Estimation

Parametic family of distributions:

Let X be a random variable denoting some measurement of a population.

$X \sim F_\theta$. F is distribution function.

θ is the parameter of the distribution.

$f_\theta(x)$ pdf (continuous) OR $f_\theta(x)$ p.m.f (discrete).

F ~~or~~ and pdf/pmf changes as θ changes.

Parametric inference

→ Assumption: ① Data given.
② Distribution of X is assumed to be known.

→ problem : Identify the value of parameter(s) given the ① data and ② distribution.

$$\mathcal{N} = \left\{ N(\mu, \sigma^2) \;\middle|\; \underset{\sim}{\theta} = (\mu, \sigma^2) \in \mathcal{H} = \mathbb{R} \times \mathbb{R}^+ \right\} \to f_{\underset{\sim}{\theta}}(x) = \frac{e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}}{\sqrt{2\pi}\,\sigma} \quad x \in \mathbb{R}$$

$$\mathbb{P} = \left\{ \text{pois}(\lambda) \;\middle|\; \theta = \lambda \in \mathcal{H} = \mathbb{R}^+ \right\} \to f_\lambda(x) = \frac{e^{-\lambda} \lambda^x}{x!} \quad x \in \mathbb{N} \cup \{0\}$$

$X_1 X_2 \cdots X_n \overset{iid}{\sim} N(\mu, \sigma^2)$.

$T_1(\underset{\sim}{X}) = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}$ $\quad E(\bar{X}) = \mu$.

$\Rightarrow \bar{X}$ is an unbiased estimator of μ.

$$\begin{cases} \hat{\mu} = \bar{x} \\ \hat{\sigma}^2 = s_1^2 \end{cases}$$

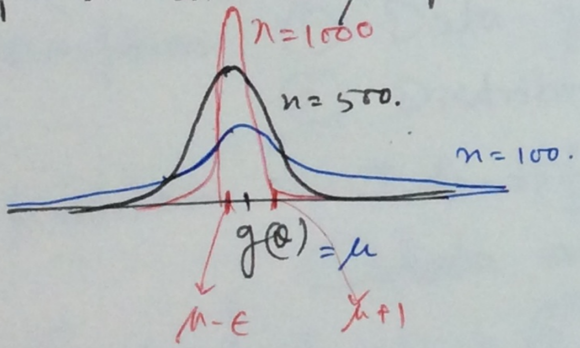
$s_1 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$ $\quad E(S_1) = \frac{\sigma^2 (n-1)}{n} = \left(1 - \frac{1}{n}\right)\sigma^2 \neq \sigma^2$.

S_1 is not an unbiased estimator of σ^2.

$E(S_1) \longrightarrow \sigma^2$ as $n \uparrow \infty$.

S_1 is an asymptotically unbiased estimator

$f(y) = \frac{\lambda^{\alpha} e^{-\lambda x} x^{\alpha - 1}}{\Gamma \alpha}$



Method of Moments.

$Y \sim G(\alpha, \lambda)$ $\quad \underset{\sim}{\theta} = (\alpha, \lambda)$

$\alpha > 0$

$\lambda > 0$.

$\alpha = g(\underset{\sim}{\theta}) = g(\alpha, \lambda)$

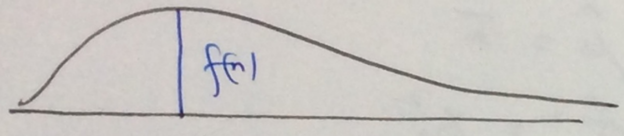
$E(Y) = \alpha/\lambda = \frac{1}{n}\sum_{i=1}^{n} y_i = \bar{x}$ known. from data.

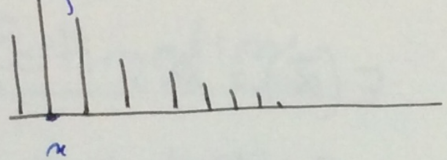
$V(Y) = \alpha/\lambda^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2 = s_1^2$

theoritical moments. $\quad$ sample moments.

$\hat{\alpha} = \frac{\bar{x}^2}{s_1^2}$

or $\hat{\lambda} = \frac{\bar{x}}{s_1^2}$



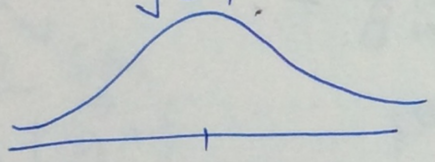


mode.

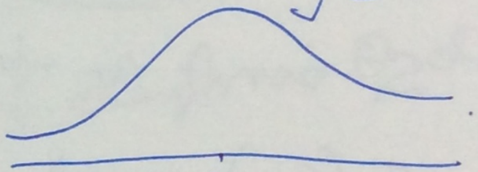
given $\underset{\sim}{x} = (x_1, \dots, x_n)$ is considered to be the mode of joint density $\prod_{i=1}^{n} f_\theta(x_i)$

$$f(x) = \frac{e^{-\frac{1}{2}(x-\mu)^2}}{\sqrt{2\pi}}.$$

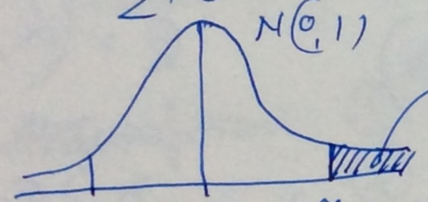
$$L(\mu \mid x = 5.45) = \frac{e^{-\frac{1}{2}(5.45-\mu)^2}}{\sqrt{2\pi}}.$$



$\longrightarrow x.$



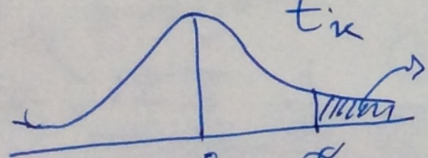
$\longrightarrow \mu$



$P(Z > Z_\alpha) = \alpha.$

$P(t_k > t_{\alpha,k}) = \alpha.$

upper α-values.



$|Z_\alpha| = |Z_{1-\alpha}|$

MLE.

$x_1\, x_2 \cdots x_n \sim \text{bin}(10, p)$

$p \in (0,1)$

$x_i \in \{0,1,2, \cdots 10\}$

Joint density is $\prod_{i=1}^{n} f(x_i) = \prod_{i=1}^{n} \binom{10}{x_i} p^{x_i} (1-p)^{10-x_i}$

likelihood function. $l(p) = \prod_{i=1}^{n} \binom{10}{x_i} p^{x_i} (1-p)^{10-x_i}$

$$= \left(\prod_{i=1}^{n} \binom{10}{x_i} \right) p^{\sum_{i=1}^{n} x_i} (1-p)^{n10 - \sum_{i=1}^{n} x_i}$$

$$L(p) = \log_e l(p) = \log_e \left(\prod_{i=1}^{n} \binom{10}{x_i} \right) + \sum_{i=1}^{n} x_i \log_e p + \left(n10 - \sum_{i=1}^{n} x_i \right) \log_e (1-p)$$

$$\frac{\partial L(p)}{\partial p} = 0 \qquad \hat{p}_{mle} = \frac{\sum_{i=1}^{n} x_i}{10n} \qquad E(\hat{p}) = p. \qquad V(\hat{p}) = \frac{n10\, p(1-p)}{(n10)^2}$$

$H_0: p = 0.5$

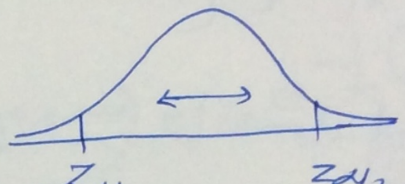
$H_1: p \neq 0.5$

$$\frac{\sqrt{10n}\,(\hat{p} - p)}{\sqrt{\hat{p}(1-\hat{p})}} \overset{a}{\sim} N(0,1). \quad \text{as } n \uparrow \infty.$$

Interval estimation. CI of μ.

$$x_i \overset{iid}{\sim} N(\mu, \sigma^2) \qquad \bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right).$$

$$\Rightarrow Z = \frac{\sqrt{n}(\bar{X}-\mu)}{\sigma} \sim N(0,1).$$



$$\Rightarrow P\left(-Z_{\alpha/2} < Z < Z_{\alpha/2}\right) = 1-\alpha.$$

$$\Rightarrow P\left(-Z_{\alpha/2} < \frac{\sqrt{n}(\bar{X}-\mu)}{\sigma} < Z_{\alpha/2}\right) = 1-\alpha.$$

$$\Rightarrow P\left(\bar{X} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = 1-\alpha.$$

$$\Rightarrow \begin{cases} L(\underline{x}) = \bar{X} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}. \\ U(\underline{x}) = \bar{X} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}. \end{cases}$$

When σ^2 known.

Length $2 Z_{\alpha/2}\, \sigma/\sqrt{n} \longrightarrow 0$ as $n \to \infty$

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

$$T = \frac{\sqrt{n}(\bar{X}-\mu)}{\hat{\sigma}} \sim t_{n-1}$$

When σ^2 unknown CI of μ

$$L(\underline{x}) = \bar{X} - t_{\alpha/2}\frac{\hat{\sigma}}{\sqrt{n}}$$

$$U(\underline{x}) = \bar{X} + t_{\alpha/2}\frac{\hat{\sigma}}{\sqrt{n}}.$$

Length. $2\, t_{\alpha/2}\, \hat{\sigma}/\sqrt{n}$.
$\longrightarrow 0$ as $n \to \infty$

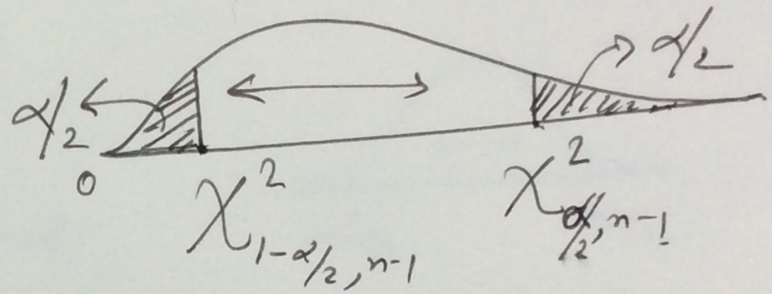
95% CI of σ^2 $\quad X_i \overset{iid}{\sim} N(\mu, \sigma^2)$.

$\alpha = 0.05$

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

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

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



$$\Rightarrow P\left(\chi^2_{1-\alpha/2, n-1} < \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{\sigma^2} < \chi^2_{\alpha/2, n-1} \right) = 1 - \alpha$$

$$\Rightarrow P\left(\underbrace{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{\chi^2_{\alpha/2, n-1}}}_{L(\underset{\sim}{x})} < \sigma^2 < \underbrace{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{\chi^2_{1-\alpha/2, n-1}}}_{U(\underset{\sim}{x})} \right) = 1 - \alpha.$$

Testing of Hypothesis.

We are interested to infer about some variable of a population.

→ Random variables of a population is specified by its distribution.

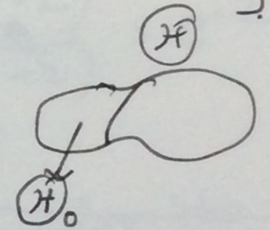
→ A distribution is uniquely identified by its cdf or pdf/pmf.

→ cdf ~~or~~ and pdf/pmf can be regulated with parameters.

→ Collection of all parameters ~~s~~ for a particular distribution is known as parameter space ~~Θ~~ Θ.

e.g. $\mathcal{N} = \{ N(\mu, \sigma^2) \mid \mu \in \mathbb{R}, \sigma^2 > 0, \Theta = \mathbb{R} \times \mathbb{R}^+ \}$

→ Null hypothesis (H_0) specifies a subset ~~is~~ $\Theta_0 \subset \Theta$



If $|\Theta_0| = 1$ then H_0 is called a simple null

if $|\Theta_0| > 1$ then H_0 is called a composit null.

~~Ho1.~~ Assuming $X \sim N(\mu, \sigma^2)$ $H_0: \mu = 1$ then it is a composit null as $H_0 \longrightarrow \Theta_0 = 1 \times \mathbb{R}^+$

Assuming $X \sim N(\mu, 1)$ $H_0: \mu = 1$ then it is a simple null. as $H_0 \longrightarrow \Theta_0 = \{(1, 1)\}$.

Alternative hypothesis is a subset of $\mathcal{H}$ which is denoted by $\mathcal{H}_1$ or $\mathcal{H}_A$ satisfying

$$\mathcal{H}_A \subset \mathcal{H} \quad \text{and} \quad \mathcal{H}_A \cap \mathcal{H}_0 = \varphi.$$

<u>need not be</u> $\mathcal{H}_A \cup \mathcal{H}_0 = \mathcal{H}$.

With a statistical test (a) we don't reject the null Hypothesis.

or (b) we reject the null hypothesis in favour of H_1

$H_0: \mu = 0$ vs $H_1: \mu \neq 0$. $X_i \overset{iid}{\sim} N(\mu, 1)$.

$H_0: \mu = 0$ vs $H_1: \mu = 0.25$ $X_i \overset{iid}{\sim} (\mu, 1)$

$H_0: \mu = 0$ vs $H_1: \mu = 3.75$ $X_i \overset{iid}{\sim} (\mu, 1)$

$H_0: \mu = 25$ vs $H_1: \mu \neq 25$ $x_i \overset{iid}{\sim} N(\mu, \sigma^2)^2$ $i = 1, 2, \ldots 20$.

1 $\hat{\mu} = \bar{x} = 25.51$

2 $\hat{\sigma} = s = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 = 2.1933.$

3 $t = \dfrac{\sqrt{n}(\bar{x} - \mu_0)}{s} = 1.04.$ <u>(observed)</u>.

$t \sim t_{19}$ under <u>H_0</u>

S5 $\alpha = 0.05$. $\tau_{\alpha/2, 19} = 2.093$.

S6 $t_{obs} = 1.04 < \tau_{\alpha/2, 19}$

Given the data we cannot reject H_0 in favour of H_1 at level <u>$\alpha = 0.05$</u>

Ex 163 $X_1 X_2 \cdots X_n \overset{iid}{\sim} N(\mu_1, \sigma^2)$ \
$Y_1 Y_2 \cdots Y_m \overset{iid}{\sim} N(\mu_2, \sigma^2)$ $\Big\rangle$ indept.

$H_0: \mu_1 = \mu_2$ \
$H_1: \mu_1 \neq \mu_2$

Let
$$\hat{\sigma}^2 = \left(\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2 + \sum_{i=1}^{m} (y_i - \bar{y})^2}{n+m-2} \right)$$

$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$ \
$\bar{y} = \frac{1}{m} \sum_{i=1}^{n} y_i$

unbiased estimator of σ^2

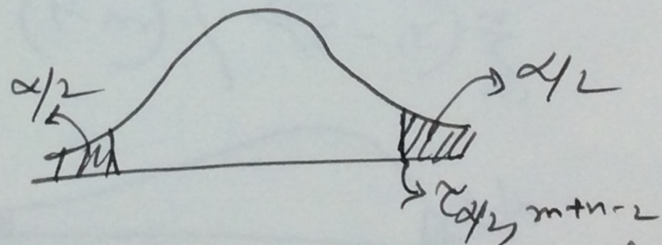
$$\sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{\sigma^2} \sim \chi^2_{n-1}$$

$$\sum_{i=1}^{m} \frac{(y_i - \bar{y})^2}{\sigma^2} \sim \chi^2_{m-1}$$

$$\frac{\sum (x_i - \bar{x})^2 + \sum (y_i - \bar{y})^2}{\sigma^2} \sim \chi^2_{m+n-2}$$

$$E\left(\frac{\sum_1^n (x_i - \bar{x})^2 + \sum_1^m (y_i - \bar{y})^2}{m+n-2} \right) = \frac{\sigma^2 (m+n-2)}{(m+n-2)} = \sigma^2 .$$

$$T = \frac{\bar{x} - \bar{y}}{\sqrt{\hat{\sigma}^2}} \sim t_{n+m-2} \text{ under } H_0$$



If $|T_{(obs)}| > t_{\alpha/2, m+n}$ reject H_0 in favour of H_1.

Ex 169 (X_i, Y_i) $i = 1, 2, \cdots n$.

paired T-test

$d_i = X_i - Y_i$ for $i = 1, 2, \cdots n$.

$E(d_i) = 0$ under H_0.

$V(d_i) = \sigma_d^2$ unknown.

Ex 166.

$$\left.\begin{array}{l} \dfrac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{\sigma^2} \sim \chi^2_{n-1} \\[2ex] \dfrac{\sum_{i=1}^{m} (Y_i - \bar{Y})^2}{\sigma^2} \sim \chi^2_{m-1} \end{array}\right\} \text{ind.}$$

$$F = \frac{\sum_{i=1}^{n} (X_i - \bar{X})^2 / (n-1)}{\sum_{i=1}^{m} (Y_i - \bar{Y})^2 / (m-1)} \sim F_{n-1,\, m-1} \text{ under } H_0.$$