参考 クラメールラオの下限の不等式の証明

ここではクラメールラオの下限 (Cramer Rao Lower Bound) の不等式について解説する。まずは基本的な事柄から確認していこう。読者は今までに色々な確率分布に出会ったはずだ。例えば離散型では二項分布 $Bin(n,\, p)$、ポアソン分布 $Po(\lambda)$、連続型では正規分布 $N(\mu,\, \sigma^2)$、指数分布 $e(\beta)$ など。ここで $p,\, \lambda,\, \mu,\, \sigma,\, \beta$ などは母数またはパラメータと呼ばれ、それらは確率分布を特徴づける重要な要素であり、我々はしばしばその値に関心を持っている。

次に不偏推定量の意味について確認しておこう。例えば今、会社に 1 時間あたりにかかってくる電話の回数に興味があるとする。1 時間あたりにかかってくる電話の回数はパラメータ λ のポアソン分布に従うとする。ただし λ の値は時間帯等に関係なく常に一定であるとする。今、あなたは λ の値が何なのか知りたいと思っている。(λ の真の値は神様のみが知りうるので、我々はせいぜい推定するしかない) そこで 1 時間毎に、かかってくる電話の回数を記録し、それらの回数を $X_1,\, X_2,\, ...,\, X_n$ というように、合計 n 時間分の記録をとった。この時、実際に観測された $X_1,\, X_2,\, ...,\, X_n$ の値の平均、つまり $\dfrac{X_1 + X_2 + ... + X_n}{n} = \bar{X}$ を λ の値として推測することが考えられる。ここで $\bar{X}$ の期待値を計算すると $E[\bar{X}] = E[\dfrac{X_1 + X_2 ... + X_n}{n}] = \dfrac{E[X_1] + E[X_2] + ... + E[X_n]}{n} = \dfrac{\lambda + \lambda + ... + \lambda}{n} = \lambda$ となる。$\bar{X}$ のように観測されたデータから構成され (λ を含まない)、パラメータの推測に用いるものを「統計量」と呼ぶ。さらに統計量のうち、その期待値がパラメータに一致するようなものを「不偏推定量」と呼ぶ。もちろん $E[X_1] = \lambda$ なので、X_1 も不偏推定量である。(つまりその場合 $X_2,\, ...,\, X_n$ の値は観測したにも関わらず無視したことになる)

X_1 と $\bar{X}$ はどちらも不偏推定量であるが、後者の方がよりよい推定量である。それは推定量の分散を比較すれば分かる。$V[X_1] = \lambda$ なのに対して $V[\bar{X}] = \dfrac{V[X_1] + V[X_2] + ... + V[X_n]}{n^2} = \dfrac{n\lambda}{n^2} = \dfrac{\lambda}{n}$ である。つまり後者の方が分散が小さい。分散はデータのバラツキを表していたことを思い出そう。$\bar{X}$ の方がバラツキつまり分散が小さいということは実際に観測される値は λ からそんなにかけ離れていないのに対し、X_1 の方はバラツキが大きくなるため、偶然 λ とは全然違った数値を観測してしまう可能性が高くなる。そのため推定精度の観点からいうと $\bar{X}$ の方が優れた推定量と言えるのである。

さてここから本題に入るとしよう。今、先ほどと同様にある確率分布のパラメータ θ を推定するという状況を考える。ただし、より一般性を持たせるため、いま θ の関数で表される $\tau(\theta)$ を推定するものとして話を進める。いまその確率分布の確率密度関数は $f(x|\theta)$ で表されているとする。(もし $f(x|\theta)$ の $|\theta$ の意味がややこしければ無視して、$f(x)$ と思っていただいても差し支えない。ただしこの関数には θ も含まれていることは意識しながら読んでいただきたい) そして、その確率分布に従う標本 $X_1,\, X_2,\, ...,\, X_n$ を観測したとする。今 T が $\tau(\theta)$ の不偏推定量であるとする。つまり T は $X_1,\, X_2,\, ...,\, X_n$ から構成される関数であり、また $E[T] = \tau(\theta)$ が成立する。

この時 $V[T] \geqq \dfrac{(\tau'(\theta))^2}{I_n(\theta)}$ が成り立つ。($I_n(\theta) = nI(\theta)$) これをクラメールラオの不等式と呼ぶ。$I(\theta)$ はフィッシャー情報量と呼ばれるもので $E[(\dfrac{\partial}{\partial\theta} logf(x|\theta))^2]$ で計算される。この不等式に一体何の意味があるかと一言で言えば、推定精度の限界を表していると言える。先ほどから推定量の分散は小さければ小さいほど推定精度は良いと言っていた。この不等式では、$\tau(\theta)$ の不偏推定量 T の分散について「これ以上は小さくすることは出来ませんよ」という限界の目安を与えている。もしある推定量の分散がこの不等式の右辺に一致する場合は (つまり等号成立の場合) 有効推定量と呼ばれる。

ただし等号を満たすような不偏推定量が常に存在するとは限らない。また一様分布のように一部の分布ではこの不等式は成立しない。

次にこの不等式の成立を証明しよう。この不等式を変形すると $E[(T-\tau(\theta))^2]I_n(\theta) \geqq (\tau'(\theta))^2$ となる。さらに $I_n(\theta)$ の部分を変形すると $E[(T-\tau(\theta))^2]E[(\frac{\partial}{\partial\theta}logf(x_1, x_2, ...x_n|\theta))^2] \geqq (\tau'(\theta))^2$ となる。※ ただし $f(x_1, x_2, ...x_n|\theta)$ は同時確率密度関数を表しており $f(x_1|\theta)f(x_2|\theta)...f(x_n|\theta)$ の意味である。また上で $I_n(\theta) = nI(\theta)$ と書いたように、$nE[(\frac{\partial}{\partial\theta}logf(x|\theta))^2]$ とも書けるが、$E[(\frac{\partial}{\partial\theta}logf(x_1, x_2, ...x_n|\theta))^2]$ もまた同じである。n 個の標本から得られるフィッシャー情報量も、1 つあたりの標本から得られるフィッシャー情報量の n 倍も同じだからである。

ここで Cauch-Schwaltz の不等式を適用する。つまり $E[X^2]E[Y^2] \geqq E[XY]^2$ である。これを式の左辺に適用すると、左辺 $\geqq E[(T-\tau(\theta))((\frac{\partial}{\partial\theta}logf(x_1, x_2, ..x_n|\theta)))]^2$ である。ここで右辺について $E[(T-\tau(\theta))\frac{f'(x_1, x_2, ..x_n|\theta)}{f(x_1, x_2, ..x_n|\theta)}]$ と書き換えられる。(注意: f' は θ による微分である) さらに $E[]$ 内は ($X_1, X_2, ..., X_n$) の関数であることに注意して $\int(T-\tau(\theta))\frac{f'(x_1, x_2, ..., x_n|\theta)}{f(x_1, x_2, ..., x_n|\theta)}f(x_1, x_2, ..., x_n|\theta)dx_1dx_2...dx_n$ となる。式を整理して $\int(T-\tau(\theta))f'(x_1, x_2, ..., x_n|\theta)dx_1dx_2...dx_n$ となる。さらに $\int Tf'(x_1, x_2, ..., x_n|\theta)dx_1dx_2...dx_n - \tau(\theta)\int f'(x_1, x_2, ..., x_n|\theta)dx_1dx_2...dx_n$ と変形できる。ここで f' は θ による微分であるから $\frac{\partial}{\partial\theta}f$ と書き直し、また $\int$ の前に置く。よって $\frac{\partial}{\partial\theta}\int Tf(x_1, x_2, ..., x_n|\theta)dx_1dx_2...dx_n - \tau(\theta)\frac{\partial}{\partial\theta}\int f(x_1, x_2, ..., x_n|\theta)dx_1dx_2...dx_n$ となる。さらにこれは $\frac{\partial}{\partial\theta}E[T] - \tau(\theta)\frac{\partial}{\partial\theta}(1) = \frac{\partial}{\partial\theta}\tau(\theta) - 0 = \tau'(\theta)$ となる。以上のことからクラメールラオの不等式が成り立つことが示された。

次に等号成立条件について検討してみよう。上記の Caucy-Schwaltz の不等式において $Y = kX$(k は定数) が成り立つ時に等号成立する。$X = (T-\tau(\theta)), Y = \frac{\partial}{\partial\theta}logf(x_1, x_2, ..x_n|\theta)$ であった。よって $(T-\tau(\theta)) = K(n,\theta)\frac{\partial}{\partial\theta}logf(x_1, x_2, ..x_n|\theta)$ の形で書ける時にクラメールラオの下限を満足する。