ECONOMICS 150:
Quantitative Methods for Economics
Class 5

- Elisa Reading Discussion
- Random Variables

Announcements

- Online Quiz 5 is due before next class, at 9am
- In-class test 1 is a week from today (Class 7);
  - covers material from the first 6 classes
  - I will post a <u>sample</u> in-class test 1 on Canvas later this week
- Reminder: my office hours are WED 1-3pm on Zoom (link on Canvas)

Elisa Test Reading Summary

SUMMARIZE INFORMATION

- The Events:
  - HIV: the person is HIV positive
  - T^+: ELISA test's result is POSITIVE
  - T^-: ELISA test's result is NEGATIVE
- Test's Correctness
  - 99% on HIV positive person "sensitivity" $P(T^+|HIV) = 0.99$
    - 1% false negative
  - 95% on HIV negative person "specificity" $P(T^-|HIV^C) = 0.95$
    - 5% false positive
- Population
  - "low risk," 1 in 500 HIV positive **"prevalence"**

Question 1, Elisa testing

Based on this information, and before Pat took the test, what was the doctor's probability assessment that Pat is HIV positive?

The only information available before the test is that Pat belongs to a "low risk" group

- P(HIV)=1/500= .002

Elisa Testing, Question 2

Given the positive test result, what is the doctor's (revised) probability assessment that Pat is HIV positive?

- What we know:

$$P(T^{+}|HIV) = 0.99$$

$$P\left(T^{-}|\mathrm{HIV}^{C}\right) = 0.95$$

$$P(\mathrm{HIV}) = 1/500 = 0.002$$

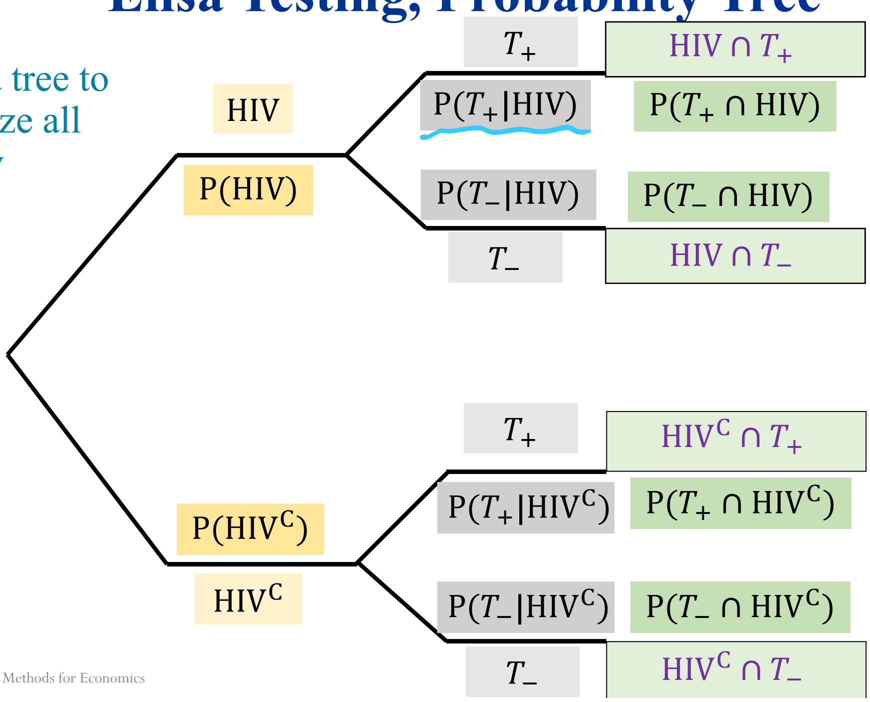
- What we want:

$$P(\mathrm{HIV}\ |T^{+}) = \frac{P(\mathrm{HIV} \cap T^{+})}{P(T^{+})}$$

We need to find numerator and denominator

Elisa Testing, Probability Tree

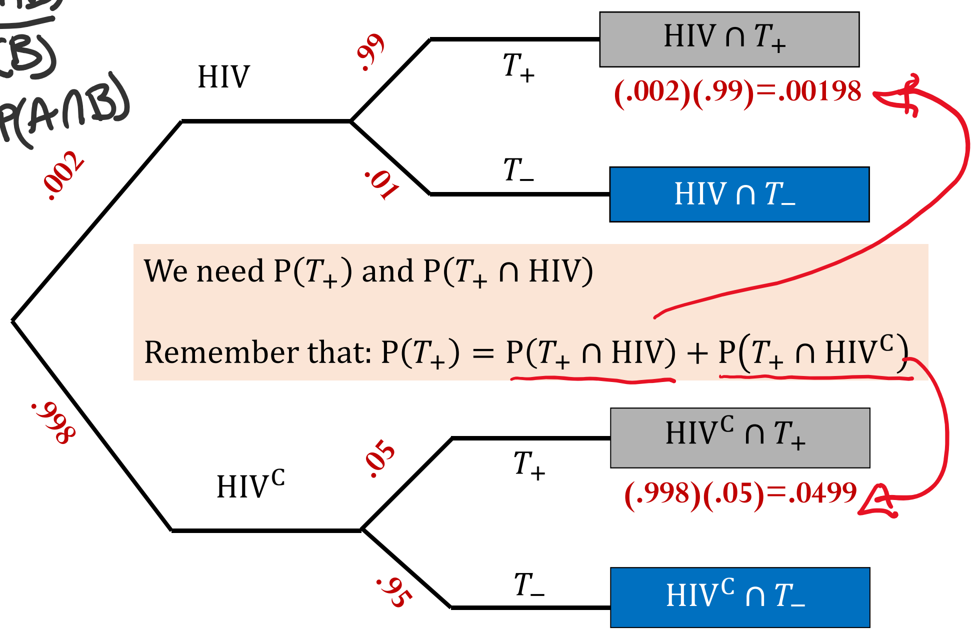
We use a tree to summarize all we know



Question 2, Elisa testing

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

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



We need $P(T_+)$ and $P(T_+ \cap \text{HIV})$

Remember that: $P(T_+) = P(T_+ \cap \text{HIV}) + P(T_+ \cap \text{HIV}^C)$

Question 2, Elisa testing

From the probability tree we have:

$$P(T_+ \cap \text{HIV}) = .00198 \qquad P(T_+ \cap \text{HIV}^C) = .0499$$

Therefore,

$$\begin{aligned} P(T_+) &= P(T_+ \cap \text{HIV}) + P(T_+ \cap \text{HIV}^C) \\ &= .00198 + .0499 \\ &= .05188 \end{aligned}$$

Question 2, Elisa testing

$$P(HIV|T_+) = \frac{P(T_+ \cap HIV)}{P(T_+)} = \frac{.00198}{.05188} = \mathbf{.03816}$$

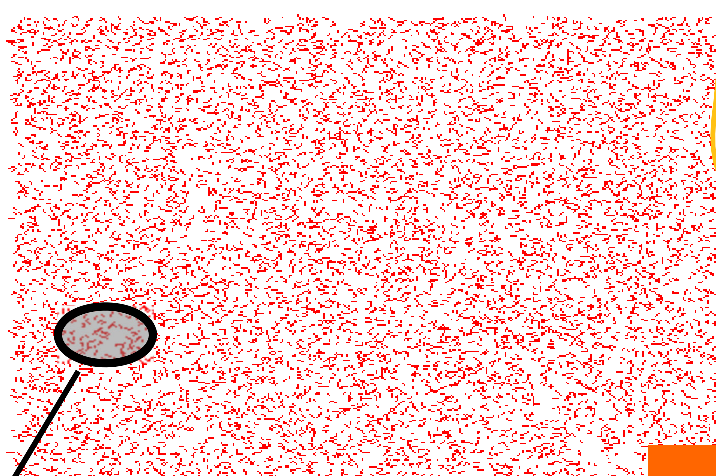
CONCLUSION:

The probability of being HIV positive given that the ELISA test gave a positive result is **less than 4%**

Surprising? YES

Question 2, Elisa testing

- What is going on? Why is the probability of being HIV positive after observing a positive test so low?
- Suppose one tests **1,000,000** individuals
- The picture tries to give some intuition about why the conditional probability is a small number



LOTS OF FALSE POSITIVE

2,000 are HIV$^+$

99% of 2,000 =1,980

998,000 are not HIV$^+$

5% of 998,000 =49,900

Question 3, Elisa testing

What is the purpose of the ELISA test?

- Screening patients at a low cost (identify those who do not have the disease)
  - After a negative test, the confidence the patient is healthy is very high: $\mathrm{P}(\mathrm{HIV}|T_{-}) = 0.00002109$

What should the doctor recommend and why?

- Further investigation:
- another ELISA test? (what is the probability of being HIV positive after two positives?)
  - a second ELISA test: $\mathrm{P}(\mathrm{HIV}\,|T_{+}^{1} \cap T_{+}^{2}) = ???$

Question 3, how about a second ELISA?

Now we want to compute:

$$\mathrm{P}(\mathrm{HIV}\,|T_{+}^{1} \cap T_{+}^{2})$$

- $\mathrm{P}(\mathrm{HIV} \cap T_{+}^{1} \cap T_{+}^{2}) = .00196$
- $\mathrm{P}(T_{+}^{1} \cap T_{+}^{2}) = .0045$
- $\mathrm{P}(\mathrm{HIV}\,|T_{+}^{1} \cap T_{+}^{2}) = .00196\,/\,.0045 = .436$

  Still not enough confidence even after two positive tests.

- $\mathrm{P}(\mathrm{HIV}\,|T_{+}^{1} \cap T_{-}^{2}) = .0000198\,/\,.047425 = .0004$

  Still serve as a good screening test.

Question 3, Elisa testing

What is the purpose of the ELISA test?

- Screening patients at a low cost (identify those who do not have the disease)
  - After a negative test, the confidence the patient is healthy is very high: $\mathrm{P}(\mathrm{HIV}|T_-) = 0.00002109$

What should the doctor recommend and why?

- Further investigation:
- another ELISA test? (what is the probability of being HIV positive after two positives?)
  - a second ELISA test may not help that much: $\mathrm{P}(\mathrm{HIV}\ |T_+^1 \cap T_+^2) = .436$
    - but $\mathrm{P}(\mathrm{HIV}\ |T_+^1 \cap T_-^2) = .0004$
- a more precise test (likely more expensive)? This is most often the answer

HIV Testing In Practice

- New more precise tests became available but ELISA-like tests are still used as initial screening given their low cost and high speed.

- After a positive "screening" test like ELISA, one runs a "confirmatory" test to rule out false positives. Typically more expensive, but specificity is higher that 99.7%.
  - In the case of HIV, Western Blot is one of the confirmatory tests.

- What does ELISA mean? Enzyme-linked immunosorbent assay.

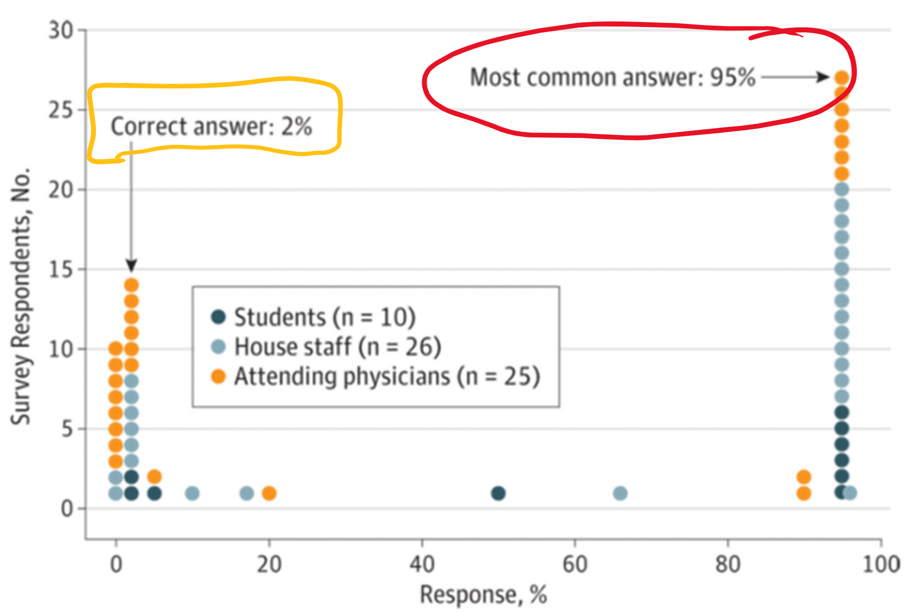
Conditional Probability Can Be Tricky

- The less than 4% chance of being sick given a positive must take into account the initial information (prevalence)
- Because prevalence is very low, the conditional probability cannot be that large
- On the other hand, the conditional probability is much larger than the initial probability assessment
  - From 0.002 to 0.038

Doctors Get This Wrong

In this survey:

- prevalence=1/1000
- sensitivity=0.95
- specificity=0.95



Elephant in the Room

What do we know about Covid-19 testing? Not enough

- Prevalence (how common is the disease) is difficult to know
  - At some point the Omicron variant was believed to be in about 20% of the population in certain areas
- Sensitivity and specificity depend on the test
  - The PCR test is very accurate: more than 99.9% sensitivity and specificity in ideal conditions
  - Fast antigen tests are less accurate.
- The best one can do is build a model with different assumptions and see what makes sense: go to Excel
  - Suppose prevalence is 0.2, sensitivity and specificity are both 0.9.
    - Then the probability of being sick after a positive test is about 70%, while the probability of being sick after a negative test is about 3%

Top Hat Attendance Code for Today

8107

Announcement

Next class will be in person

- Top Hat attendance tracker includes geo-location, so you will need to be in the classroom to register as present
- As usual, I will take attendance a few minutes after the beginning of class

Roadmap: what next?

REGRESSION ANALYSIS

CORRELATION

ESTIMATION | CONFIDENCE INTERVALS | HYPOTHESIS TESTING

SAMPLING | CENTRAL LIMITT THEOREM

NORMAL DISTRIBUTION | OTHER DISTRIBUTIONS

RANDOM VARIABLES

PROBABILITY THEORY

Random Variables

So far, basic events were usually described in words

From now, on we look at a special case in which each outcome is associated with a number

Random Variables

Outcome is uncertain **Outcome is numerically valued**

- Random Variables assign a number to every element in the sample space
  - In math, these are functions from each element of S to numbers

- Example: the value of a farmer's crop:

$$value\ of\ the\ crop = \begin{cases} \$500{,}000 & if \quad good\ weather \\ \$100{,}000 & if \quad bad\ weather \end{cases}$$

Notation

- X denotes a random variable
- x denotes a particular value or realization of X
  - Example: X is value of crop, which can be \$100,000 or \$500,000

- Given some x, write $P(X = x)$ for the probability that $X = x$
  - If all value of X can be listed, write $P(x)$ for short

P(bad weather)

  - The crop's value (X) equals \$100,000 with probability $^1/_3$ and equals \$500,000 with probability $^2/_3$
    - In short: P(X=\$100,000) $= {}^1/_3$ and P(X=\$500,000) $= {}^2/_3$

Discrete Random Variables

A <u>discrete</u> random variable takes a <u>finite</u> number of values

- One can make a list of all the possible values

The **<u>probability distribution</u>** of a discrete random variable lists the probability associated with each of its possible values

- One can make a list of all the possible probabilities

- The probability distribution of X gives $P(X = x)$ **<u>for each possible value</u>** of x
  - Obviously: $0 \leq P(X = x) \leq 1$ and

    $\sum_{all\ values\ of\ x} P(X = x) = 1$

- Example: the probability distribution X = crop's value is: $P(X = \$100{,}000) = \frac{1}{3}$ $\quad and \quad$ $P(\text{X=}\$500{,}000) = \frac{2}{3}$

Discrete random variables

Example:

X can have 5 different values:

X=1

X=2

X=3

X=4

X=5

the probabilities are:

P(X=1)=0.12

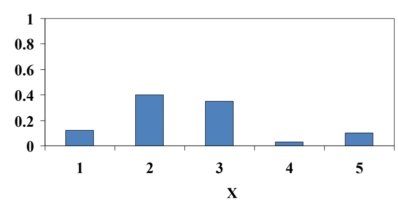
P(X=2)=0.4

P(X=3)=0.35

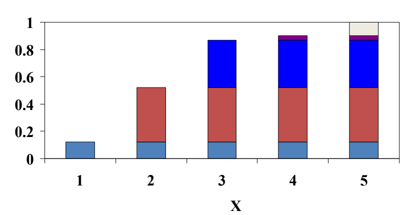
P(X=4)=0.03

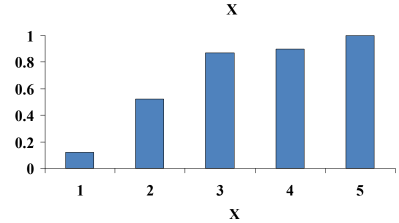
P(X=5)=0.1

Probability Distribution of X



Cumulative Probability Distribution of X is the probability of values smaller or equal than x





Continuous Random Variables

A <u>continuous</u> random variable takes an <u>infinite</u> number of values

- One cannot make a list of all the possible values

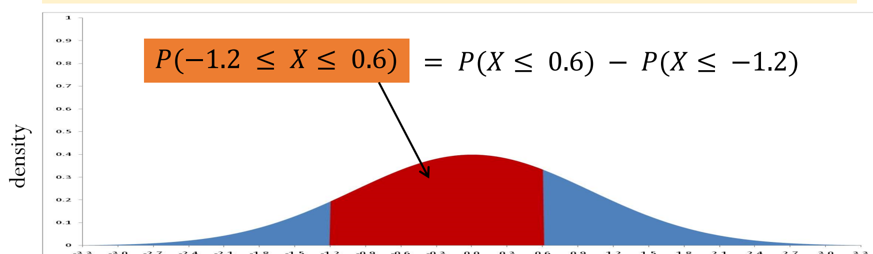
A continuous random variable is described by a **<u>probability density function $f(x)$</u>** of a discrete lists the probability associated with each of its possible values

- For any number a, the area under $f(x)$ gives the probability that $X \le a$
  - Using math: $\text{Prob}(X \le a) = \int_{-\infty}^{a} f(x)dx$
- Obviously: $f(x) \ge 0$ and
  $\int_{-\infty}^{+\infty} f(x)dx = 1$

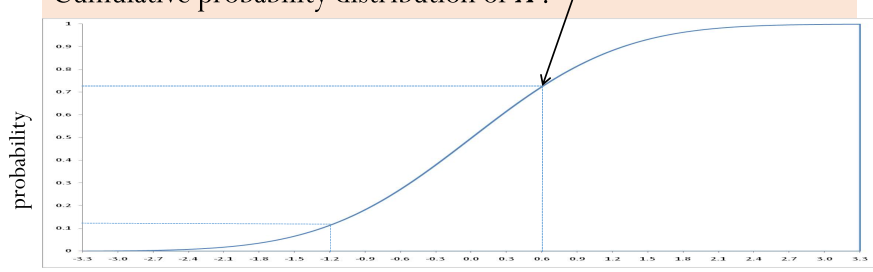
- We can describe the density function using figures

Continuous Random Variables

Probability density function of X:



Cumulative probability distribution of X:



Summary Statistics for Random Variables

- Why Summarize? Because we want to describe a random variable with a few numbers.

- Types of summary measures for random variables
  - Measures of Location
  - Measures of Dispersion, or Spread

- This looks familiar: the main difference is that now we use information about the likelihood of each outcome of the random variable

- Instead of using data, we use information contained in the probability distribution (or in the density function).

Expected Value of a Random Variable

$$E(X) = \mu_X = \sum_{\substack{\text{all possible} \\ \text{values of } x}} x \times P(X = x)$$

- Multiply each value by its probability, and then sum everything up
  - This is a "smart" average in that each outcome is weighted by its likelihood

- A measure of the center of the distribution

  (weighted average of the outcomes where the weights are given by the probabilities)

- The "expected value" is sometimes called the "mean"

- Note: The **expected value** does not need to be a possible outcome.

Next time

- Practice Problems for Test 1
  - No new topics:
    - we will work on selected problems from Practice Problems Sets 1-3