1/3/17

Let X be a r.v. which realizes to a datum "x."

$X \sim \text{Bernoulli}(\frac{1}{2})$ $\quad x = 1$ $\quad$ r.v.'s have "supports"

Two Types of r.v.'s $\quad$ $|Supp(X)|$ is all possible realization variables

· Discrete have $|Supp(X)| \leq |\mathbb{N}|$

at most... countable

If $|Supp(X)| = 1 \Rightarrow X \sim (\text{Degenerate})$

$Deg(c) := \{ c \text{ wp } 1$

$\exists\ p(x) := P(X = x)$ called the probability mass function (PMF) (probability to realize level)

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

$F(x) := P(X \leq x)$ cumulative distribution function (CDF)

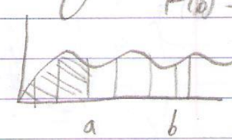
· Continuous r.v.'s have $|Supp(X)| = |\mathbb{R}|$

$\exists\ f(x) := F'(x)$ called the probability density function (pdf)

Domain is also support $\quad f : Supp(X) \to (0, \infty)$

not a pmf b/c not getting back probability

<u>density</u>



$P(X \leq [a, b]) = F(b) - F(a)$

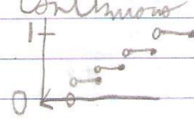
(CDF of b − CDF of a)

$= \int_a^b \underbrace{f(x)}_{F'(x)} dx$

Discrete r.v.'s are defined by their PDFs\CDFs

Continuous r.v.'s " PDFs\CDFs

Moment-generating Characteristic



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

← real analysis fact: if uncountable, sum $= \infty$

$X \sim Bern(p)$ $\quad X \sim Binom(n, p)$ $\quad$ } Discrete

$X \sim Exp(\lambda)$ $\quad X \sim N(\mu, \sigma^2)$ $\quad$ } Continuous

Parametric Families
Parameter Space

Given - Bayesian
conditional

Bern $:= p^x(1-p)^{1-x}$
Bin $:= \binom{n}{x} p^x (1-p)^{n-x}$
Exp $:= \lambda e^{-\lambda x}$ (Memoryless)
Normal $:= \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$

These are called parameters

$x \in Supp(X) = \{0, 1\}$
$x \in Supp(X) = \{0, 1, \ldots, n\}$
$x \in Supp(X) = (0, \infty)$
$x \in Supp(X) = \mathbb{R} = (-\infty, \infty)$

Parameter Space of Bernoulli

$p \in (0, 1)$ — always a success
never a success

From now on, parameters are denoted θ and param. spaces Θ (capital theta) $\theta \in \Theta$

$X \sim Bern(\theta) := \theta^x (1-\theta)^{1-x}$
$X \sim binom(n, \theta) := \binom{n}{x} \theta^x (1-\theta)^{n-x}$
$X \sim Exp(\theta) := \theta e^{-\theta x}$
$X \sim N(\theta, \theta^2) := \frac{1}{\sqrt{2\pi\theta^2}} e^{-\frac{1}{2\theta^2}(x-\theta)^2}$ $\vec{\theta} = \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix}$

Parametric Model $\mathcal{F}$
is a set of r.v. models w/ finite parameters

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

either a PMF or PDF

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

prob. of x assuming the specific value of the parameter(s) θ.

If $x_1, x_2, \ldots, x_n$ are i.i.d. realizations then iden. distr.

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

joint PMF/PDF ind.

Same PMF b/c same process each time
In the real world you "observe" data

not model selection

Standard Frequentist flavor — throughout most of course, essentially

$x = \langle 0, 0, 1, 0, 1, 0 \rangle$ and assume i.i.d. then you pick an F, but θ unknown.
Do not know θ (mother nature)
Figuring out θ is the goal of statistical inference.
There are three main objectives:
(1) point estimation (best guess of θ)
(2) confidence set... a set of "likely" θ's.
(3) Theory testing (θ value testing)
(Hypothesis testing)

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

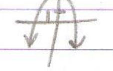
$$p(001010; \theta = \tfrac{1}{2}) = 0.5^6 = 0.0156$$
$$p(001010; \theta = \tfrac{1}{4}) = (\tfrac{1}{4})^2(\tfrac{3}{4})^4 = 0.0198$$
more likely b/c pdf matches better

$$\underset{\text{joint density}}{\overset{\text{changes}}{p(x_1, \ldots, x_n}}; \underset{\leftarrow \text{locked}}{\theta}) = \mathcal{L}(\underset{\text{changes}}{\theta}; \overset{\text{locked}}{x_1, \ldots, x_n}) \text{ likelihood}$$

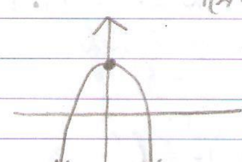
Mathematically same, conceptually different (perspective)

$$\underset{\theta \in \Theta}{\operatorname{argmax}} \{\mathcal{L}(\theta; x_1, \ldots, x_n)\} = \hat{\theta}_{MLE}$$
↑ maximum likelihood (#) estimate / estimator (r.v.)

$f(x)+2$



$f(x) = 1 - x^2$



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

If you take a 1:1 increasing function of $\mathcal{L}$, $\hat{\theta}_{MLE}$ won't change.

$$\ell(\theta; x_1, \ldots, x_n) := \ln(\mathcal{L}(\theta; x_1, \ldots, x_n)) \text{ "log-likelihood"}$$

2/2/17

$$P(X_1, \ldots, X_n; \theta) = \mathcal{L}(\theta; X_1, \ldots, X_n)$$
$$\hat{\theta}_{MLE} := \underset{\theta \in \Theta}{\operatorname{argmax}} \{\mathcal{L}(\theta; X_1, \ldots, X_n)\}$$
$$= \underset{\theta \in \Theta}{\operatorname{argmax}} \{\ell(\theta; x_1, \ldots, x_n)\}$$