MA181 INTRODUCTION TO STATISTICAL MODELLING
NORMAL DISTRIBUTION

DIAGRAMS

Origins The normal distribution was discovered, in a discrete form, by de Moivre in 1733 as an approximation to the binomial distribution. It was later shown, in 1812, to be the limiting distribution of a sample mean by Laplace. Meanwhile, in 1809, Gauss derived the normal as the distribution of errors in astronomical observations.

Formulation Let Y be a random variable with the probability density function (pdf)

$$f(y) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(y-\mu)^2}{2\sigma^2}\right], \ -\infty < y < \infty.$$

Then Y is said to follow a *normal* (or *Gaussian*)*distribution* with parameters μ and σ^2, the shorthand for which is $Y \sim N(\mu, \sigma^2)$. The cumulative distribution function (cdf) of Y is

$$F(y) = P(Y \leq y = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{y} \exp\left[-\frac{(w-\mu)^2}{2\sigma^2}\right] dw, \ -\infty < y < \infty.$$

Clearly, $f(-\infty) = 0$, and it can be shown that $F(\infty) = 1$.

Moments The moment generating function of Y is given by

$$\begin{aligned} M(t) &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{ty} \exp\left[-\frac{(y-\mu)^2}{2\sigma^2}\right] dy \\ &= e^{ut+\frac{1}{2}\sigma^2 t^2} \times \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left\{-\frac{[y-(\mu+\sigma^2 t)]^2}{2\sigma^2}\right\} dy \\ &= e^{\mu t+\frac{1}{2}\sigma^2 t^2} \\ &= 1 + (\mu t + \frac{1}{2}\sigma^2 t^2) + \frac{1}{2}(\mu t + \frac{1}{2}\sigma^2 t^2)^2 + \ldots \end{aligned}$$

Consequently, $E(Y) = \mu$ and $E(Y^2) = \sigma^2 + \mu^2$ so that $\text{var}(Y) = \sigma^2$ and σ is the standard deviation of Y.

As the normal distribution is symmetric $[f(\mu - c) = f(\mu + c)$ for all $c]$, all its odd central moments are zero. On the other hand, $\mu_4 = E(Y-\mu)^4] = 3\sigma^4$.

Probabilities Whatever the values of μ and σ, the following properties hold:

> $\mu \mp \sigma$ contains 68.27% of the distribution, $\mu \mp 1.960\sigma$ contains 95%, $\mu \mp 2.576\sigma$ 99% and $\mu\, mp 3.291\sigma$ 99.9%.

Standardisation If $Y' = a + bY$, then $Y' \sim N(a + b\mu, b^2\sigma^2)$. Hence. if $Z = (Y - \mu)/\sigma$, then $Z \sim N(0, 1)$, which is known as the *standard normal distribution* and is the only one tabulated. The pdf of Z is often denoted by $\phi(x)$ and the cdf by $\phi(z)$.