Review of Probability

CS1538: Introduction to Simulations

Probability and Statistics in Simulation

- Why do we need probability and statistics in simulation?
  - Needed to validate the simulation model
  - Needed to determine / choose the input probability distributions
    - Needed to generate random samples / values from these distributions
  - Needed to analyze the output data / results
  - Needed to design correct / efficient simulation experiments

Experiments and Sample Space

- Experiment
  - A process which could result in several different outcomes
- Sample Space
  - The set of possible outcomes of a given experiment

- Example:
  - Experiment: Rolling a single die
  - Sample Space: {1, 2, 3, 4, 5, 6}
- Another example?

Random Variables

- Random Variable
  - A function that assigns a real number to each point in a sample space
  - Example:
    - Let X be the value that results when a single die is rolled
    - Possible values of X are 1, 2, 3, 4, 5, 6
- Discrete Random Variable
  - A random variable for which the number of possible values is finite or countably infinite
    - Example above is discrete – 6 possible values
    - Countably infinite means the values can be mapped to the set of integers
      - Ex: Flip a coin an arbitrary number of times. Let X be the number of times the coin comes up heads

Random Variables and Probability Distribution

- ## Probability Distribution
  - For each possible value, x_i, for discrete random variable X, there is a probability of occurrence, $P(X = x_i) = p(x_i)$
  - $p(x_i)$ is the probability mass function (pmf) of X, and obeys the following rules:
    1) $p(x_i) >= 0$ for all I
    2) $\sum_{all\ i} p(x_i) = 1$

Random Variables and Probability Distribution

- Probability Distribution
  - The set of pairs $(x_i, p(x_i))$ is the probability distribution of X
  - Examples:
    - For the die example (assuming a fair die):
      - Probability Distribution:
        - {(1, 1/6), (2, 1/6), (3, 1/6), (4, 1/6), (5, 1/6), (6, 1/6)}

Cumulative Distribution

- Cumulative Distribution Function
  - The pmf gives probabilities for individual values x_i of random variable X
  - The cumulative distribution function (cdf), F(x), gives the probability that the value of random variable X is <= x, or

    $F(x) = P(X <= x)$
  - For a discrete random variable, this can be calculated simply by addition:

    $$F(x) = \sum_{x_i \leq x} p(x_i)$$

Cumulative Distribution

- Properties of cdf, F:
  1) F is non-decreasing
  2) $\lim_{x\to\infty}(F(x))=1$
  3) $\lim_{x\to-\infty}(F(x))=0$

and

$P(a < X \leq b) = F(b) - F(a)$ for all $a < b$

- Ex: Probability that a roll of two dice will result in a value > 7?
  - Discuss
- Ex: Probability that 10 flips of a fair coin will yield between 6 and 8 (inclusive) heads?
  - Discuss

Probability Distributions (Joint & Conditional)

- Joint – Probability that A=a ***and*** B=b

$$P(A=a, B=b)$$

- Conditional – Probability that A=a ***given*** B=b
  - Since we know B=b, what's the probability that A=a?

$$P(A=a \mid B=b) = \frac{P(A=a \cap B=b)}{P(B=b)} = \frac{P(A=a, B=b)}{P(B=b)}$$

Probability Distributions (Joint & Conditional)

- Pastries example (from survey)

Topping	Filling	Number
Powder	Apple	1
Powder	Cream	6
Sprinkles	Apple	3
Sprinkles	Cream	2

- What’s the probability that I get powder and cream if I pick at random?

Probability Distributions (Joint & Conditional)

- Pastries example (from survey)

Topping	Filling	Number
Powder	Apple	1
Powder	Cream	6
Sprinkles	Apple	3
Sprinkles	Cream	2

- What's the probability that I get cream if I can feel it has powder but I'm otherwise picking at random?

Independent & Dependent Variables

- Independent Variables
  - Changing one has no effect on the other
  - A and B are independent iff:
    - $P(A \cap B) = P(A)P(B)$

- Dependent Variables
  - Changing one has an effect on the other

Example Revisited

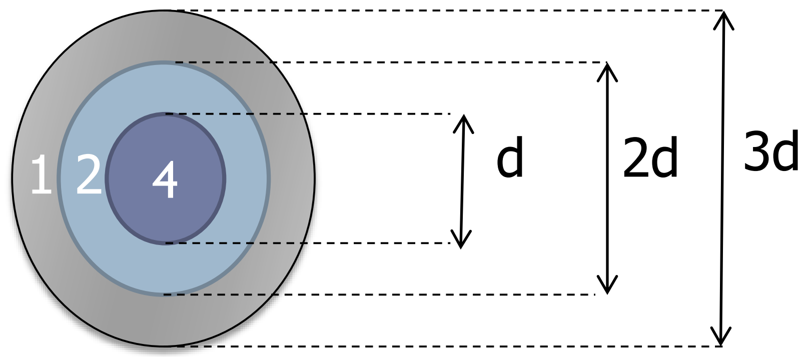
Probability that a roll of two dice will result in a value > 7?

Some Examples

- For the experiments below, provide the sample space and random variable(s)
  - Experiment: The arrival of the next customer in our single-server restaurant example
  - Experiment: The temperature in Pittsburgh on 12/1/2011

In Class Exercise

In a game of darts, one throws two darts at the board as configured below:



The score from the game is the product of the scores gotten from the two throws. Assume that when a person throws a dart, it always hits the board. Assume that the chance of a dart landing at any particular location is uniformly distributed with respect to the area of the dart board.

Set up the problem to figure out the chance that the person would score a 4 on his game.

Expected Value

- Expected Value – mean or average value
  - If X is a discrete random variable that can take on possible values x1, x2, …, then the expected value of X is defined as:

    $E[X] = \Sigma\, p(x_i) * x_i$
  - If X is a continuous random variable with a probability density function F, then the expectation is:

    $E[X] = \int_{-inf}^{inf} f(x) * x \; dx$
  - Notation wise, we often use the Greek letter μ to represent the mean. That is, $\mu = E[X]$

Expected Value

- $E[c] = c$
  - c is constant
- $E[c\,X] = c\,E[X]$
  - X is a random variable
- $E[X_1 + X_2 + X_n] = E[X_1] + E[X_2] + \ldots + E[X_n]$
  - $X_1, X_2, \ldots, X_n$, are all random variables

Variance

- Variance – how different is each r.v. outcome from the mean?
  - Expectation of how much difference there is between each possible outcome and the mean

$$Var(X) = E[\,(X - E[X])^2] = E[X^2] - (E[X])^2$$

Variance and Covariance

- Covariance of two random variables X, Y

  $$Cov(X,Y) = E[(X - \mu_x)(Y - \mu_y)] = E[XY] - E[X]E[Y]$$

- $Var(X + Y) = E[(X+Y - \mu_x - \mu_y)^2]$

  $$= Var(X) + Var(Y) + 2Cov(X,Y)$$

  - When X and Y are independent, $Cov(X,Y) = 0$

- The correlation between X, Y is defined as
  - $Corr(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}$