03/08/19 Session 2 Readings

Ch 5: Probability

1. Introduction & Definition
2. Computing P(n)
3. Base Rates

1. REMARKS ON THE CONCEPT OF "PROBABILITY"
- symmetrical outcomes
- relative frequencies
- probability being thought of as "subjective" → sometimes lack of objective criteria
- "nondogmatic" → neither 0 nor 1

2. BASIC CONCEPTS

A) PROBABILITY OF A SINGLE EVENT
- "favorable outcomes"
- if all outcomes are equally likely, (important assumption)

$$\boxed{\text{probability} = \frac{\#\text{ of favorable outcomes}}{(\#\text{ of possible equally outcomes}) = \{\text{equally likely outcomes}\}}}$$

B) PROBABILITY OF TWO (OR MORE) INDEPENDENT EVENTS
- "independent" - the P(B) is the same whether or not A occurs.
- probability of A & B
  - $P(A \text{ and } B) = P(A) \times P(B)$
- probability of A or B
  - $P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$
  - "inclusive or"
  - Deriving $P(A \cup B)$ from $P(A \cap B)$
    - ↳ $A \cap B$ happens
    - ↳ $A \cap (1-B)$
    - ↳ $(1-A) \cap B$

honestly just over complicating it

$P(A) + P(B) = P(A \& B) + P(A \& \text{not } B)) + P(A \& B) + P(A \text{ not } \& B)$

- can use $P(A-1)$ also (prob of NOT getting this)
- conditional probabilities
  - the prob of one event given that another event occured (e.g. not indpt)
  - $P(A|B)$ → P(A) given P(B)
  - $P(A \cap B) = P(A) \cdot P(B|A)$ → if P(A) & P(B) not independent
- Birthday Problem:
  - "If there are 25 people in a room, what's the probability that at least 2 of them will have the same birthday?"

(6)

- (the Birthday problem continued):
  - "what is the probability that no 2 people have the same Birthday?"

$P2 = \frac{364}{365 \text{ Days}}$ ⟶ diff. from the first person's Bday

P2 = prob that 2nd person doesn't have same Bday as person 1

P3 = prob that 3rd person drawn doesn't share Bday w/ anyone drawn given P2

└→ "conditional probability"

$P3 = \frac{363}{365}$

$P4 = \frac{362}{365}$ ...... $P25 = \frac{341}{365}$

↑ 25 people in the room

so: ~~P1~~ $P2 \cdot P3 \cdots P25 = 0.431$, so $1 - 0.431 = \boxed{0.569}$

↑ why don't we multiply P1? = $P1 = 1 = \frac{365}{365}$

- the Gambler's Fallacy ⟶ belief about sequences of independent events
  - coinflip: probability of heads after 5 flips - 1/2
  - no consistent pattern

* conditional probability Demonstration
* Gambler's Fallacy Demonstration
* Birthday Demo

JOINT PROBABILITY: $P(A \text{ and } B) = P(A) \times P(B)$ when A & B are independent ($\cap$)

UNION OF EVENTS: $P(A \text{ OR } B) = P(A) + P(B) - P(A \cap B)$ ($\cup$)

CONDITIONAL PROBABILITY: $P(A|B) = \frac{P(A \cap B)}{P(B)}$

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$

3 PERMUTATIONS AND COMBINATION

A) possible orders:

Number of orders $= n!$

B) multiplication Rule:

$m \times n$

C) permutations — order counts

$${}_nP_r = \frac{n!}{(n-r)!}$$

n things taken r at a time

D) combinations — order does not matter

$${}_nC_r = \frac{n!}{r!\,(n-r)!}$$

4 BINOMIAL DISTRIBUTION

→ (independent trials!) the experiment consists of a sequence of n smaller experiments called trials, where n is fixed in advance of the experiment

- probability distributions for w/c there are only 2 possible outcomes w/ fixed probability = 1
- coin flips: (2x)

$\boxed{H \quad H}$ = 1/4

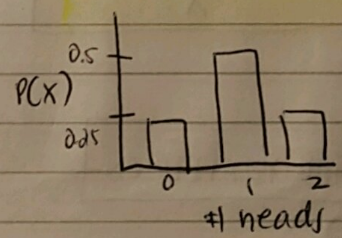
H T

T T

T H

4 possible outcomes for a coin flips

2 heads = 1/4



π = probability of occuring

$$\boxed{P(X) = \frac{N!}{x!\,(N-X)!}\,\pi^{X}(1-\pi)^{N-X}}$$

? diff b/n x and X?

where P(X) is prob of X success out of N trials,

"If you flip a coin twice, what is the prob of getting 1 or more heads?"

π = 0.5 one head + 0.25 two heads

= 0.75

⑤ CUMULATIVE PROPERTIES

- coin toss 12x ⟶ probability of 0 to 3 heads?
  = prob 0 + P(1) + P(2) + P(3)
  "cumulative Binomial probabilities"

⑥ MEAN & STANDARD DEVIATION OF BINOMIAL DISTRIBUTION

$\mu = N\pi$ where π = probability of success on each trial

$\sigma^2 = N\pi(1-\pi)$

⊛ Binomial Demonstration.

$\sigma = \sqrt{N\pi(1-\pi)}$

⑦ POISSON DISTRIBUTION

⟶ when $n \to \infty$, Binomial approaches poisson
$\mu = \sigma^2$ = Mean

- Assumes independence of events
- used to calculate the probabilities of various #'s of successes based on the mean number of successes

$p(x;\mu)$ given $\mu > 0$
$x \geq 0$

$$\boxed{p = \frac{e^{-\mu}\mu^x}{x!}}$$

where $e = \ln(1) \approx 2.7183$
μ = mean # of successes
x = # of successes in question
also var

⑧ MULTINOMIAL DISTRIBUTION

- For computing the prob of obtaining a given # of binary outcomes
- eg. getting 6 heads out of 10 coin flips

$$p = \frac{n!}{(n_1!)(n_2!)\dots(n_k!)} p_1^{n_1} p_2^{n_2} \dots p_k^{n_k}$$

where n = # of total events
n_1 = # of times outcome 1 occurs
p_1 = prob of outcome 1

9 HYPERGEOMETRIC DISTRIBUTION

- used to calculate prob when sampling w/o replacement

$$p = \frac{{}_kC_x \; {}_{(N-k)}C_{(n-x)}}{{}_NC_n}$$ where k = # of successes in pop

$$mean = \frac{(n)(k)}{N}$$

$$\sigma = \sqrt{\frac{(n)(k)(N-k)(N-n)}{N^2(N-1)}}$$

10 BASE RATES

- types of errors:
  1. misses ⟶ have it & failed to detect it (eg. pregnant but negative)
  2. false positive rates ⟶ don't have, indicate that you did (eg. male & preg)
- Base Rate: proportion of people having the disease
- Bayes theorem

$$P(D|T) = \frac{P(T|D)\,P(D)}{P(T|D)\,P(D) + P(T|D')\,P(D')}$$ $P(D') = 1 - P(D)$

11 STATISTICAL LITERACY

- Monty Hall Problem - 3 doors
- school shootings

(10)

CH 9: SAMPLING DISTRIBUTIONS

- inferential statistics ⟶ sample ⇒ population

① INTRODUCTION TO SAMPLING DISTRIBUTIONS

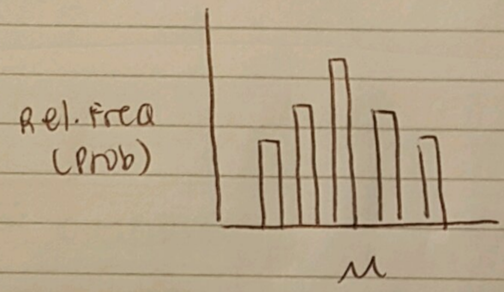
⊛ Basic demo

⊛ sampling size demo

⊛ Central limit theorem demo

} demos

A) Discrete Distributions:
- sampling w/ replacement
- Frequency = what # show up
- Relative Frequency ⟶ $\frac{\text{freq}}{N}$



← sampling distribution of the mean (for a sample size N)

"Discrete distributions"

- "As the $N \to \infty$, the rel freq dist will approach sampling dist."

B) Continuous Distributions

C) Sampling Distributions & inferential statistics
- standard deviation of the sampling dist of the mean (stand. error)

② SAMPLING DISTRIBUTIONS OF THE MEAN

A) mean
- mean of samp dist = mean of pop.

$\mu_M = \mu$

↖ denotes "of the mean"

B) variance

$$\sigma_M^2 = \frac{\sigma^2}{N}$$ ← population var

↑ sample var

can be derived from var sum law

$N\sigma^2$ → var of n #'s

$\frac{1}{N} \cdot$ sum = mean

$$\frac{1}{N^2} \cdot \frac{N\sigma^2}{} = \frac{\sigma^2}{N}$$

$\sigma_M = \frac{\sigma}{\sqrt{N}} \longrightarrow$ standard error

$\mu_{\bar{x}} = \mu$

c) Central Limit theorem $\longrightarrow$ Distribution of means approaches normal

- "given a population w/ a finite mean μ and a finite non-zero var σ^2, the sampling dist of the mean approaches a normal dist w/ a mean of μ and a var of σ^2/N as N, the sample size, increases."

$\boxed{\text{③ SAMPLING DISTRIBUTION OF DIFFERENCE BETWEEN MEANS}}$

$$\mu_{m_1-m_2} = \mu_1 - \mu_2$$

$$\sigma^2_{m_1-m_2} = \sigma^2_{m_1} + \sigma^2_{m_2}$$

var for sampling dist of the diff btn means:

$$\sigma^2_{m_1-m_2} = \frac{\sigma^2_1}{n_1} + \frac{\sigma^2_2}{n_2}$$

standard error

$$\sigma_{m_1-m_2} = \sqrt{\frac{\sigma^2_1}{n_1} + \frac{\sigma^2_2}{n_2}}$$

- use a z table

$$\sigma_{m_1-m_2} = \sqrt{\frac{\sigma^2_1}{n_1} + \frac{\sigma_2^2}{n}}$$

$$= \sqrt{\frac{\sigma^2}{n} + \frac{\sigma^2}{n}}$$

$$= \sqrt{\frac{2\sigma^2}{n}}$$

$\boxed{\text{④ SAMPLING DISTRIBUTION OF PEARSON } r}$

- Fisher: transform r to a var that's normally distributed $\to \frac{z}{SE}$
- r = sample corr

$$z' = 0.5 \ln\left(\frac{1+r}{1-r}\right)$$

$$z \text{ has } SE = \frac{1}{\sqrt{N-3}}$$

- Determine prob. of getting sample corr > 0.75 in sample of 12 w/ pop corr 0.6
  1. Convert them to z scores
  2. SE of z' for N = 12 is 0.33
  3. "Given normal dist, $\mu = 0.693$, $\sigma = 0.33$, what is the prob of obtaining value of 0.973 >?

$$\boxed{z = \frac{X - \mu}{\sigma}}$$

⑤ SAMPLING DISTRIBUTION OF p

- distribution of p closely related to the Binomial dist
  "distribution of p is the dist of the mean number of successes"
- Binomial:
  $$\mu = N\pi$$
- p sampling dist
  $$\mu_p = \pi$$
- $\sigma_{Binomial} = \sqrt{N\pi(1-\pi)}$
- $\sigma_p = \sqrt{\frac{N\pi(1-\pi)}{N}} = \sqrt{\frac{\pi(1-\pi)}{N}}$

  (population proportion vs sample proportion)
- sampling dist of p is discrete rather than continuous

- the sampling dist of p is approx normally dist if N is large and π is not close to 0 or 1.
- Rule of thumb: approx is good if both $N\pi$ and $N(1-\pi)$ are >10.

⑥ STATISTICAL LITERACY

①

03/11/19

Linear Transformations

$T(u_1 + u_2) = T(u_1) + T(u_2)$ for all $u_1, u_2 \in U$

$T(\alpha u) = \alpha T(u)$ for all $u \in U$ and all $\alpha \in \mathbb{C}$

Questions:

① what is the differenle between Random sampling & probability sampling?

② How do you carry out a linear transformation on measures of central tendeny?

③ what are examples for when we would use the multiplication rule for $P(A \cap B)$?

④ Is there a mathematical reason for dividing permutations by the # of ways in which we can re order the numbers to get a combination?

2

ECON 1620:

1) Introduction to Descriptive Statistics
2) Primer on Probability
3) Random Variables
4) Sampling Distribution
5) Intro to stat inference - estimation
6) Stat inference - confidence interval
7) Stat inference: Hypothesis testing

STAT BOOTCAMP:

1
- Dist. of data
- Graphs
- Measures of central tendency & disp
- Corr

2
- Concept of prob
- Bayes theorem
- Using theoretical dist to assign P to events
- Sampling dist

Readings:

1) SW 1, ASW 1, 3
2) LM 2.1-2.3, 2.6, 2.7
3) LM Ch 2.4, 2.5
4) LM 3.1, 3.3, 3.4, SW 2.1
5) LM 3.5, 3.6, 3.9, & SW 2.2, 2.3

1) INTRODUCTION TO DESCRIPTIVE STATISTICS

- Econometrics – application of stats to study economic data & problems
- Relationship & causality
- uncertainty as the major aspect of data analysis

Data
- Categorical or Qualitative
- Numerical or Quantitative
  - Discrete
  - continuous

Statistical methods
- Descriptive statistics
  - collect, present, and describe data
- Probability theory
  - population → sample
  - population parameters used to predict sample
- Statistical inference
  - sample → population
  - sample statistics used to infer about the population
  - "Drawing conclusions or making decisions about a population based on sample results"
  - estimation and confidence intervals
  - hypothesis testing
  - Random sampling

2) DESCRIPTIVE STATS FOR NUMERICAL DATA

- central tendency → mean vs median:
  - statistical inference:
  - mean – more precise estimator
  - median – more robust measure
- dispersion – range, var, σ, coeff of var
- shape – quantiles, skewness, kurtosis ↳ "peakedness"
- relatedness – corr, cov

var vs sd
- var measured in units of x^2

* coefficient of variation – relative measure of var

③

- covariance (in the population)

$cov(X,Y) = \sigma_{xy}$

$$= \frac{1}{N} \cdot \sum_{i=1}^{N} ((X_i - \mu_x) \cdot (y_i - \mu_y))$$

sample cov: $cov(X,Y) = S_{xy}$

$$= \frac{1}{n-1} \cdot \sum_{i=1}^{n} ((X_i - \bar{x}) \cdot (y_i - \bar{y}))$$

if $cov(X,Y) > 0 \longrightarrow$ x and y tend to move in the same direction

$< 0 \longrightarrow$ opposite directions

$= 0 \longrightarrow$ uncorrelated; no linear relationship

- coefficient of correlation:

population: $\rho_{XY} = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$ — has no unit of measurement so we can compare across diff pairs of vars

sample: $r_{xy} = \frac{S_{xy}}{S_x S_y}$

3) A PRIMER ON PROBABILITY THEORY

- random process
- mutually exclusive
- sample space (S)
- event
  - union
  - intersection
  - complement $\longrightarrow$ not in A
  - subset $\longrightarrow B \subset A$

"partition on sample space":
  - Disjoint / mutually exclusive: $A \cap B = \phi$
  - collectively exhaustive: $E_1 \cup E_2 \cup E_n \dots = S$

- Probability - a numerical measure of the likelihood that an event will occur as a result of a random process
- Kolmogorov Axioms:
  1) $0 \leq p(A)$
  2) $p(S) = 1$
  3) $P(E_1 \cup E_2 \cup E_3 \cup \dots) = \sum_i P(E_i)$

- consequences of the Axiom
  1) $P(E_i) = \frac{1}{n}$ for $i = 1 \dots, n$
  2) $P(A) = \frac{n_A}{n}$
  3) $Pr(\emptyset) = 0$
  4) $P(B) \le P(A)$ if B is a subset of A
  5) $0 \le P(A) \le 1$
  6) Addition Rule: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
  7) $P(A^c) = 1 - P(A)$
- probability counting methods:
  - Multiplication Rule $\longrightarrow$ $m \times n$ outcomes
  - permutations (order)

$$P_x^n = \frac{n!}{(n-x)!}$$

  - combinations (no order)

$$C_x^n = \frac{P_x^n}{x!} = \frac{n!}{x!(n-x)!}$$

4) CONDITIONAL PROBABILITY & STATISTICAL INDEPENDENCE ✳ — Refresher on this!

- conditional probability
  - "updated" probability of an event

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

$$P(A \cap B) = P(A|B) \cdot P(B)$$
$$= P(B|A) \cdot P(A)$$

eg: 1, 2, 4, 6 — P(odd) — 6, 3
2, 1, 3, 5 — 24/36 — 3, 6
3, 2, 4, 6 — 5, 4
4, 1, 3, 5 — 20/36 — 4, 5
5, 2, 4, 6
6, 1, 3, 5

$$P(ABC) = P(AB|C) \cdot P(C)$$
$$= P(A|BC) \cdot P(B|C) \cdot P(C)$$

- statistical independence:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = P(A) \quad , \quad P(A \cap B) = P(A) \cdot P(B)$$

⑤

@intmyhipopp?

- Bivariate probabilities: A_i B_j
- Joint & marginal probabilities

5) Bayes theorem → how we should revise prior probabilities in light of new info
- prior / initial probabilities
- revised / posterior probabilities

* disjoint: mutually exclusive

- Law of total probability:

$$P(B) = \sum_{i=1}^{k} P(B/A_i) \cdot P(A_i)$$

conditional version:

$$P(B/C) = \sum_{i=1}^{k} P(B/A_i, C) \cdot P(A_i/C)$$

- Bayes theorem:

$$P(A_j/B) = \frac{P(B/A_j) \cdot P(A_j)}{P(B)}$$

$$= \frac{P(B/A_j) \cdot P(A_j)}{\sum_{i=1}^{k} P(B/A_j) \cdot P(A_i)}$$

- can also be solved in a tabular approach

6) RANDOM VARIABLES & PROBABILITY DISTRIBUTIONS — INTRODUCTION

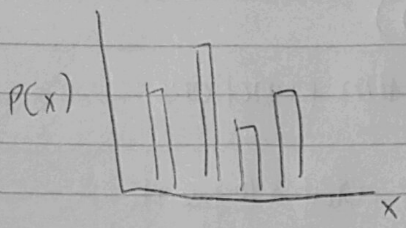
- Random variables → can redefine sample space using function random vars
  - ↳ stochastic var whose numerical value is determined by the outcome of a random process
  - can be discrete or continuous
- Probability mass function (pmf) (probability distribution)
  - describe discrete random variable
  - lists all possible values of RV and probability that each will occur

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

- Probability Distribution

$f(x) = P(X = x) \geq 0$

$\sum_x f(x) = 1$



can be visualized through a histogram

- Cumulative Distribution Function (cdf)
  - discrete or continuous random variable
  - $F(x)$
  - prob that X does not exceed a certain value x

$F(x) = P(X \leq x)$ for $-\infty < x < \infty$

  - we can determine the prob that X will lie in a specified interval

$P(a < X \leq b) = F(b) - F(a)$

$P(X > x) = 1 - F(x)$

- Examples of discrete probability distributions:
  - Bernoulli

$Pr(X = 1) = p$

$Pr(X = 0) = 1 - p$

  - Uniform

$Pr(x) = \frac{1}{k}$

  - Binomial

$Pr(x) = \binom{n}{x} \cdot p^x \cdot (1-p)^{n-x}$

- Continuous random variables:
  - probabilities of values in an interval → CDF and PDF

$F(x) = P(X \leq x)$

$f(x) = \frac{dF(x)}{dx}$

$P(a < X < b) = \int_a^b f(x)$

$\int_{-\infty}^{\infty} f(x)dx = 1$

7

7) JOINT, MARGINAL, & CONDITIONAL DISTRIBUTIONS & INDEPENDENCE

- Joint cumulative distribution function $F(x,y)$ defines the probability that $X \leq x$ and $Y \leq y$

$$F(x,y) = P((X \leq x) \cap (Y \leq y))$$

$F_1(x)$ and $F_2(y) \rightarrow$ marginal distribution functions

$$F_1(x) = \lim_{y \to \infty} F(x,y)$$ } marginal cdf related to joint cdf

- Law of total probability:

$$f_1(x) = \sum_{\text{all } y} f(x,y) = \sum_{\text{all } y} g_1(x|y) \cdot f_2(y)$$

- Bayes theorem:

$$g_1(x|y) = \frac{f(x,y)}{f_2(y)}$$

- Expected value:

$$\mu_x = E(x) = \sum x \cdot f(x)$$

$$\mu_x = E(x) = \int x \cdot f(x)\,dx$$

- the Law of iterated expectation

$$E(x) = E(E(x|y))$$

- mode: the value that maximizes $f(x)$

$$\text{mode}(x) = \text{argmax}_x f(x)$$

- $var(x) = \sigma^2_x$

$$= E((x-\mu_x)^2)$$

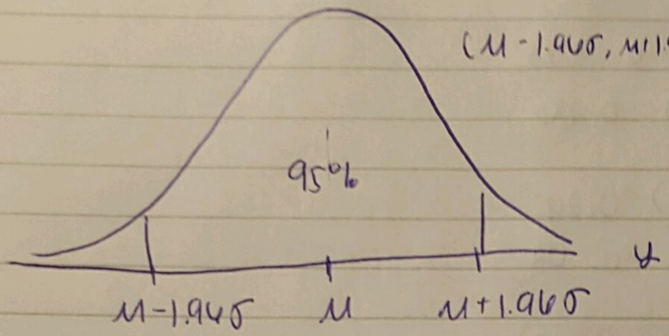
Properties of variance:

$$var(x) = E(x^2) - [E(x)]^2$$

$$var(a \cdot x + b) = a^2 \cdot var(x)$$

$$sd(a \cdot x + b) = |a|\, sd(x)$$

8) Uniform, Binomial, and NORMAL DISTRIBUTION



$(\mu - 1.96\sigma, \mu + 1.96\sigma)$ $\mathcal{N}(\mu, \sigma^2)$

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

PDF

- central limit theorem
- 0 skewness, kurtosis -3

$X \sim N(\mu, \sigma^2)$

standard normal — $\mu = 0$, $\sigma^2 = 1$

use Z to describe a standard normal random variable $Z \sim N(0,1)$

$Pr(Z \le c) = \phi(c)$

$\phi(x) + \phi(-x) = 1$

- If X is distributed normally, then any linear function of x is also distributed normally.

$$x \sim N(\mu, \sigma^2) \Rightarrow (aX + b) \sim N(a\mu + b, a^2\sigma^2) \quad a \ne 0$$

9) RANDOM SAMPLING & SAMPLING DISTRIBUTIONS

10) STATISTICAL INFERENCE - confidence intervals & sample size CALCULATION