THE NORMAL DISTRIBUTION

The *Normal* or *Gaussian* distribution is the most commonly used distribution in statistics, and is a model for random variables/obervations taking values on the whole of $\mathbb{R}$. The functional form of the Normal pdf is constructed from the non-negative, integrable function $f(x) = \exp\left\{-x^2\right\}$, but it has deeper origins in empirical data analysis, and can be derived as a limiting pdf.

First, we compute the integral of the function $f(x)$ Suppose that, say,

$$\int_{-\infty}^{\infty} \exp\left\{-x^2\right\}\,dx = c$$

(we know by inspection that f is integrable, and $c > 0$). Then

$$\begin{aligned} c &= \int_{-\infty}^{\infty} \exp\left\{-x^2\right\}\,dx \\ &= \frac{1}{c}\int_{-\infty}^{\infty} c\exp\left\{-x^2\right\}\,dx \\ &= \frac{1}{c}\int_{-\infty}^{\infty}\left\{\int_{-\infty}^{\infty} \exp\left\{-y^2\right\}\,dy\right\}\exp\left\{-x^2\right\}\,dx \\ &= \frac{1}{c}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp\left\{-\left(x^2+y^2\right)\right\}\,dy\,dx \end{aligned}$$

Now, we can make a transformation to polar coordinates in this double integral, that is, set

$$\begin{aligned} x &= r\cos\theta \\ y &= r\sin\theta \end{aligned}$$

for which the Jacobian or "change of variables" term is the determinant

$$\left|\left[\begin{array}{cc} \dfrac{\partial x}{\partial r} & \dfrac{\partial x}{\partial \theta} \\ \dfrac{\partial y}{\partial r} & \dfrac{\partial y}{\partial \theta} \end{array}\right]\right| = \left|\left[\begin{array}{cc} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{array}\right]\right| = r\cos^2\theta + r\sin^2\theta = r$$

so that the integral above becomes

$$\begin{aligned} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp\left\{-\left(x^2+y^2\right)\right\}\,dy\,dx &= \int_{0}^{\infty}\int_{-\pi}^{\pi} \exp\left\{-\left(r^2\cos^2\theta + r^2\sin^2\theta\right)\right\}\,r\,d\theta\,dr \\ &= \int_{0}^{\infty}\int_{-\pi}^{\pi} \exp\left\{-r^2\right\}\,r\,d\theta\,dr \\ &= \left\{\int_{0}^{\infty} \exp\left\{-r^2\right\}\,r\,dr\right\}\left\{\int_{-\pi}^{\pi} d\theta\right\} \\ &= \left\{\left[-\frac{1}{2}\exp\left\{-r^2\right\}\right]_0^{\infty}\right\}\{2\pi\} \\ &= \pi \end{aligned}$$

and hence we deduce that

$$c = \frac{\pi}{c} \qquad \text{so that} \qquad c = \sqrt{\pi}$$

Hence we conclude that the specification

$$f_X(x) = \frac{1}{c}\exp\left\{-x^2\right\} = \sqrt{\frac{1}{\pi}}\exp\left\{-x^2\right\} \qquad x \in \mathbb{R}$$

is a valid pdf. Note also from the above that, using another change of variables in the integral, we have that for $\lambda > 0$

$$\int_{-\infty}^{\infty} \exp\left\{-\lambda x^2\right\} \, dx = \sqrt{\frac{\pi}{\lambda}}$$

and further, for any constant μ

$$\int_{-\infty}^{\infty} \exp\left\{-\lambda (x-\mu)^2\right\} \, dx = \sqrt{\frac{\pi}{\lambda}}$$

The *standard Normal* pdf is obtained in this formulation when $\lambda = 1/2$ and $\mu = 0$, that is, for random variable X

$$f_X(x) = \sqrt{\frac{1}{2\pi}}\exp\left\{-\frac{1}{2}x^2\right\}$$

The general Normal pdf is obtained by a *location/scale* transformation of X where we set random variable $Y = \mu + \sigma X$ for $\sigma > 0$, so that

$$F_Y(y) = \mathrm{P}\left[Y \leq y\right] = \mathrm{P}\left[\mu + \sigma X \leq y\right] = \mathrm{P}\left[X \leq \frac{y-\mu}{\sigma}\right] = F_X\left(\frac{y-\mu}{\sigma}\right)$$

from which we obtain on differentiation on both sides

$$f_Y(y) = \frac{1}{\sigma} f_X\left(\frac{y-\mu}{\sigma}\right) = \sqrt{\frac{1}{2\pi\sigma^2}}\exp\left\{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2\right\}$$

Note that the pdf is symmetric about μ.

The expectation of Y is computed as follows:

$$\mathrm{E}_{f_Y}\left[Y\right] = \int_{-\infty}^{\infty} y\sqrt{\frac{1}{2\pi\sigma^2}}\exp\left\{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2\right\} \, dy$$

which, by transforming to $t = (y-\mu)/\sigma$ in the integral becomes

$$\begin{aligned}
\mathrm{E}_{f_Y}\left[Y\right] &= \textstyle\int_{-\infty}^{\infty} (\mu + \sigma t)\sqrt{\frac{1}{2\pi}}\exp\left\{-\frac{1}{2}t^2\right\} \, dt \\
&= \mu \int_{-\infty}^{\infty} \exp\left\{-\tfrac{1}{2}t^2\right\} \, dt + \sigma\sqrt{\frac{1}{2\pi}}\int_{-\infty}^{\infty} t\exp\left\{-\tfrac{1}{2}t^2\right\} \, dt \\
&= \mu + \sigma\sqrt{\frac{1}{2\pi}}\left[-\exp\left\{-\tfrac{1}{2}t^2\right\}\right]_{-\infty}^{\infty} \\
&= \mu
\end{aligned}$$

This result is intuitively reasonable; the pdf is symmetric about μ, and the expectation of a random variable is the "centre of probability".