Lesson 16. Stochastic Dynamic Programming, cont.

1 Last time...

- Last time, we considered the following problem:

Problem 1. The Dijkstra Brewing Company is planning production of its new limited run beer, Primal Pilsner, over the next 2 months. Based on some market analysis studies, the company has determined that the demand for the new beer in each month will be:

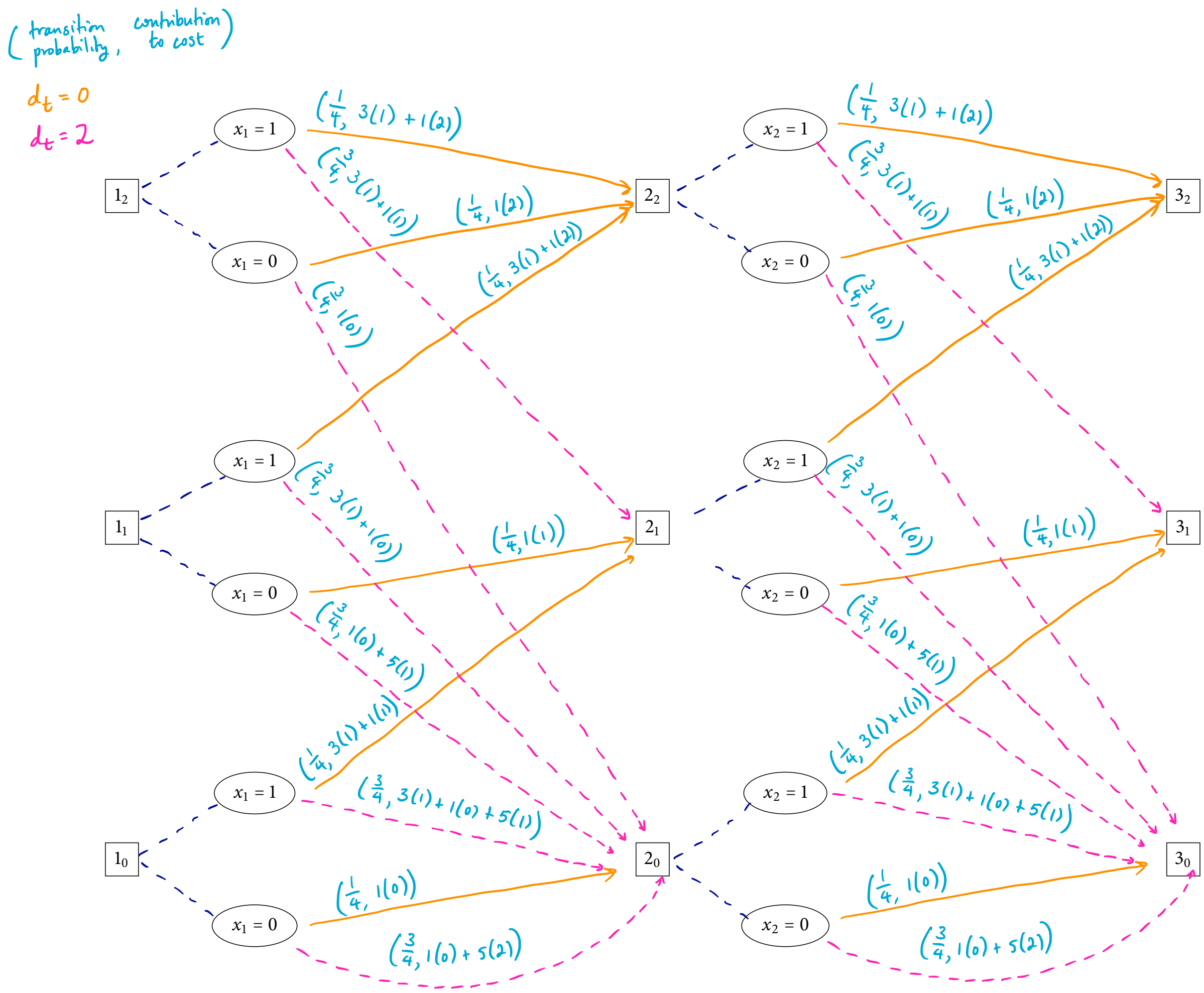
Demand (batches)	Probability
0	1/4
2	3/4

Each batch of beer costs \$3,000 to produce. Batches can be held in inventory at a cost of \$1,000 per batch per month. Each month, the company can produce either 0 or 1 batches, due to capacity limitations. In addition, the size of the company's warehouse restricts the ending inventory for each month to at most 2 batches. The company has 1 batch ready to go in inventory.

Due to contractual obligations, there is a penalty of \$5,000 for each batch of demand not met. Any batches produced that cannot be stored in the company's warehouse gets thrown away, and cannot be used to meet future demand.

The company wants to find a production plan that will minimizes its total production and holding costs over the next 3 months.

- Note that the state in the next stage depends on (i) the decision and (ii) randomness
  - We had not encountered randomness in our previous dynamic programming models
- We drew a diagram the entire 2-month process
- Let:
  - Node t_n represent month t with n batches in inventory
  - x_t represent the number of batches to produce in month t
  - d_t represent the number of batches in demand in month t



- We also computed the expected cost of a production policy:
  - In month 1, produce 1 batch
  - In month 2:
    - If there are 2 batches in inventory, produce 0 batches
    - If there are 0 batches in inventory, produce 1 batch
- Working backwards:
  - Expected cost in month 2 with 2 batches in inventory (node 2_2):

$$\frac{1}{4}\big[1(2)\big] + \frac{3}{4}\big[1(0)\big] = \frac{1}{2}$$

  - Expected cost in month 2 with 0 batches in inventory (node 2_0):

$$\frac{1}{4}\big[3(1) + 1(1)\big] + \frac{3}{4}\big[3(1) + 1(0) + 5(1)\big] = 7$$

  - Expected cost in month 1 (node 1_1):

$$\frac{1}{4}\left[3(1) + 1(1) + \frac{1}{2}\right] + \frac{3}{4}\left[3(1) + 1(0) + 7\right] = 8.875$$

2 Writing down the model

- What we really want: a production policy with minimum expected cost
- Let's write down the recursive representation of our model/diagram
- We can then solve this recursion by working backwards and determine the minimum expected cost and associated optimal policy
- Stages:

- States:

- Transition probability $p(m \mid n, t, x_t)$ of moving from state n to state m in stage t under decision x_t:

- Contribution $c(m \mid n, t, x_t)$ of moving from state n to state m in stage t under decision x_t:

- Allowable decisions x_t at stage t and state n:

- In words, the cost-to-go $f_t(n)$ at stage t and state n is:

- Boundary conditions:

- Cost-to-go recursion:

- Desired cost-to-go function value:

3 Interpreting the value-to-go function

- We can solve this recursion just like with a deterministic DP: start at the boundary conditions and work backwards
- For this problem, we get the following cost-to-go function values $f_t(n)$ for $t = 1, 2$ and $n = 0, 1, 2$, as well as the decision x_t^* that attained each value:

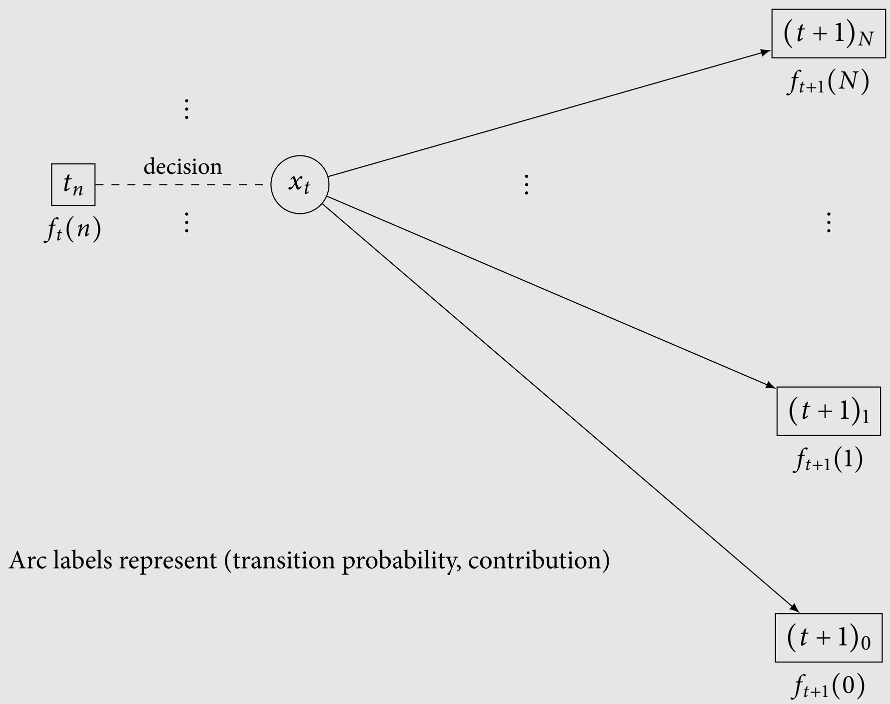
t	n	$f_t(n)$	x_t^*
1	0	13.125	1
1	1	8.875	1
1	2	5.875	0
2	0	7	1
2	1	3.5	1
2	2	0.5	0

- Based on this, what should the company's policy be?
- What is the company's total expected cost?

4 Stochastic dynamic programs, more generally

Stochastic dynamic program

- **Stages** $t = 1, 2, \dots, T$ and **states** $n = 0, 1, 2, \dots, N$
- Allowable **decisions** x_t at each stage t and state n
- **Transition probability** $p(m \mid n, t, x_t)$ of moving from state n to state m in stage t under decision x_t
- **Contribution** $c(m \mid n, t, x_t)$ for moving from state n to state m in stage t under decision x_t



- **Value-to-go** function $f_t(n)$ at each stage t and state n
- **Boundary conditions** on $f_T(n)$ for each state n
- **Recursion** on $f_t(n)$ at stage t and state n

$$f_t(n) = \min_{x_t \text{ allowable}} \left\{ \sum_{m=0}^{N} p(m \mid n, t, x_t) \Big[c(m \mid n, t, x_t) + f_{t+1}(m) \Big] \right\}$$

for $t = 1, 2, \dots, T-1$ and $n = 0, 1, \dots, N$

- **Desired value-to-go**, usually $f_1(m)$ for some state m