# GMBA 7098: Statistics and Data Analysis (Fall 2014)

## Introduction to Probability (1)

Ling-Chieh Kung

Department of Information Management
National Taiwan University

October 6, 2014

# An example of statistical inference

- Quality control: For all LED lamps of brand IM, we are interested in $\mu$, the average number of hours of luminance.
- Let's select a random sample of 40 lamps. A test shows that the sample average is $\bar{x} = 28000$ hours.
  - If I estimate that $\mu = 28000$, how likely I will be right?
  - If I estimate that $\mu \in [27000, 29000]$, how likely I will be right?
  - How about $\mu \in [26000, 30000]$?
- To assess these probabilities, we need to study Probability.

# Road map

- **Basic concepts**.
- Independent events.
- Random variables.
- Descriptive measurements.

# Experiments and events

- An **experiment** is a process that produces **(random) outcomes**.
  - Tossing a coin. Outcomes: head or tail.
  - Testing a new drug on a patient: Outcomes: Effective, not effective, getting worse.
  - Interviewing 20 consumers regarding how many will buy a new product. Outcomes: 10, 15, 0, etc.
  - Sampling every 200th bottle of ketchup for its weight. Outcome?
- An **event** is an outcome of an experiments.
- Each event has its **probability** to occur.
  - Tossing a fair coin: $\frac{1}{2}$ for head and $\frac{1}{2}$ for tail.
  - Rolling a fair dice: $\frac{1}{6}$ for each possible outcome.
- Let $A$ be an event of an experiment, we write $\Pr(A)$ to denote the probability for $A$ to occur.
  - Let $A$ be getting a head when tossing a fair coin, then $\Pr(A) = \frac{1}{2}$.

# Elementary events

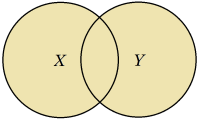
- An **elementary event** is an event that **cannot be decomposed** into smaller events.
- Consider the experiment of rolling a dice.
  - Getting 3 is an elementary event.
  - How about getting a number larger than 3?
  - The event of getting larger than 3 can be decomposed into three elementary events: getting 4, 5, and 6.
  - How about getting an even number?
- For asking Jane, Mary, Melissa, and Lucy about a new product:
  - Is "one is willing to buy" an elementary event?
  - How about "Mary is willing to buy but all the other three are not?"

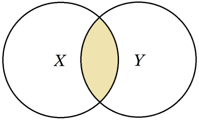
# Sample spaces

- The **sample space** of an experiment is the collection of all elementary events.
  - A sample space contains "all basic things that may happen."
  - Nothing outside the sample space can occur.
- What is the sample space of:
  - Rolling a dice?
  - Rolling two dices?
  - Asking 20 consumers?
  - Testing a new drug?
- If $S$ is a sample space, we have $\Pr(S) = 1$.
- A sample space is a **set**. Elementary elements are **elements** of the set. Events are **subsets** of the set.
  - If $x$ is an elementary event of an event $X$, we write $x \in X$.
  - E.g., "getting 2" $\in$ "getting an even number."

# Unions and intersections

- Let $A$ and $B$ be two events and $S$ be the sample space.
- The **union** of $A$ and $B$, denoted by $A \cup B$, contains elementary events in $A$ **or** $B$.
  - $A \cup B = \{x | x \in A \text{ or } x \in B\}$.
  - E.g., $\{2, 3, 5\} \cup \{1, 5, 6\} = \{1, 2, 3, 5, 6\}$.
- The **intersection** of $A$ and $B$, denoted by $A \cap B$, contains elementary events that are in $A$ **and** $B$.
  - $A \cap B = \{x | x \in A \text{ and } x \in B\}$.
  - E.g., $\{2, 3, 5\} \cap \{1, 5, 6\} = \{5\}$.



# Unions and intersections

- The union of two (or more) events is also an event.
  - Consider rolling a fair dice.
  - Let event $A$ be getting an even number. We have $\Pr(A) = \frac{1}{2}$.
  - Let event $B$ be getting larger than three. We have $\Pr(B) = \frac{1}{2}$.
  - The **union probability** of $A$ and $B$ is

$$\Pr(A \cup B) = \Pr(\text{getting 2, 4, 5, or 6}) = \frac{2}{3}.$$

- The intersection of two (or more) events is also an event.
  - Consider rolling a fair dice.
  - The **joint probability** of $A$ and $B$ is

$$\Pr(A \cap B) = \Pr(\text{getting 4 or 6}) = \frac{1}{3}.$$

- In fact, $A$ and $B$ are both unions of multiple elementary events.

# Two special cases

- Events are **mutually exclusive** if there is no intersection.
  - $A \cap B = \emptyset$ (empty).
  - Events are mutually exclusive if all their elementary events are different.
  - E.g., for rolling a dice, getting an even number and getting 5 are mutually exclusive.
- Events are **collectively exhaustive** if they together cover the whole sample space.
  - $S = A \cup B$.
  - Events are collectively exhaustive if one of them must occur.
  - E.g., for rolling a dice, getting an even number and getting smaller than six are collectively exhaustive.
  - Two collectively exhaustive sets are **not necessarily** mutually exclusive!

# Complements

- The **complement** of $X$, denoted by $X'$, contains all elements not contained in $X$.
  - $X' = \{x | x \notin X\}$, where $x \notin X$ means $x$ is not an element of $X$.
  - Graphically:
  - E.g., for rolling a dice, getting less than three and getting greater than two are complements.
  - E.g., for rolling a dice, getting less than three and getting greater than three are not complements.
- For any set $X$, $X$ and its complement $X'$ are mutually exclusive and collectively exhaustive, i.e., $X \cap X' = \emptyset$ and $X \cup X' = S$.
- Intuitively, $\Pr(X') = 1 - \Pr(X)$.

# Road map

- Basic concepts.
- **Independent events**.
- Random variables.
- Descriptive measurements.

# Independent events

- Two events are **independent** if **whether one occurs** does not affect **the probability** for the other one to occur.
- Two events are **dependent** if they are not independent.
- A set of events are independent if **all** pairs of events are independent.
- Are the following pairs of events independent?
  - Rolling two today and rolling three tomorrow with a fair dice.
  - A customer is a man and he likes watching baseball.
  - One's phone number contains "7" and she was born on July.
  - A laptop is defective and it has a 14-inch screen.

# Mathematical property

- For independent events, calculating the joint probability is easy:

**Proposition 1**

*For any two independent events $A$ and $B$, we have*

$$\Pr(A \cap B) = \Pr(A)\Pr(B).$$

- E.g., suppose we toss an unfair coin whose probability of head is $\frac{2}{3}$.
  - Let $H$ be getting a head and $T$ be getting a tail in one toss: $\Pr(H) = \frac{2}{3}$ and $\Pr(T) = \frac{1}{3}$.
  - Let $HH$ be getting two heads, $TT$ be getting two tails, $HT$ be getting a head then a tail, and $TH$ be getting a tail then a head in two tosses:

$$\Pr(HH) = \Pr(H)\Pr(H) = \frac{4}{9}, \Pr(HT) = \Pr(H)\Pr(T) = \frac{2}{9}, \text{etc.}$$

# Joint probability tables

- Two experiments may be presented by a **joint probability table**.
  - Events of experiment 1 are listed in the first **column**.
  - Events of experiment 2 are listed in the first **row**.
  - A column and a row at the margin for **totals**.
- For the previous example of an unfair dice:

| 1st | 2nd | | Total |
|---|---|---|---|
| | $H$ | $T$ | |
| $H$ | ? | ? | $\frac{2}{3}$ |
| $T$ | ? | ? | $\frac{1}{3}$ |
| Total | $\frac{2}{3}$ | $\frac{1}{3}$ | 1 |

  - The last column records the probabilities of $H$ and $T$ for the first toss.
  - The last row records the probabilities of $H$ and $T$ for the second toss.
- How to find the joint probabilities?

# Calculating joint probabilities

- To find the **joint probabilities** of two independent events $A$ and $B$, simply apply $\Pr(A \cap B) = \Pr(A)\Pr(B)$.
  - For the previous example of an unfair dice:

| 1st | 2nd $H$ | 2nd $T$ | Total |
|---|---|---|---|
| $H$ | $\frac{4}{9}$ | $\frac{2}{9}$ | $\frac{2}{3}$ |
| $T$ | $\frac{2}{9}$ | $\frac{1}{9}$ | $\frac{1}{3}$ |
| Total | $\frac{2}{3}$ | $\frac{1}{3}$ | 1 |

  - Each entry records a joint probability.
- Two joint events corresponding to two entries are mutually exclusive.
  - The **union probability** can be found by summing up joint probabilities.
  - E.g., the probability of "getting exactly one head" is

$$\Pr(HT \text{ or } TH) = \frac{2}{9} + \frac{2}{9} = \frac{4}{9}.$$

# Joint probability tables with dependent events

- Events are not always independent.

| | Supporting KMT | Supporting DPP | Neither |
|---|---|---|---|
| Will vote for Ko | 17% | 85% | 37% |
| Will vote for Lien | 71% | 4% | 20% |

| | Women | Men |
|---|---|---|
| Will vote for Ko | 36% | 39% |
| Will vote for Lien | 54% | 30% |

(http://www.chinatimes.com/newspapers/20140929000800-260302)

# Road map

- Basic concepts.
- Independent events.
- Random variables.
- **Descriptive measurements**.

# Random variables

- A **random variable** (RV) is a variable whose outcomes are random.
- Examples:
  - The outcome of tossing a coin.
  - The outcome of rolling a dice.
  - The number of people preferring Pepsi to Coke in a group of 25 people.
  - The number of consumers entering a store at 7-8pm.
  - The temperature of this classroom at tomorrow noon.
  - The average studying hours of a group of 10 students.

# Discrete and continuous random variables

- A random variable can be discrete or continuous.
- For a discrete RV, its value is **counted**.
  - The outcome of tossing a coin.
  - The outcome of rolling a dice.
  - The number of people preferring Pepsi to Coke in a group of 25 people.
  - The number of consumers entering a store at 7-8pm.
- For a continuous RV, its value is **measured**.
  - The temperature of this classroom at tomorrow noon.
  - The average studying hours of a group of 10 students.
- A discrete RV has **gaps** among its possible values; a continuous RV's possible values typically form an **interval**.

# Discrete and continuous distributions

- How to describe a random variable?
  - Writing down all possible values (the **sample space**) is not enough.
  - For each possible value, we must also describe **how likely** it will occur.
- The likelihoods for all outcomes of a random variable to be realized are summarized by **probability distributions**, or simply distributions.
- As variables can be either discrete or continuous, distributions may also be either discrete or continuous.
  - Today we study discrete distributions.
  - In the next week we study continuous distributions.

# Describing a discrete distribution

- For a discrete random variable, we may **list** all possible outcomes and their probabilities.
  - Let $X$ be the result of tossing a fair coin:

| $x$ | Head | Tail |
|---|---|---|
| $\Pr(X = x)$ | $\frac{1}{2}$ | $\frac{1}{2}$ |

  - Let $X$ be the result of rolling a fair dice:

| $x$ | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| $\Pr(X = x)$ | $\frac{1}{6}$ | $\frac{1}{6}$ | $\frac{1}{6}$ | $\frac{1}{6}$ | $\frac{1}{6}$ | $\frac{1}{6}$ |

- The function $\Pr(X = x)$, sometimes abbreviated as $\Pr(x)$, for all $x \in S$, is called the **probability mass function** (pmf) or probability function of $X$.
- For any random variable $X$, we have $\sum_{x \in S} \Pr(X = x) = 1$.

# Describing a discrete distribution: an example

- Let $X_1$ be the result of tossing a fair coin for the first time.
- Let $X_2$ be the result of tossing a fair coin for the second time.
- Let $Y$ be the **number of heads** obtained by tossing a fair coin twice.
- What is the distribution of $Y$?
  - Possible values: 0, 1, and 2.
  - Probabilities: What are $\Pr(Y = 0)$, $\Pr(Y = 1)$, and $\Pr(Y = 2)$?
- According to the joint probability table:

| | $X_2 =$ Head | $X_2 =$ Tail |
|---|---|---|
| $X_1 =$ Head | $\frac{1}{4}$ | $\frac{1}{4}$ |
| $X_1 =$ Tail | $\frac{1}{4}$ | $\frac{1}{4}$ |

| $y$ | 0 | 1 | 2 |
|---|---|---|---|
| $\Pr(Y = y)$ | $\frac{1}{4}$ | $\frac{1}{2}$ | $\frac{1}{4}$ |

- How would you find the distribution of $Z$, the number of heads obtained by tossing a fair coin for three times?

# Road map

- Basic concepts.
- Independent events.
- Random variables.
- **Descriptive measurements**.

# Descriptive measurements

- Consider a discrete random variable $X$ with a sample space $S$, realizations $\{x_i\}_{i \in S}$, and a pmf $\Pr(\cdot)$.
- The **expected value** (or mean) of $X$ is

$$\mu \equiv \mathbb{E}[X] = \sum_{i \in S} x_i \Pr(x_i).$$

- The **variance** of $X$ is

$$\sigma^2 \equiv \text{Var}(X) \equiv \mathbb{E}\big[(X - \mu)^2\big] = \sum_{i \in S} (x_i - \mu)^2 \Pr(x_i).$$

- The **standard deviation** of $X$ is $\sigma \equiv \sqrt{\sigma^2}$.

# Descriptive measurements: example 1

- Let $X$ be the outcome of rolling a dice, then the pmf is $\Pr(x) = \frac{1}{6}$ for all $x = 1, 2, ..., 6$.
  - The expected value of $X$ is

$$\mathbb{E}[X] \equiv \sum_{i=1}^{6} x_i \Pr(x_i) = \frac{1}{6}(1 + 2 + \cdots + 6) = 3.5.$$

  - The variance of $X$ is

$$\begin{aligned}\mathrm{Var}(X) &\equiv \sum_{i \in S} (x_i - \mu)^2 \Pr(x_i) \\ &= \frac{1}{6}\left[(-2.5)^2 + (-1.5)^2 + \cdots + 2.5^2\right] \approx 2.92.\end{aligned}$$

  - The standard deviation of $X$ is $\sqrt{2.92} \approx 1.71$.

# Descriptive measurements: example 2

- Let $X$ be the outcome of rolling an unfair dice:

| $x_i$ | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| $\Pr(x_i)$ | 0.2 | 0.2 | 0.2 | 0.15 | 0.15 | 0.1 |

  - The expected value of $X$ is

$$\begin{aligned}\mathbb{E}[X] &\equiv \sum_{i=1}^{6} x_i \Pr(x_i)\\ &= 1 \times 0.2 + 2 \times 0.2 + 3 \times 0.2 + 4 \times 0.15 + 5 \times 0.15 + 6 \times 0.1\\ &= 3.15.\end{aligned}$$

  - Note that $3.15 < 3.5$, the expected value of rolling a fair dice. Why?

# Descriptive measurements: example 2

- Let $X$ be the outcome of rolling an unfair dice:

| $x_i$ | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| $\Pr(x_i)$ | 0.2 | 0.2 | 0.2 | 0.15 | 0.15 | 0.1 |

  - The expected value of $X$ is $\mu = 3.15$.
  - The variance of $X$ is

$$\begin{aligned}\text{Var}(X) &\equiv \sum_{i \in S} (x_i - \mu)^2 \Pr(x_i) \\ &= (-2.15)^2 \times 0.2 + (-1.15)^2 \times 0.2 + (-0.15)^2 \times 0.2 \\ &\quad + 0.85^2 \times 0.15 + 1.85^2 \times 0.15 + 2.85^2 \times 0.1 \\ &\approx 2.6275.\end{aligned}$$

  - Note that $2.6275 < 2.92$, the variance of rolling a fair dice. Why?
  - The standard deviation of $X$ is $\sqrt{2.6275} \approx 1.62$.

# Descriptive measurements: using R

- Let $X$ be the outcome of rolling an unfair dice:

| $x_i$ | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| $\Pr(x_i)$ | 0.2 | 0.2 | 0.2 | 0.15 | 0.15 | 0.1 |

```
> x <- 1:6
> p <- c(0.2, 0.2, 0.2, 0.15, 0.15, 0.1)
> mu.x <- sum(x * p) # expected value
> var.x <- sum((x - mu.x) ^ 2 * p) # variance
> sd.x <- sqrt(var.x) # standard deviation
```