

SECTION 1 | QUESTIONS WITH SOLUTIONS

Same Prompt, Twice

Probability, randomness and temperature in a toy language model

THE THREE RULES USED (REVIEW)

NOT

$$P(E^C) = 1 - P(E)$$

OR

exclusive: $P(E) + P(F)$
 otherwise: $P(E) + P(F) - P(E \text{ and } F)$

AND

independent: $P(E) \cdot P(F)$
 otherwise: $P(E) \cdot P(F | E)$

THE TOY MODEL

Prompt: A sleepy robot dreams about. The model has exactly five possible outputs. Each call picks one output at random, and calls do not affect each other (they are independent).

i	Output	Probability p_i
1	electric sheep	0.44
2	counting stars	0.20
3	sample spaces	0.14
4	counting outcomes	0.13
5	robot puppies	0.09

Check: $0.44 + 0.20 + 0.14 + 0.13 + 0.09 = 1.00$. The subscript in p_i means “the i -th row”; for example $p_4 = 0.13$.

QUESTION 1

If prompted twice, what is the probability the model produces two different outputs?

SOLUTION

Let E be the event that both calls produce *the same* output, and let B_i be the event that both calls produce output i .

Step 1: AND rule (calls are independent)

$$P(B_i) = p_i \cdot p_i = p_i^2$$

Step 2: OR rule (the B_i are mutually exclusive, so we add)

$$P(E) = P(B_1 \text{ or } B_2 \text{ or } \dots \text{ or } B_5) = \sum_i p_i^2$$

Step 3: NOT rule

$$P(E^C) = 1 - \sum_i p_i^2$$

Numbers

$$\begin{aligned} \sum_i p_i^2 &= 0.44^2 + 0.20^2 + 0.14^2 + 0.13^2 + 0.09^2 \\ &= 0.1936 + 0.0400 + 0.0196 + 0.0169 + 0.0081 = 0.2782 \\ P(E^C) &= 1 - 0.2782 = 0.7218 \approx 0.72 \end{aligned}$$

Note. Sanity check: if all five outputs were equally likely, $\sum_i p_i^2 = 5(0.2)^2 = 0.2$ and the answer would be 0.8, the largest possible. Our 0.72 is close to that, so the model is quite spread out.

ANSWER $P(\text{two different outputs}) = 1 - \sum_i p_i^2 \approx 0.72$

QUESTION 2

extra time

Each call says “robot puppies” with chance 0.09. If you prompt twice, what is the chance at least one call says it? Let R_i be the event that call i says “robot puppies”.

SOLUTION

“At least one” is the complement of “none”, and $P(R_i^C) = 1 - 0.09 = 0.91$.

$$\begin{aligned} P(\text{at least one}) &= 1 - P(R_1^C \text{ and } R_2^C) \\ &= 1 - P(R_1^C) \cdot P(R_2^C) \\ &= 1 - 0.91 \cdot 0.91 = 1 - 0.8281 = 0.1719 \approx 0.17 \end{aligned}$$

Note. Second route (OR rule, events overlap): $0.09 + 0.09 - 0.09 \cdot 0.09 = 0.1719$. The common mistake $0.09 + 0.09 = 0.18$ forgets to subtract the overlap 0.0081.

ANSWER $P(\text{at least one in 2 calls}) \approx 0.17$

QUESTION 3

extra time, advanced

Each call says “robot puppies” with chance 0.09. How many calls until the chance of seeing it is more than 50%? Let R_i be the event that call i says “robot puppies”.

SOLUTION

Step 1: complement, then independence

$$\begin{aligned}
 P(\text{at least one}) &= 1 - P(R_1^C \text{ and } \cdots \text{ and } R_n^C) \\
 &= 1 - P(R_1^C) \cdots P(R_n^C) = 1 - 0.91^n
 \end{aligned}$$

Step 2: solve $1 - 0.91^n > 0.5$, that is $0.91^n < 0.5$

Calls n	0.91^n	$P(\text{at least one}) = 1 - 0.91^n$
1	0.9100	0.09
5	0.6240	0.38
6	0.5679	0.43
7	0.5168	0.48
8	0.4703	0.53
10	0.3894	0.61
32	0.0489	0.95

Log check

$$0.91^n < 0.5 \iff n > \frac{\ln 0.5}{\ln 0.91} = \frac{-0.6931}{-0.0943} \approx 7.35,$$

so the smallest whole number is $n = 8$.

Note. Wrong shortcut: $0.09n = 0.5$ gives $n \approx 5.6$. Adding chances only works for mutually exclusive events, and the R_i can happen together. At $n = 6$ the true chance is just 0.43.

ANSWER 8 calls (7 calls give 0.48, 8 calls give 0.53).

QUESTION 4

Turn the temperature down. Will two calls give the same answer more often, or less often?

SOLUTION

Cold: the top choice takes over, so two calls almost always match.

Hot: the chances even out, so two calls differ more often.

Numerical check

Use the first-word chances q_i computed in Question 5. Two calls give the same first word with chance $\sum_i q_i^2$ (the same idea as Question 1).

$$T = 0.5 : 0.5863^2 + 0.3298^2 + 0.0594^2 + 0.0245^2 = 0.3438 + 0.1088 + 0.0035 + 0.0006 = 0.4567$$

$$T = 1 : 0.44^2 + 0.33^2 + 0.14^2 + 0.09^2 = 0.1936 + 0.1089 + 0.0196 + 0.0081 = 0.3302$$

$$T = 2 : 0.3469^2 + 0.3005^2 + 0.1957^2 + 0.1569^2 = 0.1203 + 0.0903 + 0.0383 + 0.0246 \approx 0.27$$

As T goes down, $\sum_i q_i^2$ goes up. In the limit $T \rightarrow 0$ the top option has chance 1 and two calls always match. As $T \rightarrow \infty$ the four options become equally likely, and the same-word chance falls to $4 \cdot (1/4)^2 = 0.25$.

ANSWER Colder means the same answer **more often**. Hotter means the same answer **less often**.

QUESTION 5

advanced

What does temperature, T , do to the chances?

SOLUTION**Notation (defined first)**

p_i is the old chance of option i , q_i is the new chance after applying temperature T , and $\sum_j$ means add over all options j .

$$q_i = \frac{p_i^{1/T}}{\sum_j p_j^{1/T}}$$

At $T = 0.5$ we have $1/T = 2$: square each chance, then rescale so they add to 1.

Step 1: chances for the first word

The chances on the slide are for the first word. "Counting" starts two outputs (counting stars and counting outcomes). They cannot happen together, so we add:

$$P(\text{counting}) = 0.20 + 0.13 = 0.33$$

First word	electric	counting	sample	robot	Sum
p_i	0.44	0.33	0.14	0.09	1.00

Step 2: square each chance

$$0.44^2 = 0.1936, \quad 0.33^2 = 0.1089, \quad 0.14^2 = 0.0196, \quad 0.09^2 = 0.0081$$

Step 3: add them (the denominator)

$$0.1936 + 0.1089 + 0.0196 + 0.0081 = 0.3302$$

Step 4: divide each square by 0.3302

$$q_{\text{electric}} = \frac{0.1936}{0.3302} = 0.5863 \approx 0.59$$

$$q_{\text{counting}} = \frac{0.1089}{0.3302} = 0.3298, \quad q_{\text{sample}} = \frac{0.0196}{0.3302} = 0.0594, \quad q_{\text{robot}} = \frac{0.0081}{0.3302} = 0.0245$$

Check: $0.5863 + 0.3298 + 0.0594 + 0.0245 = 1.0000$.

ANSWER $q_{\text{electric}} = \frac{0.44^2}{0.44^2 + 0.33^2 + 0.14^2 + 0.09^2} = \frac{0.1936}{0.3302} \approx 0.59$

Effect of cold and hot temperature (same method, exponent $1/T$)

Temperature	electric	counting	sample	robot
$T = 0.5$ (cold)	0.5863	0.3298	0.0594	0.0245
$T = 1$ (original)	0.4400	0.3300	0.1400	0.0900
$T = 2$ (hot)	0.3469	0.3005	0.1957	0.1569

For $T = 2$ the exponent is $1/2$, so we take square roots: 0.6633, 0.5745, 0.3742, 0.3000, which add to about 1.912.

ANSWER A low temperature makes the most likely output even more likely ($0.44 \rightarrow 0.59$), so two calls match more often. A high temperature evens the chances out, so they differ more often.

Note. Temperature does not multiply chances by T . It raises them to the power $1/T$ and then rescales. Rare options become rarer when cold ($0.09 \rightarrow 0.02$) but never impossible.