Review of counting and probability

The core of modern statistical data analysis and machine learning is probability theory. This notebook contains a couple problems that we hope will refresh your memory of some basic, but important, concepts.

Exercise 0: Odd Person Out. Suppose you roll two, fair, six-sided dice. What is the probability that their sum is even?

> Here, "fair" means unbiased or having uniformly random outcomes.

Answer. The answer is one-half ($\frac{1}{2}$). This exercise is intended as a review of basic counting as it occurs in probability theory.

Here is one way to see why. With two dice and six sides each, there are $6^2 = 36$ possible outcomes. The only ways to get an even sum are to roll both evens or both odds, e.g., $\{2, 4, 6\}$ on both dice or $\{1, 3, 5\}$ on both dice. There are $3^2 = 9$ ways to get both evens, and the same for odds---in other words, 18 ways to get an even sum. Thus, the chance of an even sum is $\frac{18}{36} = \frac{1}{2}$.

Exercise 1: Combinaisons. Consider a standard "French" deck of 52 cards: there are four (4) suits (diamonds, clubs, hearts, and spades) and, within each suit, there are 13 ranks (cards numbered 2-10 plus a Jack, a Queen, a King, and an Ace).

- **Part a)** Suppose you shuffle the deck and then select a *hand*, consisting of five (5) cards. (That is, you draw 5 cards uniformly at random from the deck of 52 without replacement.) How many distinct hands are there in total? "Distinct" means the ordering of cards within the hand does not matter.
- **Part b)** A *three-of-a-kind* hand has three (3) cards of the same rank, plus two more cards of *different* ranks. For instance, (7, 7, 7, Queen, 2) is a three-of-a-kind but (7, 7, 7, Queen, Queen) is not. How many possible three-of-a-kind hands are there in the deck?
- **Part c)** From the above, what is the probability of drawing a three-of-a-kind on the first try from a standard French deck?

> Hint: As the name of this exercise suggests, you will find the calculational tool of a combination (https://en.wikipedia.org/wiki/Combination) handy.

Answer. The main tool you need is a combination, which in the standard notation is defined to be,

$$\binom{n}{k} \equiv \frac{n!}{k!(n-k)!}.$$

This object is also known as the *binomial coefficient*. It counts the number of subsets of size k in a set consisting of n objects. Recall that a mathematical "set" or "subset" is not ordered, so we can apply it to this problem where, for instance, the ordering of cards within a hand is not important.

Part a) The fastest method for this problem is, of course, to Google it. But a more principled way is to recognize that a hand is simply a subset of 5 cards taken from a deck of 52. Therefore, the number of such subsets is

$$\binom{52}{5} = \frac{52!}{5!47!} = \frac{52 \cdot 51 \cdot 50 \cdot 49 \cdot 48}{5 \cdot 4 \cdot 3 \cdot 2} = 2,598,960 \text{ hands.}$$

Part b) Again, we can apply combinations. In this case, the hand of interest, a three-of-a-kind, is structured: it consists of a subset of size 3 of cards with the same rank, followed by any two cards of differing ranks. Let's consider each of these parts in turn.

- Recall that there are 13 ranks. Given a rank, we need three of the same kind. Therefore, there are $13 \times \binom{4}{3} = 52$ possible subsets of 3 cards having the same rank.
- The next two cards must have different ranks. Having "chosen" a rank for the three-of-a-kind, there are 12 ranks remaining. The number of ways to choose a pair of distinct ranks from these is $\binom{12}{2} = 66$.
- The suit of the remaining two cards does not matter. There are 4 possibilities for each card, or 16 possibilities for the pair.

Thus, the total number of three-of-a-kind hands is the product of these values, or $52 \times 66 \times 16 = 54,912$ hands.

Part c) This probability is simply the number of three-of-a-kind hands divided by the total number of possible hands, or about $(54,912)/(2,598,960) \approx 0.02$ (2%).

Exercise 2: Taxi hit-and-run. (This problem is adapted from the book, Thinking, Fast and Slow (http://www.nytimes.com/2011/11/27/books/review/thinking-fast-and-slow-by-daniel-kahneman-book-review.html?mcubz=0).)

Consider a city in which 85% of the taxis are green and 15% are blue.

One night, a taxi hits a parked car and then flees the scene. A witness claims the taxi was blue. However, it is also known that the reliability of witnesses at night is imperfect; a recent study suggests that the chance a witness makes a mistake is 20%.

What is the probability that the taxi was actually blue?

> Hint: This problem intends to exercise your knowledge of conditional probabilities and Bayes' rule (https://www.cs.ubc.ca/~murphyk/Bayes/bayesrule.html).

Answer. The answer is that there is a probability of 0.41 (41%) that the taxi was actually blue.

This problem relies on the idea of *conditional probabilities*, that is, the probability of one event *given* that another event has occurred. From them, you can apply Bayes' rule (or "law" or "theorem") and the law of total probability to calculate the solution. The key starting point for this problem is to identify the known and unknown conditional probabilities.

In this problem, there are two kinds of probabilistic events: a taxi having a certain color (green or blue) and the witness perceiving a certain color (green or blue). Let's associate the former with a random variable, T for taxi, and the latter with another random variable, W for witness.

What we want to compute is the conditional probability that the taxi was blue *given* that the witness reported blue, i.e.,

$$\Pr[T = \text{blue} \mid W = \text{blue}].$$

And here is what we are given in the problem statement:

- The *prior probabilities* of a taxi being green or blue, $\Pr[T = \text{green}] = 0.85$ and $\Pr[T = \text{blue}] = 0.15$.
- The chances that the witness got it right, i.e., $\Pr[W = \text{blue} \mid T = \text{blue}] = \Pr[W = \text{green} \mid T = \text{green}] = 0.80$.
- The chances that the witness got it wrong, i.e., $\Pr[W = \text{blue} \mid T = \text{green}] = \Pr[W = \text{green} \mid T = \text{blue}] = 0.20$.

Bayes' rule, in the abstract, says for two events A and B,

$$\Pr[A \mid B] = \frac{\Pr[B \mid A] \cdot \Pr[A]}{\Pr[B]}.$$

So for our problem, that is

$$\Pr[T = \text{blue} \mid W = \text{blue}] = \frac{\Pr[W = \text{blue} \mid T = \text{blue}] \cdot \Pr[T = \text{blue}]}{\Pr[W = \text{blue}]}.$$

While we aren't given the value of the denominator, $\Pr[W = \text{blue}]$, explicitly, we *can* derive it from the information we do have using the law of total probability (https://en.wikipedia.org/wiki/Law_of_total_probability), which tells us how to compute the total probability of an event B from its known probabilities conditional on a different event A along with A's prior:

$$\Pr[B] = \sum_a \Pr[B \,|\, A = a] \cdot \Pr[A = a].$$

Applying this rule to our problem yields,

$$\begin{aligned}\Pr[W = \text{blue}] &= (\underbrace{\Pr[W = \text{blue} \mid T = \text{blue}]}_{=0.80} \cdot \underbrace{\Pr[T = \text{blue}]}_{=0.15}) + (\underbrace{\Pr[W = \text{blue} \mid T = \text{green}]}_{=0.20} \cdot \underbrace{\Pr[T = \text{green}]}_{=0.85}) \\ &= 0.29.\end{aligned}$$

Thus,

$$\begin{aligned}\Pr[T = \text{blue} \mid W = \text{blue}] &= \frac{\Pr[W = \text{blue} \mid T = \text{blue}] \cdot \Pr[T = \text{blue}]}{\Pr[W = \text{blue}]} \\ &= \frac{0.80 \cdot 0.15}{0.29} \\ &\approx 0.41,\end{aligned}$$

or roughly 41%. Thus, there is a significant chance that the taxi was not blue!

Calculus practice problems

Exercise 0: Chain (and product) of fools.... Compute the derivative $\frac{df}{dx}$ where $f(x) \equiv x \cdot e^{-x^2}$.

Answer. This problem contains applications of two basic rules when taking derivatives.

The first rule is the *product rule*,

$$\frac{d}{dx}[a(x) \cdot b(x)] = \frac{da}{dx} \cdot b + a \cdot \frac{db}{dx}.$$

The second rule is the *chain rule*,

$$\frac{d}{dx} g\left(h(x)\right) = \frac{dg(y)}{dy} \cdot \frac{dy}{dx},$$

where $y = h(x)$.

Let

$$\begin{aligned} a(x) &\equiv x \\ b(x) &\equiv e^x \\ c(x) &\equiv -x^2. \end{aligned}$$

Then, we can rewrite $f(x)$ as

$$f(x) = a(x) \cdot b(c(x)).$$

Thus,

$$\begin{aligned} \frac{df}{dx} &= \frac{da}{dx} \cdot b(c(x)) + a \cdot \left[\frac{db(y)}{dy} \cdot \frac{dy}{dx}\right] \\ \text{where} \quad y(x) &= c(x). \end{aligned}$$

Let's calculate each of the relevant pieces.

$$\begin{aligned} \frac{da}{dx} &= 1 \\ \frac{db(y)}{dy} &= e^y \\ \text{and} \quad \frac{dy}{dx} = \frac{dc}{dx} &= -2x. \end{aligned}$$

Plugging these back in yields the final result,

$$\begin{aligned} \frac{df}{dx} &= e^{-x^2} + x \cdot \left[e^{-x^2} \cdot (-2x)\right] \\ &= \left(1 - 2x^2\right) e^{-x^2}. \end{aligned}$$

Exercise 1: Gamma, gamma, go gamma. Consider the following function.

$$\Gamma(a) \equiv \int_0^\infty t^{a-1} e^{-t} dt,$$

where a is a real number. (In fact, a can be any complex number other than the negative integers, but for this problem, let's not consider that.)

Show the following.

$$\Gamma(a) = (a-1)\Gamma(a-1).$$

> Bonus: What is $\Gamma(n)$ when n is a positive integer?

Answer. This function is known as the *gamma function*, and generalizes the factorials from integers to the complex plane.

The standard way to solve this problem is to apply integration by parts:

$$\int u dv = uv - \int v du.$$

Let

$$\begin{aligned} u \equiv t^{a-1} \;\; dv = e^{-t} dt \\ du = (a-1)t^{a-2} \;\; v = -e^{-t}. \end{aligned}$$

Then,

$$\begin{aligned} \int_0^\infty t^{a-1} e^{-t} dt &= \int_0^\infty u dv \\ &= uv - \int_0^\infty v du \\ &= \underbrace{-\left.\frac{t^{a-1}}{e^t}\right|_0^\infty}_{=0} + (a-1) \underbrace{\int_0^\infty t^{a-2} e^{-t} dt}_{=\Gamma(a-1)} \\ &= (a-1)\Gamma(a-1). \end{aligned}$$

`w00t!`

Linear Algebra Review

The goal of this lab (homework) is to force you to refresh your linear algebra chops. Before starting it, please read the linear algebra review notes linked to below. These were originally written by Da Kuang (http://math.ucla.edu/~dakuang/), who taught CSE 6040 in Fall 2014.

- Review notes (https://cse6040.gatech.edu/fa17/kuang-linalg-notes.pdf)

Part 1: Plug and chug

While the following problems can be done using a computer, please do them *by hand*. The purpose is to reinforce the key definitions from the linear algebra notes . that we will use repeatedly.

Let

$$x = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$$

and

$$A = \begin{pmatrix} 2 & 1 & 0 \\ 4 & 3 & 1 \\ 6 & 7 & 8 \end{pmatrix}.$$

Compute the following.

Exercise 0 (0 points). Compute $x + x$ of the numerical x given above.

Sample answer 1.

$$x + x = 2x = 2 \cdot \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} = \begin{pmatrix} 2 \\ 4 \\ 6 \end{pmatrix}.$$

Sample answer 2.

$$\begin{aligned} x + x &= 2 \cdot x \\ &= 2 \cdot \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \\ &= \begin{pmatrix} 2 \\ 4 \\ 6 \end{pmatrix}. \end{aligned}$$

Exercise 1 (2 points). Compute the *inner-product* (or *dot-product*), $x^T x$, of the numerical x given above.

$$\begin{pmatrix} 1 & 2 & 3 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} = 1 \cdot 1 + 2 \cdot 2 + 3 \cdot 3 = 1 + 4 + 9 = 14$$

Exercise 2 (2 points). Compute the *outer-product*, xx^T, of the numerical x given above.

$$\begin{aligned} xx^T &= \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \cdot \begin{pmatrix} 1 & 2 & 3 \end{pmatrix} \\ &= \begin{pmatrix} 1 \cdot 1 & 1 \cdot 2 & 1 \cdot 3 \\ 2 \cdot 1 & 2 \cdot 2 & 2 \cdot 3 \\ 3 \cdot 1 & 3 \cdot 2 & 3 \cdot 3 \end{pmatrix} \\ &= \begin{pmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ 3 & 6 & 9 \end{pmatrix}. \end{aligned}$$

Exercise 3 (2 points). Compute the *vector 1-norm*, $\|x\|_1$, of the numerical x given above.

$$\|x\|_1 = \sum_{i=1}^{3} |x_i| = 1 + 2 + 3 = 6.$$

Exercise 4 (2 points). Compute the *vector 2-norm* (or *Euclidean length* or *distance*), $\|x\|_2$, of the numerical x given above.

$$\|x\|_2 = \left(\sum_{i=1}^{3} |x_i|^2 \right)^{1/2} = \sqrt{1 + 4 + 9} = \sqrt{14} \approx 3.74.$$

Exercise 5 (2 points). Compute the *infinity-norm*, $\|x\|_\infty$, of the numerical x given above.

$$\|x\|_\infty = \max_{i=1}^{3} |x_i| = 3.$$

Exercise 6 (2 points). Compute the Frobenius norm, $\|A\|_F$, of the numerical matrix A given above.

$$
\begin{aligned}
\|A\|_F &= \left\| \begin{pmatrix} 2 & 1 & 0 \\ 4 & 3 & 1 \\ 6 & 7 & 8 \end{pmatrix} \right\|_F \\
&= \left(\sum_{i=1}^{3} \sum_{j=1}^{3} |a_{ij}|^2 \right)^{\frac{1}{2}} \\
&= \sqrt{2^2 + 1^2 + 0^2 + 4^2 + 3^2 + 1^2 + 6^2 + 7^2 + 8^2} \\
&= \sqrt{4 + 1 + 0 + 16 + 9 + 1 + 36 + 49 + 64} \\
&= \sqrt{180} \\
&= \sqrt{9 \cdot 4 \cdot 5} \\
&= 6\sqrt{5} \\
&\approx 13.4.
\end{aligned}
$$

Exercise 7 (2 points). Compute the product, $A^T A$, of the numerical matrix A given above.

$$
\begin{aligned}
A^T A &= \begin{pmatrix} 2 & 4 & 6 \\ 1 & 3 & 7 \\ 0 & 1 & 8 \end{pmatrix} \cdot \begin{pmatrix} 2 & 1 & 0 \\ 4 & 3 & 1 \\ 6 & 7 & 8 \end{pmatrix} \\
&= \begin{pmatrix} 2 \cdot 2 + 4 \cdot 4 + 6 \cdot 6 & 2 \cdot 1 + 4 \cdot 3 + 6 \cdot 7 & 2 \cdot 0 + 4 \cdot 1 + 6 \cdot 8 \\ 1 \cdot 2 + 3 \cdot 4 + 7 \cdot 6 & 1 \cdot 1 + 3 \cdot 3 + 7 \cdot 7 & 1 \cdot 0 + 3 \cdot 1 + 7 \cdot 8 \\ 0 \cdot 2 + 1 \cdot 4 + 8 \cdot 6 & 0 \cdot 1 + 1 \cdot 3 + 8 \cdot 7 & 0 \cdot 0 + 1 \cdot 1 + 8 \cdot 8 \end{pmatrix} \\
&= \begin{pmatrix} 56 & 56 & 52 \\ 56 & 59 & 59 \\ 52 & 59 & 65 \end{pmatrix}.
\end{aligned}
$$

Exercise 8 (2 points). Compute the product, AA^T, of the numerical matrix A given above.

$$
\begin{aligned}
AA^T &= \begin{pmatrix} 2 & 1 & 0 \\ 4 & 3 & 1 \\ 6 & 7 & 8 \end{pmatrix} \cdot \begin{pmatrix} 2 & 4 & 6 \\ 1 & 3 & 7 \\ 0 & 1 & 8 \end{pmatrix} \\
&= \begin{pmatrix} 2 \cdot 2 + 1 \cdot 1 + 0 \cdot 0 & 2 \cdot 4 + 1 \cdot 3 + 0 \cdot 1 & 2 \cdot 6 + 1 \cdot 7 + 0 \cdot 8 \\ 4 \cdot 2 + 3 \cdot 1 + 1 \cdot 0 & 4 \cdot 4 + 3 \cdot 3 + 1 \cdot 1 & 4 \cdot 6 + 3 \cdot 7 + 1 \cdot 8 \\ 6 \cdot 2 + 7 \cdot 1 + 8 \cdot 0 & 6 \cdot 4 + 7 \cdot 3 + 8 \cdot 1 & 6 \cdot 6 + 7 \cdot 7 + 8 \cdot 8 \end{pmatrix} \\
&= \begin{pmatrix} 5 & 11 & 19 \\ 11 & 26 & 53 \\ 19 & 53 & 149 \end{pmatrix}.
\end{aligned}
$$

Exercise 9 (5 points). The *trace* of a matrix M, or $\text{trace}(M)$, is defined to be the sum of M's diagonal entries, i.e., $\text{trace}(M) = \sum_i m_{ii}$.

Let A be any real-valued matrix of size $m \times n$. Show that $\text{trace}(A^T A) = \text{trace}(AA^T) = \|A\|_F^2$.

> In this question, show the desired claim for *any* $m \times n$ matrix A, **not** for the numerical matrix A given above.

Let $A \begin{pmatrix} a_1 & a_2 & \cdots & a_n \end{pmatrix}$, where a_j is one of the columns of the n columns of A and each column is a vector of length m. Then,

$$A^T A = \begin{pmatrix} a_1^T \\ a_2^T \\ \vdots \\ a_n^T \end{pmatrix} \cdot \begin{pmatrix} a_1 & a_2 & \cdots & a_n \end{pmatrix} = \begin{pmatrix} a_1^T a_1 & a_1^T a_2 & \cdots & a_1^T a_n \\ a_2^T a_1 & a_2^T a_2 & \cdots & a_2^T a_n \\ \vdots & \vdots & \vdots & \vdots \\ a_n^T a_1 & a_n^T a_2 & \cdots & a_n^T a_n \end{pmatrix}.$$

Observe that each diagonal entry j equals $a_j^T a_j$. Thus,

$$\text{trace}(A^T A) = \sum_{j=1}^{n} a_j^T a_j = \sum_{j=1}^{n} \sum_{i=1}^{m} a_{ij}^2 = \|A\|_F^2.$$

Now consider the product AA^T:

$$AA^T = \begin{pmatrix} a_1 & a_2 & \cdots & a_n \end{pmatrix} \cdot \begin{pmatrix} a_1^T \\ a_2^T \\ \vdots \\ a_n^T \end{pmatrix} = a_1 a_1^T + a_2 a_2^T + \cdots + a_n a_n^T = \sum_{i=1}^{n} a_i a_i^T.$$

Each product $a_i a_i^T$ is an $m \times m$ matrix whose diagonal entries are a_{ij}^2 for $j \in \{1, \ldots, n\}$. Thus, the i-th diagonal entry of AA^T is $\sum_{j=1}^{n} a_{ij}^2$ and

$$\text{trace}(AA^T) = \sum_{i=1}^{m} \left(i\text{-th diagonal entry of } AA^T\right) = \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij}^2 = \|A\|_F^2.$$

Thus, $\text{trace}(A^T A) = \text{trace}(AA^T) = \|A\|_F^2$.

Exercise 10 (5 points). Suppose x is now unknown and $b = (1, 1, 1)^T$. Determine a vector x such that $Ax = b$, given the numerical 3×3 matrix A above. You may use *any* method, but show your work. (Methods might include basic high school algebra starting from the definitions of a matrix-vector product, or Gaussian elimination, or LU factorization.)

The system we wish to solve is the following.

$$Ax = \begin{pmatrix} 2 & 1 & 0 \\ 4 & 3 & 1 \\ 6 & 7 & 8 \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}.$$

This system is equivalent to

$$
\begin{aligned}
2x_1 + x_2 &= 1 \\
4x_1 + 3x_2 + x_3 &= 1 \\
6x_1 + 7x_2 + 8x_3 &= 1.
\end{aligned}
$$

First, let's eliminate x_3 by multiplying the second equation by 8 and subtracting the third equation from it. Then,

$$(32 - 6)x_1 + (24 - 7)x_2 = 8 - 1,$$

or

$$26x_1 + 17x_2 = 7.$$

Next, let's eliminate x_2 by subtracting the above equation from 17 times the very first equation. Thus,

$$(34 - 26)x_1 = (17 - 7),$$

or

$$8x_1 = 10 \Longrightarrow x_1 = \frac{5}{4}.$$

Plugging this back into the very first equation yields

$$2 \cdot \left(\frac{5}{4}\right) + x_2 = 1 \Longrightarrow x_2 = 1 - \frac{5}{2} = -\frac{3}{2}.$$

Lastly, plugging $(x_1, x_2) = \left(\frac{5}{4}, -\frac{3}{2}\right)$ into, say, the second of the original equations gives

$$x_3 = 1 - 4 \cdot \frac{5}{4} - 3 \cdot \left(-\frac{3}{2}\right) = \frac{1}{2}.$$

As a sanity check, try plugging $(x_1, x_2, x_3) = \left(\frac{5}{4}, -\frac{3}{2}, \frac{1}{2}\right)$ into the last of the original equations and verify that this solution also satisfies it.

Part 2: A proof [5 points]

Exercise 11 (5 points). Let $Q \in \mathcal{R}^{m \times m}$ be an orthogonal matrix, and let $A \in \mathcal{R}^{m \times n}$ be some matrix. Prove that $\|QA\|_F^2 = \|A\|_F^2$, that is, left-multiplying by an orthogonal matrix does not change the Frobenius norm.

> Hint: Consider the definitions of an orthogonal matrix, Frobenius norm, vector 2-norm, and the results of Exercise 9.

Answer. The desired fact follows from

$$\|QA\|_F^2 = \mathrm{trace}(A^T Q^T QA) = \mathrm{trace}(A^T A) = \|A\|_F^2,$$

where $\text{trace}(A^T Q^T QA) = \text{trace}(A^T A)$ follows from the definition of orthogonality, namely, that $Q^T Q = I$, the identity matrix.