Quantitative Methods in Political Science: Fundamentals of Probability

Thomas Gschwend | Oliver Rittmann | Viktoriia Semenova | David M. Grundmanns

Week 2 - 15 September 2021

The Course

Roadmap

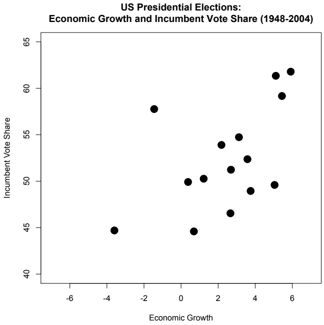
- Understand and model stochastic processes
- Understand statistical inference
- Implement it mathematically and learn how to estimate it
  - OLS
  - Maximum Likelihood
- Implement it using software
  - R
  - Basic programming skills

Overview: Week 2

- Probability Theory
  - Basics and notation
  - Definitions
- Random variables
- Probability distributions
  - Binomial distribution
  - Normal distribution
- The Central Limit Theorem

Quiz

Which of the following answers are true?



1. There is a tendency that the better the economy, the higher the incumbents' vote shares.
2. The graph makes transparent that we are facing a left-skewed distribution.
3. Scatterplots are invented to show the range of distribution.
4. There is a relationship of the economy and the vote shares of incumbent presidents.
5. Higher economic growth is causing the reelection of incumbent presidents.

Probability Theory

Probability Theory - Why should we care?

- Probability theory important tool to translate political science theories into appropriate statistical models.
- A preview - Three steps to generate a statistical model:
  (1) What is the *data-generating process* (DGP)?
  (2) Build an appropriate probability model that reflects the assumed DGP including assumptions of how Y is distributed (i.e., stochastic component)
  (3) Come-up with systematic component including a parameterization of the stuff that gets estimated (i.e., systematic component) and theory of inference to derive statistical model
- Not necessary to fit ones data to an existing but potentially inappropriate statistical model

Basic Terminology

- An experiment is a repeatable procedure for making an observation.
- An outcome is a possible result of such an experiment.
- The sample space (Ω) of an experiment is the set of all possible outcomes.
- An event is a subset of the sample space, i.e., any set of outcomes.
- The probability of an event is its long-run relative frequency.
  - If $Pr(E) = .5$, i.e., probability of event E is .5 then event E will occur approximately half of the time when the experiment is repeated infinitely.
  - If the experiment is repeated many (finite) times, then the approximation as relative frequency (proportion) is expected to improve as the number of repetitions increases.

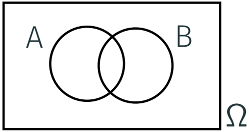
- Imagine we toss a coin two times (our experiment), resulting in heads or tails as outcomes and the following sample space: $\Omega = \{HH, HT, TH, TT\}$
- An event E could be: “Tails in second flip”.

Sets and Events

- Events can be combined to form more complicated events via a number of logical operations.

Operation	Set	Definition	Event interpretation
Union	$A \cup B$	elements either in A or B or in both	either A or B or both occur
Intersection	$A \cap B$	elements both in A and B	both A and B occur
Complement	$\bar{A}$	elements not in A	A does not occur

- If B contains A, we write $A \subseteq B$ and interpret it as: “when A occurs, so does B (but not necessarily vice versa)”.



The Concept of Probability

- Intuitively, think of probability as assigning real numbers to every element of the sample space in a way that the sum of all such numbers is 1.
  - Example: 4 people vote for candidate 'A', and 6 people vote for candidate 'B'.
  - Therefore, $Pr(A) = 0.4$ and $Pr(B) = 0.6$ i.e., probabilities express the proportion of votes cast relative to the total number of votes.
- Probabilities are real numbers $Pr(A)$ assigned to every event A of sample space Ω such that:
  (a) $Pr(A) \geq 0$: Probabilities are nonnegative.
  (b) $Pr(\Omega) = 1$: The total probability is 1.
  (c) If $A_1, \ldots, A_k$ are mutually exclusive events, then

$$Pr(A_1 \cup \cdots \cup A_k) = Pr(A_1) + \cdots + Pr(A_k)$$

Random variables

Random Variable

- A random variable is a function that assigns a number to each outcome of the sample space of an experiment.
- Imagine we toss a coin two times; this results in the following sample space: $\Omega = \{HH, HT, TH, TT\}$.
- A random variable X that counts the number of heads looks as follows:

Outcome	value x of X
HH	2
HT	1
TH	1
TT	0

- The above table shows a frequency distribution of X.

Probability Distributions

- Distributions of random variables are probability distributions if for all possible outcomes, it tells us the probabilities for these outcomes to occur.
- Probability distributions are analogous to frequency distributions, except that they are based on probability theory rather than observations in sample data.
- Two kinds of distributions exist:
    - Discrete Distributions: e.g., Bernoulli, Binomial, Poisson. (Examples?)
    - Continuous Distributions: e.g., Uniform, Normal, Logistic, t-distribution.

Distributions of Random Variables

- The probabilities $p(x)$ or $P(X = x)$ for all values of a random variable X form the *probability density function* (PDF). For example, the probability distribution for the number of heads in two coin flips is:

x	$p(x)$
0	.25
1	.50
2	.25
Total	1.00

- The probability of observing a value less or equal than x, $P(x) = P(X \leq x)$ yields the *cumulative distribution function* (CDF):

x	$P(x)$
0	.25
1	.75
2	1

Difference between PDF and CDF

- Probability density function (PDF): What is the probability that we get
  - exactly x_i (for discrete distributions)?
  - $a \leq x_i \leq b$ (for continuous distributions)?
- cumulative distribution function (CDF): What is the probability that we get some value equal to or smaller than x_i?
- The main point to understand is that areas under the density curve $p(x)$ are interpreted as probabilities, and that the height of the CDF gives the probability of observing values of X less than or equal to the value x.

Expected Value and Variance

- Expected Value: Specifies the center of the probability distribution:
  - X discrete:

$$E(X) = \sum_{all\ x} xp(x)$$

  - X continuous:

$$E(X) = \int_{-\infty}^{+\infty} xp(x)dx$$

- Variance: Specifies the spread of the probability distribution
  - X discrete:

$$Var(X) = \sum_{all\ x} (x - E(X))^2 p(x)$$

  - X continuous:

$$Var(X) = \int_{-\infty}^{+\infty} (x - E(X))^2 p(x)dx$$

Probability distributions

Binomial Distribution

- A binomial random variable K that represents the number of 'successes' in n outcomes of a binomial process.
- A binomial process is given by:
  - n independent trials.
  - Only two possible outcomes, which are arbitrarily called 'success' and 'failure'.
  - Failure and success probabilities that remain constant over trials.
- Let n be the number of trials, p the probability of success
- It has mean (expected value)

$$E(K) = np$$

  and variance

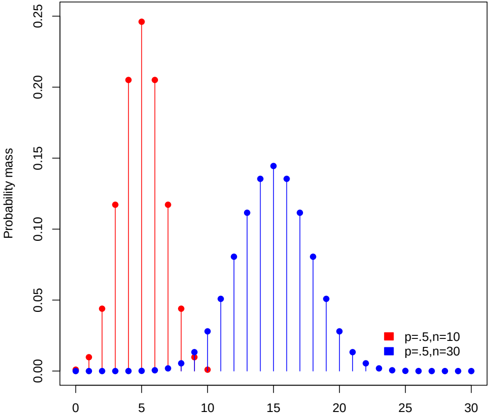
$$Var(K) = np(1-p)$$

- The expected value is interpretable as the mean score of a random variable that would be observed if we were to perform the experiment an infinite number of times.

Binomial Probability Mass Function

- The binomial probability mass function is:

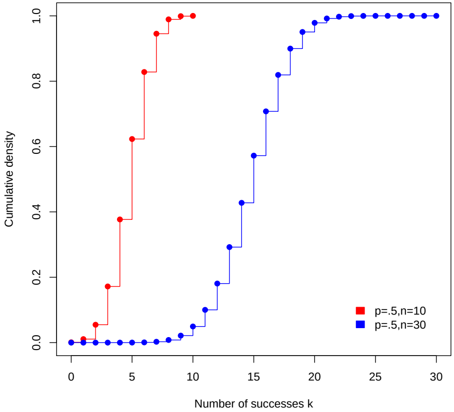
$$f(k; n, p) = P(K = k) = \binom{n}{k} p^k (1-p)^{n-k}$$



Binomial cumulative distribution function

- The binomial cumulative distribution function is:

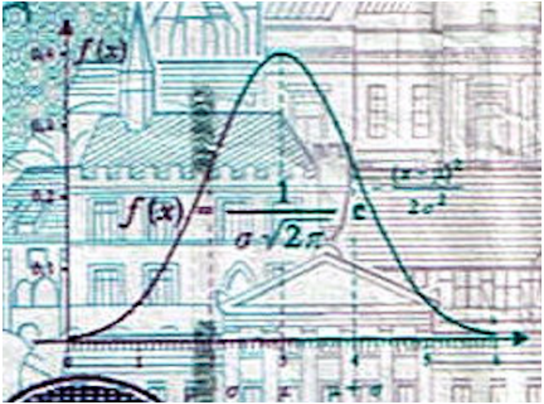
$$F(k; n, p) = P(K \leq k) = \sum_{i=0}^{k} \binom{n}{i} p^i (1-p)^{n-i}$$



Normal distribution $\mathcal{N}(\mu, \sigma^2)$

- Continuous distribution that describes data clustered around the mean.
- Uniquely determined by its mean/median/mode μ and variance σ^2.
- Importance of the normal distribution because of the Central Limit Theorem.

- Continuous distribution that describes data clustered around the mean.
- Uniquely determined by its mean/median/mode μ and variance σ^2.
- Importance of the normal distribution because of the Central Limit Theorem.

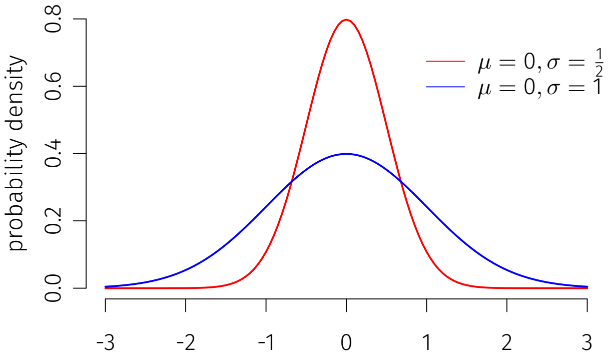


Normal Distribution: Probability Density Function

- Probability density function:

$$f(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]$$

  - $\sigma > 0$ is the standard deviation, μ is the expected value.
  - Note the difference in notation: μ = population mean, $\bar{X}$ = sample mean.

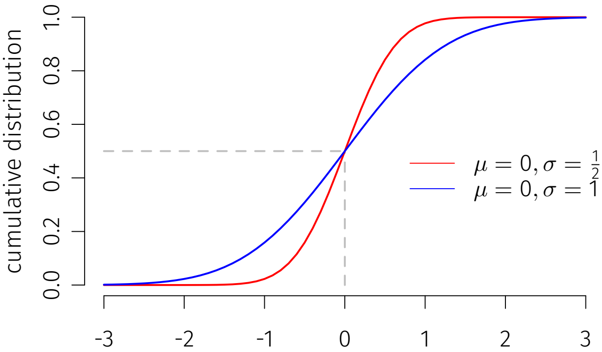


Normal Distribution: Cumulative distribution function

- Cumulative distribution function:

$$F(x;\mu,\sigma^2) = \int_{-\infty}^{x} f(t;\mu,\sigma^2)dt = \Phi\left(\frac{x-\mu}{\sigma}\right)$$

- The CDF allows us to determine the total area under the curve for any given distance from the mean μ.



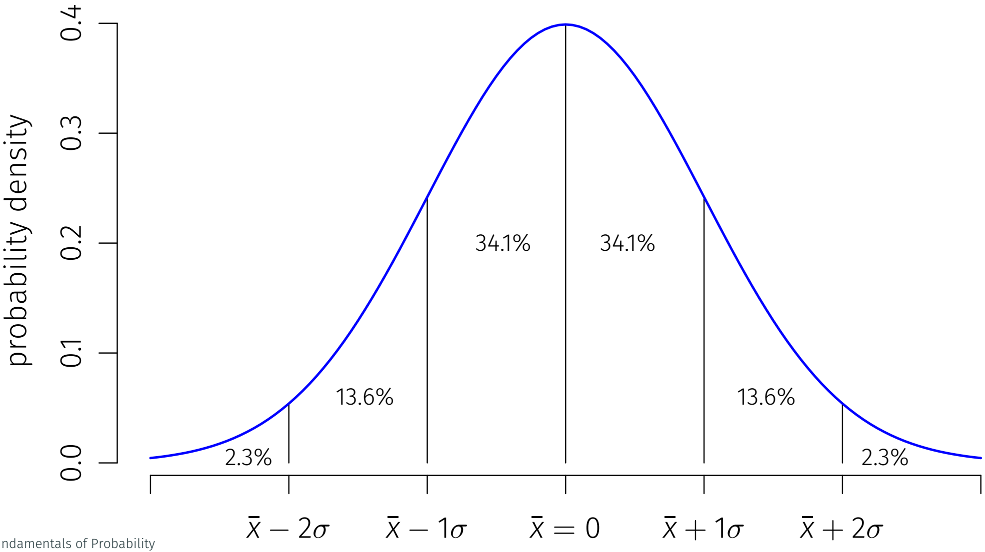
Standardizing Variables

- To compare variables from different distributions, we can standardize them by building so called *z-scores*:

$$z_i = \frac{x_i - \bar{x}}{\sigma}$$

- Standardized variables in such a way will result in a new variable with zero mean and a standard deviation of one.
- For example, we want to compare grades from two students A and B who got grades (on a scale from 1 to 10) in different classes. Student A got 8 in her class, B got 7 in his. To determine who is 'better', we can standardize grades.
- We learn that A's class has a mean grade of 7 with a standard deviation of 2. B's class has mean 6 with standard deviation 1.5.
- So, $z_A = (8 - 7)/2 = 0.5$ and $z_B = (7 - 6)/1.5 = 0.67$

The Standard Normal Distribution $\mathcal{N}(0, 1)$



The Central Limit Theorem

Central Limit Theorem: An Informal Account

- We have a population distribution (not necessarily normal distributed) with mean μ and variance σ^2 and we are interested in its mean.
- Repeatedly taking samples from that population and calculating the mean for each sample yields the sampling distribution of the mean.
- This sampling distribution approaches a normal distribution with mean μ and variance σ^2/n as n increases.
    - This holds regardless of the shape of the original population distribution
    - Basis for application of statistics to many ‘natural’ phenomena (which are the sum of many unobserved random events).
    - How? Take a sample, calculate its mean. Do the same thing again and again. The distribution of sample means will be normal even if the population distribution was not.
    - If you repeatedly draw random samples from the same population, calculate the means and plot them, you get a histogram that approaches a bell-shaped curve.