Application of the Bays' theorem to a kind of Sampling Problems.

Abstract

C. Hayashi

The estimate of population parameter is supported by the thought of confidence interval. In some problems, this estimate is not only improper but also can not give the exact meaning. Especially in the case where we want to estimate the non-existence of a label in the population by investigating the samples. In this problem, one method to solve the problem using some kinds of past knowledge is shown by applying the Bays' theorem.

It is the fundamental idea that so-called a priori probability law is estimated from its moments obtained by using the observed values and making the relations between the moments

of the a priori probability function and the moments calculated from the observed values.

This idea must be greatly used in future, I think.

② 「ない」事を知る サムプリングの一案

サムプリングの問題に於てある標識が母集團に皆無であると言ふ事を知り得る可能性について

林 知己夫

現今のサムンプリングに於ては，母集団に対する推定はみな信頼度信頼巾の形でなされてゐる。

したがつてある種の問題に対しては明確に解答を與へ得ない場合がある。 たとへば，此はある標識を示すものが母集団には存在しないと言ふ事をサムプルの知識からのべてみたいと言ふ様な問題である。

さらに又深く考へてみるに信頼巾の考へ方で満足すべき有用な知識が我々にあたへられて居るであらうか。

信頼巾信頼度の論理には隔靴搔痒の感なしとは言ひ得ぬ所がある。

それならばどの様な推定の方法がとらるべきであらうか。

当然こゝにコレクテイフの分割行爲（partition）の考へが利用せらるべきであらう。 即ち原因の確率の定理とも言はれるベイズの定理が用ひらるべきであらう。

ベイズの定理が用ひられる場合問題になるのは所謂存在の確率と言はるべきものが我々には解つて居らぬと言ふことである。

此が判明せざる限りベイズの定理の利用はナンセンスと言はざるを得ぬのである。

過去に集積されてある知識の利用によつて此の存在の確率を推定出來るならばBaysの論理を用ひて母集団に対する推定を行ふ事が出

きるのである。

直接実例について此の可能性を注意してみよう。

調査対象の各個の抽出される確率を同一とすることによつて母集団を構成する。　此の標識は１，０であり，その大さはＮとしよう。

此より大さｎのサムプルを抽出し，それらを調査し標識１の数をしらべあげるとする。　今此の数が０であつたとする。　此の時母集団が標識１を含まぬと言ふことはどれ丈の信頼性があるであらうか。

母集団は標識１をθ個もつてゐるとするときｎ個のサムプルを抽出した時x個標識１を得る確率

$$p(x|\theta)=\frac{\binom{\theta}{x}\binom{N-\theta}{n-x}}{\binom{N}{n}}$$

を得る。　我々の母集団を要素とする原母集団とも言ふべきものを考へる。　此の標識は０からＮまでの正整数であり，母集団のθの値が即ち標識となつてゐるのである。　我々が上述の様な結論を得たいと言ふ事は即ちこの様な母集団を想定して居ることになるのである。

さてこのθの分布 $P(\theta)$ が Baysの意味の存在確率と見做されるものなのである。

今、サンプル中標識１をなすものが存在しなかつたとする、この時母集団にそれが存在せぬことに帰因する確率、即ち母集団にそれが皆無だと言ふときの信頼性 g_0 は

$$g_0=\frac{P(0)\,p(0|0)}{P(0)p(0|0)+\sum_{\theta=1}^{N}p(0|\theta)P(\theta)}$$

$$=\frac{P(0)}{P(0)+\sum_{\theta=1}^{N}p(0|\theta)P(\theta)}$$

によつてあたへられる。　此が我々の求むる結論なのである。

帰因すると言ふ明析な結論ではあるまいか。

$p(\theta)$ を知らなければ q_0 は全く求められない。

$p(\theta)$ を知るために過去の知識が利用せられ得るものとしよう。

具体的にのべるならば過去にも同様の母集団から n 個のサムプルを多数回抽出、調査した結果標識1を x 個含むでゐるものの相対頻度(割合) F_x が判明してゐるものとしよう。(此の過去の調査の系列はコレクテイフの一部をなしてゐると假定する。

此の事は実際問題に於ては妥当な假定であらうと思ふ。)

しかし，$p(\theta) = F(\theta)$ でないのは申すまでもない。

F_x の知識から $p(\theta)$ を求めることを考へてみよう。

このまへに先づ次の関係式が成立することに注意しよう。

$$\sum_{x=0}^{n} x\,p(x|\theta) = \theta\,\frac{n}{N}$$

$$\sum_{x=0}^{n} x(x-1)\,p(x|\theta) = \theta(\theta-1)\,\frac{n\cdot(n-1)}{N\cdot(N-1)}$$

$$\sum_{x=0}^{n} x(x-1)(x-2)\,p(x|\theta) = \theta(\theta-1)(\theta-2)\,\frac{n(n-1)(n-2)}{N(N-1)(N-2)}$$

$$\vdots$$

今，此等の両辺に $p(\theta)$ を乗じて θ につき 0 から N まで加える。

即ち，θ について平均をとると

$$\sum_{x=0}^{n} x \sum_{\theta=0}^{N} p(x|\theta)P(\theta) = \frac{n}{N}\sum_{\theta=0}^{N}\theta P(\theta)$$

$$\sum_{x=0}^{n} x(x-1) \sum_{\theta=0}^{N} p(x|\theta)P(\theta) = \frac{n(n-1)}{N(N-1)}\sum_{\theta=0}^{N}\theta(\theta-1)P(\theta),$$

$$\sum_{x=0}^{n} x(x-1)(x-2) \sum_{\theta=0}^{N} p(x|\theta)P(\theta) = \frac{n(n-1)(n-2)}{N(N-1)(N-2)}\sum_{\theta=0}^{N}\theta(\theta-1)(\theta-2)P(\theta)$$

・
・
・
・
・

となる。　さて　$\sum_{\theta=0}^{N} p(x|\theta)P(\theta)$　は何をあらはして居るであらうか、

此はサムプル調査（サムプルの大さn）の結果の系列の中標識１をなすものの個数がx個である様なサムプルを得る確率をあらはしてゐるのである。F_x は正に此の値の推定値となつてゐるのである。

かうすると，上の式の左辺は

$$\sum_{x=0}^{n} x F_x,\quad \sum_{x=0}^{n} x(x-1)F_x,\quad \sum_{x=0}^{n} x(x-1)(x-2)F_x \text{ ------}$$

として実際に推定的に計算できるのである。

さらに一つの事が過去の知識からわかつてゐるとしよう。
即ち，$P(\theta)$は n より大な θでは十分小であると豫想されるものとしよう。

この時　$\sum_{\theta=n+1}^{N} P(\theta) \fallingdotseq 0$　と考へよう。　さうすると上に述べた関係式から直ちに

$P(0),\ P(1),$ --------- $P(n)$　を求めることが出来る。

かうすることによつて q_0 が求められてくるのである。

此が我々の到達し得た結論なのである。 q_0 は信頼性であり、我々が着目してゐる原因の妥当する確率なのである。

さて以上の様な母集団の皆無性を知らうとする場合，実際的には母集団の θがさう拡がつて居らぬと言ふ事は当然の事である。したがつて P(θ) は θ がさう大でないときからすでに十分小であると仮定出来る場合が多い。 されば P(0) から P(n) まで面倒な計算を行つて求める必要なく，求むべき P(θ) の個数は相当小さくてすむであらう。 此の辺の求むべき P(θ) の個数の程度等は夫々の実例につき試行錯誤的な Successive な方法によつて計算的に定められるべきものである。

以上によつて皆無性の推定に関する一つの可能性（サムプリングの問題に於ける）を示したものである。 こゝでのべた様に過去の知識を Bays の定理を用ひ得られる様に採用してゆく行き方は今後大いに研究せらるべきものであらうと思ふ。

サムプリングの今後の発展の方向として見棄てらるべきものでないことを注意したい。

註。 q_0 は F_x の函数である。 F_x は推定値である。したがつて q_0 の信頼度を求める必要もおこるが此は通常のサムプリング理論から容易に求められる。

なほ Richard Von Mises は此の種の興味ある推定方法に対して忘れ得ぬ理論を展開してゐる。

On the conect use of Bays' formula
Amals of Mathematical statistic 1942.