January 9, 2019 10:20 AM

Textbook: Probability and statistics - Evans & Rosenthal,

Chapter 5-9

Measuring Statistical Evidence using relative belief

- M. Evans

website: http://www.utstat.utoronto.ca/mikevans/stac58/stac58.html

Evaluations - Midterm 40%
- final 60%

January 7, 2019
11:15 AM

STAC58 - Statistical Inference

Basics: Introduction

- Statistical inference is not so much about the methods of statistics but the "why".
- what is statistics as a subject all about?
- statistical methods are used in:
  - Finance
  - Machine learning
  - medicine
  - quantum physics

    ⋮

    more!
- Furthermore, "statistical reasoning" is becoming more and more important!
- It is being used as a tool to reason about reality.
- Note: significant decisions are made based on statistical analysis.
- So we want the rules of statistical reasoning to be ____ = logical, free of contradictions, ____, etc ... so we feel confident that whatever the conclusion/Inference we draw makes sense.

January 9, 2019 10:21 AM

- Current state of statistics
    - Many different points of view about what the the correct statistical reasoning is.
    - This makes learning the subject hard.

- Purpose of this course (STAC58 - Statistical Inference)

    1.) Survey the various approaches

    2) present the outline of a logical way to develop a theory of statistical reasoning.

- Some phenomenon / context in the real world that we have questions about

- Questions like: 1) what is the value of some quantity of interest?

    eg. mean half life length of a neutron

    Answer: An estimate of assessment of its error

    2) Does a certain quantity take a particular value?

    Answer: hypothesis assessment - evidence for or against and a measure of strength.

Theory tells you how accurate estimate is.

- when can statistical inference play a role?

Statistical Problems

- The first thing we need to do is be very clear about what a statistical problem is.
- It is all based on "measuring" and counting.
- We have a population Ω = a finite set of objects of interests.

Eg. Ω = set of all students enrolled at U of T on Jan 7, 2019

- $\underbrace{\#(\Omega)}_{\text{cardinality / \# of items in the set}} < \infty$

- we have a measurement(s) defined on Ω

$X : \Omega \to \mathcal{X}$ $\quad \omega \in \Omega \quad X(\omega)$

- for $\omega \in \Omega$ = set of students at U of T.

Define

$X_1(\omega)$ = height of ω in cm (interval)

$X_2(\omega)$ = weight of ω in kg (interval)

$X_3(\omega)$ = gender of ω (categorical)

$$X(\omega) = \begin{pmatrix} X_1(\omega) \\ X_2(\omega) \\ X_3(\omega) \end{pmatrix}$$

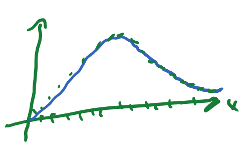
$X(\omega)$ is 3d measurements

$X : \Omega \to \underbrace{R^1 \times R^1 \times \{M, F\}}_{\mathcal{X}}$

- $X = (X_1, X_2, X_3) : \Omega \to R \times R \times \{M, F\}$
- Ω and X define relative frequency distribution over $\mathcal{X}$.

$$f_X(x) = \frac{\#\{\omega : X(\omega) = x\}}{\#(\Omega)}$$

= proportion of individuals in Ω whose X measurements is $x \in \mathcal{X}$.



simplify by introducing continuous approximation

discrete is too hard for ~~fam~~! approx it.

- note: i) $0 \le f_X(x) \le 1 \quad \forall x \in \mathcal{X}$

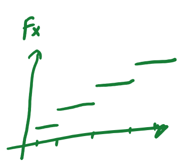
    ii) $\sum_{x \in \mathcal{X}} f_X(x) = 1$

    and only finitely many $x \in \mathcal{X}$ have $f_X(x) > 0$.

- when $\mathcal{X} = R$ (or an interval)

$$F_X(x) = \frac{\#\{\omega : X(\omega) \le x\}}{\#(\Omega)}$$ = Cumulative distributive function of X (CDF of X)

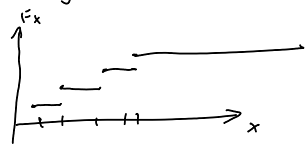
$$= \sum_{z \le x} f_X(z)$$



} Step function

$f_X(x) = F_X(x) - F_X(x - \underset{\text{epsilon}}{\epsilon})$, where $F_X(x - \epsilon) = \lim_{z \uparrow 0} F_X(z)$

- So F_X and f_X are two equivalent ways of presenting a frequency distribution.

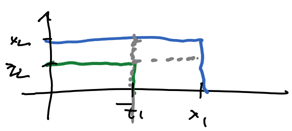


- when $\mathcal{X} = \mathbb{R}^2$

$$F_X(x_1, x_2) = \frac{\#\{\omega : x_1(\omega) \le x_1,\ x_2(\omega) \le x_2\}}{\#(\Omega)}$$

$$= \sum_{\substack{z_1 \le x_1 \\ z_2 \le x_2}} f_X(z_1, z_2)$$

$$f_X(x_1, x_2) = \lim_{z_1 \uparrow x_1} \left[F_X(x_1, x_2) - F_X(x_1, z_2) - F_X(z_1, x_2) + F_X(z_1, z_2) \right]$$



So, $F_X \Longleftrightarrow f_X$

- The whole point of any statistical analysis is to learn something about F_X.
- how do we do this?
- If possible we do a census, namely compute $X(\omega)\ \forall \omega \in \Omega$ of the form f_X.
- Typically count (return to this in a moment)
- why do we want to know F_X?

eg relationships among variables.

- Suppose (X, Y), where $X: \Omega \to \mathcal{X}$, $Y: \Omega \to \mathcal{Y}$ and we want to know if there is a relationship between X & Y on Ω.

- form the conditional relative frequency distribution.

$$f_{Y|X}(y|x) = \frac{\#\{\omega \mid X(\omega) = x,\ Y_2(\omega) = y\}}{\#\{\omega \mid X(\omega) = \Omega\}}$$

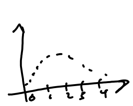
$$= \frac{f_{(X,Y)}(x, y_2)}{f_X(z)}$$

Definition: X and Y are related variables over Ω if $f_{Y|X}(\cdot|x)$ changes as x changes.

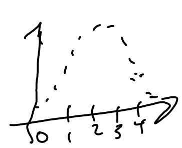
- The "form" of the relationship between X and Y is given by how $f_{Y|X}(\cdot|z)$ changes as x_2 changes.

eg. Ω = 1st year students at UofT
Y = GPA as of Dec 31, 2015.
X = gender

$f_{Y|X}(\cdot|m)$



$f_{Y|X}(\cdot|F)$

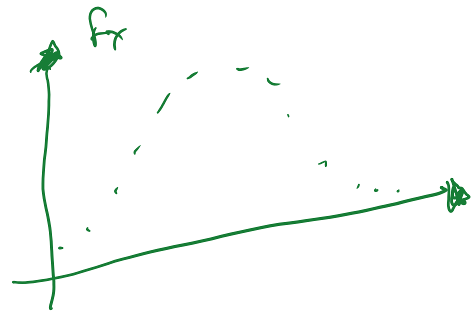


- often simplifying assumptions are introduced.
- regression assumption: $f_{Y|X}(\cdot|x)$ changes at most through its mean as x changes,

$$t(y|x)(x) \quad = \frac{1}{\#\{\omega : X(\omega) = x\}} \sum_{\{\omega : X(\omega) = x\}} Y(\omega)$$

Ex

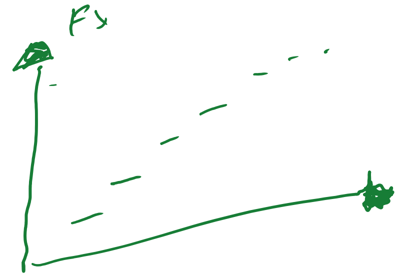
$$= \sum_y y\, f_{Y|X}(y|x)$$



Ω = a finite population

X , $\Omega \to \mathcal{X}$ a measurement , $\omega \in \Omega$
$X(\omega) \in \mathcal{X}$

ex, $f_X(x) = \dfrac{\#\{\omega : X(\omega) = x\}}{\#(\Omega)}$

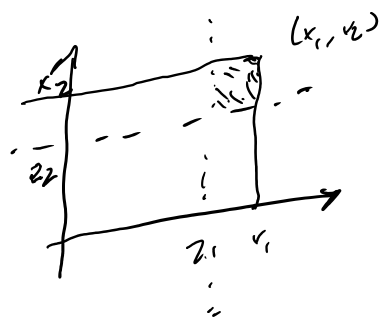


- $\mathcal{X} = \mathbb{R}^1$

$F_X(x) = \dfrac{\#\{\omega : X(\omega) \le x\}}{\#(\Omega)}$

- if you have
$F_X \Leftarrow$ cdf you can
get $f_X \Leftarrow$ pdf
$f_X \to F_X$

- $X = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} : \Omega \to \mathbb{R}^2$



$$F_{(X_1,X_2)}(x_1,x_2) = \frac{\#\{\omega : X_1(\omega) \le x_1, X_2(\omega) \le x_2\}}{\#(\Omega)}$$

$$= \sum_{\substack{z_1 \le x_1 \\ z_2 \le x_2}} f_{(X_1,X_2)}(z_1,z_2)$$

$$f_{(X_1,X_2)}(x_1,x_2) = \lim_{\substack{z_1 \uparrow x_1 \\ z_2 \uparrow x_2}} \Big(F_{(X_1,X_2)}(x_1,x_2) - F_{X_1,X_2}(x_1,z_2) - F_{(X_1,X_2)}(z_1,z_2) + F_{X_1,X_2}(z_1,z_2) \Big)$$

- The whole point is to know f_X (or F_X)
- how do we do this?
  - conduct a census boi
    - measure every $\omega \in \Omega$
    - often not possible but can comprimise
- why do we want to know F_X ?

eg. relationship among variables

- Suppose $(X, Y) : \Omega \to \mathcal{X} \times \mathcal{Y}$ and we want to know if there is any relationship between X & Y.

- form the conditional relative frequency distribution

$$f_{Y|X}(y|x) = \frac{\#\{\omega : X(\omega) = x, Y(\omega) = y\}}{\#\{\omega : X(\omega) = x\}} = \frac{f_{(X,Y)}(x,y)}{f_X(x)}$$

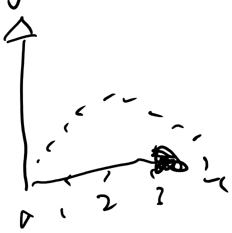
Def X & Y are related variables over a population Ω, if $f_{Y|X}(\cdot|x)$ changes as X changes.

eg. Ω = students at UofT

$X(\omega)$ = gender $\quad Y(\omega)$ = GPA

$f_{Y|X}(\cdot|m)$

$f_{Y|X}(\cdot|F)$



- The form of the relationship

if X and Y have no relationship then

$f_Y(y|x) = f_Y(y)$

$\Longleftrightarrow f_{(X,Y)}(x,y) = f_X(x) f_Y(y)$

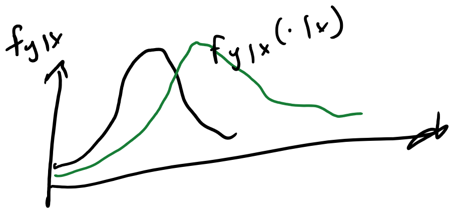
- the form of the relationship between X & Y when it exists is given by how $f_{Y|X}(\cdot|x)$ changes with x

January 9, 2019 11:01 AM

- Often simplifying assumptions are made
- regression assumptions
  - here $y = \mathbb{R}$ and we assume
    $f_{y|x}(\cdot|x)$ changes at most through μ

$$E(y|x)(x) = \frac{\sum_{\{\omega : x(\omega) = x\}}}{\# \{\omega : x(\omega) = x\}}$$

= average value of y in the sub population $\{\omega : x(\omega) = x\}$



- regression assumption
  - same distribution but shifted.

Linear regression assumption

$E(y|x) \in \mathcal{L}\{g_1, \dots, g_k\}$ where

$g_i : \mathcal{X} \to \mathbb{R}$ $i = 1, \dots, k.$

i.e $E(y|x)(x) = \beta_1 g_1(x) + \dots + \beta_k g_k(x)$, for some $\beta_1, \dots, \beta_k \in \mathbb{R}$

eg $g_1(x) = 1$, $g_2(x) = x$, $g_3(x) = x^2$