或Order Statisticの問題に就いて

小川, 潤次郎
統計數理研究所

出版情報：統計数理研究. 2 (1), pp.33-35, 1948-05-10. Research Association of Statistical Sciences

Borel 集合族として定められるから，從つて又 $\mathbf{B}^*$ は E^* の作る集合體を含む最小の正規集合族として特徴附けられるから（Saks [1] 參照），一般の $\mathbf{B}^*$ に屬する集合 B^* に對しても $T^*B^* \in \mathbf{B}^*$ となることが證明される。T^{*-1} の可測性も (20), (21) により，(19) を用いて全く同樣にして證明される。

次に T^* が保測變換であることも，全く同一に，基本集合 E^* に對して

$$m^*(E^*)=m^*(T^*E^*)$$

となることを見れば十分である。(22) に對して

$$m^*(E^*)=m(A)\cdot\prod_{j=1}^{h}\mu'(M_j)$$

であり，他方 T^*E^* に對しては Fubini の定理を用いれば

$$\begin{aligned} m^*(T^*E^*) &= \prod_{j=2}^{h}\mu'(M_j)\times(m\times\mu')(A^*) \\ &= \prod_{j=2}^{h}\mu'(M_j)\cdot\int_{M_1} m(T_lA)\mu'(dl) \\ &= \prod_{2}^{h}\mu'(M_j)\cdot m(A)\mu'(M_1) \\ &= m^*(E^*) \end{aligned}$$

となるからである。

（ハ） T^* が X^* 上の保測變換であるから，$f^*(x^*)=f(x)$ とおけば，f^* は X^* 上 m^* 可積分函數であるから，G. D. Birkhoff のエルゴード定理によつて，ある m^* 可積分函數 $F(x^*)$ に對して

$$(24)\quad \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}f^*(T^{*i}x^*)=F(x^*)$$

が X^* 上殆んどすべての x^* に對して存在する。從つて Fubini の定理によつて，殆んどすべての $k\in K$ に對して

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}f^*(T^{*i}x^*)=F_k(x)$$

が，殆んどすべての x に對して存在する。一方 $F(x^*)$ は m^* 可積分であるから，殆んどすべての k に對して $F_k(x)$ は m 可積分となる。これを組合せて，定理 1 が證明せられた。(未完)

引 用 文 獻

E. Hopf [1], Ergodentheorie (Ergebnisse), Julius Springer, Berlin, 1937.

吉田耕作 [1], エルゴード諸定理, 數物會誌, 15卷 1 號 (1939).

S. Saks [1], Theory of integral. Warsaw, 1937.

或 Order Statistic の問題に就いて

小 川 潤 次 郎

統計數理研究所

(昭和 22 年 12 月 1 日日本統計學會講演　昭和 22 年 12 月 20 日受理)

増山元三郎氏よりペニシリンを精製するカビの菌を見出す問題に關聯して，次の如き問題を提出されたことがある。即ち

ある母集團（勿論これは連續な密度函數を有するとする）より先づ第一の無作意標本 Σ_1: $X_1<X_2<\cdots\cdots<X_n$ を抽出し，その最小値 X_1，最大値 X_n とする。次に前の標本と獨立に，第二の無作意標本 Σ_2: $X_1'<X_2'<\cdots\cdots<X'_N$ を抽出し，その最小値 X_1'，最大値 X'_N とするとき，$X_N'-X_n$ の分布如何。

此問題は例へば母集團を正規母集團と假定すれば X_n 及び X'_N の分布は分つて居り，從つてその疊み込みとして X'_N-X_n の分布函數は出る筈であるが一般にはそれは未知なる母數に關係し，又その分布函數を陽に求めることは困難なので次の如く Order Statistic の問題に轉化して考へることにする。即ち

Σ_2 の最大値 X'_N が Σ_1 の最大値 X_n を越える確率如何。

問題をこのやうに轉化すれば，その答は母集團の分布とは無關係に簡單になる。又例へば，工科方面の問題でピアノ線の強度試驗[1] 等の問題では，強度の平均値よりは，實際に幾つかの試驗片を試驗して，その結果得られた知識をより有效に用ひる必要がある。それには母集團の即ち考察の對象となるピアノ線全長の最小強度がどの位になるか，又は實際の試驗から得られた最大，最小強度の間に全ピアノ線の何パーセントが含まれるであらうかと云ふ風に考へるのがより合理的であらう。

以下の方法は，このやうな問題に對する一つの approach である。

先づ第一の問題から考へる。母集團の確率密度を $f(x)$ として，次の如く Order statistic の問題と考へる。今[2]

$$1 \leq r_1 < r_2 < \cdots\cdots < r_k \leq n$$

なる正の整數列を取り，これに對應して任意の實數列

$$x_{r_1} < x_{r_2} < \cdots\cdots < x_{r_k}$$

を一つ定めたとき，第一の無作意標本 Σ_1 を大きさの順に並べて

$$X_1, \cdots\cdots, X_{r_1-1}, X_{r_1}, X_{r_1+1}, \cdots\cdots, X_{r_2-1}, X_{r_2}, X_{r_2+1}, \cdots\cdots, X_{r_k-1}, X_{r_k}, X_{r_k+1}, \cdots\cdots, X_n$$

であつて

$$(*) \quad x_{r_1} < X_{r_1} < x_{r_1} + dx_{r_1}, \quad x_{r_2} < X_{r_2} < x_{r_2} + dx_{r_2}, \cdots\cdots, x_{r_k} < X_{r_k} < x_{r_k} + dx_{r_k}$$

となる確率 P_1 を求めよう。

$$\int_{-\infty}^{x_{r_1}} f(x)dx = u_1, \int_{x_{r_1}}^{x_{r_2}} f(x)dx = u_2, \cdots\cdots,$$

$$\int_{x_{r_{k-1}}}^{x_{r_k}} f(x)dx = u_k, \int_{x_{r_k}}^{\infty} f(x)dx = u_{k+1}$$

とおくと，勿論

$$u_1 + u_2 + \cdots\cdots + u_k + u_{k+1} = 1$$

であつて，且つ

$$du_1 = f(x_{r_1})dx_{r_1}, \quad du_2 = f(x_{r_2})dx_{r_2}, \cdots\cdots, \quad du_k = f(x_{r_k})dx_{r_k}$$

求める確率 P_1 は多項分布に依つて，

$$(1) \quad P_1 = \frac{n!}{(r_1-1)!(r_2-r_1-1)!\cdots\cdots(r_k-r_{k-1}-1)!(n-r_k)!} \times u_1^{r_1-1} u_2^{r_2-r_1-1} \cdots\cdots u_k^{r_k-r_{k-1}-1} u_{k+1}^{n-r_k} du_1 \cdots\cdots du_k$$

となる。

次に，第二の無作意標本 Σ_2 が，この定められた實數列

$$x_{r_1}, x_{r_2}, \cdots\cdots x_{r_k}$$

に關して，區間

$$(-\infty, x_{r_1}), (x_{r_1}, x_{r_2}), \cdots\cdots, (x_{r_{k-1}}, x_{r_k}), (x_{r_k}, \infty)$$

に夫々，

$$N_1, N_2, \cdots\cdots N_k, N_{k+1} \qquad (**)$$

個づゝ入る確率を P_{II} とする。勿論

$$\sum_{i=1}^{k+1} N_i = N$$

である。

$$(2) \quad P_{II} = \frac{N!}{N_1!\, N_2! \cdots\cdots N_{k+1}!} u_1^{N_1} u_2^{N_2} \cdots u_k^{N_k} u_{k+1}^{N_{k+1}}$$

依つて，一つの定められた實數列 $x_{r_1}, x_{r_2}, \cdots\cdots, x_{r_k}$ に關して，第一の無作意標本 Σ_1 が (*) なる條件を滿たし，同時に第二の無作意標本 Σ_2 が (**) なる條件を滿たす確率 P' は (1) と (2) の積で與へられる。

$$(3) \quad P' = \frac{n!\, N!}{(r_1-1)!(r_2-r_1-1)!\cdots(r_k-r_{k-1}-1)!\,(n-r_k)!\, N_1!\cdots\cdots N_{k+1}!} \times u_1^{N_1+r_1-1} u_2^{N_2+r_2-r_1-1} \cdots\cdots \times u_k^{N_k+r_k-r_{k-1}-1} u_{k+1}^{N_{k+1}+n-r_k} du_1 \cdots\cdots du_k$$

從つて，これを $x_{r_1}, x_{r_2}, \cdots\cdots, x_{r_k}$ に關して積分すれば $N_1, N_2, \cdots\cdots, N_{k-1} (= N - N_1 - \cdots\cdots - N_k)$ の同時分布が得られるのであるがそれには $u_1, u_2, \cdots\cdots, u_k$ に關して

$$u_1 \geq 0, u_2 > 0, \cdots\cdots, u_k \geq 0,$$

$$u_{k+1} = 1 - u_1 - u_2 \cdots\cdots - u_k \geq 0.$$

なる範圍で積分すればよい。この範圍を D で表はすと，これはよく知られた Dirichlet 積分[3] であつて

$$\int_D \cdots\cdots \int u_1^{N_1+r_1-1} \cdots\cdots u_k^{N_k+r_k-r_{k+1}-1} (1-u_1-u_2\cdots\cdots u_k)^{N_{k+1}+n-r_k} du_1 \cdots\cdots du_k$$

$$= \frac{\Gamma(N_1+r_1)\,\Gamma(N_2+r_2-r_1) \cdots\cdots \Gamma(N_k+r_k-r_{k-1}) \times \Gamma(N_{k+1}+n-r_k+1)}{\Gamma(N+n+1)}$$

$$= \frac{(N_1+r_1-1)!(N_2+r_2-r_1-1)!\cdots\cdots \times (N_k+r_k-r_{k-1}-1)!\,(N_{k+1}+n-r_k)!}{(N+n)!}$$

從つて求むる $N_1, N_2, \cdots\cdots, N_{k+1}$ の同時分布は

$$(4) \quad P = \frac{n!\,N!(N_1+r_1-1)!\,(N_2+r_2-r_1-1)!\cdots\cdots \times (N_k+r_k-r_{k-1}-1)!(N_{k+1}+n-r_k)!}{(r_1-1)!(r_2-r_1-1)!\cdots\cdots(r_k-r_{k-1}-1)! \times (n-r_k)!N_1!N_2!\cdots\cdots N_{k+1}!(n+N)!}$$

となる。

上の結果を第一の増山氏の問題の場合に特殊化するには

$$k=1, \quad r_1 = n, \quad x_{r_1} = x_n$$

$$N_1 + N_2 = N$$

として，N_1, N_2 の同時分布を求めると，それは (4) より

$$(5) \quad \frac{n!\,N!(N_1+n-1)!\,N_2!}{(n-1)!\,N_1!\,N_2!\,(N+n)!} = \frac{n}{N+n} \frac{{}_N C_{N_2}}{{}_{N+n-1}C_{N_2}}$$

となる。從つて X'_N が X_n を越す確率 Q は次式で與へられる。

$$(6) \quad Q = \frac{n}{N+n} \sum_{r=1}^{N} \frac{{}_N C_r}{{}_{N+n-1}C_r} = \frac{N}{N+n} \text{[4]}$$

若し特に $N=n$ ならば，$Q=1/2$ でこれは常識に合致

する。

次にピアノ線の强度試驗の場合を考へよう。今問題にする一本のピアノ線は統計的に管理された製品であるとする。即ち，これから任意に切取られた試驗片(Test piece) の强度の差異は偶然的な變動であると考へ得るとする。このやうな場合に幾個かの試驗片を强度試驗をしたとき，その最小强度を知つたとき，全ピアノ線の强度に關して吾々の得る客觀的な知識は如何と云ふ問題を考へよう。

今試驗片の長さ l cm，ピアノ線の全長を L cm とし，$\left[\frac{L}{l}\right]$[5] $=N$ としよう。

今 n 個の試驗片を强度試驗した結果，吾々はその最小强度が X_1 であつたことを知つたとする。この時 N 個の試驗片の內少くとも N_0 個が X_1 より大なる强度を有する確率 R を求めよう。その爲には (4) 式で

$$k=1,\ r_1=1,\ x_{r_1}=x_1$$

$$N_1=N-\nu,\ N_2=\nu,\quad (\nu\geqq N_0)$$

とおけばよいから，求むる確率 R は

$$(7)\quad R=\sum_{\nu=N_0}^{N}\frac{n!N!(N-\nu)!(\nu+n-1)!}{(n-1)!(N-\nu)!\nu!(n+N)!}$$

$$=\frac{n\times N!}{(n+N)!}\sum_{\nu=N_0}^{N}\frac{(r+n-1)!}{\nu!}$$

そこで例へば

$$\frac{N_0}{N}\geqq 99\%$$

である如く N_0 を與へたとき

$$R\geqq\alpha(=0\ 99)$$

となる如く n を定めるか又は n, N が與へられたとき

$$P\geqq\alpha(=0.99)$$

である如く N_0 を求めるかである。

前者は，n 個の試驗片の强度試驗をして，その最小强度を知つたときピアノ線の全長の 99% がこの最小强度より大である確率が 0.99 以上である如く，n を定めると云ふのである。若し此處に何本かのピアノ線があつて，それがある一定の管理水準の製品なることが分つてゐるならば，その內から任意に上記 n 個の試驗片を取つて强度試驗をすれば，それから，吾々は 1% の危險率を以つて，その 99% は，この試驗片の最小强度より大なる强度を有すると結論出來るのである。

以上述べたことはベニシリンの場合にも，ピアノ線の場合にも未だ大なる不滿があるのであつて，吾々は標本の最大，最小値の知識から母集團の最大最小値の絕對値がどの位かを知り度いのである。勿論この爲には密度函數 $f(x)$ の函數形は既知である必要があるがこれは未だ解けてゐない。

其後東大圖書館に於て，米國寄贈の The Annals of Mathematical Statistics. Vol. 13. (1942) に於て，S. S. Wilks が Statistical Prediction with Special Reference to the Problem of Talerance Limits なる論文で，このやうな考へ方を展開してゐて，表迄作製してゐることを知つた。尙ほ第二の問題に關して N が充分大なるとき，即ち

$$(8)\quad \lim_{N\to\infty}\frac{N_0}{N}=S$$

とすれば S は

$$(9)\quad n(n-1)\int_S^1\xi^{n-2}(1-\xi)d\xi=\alpha$$

の根であることが分つてゐる。[6] この結果を用ひれば上記第二の問題は解ける譯である。

引用文献及び註

1] 東大第一工學部應用數學教室.森口助教授より提出された問題である。

2] S. S. Wilks, Mathematical Statistics, Princeton University Press, (1943), p. 94.

3] 高木貞治，解析概論，岩波書店．昭和 21年) p. 411.

4] この計算は，統計數理研究所，井草準一，丸山文行兩氏の御教示に依る。

5] $[L/l]$ はガウスの記號である。

6] S. S. Wilks, Statistical Prediction with Special Reference to The Problem of Tolerance Limits, The Annals of Mathematical Statistics. Vol. 13. (1942).