Submitted to *INFORMS Journal on Optimization*
manuscript

Tree Bounds for Sums of Bernoulli Random Variables: A Linear Optimization Approach (Online Companion)

Divya Padmanabhan
Engineering Systems Design, Singapore University of Technology and Design, divya_padmanabhan@sutd.edu.sg

Karthik Natarajan
Engineering Systems Design, Singapore University of Technology and Design, karthik_natarajan@sutd.edu.sg

Key words: probability bounds, trees, linear optimization

History: This paper was first submitted in October 2019.

---

1. Proof of Proposition 3

The proof idea is to express the probability $\mathbb{P}\left(\sum_{l\in T(i,s)} \tilde{c}_l = t, \tilde{c}_i = y\right)$, denoted by $w_{i,s,y,t}$, in terms of probabilities for smaller sub-trees of $T(i,s)$. Finally the required probability, $\mathbb{P}(\sum_{i=1}^{n} \tilde{c}_i \geq k)$ is expressed in terms of $\mathbb{P}\left(\sum_{l\in T(1,d_1)} \tilde{c}_l = \cdot, \tilde{c}_1 = \cdot\right)$ corresponding to the root node 1 (or random variable $\tilde{c}_1$). For any node with at least two children,

$$\begin{aligned}
\mathbb{P}\left(\sum_{i\in T(i,s)} \tilde{c}_l = t, \tilde{c}_i = 0\right) &= \sum_{a=0}^{N(i(s),d_{i(s)})-1} \mathbb{P}\left(\sum_{l\in V(i,s-1)} \tilde{c}_l = t-a, \tilde{c}_i = 0, \sum_{l\in V(i(s),d_{i(s)})} \tilde{c}_l = a\right) \\
&= \sum_{a=0}^{N(i(s),d_{i(s)})-1} \mathbb{P}\left(\sum_{l\in V(i,s-1)} \tilde{c}_l = t-a, \tilde{c}_i = 0\right) \mathbb{P}\left(\sum_{l\in V(i(s),d_{i(s)})} \tilde{c}_l = a | \tilde{c}_i = 0\right)
\end{aligned}$$

The first equality arises by considering all possible ways in which the sum t can be split up between the sub-trees $T(i,s-1)$ and $T(i(s),d_{i(s)})$ (the trees $T1$ and $T2$ in Figure 1 provide an illustrative example for $s=3$) and the second equality arises due to conditional independence of the random variables in $V(i(s),d_{i(s)})$ and $V(i,s-1)$ given $\tilde{c}_i$. Note that this method of splitting the sum t, is the same as what was considered in expressing $x_{i,s,0,t}$ in the quadratic knapsack problem in Section 2.3 of the paper. Next by considering all possible values of $\tilde{c}_{i(s)}$ we have,

$$\mathbb{P}\left(\sum_{l\in V(i(s),d_{i(s)})}\tilde{c}_l = a|\tilde{c}_i = 0\right) = \mathbb{P}\left(\sum_{l\in V(i(s),d_{i(s)})}\tilde{c}_l = a, \tilde{c}_{i(s)=0}|\tilde{c}_i = 0\right) + \tag{1}$$

$$\mathbb{P}\left(\sum_{l\in V(i(s),d_{i(s)})}\tilde{c}_l = a, \tilde{c}_{i(s)=1}|\tilde{c}_i = 0\right) \tag{2}$$

These can be further simplified as,

$$\mathbb{P}\left(\sum_{l\in V(i(s),d_{i(s)})}\tilde{c}_l = a, \tilde{c}_{i(s)=0}|\tilde{c}_i = 0\right) = \mathbb{P}\left(\sum_{l\in V(i(s),d_{i(s)})}\tilde{c}_l = a, \tilde{c}_{i(s)=0}, \tilde{c}_i = 0\right)/\mathbb{P}(\tilde{c}_i = 0) \tag{3}$$

$$= \mathbb{P}(\tilde{c}_i = 0, \tilde{c}_{i(s)} = 0)\mathbb{P}(\sum_{l\in V(i(s),d_{i(s)})}\tilde{c}_l = a|\tilde{c}_{i(s)} = 0, \tilde{c}_i = 0)/\mathbb{P}(\tilde{c}_i = 0) \tag{4}$$

$$= \mathbb{P}(\tilde{c}_i = 0, \tilde{c}_{i(s)} = 0)\mathbb{P}(\sum_{l\in V(i(s),d_{i(s)})}\tilde{c}_l = a|\tilde{c}_{i(s)} = 0)/\mathbb{P}(\tilde{c}_i = 0) \tag{5}$$

$$= \frac{\mathbb{P}(\tilde{c}_i = 0, \tilde{c}_{i(s)} = 0)\mathbb{P}(\sum_{l\in V(i(s),d_{i(s)})}\tilde{c}_l = a, \tilde{c}_{i(s)} = 0)}{\mathbb{P}(\tilde{c}_i = 0)\mathbb{P}(\tilde{c}_{i(s)} = 0)} \tag{6}$$

Here (5) arises as random variables corresponding to $V(i(s), d_{i(s)})$ are independent of $\tilde{c}_i$ given the value of $\tilde{c}_{i(s)}$. Similarly, it can be verified that,

$$\mathbb{P}\left(\sum_{l\in V(i(s),d_{i(s)})}c_l = a, c_{i(s)=1}|c_i = 0\right) = \frac{\mathbb{P}(\tilde{c}_i = 0, \tilde{c}_{i(s)} = 1)\mathbb{P}(\sum_{l\in V(i(s),d_{i(s)})}\tilde{c}_l = a, \tilde{c}_{i(s)} = 1)}{\mathbb{P}(\tilde{c}_i = 0)\mathbb{P}(\tilde{c}_{i(s)} = 1)} \tag{7}$$

$$\mathbb{P}\left(\sum_{l\in V(i(s),d_{i(s)})}\tilde{c}_l = a, \tilde{c}_{i(s)=0}|c_i = 1\right) = \frac{\mathbb{P}(\tilde{c}_i = 1, \tilde{c}_{i(s)} = 0)\mathbb{P}(\sum_{l\in V(i(s),d_{i(s)})}\tilde{c}_l = a, \tilde{c}_{i(s)} = 0)}{\mathbb{P}(\tilde{c}_i = 1)\mathbb{P}(\tilde{c}_{i(s)} = 0)} \tag{8}$$

$$\mathbb{P}\left(\sum_{l\in V(i(s),d_{i(s)})}\tilde{c}_l = a, \tilde{c}_{i(s)} = 1|\tilde{c}_i = 1\right) = \frac{\mathbb{P}(\tilde{c}_i = 1, \tilde{c}_{i(s)} = 1)\mathbb{P}(\sum_{l\in V(i(s),d_{i(s)})}\tilde{c}_l = a, \tilde{c}_{i(s)} = 1)}{(\mathbb{P}(\tilde{c}_i = 1)\mathbb{P}(\tilde{c}_{i(s)} = 1)} \tag{9}$$

For the case $s = 1$, by considering all possible realizations of the first child of i we get,

$$\mathbb{P}\left(\sum_{l\in V(i,1)}\tilde{c}_l = t, \tilde{c}_i = 0\right) = \mathbb{P}\left(\sum_{l\in V(i(1),d_{i(1)})}\tilde{c}_l = t, \tilde{c}_{i(1)} = 0, c_i = 0\right) + \mathbb{P}\left(\sum_{l\in V(i(1),d_{i(1)})}\tilde{c}_l = t, \tilde{c}_{i(1)} = 1, c_i = 0\right)$$

$$= \mathbb{P}(\tilde{c}_i = 0, \tilde{c}_{i(1)} = 0)\mathbb{P}\left(\sum_{l\in V(i(1),d_{i(1)})}\tilde{c}_l = t|\tilde{c}_{i(1)} = 0, c_i = 0\right)$$

$$+ \mathbb{P}(\tilde{c}_i = 0, \tilde{c}_{i(1)} = 1)\mathbb{P}\left(\sum_{l\in V(i(1),d_{i(1)})}\tilde{c}_l = t|\tilde{c}_{i(1)} = 1, c_i = 0\right)$$

$$= \mathbb{P}(\tilde{c}_i = 0, \tilde{c}_{i(1)} = 0)\mathbb{P}(\sum_{l\in V(i(1),d_{i(1)})}\tilde{c}_l = t|\tilde{c}_{i(1)} = 0) + \mathbb{P}(\tilde{c}_i = 0, \tilde{c}_{i(1)} = 1)\mathbb{P}(\sum_{l\in V(i(1),d_{i(1)})}\tilde{c}_l = t|\tilde{c}_{i(1)} = 1)$$

where the third equality is a consequence of conditional independence of the random variables corresponding to $V(i(1), d_{i(1)})$ and $\tilde{c}_i$ given $\tilde{c}_i(1)$.

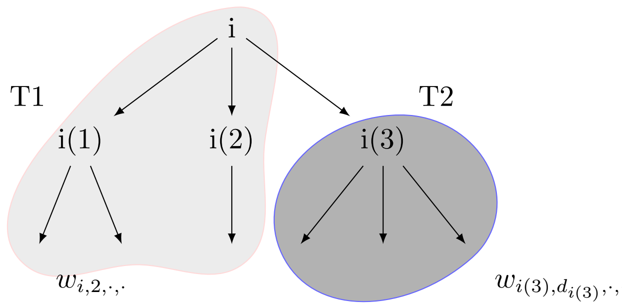

Figure 1 $w_{i,3,\cdot,t}$ **can be computed using** $w_{i,2,\cdot,t-a}$ **(corresponding to the sub-tree T1) and** $w_{i(3),d_{i(3)},\cdot,a}$ **(corresponding to the sub-tree T2) so that** $t-a$ **nodes are selected from T1 and** a **nodes are selected from T2.**