A note on Probability Distributions

Debashis Chatterjee
Department of physics

The concept of Probability Distribution plays an important role in Physics. They may be of different kinds – some discrete and some continuous. Below we shall discuss three frequently appearing distributions, one discrete, one continuous, while the third showing a transition from the former to the latter.

We shall assume that the reader is familiar with the basic concept of Probability, the basic definitions and properties.

Binomial Distribution

One-dimensional Random Walk Problem

A person standing at the origin will take 'N' steps. If the probability of his taking a step towards right is 'p' and that towards left is 'q', what is the prob. that the person takes 'n' steps towards right and the remaining (N – n) steps towards left ?

The no. of ways he can chose his 'n' steps towards right = ${}^{N}C_{n}$
(for example, the 1st step, the 3rd step and the 4th step)
For one such combination : (e.g. R L L R R),
The prob. of taking one step towards right = p
⇒ the prob. of taking 'n' step towards right = p^n
The prob. of taking one step towards left = q
⇒ the prob. of taking (N – n) step towards left = $q^{(N-n)}$
⇒ the prob. of the particular combination considered = $p^n \times q^{(N-n)}$
⇒ the prob. of **all** such combinations
(with 'n' steps towards right and (N – n) steps towards left) = $\mathbf{{}^{N}C_{n} \times p^{n} \times q^{(N-n)}}$

The above result was proved by the **Swiss mathematician Jakob Bernoulli**, which was published posthumously in **1713**.

Note that : the above expression is a term in the binomial expansion of $(p + q)^N$

$$(p + q)^N = p^N + {}^{N}C_1\, p^{n-1} q + {}^{N}C_2\, p^{n-2} q^2 + \ldots q^N$$
$$= \Sigma\, {}^{N}C_n \times p^n \times q^{(N-n)}$$

If we sum over all possible values of 'n' (from 0 to N), we exhaust all the possibilities.

∴ $\Sigma\, {}^{N}C_n \times p^n \times q^{(N-n)}$ should be 1.

Now, p + q = 1, because the a single step may either be towards left or towards right.

∴ LHS : $(p + q)^N = 1$ (proved)

Mean ⟨n⟩ :

If the value of a quantity 'x' is found to be x_i for n_i number of times, then

the mean x is : $\mathbf{\Sigma\, n_i x_i / \Sigma\, n_i}$;

but $(n_i/\Sigma\, n_i)$ gives the probability 'P_i' of obtaining the value x_i

∴ mean x : $\langle x\rangle = \mathbf{\Sigma\, x_i P_i}$

Thus, $\langle n\rangle = \Sigma\, n\, P(n)$, where $P(n) = {}^{N}C_n\, p^n q^{(N-n)}$

Now, $(p + q)^N = \Sigma\, {}^{N}C_n\, p^n q^{(N-n)}$

Now differentiate both sides partially, w.r.t. 'p' :

From the RHS : $\partial/\partial p \, \Sigma \, {}^N C_n \, p^n \, q^{(N-n)} = \Sigma \, {}^N C_n \, (np^{n-1}) \, q^{(N-n)}$

$$\Rightarrow \; p \, \partial/\partial p \, \Sigma \, {}^N C_n \, p^n \, q^{(N-n)} = \Sigma \, {}^N C_n \, (np^n) \, q^{(N-n)},$$

Which is nothing but $\Sigma \, n \, P(n) = \langle n \rangle$

From the LHS : $\partial/\partial p \, (p + q)^N = N \, (p + q)^{N-1}$

$$\text{or, } \partial/\partial p \, (p + q)^N = N \, (p + q)^{N-1}$$

$$\Rightarrow \; p \, \partial/\partial p \, (p + q)^N = Np \, (p + q)^{N-1}$$

Remember : $(p + q) = 1$.

$$\text{So, } \boldsymbol{\langle n \rangle = Np.}$$

This is somewhat expected, since 'p' s the prob. of taking a 'right' step, Np is the 'expected' number of right steps, out of a total number of N steps.

A point to remember : While differentiating partially, w.r.t. 'p', we treat p and q to be independent variables, though at the end of the calculation, we set : $q = (1 - p)$.

Mean $\langle n^2 \rangle$:

$$(p + q)^N = \Sigma \, {}^N C_n \, p^n \, q^{(N-n)}$$

$$\Rightarrow \; p \, \partial/\partial p \, \Sigma \, {}^N C_n \, p^n \, q^{(N-n)} = \Sigma \, {}^N C_n \, (np^n) \, q^{(N-n)}$$

$$\Rightarrow \; \partial/\partial p \, (p \, \partial/\partial p) \, \Sigma \, {}^N C_n \, p^n \, q^{(N-n)} = \Sigma \, {}^N C_n \, (n^2 p^{n-1}) \, q^{(N-n)}$$

$$\Rightarrow \; p \, \partial/\partial p \, (p \, \partial/\partial p) \, \Sigma \, {}^N C_n \, p^n \, q^{(N-n)} = \Sigma \, {}^N C_n \, (n^2 p^n) \, q^{(N-n)},$$

$$= \langle n^2 \rangle$$

From the LHS : $p \, \partial/\partial p \, (p + q)^N = Np \, (p + q)^{N-1}$

$$\text{or, } \partial/\partial p \, (p\partial/\partial p) \, (p + q)^N = N \, (p + q)^{N-1} + N(N-1) \, p(p + q)^{N-2}$$

$$\Rightarrow \; p \, \partial/\partial p \, (p\partial/\partial p) \, (p + q)^N = Np \, (p + q)^N + N(N-1) \, p^2 (p + q)^{N-2}$$

$$= \mathbf{Np + N(N-1) \, p^2} \quad \text{as } (p + q) = 1$$

Standard Deviation σ :

$$\langle n^2 \rangle - \langle n \rangle^2 = [Np + N^2p^2 - Np^2] - N^2p^2$$

$$= Np - Np^2$$

$$= Np \, (1 - p) = Npq$$

$$\Rightarrow \; \sigma = [\langle n^2 \rangle - \langle n \rangle^2]^{1/2} = \boldsymbol{\sqrt{(Npq)}}$$

Poisson Distribution

Poisson Distribution is named after the **French Mathematician Siméon Denis Poisson**, who discovered it in **1838**.

For a Binomial Distribution, we have seen that :

$$P(n) = {}^N C_n \, p^n \, q^{(N-n)} = {}^N C_n \, p^n \, (1 - p)^{(N-n)}$$

Let us consider a situation where 'N' is very large, 'p' is very small, but Np is finite = x, say.

The term : ${}^N C_n = N(N-1)(N-2) \ldots (N - n + 1) / n!$

As long as 'n' is finite and small compared to 'N', ${}^N C_n \approx N^n/n!$

Combined with the term p^n,

$${}^N C_n \, p^n \approx N^n p^n / n! \approx x^n / n!$$

The term : $(1 - p)^{(N-n)}$ can be approximated as : $(1 - p)^N$

$$= (1 - x/N)^N$$

and in the limit $N \to \infty$, $(1 - x/N)^N \to e^{-x}$

Explicitly:

$$(1 - p)^{(N-n)} = 1 - (N - n)p + (N - n)(N - n - 1)/2!\, p^2 + \ldots$$
$$\approx 1 - Np + (N^2/2!)\, p^2 + \ldots,$$

neglecting 'n' with respect to 'N'.

$$\Rightarrow (1 - p)^{(N-n)} \approx 1 - Np + (N^2p^2/2!) + \ldots$$
$$\approx 1 - x + x^2/2! + \ldots$$
$$= e^{-x}$$

Towards the end of the series, $[(N - n) - k]$ becomes comparable to '1', or in other words, N becomes comparable to $(n + k)$, because 'k' increases. However, combined with p^n, such terms $\to 0$, as $p \to 0$.

Thus, $P(n) = {}^{N}C_n\, p^n q^{(N-n)} \approx \mathbf{x^n e^{-x}/n!}$

Normalization :

$$\Sigma\, P(n) \quad \text{for } n = 0 \text{ to } \infty$$
$$= \Sigma\, x^n e^{-x}/n! = e^{-x}\, \Sigma\, x^n/n! = e^{-x} \times e^{x} = 1$$

Mean $\langle n \rangle$:

$$\langle n \rangle = \Sigma\, n\, P(n) = e^{-x}\, \Sigma\, nx^n/n!$$

Now, $\partial/\partial x\, \Sigma\, x^n/n! = \Sigma\, nx^{n-1}/n!$

$$\Rightarrow (x\, \partial/\partial x)\, \Sigma\, x^n/n! = \Sigma\, nx^n/n!$$

So, $\Sigma\, nx^n/n! = (x\, \partial/\partial x)\, e^x = x\, e^x$

$$\Rightarrow \boldsymbol{\langle n \rangle = e^{-x}\, \Sigma\, nx^n/n! = x}$$

Explicitly :

$$\langle n \rangle = e^{-x}\, \Sigma\, nx^n/n! \quad [\text{for for } n = 0 \text{ to } \infty]$$
$$= e^{-x} [\, x + 2x^2/2! + 3x^3/3! + \ldots]$$
$$= e^{-x} [\, x + x^2/1! + x^3/2! + \ldots]$$
$$= e^{-x}\, x\, [\, 1 + x/1! + x^2/2! + \ldots]$$
$$= e^{-x}\, xe^{x} = x$$

$\langle n^2 \rangle$:

$$\langle n^2 \rangle = \Sigma\, n^2\, P(n) = e^{-x}\, \Sigma\, n^2x^n/n!$$
$$(x\, \partial/\partial x)\, \Sigma\, x^n/n! = \Sigma\, nx^n/n!$$
$$\Rightarrow \partial/\partial x\, (x\, \partial/\partial x)\, \Sigma\, x^n/n! = \Sigma\, n^2x^{n-1}/n!$$
$$\Rightarrow (x\, \partial/\partial x)\, (x\, \partial/\partial x)\, \Sigma\, x^n/n! = \Sigma\, n^2x^n/n!$$

So, $\Sigma\, n^2x^n/n! = (x\, \partial/\partial x)^2\, e^x = (x\, \partial/\partial x)\, xe^x$
$= x\, [e^x + xe^x] = e^x\, (x^2 + x)$

$$\Rightarrow \boldsymbol{\langle n^2 \rangle = e^{-x}\, \Sigma\, n^2x^n/n! = (x^2 + x)}$$

Standard Deviation σ :

$$\langle n^2 \rangle - \langle n \rangle^2 = (x^2 + x) - x^2$$
$$= x$$

So, for Poisson Distribution, mean = variance

and **Standard Deviation σ = √x**

Gaussian Distribution

Gaussian Distribution (also called the '**Normal Distribution**') was introduced by the **German mathematician Carl Friedrich Gauss in 1809**. This is an example of a continuous Prob. Distribution Function, in contrast to the Binomial and the Poisson distribution. Hence we shall need to introduce the concept of the 'Probability Density' here. Let 'X' be a parameter and let the probability that the value of X lies between x and x + dx be **P(x) dx,** the P(x) is the **Probability Density Function**, which is basically, the probability of finding the value of X within a unit interval.

For a Gaussian Probability Distribution Function,

$$\mathbf{P(x) = N \exp\{-(x-\mu)^2/2\sigma^2\}}$$

Normalization :

$$P(x)\,dx = N \exp\{-(x-\mu)^2/2\sigma^2\}\,dx$$

Now, the value of 'X' must lie *somewhere* between $-\infty$ and $+\infty$. So the total prob. $\int P(x)\,dx$ between these limits must be 1 $\Rightarrow$

$$N\int \exp\{-(x-\mu)^2/2\sigma^2\}\,dx = I, \text{ say}$$

Subst. first : $(x-\mu) = y$

$$\Rightarrow dx = dy$$

$$\Rightarrow I = N\int \exp\{-y^2/2\sigma^2\}\,dy$$

The limits of this integral is $-\infty$ and $+\infty$, but the integrand is clearly an even function. So we can double the integrand and change the limits as : 0 and $+\infty$.

$$I = 2N\int \exp\{-y^2/2\sigma^2\}\,dy$$

Next, subst. : $y^2/2\sigma^2 = z \Rightarrow y = \sqrt{2}\sigma z^{1/2}$

$$\Rightarrow dy = \sqrt{2}\sigma(\tfrac{1}{2} z^{-1/2})\,dz$$

$$\Rightarrow I = \sqrt{2}N\sigma \int e^{-z} z^{-1/2}\,dz$$

We identify the integral over z as $\Gamma(\tfrac{1}{2})$, the value of which is $\sqrt{\pi}$.

$\Gamma(n) = \int e^{-z} z^{n-1}\,dz$, between the limits : 0 and $+\infty$.

$$\Rightarrow I = \sqrt{(2\pi)}N\sigma$$

So, for I to be unity, $\mathbf{N = 1/\sqrt{(2\pi)}\sigma}$

Mean x :

$$\langle x\rangle = \int x\,P(x)\,dx = N\int x \exp\{-(x-\mu)^2/2\sigma^2\}\,dx$$

Subst. : $(x-\mu) = y$

$$\Rightarrow \langle x\rangle = N\int (y+\mu)\exp\{-y^2/2\sigma^2\}\,dy$$

Clearly, we can split the above integral into two, The first one, $\int Ny \exp\{-y^2/2\sigma^2\}\,dy$, has an odd integrand and therefore vanish. The second integral equals :

$$N\mu\int \exp\{-y^2/2\sigma^2\}\,dy = \mu.$$

Thus, $\mathbf{\langle x\rangle = \mu}$, which furnishes an interpretation of the parameter 'μ'.

Mean x^2 :

$$\langle x^2\rangle = \int x^2 P(x)\,dx = N\int x^2 \exp\{-(x-\mu)^2/2\sigma^2\}\,dx$$

Subst. : $(x-\mu) = y$

$$\Rightarrow \langle x^2 \rangle = N \int (y+\mu)^2 \exp\{-y^2/2\sigma^2\}\, dy$$

This can be split into three integrals. The last integral :

$$N \int \mu^2 \exp\{-y^2/2\sigma^2\}\, dy = \mu^2$$

The second integral : $N \int (2\mu y) \exp\{-y^2/2\sigma^2\}\, dy$ involves an odd function as integrand and therefore vanish.

The first integral $I_1 = N \int y^2 \exp\{-y^2/2\sigma^2\}\, dy$, with the limits : $-\infty$ and $+\infty$. Again, we double the integrand to change the limits as : 0 and $+\infty$.

subst. : $y^2/2\sigma^2 = z$, as before [one may also subst. : $y^2/2\sigma^2 = z^2$]

$$\Rightarrow y = \sqrt{2}\sigma z^{1/2}$$

$$\Rightarrow dy = \sqrt{2}\sigma (\tfrac{1}{2} z^{-1/2})\, dz$$

$$\Rightarrow I_1 = \mathbf{2}N \int (2\sigma^2 z)\, e^{-z} (\sigma z^{-1/2}/\sqrt{2})\, dz \quad [\textbf{limits being : 0 and } \infty]$$

$$= 2\sqrt{2}\, N \sigma^3 \int e^{-z} z^{1/2}\, dz$$

The integral : $\int e^{-z} z^{1/2}\, dz = \Gamma(3/2) = \tfrac{1}{2}\Gamma(\tfrac{1}{2}) = \sqrt{\pi}/2$.

$$\Rightarrow I_1 = \sqrt{(2\pi)}\, N \sigma^3$$

We have found above, that $N = 1/\sqrt{(2\pi)}\sigma \Rightarrow I_1 = \sigma^2$

So, $\langle x^2 \rangle = \mu^2 + \sigma^2$, $\langle x \rangle = \mu$

$$\Rightarrow \textbf{variance} : \langle x^2 \rangle - \langle x \rangle^2 = \sigma^2$$

and standard deviation = $\sqrt{(\text{variance})} = \boldsymbol{\sigma}$

Problem Set on Probability Distributions

1. The probability that a drunkard takes a step towards right is 75% of and that towards left is 25%. If the step length is 1 meter, What is the probability that he will reach a distance of 2 meters towards right after a set of four steps ?
2. The probability of a thunder storm on a day is 20% in the month of April. What is the probability that we shall have :
   (i) Exactly 5 thunder storms in that month ?
   (ii) Exactly 6 thunder storms in that month ?
   (iii) At least 4 thunder storms in that month ?
   (iv) At most 24 thunder storms in that month ?
3. 'Mode' is defined as the value of 'x' with the max. frequency (probability), when the x-values are arranged in order. Find the mode of the Binomial distribution if :
   (i) $p = 1/4$, (ii) $p = 1/2$.
4. A number of identical bottles kept in the open, are collecting rain water. The average Number of drops collected in one bottle is 10 ($Np = 10$). What is the probability that a particular bottle collects 20 drops ?
5. Find the 'mode' (maximum) in a Gaussian Distribution.
6. The 'point of inflection' of a curve is a point where the curvature changes sign. For the curves of well-behaved functions, it requires (as a necessary cond.) that the second derivative of the function vanishes. On the basis of that, find the point of inflection of the Gaussian Distribution Function.