

# Lecture Note - 02: Random Variable, Statistic Distribution

Dihui Lai

March 29, 2020

## Contents

<b>1</b>	<b>Discrete Random Variables</b>	<b>2</b>
1.1	Probability Distribution/Mass . . . . .	2
1.2	Momentum . . . . .	2
1.3	Bernoulli Distribution . . . . .	2
1.4	Binomial Distribution . . . . .	3
1.5	Multinomial Distribution . . . . .	4
1.6	Poisson Distribution . . . . .	4
<b>2</b>	<b>Continuous Random Variable</b>	<b>6</b>
2.1	Probability Density Function . . . . .	6
2.2	Moments . . . . .	7
2.3	Gaussian Distribution . . . . .	7
<b>3</b>	<b>Algebra of Multiple Random Variables</b>	<b>8</b>

# 1 Discrete Random Variables

## 1.1 Probability Distribution/Mass

A discrete random variable  $X$  can take on  $k$  different values:  $x_1, x_2, \dots, x_k$ , each value occurs with probability  $p_1, p_2, \dots, p_k$ . The probability 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 Bernoulli Distribution

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

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

The mean of a Bernoulli random variable can be calculated by definition using the distribution function

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

A sample Bernoulli Distribution dataset of  $N$  samples is a vector composed of 0/1 e.g.  $x = [1, 0, 0, 1, \dots, 1]$  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

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

Here,  $\frac{N_0}{N_0+N_1}$  is also the occurrence 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 Bernoulli Distribution. A sample 3 Bernoulli 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 calculated 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-experiments case, the dataset looks like the following

$$\underbrace{\begin{bmatrix} 1 & 0 & 0 & \dots & 1 & 0 & 0 \\ 0 & 0 & 1 & \dots & 0 & 0 & 0 \\ 0 & 1 & 1 & \dots & 0 & 0 & 0 \\ \vdots & & & & & & \\ 0 & 0 & 0 & \dots & 1 & 0 & 0 \\ 1 & 0 & 0 & \dots & 0 & 1 & 1 \end{bmatrix}}_n$$

The value of  $X$  could vary between 0 and  $n$ . The probability that  $X = k$ , i.e.  $k$  out of  $n$  experiments have outcome 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

$$\begin{aligned} E(X) &= \sum_{k=0}^n k p(k) \\ &= \sum_{k=0}^n k \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \\ &= \sum_{k=1}^n k \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}, \text{ the } k=0 \text{ term is eliminated, the summation starts at } k=1 \\ &= \sum_{k=1}^n \frac{n!}{(k-1)!(n-k)!} p^k (1-p)^{n-k} \\ &= np \sum_{k=1}^n \frac{(n-1)!}{(k-1)!(n-k)!} p^{k-1} (1-p)^{n-k} \\ &= np \sum_{k=0}^{n-1} \frac{(n-1)!}{(k-1)!(n-k-1)!} p^k (1-p)^{n-k-1}, \text{ substitute } k \leftarrow k+1 \text{ and reindex } k \text{ from } 0 \\ &= np \end{aligned}$$

## 1.5 Multinomial Distribution

More often than not, an experiment could have  $c > 1$  outcome. For example, throwing a dice might have 6 outcomes, the weather condition next day could be sunny, cloudy or rainy etc.

Of each experiment, each outcome has different probabilities  $p_1, p_2, p_3, \dots, p_c$ . The total probability has to satisfy the condition  $\sum_{j=1}^c p_j = 1$ . If we perform  $N$  experiments, the probability distribution function can be written as

$$f(x_1, x_2, x_3, \dots, x_c) = \frac{N!}{x_1! x_2! \dots x_c!} p_1^{x_1} p_2^{x_2} \dots p_c^{x_c}$$

Here,  $x_1$  is the number of times outcome is 1,  $x_2$  is the number of times that we get outcome 2 etc. The summation of the  $x$ s has to be  $N$ , i.e.  $\sum_{j=1}^c x_j = N$

## 1.6 Poisson Distribution

**Probability distribution:** consider a random event, the probability of 1 occurrence within a unit time is  $p$ . What is the probability distribution of events occurrence within a time interval of  $\tau$  (e.g.

No. of car accidents occurs in a day in MO). Assume the probability of 1 event occurring is  $p$  in a unit time period. We then have the following probability distribution within a short interval  $dt$ , where  $dt \rightarrow 0$

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

The event occurred within a time period of  $dt$  follows a Bernoulli distribution

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

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

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

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

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

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

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

**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

$$\begin{aligned} \frac{d}{d\lambda} e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} &= 0 \\ \rightarrow -e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} + e^{-\lambda} \sum_{k=0}^{\infty} \frac{k\lambda^{k-1}}{k!} &= 0 \end{aligned}$$

The first term 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 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  $E(k) = \sum_{k=0}^{\infty} \frac{k\lambda^k}{k!} = \lambda$

$$\begin{aligned} \frac{d}{d\lambda} E(k) &= \frac{d}{d\lambda} \sum_{k=0}^{\infty} \frac{k\lambda^k}{k!} e^{-\lambda} = \frac{d}{d\lambda} \lambda \\ &\rightarrow \sum_{k=0}^{\infty} \frac{k^2 \lambda^{k-1}}{k!} e^{-\lambda} - \sum_{k=0}^{\infty} \frac{k\lambda^k}{k!} e^{-\lambda} = 1 \\ &\rightarrow \frac{E(k^2)}{\lambda} = 1 + E(k) \\ &\rightarrow E(k^2) = (1 + \lambda)\lambda \end{aligned}$$

The variance of X by definition is

$$\begin{aligned} Var(X) &= \sum_{k=0}^{\infty} (k - \lambda)^2 \frac{k\lambda^k}{k!} e^{-\lambda} \\ &= E(k^2) - 2E(k)\lambda + \lambda^2 \\ &= (1 + \lambda)\lambda - 2\lambda^2 + \lambda^2 \\ &= \lambda \end{aligned}$$

## 2 Continous Random Variable

### 2.1 Probability Density Function

A continous variable X can take on real value. The probability that  $X$  is between  $x_i - \Delta x$  and  $x_i + \Delta x$  can be described by the following probability function

$$P(x_i - \Delta x < x < x_i + \Delta x) \sim f(x_i)\Delta x$$

where  $x_i$  can be any value between  $(-\infty, \infty)$ . More precisely, we can write the relationship in integral formula

$$P(x_i - \Delta x < x < x_i + \Delta x) = \int_{x_i - \Delta x}^{x_i + \Delta x} f(x) dx$$

Here  $f(x)$  is called probability density function and has to satisfy the following constraint

$$\int_{-\infty}^{\infty} f(x) dx = 1$$

## 2.2 Moments

**Mean:** the mean of a continuous random variable is defined as

$$\bar{x} = E(X) = \int_{-\infty}^{\infty} x f(x) dx$$

**Mean of arbitrary function:**

$$E(g(x)) = \int_{-\infty}^{\infty} g(x) f(x) dx$$

**Variance:** the variance of a continuous random variable is defined as

$$Var(X) = \int_{-\infty}^{\infty} (x - \bar{x})^2 f(x) dx$$

## 2.3 Gaussian Distribution

The probability density function of a standard Gaussian function is defined as

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

$$E(X) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} x e^{-\frac{x^2}{2}} dx = 0$$

$$\begin{aligned}
Var(X) &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} x^2 e^{-\frac{x^2}{2}} dx \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} x (e^{-\frac{x^2}{2}})' dx \text{ integrate by parts} \\
&= \frac{1}{\sqrt{2\pi}} x (e^{-\frac{x^2}{2}}) \Big|_{-\infty}^{\infty} + \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \\
&= 1
\end{aligned}$$

### 3 Algebra of Multiple Random Variables

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

$$X = X_1 + X_2 + \dots + X_n$$

The expectation of the random variable can be calculated as

$$E(X) = E(X_1 + X_2 + X_3 \dots + X_n) = E(X_1) + E(X_2) + E(X_3) \dots + E(X_n)$$

The variance of the random variable is

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

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

$$Var(X) = Var(X_1) + Var(X_2) + Var(X_3) \dots + Var(X_n)$$

Consider a binomial distribution that is constructed from  $n$  Bernoulli distribution. Since each Bernoulli distribution has mean  $p$  and variance  $p(1-p)$ . The mean of the binomial distribution is the summation of the mean of  $n$  Bernoulli distributions i.e.  $E(X) = np$ . Similarly, the variance follows as  $Var(X) = np(1-p)$ .