MAT 105 - Midterm Review
Tuesday, February 16, 2016

MIDTERM REVIEW

1. **Measurement levels** for data:
   (a) *Nominal level* (in name only, qualitative in nature, describes the data).
   (b) *Ordinal level* (includes relative comparisons, there is an order, but differences do not have meaning).
   (c) *Interval level* (still ordered, but difference between data is meaningful and ratios are not meaningful).
   (d) *Ratio level* (interval level + meaningful ratios, there is a zero used as a starting point).

2. **Organizing Data:** charts, bars, graphs
   (a) A **bar graph** displays the data, using bars with uniform width and uniform spacing between bars.
   (b) A **Pareto chart** is a bar graph in which heights represent frequency (or percentage), with bars arranged according to height, from tallest to smallest.
   (c) A **pie chart** is a circular pie divided into parts according to given percentages. Note that the angles we obtain should add up to 360°.
   (d) A **histogram** is a representation of frequencies. It has the following components:

      - a number of classes, also called bins, typically between 5 and 15 of them, represented by bars
      - bar width = class width = $\frac{\text{largest element} - \text{smallest element}}{\text{number of classes}}$, rounded up to the nearest integer.
      - usually, the bars touch each other, there is no spacing between them

   (e) A **time plot** is a graph in which the x-axis describes the time and the y-axis the quantity measured.
   (f) A **scatter plot** is similar to a time-plot, but the data points are not connected by a curve
   (g) A **stem and leaf display** does not lose information like the histogram. It has the following components:

      - break the digits of each data value into two parts: a stem and a leaf.
      - you are free to choose the number of digits to be included in the stem!
      - list each stem ONCE on the left and all its leaves in the same row, to the right of the stem.

3. **Central tendency and variation:** for data points that each occur with equal probability!
   (a) The **mode** is the value or property that occurs most often (frequently) in the data.
   (b) The **median** is the middle value in the ordered list of data points (from smallest to largest).
   (c) The **mean** is the arithmetic mean of all the values in the data set. That is, if the data points are $n_1, n_2, \cdots, n_k$, then

$$\mu = \text{mean} = \frac{n_1 + n_2 + \cdots + n_k}{k}.$$

(d) The **trimmed mean** is the mean of the data set obtained by removing the smallest 5% and the largest 5% of the data. If 5% of the number of data points is not an integer, we round it to the *nearest* integer.

(e) The **range** of a data set is the difference between the largest data point and the smallest data point.

(f) The **variance** of a population is

$$\sigma^2 = \frac{\sum_{k=1}^{n}(x_k - \mu)^2}{n} = \frac{(x_1-\mu)^2 + \cdots + (x_n-\mu)^2}{n}$$

where μ is the mean of the population, n is the number of data points, with data points $x_1, x_2, \ldots, x_n$.

(g) The **standard deviation** is σ , the square root of variance

4. **Basics of counting**

Sum Rule: If there are k types and $n_1, n_2, \ldots, n_k$ objects of type $1, 2, \ldots, k$ respectively, there are

$$n_1 + n_2 + \cdots + n_k$$

ways to pick one object from the set.

Multiplication Rule: If there are k types and $n_1, n_2, \ldots, n_k$ objects of type $1, 2, \ldots, k$ respectively, there are

$$n_1 \times n_2 \times \cdots \times n_k$$

ways to pick one object of each type (note that you get a set of k objects, each of different type).

Factorial: $n! = n(n-1)(n-2)\cdots 2 \times 1$ represents the number of ways to order n objects.

Permutations: Permutations of n objects, taken k at a time count the number of ways to pick an *ordered* set of k objects out of n. It is defined as

$$P(n,k) = n(n-1)(n-2)\cdots(n-k+1) = \frac{n!}{(n-k)!}.$$

Combinations: Combinations of n objects, taken k at a time count the number of ways to pick an *unordered* set of k objects out of n. It is defined as

$$C(n,k) = \binom{n}{k} = \frac{n!}{k!(n-k)!}.$$

Multinomial: $\binom{n}{n_1, n_2, \ldots, n_k}$ counts the number of ways in which n objects can be placed in k bins of sizes $n_1, n_2, \cdots, n_k$ with $n_1 + n_2 + \cdots + n_k = n$ and is defined as

$$\binom{n}{n_1, n_2, \ldots, n_k} = \frac{n!}{n_1! \times n_2! \times \cdots \times n_k!}.$$

Remarks:

- to distinguish between when to use the sum rule and when to use the multiplication rule, think that for sum you pick A OR B, but for multiplication you pick A AND B.
- we use permutations when the desired set is ordered, and when the objects are distinct; we use combinations when the desired set is unordered, and when the objects are not distinct.

5. **Basics of probability**

   An **experiment** is an activity or a procedure that leads to *distinct* and well-defined possibilities called **outcomes**. The set of all outcomes forms the **sample space**, which we denote by S. If S is finite, let $|S|$ denote the number of outcomes in the sample space.

   An **event** is a statement about the outcome of the experiment.

   If the sample space is finite and each outcome is *equally likely*, we compute the probability of an event E by

   $$P(E) = \frac{\text{\# outcomes in } E}{\text{\# outcomes in } S} = \frac{|E|}{|S|}.$$

   A **probability** P is a function from the set of outcomes into the closed interval $[0, 1]$, $P : S \to [0, 1]$ satisfying the axioms:

   (i) $P(S) = 1$ (since the chance *something* happens is 1)

   (ii) $P(\emptyset) = 0$ (since the chance *nothing* happens is 0)

   (iii) $0 \leq P(E) \leq 1$ for any event E.

   (iv) If E^c denotes the complement of E, then $P(E^c) = 1 - P(E)$.

   (v) If E and F are *disjoint* (meaning $E \cap F = \emptyset$), then $P(E \cup F) = P(E) + P(F)$.

   Notes:

   - $\emptyset$ refers to the empty set (a set with no elements)
   - $E \cap F$ denotes the *intersection* of E and F (the overlap of E and F)
   - $E \cup F$ denotes the *union* of E and F (all outcomes that are in E, or F, or both)
   - The complement E^c contains all outcomes that are NOT in E.

   Property: $P(E \cup F) = P(E) + P(F) - P(E \cap F)$.

6. **Random variables:** a random variable $X : S \to \mathbb{R}$ is a function that takes on values from the set of outcomes and maps into the reals. X quantifies outcomes of random events.

   Example: You flip a coin. If the coin lands on Heads, you win \$10 and if the coin lands on Tails, you lose \$1. Let $X =$ the net amount of money won in this game. Then we define

   $$X(H) = 10 \qquad\qquad X(T) = -1$$

   We write $X \in \{-1, 10\}$ to mean that X can take on the values 10 or -1. We want to find probabilities for X taking on each value:

   $$P(X = 10) = P(\text{Heads}) = 1/2, \qquad P(X = -1) = P(\text{Tails}) = 1/2.$$

   This is the **probability distribution of** X, that is, a listing of probabilities for all possible values of X, and they should add up to 1.

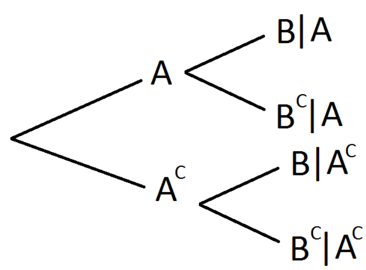
   Remark: We interpret the notation $P(X = 10)$ as "the probability that X takes on the value 10". In this example, $P(X = 10)$ means the probability that we win \$10.

7. **Expectation, variance, standard deviation:**

   The **expectation** of a discrete random variable X, also known as the **mean** or **expected value**, is given by
$$\mu = E[X] = \sum_{\text{all } a} a * P(X = a).$$
   The **variance of** X, denoted by $Var(X)$ or σ^2, is defined as
$$\sigma^2 = Var(X) = E[(X - E[X])^2] = \sum_{\text{all } a} (a - \mu)^2 * P(X = a).$$
   The **standard deviation of** X is then defined as the square root of variance $\sigma = \sqrt{Var(X)}$.

8. **Conditional Probability and Independence:**

   If A is an event so that $P(A) > 0$, we define the probability of B given A by
$$P(B|A) = \frac{P(A \cap B)}{P(A)}.$$
   Multiplying both sides by $P(A)$ gives another useful equation
$$P(A \cap B) = P(A)P(B|A).$$
   Solving conditional probability problems could be easily visualized by building a tree.



   This leads to **Bayes Formula**, which is at the base at Bayesian statistics:
$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)P(B|A)}{P(A)P(B|A) + P(A^c)P(B|A^c)}.$$
   Why is this useful? Suppose you know how the occurrence of A affects the probability of B, but we would like to find out how the occurrence of B affects the probability of A. Bayes Formula allows us to do exactly that.

   We say A and B are **independent** if
$$P(A \cap B) = P(A)P(B).$$
   Intuitively, they are independent if the occurrence of A does not influence the occurrence of B and vice-versa. Note that if A and B are independent, $P(B|A) = P(B)$, in other words, in occurrence of A has no effect on the probability of B occurring.