# Math 341 Lec 1 1/29/17

- syllabus

Let $X$ be a r.v., $x$ be a realization (data)

aka data-generating process

$x \in Supp[X]$ set of all unique values that can be realized

- $X$ can be discrete i.e. $|Supp(X)| \leq |\mathbb{N}|$ at most countably infinite different unique realizations

  $p(x) := P(X = x)$ prob. mass function (PMF)

  $p : Supp(X) \to (0, 1]$

  why? why?

  $\sum_{x \in Supp(X)} p(x) = 1$

  $F(x) := P(X \leq x)$ cumulative distr. function

  $$= \sum_{\{y \,:\, y \in Supp(X) \text{ and } y \leq x\}} p(y)$$

- Cont. $|Supp(X)| = |\mathbb{R}|$ uncountably inf. # of different realizations

  $F(x)$ same, $f(x) := F'(x)$ is the prob. dens. function (PDF)

  $\int_{Supp(X)} f(x)\,dx = \int_{-\infty}^{\infty} f(x)\,dx = F(\infty) - F(-\infty) = 1 - 0 = 1$

  by FTC calc. $P(X \in [a,b]) = P(X \leq b) - P(X \leq a) = F(b) - F(a) = \int_a^b f(x)\,dx$

  $Supp(X) = \{x : f(x) > 0\}$ $\quad f : Supp(X) \to (0, \infty)$

r.v.'s are identified by their CDF/PMF if discrete and CDF/PDF if cont. Examples:

Discrete
$$\begin{cases} X \sim \text{Bern}(p) := \underbrace{p^x(1-p)^{1-x}}_{p(x)}, & x \in \text{Supp}(x) = \{0,1\} \\ X \sim \text{Binom}(n,p) := \underbrace{\binom{n}{x} p^x(1-p)^{n-x}}_{p(x)} & \text{" " " "} = \{0,1,\dots,n\} \end{cases}$$

Cont.
$$\begin{cases} X \sim \text{Exp}(\lambda) := \underbrace{\lambda e^{-\lambda x}}_{f(x)} & x \in (0,\infty) \\ X \sim N(\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} & x \in \mathbb{R} \end{cases}$$

Using PMF's

$p(x) = p^x(1-p)^{1-x}$ $\quad P(X=1) = p(1) = p, \quad P(X=0) = p(0) = 1-p$

what is $p$? A "tunning knob" (parameter) that controls how often 0,1's realize from the process.

What are the values of $p$ that make sense $p \in (0,1)$. Why not 0 or 1?

Parameter space:

If $p=0$, $X$ only realizes 0's $\quad X \sim \text{Deg}(0) = \{0 \text{ w.p. } 1$

If $p=1$ " " " " 1's $\quad X \sim \text{Deg}(1) = \{1 \text{ w.p. } 1$

technically a r.v. but not interesting

All values of parameter that does not yield a degenerate case.

Let $\theta$ denote parameter (unknown), $\vec{\theta}$ denote multiple params (unknown) and $\Theta$ denote parameter space.

$X \sim \mathrm{Bern}(\theta) := \theta^{x}(1-\theta)^{1-x}$

$X \sim \mathrm{Bin}(n,\theta) := \binom{n}{x}\theta^{x}(1-\theta)^{n-x}$

n considered "fixed" and known. Not a parameter.

$$X \sim \mathrm{Bin}(\theta_2,\theta_1) = \binom{\theta_2}{x}\theta_1^{x}(1-\theta_1)^{\theta_2-x}$$

two parameter model $\qquad \Theta = (0,1) \times \mathbb{N}$

wait until after

$p(x)$ now denotes both

## Parametric Model

$p(x)$ PMFs and $f(x)$ PDF's

$$\mathcal{F} := \{p(x;\theta) : \theta \in \Theta\} \quad \text{s.t. } \dim[\Theta] < \infty$$

For the Bernoulli $\qquad p(x;\theta)$ prob of $x$ with $\theta$ known

$$\mathcal{F}_{Bern} = \{\theta^{x}(1-\theta)^{1-x} : \theta \in (0,1)\}$$ all possible Bernoulli models

---

$p(x_1, x_2, \ldots, x_n; \theta)$

joint mass func
joint dens. func

If $X_1, \ldots, X_n \overset{iid}{\sim}$ ← independent & identically distr.

$$\Rightarrow p(x_1,\ldots,x_n;\theta) = p(x_1;\theta)\,p(x_2;\theta)\cdot\ldots\cdot p(x_n;\theta) = \prod_{i=1}^{n} p(x_i;\theta)$$

In the real world you see $X = \langle 0,0,1,0,1,0 \rangle$, the data.

Then, you pick $\mathcal{F}$, an assumption! But you don't know $\theta$!

Figuring out $\theta$ is the goal of inference. There are generally 3 sub-goals:

1. Point estimation. Provide best guess of $\theta$
2. Confidence set. Provide a range of possible $\theta$'s.
3. Theory testing. Evaluate a theory about $\theta$.

---

e.g. imagine data above and assume $\mathcal{F}$ is iid Bernoulli

$$p(\langle 0,0,1,0,1,0 \rangle, \theta) = \left(\theta^0 (1-\theta)^1\right)\left(\theta^0 (1-\theta)^1\right) \cdots = \theta^2 (1-\theta)^4$$

if $\theta = 0.5$ $\quad = 0.5^6 = 0.0156$

if $\theta = 0.25$ $\quad = 0.25^2 \, 0.75^4 = 0.0198$

$\theta = 0.5$ is "more likely" than $\theta = 0.25$

the data is fixed, and we want to know how probable the value of $\theta$ are.

$$\mathcal{L}(\theta; x) = p(x; \theta)$$

"likelihood function": what is the likelihood of "seeing" the parameter at a certain value

prob of $\theta$ with x known

prob of data with $\theta$ known