2/18/2023

Probability Topics Continued
Discrete Probability

Experiment: an instance of a random variable (rolling a die one time, or tossing a coin one time, or a fixed number of times, etc.)

Outcome – just a set of results of the experiment

Simple event – is a single possible outcome of an experiment (trial), compound event is a collection of events that are interesting to experimenter.

List of possible outcomes is called the sample space.

Classical or theoretical probability
Experimental or empirical probabilities
Personal or subjective probabilities

Classical probabilities: they assume that everything in the sample space is equally likely, and then the probability of any simple event in the sample space is 1 divided by the size of the sample space. Equiprobable events.

If I flip a coin twice, the possible outcomes are {HH, HT, TH, TT}=sample space. So each of these has ¼ chance of occurring. If I want the probability of getting one head and one tail, then there are two outcomes that satisfy that description and so the probability is 2/4.

Experimental probabilities: probabilities are obtained by doing a large number of repeated experiments.

The Law of Large Numbers, if you conduct an experiment a large number of times, the probability obtained in the experiment will get closer to the theoretical probability of the same event, and will get closer as the number of trials increases.

Personal or subjective probabilities: they describe "chance" that an event that is difficult to repeat happens.

Could mean that the likelihood is very small. It could mean that the exact circumstances are difficult to repeat. Taking factors into account that are not well enumerated.

Maybe you walk out of a test and think you did well… you think there is a high probability that you passed the exam.

Probabilities from Two-way tables
Conditional probabilities
AND and OR events
Independence

Count of Marital Status	Column Labels		
Row Labels	Does not Own Home	Does Own Home	Grand Total
Married	1719	5147	6866
Single	3896	3297	7193
Grand Total	**5615**	**8444**	**14059**

If we were to select a random person from this sample (dataset)…

What is the probability that the person is married? $\frac{6866}{14059}$ approx. 0.48837…

What is the probability that the person Owns their own Home? $\frac{8444}{14059}$ approx. 0.6006…

What is the probability that someone in the dataset is not married? $\frac{7193}{14059}$… or $1 - \frac{6866}{14059}$

What is the probability that the person is both married and owns their own home? (intersection = AND)

$$\frac{5147}{14059}$$

What is the probability that the person is either married or owns their own home? (union = OR)

$$P(A\ or\ B) = P(A) + P(B) - P(A\ and\ B)$$

$$\frac{6866}{14059} + \frac{8444}{14059} - \frac{5147}{14059} = \frac{6866 + 8444 - 5147}{14059} = \frac{10163}{14059} \approx 0.72288\ \ldots$$

Alternatively: $\frac{1719+5147+3297}{14059} = \frac{10163}{14059}$

If the probability here goes over 100% (or 1) you've missed something.

Conditional probabilities:
If I know something about one variable, what is the probability of the other?
What is the probability that the person is married GIVEN that they own their own home? $P(A|B)$

$$\frac{5147}{8444}$$

What is the probability that the person owns their own home **given** that they are married? $P(B|A)$

$$\frac{5147}{6866}$$

Independence:

If I know the probability of one event, that does not affect my knowledge of the probability of a second event.

Let's say I toss a coin and roll a die. If I get a Head on the coin toss, does that affect my knowledge of the die roll?

These are independent events.

Mathematically, we say that if two events A and B are independent then: $P(A) = P(A|B)$. Or $P(A\ and\ B) = P(A) * P(B)$

In general: $P(A\ and\ B) = P(A|B)P(B)$

In a table: Compare the probability of being married P(A) to the probability of being married GIVEN that you own your own home P(A|B). If they are the same, then the events are independent, and if they are not identical, then the events are dependent.

P(married) = $\frac{6866}{14059}$ approx. 0.48837…
P(married|own home) = $\frac{5147}{8444}$ approx. 0.6095…

These are not equal and therefore they are dependent.

Discrete Distributions

x	1	2	3	4	5	6
$p(x)$	0.10	0.20	0.15	0.12	0.21	0.22

What is the expected value of this distribution? What is the standard deviation of this distribution?

Expected value: is the same as the mean. But we have to calculate a weighted average.

$$E(X) = \mu = \sum_{i=1}^{n} x_i p_i$$

The expected value means that if I were to conduct an experiment (many, many times) and record the outcomes of an experiment that followed the probability outcomes in the distribution, then the average of those outcomes should be about 3.8 (see Excel).

To calculate the standard deviation of outcomes:

$$V(X) = \sum_{i=1}^{n} (x_i - \mu)^2 p_i$$

$$S(X) = \sqrt{V(X)}$$

1. Find the mean E(X)
2. Subtract the mean from all the outcome values
3. Square all those values
4. Multiply by the corresponding probabilities
5. Add them up.

Binomial Distribution
1. The outcomes can be classed into just two groups: success and failure.
2. The probability of success for each trial is fixed.
3. The number of trials is fixed.

The scenario is I'm going to conduct the experiment n times and look for x successes.

$$B(x; n, p) = \binom{n}{x} p^x (1-p)^{n-x}$$

Suppose we have an unfair coin that has a 60% chance of coming up heads.
We toss the coin 15 times.

1. What is the probability of getting (exactly) 8 heads?

$$B(8; 15, 0.6) = \binom{15}{8} (0.6)^8 (0.4)^7 = 6435(.6)^8(.4)^7 = 0.17708 \ldots$$

2. What is the probability of getting 5 or fewer heads?
   The outcomes covered here are: 0 heads, 1 head, 2 heads, 3 heads, 4 heads, and 5 heads.
   In Excel, we can say the cumulative is TRUE in the formula

   The trickiest part is determining what value is the cut-off…

   5 or fewer includes 5, so 5 is the cut-off
   Less than 5 does not include 5, so 4 is cut-off

3. What is the probability of getting 13 or more heads?
   What is included here? 13, 14, 15
   Or, use the complement: what is being left out is 12 or less… use the cumulative distribution up to 12, and the subtract that result from 1.

What is the expect value of a binomial distribution? What is the mean?

What is the variance and standard deviation?

$$E(X) = np$$

For our example, the mean is $15 \times 0.6 = 9$

$$V(X) = npq = np(1-p)$$

In our example the variance is $15 \times 0.6 \times 0.4 = 3.6$

Therefore, the standard deviation is the square root of that: $\sqrt{3.6} = 1.8973 \ldots$