Exercise: Deriving the posterior predictive density for the healthy levels game

We will first consider one-dimensional "rectangles" (i.e., lines); since the dimensions are independent, we can easily generalize to 2d.

For convenience, we will follow the notation of Josh Tenenbaum's PhD thesis (**?**). In particular, let $h = \boldsymbol{\theta}$ be the unknown hypothesis or parameter vector, $\mathcal{H}$ be the set of possible hypotheses (rectangles), $\mathcal{H}_y$ be the set of hypotheses consistent with observation y (so the rectangles have to be big enough to capture y), and $\mathcal{H}_{X,y}$ be the set of hypotheses consistent with all the examples in $X = \mathcal{D}$ as well as with y.

The posterior predictive is given by

$$p(y|X) = \frac{p(y, X)}{p(X)} \tag{1}$$

where

$$\begin{aligned} p(X) &= \int_{h \in \mathcal{H}} p(h) p(X|h) dh & (2) \\ &= \int_{h \in \mathcal{H}_X} p(h)/|h|^N dh & (3) \end{aligned}$$

where we used the fact that $p(X|h) = 1/|h|^N$ if $h \in \mathcal{H}_X$ and is 0 otherwise. Similarly, $p(y, X) = \int_{h \in \mathcal{H}_{X,y}} p(h)/|h|^N dh$.

To derive the integral in Equation 3, let us assume the maximum observed value is 0 (we can pick any maximum and recenter the data, since we assume a translation invariant prior). Then the right edge of the rectangle must lie past the data, so $\ell \geq 0$. Also, if r is the range spanned by the examples, then the left most data point is at $-r$, so the left side of the rectangle must satisfy $l - s \leq -r$, where s is size of the rectangle.

1. Using these assumptions, show that

$$p(X) = \frac{1}{N(N-1)r^{N-1}} \tag{4}$$

   Hint: use integration by parts

$$I = \int_a^b f(x) g'(x) dx = [f(x)g(x)]_a^b - \int_a^b f'(x) g(x) dx \tag{5}$$

2. To compute $p(y, X)$, we just need to extend the range from r to $r + d$, where d is the distance from y to the closest observed example. Hence show that

$$p(y|X) = \frac{1}{(1 + d/r)^{N-1}} \tag{6}$$