



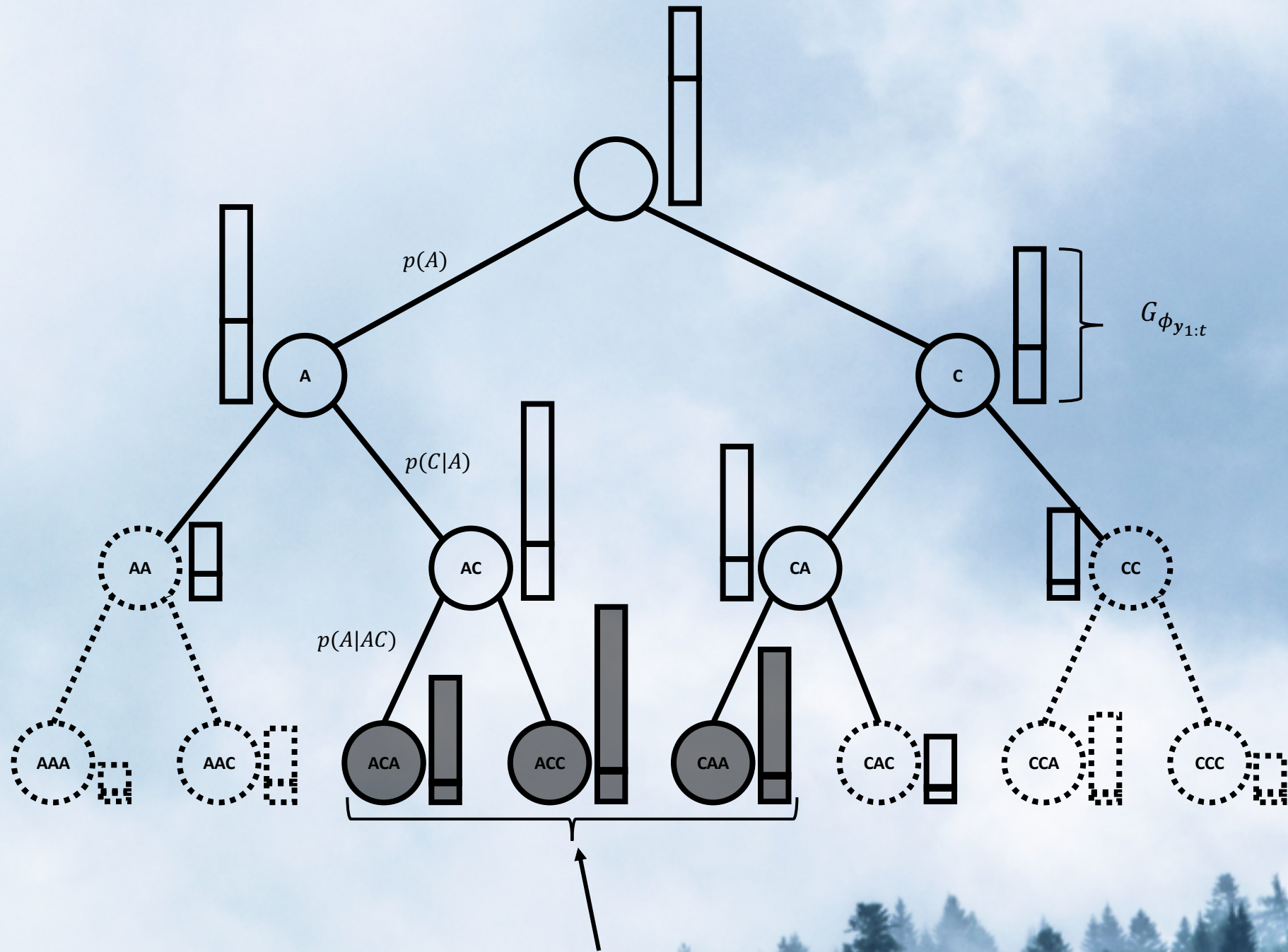
Ancestral Gumbel-Top-*k* Sampling

Ancestral Gumbel-Top- k Sampling

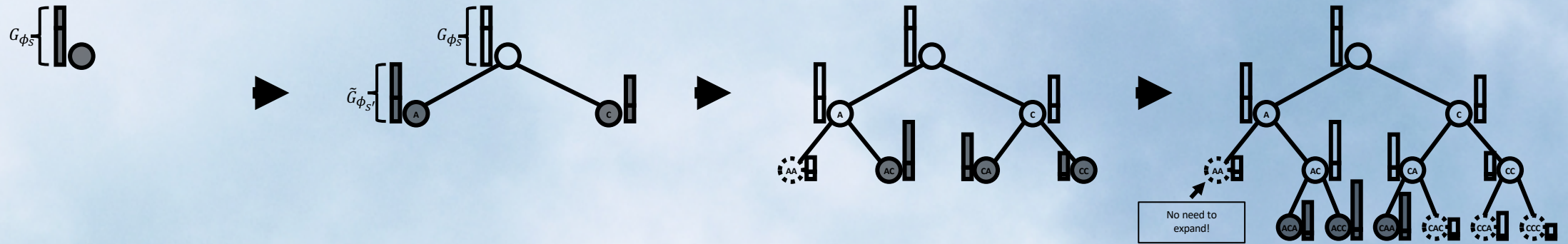
Generalizes Stochastic Beam Search

Expands $1 \leq m \leq k$ nodes per iteration

Applies to discrete valued Bayes networks

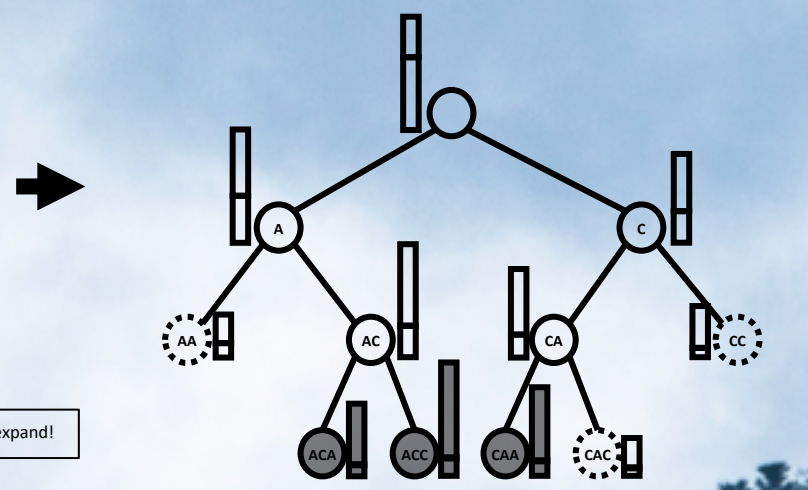
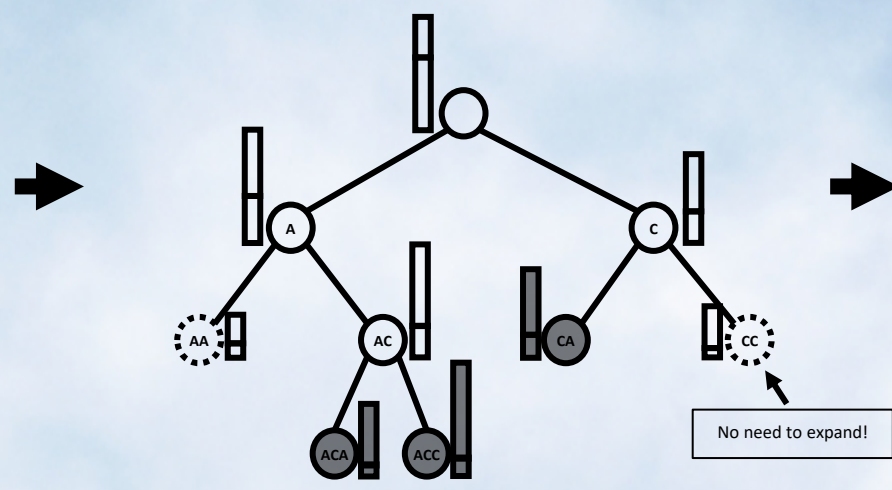
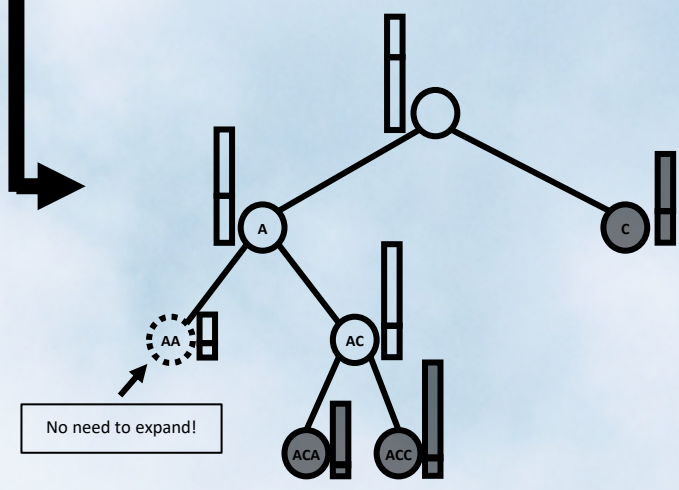
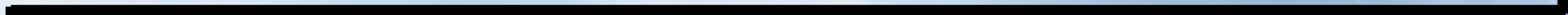
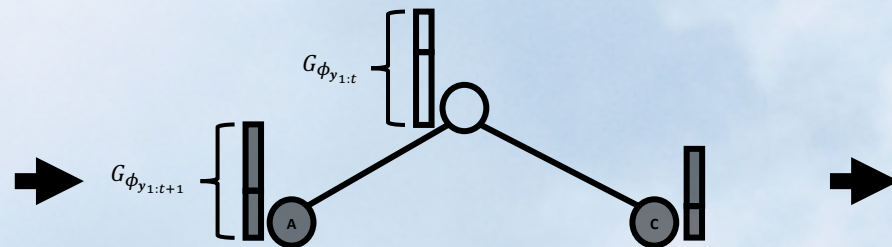
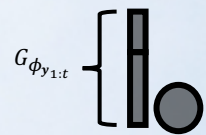


Stochastic Beam Search



$$m = k (= 3)$$

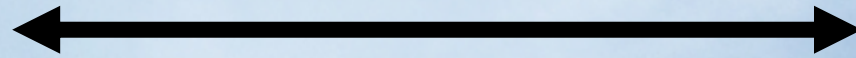
$m = 1$



Ancestral Gumbel-Top- k Sampling

$$m = 1$$

$$m = k$$



Sequential

Parallel

Incremental

Batch

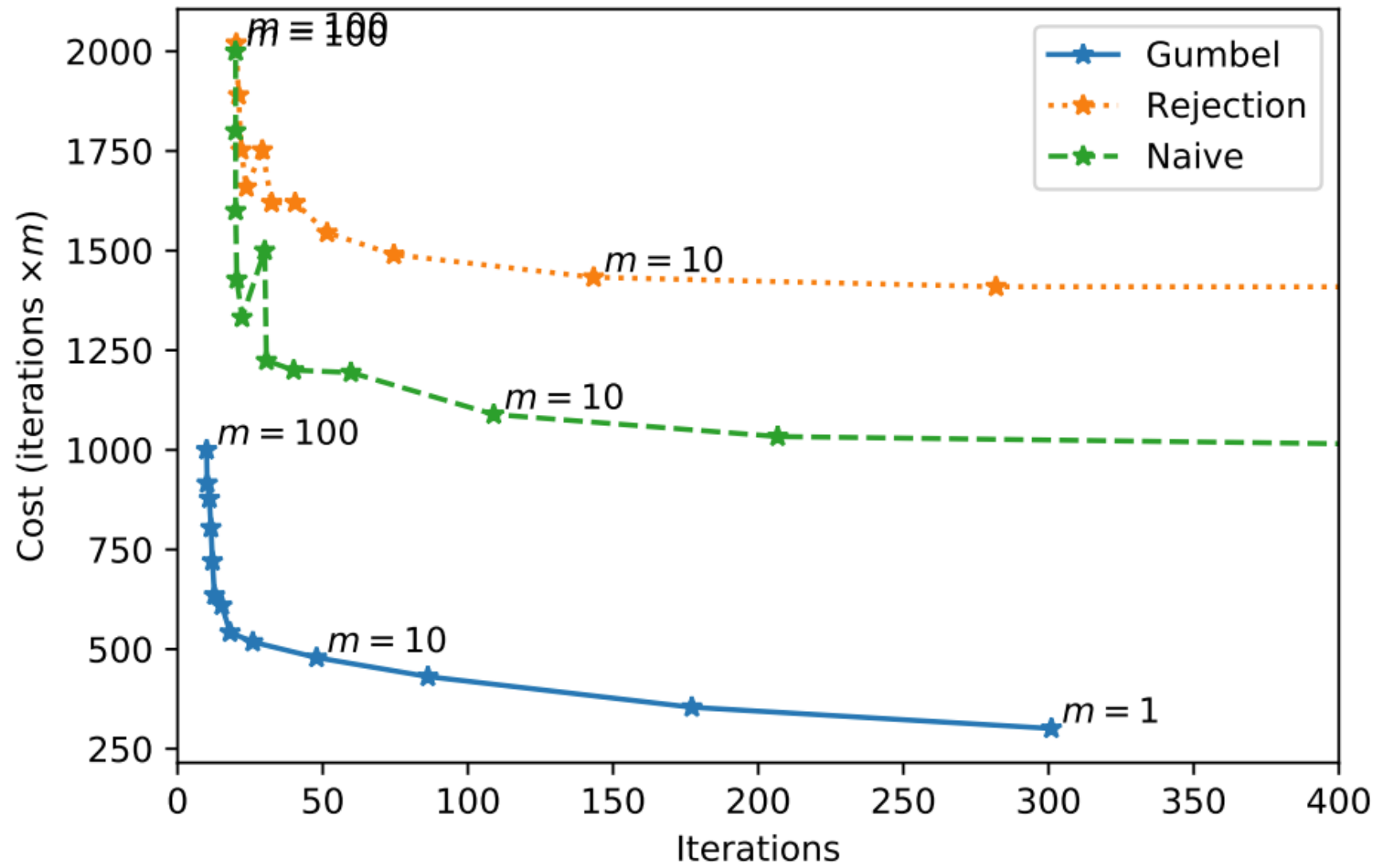
More iterations

Fewer iterations

Less computation

More computation

Cost vs. iterations ($c = 0.5, k = 100$)



Ancestral Gumbel-Top- k Sampling

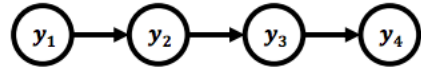
Generalizes Stochastic Beam Search

Expands $1 \leq m \leq k$ nodes per iteration

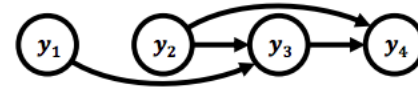
Applies to discrete valued Bayes networks



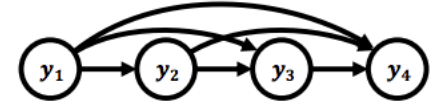
(a) Independent
 $p(\mathbf{y}) = \prod_v p(y_v)$



(b) Markov chain
 $p(\mathbf{y}) = p(y_1) \prod_{t>1} p(y_t | y_{t-1})$

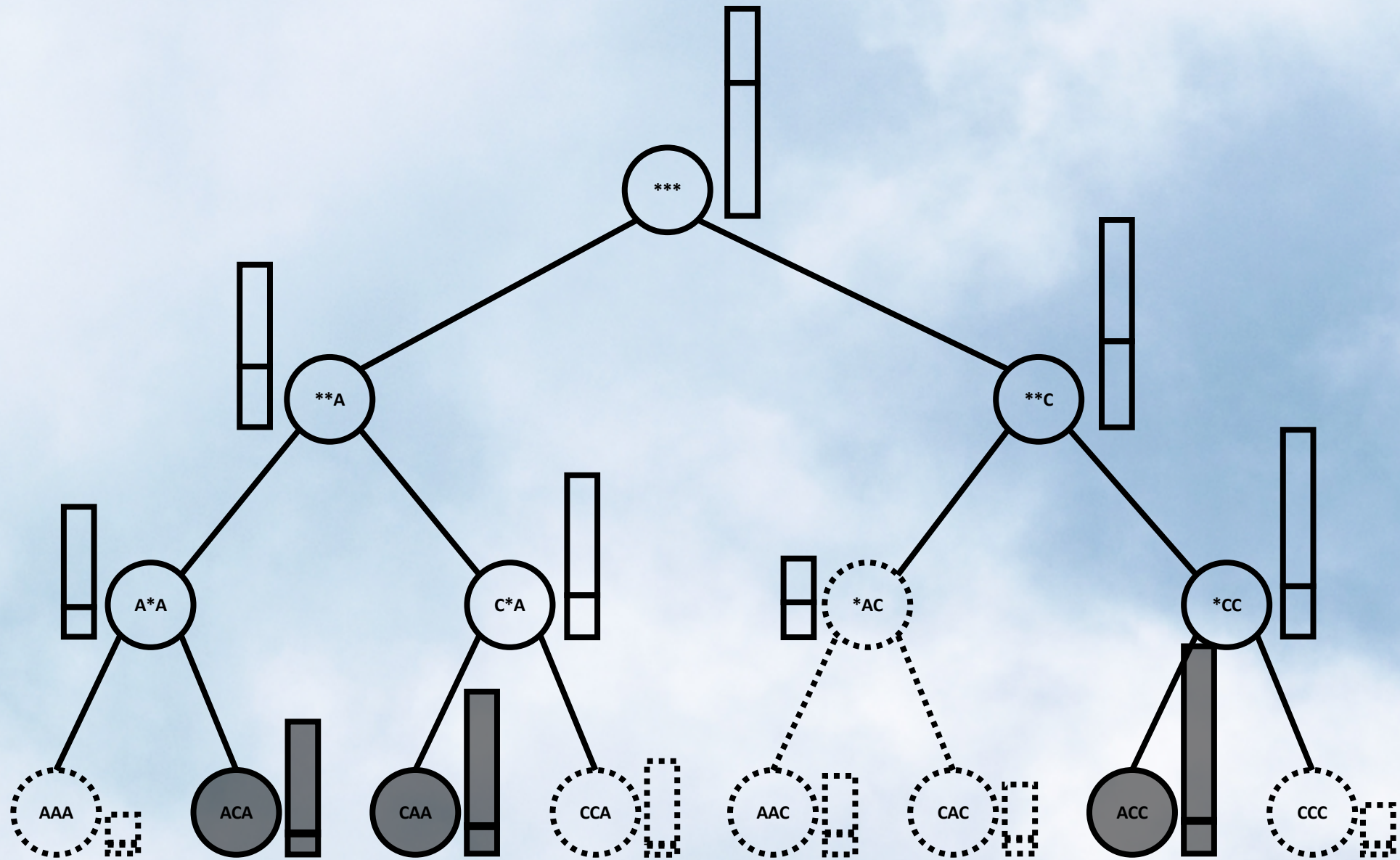


(c) General network
 $p(\mathbf{y}) = \prod_v p(y_v | \mathbf{y}_{\text{pa}(v)})$



(d) Sequence model
 $p(\mathbf{y}) = \prod_t p(y_t | \mathbf{y}_{1:t-1})$

Figure 1: Examples of Bayesian networks.



Iterations vs. connectivity c ($m = 1, k = 100$)

