Contents

Probability Distributions..cont...

G.N.Dar

July 10, 2020

0.1 Normal Distribution (Continous Distribution)

This is a Probability Distribution $f(x)$ applied to a single variable but Continous in nature say heights of students in class, weights of new born babies, time taken to complete the task, typing speed of students in the class etc.

Mathematically the Normal Distribution is given as

$$f(x) = \frac{1}{\sigma.\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$$

where σ is the variance and μ is the mean of the distribution.

Some points about Normal Distribution refere figure1 are as under

1. It is a bell shaped, symmetric about the means μ
2. As $n \to \infty, \quad x \to \infty, \quad p \neq\to 0$
3. classified by just two parameters μ and σ ; representing location (center) and spread.
4. approximately 68% data fall within $\pm 1\sigma, 95\%$ between $\pm 2\sigma$ and 99.7% between $\pm 3\sigma$
5. virtuall all data fall within $\pm 3\sigma$.
6. $$f(x) = \frac{1}{\sigma.\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx = 1$$
7. $$P(X \geq \mu) = 0.5 = 50\%$$

   Mathematically

   $$f(x) = \frac{1}{\sigma.\sqrt{2\pi}} \int_{\mu}^{\infty} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx = 0.5 = 50\%$$

   $$P(\mu - \sigma \leq X \leq \mu + \sigma) = 0.68 = 68\%$$

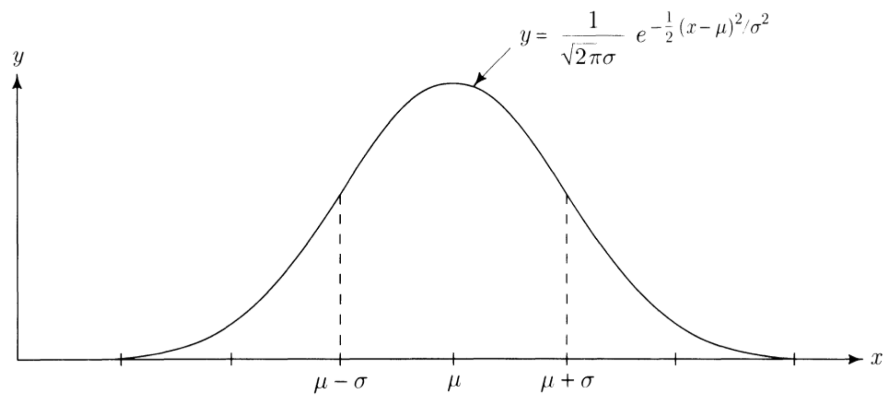

Figure 1: Normal Distribution
??

$$f(x) = \frac{1}{\sigma.\sqrt{2\pi}} \int_{\mu-\sigma}^{\mu+\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx = 0.68 = 68\%$$

$$P(\mu - 2\sigma \leq X \leq \mu + 2\sigma) = 0.95 = 95\%$$

$$f(x) = \frac{1}{\sigma.\sqrt{2\pi}} \int_{\mu-2\sigma}^{\mu+2\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx = 0.95 = 95\%$$

$$P(\mu - 3\sigma \leq X \leq \mu + 3\sigma) = 0.997 = 99.7\%$$

$$f(x) = \frac{1}{\sigma.\sqrt{2\pi}} \int_{\mu-3\sigma}^{\mu+3\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx = 0.997 = 99.7\%$$

8. shifting origin to mean that is taking $mu = 0$ we have

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

0.2 Derivation of Normal Distribution from Poisson Distribution)

We know

$$P(x; \lambda) = \frac{e^{-\lambda}\lambda^x}{x!}$$
$$np = \lambda$$

finite in Poisson distribution while as here $n \to \infty$ so λ is very very large here so will be the case with randome variable x. Taking natural lof on both side we have

$$lnP(x; \lambda) = \ln e^{-\lambda} + \ln \lambda^x - \ln x!$$

$$= -\lambda + x ln\lambda - ln\left[\sqrt{2\pi x} \quad \left(\frac{x}{e}\right)^x\right]$$

using stirling formula we have

$$lnP(x;\lambda) = -\lambda + x ln\lambda - x ln\left(\frac{x}{e}\right) - ln\sqrt{2\pi x}$$

Let us choose the devietion from the mean as new variable as $y = x - \lambda$ we have

$$y + \lambda = x$$

as

$$\lambda \to \infty$$

treat $\frac{y}{\lambda}$ as small but $\frac{y^2}{\lambda}$ finite, with this assumption we have

$$lnP(x;\lambda) = -\lambda + (y+\lambda)ln\lambda - (y+\lambda)ln(y+\lambda) + (y+\lambda).1 - ln\sqrt{2\pi(y+\lambda)}$$

treat $\frac{y}{\lambda}$ as small we have

$$= y + (y+\lambda)ln\left[1 - \frac{y}{y+\lambda}\right] - ln\sqrt{2\pi\lambda}$$

$$lnP(x;\lambda) = y + (y+\lambda)\left[-\frac{y}{y+\lambda} - \frac{y^2}{2(y+\lambda)^2}\right] - ln\sqrt{2\pi\lambda}$$

$$= -\frac{y^2}{2(y+\lambda)} - ln\sqrt{2\pi\lambda}$$

$$lnP(x;\lambda) = -\frac{y^2}{2\lambda} - ln\sqrt{2\pi\lambda}$$

$$lnP(x;\lambda) + ln\sqrt{2\pi\lambda} = -\frac{y^2}{2\lambda}$$

$$lnP(x;\lambda)\sqrt{2\pi\lambda} = -\frac{y^2}{2\lambda}$$

$$P(x;\lambda)\sqrt{2\pi\lambda} = e^{-\frac{y^2}{2\lambda}}$$

$$P(x;\lambda) = \frac{1}{\sqrt{2\pi\lambda}}e^{-\frac{y^2}{2\lambda}}$$

This is the profile of Normal Distribution in the Continous variable y with mean 0 and standar devietion $\sigma = \sqrt{\lambda}$

with mean μ and varaince σ in the variable x it can be written as

$$P(x;\lambda) = N(x,\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\ \sigma^2}}$$

0.3 Expectation Value E(X)

We know

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

Shift the orign by an amount μ that is $x \to x + \mu$, the profile remains invaraint and we have

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

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

we can solve them as

$$E(X) = I_1 + I_2$$

Now

$$I_2 = \int_{-\infty}^{\infty} \mu \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\ \sigma^2}}\, dx$$

$$I_2 = \frac{\mu}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2\ \sigma^2}}\, dx$$

It is of the type

$$I_2 = \frac{\mu}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-ax^2}\, dx$$

Therefore using Gaussian Integral we have

$$I_2 = \frac{\mu}{\sigma\sqrt{2\pi}} \sqrt{\frac{\pi}{a}}$$

Here $a = \frac{1}{2\sigma^2}$
Therefore

$$I_2 = \mu$$

which is the mean or expectation value of the normal distribution
We just need to see the other Integral I_1 is just 0 and we are done.
from the formate it is easier to understand the function is odd one from $-\infty$ to ∞ and hence the result is surely zero.
we can do in few steps as

$$I_1 = \int_{\infty}^{\infty} x \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\ \sigma^2}}\, dx$$

$$I_1 = \int_{-\infty}^{0} x \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\ \sigma^2}}\, dx + \int_{0}^{\infty} x \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\ \sigma^2}}\, dx$$

switch over limits in first part we have

$$I_1 = -\int_{0}^{-\infty} x \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\ \sigma^2}}\, dx + \int_{0}^{\infty} x \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\ \sigma^2}}\, dx$$

Now replece x by $-x$ again in ist part we have

$$I_1 = -\int_0^{\infty} (-x) \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\ \sigma^2}} (-dx) + \int_0^{\infty} x \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\ \sigma^2}} dx = 0$$

Therefore

$$E(X) = I_1 + I_2 = \mu$$

0.4 Mode of Normal Distribution

It is the value of random variable x for which $f(x)$ is maximum.

For maximum we have

$$f'(x) = 0$$

and

$$f''(x) < 0$$

we know for $X\ N(\mu, \sigma^2)$ that

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

take logrith

$$lnf(x) = c + ln[e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}] = c - \frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2$$

differential with respect x

$$\frac{1}{f(x)}.f'(x) = 0 - \frac{(x-\mu)}{\sigma^2}$$

$$f'(x) = 0 - \frac{(x-\mu)}{\sigma^2} f(x)$$

Again differentiate

$$f''(x) = -\frac{1}{\sigma^2}\left[1.f(x) + (x-\mu)f'(x)\right]$$

substiture value of $f'(x)$ we have

$$f''(x) = -\frac{1}{\sigma^2} f(x)\left[1 - \frac{(x-\mu)^2}{\sigma^2}\right]$$

Now for mode $f'(x) = 0$ since pdf $f(x) \neq 0$ we have

$$x = \mu$$

to confirm it is mode we need to see $f''(x = \mu) <)$
put $x = \mu$ we have

$$f''(\mu) = -\frac{1}{\sigma^2} f(x)$$

As $f(x) > 0$ at $x = \mu$ and σ^2 is always positive, therefore for sure

$$f''(x = \mu) < 0$$

Hence the mode of the distribution is again at μ

0.5 Median of Normal Distribution M

:

If M is the median then it divides the curve into two halves that is

$$\int_{-\infty}^{M} f(x)dx = \int_{-\infty}^{M} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\,\sigma^2}} dx = \frac{1}{2}$$

Let some μ lies in between $-\infty$ to M we have

$$\int_{-\infty}^{\mu} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\,\sigma^2}} dx + \int_{\mu}^{M} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\,\sigma^2}} dx = \frac{1}{2}$$

from mena value the first part is simply $\frac{1}{2}$

$$\frac{1}{2} + \int_{\mu}^{M} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\,\sigma^2}} dx = \frac{1}{2}$$

therefore

$$\int_{\mu}^{M} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\,\sigma^2}} dx = \frac{1}{2} - \frac{1}{2} = 0$$

this is possible only if

$$\mu = M$$

Hence proves the meadian is $M = \mu$
Conclusion
The mean mode and median of the normal distribution coincide and are equal to μ

$$Mean = Mode = Median = \mu$$

0.6 Standard Normal Distribution

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

Trick is to shift the origin or replece x as $x \rightarrow x + \mu$

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

Now use Gaussian Integral formula we have as

$$\int_{-\infty}^{\infty} x^2 e^{-ax^2} dx = 1.\frac{\sqrt{\pi}}{2a^{\frac{3}{2}}}$$

$$\sigma^2 = \frac{1}{\sigma.\sqrt{2\pi}} * 1.\frac{\sqrt{\pi}}{2\left(\frac{1}{2\sigma^2}\right)^{\frac{3}{2}}}$$

Solve this and find

$$\sigma^2 = \sigma^2$$

0.7 Some problems on Normal Distribution

copy pasted from (Library, Teaching and Learning-Lincoln University New Zealand)

1. Potassium blood levels in healthy humans are normally distributed with a mean of 17.0 mg /100 ml, and standard deviation of 1.0 mg/100 ml. Elevated levels of potassium indicate an electrolyte balance problem, such as may be caused by Addison's disease. However, a test for potassium level should not cause too many "false positives".

a) What level of potassium should we use so that only 2.5% of healthy individuals are classified as "abnormally high.

2. For a particular type of wool the number of 'crimps per 10 cm' follows a normal distribution with mean 15.1 and standard deviation 4.79.

(a) What proportion of wool would have a 'crimp per 10 cm' measurement of 6 or less?

(b) If more than 7% of the wool has a 'crimp per 10 cm' measurement of 6 or less, then the wool is unsatisfactory for a particular processing. Is the wool satisfactory for this processing?

3. The finish times for marathon runners during a race are normally distributed with a mean of195 minutes and a standard deviation of 25 minutes.

a) What is the probability that a runner will complete the marathon within 3 hours? b) Calculate to the nearest minute, the time by which the first 8% runners have completed the marathon. c) What proportion of the runners will complete the marathon between 3 hours and 4 hours?

4. The download time of a resource web page is normally distributed with a mean of 6.5 seconds and a standard deviation of 2.3 seconds.

a) What proportion of page downloads take less than 5 seconds?

b) What is the probability that the download time will be between 4 and 10 seconds? c) How many seconds will it take for 35% of the downloads to be completed?