Question:. Consider an island with n+2 inhabitants. One of them is killed, and the murderer must be one of the inhabitants of the island. Police investigators discover a DNA profile at the scene of the crime. Scientists are able to say that this particular DNA profile occurs in a fraction p of all people. Now the police starts a big screening of all the inhabitants of the island. The first person to be screened, letÂąÂŕs call her Mary Shyamala Ahmed, turns out to have this particular DNA profile. What is the probability that Mary Shyamala Ahmed is the murderer?

Yogesh asked me this question on Conditional probability question from Ronald Meester's book
https://www.springer.com/gp/book/9783034877862.
The character name has been changed in my question. It was John Smith.

Solution 1: [From Meester's book:] In order to say something about this, we need to turn the situation into a real mathematical model. We assign to every inhabitant of the island a DNA profile, and the probability that someone has the profile found at the scene of the crime is p. Different persons have independent profiles. Furthermore, we assume that each person is the murderer with probability $\frac{1}{(n+1)}$.

Denote the event that John Smith is the murderer by G, and denote the event of finding John SmithsâĂŹ particular DNA profile at the scene of the crime by E. Before seeing this DNA profile, any inhabitant of the island is equally likely to be the murderer, and we therefore have $P(G) = 1/(n + 1)$ and $P(G^c) = n/(n + 1)$. We want to compute the probability of G after the DNA evidence E, that is, we want to compute $P(G|E)$. If John Smith is the murderer, the event E occurs with probability 1, that is, $P(E|G) = 1$. If John Smith is not the murderer, then the real murderer has with probability p the same profile as John Smith, and therefore, $P(E|G^c) = p$. We can now, with this information, compute the probability that John Smith is the murderer, given the event that his DNA profile was found at the scene of the crime, that is, we can compute $P(G|E)$ from Bayes' formula to be

$$\frac{1}{1 + pn}.$$

Ritvik RR Solution 1:

Let Y be the random variable that is $\underline{1}$ if the first person is the murderer and 0 otherwise.
Let X be the number ~~of~~ ~~among the n people (other than Mary)~~ of people with the same DNA profile as the murderer (and Mary) among the n ~~re~~ remaining people on the island. Since a fracti[on] p of the population of the world has this profile, so $X \sim Bin(n, p)$. We want to compute $P(Y=1)$. ~~We~~ We know that $P(Y=1 \mid X=i) = \frac{1}{i+1}$.
So, we get:

~~Therefore~~ $P(Y=1) = \sum_{i=0}^{n} P(Y=1 \mid X=i)\, P(X=i)$

$$= \sum_{i=0}^{n} \frac{1}{i+1} \binom{n}{i} p^{i} (1-p)^{n-i}$$

$$= \frac{1}{(n+1)p} \sum_{i=0}^{n} \binom{n+1}{i+1} p^{i+1} (1-p)^{n-i}$$

$$= \frac{1}{(n+1)p} \left(1-(1-p)^{n+1}\right)$$

$$= \frac{\left(1-(1-p)^{n+1}\right)}{(n+1)p}.$$

Note: the expected number of people among the $(n+1)$ alive in the island that have this profile is $(n+1)p$, and $P(Y=1) \approx \frac{1}{(n+1)p}$.

Nitya G Solution

The DNA profile occurs in fraction p of all people

$\Rightarrow$ Every inhabitant of the island can have it with prob p.

MSA has this DNA profile.

If none of the other inhabitants of the island have the profile, MSA is the murderer with probability 1.

If one other has it, MSA is the murderer with prob $1/2$, and so on

$\therefore$

Prob(MSA is the murderer)

$$= 1 \times {}^nC_0 (1-p)^n + \frac{1}{2} \times {}^nC_1\, p(1-p)^{n-1} + \frac{1}{3} \times {}^nC_2\, p^2 (1-p)^{n-2} + \dots$$

$$+ \frac{1}{n}\, {}^nC_{n-1}\, p^{n-1} (1-p)^1 + \frac{1}{n+1}\, {}^nC_n\, p^n (1-p)^0$$

$$= \frac{1}{(n+1)p} \left[{}^{n+1}C_1 (1-p)^n p + {}^{n+1}C_2\, p^2 (1-p)^{n-1} + \dots + {}^{n+1}C_n\, p^n (1-p)^1 + {}^{n+1}C_{n+1} (p)^{n+1} \right]$$

The term in the bracket can be compared to a $\mathrm{Bin}(n+1, p)$

We know $\sum_{i=0}^{n+1} {}^{n+1}C_i\, p^i (1-p)^{n+1-i} = 1$

$\therefore$ The required prob

$$= \frac{1}{(n+1)p} \left[1 - {}^{n+1}C_0 (1-p)^{n+1} \right]$$

$$= \frac{1}{(n+1)p} \left(1 - (1-p)^{n+1} \right)$$

Paradox

$$\text{Is it } \frac{1-(1-p)^{n+1}}{(n+1)p} \text{ ?}$$

OR

$$\text{Is it } \frac{1}{1+np} \text{ ?}$$

We had a discussion in my office along with Manjunath Krishnapur. The key is what information is it that you are conditioning on ? Here are two questions:

1. Take a look at the famous Tuesday Birthday Problem .

2. Suppose you choose a point uniformly on the square. What is the probability of the x-coordinate bigger than 0.5 conditional on the event the chosen point lies on the line $y = x$. ?

We need to just condition properly.

Manjunath Krishanpur and Yogesh II : Let us say there are k gene types and each person has gene type i with probability p_i and independent of other persons.

Let person 1 be the murderer and WLOG assume his gene type is 1. Person J (chosen uniformly at random) is selected for testing. Denote U_i to be the gene type of person i.

Then the question is $P(J = 1|U_J = i, U_1 = i)$.

The numerator is $P(J = 1, U_1 = i) = p_1/(n + 1)$.

The denominator is

$$\begin{aligned} P(U_J = U_1 = i) &= \sum_{j=1}^{n+1} P(U_J = U_1 = i|J = j)P(J = j) \\ &= (p_1/(n + 1)) + P(U_2 = U_1 = 1)n/(n + 1) \\ &= (1 + np_1)(p_1/(n + 1)). \end{aligned}$$

So the overall probability is $1/(1 + np_1)$.

Ritvik RR : revised version of their solutions

Experiment :

We have $(n+1)$ people · person 1, person 2, ..., person$(n+1)$.

1. We assign each person a DNA with the appropriate probabilities.

2. We choose a person uniformly at random and call him the murderer. We observe his DNA.

3. We choose a person uniformly at random again, independent of the previous choice, and observe his DNA.

Now, we are given the following :

1. Suppose person i was the murderer. We are given that person i had DNA type d.

2. Suppose person j was chosen in 3 of the experiment. We are given that person j has DNA type d.

Let M be the event that $j = i$ and D the event that person j had DNA type d. We want to compute $P(M \mid D)$.

Let N_k be the event that k people other than person j had DNA type d after 1 in the experiment.

Then, $P(M \mid D) = \dfrac{P(M \text{ and } D)}{P(D)}$.

Note that $P(M \text{ and } D) = P(M) = 1/n+1$.

Now, $P(D) = \sum_{k=0}^{n} P(D|N_k)\, P(N_k)$

$$= \sum_{k=0}^{n} \frac{k+1}{n+1} \binom{n}{k} p^k (1-p)^{n-k}$$

$$= \frac{1}{n+1} \sum_{k=0}^{n} (k+1) \binom{n}{k} p^k (1-p)^{n-k}$$

$$= \frac{1+np}{n+1}$$

Therefore, $P(M|D) = \dfrac{1/n+1}{(1+np)/(n+1)} = \dfrac{1}{1+np}$.

<u>Fixing Our Incorrect Solution</u>

We note that $P(M|D) = \sum_{k=0}^{n} P(M|D, N_k)\, P(N_k|D)$.

But $P(N_k|D)$ is not $\binom{n}{k} p^k (1-p)^{n-k}$!

(In our original solution we just wrote $P(N_k)$)

Now, $P(M|D, N_k)$ is $1/(k+1)$, and $P(N_k|D) = \dfrac{P(N_k, D)}{P(D)}$.

First, $P(N_k, D) = \binom{n}{k} p^k (1-p)^{n-k} \times \left(\frac{k+1}{n+1}\right)$

Second, $P(D) = \dfrac{(1+np)}{(n+1)}$ (done above).

So, $P(N_k|D) = \binom{n}{k} p^k (1-p)^{n-k} \dfrac{(k+1)}{(1+np)}$.

Therefore, $P(M|D) = \sum_{k=0}^{n} \left(\frac{1}{k+1}\right)\left(\frac{k+1}{1+np}\right)\binom{n}{k} p^k (1-p)^{n-k} = \frac{1}{1+np}$.