

Conditional Probability and the Dice Room Murders

Kevin S. Van Horn

July 19, 2015

1 The puzzle

Scott Aaronson has a wonderful book called *Quantum Computing Since Democritus* [1], and in his chapter on the anthropic principle he writes about a thought experiment (attributed to John Leslie) that he calls the *Dice Room*:

Imagine that there's a very, very large population of people in the world, and that there's a madman. What this madman does is, he kidnaps ten people and puts them in a room. He then throws a pair of dice. If the dice land snake-eyes (two ones), then he simply murders everyone in the room. If the dice do not land snake-eyes, then he releases everyone, then kidnaps 100 people. He now does the same thing: he rolls two dice; if they land snake-eyes, then he kills everyone, and if they don't land snake-eyes, then he releases them and kidnaps 1000 people. He keeps doing this until he gets snake-eyes, at which point he's done.

This scenario leads to the following question: if you are kidnapped, what is your probability of dying? He gives two answers. The first:

...the dice have a $1/36$ chance of landing snake-eyes, so you should only be a "little bit" worried (considering.)

And the second:

...consider, of people who enter the room, what the fraction is of people who ever get out. Let's say that it ends at 1000. Then, 110 people get out and 1000 die. If it ends at 10,000, then 1110 people get out and 10,000 die. In either case, about $8/9$ of the people who ever go into the room die.

... We can say that we're conditioning on a specific termination point, but that no matter what that point is, we get the same answer. It could be 10 steps or 50 steps, but no matter what the termination point is, almost all of the people who go into the room are going to die, because the number of people is increasing exponentially.

(Actually, $1000/1110 \approx 9/10$ rather than $8/9$, so we'll use $9/10$ in the rest of this note.)

Aaronson comments that "If you're a Bayesian, then this kind of seems like a problem," as Bayesian reasoning seems to give two different answers to the same question.

2 Commentary

Right off the bat, you should be smelling something rotten here: Bayesian reasoning is just unadorned probability theory, nothing more and nothing less, and we know that the rules of probability theory are logically consistent—or rather, if they're not, then Zermelo-Fraenkel set theory is also inconsistent and mathematicians start throwing themselves off of tall buildings.

The other reason you should be skeptical of the above argument is that it's purely verbal. It is extremely easy to make mistakes in applying probability theory if you just give intuitive verbal arguments without actually grinding through the details of the math. It's easy to make mistakes even if you *do* most of the math except for some small step that seems obvious (see, for example, my analysis of the Marginalization Paradox [2].)

We shall see that the error in Aaronson's (Leslie's?) reasoning lies in dropping a conditioning proposition when assessing a probability; that is, an unconditional probability $\Pr(X)$ is implicitly used when a conditional probability $\Pr(X | Z)$ is called for. The conditioning proposition Z at first glance may appear irrelevant, but it greatly alters the probability of X .

3 Let's get explicit

To start the analysis, let's be crystal clear on our assumptions:

1. Define

$$s(j) = \sum_{t=1}^j 10^t = \frac{10}{9} (10^j - 1).$$

That is, $s(j)$ is the total number of people kidnapped up to the j -th step inclusive (if the process gets that far). This assumes that no person is ever kidnapped more than once, which appears to be implicit in Aaronson's description.

2. Assume that the population of the world is

$$N = s(n)$$

for some positive integer n . This assumption is just a convenience so that we don't have to deal with any "left-overs."

3. Assume that n is large enough that that it is almost certain that the madman will get snake-eyes before he runs out of people to kidnap. For example, $n = 300$ gives a probability of about 0.9998 that the madman eventually gets snake-eyes.
4. Writing $M(t)$ for "the madman murders at step t " and $\text{hm}(t)$ for "the madman has already murdered by step t ," the madman's choice process is as follows, for $1 \leq t \leq n$:

$$\begin{aligned} \text{hm}(1) &= \text{false} \\ \text{hm}(t+1) &= (\text{hm}(t) \text{ or } M(t)) \\ \Pr(M(t) \mid \text{hm}(t)) &= 0 \\ \Pr(M(t) \mid \neg\text{hm}(t)) &= p \\ &= \frac{1}{36} \end{aligned}$$

where " \neg " means "not." That is, once the madman has murdered he does not do so again, and if he has not murdered in a prior step then there is a probability p he will murder in the current step.

5. The madman selects victims in a random order. Let v be a random permutation of the numbers from 1 to N ; that is, we take all possible permutations of 1 to N to be equally probable values for v . If we imagine that the individuals in the population are numbered from 1 to N , then $v_i = 1$ means that individual i is the first selected, $v_i = 100$ means that individual i is the 100th to be selected, and so on. We can always extend the random ordering out past the point where the slaughter occurs, so we take v_i to be meaningful regardless of whether or not individual i is ever kidnapped. Since all permutations are equally likely, we have

$$\Pr(v_i = k) = \frac{1}{N}$$

for all integers $1 \leq k \leq N$.

6. The random ordering defines n batches of individuals, where batch 1 has the first 10^1 individuals, batch 2 has the next 10^2 individuals, and so on. Define $B(i)$ to be the number of the batch to which individual i belongs; specifically,

$$B(i) = t \Leftrightarrow s(t-1) < v_i \leq s(t).$$

7. Individual i dies, written $D(i)$, if the madman murders on the step corresponding to the batch to which individual i belongs:

$$D(i) = M(B(i)).$$

8. Individual i gets kidnapped, written $K(i)$, if the madman has not yet murdered on the step corresponding to the batch to which individual i belongs:

$$K(i) = \neg \text{hm}(B(i)).$$

9. You are some individual i , and the information you have available is that you have been kidnapped. Thus, the probability you are interested in is

$$\Pr(D(i) \mid K(i)).$$

4 Alternative decompositions

The first step in formalizing either of Aaronson's arguments is to decompose the problem by batch:

$$\begin{aligned} \Pr(D(i) \mid K(i)) &= \Pr(M(B(i)) \mid K(i)) \\ &= \sum_{t=1}^n \Pr((B(i) = t) \ \& \ M(t) \mid K(i)). \end{aligned}$$

That is, the probability of being murdered is the sum of the probabilities of being murdered at each of the n possible steps.

The next step is to apply the product rule to P_t , defined as

$$P_t \equiv \Pr((B(i) = t) \ \& \ M(t) \mid K(i)).$$

The product rule states that for any three propositions X , Y , and Z , we may decompose that probability that both X and Y are true, conditional on Z being true, as

$$\Pr(X \& Y \mid Z) = \Pr(X \mid Z) \Pr(Y \mid X \& Z).$$

Since $X \& Y$ is equivalent to $Y \& X$, this gives two different possible decompositions, and that is where the two analyses differ:

1. The first analysis factors P_t as

$$\Pr(B(i) = t \mid K(i)) \cdot \Pr(M(t) \mid (B(i) = t) \& K(i))$$

and implicitly claims that

$$\Pr(M(t) \mid (B(i) = t) \& K(i)) \stackrel{?}{=} \Pr(M(t) \mid \neg \text{hm}(t)) = \frac{1}{36}$$

regardless of t . (That is, knowing that i belongs to batch t and is kidnapped provides no information relevant to the probability of the madman murdering at step t beyond establishing that he has not yet murdered by step t .) Since individual i certainly belongs to one of the n batches, we then get

$$\begin{aligned} \Pr(D(i) \mid K(i)) &= \frac{1}{36} \sum_{t=1}^n \Pr(B(i) = t \mid K(i)) \\ &= \frac{1}{36}. \end{aligned}$$

2. The second analysis factors P_t as

$$\Pr(M(t) \mid K(i)) \cdot \Pr(B(i) = t \mid M(t) \& K(i))$$

and implicitly claims two things: first, that

$$\Pr(B(i) = t \mid M(t) \& K(i)) \stackrel{?}{=} \frac{10^t}{s(t)} \approx \frac{9}{10}$$

(if i is kidnapped at some point and the madman murders at step t , then i has a probability of about 9/10 of being in batch t), and second that

$$\sum_{t=1}^n \Pr(M(t) \mid K(i)) \stackrel{?}{\approx} 1$$

(it is highly probable that the madman eventually murders, and knowing that i is kidnapped does not change this). This then yields

$$\begin{aligned} \Pr(D(i) | K(i)) &\approx \frac{9}{10} \sum_{t=1}^n \Pr(M(t) | K(i)) \\ &\approx \frac{9}{10} \end{aligned}$$

in contradiction to the earlier answer of $1/36$.

To resolve the apparent paradox, we then have to determine which of these three claims are valid and which are not:

Claim 1 $\Pr(M(t) | (B(i) = t) \ \& \ K(i)) \stackrel{?}{=} \Pr(M(t) | \neg \text{hm}(t))$.

Claim 2 $\Pr(B(i) = t | M(t) \ \& \ K(i)) \stackrel{?}{=} 10^t/s(t)$.

Claim 3 $\sum_{t=1}^n \Pr(M(t) | K(i)) \stackrel{?}{\approx} 1$.

5 Conditional independence

All three of the claims involve conditional probabilities. To evaluate the claims we need to think about (conditional) independence of variables. Two propositions X and Y are independent, conditional on Z , if

$$\Pr(X | Y \ \& \ Z) = \Pr(X | Z)$$

or, equivalently, if

$$\Pr(Y | X \ \& \ Z) = \Pr(Y | Z).$$

Two variables x and y are independent, conditional on Z , if the propositions $x = X$ and $y = Y$ are independent, conditional on Z , for any possible pair of values X and Y .

Figure 1 shows the Bayesian network graph for our problem when $n = 4$. There is a node for every variable, and for every variable there are arcs to it from those variables used in its immediate definition. You can read off conditional independencies directly from this graph, using the notion of *d-separation* [3], although I won't go into all the details here.

The important point we'll use is that if every undirected path between two variables passes through a "collider" node that is not part of the information you are conditioning on (Z in the above discussion), then those two variables are conditionally independent. A collider node on an undirected path is a

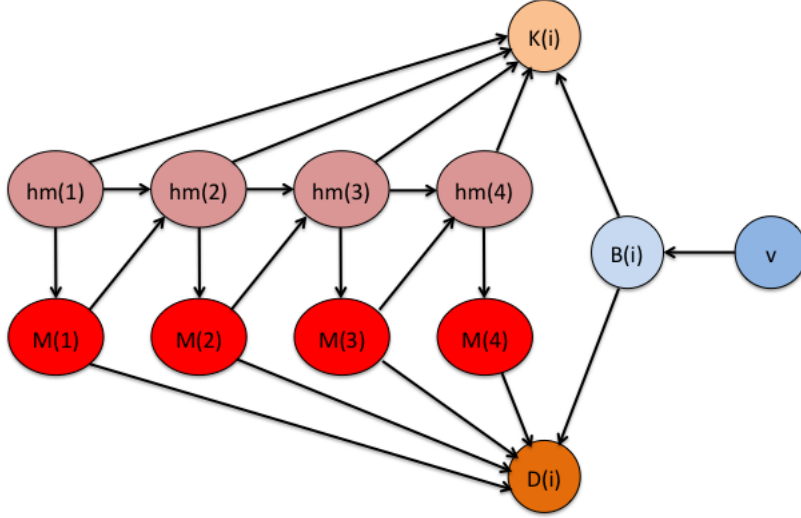


Figure 1: Bayesian network for $n = 4$

variable that has both arcs pointing in to it. For example, in the undirected path

$$\text{hm}(4) \rightarrow M(4) \rightarrow D(i) \leftarrow B(i)$$

between $\text{hm}(4)$ and $B(i)$, the variable $D(i)$ is a collider node.

6 Verifying Claim 1

We first note that, for $1 \leq t \leq n$,

$$(B(i) = t) \ \& \ K(i) \quad \Leftrightarrow \quad (B(i) = t) \ \& \ \neg \text{hm}(t).$$

If you belong to batch t , then you are kidnapped if and only if the madman has not murdered before step t . So

$$\Pr(M(t) \mid (B(i) = t) \ \& \ K(i)) = \Pr(M(t) \mid (B(i) = t) \ \& \ \neg \text{hm}(t)).$$

From Figure 1 we can see that $M(t)$ is independent of $B(i)$ conditional on $\text{hm}(t)$: all undirected paths between $M(t)$ and $B(i)$ include either the “collider”

$$\rightarrow D(i) \leftarrow$$

or the collider

$$\rightarrow K(i) \leftarrow$$

and we are not conditioning on either $D(i)$ or $K(i)$. So we have

$$\Pr(M(t) \mid (B(i) = t) \ \& \ \neg \text{hm}(t)) = \Pr(M(t) \mid \neg \text{hm}(t))$$

and Claim 1 is verified.

This tells us that the first analysis is, in fact, correct:

$$\Pr(D(i) \mid K(i)) = 1/36.$$

7 Verifying Claim 2

We first note that, for $1 \leq t \leq n$,

$$M(t) \ \& \ K(i) \quad \Leftrightarrow \quad M(t) \ \& \ (B(i) \leq t).$$

If the madman murders at step t , then you are kidnapped if and only if you are in batch t or an earlier batch. So

$$\Pr(B(i) = t \mid M(t) \ \& \ K(i)) = \Pr(B(i) = t \mid M(t) \ \& \ (B(i) \leq t)).$$

All undirected paths between $B(i)$ and $M(t)$ are again blocked by one of the colliders $\rightarrow D(i) \leftarrow$ or $\rightarrow K(i) \leftarrow$, so $B(i)$ and $M(t)$ are independent conditional on $B(i) \leq t$, giving

$$\Pr(B(i) = t \mid M(t) \ \& \ (B(i) \leq t)) = \Pr(B(i) = t \mid B(i) \leq t) = \frac{10^t}{s(t)} \approx \frac{9}{10}$$

and Claim 2 is verified.

8 Claim 3: not verified

As you have no doubt guessed by now, Claim 3 does not check out. Although it is true that

$$\sum_{t=1}^n \Pr(M(t)) \approx 1,$$

we *cannot* substitute $\Pr(M(t))$ for $\Pr(M(t) \mid K(i))$. $M(t)$ and $K(i)$ are dependent, due to the undirected path between them that run through the variables $\text{hm}(t)$ (among others.) *This is where the second analysis falls down.*

In fact, we find that

$$\sum_{t=1}^n \Pr(M(t) | K(i)) \approx \frac{10}{9} \cdot \frac{1}{36};$$

that is, if you know that you are kidnapped, but you don't know at which step, then it is highly likely that the madman never murders anyone at all! Furthermore, this holds no matter how large you make n . The intuition behind this is that it is highly likely that you are in batch n , as batch n contains about 9 times as many people as all other batches combined. But if you are in batch n and are kidnapped, then there is only a probability of $1/36$ that the madman will roll snake eyes at this final step; most likely he will not roll snake eyes and, having run out of people to kidnap, will not commit the planned murder.

9 Details on Claim 3

Having given the hand-wavy verbal argument, let's grind through the math. Using the standard identity for conditional probabilities,

$$\begin{aligned} \Pr(M(t) | K(i)) &= \Pr(M(t) \& K(i)) / \Pr(K(i)) \\ &= \Pr(M(t) \& (B(i) \leq t)) / \Pr(K(i)) \\ &= \Pr(M(t)) \Pr(B(i) \leq t) / \Pr(K(i)) \end{aligned}$$

with the last step justified by the fact that $M(t)$ and $B(i)$ are independent, as discussed in verifying Claim 2. But

$$\Pr(M(t)) = p(1-p)^{t-1}$$

and

$$\Pr(B(i) \leq t) = s(t)/s(n)$$

so

$$\begin{aligned} \Pr(M(t) | K(i)) &= \frac{p(1-p)^{t-1} s(t)}{s(n) \Pr(K(i))} \\ &= \frac{10p}{9s(n) \Pr(K(i))} (1-p)^{t-1} (10^t - 1). \end{aligned}$$

We can check from the graph that $B(i)$ and $\text{hm}(t)$ are independent, so

$$\begin{aligned}
\Pr(K(i)) &= \Pr(\neg \text{hm}(B(i))) \\
&= \sum_{t=1}^n \Pr(B(i) = t \ \& \ \neg \text{hm}(t)) \\
&= \sum_{t=1}^n \Pr(B(i) = t) \Pr(\neg \text{hm}(t)) \\
&= \sum_{t=1}^n \frac{10^t}{s(n)} (1-p)^{t-1}.
\end{aligned}$$

Putting these two equalities together gives

$$\begin{aligned}
\sum_{t=1}^n \Pr(M(t) \mid K(i)) &= \frac{10p \sum_{t=1}^n (1-p)^{t-1} (10^t - 1)}{9 \sum_{t=1}^n 10^t (1-p)^{t-1}} \\
&= \frac{10p}{9} \left(1 - \frac{\sum_{t=1}^n (1-p)^t}{\sum_{t=1}^n 10^t (1-p)^t} \right) \\
&\approx \frac{10}{9} p.
\end{aligned}$$

10 Conclusion

In summary, we have found that Aaronson’s first analysis of the Dice Room scenario is correct, and the second analysis, based on considering what proportion of the people who ever go into the Dice Room die, is not. The latter analysis implicitly (and erroneously) reduces the conditional probability $\Pr(M(t) \mid K(i))$ —the probability that the madman murders at step t , given that individual i is kidnapped—to the unconditional probability $\Pr(M(t))$. The conditioning information $K(i)$ turns out to strongly affect the probability of $M(t)$.

References

- [1] Aaronson, Scott (2013). *Quantum Computing Since Democritus*. Cambridge University Press.
- [2] Van Horn, Kevin S. (2003). “Notes on Jaynes’s Treatment of the Marginalization Paradox.” <http://ksvanhorn.com/bayes/Papers/mp.pdf>.

- [3] Geiger, Dan, Thomas Verma, and Judea Pearl (1990). “Identifying independence in Bayesian networks,” *Networks* 20, pp. 507–534. Online at http://ftp.cs.ucla.edu/pub/stat_ser/r116.pdf.