be irrelavent and are left out . Some might be relevant but it is taken for granted so they are left out. The probability that a coin comes up heads does not depend on the time of day. It does depend on the assumption that it is a fair coin - no more likely to be heads than tails - although it might not always be stated. This is a simple example of a **statistical model** for the experiment, in this case flipping a coin.

The two fundamental rules of probability theory are

$$\begin{array}{cl} P(A, B) = P(A)P(B|A) & \quad \textbf{product rule} \\ P(A) + P(\overline{A}) = 1 & \quad \textbf{sum rule} \end{array} \tag{1.2}$$

These rules are actually derivable from some basic requirements or "desidariata" of how probabilities should behave, but for our purposes we can take them to be axioms. From these two rules and logic rules we can derive all the necessary properties of probability.

There are several particularly useful results that follow from these rules. From the logical requirement that (A, B) is the same as (B, A) and the product rule we get

$$P(A|B) = \frac{P(A)P(B|A)}{P(B)} \qquad \textbf{Bayes' theorem} \tag{1.3}$$

Applying the sum rule to $(A \cup B)$ gives

$$\begin{aligned}
P(A \cup B) &= 1 - P(\overline{A \cup B}) && (1.4)\\
&= 1 - P(\overline{A}, \overline{B}) && (1.5)\\
&= 1 - P(\overline{A})P(\overline{B}|\overline{A}) && (1.6)\\
&= 1 - P(\overline{A})\left[1 - P(B|\overline{A})\right] && (1.7)\\
&= 1 - P(\overline{A}) - P(\overline{A})P(B|\overline{A}) && (1.8)\\
&= P(A) + P(\overline{A})P(B|\overline{A}) && (1.9)\\
&= P(A) + P(\overline{A}, B) && (1.10)\\
&= P(A) + P(B)P(\overline{A}|B) && (1.11)\\
&= P(A) + P(B)\left[1 - P(A|B)\right] && (1.12)\\
&= P(A) + P(B) - P(B)P(A|B) && (1.13)\\
P(A \cup B) &= P(A) + P(B) - P(B, A) \qquad \textbf{extended sum rule} && (1.14)
\end{aligned}$$

In words, the disjoint probability of two events is equal to the sum of their probabilities minus their joint probability.

If A and B are **independent** then the probability of A occurring does not depend on whether B has occurred so $P(A|B) = P(A)$ through the product rule this implies $P(B|A) = P(B)$ and

$$P(A, B) = P(A)P(B) \qquad \text{independent events} \tag{1.15}$$

If two events are **mutually exclusive**, that is they cannot occur at the same time (the first flip of a coin cannot be both heads and tails) then $P(A, B) = 0$ and the extended sum rule becomes

$$P(A \cup B) = P(A) + P(B) \qquad \text{mutually exclusive events} \tag{1.16}$$

Example: If you roll a die once the probability of getting a 6 *or* a 5 is $\frac{1}{6} + \frac{1}{6} = \frac{1}{3}$. If you roll a die twice the probability of getting a 6 *and then* a 5 is $\left(\frac{1}{6}\right)\left(\frac{1}{6}\right) = \frac{1}{36}$. The probability of getting a 6 *and* a 5 is twice this because, $\frac{1}{18}$, because there are two ways of doing this, a 6 first or a 5 first.

This second case can be calculated in an alternative way. In the first roll we must get a 5 or a 6. We have calculated that the probability of this is $\frac{1}{3}$. Once this is done in the second roll we most get whichever number we didn't get in the first roll, one number out of 6, probability $\frac{1}{6}$. The probability of these two independent events happening is then given by the product rule $\left(\frac{1}{3}\right)\left(\frac{1}{6}\right) = \frac{1}{18}$.

Now say we have a set of observations $\{A_I\}$ that are all mutually exclusive and together they include all possible outcome then

$$\begin{aligned}
1 &= P(A_1 \cup A_2 \cup A_3 \cup \ldots |B) + P(\overline{A_1 \cup A_2 \cup A_3 \cup \ldots}|B) && (1.17)\\
&= P(A_1|B) + P(A_2 \cup A_3 \cup \ldots |B) + 0 && (1.18)\\
&= P(A_1|B) + P(A_2|B) + P(A_3 \cup \ldots |B) && (1.19)\\
&= \sum_i P(A_i|B) && (1.20)
\end{aligned}$$

This is the origin of the normalization requirement on any probability distribution function (PDF). Note that I have put a B in as a condition on all the probabilities, but this would hold without them.

Another important result along these lines is

$$\sum_i P(B|A_i)P(A_i) = \sum_i P(B, A_i) = \sum_i P(A_i|B)P(B) = P(B)\sum_i P(A_i|B) = P(B) \qquad (1.21)$$

with the same requirements on the set $\{A_i\}$. This is the origin of what we will later call marginalization.

2 Some warm up problems

There are a large class of problems, classical statistical physics included, for which individual states are considered equally probable and the question is how many states out of all possible states have a certain property. The property could be the temperature, pressure or having a full house in your poker hand and states could be the position each atoms in a gas, the spin state of each atom in a metal or the identity of the five cards you are dealt in poker. Here are some very simple problems that illustrate some of the counting techniques used throughout statistics.

2.1 Rolling Dice

Say we roll a die 10 times. Lets consider the following questions:

What is the probability of getting at least one 6? This is an "or" question - What is the probability of the first roll being 6 or the second one being six or ... Lets call the event that the ith roll is a 6 A_i. The sum rule (1.14) applies, but since these are not mutually exclusive events the sum rule 1.16 does not. These are independent events since the outcome of any one does not effect the outcome of any other. We could successive apply the extended sum rule (1.14 and the product rule (1.15) to $P(A_1 \cup A_2 \cup \cdots \cup A_{10})$ to break it down into $P(A_i)$'s which we know is 1/6. However, a quicker way to the answer is to realize that the probability of at least one being 6 is 1 minus the probability that non are 6. This follows from the logical requirement that $\overline{A_1 \cup A_2 \cup \cdots \cup A_{10}} = \overline{A_1}, \overline{A_2}, \ldots, \overline{A_{10}}$. Using the original sum rule (1.2) we get symbolically

$$\begin{aligned}
P(A_1 \cup A_2 \cup \cdots \cup A_{10}) &= 1 - P(\overline{A_1 \cup A_2 \cup \cdots \cup A_{10}}) && (2.1)\\
&= 1 - P(\overline{A_1}, \overline{A_2}, \ldots, \overline{A_{10}}) && (2.2)\\
&= 1 - P(\overline{A_1})P(\overline{A_2}) \ldots P(\overline{A_{10}}) && (2.3)\\
&= 1 - P(\overline{A})^{10} && (2.4)\\
&= 1 - \left(\frac{5}{6}\right)^{10} && (2.5)\\
&= 0.838\ldots && (2.6)
\end{aligned}$$

We could also solve this problem by counting. How many combinations of rolls are there? The first roll has 6 possibilities, the second one 6, etc. so there are 6^{10} combinations. There are 5^{10} combinations with no 6s. So the fraction of the cases that have no 6s is $\left(\frac{5}{6}\right)^{10}$ so the probability of having 1 or more is $\left(\frac{5}{6}\right)^{10}$.

What is the probability of getting exactly one 6? Lets first try to solve this problem by pure symbolic logic and the rules of probability. The proposition could be stated as roll one is a 6 *and* all the others are not *or* roll two is a 6 *and* all the other are not *or* etc. Symbolically this is represented as

$$B_1 = (A_1, \overline{A_2}, .., \overline{A_{10}}) \cup (\overline{A_1}, A_2, .., \overline{A_{10}}) \cup \cdots \cup (\overline{A_1}, \overline{A_2}, .., A_{10}) \tag{2.7}$$

Each of the propositions in the parenthesis are mutually exclusive so the sum rule (1.16) can be applied to B_1 to break it up into a sum

$$P(B_1) = P(A_1, \overline{A_2}, .., \overline{A_{10}}) + P(\overline{A_1}, A_2, .., \overline{A_{10}}) + \ldots \tag{2.8}$$

Since each of the rolls are identical, the probabilities for reach of situation must be the same and each term must be the same

$$P(B_1) = 10P(A_1, \overline{A_2}, .., \overline{A_{10}}) \tag{2.9}$$
$$= 10P(A_1)P(\overline{A_2}, .., \overline{A_{10}}) \tag{2.10}$$
$$= 10\left(\frac{1}{6}\right)\left(\frac{5}{6}\right)^9 \tag{2.11}$$
$$= 0.323\ldots \tag{2.12}$$

where we use the same logic that got us from equation (2.2) to line (2.5) in the previous problem.

Now lets do this again by counting. There are 6^{10} possible combinations. If one roll is a 6 the other nine need to be less than 6. There are 5^9 combinations of nine numbers between 1 and 5. The 6 can come up on any of 10 rolls so there are in total 10^9 ways of rolling 10 times and getting one 6.

What is the probability of getting exactly four 6s? This can be confusing, but if we just look at it from a symbolic point of view we can avoid some common misunderstandings. Here we must find all the combinations of four As and six $\overline{A}$. The first A can go in one of ten slots and the second in one of the remaining 9, etc. giving $10 \times 9 \times 8 \times 7 = 10!/(10-4)!$. We have over counted here though because the order in which we place the As in the slots should not matter, it gives the same logical statement. How many orderings are there? For each selection of 4 slots there are four choices for the first one, three choices etc. - 4! orderings or **permutations**. So there are $\frac{10!}{4!(10-4)!}$ ways of having four As and six $\overline{A}$. The probability of all these combinations are equal and mutually exclusive (a roll cannot be both A and $\overline{A}$) so we can add their probabilities

$$P(B_4) = \frac{10!}{4!(10-4)!}P(A_1, A_2, A_3, A_4, \overline{A_5}, .., \overline{A_{10}}) \tag{2.13}$$
$$= \frac{10!}{4!(10-4)!}P(A_1, A_2, A_3, A_4)P(\overline{A_5}, .., \overline{A_{10}}) \tag{2.14}$$
$$= \frac{10!}{4!(10-4)!}P(A)^4 P(\overline{A})^6 \tag{2.15}$$
$$= \frac{10!}{4!(10-4)!}\left(\frac{1}{6}\right)^4\left(\frac{5}{6}\right)^6 \tag{2.16}$$
$$= 0.05\ldots \tag{2.17}$$

A confusion with this problem often arises because it is often stated or implied that all the permutations of the 6s must be considered one combination because they are indistinguishable. This might lead one to consider any two repeated numbers that are not 6s as indistinguishable and try not to over count them. This quickly becomes a very complex calculation. Although it is true that the 6s are indistinguishable this misses the point. For the purposes of this problem each roll has a binary outcome. It is either a 6 or not a 6. 6s are indistinguishable, but so are not 6s. We could have considered a different problem – "What is the probability of getting 4 rolls that are more than 4?". The calculation would be exactly the same except that the probabilities $P(A)$ and $P(\overline{A})$ would be different, $\frac{1}{3}$ and $\frac{2}{3}$ instead of $\frac{1}{6}$ and $\frac{5}{6}$.

These dice throwing problems are a special case of the **binomial distribution** which we will discuss later in more detail.

2.2 Birthday Paradox

This is another widely known problem for which many people go down the wrong path and get confused. The "paradox" is that in a relatively small group of people there is a surprisingly high

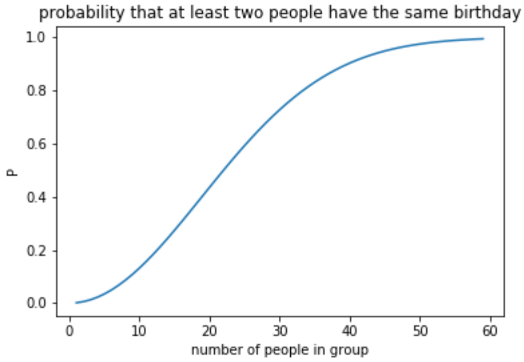

Figure 1: Probability of more than one person having the same birthday.

probability that two of them will have the same birthday.

Lets say there are n people at the party. There are 356 choices for the birthday of each person (not including leap years) so there are 356^n combinations of n birthdays. We will assume these are all equally likely. Instead of finding the number of combinations with repeat birthdays lets find the number of combinations with no repeats. There are 356 choices for the first person, then 355 choices for the second etc. until you get to the last person so the number of cases with no repeats is $356 \times 354 \times ... \times (356 - n + 1) = 356!/(356 - n)!$. So the total probability is

$$P(\text{at least two the same}) = 1 - P(\text{no two the same}) = 1 - \frac{356!}{356^n(356 - n)!}. \tag{2.18}$$

If you try to calculate this number in your directly with a computer you will find that some of these numbers are too big to store. The scipy factorial function (scipy.special.factorial) will give infinity for 356 for example. But the quotient of these numbers is something reasonable. This problem often comes up in this kind of problem. We will need an approximation to complete the calculation. Taking the log of a quotient often helps you cancel some things out. And taking Stirling's approximation ($\ln N! = N \ln N - N$) often helps simplify factorials.

$$\begin{aligned}
\ln\left(\frac{N!}{N^n(N-n)!}\right) &= \ln N! - \ln(N-n)! - n \ln N && (2.19)\\
&= N \ln N - N - (N-n)\ln(N-n) - (N-n) - n \ln N && (2.20)\\
&= (N-n)\ln N - (N-n)\ln(N-n) - n && (2.21)\\
&= (N-n)\ln\left(\frac{N}{(N-n)}\right) - n && (2.22)
\end{aligned}$$

We can then take the exponential of this to get

$$P(\text{at least two the same}) \simeq 1 - \left(\frac{N}{N-n}\right)^{N-n} e^{-n} \tag{2.23}$$

This is plotted in figure 1. For a group of 23 people there is a 50% chance that at least 2 of them will have the same birthday.

2.3 Poker

A deck of poker cards consists of 52 cards. There are four suits - diamonds ($\diamondsuit$), hearts ($\heartsuit$), spades ($\spadesuit$) and clubs ($\clubsuit$). In each suit there are an ordered sequence of 13 cards (we will take the ace to be greater than the king). A poker hand consists of 5 cards. In "five card stud" you are dealt five cards and you are not allowed to exchange any. This version of poker is almost never played because it relies too much on chance and not skill, but we will consider it here because it is simple.

What is the probability of getting a flush (five cards of the same suit) in five card stud? You might at first think this is just like the dice rolling problem and say it is $4(1/4)^5 \simeq 0.0039$, but this would be wrong because the draws are not independent. If your first card is a $\clubsuit$ there will be fewer $\clubsuit$ in the deck and the deck will be smaller so the probability of getting a club the second time will be $(13-1)/(52-1)$.

$$P(\text{flush}) = \frac{4}{4}\frac{12}{51}\frac{11}{50}\frac{10}{49}\frac{9}{48} = 0.00198\ldots \tag{2.24}$$

Significantly less probable than we would get if there where replacement.

What is the probability of a straight? This is getting five sequential cards, for example 8, 9, 10 ,J ,Q. The probability of drawing them all in a row must be the same as the probability of drawing them in any other order so we can calculate the probability of drawing them in order and then multiply by the number of permutations. First we need to draw a card below of 10 or lower or there wont be enough cards of higher value. That probability is $4 \times 9/52$. Then there are 4 cards of one higher value out of 51 remaining cards, etc.. Then for each case there are 5! permutations.

$$P(\text{straight}) = 5!\frac{36}{52}\frac{4}{51}\frac{4}{50}\frac{4}{49}\frac{4}{48} = 0.003546\ldots \tag{2.25}$$

Somewhat more likely than a flush which is why this hand is worth less. If we count the ace-low straight this is 0.00394.... This includes straight-flushes and royal-straight-flushes which are actually higher hands.

What is the probability of a full house? A full house is two of a kind (two 10's or two kings for example) and three of another kind (three aces or three twos).

Lets do this one a little differently. Lets count the total number of distinct five card hands and then count the number of distinct full houses. The probability will be the ratio of these since every hand is equally probable. Lets make this a little more abstract. There are N distinct objects (cards) we have N ways of choosing the first one. There are $N-1$ objects left when we pick the next one, etc. So there are $N \cdot (N-1) ... \cdot (N-n+1)$ distinct ways of choosing n objects out of N. This can also be written $N!/(N-n)!$. This counts combinations of objects in different orders as distinct (123 is different than 213). If we wish to count different permutations of the same objects as the same set then we need to divide by the number of permutations of n objects which is $n!$. So the number of these distinct sets is

$$\binom{N}{n} \equiv \frac{N!}{n!(N-n)!} \tag{2.26}$$

This is the **binomial coefficient**. In English this is often spoken as "N choose n." for obvious reasons. Lets use it on our problem.

There are $\binom{52}{5}$ distinct five card hands. There are four cards of each type, one for each suit, so there are $13 \cdot \binom{4}{2}$ distinct pairs of cards of the same kind. The three of a kind need to be different than the pair so there are $12 \cdot \binom{4}{3}$ of them. So the probability of a full house is

$$P(\text{full house}) = \frac{\binom{4}{2} \cdot \binom{4}{3} \cdot 13 \cdot 12}{\binom{52}{5}} = 0.00144\ldots \tag{2.27}$$

Very similar logic will lead you to the probabilities of getting two pair or four of a kind.

Calculating the probabilities for poker may seem frivolous, but the calculation of odds for gambling actually played a very important role in the development of statistics. Pascal and Fermat had a long correspondence in the 17th century in which they developed basic probability theory.

2.4 The Monty Hall Problem

This is a classic problem based on an old American TV game show. It was before my time, but apparently the host of the show was named Monty Hall. There are variations of this game show on Italian TV also. In this game the contestant is can choose between three doors. He knows that behind one of the doors is something nice like a new car and behind the other two are things that are not so nice like a chicken or an old shoe. The contestant chooses one door and then Monty eliminates one of the doors that were not chosen and shows that it has the shoe. The contestant then has a chance to change his choice or remain with his first choice. What should he do? Does it matter?

A Matrix basics

$$(\boldsymbol{ABC}\ldots)^T = \ldots \boldsymbol{C}^T\boldsymbol{B}^T\boldsymbol{A}^T \tag{A.1}$$
$$(\boldsymbol{ABC}\ldots)^{-1} = \ldots \boldsymbol{C}^{-1}\boldsymbol{B}^{-1}\boldsymbol{A}^{-1} \tag{A.2}$$
$$(\boldsymbol{A}^T)^{-1} = (\boldsymbol{A}^{-1})^T \tag{A.3}$$
$$(\boldsymbol{A}+\boldsymbol{B})^T = \boldsymbol{A}^T + \boldsymbol{B}^T \tag{A.4}$$

Some properties of the determinant

$$|\boldsymbol{A}| = \prod_i \lambda_i \tag{A.5}$$
$$|\boldsymbol{A}^{-1}| = 1/|\boldsymbol{A}| \tag{A.6}$$
$$|\boldsymbol{BA}| = |\boldsymbol{B}||\boldsymbol{A}| \tag{A.7}$$
$$|c\boldsymbol{A}| = c^n|\boldsymbol{A}| \tag{A.8}$$
$$|\boldsymbol{A}^T| = |\boldsymbol{A}| \tag{A.9}$$

Some properties of the trace

$$\mathrm{tr}(\boldsymbol{A}) = \sum_i A_{ii} \tag{A.10}$$
$$\mathrm{tr}(\boldsymbol{A}) = \sum_i \lambda_{ii} \tag{A.11}$$
$$\mathrm{tr}(\boldsymbol{A}^T) = \mathrm{tr}(\boldsymbol{A}) \tag{A.12}$$
$$\mathrm{tr}(\boldsymbol{AB}) = \mathrm{tr}(\boldsymbol{BA}) \tag{A.13}$$
$$\mathrm{tr}(\boldsymbol{A}+\boldsymbol{B}) = \mathrm{tr}(\boldsymbol{A}) + \mathrm{tr}(\boldsymbol{B}) \tag{A.14}$$

$\boldsymbol{A}$ is an **orthogonal matrix** if and only if

$$\boldsymbol{A}^T\boldsymbol{A} = \boldsymbol{A}\boldsymbol{A}^T = \boldsymbol{I} \tag{A.15}$$

An orthogonal matrix has the following properties

$$\boldsymbol{A}^T = \boldsymbol{A}^{-1} \tag{A.16}$$
$$|\boldsymbol{A}| = \pm 1 \tag{A.17}$$

The $|\lambda_i| = 1$ for all eigenvalues and the magnitude of all eigenvactors are 1.

$\boldsymbol{C}$ is a **positive definite matrix** if

$$\boldsymbol{x}^T\boldsymbol{C}\boldsymbol{x} > 0 \quad \forall \boldsymbol{x}. \tag{A.18}$$

It has the following properties

- all eigenvalues at positive
- $\mathrm{tr}(\boldsymbol{C}) > 0$
- all diagonal elements are positive, $\boldsymbol{C}_{ii} > 0, \forall i$
- $\boldsymbol{C}$ is invertible

The covariance matrix is always positive definite.

"A and B"	A, B
"A or B"	$A \cup B$
continuous random variables	x,y,x_i,y_i
vector of random variables	$\boldsymbol{x}$ or $\vec{x}$
discrete random numbers	n,m
parameters	α , β
estimator of parameter α	θ_α or $\hat{\alpha}$
data	D or d_i
indexes data or for multiple random numbers	i, j
statistical and/or theoretical model	M
Gaussian or Normal pdf	$\mathcal{G}\left(\boldsymbol{x}\,\vert\boldsymbol{\mu},\boldsymbol{C}\,\right)$
$\boldsymbol{x}$ is normally distributed	$\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{\mu},\boldsymbol{\sigma})$
x is χ^2 distributed with n degrees of freedom	$x \sim \chi^2_n$
arithmetic mean of N samples	$\bar{x}_N$
likelihood of data given model	$\mathcal{L}(\boldsymbol{D}\vert M_i)$ or $P(\boldsymbol{D}\vert M_i)$
Bayesian evidence of data	$\mathcal{E}(\boldsymbol{D})$
Heaviside function, 1 when B is true, 0 otherwise	$\Theta(B)$

Table 2: notation

B notation

Notation may vary but in general I will follow the guide in table 2

C Some Useful Integrals and mathematical definitions

$$\int_{-\infty}^{\infty} dx\; e^{-\frac{x^2}{2}} = \sqrt{2\pi} \tag{C.1}$$

$$\begin{aligned}\int_{-\infty}^{\infty} dx\; e^{-(ax^2+bx+c)} &= e^{-c}\int_{-\infty}^{\infty} dx\; e^{-\left(\sqrt{a}x+\frac{b}{2\sqrt{a}}\right)^2+\frac{b^2}{4a}} = e^{-c+\frac{b^2}{4a}}\int_{-\infty}^{\infty}\frac{dy}{\sqrt{a}}\; e^{y^2} \\ &= \sqrt{\frac{\pi}{a}}e^{-c+\frac{b^2}{4a}}\end{aligned} \tag{C.2}$$

$$\int_0^\infty dx\; x^n e^{-\frac{1}{2}Ax^2} = 2^{\frac{n-1}{2}}A^{-\frac{n+1}{2}}\Gamma\left(\frac{n+1}{2}\right) \qquad n > -1 \tag{C.3}$$

The Gamma function

$$\begin{array}{lll}\int_0^\infty dx\; x^n e^{x^2} & = \frac{1}{2}\Gamma\left(\frac{n+1}{2}\right) & \\ \Gamma(n) & = (n-1)! & n = 1, 2, \ldots \\ \Gamma\left(\frac{1}{2}+n\right) & = \frac{(2n)!}{4^n n!}\sqrt{\pi} & n = 0, 1, 2, \ldots\end{array} \tag{C.4}$$

Stirling's approximation

$$\ln N! \simeq N\ln N - N \text{ for } N \gg 1 \tag{C.5}$$

or more accurately

$$N! \simeq \sqrt{2\pi N}\left(\frac{N}{e}\right)^N \text{ for } N \gg 1 \tag{C.6}$$

Index