**Algorithm EBD**

**Input:** Dataset $D$ and parameter $\lambda$.

**Output:** An optimal Bayesian discretization of variable $X$ relative to $D$.

**Definitions of terms:**

Let $D$ be a dataset of $n$ instances consisting of the list $((X_1, Z_1), (X_2, Z_2), \ldots, (X_k, Z_k), \ldots, (X_n, Z_n))$ that is sorted in ascending order of $X_k$, where $X_k$ is a real value of the predictor variable and $Z_k$ is the associated integer value of the target variable.

Let $S_{a,b}$ be a list of the first elements in $D$, starting at the $a^{th}$ pair in $D$ and ending at the $b^{th}$ pair.

Let $T_b$ be a set that represents a discretization of $S_{1,b}$.

Let target variable $Z$ have $J$ unique values, and let $Z_j$ denote the $j^{th}$ unique value. Let $U$ be a real array of $J$ elements, and let $U_j$ denote its $j^{th}$ element. $U$ will contain the distribution of values of the target variable for some $S_{a,b}$.

Let $n'$ be the number of unique values of predictor variable $X$, and let $X_k$ denote the $k^{th}$ unique value.

Let $V$ be a real array of $n'$ elements, and let $V_y$ denote its $y^{th}$ element.

For $1 \leq k \leq n'$, let $W_k = (count_{k,1}, count_{k,2}, \ldots, count_{k,J})$ be an array such that for $1 \leq j \leq J$, the term $count_{k,j}$ is equal to the number of pairs in $D$ in which the first element has value $X_k$ and the second element (i.e., the target value) has value $Z_j$.

Let $MarginalLikelihood(U)$ be the following marginal likelihood function, which follows from Equation 7 when array $U$ is used to derive the $n_i$ and $n_{ij}$ counts:

$$MarginalLikelihood(U) := \frac{(J-1)!}{(J-1 + \sum_{j=1}^{J} U_j)!} \prod_{j=1}^{J} U_j!$$

Let $Prior(k)$ be the prior function defined in Equation 10 in the text.

**Lines of Code:**

1. $V_0 := 1$;
2. $T_0 := \{}$;
3. for $a := 1$ to $n'$
   4. $P := Prior(a)$;
   5. $V_a := 0$;
   6. $U := (0, 0, \ldots, 0)$;
   7. for $b := a$ downto 1
      8. $U := U + W_b; /* element-wise addition */$
   9. $ML := MarginalLikelihood(U);$  
10. $Score_{ba} := P \times ML$;
11. if $V_{b-1} \times Score_{ba} > V_a$
12.    then
13.        $T_a := T_{b-1} \cup \{ S_{b,a} \}$;
14.        $V_a := V_{b-1} \times Score_{ba}$;
15.    $P := P \times (1 - Prior(b-1))$;
16. return $T_{n'}$.  