

# Advanced Quantitative Methods: Probability Theory

---

Thomas Gschwend | Lisa-Marie Müller | Domantas Undzėnas

Week 3 - 25 February 2026

## Final Paper

---

# A Word on the Final Paper

- Do you have a coauthor already? (Full rather than draft paper with more than two authors)
- Track-down an article that (1) interests you and (2) has a replication data set available (e.g., AJPS dataverse)
  - Recent APSR, AJPS and JoP articles have typically the best quality (and available data sets)
  - Be efficient. Avoid collecting your own data for this paper.
- The grade of the paper is independent from the statistical model you use.
- No Compound Poisson-Gamma Regression Models and the like are necessary
- You need to make a substantive point using *advanced quantitative methods* (i.e., no OLS!)
- In Homework 4 (?) you need to provide an abstract of your paper project. Come and talk to me about this before that, i.e. rather soonish!

# Introduction

---

# What should you take home from this class today?

- You will learn that pdf's are your friends. They allow you to calculate probability statements for anything you want.
- We will also learn to appreciate simulation as a tool to calculate such probability statements.

# Probability as a Model of Uncertainty

---

# Probability Theory - Why should we care?

- Probability theory important tool to translate political science theories into appropriate statistical models.
- 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 and a theory of inference to derive statistical model
- Thus, a generalized notation for most statistical models is:
$$Y_i \sim f(y_i|\theta_i, \alpha) \quad \text{stochastic}$$
$$\theta_i = g(X_i, \beta) \quad \text{systematic}$$
  - *Estimation uncertainty*: Lack of knowledge about parameters  $(\beta, \alpha)$ .
  - *Fundamental uncertainty*: Represented by the stochastic component.
- No need to fit data to existing but inappropriate statistical model

# Probability - a model of uncertainty

- $Pr(y|M) = Pr(\text{data}|\text{Model})$ , where  $M = (f, g, X, \beta, \alpha)$
- Probabilities are real numbers  $Pr(A)$  assigned to every event  $A$  of the sample space  $\Omega$ . The sample space can be discrete (e.g., vote-choice), countably infinite (e.g., no. of conflicts), or assumed to be continuous (e.g., duration of governments).
- The sample space is relevant to us because we need to choose the *stochastic component* of our model such that it describes the sample space.
- Three axioms:
  - (a) For every  $A \in \Omega$  holds  $Pr(A) \geq 0$ : Probabilities are non-negative
  - (b)  $Pr(\Omega) = 1$ : The total probability is 1
  - (c) If  $A_1, \dots, A_k$  are mutually exclusive events, then

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

# Definition of Probabilities

- *Conditional probability*: the probability that event  $B$  occurs given that we know  $A$ :

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

- *Independent events (or joint probability of stochastically independent RVs)*:

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

- *Conditionally independent events*:

$$\Pr(A, B|C) = \Pr(A \cap B|C) = \Pr(A|C)\Pr(B|C)$$

- *Law of Total Probability*: If  $A_1, \dots, A_k$  are a disjoint partition of the sample space  $\Omega$ , then

$$\Pr(B) = \sum_{i=1}^n \Pr(B|A_i)\Pr(A_i)$$

# How to solve probability problems?

- If you like math and the problem is not too hard, do this analytically (e.g.,  $Pr(\text{rolling a 1 or rolling a 2}) = ?$ )
- You can also use simulation to solve probability problems (e.g., take 1000 draws with replacement, assign 1 or 0, sum up, divide by 1000, voila!). The result of the simulation will be close to the analytical result. In fact, take more draws to get even closer.
- Research shows that students get it right more often when using simulation rather than math.

# The Birthday Problem

Given a room with 24 randomly selected students, what is the probability that at least two have the same birthday?

# The Birthday Problem

Given a room with 24 randomly selected students, what is the probability that at least two have the same birthday?

```
sims <- 1000
  students <- 24
  days <- seq(1, 365, 1)
  sameday <- 0
for (i in 1:sims){
  room <- sample(days, students, replace = TRUE)
  if (length(unique(room)) < students)
    sameday <- sameday+1
}
cat("Pr(>=2 students same birthday):", sameday/sims, "\n")
```

# Probability Density Functions

---

# What is a Probability Density?

A probability density (probability mass) is a function,  $P(Y)$ , such that

1. Sum over all possible  $Y$  is 1
  - If  $Y$  discrete:  $\sum_Y P(Y) = 1$  (pmf)
  - If  $Y$  continuous:  $\int_{-\infty}^{\infty} P(Y) dY = 1$  (pdf)
2.  $P(Y) \geq 0$  for every  $Y$ .

The cool thing about a probability density is that once you have it we can characterize all possible outcomes with it. For instance, we can compute probability statements.

- $Pr(a \leq Y \leq b) = \int_a^b P(Y) dY$
- $Pr(Y = y) = P(Y = y)$  if  $Y$  discrete
- $Pr(Y = y) = 0$  if  $Y$  continuous

## Examples: Uniform Density on [0,1]

The DGP of  $Y_i$  is such that

- $Y_i$  falls within the interval  $[0,1]$  with probability 1:  $\int_0^1 P(y)dy = 1$
- $Pr(Y \in (a, b)) = Pr(Y \in (c, d))$  if  $a < b, c < d$ , and  $b - a = d - c$ .
  - Why is it a pdf? What is  $Pr(Y=12)$  or  $Pr(Y=.25)$ ?
  - How to simulate from this? E.g., `runif` in R! (pseudo-random generator with a seed number as starting point)

## Another Example: Bernoulli Distribution

- Characteristics about the DGP that generates  $Y_i$ :
  - $Y_i$  has 2 mutually exclusive outcomes (say  $\{0, 1\}$ )
  - Both outcomes are exhaustive ( $\Omega = \{0, 1\}$ )
- Example? How can it go wrong?
- $Pr(Y_i = 1|\pi_i) = \pi_i$ ,  $Pr(Y_i = 0|\pi_i) = 1 - \pi_i$  (Why is this a pmf?)
- The parameter  $\pi_i$  happens to be interpretable as a probability
- Thus,  $Pr(Y_i = y|\pi_i) = \pi_i^y(1 - \pi_i)^{1-y}$

## Expected value of Bernoulli (analytically)

- What will happen on average?

Expected value:

$$\begin{aligned} E(Y) &= \sum_{y \in \{0,1\}} yP(y) \\ &= 0Pr(0) + 1Pr(1) \\ &= \pi \end{aligned}$$

## Expected value of Bernoulli (analytically)

- Expected value of function  $g(Y)$

$$E[g(Y)] = \sum_{y \in \Omega} g(y)P(y)$$

or

$$E[g(Y)] = \int_{-\infty}^{\infty} g(y)P(y)dy$$

For example,

$$\begin{aligned} E[Y^2] &= \sum_{y \in \{0,1\}} y^2 P(y) \\ &= 0^2 Pr(0) + 1^2 Pr(1) \\ &= \pi \end{aligned}$$

## Variance of Bernoulli (analytically)

- By definition (first equation) and some algebra we get

$$\begin{aligned}\text{Var}(Y) &= E[(Y - E(Y))^2] \\ &= E(Y^2) - 2E(Y)E(Y) + E(Y)^2 \\ &= E(Y^2) - E(Y)^2 \\ &= \pi - \pi^2 \\ &= \pi(1 - \pi)\end{aligned}$$

- For which value of  $\pi$  would you expect to get the largest variance?

## Expectation and Variance of Bernoulli (simulated)

- Draw  $u$  from a uniform density on the interval  $[0,1]$
- Set  $\pi$  to particular value
- Set  $y = 1$  if  $u < \pi$  and  $y = 0$  otherwise

```
# Bernoulli Example in R: Simulate E(Y) & Var(Y)
```

```
sims <- 1000           # no. of simulations
pi <- .2              # set parameter
u <- runif(sims)      # draw from uniform pdf
y <- as.integer(u<pi) # compute bernoulli trials
head(y)              # a peek at the results
mean(y)              # calculate simulated mean
var(y)               # calculate simulated variance
```

# Binomial Distribution

First principles:

- $N$  Bernoulli trials  $y_1, \dots, y_N$
- Trials are *independent*
- Trials are *identically distributed*, i.e., all are Bernoulli with the same  $\pi_i = \pi$
- Only the count of those outcomes is observed, i.e.,  $Y = \sum_{i=1}^N y_i$  (Example?)

The pmf is:

$$P(Y = y | \pi) = \binom{N}{y} \pi^y (1 - \pi)^{N-y}$$

- What do we get for  $N = 1$ ?
- $\binom{N}{y} = \frac{N!}{y!(N-y)!}$  because order is not important, i.e., both (101) and (011) yield  $y = 2$ .
- $\pi^y$  is a product taken due to *iid* (independent trials and  $\pi_i = \pi$ )
- One can show that: Mean  $E(Y) = N\pi$ ; Variance  $V(Y) = N\pi(1 - \pi)$

# How to simulate from a Binomial Distribution with parameter $\pi$ and index $N$ ?

- Simulate  $N$  Bernoulli trials with parameter  $\pi$
- Add them up
- Draw samples directly from Binomial distribution using `rbinom`

```
# Simulation of Binomial as sum of 5 independent & identically distributed  
# Bernoulli RVs using rbinom
```

```
y <- rbinom(n=10000, size=5, prob=.2)      # draw from binomial  
head(y)                                   # print result - first peek  
mean(y)                                   # calculate simulated mean  
var(y)                                    # calculate simulated var
```

## Relaxing the *iid* assumption

What if *iid* (*i*ndependent *i*dentically *d*istributed) assumption is unrealistic?

- Relax identical distribution assumption ( $\pi_i = \pi$ ) such that  $\pi$  is a random variable rather than being fixed, thus we need to find  $P(\pi)$  and  $\pi$  falls in the interval  $[0,1]$ .
  - Take *Beta distribution*, i.e.,  $P = B(\rho, \gamma)$ , which can be very flexible (unimodal, bimodal, skewed). Also used to model proportions.
- One can show that *relaxing the independence assumption* by letting  $\pi$  vary according to the Beta distribution one gets the *extended Beta-Binomial distribution*  $P_{ebb}$ .
  - Combine (aka *compound*) Beta and Binomial distributions to get *extended Beta-Binomial distribution*  $P_{ebb}(y_i, \pi|\gamma)$ .  $\gamma$  represents the degree to which  $\pi$  varies across the unobserved realizations of the binary random variables. For  $\gamma = 0$  one arrives at the binomial distribution again.
- Example: [Lauderdale, Benjamin E. \(2012\). Compound Poisson-Gamma Regression Models for Dollar Outcomes That Are Sometimes Zero. \*Political Analysis\*, 20\(3\), 387–399.](#)

# Multinomial Distribution

First Principle:

- Characteristics about the DGP that generates

$Y = (y_1, \dots, y_k)' \sim \text{Multinomial}(n, \pi_1, \dots, \pi_k)$ :

- $n$  repeated, independent trials. Each trial has  $k$  mutually exclusive and exhaustive outcomes, i.e., one outcome must occur
- Probability that outcome  $j$  occurs is  $\pi_j \in [0,1]$  and  $\sum_{j=1}^k \pi_j = 1$
- Let  $y_j$  be a random variable counting how often outcome  $j$  occurs over the  $n$  trials, thus  $\sum_{j=1}^k y_j = n$ .
- The pmf is:

$$P((y_1, y_2, \dots, y_k)') = P(y|n, \pi_1, \dots, \pi_k) = \frac{n!}{y_1! y_2! \dots y_k!} \pi_1^{y_1} \pi_2^{y_2} \dots \pi_k^{y_k}$$

- Example? What happens for  $k = 2$ ?
- $E(Y_j) = n\pi_j$  and  $\text{Var}(y_j) = n\pi_j(1 - \pi_j)$

## Further Univariate Probability Distributions

There are many, many other distributions (and compounds of them) as you can imagine. Just to name a few ...

- Poisson; Negative binomial for modeling counts - discrete, countably infinite, nonnegative
- Normal - continuous, unimodal, symmetric, unbounded
- Log-Normal; Gamma - continuous, unimodal, skewed, bounded from below by zero
- Truncated-Normal - continuous, unimodal, symmetric, bounded from below or above (or both)

Remember: Pick (or construct) a probability distribution to define the stochastic component of your model that best describes the potential values of your outcome variable (i.e., the sample space).