About Chapter 1

In the first chapter, you will need to be familiar with the binomial distribution. And to solve the exercises in the text – which I urge you to do – you will need to know *Stirling's approximation* for the factorial function, $x! \simeq x^x \, e^{-x}$, and be able to apply it to $\binom{N}{r} = \frac{N!}{(N-r)!\,r!}$. These topics are reviewed below.

Unfamiliar notation?
See Appendix A, p.598.

The binomial distribution

Example 1.1. A bent coin has probability f of coming up heads. The coin is tossed N times. What is the probability distribution of the number of heads, r? What are the mean and variance of r?

Solution. The number of heads has a binomial distribution.

$$P(r \mid f, N) = \binom{N}{r} f^r (1-f)^{N-r}. \tag{1.1}$$

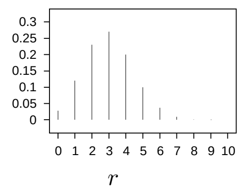

Figure 1.1. The binomial distribution $P(r \mid f = 0.3,\, N = 10)$.

The mean, $\mathcal{E}[r]$, and variance, $\mathrm{var}[r]$, of this distribution are defined by

$$\mathcal{E}[r] \equiv \sum_{r=0}^{N} P(r \mid f, N)\, r \tag{1.2}$$

$$\begin{aligned} \mathrm{var}[r] &\equiv \mathcal{E}\left[(r - \mathcal{E}[r])^2\right] & (1.3) \\ &= \mathcal{E}[r^2] - (\mathcal{E}[r])^2 = \sum_{r=0}^{N} P(r \mid f, N) r^2 - (\mathcal{E}[r])^2 \,. & (1.4) \end{aligned}$$

Rather than evaluating the sums over r in (1.2) and (1.4) directly, it is easiest to obtain the mean and variance by noting that r is the sum of N *independent* random variables, namely, the number of heads in the first toss (which is either zero or one), the number of heads in the second toss, and so forth. In general,

$$\begin{array}{rcll} \mathcal{E}[x+y] &=& \mathcal{E}[x] + \mathcal{E}[y] & \text{for any random variables } x \text{ and } y; \\ \mathrm{var}[x+y] &=& \mathrm{var}[x] + \mathrm{var}[y] & \text{if } x \text{ and } y \text{ are independent.} \end{array} \tag{1.5}$$

So the mean of r is the sum of the means of those random variables, and the variance of r is the sum of their variances. The mean number of heads in a single toss is $f \times 1 + (1-f) \times 0 = f$, and the variance of the number of heads in a single toss is

$$\left[f \times 1^2 + (1-f) \times 0^2\right] - f^2 = f - f^2 = f(1-f), \tag{1.6}$$

so the mean and variance of r are:

$$\mathcal{E}[r] = Nf \qquad \text{and} \qquad \mathrm{var}[r] = Nf(1-f). \qquad \Box \tag{1.7}$$

Approximating $x!$ and $\binom{N}{r}$

Let's derive Stirling's approximation by an unconventional route. We start from the Poisson distribution with mean λ,

$$P(r \mid \lambda) = e^{-\lambda}\frac{\lambda^r}{r!} \qquad r \in \{0, 1, 2, \ldots\}. \tag{1.8}$$

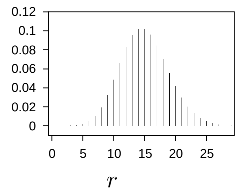

Figure 1.2. The Poisson distribution $P(r \mid \lambda = 15)$.

For large λ, this distribution is well approximated – at least in the vicinity of $r \simeq \lambda$ – by a Gaussian distribution with mean λ and variance λ:

$$e^{-\lambda}\frac{\lambda^r}{r!} \;\simeq\; \frac{1}{\sqrt{2\pi\lambda}}\, e^{-\frac{(r-\lambda)^2}{2\lambda}}. \tag{1.9}$$

Let's plug $r = \lambda$ into this formula, then rearrange it.

$$e^{-\lambda}\frac{\lambda^\lambda}{\lambda!} \;\simeq\; \frac{1}{\sqrt{2\pi\lambda}} \tag{1.10}$$

$$\Rightarrow \quad \lambda! \;\simeq\; \lambda^\lambda\, e^{-\lambda}\sqrt{2\pi\lambda}. \tag{1.11}$$

This is Stirling's approximation for the factorial function.

$$x! \;\simeq\; x^x\, e^{-x}\sqrt{2\pi x} \quad\Leftrightarrow\quad \ln x! \;\simeq\; x\ln x - x + \tfrac{1}{2}\ln 2\pi x. \tag{1.12}$$

We have derived not only the leading order behaviour, $x! \simeq x^x\, e^{-x}$, but also, at no cost, the next-order correction term $\sqrt{2\pi x}$. We now apply Stirling's approximation to $\ln\binom{N}{r}$:

$$\ln\binom{N}{r} \;\equiv\; \ln\frac{N!}{(N-r)!\,r!} \;\;\simeq\;\; (N-r)\ln\frac{N}{N-r} + r\ln\frac{N}{r}. \tag{1.13}$$

Since all the terms in this equation are logarithms, this result can be rewritten in any base. We will denote natural logarithms ($\log_e$) by 'ln', and logarithms to base 2 ($\log_2$) by 'log'.

Recall that $\log_2 x = \dfrac{\log_e x}{\log_e 2}$.

Note that $\dfrac{\partial \log_2 x}{\partial x} = \dfrac{1}{\log_e 2}\dfrac{1}{x}$.

If we introduce the *binary entropy function*,

$$H_2(x) \equiv x\log\frac{1}{x} + (1-x)\log\frac{1}{(1-x)}, \tag{1.14}$$

then we can rewrite the approximation (1.13) as

$$\log\binom{N}{r} \;\simeq\; N H_2(r/N), \tag{1.15}$$

or, equivalently,

$$\binom{N}{r} \;\simeq\; 2^{N H_2(r/N)}. \tag{1.16}$$

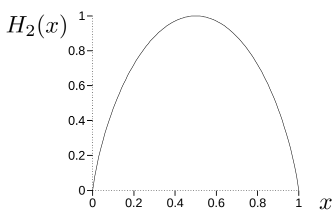

Figure 1.3. The binary entropy function.

If we need a more accurate approximation, we can include terms of the next order from Stirling's approximation (1.12):

$$\log\binom{N}{r} \;\simeq\; N H_2(r/N) - \tfrac{1}{2}\log\left[2\pi N\,\frac{N-r}{N}\,\frac{r}{N}\right]. \tag{1.17}$$