Unit 2 - Probability and distributions

Suggested reading:

* **OpenIntro Statistics, Chapter 2: Sections 2.1 and 2.2**
* **OpenIntro Statistics, Chapter 3: Sections 3.1, 3.2, and 3.4**

Suggested exercises:

* *Part 1 - Defining probability: 2.3, 2.7, 2.9, 2.11*
* *Part 2 - Conditional probability: 2.17, 2.21, 2.25*
* *Part 3 - Normal distribution: 3.3, 3.5, 3.9, 3.11, 3.19*
* *Part 4 - Binomial distribution: 3.27, 3.29, 3.31, 3.39*

---

* *Suggested Reading: Section 2.1 of OpenIntro Statistics*

LO 1. Define the probability of an outcome as the proportion of times the outcome would occur if we observed the random process that gives rise to it an infinite number of times.

LO 2. Explain why the long-run relative frequency of repeated independent events settles down to the true probability as the number of trials increases, i.e. why the law of large numbers holds.

LO 3. Define disjoint (mutually exclusive) events as events that cannot both happen at the same time:

- If A and B are disjoint, P(A and B) = 0

LO 4. Distinguish between disjoint and independent events.

- If A and B are independent, then having information on A does not tell us anything about B (and vice versa).
- If A and B are disjoint, then knowing that A occurs tells us that B cannot occur (and vice versa).
- Disjoint (mutually exclusive) events are always dependent since if one event occurs we know the other one cannot.

LO 5. Draw Venn diagrams representing events and their probabilities.

LO 6. Define a probability distribution as a list of the possible outcomes with corresponding probabilities that satisfies three rules:

- The outcomes listed must be disjoint.
- Each probability must be between 0 and 1.
- The probabilities must total 1.

LO 7. Define complementary outcomes as mutually exclusive outcomes of the same random process whose probabilities add up to 1.

- If A and B are complementary, P(A) + P(B) = 1

LO 8. Distinguish between union of events (A or B) and intersection of events (A and B).

- Calculate the probability of union of events using the (general) addition rule:
  If A and B are not mutually exclusive, $P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$
  If A and B are mutually exclusive, $P(A \text{ or } B) = P(A) + P(B)$, since for mutually exclusive events $P(A \text{ and } B) = 0$
- Calculate the probability of intersection of independent events using the multiplication rule:
  if A and B are independent, $P(A \text{ and } B) = P(A) \times P(B)$
  If A and B are dependent, $P(A \text{ and } B) = P(A|B) \times P(B)$

∗ *Test yourself:*

1. *What is the probability of getting a head on the 6th coin flip if in the first 5 flips the coin landed on a head each time?*
2. *True / False: Being right handed and having blue eyes are mutually exclusive events.*
3. *P(A) = 0.5, P(B) = 0.6, and there are no other possible outcomes in the sample space. What is P(A and B)?*

∗ *Suggested Reading: Section 2.2 of OpenIntro Statistics*

LO 9. Distinguish between marginal and conditional probabilities.

LO 10. Construct tree diagrams to calculate conditional probabilities and probabilities of intersection of non-independent events using Bayes' theorem: $P(A|B) = \frac{P(A \text{ and } B)}{P(B)}$.

∗ *Test yourself: 50% of students in a class are social science majors and the rest are not. 70% of the social science students and 40% of the non-social science students are in a relationship. Create a contingency table and a tree diagram summarizing these probabilities. Calculate the percentage of students in this class who are in a relationship.*

∗ *Suggested Reading: Section 3.1 and 3.2 of OpenIntro Statistics*

LO 11. Define the standardized (Z) score of a data point as the number of standard deviations it is away from the mean: $Z = \frac{X-\mu}{\sigma}$.

LO 12. Use the Z score

- if the distribution is normal: to determine the percentile score of a data point (using technology or normal probability tables)

  - regardless of the shape of the distribution: to assess whether or not the particular observation is considered to be unusual (more than 2 standard deviations away from the mean)

LO 13. Depending on the shape of the distribution determine whether the median would have a negative, positive, or 0 Z score keeping in mind that the mean always has a Z score of 0.

LO 14. Assess whether or not a distribution is nearly normal using the 68-95-99.7% rule or graphical methods such as a normal probability plot.

∗ *Test yourself: True/False: In a right skewed distribution the Z score of the median is positive.*

∗ *Suggested Reading: Section 3.4 of OpenIntro Statistics*

∗ *About the Milgram Experiment (examples used in this section): From RadioLab:* ***The Bad Show***

LO 15. Determine if a random variable is binomial using the four conditions.

  - The trials are independent.
  - The number of trials, n, is fixed.
  - Each trial outcome can be classified as a success or failure.
  - The probability of a success, p, is the same for each trial.

LO 16. Calculate the number of possible scenarios for obtaining k successes in n trials using the choose function: $\binom{n}{k} = \frac{n!}{k!(n-k)!}$.

LO 17. Calculate probability of a given number of successes in a given number of trials using the binomial distribution: $P(k = K) = \binom{n}{k} p^k (1-p)^{(n-k)}$.

LO 18. Calculate the expected number of successes in a given number of binomial trials ($\mu = np$) and its standard deviation ($\sigma = \sqrt{np(1-p)}$).

LO 19. When number of trials is sufficiently large ($np \geq 10$ and $n(1-p) \geq 10$), use the normal approximation to calculate binomial probabilities, and explain why this approach works.

∗ *Test yourself:*

1. *True/False: We can use the binomial distribution to determine the probability that in 10 rolls of a die the first 6 occurs on the 8th roll.*
2. *True / False: If a family has 3 kids, there are 8 possible combinations of gender order.*
3. *True/ False: When $n = 100$ and $p = 0.92$ we can use the normal approximation to the binomial to calculate the probability of 90 or more successes.*