MITx: 6.008.1x Computational Probability and Inference

Hel|

Bookmarks

Exercise: Storing the Joint Probability Table of Two Random Variables

Bookmark this page

STORING JOINTLY DISTRIBUTED RANDOM VARIABLES

Let's start with two random variables, X and Y. To infer the value of random variable X given the observed value of a random variable Y, we first model how X and Y relate. As we saw in the first part of the course, we can model the relationship between X and Y by looking at their joint probability table $p_{X,Y}$ (which we might not know directly but can piece together if, for instance, we had other tables such as p_X and $p_{Y|X}$).

Exercise: Storing the Joint Probability Table of Two Random Variables

4/4 points (graded)

Suppose X takes on k different values and Y takes on ℓ different values.

For the answer boxes below, please provide your answer as a mathematical formula (and not as Python code). Use `^` for exponentiation, e.g., `x^2` denotes x^2. Explicitly include multiplication using `*`, e.g. `x*y` is xy.

- Suppose that X and Y may possibly be dependent (i.e., we don't know if they are independent). How many table entries are in the joint probability table $p_{X,Y}$?

  k*l

  Answer: k*l

  $k \cdot l$

- Next, consider when the two random variables X and Y are known to be independent. This means that the joint probability table factorizes so that $p_{X,Y}(x,y) = p_X(x)p_Y(y)$. In other words, rather than storing $p_{X,Y}$, we could instead store the marginal distributions p_X and p_Y, and then to compute $p_{X,Y}(x,y)$, we first look up $p_X(x)$ and $p_Y(y)$ and multiply these two numbers together.

  How many table entries are in the probability table p_X?

  k

  Answer: k

  k

  How many table entries are in the probability table p_Y?

  l

  Answer: l

  l

  Note that the number of table entries needed to store p_X and p_Y without making any additional assumptions is precisely the sum of the number of table entries in p_X and the number of table entries in p_Y.

- Now suppose that p_X and p_Y are actually independent and identically distributed and so each of X and Y now take on k possible values. How many table entries are needed to store p_X and p_Y? Please provide your answer in terms of k and not ℓ.

  k

  ✔ **Answer:** k

  k

Independence can drastically reduce how many numbers we need to store especially when we look at distributions over many random variables!

Solution:

- Suppose that X and Y may possibly be dependent (i.e., we don't know if they are independent). How many table entries are in the joint probability table $p_{X,Y}$?

  Solution: There are k possibilities for X, and ℓ possibilities for Y so the joint probability table has $\boxed{k\ell}$ entries.

- Next, consider when the two random variables X and Y are known to be independent. This means that the joint probability table factorizes so that $p_{X,Y}(x, y) = p_X(x)p_Y(y)$. In other words, rather than storing $p_{X,Y}$, we could instead store the marginal distributions p_X and p_Y, and then to compute $p_{X,Y}(x, y)$, we first look up $p_X(x)$ and $p_Y(y)$ and multiply these two numbers together.

  How many table entries are in the probability table p_X?

  Solution: There are k possibilities for X, so p_X has $\boxed{k}$ entries.

  How many table entries are in the probability table p_Y?

Solution: There are ℓ possibilities for Y, so p_Y has $\boxed{\ell}$ entries.

Note that the number of table entries needed to store p_X and p_Y without making any additional assumptions is precisely the sum of the number of table entries in p_X and the number of table entries in p_Y.

- Now suppose that p_X and p_Y are actually independent and identically distributed and so each of X and Y now take on k possible values. How many table entries are needed to store p_X and p_Y? Please provide your answer in terms of k and not ℓ.

Solution: It suffices to store a single table for both p_X and p_Y with $\boxed{k}$ entries.

Submit You have used 1 of 5 attempts

✔ Correct (4/4 points)