# Math 341 Lec 1 1/29/18

- syllabus

Let $X$ be a r.v. which means it "realizes" to "data" $x$. The value $x \in Supp(X)$

It can • discrete $|Supp(X)| \le |\mathbb{N}|$ if $|Supp(X)| = 1 \Rightarrow X \sim Deg(c) := \{c \text{ w.p. } 1$ → Supp is cntbl. if and at most two vals.

$\exists\, p(x) := P(X = x)$, a prob mass function (PMF)

$Supp(X) = \{x : p(x) > 0\}$

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

$\exists\, F(x) := P(X \le x)$

• Cont. $|Supp(X)| = |\mathbb{R}|$ uncntbl. infin. each realization contains infinite information

$Supp(X) = \{x : f(x) > 0\}$

$\exists\, f(x) := F'(x)$ which is the prob density func (PDF) $:= F'(x)$

where $F(x) := P(X \le x)$, the cumul. distr. function (CDF) $\Rightarrow$ all r.v.'s have $F(x)$ but $p(x), f(x) \to$ n.a.

Note: $P(X \in [a,b]) = P(X \le b) - P(X \le a) = F(b) - F(a) = \int_a^b f(x)\,dx$ F.T.C.

if $f(x)$ exists ($F$ needs to be diff.)

r.v.'s are defined by their PMF/PDF/CDF's. Some common ones are below:

Discrete
$$\begin{cases} X \sim Bernoulli(p) := p^x (1-p)^{1-x}, & x \in Supp(X) = \{0, 1\} \\ X \sim Binomial(n, p) := \binom{n}{x} p^x (1-p)^{n-x}, & x \in Supp(X) = \{0, 1, \ldots, n\} \end{cases}$$

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

$p^x(1-p)^{1-x}$ what is $p$? If you remember the bernulli model,

$p(X=1)=p$, $p(X=0)=1-p$, $p$ is a "tuning knob" AKA a parameter

non-degenerate values of $p$ are $(0,1)$ why not $0$ or $1$?

~~Bernulli($p$) r.v. is a parametric model i.e. a model with finite parameters.~~ stochastic

Parameter space: all set of possible values of the parameter(s) in the model

From now on, parameters are denoted $\theta$ and param. space, $\Theta$.

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

$X \sim \text{Bin}(n,\theta) = \binom{n}{x}\theta^x(1-\theta)^{n-x}$ ← n considered fixed

or

$X \sim \text{Bin}(\theta_2,\theta_1) = \binom{\theta_2}{x}\theta_1^x(1-\theta_1)^{\theta_2-x}$ ← More complicated model (not done in this class)

Parametric Model ($\mathcal{F}$): is a set of r.v. models that are parametrized with finite parameters

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

s.t. $\dim[\Theta] < \infty$

$p(X;\theta)$ "prob of $X$ assuming a value of $\theta$" i.e. prob's are different depending on the specific model (specific $\theta$).

For a bernulli,

$$\mathcal{F} := \{\theta^x(1-\theta)^{1-x} : \theta \in (0,1)\}$$

from now on $p(x) = f(x)$ ... no notational difference ... just gotta know from context

What is iid?

Assume $x_1, \ldots, x_n$ are realizations from an iid model, then:

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

"jmf" or

"jdf" for continuous

plain $x$ means all $x_1 \ldots x_n$

In the real world, you see $x \stackrel{iid}{=} \langle 0, 0, 1, 0, 1, 0 \rangle$ (the data)
and if you want to model this data, you pick an $\mathcal{F}$,
a **class** of parametric models, but you don't know $\theta$.
Figuring out $\theta$ is the goal of "inference" and there are generally 3:

(1) Point estimation. Give best guess of $\theta$.

(2) Confidence set. Give a range of possible $\theta$'s.

(3) Theory testing. Evaluate whether or not a theory about $\theta$ is true

e.g. data above, let $\mathcal{F}$ = Bernoulli model, $\Theta = (0,1)$

$$p(\langle 0,0,1,0,1,0 \rangle; \theta) = \prod_{i=1}^{6} p(x_i; \theta)$$

if $\theta = 0.5$

$$= \prod_{i=1}^{6} 0.5^{x_i} (1-0.5)^{1-x_i} = 0.5^6 = 0.0156$$

if $\theta = 0.25$

$$= \prod_{i=1}^{6} 0.25^{x_i} (1-0.25)^{1-x_i} = 0.25^2\, 0.75^4 = 0.0198$$

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

We really want to know how probable the value of $\theta$ are when data are fixed:

$$\mathcal{L}(\theta; x) := P(x; \theta)$$ the "inverse question"

likelihood

the likelihood asks the question "what is the likelihood of seeing the parameter?" It is equal to the probability of the data under the param value. Higher probs $\Rightarrow$ higher lik of a given $\theta$.

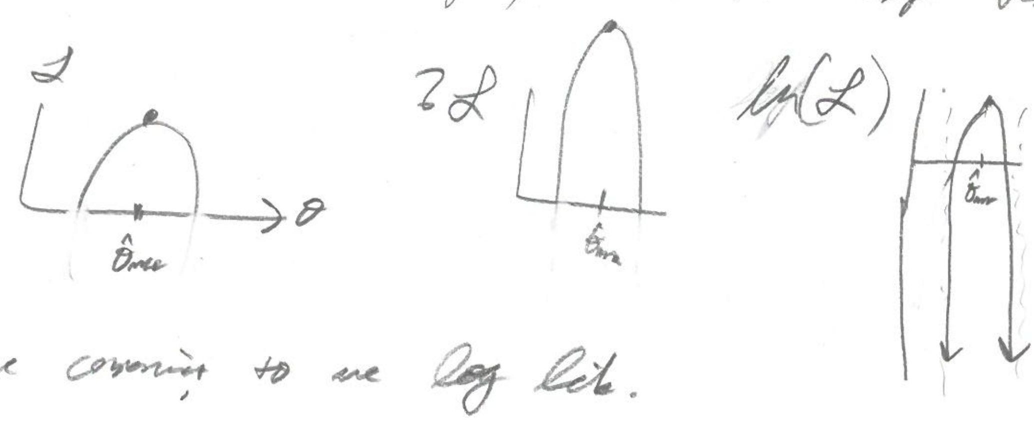
Is $\mathcal{L}$ a PMF/PDF of a r.v.? No... there is no r.v. spanning over $\theta$'s.

The most likely value of $\theta$ is..

$$\hat{\theta}_{MLE} := \underset{\theta \in \Theta}{\operatorname{argmax}} \left\{ \mathcal{L}(\theta; x) \right\}$$

maximum likelihood estimator

Note that it remains the same if you take a 1:1 monotonically increasing function of $\mathcal{L}$




Usually it's more convenient to use log lik.

$$\ell(\theta; x) := \ln\left(\mathcal{L}(\theta; x)\right)$$

$$\hat{\theta}_{MLE} = \underset{\theta \in \Theta}{\operatorname{argmax}} \left\{ \ell(\theta; x) \right\}$$