Probability and Statistics

(November-2)

Statistics (Estimation and hypothesis testing)

Generally : given data (set)

Goal: Infere key things which describe the given data in best possible way.

Definition: Let $X_1, X_2, \ldots, X_n$ be iid random variable with the common density $f_\theta(\cdot)$.

Then the collection $X_1, \ldots, X_n$ is called as random sample.

Note : The joint pdf of $X_1, \ldots, X_n$ is given by

$$f(x_1, \ldots, x_n) = \prod_{i=1}^{n} f_\theta(x_i)$$

Ex: Let $x_1, x_2, \ldots, x_{10}$ be a random sample from Bernoulli (p).

prove certain statistical properties

What do we observe??

$x_1 = 1$, $x_2 = 1$, $x_3 = 0$, $x_4 = 1$, $x_5 = 1$, $x_6 = 0$, $x_7 = 0$, $x_8 = 0$, $x_9 = 1$, $x_{10} = 0$

I have observed $X_i = x_i$

realisations of the r.v. X_i

Data: $(1, 1, 0, 1, 1, 0, 0, 0, 1, 0)$

random sample - Data

Random sample of size $n = 10$.

Ex: Let X_i : denote the time for a battery to discharge completely.

$X_1, X_2, \cdots, X_{55}$: random sample

$x_1 =$ 10 hrs , $x_2 =$ 13 hrs, $x_3 =$ 3 hrs, 22 mins, ...

$x_{55} =$ 17 hrs

$[0, \infty)$

Random sample of size $n = 55$.

Ex: Revisit the example of $X_1, \dots, X_{10}$ random sample from $\boxed{\text{Bernoulli}(p)}$.

$$\underbrace{(\overset{x_1}{1}, \overset{x_2}{1}, 0, 1, 1, 0, 0, 0, 1, \overset{x_{10}}{0})}\ x \leftarrow \text{data, observed}$$

$p \in \underbrace{[0,1]}_{\substack{\shortparallel \\ \Theta}}$

Q: Can you guess 'p' from the evidence (data)??

Obvious choice: $\hat{p} = \dfrac{\sum_{i=1}^{10} x_i}{10} = \bar{x}$

Consider a r.v. $\bar{X} = \dfrac{\sum_{i=1}^{10} X_i}{10}$

$\boxed{\bar{X} : \text{random variable}}$ and $\hat{p} = \bar{x} = \dfrac{\sum_{i=1}^{10} x_i}{10}$ is a realisation.

$\bar{X}$ is a random variable (later we will call this an estimator of parameter p) and

$\bar{x} = \dfrac{\sum_{i=1}^{10} x_i}{10}$ is a realization (later we will call this an estimate of p obtained from the data) of $\bar{X}$.

For different set of repetitions of the random expt, one would get different realizations of $\hat{p} = \bar{x} = \dfrac{\sum_{i=1}^{10} x_i}{10}$.

Q: How do we understand variation in different values of $\hat{p}$??

A: This variation can be understood from the density of $\bar{X}$.

In this particular case,

$$X_1, X_2, \ldots, X_{10} \overset{iid}{\sim} \text{Bernoulli}(p)$$

$$X_1 + \cdots + X_{10} \sim \text{Binomial}(10, p)$$

$$E(\bar{X}) = E\left(\frac{X_1 + \cdots + X_n}{n}\right) = p$$

$$\text{var}(\bar{X}) = \text{var}\left(\frac{X_1 + \cdots + X_n}{n}\right) = \frac{1}{n^2} \text{var}(X_1 + \cdots + X_n) = \frac{np(1-p)}{n^2}$$

$$= \frac{p(1-p)}{n} = \boxed{\frac{p(1-p)}{10}}$$

Few definitions.

Let $X_1, X_2, \ldots, X_n$ be a random sample from $f_\theta(\cdot)$

Def: 1 Family of distributions

$$\mathcal{F} = \{ f_\theta(\cdot) \mid f_\theta \text{ is a pdf of } X \text{ and } \theta \in \Theta \}$$

(capital theta)

Θ: parameter space

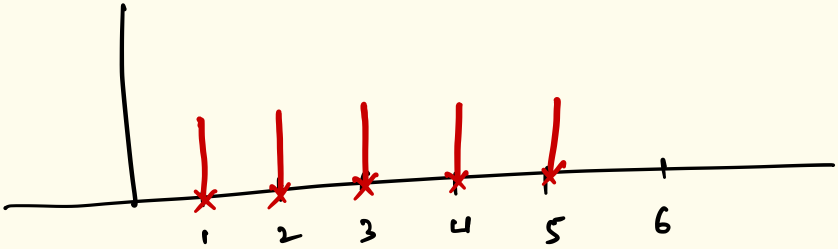
e.g. Let $X_1, \ldots, X_n$ be a random sample from Binomial (n, p)

$$\mathcal{P} = \left\{ \binom{n}{x} p^x (1-p)^{n-x} \;\middle|\; n \in \mathbb{N},\ p \in [0,1] \right\}$$

Here $\theta = (n, p)$: parameter vector

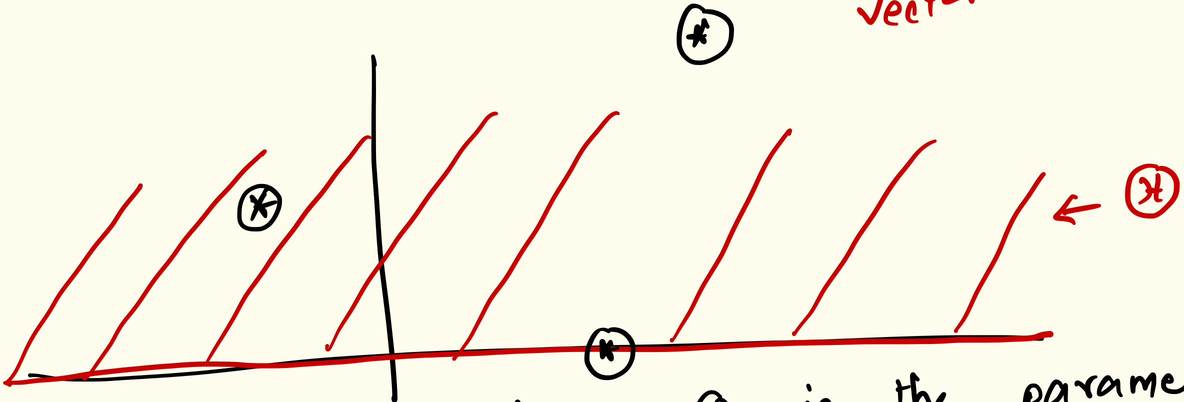
$\Theta = \mathbb{N} \times [0,1]$: parameter space.

$\subseteq \mathbb{R}^2$



Ex: Let $X_1, \dots, X_n$ be a random sample from $N(\mu, \sigma^2)$

$$\mathcal{N} = \left\{ \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} \;\middle|\; \underbrace{(\mu, \sigma^2)}_{\substack{\theta \\ \text{parameter vector}}} \in \underbrace{\mathbb{R} \times \mathbb{R}_+}_{\substack{\Theta \\ \text{parameter space}}} \right\}$$



Note: Any vector denoted as $\circledast$ in the parameter space is a valid "guess" for the parameter.

Want to learn:

i) Estimation (parametric).

ii) Definition of Estimator and estimate.

iii) Good properties of estimator.

iv) Methods of estimation

(Method of moments MoM
Maximum likelihood estimators MLE)

Non-parametric estimation.

Random sample drawn from a non-parametric pdf / pmf

Definition: Statistic (NOT to be confused with statistics)

A statistic is a function of data (random sample) but it is free from any unknown parameter.

Ex: Let $X_1, X_2, \ldots, X_n$ be a random sample from $N(\mu, \sigma^2)$

Case (i): Let μ be known. σ^2 : unknown.

$$\sum_{i=1}^{n} (X_i - \mu)^2$$ is a statistic.

<u>case (ii)</u> μ is unknown, σ^2 is unknown.

$\sum_{i=1}^{n}(x_i-\mu)^2$ is <u>NOT</u> a statistic.

<u>case (iii)</u> μ, σ^2 both unknown:

$$\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{\sigma^2}$$ is <u>NOT</u> a statistic.

$$\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n}$$ is a statistic.

Let $x_1, x_2, \dots, x_n \overset{iid}{\sim} f_\theta(\cdot)$

where $f_\theta \in \{ f_\theta \mid \theta \in \Theta \}$

Objective: We may be interested to estimate a function of parameters θ, denoted by $g(\theta)$.

Ex: In case $x_1, x_2, \dots, x_n \overset{iid}{\sim} \text{gamma}(\alpha, \lambda)$

$\theta = (\alpha, \lambda)$: parameter vector.

$g(\theta) = \text{mean} = \dfrac{\alpha}{\lambda}$

Definition:

Estimator: Let $X_1, X_2, \dots, X_n$ be a random sample from $f_\theta(\cdot)$

A statistic $T(X_1, \dots, X_n)$ when used to estimate a function of parameters $g(\theta)$ is called as an estimator of $g(\theta)$.

Note: $T(X_1, \dots, X_n)$ being a function of random variables is itself a random variable and the density of $T(X_1, \dots, X_n)$ is called as sampling density.

Ex: Recall the example of $X_1, \ldots, X_n$ being a random sample from Bernoulli (p):

Here parameter $\theta = p$

$\Theta = [0,1] =$ parameter space.

To estimate $p = g(\theta)$

Estimator: $T(X_1, \ldots, X_n) = \bar{X} = \dfrac{\sum_{i=1}^{n} X_i}{n}$

Definition (estimate): The value of $T(X_1, \ldots, X_n)$ evaluated at the given data points $X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n$ is called as an <u>estimate</u>.

Examples:

Let $x_1, x_2, \cdots, x_n$ be a random sample from $f_\theta(\cdot)$ with $E(x_i) = \mu$ and $var(x_i) = \sigma^2$ $\Big\}$ for $i = 1, 2, \cdots, n$

Then μ can be estimated by

i) $T(x_1, \cdots, x_n) = x_1$

ii) $T(x_1, \cdots, x_n) = \dfrac{5x_3 + 7x_8}{12}$

iii) $T(x_1, \cdots, x_n) = \bar{x}$

iv) $T(x_1, \cdots, x_n) = \sum_{i=1}^{n} a_i x_i$ where $\Sigma a_i = 1$, $a_i \in [0, 1]$

σ^2 can be estimated by

i) $s_1^2(x_1, \ldots, x_n) = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2$

ii) $s^2(x_1, \ldots, x_n) = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$

Properties of estimators

Unbiased estimator: Let $T(x_1, \ldots, x_n)$ be an estimator of $g(\theta)$ which satisfies

$$E[T(x_1, \ldots, x_n)] = g(\theta) \quad \forall \, \theta \in \Theta,$$

then $T(x_1, \ldots, x_n)$ is called as an unbiased estimator of $g(\theta)$.

<u>Ex:</u> let $x_1, \dots, x_n$ be a random sample from $f_\theta(\cdot)$.

Let $\mu = E(X_i)$ and $\sigma^2 = \text{var}(X_i)$

for $i = 1, 2, \dots, n$

We want to estimate μ.

Let $f_\theta = N(\mu, \sigma^2)$ $\quad\quad \theta = (\mu, \sigma^2)$

i) $T(X_1, \dots, X_n) = X_1$ ✗

$E(T(X_1, \dots, X_n)) = E(X_1) = \mu$

$\Rightarrow$ T is an unbiased estimator.

ii) $T(x_1, \ldots, x_n) = \dfrac{7X_3 + 8X_9}{15}$

$E(T(x_1, \ldots, x_n)) = E\left(\dfrac{7X_3 + 8X_9}{15}\right) = \dfrac{7}{15}E(X_3) + \dfrac{8}{15}E(X_9)$

$= \dfrac{7}{15}\mu + \dfrac{8}{15}\mu = \mu$

iii) $T(x_1, \ldots, x_n) = \dfrac{x_1 + \cdots + x_n}{n} = \bar{x}$

$E(T(x_1, \ldots, x_n)) = E(\bar{x}) = \mu$

iv) $T(X_1, \ldots, X_n) = 3X_1 \Rightarrow E(T(x_1, \ldots, x_n)) = 3\mu \neq \mu$ unless $\mu = 0$

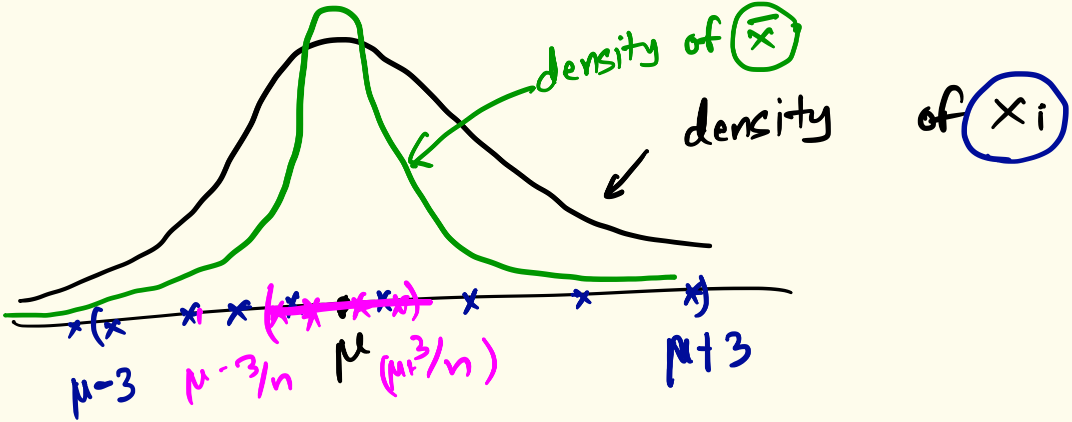
Ex: Let $X_1, X_2, \ldots, X_n$ be a random sample from $N(\mu, \sigma^2)$

Then we have already seen that $\bar{X}$ is an unbiased estimator of μ.

Here, we know

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

Consider the case: when μ is unknown and σ^2 is known.



$\bar{x} \sim N(\mu, \frac{\sigma^2}{n})$

Suppose $\sigma^2 = 1$

$\bar{x} \sim N(\mu, \frac{1}{n})$

We are going to study 3 important examples:

$x_1, x_2, \ldots, x_n$ is a random sample from $N(\mu, \sigma^2)$

i) Estimate μ when σ^2 : known

ii) Estimate μ when σ^2 : unknown ← t-distribution

iii) Estimate σ^2 ← chi-squared