二叉树

Huffman编码树：正确性

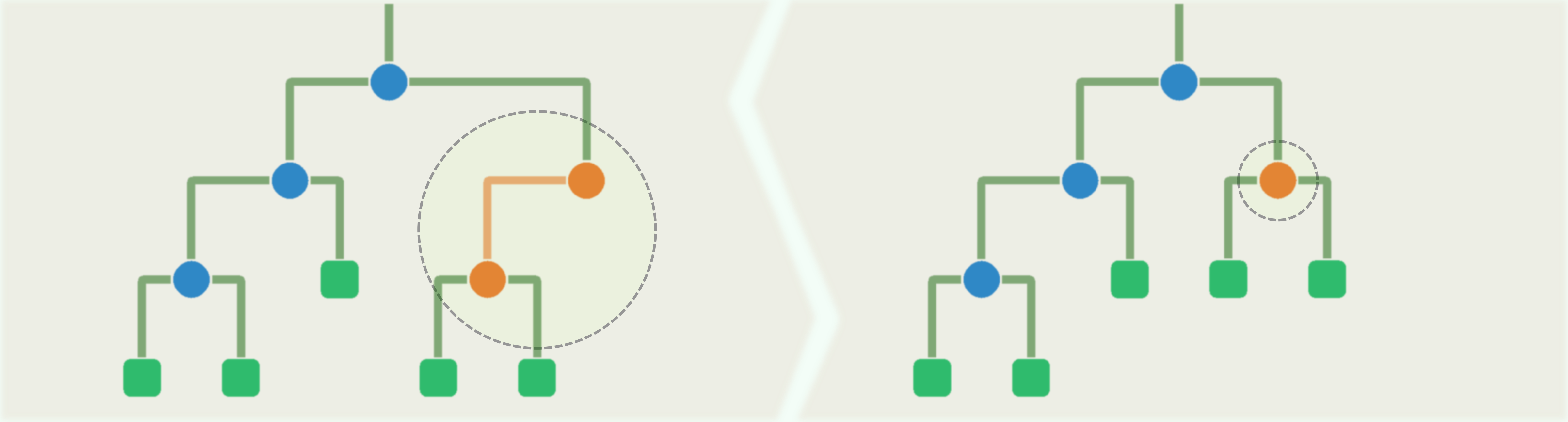
05-J2

我生来就不像我所见过的任何一个人；我敢断言，我与世上的任何一个人都迥然不同；虽说我不比别人好，但至少我与他们完全两样。

邓俊辉
deng@tsinghua.edu.cn

双子性

- ❖ 最优编码树有何特征？
- ❖ 首先，每一内部节点都有**两个**孩子——节点度数均为偶数（0或2），即**真**二叉树
- ❖ 否则，将1度节点替换为其唯一的孩子，则新树的wald将更小



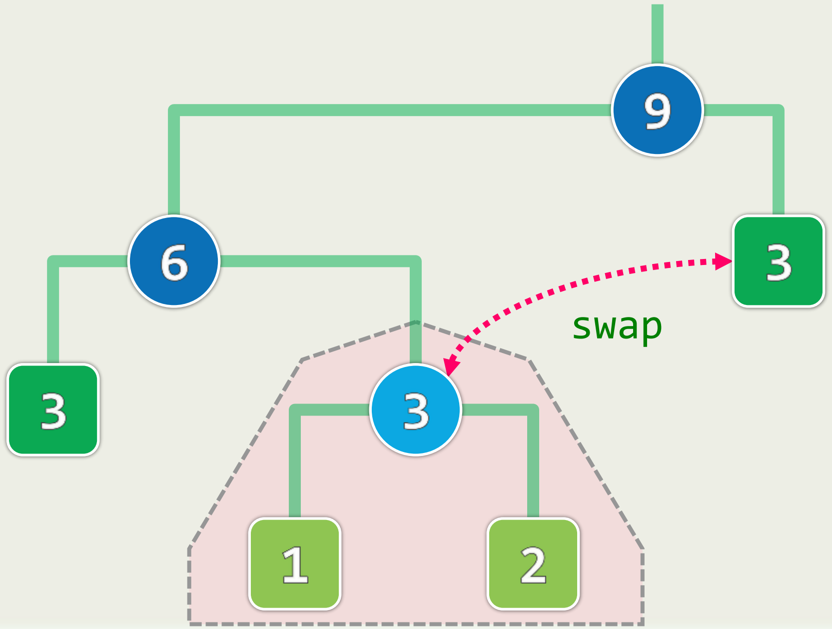
不唯一性

❖ 对任一内部节点而言

左、右子树互换之后wald不变

❖ 上述算法中，兄弟子树的次序系随机选取

故有可能...

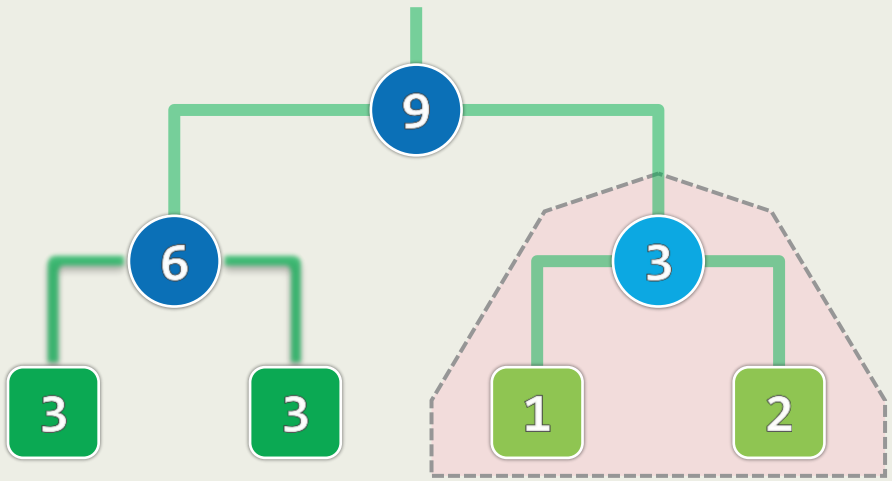


❖ 为消除这种歧义，可以（比如）

明确要求左子树的频率更低

❖ 不过，倘若

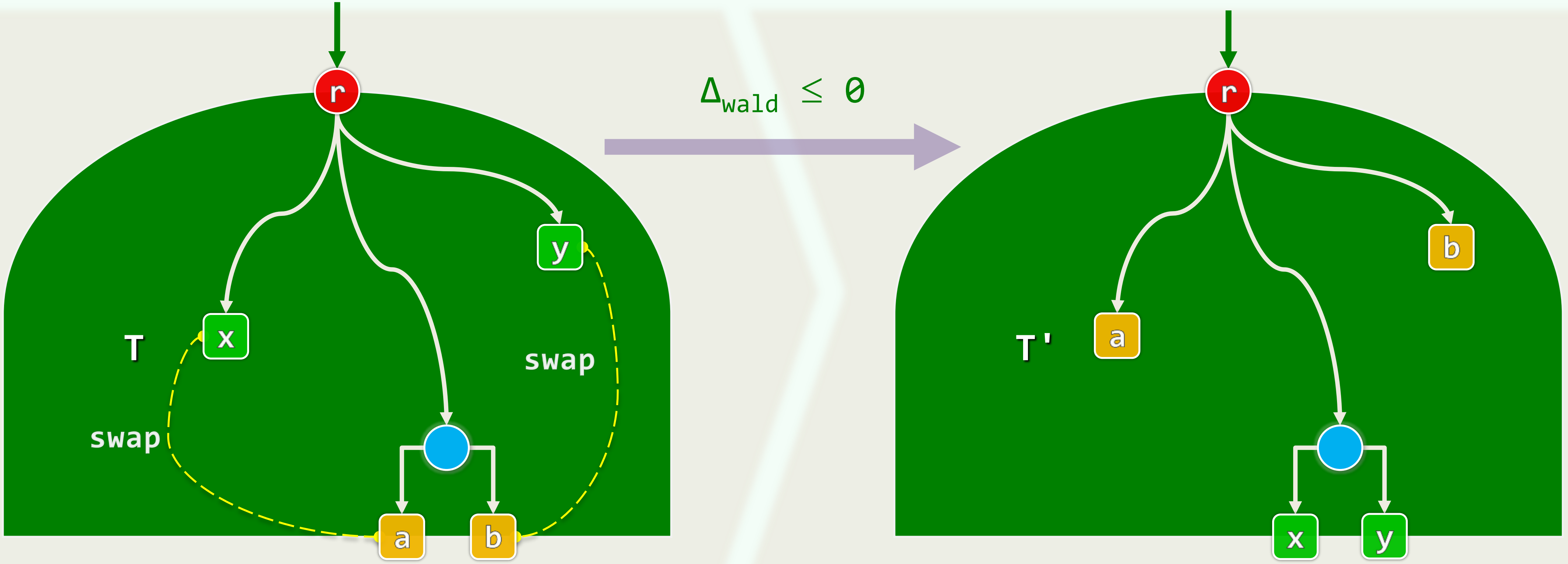
它们（甚至更多节点）的频率恰好相等...



层次性

- 出现**频率最低**的字符$\boxed{x}$和$\boxed{y}$，必在**某棵**最优编码树中处于**最底层**，且互为**兄弟**
- 否则，**任取**一棵最优编码树，并在其最底层**任取**一对兄弟$\boxed{a}$和$\boxed{b}$

  于是，$\boxed{a}$和$\boxed{x}$、$\boxed{b}$和$\boxed{y}$交换之后，wald绝不会增加

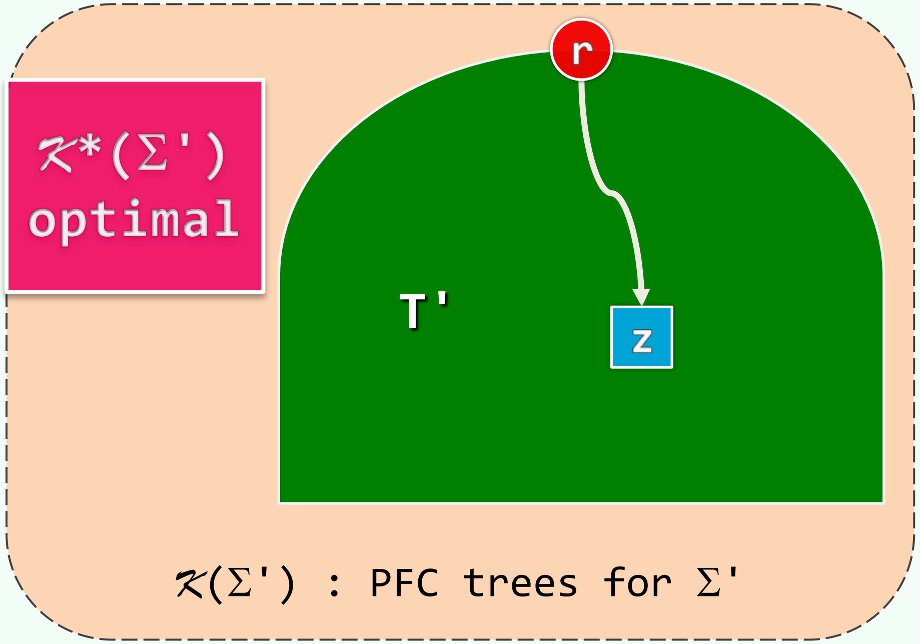


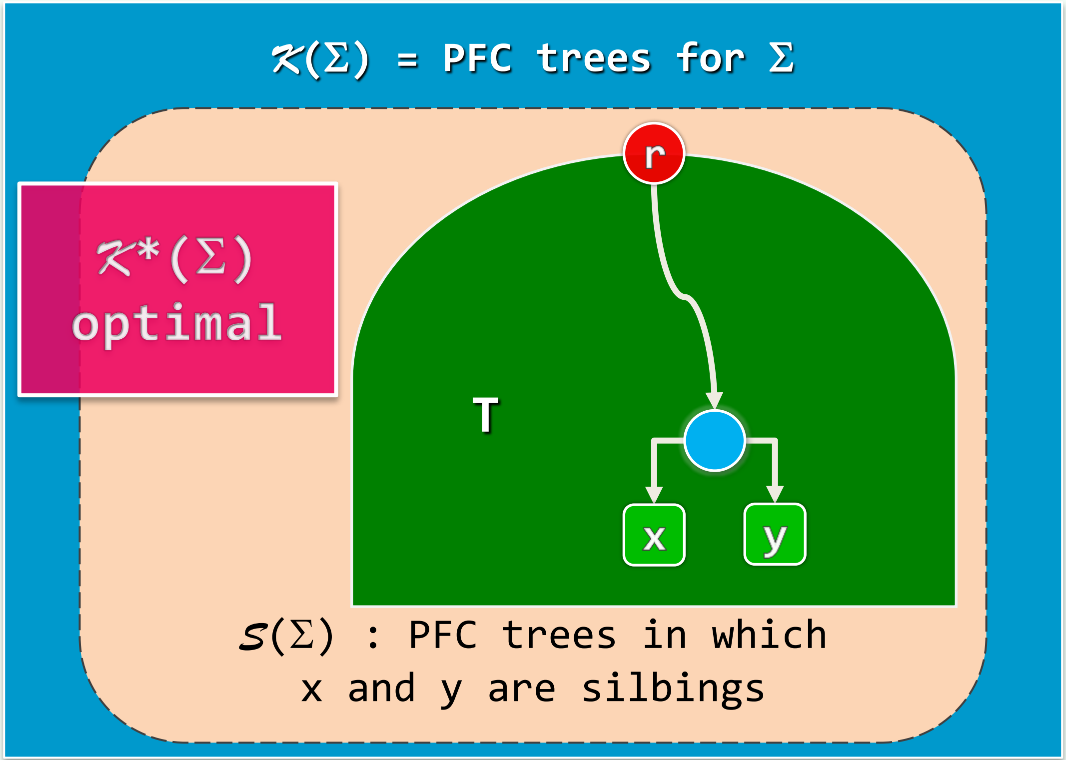
数学归纳

❖ 对|Σ|做归纳可证：Huffman算法所生成的，必是一棵最优编码树！|Σ| = 2时显然

❖ 设算法在|Σ|<n时均正确。现设|Σ|=n，取Σ中频率最低的$\boxed{x}$、$\boxed{y}$（不妨就设二者互为兄弟）

❖ 令：$\Sigma' = (\Sigma \backslash \{x, y\}) \cup \{z\}$ ，$w(z) = w(x) + w(y)$



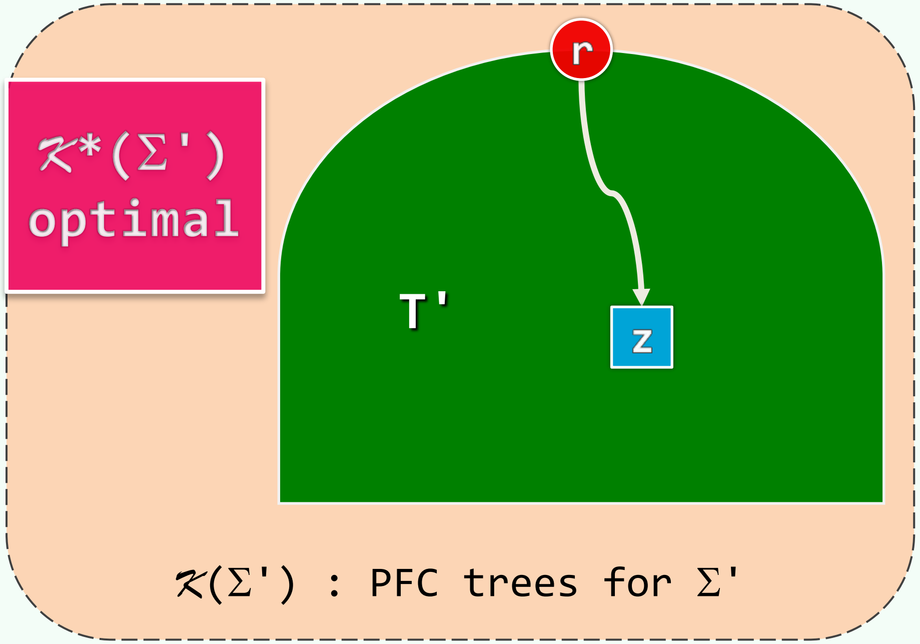


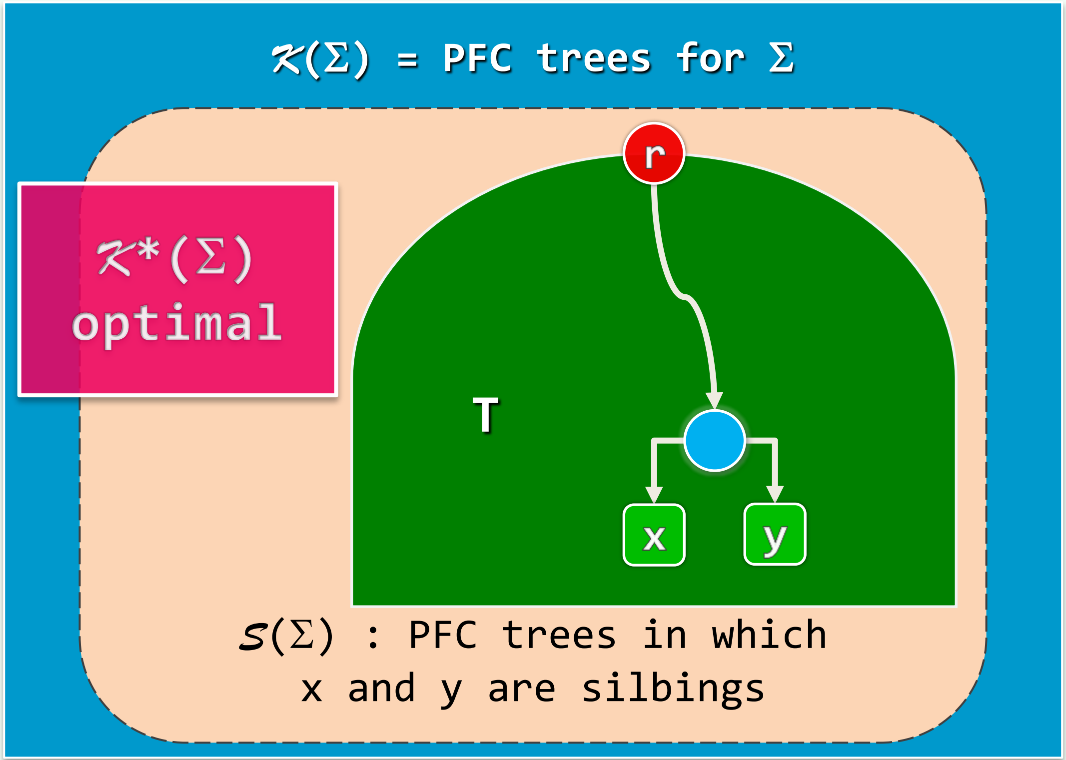
定差

❖ 对于Σ'的任一编码树T'，只要为z添加孩子x和y，即可得到Σ的一棵编码树T，且

$$wd(T) - wd(T') \ = \ w(x) + w(y) \ = \ w(z)$$

❖ 可见，如此对应的T和T'，wd之差与T的具体形态无关





最优对最优

- 因此，只要T'是Σ'的最优编码树，则T也必是Σ的最优编码树（之一）
- 实际上，Huffman算法的过程，与上述归纳过程完全一致

——每一步迭代都可视作，从某棵T转入对应的T'