4.4.5 Unbiased estimates

The entropy S can also be used to quantify subjective estimates of probabilities. In the absence of any information, the best *unbiased estimate* is that all M outcomes are equally likely. This is the distribution of maximum entropy. If additional information is available, the unbiased estimate is obtained by maximizing the entropy subject to the constraints imposed by this information. For example, if it is known that $\langle F(x)\rangle = f$, we can maximize

$$S\left(\alpha,\beta,\{p_i\}\right) = -\sum_i p(i)\ln p(i) - \alpha\left(\sum_i p(i) - 1\right) - \beta\left(\sum_i p(i)F(x_i) - f\right) , \qquad (4.4.26)$$

where the *Lagrange multipliers* α and β are introduced to impose the constraints of normalization, and $\langle F(x)\rangle = f$, respectively. The result of the optimization is a distribution $p_i \propto \exp\big(-\beta F(x_i)\big)$, where the value of β is fixed by the constraint. This process can be generalized to an arbitrary number of conditions. It is easy to see that if the first $n = 2k$ moments (and hence n cumulants) of a distribution are specified, the unbiased estimate is the exponential of an n^{th} order polynomial.

In analogy with Eq. (4.4.24), we can define an entropy for a continuous random variable ($\mathcal{S}_x = \{-\infty < x < \infty\}$) as

$$S = -\int dx\, p(x)\ \ln p(x) = -\left\langle \ln p(x)\right\rangle \quad . \qquad (4.4.27)$$

There are, however, problems with this definition, as for example S is not invariant under a one to one mapping. (After a change of variable to $f = F(x)$, the entropy is changed by $\langle |F'(x)|\rangle$.)