Nested Sampling

The goal of nested sampling is to compute a multidimensional integral

$$Z = \int_{\Omega} L(\theta)\, \mathrm{d}\mu(\theta) \tag{1}$$

for some $L(\theta) \geq 0$. We assume the total volume of parameter space is $\int_{\Omega} \mathrm{d}\mu(\theta) = \mu(\Omega) = 1$. In applications, $L(\theta)$ is usually interpreted as a likelihood and the measure as a prior

$$\mathrm{d}\mu(\theta) = \pi(\theta)\, \mathrm{d}\theta$$

which is normalised so that $\int \pi(\theta)\, \mathrm{d}\theta = 1$.

Transforming the integral to one dimension

Summary. We can rewrite eq. 1 as a one-dimensional integral

$$Z = \int_0^1 \tilde{L}(X)\, \mathrm{d}X$$

where we define:

$$\tilde{L}(\xi) = \sup\{L^* \in [0, \infty) \,|\, X(L^*) > \xi\}$$
$$X(L^*) = \mu(\{\theta \in \Omega \mid L(\theta) > L^*\})$$

We will build up to this incomprehensible result below.

Similar to the construction of a Lebesgue integral, we can rewrite eq. 1 as

$$Z = \int_0^{\infty} X(L)\, \mathrm{d}L$$

where $X(L)\, \mathrm{d}L$ is volume of a horizontal slice through $L(\theta)$ as in Figure 1. Formally, $X(L^*)$ is the the μ-measure of the L^*-super-level sets, or the volume of points in parameter space having larger likelihood than L^*.

$$X(L^*) \coloneqq \int_{L(\theta) > L^*} \mathrm{d}\mu(\theta) = \mu(\{\theta \in \Omega \mid L(\theta) > L^*\})$$

In the Bayesian interpretation, $X(L^*)$ is the prior mass of minimum likelihood L^*, or the probability that a sample drawn from the prior has likelihood at least L^*.

$$X(L^*) = \int_{L > L^*} \pi(\theta)\, \mathrm{d}\theta = \mathbb{P}_{\theta \sim \pi}[L(\theta) > L^*] \tag{2}$$

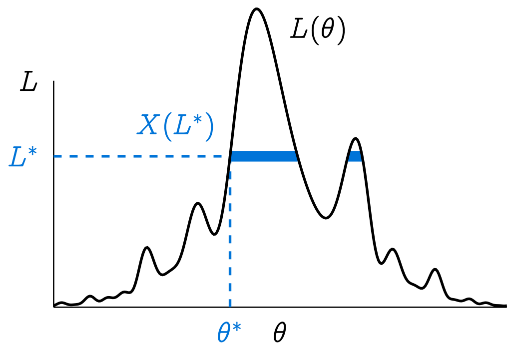

Figure 1: The integral $Z = \int L(\theta)\, \mathrm{d}\mu(\theta) = \int_0^{\infty} X(L)\, \mathrm{d}L$ can be computed by integrating together all the horizontal slices with measure $X(L)$.

Notice that the range of X is $[0, 1]$ because the total μ-measure of parameter space $X(0)$ is one (or in the Bayesian picture, because the prior is normalised). It is also clear that X is monotonically decreasing; the slices get smaller as you go up. If we assume that $L(\theta)$ is never flat (no likelihood plateaux) then X is also continuous. Together, these facts mean that $X : \mathrm{im}(L) \to [0, 1]$ is invertible, as in Figure 2.

$$X^{-1}(L^*) = \sup\{L^* \in \mathrm{im}(L) \mid X(L^*) > x\} \tag{3}$$

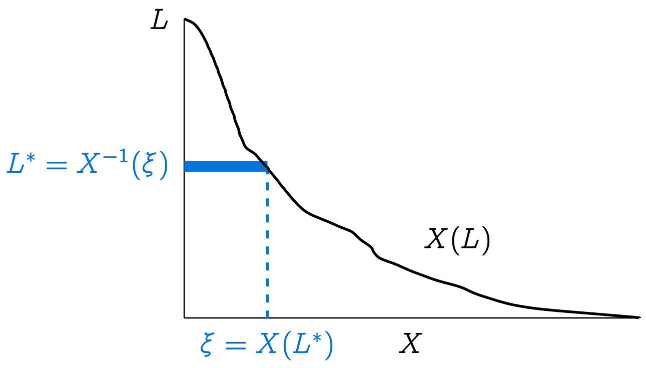

Figure 2: The measure of the L-super-level sets $X(L)$ plotted horizontally as a function of L on the vertical. Since X is monotonically decreasing and continuous, it has an inverse.

Now, we can change coordinates $L \mapsto X$ to transform the integral

$$Z = \int_0^{\infty} X(L)\, \mathrm{d}L = \int_0^1 X^{-1}(\xi)\, \mathrm{d}\xi$$

which pictorially just says that integrating vertical slices $X(L)\, \mathrm{d}L$ under the curve in Figure 2 is the same as integrating the horizontal slices, $X^{-1}(\xi)\, \mathrm{d}\xi$.

Evaluating the integral

Suppose we sample som points $\theta_i \sim \pi$ from the prior. The volumes $X(L(\theta_i))$ associated to these points are uniformly distributed.

Proof. The random variable $X(L(\theta)) \in [0, 1]$ is uniformly distributed if for any $0 \leq u \leq 1$ we have $\mathbb{P}_{\theta \sim \pi}[X(L(\theta)) < u] = u$. Applying the inverse eq. 3 (remembering that it is decreasing so the inequality is reversed) we get

$$\mathbb{P}_{\theta \sim \pi}[X(L(\theta)) < u] = \mathbb{P}_{\theta \sim \pi}\left[L(\theta) > X^{-1}(u)\right] = X(X^{-1}(u)) = u$$

where the middle step follows by substituting eq. 2. ∎