Remarks on Probability

In order to better understand theorems on "average" performance analyses, it is helpful to know a little about probability and random variables.

This will also help in understanding how to design simulation experiments.

The material is purposely simplified; no lies, but not the complete truth either. We will restrict it to *discrete* probability and random variables.

Events

A probability space (for our limited purpose) is a set of events. To each event E we assign a probability $P(E)$ ranging between 0 and 1. We can combine events in standard set-theoretic ways, viz., $E_1 \cup E_2$, $E_1 \cap E_2$, $\bar{E}$ (indeed we can even admit countable unions, etc.). So, events form an algebra of sets. Probability functions must obey some simple properties that are quite intuitive, e.g.,
$P(A \cup B) = P(A) + P(B) - P(A \cap B)$, $P(\bar{A}) = 1 - P(A)$.

$E_1 \cup E_2$ is event of either E_1 or E_2 (occuring), $E_1 \cap E_2$ is that of both E_1 and E_2, and $\bar{E}$ is the event that is possibility other than E.

What is a probability?

This is a difficult question, fraught with philosophical controversy. We skirt around it by simply taking a view that is helpful in simulation contexts, without claiming that it is the *truth*.

This can be called the *Frequentist* view. To make the interpretation concrete, consider the event space of tossing a die A. This has six outcomes, which are the events; we name them 1, 2, 3, 4, 5, 6.

What do we mean when we say, e.g., $P(5) = 1/6$?

It is that the "chance" of getting a 5 in a toss is 1/6. But what does that "really" mean?

A Gedanken Experiment Answer

The frequentist answer is an *operational* one.

Intuitively, the frequentist view is to imagine *repeating* the tossing experiment many times and define:

$S_0 = 0$; if at the i-th toss we get a 5, set $S_i = S_{i-1} + 1$, otherwise $S_i = S_{i-1}$.

This S_n is a count of how many 5's have occurred up to the n-th toss.

Then consider the ratio:

$$\frac{S_n}{n} \tag{1}$$

The limit interpretation

As the number n of tosses increases, intuitively we expect the ratio $\frac{S_n}{n}$ to gradually converge to some number.

The frequentist interpretation is that this limit is the intended probability of the event “5 shows up in a toss”; i.e.

$$P(5) = lim_{n \to \infty} \frac{S_n}{n} \tag{2}$$

A *fair dice* would be one for which all the events 1, 2, 3, 4, 5 and 6 have this same limit, i.e., $P(i) = 1/6$ for each of them.

Probabilities are OK

The frequentist interpretation is consistent with the usual properties desired of a probability measure. Let us just check out one.

$P(\bar{A}) = 1 - P(A)$: For this, call each occurrence of the event A a "success" in a long experiment of repeated trials. Define S_i as before but use it to count the number of successes. Then T_i is used to count the number of failures (i.e., $\bar{A}$ occuring).

Probabilities are OK, cont'd

Since $S_i + T_i = i$, we have

$$\begin{aligned} P(\bar{A}) &= lim_n \frac{T_n}{n} \\ &= lim_n \frac{n - S_n}{n} \\ &= 1 - lim_n \frac{S_n}{n} \\ &= 1 - P(A) \end{aligned} \tag{3}$$

Independence

Two events A and B are *independent* if the occurrence of one has no influence (positively or negatively) on the other. It can be shown from the frequentist approach that this means
$P(A \cap B) = P(A) * P(B)$.
(Comp 2711 people, reflect on this.)

So, if I toss a fair coin and a fair dice together, since the outcomes are independent $P(coin = H \ and \ dice = 6) = \frac{1}{2} * \frac{1}{6} = \frac{1}{12}$

If I toss the coin three times, each toss is independent of the others. So the probability of the outcome HHT is $P(H) * P(H) * P(T) = 1/8$.

Random Variables

A *random variable* is a function that associates a number to events. What use is this?

Example: A Payoff Problem
Suppose I play this game: A biased coin is tossed repeatedly. The probability of T is 1/3, so that of H is 2/3. I win \$$n$ if (from the start) the following sequence of events is observed: $T, T, T, \ldots, H$ where the only H is on the nth toss, all preceeding tosses yielding T.

The game stops when I win, so there are no sequences of the form $TTHTTTH\ldots$, as the game would have stopped at TTH.

What is my *expected winnings*?

Random Variables cont'd

Now, each winning event for me is associated with a payoff (my winnings) of some dollar amount, which is a number. This association will be the random variable. More formally:

Each winning event for me is some n-length sequence $TTT \ldots TTH$ which I denote by $E(n)$. To say my winnings for this event is \$$n$, we define the random variable $W : Events \rightarrow Numbers$ by:

$$\begin{aligned} W(E) &= k \;\; if \; E = E(k) \; for \; some \; k \\ &= 0 \;\; otherwise. \end{aligned} \tag{4}$$

Repeated Runs

The frequentist interpretation of the probability $P(E(n))$ of event $E(n)$ is that it is the proportion of times that $E(n)$ will occur if this game is played over and over again. So, if we play this game N times, I will win roughly this dollar amount:

$$S = 1 * P(E(1)) * N + 2 * P(E(2)) * N + \cdots + n * P(E(n)) * N + \cdots \quad (5)$$

So, the *average* amount I will win over the N games is

$$\begin{aligned} SA &= \frac{1}{N}(1 * P(E(1)) * N + \cdots + n * P(E(n)) * N + \cdots) \\ &= 1 * P(E(1)) + 2 * P(E(2)) + \cdots + n * P(E(n)) + \cdots \quad (6) \end{aligned}$$

This is the "long run average payoff for me".

Expectation = Long Run Average

The last equation can be re-written using the random variable W as:

$$SA = W(1) * P(E(1)) + W(2)P(E(2)) + \cdots \tag{7}$$

Generalizing this idea, suppose E_n is a set of events on which we can define a random variable X. Thus $X(E_n)$ is a number — the "payoff" for event E_n. The *Expectation* or *expected value* of X is the sum

$$E(X) = \sum_i X(E_i) * P(E_i) \tag{8}$$

Often we get lazy and write this simply as $E(X) = \sum_i X(i) * P(i)$.

Back to the game

So, getting back to our motivating example, what are my expected winnings? Recall that we defined a random variable $W(n)$ associated with the events $E(n)$ of outcomes which are n-length sequence $TTT \ldots TTH$.

$$\begin{aligned} W(E) &= k \;\; if\ E = E(k)\ for\ some\ k \\ &= 0 \;\; otherwise. \end{aligned} \tag{9}$$

To use equation 8 to calcalute the expectation of W (which is the expected value of my winnings), all we need to do is find $P(E(n))$.

First Visit Problems

$P(E(n))$ is $(\frac{1}{3})^{n-1}\frac{2}{3}$ as $P(T) = 1/3$ and $P(H) = 2/3$. Hence

$$E(W) = \sum_{i=0}^{i=\infty}(i+1)(\frac{1}{3})^i\frac{2}{3} \tag{10}$$

We had actually seen this before when we considered the expected time to success in linear probing!

This game is an abstract model of *First Visit Problems.* In each Bernoulli trial we have a probability of failure q, and of success p. Then the expected number of trials to the first success is $\sum_{i=0}^{i=\infty}(i+1)q^ip$, which we had shown how to evaluate in the lecture that analysed the naive model of linear probing.

Gambling

Modern probability theory originated with *Pascal* who worked out gambling odds for his friends. It was later aximatized by *Kolmogorov.* Pascal was a religious mathematician, a rare breed today. He used expectation of a random variable as an argument that one should be a theist.

Pascal's Wager:
By random variable expectation, you should believe that God exists.

Pascal's Wager "Proof"

Let G be the event that you did not believe in God, but he exists
H be the event that you did not believe in God and he does not exist
I be the event that you did believe in God, and he exists
J be the event that you did believe in God, but he he does not exist

Event G leads to a *big penalty* when you find out too late when you die; say this payoff is $-2^{1,000,000,000,000,000,000,000}$.

Event I leads to a *big reward* with you landing in Heaven surrounded by cherubs and angels, basking in eternal happiness; say this payoff is $2^{1,000,000,000,000,000,000,000}$.

Pascal's Wager "Proof" cont'd

Events H and J are neutral, you neither gain nor lose anything.

What is your expected payoff with different decisions leading to different probabilities for the events?

The expected payoff is $2^{1,000,000,000,000,000,000,000} * P(I) - 2^{1,000,000,000,000,000,000,000} * P(G) + 0 * P(H) + 0 * P(J)$

But I is actually two independent atomic events: Bel — you belive in God, AND GE — God exists, i.e $Bel \wedge GE$;
and G is correspondingly $\neg Bel \wedge GE$. So $P(I) = P(Bel) * P(GE)$ and $P(G) = P(\neg Bel) * P(GE)$.

Pascal's Wager "Proof" cont'd

By symmetry (or agnosticism) $P(GE) = P(\neg GE) = 0.5$. Suppose you believe, i.e. $P(Bel) = 1$ (so $P(\neg Bel) = 0$) and so $P(I) = 0.5$ (P(G) = 0). Your payoff is $\frac{1}{2} * 2^{1{,}000{,}000{,}000{,}000{,}000{,}000{,}000}$.

On the contrary, if you did not believe, your payoff is $-\frac{1}{2} * 2^{1{,}000{,}000{,}000{,}000{,}000{,}000{,}000}$.

This choice is under your control!

QED.

Other Examples on Expection of a RV

Pokies

You play the pokies. On each trial your probability of winning is p and losing is $q = 1 - p$. If you win, the payout is $\$P$, but it costs you $\$Q$ per trial. What is your expected payoff W?

$$W = P * p - Q * q = (P + Q) * p - Q$$

The club will so adjust P, Q, p so that your payoff is slightly *negative*!

St Petersburg Paradox
Expectations may not be finite.

A fair coin is used in Bernoulli trials. On each trial I pay \$5 to play. I win on the n-th trial if H is the result on this trial, and it is the first time that it has occurred. On such an event, call it E_n, I collect $\$2^n$. My net gain then would be $G(n) = \$2^n - 5n$. What is my expected winnings?

Define a RV X such that $X(E_n) = 2^n - 5n$. The probability of E_n is $(\frac{1}{2})^n$. Hence, the expectation of X is $\sum_{n=1}^{\infty} G(n)P(E_n)$
$= \sum_{n=1}^{\infty}(\frac{1}{2})^n(2^n - 5n)$, which is *unbounded*.