is normalized and that it has mean and variance given by

$$\begin{aligned}\mathbb{E}[x] &= \mu \qquad (3.3)\\ \operatorname{var}[x] &= \mu(1-\mu). \qquad (3.4)\end{aligned}$$

Now suppose we have a data set $\mathcal{D} = \{x_1, \ldots, x_N\}$ of observed values of x. We can construct the likelihood function, which is a function of μ, on the assumption that the observations are drawn independently from $p(x|\mu)$, so that

$$p(\mathcal{D}|\mu) = \prod_{n=1}^{N} p(x_n|\mu) = \prod_{n=1}^{N} \mu^{x_n}(1-\mu)^{1-x_n}. \qquad (3.5)$$

We can estimate a value for μ by maximizing the likelihood function or equivalently by maximizing the logarithm of the likelihood, since the log is a monotonic function. The log likelihood function of the Bernoulli distribution is given by

$$\ln p(\mathcal{D}|\mu) = \sum_{n=1}^{N} \ln p(x_n|\mu) = \sum_{n=1}^{N} \{x_n \ln \mu + (1-x_n)\ln(1-\mu)\}. \qquad (3.6)$$

Section 3.4

At this point, note that the log likelihood function depends on the N observations x_n only through their sum $\sum_n x_n$. This sum provides an example of a *sufficient statistic* for the data under this distribution. If we set the derivative of $\ln p(\mathcal{D}|\mu)$ with respect to μ equal to zero, we obtain the maximum likelihood estimator:

$$\mu_{\mathrm{ML}} = \frac{1}{N}\sum_{n=1}^{N} x_n, \qquad (3.7)$$

which is also known as the *sample mean*. Denoting the number of observations of $x = 1$ (heads) within this data set by m, we can write (3.7) in the form

$$\mu_{\mathrm{ML}} = \frac{m}{N} \qquad (3.8)$$

so that the probability of landing heads is given, in this maximum likelihood framework, by the fraction of observations of heads in the data set.

3.1.2 Binomial distribution

We can also work out the distribution for the binary variable x of the number m of observations of $x = 1$, given that the data set has size N. This is called the *binomial* distribution, and from (3.5) we see that it is proportional to $\mu^m(1-\mu)^{N-m}$. To obtain the normalization coefficient, note that out of N coin flips, we have to add up all of the possible ways of obtaining m heads, so that the binomial distribution can be written as

$$\mathrm{Bin}(m|N,\mu) = \binom{N}{m}\mu^m(1-\mu)^{N-m} \qquad (3.9)$$