Technical notes on probability: The discrete case

Yang Xu

   Probability theory provides the mathematical foundations for describing random variables and their relations under uncertainty. This note summarizes common rules for probabilities concerning discrete variables, adapted from *A philosophical essay on probabilities* by Laplace (1902; 1814).

1. A sample space is the set of all outcomes from an experiment. An event is a subset of a sample space. A random variable is a function that maps events to numbers. The probability $p(\cdot)$ of a variable X is defined as the ratio of number of outcomes that entails event $X = x$ against that of all outcomes:

$$p(X = x) = p(x) = \frac{\#I(x)}{\sum_{x'} \#I(x')},\tag{1}$$

   such that $0 \leq p(x) \leq 1$ and $\sum_x p(x) = 1$.

   Example A
   The experiment of tossing a coin has the sample space {head,tail}. Obtaining either a head or a tail is a possible event from that sample space. Now suppose you toss two fair coins and would like to know the probability of obtaining at least one head. Let 1 represent a head and 0 represent a tail. The sample space now includes four outcomes: 00, 01, 10, 11. Let X represent the sum of two tosses in an outcome. The probability of obtaining at least one head is then $p(head) = p(X \geq 1) = \frac{3}{4}$, because 3 out of 4 outcomes entail head(s).

2. The probability of a particular event x and that of all other events excluding x sum to 1: $p(x) + p(\neg x) = 1$, hence $p(x) = 1 - p(\neg x)$.

   Example B
   $p(head)$ in Example A can also be calculated by considering the case that heads are not present in either of the two tosses, which corresponds to the outcome 00. It follows then $p(head) = 1 - p(\neg head) = 1 - p(X < 1) = 1 - \frac{1}{4} = \frac{3}{4}$.

3. The joint probability of two independent events x and y under variables X and Y is the product of probabilities of the two events:

$$p(x, y) = p(x)p(y),\tag{2}$$

   such that $\sum_x \sum_y p(x, y) = 1$.

Example C
Suppose you toss two fair coins, $C1$ and $C2$, N times and record the count for each of the four possible outcomes in the following contingency table. You can calculate joint probabilities $p(C1 = 0/1, C2 = 0/1)$ by dividing these counts into N: $p(0,0) = p(0,1) = p(1,0) = p(1,1) = \frac{N}{4}/N = \frac{1}{4}$. Because both coins are fair, you know that $p(0) = .5$ and $p(1) = .5$. It is easy to show that $p(C1, C2) = p(C1)p(C2)$, hence the two coins are independent.

C1 \ C2	1	0
1	$N/4$	$N/4$
0	$N/4$	$N/4$

4. The marginal probability of an event x under variable X is equivalent to the sum of its joint probability with y for all outcomes concerning variable Y, known as *the sum rule*:

$$p(x) = \sum_y p(x, y). \tag{3}$$

Convince yourself this is the case by calculating $p(C1 = 1), p(C1 = 0), p(C2 = 1), p(C2 = 0)$ based on the contingency table in Example C - they should all be 0.5.

5. The joint probability of two (in)dependent events is the product of probability of one event and probability of the other event conditioned on the first, known as *the product rule*:

$$p(x, y) = p(x)p(y|x) = p(y)p(x|y). \tag{4}$$

Therefore the conditional probabilities can be calculated as follows:

$$p(x|y) = \frac{p(x, y)}{p(y)}; \tag{5}$$

$$p(y|x) = \frac{p(x, y)}{p(x)}.$$

Example D
Suppose two weakly magnetized coins are tossed N times and the outcomes are summarized in the following contingency table. You can tell from these counts that the coins are probably not independent, because the head of $C1$ seems to co-occur often with the tail of $C2$. More formally, you can show this dependence relation by calculating $p(C1 = 1|C2 = 0) = \frac{p(C1=1,C2=0)}{p(C2=0)} = \frac{10/16}{10/16+3/16} = \frac{10}{13}$, and $p(C2 = 0|C1 = 1) = \frac{p(C1=1,C2=0)}{p(C1=1)} = \frac{10/16}{10/16+1/16} = \frac{10}{11}$ - both are quite high (what would these probabilities be if assume the coins are independent?).

C1 \ C2	1	0
1	$N/16$	$10N/16$
0	$2N/16$	$3N/16$

6. The joint probablity of two events is no greater than the probability of either: $p(x, y) \leq p(x)$ or $p(y)$.

   Convince yourself that this is the case by making use of rules 1, 3, and 5.

7. The Bayes rule relates unknown event(s) e to some observables o:

$$p(e|o) = \frac{p(e)p(o|e)}{p(o)} = \frac{p(e)p(o|e)}{\sum_e p(e)p(o|e)}, \tag{6}$$

   where $p(e)$ is prior probability, $p(o|e)$ is likelihood, and $p(e|o)$ is posterior probability. Note that the Bayes rule may be derived from the sum and product rules.

   Example E
   Suppose you wish to predict your luck in an upcoming event by flipping a fortune-telling coin you keep. Over the past, you had observed that good luck was likely to be associated with heads facing up, more so than bad luck (e.g. $p(head|good) = 0.8$ and $p(head|bad) = 0.3$). You also have the prior belief that chances break even between having good or bad luck regardless of the state of the coin: $p(good) = 1 - p(bad) = 0.5$. The posterior probability that you would have good luck now given your past observations is then:

$$p(good|head) = \frac{p(head|good)p(good)}{p(head|good)p(good) + p(head|bad)p(bad)} = \frac{0.8 \times 0.5}{0.8 \times 0.5 + 0.3 \times 0.5} = \frac{8}{11}.$$

8. The mean or expected value of a variable is the sum of its possible values weighted by their corresponding probabilities:

$$E[X] = \sum_x xp(x). \tag{7}$$

   Convince yourself by calculating the mean for either coin in Example C - it should be 0.5.

9. The expected benefit B or cost C of a variable is the sum of benefits or costs of individual events weighted by the probabilities: $E[B] = \sum_x B(x)p(x)$, or $E[C] = \sum_x c(x)p(x)$.

   Example F
   Suppose you bet with someone that $C1$ head and $C2$ tail are of opposite magnetic polarities for the coins in Example D, such that you gain \$1 if $C1 = 1 \& C2 = 0$, and nothing otherwise. The expected gain or benefit in this case would be \$1 $\times \frac{5}{8}$ + \$0 $\times \frac{3}{8}$ = \$$\frac{5}{8}$.

10. The variance of a variable is the sum of deviations of its possible values from the mean weighted by their corresponding probabilities:

$$Var(X) = \sum_x (x - E[X])^2 p(x). \tag{8}$$

   Show that variance of $C1$ in Example C is greater than that of $C1$ in Example D - why so?

3