

1. Rules of Probability Theory

Identify which of the following statements are correct and which are incorrect. Explain in one sentence why you identified each statement as correct or incorrect.

1. $P(A, B) = P(A|B)P(B)$

True. The joint probability of event A and B concurrently happening can also be thought of as the probability that event B occurs, and that A occurs within the probability space of B.

2. $P(A) = P(A|B)P(B) + P(A|\neg B)P(\neg B)$

True. If the equation in Question 1 is true, then both sides of the statement simplify to

$$P(A) = P(A, B) + P(A, \neg B)$$

This is clearly true, as the probability of A happening is when it happens along with B and when it does without B. In other words, B can be ‘summed out’.

3. $P(A) = P(A|B)P(B) + P(A|C)P(C) + P(A|D) + P(D)$

False. The RHS simplifies to $P(A, B) + P(A, C) + P(A|D) + P(D)$.

4. $P(A|B) = \frac{P(B|A)P(A)}{P(B)}$

True. Multiplying both sides by $P(B)$ shows $P(A, B) = P(A, B)$.

5. $P(A|B)P(B) = P(B|A)P(A)$

True. This is obtained by multiplying both sides of Question 4 by $P(B)$.

2. Logarithms and Probability Distributions

What is the log of the probability density function (pdf) of each of the following probability distributions?

For the normal distribution, $P(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp(-\frac{(x-\mu)^2}{2\sigma^2})$
 $\implies \log(P(x)) = -\log(\sqrt{2\pi}\sigma) - \frac{(x-\mu)^2}{2\sigma^2}$

For the exponential distribution, $P(x) = \lambda e^{-\lambda x}$
 $\implies \log(P(x)) = \log(\lambda) - \lambda x$

For the gamma distribution, $P(x) = \frac{1}{\Gamma(k)\theta^k} x^{k-1} e^{-x/\theta}$
 $\implies \log(P(x)) = -\log(\Gamma(k)\theta) - \lambda x + (k-1)\log(x) - x/\theta$

3. Normal Distribution

If x is distributed according to the normal distribution with mean μ and standard deviation σ , and if $f(x) = x^3 + 3$,

1. Calculate the expected value of $f(x)$
 $E[x^3 + 3] = E[x^3] + 3$. The moment generating function of the normal distribution is

$$M_x(t) = E(x^{xt}) = \int e^{xt} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}(x-\mu)^2/\sigma^2} dx$$

Use the substitution $z = \frac{x-\mu}{\sigma} \implies x = z\sigma + \mu \implies dx/dz = \sigma$ to obtain

$$\begin{aligned} M_x(t) &= e^{\mu t} \int e^{z\sigma t} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}z^2} \left| \frac{dx}{dz} \right| dz \\ &= e^{\mu t} \int e^{z\sigma t} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz \\ &= e^{\mu t} e^{\frac{1}{2}\sigma^2 t^2} \end{aligned}$$

Therefore the MGF of the normal distribution is $\exp\{\mu t + \sigma^2 t^2/2\}$. Since $E[x^3]$ is the third moment of x , $E[x^3]$ is given as the third derivative of the MGF at $t = 0$:

$$E[x^3] = \frac{d^3 M_x(0)}{dt^3} = (\mu + \sigma^2 t)(3\sigma^2 + (\mu + \sigma^2 t)^2) e^{\mu t + \sigma^2 t^2/2} \Big|_{t=0} = 3\sigma^2 \mu + \mu^3$$

Thus the expected value of $f(x)$ is $3\sigma^2 + \mu^3 + 3$

2. Calculate the probability $P(f(x) > 0.5)$
 Through rearrangement, $P(x^3 + 3 > 0.5) \iff P(x > \sqrt[3]{-2.5})$, which is evaluated as:

$$\frac{1}{2} \operatorname{erf}\left(\frac{\sqrt[3]{-2.5} - \mu}{\sqrt{2}\sigma}\right) = \frac{1}{\sqrt{\pi}} \int_0^{\frac{\sqrt[3]{-2.5} - \mu}{\sqrt{2}\sigma}} e^{-t^2} dt$$

3. Write a Python script to confirm your answer to question 2. Generate a lot of random numbers from a normal distribution with a particular mean and standard deviation. Calculate $f(x)$ for each of these random numbers. How many of them are greater than 0.5? Does that match the probability you calculated in question 2?

```
from scipy.special import erfc

def function(mu, sigma):
    arr = np.random.normal(mu, sigma, 100000)
    fx = (np.power(arr, 3) + 3 > 0.5)
    proportion = sum(fx)/100000

    theoretical = 1/2 * erfc((-2.5**(1/3) - mu) / (np.sqrt(2) * sigma))
    return proportion, theoretical

print(function(1,2)) # (0.88166, 0.8807221314443912)
print(function(-23,56)) # (0.3505, 0.3495711930793632)
print(function(100,100)) # (0.84528, 0.844606509043253)
```

4. Marginal and Conditional Probabilities

In a country with high unemployment, 27.3% of the working population considered “young”, and of the young working population, 38.2% are unemployed. The unemployment rate for those in the working population who are not young, is lower at 22.4%.

Write down or calculate, as required, all the following marginal and conditional probabilities for the working population in this scenario.

$$P(Y) = 0.273 \text{ as given.}$$

$$P(\neg Y) = 1 - P(Y) = 1 - 0.273 = 0.727$$

$$P(U) = P(U|Y)P(Y) + P(U|\neg Y)P(\neg Y) = 0.382 * 0.273 + 0.224 * 0.727 = 0.267134$$

$$P(\neg U) = 1 - P(U) = 1 - 0.267134 = 0.732866$$

$$P(U|Y) = 0.382 \text{ as given.}$$

$$P(Y|U) = P(U|Y)P(Y)/P(U) = 0.382 * 0.273 / 0.267134 = 0.39038834$$

$$P(U|\neg Y) = 0.224 \text{ as given.}$$

$$P(\neg Y|U) = 1 - P(Y|U) = 1 - 0.39038834 = 0.609611656$$

5. Inference

You take a guess and think there is a 30% chance that Olivia, 5 year old girl, will be able to read at Grade 1 level by the time she turns 6. (Kids usually start learning to read later, at school, so this isn't a terrible guess.)

An educational expert tells you that of children who are already able to read at Grade 1 level by the time they turn 6 years old, 60% had training (usually by their parents) in pronouncing simple words and writing some letters, and 40% did not. On the other hand, of children who are not already able to read at Grade 1 level by age 6, only 10% had the same training.

1. After receiving this information, what is your revised estimate of the probability that Olivia will be able to read at that level if you learned that her parents are currently busy doing basic training for reading with her?

From the question, $P(R) = 0.3$, $P(T|R) = 0.6$, $P(T|\neg R) = 0.1$

$$\begin{aligned} P(R|T) &= \frac{P(T|R) * P(R)}{P(T)} \\ &= \frac{P(T|R) * P(R)}{P(T|R) * P(R) + P(T|\neg R) * P(\neg R)} \\ &= \frac{0.6 * 0.3}{0.6 * 0.3 + 0.1 * (1 - 0.3)} = 0.72 \end{aligned}$$

2. The above is a very simple statistical model. In a paragraph of 80 to 120 words outline the structure of the model. What are the variables? How are these variables related?

In this simple model, we essentially have a guess, or a prior probability of 0.3, which is updated to a posterior distribution. We are given two variables;

- (a) R : The ability of a 5 year old will be able to read at Grade 1 level when they turn 6,
- (b) T : The training said child may receive which may increase R .

The question in Part I is essentially asking, of all the times children undergo training (of which Olivia is one), what proportion successfully read by the time they turn 6?

$$P(R|T) = \frac{P(R, T)}{P(T)}$$

While neither the numerator nor denominator are directly given, they can be further expanded into subprobabilities for which we do have data. $P(R, T)$ is easily expanded as $P(T|R) \times P(R)$. Clearly, the probability that a child both reads and receives training is when it receives training, and that it reads after receiving training $P(T) * P(R|T)$, but logically that it is the probability that it can read, and retrospectively the likelihood it receives training is true as well, and it is only for the latter that we have data.

Similarly for the denominator, expanding the probability of a randomly selected child having undergone training can be divided, seemingly trivially, as the probability of when it first receives training and when the child can successfully read, as well as when it first receives training and did not, $P(T) = P(T, R) + P(T, \neg R)$. The first term is the same as the numerator, and the second is similarly expandable.

6. Python Script

1. How many students are in your CS146 class (including yourself)?
16, but this will be generalizable as N students.
2. Assuming you know nothing about your classmates,
 - (a) What is the probability that you were born after all of them? $1/N$
 - (b) What is the probability that you were born before all of them? $1/N$
 - (c) What is the probability that you were born after at least half of them?
 $1/2$ if N is even, and $(N - 1)/2N$ if N is odd.

7. Python Script

Confirm your solution to Exercise 6 numerically by writing a Python script. They're harder than you think and there's a good chance you got them wrong without realizing it.

Generate a list of uniformly distributed random numbers using the `random.uniform()` function in the `numpy` module.

1. Check whether the first number is greater than all other numbers
2. Check whether the first number is less than all other numbers.
3. Check whether the first number is greater than at least half of the other numbers.

Do this 1,000,000 times and count how often each of the above three events happen. Use your counts to report the approximate probability for each event. These values should match your answers to questions 6ac in Exercise 6 above. If they do, congratulations! your intuition for probability problems is well-developed. If your answers do not match, try to figure out where you went wrong and update your solutions to Exercise 6.

```
def greater(n, length):
    # creates a N * length dimensional array.
    # each column therefore contains the numbers 1 - N only once.
    x = np.argsort(np.random.rand(n,length),axis=0)+1
    return np.all(x[0] > x[1:], axis = 0)

sum(greater(18, 1000000))/1000000 # 0.05531, expected 1/18 = 0.0555..

def less(n, length):
    x = np.argsort(np.random.rand(n,length),axis=0)+1
    return np.all(x[0] < x[1:], axis = 0)

sum(less(18, 1000000))/1000000 # 0.055409, expected 1/18 = 0.0555..

def atleasthalf(n, length):
    x = np.argsort(np.random.rand(n,length),axis=0)+1
    return sum(x[0] > x[1:]) >= n/2

sum(atleasthalf(17, 1000000))/1000000 # 0.470272, expected 8/17 = 0.47058
sum(atleasthalf(18, 1000000))/1000000 # 0.500845, expected 1/2
```

8. Logarithms

If $p \in [0, 1]$ is a probability, $\log(\frac{p}{1-p})$, is known as the log-odds or the logit of p . This function comes up a lot in data modeling.

1. If an event has 10:1 odds of happening, what is the probability of the event? What is the logit of this probability when calculated using base 10, base 2, and base e ?

$$\frac{p}{1-p} = X \implies p = \frac{X}{X+1}$$

For $X = 10$, $p = 10/11$. $\log_{10}(\frac{p}{1-p}) = 1$, $\log_2(\frac{p}{1-p}) = 3.3219$, $\ln(\frac{p}{1-p}) = 2.302$

2. If an event has 100:1 odds of happening, what is the probability of the event? What is the logit of this probability when calculated using base 10, base 2, and base e ?

For $X = 100$, $p = 100/101$. $\log_{10}(\frac{p}{1-p}) = 2$, $\log_2(\frac{p}{1-p}) = 6.6438$, $\ln(\frac{p}{1-p}) = 4.605$

3. Why does it not really matter match which base we use when calculating the logit function, as long as we always use the same base in our calculations?

This is because of the ratio between the probabilities are the same. For example, The answers in part 1 have the ratio $1 : 3.32 : 2.3$, which is equal to the ratio in part 2 of $2 : 6.64 : 4.6$.

9. Bags and Cookies

There are two bags containing 3 types of cookies chocolate, vanilla, and caramel. The first bag has 1 chocolate cookie and 2 vanilla cookies. The second bag has 1 chocolate cookie and 3 caramel cookies. Without looking inside, you stick your hand in one of the two bags and take out a cookie, which turns out to be chocolate. You proceed to eat the cookie.

1. What is the probability that you picked the bag containing the vanilla cookies?

$$\begin{aligned} \text{From Bayes' Theorem, } P(\text{Bag}_V | \text{Chocolate}) &= \frac{P(\text{Chocolate} | \text{Bag}_V)P(\text{Bag}_V)}{P(\text{Chocolate})} \\ &= \frac{1/3 * 1/2}{1/3 * 1/2 + 1/2 * 1/4} = 4/7 \end{aligned}$$

2. If you now take a cookie out of the other bag, what is the probability that the cookie is chocolate, vanilla, or caramel?

For clarity, $P(\text{Chocolate})$, $P(\text{Vanilla})$, and $P(\text{Caramel})$ will now mean the probability of choosing a chocolate, vanilla and caramel cookie respectively from the **second** bag. $P(\text{Bag}_V)$ and $P(\text{Bag}_C)$ will similarly now mean the probability that the **first** bag chosen was the bag containing vanilla and chocolate cookies respectively. For example, now $P(\text{Bag}_V) = 4/7$ as given in Part 1.

$$\begin{aligned} P(\text{Chocolate}) &= P(\text{Chocolate} | \text{Bag}_C)P(\text{Bag}_C) + P(\text{Chocolate} | \text{Bag}_V)P(\text{Bag}_V) \\ &= 1/3 * (1 - 4/7) + 1/4 * 4/7 = 2/7 \end{aligned}$$

$$\begin{aligned} \text{Similarly, } P(\text{Caramel}) &= P(\text{Caramel} | \text{Bag}_C)P(\text{Bag}_C) + P(\text{Caramel} | \text{Bag}_V)P(\text{Bag}_V) \\ &= 0 * (1 - 4/7) + 3/4 * 4/7 = 3/7 \end{aligned}$$

$$\begin{aligned} P(\text{Vanilla}) &= P(\text{Vanilla} | \text{Bag}_C)P(\text{Bag}_C) + P(\text{Vanilla} | \text{Bag}_V)P(\text{Bag}_V) \\ &= 2/3 * (1 - 4/7) + 0 * 4/7 = 2/7 \end{aligned}$$