Day 5: Probability

Ikuma Ogura

Ph.D. student, Department of Government, Georgetown University

August 23, 2019

Today

- Today
  - Probability
  - Random variable
  - Probability distribution
  - (if time permits) Probability distributions in R

Why Probability?

- Foundation of statistical inference
- Formal modeling/game theory: describe uncertainty

Probability and Statistics

$$\boxed{\text{Data generating process}} \underset{\text{Statistical inference}}{\overset{\text{Probability}}{\rightleftarrows}} \boxed{\text{Event}}$$

- Probability: explains how likely each event occurs based on the known data generating process (DGP)

- Statistical inference: infer the DGP based on the data (= collection of events)

Probability

- **Sample space** (S): A set/collection of all possible outcomes from some process.
    - Outcomes of sample space can be countable (discrete) or uncountable (continuous).

- **Event**: Any set of possible outcomes from the process. Any subset of the full set of possibilities ($=$ sample space), including the full set itself.

- **Partition**: a sequence of disjoint events $A_1, A_2, \cdots, A_n$ where

$$A_1 \cup A_2 \cup \cdots \cup A_n = S$$

Probability (cont.)

- **Probability** is a function that maps events to a real number and follows the **axioms of probability** below.
    1. For any event A, $\Pr(A) \geq 0$.
    2. $\Pr(S) = 1$.
    3. For any sequence of disjoint events $A_1, A_2, \cdots$

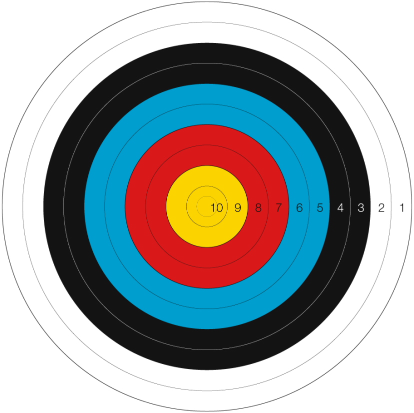
$$\Pr(A_1 \cup A_2 \cup \cdots) = \Pr(\bigcup_{i=1} A_i) = \sum_{i=1} \Pr(A_i)$$

Probability (cont.)

- Example 1: Let A_i denote the event of getting hand i in the game of poker. Then, probability of hand i can be defined as

$$\Pr(A_i) = \frac{\text{Frequency of getting hand } i}{\text{Number of all possible outcomes}}$$

Probability (cont.)



- Example 2: Ikuma is novice in archery. Assuming that he can always hit the target (!), what is the probability of scoring point i?

- Let A_i denote the event of scoring i,

$$\Pr(A_i) = \frac{\text{Area of region } i}{\text{Total area}}$$

Probability (cont.)

- Properties of probability operation
    1. $\Pr(\emptyset) = 0.$
    2. $0 \leq \Pr(A) \leq 1$
    3. $\Pr(A^c) = 1 - \Pr(A)$
    4. If $A \subseteq B$, $\Pr(A) \leq \Pr(B)$

Counting

- Counting *with replacement* or *without replacement*
- *Ordering* is important or not
- Below let's think of a case in which we select k out of n.

Counting (cont.)

- **Ordered, with replament**: in this case, the number of different outcomes is

$$n^k$$

- **Ordered, without replament**: in this case the number of different outcomes is

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

Counting (cont.)

- **Unordered, without replament**: as there is $k!$ ways to order k objects, the number of different outcomes in this case is

$$\frac{n \times (n-1) \times (n-2) \times \cdots \times (n-k+1)}{k!} = \frac{n!}{k!(n-k)!} = \binom{n}{k}$$

  - $\binom{n}{k}$ is called the binomial coefficient

Counting (cont.)

- Example: What is the probability of getting full house?
  - Denominator: the number different ways to select 5 cards out of 52 is $\binom{52}{5}$.
  - Numerator: we need to choose 3 cards from one face value, and 2 cards from another face value. Therefore, the number of distinct hands is

$$\binom{13}{1}\binom{4}{3}\binom{12}{1}\binom{4}{2}$$

  - Therefore, the probability of getting full house is

$$\frac{13 \cdot 4 \cdot 12 \cdot 6}{\frac{52 \cdot 51 \cdot 50 \cdot 49 \cdot 48}{5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}} = \frac{6}{4165} \approx 0.00144$$

Counting: Exercises

1. (de Mere's problem) Which is higher:
    1.1 the probability of getting at least one "6" in 4 rolls of a 6-sided die
    1.2 the probability of at least one double-six in 24 rolls of two dice
2. (Birthday problem) Suppose 20 individuals are randomly selected. Find the probability that at least two of them have the same birthday.

Conditional Probability

- **Conditional probability** $\Pr(A|B)$ is the probability of event A given that the event B occurred, which is defined as

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

- If $\Pr(B) \neq 0$, $\Pr(A \cap B) = \Pr(A|B)\Pr(B)$
- If $\Pr(A|B) = \Pr(A)$, events A and B are said to be **independent**.
  - Equivalently, if A and B are independent, $\Pr(A \cap B) = \Pr(A)\Pr(B)$

Conditional Probability (cont.)

- Example: what is the probability of choosing the Ace of heart given that the selected card is red? Let

$$\begin{aligned} A &= \{\text{Choose Ace of Heart}\} \\ R &= \{\text{Choose Red}\}, \end{aligned}$$

the probability of interest is

$$\Pr(A|R) = \frac{\Pr(A \cap R)}{\Pr(R)} = \frac{\Pr(A)}{\Pr(R)} = \frac{1}{26}$$

Bayes Rule

- **Law of Total Probability**: Let $B_1, B_2, \cdots, B_n$ be the partition of S. For some event A in S, we can represented the set as the union of disjoint subsets

$$A = (A \cap B_1) \cup (A \cap B_2) \cup \cdots \cup (A \cap B_n).$$

Therefore, from the axiom of probability,

$$\Pr(A) = \Pr(A \cap B_1) + \Pr(A \cap B_2) + \cdots + \Pr(A \cap B_n) = \sum_{i=1}^{n} \Pr(A \cap B_i)$$

Bayes Rule (cont.)

- Example: the probability of choosing a heart is

$$
\begin{aligned}
\Pr(\text{Choose Heart}) &= \Pr(\text{Choose Heart} \cap \text{Choose A}) + \\
&= \Pr(\text{Choose Heart} \cap \text{Choose 2}) + \\
&= \cdots + \\
&= \Pr(\text{Choose Heart} \cap \text{Choose King})
\end{aligned}
$$

Bayes Rule (cont.)

- **Bayes rule**: Let $B_1, B_2, \cdots, B_n$ be the partition of S. Based on the defition of conditional probability and the law of total probability,

$$\Pr(B_k|A) = \frac{\Pr(A \cap B_k)}{\Pr(A)} = \frac{\Pr(A|B_k)\Pr(B_k)}{\sum_{i=1}^{n}\Pr(A|B_i)\Pr(B_i)}$$

  - $\Pr(B_k)$ is called the prior probability.
  - Bayes rule describes how $\Pr(B_k)$ changes with additional information.

Bayes Rule: Example

- Example: (Monty Hall problem) You are on a game show and asked to choose between three doors. Behind each door, there is either a car or a goat. After you choose a door, the host, Monty, opens another door, shows a goat, and gives you a chance to change your choice. Should you change your choice?

  - Answer: let's name the three doors A, B, and C, and represent the event that a car is behind the corresponding door. Let X_M $(X \in \{A, B, C\})$ denote the event that Monty opens the door X. For generality, let's assume you pick the door A and Monty opens the door B. Therefore, we need to compare the probability $\Pr(A|B_M)$ and $\Pr(C|B_M)$. Here, let's also assume that

$$\Pr(A) = \Pr(B) = \Pr(C) = \frac{1}{3}.$$

(*continue from the previous slide*)

Applying the Bayes rule,

$$\begin{aligned}
\Pr(A|B_M) &= \frac{\Pr(B_M|A)\Pr(A)}{\Pr(B_M|A)\Pr(A)+\Pr(B_M|B)\Pr(B)+\Pr(B_M|C)\Pr(C)} \\
&= \frac{\frac{1}{2}\cdot\frac{1}{3}}{\frac{1}{2}\cdot\frac{1}{3}+0+1\cdot\frac{1}{3}} = \frac{1}{3}
\end{aligned}$$

and

$$\begin{aligned}
\Pr(C|B_M) &= \frac{\Pr(B_M|C)\Pr(C)}{\Pr(B_M|A)\Pr(A)+\Pr(B_M|B)\Pr(B)+\Pr(B_M|C)\Pr(C)} \\
&= \frac{1\cdot\frac{1}{3}}{\frac{1}{2}\cdot\frac{1}{3}+0+1\cdot\frac{1}{3}} = \frac{2}{3}
\end{aligned}$$

Therefore, you should switch the door you choose from A to C.

Bayes Rule: Exercise

- In Orange County, 51% of the adults are men. One adult is randomely selected for a survey, and the selected survey subjest was smoking cigarettes. Based on administrative data, we know that 16% of men smoke cigarettes, whereas 12% of women do so. What is the probability that the selected subject is a man?

Random Variable

- **Random variable** (or **stochastic variable**) is a function from a sample space to real numbers.

- We use capital letters (e.g., X) to represent random variables and small letters (e.g., x) to denote their realizations (concrete values they take).

- We can also define probability for a random variable.

$$\Pr(X = x_i) = \Pr(\{s_j \in S | X(s_j) = x_i\})$$

- Discrete v. Continuous
  - A random variable X is **discrete** if it takes finite or countably infinite number of values
  - A random variable X is **continuous** if it can take any real numbers in the domain
  - continuous sample space does not nessarily lead to contious random varialbe (see example)

Random Variable (cont.)

- Example 1: we can assign scores to poker hands as follows:

Hand	Score (X)
No pair	0
One pair	1
Two pair	2
Three of a kind	3
Straight	4
Flush	5
Full house	6
Four of a kind	7
Straight flush	8

Here, outcomes in the sample space ($=$ combination of 5 cards) are mapped to real numbers (score X).

Random Variable (cont.)

- Example 2: For the example of archery we saw earlier, we can define the random variable X as

$$X(\{\text{Arrow hits the region } i\}) = i.$$

In this case, even though the sample space is continuous, the random variable X is discrete.

Random Variable (cont.)

- Example 3: Define the random variable T to denote the time until the next train arrives at the station. Here, T is a continuous random variable since it can take any real numbers.

Random Variable (cont.)

- Randomness does not mean lack of pattern/structure/order.
  - Randomness means that the outcome of some process/realization of the variable is not deterministic.
  - Pattern/structure/order of the process is characterized by probability.

Probability Distribution

- Probability distributions specify the relationship between the random variable and probabilities.
- Below I'll introduce two types of functions describing a probability distribution.

Probability Mass/Density Function

- **Probability mass function (PMF)** for a discrete random variable X describes the probability that X takes a specific value x.

$$f(x) = \Pr(X = x)$$

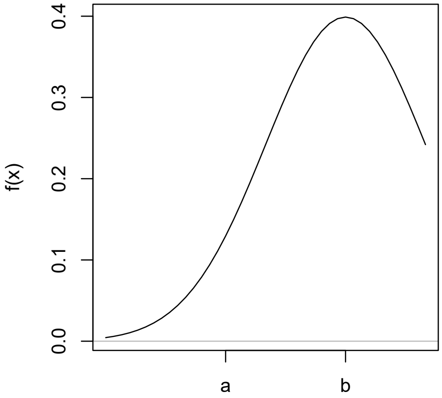
- We can also consider similar function $f(x)$ for a continuous random variable X, which is called the **probability density function (PDF)**.

Probability Mass/Density Function (cont.)

- Example: Let X be the sum of numbers we get by rolling two dice. Assuming that these dice are fair, the PMF $f(x)$ is

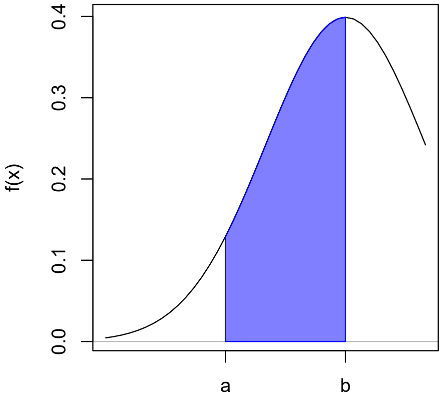
$$f(x) = \Pr(X = x) = \frac{6 - |7 - x|}{36}$$

Probability Mass/Density Function (cont.)



- For a continuous random variable, we cannot allow a point $X = x$ to have probability larger than 0!
- $f(x)$ describes the *relative likelihood* that X equals to x.
- Instead, for a continuous distribution we define probability for an interval in the domain.

Probability Mass/Density Function (cont.)



- The probability of X falling between (a, b) equal to

$$\Pr(a < x < b) = \int_a^b f(x)dx$$

- As the probability of a single point is 0,

$$\Pr(a < x < b) = \Pr(a \leq x \leq b)$$

Cumulative Density Function

- **Cumulative density function** (CDF) describes the probability that the random variable X is smaller than or euqal to x.
  - For a discrete random variable, CDF is defined as

$$F(x) = \Pr(X \leq x) = \sum_{i \leq x} f(i)$$

  - For a continuous random variable, CDF is defined as

$$F(x) = \Pr(X \leq x) = \int_{-\infty}^{x} f(t)dt$$

Cumulative Density Function (cont.)

- For a continuous random variable, from the fundamental theorem of calculus,
$$F'(x) = f(x)$$

- CDF of a discrete distribution is a step function, whereas CDF of a continuous distribution is a continuous function.

- Properties of CDF
  1. $F(x)$ is a non-decreasing function of x
  2. $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to \infty} F(x) = 1$
  3. $F(x)$ is right-continuous, meaning that, for every x_0 in the domain $\lim_{x \to x_0^+} F(x) = F(x_0)$

Cumulative Density Function: Exercise

- Draw the CDF of the random variable for a single roll of a fair die.

Joint Probability Distribution

- A probability distribution where multiple random variables are considered together is called a joint probability distribution.

- If
$$f(x,y) = f(x)f(y)$$
for all x and y in the support, we say X and Y are **independent**.

Joint Probability Distribution (cont.)

- Example: Let's define a random variable X as the number you get from a single roll of die, and Y is the face value you get when you pick a card from a deck. Their distributions are independent, as

$$\begin{aligned}
\Pr(X=1,Y=1) &= \Pr(X=1)\cdot\Pr(Y=1)=\frac{1}{6}\cdot\frac{1}{13}\\
\Pr(X=1,Y=2) &= \Pr(X=1)\cdot\Pr(Y=2)=\frac{1}{6}\cdot\frac{1}{13}\\
&\vdots\\
\Pr(X=2,Y=1) &= \Pr(X=2)\cdot\Pr(Y=1)=\frac{1}{6}\cdot\frac{1}{13}\\
&\vdots\\
\Pr(X=6,Y=13) &= \Pr(X=1)\cdot\Pr(Y=13)=\frac{1}{6}\cdot\frac{1}{13}
\end{aligned}$$

Measures Characterizing Distributions

- Expectation
- Variance
- Covariance

Expectation

- Expected value of random variable X is the average of its values weighted by their probability/density.
  - Expected value of a discrete random variable X is given by

$$\mathrm{E}(X) = \sum_x x \Pr(X = x)$$

  - Expected value of a continuous random variable X is given by

$$\mathrm{E}(X) = \int_x x f(x) dx$$

  - We can consider expected value of a function of X as

$$\begin{aligned} \mathrm{E}[g(X)] &= \sum_x g(x) \Pr(X = x) \; (X \text{ is discrete}) \\ \mathrm{E}[g(X)] &= \int_x g(x) f(x) dx \; (X \text{ is continuous}) \end{aligned}$$

Expectation (cont.)

- Example: Let X be the sum of numbers from two rolls of dice. Then

$$\mathrm{E}(X) = 2 \times \frac{1}{36} + 3 \times \frac{2}{36} + \cdots + 12 \times \frac{1}{36} = 7$$

and

$$\begin{aligned}\mathrm{E}(X^2) &= (2)^2 \times \frac{1}{36} + (3)^2 \times \frac{2}{36} + \cdots + (12)^2 \times \frac{1}{36} \\ &= \frac{1974}{36} = \frac{329}{6}\end{aligned}$$

Expectation (cont.)

- Properties of expectation operator
  1. $\mathrm{E}(c) = c$.
  2. $\mathrm{E}(aX + bY) = a\mathrm{E}(X) + b\mathrm{E}(Y)$
  3. if X and Y are independent, $\mathrm{E}(XY) = \mathrm{E}(X)\mathrm{E}(Y)$

Variance

- Variance describes the degree of spread of a distribution.
  - For a discrete random variable X, its variance is defined as

$$\operatorname{Var}(X) = \sum_x (x - \operatorname{E}(X))^2 \Pr(X = x)$$

  - For a discrete random variable X, its variance is defined as

$$\operatorname{Var}(X) = \int_x (x - \operatorname{E}(X))^2 f(x)$$

- Alternatively, $\operatorname{Var}(X)$ is defined as

$$\begin{aligned} \operatorname{Var}(X) &= \operatorname{E}[(X - \operatorname{E}(X))^2] \\ &= E(X^2) - (E(X))^2 \end{aligned} \tag{1}$$

- Square root of $\operatorname{Var}(X)$ is called the **standard deviation**.

Variance (cont.)

- Example: Let's use the example of rolling two dice again. The variance of X can be calculated as

$$\begin{aligned}\mathrm{Var}(X) &= (2-7)^2 \times \frac{1}{36} + (3-7)^2 \times \frac{2}{36} + \cdots + (12-7)^2 \times \frac{1}{36} \\ &= \frac{210}{36} = \frac{35}{6}.\end{aligned}$$

Equivalently,

$$\mathrm{Var}(X) = \frac{329}{6} - (7^2) = \frac{35}{6}.$$

Variance (cont.)

- Let's derive the equation (1) we saw earlier.

$$
\begin{aligned}
\mathrm{Var}(X) &= \mathrm{E}[(X-\mathrm{E}(X))^2] \\
&= \mathrm{E}[(X-\mathrm{E}(X))(X-\mathrm{E}(X))] \\
&= \mathrm{E}[X^2 - 2E(X)X + (E(X))^2] \\
&= \mathrm{E}(X^2) - \mathrm{E}[2\mathrm{E}(X)X] + \mathrm{E}[(E(X))^2] \\
&= \mathrm{E}(X^2) - 2\mathrm{E}(X)\mathrm{E}(X) + (E(X))^2 \\
&= E(X^2) - (E(X))^2
\end{aligned}
$$

Covariance

- Covariance describes the degree to which two random variables vary together, which is defined as

$$\begin{aligned}\mathrm{Cov}(X,Y) &= \mathrm{E}[(X-\mathrm{E}(X))(Y-\mathrm{E}(Y))] \\ &= \mathrm{E}(XY)-\mathrm{E}(X)\mathrm{E}(Y)\end{aligned}$$

  - By definition, $\mathrm{Cov}(X,X)=\mathrm{Var}(X)$.
  - If X and Y are independent, $\mathrm{Cov}(X,Y)=0$.

Properties of Variance & Covariance

- Properties of variance & covariance operators
  1. $\text{Var}(c) = 0$
  2. $\text{Var}(cX) = c^2\text{Var}(X)$
  3. $\text{Var}(aX + bY) = a^2\text{Var}(X) + b^2\text{Var}(Y) + 2ab\text{Cov}(X, Y)$
  4. $\text{Cov}(aX, bY) = ab\text{Cov}(X, Y)$
  5. $\text{Cov}(X + a, Y + b) = \text{Cov}(X, Y)$
  6. $\text{Cov}(X + Z, Y + W) =$
     $\text{Cov}(X, Y) + \text{Cov}(X, W) + \text{Cov}(Z, Y) + \text{Cov}(Z, W)$

Measures Characterizing Distributions: Exercise

- A random variable X follows a distribution whose PDF is defined as

$$f(x) = \begin{cases} \frac{1}{b-a} \ (x \in [a, b]) \\ 0 \ (\text{otherwise}) \end{cases}$$

Find the expectation and variance of this distribution.

Distributions Frequently Used in Social Science

- Here I introduce some of the probability distributions often used in social science research.
- We can determine the exact forms of their PMF/PDF by specifying the values of the **parameters**.
- Additional terms
    - We say random variable X **follows** some specific probability distribution to mean that the values of X are determined according to the PMF/PDF of the corresponding probability distribution
    - **Support** of the probability distribution is the subset of its domain where probability/density is not 0.

Binomial Distribution

- Binomial random variable X describes the number of "successes" $(x = 1)$ out of n identical trials, where the probability of success is p (so the probability of "failure" is $1 - p$).

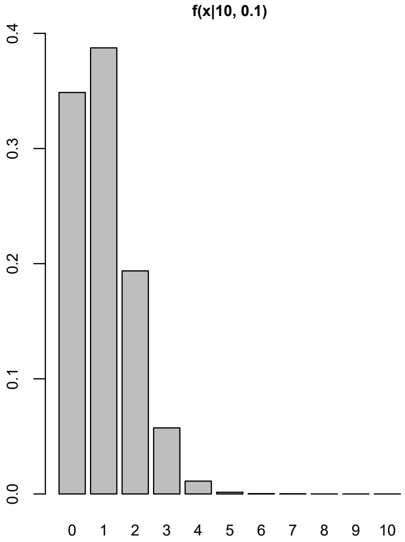
- Binomial PMF is defined as

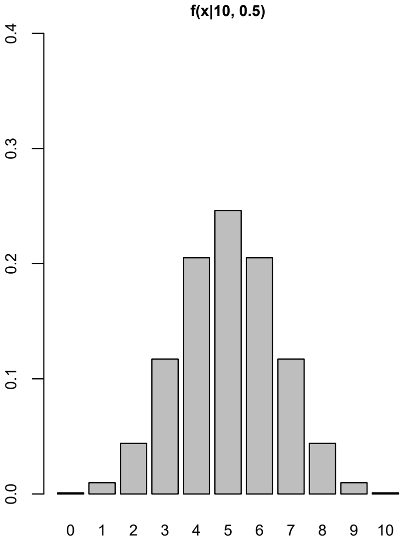
$$f(x|n,p) = \Pr(X = x|n,p) = \binom{n}{x} p^x (1-p)^{n-x}$$

- Binomial distribution with parameters n and p is often denoted as $B(n,p)$
  - $X \sim B(n,p)$ means that the random variable X follows the binomial distribution with parameters n and p.

- Expectation and variance
  - $\mathrm{E}(X) = np$
  - $\mathrm{Var}(X) = np(1-p)$

Binomial Distribution (cont.)





Binomial Distribution (cont.)

- Example: Suppose 14% of US adults smoke cigarettes. If we randomly sample 1,000 individuals from US adults, the number of smokers in the sample, X, is modeled using the binomial distribution as

$$f(x|1000, 0.14) = \Pr(X = x|1000, 0.14) = \binom{1000}{x}(0.14)^x \cdot (0.86)^{1000-x}$$

Poisson Distribution

- A random variable X follows a Poisson distribution if itsu PMF is

$$f(x|\lambda) = \Pr(X = x|\lambda) = \frac{e^{-\lambda}\lambda^x}{x!}$$

- Poisson distribution with parameter λ is often denoted as $\mathrm{Pois}(\lambda)$.
- The support of a Poisson random variable is $\mathbb{N}$, so it is often used to model counts.
- Expectation and variance
  - $\mathrm{E}(X) = \mathrm{Var}(X) = \lambda$

Uniform Distribution

- A random variable X follows a continuous uniform distribution on the interval (a, b) if its PDF is given by

$$f(x|a,b) = \begin{cases} \frac{1}{b-a} \ (x \in [a,b]) \\ 0 \ (\text{otherwise}) \end{cases}$$

- Applications
  - model lack of information
  - random number generation

Normal Distribution

- A random variable X follows a normal distribution (or Gaussian distribution) with expectation μ and variance σ^2 if its PDF is

$$f(x|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big(- \frac{(x-\mu)^2}{2\sigma^2}\Big)$$

- Normal distribution with parameters μ and σ^2 is often denoted as $N(\mu,\sigma^2)$

Normal Distribution (cont.)

- Normal distribution with $\mu = 0$ and $\sigma = 1$ is called the **standard normal distribution**, and we often denote its PDF as $\phi(x)$

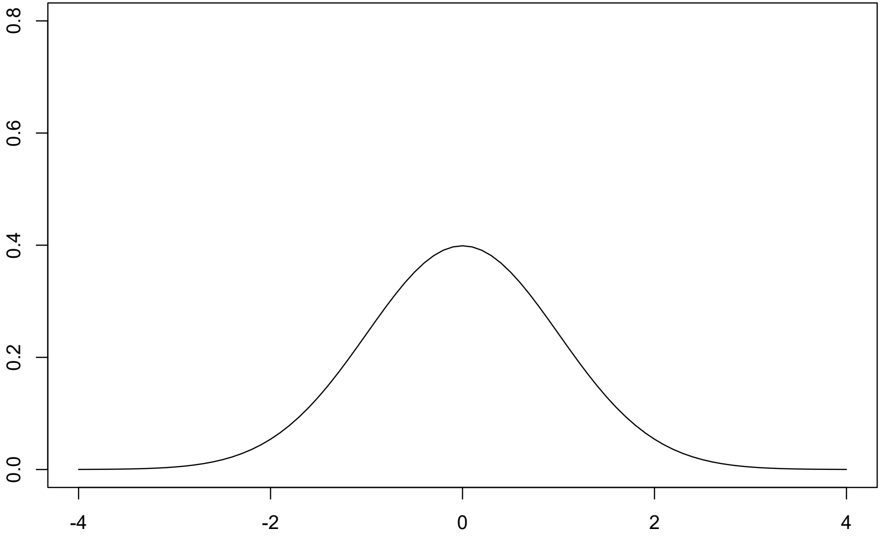
$$\phi(x) = \frac{1}{\sqrt{2\pi}} \exp\Big(-\frac{x^2}{2}\Big)$$

  and its CDF as $\Phi(x)$

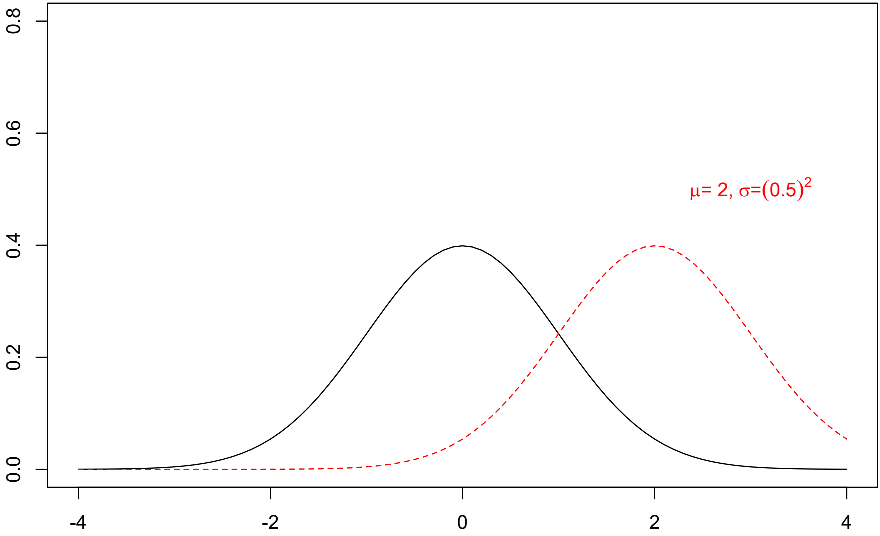
$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp\Big(-\frac{t^2}{2}\Big) dt$$

  - Using this notation, the PDF of $N(\mu, \sigma^2)$ can be written as $\frac{1}{\sigma}\phi(\frac{x-\mu}{\sigma})$

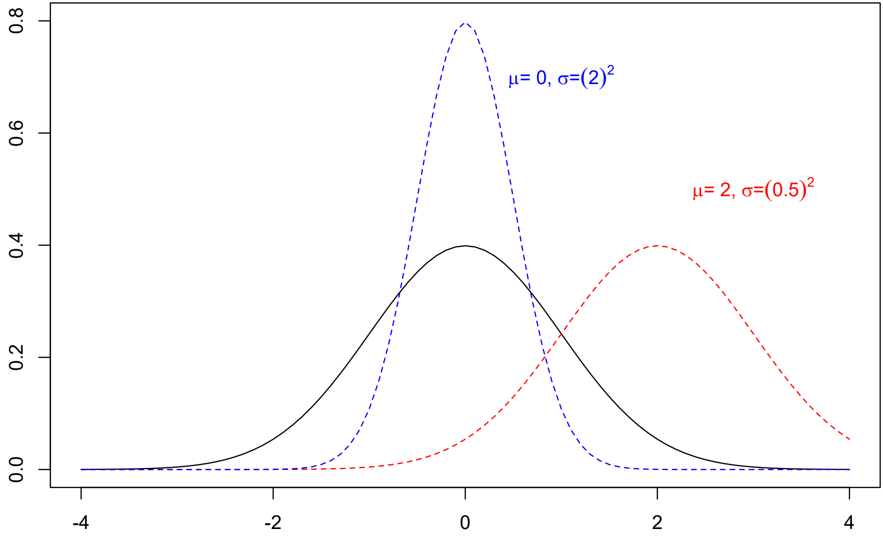
Normal Distribution (cont.)



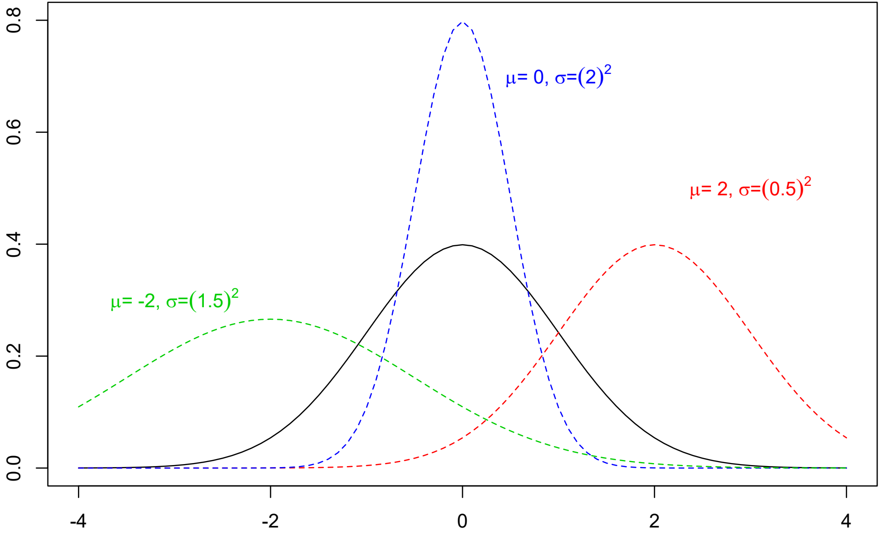
Normal Distribution (cont.)



Normal Distribution (cont.)



Normal Distribution (cont.)



Other Distributions Often Used in Social Science

- Negative binomial distribution
- Multinomial distribution
- t distribution
- F distribution
- χ^2 distribution
- Exponential distribution
- Weibull distribution
- Gamma distribution
- Beta distribution
- Dirichlet distribution
- ...

Wrapping Up...

- We learned a lot!
  - Basics
    - Exponential & log functions
    - Summation & product operators
    - ...
  - Calculus
    - Limit
    - Derivative
    - Integral
    - (Unconstrained) Optimization
  - Matrix Algebra
  - Probability theory

Now They Look Familiar! (Right?)

Specifically, the expression for the sum of squared residuals for any given estimates of $\hat{\beta}_0$ and $\hat{\beta}_1$ is

$$\sum_{i=1}^{N} \hat{\epsilon}_i^2 = \sum_{i=1}^{N} (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)^2$$

Now They Look Familiar! (Right?)

where $y_i(e_i) = e_i + \varepsilon_i$. The first-order condition for (2.2.4) is

$$(w_H - w_L)\frac{\partial \text{Prob}\{y_i(e_i) > y_j(e_j^*)\}}{\partial e_i} = g'(e_i). \qquad (2.2.5)$$

That is, worker i chooses e_i such that the marginal disutility of extra effort, $g'(e_i)$, equals the marginal gain from extra effort, which is the product of the wage gain from winning the tournament, $w_H - w_L$, and the marginal increase in the probability of winning.

By Bayes' rule,[12]

$$\begin{aligned}
\text{Prob}\{y_i(e_i) > y_j(e_j^*)\} &= \text{Prob}\{\varepsilon_i > e_j^* + \varepsilon_j - e_i\} \\
&= \int_{\varepsilon_j} \text{Prob}\{\varepsilon_i > e_j^* + \varepsilon_j - e_i \mid \varepsilon_j\} f(\varepsilon_j)\, d\varepsilon_j \\
&= \int_{\varepsilon_j} [1 - F(e_j^* - e_i + \varepsilon_j)] f(\varepsilon_j)\, d\varepsilon_j,
\end{aligned}$$

Now They Look Familiar! (Right?)

orm. The least squares estimate of $a^T\beta$ is

$$\hat{\theta} = a^T\hat{\beta} = a^T(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}. \tag{3.17}$$

Considering $\mathbf{X}$ to be fixed, this is a linear function $\mathbf{c}_0^T\mathbf{y}$ of the response vector $\mathbf{y}$. If we assume that the linear model is correct, $a^T\hat{\beta}$ is unbiased ince

$$\begin{aligned} \mathrm{E}(a^T\hat{\beta}) &= \mathrm{E}(a^T(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}) \\ &= a^T(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{X}\beta \\ &= a^T\beta. \end{aligned} \tag{3.18}$$

Topics Not Covered in This Class...

- Constrained optimization
- Multiple integral
- Matrix decomposition
- More probability theory
- ...

Last Word

- Review materials as often as possible
- Enjoy studying statistics & formal modeling!