L15 Likelihood equation and maximum likelihood estimator (MLE)

1. Likelihood equation
   $X_1, ..., X_n$ is a random sample from X with pdf or pmf $f_X(x; \theta)$ where $\theta \in \Omega \subset R^k$.

   (1) Likelihood function
   The joint pdf or pmf for $X_1, ..., X_n$ treated as a function θ is the likelihood function.

   $$L(\theta|X_1, ..., X_n) = f(X_1, ..., X_n; \theta) = f_X(X_1; \theta) \cdots f_X(X_n; \theta)$$

   $\widehat{\theta}$ is the maximum likelihood estimator (MLE) for θ

   $$\stackrel{def}{\Longleftrightarrow} L(\theta|X_1, ..., X_n) \leq L(\widehat{\theta}|X_1, ..., X_n) \ \forall \theta \in \Omega$$

   $\widehat{\theta}$ is also denoted as $\widehat{\theta}_n$ to indicate the sample size n.

   (2) Log-likelihood function

   $$\ln L(\theta|X_1, .., X_n) = \ln f(X_1, ..., X_n; \theta) = \ln f_X(X_1; \theta) + \cdots + \ln f_X(X_n; \theta)$$

   is the log-likelihood function.

   $$\widehat{\theta}_n \text{ is the MLE for } \theta \Longleftrightarrow \ln f(X_1, .., X_n; \theta) \leq \ln f(X_1, .., X_n; \widehat{\theta}) \ \forall \theta \in \Omega.$$

   (3) Likelihood equation
   $\frac{\partial}{\partial \theta} \ln f(X_1, .., X_n; \theta) = 0$ is the likelihood equation.
   If $\widehat{\theta}$ is the MLE for θ, then $\widehat{\theta}$ is a solution to the likelihood equation.
   If MLE exists and the solution of the likelihood equation is unique, then

   $$\widehat{\theta} \text{ is the MLE of } \theta \Longleftrightarrow \widehat{\theta} \text{ is the solution to the likelihood equation}$$

   The likelihood equation can be written as

   $$\frac{\partial}{\partial \theta} \ln f_X(X_1; \theta) + \cdots + \frac{\partial}{\partial \theta} \ln f_X(X_n; \theta) = 0$$

   Comment: $f_X(X_1; \theta)$,...,$f_X(X_n; \theta)$ are iid with a common distribution, the distribution of $f_X(X; \theta)$;
   $\ln f_X(X_1; \theta)$,...,$\ln f_X(X_n; \theta)$ are iid with a common distribution, the distribution of $\ln f_X(X; \theta)$;
   $\frac{\partial}{\partial \theta} \ln f_X(X_1; \theta)$,...,$\frac{\partial}{\partial \theta} \ln f_X(X_n; \theta)$ are iid with a common distribution, the distribution of $\frac{\partial}{\partial \theta} \ln f_X(X; \theta) \sim (0, I(\theta))$.

2. Solving likelihood equation when $\theta \in R$

   (1) Newton-Raphson method for solving $f(x) = 0$

   $$\boxed{\begin{array}{l} x_0 \\ x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)} \end{array}}$$

   x_0 is the first conjectured solution.
   Let x_k be the kth suggested solution. Approximate $f(x)$ in a neighborhood of x_k by the local linear approximation $f(x) \approx f(x_k) + (x - x_k)f'(x_k)$ and solve the equation to get the $(k+1)$th suggested solution $x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)}$.

(2) Solving the likelihood equation

Let $g(\theta) = 0$ be the likelihood equation. Then $\boxed{\begin{array}{l}\tilde{\theta}_0 \\ \tilde{\theta}_{k+1} = \tilde{\theta}_k - \frac{g(\tilde{\theta}_k)}{g'(\tilde{\theta}_k)}\end{array}}$

(3) Modifications

In the likelihood equation $g(\theta) = 0$, $g(\theta) = \frac{\partial}{\partial \theta} \ln f(X_1, ..., X_n; \theta)$.

So $E[g'(\theta)] = E\left[\frac{\partial^2}{\partial \theta^2} \ln f(X_1, ..., X_n; \theta)\right] = -nI(\theta)$.

Replace $g'(\tilde{\theta}_k)$ by its mean $-nI(\tilde{\theta}_k)$. We have $\boxed{\begin{array}{l}\tilde{\theta}_0 \\ \tilde{\theta}_{k+1} = \tilde{\theta}_k + \frac{g(\tilde{\theta}_k)}{nI(\tilde{\theta}_k)}\end{array}}$

3. Convex functions and concave functions

(1) Convex combination

For x and y in a LS, with $\lambda \in [0, 1]$, $\lambda x + (1 - \lambda)y$ is a convex combination of x and y.

A set is a convex set if it is closed under convex combinations.

(2) Convex functions

$f(\cdot)$ is a convex function $\overset{def}{\Longleftrightarrow} f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda) f(y)\ \forall\, \lambda \in [0, 1]$.

$f(\cdot)$ is a strictly convex function $\overset{def}{\Longleftrightarrow} f(\lambda x + (1 - \lambda)y) < \lambda f(x) + (1 - \lambda) f(y)$ $\forall\, x \neq y$ and $\lambda \in (0, 1)$.

f is a strictly convex function $\Longrightarrow f$ is a convex function

(3) Concave functions

$f(\cdot)$ is a concave function $\overset{def}{\Longleftrightarrow} f(\lambda x + (1 - \lambda)y) \geq \lambda f(x) + (1 - \lambda) f(y)\ \forall\, \lambda \in [0, 1]$.

$f(\cdot)$ is a strictly concave function $\overset{def}{\Longleftrightarrow} f(\lambda x + (1 - \lambda)y) > \lambda f(x) + (1 - \lambda) f(y)$ $\forall\, x \neq y$ and $\lambda \in (0, 1)$.

f is a strictly concave function $\Longrightarrow f$ is a concave function.

(4) f is a convex function $\Longleftrightarrow -f$ is a concave function;

f is a strictly convex function $\Longleftrightarrow -f$ is a strictly concave function.

Ex1: $f(x) = \ln x$, $x > 0$ is a strictly concave function

For $x > 0$, $y > 0$ and $0 < \lambda < 1$, $\lambda x + (1 - \lambda)y$ is a weighted arithmetic mean of x and y; $x^\lambda y^{1-\lambda}$ is a geometric mean of x and y. It is well known that $\lambda x + (1 - \lambda)y > x^\lambda y^{1-\lambda}$ when $x \neq y$. Thus $\ln[\lambda x + (1 - \lambda)y] > \lambda \ln x + (1 - \lambda) \ln y$. So $f(x) = \ln x$ is a strictly concave function.

L16: Jensen's inequality

1. Iff conditions for convex/concave functions

   (1) Using argument average and function average
   With $\sum_{i=1}^k \lambda_i = 1$ where $\lambda_i \in (0,\, 1)$, $a_k = \sum_{i=1}^k \lambda_i x_i$ is an average of $x_1 < \cdots < x_k$ and $b_k = \sum_{i=1}^k \lambda_i f(x_i)$ is the corresponding average of $f(x_i)$, $i = 1, .., k$.

(i)	$f(x)$ is convex	$\iff$	for all $k \geq 2$, $f(a_k) \leq b_k$
(ii)	$f(x)$ is strictly convex	$\iff$	for all $k \geq 2$, $f(a_k) < b_k$
(iii)	$f(x)$ is concave	$\iff$	for all $k \geq 2$, $f(a_k) \geq b_k$
(iv)	$f(x)$ is strictly concave	$\iff$	for all $k \geq 2$, $f(a_k) > b_k$

   Pf: We show (i) only.

   $\Leftarrow$: With $k = 2$, by the definition, $f(x)$ is a convex function.

   $\Rightarrow$: By the definition of convexity of $f(x)$, $f(a_2) \leq b_2$. Assume $f(a_t) \leq b_t$. We show $f(a_{t+1}) \leq b_{t+1}$.

   $$\begin{aligned} f(a_{t+1}) &= f(\textstyle\sum_{i=1}^{t+1} \lambda_i x_i) = f((1-\lambda_{t+1}) \sum_{i=1}^{t} \frac{\lambda_i}{1-\lambda_{t+1}} x_i + \lambda_{t+1} x_{t+1}) \\ &\leq (1-\lambda_{t+1}) f(\textstyle\sum_{i=1}^{t} \frac{\lambda_i}{1-\lambda_{t+1}} x_i) + \lambda_{t+1} f(x_{t+1}) \\ &\leq (1-\lambda_{t+1}) \textstyle\sum_{i=1}^{t} \frac{\lambda_i}{1-\lambda_{t+1}} f(x_i) + \lambda_{t+1} f(x_{t+1}) \\ &= \textstyle\sum_{i=1}^{t+1} \lambda_i f(x_i) = b_{t+1}. \end{aligned}$$

   By induction $f(a_k) \leq b_k$ for all $k \geq 2$. $\square$

   (2) Using the monotonicity of the slops of secant lines
   Secant line passing $(x_i,\, f(x_i))$ and $(x_j,\, f(x_j))$ has slope $k_{ij} = \frac{f(x_i)-f(x_j)}{x_i - x_j}$.
   With $x_1 < x_2 < x_3$ there are three k_{ij}, $i < j$: k_{12}, k_{13}, k_{23}.

(i) $f(x)$ is convex	$\iff$	$k_{12} \leq k_{13} \iff k_{12} \leq k_{23} \iff k_{13} \leq k_{23}$
(ii) $f(x)$ is strictly convex	$\iff$	$k_{12} < k_{13} \iff k_{12} < k_{23} \iff k_{13} < k_{23}$
(iii) $f(x)$ is concave	$\iff$	$k_{12} \geq k_{13} \iff k_{12} \geq k_{23} \iff k_{13} \geq k_{23}$
(iv) $f(x)$ is strictly concave	$\iff$	$k_{12} > k_{13} \iff k_{12} > k_{23} \iff k_{13} > k_{23}$

   Pf: Show the first $\iff$ in (ii). With $x_1 < x_2 < x_3$,
   $x_2 = \frac{x_3-x_2}{x_3-x_1} x_1 + \frac{x_2-x_1}{x_3-x_1} x_3 = \lambda x_1 + (1-\lambda) x_3$ where $\lambda = \frac{x_3-x_2}{x_3-x_1} \in (0, -, 1)$.

   $\Rightarrow$: By the convexity of $f(x)$, $f(x_2) \leq \lambda f(x_1) + (1-\lambda) f(x_3)$. So
   $f(x_2) - f(x_1) \leq (1-\lambda)[f(x_3) - f(x_1)] = \frac{x_2-x_1}{x_3-x_1}[f(x_3) - f(x_1)]$.
   Thus $\frac{f(x_2)-f(x_1)}{x_2-x_1} \leq \frac{f(x_3)-f(x_1)}{x_3-x_1}$, i.e., $k_{12} \leq k_{13}$.

   $\Leftarrow$: $k_{12} \leq k_{13} \Longrightarrow \frac{f(x_2)-f(x_1)}{x_2-x_1} \leq \frac{f(x_3)-f(x_1)}{x_3-x_1}$.
   So $f(x_2) \leq \frac{x_3-x_2}{x_3-x_1} f(x_1) + \frac{x_2-x_1}{x_3-x_1} f(x_3) = \lambda f(x_1) + (1-\lambda) f(x_3)$.
   Thus $f(x)$ is convex. $\square$

2. Iff-conditions for f to be convex/concave when f is differentiable

   (1) $f'(x)$ exists

(i)	$f(x)$ is convex	$\iff$	$f'(x)$ is non-decreasing
(ii)	$f(x)$ is strictly convex	$\iff$	$f'(x)$ is increasing
(iii)	$f(x)$ is convave	$\iff$	$f'(x)$ is non-increasing
(iv)	$f(x)$ is strictly concave	$\iff$	$f'(x)$ is decreasing.

   Pf: Show the first $\iff$

$\Rightarrow$**:** We show that $x_1 < x_2 \Longrightarrow f'(x_1) \le f'(x_2)$. Let $x_1 < x_1 + h_1 < x_2 < x_2 + h_2$.
By (i) of (2) in 1,
$x_1 < x_1 + h_1 < x_2 \Longrightarrow \frac{f(x_1+h_1)-f(x_1)}{h_1} \le \frac{f(x_2)-f(x_1)}{x_2-x_1} \Longrightarrow f'(x_1) \le \frac{f(x_2)-f(x_1)}{x_2-x_1}$.
$x_1 < x_2 < x_2 + h_2 \Longrightarrow \frac{f(x_2)-f(x_1)}{x_2-x_1} \le \frac{f(x_2+h_2)-f(x_2)}{h_2} \Longrightarrow \frac{f(x_2)-f(x_1)}{x_2-x_1} \le f'(x_2)$.
Thus $f'(x_1) \le f'(x_2)$.

$\Leftarrow$**:** With $x_1 < x_2 < x_3$, we show that $k_{12} \le k_{23}$.
$k_{12} = \frac{f(x_2)-f(x_1)}{x_2-x_1} = f'(\xi_1)$ where $\xi_1 \in (x_1, x_2)$
and $k_{23} = \frac{f(x_3)-f(x_2)}{x_3-x_2} = f'(\xi_2)$ where $\xi_2 \in (x_2, x_3)$.
With non-decreasing $f'(x)$, $k_{12} = f'(\xi_1) \le f'(\xi_2) = k_{23}$. Thus $f(x)$ is convex. □

(2) f'' exists.

(i)	$f(x)$ is convex	$\Longleftrightarrow$	$f''(x) \ge 0$
(ii)	$f(x)$ is strictly convex	$\Longleftrightarrow$	$f''(x) > 0$
(iii)	$f(x)$ is convave	$\Longleftrightarrow$	$f''(x) \le 0$
(iv)	$f(x)$ is strictly concave	$\Longleftrightarrow$	$f''(x) < 0$.

Pf: Trivial.

3. Jensen's inequality

(1) Theorem
X is a non-degenerated random variable, i.e., X is not a non-random constant. Then

(i)	$f(x)$ is convex	$\Longrightarrow$	$f(E(X)) \le E[f(X)]$
(ii)	$f(x)$ is strictly convex	$\Longrightarrow$	$f(E(X)) < E[f(X)]$
(iii)	$f(x)$ is concave	$\Longrightarrow$	$f(E(X)) \ge E[f(X)]$
(iv)	$f(x)$ is strictly concave	$\Longrightarrow$	$f(E(X)) > E[f(X)]$.

Pf: Prove (i) in two cases

(a) Case I: X has discrete distribution with $P(X = x_i) = p_i$, $i = 1, ..., k$.
$f(E(X)) = f(\sum_{i=1}^k p_i x_i) \le \sum_{i=1}^k p_i f(x_i) = E[f(X)]$.

(b) Case II: Convex function $f(x)$ has second order derivative.
By Taylor expansion,

$$f(X) = f(E(X)) + [X - E(X)]f'(E(X)) + \frac{1}{2!}[X - E(X)]^2 f''(\xi).$$

Hence

$$E[f(X)] = f(E(X)) + 0 \cdot f'(E(X)) + E\left\{\frac{1}{2}E[X - E(X)]^2 f''(\xi)\right\} \ge f(E(X)). \square$$

(2) Examples

Ex1: $f(x) = \ln x$ on $x > 0$ has $f'(x) = \frac{1}{x}$ which is a decreasing function on $x > 0$.
Thus $f(x) = \ln x$ is a strictly concave function on $x > 0$.

Ex2: $f(x) = \ln x$ on $x > 0$ has $f''(x) = \frac{-1}{x^2} < 0$.
Thus $f(x) = \ln x$ is a strictly concave function on $x > 0$.

Ex3: For non-degenerated X, $\ln E(X) > E(\ln X)$.