Chapter 14

The Normal Distribution

Lecture 14 | Devore 9e: 4.3

Recall the continuous toolkit from last lecture:

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x f(x)\,\mathrm{d}x, \quad \mathbb{E}[h(X)] = \int_{-\infty}^{\infty} h(x) f(x)\,\mathrm{d}x, \quad \mathrm{Var}(X) = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\,\mathrm{d}x = \mathbb{E}[X^2]-\mu^2.$$

14.1 Why this distribution and not another

Take $X_1, X_2, \ldots, X_N$ independent and identically distributed — *whatever* their common distribution — and form the average

$$Z = \frac{1}{N}\sum_{i=1}^{N} X_i.$$

As N grows, the distribution of Z looks like the same bell-shaped curve, no matter what the X_i were. Complicated inputs, one universal output. That is the central limit theorem (Lecture 21), and it is the reason the normal distribution is everywhere: any quantity that is the accumulation of many small independent effects — measurement error, height, exam totals — comes out approximately normal.

14.2 Definition

Definition 14.1 (Normal distribution)**.** $X \sim \mathrm{Normal}(\mu, \sigma^2)$ if its pdf is

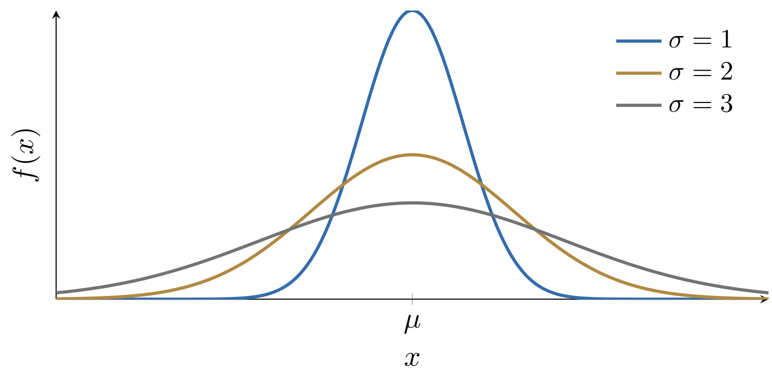
$$f(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi}\,\sigma}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}} = \frac{1}{\sqrt{2\pi}\,\sigma}\,\exp\left(-\frac{1}{2}\Big(\frac{x-\mu}{\sigma}\Big)^2\right), \qquad x \in \mathbb{R},$$

where $\mu \in \mathbb{R}$ and $\sigma > 0$.

Three facts to check (all are Homework):

$$\int_{-\infty}^{\infty} f(x;\mu,\sigma)\,\mathrm{d}x = 1, \qquad \mathbb{E}[X] = \mu, \qquad \mathrm{Var}(X) = \sigma^2.$$

The factor $\frac{1}{\sqrt{2\pi}\sigma}$ is exactly the *normalization* needed to make the total area 1; the shape is carried entirely by $\exp\big(-\frac{1}{2}z^2\big)$ with $z = \frac{x-\mu}{\sigma}$.



The curve is symmetric about μ (so the mean, median and mode all coincide there), and larger σ means more spread. Note that the peak height is $\frac{1}{\sqrt{2\pi}\sigma}$, so a very small σ gives a very tall narrow spike — again a reminder that a density is not a probability and may exceed 1.

14.3 The standard normal, and standardization

Definition 14.2 (Standard normal)**.** $Z \sim \text{Normal}(0, 1)$, with pdf and cdf

$$\varphi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}, \qquad \Phi(z) = \mathbb{P}\{Z \le z\} = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\,\mathrm{d}t.$$

Remark 14.3. Φ has no closed form in elementary functions — you cannot antidifferentiate $e^{-t^2/2}$ by hand. That is why the standard normal table exists, and why every normal computation is routed through it.

Theorem 14.4 (Standardization)**.** *If* $X \sim \text{Normal}(\mu, \sigma^2)$ *and*

$$Z := \frac{X - \mu}{\sigma},$$

then $Z \sim \text{Normal}(0, 1)$.

Proof. $F_Z(z) = \mathbb{P}\{\frac{X-\mu}{\sigma} \le z\} = \mathbb{P}\{X \le \mu + \sigma z\} = F_X(\mu + \sigma z)$. Differentiate in z: $f_Z(z) = \sigma f_X(\mu + \sigma z) = \sigma \cdot \frac{1}{\sqrt{2\pi}\sigma} e^{-z^2/2} = \varphi(z)$. □

This single fact reduces every normal question to a table lookup:

$$\boxed{\mathbb{P}\{X \le x\} = \mathbb{P}\left\{\frac{X - \mu}{\sigma} \le \frac{x - \mu}{\sigma}\right\} = \Phi\left(\frac{x - \mu}{\sigma}\right)}$$

and consequently

$$\mathbb{P}\{a \le X \le b\} = \Phi\left(\frac{b - \mu}{\sigma}\right) - \Phi\left(\frac{a - \mu}{\sigma}\right).$$

Remark 14.5 (Reading the table backwards)**.** Percentiles work the same way. If z_p solves $\Phi(z_p) = p$, then the p-th percentile of X is $x_p = \mu + \sigma z_p$. Values worth memorizing:

$$z_{0.90} = 1.28, \qquad z_{0.95} = 1.645, \qquad z_{0.975} = 1.96, \qquad z_{0.99} = 2.33.$$

By symmetry $\Phi(-z) = 1 - \Phi(z)$, which is how negative arguments are handled.

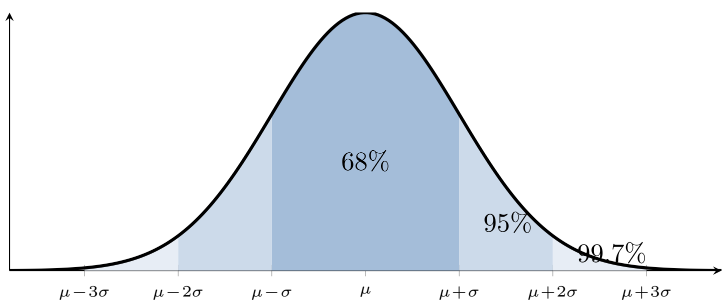
Example 14.6. Exam scores are Normal$(72, 8^2)$. What fraction of students score between 60 and 90?

$$\mathbb{P}\{60 \le X \le 90\} = \Phi\left(\frac{90-72}{8}\right) - \Phi\left(\frac{60-72}{8}\right) = \Phi(2.25) - \Phi(-1.5) = 0.9878 - 0.0668 = 0.9210.$$

14.4 The empirical rule

Theorem 14.7. *If* $X \sim \text{Normal}(\mu, \sigma^2)$ *then*

$$\begin{aligned} \mathbb{P}\{\mu - \sigma \le X \le \mu + \sigma\} &\approx 68\%, \\ \mathbb{P}\{\mu - 2\sigma \le X \le \mu + 2\sigma\} &\approx 95\%, \\ \mathbb{P}\{\mu - 3\sigma \le X \le \mu + 3\sigma\} &\approx 99.7\%. \end{aligned}$$



Remark 14.8. This is worth internalizing because it turns σ into an interpretable number. If a process has $\mu = 60$ and $\sigma = 10$, then roughly 95% of output lies in $[40, 80]$, and a value of 100 is a four-sigma event — about 1 in 30,000. The rule is also the fastest sanity check on an exam: if your answer says $\mathbb{P}\{X > \mu + \sigma\} = 0.40$, you have made an arithmetic error, because it must be about 0.16.

14.5 Looking ahead: the normal approximation to the binomial

If $X \sim \text{Binomial}(n, p)$ then $\mu = np$ and $\sigma^2 = np(1-p)$. For large n the binomial pmf traces out a normal curve, and

$$\mathbb{P}\{X \le x\} \approx \Phi\left(\frac{x + 0.5 - np}{\sqrt{np(1-p)}}\right).$$

The $+0.5$ is the **continuity correction**: we are approximating a lattice-valued variable by a continuous one, so the discrete point x is represented by the interval $[x - 0.5,\, x + 0.5]$. The usual rule of thumb is that the approximation is good when $np \ge 10$ and $n(1-p) \ge 10$.

We will see in Lecture 21 that this is just the central limit theorem applied to $X = \mathbf{1}_1 + \cdots + \mathbf{1}_n$, where $\mathbf{1}_i$ is the indicator of a success on the i-th trial: a binomial count *is* a sum of n i.i.d. terms.