Exercise: Fayyad-Irani binning

The standard definition of Fayyad-Irani binning is as follows. Define a cost function for a given proposed binary partition of a continuous scalar x, based on a set S of instances, as follows:

$$\text{cost}(b_1, b_2) \triangleq \frac{|S_1|}{|S|} H(S_1) + \frac{|S_2|}{|S|} H(S_2) \tag{1}$$

where $H(S_k)$ is the entropy of of the class labels in the induced partition. Show that $\text{cost}(b_1, b_2) = -\mathbb{I}(B; Y) + \text{const}$, where $B \in \{b_1, b_2\}$ is a binary variable specifying which bin to use, and const is a constant independent of B. (Use a plugin approximation based on the MLE for all unknown probabilities.)