Statistics.

Statistical inference. (Eestimation & Testing of Hypothesis).

Let $(X_1 X_2 \cdots X_n)$ be random samples from. $f_\theta(\ )$ or $F_\theta(\ )$. where f is the p.d.f. and. F is the c.d.f. of X. Here random sample. $X_1 X_2 \cdots X_n$ are iid $f_\theta(\ )$. i.e. joint pdf of $(X_1 \cdots X_n)$ at $(x_1, x_2 \cdots, x_n)$. is

$$f(x_1, x_2, \cdots x_n) = \prod_{i=1}^{n} f_\theta(x_i)$$

The family of distributions is defined as.

$$\mathcal{F} = \left\{ f_\theta(\ ) \mid f_\theta(\ ) \text{ is the pdf of } X \text{ and. } \theta \in \mathcal{X} \right\}$$

$$\text{or } = \left\{ F_\theta(\ ) \mid F_\theta(\ ) \text{ is the cdf of } X \text{ and } \theta \in \mathcal{X} \right\}.$$

e.g.. $\mathcal{B} = \left\{ \binom{n}{x} p^{x} (1-p)^{n-x} \mid n \in \mathbb{N},\ p \in [0,1] \right\}$

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

$$\mathcal{N} = \left\{ \frac{e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}}{\sqrt{2\pi}\,\sigma} \;\middle|\; (\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}^+ = \mathcal{X} \right\}$$

In parametric estimation problem we want (2)
to identify the parameter values of the
distribution from given data set.

→ In non-parametric estimation problem we want
to identify p.d.f or c.d.f directly from
the data without any parametric - setup.

Point estimation in parametric setup.

(1) Dfn of estimation.
(ii) Good properties of an estimator.
(iii) Methods of estimation. (method of Moment; MLE)

Statistic: A statistic is a function of data but it is free from any unknown parameter.

Let $x_1 \ldots x_n$ be random sample from $N(\mu, \sigma^2)$.

(1) μ known: $\Sigma(x_i - \mu)^2$ is a statistic.

(2). μ unknown: (a) $\Sigma(x_i - \mu)^2$ is not a statistic.

(b) $\Sigma(x_i - \bar{x})^2$ is a statistic.

(3) μ, σ^2 unknown. $\dfrac{\Sigma(x_i - \bar{x})^2}{\sigma^2}$ is not a statistic.

$X \sim N(0,1)$

$X^2 \sim \chi^2_1 \equiv G(1/2, 1/2)$.

$\sum_{i=1}^{n} (x_i-\mu)^2 \sim \sigma^2 \chi^2_n = \sigma^2 G(\frac{n}{2}, 1/2)$.

$\sum_{i=1}^{n} (x_i-\bar{x})^2 \sim \sigma^2 \chi^2_{n-1} = \sigma^2 G(\frac{n-1}{2}, 1/2)$.

$\frac{1}{\sigma^2}\sum_{i=1}^{n} (x_i-\bar{x})^2 \sim \chi^2_{n-1} = G(\frac{n-1}{2}, 1/2)$.

Let $x_1 \cdots x_n \overset{iid}{\sim} f_{\underset{\sim}{\theta}} \in \mathcal{F} = \{ f_{\underset{\sim}{\theta}} \mid \underset{\sim}{\theta} \in \mathcal{H} \}$

We may be intereted to estimat a function of $\underset{\sim}{\theta}$. say $g(\underset{\sim}{\theta})$. from data..

$g(\mu, \sigma^2) = \mu$ for $N(\mu, \sigma^2)$.

$g(\mu, \sigma^2) = \sigma^2$ for $N(\mu, \sigma^2)$

$g(\mu, \sigma^2) = \frac{\sigma^2}{\mu}$. if . $\mu \neq 0$. [coefficient of variation.]

$g(\alpha, \lambda) = \frac{\alpha}{\lambda}$ for $G(\alpha, \lambda)$

$g(\alpha, \lambda) = \frac{\alpha}{\lambda^2}$ for $G(\alpha, \lambda)$.

$g(\alpha, \lambda) = \lambda$. for $G(\alpha, \lambda)$.

Estimator: A statistic $T(\underset{\sim}{X})$ when used to estimate a parametric function $g(\underset{\sim}{\theta})$, is called an estimator of $g(\underset{\sim}{\theta})$.

As $T(\underset{\sim}{X})$ is a function of random variables $\underset{\sim}{X} = (x_1 \cdots \cdot x_n)$, it is also a random variable.

For a given set of data.

$(X_1 = \overset{1.5}{x_1},\ X_2 = \overset{2.3}{x_2} \cdots \quad X_n = \overset{1.7}{x_n})$.

the value of $T(\underset{\sim}{x})$ is known as an estimate of $g(\underset{\sim}{\theta})$.

$$g(\underset{\sim}{\theta}) \hat{=} T(\underset{\sim}{X}) \quad \text{or} \quad \widehat{g(\theta)} = T(\underset{\sim}{X}).$$

$$\mu \hat{=} \overline{X} \quad \text{or} \quad \hat{\mu} = \overline{X}.$$

$$\sigma^2 \hat{=} \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 \qquad \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

Eg. 1 Let $X_1 \cdots \cdots X_n$ be iid random variables. with $E(X_i) = \mu$, $Var(X_i) = \sigma^2$

μ can be estimate by (I) $T_1(\underset{\sim}{X}) = X_5$

(II) $T_2 = (5X_1 + 3X_2)/8$ ~~(III)~~

(III) $T_3 = \sum_{i=1}^{n} a_i x_i$ such that $\sum a_i = 1$

(IV) $T_4 = \frac{1}{n}\sum x_i = \overline{X}$

eg. σ^2 can be estimated by

$$S_1^2(\underset{\sim}{X}) = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \frac{1}{2n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}(X_i - X_j)^2$$

$$S_2^2(\underset{\sim}{X}) = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2.$$

HW.

$$\boxed{E\left(S_2^2(\underset{\sim}{X})\right) = \sigma^2.}$$

In particular. if

$$\sum(X_i - \bar{X})^2 \sim \sigma^2 \chi^2_{n-1}.$$

then $E\left(\sum_{i=1}^{n}(X_i - \bar{X})^2\right) = \sigma^2(n-1).$

$$\Rightarrow E\left(\frac{1}{n-1}\sum(X_i - \bar{x})^2\right) = \sigma^2.$$

Properties of an estimator.

Unbiased Estimator: Let $T(\underset{\sim}{X})$ be an estimator of $g(\underset{\sim}{\theta})$, which satisfies.

$$E\left(T(\underset{\sim}{X})\right) = g(\underset{\sim}{\theta}) \quad \forall \underset{\sim}{\theta} \in \textcircled{H}$$

then. $T(\underset{\sim}{X})$ is said to be <u>an</u> unbiased estimator of $g(\underset{\sim}{\theta})$.

Let $X_1 \sim N(0,1)$.

$$T(\underset{\sim}{x}) = 3X_1$$

is $T(\underset{\sim}{x})$ an unbiased estimator of mean?

Not an unbiased estimator.

Bias of an estimator is defined as.

$$B_{g(\theta)}(T(\underset{\sim}{x})) = E\left(T(\underset{\sim}{x}) - g(\underset{\sim}{\theta})\right)$$
$$= E\left(T(\underset{\sim}{x})\right) - g(\underset{\sim}{\theta}).$$

<u>Ex1.</u> Show that $T_1\ T_2 \dots T_4$ all are unbiased estimators of μ in Example 1.

<u>Ex2</u> Show that $S_2^2(\underset{\sim}{X})$ is an unbiased estimator of σ^2 but $S_1^2(\underset{\sim}{X})$ is not an unbiased estimator of σ^2 in <u>Example 2</u>.

<u>Asymptotically unbiased estimator.</u>

If the bias of an estimator $T(\underset{\sim}{x})$ goes to zero when samplesize $n \uparrow \infty$ then $T(\underset{\sim}{x})$ is call an asymptotically unbiased estimator.

Ex 3: $S_1^2(\underset{\sim}{X})$ is an asymptotically unbiased estimator of σ^2 in example 2.

Mean squared error: (MSE).

If $g(\underset{\sim}{\theta})$ is a parametric function and it is estimated by $T(\underset{\sim}{X})$, then the mean squared error of $T(\underset{\sim}{X})$ for $g(\underset{\sim}{\theta})$ is defined as.

$$MSE\left(T(\underset{\sim}{X})\right) = E\left(T(\underset{\sim}{X}) - g(\underset{\sim}{\theta})\right)^2.$$

$$MSE\left(T(\underset{\sim}{X})\right) = Var\left(T(\underset{\sim}{X})\right) + \left(Bias\, T(\underset{\sim}{X})\right)^2$$

If $T(\underset{\sim}{X})$ is an unbiased estimator of $g(\underset{\sim}{\theta})$.

then $Bias\left(T(\underset{\sim}{X})\right) = 0$.

$$\Rightarrow MSE\left(T(\underset{\sim}{X})\right) = var\left(T(\underset{\sim}{X})\right).$$

o/w. $MSE\left(T(\underset{\sim}{X})\right) > Var\left(T(\underset{\sim}{X})\right)$.

$$E\left(T(\underset{\sim}{X}) - g(\underset{\sim}{\theta})\right)^2 = E\left\{\left(T(\underset{\sim}{X}) - E\,T(\underset{\sim}{X})\right) + \left(E\,T(\underset{\sim}{X}) - g(\underset{\sim}{\theta})\right)\right\}^2$$

$$= E\left(T(\underset{\sim}{X}) - E\,T(\underset{\sim}{X})\right)^2 + \left(E\,T(\underset{\sim}{X}) - g(\underset{\sim}{\theta})\right)^2 + 0.$$

$$= var\left(T(\underset{\sim}{X})\right) + \left(Bias\, T(\underset{\sim}{X})\right)^2.$$

consistent estimator.

An estimator $T_n(\underline{x}) = T(x_1 \cdots x_n)$ of $g(\theta)$ is said to be a consistent estimator if

$$\lim_{n \uparrow \infty} P\left(|T_n(\underline{x}) - g(\theta)| > \epsilon\right) = 0.$$

or $$\lim_{n \uparrow \infty} P\left(|T_n(\underline{x}) - g(\theta)| < \epsilon\right) = 1.$$

Ex. 4. If the MSE of $T_n(\underline{x})$ for $g(\theta)$ is going to zero as $n \uparrow \infty$. then $T_n(\underline{x})$ is a consistent estimator of $g(\theta)$.

$$T_3 = \sum_{i=1}^{n} a_i x_i \qquad \sum_{i=1}^{n} a_i = \mu.$$

For what values of a_i T_3 has minimum variance.

$$\begin{aligned} Var(T_3) &= Var\left(\sum_{i=1}^{n} a_i x_i\right) \\ &= \sum_{i=1}^{n} a_i^2 \, var(x_i) \\ &= \sigma^2 \sum_{i=1}^{n} a_i^2 \end{aligned}$$

$$Var(T_3) = \sigma^2 \sum_{i=1}^{n} a_i^2$$

$$= \frac{\sigma^2}{n} \left(\sum_{i=1}^{n} a_i^2\right)(n)$$

$$= \frac{\sigma^2}{n} \left(\sum_{i=1}^{n} a_i^2\right)\left(\sum_{i=1}^{n} 1\right)$$

C-S. $\sum a_i^2 \sum b_i^2 \geq \left(\sum a_i b_i\right)^2$

'=' holds. $a_i \propto b_i$

$\Rightarrow a_i = k b_i$

$$\geq \frac{\sigma^2}{n} \left(\sum_{i=1}^{n} a_i . 1\right)^2$$

$\sum a_i = 1$

$$= \frac{\sigma^2}{n}$$

'=' holds if $a_i \propto 1$.

$\Rightarrow a_i = k$.

$\Rightarrow \sum a_i = nk$.

$\Rightarrow 1 = n.k$.

$\Rightarrow k = 1/n$.

minimum variance unbiased estimator of μ.

is $\frac{1}{n} \sum_{i=1}^{n} x_i = \overline{X}$.