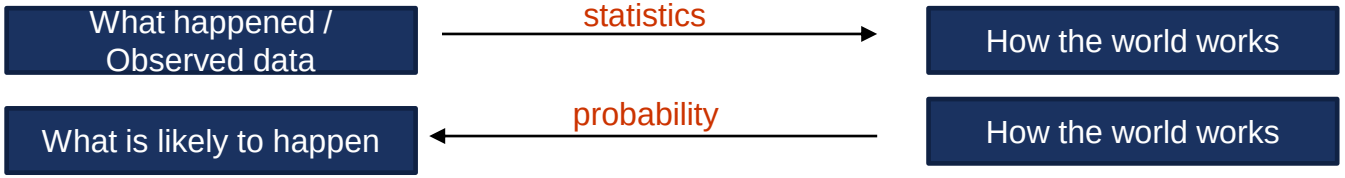
LECTURE 4
MONTE CARLO SIMULATION BY HAND

AGENDA

- Randomness
- Random Observations (or Variates)
- Random Numbers
- Monte Carlo Simulation

BASIC STATISTICS

- Statistics is the art and science of collecting and understanding data. It is the inverse of probability.



- Sample
  - A set of data that is randomly collected from a population in order to know how the world works (e.g., to understand more about Hong Kong population).
- Sample mean
  - The average of the data of a sample serves as an estimate to the mean of the associated population.
  - Sample mean itself is a random variable, because different samples yield different sample means.
- Confidence interval
  - An interval characterized by a probability (usually, 90%, 95%, 99%), which quantifies how likely it includes a population parameter (e.g. the mean of a population).

MOTIVATION

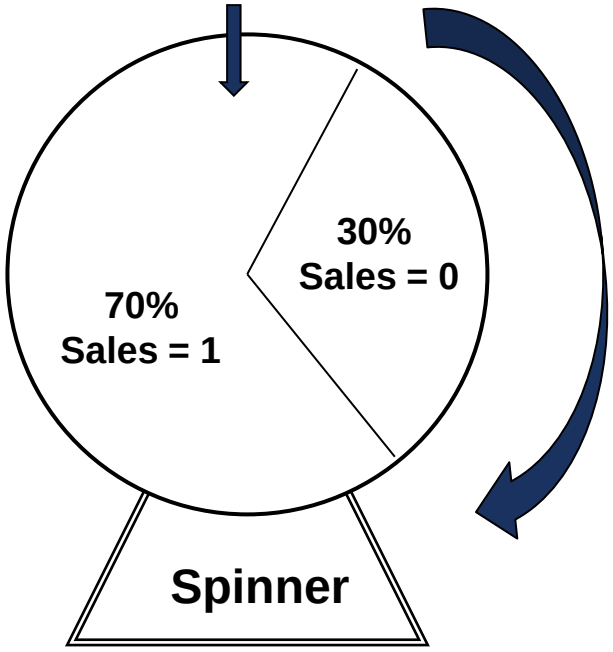
- Suppose you want to purchase a boutique that sells hand-made leather jackets.
- The current owner sells the jackets at $7,000 each.
- Daily operating cost is $5,500.
- Past records tells us that on each day:
  - One customer arrives to purchase a jacket with probability 70%.
  - No customer arrives with probability 30%.
- Do you expect the boutique to be profitable in the long-run?

PROBABILITY DISTRIBUTION OF NET INCOME

	Probability
Sell 1 jacket Income = $1,500	70%
Sell 0 jackets Income = $-5,500	30%

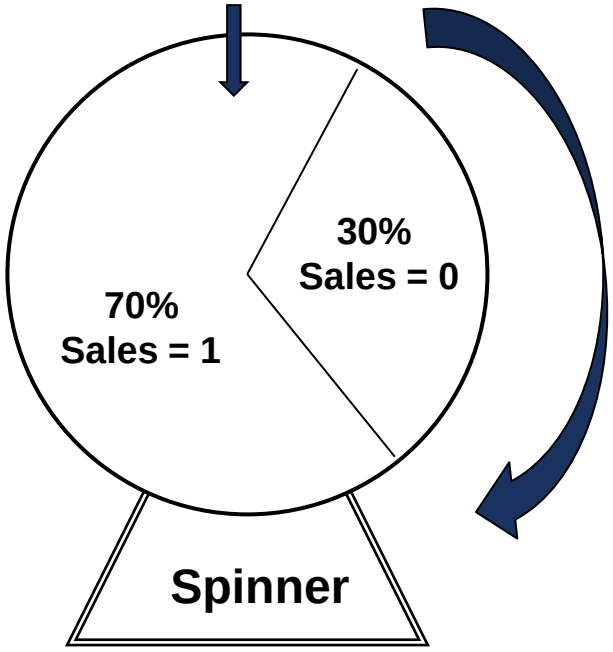
- We know: revenue = $7,000 per jacket; daily expenses = $5,500
- Is there any tool which can help you decide whether the boutique is expected to be profitable in the long-run?
- Possibility: Observe the boutique's business for the next year?

THE SPINNER



- Turning the spinner (as a simulator) once to "predict" sales for 1 day.
- Regard "sales" as a random variable.
- Get "1 unit sold" or "0 units sold" as a random observation (or variate).

THE SPINNER



- Suppose instead that the owner sells TWO versions of the jacket:
  - Standard = $7,000
  - Premium = $10,000
- 90% of customers would buy the standard jacket. The remaining customers purchase the premium jacket.
- How do you modify the spinner to include these two properties?

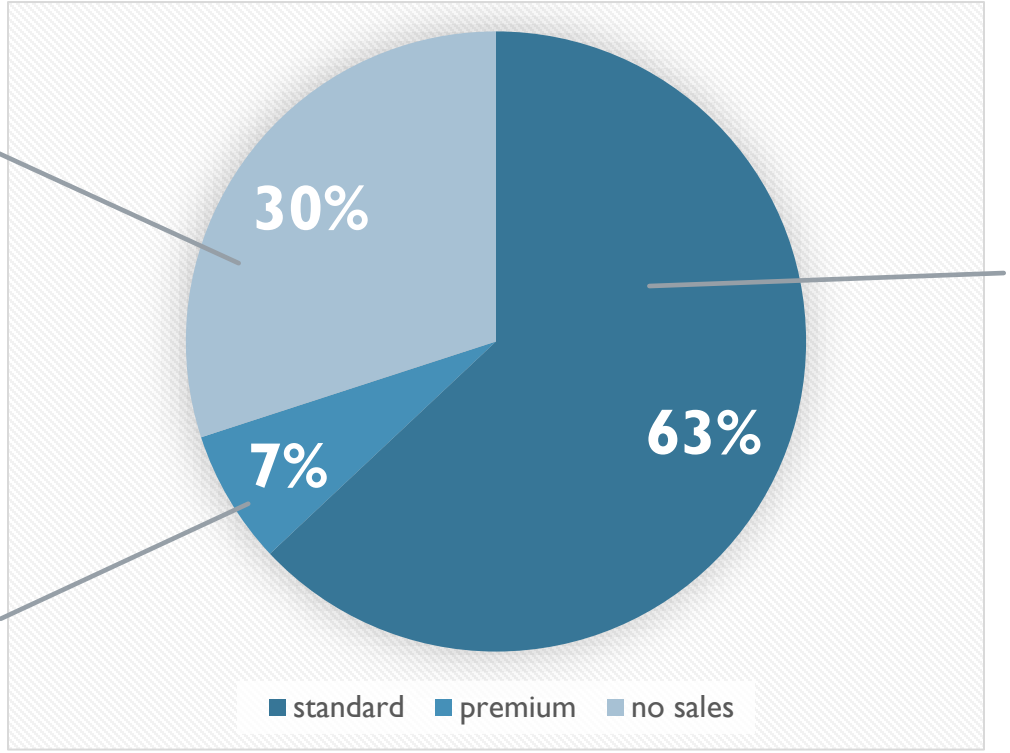
PROBABILITY DISTRIBUTION OF NET INCOME

Prob. of selling 1 or 0 jackets		Prob. of selling style X, conditional on purchase		Joint prob.
Sell1 jacket	70%	Standard Income = $1,500	90%	63%
		Premium Income = $4,500	10%	7%
Sell 0 jackets Income = $-5,500	30%	n/a		30%

- Recall: daily expenses $5,500
  - Sell standard jacket: $7,000 – 5,500 = 1,500
  - Sell premium jacket: $10,000 – 5,500 = 4,500
  - Sell 0 jackets: $0 – 5,500 = -5,500

THE SPINNER – METHOD 1

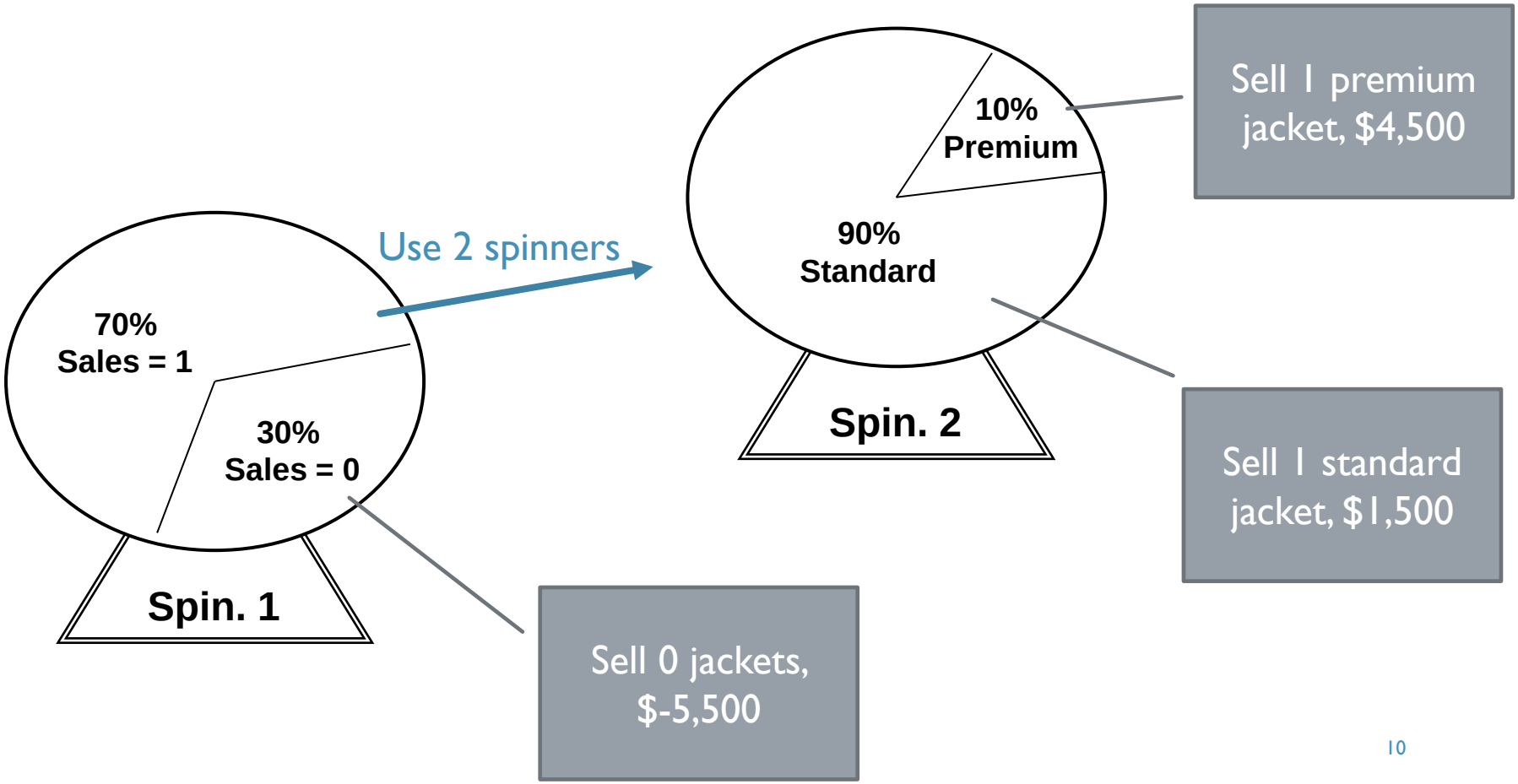
Sell 0 jackets, $-5,500



Sell 1 standard jacket, $1,500

Sell 1 premium jacket, $4,500

THE SPINNER – METHOD 2



GENERATION OF RANDOMNESS AND RANDOM OBSERVATIONS

- Generation of randomness:
  - A fundamental question in simulation: Given a probability distribution, how do you generate random observations (or variates)?
  - You aim to predict what / which is likely to happen, given some knowledge about part of a system.
- Generation of random observations:
  - A spinner is a "random observation generator" or "random variate generator".

RANDOM NUMBERS

- A spinner is a "random observation generator".
- Turning a spinner to generate "random observations" is not practical.
- Instead, you can use "random numbers".
- For example, you can use 100 double-digit numbers, namely 00 – 99.
  - Align the distribution of 00 – 99 numbers with the distribution of sales.

RANDOM NUMBERS

(Joint) Prob. of events		CUMULATIVE probability	Random number assignment
Standard Income = $1,500	63%	63%	00-62
Premium Income = $4,500	7%	70%	63-69
Sell 0 jackets Income = $-5,500	30%	100%	70-99

- Prepare a "black-box" that contains 100 pieces of folded paper. Write the numbers 00, 01, 02, …, 99 on each piece of paper.
- "Randomly" select one piece of paper from the box. The corresponding number represents the event according to the pre-set distribution of random numbers. Return the piece to the box afterwards.
- If our chosen piece of paper says "14", then we "sold" a standard jacket.

RANDOM NUMBERS

(Joint) Prob. of events		CUMULATIVE probability	Random number assignment
Standard Income = $1,500	63%	63%	01-63
Premium Income = $4,500	7%	70%	64-70
Sell 0 jackets Income = $-5,500	30%	100%	71-00

- Another possible assignment of random numbers:
  - Start first row at 01.
  - Treat 00 as "100".
  - Then upper limit of each range is equal to the cumulative probability.

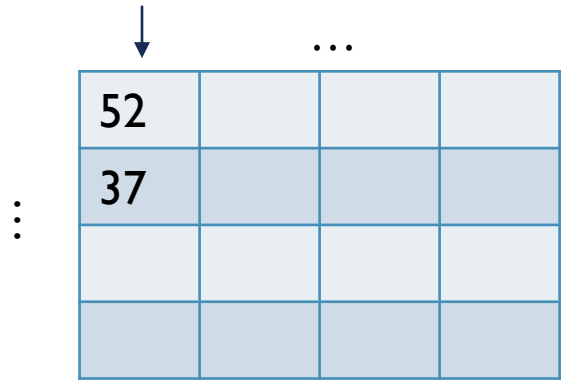
RANDOM NUMBERS

(Joint) Prob. of events		CUMULATIVE probability	Random number assignment
Standard Income = $1,500	63%	63%	01-53 91-00
Premium Income = $4,500	7%	70%	54-60
Sell 0 jackets Income = $-5,500	30%	100%	61-90

- Any assignments of 00 to 99 is allowed, as long as we match the probabilities of each event occurring

RANDOM NUMBER TABLE

- A person mimics this process of selection for 306 (= 17 rows×18 columns) times, and tabulates all 306 selected "2-digit random numbers" column by column.
- In particular, he places
  - 1st random number "52" in entry (1,1)
  - 2nd "37" in (2,1)
  - …
  - 17th "30" in (17,1)
  - 18th "06" in (1,2)
  - …
  - 306th "73" in (17,18).



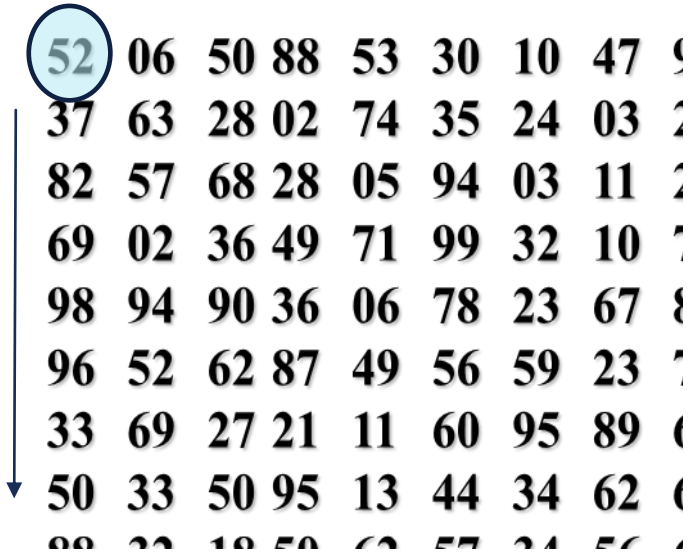
- The table (shown on next slide) is regarded as a "2-digit random number table".
- This process can be carried on to produce other pages of 2-digit random numbers.

RANDOM NUMBER TABLE

52	06	50	88	53	30	10	47	99	37	66	91	35	32	00	84	57	07
37	63	28	02	74	35	24	03	29	60	74	85	90	73	59	55	17	60
82	57	68	28	05	94	03	11	27	79	90	87	92	41	09	25	36	77
69	02	36	49	71	99	32	10	75	21	95	90	94	38	97	71	72	49
98	94	90	36	06	78	23	67	89	85	29	21	25	73	69	34	85	76
96	52	62	87	49	56	59	23	78	71	72	90	57	01	98	57	31	95
33	69	27	21	11	60	95	89	68	48	17	89	34	09	93	50	44	51
50	33	50	95	13	44	34	62	64	39	55	29	30	64	49	44	30	16
88	32	18	50	62	57	34	56	62	31	15	40	90	34	51	95	26	14
90	30	36	24	69	82	51	74	30	35	36	85	01	55	92	64	09	85
50	48	61	18	85	23	08	54	17	12	80	69	24	84	92	16	49	59
27	88	21	62	69	64	48	31	12	73	02	68	00	16	16	46	13	85
45	14	46	32	13	49	66	62	74	41	86	98	92	98	84	54	33	40
81	02	01	78	82	74	97	37	45	31	94	99	42	49	27	64	89	42
66	83	14	74	27	76	03	33	11	97	59	81	72	00	64	61	13	52
74	05	82	82	93	09	96	33	52	78	13	06	28	30	94	23	37	39
30	34	87	01	74	11	46	82	59	94	25	34	32	23	17	01	58	73

USE OF RANDOM NUMBER TABLE

Event	Cumulative Probability	Random number assignment
Standard Income = $1,500	0.63	01-63
Premium Income = $4,500	0.70	64-70
Sell 0 jackets Income = $-5,500	1.00	71-00



- We can use this table to simulate future sales.
- For the first run (or attempt), the random number "52" is in the range of "01-63". So a standard jacket is sold and the boutique earns $1,500.
- For the second run, "37" → "standard jacket, $1,500".

USE OF RANDOM NUMBER TABLE

Event	Cumulative Probability	Random number assignment
Standard Income = $1,500	0.63	01-63
Premium Income = $4,500	0.70	64-70
Sell 0 jackets Income = $-5,500	1.00	71-00

Day		Event		Income
1	52	standard		1,500
2	37	standard		1,500
3	82	no sales	-	5,500
4	69	premium		4,500
5	98	no sales	-	5,500
6	96	no sales	-	5,500
7	33	standard		1,500
8	50	standard		1,500
9	88	no sales	-	5,500
10	90	no sales	-	5,500
		Average	-	1,700

- Repeat this process for 10 rounds.
- We would SIMULATE the sales over 10 days.
- What happens if we use another column of the random number table?

ALTERNATIVE SOLUTION

- Simulate # of jackets sold on each day. For each sale, simulate the jacket style → Recall Spinner Method 2.
- Use column 1 of random number table for # of sales.
- Use column 2 for jacket style.
- Step 1:

# of jackets sold	Cumulative Probability	Random number assignment
1 jacket	0.70	01-70
0 jackets	1.00	71-00

Day	# of jackets		Income
1	52	1	
2	37	1	
3	82	0	- 5,500
4	69	1	
5	98	0	- 5,500
6	96	0	- 5,500
7	33	1	
8	50	1	
9	88	0	- 5,500
10	90	0	- 5,500

ALTERNATIVE SOLUTION

- Step 2:

Jacket style	Cumulative Probability	Random number assignment
Standard	0.90	01-90
Premium	1.00	91-00

Day	# of jackets	Jacket style	Income
1	1	standard	1,500
2	1	standard	1,500
3	0		- 5,500
4	1	standard	1,500
5	0		- 5,500
6	0		- 5,500
7	1	standard	1,500
8	1	premium	4,500
9	0		- 5,500
10	0		- 5,500
		Average	- 1,700

06
63
57
02
94
52
69
33
32
30

GENERATION OF RANDOMNESS

- A spinner or a random number table is a "random observation generator".
- Using a spinner to generate "random observations" (also called "random variates") is not practical.
- Instead, other tools operated by computers are developed.
- To generate random observations includes
  - The generation of random numbers which are uniformly distributed over [0, 1) and,
  - The generation of random observations which aligns with the random numbers.
- Why use this approach?
  - Creating "randomness" is a difficult task.
  - Given the random numbers, the corresponding random observations are relatively easy to obtain.
  - A unified approach.

USE OF SIMULATION

- We should NOT use one round of simulation to estimate tomorrow's income (i.e. one day's income).
- Run the simulation MANY times to estimate long-run's average income.

EXAMPLE 1: MONTE CARLO SIMULATION

- Higgins Plumbing and Heating maintains a stock of 30-gallon hot water heaters that it sells and installs for homeowners. Jerry Higgins (the owner) likes the idea of having a large supply on hand to meet customer demand, but he also recognizes that it is expensive to do so. He examines hot water heater sales over the past 50 weeks and notes the following:

Sales of heaters (per week)	4	5	6	7	8	9	10
# of weeks with listed sales	6	5	9	12	8	7	3

EXAMPLE 1

- Question: Using Monte Carlo simulation, what is the average sales per week (including shortages) over the 20-week period?
- We will use 17 random numbers in column 7 (i.e. (1,7) to (17, 7)) and 3 random numbers in column 8 (i.e. (1,8) to (3,8)) of our 2-digit random number table to conduct the simulation.
- Answer: Total sales / 20 weeks = _____

EXAMPLE 1

- Step 1: Calculate the cumulative probability based on historical data and assign the corresponding 2-digit random numbers

Heater Sales	Probability	Cumulative Probability	Random Number Assignment
4	6/50=0.12		
5	5/50=0.10		
6	9/50=0.18		
7	12/50=0.24		
8	8/50=0.16		
9	7/50=0.14		
10	3/50=0.06		
Total	**1.00**		

EXAMPLE 1

- Step 2: Prepare a Monte Carlo simulation table using random numbers selected from column 7 to column 8.
- Average sales using sales = total sales/20 weeks.

Week	Random number	Simulated sales	Week	Random number	Simulated sales
1			11		
2			12		
3			13		
4			14		
5			15		
6			16		
7			17		
8			18		
9			19		
10			20		
				Total	

EXAMPLE 1

- Question: Using an analytic technique, what is the theoretical average sales (or called expected sales) per week? How does this compare with the simulated answer?
- Answer:

$$\mathrm{E[sales]} = 0.12 \times 4 + 0.10 \times 5 + 0.18 \times 6 + 0.24 \times 7 + 0.16 \times 8 + 0.14 \times 9 + 0.06 \times 10$$

$$= 6.88 \text{ heaters per week}$$

- With a larger number of runs (e.g. 100 weeks) in a simulation, these two methods lead to closer values.
- In general, the larger the number of runs in a simulation, the closer the value to the actual or theoretical one.

EXAMPLE 1

- Question: If Jerry maintains a constant supply of 8 hot water heaters in any given week, how many times will he be out of stock during the above-simulated 20 weeks?
- Answer: Jerry will be out of stock three times (in weeks 7, 14, and 16).
- Notes:
  1. Jerry stores unsold heaters in a week and tops up to 8 heaters for the following week, e.g. Jerry stores 4 unsold heaters in week 1 and orders 4 more for week 2.
  2. All homeowners wait when shortages occur in a week, and Jerry orders the sum of shortages and 8 heaters for the following week, e.g. Jerry orders 10 heaters for week 8.

LEARNING POINTS AFTER THE EXAMPLE: VARIANTS OF USING RANDOM NUMBER TABLES

- There are different kinds of random number tables.
- Different assignments of random-number ranges yield different results (even using the same random number table).
- Method: multiple simulations.

WHY SIMULATION INSTEAD OF ANALYTICAL SOLUTION?

- In the previous example, it was easy to compute the expected number of heaters sold.
- Simulation is useful when the reward/cost is hard to compute.
- Be careful with how you define your reward/cost function and the observations!

WHY SIMULATION INSTEAD OF ANALYTICAL SOLUTION?

- Suppose you are responsible for buying refreshments (cookies) for a company event. You want to know if a platter of 500 cookies will be enough.
- Based on past experiences, the number of employees that will show up can be described by the following probability distribution:

# of employees who arrive	Probability
100	15%
250	45%
450	35%
600	5%

- Q: How many cookies does each person get to eat on average?
  - Average # of employees = 15% x 100 + 45% x 250 + 35% x 450 + 5% x 600 = 315, so average # of cookies per employee = 500/315 = 1.59? WRONG

WHY SIMULATION INSTEAD OF ANALYTICAL SOLUTION?

- Do 10 simulation runs by using the last column of the random number table.
- A simulation trial should report the **observed # of cookies per person** rather than the number of people at the event.

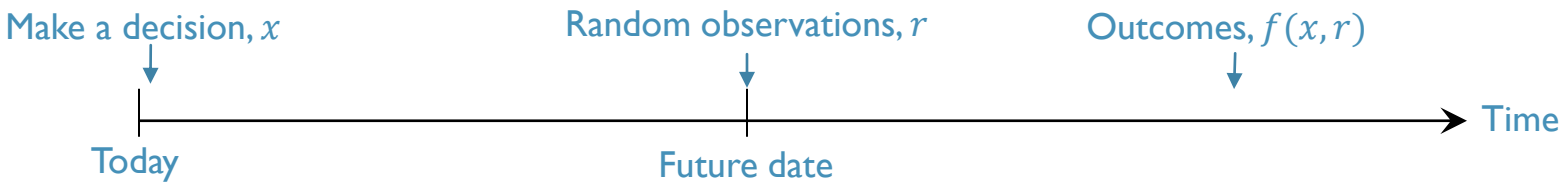
# of employees who arrive	Cumulative Probability	Random number assignment
100	15%	01-15
250	60%	16-60
450	95%	61-95
600	100%	96-00

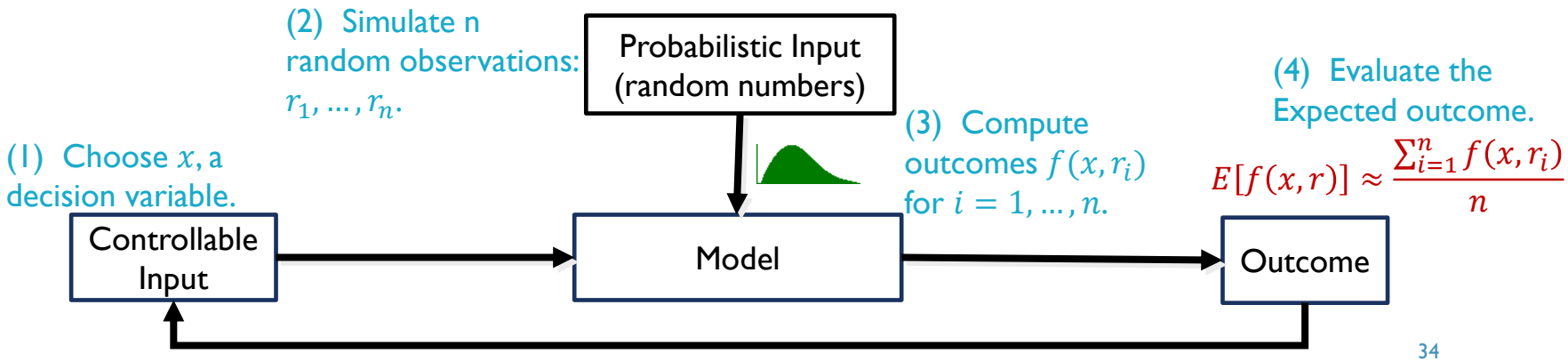
Run	Rand #	# of emp.	# of cookies/emp.
1	7	100	5.00
2	60	250	2.00
3	77	450	1.11
4	49	250	2.00
5	76	450	1.11
6	95	450	1.11
7	51	250	2.00
8	16	250	2.00
9	14	100	5.00
10	85	450	1.11
Average		300.00	2.24

07
60
77
49
76
95
51
16
14
85
59
85
40
42
52
39
73

DECISION MAKING USING SIMULATION

- A model simulates the behaviour of a (stochastic) system.





(5) If $\mathrm{E}[f(x,r)]$ is not a satisfactory result, then go to (1) and repeat for another value of x.

SIMULATION

- Simulation modeling process
  - Identify a decision variable x, and a random input r.
  - Given x and r, express the formula of an outcome $f(x, r)$.
  - Conduct a pre-specified number of runs in a simulation to generate random observations $r_1, r_2, \ldots, r_n$.
  - Compute the corresponding outcomes $f(x, r_1), \ldots, f(x, r_n)$.
  - Compute the summary statistics and make inferences about the problem under study.

$$E[f(x,r)] \approx \frac{\sum_{i=1}^{n} f(x, r_i)}{n}$$

APPLICATIONS

- Non-Business applications
  - Flight simulator
  - Aircraft design
  - War games in military
- Business applications
  - P&G uses simulation to hedge exchange rate risk.
  - GM uses simulation to forecast net income and predict purchasing costs.
  - Uber uses simulations to decide which drivers to dispatch to customers.[1]

[1] Gaining Insights in a Simulated Marketplace with Machine Learning at Uber. 2019. Retrieved Dec. 4, 2019. https://eng.uber.com/simulated-marketplace/

SUMMARY

- What is simulation?
- Randomness
- Random observations (or variates)
- Random numbers
- Next week: Monte Carlo simulation by Excel