Lesson 13. Introduction to Stochastic Dynamic Programming

1 Motivation

- In the dynamic programs we have studied so far, the transitions from one state to the next are **deterministic**
- For example, the knapsack problem:
  - Suppose we are in stage t and state n (deciding whether to take metal t with n kg of space remaining)
  - If we decide to take metal t in stage t, we know exactly what state we will be in stage t+1: $n - \left(\text{weight of metal } t\right)$
- What if the transitions between states are subject to some randomness or **stochasticity**?

2 A production and inventory problem with stochastic demand

Example 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.

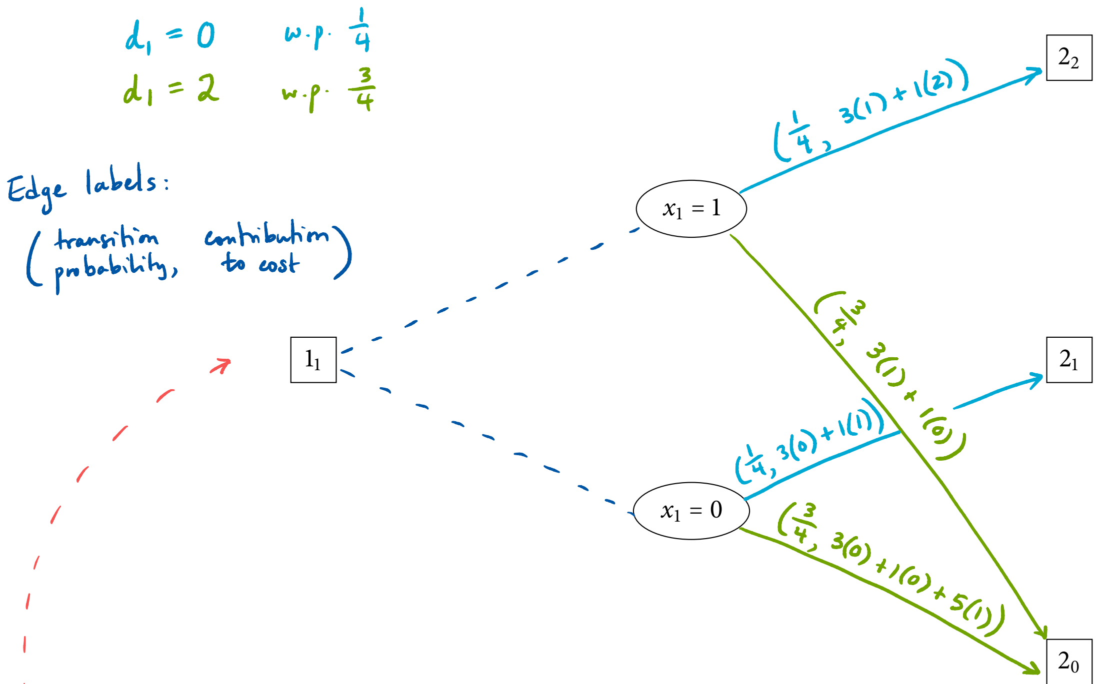
Due to contractual obligations, there is a penalty of \$5,000 for each batch of demand not met. Any batches produced that cannot be stored in the company's warehouse gets thrown away, and cannot be used to meet future demand.

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

2.1 Modeling the problem visually

- Let's think about the decision-making process starting at month 1
- 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 can draw the following diagram (that looks like a graph) that models the decision-making process

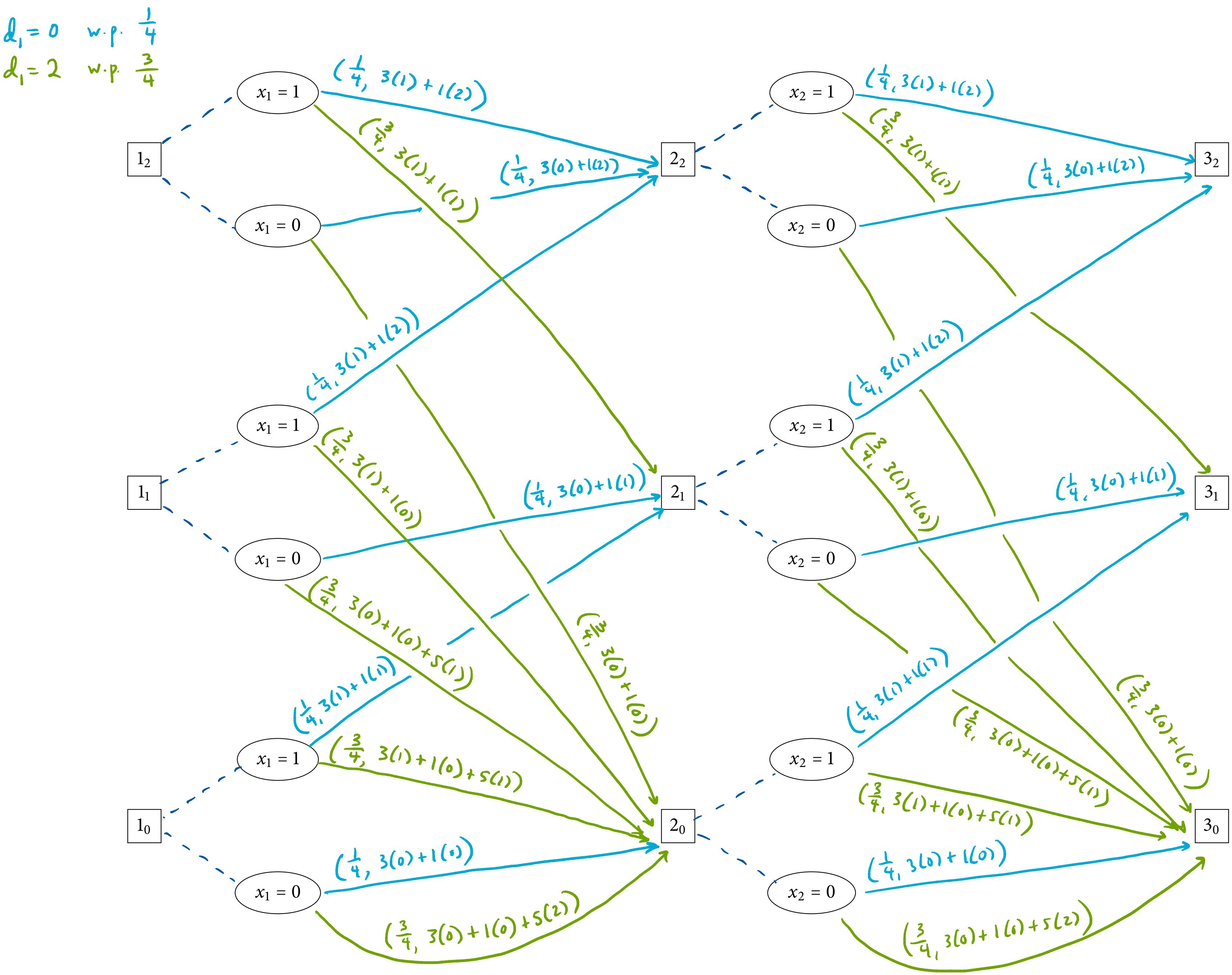


- We can diagram the entire 2-month process in a similar fashion:

$$f_1(1) = \min \left\{ \begin{array}{l} \underbrace{\frac{1}{4}\left[3(0) + 1(1) + f_2(1)\right] + \frac{3}{4}\left[3(0) + 1(0) + 5(1) + f_2(0)\right]}_{x_1 = 0}, \\ \underbrace{\frac{1}{4}\left[3(1) + 1(2) + f_2(2)\right] + \frac{3}{4}\left[3(1) + 1(0) + 5(0) + f_2(0)\right]}_{x_1 = 1} \end{array} \right\}$$

plug in $t=1$, $n=1$

$$f_t(n) = \min_{x_t \in \{0,1\}} \left\{ \begin{array}{l} \frac{1}{4}\left[3x_t + 1 \cdot \min\{n + x_t, 2\} + f_{t+1}(\min\{n + x_t, 2\})\right] \\ + \frac{3}{4}\left[3x_t + 1 \cdot \max\{n + x_t - 2, 0\} + 5\max\{2 - (n + x_t), 0\} + f_{t+1}(\max\{n + x_t - 2, 0\})\right] \end{array} \right\}$$



d1 = 0 w.p. 1/4
d1 = 2 w.p. 3/4
1_2
1_1
1_0
2_2
2_1
2_0
3_2
3_1
3_0
x1 = 1
x1 = 0
x2 = 1
x2 = 0
(1/4, 3(1) + 1(2))
(3/4, 3(1) + 1(1))
(1/4, 3(0) + 1(2))
(1/4, 3(1) + 1(1))
(3/4, 3(1) + 1(0))
(1/4, 3(0) + 1(1))
(3/4, 3(0) + 1(0) + 5(1))
(3/4, 3(0) + 1(0))
(3/4, 3(1) + 1(0) + 5(1))
(1/4, 3(0) + 1(0))
(3/4, 3(0) + 1(0) + 5(2))

- Consider the following 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
- What is the expected cost of this policy?
- Working backwards:
  - Expected cost in month 2 with 2 batches in inventory (node 2_2):

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

  - 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):

$$\frac{1}{4}\Big(3(1)+1(2)+\underset{\substack{\uparrow\\ \text{Expected cost in}\\ \text{month 2 w/ 2 batches}\\ \text{in inventory (node } 2_2)}}{\frac{1}{2}}\Big) + \frac{3}{4}\Big(3(1)+1(0)+\underset{\substack{\uparrow\\ \text{Expected cost in}\\ \text{month 2 w/ 0 batches}\\ \text{in inventory (node } 2_0)}}{7}\Big) = 8.875$$

2.2 Things to think about

- The policy above gives **contingency plans**
- The diagram we drew on page 3 sort of looks like a shortest path problem, but it's not!
- We cannot solve this example as a shortest path problem, since the edges "are random"
- We can, however, still write a recursion to represent this example problem, and others like it
- We'll explore this next...

2.3 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:

$$\text{Stage } t \leftrightarrow \begin{cases} \text{month } t & (t=1,2) \\ \text{end of decision-making process} & (t=3) \end{cases}$$

- States:

$$\text{State } n \leftrightarrow \text{having } n \text{ batches in inventory at the beginning of the month} \quad (n=0,1,2)$$

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

warehouse capacity

unmet demand?

$$p(m \mid n,t,x_t) = \begin{cases} \frac{1}{4} & \text{if } m = \min\{n+x_t, 2\} & d_t = 0 \text{ w.p. } \frac{1}{4} \\ \frac{3}{4} & \text{if } m = \max\{n+x_t-2, 0\} & d_t = 2 \text{ w.p. } \frac{3}{4} \\ 0 & \text{o/w} \end{cases}$$

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

$$c(m \mid n,t,x_t) = \begin{cases} 3x_t + 1 \cdot \min\{n+x_t, 2\} & \text{if } m = \min\{n+x_t, 2\} & d_t = 0 \text{ w.p. } \frac{1}{4} \\ 3x_t + 1 \cdot \max\{n+x_t-2, 0\} & \text{if } m = \max\{n+x_t-2, 0\} & d_t = 2 \text{ w.p. } \frac{3}{4} \\ \quad + 5 \cdot \max\{\underline{2-(n+x_t), 0}\} \\ \text{undefined} & \text{o/w} \end{cases}$$

$$\begin{cases} n+x_t \geq 2 \Rightarrow \text{demand can be met} \\ n+x_t < 2 \Rightarrow \text{unmet demand} = 2-(n+x_t) \end{cases}$$

- Allowable decisions x_t at stage t and state n:

Let x_t = number of batches to produce in month t

x_t must satisfy: $x_t \in \{0, 1\}$ for $t = 1, 2$; $n = 0, 1, 2$.

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

$f_t(n)$ = minimum total expected production and inventory cost for months $t, \ldots, 2$ w/ n batches in inventory.

for $t = 1, 2, 3$; $n = 0, 1, 2$

- Boundary conditions:

$$f_3(n) = 0 \qquad \text{for } n = 0, 1, 2$$

- Value-to-go recursion:

$$f_t(n) = \min_{x_t \in \{0,1\}} \left\{ \frac{1}{4}\left[3x_t + 1 \cdot \min\{n + x_t, 2\} + f_{t+1}\left(\min\{n + x_t, 2\}\right) \right] + \frac{3}{4}\left[3x_t + 1 \cdot \max\{n + x_t - 2, 0\} + 5\max\{2 - (n + x_t), 0\} + f_{t+1}\left(\max\{n + x_t - 2, 0\}\right) \right] \right\}$$

for $t = 1, 2$; $n = 0, 1, 2$

- Desired value-to-go function value: $f_1(1)$

2.4 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 value-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?

$t=1$ (month 1) : $n = 1 \Rightarrow \left.\begin{array}{l} f_1(1) = 8.875 \\ x_1^* = 1 \end{array}\right\} \Rightarrow$ produce 1 batch.

$t=2$ (month 2): If $n=2 \Rightarrow \left.\begin{array}{l} f_2(2) = 0.5 \\ x_2^* = 0 \end{array}\right\} \Rightarrow$ produce 0 batches

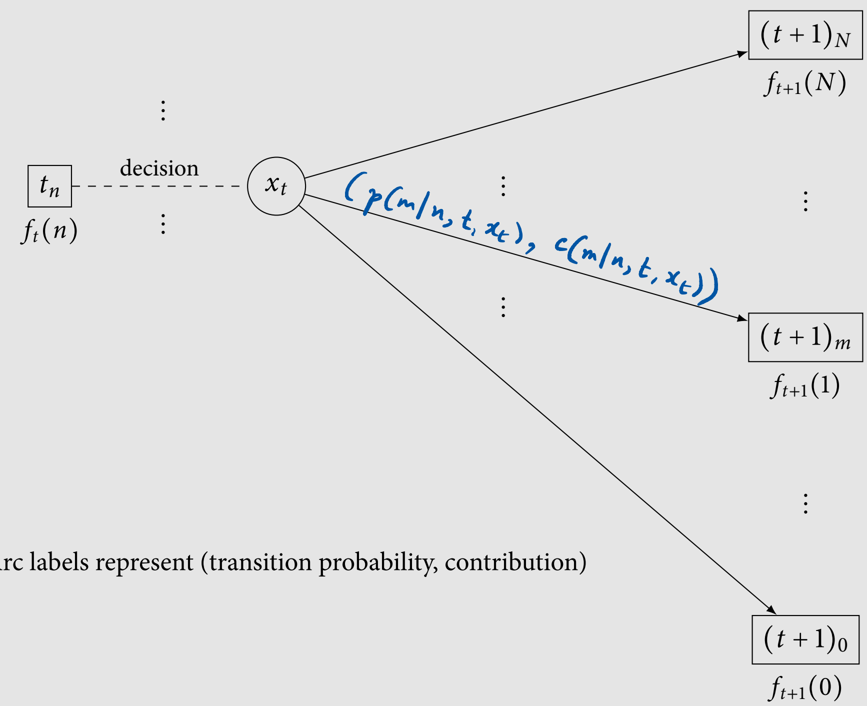
If $n=0 \Rightarrow \left.\begin{array}{l} f_2(0) = 7 \\ x_2^* = 1 \end{array}\right\} \Rightarrow$ produce 1 batch

- What is the company's total expected cost? $f_1(1) = 8.875$

3 Stochastic dynamic programs, more generally

Stochastic dynamic program

- **Stages** $t = 1, 2, \ldots, T$ and **states** $n = 0, 1, 2, \ldots, N$
- 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



Arc labels represent (transition probability, contribution)

- **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

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

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

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

4 A precision manufacturing problem

Example 2. The Hit-and-Miss Manufacturing Company has received an order to supply one item of a particular type. However, manufacturing this item is difficult, and the customer has specified such stringent quality requirements that the company may have to produce more than one item to obtain an item that is acceptable.

The company estimates that each item of this type will be acceptable with probability 1/2 and defective with probability 1/2. Each item costs \$100 to produce, and excess items are worthless. In addition, a setup cost of \$300 must be incurred whenever the production process is setup for this item. The company has time to make no more than 3 production runs, and at most 5 items can be produced in each run. If an acceptable item has not been obtained by the end of the third production run, the manufacturer is in breach of contract and must pay a penalty of \$1600.

The objective is to determine how many items to produce in each production run in order to minimize the total expected cost.

4.1 Warm up

- Suppose the manufacturer produces x items in a single production run.
- What is the probability that at least one of these items is acceptable?

$$\Pr\{\#\text{acceptable} \geq 1\} = 1 - \Pr\{\#\text{acceptable} = 0\} = 1 - \left(\tfrac{1}{2}\right)^x$$

- What is the expected number of acceptable items?

$$E[\#\text{acceptable}] = \tfrac{1}{2}x$$

4.2 Modeling the problem

- Stages:

$$\text{Stage } t \longleftrightarrow \begin{cases} \text{production run } t & (t = 1, 2, 3) \\ \text{end of decision-making process} & (t = 4) \end{cases}$$

- States:

$$\text{State } n \longleftrightarrow n \text{ acceptable items still needed} \quad (n = 0, 1)$$

- Allowable decisions x_t at stage t and state n:

Let x_t = number of items to produce in production run t

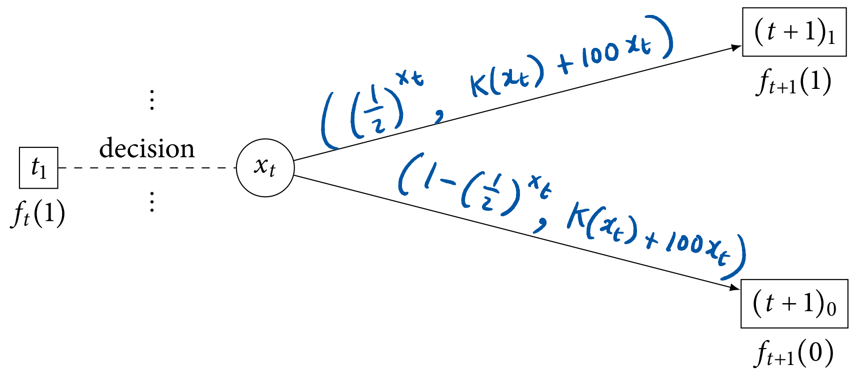
x_t must satisfy: $x_t \in \{0, 1, 2, 3, 4, 5\}$ for $t = 1, 2, 3$

- Sketch of basic structure:

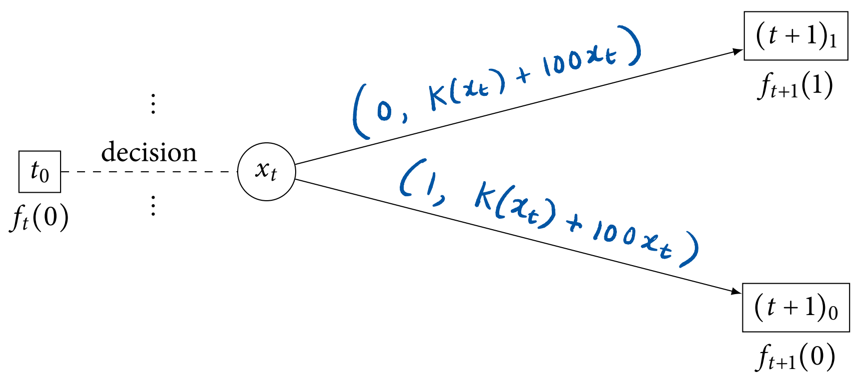
  - When the state $n = 1$:

Define:

$$K(x_t) = \begin{cases} 300 & \text{if } x_t > 0 \\ 0 & \text{o/w} \end{cases}$$



  - When the state $n = 0$:



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

$f_t(n)$ = minimum total expected cost for production runs $t, \ldots, 3$ starting with n acceptable items still needed

for $t = 1, 2, 3, 4$
$n = 0, 1$

- Boundary conditions:

$f_4(0) = 0$ $\quad$ $f_4(1) = 1600$

- Value-to-go recursion:

$$f_t(n) = \min_{x_t \text{ allowable}} / \max \left\{ \sum_{m \text{ state}} p(m \mid n, t, x_t) \Big[c(m \mid n, t, x_t) + f_{t+1}(m) \Big] \right\} \quad \text{for stages } t \text{ and states } n$$

$$f_t(1) = \min_{x_t \in \{0,1,2,3,4,5\}} \left\{ \left(\tfrac{1}{2}\right)^{x_t} \left[K(x_t) + 100x_t + f_{t+1}(1) \right] + \left(1 - \left(\tfrac{1}{2}\right)^{x_t}\right) \left[K(x_t) + 100x_t + f_{t+1}(0) \right] \right\}$$

for $t = 1, 2, 3$

$$f_t(0) = \min_{x_t \in \{0,1,2,3,4,5\}} \left\{ 0 \left[K(x_t) + 100x_t + f_{t+1}(1) \right] + 1 \left[K(x_t) + 100x_t + f_{t+1}(0) \right] \right\}$$

for $t = 1, 2, 3$

- Desired value-to-go function value: $f_1(1)$

4.3 Interpreting the value-to-go function

- Solving the recursion, we get the following value-to-go function values $f_t(n)$ for $t = 1, 2, 3$ and $n = 0, 1$, as well as the decision x_t^* that attained each value:

t	n	$f_t(n)$	x_t^*
1	0	0	0
1	1	675	2
2	0	0	0
2	1	700	2
3	0	0	0
3	1	800	3

- Based on this, what should the company's policy be?

Run 1: produce 2

Run 2: If $n = 0$, produce 0
If $n = 1$, produce 2

Run 3: If $n = 0$, produce 0
If $n = 1$, produce 3

- What is the company's total expected cost? $f_1(1) = 675$

5 A small investment problem

Example 3. Suppose you have \$5,000 to invest. Over the next 3 years, you want to double your money. At the beginning of each of the next 3 years, you have an opportunity to invest in one of two investments: A or B. Both investments have uncertain profits. For an investment of \$5,000, the profits are as follows:

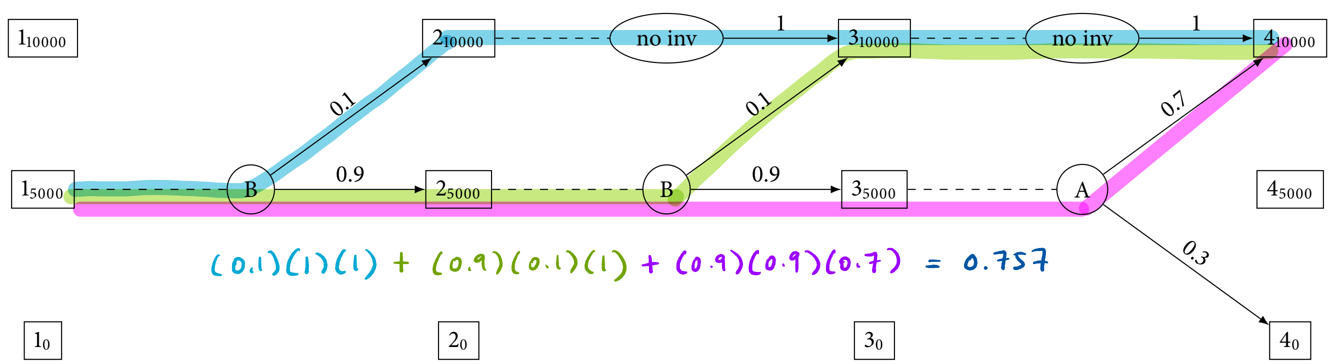
Investment	Profit (\$)	Probability
A	−5,000	0.3
	5,000	0.7
B	0	0.9
	5,000	0.1

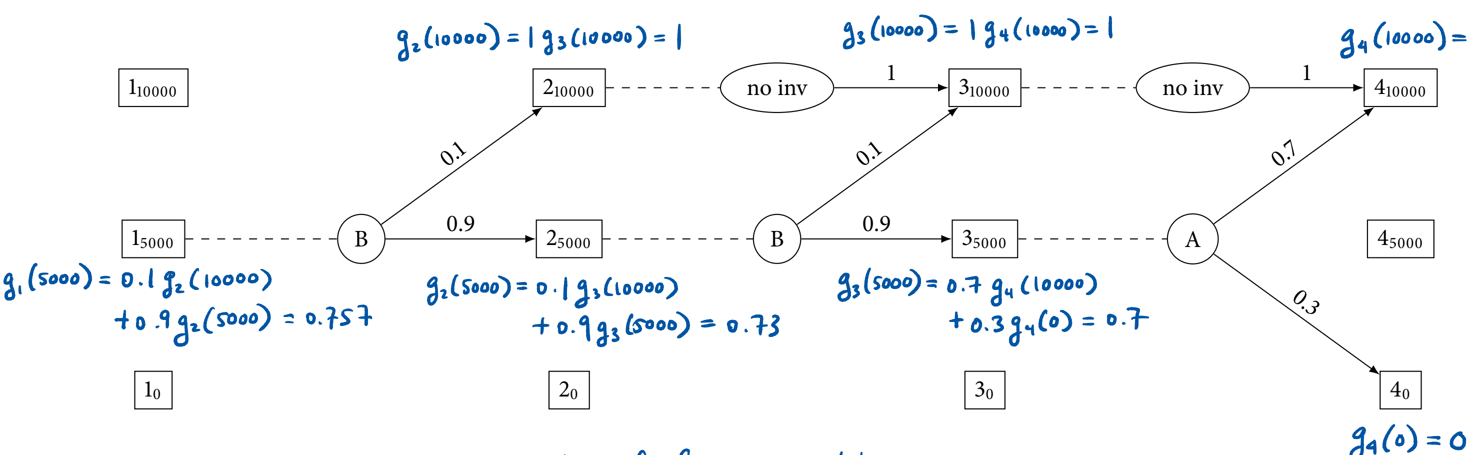
You are allowed to make at most one investment each year, and can invest only \$5,000 each time. Any additional money accumulated is left idle. Once you've accumulated \$10,000, you stop investing.

Formulate a stochastic dynamic program to find an investment policy that maximizes the probability you will have \$10,000 after 3 years.

5.1 Warm up

Consider the following investment policy. What is the probability of having at least \$10,000?





$g_t(n)$ = prob. of finishing w/ \$10000 starting at year t w/ n dollars

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

5.2 Formulating the stochastic dynamic program

- Stages:

$$\text{Stage } t \longleftrightarrow \begin{cases} \text{beginning of year } t & (t = 1, 2, 3) \\ \text{end of the decision-making process} & (t = 4) \end{cases}$$

- States:

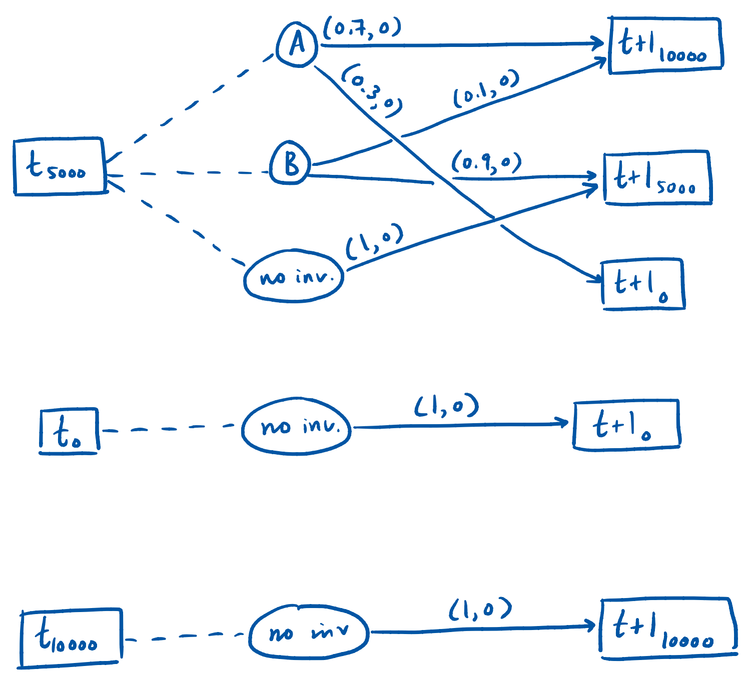
$$\text{State } n \longleftrightarrow \text{having } n \text{ dollars} \qquad (n = 0, 5000, 10000)$$

- Allowable decisions x_t at stage t and state n:

Let x_t = investment to make in stage t

$$x_t \text{ must satisfy:} \qquad x_t \in \begin{cases} \{A, B, \text{no inv.}\} & \text{if } n = 5000 \\ \{\text{no inv.}\} & \text{if } n = 0, 10000 \end{cases}$$

- Sketch of basic structure – transition probabilities and contributions:



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

$f_t(n)$ = maximum probability of finishing w/ \$10000 starting at year t w/ n dollars

for $t = 1, 2, 3, 4$
$n = 0, 5000, 10000$

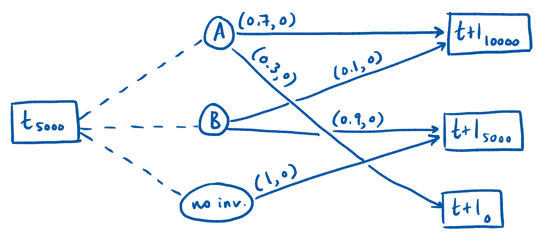
- Value-to-go recursion

$$f_t(n) = \min_{x_t \text{ allowable}} / \max \left\{ \sum_{m \text{ state}} p(m \mid n, t, x_t)\Big[c(m \mid n, t, x_t) + f_{t+1}(m)\Big] \right\} \quad \text{for stages } t \text{ and states } n$$

$$f_t(5000) = \max \Big\{ \underbrace{0.7 f_{t+1}(10000) + 0.3 f_{t+1}(0)}_{x_t = A},\ \underbrace{0.1 f_{t+1}(10000) + 0.9 f_{t+1}(5000)}_{x_t = B},\ \underbrace{1 f_{t+1}(5000)}_{x_t = \text{no inv}} \Big\}$$

$$f_t(0) = \max \Big\{ \underbrace{1 f_{t+1}(0)}_{x_t = \text{no inv}} \Big\} = f_{t+1}(0)$$

$$f_{t+1}(10000) = \max \Big\{ \underbrace{1 f_{t+1}(10000)}_{x_t = \text{no inv}} \Big\} = f_{t+1}(10000)$$



- Boundary conditions:

$f_4(10000) = 1$ $\qquad f_4(5000) = 0$ $\qquad f_4(0) = 0$

- Desired value-to-go function value: $f_1(5000)$

5.3 Interpreting the value-to-go function

- Solving the recursion on $f_t(n)$, we obtain:

t	n	$f_t(n)$	x_t^*
1	0	0	no investment
1	5000	0.757	B
1	10000	1	no investment
2	0	0	no investment
2	5000	0.73	B
2	10000	1	no investment
3	0	0	no investment
3	5000	0.7	A
3	10000	1	no investment

- Based on this, what should your investment policy be?

Year 1: invest in B

Year 2: If $n = 5000$, invest in B
If $n = 10000$, no inv.

Year 3: If $n = 5000$, invest in A
If $n = 10000$, no inv.

- What is your probability of having $10,000?

$$f_1(5000) = 0.757$$

A Problems

Problem 1 (Baytheon)**.** Baytheon has received an order to supply 2 guided missiles. In order to meet stringent quality requirements, the company may have to manufacture more than one missile to obtain an missile that is acceptable. The company has time to make no more than 3 production runs, and at most 2 missiles can be produced in each run. The probability distribution of acceptable missiles in a given run depends on how many missiles are produced:

	Probability of acceptable missiles		
Number of missiles produced	0	1	2
0	1	0	0
1	1/3	2/3	0
2	1/4	1/2	1/4

Each missile costs \$100,000 to produce, and excess missiles are worthless. In addition, a setup cost of \$50,000 must be incurred whenever the production process is setup for this item. If 2 acceptable missiles have not been obtained by the end of the third production run, Baytheon is in breach of contract and must pay a penalty of \$25,000. The objective is to determine how many missiles to produce in each production run in order to minimize the total expected cost.

Formulate this problem as a stochastic dynamic program.

Problem 2 (Farkas Investments)**.** You have recently been hired as a junior analyst at Farkas Investments. You have been given \$4 million to invest over the next 3 years. At the beginning of each of the next 3 years, you can invest in one of two investments: A or B.

Investment	Cost (\$ millions)	Profit (\$ millions)	Probability
A	3	2	0.5
		-2	0.5
B	5	3	0.1
		-1	0.9

You are allowed to make at most one investment each year. Any additional money accumulated is left idle. You may not borrow money to invest; that is, you cannot buy into an investment if it costs more than you currently have.

Formulate a stochastic dynamic program to find an investment policy that maximizes the probability you will have at least \$10 million at the end of 3 years.