Lecture Note - 02

Dihui Lai

March 1, 2020

Contents

1   Discrete Random Variables

1.1   Probability Distribution/Mass

A discrete random variable X can take on k different values: x_1, x_2, ..., x_k, each value occurs with probablity p_1, p_2, ..., p_k. The probablity distribution function or probability mass can be denoted as $X : P(x_i) = p_i$

1.2   Momentum

Mean/First Momentum:

$$E(X) = \bar{x} = \sum_{i=1}^{k} x_i p_i$$

n-th Momentum:

$$E(X^n) = \sum_{i=1}^{k} x_i^n p_i$$

Variance

$$Var(X) = \sum_{i=1}^{k} (x_i - \bar{x})^2 p_i$$

Mean of a general function $f(x)$

$$E(f(X)) = \sum_{i=1}^{k} f(x_i)p_i$$

1.3 Bernouli Distribution

A random variable that follows Bernouli Distribution can have two possible values 1 or 0, with probability p and $1 - p$, respectively.

$$P(x) = \begin{cases} p, & x = 1 \\ 1 - p, & x = 0 \end{cases} \tag{1}$$

The mean of a Bernouli random variable can be calculated as

$$\bar{x} = 1 \cdot p + 0 \cdot (1 - p) = p$$

A sample Bernouli Distribution dataset of N samples is a vector composed of 0/1 e.g. $x = [1, 0, 0, 1, ...1]$ and has N elements. Let's assume there are N_1 1s and N_0 0s in the dataset. Numerically, the average of the dataset can be calculated as

$$\bar{x} = \frac{1 + 0 + 0 + 1...1}{N} = \frac{N_0}{N} = \frac{N_0}{N_0 + N_1}$$

Here, $\frac{N_0}{N_0+N_1}$ is also the occurence probability of 1 in the dataset i.e. $\frac{N_0}{N_0+N_1} \sim p$

1.4 Binomial Distribution

Assume we do 3 experiments and each experiment has an outcome of either 1 or 0. In another word, each experiment's outcome follows Bernouli Distribution. A sample 3 bernouli experiment dataset looks like the following

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 1 \\ \vdots \\ 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 1 \end{bmatrix}$$

The total number of experiments that have an outcome of 1 can be calclulted as the summation of 3 random variables. X can have 4 possible values 0, 1, 2, 3 out of 8 configurations

$$X = \begin{cases} x = 0, [0, 0, 0] \\ x = 1, [1, 0, 0], [0, 1, 0], [0, 0, 1] \\ x = 2, [1, 1, 0], [0, 1, 1], [1, 0, 1] \\ x = 3, [1, 1, 1] \end{cases}$$

The probability distribution of each value is therefore

$$X = \begin{cases} x = 0, p^3 \\ x = 1, 3p(1-p)^2 \\ x = 2, 3p^2(1-p) \\ x = 3, (1-p)^3 \end{cases}$$

In an n-expemriments case, the dataset looks like the following

$$\underbrace{\begin{bmatrix} 1 & 0 & 0 & \cdots & 1 & 0 & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 & 0 \\ 0 & 1 & 1 & \cdots & 0 & 0 & 0 \\ \vdots & & & & & & \\ 0 & 0 & 0 & \cdots & 1 & 0 & 0 \\ 1 & 0 & 0 & \cdots & 0 & 1 & 1 \end{bmatrix}}_{n}$$

The value of X could vary between 0 and n. The probability that $X = k$, i.e. k out n experiments have out come 1, can have $\frac{n!}{k!(n-k)!}$ configurations. The probability distribution is therefore

$$P(k) = \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}$$

The expectation of X can be calculated as

$$E(X) = \sum_{k=0}^{n} kp(k)$$

$$= \sum_{k=0}^{n} k \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}$$

$$= \sum_{k=1}^{n} k \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}, \text{ the k=0 term is eliminated, the summation starts at k=1}$$

$$= \sum_{k=1}^{n} \frac{n!}{(k-1)!(n-k)!} p^k (1-p)^{n-k}$$

$$= np \sum_{k=1}^{n} \frac{(n-1)!}{(k-1)!(n-k)!} p^{k-1} (1-p)^{n-k}$$

$$= np \sum_{k=0}^{n-1} \frac{(n-1)!}{(k-1)!(n-k-1)!} p^k (1-p)^{n-k-1}, \text{ substitute } k \leftarrow k+1 \text{ and reindex k from 0}$$

$$= np$$

2 Poisson Distribution

2.1 Probability Distribution

Consider a random event, the probability of 1 occurence within a unit time is p. What's the probability distribution of events occurence within time interval of τ (e.g. No. of car accidents

occurs in a day in MO). Assume the probablity of 1 event occurring is p in a unit time period. We then have the following probability distribution within a short interval dt, where $dt \to 0$

$$P = \begin{cases} 1 - pdt, \text{no event} \\ pdt, 1 \text{ event} \\ (pdt)^n \approx 0, \text{n events} \end{cases}$$

The event occured within a time period of dt follows a Bernouli distribution

$$\Delta = \begin{cases} 0, 1 - pdt \\ 1, pdt \end{cases}$$

For an period of τ, we can construct a random variable X that is a summation of N random variables $X = \Delta_1 + \Delta_2 + ...\Delta_N$. Each variable on the right hand side follows the Bernouli distribution described above. Becase $dt \to 0$ we need to have $N \to \infty$ to have $Ndt = \tau$. In another word, X is the summation of infinite number of Bernouli distribution or a binomial distribution when $N \to \infty$

$$X = \Delta_1 + \Delta_2 + ...\Delta_i + ...$$

The distribution of X can be derived by taking the limit of a Binomial distribution.

$$\begin{aligned} P(k) &= \lim_{N \to \infty} \frac{N!}{k!(N-k)!}(pdt)^k(1 - pdt)^{N-k} \\ &= \lim_{N \to \infty} \frac{N!}{k!(N-k)!}(\frac{p\tau}{N})^k(1 - \frac{p\tau}{N})^{N-k} \\ &= \lim_{N \to \infty} \frac{N!}{k!(N-k)!}(\frac{\lambda}{N})^k(1 - \frac{\lambda}{N})^{N-k} \text{ where } \lambda = p\tau \\ &= \lim_{N \to \infty} \frac{N(N-1)...(N-k+1)}{k!}(\frac{\lambda}{N})^k(1 - \frac{\lambda}{N})^{N-k} \\ &= \frac{\lambda^k}{k!}e^{-\lambda} \end{aligned}$$

According to Taylor expasion of a exponential function, the summation of $P(k)$ is

$$\sum_{k=0}^{\infty} P(k) = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{-\lambda}e^{\lambda} = 1$$

2.2   Moments

Mean: there are many ways to calculate the mean of Poisson distribution. Here we use a trick by taking the derivative of the equation above, $e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = 1$, we then have

$$\frac{d}{d\lambda}e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = 0$$

$$\to -e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} + e^{-\lambda} \sum_{k=0}^{\infty} \frac{k\lambda^{k-1}}{k!} = 0$$

The first of the equation above is nothing but 1, we therefore have

$$-1 + e^{-\lambda} \sum_{k=0}^{\infty} \frac{k\lambda^{k-1}}{k!} = 0$$

Organize the equation a little bit we have

$$\frac{1}{\lambda} e^{-\lambda} \sum_{k=0}^{\infty} \frac{k\lambda^k}{k!} = 1$$

By definition, $E(X) = e^{-\lambda} \sum_{k=0}^{\infty} \frac{k\lambda^k}{k!}$, we therefore have

$$\frac{1}{\lambda} E(X) = 1 \text{ or } E(X) = \lambda$$

Variance: To calculate the variance we use the same trick of derivative. Here, we take the derivative of the equation $E(k) = \sum_{k=0}^{\infty} \frac{k\lambda^k}{k!} = \lambda$

$$\frac{d}{d\lambda} E(k) = \frac{d}{d\lambda} \sum_{k=0}^{\infty} \frac{k\lambda^k}{k!} e^{-\lambda} = \frac{d}{d\lambda} \lambda$$

$$\rightarrow \sum_{k=0}^{\infty} \frac{k^2 \lambda^{k-1}}{k!} e^{-\lambda} - \sum_{k=0}^{\infty} \frac{k\lambda^k}{k!} e^{-\lambda} = 1$$

$$\rightarrow \frac{E(k^2)}{\lambda} = 1 + E(k)$$

$$\rightarrow E(k^2) = (1 + \lambda)\lambda$$

The variance of X by definition is

$$Var(X) = \sum_{k=0}^{\infty} (k - \lambda)^2 \frac{k\lambda^k}{k!} e^{-\lambda}$$

$$= E(k^2) - 2E(k)\lambda + \lambda^2$$

$$= (1 + \lambda)\lambda - 2\lambda^2 + \lambda^2$$

$$= \lambda$$

3 Algebra of Multiple Random Variables

Assume the value of a random variable X is the summation of n random varaibles, X_1, X_2, ... X_n. It can be denoted as

$$X = X_1 + X_2 + ... + X_n$$

The expectation of the random variable can be calculated as

$$E(X) = E(X_1 + X_2 + X_3... + X_n) = E(X_1) + E(X_2) + E(X_3)... + E(X_n)$$

The variance of the random variable is

$$Var(X) = Var(X_1 + X_2 + X_3... + X_n) = Var(X_1) + Var(X_2) + Var(X_3)... + Var(X_n) + \sum_{i \neq j} Cov(X_i, X_j)$$

When X_1, X_2, ... X_n are independent, we have $Cov(X_i, X_j) = 0$, therefore,

$$Var(X) = Var(X_1) + Var(X_2) + Var(X_3)... + Var(X_n)$$