Topics in Econometrics— Discrete Choice Models
Maximum Likelihood Estimation

Yao Thibaut Kpegli

Master II Advanced Economics, ENS Lyon

2022–2023

Intuition

Context

1. You have a sample $\boldsymbol{y} = \{y_1, ..., y_n\}$
2. You know that the sample comes from a random variable Y with a vector of parameters $\boldsymbol{\theta} \in \mathbb{R}^K$ whose true value is $\boldsymbol{\theta}_0$
3. You don't know the true value $\boldsymbol{\theta}_0$

Objective

Provide an estimate of $\boldsymbol{\theta}_0$

Intuition

$\hat{\boldsymbol{\theta}}_{MLE}$ = the value of $\boldsymbol{\theta}$ that is such that the probability of having observed $\boldsymbol{y}$ is the highest possible.

Intuition

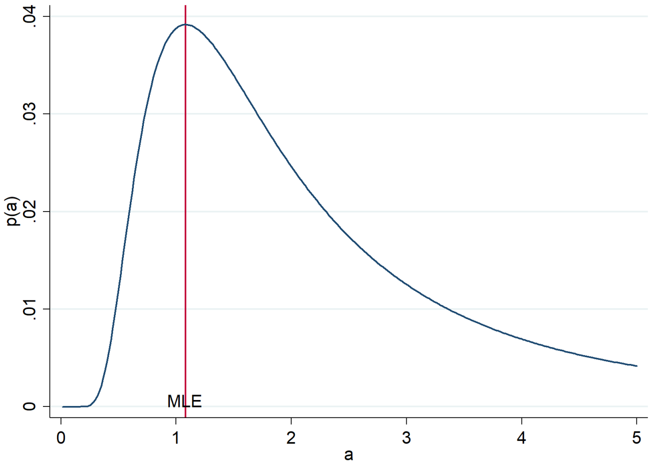
Example (a continuous case)

1. Assume that lifetime of an electronic equipment is a $r.v$ following an exponential distribution with parameter $a > 0$.
2. The density of exponential distribution with parameter $a > 0$ is $f(y; a) = \frac{1}{a}\exp\left(-\frac{y}{a}\right)$
3. We have observed randomly the lifetime 3 times, thereby constituting a sample $y_1 = 1, y_2 = 0.5$ and $y_3 = 1.75$.
4. You want to estimate a, that is the vector of parameters is simply $\boldsymbol{\theta} = a$
5. The joint density of having observed $\{y_1, y_2, y_3\}$ is

$$p(a) = \frac{1}{a}\exp\left(-\frac{1}{a}\right) \times \frac{1}{a}\exp\left(-\frac{0.5}{a}\right) \times \frac{1}{a}\exp\left(-\frac{1.75}{a}\right) = \frac{1}{a^3}\exp\left(-\frac{3.25}{a}\right)$$

Intuition

FIGURE: probability density of observing the sample as function of a



Intuition

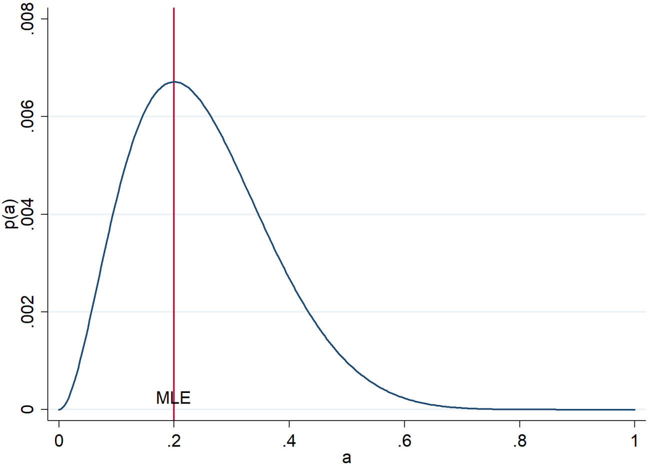
Example (a discrete case)

1. Assume that the fact that a resident of a town has a specific disease is a $r.v$ following a Bernoulli distribution with parameter $a \in (0,1)$, i.e. the probability of having the disease is a
2. We have randomly 10 people, thereby constituting a sample $\{y_1, y_2, ..., y_{10}\} = \{0,1,0,0,1,0,0,0,0,0\}$.
3. You want to estimate a, that is the vector of parameters is simply $\boldsymbol{\theta} = a$
4. The joint probability of observing $\{y_1, ..., y_{10}\}$ is

$$p(a) = (1-a)a(1-a)(1-a)a(1-a)(1-a)(1-a)(1-a)(1-a) = a^2(1-a)^8$$

Intuition

Figure: probability of observing the sample as function of a



Notations and Definitions

Notations

1. Y is a random variable with a distribution depending on a vector of parameters $\boldsymbol{\theta}$
2. An observed sample $\boldsymbol{y} = \{y_1, ..., y_n\}$ from n **independent** draws according to the distribution of Y
3. The probability $g(\boldsymbol{y}; \boldsymbol{\theta})$ of observing the sample $\boldsymbol{y}$ of Y is given by:

$$g(y_1, ..., y_n; \boldsymbol{\theta}) = \prod_{i=1}^{n} f(y_i; \boldsymbol{\theta})$$

with:

- $f(y; \boldsymbol{\theta})$ the probability density function (p.d.f) if Y is a continuous $r.v$
- the probability $P(Y = y_i; \boldsymbol{\theta})$ of the value y_i if Y is a discrete $r.v.$

Notations and Definitions

Definition (Likelihood function)

The likelihood function $\mathcal{L}$ is:

$$\begin{array}{ccc} \mathbb{R}^K & \longrightarrow & [0,1] \\ \theta & \mapsto & \mathcal{L}(\boldsymbol{\theta};\boldsymbol{y}) = g(\boldsymbol{y};\boldsymbol{\theta}) \end{array}$$

Definition (Log-likelihood function)

The log-likelihood function is:

$$log\mathcal{L}(\boldsymbol{\theta};\boldsymbol{y}) = \sum_{i=1}^{n} log f(y_i;\boldsymbol{\theta})$$

Notations and Definitions

Example

1. if $Y_i \rightsquigarrow \mathcal{B}(p)$, then $\boldsymbol{\theta} = p$ and
$$log\mathcal{L}(\boldsymbol{\theta};\boldsymbol{y}) = (\sum_{i=1}^{n} y_i)log(p) + (n - \sum_{i=1}^{n} y_i)log(1-p)$$
2. if $Y_i \rightsquigarrow \mathcal{N}(m, \sigma^2)$, then $\boldsymbol{\theta} = [m, \sigma^2]'$ and
$$log\mathcal{L}(\boldsymbol{\theta};\boldsymbol{y}) = -\frac{1}{2}\sum_{i=1}^{n}\left(log\sigma^2 + log(2\pi) + \frac{(y_i - m)^2}{\sigma^2}\right)$$

Notations and Definitions

Definition (Score function)

The score function is:

$$S(y;\boldsymbol{\theta}) = \frac{\partial log\ f(y,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$$

Example

1. For $Y \rightsquigarrow \mathcal{B}(p)$
$$S(y,\boldsymbol{\theta}) = \frac{y}{p} - \frac{1-y}{1-p}$$
2. For $Y \rightsquigarrow \mathcal{N}(m,\sigma^2)$
$$S(y,\boldsymbol{\theta}) = \begin{bmatrix} \dfrac{y-m}{\sigma^2} \\ \\ \dfrac{1}{2\sigma^2}\left(\left(\dfrac{y-m}{\sigma}\right)^2 - 1\right) \end{bmatrix}$$

Notations and Definitions

Proposition 1 (Score expectation)

The expectation of the score is zero

Proof:

$$\mathbb{E}[S(Y,\boldsymbol{\theta})] = \int \frac{\partial\, log\, f(y;\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} f(y;\boldsymbol{\theta})dy = \int \frac{\partial\, f(y;\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} \frac{1}{f(y;\boldsymbol{\theta})} f(y;\boldsymbol{\theta})dy =$$

$$\frac{\partial \int f(y;\boldsymbol{\theta})dy}{\partial\boldsymbol{\theta}} = \frac{\partial(1)}{\partial\boldsymbol{\theta}} = \mathbf{0}$$

Definition (Information matrix)

$$\mathcal{I}_Y(\boldsymbol{\theta}) = -\mathbb{E}\left(\frac{\partial^2\, log\, f(Y,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right)$$

Notations and Definitions

Remark (additivity of the information matrix)

The information matrix of two independent experiments is:

$$\mathcal{I}_{X,Y}(\boldsymbol{\theta}) = \mathcal{I}_X(\boldsymbol{\theta}) + \mathcal{I}_Y(\boldsymbol{\theta})$$

Proposition 2 (Variance of the score)

The variance of the score is equal to the information matrix

$$V\Big(S(Y;\boldsymbol{\theta})\Big) \equiv \mathbb{E}\left(\left(\frac{\partial \log f(Y,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)\left(\frac{\partial \log f(Y,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)'\right) = \mathcal{I}_Y(\boldsymbol{\theta})$$

Proof:

Note that $\dfrac{\partial^2 \log f(y,\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} = \dfrac{\partial^2 f(y;\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \dfrac{1}{f(y;\boldsymbol{\theta})} - \dfrac{\partial \log f(y,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \dfrac{\partial \log f(y;\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'}$

Also $\mathbb{E}\left[\dfrac{\partial^2 f(y;\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \dfrac{1}{f(y;\boldsymbol{\theta})}\right] = \dfrac{\partial^2 (1)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} = \mathbf{0}$. This leads to the result.

Notations and Definitions

Example

1. For $Y \rightsquigarrow \mathcal{B}(p)$

$$\mathcal{I}_Y(\boldsymbol{\theta}) = \mathbb{E}\left[\frac{Y}{p^2} + \frac{1-Y}{(1-p)^2}\right] = \frac{1}{p(1-p)}$$

2. For $Y \rightsquigarrow \mathcal{N}(m, \sigma^2)$

$$\mathcal{I}_Y(\boldsymbol{\theta}) = \mathbb{E}\begin{bmatrix} \frac{1}{\sigma^2} & \frac{y-m}{\sigma^4} \\ \frac{y-m}{\sigma^4} & \frac{(y-m)^2}{\sigma^6} - \frac{1}{2\sigma^4} \end{bmatrix} = \begin{bmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{1}{2\sigma^4} \end{bmatrix}$$

Notations and Definitions

Theorem 1 (Cauchy-Schwarz inequality)

Let X and Y be two random variables. Then:

$$|Cov(X,Y)| \leq \sqrt{V(X)V(Y)}$$

Proof: Let $Z = Y - \frac{Cov(X,Y)}{V(X)}X$. Then,

$Cov(X,Z) = Cov(X,Y) - Cov(X,X)\frac{Cov(X,Y)}{V(X)} = 0$. Then,

$V(Y) = V\left(Z + \frac{Cov(X,Y)}{V(X)}X\right) = V(Z) + \left(\frac{Cov(X,Y)}{V(X)}\right)^2 V(X) \geq \frac{(Cov(X,Y))^2}{V(X)}$.

Hence, $V(Y)V(X) \geq (Cov(X,Y))^2$ which leads to the result.

Notations and Definitions

Theorem 2 (Fréchet-Darmois-Cramér-Rao bound)

Consider an unbiased estimator $\hat{\boldsymbol{\theta}}(\boldsymbol{Y})$ of $\boldsymbol{\theta}$. The variance of the estimator $\hat{\boldsymbol{\theta}}(\boldsymbol{Y})$ has a lower bound:

$$V\left(\hat{\boldsymbol{\theta}}(\boldsymbol{Y})\right) \geq \mathcal{I}_Y(\boldsymbol{\theta})^{-1} \equiv B_F(\boldsymbol{\theta})$$

Proof: First, $Cov\left(\hat{\boldsymbol{\theta}}(\boldsymbol{Y}), S(Y;\boldsymbol{\theta})\right) = \mathbb{E}\left[(\hat{\boldsymbol{\theta}}(\boldsymbol{Y}) - \boldsymbol{\theta})S(Y;\boldsymbol{\theta})\right] = \mathbb{E}\left[\hat{\boldsymbol{\theta}}(\boldsymbol{Y})S(Y;\boldsymbol{\theta})\right] = \int \hat{\boldsymbol{\theta}}(\boldsymbol{y})S(Y;\boldsymbol{\theta})f(y;\boldsymbol{\theta})dy = \int \hat{\boldsymbol{\theta}}(\boldsymbol{y})\frac{\partial f(y;\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\frac{1}{f(y;\boldsymbol{\theta})}f(y;\boldsymbol{\theta})dy = \frac{\partial \int \hat{\boldsymbol{\theta}}(\boldsymbol{y})f(y;\boldsymbol{\theta})dy}{\partial \boldsymbol{\theta}} = \frac{\partial \mathbb{E}\left(\hat{\boldsymbol{\theta}}(\boldsymbol{Y})\right)}{\partial \boldsymbol{\theta}} = \frac{\partial \boldsymbol{\theta}}{\partial \boldsymbol{\theta}} = \mathbf{1}$. Hence, by Cauchy-Schwarz inequality we have $V\left(\hat{\boldsymbol{\theta}}(\boldsymbol{Y})\right)V\left(S(Y;\boldsymbol{\theta})\right) \geq \mathbf{1}$. The result follows as $V\left(S(Y;\boldsymbol{\theta})\right) = \mathcal{I}_Y(\boldsymbol{\theta})$.

Maximum Likelihood Estimation (MLE)

Definition (Identification)

The vector of parameters $\boldsymbol{\theta}$ is identifiable if, for any other vector $\boldsymbol{\theta}^*$:

$$\boldsymbol{\theta}^* \neq \boldsymbol{\theta} \Longrightarrow log\mathcal{L}(\boldsymbol{\theta}^*; \boldsymbol{y}) \neq log\mathcal{L}(\boldsymbol{\theta}; \boldsymbol{y})$$

Definition (Likelihood equation)

Necessary condition for maximizing the likelihood function:

$$\frac{\partial log\mathcal{L}(\boldsymbol{\theta}; \boldsymbol{y})}{\partial \boldsymbol{\theta}} = \sum_{i=1}^{n} S(y_i; \boldsymbol{\theta}) = \mathbf{0}$$

Definition (Maximum Likelihood Estimator (MLE))

The maximum likelihood estimator $\boldsymbol{\theta}_{MLE}$ is the vector θ that maximizes the likelihood function. Formally:

$$\boldsymbol{\theta}_{MLE} = \arg\max_{\boldsymbol{\theta}} log\mathcal{L}(\boldsymbol{\theta}; \boldsymbol{y}) = \arg\max_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta}; \boldsymbol{y})$$

Maximum Likelihood Estimator (MLE)

Example

1. For $Y \rightsquigarrow \mathcal{B}(p)$

$$\boldsymbol{\theta}_{MLE} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

2. For $Y \rightsquigarrow \mathcal{N}(m, \sigma^2)$

$$\boldsymbol{\theta}_{MLE} = \begin{bmatrix} \frac{1}{n}\sum_{i=1}^{n} y_i \\ \frac{1}{n}\sum_{i=1}^{n} (y_i - \boldsymbol{\theta}_{MLE})^2 \end{bmatrix}$$

Regularity conditions

Regularity conditions

1. The support of Y does not depends on $\boldsymbol{\theta}$
2. $\boldsymbol{\theta}_0$ is identified
3. The log-likelihood function is continuous in $\boldsymbol{\theta}$
4. $\mathbb{E}(log f(Y;\boldsymbol{\theta}_0))$ exists
5. The log-likelihood function is twice continuously differentiable
6. The information matrix at $\boldsymbol{\theta}_0$ $\mathcal{I}_Y(\boldsymbol{\theta}_0) = -\mathbb{E}\left(\frac{\partial^2 \; log \; f(Y, \boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right)$ exists and is nonsingular

Proposition 3 (Properties of MLE)

Under regularity conditions, the MLE is (i) consitent, (ii) asymptotically normally distributed and (iii) asymptotically efficient.

Consistency

Kullback-Leibler divergence

If $f_{\boldsymbol{\theta}_0}(y)$ and $f_{\boldsymbol{\theta}_1}(y)$ are two densities, the Kullback-Leibler divergence of $f_{\boldsymbol{\theta}_1}$ w.r.t $f_{\boldsymbol{\theta}_0}$ is

$$KL(f_{\boldsymbol{\theta}_1} \| f_{\boldsymbol{\theta}_0}) = \mathbb{E}_{\boldsymbol{\theta}_0}\left[log \frac{f(Y, \boldsymbol{\theta}_0)}{f(Y, \boldsymbol{\theta}_1)} \right] = \int f(y, \boldsymbol{\theta}_0) log \frac{f(y, \boldsymbol{\theta}_1)}{f(y, \boldsymbol{\theta}_0)} dy$$

Proposition 3-0

1. $KL(f_{\boldsymbol{\theta}_1} \| f_{\boldsymbol{\theta}_0}) \geq 0$
2. $KL(f_{\boldsymbol{\theta}_1} \| f_{\boldsymbol{\theta}_0}) = 0 \qquad \text{iff} \qquad f_{\boldsymbol{\theta}_0} = f_{\boldsymbol{\theta}_1}$

Proof: First, $-log(x)$ is a convex function. By Jensen's inequality,
$KL(f_{\boldsymbol{\theta}_1} \| f_{\boldsymbol{\theta}_0}) = \mathbb{E}_{\boldsymbol{\theta}_0}\left[-log \frac{f(Y, \boldsymbol{\theta}_1)}{f(Y, \boldsymbol{\theta}_0)} \right] \geq -log \mathbb{E}_{\boldsymbol{\theta}_0}\left[\frac{f(Y, \boldsymbol{\theta}_1)}{f(Y, \boldsymbol{\theta}_0)} \right] =$
$-log \int f(y, \boldsymbol{\theta}_1) dy = 0.$
Second, because $-log(x)$ is strictly convex, equality holds if and only if $f(y, \boldsymbol{\theta}_1)/f(y, \boldsymbol{\theta}_0)$ is constant.

Consistency

Proposition 3-1 (Consistency of MLE)

Under regularity conditions, $\boldsymbol{\theta}_{MLE}$ converge in probability to the true value $\boldsymbol{\theta}_0$:

$$\text{plim } \boldsymbol{\theta}_{MLE} = \boldsymbol{\theta}_0$$

Informal argument:

$$\begin{aligned}
\boldsymbol{\theta}_{MLE} &= \arg\max_{\boldsymbol{\theta}} \frac{1}{n}\sum_{i=1}^{n} log f(Y_i;\boldsymbol{\theta}) \\
&= \arg\min_{\boldsymbol{\theta}} -\frac{1}{n}\sum_{i=1}^{n} log f(Y_i;\boldsymbol{\theta}) \\
&= \arg\min_{\boldsymbol{\theta}} \frac{1}{n}\sum_{i=1}^{n} log f(Y_i;\boldsymbol{\theta_0}) - \frac{1}{n}\sum_{i=1}^{n} log f(Y_i;\boldsymbol{\theta}) \\
&\simeq_{n\longrightarrow+\infty} \arg\min_{\boldsymbol{\theta}} \mathbb{E}_{\boldsymbol{\theta}_0} log f(Y;\boldsymbol{\theta}_0) - \mathbb{E}_{\boldsymbol{\theta}_0} log f(Y;\boldsymbol{\theta}) \\
&\simeq_{n\longrightarrow+\infty} \arg\min_{\boldsymbol{\theta}} KL(f_{\boldsymbol{\theta}_0} \| f_{\boldsymbol{\theta}}) = \boldsymbol{\theta}_0 \\
&\simeq_{n\longrightarrow+\infty} \boldsymbol{\theta}_0
\end{aligned}$$

with $\simeq_{n\longrightarrow+\infty}$ stands for Law of Large Numbers

Consistency

Example

1. For $Y \rightsquigarrow \mathcal{B}(p)$

$$\boldsymbol{\theta}_{MLE} = \frac{1}{n}\sum_{i=1}^{n} Y_i \quad \xrightarrow{p} \quad E[Y] = p$$

2. For $Y \rightsquigarrow \mathcal{N}(m, \sigma^2)$

$$\boldsymbol{\theta}_{MLE} = \begin{bmatrix} m_{MLE} \\ \sigma^2_{MLE} \end{bmatrix}$$

$$m_{MLE} = \frac{1}{n}\sum_{i=1}^{n} Y_i \quad \xrightarrow{p} \quad E[Y] = m$$

$$\sigma^2_{MLE} = \frac{1}{n}\sum_{i=1}^{n} Y_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} Y_i\right)^2 \quad \xrightarrow{p} \quad E[Y^2] - E[Y]^2 = \sigma^2$$

Consistency

Remark

1. Regularity conditions are sufficient but not necessary conditions to have consistency
   - It is therefore possible for the MLE to be consistent even in situations that do not meet regularity conditions
   - e.g: if $Y \rightsquigarrow \mathcal{U}[0, \theta]$, we have $\boldsymbol{\theta}_{MLE} = \max\{Y_1, ..., Y_n\}$

$$\lim_{n \longrightarrow +\infty} \mathbb{P}(|\boldsymbol{\theta}_{MLE} - \boldsymbol{\theta}| > \epsilon) = \lim_{n \longrightarrow +\infty} \left(1 - \frac{\epsilon}{\theta}\right)^n = 0$$

Asymptotic normality

Proposition 2 (Asymptotic normality)

The MLE estimator $\boldsymbol{\theta}_{MLE}$ is normally distributed asymptotically:

$$\sqrt{n}\left(\boldsymbol{\theta}_{MLE}-\boldsymbol{\theta}_0\right) \xrightarrow{dist.} \mathcal{N}\left(0, \mathcal{I}_Y(\boldsymbol{\theta}_0)^{-1}\right)$$

Proof:

- First order Taylor expansion of the first derivative of $log\ \mathcal{L}(\boldsymbol{\theta};\boldsymbol{y})$
  $$\frac{\partial log\ \mathcal{L}(\boldsymbol{\theta};\boldsymbol{y})}{\partial\boldsymbol{\theta}} \simeq \frac{\partial log\ \mathcal{L}(\boldsymbol{\theta}_0;\boldsymbol{y})}{\partial\boldsymbol{\theta}} + \frac{\partial^2 log\ \mathcal{L}(\boldsymbol{\theta}_0;\boldsymbol{y})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}(\boldsymbol{\theta}-\boldsymbol{\theta}_0)$$
- Evaluate at $\boldsymbol{\theta}=\boldsymbol{\theta}_{MLE}$
  $$0 \simeq \frac{\partial log\ \mathcal{L}(\boldsymbol{\theta}_0;\boldsymbol{y})}{\partial\boldsymbol{\theta}} + \frac{\partial^2 log\ \mathcal{L}(\boldsymbol{\theta}_0;\boldsymbol{y})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}(\boldsymbol{\theta}_{MLE}-\boldsymbol{\theta}_0)$$

Asymptotic normality

- Rearrange

$$\sqrt{n}(\boldsymbol{\theta}_{MLE} - \boldsymbol{\theta}_0) \simeq \sqrt{n}\frac{\frac{\partial log\ \mathcal{L}(\boldsymbol{\theta}_0; \boldsymbol{y})}{\partial \boldsymbol{\theta}}}{-\frac{\partial^2 log\ \mathcal{L}(\boldsymbol{\theta}_0; \boldsymbol{y})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}} = \sqrt{n}\frac{\frac{1}{n}\sum_{i=1}^{n}\frac{\partial log\ f(y_i; \boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}}{-\frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2 log\ f(y_i; \boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}}$$

- Central Limit Theorem

$$A_n = \frac{1}{n}\sum_{i=1}^{n}\frac{\partial log\ f(y_i; \boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \xrightarrow{dist.} \mathcal{N}\left(E[S(Y, \boldsymbol{\theta}_0)], \frac{V\left(S(Y, \boldsymbol{\theta}_0)\right)}{n}\right)$$

- Law of Large Number

$$B_n = -\frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2 log\ f(y_i; \boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \xrightarrow{p} -E\left[\frac{\partial^2 log\ f(Y; \boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}\right] = \mathcal{I}_Y(\boldsymbol{\theta}_0)$$

- Remembering that $E[S(Y, \boldsymbol{\theta}_0)] = 0$ and $V\left(S(Y, \boldsymbol{\theta}_0)\right) = \mathcal{I}_Y(\boldsymbol{\theta}_0)$, Slutsky Theorem leads to

$$\sqrt{n}(\boldsymbol{\theta}_{MLE} - \boldsymbol{\theta}_0) = \sqrt{n}\frac{A_n}{B_n} = \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \mathcal{I}_Y(\boldsymbol{\theta}_0)^{-1}\right)$$

Asymptotically efficient

Proposition (Asymptotically efficient)

$\boldsymbol{\theta}_{MLE}$ is asymptotically efficient, *i.e.* achieves the FDCR lower bound for consistent estimators.

Proof: We have

$$\sqrt{n}(\boldsymbol{\theta}_{MLE} - \boldsymbol{\theta}_0) \xrightarrow{dist.} \mathcal{N}\left(0, \mathcal{I}_Y(\boldsymbol{\theta}_0)^{-1}\right)$$

This means that

$$V\left(\boldsymbol{\theta}_{MLE}\right) = \mathbb{I}(\boldsymbol{\theta}_0)^{-1}$$

with $\mathbb{I}(\boldsymbol{\theta}_0) = n\mathcal{I}_Y(\boldsymbol{\theta}_0)$ the information matrix associated to $\{Y_1, ..., Y_n\}$.

Remark:

- The asymptotic variance-covariance matrix $\mathbb{I}(\boldsymbol{\theta}_0)$ of the MLE depends on the unknown value of $\boldsymbol{\theta}_0$
- In practice, the matrix is evaluated at $\boldsymbol{\theta}_{MLE}$

Sum up

Sum up

1. You have a sample $\boldsymbol{y} = (y_1, ..., y_n)$
2. You know that the sample comes from a random variable Y with a vector of parameters $\boldsymbol{\theta} \in \mathbb{R}^K$ whose true value $\boldsymbol{\theta}_0$ is unknown
3. The log-likelihood of one observation y_i is computed analytically (as a function of $\boldsymbol{\theta}$): $l_i(\boldsymbol{\theta}; y_i) = log f(y_i; \boldsymbol{\theta})$
4. The log-likelihood of the sample is $log\mathcal{L}(\boldsymbol{\theta}; \boldsymbol{y}) = \sum_i l_i(\boldsymbol{\theta}; y_i)$
5. The The MLE estimator results from the optimization problem
$$\boldsymbol{\theta}_{MLE} = \arg\max_{\boldsymbol{\theta}} log\mathcal{L}(\boldsymbol{\theta}; \boldsymbol{y})$$
6. We have: $\boldsymbol{\theta}_{MLE} \rightsquigarrow \mathcal{N}\left(0, \mathbb{I}(\boldsymbol{\theta}_0)^{-1}\right)$, where $\mathbb{I}(\boldsymbol{\theta}_0)$ is estimated as $\mathbb{I}(\boldsymbol{\theta}_{MLE})$.

A replication of Gonzalez and Wu (1999)

- Let $L = (h, s; p, 1 - p)$ be a lottery that yields h with proba p and s otherwise, with $h > s \geq 0$
- The certainty equivalent y of a lottery L is the sure amount that provides the same level of utility (satisfaction)
- We observe from the experiment of Gonzalez and Wu (1999) a sample of certainty equivalent $\boldsymbol{y} = \{y_1, ..., y_n\}$ associated to lotteries $L_i = (h_i, s_i; p_i, 1 - p_i)$, $i = 1, .., n$
- Assume that y_i is a realisation of $Y_i \rightsquigarrow \mathcal{N}(m(h_i, s_i, p_i), \sigma^2)$
- According to Prospect Theory (Tversky et Kahneman, 1992), we assume

$$m(h_i, s_i, p_i) = u^{-1}[w(p_i)u(h_i) + (1 - w(p_i))u(s_i)]$$

with $u(.)$ and $w(.)$ the utility and weighting functions

A replication of Gonzalez and Wu (1999)

- According to Gonzalez and Wu (1999), we assume

$$u(z) = z^{\alpha} \quad \text{and} \quad w(p) = \frac{\delta p^{\gamma}}{\delta p^{\gamma} + (1-p)^{\gamma}}$$

so that

$$m(h_i, s_i, p_i) = \left[(h_i^{\alpha} - s_i^{\alpha}) \frac{\delta p_i^{\gamma}}{\delta p_i^{\gamma} + (1-p_i)^{\gamma}} + s_i^{\alpha} \right]^{\frac{1}{\alpha}}$$

- The log-likelihood $l_i(\boldsymbol{\theta}; \boldsymbol{y})$ associated to y_i is then

$$l_i(\boldsymbol{\theta}; y_i) = -\frac{1}{2} \left(log \sigma^2 + log(2\pi) + \frac{(y_i - m(h_i, s_i, p_i))^2}{\sigma^2} \right)$$

with $\boldsymbol{\theta} = [\alpha, \delta, \gamma, \sigma^2]'$

A replication of Gonzalez and Wu (1999)

Programming of the the MLE in STATA (ML routine)

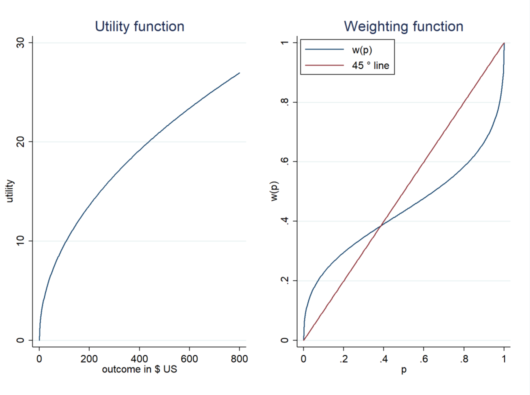
A replication of Gonzalez and Wu (1999)

Table: Estimation results : ml

id	Coefficient	(Std. Err.)	id	Coefficient	(Std. Err.)
	α			γ	
1	0.676	(0.102)	1	0.392	(0.027)
2	0.227	(0.069)	2	0.651	(0.045)
3	0.654	(0.130)	3	0.388	(0.027)
4	0.590	(0.053)	4	0.154	(0.021)
5	0.401	(0.078)	5	0.273	(0.022)
6	0.676	(0.059)	6	0.890	(0.025)
7	0.604	(0.059)	7	0.204	(0.016)
8	0.390	(0.077)	8	0.373	(0.042)
9	0.518	(0.074)	9	0.863	(0.039)
10	0.455	(0.090)	10	0.501	(0.034)
11	0.493	(0.043)	11	0.435	(0.014)
	δ			σ^2	
1	0.461	(0.115)	1	421.467	(46.402)
2	1.508	(0.517)	2	873.034	(96.118)
3	1.449	(0.379)	3	621.578	(68.433)
4	0.211	(0.039)	4	138.202	(15.216)
5	1.191	(0.303)	5	966.502	(106.408)
6	1.331	(0.155)	6	97.057	(10.686)
7	0.376	(0.065)	7	226.450	(24.931)
8	0.381	(0.122)	8	567.347	(62.463)
9	0.896	(0.178)	9	282.749	(31.130)
10	0.935	(0.253)	10	715.257	(78.747)
11	0.766	(0.097)	11	153.888	(16.943)

A replication of Gonzalez and Wu (1999)

Figure: Plot for the median subject (id = 11)



- Concave utility function (in gain domain)
- Overweighting (resp. underweighting) of small (resp. intermedaite and high) probability of winning

Next: Binary Outcome Models!