

Exercise 2.1: Is it possible to find a general formula for $p(C|A + B)$, analogous to (2 – 48), from the product and sum rules? If so, derive it; if not, explain why this cannot be done.

Answer:

Sure. We've derived Bayes theorem without naming it, which is simply product rule and commutativity, so we can say

$$\begin{aligned} p(C|A + B) &= \frac{p(C) p(A + B|C)}{p(A + B)} \\ &= \frac{p(C) (p(A) + p(B) - p(AB|C))}{p(A) + p(B) - p(AB)}. \end{aligned}$$

Exercise 2.2: Now suppose we have a set of propositions $\{A_1, \dots, A_n\}$ which on information X are mutually exclusive: $p(A_i A_j|X) = p(A_i|X)\delta_{ij}$. Show that $p(C|(A_1 + A_2 + \dots + A_n)X)$ is a weighted average of the separate plausibilities $p(C|A_i X)$:

$$p(C|(A_1 + \dots + A_n)X) = p(C|A_1 X + A_2 X + \dots + A_n X) = \frac{\sum_i p(A_i|X)p(C|A_i X)}{\sum_i p(A_i|X)}.$$

Answer: For $i \neq j$,

$$\begin{aligned} p(C|A + B) &= \frac{p(A + B|C)p(C)}{p(A + B)} \\ &= \frac{(p(A) + p(B) - p(AB|C))p(C)}{p(A) + p(B) - p(AB|C)}. \end{aligned}$$

Iterating this over $\sum_i A_i$ we get

$$p\left(\sum_i A_i \middle| X\right) = \sum_i p(A_i|X).$$

Thus, and in a similar manner to the above exercise,

$$\begin{aligned} p\left(C \middle| \sum_i A_i X\right) &= \frac{p(\sum_i A_i|CX) p(C|X)}{p(\sum_i A_i|X)} \\ &= \frac{\sum_i p(A_i|CX)p(C|X)}{\sum_i p(A_i|X)} \\ &= \frac{\sum_i p(A_i|X)p(C|A_i X)}{\sum_i p(A_i|X)}. \end{aligned}$$

Exercise: Let A_i mutually exclusive ($P(A_i A_j) = P(A_i)\delta_{ij}$). Then, $P(\sum_i A_i) = \sum_i P(A_i)$.

Answer:

If $i \in \{1, 2\}$, then

$$P(\sum_i A_i) = P(A_1 + A_2) \tag{1}$$

$$= P(A_1) + P(A_2) - \cancel{P(A_1 A_2)} \xrightarrow{0} \tag{2}$$

$$= \sum_i P(A_i). \tag{3}$$

Then, by induction,

$$P\left(\sum_{i=1}^{N+1} A_i\right) = P\left(\sum_{i=1}^N A_i\right) + P(A_{N+1}) - P\left(\sum_{i=1}^N A_i A_{N+1}\right) \quad (4)$$

$$= \sum_{i=1}^N P(A_i) + P(A_{N+1}) - \sum_{i=1}^N \underbrace{(A_i A_{N+1})}_0 \quad (5)$$

$$= \sum_{i=1}^N P(A_i). \quad (6)$$

Exercise 2.3: Limits on Probability Values: As soon as we have the numerical values $a = P(A|C)$ and $b = P(B|C)$, the product and sum rules place some limits on the possible numerical values for their conjunction and disjunction. Supposing that $a \leq b$, show that the probability of the conjunction cannot exceed that of the least probable proposition: $0 \leq P(AB|C) \leq a$, and the probability of the disjunction cannot be less than that of the most probable proposition: $b \leq P(A + B|C) \leq 1$. Then show that, if $a + b > 1$, there is a stronger inequality for the conjunction; and if $a + b < 1$ there is a stronger one for the disjunction. These necessary general inequalities are helpful in detecting errors in calculations.

Answer:

For the conjunction,

$$P(AB|C) = P(A|C)P(B|AC) \quad (7)$$

$$= aP(B|AC) \leq a. \quad (8)$$

Since $P(AB|C) = b + a - P(A + B|C)$, if $b + a > 1$ then $P(AB|C) > 1 - P(A + B|C) = P(\overline{A + B}|C)$.

For the disjunction,

$$P(A + B|C) = P(A|C) + P(B|C) - P(AB|C) \quad (9)$$

$$= b + a - P(AB|C) \quad (10)$$

$$\geq b. \quad (11)$$

And also, $b + a < 1$ implies $P(A + B|C) < 1 - P(AB|C) = P(\overline{AB}|C)$.

Exercise 3.1: Why isn't the multiplicity factor (3-16) just $n!$? After all, we started this discussion by stipulating that the balls, in addition to having colors, also carry labels $(1 \cdots N)$, so that different permutations of the red balls among themselves, which give the $r!$ in the denominator of (3-16), are distinguishable arrangements.

Answer:

We've decomposed the event "drawing r red balls in n draws" as the union of events of the form "draw r red balls in n draws with the particular permutation of orders p " for permutations p , which are mutually exclusive and identically plausible. At doing this, we're using the events R_i and W_i , with the probability given by (3-14). If we were to use "drawing the following n balls, where r are red", we should be accounting for the probability B_{ij} of drawing the i th ball after j draws, which would incorporate the missing factors.

Exercise 3.2: Probability of a Full Set: Suppose an urn contains $N = \sum_i N_i$ balls, N_1 of color 1, N_2 of color 2, \dots , N_k of color k . We draw m balls without replacement; what is the probability that we have at least one of each color? Supposing $k = 5$, all $N_i = 10$, how many do we need to draw in order to have at least 90% probability of getting a full set?

Answer:

We're asked $P(\sum_{r \in I} \prod_i r_i)$, where $I = \{(r_i) : r_i \geq 1 \forall i\}$. We can decompose this as

$$P\left(\sum_{r \in I} \prod_i r_i | B\right) = \sum_{r \in I} P(r_1 \cdots r_k | B) \quad (12)$$

$$= \frac{1}{\binom{N}{m}} \sum_{r \in I} \prod_i \binom{N_i}{r_i} \quad (13)$$

A calculation is probably better done by calculating the missing terms instead, which are those for which at least one of those are 0, of which there are less, since at least one of the summands is already 0, the rest varying between the same intervals.

The “3.2.py” script does the maths, and outputs $m = 15$ as the solution, with a probability of about 0.91.

Exercise 3.3: Reasoning Backwards: Suppose that in the previous exercise k is initially unknown, but we know that the urn contains exactly 50 balls. Drawing out 20 of them, we find 3 different colors; now what do we know about k ? We know from deductive reasoning (i.e., with certainty) that $3 \leq k \leq 33$; but can you set narrower limits $k_1 \leq k \leq k_2$ within which it is highly likely to be?

Answer:

I suppose this is the first use case of Bayes theorem as it's usually used. By letting the hypotheses $H_k =$ “There are k different colors in the urn” and the data $D =$ “3 different colors have been seen after 20 draws”, we want $P(H_k|D)$, but what our model tells us with certainty is $P(D|H_k)$, so we'll use Bayes' theorem for this:

$$P(H_k|D) = \frac{P(D|H_k)P(H_k)}{\sum_k P(D|H_k)P(H_k)}. \quad (14)$$

I will assume each H_k is equally likely, since I've got no information on anything. Thus,

$$P(H_k|D) = \frac{P(D|H_k)}{\sum_k P(D|H_k)}. \quad (15)$$

We're now left with the task of finding the probability that 3 different colors are seen in a set of 20 taken from 50, given there are k different colors. To this end, we split $D = \sum_N DH_N$, where $N \in \mathbb{N}^k$ is such that N_i is the number of balls with the i th color, and get

$$P(D|H_k) = P\left(\sum_N DH_N \middle| H_k\right) \quad (16)$$

$$= \sum_N P(DH_N|H_k) \quad (17)$$

$$= \sum_N P(D|H_N H_k) P(H_N|H_k). \quad (18)$$

We once again propose an homogeneous distribution over the N , which gives $P(H_N|H_k) = \binom{50-k}{k-1}^{-1}$.

Now, let's think of the probability of drawing from only the colors 1, 2, and 3 in all draws. For the m th draw, the probability is

$$\frac{51 - m - \sum_{l=3}^{k-2} N_l}{51 - m} = \frac{51 - m - N_{\text{rem}}}{51 - m}, \quad (19)$$

so for the 20 draws, we get

$$\frac{(50 - N_{\text{rem}})! 30!}{(30 - N_{\text{rem}})! 50!}. \quad (20)$$

So, the chances of drawing only the first 3 colors are the sum over all of the N of this term. For each choice of N_{rem} , there are $\binom{47 - N_{\text{rem}}}{2} = (47 - N_{\text{rem}})(46 - N_{\text{rem}})$ such terms. And, by symmetry of label permutation, for each choice of 3 colors the result is the same, so finally, by splitting D into the sum of “drawing only for these 3 particular colors”, we sum over these $k!/(k-3)!$ choices, and get our final result:

$$P(D|H_k) = \frac{k!}{(k-3)!} \sum_{N_{\text{rem}}=k-3}^{47} (47 - N_{\text{rem}})(46 - N_{\text{rem}}) \frac{(50 - N_{\text{rem}})!}{(30 - N_{\text{rem}})!} \frac{30!}{50!} \binom{50 - k}{k - 1}^{-1} \quad (21)$$

The “3.3.py” script implements this solution, giving the probability distribution shown in Figure ??

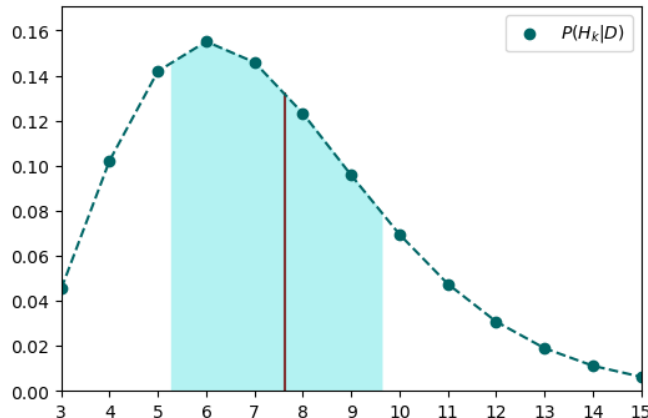


Figure 1: Probability distribution $P(H_k|D)$ for different values of k . The shaded region goes between first and third quartile, and the vertical line denotes the (linearly extrapolated) median.

Exercise 3.4: Matching: The M urns are now numbered 1 to M , and M balls, also numbered 1 to M , are thrown into them, one in each urn. If the numbers of a ball and its urn are the same, we have a match. Show that the probability of at least one match is

$$P = \sum_{k=1}^M (-1)^{k+1} / k!.$$

As $M \rightarrow \infty$, this converges to $1 - 1/e = 0.632$. The result is surprising to many, because however large M is, there remains an appreciable probability of no match at all.

Answer:

Indeed, it surprises me. But, we’re throwing infinite uniquely labeled balls to infinite uniquely labeled urns. It’s no surprise it doesn’t help a lot. It’s also surprising that the probability lowers when going from odd k to even k . The proposition \mathcal{M} = “There is at least one match” can be decomposed as the sum of \mathcal{M}_i =

“There is a match in urn i ”, so

$$P(\mathcal{M}) = P\left(\sum_{i=1}^M \mathcal{M}_i\right) \quad (22)$$

$$= P(\mathcal{M}_1) + P\left(\sum_{i=2}^M \mathcal{M}_i\right) - P\left(\mathcal{M}_1 \sum_{i=2}^M \mathcal{M}_i\right) \quad (23)$$

$$= \dots \quad (24)$$

$$= \sum_{i=1}^M P(\mathcal{M}_i) - \sum_{k=2}^M P\left(\mathcal{M}_{k-1} \sum_{i=k}^M \mathcal{M}_i\right) \quad (25)$$

$$= \sum_{i=1}^M P(\mathcal{M}_i) - \sum_{k=2}^M P(\mathcal{M}_{k-1}) P\left(\sum_{i=k}^M \mathcal{M}_i \middle| \mathcal{M}_{k-1}\right). \quad (26)$$

Now, since the set of propositions $\mathcal{B}_{n,i}$ = “The ball n falls into the urn i ” is exhaustive, mutually exclusive, and symmetrical for every i , we have $P(\mathcal{B}_{n,i}) = 1/M$. By observing that $\mathcal{M}_i = \mathcal{B}_{i,i}$, we see $P(\mathcal{M}_i) = 1/M$, so that

$$P(\mathcal{M}) = 1 - \frac{1}{M} \sum_{k=2}^M P\left(\sum_{i=k}^M \mathcal{M}_i \middle| \mathcal{M}_{k-1}\right) \quad (27)$$

$$= 1 - \frac{1}{M} \quad (28)$$

The proposition \mathcal{M} = “There is at least one match” is the negation of \mathcal{N} = “There is no match at all”, which in turn can be decomposed as the product of propositions \mathcal{N}_i = “There is no match on box i ”. In terms of this negation, the exercise takes the form $P(\mathcal{N}) = \sum_{k=2}^M (-1)^{k+1}/k!$.

In these terms,

$$P(\mathcal{N}) = P\left(\prod_{i=1}^M \mathcal{N}_i\right) \quad (29)$$

$$= P(\mathcal{N}_1) P\left(\prod_{i=2}^M \mathcal{N}_i \middle| \mathcal{N}_1\right). \quad (30)$$

We can see that the propositions $\mathcal{B}_{i,j}$ = “Ball i goes to urn j ” for fixed j are exhaustive and equally likely, so that $P(\mathcal{B}_{i,j}) = 1/M$, and so $P(\mathcal{N}_1) = P\left(\sum_{i=2}^M \mathcal{B}_{i,1}\right) = (1 - 1/M)$. We keep on,

$$P(\mathcal{N}) = P(\mathcal{N}_1) P\left(\prod_{i=2}^M \mathcal{N}_i \middle| \mathcal{N}_1\right) \quad (31)$$

$$= (1 - 1/M) P(\mathcal{N}_2 | \mathcal{N}_1) \left(\prod_{i=3}^M \mathcal{N}_i \middle| \mathcal{N}_1 \mathcal{N}_2\right). \quad (32)$$

Now, since by symmetry $P(\mathcal{N}_2 | \mathcal{N}_1) = P(\mathcal{N}_1 | \mathcal{N}_2)$