DSFA

Spring 2019

Lecture 12

Probability and Sampling

Announcements

- Homework 4: due Monday
-

Recipes

INGREDIENTS

2 1/4 cups all-purpose flour

1 teaspoon baking soda

1 teaspoon salt

1 cup (2 sticks) butter, softened

3/4 cup granulated sugar

3/4 cup packed brown sugar

1 teaspoon vanilla extract

2 large eggs

2 cups (12-oz. pkg.) NESTLÉ® TOLL HOUSE® Semi-Sweet Chocolate Morsels

1 cup chopped nuts

IN THIS RECIPE

INSTRUCTIONS

VIEW: TEXT VIDEO

PREHEAT oven to 375° F.

COMBINE flour, baking soda and salt in small bowl. Beat butter, granulated sugar, brown sugar and vanilla extract in large mixer bowl until creamy. Add eggs, one at a time, beating well after each addition. Gradually beat in flour mixture. Stir in morsels and nuts. Drop by rounded tablespoon onto ungreased baking sheets.

BAKE for 9 to 11 minutes or until golden brown. Cool on baking sheets for 2 minutes; remove to wire racks to cool completely.

PAN COOKIE VARIATION: Preheat oven to 350° F. Grease 15 x 10-inch jelly-roll pan. Prepare dough as above. Spread into prepared pan. Bake for 20 to 25 minutes or until golden brown. Cool in pan on wire rack. Makes 4 dozen bars.

SLICE AND BAKE COOKIE VARIATION:
PREPARE dough as above. Divide in half; wrap in waxed paper. Refrigerate for 1 hour or until firm. Shape each half into 15-inch log; wrap in wax paper. Refrigerate for 30 minutes.* Preheat oven to 375° F. Cut into 1/2-inch-thick slices; place on ungreased baking sheets. Bake for 8 to 10 minutes or until golden brown. Cool on baking sheets for 2 minutes; remove to wire racks to cool completely. Makes about 5 dozen cookies.

Recipe instructions

COMBINE flour, baking soda and salt in small bowl. Beat butter, granulated sugar, brown sugar and vanilla extract in large mixer bowl until creamy. Add eggs, one at a time, beating well after each addition. Gradually beat in flour mixture. Stir in morsels and nuts. Drop by rounded tablespoon onto ungreased baking sheets.

BAKE for 9 to 11 minutes or until golden brown. Cool on baking sheets for 2 minutes; remove to wire racks to cool completely.

Algorithm

Rules or *a recipe* for performing computation

Ideas we see in cookie recipe:

- Iteration: do something many times
- Conditionals: decide whether something is true, and maybe do something different
- Variability or randomness: some tasks might not be completely predictable

Control Statements

Control Statements

These statements *control* the sequence of computations that are performed in a program

- The keywords `if` and `for` begin control statements
- The purpose of `if` is to define computations that can choose different behaviors
- The purpose of `for` is to perform a computation for every element in a collection

(Demo)

Random Selection

Random Selection

`np.random.choice`

- Selects at random
- with replacement
- from an array
- a specified number of times

`np.random.choice(some_array, sample_size)`

`t.sample(sample_size)`

- Sample rows from a table

Probability

Probability

- Lowest value: 0
  - Chance of event that is impossible
- Highest value: 1 (or 100%)
  - Chance of event that is certain
  - In general: 0 <= P(A) <= 1
- If an event has chance 70%, then the chance that it doesn't happen is
  - 100% - 70% = 30%
  - 1 - 0.7 = 0.3
  - In general: P(not A) = 1 - P(A)

Equally Likely Outcomes

Assuming all outcomes are equally likely, the chance of an event A is:

$$P(A) = \frac{\text{number of outcomes that make A happen}}{\text{total number of outcomes}}$$

(Demo)

Multiplication Rule

Chance that two events *A* and *B* both happen

= P(*A* happens)

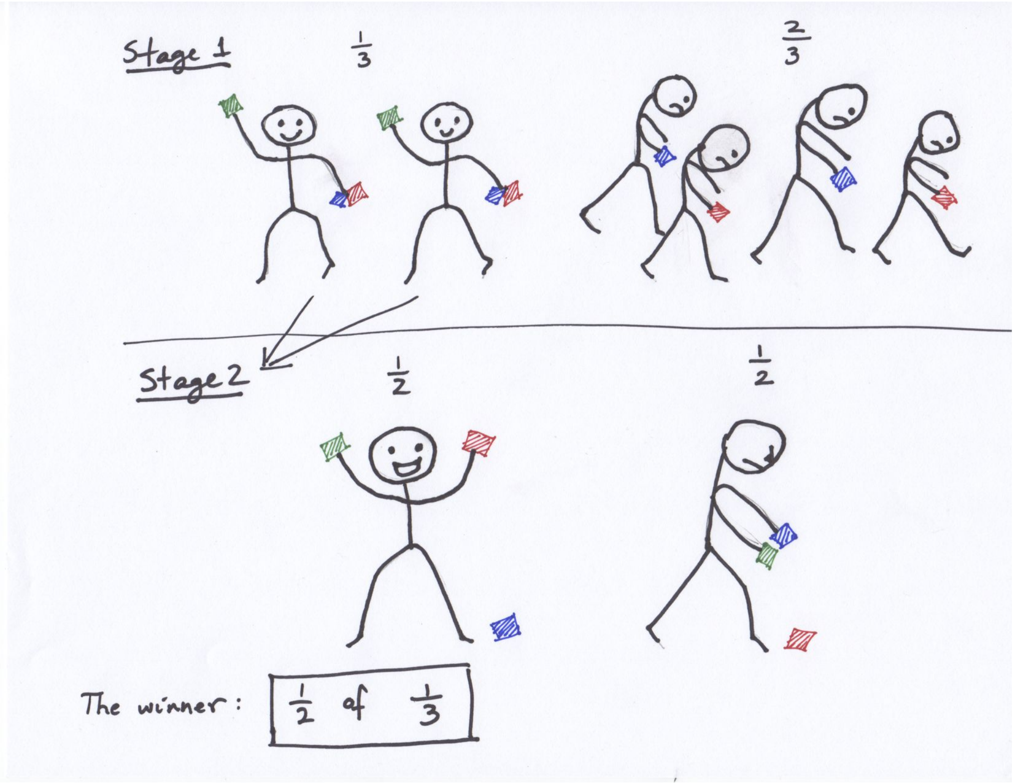
x P(*B* happens **given that** *A* has happened)

= P(A happens) x P(B happens) if A and B **independent**

- The answer is *less than or equal to* each of the two chances being multiplied

(Demo)

Fraction of a Fraction



Addition Rule

If event *A* can happen in *exactly one* of two ways, then

$$P(A) = P(\text{first way}) + P(\text{second way})$$

- The answer is *greater than or equal to* the chance of each individual way

Example: At Least One Head

- In 3 tosses:
  - Any outcome *except* TTT
  - P(TTT) = (½) x (½) x (½) = $(½)^3$ = ⅛
  - P(at least one head) = 1 - P(TTT) = ⅞ = 87.5%

- In 10 tosses:
  - P(TTTTTTTTTT) = $(½)^{10}$
  - P(at least one head) = 1 - $(½)^{10}$ = 99.90%

(Demo)

Sampling

Sampling

Observe some *individuals* from a *population*

a. Examine 10 rolls of a d6 (six-sided die)
b. Coat color of the first 20 people who walk through door
c. Survey 1000 students living in campus dorms, where every student on campus is equally likely to be chosen, and ask them about their perspective on gun control

Sampling

- Deterministic sample:
  - Sampling scheme doesn’t involve chance

- Probability (random) sample:
  - Before the sample is drawn, you have to know the selection probability of every group of people in the population
  - Not all individuals have to have equal chance of being selected

(Demo)

Sample of Convenience

- Example: sample consists of whoever walks by
- Just because you think you're sampling "at random", doesn't mean you are. If you can't figure out ahead of time
    - what's the population
    - what's the chance of selection, for each group in the population

then you don't have a random sample

Does sample look like population?

(Demo)

Large Random Samples

If the sample size is large,

then the empirical distribution of a simple random sample

resembles the population distribution,

with high probability.

Distribution

- A distribution is a description of the likelihood of *events or outcomes*

- Empirical distribution:
  - Experimental: made from observations
  - Proportion of each event in sample

vs.

- Probability distribution:
  - Theoretical: made from mathematics
  - Probability of each event

Law of Large Numbers

If an experiment is repeated many times,
independently and under the same conditions,
then the proportion of times that an event occurs
gets closer to the theoretical probability of the event

Sometimes called *Law of Averages*