WHAT IS A PROBABILISTIC PROOF?

Paul McGill

This note is aimed at those who ask naive, and sometimes not so naive, questions about 'probability'. I try to give the flavour of the approach. For that is what it is. A way of looking at problems 'probabilistically'.

Probabilistic arguments arise in all sorts of different situations. For example one comes across them in combinatorics, statistical physics, differential geometry, and especially in analysis. It is this last that I shall concentrate on, in an attempt to clarify the difference between a probabilistic and an analytic proof of the same result. One confusion is that an analytic proof for one person may be a probabilistic proof to another. My definition is the very purest of all. Namely that a probabilistic proof is one which is motivated in terms of the sample path (or individual trial).

I have found it helpful to think of 'probability' as a factorisation.

$$\textit{Problem} \mapsto \textit{Probabilistic Formulation} \overset{E}{\mapsto} \textit{Solution}$$

where E is of course the expectation operator. So, roughly speaking, one argues in terms of the sample path, then integrates to obtain the (analytic) answer. It is not claimed that this factoring is the easiest solution, but rather that it is sometimes more 'intuitive' (whatever that means) or maybe more 'natural'.

Example 1. Consider the probability density in t

$$q_x(t) \;=\; \frac{x}{\sqrt{2\pi t^3}}\exp\left(-\frac{x^2}{2t}\right) \qquad (t > 0)$$

where $x > 0$ is a parameter. It is not immediately obvious that

$$q_x * q_y \;=\; q_{x+y} \qquad (\dagger)$$

with $*$ denoting convolution. So we see that an **analytic** proof of ($\dagger$) is the computation of the Laplace Transform of q_x, the result being

$$\int_0^\infty e^{-\lambda t} q_x(t)\,dt \;=\; e^{-\sqrt{2\lambda}x},$$

and NOW the answer is clear. But recall that if we add two independent random variables then the law of their sum is given by the convolution of the separate laws (we conveniently omit the proof!). Hence a **probabilistic** proof of ($\dagger$) is possible if one can find two independent random variables H_x and H_y, such that $H_x + H_y = H_{x+y}$, where H_z has law q_x for all $z > 0$.

So there it is. All that needs to be done is to find the appropriate probabilistic setting in which the result will be obvious. To set up the answer we digress a little, and introduce the currently fashionable theory of martingales.

Example 2. Suppose that X_n is a sequence of i.i.d. (independent identically distributed) random variables such that X_1 has values ± 1 with $P[X_1 = 1] = p$. We define the simple random walk as $s_n = \sum_{i=1}^n X_i$. One of the things to notice about this is the way the definition is sequential. Thus we define the sum s_n when we have observed the variables $X_1, X_2, \dots X_n$ already. One thinks of this as tossing a (biased) coin successively, and the picture is one of dynamic probability (the universe unfolding, etc.). It is natural from this point of view to think not just of the process s_n itself, but of the pair consisting of the process s_n and the information which it has accumulated up to the time n, which we represent by the σ-algebra $S_n = \sigma(X_1, X_2, \dots, X_n)$. Now THINK. Suppose we are betting on the value of s_n. Clearly it is more advantageous to know the value of s_{n-1} than it is to know that $s_0 \doteq 0$.

On the basis of 'latest is best' (intuition!) we agree that

$$E[s_n | S_{n-1}] = E[s_n | s_{n-1}],$$

and by definition

$$E[s_n | s_{n-1}] = s_{n-1} + E[X_n | s_{n-1}] = s_{n-1} + (2p - 1).$$

Iterating one obtains

$$E[s_n - (2p - 1)s_n | S_m] = s_m - (2p - 1)m. \qquad (m < n)$$

We have written it in this way to emphasise that the process (new word!) $s_n - (2p - 1)n$ is stable under the operation of taking the conditional expectation. Before leaving this example we introduce the notion of a random time. Consider the first time τ_1 that the random walk goes strictly positive (sometimes called the hitting time of 1). Then we might be interested in computing the distribution of τ_1. The question is how.

Definitions (1) An increasing family $\mathcal{F}_n$ of σ-algebras of events in a probability space is called a filtration.

(2) A sequence of random variables X_n is said to be adapted to $\mathcal{F}_n$ if each X_n is measurable w.r.t. $\mathcal{F}_n$.

Thus 'adapted' has connotations of being observable in the filtration at the appropriate time. The filtration S_n defined in Example 2 above is called the natural filtration of the random walk s_n.

A martingale M_n in the filtration $\mathcal{F}_n$ is a process which is adapted and stable under the conditional expectation operation, i.e.

$$E[M_n | \mathcal{F}_m] = M_m. \qquad (m < n)$$

Recall how this means that $E[M_n 1_A] = E[M_m 1_A]$ for every $A \in \mathcal{F}_m$.

In words: 'averaging M_n over members of $\mathcal{F}_m$ yields M_m'. The only way to understand this definition is to work with it. We do this later on. But for the moment be content with a few examples.

Examples 3

(a) Suppose that the probability space is $([0,1],\mathcal{B},m)$ where $\mathcal{B}$ is the Borel σ-algebra and m is Lebesgue measure. Consider the sequence of Rademacher functions $f_n = \operatorname{sgn}(\sin 2^n \pi x)$, each of which has mean zero. We define g_n almost everywhere by

$$g_0 = 0$$

$$g_1 = 1_{[0,\frac{1}{2}]} - 1_{[\frac{1}{2},1]} = f_1$$

$$g_2 = \frac{3}{2} 1_{[0,\frac{1}{4}]} + \frac{1}{2} 1_{[\frac{1}{4},\frac{1}{2}]} - \frac{1}{2} 1_{[\frac{1}{2},\frac{3}{4}]} - \frac{3}{2} 1_{[\frac{3}{4},1]} = g_1 + 2^{-1} f_2$$

$$\vdots$$

The general formula being $g_n = g_{n-1} + 2^{-(n-1)} f_n$.
Let $\mathcal{F}_n = \sigma(f_i : 1 \le i \le n)$. Up to null sets $\mathcal{F}$ is just unions of the dyadic intervals $\{[k2^{-n}, (k+1)2^{-n}] : 0 \le k \le 2^n - 1\}$. Then

$$E[g_n | \mathcal{F}_{n-1}] = g_{n-1} + 2^{-(n-1)} E[f_n | \mathcal{F}_{n-1}] = g_{n-1}$$

so by induction g_n is an $\mathcal{F}_n$ bounded martingale.

(b) If X is an integrable random variable (so that one can define conditional expectations) then the sequence

$$X_n = E[X | \mathcal{F}_n]$$

is a martingale in the filtration $\mathcal{F}_n$. This is an example of a closed martingale.

(c) Let u_n be a random walk whose i.i.d. steps are now normal $N(0,1)$. By using the formula

$$\frac{1}{\sqrt{2\pi t}}\int_{-\infty}^{\infty} \exp\left(-\frac{x^2}{2t} + \sqrt{2\lambda}x\right)dx = e^{\lambda t}$$

one sees that $\exp\{\sqrt{2\lambda}u_n - \lambda n\}$ is a martingale for the natural filtration of u_n.

Definition An integer-valued random variable $T \geq 0$ is said to be a stopping time for the filtration $\mathcal{F}_n$ if $\{T = n\} \in \mathcal{F}_n$ for all $n \geq 0$.

Example 4 The first passage time τ_1 defined in Example 2 is a stopping time for the filtration S_n. To see why note that

$$\{\tau_1 = n\} = \{s_1 < 1, s_2 < 1, \ldots, s_{n-1} < 1, s_n = 1\} \in S_n$$

Thus a stopping time is one which can be observed 'as soon as it happens'. Notice that the last zero before time τ_1 cannot. Nor can the minimum before τ_1. Both of these facts are all too familiar to gamblers.

The most important thing about martingales is not so much the celebrated martingale convergence theorem, but rather the fact that the definition can be made to work for stopping times also. Notice that, by definition of the conditional expectation, if M_n is a martingale then $E[M_n] = E[M_0]$.

Doob's Optional Stopping Theorem If M_n is an $\mathcal{F}_n$ martingale which is uniformly bounded up to the $\mathcal{F}_n$ stopping time T then

$$E[M_T] = E[M_0].$$

Let us now construct a martingale which gives us the law of τ_1. We will look for a function f such that $M_n = e^{-\lambda n}f(s_n)$ is a martingale (for the natural filtration S_n of s_n). Let

us suppose that $f(x) = e^{\mu x}$. Then computing the conditional expectation we have

$$E[f(s_n)e^{-\lambda n}|S_{n-1}] = e^{-\lambda(n-1)}e^{\mu s_{n-1}}[pe^{\mu} + e^{-\mu}(1 - p)]e^{-\lambda}.$$

From which the condition for a martingale is that

$$e^{\lambda} = pe^{\mu} + (1 - p)e^{-\mu}.$$

This is a quadratic equation in e^{μ}, with two solutions. We want our martingale to be bounded up to the time τ_1 so choose (for $\mu > 0$) the positive square root

$$\mu = \mu(\lambda) = \log\left(\frac{e^{\lambda} + \sqrt{e^{2\lambda} - 4p(1 - p)}}{2p}\right).$$

With this choice of μ we can apply the Doob Theorem at the stopping time τ_1 and get

$$E[e^{-\lambda\tau_1 + \mu s_{\tau_1}}] = 1$$

Notice how we ignore the set $\{\tau_1 = +\infty\}$ since the martingale is zero there. But if $\{\tau_1 < +\infty\}$ then $s_{\tau_1} = 1$ and so $E[e^{-\lambda\tau_1}] = e^{-\mu(\lambda)}$. From this information we make various computations. Note that

$$\mu(0) = \log\frac{1 + |2p - 1|}{2p}$$

which gives $P[\tau_1 < +\infty] = e^{-\mu(0)} = (p/(p-1)) \wedge 1$. If we look at the (interesting) balanced case $p = \frac{1}{2}$ then we compute that

$$E[e^{-\lambda\tau_1}] = \frac{1}{e^{\lambda} + \sqrt{e^{2\lambda} - 1}}$$

so by differentiation, putting $\lambda = 0$, we get $E[\tau_1] = +\infty$. Thus the expected waiting time for first positive passage is infinite, although the time itself is finite.

This is an example of the probabilistic method. It is clearly formulated in terms of the sample path, and in the end the answer comes by taking an expectation.

Important Remark   The boundedness condition in Doob's Theorem is essential. Consider the example of the simple random walk when $p = 1/2$. Then τ_1 is a stopping time but we have

$$1 = E[s_{\tau_1}] \neq E[s_0] = 0.$$

We are now ready to finish off this circle of ideas. We begin with the martingale of Example 3(c) above $M_n = \exp(\sqrt{2\lambda}u_n - \lambda n)$. There is a continuous time analogue of this martingale $M_t = \exp(\sqrt{2\lambda}B_t - \lambda t)$, where B_t is called Brownian motion (and we take $B_0 = 0$ here). There are two structural facts that we need, both of them difficult to prove.

(1) The process B_t varies continuously with time. This result is due to Wiener.

To state the second one we define the random time

$$T_x = \inf(t > 0 : B_t = x).$$

(2) $(B_{T_x+t} - x : t \geq 0)$ is a process with the same law as B_t which is independent of the process $B_{T_x \wedge t}$. This is a particular case of the strong Markov property.

It is a FACT that we can apply the Doob theorem at time T_x to the martingale M_t. Which gives us

$$E[M_{T_x}] = E[M_0] = 1 = E[\exp(\sqrt{2\lambda}B_{T_x} - \lambda T_x)]$$

But using (1) $B_{T_x} = x$ (at least when T_x is finite) so we get

$$E[e^{-\lambda T_x}] = e^{-\sqrt{2\lambda}x}.$$

Going back to Example 1 we find that T_x has law q_x. But now (2) shows that $\tilde{T}_y = \inf(t > 0 : B_{T_x+t} - x > y)$ has the same law as T_y, while at the same time being independent of T_x. Since we have the sample path identity

$$T_{x+y} = T_x + \tilde{T}_y$$

the conclusion $q_x * q_y = q_{x+y}$ is immediate.

As we have written it here the probabilistic proof seems to depend on the analytic proof. However one can see that T_x has law q_x directly, by using (2) and the reflection argument of Désiré André. This reasoning is too subtle for the casual reader.

In conclusion we point out how this typifies the ingredients of a probabilistic proof. It is certainly harder than the original, but has an undeniable charm and utility since we have a diagram for 'why' the result holds.

Department of Mathematics,
Maynooth College,
Co. Kildare