Chapter 2

Expectations

Just as events have (conditional) probabilities attached to them, with possible interpretation as a long-run frequency, so too do random variables have a number interpretable as a long-run average attached to them. Given a particular piece of information, the symbol

$$\mathbb{E}\big(X \mid \text{information}\big)$$

•(conditional) expected value

denotes the (CONDITIONAL) EXPECTED VALUE or (CONDITIONAL) EXPECTATION of the random variable X (given that information).

When the information is taken as understood, the expected value is abbreviated to $\mathbb{E}X$.

Expected values are not restricted to lie in the range from zero to one.

As with conditional probabilities, there are convenient abbreviations when the conditioning information includes something like {event F has occurred}:

$$\mathbb{E}\big(X \mid \text{information and “}F\text{ has occurred”}\big)$$
$$\mathbb{E}\big(X \mid \text{information, } F\big)$$

Unlike many authors, I will take the expected value as a primitive concept, not one to be derived from other concepts. All of the methods that those authors use to *define* expected values will be *derived* from a small number of basic rules. You should provide the interpretations for these rules as long-run averages of values generated by independent repetitions of random experiments.

Rules for (conditional) expectations

Let X and Y be random variables, c and d be constants, and $F_1, F_2, \ldots$ be events. Then:

(E1) $\mathbb{E}\big(cX + dY \mid \text{info}\big) = c\mathbb{E}\big(X \mid \text{info}\big) + d\mathbb{E}\big(Y \mid \text{info}\big)$;

(E2) if X can only take the constant value c under the given “info” then $\mathbb{E}\big(X \mid \text{info}\big) = c$;

(E3) if the given “info” forces $X \le Y$ then $\mathbb{E}\big(X \mid \text{info}\big) \le \mathbb{E}\big(Y \mid \text{info}\big)$;

(E4) if the events $F_1, F_2, \ldots$ are disjoint and have union equal to the whole sample space then
$$\mathbb{E}\big(X \mid \text{info}\big) = \sum_i \mathbb{E}\big(X \mid F_i,\text{info}\big)\mathbb{P}\big(F_i \mid \text{info}\big).$$

Only rule E4 should require much work to interpret. It combines the power of both rules P4 and P5 for conditional probabilities. Here is an interpretation for the case of two disjoint events F_1 and F_2 with union S.

Repeat the experiment a very large number (N) of times, noting for each repetition the value taken by X and which of F_1 or F_2 occurs.

	1	2	3	4	...				$N-1$	N	total
F_1 occurs	✓	✓		✓	...				✓	✓	N_1
F_2 occurs			✓		...	✓	✓	✓			N_2
X	x_1	x_2	x_3	x_4	...				x_{N-1}	x_N	

© David Pollard

Those trials where F_1 occurs correspond to conditioning on F_1:

$$\mathbb{E}\big(X \mid F_1, \text{info}\big) \approx \frac{1}{N_1} \sum_{F_1 \text{occurs}} x_i.$$

Similarly,

$$\mathbb{E}\big(X \mid F_2, \text{info}\big) \approx \frac{1}{N_2} \sum_{F_2 \text{occurs}} x_i$$

and

$$\mathbb{P}\big(F_1 \mid \text{info}\big) \approx N_1/N$$
$$\mathbb{P}\big(F_2 \mid \text{info}\big) \approx N_2/N.$$

Thus

$$\begin{aligned}
&\mathbb{E}\big(X \mid F_1, \text{info}\big)\mathbb{P}\big(F_1 \mid \text{info}\big) + \mathbb{E}\big(X \mid F_2, \text{info}\big)\mathbb{P}\big(F_2 \mid \text{info}\big) \\
&\approx \left(\frac{1}{N_1} \sum_{F_1 \text{occurs}} x_i\right)\left(\frac{N_1}{N}\right) + \left(\frac{1}{N_2} \sum_{F_2 \text{occurs}} x_i\right)\left(\frac{N_2}{N}\right) \\
&= \frac{1}{N} \sum_{i=1}^{N} x_i \\
&\approx \mathbb{E}\big(X \mid \text{info}\big).
\end{aligned}$$

As N gets larger and larger all approximations are supposed to get better and better, and so on.

Rules E2 and E5 imply immediately a result that is used to calculate expectations from probabilities. Consider the case of a random variable Y expressible as a function $g(X)$ of another random variable, X, which takes on only a discrete set of values $c_1, c_2, \ldots$. (I will return later to the case of so-called continuous random variables.) Let F_i be the subset of S on which $X = c_i$, that is,

$$F_i = \{X = c_i\}.$$

Then by E2,

$$\mathbb{E}\big(Y \mid F_i, \text{info}\big) = g(c_i),$$

and by E5,

$$\mathbb{E}\big(Y \mid \ \text{info}\big) = \sum_i g(c_i)\mathbb{P}\big(F_i \mid \text{info}\big).$$

More succinctly,

<2.1> $$\boxed{\mathbb{E}\big(g(X) \mid \ \text{info}\big) = \sum_i g(c_i)\mathbb{P}\big(X = c_i \mid \text{info}\big)}$$

In particular,

<2.2> $$\mathbb{E}\big(X \mid \ \text{info}\big) = \sum_i c_i \mathbb{P}\big(X = c_i \mid \text{info}\big).$$

Formulas <2.1> and <2.2> apply to random variables that take values in the "discrete set" $\{c_1, c_2, \ldots\}$. If the range of values includes an interval of real numbers, an approximation argument (see Chapter 4) replaces sums by integrals.

<2.3> **Example.** The "HHH versus TTHH" Example in Chapter 1 solved the following problem:

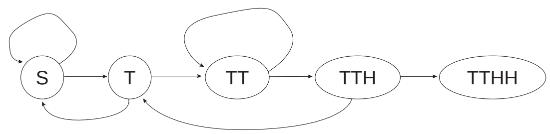
> Imagine that I have a fair coin, which I toss repeatedly. Two players, M and R, observe the sequence of tosses, each waiting for a particular pattern on consecutive tosses: M waits for hhh, and R waits for tthh. The one whose pattern appears first is the winner. What is the probability that M wins?

The answer—that M has probability 5/12 of winning—is slightly surprising, because, at first sight, a pattern of four appears harder to achieve than a pattern of three.

 © David Pollard

A calculation of expected values will add to the puzzlement. As you will see, if the game is continued until each player sees his pattern, it takes tthh longer (on average) to appear than it takes hhh to appear. However, when the two patterns are competing, the tthh pattern is more likely to appear first. How can that be?

For the moment forget about the competing hhh pattern: calculate the expected number of tosses needed before the pattern tthh is obtained with four successive tosses. That is, if we let X denote the number of tosses required then the problem asks for the expected value $\mathbb{E}X$.



The Markov chain diagram keeps track of the progress from the starting state (labelled S) to the state TTHH where the pattern is achieved. Each arrow in the diagram corresponds to a transition between states with probability 1/2.

Once again it is easier to solve not just the original problem, but a set of problems, one for each starting state. Let

$$\begin{aligned}\mathcal{E}_S &= \mathbb{E}(X \mid \text{start at S})\\ \mathcal{E}_H &= \mathbb{E}(X \mid \text{start at H})\\ &\vdots\end{aligned}$$

Then the original problem is asking for the value of $\mathcal{E}_S$.

Condition on the outcome of the first toss, writing $\mathcal{H}$ for the event {first toss lands heads} and $\mathcal{T}$ for the event {first toss lands tails}. From rule E4 for expectations,

$$\mathcal{E}_S = \mathbb{E}(X \mid \text{start at S}, \mathcal{T})\mathbb{P}(\mathcal{T} \mid \text{start at S}) + \mathbb{E}(X \mid \text{start at S}, \mathcal{H})\mathbb{P}(\mathcal{H} \mid \text{start at S})$$

Both the conditional probabilities equal 1/2 ("fair coin"; probability does not depend on the state). For the first of the conditional expectations, count 1 for the first toss, then recognize that the remaining tosses are just those needed to reach TTHH starting from the state T:

$$\mathbb{E}(X \mid \text{start at S}, \mathcal{T}) = 1 + \mathbb{E}(X \mid \text{start at T})$$

Don't forget to count the first toss. An analogous argument leads to an analogous expression for the second conditional expectation. Susbtitution into the expression for $\mathcal{E}_S$ then gives

$$\mathcal{E}_S = \tfrac{1}{2}(1 + \mathcal{E}_T) + \tfrac{1}{2}(1 + \mathcal{E}_S)$$

Similarly,

$$\begin{aligned}\mathcal{E}_T &= \tfrac{1}{2}(1 + \mathcal{E}_{TT}) + \tfrac{1}{2}(1 + \mathcal{E}_S)\\ \mathcal{E}_{TT} &= \tfrac{1}{2}(1 + \mathcal{E}_{TT}) + \tfrac{1}{2}(1 + \mathcal{E}_{TTH})\\ \mathcal{E}_{TTH} &= \tfrac{1}{2}(1 + 0) + \tfrac{1}{2}(1 + \mathcal{E}_T)\end{aligned}$$

What does the zero in the last equation represent?

The four linear equations in four unknowns have the solution $\mathcal{E}_S = 16$, $\mathcal{E}_T = 14$, $\mathcal{E}_{TT} = 10$, $\mathcal{E}_{TTH} = 8$. Thus, the solution to the original problem is that the expected number of tosses to achieve the tthh pattern is 16.

On Problem Sheet 2 you are asked to show that the expected number of tosses needed to get hhh, without competition, is 14. The expected number of tosses for the game with competition between hhh and tthh is $9\tfrac{1}{3}$ (see Matlab m-file solve_hhh_tthh.m). Notice that the expected value for the game with competition is smaller than the minimum of the expected values for the two games. Why must it be smaller?

□

The calculation of an expectation is often a good way to get a rough feel for the behaviour of a random process. It is helpful to remember expectations for a few standard mechanisms, such as coin tossing, rather than have to rederive them repeatedly.

 © David Pollard

<2.4> **Example.** For independent coin tossing, what is the expected number of tosses to get the first head?

Suppose the coin has probability $p > 0$ of landing heads. (So we are actually calculating the expected value for the geometric(p) distribution.) I will present two methods.

Method A.

Condition on whether the first toss lands heads (H) or tails (T). With X defined as the number of tosses until the first head,

$$\begin{aligned}\mathbb{E}X &= \mathbb{E}(X \mid H)\mathbb{P}H + \mathbb{E}(X \mid T)\mathbb{P}T \\ &= (1)p + (1 + \mathbb{E}X)(1-p).\end{aligned}$$

The reasoning behind the equality

$$\mathbb{E}(X \mid T) = 1 + \mathbb{E}X$$

is: After a tail we are back where we started, still counting the number of tosses until a head, except that the first tail must be included in that count.

Solving the equation for $\mathbb{E}X$ we get

$$\mathbb{E}X = 1/p.$$

Does this answer seem reasonable? (Is it always at least 1? Does it increase as p increases? What happens as p tends to zero or one?)

Method B.

By the formula <2.1>,

$$\mathbb{E}X = \sum_{k=1}^{\infty} k(1-p)^{k-1}p.$$

There are several cute ways to sum this series. Here is my favorite. Write q for $1-p$. Write the kth summand as a a column of k terms pq^{k-1}, then sum by rows:

$$\begin{aligned}\mathbb{E}X = p + pq + pq^2 + pq^3 + \dots \\ +pq + pq^2 + pq^3 + \dots \\ +pq^2 + pq^3 + \dots \\ +pq^3 + \dots \\ \vdots\quad\end{aligned}$$

Each row is a geometric series.

$$\begin{aligned}\mathbb{E}X &= p/(1-q) + pq/(1-q) + pq^2/(1-q) + \dots \\ &= 1 + q + q^2 + \dots \\ &= 1/(1-q) \\ &= 1/p,\end{aligned}$$

□ same as before.

Probabilists study standard mechanisms, and establish basic results for them, partly in the hope that they will recognize those same mechanisms buried in other problems. In that way, unnecessary calculation can be avoided, making it easier to solve more complex problems. It can, however, take some work to find the hidden mechanism.

<2.5> **Example.** (coupon collector's problem) In order to encourage consumers to buy many packets of cereal, a manufacurer includes a Famous Probabilist card in each packet. There are 10 different types of card: Chung, Feller, Levy, Kologorov, ..., Doob. Suppose that I am seized by the

 © David Pollard

desire to own at least one card of each type. What is the expected number of packets that I need to buy in order to achieve my goal?

Assume that the manufacturer has produced enormous numbers of cards, the same number for each type. (If you have ever tried to collect objects of this type, you might doubt the assumption about equal numbers. But, without it, the problem becomes exceedingly difficult.) The assumption ensures, to a good approximation, that the cards in different packets are independent, with probability 1/10 for a Chung, probability 1/10 for a Feller, and so on.

The high points in my life occur at random "times" T_1, T_1+T_2, $\ldots$, $T_1+T_2+\ldots+T_{10}$, when I add a new type of card to my collection: After $T_1 = 1$ card I have my first type; after another T_2 cards I will get something different from the first card; after another T_3 cards I will get a third type; and so on.

The question asks for $\mathbb{E}(T_1 + T_2 + \ldots + T_{10})$, which rule E1 (applied repeatedly) reexpresses as $\mathbb{E}T_1 + \mathbb{E}T_2 + \ldots + \mathbb{E}T_{10}$.

The calculation for $\mathbb{E}T_1$ is trivial because T_1 must equal 1: we get $\mathbb{E}T_1 = 1$ by rule E2. Consider the mechanism controlling T_2. For concreteness suppose the first card was a Doob. Each packet after the first is like a coin toss with probability 9/10 of getting a head (= a non-Doob), with T_2 like the number of tosses needed to get the first head. Thus

$$T_2 \text{ has a geometric(9/10) distribution.}$$

Deduce from Example <2.4> that $\mathbb{E}T_2 = 10/9$, which is slightly larger than 1.

Now consider the mechanism controlling T_3. Condition on everything that was observed up to time $T_1 + T_2$. Under the assumption of equal abundance and enormous numbers of cards, this conditioning information is acually irrelevent; the mechanism controlling T_3 is independent of the past information. (Hard question: Why would the T_2 and T_3 mechanisms not be independent if the cards were not equally abundant?) So what is that T_3 mechanism? I am waiting for any one of the 8 types I have not yet collected. It is like coin tossing with probability 8/10 of heads:

$$T_3 \text{ has geometric (8/10) distribution,}$$

and thus $\mathbb{E}T_3 = 10/8$. And so on, leading to

$$\mathbb{E}T_1 + \mathbb{E}T_2 + \ldots + \mathbb{E}T_{10} = 1 + 10/9 + 10/8 + \ldots + 10/1 \approx 29.3.$$

□ I should expect to buy about 29.3 packets to collect all ten cards.

The independence between packets was not needed to justify the appeal to rule E1, to break the expected value of the sum into a sum of expected values. It did allow us to recognize the various geometric distributions without having to sort through possible effects of large T_2 on the behavior of T_3, and so on.

You might appreciate better the role of independence if you try to solve a similar problem with just two sorts of card, not in equal proportions.

For the coupon collectors problem I assumed large numbers of cards of each type, in order to justify the analogy with coin tossing. Without that assumption the depletion of cards from the population would have a noticeable effect on the proportions of each type remaining after each purchase. The next example illustrates the effects of sampling from a finite population without replacement, when the population size is not assumed very large.

<2.6> **Example.** Suppose an urn contains r red balls and b black balls, all balls identical except for color. Suppose balls are removed from the urn one at a time, without replacement. Assume that the person removing the balls selects them at random from the urn: if k balls remain then each has probability 1/k of being chosen.

Question: What is the expected number of red balls removed before the first black ball?

The problem might at first appear to require nothing more than a simple application of formula <2.1> for deriving expectations from probabilities. We shall see. Let T be the number of

©David Pollard

reds removed before the first black. Find the distribution of T, then appeal to the formula <2.1> to get

$$\mathbb{E}T = \sum_k k\mathbb{P}\{T = k\}.$$

Sounds easy enough.

Define $R_i = \{i\text{th ball red}\}$ and $B_i = \{i\text{th ball black}\}$. The possible values for T are $0, 1, \ldots, r$. For k in this range,

$$\begin{aligned}\mathbb{P}\{T = k\} &= \mathbb{P}\{\text{first k balls red, (k+1)st ball is black}\} \\ &= \mathbb{P}(R_1 R_2 \ldots R_k B_{k+1}) \\ &= (\mathbb{P}R_1)\mathbb{P}(R_2 \mid R_1)\mathbb{P}(R_3 \mid R_1 R_2)\ldots\mathbb{P}(B_{k+1} \mid R_1 \ldots R_k) \\ &= \frac{r}{r+b}.\frac{r-1}{r+b-1}\cdots\frac{b}{r+b-k}.\end{aligned}$$

The dependence on k is fearsome. I wouldn't like to try multiplying by k and summing. If you are into pain you might continue the argument from here.

There is a much easier way to calculate the expectation, by breaking T into a sum of much simpler random variables for which formula <2.1> is trivial to apply. This approach is sometimes called the METHOD OF INDICATORS.

•method of indicators

Suppose the red balls are labelled $1, \ldots, r$. Let T_i equal 1 if red ball number i is sampled before the first black ball. (Be careful here. The black balls are not thought of as numbered. The first black ball is not a ball bearing the number 1; it might be any of the b black balls in the urn.) Then $T = T_1 + \ldots + T_r$. By symmetry—it is assumed that the numbers have no influence on the order in which red balls are selected—each T_i has the same expectation. Thus

$$\mathbb{E}T = \mathbb{E}T_1 + \ldots + \mathbb{E}T_r = r\mathbb{E}T_1.$$

For the calculation of $\mathbb{E}T_1$ we can ignore most of the red balls. The event $\{T_1 = 1\}$ occurs if and only if red ball number 1 is drawn before all b of the black balls. By symmetry, the event has probability $1/(b+1)$. (If $b+1$ objects are arranged in random order, each object has probability $1/(1+b)$ of appearing first in the order.)

If you are not convinced by the appeal to symmetry, you might find it helpful to consider a thought experiment where all $r+b$ balls are numbered and they are removed at random from the urn. That is, treat all the balls as distinguishable and sample until the urn is empty. (You might find it easier to follow the argument in a particular case, such as all $120 = 5!$ orderings for five distinguishable balls, 2 red and 3 black.) The sample space consists of all permutations of the numbers 1 to $r+b$. Each permutation is equally likely. For each permutation in which red 1 precedes all the black balls there is another equally likely permutation, obtained by interchanging the red ball with the first of the black balls chosen; and there is an equally likely permutation in which it appears after two black balls, obtained by interchanging the red ball with the second of the black balls chosen; and so on. Formally, we are partitioning the whole sample space into equally likely events, each determined by a relative ordering of red 1 and all the black balls. There are $b+1$ such equally likely events, and their probabilities sum to one.

Now it is easy to calculate the expected value for red 1.

$$\mathbb{E}T_1 = 0\,\mathbb{P}\{T_1 = 0\} + 1\,\mathbb{P}\{T_1 = 1\} = 1/(b+1)$$

□ The expected number of red balls removed before the first black ball is equal to $r/(b+1)$.

Problem Sheet 3 outlines another way to solve the problem.

Compare the solution $r/(b+1)$ with the result for sampling with replacement, where the number of draws required to get the first black would have a geometric$(b/(r+b))$ distribution. With replacement, the expected number of reds removed before the first black would be

$$(b/(r+b))^{-1} - 1 = r/b.$$

Replacement of balls after each draw increases the expected value slightly. Does that make sense?

© David Pollard

You could safely skip the remainder of this Chapter. It contains a discussion of a tricky little problem, that can be solved by conditioning or by an elegant symmetry argument.

✠ ✠ ✠ ✠ ✠ ✠ ✠ ✠ ✠ ✠ ✠ ✠ ✠ ✠ ✠ ✠

My interest in the calculations in the last Example was kindled by a problem that appeared in the August-September 1992 issue of the American Mathematical Monthly. My solution to the problem—the one I first came up with by application of a straightforward conditioning argument—reduces the calculation to several applications of the result from the previous Example. The solution offered by two readers of the Monthly was slicker.

<2.7> **Example.** (The problem of the Big Pills and Little Pills)

E 3429 [1991, 264]. *Proposed by Donald E. Knuth and John McCarthy, Stanford University, Stanford, CA.*

A certain pill bottle contains m large pills and n small pills initially, where each large pill is equivalent to two small ones. Each day the patient chooses a pill at random; if a small pill is selected, (s)he eats it; otherwise (s)he breaks the selected pill and eats one half, replacing the other half, which thenceforth is considered to be a small pill.

(a) What is the expected number of small pills remaining when the last large pill is selected?

(b) On which day can we expect the last large pill to be selected?

Solution from AMM:

Composite solution by Walter Stromquist, Daniel H. Wagner, Associates, Paoli, PA and Tim Hesterberg, Franklin & Marshall College, Lancaster, PA. The answers are (a) $n/(m+1)+\sum_{k=1}^{m}(1/k)$, and (b) $2m+n-(n/(m+1))-\sum_{k=1}^{m}(1/k)$. The answer to (a) assumes that the small pill created by breaking the last large pill is to be counted. A small pill present initially remains when the last large pill is selected if and only if it is chosen last from among the $m+1$ element set consisting of itself and the large pills—an event of probability $1/(m+1)$. Thus the expected number of survivors from the original small pills is $n/(m+1)$. Similarly, when the kth large pill is selected ($k=1,2,\dots,m$), the resulting small pill will outlast the remaining large pills with probability $1/(m-k+1)$, so the expected number of created small pills remaining at the end is $\sum_{k=1}^{m}(1/k)$. Hence the answer to (a) is as above. The bottle will last $2m+n$ days, so the answer to (b) is just $2m+n$ minus the answer to (a), as above.

I offer two methods of solution for the problem. The first method uses a conditioning argument to set up a recurrence formula for the expected numbers of small pills remaining in the bottle after each return of half a big pill. The equations are easy to solve by repeated substitution. The second method uses indicator functions to spell out the Hesterberg-Stromquist method in more detail. Apparently the slicker method was not as obvious to most readers of the Monthly (and me):

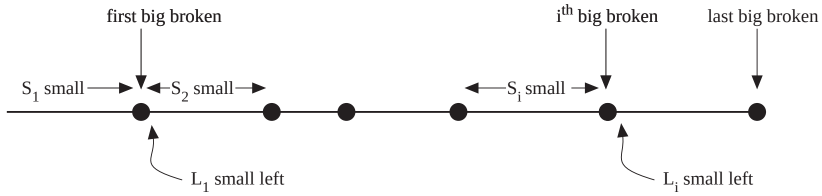
> *Editorial comment.* Most solvers derived a recurrence relation, guessed the answer, and verified it by induction. Several commented on the origins of the problem. Robert High saw a version of it in the MIT Technology Review of April, 1990. Helmut Prodinger reports that he proposed it in the Canary Islands in 1982. Daniel Moran attributes the problem to Charles MacCluer of Michigan State University, where it has been know for some time.
>
> Solved by 38 readers (including those cited) and the proposer. One incorrect solution was received.

Conditioning method.

Invent random variables to describe the depletion of the pills. Initially there are $L_0 = n$ small pills in the bottle. Let S_1 small pills be consumed before the first large pill is broken. After

© David Pollard

the small half is returned to the bottle let there be L_1 small pills left. Then let S_2 small pills be consumed before the next big pill is split, leaving L_2 small pills in the bottle. And so on.



With this notation, part (a) is asking for $\mathbb{E}L_m$. Part (b) is asking for $2m + n - \mathbb{E}L_m$: If the last big pill is selected on day X then it takes $X + L_m$ days to consume the $2m + n$ small pill equivalents, so $\mathbb{E}X + \mathbb{E}L_m = 2m + n$.

The random variables are connected by the equation

$$L_i = L_{i-1} - S_i + 1,$$

the $-S_i$ representing the small pills consumed between the breaking of the $(i-1)$st and ith big pill, and the $+1$ representing the half of the big pill that is returned to the bottle. Taking expectations we get

<2.8> $$\mathbb{E}L_i = \mathbb{E}L_{i-1} - \mathbb{E}S_i + 1.$$

The result from Example <2.6> will let us calculate $\mathbb{E}S_i$ in terms of $\mathbb{E}L_{i-1}$, thereby producing the recurrence formula for $\mathbb{E}L_i$.

Condition on the pill history up to the $(i-1)$st breaking of big pill (and the return of the unconsumed half to the bottle). At that point there are L_{i-1} small pills and $m - (i-1)$ big pills in the bottle. The mechanism controlling S_i is just like the urn problem of Example <2.6>, with

$$\begin{aligned} r &= L_{i-1} \text{ red balls (= small pills)} \\ b &= m - (i-1) \text{ black balls (= big pills).} \end{aligned}$$

From that Example,

$$\mathbb{E}\big(S_i \mid \text{history to } (i-1)\text{st breaking of a big pill}\big) = \frac{L_{i-1}}{1 + m - (i-1)}.$$

To calculate $\mathbb{E}S_i$ we would need to average out using weights equal to the probability of each particular history:

$$\mathbb{E}S_i = \frac{1}{1 + m - (i-1)} \sum_{\text{histories}} \mathbb{P}\{\text{history}\}(\text{value of } L_{i-1} \text{ for that history}).$$

The sum on the right-hand side is exactly the sum we would get if we calculated $\mathbb{E}L_{i-1}$ using rule E4, partitioning the sample space according to possible histories up to the $(i-1)$st breaking of a big pill. Thus

$$\mathbb{E}S_i = \frac{1}{2 + m - i}\mathbb{E}L_{i-1}.$$

Now we can eliminate $\mathbb{E}S_i$ from equality <2.8> to get the recurrence formula for the $\mathbb{E}L_i$ values:

$$\mathbb{E}L_i = \left(1 - \frac{1}{2 + m - i}\right)\mathbb{E}L_{i-1} + 1.$$

If we define $\theta_i = \mathbb{E}L_i/(1 + m - i)$ the equation becomes

$$\theta_i = \theta_{i-1} + \frac{1}{1 + m - i} \qquad \text{for } i = 1, 2, \dots, m,$$

with initial condition $\theta_0 = \mathbb{E}L_0/(1 + m) = n/(1 + m)$. Repeated substitution gives

$$\theta_1 = \theta_0 + \frac{1}{m}$$

© David Pollard

$$
\begin{aligned}
\theta_2 &= \theta_1 + \frac{1}{m-1} = \theta_0 + \frac{1}{m} + \frac{1}{m-1} \\
\theta_3 &= \theta_2 + \frac{1}{m-2} = \theta_0 + \frac{1}{m} + \frac{1}{m-1} + \frac{1}{m-2} \\
&\vdots \\
\theta_m &= \ldots = \theta_0 + \frac{1}{m} + \frac{1}{m-1} + \ldots + \frac{1}{2} + \frac{1}{1}.
\end{aligned}
$$

That is, the expected number of small pills left after the last big pill is broken equals

$$
\begin{aligned}
\mathbb{E}L_m &= (1+m-m)\theta_m \\
&= \frac{n}{1+m} + 1 + \frac{1}{2} + \ldots + \frac{1}{m}.
\end{aligned}
$$

Rewrite of the Stromquist-Hesterberg solution.

Think in terms of half pills, some originally part of big pills. Number the original half pills $1, \ldots, n$. Define

$$
H_i = \begin{cases} +1 & \text{if original half pill } i \text{ survives beyond last big pill} \\ 0 & \text{otherwise.} \end{cases}
$$

Number the big pills $1, \ldots, m$. Use the same numbers to refer to the half pills that are created when a big pill is broken. Define

$$
B_j = \begin{cases} +1 & \text{if created half pill } j \text{ survives beyond last big pill} \\ 0 & \text{otherwise.} \end{cases}
$$

The number of small pills surviving beyond the last big pill equals

$$
H_1 + \ldots + H_n + B_1 + \ldots + B_m.
$$

By symmetry, each H_i has the same expected value, as does each B_j. The expected value asked for by part (a) equals

$$
n\mathbb{E}H_1 + m\mathbb{E}B_1 = n\mathbb{P}\{H_1 = 1\} + m\mathbb{P}\{B_1 = 1\}. \tag{<2.9>}
$$

For the calculation of $\mathbb{P}\{H_1 = +1\}$ we can ignore all except the relative ordering of the m big pills and the half pill described by H_1. By symmetry, the half pill has probability $1/(m+1)$ of appearing in each of the $m+1$ possible positions in the relative ordering. In particular,

$$
\mathbb{P}\{H_1 = +1\} = \frac{1}{m+1}.
$$

For the created half pills the argument is slightly more complicated. If we are given that big pill number 1 the kth amongst the big pills to be broken, the created half then has to survive beyond the remaining $m-k$ big pills. Arguing again by symmetry amongst the $(m-k+1)$ orderings we get

$$
\mathbb{P}\big(B_1 = +1 \mid \text{big number 1 chosen as kth big}\big) = \frac{1}{m-k+1}.
$$

Also by symmetry,

$$
\mathbb{P}\{\text{big 1 chosen as kth big}\} = \frac{1}{m}.
$$

Average out using the conditioning rule E4 to deduce

$$
\mathbb{P}\{B_1 = +1\} = \frac{1}{m}\sum_{k=1}^{m} \frac{1}{m-k+1}.
$$

Notice that the summands run through the values $1/1$ to $1/m$ in reversed order.

When the values for $\mathbb{P}\{H_1 = +1\}$ and $\mathbb{P}\{B_1 = +1\}$ are substituted into <2.9>, the asserted answer to part (a) results. $\square$

© David Pollard