15.063: Communicating with Data

Summer 2003

Recitation 2: Probability

Today’s Goal

Review

Laws of Probability

and

Discrete Random Variables

Content

- Laws of Probability
- Conditional Probability: Problem 2.7
- Independence: Rolling two dice
- Discrete Random Variables

Events

- An ***event*** is a collection of ***outcomes***
- ***First law of probability*** : the probability of an event is between **0** and **1**
- Two events are ***disjoint*** when they have no ***common outcome***.

  Question : consider tossing a quarter and a penny. Let event **A** be "the quarter landed heads" and event **B** be "the penny landed heads". Are they disjoint?

Laws of Probability

- p(**A** or **B**) = p(**A**) + p(**B**) when **A** and **B** are *disjoint*.
  Question : If the probability of a plane crash somewhere in the world during a whole year is 0.0001, is the probability of a crash during 2 years 0.0002?
- p(**A** and **B**) = p(**A**) p(**B**) iff **A** and **B** are *independent*.
  Question : If different years are independent, what is the probability of no crash in 2 years ?
- *Remember* : **if independent $\Rightarrow$ not disjoint**

Conditional Probability

- For events **A** and **B**, the probability that **A** occurs given that **B** occurred is:

  $$p(\mathbf{A} / \mathbf{B}) = p(\mathbf{A} \,\&\, \mathbf{B}) / p(\mathbf{B})$$

  Note : independence not required

- *Question* : Given that a plane did not crash the first year, what is the probability of a crash during the second year?

Conditional Probability

- See problem 2.7 in the course textbook:

Data, Models, and Decisions: The Fundamentals of Management Science by Dimitris Bertsimas and Robert M. Freund, Southwestern College Publishing, 2000.

Rolling two dice

- *Exercise* : Two fair six-sided dice are tossed.

  Find the probability of:
  - a. the sum of the dice is exactly 2
  - b. the sum of the dice exceeds 2
  - c. both dice come up with the same number
  - d. event (a) occurs given that event (c) does
  - e. both dice come up with odd numbers

Independence

- Two events **A** and **B** are said to be ***independent*** when

$$p(\mathbf{A} \cap \mathbf{B}) = p(\mathbf{A})\, p(\mathbf{B}) \, .$$

- Or, using the definition of conditional probability:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)P(B)}{P(B)} = P(A)$$

- *Intuition* : The fact that you know that event B happened, does not change the likelihood of event A.

Independence

- *Example* : For the two dice, is the event of getting a double independent of the number of the first die?
- Let **D** be the event of obtaining a double.
  Let **A** and **B** be the numbers of the first and second die.
- The possible outcomes are the following 36 equally likely combinations:

11	12	13	14	15	16
21	22	23	24	25	26
31	32	33	34	35	36
41	42	43	44	45	46
51	52	53	54	55	56
61	62	63	64	65	66

The probability of each outcome is 1/36

Independence

- p(**D**)= p(A=1 & B=1) + ... + p(A=6 & B=6) = 6 (1/36)

  = 1/6.

- p(**D**/A=1) = p(B=1) = 1/6,

  p(**D**/A=2) = p(B=2) = 1/6 and so on.

- Therefore **p(D/A=i) = p(D)** for i=1,2,...,

  which means that the event of getting a double is *independent* of the outcome of the first die.

Independence

- If we are conducting this same experiment in Las Vegas, where its very hard to find a ‘fair’ die, the conclusion changes.
- Suppose the actual probability distribution of the two dice we will use is:
- The 36 possible outcomes are not equally likely, but we can compute their probability because the two dice are still independent.
- For example:

  $p(A=2 \ \& \ B=5)=p(A=2)\ p(B=5) = 0.2 \times 0.1$

  $= 0.02$

xi	pi
1	0.1
2	0.2
3	0.3
4	0.2
5	0.1
6	0.1

Independence

- Now, $p(D) = p(A=1 \& B=1) + \ldots + p(A=6 \& B=6)$

  $= p(A=1)\, p(B=1) + \ldots + p(A=6)\, p(B=6)$

  $= p(A=1)^2 + \ldots + p(A=6)^2$

  $= .1^2 + .2^2 + .3^2 + .2^2 + .1^2 + .1^2 = 0.2$

- Lets compute the conditional probabilities:

  $p(D/A=1) = p(B=1) = 0.1$,

  $p(D/A=2) = p(B=2) = 0.2$, and so on.

- As **$p(D/A=i) \neq p(D)$**, the likelihood of drawing a double *depends* on the outcome of the first die.

Discrete Random Variables

- A random variable assigns a value *(probability)* to each possible outcome of a probabilistic experiment.
- A discrete RV can take only distinct, separate values.
- Used to model discrete situations and compute expected values and variances.

Selling Newspapers

- A city newsstand has been keeping records for the past year of the number of copies of the newspapers sold daily. Records were kept for 200 days.

Number of copies	Frequency
0	24
1	52
2	38
3	16
4	37
5	18
6	13
7	2

Selling Newspapers

(a) What is the mean of the distribution?

Recall that $\mu_x = E(X) = \Sigma_i \, p_i x_i$

$$\mu_x = 0.12 \times 0 + 0.26 \times 1 + 0.19 \times 2 + 0.08 \times 16 + 0.185 \times 4 + 0.09 \times 5 + 0.065 \times 6 + 0.01 \times 7$$

= 2.53

Number of copies	Frequency	Probability p_i
0	24	0.120
1	52	0.260
2	38	0.190
3	16	0.080
4	37	0.185
5	18	0.090
6	13	0.065
7	2	0.010
		1.000

Selling Newspapers

(b) What is the standard deviation of the distribution?

Recall that $\mu_x = 2.53$ and

$$\sigma^2_x = VAR(X) = \Sigma\ p_i(x_i - \mu_x)^2$$

$$\sigma_x = \sqrt{[(0.12)(0 - 2.53)^2 + (0.26)(1-2.53)^2 + (0.19)(2-2.53)^2 + \ldots + (0.01)(7-2.53)^2]}$$

$$= 1.838$$

Number of copies	Frequency	Probability p_i
0	24	0.120
1	52	0.260
2	38	0.190
3	16	0.080
4	37	0.185
5	18	0.090
6	13	0.065
7	2	0.010
		1.000

Selling Newspapers

(c) Find the probability that at least 2 but no more than 6 copies are sold in a day.

$p(2 <= X <= 6)$

$= p(X=2) + ... + p(X=6)$

$= 0.19 + 0.08 + 0.185$

$+ 0.09 + 0.065$

$= 0.87$

Number of copies	Frequency	Probability p_i
0	24	0.120
1	52	0.260
2	38	0.190
3	16	0.080
4	37	0.185
5	18	0.090
6	13	0.065
7	2	0.010
		1.000

Binomial Distribution

- **Count** the number of times something happens
- Events have to be **repeated** and **independent**
- Allow us to compute **expectation**, **variance** and **probability of outcomes**
- Described by: **# trials** and **success probability**

Binomial Distribution

- *Example* : flipping a coin 10 times.
  RV number of tails is binomial
  - # trials: 10
  - success probability in each trial: 1/2
- *Question* : Is the number of aces we get in a poker hand a binomial RV?

Binomial Distribution

- *Example* : flipping a coin 10 times.

  X: RV number of tails is binomial(10,1/2)

  - E(**X**) = *np* = 10/2 = 5 (expected # of tails)
  - V(**X**) = *np* (1-*p*) = 10 / 4 = 2.5
  - stdev(**X**) = $\sqrt{np(1-p)} = \sqrt{2.5} = 1.6$
  - p(a single tail) = 5! / 4! x $(.5)^9 (.5)^1$

The End.