ការប៉ាន់ស្មាន (Estimation)

១ សេចក្តីផ្តើម

- ឧបមាថាយើងមានសំនុំសាកល (population) ដែលមានទំហំ N មធ្យម μ និង វ៉ារ្យ៉ង់ σ^2។ μ និង σ^2 ត្រូវបានគេហៅថាប៉ារ៉ាម៉ែត្រ (parameter)។
- ចេញពីសំនុំសាកលនេះ យើងរើសគំរូតាង (sample) ដោយចៃដន្យដែលមានទំហំ n ជាមួយ $n \leq N$។ តាមរយៈគំរូតាងដែលយើងរើសបាន យើងអាចគណនា មធ្យម $\bar{X}_n$ និង វ៉ារ្យ៉ង់ S_n^2។ $\bar{X}_n$ និង S_n^2 ត្រូវបានគេហៅថា statistics ។
- *វត្ថុបំណងនៃមេរៀននេះគឺ យើងប្រើ statistics $(\bar{X}_n, S_n^2)$ ដើម្បីប៉ាន់ស្មាន parameter (μ, σ^2) ។ ក្នុងការប៉ាន់ស្មាននេះ យើងនឹងប្រើវិធី Maximum Likelihood Estimation (MLE)។*
- ជាធម្មតា គំរូតាងមួយបង្កើតបាន statistics មួយ (ចេញពីសំនុំសាកល (population) ដែលមានទំហំ N យើងអាចរើសបានគំរូតាង (sample) ដោយចៃដន្យដែលមានទំហំ n ច្រើនអនេក)។ ដោយសារតែយើងមានគំរូតាងច្រើន នោះ statistics ក៏មានតម្លៃច្រើនដែរ ដែលជាទូទៅតម្លៃនីមួយៗមិនស្មើគ្នាទេ។ ចុះបើ statistics ($\bar{X}_n$) មានតម្លៃច្រើនបែបនេះ តើយើងគួរប្រើតម្លៃមួយណាដើម្បីប៉ាន់ស្មាន parameter (μ) ?
- ក្នុងជាក់ស្តែង យើងមានលទ្ធភាពរើសយកគំរូតាងតែមួយទេ (រើសច្រើនដង ចំណាយច្រើន)។

២ ការប៉ាន់ស្មានជាចំនុច (Point Estimation)

២.១ ម៉ូដែលស្ថិតិ (statistical model)

នៅជំពូកមុនយើងបានសិក្សាពីអថេរចៃដន្យ លក្ខណៈនានារបស់វា និង អថេរចៃដន្យមួយចំនួន។ ជាធម្មតាអថេរចៃដន្យត្រូវបានកំណត់លក្ខណៈ (សំគាល់) ដោយអនុគមន៍ដង់ស៊ីតេ ដែលជាធម្មតាអនុគមន៍ដង់ស៊ីតេអាស្រ័យទៅនឹងប៉ារ៉ាម៉ែត្រ θ។ បើ $\theta \in \mathbb{R}^k$ នោះយើងសិក្សានៅក្នុងស្ថិតិប៉ារ៉ាម៉ែត្រ (Parametric Statistics) ហើយបើ $\theta \notin \mathbb{R}^k$ នោះយើងសិក្សានៅក្នុងស្ថិតិមិនប៉ារ៉ាម៉ែត្រ (Nonparametric Statistics)។ ក្នុងវគ្គនេះយើងនិងសិក្សាតែក្នុងករណី ដែល $\theta \in \mathbb{R}^k$។ វត្ថុបំណងក្នុងចំណុចនេះ គឺការប៉ាន់ស្មាន θ ដោយប្រើប្រាស់សំនុំទិន្នន័យ $x_1, x_2, \ldots, x_n$។

ឧទាហរណ៍ ១. *ឧបមាថាយើងបោះកាក់មួយ n ដង ដែលក្នុងការបោះមួយដងៗឪកាសនៃការចេញមុខគឺ $p, 0 < p < 1$ ។ តាង X ជាចំនួនមុខដែលផ្ងារឡើង។ នោះគេថា X ជាអថេរចៃដន្យមានបំណែងចែកទ្វេធាដែលអនុគមន៍ដង់ស៊ីតេអោយដោយរូបមន្តៈ*

$$P_p(X = k) = C_n^k p^k (1-p)^{(n-k)}, \ \ k = 0, 1, 2, \ldots, n$$

តាមរូបមន្តនេះយើងឃើញថា អនុគមន៍ដង់ស៊ីតេអាស្រ័យទៅនឹង p ដែល p ហៅថាប៉ារ៉ាម៉ែត្រ។

និយមន័យ ១. *ម៉ូដែលស្ថិតិ (Statistical Model) គឺជាគ្រួសារនៃអនុគមន៍ដង់ស៊ីតេ $\{P_\theta, \theta \in \Theta \subset \mathbb{R}^k\}$ ដែល P_θ ជាអនុគមន៍ដង់ស៊ីតេរបស់អថេរចៃដន្យណាមួយ។*

ឧទាហរណ៍ ២. *ឧទាហរណ៍នៃម៉ូដែលស្ថិតិ*

- *បំណែងចែកប៊ែរនូយី (Bernoulli Distribution):*
  $\{P_\theta, \theta \in \Theta \subset \mathbb{R}^k\} = \{p^x(1-p)^{1-x}, p \in]0,1[\}$
- *បំណែងចែកទ្វេធា (Binomial Distribution):*
  $\{P_\theta, \theta \in \Theta \subset \mathbb{R}^k\} = \{C_n^x p^x(1-p)^{(n-x)}, p \in]0,1[\}$
- *បំណែងចែកព័រសុង (Poisson Distribution):*
  $\{P_\theta, \theta \in \Theta \subset \mathbb{R}^k\} = \{\frac{e^{-\lambda}\lambda^x}{x!}, \lambda \in]0,\infty[\}$
- *បំណែងចែកន័រម៉ាល់ (Normal Distribution):*
  $\{P_\theta, \theta \in \Theta \subset \mathbb{R}^k\} = \{\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}, (\mu,\sigma^2) \in \mathbb{R}\times]0,\infty[, k=2\}$

២.២ អនុគមន៍ likelihood (Likelihood Function)

ក្នុងចំណុចនេះយើងនឹងសិក្សាអំពីការគណនាការប៉ាន់ស្មានជាចំណុច (point estimator) តាមរយៈអនុគមន៍ likelihood។

និយមន័យ ២. *X ជាអថេរចៃដន្យដែលមានអនុគមន៍ដង់ស៊ីតេ $f_\theta(x)$ រឺ $P_\theta(X=x)$ នោះអនុគមន៍ likelihood ត្រូវបានគេអោយនិយមន័យដោយ*

$$L_x(\theta) = \Pi_{i=i}^n f_\theta(x_i), \theta \in \Theta$$

ហើយ អនុគមន៍ log-likelihhod វូបានគេអោយនិយមន័យដោយ

$$l_x(\theta) = ln(L_x(\theta)), \theta \in \Theta$$

សម្គាល់ ១.

- *f_θ ជាអនុគមន៍នៃ x ហើយ L_x និងl_x ជាអនុគមន៍នៃ θ ។*
- *ដោយលោការីតជាអនុវត្តន៍មួយទល់មួយ ដូចនេះដើម្បីសិក្សាពីការប៉ាន់ស្មាន យើងប្រើ $L(\theta)$ និង $l(\theta)$ អោយលទ្ធផលដូចគ្នា។*
- *i.i.d = Independent and identically distributed random variables*

ឧទាហរណ៍ ៣. *មាន $X_i \sim i.i.d \quad Ber(p), 0<p<1$ ដែល $i=1,2,\dots,n$ ។ ចូរកំណត់ $L(p)$ និង $l(p)$ ។*

ចំណុចបន្ទាប់ យើងនឹងសិក្សាពីការប៉ាន់ស្មាន θ ដោយប្រើប្រាស់សំនុំទិន្នន័យ $x_1, x_2, \dots, x_n$។ តំលៃប៉ាន់ស្មាននៃប៉ារ៉ាម៉ែត្រ θ (estimator) គឺជាតំលៃ(statistic) ទាំងឡាយណាដែលធ្វើអោយ l មានតំលៃអតិបរមា។

និយមន័យ ៣. *យើងថា $\hat{\theta}$ ជាតំលៃប៉ាន់ស្មាននៃប៉ារ៉ាម៉ែត្រ θ (estimator) បើ*

$$\hat{\theta} = Argmax_{\theta \in \Theta} l(\theta)$$

ឧទាហរណ៍ ៤. *ចូរកំណត់តំលៃប៉ាន់ស្មាននៃប៉ារ៉ាម៉ែត្រ θ (estimator) ក្នុងករណី៖*

- *បំណែងចែក Bernoulli, Poisson, Geometric*
- *បំណែងចែក Uniform[0, θ], Exponential, Normal*

ទ្រឹស្តីបទ ១. *មាន X_i អថេរចៃដន្យ $i.i.d$ និង ដង់ស៊ីតេ $f_\theta(x)$ ។ ឧបមាថាមានអនុគមន៍ g ដែល $\lambda = g(\theta)$ នោះតំលៃប៉ាន់ស្មានរបស់ λ គឺ $\hat{\lambda} = g(\hat{\theta})$ ដែល $\hat{\theta}$ ជាតំលៃប៉ាន់ស្មាននៃប៉ារ៉ាម៉ែត្រ θ ។*

សម្រាយ

dfdsgsdfgsdfhgdfgsfd □

ឧទាហរណ៍ ៥. *ចូរគណនាតំលៃប៉ាន់ស្មានរបស់* $V(x)$ *បើ* $X \sim Ber(p)$ ។

សម្គាល់ ២.

- *ជាធម្មតាតំលៃប៉ាន់ស្មានរបស់ប៉ារ៉ាម៉ែត្រ គេតែងតែថែមសន្ទស្សន៍* n, $\hat{\theta}_n$
- *យើងឃើញថាតំលៃប៉ាន់ស្មាន* $\hat{\theta}_n$ *ជាអនុគមន៍នៃ* $X_1, X_2, \ldots, X_n$ *ដែលសុទ្ធតែជាអថេរចៃដន្យ នោះ* $\hat{\theta}_n$ *ក៏ជាអថេរចៃដន្យដែរ។*

ការសន្មត ១. *(Regular Condition)*

- *Identifiability: បើ* $\theta \neq \theta'$ *នោះ* $f_\theta(x) \neq f_{\theta'}(x)$
- *Support: គ្រប់អនុគមន៍ដង់ស៊ីតេមានទម្រង់ដូចគ្នា,* $\forall \theta \in \Theta$

សម្គាល់ ៣. *យើងក៏មានវិធីសាស្ត្រផ្សេងទៀតដើម្បីកំណត់តំលៃប៉ាន់ស្មាននៃប៉ារ៉ាម៉ែត្រ* θ *(estimator) គឺវិធីម៉ូម៉ង់ (Moment Method)។ ឧបមាថាមានអថេរចៃដន្យ X ដែលមានម៉ូម៉ងកំណត់ (finite* k^{th} *moment) មានន័យថា* $E[|x^k|] < \infty$ *នោះតំលៃប៉ាន់ស្មានរបស់* $\theta_k = E[x^k]$ *អោយដោយរូបមន្ត*

$$\hat{\theta}_{k,n} = \frac{1}{n}\sum_{i=i}^{n} X_i^k$$

ឧទាហរណ៍ ៦. *ចូរគណនាតំលៃប៉ាន់ស្មានរបស់* $E(X)$ *និង* $V(x)$ *តាមវិធី Likelihood និង ម៉ូម៉ង់បើ៖*

- $X \sim Ber(p)$
- $X \sim N(\mu, \sigma^2)$

២.៣ លក្ខណៈនៃការប៉ាន់ស្មានជាចំណុច (Properties of Point Estimator)

បន្ទាប់ពីនិយាយពីតំលៃប៉ាន់ស្មានរួចហើយ ជាបន្ទាប់យើងនឹងនិយាយពីលក្ខណៈនានារបស់តំលៃប៉ាន់ស្មាននោះ។

និយមន័យ ៤.

១. *(Unbiased) : យើងថា* $\hat{\theta}_n$ *ជាតំលៃប៉ាន់ស្មានមិនល្អៀង (unbiased estimator) របស់* θ *បើ*

$$E(\hat{\theta}_n) = \theta$$

២. *(Efficiency): យើងថា* $\hat{\theta}_n$ *ជាតំលៃប៉ាន់ដ៏មានប្រសិទ្ធិភាព (efficient estimator) របស់* θ *បើ*

$$V(\hat{\theta}_n) \xrightarrow{n->\infty} 0$$

៣. *(Cosistency): យើងថា* $\hat{\theta}_n$ *ជាតំលៃប៉ាន់កុងស៊ីស្តង់ (consistent estimator) របស់* θ *បើវា* $\hat{\theta}_n$ *រួមកេប៉ារ៉ាម៉ែត្រ* θ *ក្នុងប្រូបាប ដែលគេកំណត់សរសេរដោយ*

$$\hat{\theta}_n \xrightarrow[n->\infty]{P} \theta$$

៤. *[Mean Square Error (MSE)]: អោយដោយរូបមន្ត*

$$MSE(\hat{\theta}_n) = E[(\hat{\theta}_n - \theta)^2] = V(\hat{\theta}_n) - (Bias(\hat{\theta}_n))^2$$

$$\text{ដែល } Bias(\hat{\theta}_n) = E(\hat{\theta}_n) - \theta$$

ឧទាហរណ៍ ៧. *ចូរគណនាពាក្យក្នុងនិយមន័យទាំង ៤ ខាងលើចំពោះ $\hat{\theta}_n$ ក្នុងបំណែងចែក Bernoulli ។*

ការសន្មត ២. *(derivative condition): មាន X_i អថេរចៃដន្យ $i.i.d$ និង ដង់ស៊ីតេ $f_\theta(x)$ ។*

- *f មានដេរីវេលំដាប់ពីរធៀបនឹង θ*
- *$\int f_\theta(x)dx$ មានដេរីវេលំដាប់ពីរក្រោមសញ្ញាអាំងតេក្រាលធៀបនឹង θ*

បទគន្លឹះ ១. *: ឧបមាថា X (អថេរចៃដន្យតែមួយ) មានអនុគមន៍ដង់ស៊ីតេ $f_\theta(x)$ ។ ក្រោមការសន្មតទាំងពីរខាងលើយើងបាន៖*

- $E(\dfrac{\partial log(f_\theta(x))}{\partial\theta}) = 0$
- $E(\dfrac{\partial^2 log(f_\theta(x))}{\partial\theta^2}) = -E[(\dfrac{\partial log(f_\theta(x))}{\partial\theta})^2]$

សម្រាយ

dfdsgsdfgsdfhgdfgsfd □

និយមន័យ ៥. *Fisher Information ត្រូវបានគេអោយនិយមន័យដោយ*

$$I(\theta) = -E(\frac{\partial^2 log(f_\theta(x))}{\partial\theta^2}) = E[(\frac{\partial log(f_\theta(x))}{\partial\theta})^2] = V[\frac{\partial log(f_\theta(x))}{\partial\theta}]$$

ឧទាហរណ៍ ៨. *ចូរគណនា $I(\theta)$ នៃ $X \sim Ber(p)$ ។*

ករណីទូទៅ: ឧបមាថា $X_1, X_2, \dots, X_n, i.i.d$ មានអនុគមន៍ដង់ស៊ីតេ $f_\theta(x)$។ នោះយើងបាន៖

- $E(\dfrac{\partial l}{\partial\theta}) = 0$
- $E(\dfrac{\partial^2 l}{\partial\theta^2}) = -E[(\dfrac{\partial l}{\partial\theta})^2]$
- $V[\dfrac{\partial l}{\partial\theta}] = -E(\dfrac{\partial^2 l}{\partial\theta^2}) = E[(\dfrac{\partial l}{\partial\theta})^2] = nI(\theta)$

ទ្រឹស្តីបទ ២. *(Rao-Cramér Lower Bound): ឧបមាថា $X_1, X_2, \dots, X_n, i.i.d$ មានអនុគមន៍ ដង់ស៊ីតេ $f_\theta(x)$ ។ ឧបមាថាយើងមានការសន្មតខាងលើ។ តាង $Y = u(X_1, X_2, \dots, X_n)$ ដែល $E(Y) = k(\theta)$ នោះយើងបាន ៖*

$$V(Y) \geq \frac{[k'(\theta)]^2}{nI(\theta)}$$

សម្រាយ

fgfdgfd □

៣ ទ្រឹស្តីបទលីមីត (Limit Theorem)

៣.១ ភាពរួមរបស់ស្វីតនៃអថេរចៃដន្យ (Convergences of Sequence of Random Variables)

៣.២ ច្បាប់នៃចំនួនធំ (Law of Large Number:LLN)

៣.៣ ទ្រឹស្តីបទគំលែកណ្តាល (Central Limit Theorem:CLT)

៤ ការប៉ាន់ស្មានជាចន្លោះ (Interval Estimation)

៤.១ ការប៉ាន់ស្មានជាចន្លោះចំពោះសមាមាត្រ (Interval Estimation for Proportion)

៤.២ ការប៉ាន់ស្មានជាចន្លោះចំពោះមធ្យម (Interval Estimation for Mean)

៤.៣ ការប៉ាន់ស្មានជាចន្លោះចំពោះវ៉ារ្យង់ (Interval Estimation for Variance)