HOMEWORK 1 : KALPAK SEAL (8241-7219)

1. (Probability distribution function)
   The uniform distribution for a continuous variable x is

$$p(x; a, b) = \frac{1}{b-a}, \quad a \leq x \leq b$$

   Verify that this distribution is normalized (integrats to one), and find expressions for its mean and variance.

Given, the uniform distribution for a continuous random variable is

$$p(x; a, b) = 1/b-a \; ; \quad a \leq x \leq b$$

Now, $\int_a^b p(x)\,dx = \int_a^b \frac{dx}{b-a} = \frac{1}{b-a}[x]_a^b$

$$= \frac{b-a}{b-a}$$

$$= \underline{\underline{1}} \quad (\text{Ans})$$

Now, $E[X] = \int_{-\infty}^{\infty} x f_X(x)\,dx$

$$= \int_a^b x \cdot \frac{1}{b-a}\,dx = \frac{1}{b-a}\left[\frac{x^2}{2}\right]_a^b$$

$$= \frac{\cancel{(b-a)}(b+a)}{2\cancel{(b-a)}} = \frac{a+b}{2} \quad \text{(Ans)}$$

Also,

$$E[X^2] = \int_a^b x^2 \cdot \frac{1}{b-a} \cdot dx = \frac{1}{(b-a)} \cdot \left(\frac{b^3}{3} - \frac{a^3}{3}\right)$$

We know,

$$Var(X) = E[X^2] - (E[X])^2$$

$$= \frac{1}{b-a}\left\{\frac{b^3}{3} - \frac{a^3}{3}\right\} - \left(\frac{a+b}{2}\right)^2$$

$$= \frac{b^2 + ab + a^2}{3} - \frac{a^2 + 2ab + b^2}{4}$$

$$= \frac{4b^2 + 4ab + 4a^2 - 3a^2 - 6ab - 3b^2}{12}$$

$$= \frac{a^2 - 2ab + b^2}{12} = \frac{(a-b)^2}{12} \quad \text{Ans}$$

2. (Probability)
Probabilities are sensitive to the form of the question that was used to generate the answer (Source: Minka.) My neighbor has two children. Assuming that the gender of a child is like a coin flip, it is most likely, a priori, that my neighbor has one boy and one girl, with probability 1/2. The other possibilities - two boys or two girls-have probabilities 1/4 and 1/4

(a) Suppose I ask him whether he has any boys, and he says yes. What is the probability that one child is a girl?

(b) Suppose instead that I happen to see one of his children run by, and it is a boy. What is the probability that the other child is a girl?

a) The sample space looks like $\{BB, BG, GB, GG\}$

Now, Since, one child is a girl,
we need to omit BB.

So, the sample space becomes $\{BG, GB, GG\}$

Now, The probability that one child is a girl
is therefore $\frac{2}{3}$ (Ans)

b) Now, if it is given that one of his children is a boy.

The probability that the other child is a girl
is $P(G|B) = \dfrac{P(G \cap B)}{P(B)} = \dfrac{\frac{1}{2} \cdot \frac{1}{2}}{\frac{1}{2}} = \frac{1}{2}$ (Ans)

3. (Random variables and variance)
Show that the variance of a sum is $\mathrm{var}[X+Y] = \mathrm{var}[X] + \mathrm{var}[Y] + 2\,\mathrm{cov}[X,Y]$ where $\mathrm{cov}[X,Y]$ is the covariance between X and Y

$$\mathrm{var}[X+Y] = E\left[\left(X+Y - E[X+Y]\right)^2\right]$$

$$= E\left[\left\{(X - E[X]) + (Y - E[Y])\right\}^2\right]$$

By linearity of Expectation

$$= E\left[(X-E[X])^2 + (Y-E[Y])^2 + 2(X-E[X])(Y-E[Y])\right]$$

$$= \mathrm{var}(X) + \mathrm{var}(Y) + 2\,\mathrm{cov}(X,Y).$$

4. (Variables and covariance)
Show that if two variables x and y are independent, then their covariance is zero.

$$\mathrm{cov}(X,Y) = E\left[(X-E[X]).(Y-E[Y])\right]$$

$$= E[XY] - E[XE[Y]] - E[E[X]Y] + E[E[X]E[Y]]$$

$$= E[XY] - E[X][Y]$$

Since, X and Y are independent, $E[XY] = E[X][Y]$

$$\therefore \mathrm{cov}(X,Y) = E[X]E[Y] - E[X][Y] = 0 \text{ (Ans)}$$

5. (Independent variables and variance)
Suppose that the two variables x and z are statistically independent. Show that the mean and variance of their sum satisfies

$$\mathbb{E}[x+z] = \mathbb{E}[x] + \mathbb{E}[z]$$
$$\mathrm{var}[x+z] = \mathrm{var}[x] + \mathrm{var}[z]$$

- Let. $g(x,z) = x+z$

So, $E[X+Z] = E[g(x,z)] = \sum_x \sum_z (x+z)\, P_{X,Z}(X,Z)$

$= \sum_x \sum_z x P_{X,Z}(x,z) + \sum_x \sum_z z P_{X,Z}(x,z)$

$= \sum_x x \sum_z P_{x,z}(x,z) + \sum_z z \sum_x P(x,z) = \sum_x x P_X(x) + \sum_z z P_z(z)$

$= E[X] + E[Z]$ (Ans)

- $\mathrm{Var}(X+Z) = E[(X+Z)^2] = E[X^2+2XZ+Z^2]$

$= E[X^2] + 2E[XZ] + E[Z^2]$

Now, since X and Z are statistically independent,

$E[X] = E[Z] = 0$

$E[XZ] = E[X] \cdot E[Z] = 0$

So, $\mathrm{Var}(X+Z) = E[X^2] + E[Z^2]$

$= \mathrm{var}(x) + \mathrm{var}(z)$

6. (Concept class/Objective function)
Alex goes to a casino and finds out about a new type of gambling game.

In the game, Alex can randomly choose an integer, let's say x and feed that number to a machine. The machine in turn runs a program which takes the given number as input and transforms that number using a fixed function $\mathcal{F} : \mathbb{Z} \rightarrow \mathbb{Z}$ (which takes an integer as argument and returns an integer) and then adds some integer noise ϵ to that and produces the final output. So given an integer x the overall output of the program is of the form: $y = \mathcal{F}(x) + \epsilon,$ where $x, y, \epsilon \in \mathbb{Z}$.
Now, Alex's job is to predict the output y. Based on how accurate his prediction is, he gets some reward. It is evident that Alex does not know the exact functional form of $\mathcal{F}$ and hence is planning to employ a machine learning model to improve his predictions in order to maximize his reward. Fortunately, Alex somehow has the some information about the functional form of $\mathcal{F}$ that the program uses to generate its output. The function $\mathcal{F}$ has the following property:

$$\mathcal{F}(\alpha x_1) + \alpha \mathcal{F}(x_2) = \mathcal{F}(\mathcal{F}(x_1 + x_2)) \forall x_1, x_2 \in \mathbb{Z} \tag{1}$$

where α is an integer constant unknown to Alex.
Naturally, Alex chooses the set of all functions that satisfy the equation 1 as the concept class $\mathcal{C}$ of his machine learning model. For training the model Alex plays against the machine N times. Let $x_1, x_2, ...x_N$ be the input values in those N trails and $y_1, y_2, ...y_N$ be the corresponding outputs from the machine. The model is trained with those $(x_i, y_i), i \in 1, 2, ..., N$ pairs. Alex uses mean squared loss as objective for training the model. So basically Alex's model would choose the function $\mathcal{G}$ for its prediction, which minimizes the mean squared error over the training data $((x_i, y_i), i \in 1, 2, ..., N)$ among all the functions in $\mathcal{C}$. Formally,

$$\mathcal{G} = \arg \underset{f \in \mathcal{C}}{\overset{\min}{\cancel{\max}}} \Sigma_{i=1}^{N} (y_i - f(x_i))^2$$

Given a random integer $\hat{x}$ as input what should be the prediction of the model after training is done ?
Hint: Plug $x_1 = 0, x_2 = n$ and $x_1 = 1, x_2 = n - 1$. Can you infer the functional form of $\mathcal{F}$ from this?

Given,

$$F(\alpha x_1) + \alpha F(x_2) = F(F(x_1 + x_2)); \ \forall x_1, x_2 \in Z$$

Put, $x_1 = 0$, $x_2 = n$

So, $$F(0) + \alpha (F(n)) = F(F(n)) \quad \text{--- i}$$

And, for $x_1 = 1$, $x_2 = n-1$

$$F(\alpha) + \alpha F(n-1) = F(F(n)) \quad \text{--- ii}$$

Now

Equating i) and ii)

$$F(0) + \alpha F(n) = F(\alpha) + \alpha F(n-1)$$

$$\Rightarrow F(n) = F(n-1) + \frac{1}{\alpha}\{F(\alpha) - F(0)\}$$

Now, Since 'n' is a natural number.

This $F(n)$ will contain multiple function viz

straight line, parabola, ..., convex function.

Now, if I choose the convex functions to help me guarantee a reach to the global minima.

- f is called **convex** if:
  $\forall x_1, x_2 \in X, \forall t \in [0,1]: \qquad f(tx_1 + (1-t)x_2) \leq tf(x_1)$

Thus I will choose functions that is of this form.

So,
To compute the prediction,
we need to compute the derivative w.r.t 't'.
and according solve and get the prediction for the corresponding input.

7. (Probability)
Let X be a set and let $A = (A_n)_{n=1}^{\infty}$ whose union is X where all $A_i, i \in \mathbb{N}$ are mutually disjoint. You can convince yourself that the set of all finite or countable unions of the members of A, including the empty set, ϕ, is σ-algebra (you don't need to prove this). A σ-algebra of this type, generated by taking all the finite or countable unions of a set of atoms

(A_i's in this case), is called atomic. Prove that the Borel σ-algebra $\mathcal{B}_R$ is not atomic.

Let, $\mathcal{B}_{\mathbb{R}}$ be a collection of all finite/countable unions of sets in $A = (A_n)_{n=1}^{\infty}$, where the A_i's are mutually disjoint, non-empty subsets of $\mathbb{R}$, unioning to $\mathbb{R}$.

Particularly, each set in A is a Borel set in itself in $\mathbb{R}$.

Now,

Let $\mathcal{U} = \{ \{p\} : 0 < p < 1 \}$.

Then each singleton in $\mathcal{U}$ is a Borel set, being closed w.r.t $\mathbb{R}$.

So, for each $p \in (0,1)$, $\{p\}$ is a union of a finite or countable subset of A_i.

- But since the A_i are non-empty by itself, $\{p\}$ cannot be a union of more than one A_i, else there would be more than one element of $\{p\}$.

Hence, for each $p \in (0,1)$, there can be a injective association of a set $A_i \in A$ with $\{p\} = A_i$.

But, $\mathcal{V}$ is uncountable and A is countable.

$\therefore$ A is not atomic //

(A_i's in this case), is called atomic. Prove that the Borel σ-algebra $\mathcal{B}_R$ is not atomic.

8. Your friend, Carlsen performs an experiment in which he repeatedly flips a biased coin until a head appears. He notes down the outcome of each trial of his experiment. Naturally the set of outcomes (call it Ω) contains elements like H, TH, TTH and so on (where H denotes the occurrence of a Head and T is for a Tail). Now Carlsen needs help to equip himself with the required mathematical framework to deal with probabilities.

   (a) Construct a σ-algebra ($\mathcal{F}$) on Ω so that Carlsen will be able to assign probabilities to (at the very least) all possible singleton sets $\{\omega\}$ where $\omega \in \Omega$, and prove that $\mathcal{F}$ is indeed a σ-algebra.

   (b) Is the set $\mathcal{F}$ countable? Justify your answer.

   (c) Now Define a probability measure $P : \mathcal{F} \to [0, 1]$. Justify why you think P is a probability measure.

   (d) Give a random variable X on the probability space $(\Omega, \mathcal{F}, P)$ as defined by you so far. Now use the same probability as defined in (c) induced by the random variable X to show that

$$\lim_{x \to +\infty} P(X \leq x) = 1$$

   Also calculate the value of E(X) for the above setup.

$\Omega = \{H, TH, TTH, \ldots\}$

a) Creating a σ-algebra on this sample space.

- thus both $\Omega \in F$ and $\phi \in F$
- Considering any singleton set like $\{TH\} \in F$ and thereby, its complement $\{TH\}^c \in F$.
- Also, $\bigcup_{i=1}^{\infty} \{P_i\} \in F$ where $P_i \cap P_{i+1} = \phi$.

So, F is our required σ-algebra.

b) Now our sample space 'Ω' is a countably finite.

The power set of Ω is 2^{Ω} which is uncountably infinite.

Since the set of all possible subsets of Ω lies in $\mathcal{F}$, therefore, $\mathcal{F}$ is not countable.

c) Consider $\{\Omega, \mathcal{F}, P\}$ to be a probability measure.

For any event $\omega \in \Omega$ of size 'i',

$$P(\omega) = (1-p)^{i-1} \cdot p.$$

Now,

$$\sum_{\omega \in \Omega} P(\omega) = \sum_{i=1}^{\infty} (1-p)^{i-1} \cdot p.$$

$$= p \sum_{i=0}^{\infty} (1-p)^{i} = p \cdot \frac{1}{[1-(1-p)]}$$

Since,

$P(\Omega)$ is 1,

$$= 1$$

$\{\Omega, \mathcal{F}, P\}$ is a probability measure.

d) The samples of our Ω are disjoint and $\sum_{i=1}^{\infty} P(i) = P(\Omega)$.

$P(X=i) = p \cdot (1-p)^{i-1}$. $\forall i = 1, 2, 3$. $E[X] = \sum_{i=1}^{\infty} i (1-p)^{i-1} \cdot p = 1/p$ (Ans).

9. Let $A_1, A_2, A_3, \ldots$ be a sequence of events. Prove that, $P(\lim_{n\to+\infty} inf\, A_n) \leq \lim_{n\to+\infty} inf\, P(A_n)$

Let's assume that $(\Omega, \mathcal{F}, P)$ is the probability space.

Now, $A_n \in \mathcal{F}$ for all $n \geq 1$.

We know. $\liminf_{n\to\infty} A_n = \bigcup_{n\geq 1} \bigcap_{k\geq n} A_k \in \mathcal{F}$

and $\bigcap_{k\geq n} A_k \subset \bigcap_{k\geq n+1} A_k$ for any $n \geq 1$.

By continuity of probability P,

$$\bigcap_{k\geq n} A_k \subset A_n$$

So, $$P\left(\liminf_{n\to\infty} A_n\right) = P\left(\bigcup_{n\geq 1} \bigcap_{k\geq n} A_k\right)$$

$$= \lim_{n\to\infty} P\left(\bigcap_{k\geq n} A_k\right)$$

$$= \liminf_{n\to\infty} P\left(\bigcap_{k\geq n} A_k\right)$$

$$\leq \liminf_{n\to\infty} P(A_n)$$