# Probability review

INFO 1301

Prof. Michael Paul

Prof. William Aspray

# R1

- The probability of an outcome is the proportion of times the outcome would occur if we observed the random process an infinite number of times.
- Law of Large Numbers – If you repeat an experiment (e.g. flipping a coin) enough times, you will get closer and closer to the actual probability of that event, .5 for heads when flipping a coin or 1/6 for getting a 3 on a die.
- A **distribution X** is a table of the probabilities of all possible outcomes of a random variable.
- The sum of all probabilities in a distribution must equal 1 (or 100%)
- For random variables, the standard measure of central tendency is the **expected value.**
- E(x) = expected value of x is given by the formula $\mathbf{E}[X] = \Sigma_x P(X = x)\, x$

# R2

- If you take the average of multiple outcomes of a random variable, the average will most often be close to the expected value
- **Central Limit Theorem** More formally, the theorem states that if you take the average of multiple random outcomes multiple times, the averages will form a bell curve where the mean is the expected value of that random variable
- If two or more outcomes cannot all be true at once, they are called **disjoint** or **mutually exclusive**
- The **complement** of an outcome is the set of all other outcomes in the sample space. P(not(x=a)) = 1 – P(x=a)

# R3

- The probability that multiple outcomes are true can be described with an AND expression – and is measured by the product if the events are independent of one another.
- P(H and 5) = P(heads) x P(5)
- The probability that any of a set of multiple outcomes can be true can be described with an OR expression.
- **If outcomes are disjoint**, the probability that any of them are true is the *sum* of their individual probabilities
- P(3 or 5) = P(3) or P(5)

**If not disjoint**, the probability that either outcome is true is the *sum* of their individual probabilities, *minus* the probability that they are both true

- i.e., P($X$ OR $Y$) = P($X$) + P($Y$) – P($X$ AND $Y$)

# R4

- The probability of exactly one outcome is sometimes called a **marginal probability**
  - For example from class, Let *X* be the health status of an individual and *Y* be the insurance status of the individual

P(*X* = Excellent) **marginal probability**

- The probability that two or more outcomes are all true is called a **joint probability**
  - For example, P(*X* = Excellent AND *Y* = Yes) **joint probability**. Also written, P(*X* = Excellent, *Y* = Yes)
- P(*X* = Excellent | *Y* = Yes) **conditional** probability

The probability of an outcome, given that one or more other outcomes are true, is a **conditional probability**

- In this example, we would say that the probability of *X* is *conditioned* on *Y*
- In other words, the probability of X **if** we know Y is true.

# R5

- If you know the value of two of these 3 types of probabilities (marginal, joint, conditional), you can calculate the third.
  - **Marginalization:** The marginal probability of an outcome can be calculated by summing over all joint probabilities that include the outcome
- Rules

  For any two random variables $X$ and $Y$ with values a and b:

  $P(X = a) = \sum_b P(X = a, Y = b)$

  $P(X = a, Y = b) = P(X = a \mid Y = b) \times P(Y = b)$

  $P(X = a \mid Y = b) = P(X = a, Y = b) / P(Y = b)$

# R6

- Two random variables are **independent** if knowing the outcome of one does not change the probability of the other
  - If $X$ and $Y$ are independent then: $P(X = a, Y = b) = P(X = a) \times P(Y = b)$
- **Entropy** is a measurement of how evenly distributed a probability distribution is
- Entropy of a random variable $X$ is denoted $H(X)$
  - Entropy is non-negative (0 or higher)

    Lower entropy means it is less even, more certain
    Higher entropy means it is more even, less certain
- The lowest possible value of entropy is 0
  - This occurs when a distribution gives 0 probability to all but one outcome
- The highest possible value of entropy occurs when the distribution is **uniform**

# R7

How to calculate $H(X)$? A mess!

1. For every outcome of $X$, calculate:
$P(X{=}a) \times \log_2 P(X{=}a)$

2. Then sum the results for each outcome:
$P(X{=}a) \times \log_2 P(X{=}a) + P(X{=}b) \times \log_2 P(X{=}b)$

3. Then multiply the final result by –1:
$- P(X{=}a) \times \log_2 P(X{=}a) - P(X{=}b) \times \log_2 P(X{=}b)$

General formula:
$H(X) = - \sum_a P(X = a) \log_2 P(X = a)$