CS146 Assignment 1

Diagnostic assessment
Oscar Engelbrektson, Fall 2019

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)

Correct. The denominator in Bayes Theorem's can be written as either.

2. P(A) = P(A | B) P(B) + P(A | not B) P(not B)

Correct. P(B) + P(notB)=1. From Eq.1, P(A | B) P(B) + P(A |not B) P(not B) = P(A, B) + P(A, notB), which is equivalent to P(A) if A and P are independent.

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

Incorrect. There is nothing to suggest that P(A, B) + P(A, C)... should sum to A. We know nothing of D, C, B to justify that type of conclusion.

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

Correct. Bayes' Theorem for two variables.

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

Correct. Multiplying Eq.4 by P(B) yields Eq.5.

2. Logarithms and probability distributions

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

1. Normal Distribution

Starting form the PDF of the normal distribution,

In[1]:=

$$\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Out[1]=

$$\frac{e^{-\frac{(x-\mu)^2}{2\sigma^2}}}{\sqrt{2}\sqrt{\pi\sigma^2}}$$

In[2]:=

```
ClearAll[μ, σ, x]
(*By the Logarithm qoutient and product rules*)
```

$$\text{Log}\left[\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\right] == \left(\text{Log}[1] - \text{Log}\left[\sqrt{2\pi\sigma^2}\right]\right) + \text{Log}\left[e^{-\frac{(x-\mu)^2}{2\sigma^2}}\right]$$

```
(*Log[1]=0; Log[e]=1*)
```

$$== -\text{Log}\left[\sqrt{2\pi\sigma^2}\right] - \frac{(x-\mu)^2}{2\sigma^2}$$

```
(*By the Logarithm power rule*)
```

$$= -\frac{1}{2}\text{Log}\left[2\pi\sigma^2\right] - \frac{(x-\mu)^2}{2\sigma^2}$$

```
(*by the logarithm product rule*)
```

$$= -\frac{1}{2}\text{Log}[2\pi] + \text{Log}[\sigma] - \frac{(x-\mu)^2}{2\sigma^2}$$

Out[3]=

$$\text{Log}\left[\frac{e^{-\frac{(x-\mu)^2}{2\sigma^2}}}{\sqrt{2}\sqrt{\pi\sigma^2}}\right] == \text{Log}\left[e^{-\frac{(x-\mu)^2}{2\sigma^2}}\right] - \text{Log}\left[\sqrt{2}\sqrt{\pi\sigma^2}\right]$$

2. Gamma distribution

Starting form the PDF of the Gamma distribution,

In[4]:=
$$\frac{1}{\Gamma(k)\theta^k} x^{k-1}\, e^{\frac{-x}{\theta}}$$

(*By the Logarithm qoutient and product rules; simplifying Log[1]=0*)

$$-\mathrm{Log}[\Gamma(k)\theta^k] + \mathrm{Log}[x^{k-1}] + \mathrm{Log}[e^{\frac{-x}{\theta}}]$$

(*Log[e]=1*)

$$-\mathrm{Log}[\Gamma(k)\theta^k] + \mathrm{Log}[x^{k-1}] + \frac{-x}{\theta}$$

(*By the Logarithm product and power rules*)

$$-\mathrm{Log}[\Gamma(k)] + k\mathrm{Log}[\theta] + (k-1)\mathrm{Log}[x] + \frac{-x}{\theta}$$

Out[4]= $\frac{e^{-\frac{x}{\theta}}\, x^{-1+k}\, \theta^{-k}}{k\,\Gamma}$

Out[5]= $\mathrm{Log}[e^{-\frac{x}{\theta}}] + \mathrm{Log}[x^{-1+k}] - \mathrm{Log}[k\,\Gamma\,\theta^k]$

Out[6]= $-\frac{x}{\theta} + \mathrm{Log}[x^{-1+k}] - \mathrm{Log}[k\,\Gamma\,\theta^k]$

Out[7]= $-\frac{x}{\theta} + k\mathrm{Log}[\theta] + (-1+k)\,\mathrm{Log}[x] - \mathrm{Log}[k\,\Gamma]$

3. Beta distribution

PDF	$\frac{x^{\alpha-1}(1-x)^{\beta-1}}{\mathrm{B}(\alpha,\beta)}$ where $\mathrm{B}(\alpha,\beta)=\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$ and Γ is the Gamma function.

Starting form the PDF of the Beta distribution,

In[8]:=
$$\frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)}$$

(*y the Logarithm qoutient and product rules*)

$$\mathrm{Log}\left[\frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)}\right] == \mathrm{Log}[x^{\alpha-1}] + \mathrm{Log}[(1-x)^{\beta-1}] - \mathrm{Log}[B(\alpha,\beta)]$$

(*By the Logarithm power rule*)

$$(\alpha-1)\mathrm{Log}[x] + (\beta-1)\mathrm{Log}[1-x] - \mathrm{Log}[B(\alpha,\beta)]$$

3. Normal distribution

If x is distributed according to the normal distribution with mean μ and standard deviation σ, *and if* $f(x) = x^3 + 2x + 1$

1. Calculate the expected value of f(x).

Expected value of $f(x) = x^3 + 2x + 1$, if *x is distributed according to the normal distribution with mean* μ *and standard deviation* σ*:*

In[8]:=
```
Expectation[x^3+2x+1, x≈NormalDistribution[μ,σ]]
```

Out[8]= $1 + 2\mu + \mu^3 + 3\mu\sigma^2$

2. Calculate the probability P(f(x) > 1).

Probability that $f(x) = x^3 + 2x + 1 > 1$, if *x is distributed according to the normal distribution with mean* μ *and standard deviation* σ*:*

In[9]:=
```
Probability[x^3+2x+1>1,x≈NormalDistribution[μ,σ]]
```

Out[9]= $1 - \frac{1}{2}\,\mathrm{Erfc}\left[\frac{\mu}{\sqrt{2}\,\sigma}\right]$

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 1?
Does that match the probability you calculated in question 2?

```python
from scipy import stats

#Define f(x)
def f(x):
    return x**3 + 2*x +1

#Normal distribution with \[Mu] = 0 and \[Sigma] = 1
dist = stats.norm(0, 1)

samples = 100000
success_counter = 0

for i in range(samples):

    #Draw random sample x from normal distribution
    x = dist.rvs()

    #If f(x) > 1, increment counter
    if f(x) > 1:
        success_counter += 1

print("Number of times f(x)>1 over 100 000 samples:", success_counter,
      "\n Fraction of samples x for which f(x) > 1:", success_counter/samples)
```

```
Number of times f(x)>1 over 100 000 samples: 50072
 Fraction of samples x for which f(x) > 1: 0.50072
```

The particular mean and standard deviation chosen are 0 and 1, respectively.
To evaluate if this matches the probability we got in question 2, we plugging the same parameters into our analytical solution from question 2:

$\mathrm{Erfc}\left[\frac{0}{\sqrt{2}}\right] = 1$ thus,

$1 - \frac{1}{2}\,\mathrm{Erfc}\left[\frac{0}{\sqrt{2}\,1}\right] = 1 - \frac{1}{2} = 0.5$

This is consistent with the result of the python simulation in which f(x)>1 50.074% of the time over 100 000 samples, and vice versa.

4. Marginal and conditional probabilities

In a country with high unemployment, 25.2% of the working population is considered "young", and of the young working population, 37.7% are unemployed. The unemployment rate for those in the working population who are not young, is lower at 21.5%.

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

1. P(young)

"*25.2% of the working population is considered "young""* ⟶ P(young) = 0.252

2. P(not young)

P(young) + P(not young) = 1 ⟶ P(not young) = 0.748

3. P(unemployed)

P(unemployed) = P(Unemployed|young)P(young) + P(unemployed|not young)P(not young)
= 0.377*0.252 + 0.748*0.215
⟶ P(unemployed) = 0.256

4. P(not unemployed)

P(unemployed)+ P(not unemployed)= 1 ⟶ P(not young) = 1- 0.256 = 0.744

5. P(unemployed | young)

"of the young working population, 37.7% are unemployed" ⟶ P(unemployed) = 0.377

6. P(young | unemployed)

$$\text{P(young|unemployed)} = \frac{P(\text{unemployed | young})\, P(\text{young})}{P(\text{unemployed})}$$

$$= \frac{0.377 * 0.252}{0.255824}$$

⟶ P(young | unemployed) = 0.371

7. P(unemployed | not young)

"*The unemployment rate for those in the
working population who are not young, is lower at 21.5%.*"
⟶ P(unemployed | young) = 0.215

8. P(not young | unemployed)

From P(A|B)+P(notA|B)=1, we have P(not young | unemployed) = 1 - P(young | unemployed)
⟶ P(not young | unemployed) = 0.629

5. Inference

You take a guess and think there is a 33% 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, 65% had training (usually by their parents) in pronouncing simple words and writing some letters, and 35% 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?

R is a random variable indicating Olivia's ability to read at Grade 1 level by age 6: 1 if true, 0 otherwise.
T is a random variable indicating whether whether Olivia has had training (usually by their parents) in pronouncing simple words
and writing some letters: 1 if true, 0 otherwise.

From "*You take a guess and think there is a 33% chance that Olivia, 5 year old girl, will be able to read at*
Grade 1 level by the time she turns 6.", we have P(R) = 0.33.
As P(A)+P(notA)=1, we infer P(notR) = 1-0.33 = 0.67.

From "*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, 65% had training (usually by their parents) in pronouncing simple words and writing some letters, and 35% did not.", we have P(T|R)=0.65 and P(notT|R) = 0.35.

From "*of children who are not already able to read at Grade 1 level by age 6,*
only 10% had the same training", we have P(T|notR)=0.1.
As P(A|B)+P(notA|B)=1, we infer P(notT|notR) = 1-0.1 = 0.9

Armed with this information, we revise our probability of R given T:

In[10]:=

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

$$= \frac{P(T|R)P(R)}{P(T|R)P(R)+P(T|\text{notR})P(\text{notR})} =$$

$$= \frac{0.65*0.33}{0.65*0.33+0.1*0.67} = 0.76$$

$RecursionLimit: Recursion depth of 1024 exceeded during evaluation of P[R].

Out[10]= $\text{Hold}\left[\frac{P[R] \times P[T \mid R]}{P[T]}\right]$

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

(Not sure what you want me to get at here, above and beyond the above, so I'll keep it high level:)
This model is an application of Bayes' Theorem to calculate the conditional probability P(R|T), where R and T are random indicator variables as defined above. It gives us a formal method for updating our
prior belief in whether Olivia will be able to read at Grade 1 level by age 6, P(R), in the face of additional information:
a statement by an educational expert on the role of parental training in achieving grade 1 level at age 6,
and the assumption that Olivia had indeed received such training by her parents.

6. Calculating probabilities

1. How many students are in your CS146 class (including yourself)?

18 students by my count, N = {1, 2...n=18}.

2. Assuming you know nothing about your classmates —

a. What is the probability that you were born after all of them?

I was born January 29. That is the 29th day out of the 365 days in a year.
Assuming that (a) that each of my class mates is equally likely to have been born on any of the 365 days in the year (discrete uniform distribution) and that (b) each birth date is independent. Thus, the probability that student j was born on day i is given by $P(X_{ij}) = \frac{1}{365}$. Consequently, the probability that any given student was born before me is p = $\frac{1}{28}$. We can thus describe this problem as a Binomial distribution with n=17 trials and $p = \frac{1}{18}$, the probability of p(X=17) | x~Binomial(17, $\frac{1}{28}$).

In[22]:=
```
Probability[x==17, x≈BinomialDistribution[17, 1/28]]
```

Out[22]= $\frac{1}{3\,996\,561\,798\,506\,898\,509\,529\,088}$

b. What is the probability that you were born before all of them?

Using the approach to a), we ask the P(x=0) | x~Binomial(17, $\frac{1}{28}$).

In[21]:=
```
Probability[x==0, x≈BinomialDistribution[17, 1/28]] // N
```

Out[21]= 0.538887

c. What is the probability that you were born after at least half of them?

And again, we ask $P(x = 9)$ | x~Binomial(17, $\frac{1}{28}$).

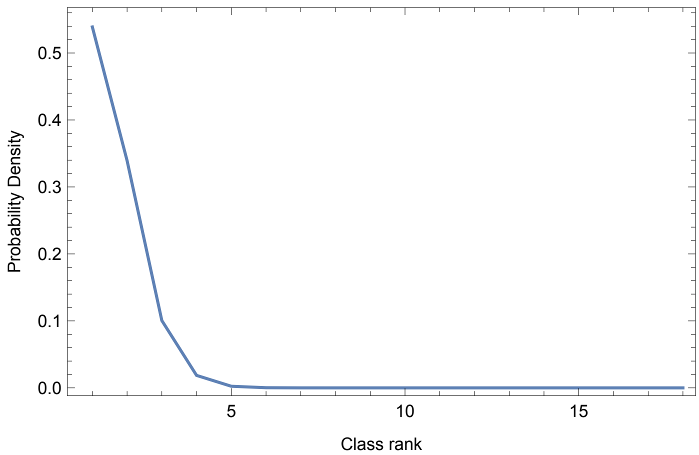
In[25]:=
```
Probability[x==9, x≈BinomialDistribution[17, 1/28]] //N
```

Out[25]= 1.71794×10^{-9}

More generally, we can plot the probability of any give rank in the class given my birthdate using the PMF of the binomial distribution,

```
ListLinePlot[Table[PDF[BinomialDistribution[17, 1/28], x], {x, 0, 17}], PlotRange→All, Fr
```

Out[30]=



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?*

There are two bags, A and B.
There are three types of cookies - chocolate c, vanilla v, and caramel l.

A = {c, v, v}
B = {c, l, l, l}

C is the random variable taking the value 1 if a random draw from one of the bags is chocolate.
We assume equal probability of either bag being selected, P(A)=P(B)=0.5.
If a cookies is not drawn from bag A, it must have been drawn from bag B. ⟶ notA=B ⟶ P(notA) = P(B)
1/3 cookies are chocolate in bag A ⟶ P(C|A) = 1/3
1/3 cookies are chocolate in bag B ⟶ P(C|notA) = 1/4
We then have,

$$P(A|C) = \frac{P(C|A)P(A)}{P(C|A)P(A)+P(C|\text{notA})P(\text{notA})}$$

In[12]:=
$$\frac{\frac{1}{3}*\frac{1}{2}}{\frac{1}{3}*\frac{1}{2}+\frac{1}{4}*\frac{1}{2}}$$

Out[12]=
$$\frac{4}{7}$$

$P(A|C) = \frac{4}{7}$
$\longrightarrow P(B|C) = 1-\frac{4}{7} = \frac{3}{7}$

Answer: There is a $\frac{4}{7}$ probability that I picked the bag containing the vanilla cookies, bag A.

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

World 1: If we are in the world where we drew the chocolate cookie from bag A, then
A_1 = {v,v}
B_1 = {c, l, l, l}

World 2: If we are in the world where we drew the chocolate cookie from bag B, then
A_2 = {c, v, v}
B_2 = {l, l, l}

We are in world 1 with probability $\frac{4}{7}$(P(A|C)) and World 2 with probability $\frac{3}{7}$ (P(B|C)).
V is a random variable taking the value 1 if a vanilla cookie is drawn the round after a chocolate cookie was randomly sampled from one of the bags, 0 otherwise.
Therefore,

In[13]:=

$$P(V) = P(A|C) * (P(v\ |A_1\ )P(A_1) + P(v\ |\ B_1)P(B_1)) + P(B|C) * (P(v|A_2)P(A_2) + P(v\ |\ B_2)P(B_2))$$

$$= \frac{4}{7}\left(1 * \frac{1}{2} + 0 * \frac{1}{2}\right) + \frac{3}{7}\left(\frac{2}{3} * \frac{1}{2} + 0 * \frac{1}{2}\right) = \frac{3}{7}$$

$$P(V) = \frac{3}{7}$$

Similar procedure for chocolate and caramel:
C* is a random variable taking the value 1 if a vanilla cookie is drawn the round after a chocolate cookie was randomly sampled from one of the bags, 0 otherwise.

In[13]:=

$$P(C^*) = P(A|C) * (P(v\ |A_1\ )P(A_1) + P(v\ |\ B_1)P(B_1)) + P(B|C) * (P(v|A_2)P(A_2) + P(v\ |\ B_2)P(B_2))$$

$$= \frac{4}{7}\left(0 * \frac{1}{2} + \frac{1}{4} * \frac{1}{2}\right) + \frac{3}{7}\left(\frac{1}{3} * \frac{1}{2} + 0 * \frac{1}{2}\right) = \frac{1}{7}$$

$$P(C^*) = \frac{1}{7}$$

L is a random variable taking the value 1 if a vanilla cookie is drawn the round after a chocolate cookie was randomly sampled from one of the bags, 0 otherwise.

In[13]:=

$$P(L) = P(A|C) * (P(v\ |A_1\ )P(A_1) + P(v\ |\ B_1)P(B_1)) + P(B|C) * (P(v|A_2)P(A_2) + P(v\ |\ B_2)P(B_2))$$

$$= \frac{4}{7}\left(0 * \frac{1}{2} + \frac{3}{4} * \frac{1}{2}\right) + \frac{3}{7}\left(0 * \frac{1}{2} + 1 * \frac{1}{2}\right) = \frac{3}{7}$$

$$P(L) = \frac{3}{7}$$

The probability that the cookie is chocolate, vanilla, or caramel is $\frac{1}{7}$, $\frac{3}{7}$, *and* $\frac{3}{7}$, respectively.