650.3 ↦ Adam Kapelner 1/31/17

- Let X be a r.v. which ~~...~~ realizes to a datum "x".
  coin toss : $X \sim \text{Bernoulli}\left(\frac{1}{2}\right) \Rightarrow x = \underset{\text{of head}}{\text{success}} = 1$

- $\exists$ 2 types of r.v.'s.
  1. Discrete : Random variables that have $|\text{supp}(X)| \leq |\mathbb{N}|$ (at most countable)
     "size of support" "size of $\mathbb{N}$"

     *Def: Support of X $(\text{supp}(X))$ is all possible realization values.

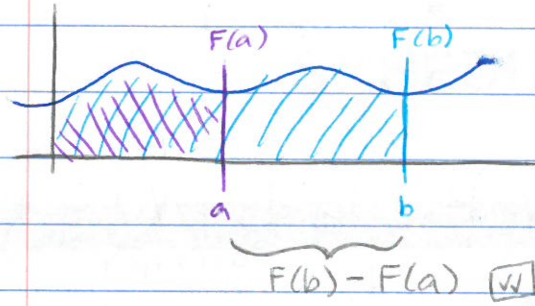
     : If $|\text{Supp}(X)| = 1 \Rightarrow X \sim \text{Deg}(c) = \{c \ 1$
     Degenerate = only 1 outcome
     : ~~$\exists \ p(x) = P\{X = x\}$~~
     : $\exists \ p(x) = P\{X = x\}$ is called the Probability mass function (PMF)
       ↳ Domain of PMF is the Support (if you put anything $\neq$ in support, you get PMF = 0)
       ↳ Range of PMF : $(0,1]$ $(f : \text{Support}(X) \to (0,1])$

- Cumulative Distribution Function (CDF) $\Rightarrow F(X) = P(X \leq x)$

  2. Continuous : Random Variables that have $|\text{Supp}(X)| = |\mathbb{R}|$ (uncountable infinity)
     : If $|\text{Supp}(X)| = 1 \rightarrow X \sim \text{Deg}(c) = \{c \ 1$
     : $\exists \ f(x) = F'(x)$ called the probability density function (PDF).
       ↳ Domain of PDF is the Support
       ↳ Range : $(0, +\infty)$ $(f : \text{Supp}(X) \to (0, +\infty))$
       ↳ Represents an area ✓

* Connection between PDF & PMF :

$P\{X \in [a,b]\} = F(b) - F(a) = \int_a^b \underbrace{f(x)}_{F'(x)} dx$



connecting density to actual probability

- Discrete r.v.'s are defined by their PMFs \ cdfs.
- Continuous r.v.'s are defined by their PDFs \ cdfs.
  - +Discrete r.v.'s ≠ have pdfs ⇒
  - Continuous r.v.'s ≠ have pmfs. ⇒ bc it would be zero everywh

Example distributions : all parameters : From now on, we denote random variable by parameters by θ and parameter spaces Θ

<u>Discrete</u>
- $X \sim$ Bernoulli (p) : $p^x(1-p)^{1-x} \xrightarrow{p=\theta} \theta^x(1-\theta)^{1-x}$
- $X \sim$ Binomial (n,p) : $\binom{n}{x} p^x (1-p)^{n-x} \xrightarrow{p=\theta} \binom{n}{x}\theta^x(1-\theta)^{n-x}$

<u>Continuous</u>
- $X \sim$ Exponential (λ) : $\lambda e^{-\lambda x} \xrightarrow{\lambda=\theta} \theta e^{-\theta x}$
- $X \sim$ Normal (μ, σ^2) : $\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$

$\mu = \theta_1$, $\sigma^2 = \theta_2^2$: $\frac{1}{\sqrt{2\pi\theta_2^2}} e^{-\frac{1}{2\theta_2^2}(x-\theta_1)^2}$

- $X \sim$ Bernoulli (θ) : $X \in Supp(X) = \{0,1\}$
- $X \sim$ Binomial (n,θ) : $X \in Supp(X) = \{0,1,\dots,n\}$
- $X \sim$ Exponential (θ) : $X \in Supp(X) = (0,+\infty)$
- $X \sim$ Normal (μ,σ^2) : $X \in Supp(X) = (-\infty,+\infty) = \mathbb{R}$

Def: A Parametric Model $\mathcal{F}$ is a set of r.v. models with finite parameters

: $\mathcal{F} = \{ \underbrace{p(x;\theta)} : \theta \in \Theta \}$

meaning the probability of x assuming the specific value of the parameter(s) θ. s.t. $\theta \in$ Parameter space Θ

EX: $\mathcal{F} = \{ \theta^x (1-\theta)^{1-x} : \theta \in (0,1) \} \Rightarrow$ all possible Bernoulli.

* assume the model from the getgo.

joint pmf/pdf

- If $x_1, x_2, \cdots$ are iid realizations then $p(x_1, x_2, \cdots x_n; \theta) \overset{\text{by Independence}}{=}$ product of marginals $= p(x_1)p(x_2)\cdots p(x_n) \overset{\text{by iid}}{=} \prod_{i=1}^{n} p(x_i;\theta)$

* Think of coin flips, → They ≠ talk to each other.
Same coin, same θ.
Same process each time.

- In the Real World, you "observe" data $x = (0,0,1,0,1,0)$ and assume iid then you pick an $\hat{\mathcal{F}}$. In general you ≠ know $\theta \rightarrow \theta$ is unknow. (≠ know if the coin is fair)
θ = "Mother Nature herself"

Figuring out θ is the goal of statistical inference. $\exists$ 3 main objectives:
(1) Point estimation : best guess of θ
(2) Confidence Set : A set of "likely" θ's (interval)
(3) Theory Testing = Value Testing = Hypothesis Testing

4

EX: Assume Bernoulli

$p(001010;\theta) = \prod_{i=1}^{6} p(x_i;\theta) = \prod_{i=1}^{6} \theta^{x_i}(1-\theta)^{1-x_i}$

if $\theta = \frac{1}{2}$: $p(001010; \theta=\frac{1}{2}) = 0.5^6 = 0.0156 \approx 1.5\%$

if $\theta = \frac{1}{4}$: $p(001010; \theta=\frac{1}{4}) = (\frac{1}{4})^2(\frac{3}{4})^4 = 0.0198 \approx 1.9\%$

Better
higher than other model

$P(x_1,\dots,x_n;\theta) = \mathcal{L}(\theta; x_1,\dots,x_n)$ ← Likelihood func.

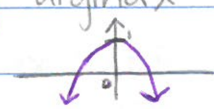
know θ, get data — given data, find θ

point of view of data — point of view of θ

(optimizing) $\underset{\theta \in \Theta}{\operatorname{argmax}} \{\mathcal{L}(\theta; x_1, x_2, \dots, x_n)\} = \hat{\theta}_{MLE}$

Max. Likelihood Estimator (r.v)
Max. Likelihood Estimate (#)

Def: argmax : the value of x to give us the max. value



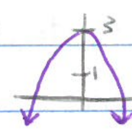
← $f(x) = 1 - x^2$

$\max\{f(x)\} = 1$

$\operatorname{argmax}\{f(x)\} = 0$ vs $(\underline{0},1)$ max

- If you take a one-to-one ↑ing function of $\mathcal{L}$, $\hat{\theta}_{MLE} \neq \Delta$

EX: take $f(x)+2$:



argmax = still $\underline{0}$

- A typical $\ell(\theta; x_1, x_2, \dots, x_n) = \ln(\mathcal{L}(\theta; x_1, \dots, x_n))$

= log likelihood Function.

$\hat{\theta}_{MLE} = \underset{\theta \in \Theta}{\operatorname{argmax}} \{\mathcal{L}(\theta; x_1, \dots, x_n)\} = \underset{\theta \in \Theta}{\operatorname{argmax}} \{\ell(\theta; x_1, \dots x_n)\}$