DSCI 551 Review

Jakob Thoms

2022-10-04

Lecture 1

In general, the probability of an event A occurring is denoted as $P(A)$ and is defined as

$$P(A) = \frac{\text{Number of times event } A \text{ is observed}}{\text{Total number of events observed}}$$

as the number of events goes to infinity.

- We heavily rely on the "frequency of events" to make estimations of specific parameters of interest in a population or system.
- This is basically the foundation of a frequentist approach: relying on the frequency (or "number"!) of events to estimate your parameters of interest.

Law of total probability: When partitioning the sample space (the set of all possible events), the sum of the probabilities of each event should be one.

$$\sum_{E \in \Omega} P(E) = 1.$$

- In general, for a given event A, the law implies that

$$1 = P(A) + P(A^c).$$

Inclusion-exclusion principle:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B),$$

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(B \cap C) - P(A \cap C) + P(A \cap B \cap C),$$

etc.

Odds: are quite helpful in comparing the probability of two events.

$$o = \frac{p}{1-p},$$

where p is the probability of an event.

- This implies

$$p = \frac{o}{o+1}.$$

Central tendency: a measure denoting a "typical" value in a random variable.

Uncertainty:a measure of how "spread" a random variable is

- Called *parameters** when it comes to a population

- Are estimated via *sample statistics**

Mode: the outcome having the highest probability.

Entropy: a measure of uncertainty defined by

$$H(X) = \sum_x P(X = x) \ln \left(\frac{1}{P(X = x)} \right)$$

or

$$H(X) = \int_x f_X(x) \ln \left(\frac{1}{f_X(x)} \right) \mathrm{d}x.$$

- Always non-negative in the discrete case
- $H(X) = 0 \iff X$ is constant in the discrete case.

Expectation:

$$\mathbb{E}(X) = \sum_x x \cdot P(X = x).$$

or

$$\mathbb{E}(X) = \int_x x \cdot f_X(x)$$

- Can usually be estimated via the **sample mean**

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

Variance:

$$\mathrm{Var}(X) = \mathbb{E}\{[X - \mathbb{E}(X)]^2\}.$$

$$\implies \mathrm{Var}(X) = \mathbb{E}(X^2) - [\mathbb{E}(X)]^2.$$

- the variance is an expectation (specifically, the squared deviation from the mean)
- can usually be estimated via the *sample variance**

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$$

- Always non-negative, and $\mathrm{Var}(X) = 0 \iff X$ is constant

Standard deviation: The square root of the variance,

$$\sigma_X = \sqrt{\mathrm{Var}(X)}.$$

Lecture 2

- To maximize entropy, you need equal probabilities for all the outcomes in the sample space. This indicates we have a uniform uncertainty over the whole range of possible outcomes.
- Helpful univariate distribution guide: http://www.math.wm.edu/~leemis/chart/UDR/UDR.html

Binomial distribution:

$$X \sim \text{Binomial}(n, \pi)$$

- X is the number of successes in n trials in which each trial has probability π of success, independent of all other trials.
- PMF:
$$P(X = x \mid n, \pi) = \binom{n}{x} \pi^x (1-\pi)^{n-x} \quad \text{for} \quad x = 0, 1, \ldots, n.$$
- Expected value:
$$\mathbb{E}(X) = n\pi$$
- Variance:
$$\text{Var}(X) = n\pi(1-\pi)$$

Families and Parameters:

- We refer to the entire set of Binomial probability distributions as the *Binomial family of distributions**.
- Specifying a value for both π and n results in a unique Binomial distribution.
- Since π and n fully specify a Binomial distribution, we call them *parameters** of the Binomial family, and we call the Binomial family a *parametric family** of distributions.
- There are other ways we can specify the distribution. For instance, specifying the mean and variance is enough to identify a Binomial distribution.
- Exactly which variables we decide to use to identify a distribution within a family is called the family's parameterization.
- The parameterization you use in practice will depend on the information you can more easily obtain

Geometric distribution:

$$X \sim \text{Geometric}(\pi)$$

X is the number of trials **before** experiencing a success, where each trial has probability π of success, independent of all other trials.
- PMF:
$$P(X = x \mid \pi) = \pi(1-\pi)^x \quad \text{for} \quad x = 0, 1, \ldots$$

- Since there is only one parameter, this means that if you know the mean, you also know the variance!
- Expected value:
$$\mathbb{E}(X) = \frac{1-\pi}{\pi}$$
- Variance:
$$\text{Var}(X) = \frac{1-\pi}{\pi^2}$$

Negative Binomial Distribution:

$$X \sim \text{Negative Binomial}(k, \pi)$$

- X is the number of failed trials before experiencing k successes, where each trial has probability π of success, independent of all other trials. - PMF:

$$P(X = x \mid k, \pi) = \binom{k-1+x}{x} \pi^k (1-\pi)^x \quad \text{for} \quad x = 0, 1, \ldots$$

- The Geometric family results with $k = 1$.
- Expected value:

$$\mathbb{E}(X) = \frac{k(1-\pi)}{\pi}.$$

- Variance:

$$\text{Var}(X) = \frac{k(1-\pi)}{\pi^2}.$$

Poisson Distribution:

$$X \sim \text{Poisson}(\lambda)$$

- X is number of events occurring in a fixed interval of time or space, assuming that these events occur with a known constant mean rate (e.g. 3 events per minute or 5 events per meter) and independently of the time since the last event
- PMF

$$P(X = x \mid \lambda) = \frac{\lambda^x \exp(-\lambda)}{x!} \quad \text{for} \quad x = 0, 1, \ldots$$

- Expected value:

$$\mathbb{E}(X) = \lambda.$$

- Variance:

$$\text{Var}(X) = \lambda.$$

Bernoulli Distribution:

$$X \sim \text{Bernoulli}(\pi)$$

- X is equal to one with probability π and equal to zero with probability $1 - \pi$.
- Basically a weighted coin-flip
- A special case of the Binomial family ($n = 1$)
- PMF:

$$P(X = x \mid \pi) = \pi^x (1-\pi)^{1-x} \quad \text{for} \quad x = 0, 1.$$

- Expected value:

$$\mathbb{E}(X) = \pi.$$

- Variance:

$$\text{Var}(X) = \pi(1-\pi).$$