Introduction to Deep Learning M2177.0043
Seoul National University

Homework #(**1**)
Seungyong Moon

INSTRUCTIONS

- Anything that is received after the deadline will be considered to be late and we do not receive late homeworks. We do however ignore your lowest homework grade.
- Answers to every theory questions need to be submitted electronically on ETL. Only PDF generated from LaTex is accepted.
- Make sure you prepare the answers to each question separately. This helps us dispatch the problems to different graders.
- Collaboration on solving the homework is allowed. Discussions are encouraged but you should think about the problems on your own.
- If you do collaborate with someone or use a book or website, you are expected to write up your solution independently. That is, close the book and all of your notes before starting to write up your solution.

1 Learning LaTex

1. Insert a 2 by 2 table in LaTex and fill in (first name, last name, major, student id) in row major format.

sol)

Seungyong	Moon
College of Liberal Studies	2011-13285

Table 1: Seungyong's profile

2. Insert a jpeg or png image of your profile photo in Latex and center align the image.

sol)

Figure 1: Seungyong's profile photo

2 Russian roulette

Let's take this opportunity to brush up on basic probability and start by analyzing various survival conditions on a six chamber revolver. For the following four games, write down your probabilistic reasoning.

Introduction to Deep Learning M2177.0043
Seoul National University

Homework #(**1**)
Seungyong Moon

---

1. Game 1. One bullet is put into the chamber and the barrel is spun randomly. If two players take turns to pull the trigger, who has the higher probability of survival? Should you go first or second?

sol) There is no difference between two.
Without loss of generality, we can assume that the bullet is put into 1st chamber and first player starts the game at ith(random) chamber, since the chambers are circular and the barrel is spun randomly. Then, if I go first, the probability of death is 1/2 (i = 1, 3, 5). Also, if I go second, the probability of death is 1/2 (i = 2, 4, 6). Therefore, the probability of death is the same whether I go first or second.

2. Game 2. Now if we spin the barrel after each trigger pull, how does the analysis change? Should you go first or second?

sol) It is better to go second.
Since we spin the barrel after each trigger pull, all trials(pulling the trigger) are independent. That is, at each trigger, the probability of firing is 1/6.
Let's calculate the probability that first player(P1) wins. Let p be the probability that P1 wins. At the first attempt, P1 dies with the probability of 1/6, and the game continues with the probability of 5/6. Since each trigger is independent, the probability that second player(P2) wins conditioned on P1 survives at the first attempt, is the same as the probability that P1 wins.
Also, the probability of P1 dies is equal to the sum of

1. The probability that P1 dies at the first attempt
2. The probability that P1 survives at the first attempt multiplied by The probability that P2 wins conditioned on P1 survives at the first attempt (since P1 dies if P2 wins)

Therefore, we can make an equation as follows.

$$1 - p = \frac{1}{6} + \frac{5}{6} \times p \tag{1}$$

Then $p = \frac{5}{11}$ and P1 is more likely to die.

3. Game 3. Now we'll put two bullets as opposed to one bullet randomly into the chamber. Your opponent played first and survived. You have the option to spin the barrel. Should you spin the barrel? Why or why not?

sol) It is better to spin the barrel.
Suppose that I will not spin the barrel. Since the first player survived, 2 out of 5 chambers left are filled with bullet. So the probability that I die is 2/5
Suppose that I will spin the barrel. Then the probability that I die is definitely 1/3, since 2 chambers out of 6 are filled with bullets. Therefore, It is better to spin the barrel.

4. Game 4. What if the two bullets are randomly put in two consecutive positions? If your opponent survived his first round, should you spin the barrel? Why or why not?

sol) It is better not to spin the barrel.
As in problem 2.1, we can assume that the bullets are put at 1st chamber and 2nd chamber, first player starts the game at ith(random) chamber. Since the first player survived, i must be 3, 4, 5, 6 with the same probability 1/4.
Suppose that I will not spin the barrel. Then if i is 3, 4 or 5, then I will survive and if i is 6, then I will die. Therefore, the probability that I die is 3/4.
Suppose that I will spin the barrel. Then the probability that I die is definitely 1/3, since 2 chambers out of 6 are filled with bullets. Therefore, It is better not to spin the barrel.

Introduction to Deep Learning M2177.0043
Seoul National University

Homework #(**1**)
Seungyong Moon

---

3 Inverse transform sampling

1. A random variable X has a continuous pdf $f_X(x)$ and cdf $F_X(x)$. Prove that the random variable $Y = F_X(X)$ is uniformly distributed in [0, 1].

sol) It is enough to show that cdf $F_Y(k)$ of Y is equal to k in [0, 1].
Suppose that $F_X(X)$ has an inverse. Then, there exists the unique X_0 satisfying $F_X(X_0) = k$. Also, because $F_X(x)$ is a monotone increasing, $F_X(X) \leq F_X(X_0)$ implies $X \leq X_0$ and vice versa. Therefore

$$\begin{aligned} F_Y(k) &= P[Y|Y \leq k] \\ &= P[X|F_X(X) \leq k] \\ &= P[X|X \leq X_0] \\ &= F_X(X_0) = k \end{aligned} \tag{2}$$

Now, assume that the inverse image of k under $F_X(X)$ is not a point. Knowing that the inverse image of an interval under a monotone function is also a interval, the inverse image of the point k, which can be expressed as a intersection of nested closed interval, is a closed interval, denoted $[a, b]$. Then the followings are satisfied.

$$\begin{aligned} F_Y(k) &= P[Y|Y \leq k] \\ &= P[X|F_X(X) \leq k] \\ &= P[X|X \leq b] \\ &= F_X(b) = k \end{aligned} \tag{3}$$

It is definitely true that $F_Y(k) = 0$ if $k < 0$ and $F_Y(k) = 1$ if $k > 1$, since $Y = F_X(X)$ is a random variable between 0 and 1. Therefore Y is uniformly distributed in [0, 1].

2. A random variable X has a continuous pdf $f_X(x)$ and cdf $F_X(x)$. Prove (without using the results above) that if the random variable U is uniformly distributed in [0,1], then $F_X^{-1}(U)$ has $F_X(x)$ as its CDF.

sol) First, suppose that $F_X^{-1}(U)$ exists. Then,

$$\begin{aligned} P[U|F_X^{-1}(U) \leq k] &= P[U|U \leq F_X(k)] \quad (\because f_X(x) \text{ is monotone increasing}) \\ &= F_X(k) \quad (\because U \text{ is uniformly distributed in } [0, 1]) \end{aligned} \tag{4}$$

Therefore, $F_X^{-1}(U)$ has $F_X(x)$ as its CDF.
Now, assume that $F_X(x)$ is not one-to-one. Suppose that there exists $u \in [0, 1]$ such that $F_X^{-1}(u)$ is a interval containing k. For all positive integer n, the following inequality satisfies, since the inverse image of $\{U|U \leq u - \frac{1}{n}\}$ does not contain k, and the inverse image of $\{U|U \leq u + \frac{1}{n}\}$ always contain k.

$$P[U|U \leq u - \frac{1}{n}] \leq P[U|F_X^{-1}(U) \leq k]] \leq P[U|U \leq u + \frac{1}{n}] \tag{5}$$

Then, as n goes to infinity, we can get the following result.

$$\begin{aligned} P[U|F_X^{-1}(U) \leq k]] &= P[U|U \leq u] \\ &= P[U|U \leq F_X(k)] \\ &= F_X(k) \quad (\because U \text{ is uniformly distributed in } [0, 1]) \end{aligned} \tag{6}$$

3. Use the results above to simulate the draw of 1 million samples from exponential distribution $f_X(x) = \lambda e^{-\lambda x}$ with λ = 1.0. Draw a figure overlaying following two plots: (1) analytical CDF of exponential distribution. (2) normalized histogram of your samples (use 500 bins). X-axis should be the value of X and the Y-axis should be the probability. Attach both the code in NumPy and the figure.

sol) First, get the analytic CDF of exponential distribution (denote $F(x)$). And then, pick a random number from [0, 1] and take inverse function $F^{-1}(x)$ on the number. By 3.2, the output has $F(x)$ as its CDF, and has the same CDF with exponential distribution. Therefore, we can conclude that the output has the same distribution with exponential distribution.

```
import numpy as np
import matplotlib.pyplot as plt

coef = 1.0 # lambda
numofsample = 1000000 # the number of samples

def exponential_cdf(x): # cdf of exponential distribution
    return 1-np.exp(-1*coef*x)

def exponential_cdf_inverse(x): # the inverse of the cdf
    return -1/coef*np.log(1-x)

# pick points from [0, 1] uniformly
sample = np.random.uniform(0, 1, numofsample)

x = np.linspace(0, 15, 501);

# plot the analytic cdf of exponential distribution
plt.plot(x, exponential_cdf(x))

# plot a normalized histogram of the samples
plt.hist(exponential_cdf_inverse(sample), bins = 500, density = True)

plt.xlabel("value of X")
plt.ylabel("probability")

plt.show() # draw the plots
```

4 Optimal hedge ratio

You own some shares of stock *B* and you just bought one share of stock *A*. How many shares of *B* do you need to sell to minimize the variance of total stocks you own? Assume you know the variance of each stocks σ_A^2, σ_B^2 and their correlation coefficient ρ.

sol) We know that $A \sim (\mu_A, \sigma_A^2)$ and $B \sim (\mu_B, \sigma_B^2)$. Suppose I have *x* amounts of stock *B*. Then the variance of my stocks is as follows.

$$\begin{aligned} Var(A + xB) &= Var(A) + x^2 Var(B) + 2xCov(A, B) \\ &= \sigma_B^2 x^2 + 2\rho\sigma_A\sigma_B x + \sigma_A^2 \end{aligned} \tag{7}$$

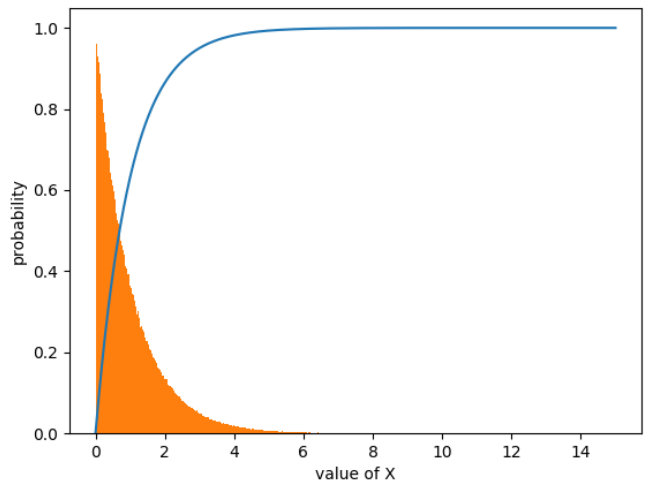

Figure 2: CDF (blue) and appoximate PDF (orange) of exponential distribution

Since the variance function is concave upward, We can get the optimal x that minimizes the variance by differentiating in x,

$$\begin{gathered}\frac{d}{dx}Var(A+xB) = 2\sigma_B^2 x + 2\rho\sigma_A\sigma_B = 0 \\ x^* = -\rho\frac{\sigma_A}{\sigma_B}\end{gathered} \tag{8}$$

There are 3 cases.

1. $\rho > 0$: Since the optimal x^* is negative and the variance function is quadratic, the variance function is increasing in $x \in [0, \infty)$. So, if I sell all the stocks of B, then the variance will be minimized.
2. $\rho = 0$: Since the optimal is zero, I need to sell all the stocks of B.
3. $\rho < 0$
   i) $-\rho\frac{\sigma_A}{\sigma_B} \geq x$: The variance function is decreasing in $x \in [0, -\rho\frac{\sigma_A}{\sigma_B}]$. So if I sell some stocks of B, then the variance will increase. Therefore. I should keep all the stocks of B (or buy $-\rho\frac{\sigma_A}{\sigma_B} - x$ stocks of B).
   ii) $-\rho\frac{\sigma_A}{\sigma_B} < x$: The variance function is decreasing in $x \in [-\rho\frac{\sigma_A}{\sigma_B}, x]$. So I should sell $x + \rho\frac{\sigma_A}{\sigma_B}$ of stock B.

Also, there are some exceptional cases, which don't satisfy the previous equation.

1. $\sigma_B = 0$: The variance of total stocks is not affected by how many stocks of B I have. So, there is no way to minimize the variance.
2. $\sigma_A = 0$: The variance of total stocks is determined only by how many stocks of B I have. So, I have to sell all the stocks of B.

Introduction to Deep Learning M2177.0043
Seoul National University

Homework #(**1**)
Seungyong Moon

---

5 Chebyshev inequality with bounded random variable

1. In the lecture, we saw the Chebyshev inequality requires us to know the mean and variance of the random variable. In practice, we may not always have access to the variance estimate. In this case, we need to come up with a conservative estimate of the variance if possible so we can still use the Chebyshev inequality. If a random variable X is known to take values in a range [a,b], derive the most conservative upper bound on the variance of X.

sol) First, let's find the optimal value that minimize $f(t) = E[(X-t)^2]$. After differentiating twice, we can get the followings

$$\begin{aligned} f'(t) &= -2E[(X-t)] = -2E[X] + 2t \\ f''(t) &= 2 \end{aligned} \tag{9}$$

Since $F(t)$ is concave upward, $t = E[X]$ is the optimal value that minimizes $f(t)$ and its minimum value is the variance of X. Now, plug $\frac{a+b}{2}$ into $f(t)$ for t, then we can get an inequality

$$\begin{aligned} Var(X) = f(E[X]) &\leq f(\frac{a+b}{2}) \\ &= E[(X - \frac{a+b}{2})^2] \\ &= \frac{1}{4}E[(2X-(a+b))^2] \\ &= \frac{1}{4}E[((X-a)+(X-b))^2] \end{aligned} \tag{10}$$

We know that $X - a \geq 0$ and $X - b \leq 0$. Therefore the followings are satisfied.

$$\begin{aligned} \frac{1}{4}E[((X-a)+(X-b))^2] &\leq \frac{1}{4}E[((X-a)+(b-X))^2] \\ &= \frac{1}{4}E[(b-a)^2] \\ &= \frac{1}{4}(b-a)^2 \end{aligned} \tag{11}$$

Finally we can get the inequality.

$$Var(X) \leq \frac{1}{4}(b-a)^2 \tag{12}$$

2. When does the variance inequality hold with equality? Define the probability distribution of X when this happens and prove the variance inequality holds with equality.

sol) There are two gaps in the inequality.

i) $f(E[X]) \leq f(\frac{a+b}{2})$
ii) $\frac{1}{4}E[((X-a)+(X-b))^2] \leq \frac{1}{4}E[((X-a)+(b-X))^2]$

So, for the variance inequality to holds with equality, two conditions should be satisfied.

i) $E[X] = \frac{a+b}{2}$
ii) $|(X-a)+(X-b)| = |X-a| + |X-b| \rightarrow X = a$ or b ($\because$ triangular inequality)

There is the only one distribution that satisfies the conditions.

$$X = \begin{cases} a & \text{with probability } \frac{1}{2} \\ b & \text{with probability } \frac{1}{2} \end{cases} \tag{13}$$

Let's calculate the variance of the distribution. The mean is trivially $\frac{a+b}{2}$, and the variance is,

$$\frac{1}{2}\{(\frac{a+b}{2} - a)\}^2 + \frac{1}{2}\{(\frac{a+b}{2} - b)\}^2 = \frac{1}{4}(b-a)^2 \tag{14}$$