Course Material on Probability Theory

Introduction

Define : An event is called random if its outcome cannot be predicted from its initial state. For example tossing a coin whether its heads or tail or rolling of a a dice. Thus the outcome of an event can be assigned chances (probability) of happening.
Example : For tossing of coin there is a chance of either heads coming as outcome or tails $\therefore$ chance of head or probability of head :

$$\mathbf{P}(\text{head}) = \frac{\text{1 possible outcome (head)}}{\text{2 possible total outcomes (head+tails)}} \tag{1}$$

And for rolling a dice, chances of getting any given number, say 1 :

$$\mathbf{P}(1) = \frac{\text{1 possible outcome (number 1 appears)}}{\text{6 possible total outcomes (1,2,3,4,5,6)}} \tag{2}$$

Note that : $\mathbf{P}(\text{head})+\mathbf{P}(\text{tail})=1$ and $\mathbf{P}(1)+\mathbf{P}(2)+\mathbf{P}(3)+\mathbf{P}(4)+\mathbf{P}(5)+\mathbf{P}(6)=1$.
Generically : One defines the theoretical probability $\mathbf{P}$ of an event with outcome x_i as :

$$\mathbf{P}(\mathbf{x_i}) = \frac{\text{number of outcomes of the event } \mathbf{x_i}}{\text{total number of events}} \tag{3}$$

Let us look at the case of 20 students in a class and let the event be the year of birth of the student. Then given the tabulation of number of students born in each year :

$$\begin{aligned}
1998 &: 1 \ \mathbf{P(1998)} = \mathbf{\frac{1}{20}} \\
1999 &: 4 \ \mathbf{P(1999)} = \mathbf{\frac{4}{20}} \\
2000 &: 8 \ \mathbf{P(2000)} = \mathbf{\frac{8}{20}} \\
2001 &: 6 \ \mathbf{P(2001)} = \mathbf{\frac{6}{20}} \\
2002 &: 1 \ \mathbf{P(2002)} = \mathbf{\frac{1}{20}}
\end{aligned} \tag{4}$$

Mutually exclusive events are those, such that given one particular event did occur, the others could not have occurred.
Example: If head happens tails cannot happen or a student born in a particular year can not be born in any other year.

Properties

Now the following is always true :

- Probability always satisfies :

$$0 \leq P \leq 1 \tag{5}$$

  where $P = 0$ means impossible and $P = 1$ means certainty.

- $\sum_i P_i = 1$, which means probability of any event (not a particular event) happening $= 1$. Example : either head or tail or either getting result $1 - 6$ in a dice roll.

- Probability of mutually exclusive events add and probability of same vent multiply. Example : if we have 2 coin tosses then the probability of getting both heads is :

$$\begin{aligned} &= \frac{\text{1 event i.e. getting both heads : HH}}{\text{4 outcomes (HH, HT, TH, TT)}} \\ &= \text{chances of head in 1st toss} \times \text{chances of head in 2nd toss} \\ &= \frac{1}{2} \times \frac{1}{2} \end{aligned} \tag{6}$$

  while if we want to calculate the probability of 1 head in two coin tosses :

$$\begin{aligned} \mathbf{P}(\mathbf{1}\text{ heads}) &= \frac{\text{HT or TH, 2 chances}}{\text{4 outcomes (HH, HT, TH, TT)}} \\ &= \frac{1}{4}(\text{H first and then T}) + \frac{1}{4}(\text{T first and then H}) \\ &= \frac{1}{2} \end{aligned} \tag{7}$$

Till now we have only discusses discrete variables. Probability can also be defined for continuous variables, which can be understood as limit of discrete variables. As an example take the case of sum of outcome of $2-$ dice throws :

$$
\begin{aligned}
&\mathbf{P}(1+1=2)=\frac{1}{36}\\
&\mathbf{P}(1+2/2+1=3)=\frac{2}{36}\\
&\mathbf{P}(1+3/2+2/3+1=4)=\frac{3}{36}\\
&\mathbf{P}(1+4/2+3/3+2/4+1=5)=\frac{4}{36}\\
&\mathbf{P}(1+5/2+4/3+3/4+2/5+1=6)=\frac{5}{36}\\
&\mathbf{P}(1+6/2+5/3+4/4+3/5+2/6+1=7)=\frac{6}{36}\\
&\mathbf{P}(2+6/3+5/4+4/5+3/6+2=8)=\frac{5}{36}\\
&\mathbf{P}(3+6/4+5/5+4/6+3=9)=\frac{4}{36}\\
&\mathbf{P}(4+6/5+5/6+4=10)=\frac{3}{36}\\
&\mathbf{P}(5+6/6+5=11)=\frac{2}{36}\\
&\mathbf{P}(6+6=12)=\frac{1}{36}
\end{aligned}
\tag{8}
$$

Thus for discrete variables, probability $P(x_i)=p_i\geq 0$ for random variable x where x takes value x_i due to an event.

Probability for continuous variables

Similar to the above case, for continuous variables (random) one can define a distribution function $f(x)$ which will give the chance of the variable.

This is defined as follows : If say $x\leq X\leq x+dx$ where dx is small, which is to say the random variable X lies in the range $(x,x+dx)$, the probability of $\mathbf{P}(x\leq X\leq x+dx)=\int_x^{x+dx}f(x)dx\sim f(X)dx$. For the discrete example of rolling od two dices above we immediately see : X be 6 i.e. sum of outcomes = 6

$$
\begin{aligned}
&\therefore \mathbf{P}(x\leq 6\leq x+dx)=\int_x^{x+dx}f(x)dx=\frac{5}{36}\quad \text{let x=}6-\epsilon \text{ and } dx=2\epsilon\\
&\Rightarrow \mathbf{P}(6-\epsilon\leq 6\leq 6+\epsilon)=\lim_{\epsilon\to 0}\int_{6-\epsilon}^{6+\epsilon}f(x)dx=\frac{5}{36}
\end{aligned}
\tag{9}
$$

This is only true if $f(x) \sim \frac{5}{35}\delta(x-6)$. This gives a simple way of writing probability function which in the case of sum of outcome of two dices give:

$$\begin{aligned} f(x) = \quad & \frac{1}{36}[\delta(x-2)+\delta(x-12)] + \frac{2}{36}[\delta(x-3)+\delta(x-11)] + \frac{3}{36}[\delta(x-4)+\delta(x-10)] \\ & + \frac{4}{36}[\delta(x-5)+\delta(x-9)] + \frac{5}{36}[\delta(x-6)+\delta(x-8)] + \frac{6}{36}\delta(x-7) \end{aligned} \tag{10}$$

where we see that $\mathbf{P}(x_i) = \int_{x_i-\epsilon}^{x_i+\epsilon} f(x)dx$. $f(x)$ is called the probability distribution function and for discrete case it is easy to see, its given by sum δ function. If x can take any value, not necessarily discrete,we have the generalization:

$$\begin{aligned} & \mathbf{P}(x_i) \geq 0 \rightarrow f(x) \geq 0 \\ & \sum_i \mathbf{P}(x_i) = 1 \rightarrow \int f(x)dx \end{aligned} \tag{11}$$

Using these basic definitions one can now do most things which we did for discrete variables. As an application we have Quantum Mechanics. In Quantum Mechanics we look at say position variable which is taken as a random variable i.e. for a given particle if one makes many measurements, one will get random outcomes of x (position) and the probability distribution function $f(x)$ tells us the probability that the measurement of x lies between x and $x + dx$. Let us do an example, say $f(x) = Ne^{-\frac{2x}{a}}$ for $0 \leq x \leq \infty$ be the probability distribution function of a quantum particle, then :

$$\begin{aligned} & \int_0^\infty f(x)dx = 1 = \mathbf{P}(0 \leq x \leq \infty) \\ & \text{which is the total probability of finding the particle as it must be inside } (0,\infty) \\ & \Rightarrow \int_0^\infty Ne^{-\frac{2x}{a}}dx = 1 \\ & \Rightarrow N\frac{e^{-2\frac{x}{a}}}{-\frac{2}{a}}\Bigg|_0^\infty = 1 \\ & \Rightarrow \frac{Na}{2} = 1 \Rightarrow N = \frac{2}{a} \end{aligned} \tag{12}$$

Thus N has to be chosen such that the total probability of the particle being in the allowed domain is 1, i.e. particle has to exist somewhere in the given domain. Hence we obtain the normalized probability distribution function $f(x) = \frac{2}{a}e^{-\frac{2x}{a}}$ for $0 \leq x \leq \infty$. From this we can obtain the probability of finding the particle in the interval $[x_1, x_2]$, which is :

$$\Rightarrow \mathbf{P}(x_1 \leq x \leq x_2) = \int_{x_1}^{x_2} \frac{2}{a}e^{-2\frac{x}{a}}dx \tag{13}$$

i.e. chance that x will lie between $[x_1, x_2]$. This means that when we make large number of measurements N, then only in certain cases say N' among them will x lie between $[x_1, x_2]$, for other cases it will lie outside. Hence :

$$P(x_1 \leq x \leq x_2) = \lim_{N\to\infty} \frac{N'}{N} \tag{14}$$

and we will observe that as $N \to \infty$, $N' \to \mathbf{P}(x_1 \leq x \leq x_2)N$

The Average Value

The average value of the random variable is given by :$\langle x \rangle = \sum_i x_i p_i$. Let us compute the average year of birth of students born in a class with the following data :

1998 $\to$ 1 student, 1999 $\to$ 4 students, 2000 $\to$ 8 students, 2001 $\to$ 6 students, 2002 $\to$ 1 student

From which we calculate :

$$\mathbf{P}(1998) = \frac{1}{20}, \mathbf{P}(1999) = \frac{4}{20}, \mathbf{P}(2000) = \frac{8}{20}, \mathbf{P}(2001) = \frac{6}{20}, \mathbf{P}(2002) = \frac{1}{20}$$

Then the average year of birth is :

$$\begin{aligned}
&1998 \times \frac{1}{20} + 1999 \times \frac{4}{20} + 2000 \times \frac{8}{20} + 2001 \times \frac{6}{20} + 2002 \times \frac{1}{20} \\
&= \frac{1}{20}(1 \times 1998 + 4 \times 1999 + 8 \times 2000 + 6 \times 2001 + 1 \times 2002) \\
&= \frac{1}{N}\sum_i x_i
\end{aligned} \tag{15}$$

where each out come of the event (year of student's birth) is either countd separately or each outcome is assigned multiplicities with which it occurs.
For the discrete case of sum of 2 dices, the average outcome is :

$$\begin{aligned}
\sum_i x_i p_i \quad &= 2 \times \frac{1}{36} + 3 \times \frac{2}{36} + 4 \times \frac{3}{36} + 5 \times \frac{4}{36} + 6 \times \frac{5}{36} \\
&+7 \times \frac{6}{36} + 8 \times \frac{5}{36} + 9 \times \frac{4}{36} + 10 \times \frac{3}{36} + 11 \times \frac{2}{36} + 12 \times \frac{1}{36} \\
&= \int_0^{13} x[\frac{1}{36}[\delta(x-2) + \delta(x-12)] + \frac{2}{36}[\delta(x-3) + \delta(x-11)] + \frac{3}{36}[\delta(x-4) + \delta(x-10)] \\
&+\frac{4}{36}[\delta(x-5) + \delta(x-9)] + \frac{5}{36}[\delta(x-6) + \delta(x-8)] + \frac{6}{36}\delta(x-7)] \\
&= \int_0^{13} x f(x) dx
\end{aligned} \tag{16}$$

where we have gone from the discrete to continuous variable x, using the delta function. Thus we automatically find the generalization of the mean or average value of the random variable using the distribution function :

$$\langle x \rangle_{(a,b)} = \int_a^b x f(x) dx \quad \text{average value in a given region } (a,b) \tag{17}$$

or over the entire support (support is the range of values the random variable can take) of the random variable x :

$$\langle x \rangle_{total} = \int_{support} x f(x) dx \tag{18}$$

Let us do an example : for given $f(x) = \frac{2}{a} e^{-2\frac{x}{a}}$, the mean of the random variable x is computed over the support $(0, \infty)$:

$$\begin{aligned} \langle x \rangle &= \int_0^\infty x \left(\frac{2}{a} e^{-2\frac{x}{a}} \right) dx \\ &= \frac{2}{a} \frac{x e^{-2\frac{x}{a}}}{-\frac{2}{a}} \bigg|_0^\infty - \frac{2}{a} \int_0^\infty \frac{e^{-2\frac{x}{a}}}{-\frac{2}{a}} dx \\ &= \frac{a}{2} \end{aligned} \tag{19}$$

One can define something called the "deviation" which is the difference of the random variable from the mean : $v_i = x_i - \langle x \rangle$. Obviously : $\langle v \rangle = \sum_i v_i p_i = \sum_i (x_i - \langle x \rangle) p_i = \sum_i x_i p_i - \langle x \rangle \sum_i p_i = 0$. Or analogously over continuous variables :

$$\begin{aligned} \langle v \rangle &= \int_{support} (x - \langle x \rangle) f(x) dx \\ &= \int_{support} x f(x) dx - \langle x \rangle \int_{support} f(x) dx = 0 \end{aligned} \tag{20}$$

where we have taken $\langle x \rangle$ outside the integral since its a constant. More importantly one can define a quantity which tells how some random variable is spread called the standard deviation :

$$\begin{aligned} \sigma &= \sqrt{\langle v^2 \rangle} \\ &= \sqrt{\frac{1}{N} \sum_i (x_i - \langle x \rangle)^2} \\ &= \sqrt{\sum_i (x_i - \langle x \rangle)^2 p_i} \end{aligned} \tag{21}$$

where in the second equality the counting of the sum is over each event/value of random variable separately and the third equality is when we rewrite it using the probability of occurrence of each event/random variable. Now it is easy to check that :

$$\begin{aligned} \sigma^2 &= \sum_i (x_i - \langle x \rangle)^2 p_i \\ &= \sum_i (x_i^2 p_i - 2 x_i \langle x \rangle p_i + \langle x \rangle^2 p_i) \\ &= \langle x^2 \rangle - 2 \langle x \rangle \langle x \rangle + \langle x \rangle^2 \\ &= \langle x^2 \rangle - \langle x \rangle^2 \end{aligned} \tag{22}$$

While for continuous distributions analogously we have:

$$\sigma^2 = \int (x - \langle x \rangle)^2 f(x) dx \tag{23}$$

Transformation between variables

Now say there is another random variable $y = ax + b$ related to x, where a, b are constants :

$$\therefore \int_{support} f(x)dx = 1 \quad \int_{support'} g(y)dy = 1 \tag{24}$$

where " support " and " support' " are the domain over which the integration is defined, i.e. the values the two random variables x and y take. Hence using the transformation:

$$\begin{aligned} \int_{support'} g(ax+b)d(ax+b) = 1 \\ \Rightarrow \int_{support} g(ax+b)adx = \int_{support} f(x)dx \end{aligned} \tag{25}$$

Comparing we have $g(ax+b)adx = f(x)dx$. This gives the relation between the probability measure. This can also be seen from the notion of infinitesimal probability in an interval. Note that the probability of the value of the random variable y to lie in the range $[y', y' + \Delta y] = [ax'+b, a(x'+\Delta x)+b]$ is the same if the random variable x lies in the range $[x', x'+\Delta x]$ simply because for each value of the random varaible x there is an unique value of the random varaible y and when ever x takes value in this range y takes value in the corresponding range related by the linear transformation. Hence we must have:

$$\int_{x'}^{x'+\Delta x+b} g(ax+b)adx = \int_{x'}^{x'+\Delta x} f(x)dx \tag{26}$$

Taking Δx small and using Mean value theorem :

$$g(ax'+b)a\Delta x \sim f(x')\Delta x \tag{27}$$

Which shows the equality of infinitesimal probability under linear transformation. Now the mean $\langle y \rangle$ is then for say $x \in (-\infty, \infty)$

$$\begin{aligned} \langle y \rangle &= \int_{-\infty}^{\infty} yg(y)dy \\ &= \int_{-\infty}^{\infty} (ax+b)g(ax+b)d(ax+b) \\ &= \int_{-\infty}^{\infty} (ax+b)f(x)dx \\ &= a\langle x \rangle + b \end{aligned} \tag{28}$$

Similarly for variance we have :

$$\begin{aligned}\sigma_y^2 &= \int_{-\infty}^{\infty} (y - \langle y \rangle)^2 g(y) dy \\ &= \int_{-\infty}^{\infty} (ax + b - a\langle x \rangle - b)^2 f(x) dx \\ &= \int_{-\infty}^{\infty} a^2 (x - \langle x \rangle)^2 f(x) dx \\ &= a^2 \sigma_x^2 \end{aligned} \tag{29}$$

We can have cases where we have a non linear but monotonic/one-one map between two random variable. Let us do an example : if the normalized probability distribution of the random variable $x \in (0, \infty)$ is given by $f(x) = \frac{2}{a} e^{-2\frac{x}{a}}$ and given $y = x^2$ a monotonic map,

$$= \int_0^{\infty} \frac{2}{a} e^{-2\frac{x}{a}} dx = \int_0^{\infty} \frac{2}{a} e^{-2\frac{\sqrt{y}}{a}} \frac{dy}{2\sqrt{y}} \tag{30}$$

which means the probability measure for y is given by : $\frac{1}{a\sqrt{y}} e^{-2\frac{\sqrt{y}}{a}}$. This can also be seen from the fact that since this is a monotonic map we have the consequence that if $x \in [x', x' + \Delta x]$ we must $y \in [x'^2, (x' + \Delta x)^2]$, hence the probability of the random variables to occur within the given interval must be same :

$$\int_{y'}^{y' + \Delta y} g(y) dy = \int_{x'^2}^{x'^2 + 2x'\Delta x} g(y) dy = \int_{x'}^{x' + \Delta x} f(x) dx \tag{31}$$

where in the upper limit we have used $y' + \Delta y \sim x'^2 + 2x'\Delta x$ by keeping first order of $(x' + \Delta x)^2$. Then using the Mean value theorem for infinitesimal Δx :

$$\begin{aligned} & g(y')\Delta y = f(x')\Delta x \\ & \Rightarrow g(y') 2x' \Delta x = f(x') \Delta x \\ & \Rightarrow g(y') = \frac{f(x')}{2x'}, \quad x' = \sqrt{y'} \end{aligned} \tag{32}$$

Thus it is important to incorporate the Jacobian of transformation coming from the change of variables in the integration while defining the new probability density function.

Examples of Distribution functions

The Gaussian Distribution functions

We define the Normalized Gaussian distribution as follows:

$$f(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{33}$$

- To check normalization :

$$\begin{aligned}
&\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{\pi}} e^{-\left(\frac{x-\mu}{\sqrt{2}\sigma}\right)^2} d\left(\frac{x-\mu}{\sqrt{2}\sigma}\right) \\
&= 2\int_{0}^{\infty} \frac{1}{\sqrt{\pi}} e^{-t^2} dt \\
&= \frac{1}{\sqrt{\pi}} \int_{0}^{\infty} e^{-\omega} \frac{d\omega}{\sqrt{\omega}} \quad t^2 = \omega \\
&= \frac{1}{\sqrt{\pi}} \int_{0}^{\infty} \omega^{\frac{1}{2}-1} e^{-\omega} d\omega \\
&= \frac{1}{\sqrt{\pi}} \Gamma(\frac{1}{2}) = 1, \quad \Gamma(\frac{1}{2}) = \sqrt{\pi}
\end{aligned} \tag{34}$$

- Now let us check the mean of this given distribution :

$$\begin{aligned}
\langle x \rangle \quad &= \int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \\
&= \int_{-\infty}^{\infty} (x-\mu) \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx + \mu \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \\
&= \int_{-\infty}^{\infty} t \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{t^2}{2\sigma^2}} + \mu \\
&= \mu
\end{aligned} \tag{35}$$

- Square of the standard deviation (variance):

$$\begin{aligned}
\langle (x - \langle x \rangle)^2 \rangle = \quad &\langle x^2 \rangle - \langle x \rangle^2 \\
&= \int_{-\infty}^{\infty} (x - \langle x \rangle)^2 \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \\
&= 2\sigma^2 \int_{-\infty}^{\infty} \frac{(x-\mu)^2}{2\sigma^2} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} d\frac{(x-\mu)}{\sqrt{2\sigma^2}}, \quad \langle x \rangle = \mu \\
&= 4\sigma^2 \int_{0}^{\infty} t^2 \frac{1}{\sqrt{\pi}} e^{-t^2} dt, \quad t^2 = \omega \\
&= \frac{4\sigma^2}{\sqrt{\pi}} \int_{0}^{\infty} \omega e^{-\omega} \frac{d\omega}{2\sqrt{\omega}} \\
&= \frac{2\sigma^2}{\sqrt{\pi}} \int_{0}^{\infty} \omega^{\frac{3}{2}-1} e^{-\omega} d\omega \\
&= \frac{2\sigma^2}{\sqrt{\pi}} \Gamma(\frac{3}{2}) = \frac{2\sigma^2}{\sqrt{\pi}} \frac{1}{2} \Gamma(\frac{1}{2}) = \sigma^2
\end{aligned} \tag{36}$$

1. Plot the Gaussian distribution, pointing out the values of the function at $x = \mu \pm n\sigma$ for $n = 0, 1, 2$.

Binomial Distribution function

Let a random variable X be number of successes in $n \in$ natural numbers, repeated independent trials, then the probability of k successes in n trials where p is the probability of success and $(1-p)$ the probability of failure $= {}^nC_k p^k (1-p)^{n-k}$. Here k successes occur with probability p^k and $n-k$ failure with probability $(1-p)^{n-k}$. However there are ${}^nC_k$ ways of distributing k successes in a sequence on n trials. Note that the above is the k^{th} binomial coefficient of $(p+(1-p))^n$ expansion. Hence the name.

Examples: The probability for getting $k = 0, 1, \ldots, 6$ heads in 6 trials for an unbiased coin. Here $p = \frac{1}{2}$

$$\begin{aligned} \mathbf{P}(0) &= {}^6C_0 \left(\frac{1}{2}\right)^0 \left(\frac{1}{2}\right)^6 \\ \mathbf{P}(1) &= {}^6C_1 \left(\frac{1}{2}\right)^1 \left(\frac{1}{2}\right)^5 \end{aligned} \tag{37}$$

and so on. If say the coin in biased and $p = 0.3$ for heads, we have:

$$\begin{aligned} \mathbf{P}(0) &= {}^6C_0 (0.3)^0 (0.7)^6 \\ \mathbf{P}(1) &= {}^6C_1 (0.3)^1 (0.7)^5 \end{aligned} \tag{38}$$

and so on.

- Let us compute the mean for this distribution :

$$\langle x \rangle = \sum_{i=0} x_i p_i \tag{39}$$

Here $x_i = k$ the number of successes and $p_i = \mathbf{P}(k)$. Then mean :

$$\begin{aligned} & \sum_{k=0}^{n} k\, {}^nC_k p^k (1-p)^{n-k} \\ &= \sum_{k=0}^{n} k \frac{n!}{(n-k)!k!} p^k (1-p)^{n-k} \\ &= \sum_{k=1}^{n} \frac{n(n-1)!}{(n-1-k+1)!(k-1)!} p p^{k-1} (1-p)^{n-1-k+1} \\ &= np \sum_{k-1=l=0}^{n-1} \frac{(n-1)!}{(n-1-l)!l!} p^l (1-p)^{n-1-l} \\ &= np(p+1-p)^{n-1} = np \end{aligned} \tag{40}$$

- Now let us calculate the variance :

$$\begin{aligned}
\langle x^2\rangle &= \sum_i^n x_i^2 p_i \\
&= \sum_{k=0}^n \frac{k^2 n!}{(n-k)!k!} p^k (1-p)^{n-k} \\
&= \sum_{k=1}^n \frac{kn(n-1)!}{(n-k)!(k-1)!} p^k (1-p)^{n-k} \\
&= \sum_{k=1}^n \frac{(k-1)np^2(n-1)(n-2)!}{(n-2-k+2)!(k-1)!} p^{k-2} (1-p)^{n-k} \\
&+ \sum_{k=1}^n \frac{np(n-1)!}{(n-1-k+1)!(k-1)!} p^{k-1} (1-p)^{n-k} \\
&= \sum_{k=2}^n \frac{n(n-1)p^2(n-2)!}{(n-2-k+2)!(k-2)!} p^{k-2} (1-p)^{n-2-k+2} + np(p+1-p)^{n-1} \\
&= n(n-1)p^2(p+1-p)^{n-2} + np \\
&= np + n(n-1)p^2
\end{aligned} \tag{41}$$

Thus σ^2 or variance is given :

$$\begin{aligned}
\sigma^2 &= \langle x^2\rangle - \langle x\rangle^2 \\
&= np + n(n-1)p^2 - (np)^2 \\
&= np(1-p)
\end{aligned} \tag{42}$$

Special case : If we take $n \to \infty$ and $p \to 0$ with $np = \mu$, fixed i.e. the mean is kept fixed then : ($k = x$ here and $0 \le x \le \infty$)

$$\lim_{n\to\infty;p\to 0} {}^nC_x p^x (1-p)^x \tag{43}$$

for large n we have $\ln(n!) \to n\ln n - n$, $\ln(n-x)! \to (n-x)\ln(n-x) - (n-x)$, using the Stirling approximation :

$$
\begin{aligned}
\frac{n!}{(n-x)!} \quad & \sim \frac{e^{n\ln n - n}}{e^{(n-x)\ln(n-x)-(n-x)}} \\
& = \frac{n^n e^{-n}}{(n-x)^{n-x} e^{-(n-x)}} \\
& = \frac{n^{n-x} n^x}{(n-x)^{n-x} e^x} \\
& = \left(\frac{n}{e}\right)^x \left(\frac{n}{n-x}\right)^{n-x} \\
& = \left(\frac{n}{e}\right)^x \left(\frac{n-x+x}{n-x}\right)^{n-x} \\
& = \left(\frac{n}{e}\right)^x \left(1+\frac{x}{n-x}\right)^{n-x}
\end{aligned} \tag{44}
$$

Now since $n >> x$ in this limit we can write above as:

$$
\sim \left(\frac{n}{e}\right)^x \left(1+\frac{x}{n}\right)^n \sim \left(\frac{n}{e}\right)^x e^x \sim n^x \tag{45}
$$

where we have used the fact that

$$
\lim_{\Lambda \to \text{large number}} \left(1+\frac{x}{\Lambda}\right)^\Lambda \sim e^x \tag{46}
$$

Again :

$$
\begin{aligned}
(1-p)^{n-x} \quad & = \left(1-\frac{pn}{n}\right)^{n-x} \\
& \Rightarrow \lim_{n\to\infty; p\to 0; pn=\mu} \left(1-\frac{\mu}{n}\right)^{n-x} \quad n >> x \\
& = e^{-\mu}
\end{aligned} \tag{47}
$$

gathering terms :

$$
\lim_{n\to\infty; p\to\infty} {}^nC_x p^x (1-p)^x \to \frac{n^x p^x}{x!} e^{-\mu} = \frac{\mu^x}{x!} e^{-\mu} \tag{48}
$$

The above is called the Poisson distribution function and is the large number limit of the binomial distribution. We can check the following statements regarding this distribution then :

- Normalization :

$$\begin{aligned}
\sum_{x=0}^{\infty} P(x) \quad &= \sum_{x=0}^{\infty} \frac{\mu^x}{x!} e^{-\mu} \\
&= e^{-\mu} \sum_{x=0}^{\infty} \frac{\mu^x}{x!} \\
&= e^{-\mu} e^{\mu} = 1
\end{aligned} \tag{49}$$

- Mean :

$$\begin{aligned}
\langle x \rangle \quad &= e^{-\mu} \sum_{x=0}^{\infty} x \frac{\mu^x}{x!} \\
&= e^{-\mu} \sum_{x=1}^{\infty} x \frac{\mu^x}{x!} \\
&= e^{-\mu} \sum_{x=1}^{\infty} \frac{\mu^x}{(x-1)!} \\
&= e^{-\mu} \mu \sum_{x=1}^{\infty} \frac{\mu^{x-1}}{(x-1)!} \\
&= \mu e^{-\mu} e^{\mu} = \mu = \lim_{n \to \infty; p \to 0} np
\end{aligned} \tag{50}$$

where the first term in the series is zero hence the sum starts from $x = 1$.

- Variance ,σ^2 :

$$\begin{aligned}
&\lim_{n \to \infty; p \to 0; np = \mu} np(1-p) \\
&= \lim_{p \to 0} \mu - \mu p = \mu
\end{aligned} \tag{51}$$

Hence we see that mean and variance are the same.

- We can also take the continuous limit of x. Then :

$$\mathbf{P}(x) = \frac{\mu^x e^{-\mu}}{\Gamma(x+1)}, \quad x! \to Gamma(x+1) \tag{52}$$

Again we can check the mean:

$$\langle x \rangle \quad = \int_0^{\infty} \frac{x \mu^x e^{-\mu}}{\Gamma(x+1)} dx \tag{53}$$

Note that :

$$\begin{aligned} x\mu^x e^{-\mu} \quad &= \mu \frac{d(\mu^x)}{d\mu} e^{-\mu} \\ &= \mu \frac{d}{d\mu}(\mu^x e^{-\mu}) + \mu\mu^x e^{-\mu} \end{aligned} \tag{54}$$

Plugging this in the integral we obtain :

$$\begin{aligned} &\int_0^\infty \left[\frac{\mu \frac{d}{d\mu}(\mu^x e^{-\mu}) + \mu\mu^x e^{-\mu}}{\Gamma(x+1)} \right] dx \\ &= \mu \frac{d}{d\mu} \left[\int_0^\infty \frac{\mu^x e^{-\mu}}{\Gamma(x+1)} dx \right] + \mu \int_0^\infty \frac{\mu^x e^{-\mu}}{\Gamma(x+1)} dx \\ &= \mu \frac{d}{d\mu}(1) + \mu(1) = \mu \end{aligned} \tag{55}$$

similarly calculate σ^2

Lorentzian

We define the Normalized Lorentzian distribution as follows:

$$f(;x, x_0, \gamma) = \frac{1}{\pi\gamma} \left[\frac{\gamma^2}{(x-x_0)^2 + \gamma^2} \right] \tag{56}$$

- Normalization:

$$\begin{aligned} \int_{-\infty}^{\infty} f(x; x_0, \gamma) \quad &= \int_{-\infty}^{\infty} dx \frac{1}{\pi\gamma} \left[\frac{\gamma^2}{(x-x_0)^2 + \gamma^2} \right] \\ &= \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{1}{\frac{(x-x_0)^2}{\gamma^2} + 1} d\left(\frac{x-x_0}{\gamma} \right) \\ &= \frac{1}{\pi} \tan^{-1} \left(\frac{x-x_0}{\gamma} \right) \Big|_{-\infty}^{\infty} \\ &= \frac{1}{\pi} [\frac{\pi}{2} - (-\frac{\pi}{2})] = 1 \end{aligned} \tag{57}$$

- Mean :

$$\begin{aligned} \langle x \rangle \quad &= \frac{1}{\pi\gamma} \int_{-\infty}^{\infty} x \left[\frac{\gamma^2}{(x-x_0)^2 + \gamma^2} \right] dx \\ &= \frac{1}{\pi\gamma} \int_{-\infty}^{\infty} (x-x_0) \left[\frac{\gamma^2}{(x-x_0)^2 + \gamma^2} \right] d(x-x_0) + \frac{1}{\pi\gamma} \int_{-\infty}^{\infty} x_0 \left[\frac{\gamma^2}{(x-x_0)^2 + \gamma^2} \right] dx \end{aligned} \tag{58}$$

The first term is zero because its an odd integral and the integral in the second term gives x_0 using the normalization condition. But this is a bit misleading since the integral is not pice-wise finite under splitting at any point on the real axis. Hence in the proper sense the mean is not defined.

- Variance :

$$\begin{aligned}
\sigma^2 = \langle (x - \langle x \rangle)^2 \rangle \qquad &= \int_{-\infty}^{\infty} dx \frac{1}{\pi\gamma} \left[\frac{(x - x_0)^2 \gamma^2}{(x - x_0)^2 + \gamma^2} \right] \\
&= \frac{\gamma}{\pi} \int_{-\infty}^{\infty} dx \left[\frac{(x - x_0)^2 + \gamma^2 - \gamma^2}{(x - x_0)^2 + \gamma^2} \right] \\
&= \frac{\gamma}{\pi} \int_{-\infty}^{\infty} dx \left[\frac{(x - x_0)^2 + \gamma^2}{(x - x_0)^2 + \gamma^2} \right] + \frac{\gamma}{\pi} \int_{-\infty}^{\infty} dx \left[\frac{-\gamma^2}{(x - x_0)^2 + \gamma^2} \right]
\end{aligned} \tag{59}$$

As is evident the first integral is undefined and hence in this case σ^2 does not exist.