Probability of a match between a random binary vector and a union of random binary vectors

Asked yesterday Active yesterday Viewed 39 times

1

Suppose I have M random binary vectors, each of length L, each consisting of zeros except exactly W randomly distributed on-bits (ones). Then I make a union (take OR) of these M vectors, such that the resulting vector **u** is much denser (i.e. has more ones) than each of the original M vectors. Finally, I match (take AND of) this union **u** with another random binary vector **t** (again, of size L, all zeros except W random ones). The question I cannot solve is this: what is the probability that all the on-bits in **t** will overlap with on-bits in **u**? If you can help, and perhaps illustrate your solution on a small example (say, $M = 3$, $L = 4$ and $W = 2$), you're my hero.

Edit: **Example:**

W = 2, L = 5, M = 2

vector 1: 10100

vector 2: 01100

union (u): 11100

example t: 11000 (match with u)

example t: 00101 (not a match with u)

With very small vectors (and small values of M) it is possible to write out the full space of possible combinations of vectors, manually count the combinations whose union will give a match with the target vector t and compute the probability (by dividing this number by the total number of possible combinations). But obviously this isn't feasible with large values of L and M. Isn't there a formula or some trick to compute the probability?

probability combinations combinations

edited yesterday

asked yesterday
MegaNightdude
107 8

2 Answers

2

This can be done using the principle of inclusion exclusion. We start with the complementary probability, that at least one of the W bits of **t** is "missed" by the corresponding bits in all the vectors making up **u**. The probability that a particular vector does not have a particular bit set is $\binom{L-1}{W}/\binom{L}{w}$. We then raise this to the M^{th} power to account for all vectors not having this bit set. We then add this up for each of the possible W bits of $\mathbf{t}$ being missed, which corresponds to multiplying by W.

However, we are not done. The event where *two* bits of $\mathbf{t}$ are missed in $\mathbf{u}$ has been double counted, and now needs to be subtracted out. Once these have been subtracted out, the triple intersections where three bits are missed now have to be added in, and so on. (To understand the details of this part, you need to familiarize yourself with the Principle of Inclusion Exclusion. This Brilliant article may be a good resource).

The final result is

$$\sum_{k=0}^{W}(-1)^k\binom{W}{k}\left(\frac{\binom{L-k}{W}}{\binom{L}{W}}\right)^M$$

For example, with $M = 3, L = 4, W = 2$, you get a probability of

$$1-2\cdot\left(\frac{\binom{3}{2}}{\binom{4}{2}}\right)^3+1\cdot\left(\frac{\binom{2}{2}}{\binom{4}{2}}\right)^3 = 1-2\cdot(\tfrac{1}{2})^3+(\tfrac{1}{6})^3 \approx 75.46\%$$

answered yesterday
Mike Earnest
34.9k 3 29 61

Great answer! :) – Pspl yesterday

0

Your problem is quite confusing. So, to don't waste much time on a wrong interpretation, I picked $M = 2$, $L = 2$ and $W = 1$ and you will tell us if I got it right.

Let's begin by construct the distribution of the "union" of two arbitrary vectors $M_1 \in \{(0,1),(1,0)\}$ and $M_2 \in \{(0,1),(1,0)\}$. Your definition of union (let's call it $\cup$) tells me $\cup = ((0,1),(1,0),(1,1))$ and the probabilities are $P(\cup) = (\frac{1}{4},\frac{1}{4},\frac{1}{2})$. you can verify this with a simple tree diagram.

Now, let's consider another vector $t \in \{(0,1),(1,0)\}$.

If $t = (0,1)$, then the on-bits of t overlap with the on-bits of $(0,1)$ and $(1,1)$ of $\cup$ so, the probability in this case is $P_1 = \frac{1}{2}\times\left(\frac{1}{4}+\frac{1}{2}\right) = \frac{3}{8}$.

On the other hand, if $t = (1,0)$, then the on-bits of t overlap with the on-bits of $(1,0)$ and $(1,1)$ of $\cup$ so, the probability in this case is $P_2 = \frac{1}{2}\times\left(\frac{1}{4}+\frac{1}{2}\right)$ which is also $\frac{3}{8}$.

The final answer for $M = 2$, $L = 2$ and $W = 1$ is then $P = P_1 + P_2 = \frac{3}{4}$.

Did I get it right?

answered yesterday
Pspl
518 9

This approach becomes impractical if you slightly increase W and L. You can see that if you take $M = 3$, $L = 4$ and $W = 2$ – MegaNightdude yesterday

I have added an example to make things clear. – MegaNightdude yesterday

@MegaNightdude, I was just trying to get your feedback to see if I understand your question correctly. Maybe the approach is impractical, yes. But allows me (and maybe others) to understand... – Pspl yesterday

Sorry, your understanding was correct. I added and example in the main post. – MegaNightdude yesterday

And the Brilliant article is great too. – MegaNightdude 22 hours ago