Definition 0.1. Random Variable: realizes to a data "x," denoted by X

Definition 0.2. Supports: all possible realization values, denoted by $\text{Supp}(X)$

Note: Real variables have "supports."

Two Types of Random Variables:

- Discrete:
$$|\text{Supp}(X)| \leq |\mathbb{N}|$$
where it is countable,
If $\text{Supp}(X) = 1$, then $X \sim \text{Deg}(c) = \{1 \text{ outcome}\}$.

  There exists $p(x) = P(X = x)$ called the probability mass function or pmf which relates $\text{Supp}(X) \to \big(0, 1\big)$.

  $F(x) = P(X \leq x)$ is called the cumulative density function (cdf)

- Continuous:
$$|\text{Supp}(X)| \leq |\mathbb{R}|$$
There exists $f(x) = F'(x)$ called the probability density function (pdf) where $f : \text{Supp}(X) \to \big(0, 1\big)$. The cumulative density function is denoted $P(X \in \big[a, b\big])$ which is equal to
$$\int_a^b \underbrace{f(x)}_{F'(x)} dx = F(b) - F(a)$$

Note: Discrete random variables are defined by their pmf and cdf whereas continuous random variables are defined by their pdf and cdf. Types of Distributions:

- Discrete
  - $X \sim \text{Bern}(p) = p^x(1-p)^{1-x}$ where $x \in \text{Supp}(X) = \{0, 1\}$.
  - $X \sim \text{Bern}(n, x) = \binom{n}{p} p^x 1 - p^{1-x}$ where $x \in \text{Supp}(X) = \{0, 1, 2, \ldots, n\}$.
- Continuous
  - $X \sim \text{Exp}(\lambda) = \lambda e^{-\lambda x}$ where $x \in \text{Supp}(X) = \big[0, \infty\big)$.
  - $X \sim \text{N}(\mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$ where $x \in \text{Supp}(X) = \big(-\infty, \infty\big)$.

From now on, parameters will be denoted by θ and parameter spaces will be denoted Θ (capital θ). This transforms the above distributions to the following:

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

Definition 0.3. Parametric Models: a set of random variable models with finite parameters, denoted by $\mathcal{F}$

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

where $p(x;\theta)$ is the probability of assuming the value of the parameter θ.

Example 0.1. Let's say we want to model the parameters for a normal distribution. We can represent this as follows:

$$\hat{\theta} = \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix} = \begin{bmatrix} \mu \\ \sigma \end{bmatrix}$$

Note: Parametric models can be either pmf or pdf.

If $x_1, x_2, \ldots, x_n$ are realizable, then

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

In the real world, let's say we "observe" data as follows: $x = \langle 0, 0, 1, 0, 1, 0 \rangle$ and we assume IID. Then you pick a parametric model, $\mathcal{F}$, but θ is not known. Figuring out θ is the point of statistical inference.

Three Main Objectives:

- Point Estimation: best guess of θ
- Confidence Set: a set of "likely" θ's
- Theory Testing: θ value testing, also called hypothesis testing

Let's say we assume a Bernoulli distribution for the data set $x = \langle 0, 0, 1, 0, 1, 0 \rangle$. Then

$$p(0, 0, 1, 0, 1, 0) = \prod_{i=1}^{6} \theta^x(1-\theta)^{1-x}$$

For example. let's take $\theta = \frac{1}{2}$, then

$$p(x_1, x_2, \ldots, x_6; \frac{1}{2}) = 0.5^6 = 0.0156$$

Let's take $\theta = \frac{1}{4}$, then

$$p(x_1.x_2.\ldots,x_6;\frac{1}{4}) = (\frac{1}{4})^2(\frac{3}{4})^4 = 0.0198$$

Out of the two choices for θ, the second one is more likely since the second model has a higher probability than the first one. But we can take an infinite number of guess for θ. There has to be a better way to figure out θ.

Definition 0.4. Likelihood Function:

$$p(x_1,x_2,\ldots,x_n;\theta) = \mathcal{L}(\theta;x_1,x_2,\ldots,x_n)$$

where the joint density function on the left hand side is in perspective of $x_1,x_2,\ldots,x_n$ and allowing it to change whereas the likelihood function on the right hand side is in perspective of θ and allowing it to change.

To get the best model, we must optimize $\text{argmax}\{\mathcal{L}(\theta;x_1,x_2,\ldots,x_n)\}$.

Definition 0.5. $\overset{n}{\theta}_{MLE}$: maximum likelihood estimate or maximum likelihood estimate, must be within Θ

Example 0.2. If $f(x) = 1 - x^2$, then $\max\{f(x)\} = 1$ but $\text{argmax}\{f(x)\} = 0$.

Note: If you taken an increasing 1-1 function of $\mathcal{L}$, then θ_{MLE} won't change.

Example 0.3. Let $l(\theta;x_1,x_2,\ldots,x_n\} = \ln(\mathcal{L}(\theta;x_1,x_2,\ldots,x_n))$ be a log-likelihood function. Then

$$\hat{\theta}_{MLE} = \underset{\theta\in\Theta}{\text{argmax}}\{l(\theta;x_1,x_2,\ldots,x_n)\}$$

or

$$\hat{\theta}_{MLE} = \underset{\theta\in\Theta}{\text{argmax}}\ln(\mathcal{L}(\theta;x_1,x_2,\ldots,x_n))$$