Machine Learning Assignment 49

Elijah Tarr

February 24, 2021

Problem 1

(a) *Given that $X \sim p(x)$, where $p(x)$ is a continuous distribution, prove that for any real number a we have $E[aX] = aE[X]$.*

Since a is a constant, we can literally just take it out of the integral.

$$\begin{aligned} E[aX] &= \int_0^1 (ax) * P(x)\mathrm{d}x \\ &= a\int_0^1 x * P(x)\mathrm{d}x \\ &= aE[X] \end{aligned}$$

(b) *Given that $X_1, X_2 \sim p(x)$, where $p(x)$ is a continuous distribution, prove that $E[X_1 + X_2] = E[X_1] + E[X_2]$.*

$$\begin{aligned} E[X_1 + X_2] &= \int_0^1 (x_1 + x_2) * p(x)\mathrm{d}x \\ &= \int_0^1 x_1 * p(x) + x_2 * p(x)\mathrm{d}x \\ &= \int_0^1 x_1 * p(x)\mathrm{d}x + \int_0^1 x_2 * p(x)\mathrm{d}x \\ &= E[X_1] + E[X_2] \end{aligned}$$

(c) *Given that $X \sim p(x)$ where $p(x)$ is a continuous probability distribution, prove the identity $Var[X] = E[X^2] - E[X]^2$.*

$$
\begin{aligned}
Var[X] &= E[(X - \overline{X})^2]\mathrm{d}x \\
&= \int_0^1 (x - E[X])^2 * p(x)\mathrm{d}x \\
&= \int_0^1 (x^2 - 2xE[X] + E[X]^2) * p(x)\mathrm{d}x \\
&= \int_0^1 x^2 * p(x)\mathrm{d}x + E[X]\int_0^1 (E[X] - 2x) * p(x)\mathrm{d}x \\
&= E[X^2] + E[X](E[X]\int_0^1 p(x)\mathrm{d}x - 2\int_0^1 x * p(x)\mathrm{d}x \\
&= E[X^2] + E[X](E[X] - 2E[X]) \\
&= E[X^2] - E[X]^2
\end{aligned}
$$

(d) *Use bisection search to estimate $\sqrt{5}$ to 4 decimal places by hand, showing your work at each step of the way. See problem 5-2 for a refresher on bisection search.*

Start out with interval $[2, 3]$ since $2^2 = 4$ and $3^2 = 9$

The function with $\sqrt{5}$ as a root can just be $f(x) = x^2 - 5$

Middle is 2.5. f(2.5) = 1.25. It is less than 0 so the guess is too low. We must lower the bounds

Middle is 2.25. f(2.25) = 0.0625. It is less than 0 so the guess is too low. We must lower the bounds

Middle is 2.125. f(2.125) = -0.484375. It is greater than 0 so the guess is too high. We must raise the bounds

Middle is 2.1875. f(2.1875) = -0.21484375. It is greater than 0 so the guess is too high. We must raise the bounds

Middle is 2.21875. f(2.21875) = -0.0771484375. It is greater than 0 so the guess is too high. We must raise the bounds

Middle is 2.234375. f(2.234375) = -0.007568359375. It is greater than 0 so the guess is too high. We must raise the bounds

Middle is 2.2421875. f(2.2421875) = 0.02740478515625. It is less than 0 so the guess is too low. We must lower the bounds

Middle is 2.23828125. f(2.23828125) = 0.0099029541015625. It is less than 0 so the guess is too low. We must lower the bounds

Middle is 2.236328125. f(2.236328125) = 0.001163482666015625. It is less than 0 so the guess is too low. We must lower the bounds

Middle is 2.2353515625. f(2.2353515625) = -0.0032033920288085938. It is greater than 0 so the guess is too high. We must raise the bounds

Middle is 2.23583984375. f(2.23583984375) = -0.001020193099975586. It is greater than 0 so the guess is too high. We must raise the bounds

Middle is 2.236083984375. f(2.236083984375) = 7.158517837524414e-05. It is less than 0 so the guess is too low. We must lower the bounds

Middle is 2.2359619140625. f(2.2359619140625) = -0.00047431886196136475. It is greater than 0 so the guess is too high. We must raise the bounds

Middle is 2.23602294921875. f(2.23602294921875) = -0.00020137056708335876. It is greater than 0 so the guess is too high. We must raise the bounds

Middle is 2.236053466796875. f(2.236053466796875) = -6.489362567663193e-05. It is greater than 0 so the guess is too high. We must raise the bounds

Now, f(x) rounded to 4 decimal places is 0. Therefore, the most accurate 4 decimals we have is:

$$\sqrt{5} \approx 2.2361$$

(e) *Use "merge sort" to sort the list* $[4, 8, 7, 7, 4, 2, 3, 1]$. *Do the problem by hand and show your work at each step of the way. See problem 23-3 for a refresher on merge sort.*

input list:[4,8,7,7,4,2,3,1] break it in half: [4,8,7,7] 4,2,3,1] sort the two halves

input list: [4,8,7,7] break it in half: [4,8] [7,7] sort the two halves

input list: [4,8] break it in half: [4] [8] the two halves have only one element each, so they are already sorted so we can combine them to get [4,8]

input list: [7,7] break it in half: [7] [7] the two halves have only one element each, so they are already sorted so we can combine them to get [7,7]

now we have two sorted lists [4,8] and [7,7] so we can combine them to get [4,7,7,8]

input list: [4,2,3,1] break it in half: [4,2] [3,1] sort the two halves

input list: [4,2] break it in half: [4] [2] the two halves have only one element each, so they are already sorted so we can combine them to get [2, 4]

input list: [3,1] break it in half: [3] [1] the two halves have only one element each, so they are already sorted so we can combine them to get [3, 1]

now we have two sorted lists [2, 4] and [3, 1] so we can combine them to get [1, 2, 3, 4]

now we have two sorted lists [4,7,7,8] and [1, 2, 3, 4] so we can combine them to get [1, 2, 3, 4, 4, 7, 7, 8]

1 Problem 2

(a) *Other people make* 1000 *successful trips through the wormhole with no failures. What is the likelihood function for* k *given these* 1000 *successful trips?*

data = [1000 successful]

$L(k|[1000 successful]) = k^{1000}$

(b) *What is the posterior distribution for* k *given these* 1000 *successful trips? (This is the same as just normalizing the likelihood function).*

$$
\begin{aligned}
1 &= \int_0^1 c * k^{1000} \mathrm{d}k \\
&= c \left. \frac{k^{1001}}{1001} \right|_0^1 \\
c &= 1001 \\
P(k|1000\,\text{successes}) &= 1001k^{1000} \\
P(0.99|100\text{successes}) &=\sim 0.043
\end{aligned}
$$

(c) *Assuming that you will use the wormhole* 500 *times per year, what is the posterior probability that the risk of disappearing forever into the wormhole is no more than your risk of dying from a car crash in a given year (1 in 10000)? In other words, what is* $P(1k^{500} \leq 110000|1000 \textit{successes})$

$$
\begin{aligned}
1 - k^{500} &\leq \frac{1}{10000} \\
k^{500} &\geq \frac{9999}{10000} \\
k &\geq \sqrt[500]{\frac{9999}{10000}} \\
P(k \geq \sqrt[500]{\frac{9999}{10000}}) &= 1 - \int_0^k 1001k^{1000} \mathrm{d}k \\
&= 1 - k^{1001}\Big|_0^{\sqrt[500]{\frac{9999}{10000}}} \\
&=\sim 0.000200
\end{aligned}
$$

1.1 Part 2

(a) *Looking at the given posterior distribution, how many successes have you counted?*

5000, because k is raised to that power, implying k had to have happened 5000 times.

(b) *Suppose you observe* 2000 *more successes. What is the posterior distribution now?*

Using similar reasoning to part 1-b, we can see that the new posterior distribution will be:

$$P(k|7000 \text{ successes}) = 7001k^{7000}$$

$$P(0.999|7000 \text{ successes}) =\sim 6.362$$

1.2 Part 3

(a) *Suppose you observe* 2000 *more successes. Fill in the blanks:*
likelihood = k^{2000}

$$p(2000 successes|k = 0.999) =\sim 0.1351$$

prior * likelihood =

$$5001k^{2000+5000}$$

after re-normalizing the prior * likelihood function, we get:
posterior distribution = $7001k^{7000}$
(with the re-normalization constant being $\frac{7001}{5001}$)

The reason why this is the same is that the posterior distribution is proportional to k^{7000}, and so is the likelihood function for $P(7000 successes|k)$. Since both likelihood functions are equal, normalizing each will give the exact same results.

1.3 Part 4

(a) *Assuming that you will use the wormhole* 500 *times per year, how many more people do you need to observe successfully come out of the wormhole to be* 99% *sure the risk of disappearing forever into the wormhole is no more than your risk of dying from a car crash in a given year (*1 *in* 10000*)?*

$$\mathcal{L} = k^{5000+n}$$

Through similar reasoning in Part 3a, we have:

$$P(5000 + N) = (5001 + n)k^{5000+n}$$

And to check if this function is right,

$$\begin{aligned}
a &= \sqrt[500]{\frac{9999}{10000}} \\
0.99 &= P(k \geq a)|5000 + N successes) \\
&= 1 - \int_0^a (5001 + n)k^{5000+n} \mathrm{d}k \\
&= 1 - k^{5001+n}\Big|_0^a \\
0.01 &= (\sqrt[500]{\frac{9999}{10000}})^{5001+n} \\
n &= \frac{\ln 0.01}{\ln \sqrt[500]{\frac{9999}{10000}}} - 5001 \\
&=\sim 23019699
\end{aligned}$$