Lecture 2

Name: Harsh Sanjay Roniyar

Roll Number: 22B3942

Random Variable

Example that we are considering is tossing a biased coin, hence
$\mu = P(X = 1) = 1 - P(X = 0)$ and they are not equal to $\frac{1}{2}$. The random variable here is X.

Probability Mass Function

Cumulative Distribution Function

Some common PMFs

1. Bernoulli Distribution

$x \in \{0, 1\}$
$\mu = P(X = 1)$

Hence, $Bern(x|\mu) = \mu^x(1-\mu)^{1-x}$

2. Binomial Distribution

N coin tosses, with each toss with $P(X = 1) = \mu$

$Bin(y|\mu, N) = \binom{N}{y}\mu^y(1-\mu)^{N-y}$

The event space $y \in \{0, 1, \ldots, N\}$
Then,

$$\bar{y} = \mathbb{E}(y) = \sum_{y=0}^{N} Bin(y|\mu, N) \cdot y$$

$$\mathbb{E}(f(y)) = \sum_{y} p(y)f(y)$$

Entropy

The entropy of y, would be defined as -

$$\mathbb{E}[-\log(p(y))]$$

which is equal to -

$$= -\mu \log \mu - (1-\mu)\log(1-\mu)$$

Joint, Conditional and Marginal Probability

- Joint: $p(y, z)$ such that $\sum_y \sum_z p(y, z) = 1$
- Conditional: $p(y|z) = \frac{p(y,z)}{p(z)}$
- Marginal: $p(z) = \sum_y p(y, z)$ and $p(y) = \sum_z p(y, z)$

Continuous Random Variable

For continuous random variable, $P(X = x) = 0$.

A pdf (probability density function) $p_X(x)$ describes the random variable.

Probability Density Function

- $p(x) \geq 0$
- $\int_x p(x)dx = 1$
- $p_X(x) = \frac{dP_X(x)}{dx}$

Cumulative Distribution Function

The CDF is describes as -

$$P_X(x) = \int_{-\infty}^{x} p_X(x)dx$$

The value of the CDF always reaches 1 at ∞ and starts from 0 at $-\infty$

Some common PDFs

1. Uniform Distribution

$$\mathrm{U}(x|a, b) - \begin{cases} \frac{1}{b-a} & a \leq x \leq b \\ 0 & \text{otherwise} \end{cases}$$

2. Gaussian/Normal Distribution

$$\mathrm{N}(x|\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

3. Beta, etc.

Empirical Distribution

Empirical means *by experiment*.

Dirac-Delta function

$$\delta(x) = \lim_{\Delta x \to 0} \begin{cases} \frac{1}{\Delta x} & \text{if } x = 0 \\ 0 & \text{otherwise} \end{cases}$$

The empirical PDF is thus -

$$p_X(x) = \sum_{i=1}^{N} \frac{1}{N} \delta(x - x_i)$$

Integrating the PDF gives the CDF and similarly, differentiating CDF gives us the PDF.

Problems with Empirical Dist.

- Rote Learning
- No generalization

IID (independent and identically distributed)

- **Independent**: $p(x_i, x_j) = p(x_i)p(x_j)$
- **Identically**: $p_{X_i}(x_i = x) = p_{X_j}(x_j = x)$
  - $x_i \sim p_X$
  - $x_j \sim p_X$

Now, if we have $X = (x_1, x_2, \ldots, x_N)$
Then,

$$p(X) = p(x_1, x_2, \ldots, x_N)$$

$$= p_X(x_1)p_X(x_2)\ldots p_X(x_N)$$

$$= \prod_{i=1}^{N} p_X(x_i)$$

$$\log p(x) = \sum_i \log(p_X(x_i))$$

MLE (Maximum Likelihood Estimator)

For random variables, $x_1, x_2, \ldots x_N$
ML Estimate of $\text{Gaussian}(\mu, \sigma)$

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

Take log, then differentiate and equate to zero, to get μ_{ML} -

$$\mu_{ML} = \frac{1}{N} \sum_i x_i$$

For uniform distribution $U(x|a, b)$:

$$p(X) = \prod_i \begin{cases} \frac{1}{b-a} & \text{if } a \le x \le b \\ 0 & \text{otherwise} \end{cases}$$

On applying MLE, we get $a = \min x_i$ and $b = \max x_i$.