Chapter 3. Probability

3.1 Defining probability

- Definition: probability (p) of an outcome is the proportion of times the outcome would occur if we observed the random process an infinite number of times.
- A value between 0 and 1.
- P(rolling a 1) = P(1)

Law of Large Numbers

- As more observations are collected, the proportion ($\hat{p}_n$ or p-hat) of occurrences with a particular outcome converges to the probability of that outcome.
- Common misunderstanding of LLN: gambler's fallacy / law of averages.
- **Disjoint** or **mutually exclusive** outcomes cannot both happen.

<u>**Addition Rule, *P*(A or B)**</u>

- To get the probability that **one of them ("or" situation)** will occur: add probabilities together.
  P(A or B) = P(A) + P(B),
  <u>where outcomes are disjoint.</u>

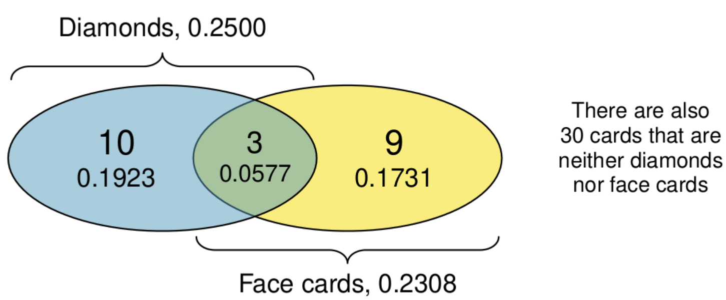

Figure 3.4: A Venn diagram for diamonds and face cards.

<u>**General Addition Rule, *P*(A or B)**</u>

- To get the probability that **one of them ("or" situation)** will occur: use the Venn diagram.
  - Add probabilities of 2 events.
  - Probability of common events are counted twice, so subtract it.

  P(A or B) = P(A) + P(B) - P(A and B),
  <u>where outcomes are disjoint or not.</u>

Probability distribution

- A list of possible outcomes with corresponding probabilities that satisfies 3 rules:
  a. Outcomes listed must be disjoint.

b. Each probability must be between 0 and 1.
c. Probabilities must total 1.

- Sample space: set of all possible outcomes.
- Complement of D represents all outcomes in our sample space that are not in D.
- Independence: two processes are independent if knowing the outcome of one provides no useful info about the outcome of the other.

Multiplication Rule, *P*(A and B)

- To get the probability that **both A and B ("and" situation)** occur: multiply their separate probabilities.
  P(A and B) = *P*(A) x *P*(B),
  where outcomes are independent.

3.2 Conditional probability

- Contingency table and Venn diagram

		truth		
		fashion	not	Total
mach_learn	pred_fashion	197	22	219
	pred_not	112	1491	1603
	Total	309	1513	1822

Figure 3.11: Contingency table summarizing the photo_classify data set.

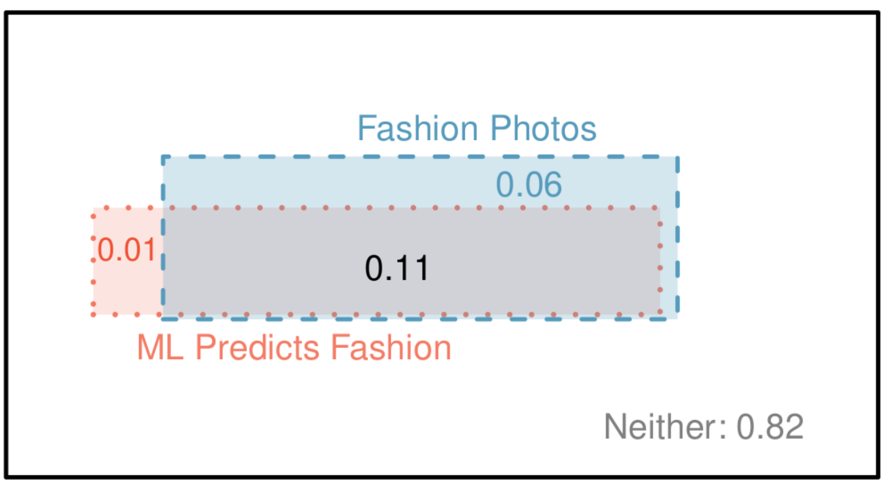

Figure 3.12: A Venn diagram using boxes for the photo_classify data set.

- **Marginal probabilities**: based on a single variable without regard to any other variables (e.g., probability that mach_learn predicted fashion = 0.12).

- **Joint probability**: a probability of outcomes for 2 or more variables or processes (e.g., probability that mach_learn predicted fashion and truth is fashion = 0.11).
- "And" = comma.
- **Table proportions** to summarize joint probabilities.

Conditional probability

- Probability of an event under a condition (e.g., probability that truth is fashion given mach_learn predicted fashion).
- 2 parts:
    - Outcome of interest
    - Condition: information we know to be true ("|" or "given")
- *P*(truth is fashion | mach learn is pred fashion).
- **Conditional probability equation**
  Outcome A given condition B is:
  P(A|B) = *P*(A and B) / *P*(B).

General Multiplication Rule, *P*(A and B)

- To get the probability that **both A and B ("and" situation)** occur: multiply their separate probabilities.
  P(A and B) = *P*(A|B) x *P*(B),
  <u>where toucomes or events may not be independent (A is outcome, B is condition).</u>
- Simply a rearrangement of the conditional probability equation.
- Example: 96% not vaccinated, 85% of not vaccinated people ended up surviving. What is the probability that a person was not vaccinated and survived? 0.85 x 0.96 = 0.816.

Tree diagrams

- Tool to organize outcomes and probabilities around the structure of the data.
- Most useful when 2 or more processes occur in a sequence and each process is conditioned on its predecessors.

Figure 3.17: A tree diagram of the `smallpox` data set.

- Put joint probabilities at the end of each branch using General Multiplication Rule.

Bayes' Theorem

- We are given conditional probability of *P*(variable 1 | variable 2) but we want to know the inverted conditional probability of *P*(variable 2 | variable 1).
- A very useful and general formula Bayes' Theorem can be used when there are so many scenarios that a tree diagram would be too complex.

BAYES' THEOREM: INVERTING PROBABILITIES

Consider the following conditional probability for variable 1 and variable 2:

$$P(\text{outcome } A_1 \text{ of variable 1} \mid \text{outcome } B \text{ of variable 2})$$

Bayes' Theorem states that this conditional probability can be identified as the following fraction:

$$\frac{P(B|A_1)P(A_1)}{P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + \cdots + P(B|A_k)P(A_k)}$$

where A_2, A_3, ..., and A_k represent all other possible outcomes of the first variable.

- To apply Bayes' Theorem correctly, 2 prep steps:
  - Identify marginal probabilities of each possible outcome of the first variable $P(A_1)$, $P(A_2)$, … , $P(A_k)$.
  - Then identify probability of outcome B, conditioned on each possible scenario for first variable: $P(B|A_1)$, $P(B|A_2)$, … , $P(B|A_k)$.
- Bayesian statistics: strategy of updating beliefs using Bayes' Theorem

3.3 Sampling from a small population

- Sampling from a small population **without replacement**, we don't have independence between observations.
- Sampling **with replacement**, we still have independent observations.
- When sample size is only a small fraction of population (<10%) observations are nearly independent even when sampling without replacement.
- Guided Practice 3.52 stumped me! See end of this chapter

3.4 Random variables

- Definition: a random process or variable with a numerical outcome.
- Usually represent it with a capital letter, e.g., X, Y, or Z.
- **Expected value** of a discrete random variable = average outcome, computed by sum of each outcome multiplied by its corresponding probability.
- Expectation = center of gravity; this idea expands to continuous probability distributions.
- **General variance formula**: variance is weighted sum of squared deviations (weighted by their corresponding probabilities) and standard deviation is square root of variance.

Thus, the expected value is $\mu = 117.85$, which we computed earlier. The variance can be constructed by extending this table:

i	1	2	3	Total
x_i	\$0	\$137	\$170	
$P(X = x_i)$	0.20	0.55	0.25	
$x_i \times P(X = x_i)$	0	75.35	42.50	117.85
$x_i - \mu$	-117.85	19.15	52.15	
$(x_i - \mu)^2$	13888.62	366.72	2719.62	
$(x_i - \mu)^2 \times P(X = x_i)$	2777.7	201.7	679.9	3659.3

The variance of X is $\sigma^2 = 3659.3$, which means the standard deviation is $\sigma = \sqrt{3659.3} = \$60.49$.

- **Linear combination** of 2 random variables X and Y:
  aX + bY
  where a and b are some fixed and known numbers.
- Average value of a linear combination of random variables:
  A x E(X) + b x E(Y)
  where expected value is same as the mean e.g., E(X) = μX.
- Variance in linear combinations of random variables:
  Var(aX + bY) = $a^2 \times$ Var(X) + $b^2 \times$ Var(Y)
  where random variables are independent of each other.

3.5 Continuous distributions

- Probability density function: smooth curve representing the continuous probability distribution (with total area under the curve of 1).
- Use area under the curve within a selected region to find probability of an event.

Chapter exercises

Guided Practice 3.52. <- this one really stumped me but got same answer using diff strategy!
Your department is holding a raffle. They sell 30 tickets and offer seven prizes. (a) They place the tickets in a hat and draw one for each prize. The tickets are sampled without replacement, i.e. the selected tickets are not placed back in the hat. What is the probability of winning a prize if you buy one ticket? (b) What if the tickets are sampled with replacement?

a) Without replacement - Your ticket is 1 out of 30. There are 7 observations out of 30 in the population = 7/30 = 23%, so this is a “sampling from a small population” situation and observations are not independent.
   <u>One way to solve (in textbook):</u> Probability of not winning on any round, subtract from 1:
      P(1st = 0) = 29/30 (where 0 = lose)
      P(2nd = 0) = 28/29
      P(3rd = 0) = 27/28 …. (4th: 27; 5th: 26, 6th: 25) ...
      P(7th = 0) = 23/24
   Question about probability that all of those scenarios happen (“and” situation) -> multiply!
   29/30 x 28/29 x 27/28 …. 24/25 x 23/24 = 23/30 = 0.7666
   1 - 0.7666 = 0.2333
   <u>Another way to solve (not in textbook):</u> Probability of winning on some round:
      1, 0, 0, 0, 0, 0 (where 1 = win, 0 = lose): 1/30 x 29/29 x 28/28… -> 1/30
      0, 1, 0, 0, 0, 0: 29/30 x 1/29 x 28/28 … -> 29/30 x 1/29 -> 1/30
      0, 0, 1, 0, 0, 0: 29/30 x 28/29 x 1/28 x 27/27 … -> 1/30
      0, 0, 0, 1, 0, 0: -> 1/30
      0, 0, 0, 0, 1, 0: -> 1/30
      0, 0, 0, 0, 0, 1: -> 1/30
   Question is about probability that any of these scenarios happen (“or” situation) -> add!
   1/30 + 1/30 + …. 1/30 = 7/30 = 0.2333

b) With replacement - observations are independent. This is a much complicated problem to solve when you try to compute the probability of winning 1 or many prizes!
   <u>One way to solve (in textbook):</u> Probability of not winning on any round, subtract from 1:
      Not win on 1st round: 29/30
      Not win on 2nd round: 29/30
      … on 7th round: 29/30
   Question about probability that all of these scenarios happen (“and” situation) -> multiply!
   $(29/30)^7$ = 0.7887
   1 - 0.7887 = 0.2112

3.42 Twins.
Draw tree diagram:

Identical 30% ---- MM 50% x 30% = 15%
---- FF 15%
Fraternal 70% ---- MM 25% x 70% = 17.5%
---- FF 17.5%
---- MF 50% x 70% = 35%

Denominator (total probability that you have FF) = 15% + 17.5% = 32.5%
Numerator (probability that they are identical) = 15%
Answer = 15%/32.5% = 46.15%