CMPE 343: Introduction to Probability and Statistics for Computer Engineers (Fall 2021)

Homework #1

Due November 29, 2021 by 11:59pm on Moodle

Note: Please type your answers and submit your homework as PDF. To get full points, you need to show your steps clearly. If you use a theorem, rule, definition, derivation, etc. that were not covered in the lectures, you need to cite your resources. If you fail to cite your references, if you plagiarize, if you give your answers to another person, if you copy someone else's answers, your grade will be -100.

1. A function $P : A \subset \Omega \to \mathbb{R}$ is called a **probability law** over the sample space Ω if it satisfies the following three probability axioms.

   - (Nonnegativity) $P(A) \geq 0$, for every **event** A.
   - (Countable additivity) If A and B are two disjoint events, then the probability of their union satisfies

   $$P(A \cup B) = P(A) + P(B).$$

   More generally, for a countable collection of disjoint events $A_1, A_2, ...$ we have

   $$P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i).$$

   - (Normalization) The probability of the entire sample space is 1, that is, $P(\Omega) = 1$.

   (a) (5 pts) Prove, using only the axioms of probability given, that $P(A) = 1 - P(A^c)$ for any event A and probability law P where A^c denotes the complement of A.

   (b) (5 pts) Let $E_1, E_2, .., E_n$ be disjoint sets such that $\bigcup_{i_1}^{n} E_i = \Omega$ and let P be a probability law over the sample space Ω. Show that, for any event A we have

   $$P(A) = \sum_{i=1}^{n} P(A \cap E_i).$$

   (c) (5 pts) Prove that for any two events A, B we have

   $$P(A \cap B) \geq P(A) + P(B) - 1.$$

2. (10 pts) Two fair dice are thrown. Let

$$X = \begin{cases} 1, & \text{if the sum of the numbers} \leq 5 \\ 0, & \text{otherwise} \end{cases}$$

$$Y = \begin{cases} 1, & \text{if the product of the numbers is odd} \\ 0, & \text{otherwise} \end{cases}$$

What is $\mathrm{Cov}\,(X, Y)$? Show your steps clearly.

3. (10 pts) Derive the mean of Poisson distribution.

4. In this problem, we will explore certain properties of probability distributions and introduce new important concepts.

(a) (5 pts) Recall Pascal's Identity for combinations: $\binom{N}{m} + \binom{N}{m-1} = \binom{N+1}{m}$

Use the identity to show the following

$$(1+x)^N = \sum_{m=0}^{N} \binom{N}{m} \cdot x^m$$

which is called the *binomial theorem*. *Hint*: You can use induction.

Finally, show that the binomial distribution with parameter p is normalized, that is

$$\sum_{m=0}^{N} \binom{N}{m} \cdot p^m \cdot (1-p)^{(N-m)} = 1$$

(b) (5 pts) Suppose you wish to transmit the value of a random variable to a receiver. In Information Theory, the average amount of information you will transmit in the process (in units of "nat") is obtained by taking the expectation of $\ln p(x)$ with respect to the distribution $p(x)$ of your random variable and is given by

$$H(x) = -\int_x p(x) \cdot \ln p(x) \cdot dx$$

This quantity is the *entropy* of your random variable. Calculate and compare the entropies of a uniform random variable $x \sim U(0,1)$ and a Gaussian random variable $z \sim \mathcal{N}(0,1)$.

(c) In many applications, e.g. in Machine Learning, we wish to approximate some probability distribution using function approximators we have available, for example deep neural networks. This creates the need for a way to measure the similarity or the *distance* between two distributions. One proposed such measure is the *relative entropy* or the Kullback-Leibler divergence. Given two probability distributions p and q the KL-divergence between them is given by

$$KL(p||q) = \int_{-\infty}^{\infty} p(x) \cdot \ln \frac{p(x)}{q(x)} \cdot dx$$

i. (2 pts) Show that the KL-divergence between equal distributions is zero.

ii. (2 pts) Show that the KL-divergence is not symmetric, that is $KL(p||q) \neq KL(q||p)$ in general. You can do this by providing an example.

iii. (16 pts) Calculate the KL divergence between $p(x) \sim \mathcal{N}\left(\mu_1, \sigma_1^2\right)$ and $q(x) \sim \mathcal{N}\left(\mu_2, \sigma_2^2\right)$ for $\mu_1 = 2, \mu_2 = 1.8, \sigma_1^2 = 1.5, \sigma_2^2 = 0.2$. First, derive a closed form solution depending on $\mu_1, \mu_2, \sigma_1, \sigma_2$. Then, calculate its value. (Only numerical answer without clearly showing your steps will not be graded.)

Remark: We call this measure a *divergence* since a proper *distance* function must be symmetric.

5. In this problem, we will explore some properties of random variables and in particular that of the Gaussian random variable.

(a) (7 pts) The convolution of two functions f and g is defined as

$$(f * g)(t) = \int_{-\infty}^{\infty} f(\tau)g(t-\tau)d\tau$$

One can calculate the probability density function of the random variable $Z = X + Y$ using convolution operation with X and Y independent and continuous random variables. In fact,

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(\tau)f_Y(z-\tau)d\tau$$

Using this fact, find the probability density function of $Z = X + Y$, where X and Y are independent standard Gaussian random variables. Find μ_Z, σ_Z. Which distribution does Z belong to? (Hint: use $\sqrt{\pi} = \int_{-\infty}^{\infty} e^{-x^2} dx$)

(b) (5 pts) Let X be a standard normal Gaussian random variable and Y be a discrete random variable taking values $\{-1, 1\}$ with equal probabilities. Is the random variable $Z = XY$ independent of Y? Give a formal argument(proof or counter example) justifying your answer.

(c) (8 pts) Let X be a non-negative random variable. Let k be a positive real number. Define the binary random variable $Y = 0$ for $X < k$ and $Y = k$ for $X \geq k$. Using the relation between X and Y, prove that $P(X \geq k) \leq \dfrac{E[X]}{k}$. (Hint: start with finding $E[Y]$).

6. In this problem, we will empirically observe some of the results we obtained above and also the convergence properties of certain distributions. You may use the python libraries Numpy and Matplotlib.

(a) (5 pts) In *3.a* you have found the distribution of Z = X+ Y. Let X and Y be Gaussian random variables with $\mu_X = -1$, $\mu_Y = 3$, and $\sigma_X^2 = 1$ and $\sigma_Y^2 = 4$. Sample 100000 pairs of X and Y and plot their sum Z = X+Y as a histogram. Is the shape of X and its apparent mean consistent with *what you have learned in the lectures*?

(b) (5 pts) Let X $B(n, p)$ be a binomially distributed random variable. One can use the normal distribution as an approximation to the binomial distribution when n is large and/or p is close to 0.5. In this case, $X \approx N(np, np(1 - p))$. Show how such approximation behaves by drawing 10000 samples from binomial distribution with $n = 5, 10, 20, 30, 40, 100$ and $p = 0.2, p = 0.33, 0.50$ and plotting the distributions of samples for each case as a histogram. Report for which values of n and p the distribution resembles that of a Gaussian?

(c) (5 pts) You were introduced to the concept of KL-divergence analytically. Now, you will estimate this divergence $KL(p||q)$. Where $p(x) = N(0, 1)$ and $q(x) = N(0, 4)$. Sample 1000 samples from a Gaussian with mean 0 and variance 1. Call them $x_1, x_2, ..., x_{1000}$. Estimate the KL divergence as

$$\frac{1}{1000} \sum_{i=1}^{1000} \ln \frac{p(x_i)}{q(x_i)}$$

where $p(x) = N(0, 1)$ and $q(x) = N(0, 4)$. Calculate the divergence analytically for $KL(p||q)$. Is the result consistent with your estimate?