

經濟論叢

第九十九卷 第二號

ペイズの決定関数	鎌倉昇	1
セールスマンシップ論の成立	橋本勲	15
イギリス産業革命期における 製鉄業技術の発展段階	坂本和一	33
「貨幣の資本への転化」と宇野理論	松石勝彦	51

昭和四十二年二月

京都大學經濟學會

ベイズの決定関数

——選択理論の最近における発展 (2)——

鎌 倉 昇

本稿は「経済論叢」昭和40年12月号所収「ベルヌーイの効用指標」にただちに接続するものであり、さきに記したように、昭和38年以来3カ年連続して京都大学大学院でおこなった講義「数理経済学特論」が基礎になっている。

2. 1. 抜取り検査の例

統計的決定の考え方を説明するにあたって、まず一つの例をかかげよう。抜取り検査によって品質の管理をおこなっている工場を念頭におけばよい。罐詰会社でもよいし、電球工場でもよい。すべての製品について検査を行なうことは無意義であり、一部の製品を抜取って調査することにより、全体の製品が一定の水準に達しているか否かを判定するのである。

罐詰工場を考えて話を進めることにする。罐詰はベルト・コンベアーに乗ってつぎつぎに生産されてくる。そのすべてを調査するのではないから、そのうちの一部について調査を行なうわけである。いまかりに製品のうち、ある時点に出来上がってくる塊りをロットとよぶことにする。1,000個単位にロットが作られているとしよう。いまこの1,000個だけを調査して、それによって、それ以外のロットの製品も品質検査の基準にあうかどうかをみようとするのである。

この場合、このロットが検査に合格するか否かによって、残りのすべての製品を合格とみるか、不合格とみて破棄するかの分れ目になるわけであり、手順はつぎのような形をとる。いま1,000個からなるロットを調べてみて、もしそのうち n 個が不合格の場合、全体の製品をかりに合格あるいは不合格と決断をくだしたとして、効用の高さはどのようになるかを考えるわけである。この場

合の効用としてベルヌーイの効用指標を利用し得ることはいうまでもない。

いま単純化のために、ベルヌーイの効用指標が第1表のような形でしめされるとしよう。1,000個のロットの中、100個が不合格品であった場合を、いちおうの分れ目とし、この場合の効用をゼロとする。もし1,000個からなるロットのなかに不合格品がぜんぜんなかった場合の効用を100としている。1,000個からなるロットのすべてが不合格である場合に、もしほかのものを合格にした場合の効用はマイナス200とする。

第1表

企業のとるべき選択

効 用	1ロットの中の不合格品の数
	0……………100……………1,000
決定Ⅰ：合格とする	100……………0……………-200
決定Ⅱ：不合格とする	0……………0……………0

の余地はいま二つしかないと考えられている。

他の全製品を合格品とみなして売出すか、そ

れとも不合格品とみなして研究するか二つがそれである。決定Ⅰと書いたのは合格品として売出である。決定Ⅱと書いたのは不合格品としてすべて破棄する場合である。決定Ⅱの場合に効用をすべてゼロにしているのは、この場合にはすべて破棄してしまうのであるから、企業には1円の収入も入らないからゼロの効用しか与えないと考えるのである。もちろん生産のためにコストがかかっているし、破棄するためにも費用がいる。したがって、それだけはマイナスになるのであるが、いまはその点は不問に付しておく。すでに指摘したように、ベルヌーイの効用指標は絶対的な高さをしめすものでなく、比較の基準をしめすものであるから、このように考えても誤りはおこらないはずである。

これに対して、決定Ⅰ、すなわち合格とみなして売出す場合には、種々の場合がおこってくる。もし1,000個のロットのうちすべてが不合格であるような場合に、他のすべての製品が合格であるという結論を下して売出したとすれば、その場合には、この罐詰を買った消費者あるいは、それを販売している卸売業者、小売業者などから、つぎつぎに苦情が出てくるかも知れない。罐詰を生産している会社の信用はがた落ちになって、将来に対して大きな損害をこうむる

かも知れない。したがって、この場合にはマイナス 200 というきびしい点をつけたのである。また、1,000 個のロットのなかで、ぜんぜん不合格品が存在しない場合に売出したとしよう。売上げによる収入は得られるし、苦情も起ってこないと考えられるから、100 という効用をつけておくのである。1,000 個のうち 100 個ぐらいが不合格品であるような場合は、このちょうど中間であり、いまかりにその効用をゼロとしたのである。いうまでもなく、不合格品の数が 0 から 1, 2, 3, 4 としだいに多くなるにしたがって、効用の値は 100 から順次に低い値に移って行くわけである。不合格品が 100 個をこえたところから効用の値はマイナスの値を取り、最後にマイナス 200 の水準まで下がると考えるわけである。

問題の困難さは、いままたま取出したロットが全体を完全に代表しているか否かを保証し得ない点にある。このロットの中に 100 個の不合格品があったからといって、他のロットがすべてこのような構成になっているとはいえない。また、かりにこのロットに不合格品がなかったからといって、他のロットにもすべて不合格品がないとは断言できない。そこでまずこの一つのロットを取出して、それによって、他の多くのロットを推定し得る確率について考えてみる必要がある。

2. 2. 問題の性格

問題を記号化してしめしておくのが、推論を進めるにあたって便利であろう。

ロットを構成している罐詰の数（標本）を N であらわす。いまの例では $N = 1,000$ である。このロットを抜き出して検査をしたところ不合格品が n 個あったとする。この n は 1 から 1,000 のあいだの整数になる。さて N からなるロットに n 個の不合格品があった場合に、全部の製品を適格とみなした場合の効用を u_n であらわす。さきにかかげた仮設例では、 $u_0 = 100$, $u_{100} = 0$, $u_{1,000} = -200$ というのであった。

さて N 個からなるロットのうち n 個の不合格品のあらわれる確率は二項分布

$$P_n = \binom{N}{n} p^n (1-p)^{N-n}$$

によって求められる。したがってこの場合、全部を合格とする場合および全部を不合格とする場合の、それぞれの効用の期望値は、

$$u(\text{合格}) = \sum_{n=0}^N u_n \binom{N}{n} p^n (1-p)^{N-n} = \sum_{n=0}^N u_n P_n$$

$$u(\text{不合格}) = 0$$

としてしめすことが出来る。決定のルールはこの二つの期望値のいずれが大きいかを比較して、大きい方の行動をとることになる。さきにかかげた例についていうと、 n が100のとき $u(\text{合格})$ と $u(\text{不合格})$ が等しくなる。 u が100以下のときは $u(\text{合格})$ の方が $u(\text{不合格})$ より値が大で、合格の判定を下すことが有利であるとみられる。

この簡単な例を一般化して考えてみよう。効用は決定を下す人の主観的判断をかなりの程度反映したものであるが、さきにベルヌーイの効用指標について述べたような性質をみたすものでなければならない。さて、種々の可能性のあらわれる確率分布を想定し得れば、それから効用の数学的期望値を求め得るはずであるし、またその数学的期望値の比較によって決定の基準が得られるはずである。

2. 3. ベイズの定理

この問題を一般的に拡張して解くために考えられたのが、ベイズの決定関数 (Bayes' decision function) である。これがベイズの決定関数といわれるのは、トーマス・ベイズ (Thomas Bayes) が1763年に考え出したベイズの定理というのを利用するからである。ベイズの定理というのは、初等的な確率論の教科書にも出ており、すでによく知られているが、まずそれを簡単に説明しておいたほうがベイズの決定関数を理解するのに便利であろう。

例をあけて説明しよう。ヒズミのないサイコロを投げて1の目が出る確率は、いうまでもなく1/6である。いま投げられたサイの目が奇数であることがわかっており、ただそれが1であるか3であるか5であるかはわからないとす

る。この場合 1 である確率というまでもなく $1/3$ である。ベイズの定理はこのようにある条件が与えられていて、そのもとでの確率と、条件の存在しない場合における確率の関係について明らかにしたものである。条件つき確率(conditional probability)というのは、当面の例でいえば奇数が出ているという条件のもとで、1 である確率ということである。一般に θ という条件が与えられていて、 x という事象の現われる確率は $p(x|\theta)$ とあらわす。一般にサイを振って奇数が出る確率は $q(\theta)$ であらわす。この場合、つぎのような関係になることはよく知られている。

$$p(x|\theta) \cdot q(\theta) = r(x, \theta)$$

ここで $r(x, \theta)$ は、 x および θ が同時に現われる確率であり、いまの例でいうと、奇数であり同時に 1 である確率である。実際の数字を入れればただちに明らかである。 $r(x, \theta)$ は前述のように $1/6$ 、 $q(\theta)$ は奇数の現われる確率であり $1/2$ 、 $p(x|\theta)$ は奇数であることがわかっていて 1 である確率、すなわち $1/3$ である。したがって、

$$\frac{1}{3} \times \frac{1}{2} = \frac{1}{6}$$

となる。同じようなことはクジ引きの例で考えてもよい。10,000 枚の抽選券が発行されているとして、まず $1/10$ の人に第 1 次の賞品を与える。さらにその第 1 次の賞品を得た人を対象にして第 2 回の抽選を行ない、 $1/50$ の人に第 2 次の賞品を与えるというルールである。したがって最終当選者は $1/500$ すなわち 20 人である。この場合、 $p(x|\theta)$ は $1/50$ 、 $q(\theta)$ は $1/10$ 、 $r(x, \theta)$ すなわち第 1 次および第 2 次の抽選にいずれも当選する確率は、その積 $1/500$ である。

ベイズはこの簡単な事柄を、まず $1/50$ を選び出し、それからさらに $1/10$ を選んでも、結局は $1/500$ の最終当選者が決まるという点に着目するのである。最初に抽選で $1/50$ 選び出す確率は $t(x)$ とし、 x が決まったうえでさらに θ を選び出す条件つき確率を $s(\theta|x)$ とすると、やはり、

$$s(\theta|x) \cdot t(x) = r(x, \theta)$$

となる。これを組合せて、

$$p(x|\theta) \cdot q(\theta) = r(x, \theta) = s(\theta|x) \cdot t(x)$$

としたものがベイズの定理である。こゝる定理をたくみに利用することによって決定関数を作ったから、ベイズの決定関数という名称がうまれたのである。

2. 4. ベイズの決定関数

さきにあげた抜取り検査の例では、決定は合格とみるか不合格とみるか、二つの選択しかなかった。しかし実際には選択の対象が二種しか存在しない場合はむしろまれであり、多数の可能性のなかから選び出すことのほうが多い。いまその多数の可能性を A とし、そのなかから選び出す決定を a とする。記号的には、 $a \in A$ と書く。これは集合論でしばしば使われる記号法で、「 a は A にふくまれる」とよむ。つぎに決定をくださる人に与えられた知識を X であらわす。これは通常、標本(サンプル)とか、観察値などといわれるものである。さきの抜取り検査の例でいうと、標本として取り出したロットのなかの不合格品がこの x にあたる。さてこの x はなんらかの意味でその標本をふくむ全体の集団を代表しているものである、と考えるのであるが、そのもとの集団を母集団(population)といい、その母集団のもっている性格を θ であらわす。まえの例に即して言えば、全体の製品が合格であるか否かの性質である。

x はなんらかの意味で θ を代表しているという点は、別の言葉でいえば母集団に θ の性質があり、たまたま標本に x という観察値として現われたと考えればよい。いうまでもなく θ がそのまま縮小されて x になっているわけではない。 θ はわからないのであるが、しかし決定をくださる人の目につかぬところですでに存在している。その θ が x となって人々の目にふれる確率は、 $p(x|\theta)$ という条件つき確率であらわすことができる。現実には直面する問題は、 θ そのものを知らない人が、 x という観察値にもとづいて判断をくださねばならないことから発生してくる。 x にもとづいて決定 a をくださすということを、 $a = d(x)$ という形であらわす。これは a は x に依存するとよめばよい。

問題は $a = d(x)$ の具体的な形を決めることであるが、そのためには効用指

標を決めておかなければならない。ここでは取扱いの便宜上、効用の代りに損失 (loss) という概念を使う。損失はマイナスの効用であり、したがって効用の極大を求める代りに、損失の極小を求めればよい。損失は $L(a, \theta)$ と定義する。いうまでもなく母集団が θ という特徴をもっているときに a という決定を行なうのであるから、損失の大きさは θ と a の双方に依存することになる。

いままでの点を要約して記号を一覧表の形で記しておこう。

決 定	$a \in A$	(A から a をえらぶ)
パラメーター	θ	(母集団の特徴)
損 失	$L(a, \theta)$	(マイナスの効用)
標本もしくは観察値	x	
x の現われる確率	$p(x \theta)$	
決定関数	$a = d(x)$	(x にもとづく a の決定)

これだけのことを念頭において、もう一度整理すると、 $a = d(x)$ という形で決定を行なうと、 $L(a, \theta)$ の損失が起こることになる。ところが x は $p(x|\theta)$ という分布をもって散らばっているのであるから、 L の数学的期望値を求めて、その極小をはかればよい。 L の数学的期望値は危険 (risk) といわれる。要するに損失の平均値であり、保険会社が危険とよんでいる概念とまったくおなじものである。

危険を R であらわすと、

$$R[d(x), \theta] = \int L[d(x), \theta] \cdot p(x|\theta) dx$$

となる。積分記号を使っているが、これは期望値計算の場合の Σ と事実上おなじことであるが、ただいまの場合に x の分布に連続性を想定しているから、積分になったのである。単純に総和と考えておけばよい。なお R および L のカッコのなかで、 a を書くべき部分を $d(x)$ としてあるが、これは $a = d(x)$ の関係を代入したものである。

さて、つぎに θ の分布は $q(\theta)$ という確率で示されると考える。そうすると危険 R は θ の分布にしたがって分布していることになり、そこから R の期望値

を定義することができる。

$$\int R[d(x), \theta] q(\theta) d\theta$$

がそれである。 R の定義式をこれに代入すると、

$$\int R[d(x), \theta] q(\theta) d\theta = \iint L[d(x), \theta] p(x|\theta) q(\theta) dx d\theta$$

になる。ベイズの定理が登場するのはこの点においてである。

$$p(x|\theta) q(\theta) = r(x, \theta) = s(\theta|x) t(x)$$

が、そのベイズの定理であった。これを代入すると、

$$\int R[d(x), \theta] q(\theta) d\theta = \iint L[d(x), \theta] s(\theta|x) t(x) d\theta dx$$

となる。これを極小にすればよいのである。

こんな複雑な式の極小はどうすれば求められるか、途方にくれる人も少なくないであろう。しかし実際は見かけよりはるかに簡単なのである。この式の右辺をもうすこし変形すれば、かなり単純化されるからである。実際にやってみよう。

$$\begin{aligned} & \iint L[d(x), \theta] s(\theta|x) t(x) d\theta dx \\ &= \int t(x) \{ \int L[d(x), \theta] s(\theta|x) d\theta \} dx \end{aligned}$$

x はわかっているのであるから、それにだけ関係している部分を右側および左側にくり出したのである。それに $t(x)$ は確率であり、正の値をとことはわかっている。したがって危険の期望値を極小にするのは、まん中に残された

$$\int L[d(x), \theta] s(\theta|x) d\theta$$

を極小にすればよい。これが極小になるように決定 α すなわち $d(x)$ を選ばせようのである。こうして選ばれた $d(x)$ の形がベイズの決定関数といわれるものである。

2. 5. 点推計の問題

たとえば x が $p(x|\theta)$ という分布をもっているときに、母集団のパラメータ θ についての推定を行なうという問題におきかえて考えてみよう。これは典

型的な点推計 (point estimation) の問題である。この場合、かりに誤差の自乗 $[d(x) - \theta]^2$ が損失であると考えたとする。これはパラメーターの真の値は θ であるにもかかわらず、 $d(x)$ という推計を行なったことから生ずる損失である。

推計についての危険関数 (risk function) は

$$R[d(x), \theta] = E[d(x) - \theta]^2 = \int [d(x) - \theta]^2 p(x|\theta) dx$$

という形になる。この $R[d(x), \theta]$ を極小にするような推計値 ($d(x)$) をえることが、点推計の問題であるといつてよい。

点推計の問題はこのように設定された統計的決定の問題にただちにそのまま対応する。まず明らかなことは、点推計は統計的決定の特殊ケースであるということである。どのような特殊ケースであるかについて、いわゆる heuristic method を利用して明らかにしていくことにしようと思う。heuristic method というのは、厳密な定義や証明を行なうのではなく、若干のケースを考え、結局どういふものであるかの見当をつける方法であり、一種の試行錯誤の過程を通じて真実を発見する方法と考えてよい。

かりに、統計的決定の問題を決定空間 D の性質によって分類するとしよう。その一つの方法として、 D は実数からなる集合である場合を考えてみる。そのような決定空間における決定の問題は点集合の問題に帰着するかどうかを考えようというのである。ただちに明かなように、このような決定空間における決定は点集合のみならず、たとえば仮設検定 (testing hypothesis) をもふくむことになる。

いま決定空間を $D = (0, 1)$ とし、0 を仮設 H を受入れる決定をあらわすものとし、1 を H にかわる仮設 H' を受入れる決定をあらわすものと考えれば、仮設検定もまた決定空間 D が実数のみからなる決定にふくまれることは明らかである。

つぎに、決定空間 D を実数の区間 (interval of real numbers) である場合と考え、そのような空間における決定が点推計であると、かりに考えてみよう。この場合にも不都合はおこる。空間をこう定義すると、ほとんどの点推計の問題

はふくまれるし、また仮設検査の問題が入りこむことはないが、さきに例にしたような、あたえられたロットの中の不合格品を推定するような問題は、このように定義された問題からは逸脱することになる。

このように見てくると、決定空間を限定することによって点推計を統計的決定関数の特殊ケースとして定義することは不可能であったようである。もう一つの接近方法は損失関数 L をどう限定するかという角度からである。

- (1) 損失関数の定め方は正確にというよりは、なるべく近い値を求めるという方針で定めねばならない。

まずこの点の説明をしよう。ここにいうことを別の表現ですれば、われわれが求めるものは確率変数であるといってもよい。仮設検定の問題とするのは確率変数ではない。しかし点推計の問題とするのは、一定の確率分布をもち、したがって標準偏差をもつ平均値あるいは中位数であると考えればよい。

- (2) 損失関数の対称性について、より多くの評価をあたえる。

たとえば仮設検定の場合、損失関数の非対称性を考えるのが普通である。たとえばコロンブスの例をもう一度想起していただきたい。近くに大陸があるという仮設を受入れるか、それとも逆に近くに大陸はないという仮設を受入れるかどうかによって損失関数の値はまったく非対称的である。これに対して、点推計の問題は通常の場合、推計値が上に誤差を生じるときと、下に誤差を生じるときと、それにともなう損失はおなじであるという形で処理する。もちろん、そのような処理がつねによりかどうかには問題がある。たとえば需要予測を誤って、生産を多くしすぎた場合と、生産を少くしすぎた場合とでは、それにともなう損失に差異があり得る。しかし点推計そのものとしては、このような損失の非対称性は考慮のそとにおくことになっている。

- (3) 損失関数のきめ方はある意味で恣意的である。

点推計の損失関数として $[d(x) - \theta]^2$ というように誤差の自乗を使うのが通例になっている。これはガウス以来¹⁾、一般に広くとられている方法であるが、

1) C. F. Gauss, *Theorie den kleinsten Fahren unterworfenen Kombination der Beobachtungen*, 1821, にこの方法が提唱されているといわれている。

これでなければならないという論理的必然性はない。

ガウスの考え方を要約しておこう。いまなにか（たとえば実験結果の測定）の測定誤差が x であるとする。何回か測定をくりかえしたとして誤差の分布は $\phi(x)$ という density function でしめされたとする。この場合、誤差の期待値は $\int x\phi(x)dx$ である。ガウスが提唱し、現在もっとも普通に使われている不確定性の測度は $\int x^2\phi(x)dx$ であり、この $\int x^2\phi(x)dx$ の値の大小によって測定の正確性あるいは不正確さの度合いを知ろうとするのであるが、しかしこの測度はある意味ではかなり恣意的である。 x の絶対値の積分をとってもよいし、 x^4 の積分をとってもさしつかえないはずである、これでなければ困るという絶対的な規準がないから、そのうち最も便利な x^2 の積分という方法がとられるだけのことである。

2. 6. 点推計

さて点推計の問題であるが、もっとも広く使われる正規分布について考える。いま

$$N(\mu, \sigma)$$

で平均 μ 、標準偏差 σ の正規分布をあらわすものとする。この場合

標	本	X
行	動	$\hat{\mu}$ (実数)
損	失	$L(\hat{\mu}, \mu) = (\hat{\mu} - \mu)^2$

とおけばよい。損失関数をこういう形におく理由は前節で説明したから、くりかえさない。なお便宜上、ここでの分布は $N(\mu, \sigma)$ とする。記号上の混乱をさけるために一言すると、前節で θ としたものが、ここでは平均をしめす記号として広くもちいられている μ にかわったのであり、 $\wedge$ の記号は、これも慣例にしたがって測定をあらわすものとして使ったのである。したがって $L(\hat{\mu}, \mu)$ は実際の（母集団の）平均値が μ であるにもかかわらず、標本からそれを $\hat{\mu}$ と推定したことによる損失と考えればよい。まず $\sigma=1$ の場合から考えよう。

標本 X は $N(\mu, 1)$ という分布をもつから

$$p(x|\mu) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2}}$$

$$q(\mu) = \frac{1}{\sqrt{2\pi}\sigma_\mu} e^{-\frac{(\mu-\bar{\mu})^2}{2\sigma_\mu^2}}$$

となる。したがって問題は

$$\int [\hat{\mu}(x) - \mu]^2 \frac{1}{2\pi} e^{-\frac{(x-\mu)^2}{2}} \cdot \frac{1}{\sqrt{2\pi}\sigma_\mu} e^{-\frac{(\mu-\bar{\mu})^2}{2\sigma_\mu^2}} d\mu$$

を極小にするように $\hat{\mu}(x)$ をえらぶという形に帰着する。

手順として目的関数の簡略化をはからねばならない。

$$\begin{aligned} & \int [\hat{\mu}(x - \mu)]^2 \frac{1}{2\pi} e^{-\frac{(x-\mu)^2}{2}} \cdot \frac{1}{\sqrt{2\pi}\sigma_\mu} e^{-\frac{(\mu-\bar{\mu})^2}{2\sigma_\mu^2}} d\mu \\ &= \frac{1}{2\pi\sigma_\mu} \int (\hat{\mu} - \mu)^2 e^{-\frac{\sigma_\mu^2(x-\mu)^2 + (\mu-\bar{\mu})^2}{2\sigma_\mu^2}} d\mu \\ &= \frac{1}{2\pi\sigma_\mu} \int (\hat{\mu} - \mu)^2 e^{-\frac{\sigma_\mu^2(x-\mu)^2 + (\mu-\bar{\mu})^2}{2\sigma_\mu^2}} d\mu \\ &= \frac{1}{2\pi\sigma_\mu} \int (\hat{\mu} - \mu)^2 e^{-\frac{(1+\sigma_\mu^2)(\mu-\bar{\mu})^2 + (1+\sigma_\mu^2)K}{2\sigma_\mu^2}} d\mu \end{aligned}$$

$$\text{ここで } \bar{\mu} = \frac{\sigma_\mu^2 + \bar{\mu}}{1 + \sigma_\mu^2} \text{ である。}$$

$$\begin{aligned} &= \frac{1}{2\pi\sigma_\mu} \int (\hat{\mu} - \mu)^2 e^{-\frac{(1+\sigma_\mu^2)(\mu-\bar{\mu})^2}{2\sigma_\mu^2}} \cdot e^{-\frac{(1+\sigma_\mu^2)K}{2\sigma_\mu^2}} d\mu \\ &= \frac{1}{2\pi\sigma_\mu} e^{-\frac{(1+\sigma_\mu^2)K}{2\sigma_\mu^2}} \int (\hat{\mu} - \mu)^2 e^{-\frac{(1+\sigma_\mu^2)(\mu-\bar{\mu})^2}{2\sigma_\mu^2}} d\mu \\ &= K' \int (\hat{\mu} - \mu)^2 e^{-\frac{(\mu-\bar{\mu})^2}{2\sigma_\mu^2}} d\mu \end{aligned}$$

ここで

$$K' = \frac{1}{2\pi\sigma_\mu} e^{-\frac{(1+\sigma_\mu^2)K}{2\sigma_\mu^2}} > 0$$

$$\bar{\sigma}_\mu^2 = \frac{\sigma_\mu^2}{1 + \sigma_\mu^2}.$$

$$\begin{aligned} 2) \quad & \sigma_\mu^2(x-\mu)^2 + (\mu-\bar{\mu})^2 = \sigma_\mu^2 x^2 - 2\sigma_\mu^2 \mu x + \sigma_\mu^2 \mu^2 + \mu^2 - 2\mu\bar{\mu} + \bar{\mu}^2 \\ &= (1+\sigma_\mu^2)\mu^2 - 2(\sigma_\mu^2 x + \bar{\mu})\mu + \sigma_\mu^2 x^2 + \bar{\mu}^2 \\ &= (1+\sigma_\mu^2) \left\{ \mu^2 - 2 \left(\frac{\sigma_\mu^2 x + \bar{\mu}}{1+\sigma_\mu^2} \right) \mu + \frac{\sigma_\mu^2 x^2 + \bar{\mu}^2}{1+\sigma_\mu^2} \right\} \\ &= (1+\sigma_\mu^2) \left\{ \left(\mu^2 - \frac{\sigma_\mu^2 x + \bar{\mu}^2}{1+\sigma_\mu^2} \right) + K(\bar{\mu}, 1) \right\} \end{aligned}$$

したがって点推計の問題はいちじるしく簡略化される。

$$-\frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} (\hat{\mu} - \mu)^2 e^{-\frac{(\mu - \hat{\mu})^2}{2\sigma^2}} d\mu$$

を極小にすればよいのである。それには $\hat{\mu}$ で微分してゼロに等しいとおけばよい。

$$-\frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} 2(\hat{\mu} - \mu)^2 e^{-\frac{(\mu - \hat{\mu})^2}{2\sigma^2}} d\mu = 0$$

$$\therefore -\frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} \hat{\mu} e^{-\frac{(\mu - \hat{\mu})^2}{2\sigma^2}} d\mu = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} \mu e^{-\frac{(\mu - \hat{\mu})^2}{2\sigma^2}} d\mu$$

ところで

$$-\frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} e^{-\frac{(\mu - \hat{\mu})^2}{2\sigma^2}} d\mu = 1$$

であるから

$\hat{\mu} = \mu$ の期望値

$$= \bar{\mu} = \frac{\sigma_{\mu}^2 X + \bar{\mu}}{1 + \sigma_{\mu}^2} = \frac{X + \frac{1}{\sigma_{\mu}^2} \bar{\mu}}{1 + \frac{1}{\sigma_{\mu}^2}}$$

になる。

ただ、いままでのところでは標準偏差を 1 と考えていた。いまそれを σ_x であると考えればどうなるか。この一般の場合について検討しよう。この場合、

$$p(x_i|\mu) = \frac{1}{\sqrt{2\pi}\sigma_x} e^{-\frac{(x_i - \mu)^2}{2\sigma_x^2}}, \quad (i=1, 2, \dots, N)$$

$$q(\mu) = \frac{1}{\sqrt{2\pi}\sigma_{\mu}} e^{-\frac{(\mu - \bar{\mu})^2}{2\sigma_{\mu}^2}}$$

$$L = (\hat{\mu} - \mu)^2$$

となる。ここで $X = (x_1, x_2, \dots, x_N)$ は標本の観察値で、 x_i はいずれも相互に独立であると仮定する。そうすると

$$\begin{aligned} p(X|\mu) &= \prod_{i=1}^N \frac{1}{\sqrt{2\pi}\sigma_x} e^{-\frac{(x_i - \mu)^2}{2\sigma_x^2}} \\ &= \left(\frac{1}{\sqrt{2\pi}\sigma_x} \right)^N \cdot e^{-\frac{\sum_{i=1}^N (x_i - \mu)^2}{2\sigma_x^2}} \end{aligned}$$

また

$$x_i - \mu = (x_i - \bar{x}) + (\bar{x} - \mu),$$

ここで $\bar{x} = \frac{\sum_{i=1}^N x_i}{N}$ すなわち x の平均であるから、

$$\begin{aligned} \sum_{i=1}^N (x_i - \mu)^2 &= \sum_{i=1}^N [(x_i - \bar{x}) + (\bar{x} - \mu)]^2 \\ &= \sum_{i=1}^N [(x_i - \bar{x})^2 + 2(x_i - \bar{x})(\bar{x} - \mu) + (\bar{x} - \mu)^2] \\ &= \sum_{i=1}^N (x_i - \bar{x})^2 + 2\sum_{i=1}^N (x_i - \bar{x})(\bar{x} - \mu) + \sum_{i=1}^N (\bar{x} - \mu)^2 \\ &= \sum_{i=1}^N (x_i - \bar{x})^2 + 0 + N(\bar{x} - \mu)^2 \quad 3) \end{aligned}$$

これらの点を考慮すれば

$$p(x|\mu) = \left(\frac{1}{\sqrt{2\pi}\sigma_x} \right)^N e^{-\frac{\sum_{i=1}^N (x_i - \bar{x})^2}{2\sigma_x^2}} \cdot e^{-\frac{N(\bar{x} - \mu)^2}{2\sigma_x^2}}$$

となる。したがって問題は

$$\begin{aligned} &\int (\hat{\mu} - \mu)^2 p(x|\mu) q(\mu) d\mu \\ &= \int (\hat{\mu} - \mu)^2 K e^{-\frac{(\bar{x} - \mu)^2}{2\sigma_x^2 N}} \cdot \frac{1}{\sqrt{2\pi}\sigma_\mu} e^{-\frac{(\mu - \bar{\mu})^2}{2\sigma_\mu^2}} d\mu \end{aligned}$$

の極小を求めることに帰着する。ここで

$$K = \left(\frac{1}{\sqrt{2\pi}\sigma_x} \right)^N e^{-\frac{\sum_{i=1}^N (x_i - \bar{x})^2}{2\sigma_x^2}}$$

であることはいうまでもない。

(未完)

3) ここで

$$\sum_{i=1}^N (x_i - \bar{x})(\bar{x} - \mu) = (\bar{x} - \mu) \sum_{i=1}^N (x_i - \bar{x})$$

であり

$$\sum_{i=1}^N (x_i - \bar{x}) = 0.$$