A brief overview of probability

Random or stochastic variables

A random or stochastic variable is a type of variable whose possible values depend on the outcome of a ~~R~~ certain ~~grand~~ random phenomenon. Each random variable posseses a specific probability distribution that represents the probabilities of occurance of all possible outcomes.

We can have two differenet type of random variables.

(1) Discrete random variable: This type of random variables can only take a discrete, finite number of values. For example, ~~a co~~ the outcome of a coin toss contains two possible values - either head or tail. If you associate numbers with this, ~~then~~ say, 1 for H and -1 for tail, then this is a random variable with probability distribution given by,

$$P(1) = \frac{1}{2} \quad \text{and} \quad P(-1) = \frac{1}{2}.$$

Another example can be the number obtained when you throw a die. The random variable associated with this can have any values within $\{1, 2, 3, 4, 5, 6\}$. The probability distribution is -

$$P(1) = \frac{1}{6}, \quad P(2) = \frac{1}{6}, \quad P(3) = \frac{1}{6}, \quad P(4) = \frac{1}{6}, \quad P(5) = \frac{1}{6} \text{ and } P(6) = \frac{1}{6}.$$

Instead of a fair coin, we could have an unfair one, whee,

$$P(1) = 0.3 \quad \text{and} \quad P(-1) = 0.7.$$

All these represents discrete random variables with discrete probability distributions.

If x is a discrete random variable which takes values x_i with probability P_i, then we require,

$$\sum_{i=1}^{N} P_i = 1$$

since all the probabilities must add up to 1.

The mean or average of a discrete random variable x is defined as,

$$\langle x \rangle = \sum_i x_i P_i$$

So, you are giving weight to the value of the random variable by the probability of its occuring.

Remember, the mean value of x may not be a number that the random variable can take. For example, the mean of the coin toss is -

$$\langle x \rangle = 1\times\frac{1}{2} + (-1)\times\frac{1}{2} = 0$$

which is not inclueded in the possible outcomes.

We can define the mean squared value of x using,

$$\langle x^2 \rangle = \sum_i x_i^2 P_i$$

For any function of x, the mean can be defined as -

$$\langle f(x) \rangle = \sum_i f(x_i) P_i$$

We will soon see that the mean squared value of x is important in many calculations.

Example

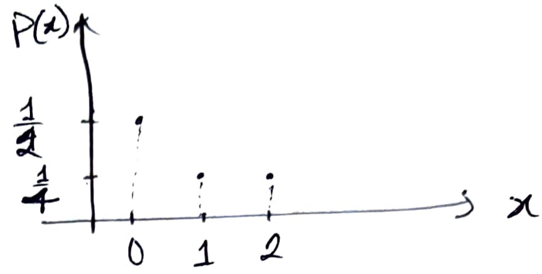
x can take the values 0, 1, 2 with probabilities $\frac{1}{2}, \frac{1}{4}, \frac{1}{4}$. Calculate the mean $\langle x \rangle$ and mean squared value $\langle x^2 \rangle$, and $\langle x^3 \rangle$.
H.W.

$$\langle x \rangle = 0\times\frac{1}{2} + 1\times\frac{1}{4} + 2\times\frac{1}{4} = \frac{1}{4}+\frac{1}{2} = \frac{3}{4}$$

$$\langle x^2 \rangle = 0^2\times\frac{1}{2} + 1^2\times\frac{1}{4} + 2^2\times\frac{1}{4} = \frac{1}{4}+1 = \frac{5}{4}$$

Mean squared value for coin toss is, - $\langle x^2 \rangle = (1)^2\times\frac{1}{2} + (-1)^2\times\frac{1}{2}$
$= 1$

The probability distribution of the example looks like the following graph.



Continuous probability distribution

Continuous random variable

A continuous random variable can take an uncountably infinite number of values. It can take a range of possible values. But since a continuous random variable can take an infinite, number of values, the probability of any one of them occuring is zero. So, we do not talk about the probability of a particular value of x to be $P(x)$. Rather, we say, $P(x)dx$ is the probability of the variable to has a value between x and $x+dx$. In this sense $P(x)$ describes the probability density, rather than exact probability. You then multiply with the interval to find the probability that the value of the variable lies in that interval.

Its like the mass density in mechanics. In 1D, the mass density is ~~defined~~ denoted by λ, which might be a a function of x, so $\lambda(x)$. If you want to calculate the mass between $x=a$ and $x=b$, its just given by,

$$\text{mass} = \int_a^b \lambda(x)\,dx$$

The same is true here, except for the dimension. The dimension of probability density function $P(x)$ is probability/length.

The PDF will always satisfy the following property:

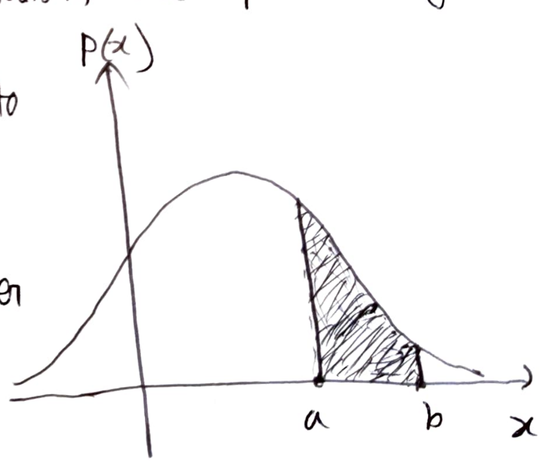
(i) $P(x) \geq 0$ (ii) $\int_{-\infty}^{\infty} P(x)\,dx = 1$

However, since $P(x)$ is not probability there is no restriction that is has to be less than 1, as long as (ii) is satisfied.

The probability that a continuous random variable x has a value between a and b is given by,

$$P(a \leq x \leq b) = \int_a^b P(x)\,dx$$

In terms of graphical representations, the probability of a continuous random variable to has values within a and b is given by the area under the curve of $P(x)$ vs x.



So, the total area under the curve of $P(x)$ vs x will always be equal to 1.

For continuous random variable, the mean is defined as

$$\langle x \rangle = \int x P(x)\, dx$$

Mean squared value, $\langle x^2 \rangle = \int x^2 P(x)\, dx$ and mean of any function of x is given by,

$$\langle f(x) \rangle = \int f(x) P(x)\, dx$$

Example 1

Consider a random variable x within the range $[-0,1]$. and the PDF is given by $P(x) = Cx^2$. Calculate the value of C.

Since the probability is normalized,

$$\int_0^1 P(x)\, dx = 1 \Rightarrow C\int_0^1 x^2 dx = 1$$

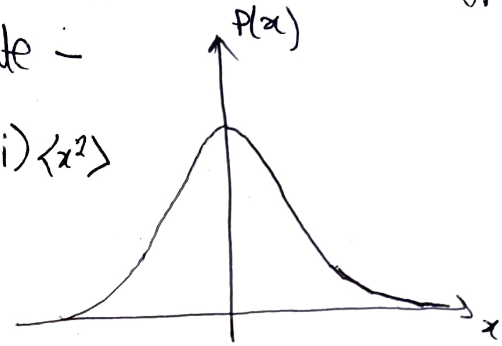
$$\Rightarrow C \left.\frac{x^3}{3}\right|_0^1 = 1 \Rightarrow \frac{C}{3}(1^3 - 0) = 1$$

$$\therefore C = 3$$

Here, the constant C is needed to normalize the probability.

Example

Consider the PDF $P(x) = Ce^{-\frac{x^2}{2a^2}}$ where C and a are constants. The probability density curve of this type is called a Gaussian. Calculate –



(i) C (ii) $\langle x \rangle$ and (iii) $\langle x^2 \rangle$

(i) H.W.

You will find it to be $\frac{1}{\sqrt{2\pi a^2}}$.

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

Now, $$\langle x \rangle = \frac{1}{\sqrt{2\pi a^2}} \int_{-\infty}^{\infty} x e^{-\frac{x^2}{2a^2}} dx$$

Now, $e^{-\frac{x^2}{2a^2}}$ is an even function of x with respect to $x=0$. On the other hand, x is an odd function. The multiplication of an even function with an odd function is again an odd function, and the integral of an odd function over symmetric interval is always zero.

$$\therefore \int_{-\infty}^{\infty} x e^{-\frac{x^2}{2a^2}} dx = 0$$

$\therefore \ \langle x \rangle = 0$

Now,
$$\langle x^2 \rangle = \frac{1}{\sqrt{2\pi a^2}} \int_{-\infty}^{\infty} x^2 e^{-\frac{x^2}{2a^2}} dx$$

Gaussian integrals

$$I = \int_{-\infty}^{\infty} e^{-\alpha x^2} dx = ?$$

Consider the integral first — $\int_{-\infty}^{\infty} e^{-\alpha(x^2+y^2)} dx\, dy$

$$= \int_{-\infty}^{\infty} e^{-\alpha x^2} \cdot e^{-\alpha y^2} dx\, dy = \int_{-\infty}^{\infty} e^{-\alpha x^2} \left(\int e^{-\alpha y^2} dy \right) dx$$

$$= \int_{-\infty}^{\infty} e^{-\alpha x^2} dx \int_{-\infty}^{\infty} e^{-\alpha y^2} dy$$

$$= I \times I = I^2$$

Now, $I^2 = \int_{-\infty}^{\infty} e^{-\alpha(x^2+y^2)} dx\, dy$

In polar coordinate, $x^2+y^2 = r^2$

In polar coordinate, the integral becomes,

$$I^2 = \int e^{-\alpha r^2} r\, dr\, d\theta = \int_0^{2\pi} d\theta \int_0^{\infty} dr\, r e^{-\alpha r^2}$$

Let, $z = \alpha r^2 \Rightarrow \frac{dz}{dr} = 2\alpha r \Rightarrow dz = 2\alpha r dr$

$$\therefore I^2 = \int_0^{2\pi} d\theta \int_0^{\infty} \frac{1}{2\alpha} e^{-z} dz = \frac{1}{2\alpha} \times 2\pi \cdot \left[-e^{-z} \right]_0^{\infty}$$

$$= \frac{\pi}{\alpha}$$

$\therefore I = \sqrt{\frac{\pi}{a}}$.

- Commutes means:
$AB = BA$

$$\therefore \int_{-\infty}^{\infty} e^{-\alpha x^2} dx = \sqrt{\frac{\pi}{\alpha}} \quad \text{——} (1)$$

If we differentiate both sides of (1) with respect to α,

The $\frac{d}{d\alpha}$ and $\int$ commutes

$$\int_{-\infty}^{\infty} -x^2 e^{-\alpha x^2} dx = \sqrt{\pi}\cdot\left(-\frac{1}{2}\right) \alpha^{-\frac{1}{2}-1}$$

Leibnitz integral rule

$$\boxed{\therefore \int_{-\infty}^{\infty} x^2 e^{-\alpha x^2} dx = \frac{1}{2}\sqrt{\frac{\pi}{\alpha^3}}}$$

We can repeat this process to find, $\int_{-\infty}^{\infty} x^4 e^{-\alpha x^2} dx = \frac{3}{4}\sqrt{\frac{\pi}{\alpha^5}}$

and in general,

$$\int_{-\infty}^{\infty} x^{2n} e^{-\alpha x^2} dx = \frac{2n!}{n!\, 2^{2n}} \sqrt{\frac{\pi}{\alpha^{2n+1}}}$$

Since these integrands are even functions, the integral from 0 to ∞ is just half of these.

$$\therefore \int_{0}^{\infty} \cancel{x} e^{-\alpha x^2} dx = \frac{1}{2}\sqrt{\frac{\pi}{\alpha}}$$

$$\int_{0}^{\infty} x^2 e^{-\alpha x^2} dx = \frac{1}{4}\sqrt{\frac{\pi}{\alpha^3}} \quad \text{and so on.}$$

Integrating $x^{2n+1} e^{-\alpha x^2}$ is easy, since the integrand is odd and so the integral from $-\infty$ to $+\infty$ is just 0.

Finding $\int_{0}^{\infty} x e^{-\alpha x^2} dx$ is surprisingly easy.

Let, $+\alpha x^2 = z \Rightarrow \frac{dz}{dx} = +2\alpha x$

$\therefore dz = +2\alpha x \, dx$

$$\therefore I = \int_0^\infty x \cdot e^{-z} \cdot \frac{dz}{2\alpha x} = \frac{1}{2\alpha}\int_0^\infty e^{-z}\, dz$$

$$\therefore I = \frac{1}{2\alpha}$$

$$\therefore \int_0^\infty x e^{-\alpha x^2} dx = \frac{1}{2\alpha}$$

Differentiating both sides w.r.t. α gets us to,

$$\int_0^\infty x^3 e^{-\alpha x^2} dx = \frac{1}{2\alpha^2}$$ and so on. In general,

$$\int_0^\infty x^{2n+1} e^{-\alpha x^2} dx = \frac{n!}{2\alpha^{n+1}}.$$

Back to Example:

$$\langle x^2 \rangle = \frac{1}{\sqrt{2\pi a^2}} \int_{-\infty}^{\infty} x^2 e^{-\frac{x^2}{2a^2}} dx$$

$$= \frac{1}{\sqrt{2\pi a^2}} \times \frac{1}{2}\sqrt{\frac{\pi}{\left(\frac{1}{2a^2}\right)^3}}$$

$$= \frac{1}{\sqrt{2\pi a^2}} \times \frac{1}{2} \times \sqrt{8\pi a^6}$$

$$\boxed{\therefore \langle x^2 \rangle = a^2}$$

Linear transformation

If you have a random variable x, you can find a second random variable y by just performing a linear transformation on x given by,

$$y = ax + b \quad \text{with } a \text{ and } b \text{ being constant.}$$

$$\langle y \rangle = \langle ax+b \rangle = a\langle x \rangle + \langle b \rangle = a\langle x \rangle + b$$

$$\langle ax \rangle = \int_{-\infty}^{\infty} ax\, P(x)\, dx = a \int_{-\infty}^{\infty} x\, P(x)\, dx = a\langle x \rangle .$$

Variance

Say, we have calculated the average of a set of values. How do we then determine the spread of these values from the average? We can define the "deviation" from the mean as –

$$D = x - \langle x \rangle$$

can be considered as a linear transformation

This quantity tells you how much a particular value is above or below the mean value. To get a sense of how the values are spread out, we might want to take the average of the deviation, given by,

$$\langle D \rangle = \langle x - \langle x \rangle \rangle = \langle x \rangle - \langle \langle x \rangle \rangle = \langle x \rangle - \langle x \rangle = 0$$

So, the average deviation gives us no idea about the average spread of values. The problem is that, the deviations are positive sometimes and negative sometimes and they cancells out. So, we need something else. We can consider the modulus of deviation.

$$|D| = |x - \langle x \rangle|$$

But the modulus sign is a bit confusing and not so good to compute algebraic calculations. So, people came up with another quantity, which is always positive like the mod, the square of deviation — $(x - \langle x \rangle)^2$. The average of this square of deviation is given a special name, the variance, or the mean squared deviation, denoted by σ_x^2.

$$\therefore \sigma_x^2 = \langle (x - \langle x \rangle)^2 \rangle \rightarrow$$

Gives us a measure of width of a distribution

The root of this quantity is called the standard deviation or the root mean square (rms) deviation, given by,

$$\sigma_x = \sqrt{\langle (x - \langle x \rangle)^2 \rangle}$$

Now,

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

$$= \langle x^2 \rangle - 2\langle x \rangle \langle x \rangle + \langle x \rangle^2$$

$$\boxed{\therefore \sigma_x^2 = \langle x^2 \rangle - \langle x \rangle^2}$$

We will use this identity often.

In general one can define the k^{th} moment about the mean as $\langle (x-\langle x\rangle)^k \rangle$. The first moment gives the mean deviation, second moment gives variance, third moment is known as skewness etc.

Variance after linear transformation

$$y = ax+b \qquad \langle y\rangle = a\langle x\rangle + b$$

$$\sigma_y^2 = \langle (y-\langle y\rangle)^2 \rangle = \langle (ax+b-a\langle x\rangle-b)^2 \rangle$$

$$= \langle a^2x^2 - 2a^2x\langle x\rangle + a^2\langle x\rangle^2 \rangle$$

$$= a^2\langle x^2\rangle - 2a^2\langle x\rangle\langle x\rangle + a^2\langle x\rangle^2$$

$$= a^2\left(\langle x^2\rangle - \langle x\rangle^2\right) = a^2\sigma_x^2$$

$$\therefore \quad \sigma_y^2 = a^2\sigma_x^2 \quad \text{and} \quad \sigma_y = a\sigma_x$$

So, variance after linear transformation depends on a and not b. Think why.

Independent variables

Two random variables are independent if knowing the value of one of them yields no information about the other. For example, the height of a person chosen at random from a city and the amount of raintall on a particular day in that city are two independent random variables.

For two random variables u and v, the probability that u is in the range from u and u+du and v is in the range from v and v+dv is given by

$$P_u(u)\,du\;P_v(v)\,dv$$

The average of the product uv is then,

$$\langle uv\rangle = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} P_u(u)\,du\;P_v(v)\,dv = \int_{-\infty}^{\infty} P_u(u)\,du \int_{-\infty}^{\infty} P_v(v)\,dv$$

since they are independent.

$$\therefore \langle uv\rangle = \langle u\rangle\langle v\rangle$$

Use of the result above

Consider that you are measuring a quantity x. You repeated the experiment n times, each time with independent rms errors σ_x.

Example

Say, there are n independent random variables X_i, each with mean $\langle x \rangle$ and variance σ_x^2. If Y is the sum of the random variables, ~~then~~ so that, $Y = X_1 + X_2 + \cdots + X_n$, then find the mean and variance of Y.

$$\langle Y \rangle = \langle X_1 \rangle + \langle X_2 \rangle + \cdots + \langle X_n \rangle$$
$$= \langle x \rangle + \langle x \rangle + \cdots + \langle x \rangle = n \langle x \rangle$$

And,

$$\sigma_Y^2 = \langle Y^2 \rangle - \langle Y \rangle^2$$

$$\langle Y^2 \rangle = \langle X_1^2 + X_2^2 + \cdots + X_n^2 + X_1 X_2 + X_2 X_1 + X_1 X_3 + \cdots \rangle$$
$$= \langle X_1^2 \rangle + \langle X_2^2 \rangle + \cdots + \langle X_n^2 \rangle + \langle X_1 \rangle \langle X_2 \rangle + \langle X_2 \rangle \langle X_1 \rangle + \cdots$$
$$= n \langle x^2 \rangle + n(n-1) \langle x \rangle^2$$

this much $\langle x \rangle \langle x \rangle$ terms. check for 3

$$\therefore \sigma_Y^2 = n \langle x^2 \rangle + n(n-1) \langle x \rangle^2 - n^2 \langle x \rangle^2$$
$$= n \langle x^2 \rangle + \cancel{n^2 \langle x \rangle^2} - n \langle x \rangle^2 - \cancel{n^2 \langle x \rangle^2}$$
$$= n \langle x^2 \rangle - n \langle x \rangle^2 .$$

$$\therefore \sigma_Y^2 = n \sigma_x^2$$

$$\boxed{\therefore \sigma_Y = \sqrt{n}\, \sigma_x}$$

Example

Say, in an experiment, a quantity X is measured n times, with an independent rms error of σ_x each time. If you add these up to get $Y = \sum_i X_i$, then the rms error in Y is $= \sqrt{n}\,\sigma_x$. If you now want to get a good estimate of X by calculating $\frac{\sum_i X_i}{n}$, then the rms error is $= \frac{\sqrt{n}\,\sigma_x}{n} = \frac{\sigma_x}{\sqrt{n}}$.

Binomial distribution

Bernoulli trial

Bernoulli trial is an experiment with two possible outcomes — one with probability p and another with 1-p.

success failure

An example is the obvious one — a coin toss.

The binomial distribution is a discrete probability distribution that gives you the probability of k number of successes in a total of n trials.

Now, the probability of k successes $= p^k$
and at the same time of n-k failure $= (1-p)^{n-k}$

But, you can choose these k succeses out of trials in ${}^nC_k$ possible ways. You have to count all of them. The $P(n,k)$ is just the product of all these factors.

$$\therefore P(n,k) = {}^nC_k \, p^k (1-p)^{n-k}$$

Consider a total of three trials of a coin toss. The probability of head is 0.5 and tail is also 0.5. First, think of zero heads,

Zero ~~head~~ head $\rightarrow P(3,0) = (0.5)^0 (0.5)^3 = 0.125$ (only one possible way)

One head $\rightarrow P(3,1) =$ HTT, THT, TTH } 3 Possible ways

$$\therefore P(3,1) = {}^3C_1 (0.5)^1 (0.5)^2 = 0.375$$

$$P(3,2) = \text{HHT, HTH, THH} = {}^3C_2 (0.5)^2 (0.5)^1 = 0.375$$

$$P(3,3) = {}^3C_3 (0.5)^3 (0.5)^0 = 0.125$$

Total, $\sum_{k=0}^{3} P(n,k) = 1$.

So, in general, $P(n,k) = {}^nC_k \, p^k (1-p)^{n-k}$

Again, the binomial theorem of elementary algebra states that,

$$(x+y)^n = \sum_{k=0}^{n} {}^nC_k \, x^k y^{n-k}$$

Using $x = p$ and $y = 1-p$ we see,

$$\sum_{k=0}^{n} {}^{n}C_{k}\, p^{k} (1-p)^{n-k} = (p+1-p)^{n} = 1$$

$$\therefore \sum_{k=0}^{n} {}^{n}C_{k}\, p^{k} (1-p)^{n-k} = 1$$

$$\boxed{\therefore \sum_{k=0}^{n} P(n,k) = 1}$$

Mean and variance of binomial distribution

$$\langle k \rangle = \sum_{k=0}^{n} k\, P(n,k) = \sum_{k=0}^{n} k\, {}^{n}C_{k}\, p^{k} (1-p)^{n-k}$$

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

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

$$= np \sum_{k=0}^{n} {}^{n-1}C_{k-1}\, p^{k-1} (1-p)^{(n-1)-(k-1)} = np\,(p+1-p)^{n-1}$$

$$\therefore \langle k \rangle = np$$

We could reach this result by a different and easy approach ~~of~~ using the idea of independent variables.

Consider each Bernoulli trial to be an independent ~~variable,~~ event (which is obviously true). For a particular trial, the average value ~~is =~~ of k is the probability of getting it, which is $= p$,

For n trials, we know,

$$\langle k \rangle = \langle k_1 \rangle + \langle k_2 \rangle + \cdots + \langle k_n \rangle = p + p + \cdots + p = np$$

For a particular trial, if you associate the success with the number ~~zero~~ 1 and failure as 0, then, for one trial, $\langle k \rangle = 1 \times P + 0 \times (1-P) = P$

The variance for a Bernoulli trial, $\cancel{\sigma_k^2 = 1^2 \times P +}$

$$\sigma^2 = \langle k^2 \rangle - \langle k \rangle^2$$

$$= \left[1^2 \times P + 0^2 \times (1-P)\right] - P^2 = P - P^2 = P(1-P)$$

$\therefore$ Variance of the Binomial distribution is –

$$\sigma_k^2 = n\sigma = nP(1-P)$$