- Maximum Likelihood.

- Example. $Y_1 \cdots Y_n \overset{iid}{\sim}$ Poisson (λ)

$$L(\lambda) = \prod_{i=1}^{n} \frac{1}{y_i!} e^{-\lambda} \lambda^{y_i}$$

log-likelihood: $l(\lambda) = \sum_{i=1}^{n} (-\lambda + Y_i \log\lambda - \log Y_i!)$

$$= -n\lambda + \left(\sum_{i=1}^{n} Y_i\right)\log\lambda - \sum_{i=1}^{n}\log Y_i!$$

$l'(\lambda) = 0 \quad \Rightarrow \quad \hat{\lambda} = \bar{Y}_n \qquad$ (check $l''(\hat{\lambda}) < 0$)

- Set $\quad X_i = \begin{cases} 0 & \text{if } Y_i = 0 \\ 1 & \text{if } Y_i > 0 \end{cases}$

Goal is to use X_i's to estimate λ.

$$P(X_i = 0) = P(Y_i = 0) = e^{-\lambda}$$

$$P(X_i = 1) = 1 - e^{-\lambda} = p \qquad \hat{p} = \bar{X}_n = 1 - e^{-\lambda} \Rightarrow \hat{\lambda}$$

$$L(\lambda) = \prod_{i=1}^{n} \{e^{-\lambda}\}^{I(X_i=0)} \{1 - e^{-\lambda}\}^{I(X_i=1)}$$

$$l(\lambda) = \sum_{i=1}^{n} \{I(X_i=0)\log(e^{-\lambda}) + I(X_i=1)\log(1-e^{-\lambda})\}$$

$$= \sum_{i=1}^{n} \{(1-X_i)\log(e^{-\lambda}) + X_i\log(1-e^{-\lambda})\}$$

$$= \left(n - \sum_{i=1}^{n} X_i\right)(-\lambda) + \left(\sum_{i=1}^{n} X_i\right)\log(1-e^{-\lambda})$$

$$l'(\lambda) = -\left(n - \sum_{i=1}^{n} X_i\right) + \sum_{i=1}^{n} X_i \frac{e^{-\lambda}}{1-e^{-\lambda}} = 0$$

$$\frac{e^{-\lambda}}{1-e^{-\lambda}} = \frac{n - \sum_{i=1}^{n} X_i}{\sum_{i=1}^{n} X_i}$$

$$\Rightarrow \hat{\lambda} = \log\frac{n}{n - \sum_{i=1}^{n} X_i} = \underline{-\log(1 - \bar{X}_n)}$$

when n is finite. it is possible to observe $X_1 = X_2 = X_3 = \cdots = X_n = 1$

Then $\hat{\lambda} = -\log(1-1) = -\log 0$???

How likely? $P(X_1 = \cdots = X_n = 1)$

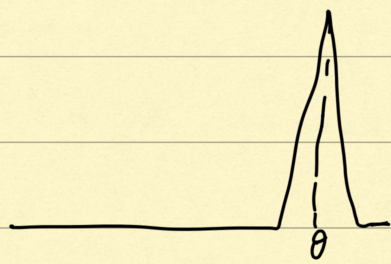
$$= (1-e^{-\lambda})^n \to 0 \quad \text{as } n \to +\infty$$

$\hookrightarrow 1$ as $\lambda \to +\infty$ when n is fixed.

(Asymptotically, in the sense of $n \to +\infty$, MLE works as expected

but in finite sample case, MLE may not exist)

- A homework problem. (HW 1) $\quad X_1, \dots, X_n \overset{iid}{\sim} \begin{cases} \text{Uniform}(-1,1) & \text{with prob } \theta \\ \frac{1}{C(\theta)}\left[1 - \frac{|x-\theta|}{C(\theta)}\right] & 1-\theta \end{cases}$



In this example. MLE $\hat{\theta}$ always exists.

but no matter what the true value of θ, $\hat{\theta} \to 1$ as $n \to +\infty$

Θ . MLE: $\hat{\theta}_n \to X_1, \dots, X_n \overset{iid}{\sim} \{P_\theta, \theta \in \Omega\}$

↳ univariate

$\boxed{\text{Scale}}$

Univariate

For the existence of a consistent $\hat{\theta}_n$ we make the following assumptions

(C1) Identifiability: if $P_{\theta_1} = P_{\theta_2}$, then $\theta_1 = \theta_2$

(C2) The parameter space Ω is an open interval $(\underline{\theta}, \bar{\theta})$

$$\Omega: \ -\infty \leq \underline{\theta} < \theta < \bar{\theta} \leq +\infty$$

(C3) iid sample from P_θ

(C4) The $\underbrace{\text{set}}_{}$ (support) $A: \{x : f_\theta(x) > 0\}$ (if continuous, density) or $A = \{x : P_\theta(x) > 0\}$ (if discrete, p.m.f)

is independent of θ

(C5) for all x in A, $f_\theta(x)$ or $P_\theta(x)$ is differentiable with respect to θ

Remark. (C1): $Y_1, \cdots, Y_n \overset{iid}{\sim} N(\mu, 1)$. but we observed $X_i = |Y_i|$

Then μ is not identifiable, but $|\mu|$ is!

Theorem. Let $X_1, \cdots, X_n$ be iid fom a distribution satisfying (C1)-(C5).
Then, there exists a sequence $\hat{\theta}_n = \hat{\theta}_n(X_1, \cdots, X_n)$ of local maxima of the likelihood function $L_n(\theta)$ which is consistent; $\hat{\theta}_n \xrightarrow{P} \theta$ for all $\theta \in \Omega$

$l_n(\theta) = \log L_n(\theta)$

- local maxima is a root of $l_n'(\theta) = 0$

Corollary: Under the same assumptions, if the likelihood equation. $l_n'(\theta) = 0$ has a unique root $\hat{\theta}_n$ for each n. and all $(x_1, \cdots, x_n)$.
then (i) $\hat{\theta}_n$ is a consistent estimator of θ
(ii) with probability tending to 1 as $n \to +\infty$, $\hat{\theta}_n$ is the MLE

Example (Normal) $X_1, \cdots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$ $\quad l_n(\mu, \sigma^2) = -n \log\sqrt{2\pi} - n\log\sigma - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(X_i - \mu)^2$

(i) σ^2 is known. esti μ. $(\theta = \mu)$

l_n reduces to $-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(X_i - \mu)^2$

Likelihood equation $l_n'(\theta) = 0$ reduces to $\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \mu) = 0$

which only has one root $\hat{\mu} = \bar{X}_n$

(ii) μ is known. esti σ^2 $(\theta = \sigma^2)$

Likelihood equation is $-\frac{n}{\sigma} + \frac{1}{\sigma^3}\sum_{i=1}^{n}(X_i - \mu)^2 = 0$

which also has a unique root $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2$

(iii) suppose $\frac{\mu}{\sigma}$ is known. say $\mu = \overset{\text{known}}{a}\,\sigma$, esti σ.

$l_n(\theta) = l_n(a\sigma, \sigma^2)$ reduces to $-n\log(a\sigma) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(X_i - a\sigma)^2$

take derivative to get the likelihood

$$-n\frac{a}{a\sigma} - \frac{1}{2\sigma^2}\left(2\sum_{i=1}^{n}(x_i-a\sigma)\times(-a)\right) - \frac{1}{2}\times(-2)\,\sigma^{-3}\sum_{i=1}^{n}(x_i-a\sigma)^2 = 0$$

$$-\frac{n}{\sigma} + \frac{a\sum_{i=1}^{n}(x_i-a\sigma)}{\sigma^2} + \frac{\sum_{i=1}^{n}(x_i-a\sigma)^2}{\sigma^3} = 0$$

$$-n + \frac{a\sum_{i=1}^{n}x_i - na^2\sigma}{\sigma} + \frac{\sum_{i=1}^{n}(x_i^2-2a\sigma x_i+a^2\sigma^2)}{\sigma^2} = 0$$

$$\frac{\sum_{i=1}^{n}x_i^2}{\sigma^2} - \frac{a\sum_{i=1}^{n}x_i}{\sigma} = n$$

$$\frac{1}{\sigma^2} - \frac{1}{\sigma}\times\frac{a\sum_{i=1}^{n}x_i}{\sum_{i=1}^{n}x_i^2} - \frac{n}{\sum_{i=1}^{n}x_i^2} = 0 \quad \text{has roots}$$

$$\frac{1}{\sigma} = \frac{1+\sqrt{\left(\frac{a\sum x_i}{\sum x_i^2}\right)^2 + \frac{4n}{\sum_{i=1}^{n}x_i^2}}}{2}$$

Asymptotic Normality. under ??? assumptions, we know

$$\sqrt{n}(\theta_n-\theta) \xrightarrow{d} N\left(0, \frac{1}{I(\theta)}\right)$$

$$Var(\hat{\theta}_n) = \frac{1}{nI(\theta)}$$

Theorem (i) Suppose (C1-C5) hold

(C6). the derivative with respect to θ of the left side of the equation

$$\int f_\theta(x)\,dx = 1 \qquad \left(\frac{\partial}{\partial\theta}\int f_\theta(x)\,dx = \int\frac{\partial f_\theta(x)}{\partial\theta}\,dx\right)$$

can be obtained by differentiating under the integral sign.

Then $E\left[\frac{\partial}{\partial\theta}\log f_\theta(X)\right] = 0$ $\qquad \int\frac{f_\theta'(x)}{f_\theta(x)}f_\theta(x)\,dx$

$\Rightarrow$ $I(\theta) = Var_\theta\left[\frac{\partial}{\partial\theta}\log f_\theta(X)\right]$

(ii) Suppose that also

(C6') The first two derivatives with respect to θ of $f_\theta(x)$ exist for all $x\in A$ and all θ, and the corresponding derivative with respect to θ of the left side of $\int f_\theta(x)\,dx = 1$ can be obtained by differentiation under the integral sign.

Then also $\quad \underbrace{I(\theta)} = -E_\theta\left[\frac{\partial^2}{\partial\theta^2}\log f_\theta(X)\right]$

Maximum information about θ contained by X.

Aditivity: ① $X_1, \cdots, X_n$. Maximum information about θ contained by $X_1, \cdots, X_n$ is $nI(\theta)$

② If $X \sim f_\theta$, $Y \sim g_\theta$, $\underline{X \perp Y}$ (independent) $\to h_\theta = f_\theta g_\theta$

$$\begin{array}{cc} X & Y \\ \downarrow & \downarrow \\ I_1(\theta) & I_2(\theta) \end{array}$$

$\{X, Y\} \Rightarrow \underbrace{I(\theta)} = I_1(\theta) + I_2(\theta)$

$$\begin{aligned} I(\theta) &= -E_\theta\left[\frac{\partial^2}{\partial\theta^2}\log(h_\theta(X,Y))\right] \quad \text{where } h_\theta \text{ is the joint density of } (X,Y) \\ &= -E_\theta\left[\frac{\partial^2}{\partial\theta^2}\left\{\log\left[f_\theta(X)g_\theta(Y)\right]\right\}\right] \\ &= -E_\theta\left[\frac{\partial^2}{\partial\theta}\left\{\log f_\theta(X) + \log g_\theta(Y)\right\}\right] \\ &= \underbrace{-E_\theta\left[\frac{\partial^2}{\partial\theta}\log f_\theta(X)\right]} + \left(\underbrace{-E_\theta\left[\frac{\partial^2}{\partial\theta^2}\log g_\theta(Y)\right]}\right) \\ &= I_1(\theta) + I_2(\theta) \end{aligned}$$

③ If $Y = g(X)$ Let $I^*(\theta)$ be the Fisher Info from Y

$I^*(\theta) = ?$

Suppose g is strictly increasing

$$P(Y \le y) = P(g(X) \le y) = P(X \le g^{-1}(y))$$

$$f_y(y) = f_X(g^{-1}(y)) \cdot \frac{\partial g^{-1}(y)}{\partial y} \qquad g(g^{-1}(y)) = y$$

$$f_X = f(\ |\theta) \qquad = f_X(g^{-1}(y))\frac{1}{g'(g^{-1}(y))} \qquad g'(g^{-1}(y)) \cdot \frac{\partial g^{-1}(y)}{\partial y} = 1$$

$$\log f_Y(y) = \log \underset{\downarrow \atop \theta}{f_X}(g^{-1}(y)) - \log(\underbrace{g'(g^{-1}(y))}_{\text{no } \theta})$$

$$I^*(\theta) = -E\left[\frac{\partial^2}{\partial\theta^2} \log f_Y(y)\right] = -E\left[\frac{\partial^2}{\partial\theta^2} \log f_X(X)\right] = I(\theta)$$

④ $X \to I(\theta)$

we are interested in $\eta = g(\theta)$, g is strictly increasing

$I(\eta)$?

$$f(x|\theta) = f(x|g^{-1}(\eta))$$

$$I(\eta) = Var\left[\frac{\partial}{\partial\eta} \log f(X|g^{-1}(\eta))\right]$$

$$Var\left[\frac{\partial}{\partial\eta} \log f(x|g^{-1}(\eta))\right] = Var\left[\underbrace{\frac{\partial}{\partial g^{-1}(\eta)} \log f(x| g^{-1}(\eta))}\times \frac{\partial g^{-1}(\eta)}{\partial\eta}\right]$$

$$= Var\left[\underbrace{\frac{\partial}{\partial\theta} \log f(x|\theta)} \times \underbrace{\frac{1}{g'(g^{-1}(\eta))}}_{\text{no random}}\right]$$

$$= \left[\frac{1}{g'(\theta)}\right]^2 \times I(\theta)$$

$$\boxed{I(\eta) = I(\theta) \Big/ [g'(\theta)]^2, \text{ where } \eta = g(\theta)}$$

Theorem: Let $X_1, \cdots, X_n \overset{iid}{\sim} f_\theta(x)$ satisfying (C1 − C5)

In addition, we assume

(C6)'' for all $x \in A$, the density $f_\theta(x)$ is three times differentiable with respect to θ and the third derivative is continuous. The corresponding derivative of the integral $\int f_\theta(x)dx$ can be obtained by differentiating under the integral sign.

(C7) If θ_0 denote the true value of θ, there exists a positive number $c(\theta_0)$ and a function $M_{\theta_0}(x)$ such that

$$\left| \frac{\partial^3}{\partial\theta^3} \log f_\theta(x) \right| \le M_{\theta_0}(x) \quad \text{for all } x \in A, \ |\theta-\theta_0| < c(\theta_0)$$

and $$E_\theta [M_{\theta_0}(X)] < \infty$$

Then any consistent sequence $\hat{\theta}_n = \hat{\theta}_n(X_1, \cdots, X_n)$ of roots of the likelihood equation satisfies

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N\left(0, \frac{1}{I(\theta_0)}\right)$$

provided $0 < I(\theta_0) < \infty$

Proof: Taylor Expansion

Theorem Assume (C1 − C5), (C6''), (C7) hold.

Then. ① If $\hat{\theta}_n$ is a consistent sequence of roots of the likelihood equation, the probability tends to 1 as $n \to +\infty$ that $\hat{\theta}_n$ is a local maximum of the log likelihood

② If $\hat{\theta}_{1n}$ and $\hat{\theta}_{2n}$ are two consistent sequences of roots of the likelihood equation, then

$$P(\hat{\theta}_{1n} = \hat{\theta}_{2n}) \to 1 \quad \text{as } n \to +\infty$$

- Sometimes. MLE is very hard to find numerically.

"One-Step Estimator"

we start with $\tilde{\theta}_n$ a reasonable estimator of θ. (MoM)

$\sqrt{n}(\tilde{\theta}_n - \theta)$ is bounded in probability. $= O_p(1)$

Then
$$\boxed{\delta_n = \tilde{\theta}_n - \frac{l_n'(\tilde{\theta}_n)}{l_n''(\tilde{\theta}_n)}}$$

under C1–C5, C6'', C7)

$$\sqrt{n}(\delta_n - \theta) \xrightarrow{d} N\left(0, \frac{1}{I(\theta)}\right)$$

Proof:
$$\delta_n = \tilde{\theta}_n - \frac{l_n'(\tilde{\theta}_n)}{l_n''(\tilde{\theta}_n)}$$

Let θ_0 be the true value of θ that generates the data

Taylor E.
$$l_n'(\tilde{\theta}_n) = l_n'(\theta_0) + (\tilde{\theta}_n - \theta_0) l''(\theta_0) + \frac{1}{2}(\tilde{\theta}_n - \theta_0)^2 l'''(\theta_n^*)$$

where θ_n^* is between $\tilde{\theta}_n$ and θ_0

$$\delta_n = \tilde{\theta}_n - \frac{l_n'(\theta_0) + (\tilde{\theta}_n - \theta_0) l''(\theta_0) + \frac{1}{2}(\tilde{\theta}_n - \theta_0)^2 l'''(\theta_n^*)}{l_n''(\tilde{\theta}_n)}$$

$$\sqrt{n}(\delta_n - \theta_0) = \sqrt{n}(\tilde{\theta}_n - \theta_0) - \sqrt{n}\frac{l_n'(\theta_0)}{l_n''(\tilde{\theta}_n)} - \sqrt{n}(\tilde{\theta}_n - \theta_0)\frac{l''(\theta_0)}{l_n''(\tilde{\theta}_n)} + \frac{\sqrt{n}}{2}\cdot\frac{(\tilde{\theta}_n - \theta_0)^2 l'''(\theta_n^*)}{l_n''(\tilde{\theta}_n)}$$

$$= \sqrt{n}(\tilde{\theta}_n - \theta_0)\underbrace{\left[\underbrace{1 - \frac{l''(\theta_0)}{l_n''(\tilde{\theta}_n)}}_{\xrightarrow{p} 0} + \frac{1}{2}\underbrace{\frac{(\tilde{\theta}_n - \theta_0)\cdot l'''(\theta_n^*)}{l_n''(\tilde{\theta}_n)}}_{\xrightarrow{p} 0}\right]}_{} \quad (*)$$

$$- \sqrt{n}\cdot\frac{l_n'(\theta_0)}{l_n''(\tilde{\theta}_n)}$$

① $\boxed{\sqrt{n}(\tilde{\theta}_n - \theta_0) = O_p(1)}$ $\quad \tilde{\theta}_n \xrightarrow{p} \theta_0 \quad$ $l_n''(\tilde{\theta}_n) \xrightarrow{p} l''(\theta_0)$

$$1 - \frac{l''(\theta_0)}{l_n''(\tilde{\theta}_n)} \xrightarrow{P} 1 - 1 = 0$$

$l'''(\theta_n^*)$ $\quad \tilde{\theta}_n \xrightarrow{P} \theta_0 \quad \theta_n^* \in \hat{\theta}_n$ and θ_0

$\theta_n^* \xrightarrow{P} \theta_0$

② $$\frac{1}{2} \frac{(\tilde{\theta}_n - \theta_0)\, l'''(\theta_n^*)}{l_n''(\tilde{\theta}_n)} \xrightarrow{P} 0$$

$$(*) = O_p(1) \times o_p(1) = o_p(1) \qquad (*) \xrightarrow{P} 0$$

③ $$-\sqrt{n} \frac{l_n'(\theta_0)}{l_n''(\tilde{\theta}_n)} \xrightarrow{P} N\left(0, \frac{1}{I(\theta)}\right)$$

$$\boxed{-\sqrt{n} \frac{l_n'(\theta_0)}{l_n''(\theta_0)}} \times \boxed{\frac{l_n''(\theta_0)}{l_n''(\tilde{\theta}_n)}} \xrightarrow{P} 1$$

Y_i

$$l_n'(\theta_0) = \frac{d}{d\theta}\left(\sum_{i=1}^{n} \log f_{\theta_0}(X_i)\right) = \sum_{i=1}^{n} \boxed{\frac{\frac{d}{d\theta} f_{\theta_0}(X_i)}{f_{\theta_0}(X_i)}}$$

CLT

$$E(Y_i) = \int \frac{\frac{d}{d\theta} f_{\theta_0}(X_i)}{\cancel{f_{\theta_0}(X_i)}} \cancel{f_{\theta}(X_i)}\, dX_i$$

$$= \frac{d}{d\theta} \int f_{\theta_0}(X_i)\, dX_i = \frac{d}{d\theta} 1 = 0$$

$$Var(Y_i) = E(Y_i^2)$$

$$= \int \left[\frac{\frac{d}{d\theta} f_{\theta_0}(X_i)}{f_{\theta_0}(X_i)}\right]^2 f_{\theta}(X_i)\, dX_i$$

$$= \int \frac{\left(\frac{d}{d\theta} f_{\theta_0}(x_i)\right)^2}{f_{\theta_0}(x_i)} dx_i$$

$$= I(\theta_0)$$

$$\sqrt{n}\, l_n'(\theta_0) \xrightarrow{d} N(0, I(\theta_0))$$

$$-l_n''(\theta_n) \xrightarrow{P} I(\theta_0)$$

$$-\sqrt{n} \frac{l_n'(\theta_0)}{l_n''(\theta_n)} \xrightarrow{d} N\left(0, \frac{I(\theta_0)}{I^2(\theta_0)}\right) = N\left(0, \frac{1}{I(\theta_0)}\right) \#$$

Example. logistic distribution. $X_1, \dots, X_n \overset{iid}{\sim} f_\theta(x) = \frac{e^{-(x-\theta)}}{(1+e^{-(x-\theta)})^2}$ $-\infty < x < +\infty$

$$f_\theta(\theta - x) = \frac{e^{-(\theta-x-\theta)}}{(1+e^{-(\theta-x-\theta)})^2} = \frac{e^{-(-x)}}{(1+e^{-(-x)})^2}$$

$$= \frac{e^x}{(1+e^x)^2}$$

$$f_\theta(\theta+x) = \frac{e^{-x}}{(1+e^{-x})^2} = \frac{e^{-x} \times e^{2x}}{(1+e^{-x})^2 \times e^{2x}} = \frac{e^x}{(1+e^x)^2}$$

$$f_\theta(\theta-x) = f_\theta(\theta+x)$$

$$E(x) = \theta$$

$$\theta \xleftarrow{MOM} \overline{X_n} \qquad \sqrt{n}(\overline{X_n} - \theta) = O_p(1)$$

$$\delta_n = \overline{X_n} - \frac{l_n'(\overline{X_n})}{l_n''(\overline{X_n})}$$

$$l_n(\theta) = n\theta - \sum X_i - 2\sum \log\left(1+e^{\theta - X_i}\right)$$

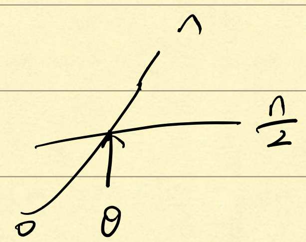
Set $\dfrac{\partial l_n(\theta)}{\partial \theta} = n - 2\sum \dfrac{e^{\theta - X_i}}{1+e^{\theta - X_i}} = 0$

unique root.

$$\Leftrightarrow \quad \sum \frac{e^{\theta - X_i}}{1+e^{\theta - X_i}} = \frac{n}{2}$$

θ $-\infty$ to $+\infty$

$\underline{g(\theta)}$ monotone $\nearrow$ fro $\underline{0 \text{ to } n}$



Correlation Coefficient.

$\rho = \operatorname{cov}(X,Y)$
$= E(XY) - E(X)E(Y)$

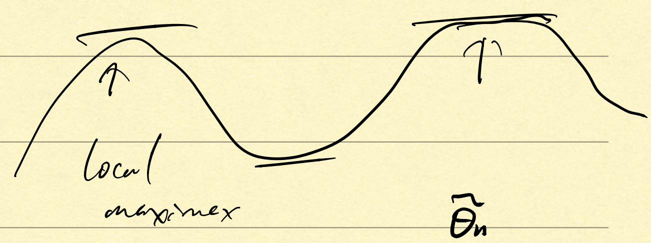
$$(X_1, Y_1), \ldots, (X_n, Y_n) \overset{iid}{\sim} N\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}1 & \rho\\ \rho & 1\end{pmatrix}\right)$$

$$f_\rho(x,y) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left\{-\frac{x^2 - 2\rho xy + y^2}{2(1-\rho^2)}\right\}, \quad -1 < \rho < 1$$

$$l_n(\rho) = -\frac{1}{2(1-\rho^2)} \sum_{i=1}^{n} \left(X_i^2 - 2\rho X_i Y_i + Y_i^2\right) - n \log\left(\sqrt{1-\rho^2}\right)$$

$$l_n'(\rho) = 0 \Rightarrow \quad \rho(1-\rho^2) + (1-\rho^2)\frac{\sum X_i Y_i}{n} - \rho \cdot \frac{\sum X_i^2 + \sum Y_i^2}{n} = 0$$

three roots?
two local maxima at maximum



$E X = 0 \quad E(Y) = 0 \qquad \rho = E(XY) = \dfrac{\sum X_i Y_i}{n}$ (MoM)

$\tilde{\theta}_n$

$$\to \delta_n = \begin{cases} \underbrace{\left(\tilde{\theta}_n - \dfrac{l_n'(\tilde{\theta}_n)}{l_n''(\tilde{\theta}_n)}\right)}_{*} & \text{if } * \in (-1, 1) \\ -1 & \text{if } * < -1 \\ 1 & \text{if } * > 1 \end{cases}$$

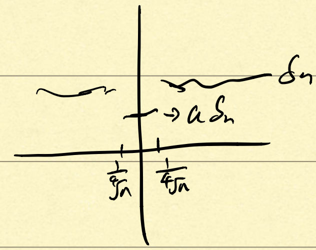
Efficiency Most Efficient Estimator.

$$\begin{cases} E(\hat{\theta}) \to \theta \\ Var(\hat{\theta}) \text{ is the smallest among all asy unbiased estimators} \end{cases}$$

Example. The Hodges Counterexample.

$$\sqrt{n}(\delta_n - \theta) \xrightarrow{d} N(0, v(\theta))$$

$$\delta_n' = \begin{cases} a\,\delta_n & \text{if } |\delta_n| \le \frac{1}{\sqrt[4]{n}} \\ \delta_n & \text{otherwise.} \end{cases}$$



$$\underline{P(\sqrt{n}|\delta_n' - \theta| \le c)} = P(\sqrt{n}|\delta_n| \le c\sqrt[4]{n},\ \sqrt{n}|\delta_n' - \theta| \le c)$$
$$+ P(\sqrt{n}|\delta_n| > c\sqrt[4]{n},\ \sqrt{n}|\delta_n' - \theta| \le c)$$

$$= P(\sqrt{n}|\delta_n| \le c\sqrt[4]{n},\ \sqrt{n}|a\delta_n - \theta| \le c)$$
$$+ P(\sqrt{n}|\delta_n| > c\sqrt[4]{n},\ \sqrt{n}|\delta_n' - \theta| \le c)$$

$\sqrt[4]{n} \to +\infty$

$\sqrt{n}|\delta_n| = O_p(1)$

If $\theta = 0$

$$P(\sqrt{n}|\delta_n' - 0| \le c) = P(\sqrt{n}|\delta_n| \le c\sqrt[4]{n},\ a\sqrt{n}|\delta_n| \le c)$$
$$+ P(\sqrt{n}|\delta_n| > c\sqrt[4]{n},\ \sqrt{n}|\delta_n| \le c)$$

$$= P(a\sqrt{n}|\delta_n| \le c) + 0$$

$\delta_n \xrightarrow{P} 0$

$\sqrt{n}(\delta_n - 0) \xrightarrow{d} N(0, v(0))$ $\quad \sqrt{n}\,\delta_n = O_p(1)$

$$\sqrt{n}(\delta_n' - 0) \xrightarrow{d} N(0, a^2 V(0))$$

If set $a<1$, $a^2 V(0) < V(0)$

δ_n' is better than δ_n for estimating $\theta = 0$

Theorem: Suppose δ_n is an efficient estimator of θ

$$\sqrt{n}(\delta_n - \theta) \xrightarrow{d} N\left(0, \frac{1}{I(\theta)}\right)$$

Then for g a differentiable function of θ

then $g(\delta_n)$ is an efficient estimator of $g(\theta)$

at all points θ for which $g'(\theta) \neq 0$.

$$\sqrt{n}\left(g(\delta_n) - g(\theta)\right) \xrightarrow{d} N\left(0, \frac{[g'(\theta)]^2}{I(\theta)}\right)$$

Multiparameter case. $\theta = (\theta_1, \cdots, \theta_k)$

Definition. $I(\theta) = [I_{ij}(\theta)]_{1 \le i, j \le k}$ Information Matrix

$$I_{ij}(\theta) = E\left[\frac{\partial}{\partial \theta_i} \log f_\theta(X) \frac{\partial}{\partial \theta_j} \log f_\theta(X)\right]$$

Exaple Normal (μ, σ^2) $\theta = (\mu, \sigma^2)$

$$I(\theta) = \begin{pmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{1}{2\sigma^4} \end{pmatrix}$$

(M1) The distributions P_θ are distinct.

$$\theta_1 \neq \theta_2 \Leftarrow P_{\theta_1} \neq P_{\theta_2}$$

Example: $N(\mu, \sigma^2)$, $\theta = \begin{pmatrix} \mu \\ \sigma \end{pmatrix}$

Single Index Model

$Y|X \sim N(g(X^T\beta), \sigma^2)$ function

fixed design.

$$\theta = \begin{pmatrix} g \\ \beta \\ \sigma \end{pmatrix} \to \text{vector}$$

$$\theta_1 = \begin{pmatrix} g_1(t) \\ \beta \\ \sigma \end{pmatrix} \quad \theta_2 = \begin{pmatrix} g_2(t) \\ -\beta \\ \sigma \end{pmatrix}$$

Consistency

$$N(g(X^T\beta), \sigma^2)$$

$$g_2(t) = g_1(-t)$$

(M2) The parameter space Ω is open.

(M3) iid observations $X_1, \cdots, X_n \overset{iid}{\sim} P_\theta$

(M4) The set A on which $f_\theta(x)$ is positive is independent of θ

(M5) for all $x \in A$, $\frac{\partial}{\partial \theta_i} f_\theta(x)$ exist for all $i = 1, \cdots, k$

(M6) The partial derivatives of the left side of

$$\int f_{(\theta_1 \cdots \theta_k)}(x)\,dx = 1$$

Calculating $I(\theta)$

exists and can be obtained by differentiating under the integral sign

---

Under (M1-M6)

$$I_{ij}(\theta) = \mathrm{Cov}\left[\frac{\partial}{\partial\theta_i}\log f_\theta(x), \frac{\partial}{\partial\theta_j}\log f_\theta(x)\right]$$

If (M6') the first two derivatives $\frac{\partial^2}{\partial\theta_i\partial\theta_j}$ exists

for all $x \in A$, and the exchange of differentiating and Intergal also holds for the second partial der.

$$I_{ij}(\theta) = -E_\theta\left[\frac{\partial^2}{\partial\theta_i\partial\theta_j}\log f_\theta(x)\right]$$

(M6)'' for all $x \in A$, the third derivatives $\frac{\partial^3}{\partial\theta_i\partial\theta_j\partial\theta_k} f_\theta(x)$ exist and are continuous. and the corresponding derivatives can be obtained by differentiating under the integral sign

$$\frac{\partial^3}{\partial\theta_i\partial\theta_j\partial\theta_k}\int f_{\theta_1,\ldots,\theta_k}(x)\,dx = 1$$

(M7) If $\theta_0 = (\theta_1^{(0)}, \cdots, \theta_k^{(0)})$ denotes the true value of θ, there exist functions $M_{ijk}(x)$ and a positive number $c(\theta_0)$ such that

$$\left| \frac{\partial^3}{\partial\theta_i \theta_j \theta_k} \log f_\theta(x) \right| \le M_{ijk}(x)$$

for all θ with $\Sigma(\theta_i - \theta_i^{(0)}) < c(\theta_0)$.

where $E_\theta[M_{ijk}(X)] < \infty$ for all i, j, k.

(M8) the elements $I_{ij}(\theta)$ of the information matrix $I(\theta) = [I_{ij}(\theta)]_{ij}$ are finite and the matrix $I(\theta)$ is positive definite.

Thm. Under (M1)-(M5), (M6)'', (M7) and M(8)
there exists a solution $\hat\theta_n = (\hat\theta_{1n}, \cdots, \hat\theta_{kn})$ of the likelihood equations which is consistent and any such solution satisfies.

$$\sqrt{n}\begin{pmatrix} \hat\theta_{1n} - \theta_1^{(0)} \\ \vdots \\ \hat\theta_{kn} - \theta_k^{(0)} \end{pmatrix} \xrightarrow{d} N\left(\underset{\sim}{0}, I^{-1}(\theta_0)\right)$$

Example (Normal) $X_1, \cdots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$ $\quad \theta = \begin{pmatrix}\theta_1 \\ \theta_2\end{pmatrix} = \begin{pmatrix}\mu \\ \sigma^2\end{pmatrix}$

$$l_n(\theta) = l_n(\mu, \sigma) = -n\log\sigma - \frac{1}{2\sigma^2}\sum_{i=1}^n (X_i - \mu)^2 - \cancel{n\log(\sqrt{2\pi})}$$

likelihood equations $\begin{cases} \frac{\partial}{\partial\mu} l_n(\theta) = 0 \\ \frac{\partial}{\partial\sigma} l_n(\theta) = 0 \end{cases} \Rightarrow \begin{cases} \frac{1}{\sigma^2}\Sigma(X_i - \mu) = 0 \\ -\frac{n}{\sigma} + \frac{1}{\sigma^3}\Sigma(X_i-\mu)^2 = 0 \end{cases}$

the unique solution: $\hat\mu = \bar{X}_n$

$\hat\sigma^2 = \frac{1}{n}\Sigma(X_i - \bar{X}_n)^2$

$$\sqrt{n}\begin{pmatrix}\hat\mu - \mu \\ \hat\sigma^2 - \sigma^2\end{pmatrix} \xrightarrow{d} N(0, \Sigma). \quad \Sigma = \begin{pmatrix}\sigma^2 & 0 \\ 0 & 2\sigma^4\end{pmatrix}$$

Example (Multinomial) $(n, p_1, p_2, p_3, \dots, p_k)$ where $\sum p_i = 1$

Observation $(n, y_1, y_2, \dots, y_k)$ where $\sum y_i = n$

$$L_n(\theta) = \frac{n!}{y_1! \cdots y_k!} \cdot p_1^{y_1} \cdots p_k^{y_k} \qquad \theta = \begin{pmatrix} p_1 \\ \vdots \\ p_k \end{pmatrix}$$

$$l_n(\theta) = \log L_n(\theta) = y_1 \log p_1 + \cdots + y_k \log p_k$$

$$= y_1 \log p_1 + \cdots + y_{k-1} \log p_{k-1} + y_k \log(1 - p_1 - \cdots - p_{k-1})$$

$$\frac{\partial l_n(\theta)}{\partial p_1} = \frac{y_1}{p_1} - \frac{y_k}{p_k} = 0 \qquad (p_k = 1 - p_1 - \cdots - p_{k-1})$$

$$\vdots$$

$$\frac{\partial l_n(\theta)}{\partial p_{k-1}} = \frac{y_{k-1}}{p_{k-1}} - \frac{y_k}{p_k} = 0$$

the unique solution: $\hat{p}_i = \dfrac{y_i}{n}$, $\underbrace{i = 1, \dots, k}$ $\quad k-1$

$$\sqrt{n}\left(\begin{pmatrix} \hat{p}_1 - p_1 \\ \vdots \\ \hat{p}_{k-1} - p_{k-1} \end{pmatrix}\right) \xrightarrow{d} N(0, I^{-1}(\theta))$$

$I^{-1}(\theta)$: a $(k-1) \times (k-1)$ matrix positive definite

If $\sqrt{n}\begin{pmatrix} \hat{p}_1 - p_1 \\ \vdots \\ \hat{p}_k - p_k \end{pmatrix} \xrightarrow{d} N(0, \Sigma)$

Σ is semi-positive definite

Eifficiency. Single Parameter. $\sqrt{n}(\hat{\theta}_1 - \theta) \xrightarrow{d} N(0, v_1(\theta))$

$\sqrt{n}(\hat{\theta}_2 - \theta) \xrightarrow{d} N(0, v_2(\theta))$

If $v_1(\theta) \le v_2(\theta)$ for all $\theta \in \Omega$

Then $\hat{\theta}_1$ is more efficient than $\hat{\theta}_2$

Multivariate. $\sqrt{n}(\hat{\underset{\sim}{\theta}}_1 - \underset{\sim}{\theta}) \xrightarrow{d} N(0, \Sigma_1(\theta))$

$\sqrt{n}(\hat{\underset{\sim}{\theta}}_2 - \underset{\sim}{\theta}) \xrightarrow{d} N(0, \Sigma_2(\theta))$

If $\Sigma_2(\theta) - \Sigma_1(\theta)$ is a $\overset{\text{semi-}}{}$positive definite matrix

then $\hat{\underset{\sim}{\theta}}_1$ is more efficient than $\hat{\underset{\sim}{\theta}}_2$

In practice: $\Sigma_1(\theta)$, $\Sigma_2(\theta)$ are too complicated to compare. theoretically

500 replications. $\hat{\underset{\sim}{\theta}}_1^{(m)}$ $m = 1, \cdots, 500$, $\hat{\underset{\sim}{\theta}}_2^{(m)}$ $m = 1, \cdots, 500$ } 500 sample

~~Do not~~ approximate $\Sigma_1(\theta)$ covariance matrix of $\hat{\underset{\sim}{\theta}}_1$

$$\underline{\hat{\Sigma}_1(\theta)} - \underline{\hat{\Sigma}_2(\theta)} \geq 0$$

Component-wise MSE: $\frac{1}{500} \sum_{m=1}^{500} \left(\theta_{1i}^{(m)} - \theta_i^{(0)} \right)^2$ for $i = 1, \cdots, k$

AMSE: $\sum_{i=1}^{K} \frac{1}{500} \sum_{m=1}^{500} \left(\theta_{i}^{(m)} - \theta_i^{(0)} \right)^2$

One-step estimation: Find $\underset{\sim}{\delta_n}$ $\sqrt{n}(\underset{\sim}{\delta_n} - \theta)$ is bounded $\xrightarrow{d} N(0, \Sigma^*)$

$$\tilde{\theta}_n = \delta_n - \left[l_n''(\delta_n) \right]^{-1} l_n'(\delta_n)$$

$\tilde{\theta}_n$ behaves as good as the MLE asympotically

Tests & Confidence Intervals using likelihood

A. The Wald test.

Suppose that $X_1, \cdots, X_n$ are iid. consider an estimator $\hat{\theta}_n$ which is efficient in the sense of satisfying $\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} N(0, I^{-1}(\theta))$

single parameter: $\sqrt{n}(\hat{\theta}_n-\theta)\times\sqrt{I(\theta)} \xrightarrow{d} N(0,1)$

$$\sqrt{n}(\hat{\theta}_n-\theta)\times\sqrt{I(\hat{\theta})} \xrightarrow{d} N(0,1)$$

$$\hat{\theta}_n-\frac{z_{\alpha/2}}{\sqrt{nI(\hat{\theta}_n)}} < \theta < \hat{\theta}_n+\frac{z_{\alpha/2}}{\sqrt{nI(\hat{\theta}_n)}}$$

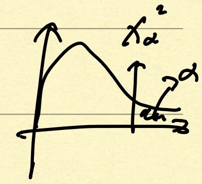
Approximately $100(1-\alpha)\%$ $\underbrace{\text{CI}}^{\text{Wald-Type}}$ for θ

$\theta=\begin{pmatrix}\theta_1\\ \vdots\\ \theta_k\end{pmatrix}$

Multiparameters: $I(\theta)$ is a matrix

$$[\sqrt{n}(\hat{\theta}_n-\theta)]'\,I(\theta)\,[\sqrt{n}(\hat{\theta}_n-\theta)] \xrightarrow{d} \chi^2_k$$

$$T_W = n\,(\hat{\theta}_n-\theta)'\,I(\hat{\theta}_n)\,(\hat{\theta}_n-\theta) \xrightarrow{d} \chi^2_k$$



Wald-Type

Asymptotic $100\times(1-\alpha)\%$ confidence Region:

$$\{\theta\in\mathbb{R}^k:\ n(\hat{\theta}_n-\theta)'\,I(\hat{\theta}_n)(\hat{\theta}_n-\theta)\le\chi^2_{k,\alpha}\}$$

Test:
- Simple. $H_0:\ \underset{\sim}{\theta}=\underset{\sim}{\theta}^*$ versus $H_1:\ \underset{\sim}{\theta}\ne\underset{\sim}{\theta}^*$
- Composite $H_0:\ h(\underset{\sim}{\theta})=\underset{\sim}{0}$ versus $H_1:\ h(\underset{\sim}{\theta})\ne 0$

  where $h:\ \mathbb{R}^k\to\mathbb{R}^q$ $\quad q\le k$

Eg: $H_0:\ \theta_1+\theta_2=0$ versus $H_1:\ \theta_1+\theta_2\ne 0$ $\quad h:\mathbb{R}^k\to\mathbb{R}$

$h(\theta_1,\dots,\theta_k)=\theta_1+\theta_2$

Typo I error: If H_0 is true. Probability of reject H_0

$\hookrightarrow \alpha$

① Find Test statistic (TS)

② Find rejection region (RR)

significance level

$\hookrightarrow P(TS\in RR \mid H_0 \text{ is true})=\alpha$

Wald-Type test for simple case:

$$T_W = n(\hat{\underset{\sim}{\theta}}_n - \underset{\sim}{\theta}^*)^T I(\hat{\underset{\sim}{\theta}}_n)(\hat{\underset{\sim}{\theta}}_n - \underset{\sim}{\theta}^*)$$

under H_0. $\underset{\sim}{\theta} = \underset{\sim}{\theta}^*$ holds

$T_W \xrightarrow{d} \chi^2_p$ where p is rank of $I(\underset{\sim}{\theta}^*)$

rejection region: reject H_0 if $T_W > \chi^2_{p,\alpha}$

Problem: Calculating $I(\underset{\sim}{\theta})$ may be difficult or Impossible.