What is its expectation, averaged under the distribution $Q = Q^*/Z_Q$ of the point $x^{(r)}$?

$$\langle w_r \rangle = \int \mathrm{d}x \, Q(x) \frac{P^*(x)}{Q^*(x)} = \int \mathrm{d}x \, \frac{1}{Z_Q} P^*(x) = \frac{Z_P}{Z_Q}. \tag{29.54}$$

So the expectation of the denominator is

$$\left\langle \sum_r w_r \right\rangle = R \frac{Z_P}{Z_Q}. \tag{29.55}$$

As long as the variance of $w_r$ is finite, the denominator, divided by $R$, will converge to $Z_P/Z_Q$ as $R$ increases. [In fact, the estimate converges to the right answer even if this variance is infinite, as long as the expectation is well-defined.] Similarly, the expectation of one term in the numerator is

$$\langle w_r \phi(x) \rangle = \int \mathrm{d}x \, Q(x) \frac{P^*(x)}{Q^*(x)} \phi(x) = \int \mathrm{d}x \, \frac{1}{Z_Q} P^*(x) \phi(x) = \frac{Z_P}{Z_Q} \Phi, \tag{29.56}$$

where $\Phi$ is the expectation of $\phi$ under $P$. So the numerator, divided by $R$, converges to $\frac{Z_P}{Z_Q}\Phi$ with increasing $R$. Thus $\hat{\Phi}$ converges to $\Phi$.

The numerator and the denominator are unbiased estimators of $RZ_P/Z_Q$ and $RZ_P/Z_Q\Phi$ respectively, but their ratio $\hat{\Phi}$ is not necessarily an unbiased estimator for finite $R$.

Solution to exercise 29.2 (p.363). When the true density $P$ is multimodal, it is unwise to use importance sampling with a sampler density fitted to one mode, because on the rare occasions that a point is produced that lands in one of the other modes, the weight associated with that point will be enormous. The estimates will have enormous variance, but this enormous variance may not be evident to the user if no points in the other modes have been seen.

Solution to exercise 29.5 (p.371). The posterior distribution for the syndrome decoding problem is a pathological distribution from the point of view of Gibbs sampling. The factor $\mathbb{1}[\mathbf{Hn} = \mathbf{z}]$ is 1 only on a small fraction of the space of possible vectors $\mathbf{n}$, namely the $2^K$ points that correspond to the valid codewords. No two codewords are adjacent, so similarly, any single bit flip from a viable state $\mathbf{n}$ will take us to a state with zero probability and so the state will never move in Gibbs sampling.

A general code has exactly the same problem. The points corresponding to valid codewords are relatively few in number and they are not adjacent (at least for any useful code). So Gibbs sampling is no use for syndrome decoding for two reasons. First, finding *any* reasonably good hypothesis is difficult, and as long as the state is not near a valid codeword, Gibbs sampling cannot help since none of the conditional distributions is defined; and second, once we are in a valid hypothesis, Gibbs sampling will never take us out of it.

One could attempt to perform Gibbs sampling using the bits of the original message $\mathbf{s}$ as the variables. This approach would not get locked up in the way just described, but, for a good code, any single bit flip would substantially alter the reconstructed codeword, so if one had found a state with reasonably large likelihood, Gibbs sampling would take an impractically large time to escape from it.

Solution to exercise 29.12 (p.380). Each Metropolis proposal will take the energy of the state up or down by some amount. The total change in energy when $B$ proposals are concatenated will be the end-point of a random walk with $B$ steps in it. This walk might have mean zero, or it might have a tendency to drift upwards (if most moves increase the energy and only a few decrease it). In general the latter will hold, if the acceptance rate $f$ is small: the mean change in energy from any one move will be some $\Delta E > 0$ and so the acceptance probability for the concatenation of $B$ moves will be of order $1/(1 + \exp(-B\Delta E))$, which scales roughly as $f^B$. The mean-square-distance moved will be of order $f^B B\epsilon^2$, where $\epsilon$ is the typical step size. In contrast, the mean-square-distance moved when the moves are considered individually will be of order $fB\epsilon^2$.

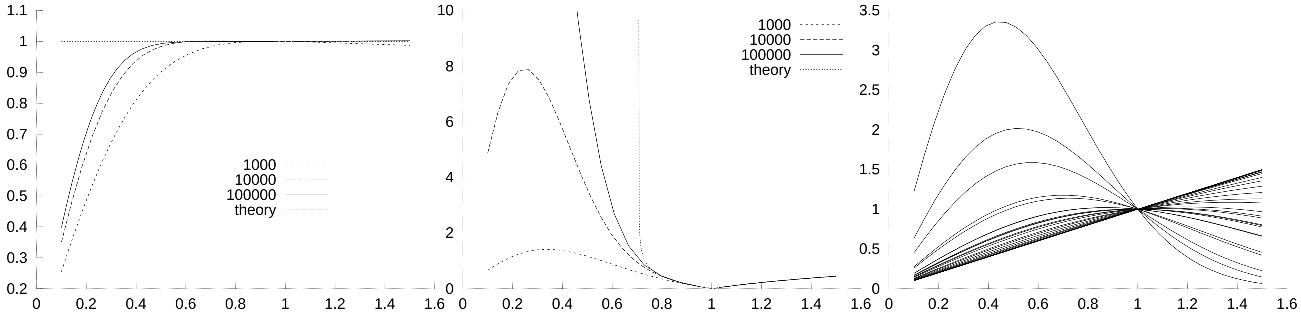

Figure 29.20. Importance sampling in one dimension. For $R = 1000$, $10^4$, and $10^5$, the normalizing constant of a Gaussian distribution (known in fact to be 1) was estimated using importance sampling with a sampler density of standard deviation $\sigma_q$ (horizontal axis). The same random number seed was used for all runs. The three plots show (a) the estimated normalizing constant; (b) the *empirical* standard deviation of the $R$ weights; (c) 30 of the weights.

Solution to exercise 29.13 (p.382). The weights are $w = P(x)/Q(x)$ and $x$ is drawn from $Q$. The mean weight is

$$\int \mathrm{d}x\, Q(x)\, [P(x)/Q(x)] = \int \mathrm{d}x\, P(x) = 1, \tag{29.57}$$

assuming the integral converges. The variance is

$$\begin{aligned} \mathrm{var}(w) &= \int \mathrm{d}x\, Q(x) \left[\frac{P(x)}{Q(x)} - 1\right]^2 & (29.58)\\ &= \int \mathrm{d}x\, \frac{P(x)^2}{Q(x)} - 2P(x) + Q(x) & (29.59)\\ &= \left[\int \mathrm{d}x\, \frac{Z_Q}{Z_P^2} \exp\left(-\frac{x^2}{2}\left(\frac{2}{\sigma_p^2} - \frac{1}{\sigma_q^2}\right)\right)\right] - 1, & (29.60) \end{aligned}$$

where $Z_Q/Z_P^2 = \sigma_q/(\sqrt{2\pi}\sigma_p^2)$. The integral in (29.60) is finite only if the coefficient of $x^2$ in the exponent is positive, i.e., if

$$\sigma_q^2 > \frac{1}{2}\sigma_p^2. \tag{29.61}$$

If this condition is satisfied, the variance is

$$\mathrm{var}(w) = \frac{\sigma_q}{\sqrt{2\pi}\sigma_p^2}\sqrt{2\pi}\left(\frac{2}{\sigma_p^2} - \frac{1}{\sigma_q^2}\right)^{-\frac{1}{2}} - 1 = \frac{\sigma_q^2}{\sigma_p\left(2\sigma_q^2 - \sigma_p^2\right)^{1/2}} - 1. \tag{29.62}$$

As $\sigma_q$ approaches the critical value – about $0.7\sigma_p$ – the variance becomes infinite. Figure 29.20 illustrates these phenomena for $\sigma_p = 1$ with $\sigma_q$ varying from 0.1 to 1.5. *The same random number seed was used for all runs,* so the weights and estimates follow smooth curves. Notice that the *empirical* standard deviation of the $R$ weights can look quite small and well-behaved (say, at $\sigma_q \simeq 0.3$) when the true standard deviation is nevertheless infinite.