Bayes' Rule

Robert Y. Lewis

CS 0220 2024

April 19, 2024

Overview

Coupon collecting

Sometimes a company will run a coupon-collecting promotion. For example, each time you make a purchase, you get a coupon at random. If you collect all n coupons, you win a large prize.

How many purchases should you expect to make to win (collect all n coupons) if coupon types are given uniformly at random at each purchase?

Solving it

Consider a sequence of coupons that has all n types at the end. We want to know the *length* of this sequence.

Let X_i be the sequence of coupons from right *after* the $(i-1)$th unique coupon was received until the ith unique coupon is received.

Examples:

$$\underbrace{3}_{X_1}\ \underbrace{4}_{X_2}\ \underbrace{5}_{X_3}\ \underbrace{3\quad 3\quad 2}_{X_4}\quad \underbrace{2\quad 5\quad 1}_{X_5}$$

The length of the entire coupon sequence is the sum of the lengths of the X_is: $1+1+1+3+3=9$.

The *expected* length of the entire sequence is the sum of the expected lengths of the X_is.

What's Xk?

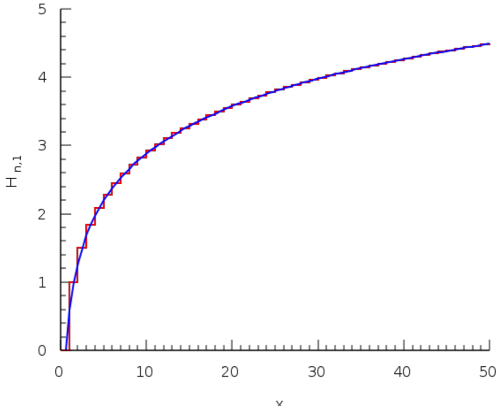
What's the length of X_k? At the beginning of segment k, we have collected $k-1$ different coupons from the n possibilities. Thus, the probability of getting one of the new ones is $p = \frac{n-k+1}{n}$.

What's the expected number of tries until a new one is found? $\frac{n}{n-k+1}$ by the "mean time to failure" analysis!

$\mathbb{E}[\sum_{k=1}^{n} \text{length}(X_k)]$	coup col
$= \sum_{k=1}^{n} \mathbb{E}[\text{length}(X_k)]$	lin of exp
$= \sum_{k=1}^{n} n/(n-k+1)$	mean fail time
$= n \sum_{i=1}^{n} 1/i$	$i = n-k+1$
$= nH(n)$	defn Harmonic number
$\approx n \ln(n)$	prop of H

Aside: harmonic numbers

$$H(n) = \sum_{i=1}^{n} \frac{1}{i} \approx \ln(n)$$



Examples

- People to poll before having representatives of each birthday: $365 \times H(365) \approx 2365$
- Ice skates before you have one of each: $2 \times H(2) = 3$
- Die rolls until you have one of each: $6 \times H(6) = 14.7$
- Number of randomly assigned pigeons until each hole is filled: $n \times H(n) \approx n \ln(n)$

Flipping the conditional probability

Theorem (Bayes' Theorem/Rule/Law): For any events A and B on any probability pace,

$$\Pr(A|B) = \frac{\Pr(B|A)\Pr(A)}{\Pr(B)}.$$

Proof:

$$\begin{aligned}
\Pr(A|B) &= \Pr(A \wedge B)/\Pr(B) && \text{defn cond prob} \\
\Pr(A|B)\Pr(B) &= \Pr(A \wedge B) && \text{mult by } \Pr(B) \\
\Pr(B|A) &= \Pr(A \wedge B)/\Pr(A) && \text{defn cond prob} \\
\Pr(B|A)\Pr(A) &= \Pr(A \wedge B) && \text{mult by } \Pr(A) \\
\Pr(A|B)\Pr(B) &= \Pr(B|A)\Pr(A) && \text{both are } \Pr(A \wedge B) \\
\Pr(A|B) &= \Pr(B|A)\Pr(A)/\Pr(B) && \text{div by } \Pr(B)
\end{aligned}$$

Interpreting Bayes' rule

Sometimes thought of in terms of *information*: measuring how the probability of a hypothesis A changes if you learn the evidence B. Reassociating:

$$\Pr(A|B) = \Pr(A)\frac{\Pr(B|A)}{\Pr(B)}.$$

- $\Pr(A)$: the *prior* probability of A (before learning B)
- $\Pr(A|B)$: the *posterior* probability of A (after learning B)
- $\frac{\Pr(B|A)}{\Pr(B)}$: the *support* B lends to A; measures the *compatibility* of the evidence B with the hypothesis A.

Virus test example

+test	person tests positive
+virus	person has the virus
−test	person tests negative
−virus	person does not have the virus

$$
\begin{aligned}
\Pr(+\text{test}| + \text{virus}) &= 0.85 && \text{true positive}\\
\Pr(+\text{test}| - \text{virus}) &= 0.10 && \text{false positive}\\
\Pr(+\text{virus}) &= 0.01 && \text{population prob}
\end{aligned}
$$

Test says you have the virus. What's the chance you have it?

$$
\begin{aligned}
&\Pr(+\text{virus}| + \text{test}) = \Pr(+\text{test}| + \text{virus}) \Pr(+\text{virus}) / \Pr(+\text{test}) && \text{Bayes rule}\\
&= 0.85 \times 0.01 / \Pr(+\text{test}) && \text{plug}
\end{aligned}
$$

$$
\begin{aligned}
&\Pr(-\text{virus}| + \text{test}) = \Pr(+\text{test}| - \text{virus}) \Pr(-\text{virus}) / \Pr(+\text{test}) && \text{Bayes rule}\\
&= 0.10 \times 0.99 / \Pr(+\text{test}) && \text{plug}
\end{aligned}
$$

Probability of positive test

Either you have the virus, or you don't. So,

$$\Pr(+\text{virus}| + \text{test}) + \Pr(-\text{virus}| + \text{test})$$
$$= 0.85 \times 0.01/\Pr(+\text{test}) + 0.10 \times 0.99/\Pr(+\text{test}) = 1$$

$\Pr(+\text{test}) = 0.85 \times 0.01 + 0.10 \times 0.99 = 0.1075$. So,

$\Pr(+\text{virus}| + \text{test}) = 0.85 \times 0.01/0.1075 \approx 8\%$.

Interpretation of this example

Positive test makes you much more likely to have the virus. But, it's still unlikely. The "ground truth" likelihood of $+$virus is very low. The vast majority of people taking the test do not have the virus, so even though false positives are rare, they're more common than true positives.

In other words: to begin with, the hypothesis that you have the virus is very unlikely to be true. The evidence of a postive test supports this hypothesis, since most people who have the virus test positive, and most people without the virus test negative. So, upon learning this evidence, it becomes more likely that the hypothesis is true—but only proportionally so.

Cookie choice

Two bowls of cookies:
Bowl 1 has 10 chocolate chip and 30 raisin
Bowl 2 has 20 chocolate chip and 20 raisin

Tyler has a probabilistic preference over bowls, but we don't know what. He'll choose a bowl and pick a cookie blindly from that bowl, show us the cookie, and put it back. We want to learn his preferences!

Intuitively: if he shows us a raisin cookie, it's more likely that he chose bowl 1. If he shows us a long sequence of cookies that's evenly mixed, it's likely that he exclusively picks bowl 2. But probably, he's picking bowl 1 with probability p. How to find p?

Cookie choice

Bowl 1 has 10 chocolate chip and 30 raisin. Bowl 2 has 20 chocolate chip and 20 raisin.

Let B_1, B_2 be the events he chose from bowl 1 and bowl 2, R be the event he showed a raisin cookie. Start by assuming Tyler chooses uniformly: $\Pr(B_1) = .5$. Suppose he shows us a raisin cookie.

$$\begin{aligned}\Pr(B_1|R) &= \frac{\Pr(R|B_1)\Pr(B_1)}{\Pr(R)} \\ &= \frac{\Pr(R|B_1)\Pr(B_1)}{\Pr(R|B_1)\Pr(B_1) + \Pr(R|B_2)\Pr(B_2)} \\ &= \frac{.75 \cdot .5}{.75 \cdot .5 + .5 \cdot .5} \\ &= .6\end{aligned}$$

Update our belief in B_1 to this new value, and keep running the experiment! Eventually $\Pr(B_1)$ will converge to his true preference.

Applications everywhere

Bayes' rule a *fundamental* building block of old (and some new!) AI/machine learning. (E.g.: spam classification, see hw10 mindbender!) This updating strategy lets us build our own model of a probabilistic world by observing pieces of that world.

The scientific method: pretty much Bayesian updating? (Philosophers of science love to discuss it.)

Decision theory, bioinformatics, probabilistic programming, …