imply (absolute) independence, that is,

\[(2.20)\quad p(x, y \mid z) = p(x \mid z) \cdot p(y \mid z) \neq p(x, y) = p(x)p(y)\]

The converse is also in general untrue: absolute independence does not imply conditional independence:

\[(2.21)\quad p(x, y) = p(x)p(y) \neq p(x, y \mid z) = p(x \mid z)p(y \mid z)\]

In special cases, however, conditional and absolute independence may coincide.

A number of probabilistic algorithms require us to compute features, or statistics, of probability distributions. The expectation of a random variable \(X\) is given by

\[(2.22)\quad E[X] = \sum_x x \cdot p(x) \quad \text{(discrete)}\]

\[(2.23)\quad E[X] = \int x \cdot p(x) \, dx \quad \text{(continuous)}\]

Not all random variables possess finite expectations; however, those that do not are of no relevance to the material presented in this book.

The expectation is a linear function of a random variable. In particular, we have

\[(2.24)\quad E[aX + b] = aE[X] + b\]

for arbitrary numerical values \(a\) and \(b\). The covariance of \(X\) is obtained as follows

\[(2.25)\quad \text{Cov}[X] = E[(X - E[X])^2] = E[X^2] - E[X]^2\]

The covariance measures the squared expected deviation from the mean. As stated above, the mean of a multivariate normal distribution \(\mathcal{N}(x; \mu, \Sigma)\) is \(\mu\), and its covariance is \(\Sigma\).

A final concept of importance in this book is entropy. The entropy of a probability distribution is given by the following expression:

\[(2.26)\quad H_p(x) = E[-\log_2 p(x)]\]

which resolves to

\[(2.27)\quad H_p(x) = -\sum_x p(x) \log_2 p(x) \quad \text{(discrete)}\]

\[(2.28)\quad H_p(x) = -\int p(x) \log_2 p(x) \, dx \quad \text{(continuous)}\]