$p(x_i)\Delta$. This gives a discrete distribution for which the entropy takes the form

$$\mathrm{H}_\Delta = -\sum_i p(x_i)\Delta \ln\left(p(x_i)\Delta\right) = -\sum_i p(x_i)\Delta \ln p(x_i) - \ln\Delta \qquad (2.90)$$

where we have used $\sum_i p(x_i)\Delta = 1$, which follows from (2.89) and (2.25). We now omit the second term $-\ln\Delta$ on the right-hand side of (2.90), since it is independent of $p(x)$, and then consider the limit $\Delta \to 0$. The first term on the right-hand side of (2.90) will approach the integral of $p(x)\ln p(x)$ in this limit so that

$$\lim_{\Delta\to 0}\left\{-\sum_i p(x_i)\Delta \ln p(x_i)\right\} = -\int p(x)\ln p(x)\,\mathrm{d}x \qquad (2.91)$$

where the quantity on the right-hand side is called the *differential entropy*. We see that the discrete and continuous forms of the entropy differ by a quantity $\ln\Delta$, which diverges in the limit $\Delta \to 0$. This reflects that specifying a continuous variable very precisely requires a large number of bits. For a density defined over multiple continuous variables, denoted collectively by the vector $\mathbf{x}$, the differential entropy is given by

$$\mathrm{H}[\mathbf{x}] = -\int p(\mathbf{x})\ln p(\mathbf{x})\,\mathrm{d}\mathbf{x}. \qquad (2.92)$$

2.5.4 Maximum entropy

We saw for discrete distributions that the maximum entropy configuration corresponds to a uniform distribution of probabilities across the possible states of the variable. Let us now consider the corresponding result for a continuous variable. If this maximum is to be well defined, it will be necessary to constrain the first and second moments of $p(x)$ and to preserve the normalization constraint. We therefore maximize the differential entropy with the three constraints:

$$\int_{-\infty}^{\infty} p(x)\,\mathrm{d}x = 1 \qquad (2.93)$$

$$\int_{-\infty}^{\infty} xp(x)\,\mathrm{d}x = \mu \qquad (2.94)$$

$$\int_{-\infty}^{\infty} (x-\mu)^2 p(x)\,\mathrm{d}x = \sigma^2. \qquad (2.95)$$

Appendix C

The constrained maximization can be performed using Lagrange multipliers so that we maximize the following functional with respect to $p(x)$:

$$-\int_{-\infty}^{\infty} p(x)\ln p(x)\,\mathrm{d}x + \lambda_1\left(\int_{-\infty}^{\infty} p(x)\,\mathrm{d}x - 1\right) + \lambda_2\left(\int_{-\infty}^{\infty} xp(x)\,\mathrm{d}x - \mu\right) + \lambda_3\left(\int_{-\infty}^{\infty} (x-\mu)^2 p(x)\,\mathrm{d}x - \sigma^2\right). \qquad (2.96)$$