Probability Basics

Mark Andrews

What is probability?

- Probability is a means to quantify uncertainty.
- If a variable can take on more than one value, probability can be used to describe the certainty that is will take each one of its possible values.
- Probabilities must lie between zero and one.
- The sum of the probabilities of all values of a variable must equal one.

What is probability? Notation

- If X is a variable with possible values $\{x_1, x_2 \ldots x_k \ldots x_K\}$, then

$$P(X = x_k) \tag{1}$$

gives the probability that X takes the value x_k.

- The rules of probability require that

$$0 \leqslant P(X = x_k) \leqslant 1 \quad \forall x_k \tag{2}$$

and

$$\sum_{k=1}^{K} P(X = x_k) = 1. \tag{3}$$

Joint probability

- If X is a variable with possible values $\{x_1, x_2 \ldots x_k \ldots x_K\}$, and Y is a variable with possible values $\{y_1, y_2 \ldots y_k \ldots y_L\}$, then

$$P(X = x_k, Y = y_l) \tag{4}$$

gives the probability that X takes the value x_k and Y takes the value of y_l.

- The rules of probability require that

$$0 \leqslant P(X = x_k, Y = y_l) \leqslant 1 \quad \forall x_k \tag{5}$$

and

$$\sum_{l=1}^{L} \sum_{k=1}^{K} P(X = x_k, Y = y_l) = 1. \tag{6}$$

Conditional, joint and marginal probability

- Conditional probability is the probability distribution of a variable when the value of another variable is known, i.e.

$$P(X = x_k | Y = y_l) \tag{7}$$

  is read as the "the probability that X is x_k *given* Y is y_l.
- Conditional probability can be derived from the joint probability as follows:

$$P(X = x_k | Y = y_l) = \frac{P(X = x_k, Y = y_l)}{P(Y = y_l)} \tag{8}$$

- Marginal probability can also be derived from the joint probability as follows:

$$P(X = x_k) = \sum_{l=1}^{L} P(X = x_k, Y = y_l). \tag{9}$$

Conditional, joint and marginal probability

- There at least three different types of probability distributions, referred to as *joint*, *marginal* or *conditional* probabilities.
- There are fundamental relationships between them.
- We can illustrate these concepts by looking at survival rates of males and females on *RMS Titanic*.

Frequency distribution of survival rates on Titanic

- The following table shows the number of men and women who died or survived on the *Titanic*:

	Men	Women
Perished	1352	109
Survived	338	316

.

- Altogether, there were 2115 onboard.

From frequencies to probabilities

- We can convert the frequencies

	Men	Women
Perished	1352	109
Survived	338	316

,

to the probabilities

	Men	Women
Perished	.64	.05
Survived	.16	.15

,

by dividing each frequency by the total number onboard, i.e. 2115.

Joint probability tables

- The table

	Men	Women
Perished	.64	.05
Survived	.16	.15

,

  is a *joint probability* table.
- It provides the probability for every combination of the two variables.
- It is just another probability distribution, like what we have met already.
- Each element in the table lies between 0 and 1 and together they must sum to 1.

Marginal probabilities from joint probabilities

- From the table

	Men	Women
Perished	.64	.05
Survived	.16	.15

,

what is the overall probability of dying?
- Recall that each element in the table provides the probability for a combination of values of the two variables, e.g. the probability of dying and being male is .64.
- Following the rules of probability, we calculate the probability of dying as follows:

$$\mathrm{P}(\text{Perished}) = \mathrm{P}(\text{Male \& Perished}) + \mathrm{P}(\text{Female \& Perished}), \tag{10}$$

$$= .64 + .05 = .69 \tag{11}$$

Conditional probabilities from joint probabilities

- From the table

	Men	Women
Perished	.64	.05
Survived	.16	.15

,

what is the probability of dying if the person is a man?

- First, the probability of being a man onboard *Titanic* is

$$\begin{aligned} \mathrm{P}(\text{Male}) &= \mathrm{P}(\text{Male \& Perished}) + \mathrm{P}(\text{Male \& Survived}), && (12)\\ &= .64 + .16 = .8\,. && (13) \end{aligned}$$

Conditional probabilities from joint probabilities (cont'd)

- Then, what is the probability of being a man and dying? This is simply .64.
- Putting these together: 80% of the ship's total was male and 64% of the total were men who died. That means the fraction of men who died is 64/80, i.e.

$$\mathrm{P}(\text{Perished}|\text{Male}) = \frac{\mathrm{P}(\text{Male \& Perished})}{\mathrm{P}(\text{Male})}, \tag{14}$$

$$= \frac{.64}{.8} = .8 \tag{15}$$

Conditional, joint and marginal: More Rules

- As we have

$$P(X = x_k | Y = y_l) = \frac{P(X = x_k, Y = y_l)}{P(Y = y_l)} \tag{16}$$

then we also have

$$P(X = x_k, Y = y_l) = P(X = x_k | Y = y_l) P(Y = y_l), \tag{17}$$

- Likewise, given that

$$P(Y = y_l | X = x_k) = \frac{P(Y = y_l, X = x_k)}{P(X = x_k)} \tag{18}$$

then

$$P(Y = y_l, X = x_k) = P(Y = y_l | X = x_k) P(X = x_k). \tag{19}$$

Chain rule

- $$\begin{aligned} P(X, Y, Z) &= P(X|Y, Z)P(Y|Z)P(Z), & (20)\\ &= P(X|Z, Y)P(Z|Y)P(Y), & (21)\\ &= P(Y|X, Z)P(X|Z)P(Z), & (22)\\ &= \dots & (23) \end{aligned}$$

Independence and conditional independence

- If $P(X, Y) = P(X)P(Y)$, then X and Y are independent.
- If X and Y are independent, their correlation is 0.
- If $P(X, Y|Z) = P(X|Z)P(Y|Z)$, then X and Y are conditionally independent, conditional on Z.
- If X and Y are conditionally independent, conditional of Z, the parial correlation of X and Y given Z is zero.