Lecture Notes

ENEE 324 – Engineering Probability

P.S. Krishnaprasad

Fall, 2002

PROBABILITY: a bit of etymology

Wahr -true, genuine
Wahrhaft - truthful, true
Wahrhaftigkeit - truthfulness
Wahrheit - truth, fact
Wahrnehmen - to perceive, observe, notice
Wahrnehmung - perception
Wahrsagen - to prophesy, tell fortunes
Wahrsager -
Wahrsagerin - soothsayer, fortune-teller
Wahrsagung - prophecy
Wahrscheinlich - probable, probably
Wahrscheinlichkeit - probability
Wahrscheinspruch - verdict
Wahrscheinlichkeitstheorie - probability theory

– Collins Gem German-English Dictionary

ENEE 324 Engineering Probability Lecture 1

Basic Concepts

This is a course on *modeling uncertainty*. Uncertainty is all about us – the outcome of a coin toss, the life-time of an electric light bulb, Nyquist-Johnson fluctuations in the measured value of current through a resistor connected to a heat bath, the chance of rain at noon on a weekday, are all valid examples of uncertain/chance/random phenomena. Yet, through study of specific contexts, and by carrying out carefully repeated experiments, it is possible to get a handle on uncertainty, sufficiently to be able to make useful predictions. A great deal of science and engineering is concerned with making predictions in the face of uncertainty. Probability theory provides the language, the techniques, and as a consequence the mathematical models that enable us to do this.

There are other ways to approach uncertainty, but probability theory is quite possibly the most wide-ranging and successful means to do this. Probability theory offers a coherent conceptual system to understand and cope with uncertainty.

Modern technology makes extensive use of probability theory. Some examples include: (a) algorithms used to route messages/data in a communication/computing network; (b) techniques used to project the yield in acceptable quality silicon wafers in a semiconductor manufacturing plant; (c) the error-correcting codes used in compact disc players; (d) performance analysis and design of a service system using the theory of queues (waiting lines).

Everyday use of the language of probability is based on built-up intuition that people have. Sometimes such intuition can prove unreliable or ill-defined. One can build correct intuition by solving certain "toy problems", such as card-shuffling. It is useful and advisable to develop a systematic approach to probability. In particular, the models of probability have to be tested for "consistency" against data (observed in experiments).

Often, costly and sensitive decision-making processes depend on probability models. Some examples: (a) the decision by a "wild-catter" to drill or not to drill for oil in a particular parcel of optioned land; (b) the decision to launch a space-shuttle based on forecasts of weather patterns; (c) the decision to attempt circum-navigation of the globe in a hot-air balloon; (d) the decision to attempt maiden voyage of a grand ocean liner in sea-lanes known to be populated by ice-bergs. The *risks* involved in such decision processes must be quantified so that an experienced and competent human can make *rational* choices. Lloyd's of London quantifies such risks all the time. How? The answer lies in probabilistic concepts.

Probability can also be used to answer (approximately) questions in fields where one normally does not expect to have to deal with uncertainty. An example of this is the *Buffon's needle* problem: suppose a needle is "tossed at random" onto a plane ruled with parallel lines a distance L apart, where by a "needle" we mean a line segment of length $l \leq L$. What is the probability of the needle intersecting one of the parallel lines?

We present a systematic approach to this important subject by beginning with fundamental concepts.

Very often, one thinks of a problem involving uncertainty as being associated to an *experiment* $\mathcal{E}$. If $\mathcal{E}$ is repeatable, so much the better. There is a whole school of thought, that insists on attaching probabilities only to repeatable experiments, known as *the frequentists*. Yet, there are problems involving uncertainty where no natural experiments can be suggested to model or deduce the uncertainty. For instance, despite having a large body of solid geophysical knowledge and experience, a geophysicist, when called upon to offer what he/she thinks of as the "likelihood" of a cataclysmic earthquake on the eastern sea-board by the year 2000, may appear to "pick a percentage out of the hat". What is going on here is that the number offered is a measure of the scientist's conviction – an example of *subjective* probability. (There is a history of raging arguments between subjectivists and frequentists. After all, is it not the goal of science to be objective and stamp out all that carries the taint of prejudice/subjectivity? We will meet on our journey, representatives of both camps,–Richard von Mises, Bruno de Finetti, John Maynard Keynes, Leonard Savage, Ronald A. Fisher,...). Whether the probabilities that we discuss below are based on (repeatable) experiments or based on an expert's conviction, the rules for working with probabilities are the same. These rules serve as a foundation for mathematical modeling of uncertainty. At a fundamental level, these are based on the language of *set theory* and *Boolean algebra*.

First, we need a set Ω, called the *sample space*. The elements of this set are the (exhaustive list of) possible *outcomes* of an experiment $\mathcal{E}$. With reference to $\mathcal{E}$, we will have the notion that Ω is a universal set, i.e., all possible outcomes of $\mathcal{E}$ are accounted for in Ω.

Examples

(i) $\mathcal{E}$ = single coin toss, $\Omega = \{H, T\}$

(ii) $\mathcal{E}$ = roll of a single die, $\Omega = \{1, 2, 3, 4, 5, 6\}$

(iii) $\mathcal{E}$ = coin toss until a first head, $\Omega = \{H, TH, TTH, ...\}$

(iv) $\mathcal{E}$ = mark a random dot on a ruler of length L. Here we take $\Omega = [0, L]$.
(Note, this is an easy experiment to repeat and there are different ways to repeat it, either via independent trials or dependent trials.)

(v) $\mathcal{E}$ = survey of all computers that are up or down at 11:00 a.m. Here Ω can be taken as simply a list of all IP addresses with a tag UP or DOWN

In example (i) and (ii) the sample space Ω is a finite set. In (v) it is finite but large (in Maryland campus)! In (iii) it is countably infinite. In (iv) it is uncountably infinite. In the beginning we will concentrate on situations wherein Ω is finite or countably infinite (a discrete sample space).

An *event* A (associated to an experiment) is simply a set of possible outcomes, i.e. a subset of Ω. The collection of all possible events is denoted as 2^{Ω} and is called the power set of Ω.

Examples (associated to above experiments)

(i) head occurs: $A = \{H\}$

(ii) even number occurs: $A = \{2, 4, 6\}$

(iii) first head occurs in at most 3 tosses: $A = \{H, TH, TTH\}$

(iv) mark within halfway point: $A = [0, 0.5L]$
(v) only one computer is down: $A = \{u_1\, d_2\, u_3, d_1\, u_2\, u_3, ...\}$

Events cannot be discussed in isolation. Thus if the event A occured, then event A^c, the complement of A, did not occur. Thus we are also thinking about A^c even as we speak of A. (Remark: We also denote A^c as $\bar{A}$.) In fact we are thinking about a whole algebra of events constructed out of the operations of set intersection and set union, respectively mirroring the logical connectives AND and OR.

We state below, the elements of set theory relevant to probablity calculations: **A set is a collection of objects.**

The set of outcomes of rolling a die, $\Omega = \{1, 2, 3, 4, 5, 6\}$.
For each experiment $\mathcal{E}$ we need to define Ω. $\emptyset$ denotes the empty set.

(1) $A \sqsubset B$ means A is a subset of B. Then, $a \in A \Rightarrow a \in B$
(2) $A \cup B = C$ means $c \in C \Rightarrow c \in A$ <u>or</u> $c \in B$ (or both).
(3) $A \cap B = C$ means $c \in C \Rightarrow c \in A$ <u>and</u> $c \in B$.
(4) $\bar{A} = C$ means $c \in C \Rightarrow c \notin A$.
(5) $A \times B = C$ denotes the Cartesian product

The cartesian product of sets means $c \in C \Leftrightarrow c = (a, b)$ where $a \in A$ $b \in B$. Note that c is an <u>ordered</u> pair.

Using these basic operations, one builds more "complicated" events from elementary events. Given an experiment $\mathcal{E}$ with sample space Ω, any member of 2^Ω could be an event, in principle. In practice, one may limit oneself to a subcollection $\mathcal{A} \subseteq 2^\Omega$.

How to choose $\mathcal{A}$?

Basic ground rules for $\mathcal{A}$ (=Boolean algebra)

$A \in \mathcal{A} \Rightarrow \bar{A} \in \mathcal{A}$
$\Omega \in \mathcal{A}$
$\emptyset \in \mathcal{A}$
A, $B \in \mathcal{A}$ then $\quad A \cup B \in \mathcal{A}, \quad A \cap B \in \mathcal{A}$
We think of $\mathcal{A}$ as a collection of *interesting events*.

Example:

$\Omega = \{1, 2, 3, 4, 5, 6\}$
$\mathcal{A} = \{\emptyset, \Omega, A_1, A_2\}$, where $A_1 = \{1, 3, 5\}, \quad A_2 = \{2, 4, 6\}$.

We define probability in a manner that *agrees* with experimental observations – mimics frequencies, and consistently for all $A \in \mathcal{A}$.

Definition: Relative frequency of event A,
$f_A \stackrel{\triangle}{=} \frac{n_A}{n}$ where $n = \#$ repetitions/trials of $\mathcal{E}$ and $n_A = \#$ occurrences of A in n trials.
Check Properties

(i) $0 \leq f_A \leq 1$
(ii) $f_A = 1$ *iff* A occurs every time in the n trials/repetitions. In particular $f_\Omega = 1$.
(iii) $f_A = 0$ *iff* A never occurs in the n trials. In particular $f_\emptyset = 0$.
(iv) If A and B are disjoint, i.e. $A \cap B = \emptyset \Rightarrow f_{A \cup B} = f_A + f_B$. In particular $f_A = 1 - f_{\bar{A}}$.
(v) As $n \to \infty$, $f_A(n) \to P(A)$ **(??)**

For probability, turn these properties into **axioms**.
Given an experiment $\mathcal{E}$, sample space Ω, and collection of *interesting events* $\mathcal{A}$, a probability law or probability measure is a function, (Here the term *measure* used in the same way as a measure of length, or area, or volume.)

$P : \mathcal{A} \to [0, 1]$, satisfying
(a) $0 \leqq P(A) \leqq 1$
(b) $P(\Omega) = 1$
(c) $A \cap B = \emptyset \Rightarrow \quad P(A \cup B) = P(A) + P(B)$ (addition rule).

Prove that

$P(\emptyset) = 0$
$P(A) = 1 - P(\bar{A})$

Some basic properties of probabilities

(1) $A \subseteq B \Rightarrow P(A) \leq P(B)$
Proof:
Let $C = \{y \in B : y \notin A\}$
Then $B = C \cup A$, and $C \cap A = \emptyset$
Thus: $P(B) = P(C \cup A) = P(C) + P(A)$
Since $P(C) \geq 0$, the result follows. □

(2) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
Proof:
$A \cup B = (A \cap B) \cup (A \cap \bar{B}) \cup (\bar{A} \cap B)$ is a decomposition into disjoint sets. By the addition of axiom of probability,

$$\begin{aligned} P(A \cup B) &= P(A \cap B) + P(A \cap \bar{B}) + P(\bar{A} \cap B) \\ &= \Big(P(A \cap B) + P(A \cap \bar{B})\Big) \\ &\quad + \Big(P(\bar{A} \cap B) + P(A \cap B)\Big) - P(A \cap B) \\ &= P\Big((A \cap B) \cup (A \cap \bar{B})\Big) \\ &\quad + P\Big((\bar{A} \cap B) \cup (A \cap B)\Big) - P(A \cap B) \end{aligned}$$

(by the addition axiom)

$$\begin{aligned} &= P\Big(A \cap (B \cup \bar{B})\Big) + P\Big((\bar{A} \cup A) \cap B\Big) - P(A \cap B) \\ &= P(A \cap \Omega) + P(\Omega \cap B) - P(A \cap B) \\ &= P(A) + P(B) - P(A \cap B) \qquad \square \end{aligned}$$

In the above, we have made use of the distributive law.

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$

One builds probability laws by recognizing what the equally *likely* events are in a given experiment. The idea of equally likely outcomes draws on symmetry. No special status is given to any particular outcome. One then applies the axioms. Finite sample space problems are key to building intuition.
What are the elementary events in experiment (iv) above? Assuming they are equally likely, what is their common probability?