(39) 或 Order Statistic の問題に就て

小川潤次郎

統計数理研究所

(昭和22年10月22日統計數理研究所談話會講演)

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関聯して、次の如き問題を提出されたことがある。即ち、

ある母集団(勿論これは連続な密度函数を有するとする)より、先づ第一の無作意標本 $\Sigma_1 : x_1<x_2<\cdots<x_n$ を抽出し、その最小値 x_1、最大値 x_n とする。次に前の標本と独立に第二の無作意標本 $\Sigma_2 : x'_1<x'_2<\cdots\cdots<x'_n$ を抽出し、その最小値 x'_1、最大値 x'_n とするとき ~~に前の標本と独立に第二の無作意標本 $\Sigma_2 : x'_1<x'_2<\cdots\cdots<x'_n$ を抽出し、その最小値 x'_1、最大値 x'_n とするとき~~、Σ_2 の最大値 x'_n が Σ_1 の最大値 x_n を越える確率は如何。

又例へば工科方面の問題で、ピアノ線の強度試験并の問題がある。(1) それはピアノ線のやうな場合には強度の平均値よりは、強度の最小が問題になることがあるからであって、それを使ふ側から云へば、その強度試験と云ふのは、この最小を control する様に design されることが望ましい訳である。

先づ第一の問題から考へる。母集団の確率密度を $f(x)$ として、次の如く、order statistic の問題を考へる。

今、

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

なる正の整数列を取り、これに対応して、任意の定数列

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

を一つ定めたとき、第一の無作為標本 Σ_1 を大きさの順に並べて、

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

であつて、

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

となる確率 P_I を求めよう。

$$\int_{-\infty}^{x_{r_1}} f(x)\,dx = u_1,\ \int_{x_{r_1}}^{x_{r_2}} f(x)\,dx = u_2, \cdots, \int_{x_{r_{k-1}}}^{x_{r_k}} f(x)\,dx = u_k,\ \int_{x_{r_k}}^{\infty} f(x)\,dx = u_{k+1}$$

とおくと、勿論

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

であつて、且つ、

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

求める確率 P_I は多項分布に依つて

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

と成る。

次に、第二の無作為標本Σ_2が、この定められた実数列、

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

に関して、区間

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

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。

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

依つて、一つの定められた実数列$x_{r_1}, x_{r_2}, \cdots, x_{r_k}$に関して、第一の無作為標本$\Sigma_1$が(*)なる条件を満たし、同時に第二の無作為標本$\Sigma_2$が(**)なる条件を満たす確率$P'$は(1)と(2)の積で与へられる。

$$(3)\ P' = \frac{n!\,N!}{(r_1-1)!(r_2-r_1-1)!\cdots(r_k-r_{k-1}-1)!(n-r_k)!N_1!\cdots N_{k+1}!} \times$$

$$u_1^{N_1+r_1-1} u_2^{N_2+r_2-r_1-1} \cdots u_k^{N_k+r_k-r_{k-1}-1} u_{k+1}^{N_{k+1}+n-r_k} du_1 \cdots du_k$$

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

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

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

なる範囲で積分すればよい。この範囲をDて表はすと、これはよく知られたDirichlet積分[3]であつて、

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

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

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

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

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

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) $$\frac{n!\,N!\,(N_1+n+1)!\,N_2!}{(n-1)!\,N_1!\,N_2!\,(N+n)!}$$

$$= \frac{n}{N+n} \frac{{}_N C_{N_2}}{{}_{N+n-1} C_{N_2}}$$

となる、従つてX'_NがX_nを越す確率Qは次式で与へ

られる。

$$(6) \quad Q=\frac{n}{N+n}\sum_{\nu=1}^{N}\frac{{}_{N}C_{\nu}}{{}_{N+n-1}C_{\nu}}=\frac{N}{N+n} \quad ^{(4)}$$

若し特に$N=n$ならば、$Q=\frac{1}{2}$でこれは常識に合致する。

次にピアノ線の強度試験の場合を考へよう。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験片(Test piece)の強度の結果は偶然的な変動であると考へ得るとする。このやうな場合に、幾個かの試験片を強度試験をしたとき、その最小強度を知つたとき、全ピアノ線の強度に関して吾々の得る客観的な知識は如何と云ふ問題を考えよう。

今試験片の長さ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個の試験片を強度試験した結果、吾々はその最小強度がx_1であつたことを知つたとする。この時N個の試験片の内少くともN_0個がx_1より大なる強度を有する確率Rを求めよう。その為には(4)式で

$$k=1,\ r_1=1,\ x_{r_1}=x_1,$$
$$N_1=N-\nu,\ N_2=\nu,\ (\nu\geqq N_0)$$

とおけばよいから、求むる確率Rは、

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

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

そこで、例へば

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

である如く N_0 を与へたとき、

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

となる如く n を定める、か又は n, N が与へられたとき

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

である如く N_0 を求めるかである。

前者は n 個の試験片の強度試験をして、その最小強度を知つたとき、ピアノ線の全長の99%がこの最小強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。若し此処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、その内から任意に、上記 n 個の試験片を取つて強度試験をすれば、これから、吾々は1%の危険率を以つて、その99%は、この試験片の最小強度より大なる強度を有すると結論出来るのである。

以上述べたことは、ペニシリンの場合にも、ピアノ線の場合に未だ大な不満があるのであつて、夫々は、標本の最大、最小値の知識から母集団の最大、最小値の絶対値がどの位かを知り度いのである。勿論この為には、密度函数 $f(x)$ の函数形は既知である必要があるがこれは未だ解けてゐない。

其后東大図館に於て、米国寄贈の The Annals of Mathematical Statistics Vol 13. (1942) に於

て、S.S. Wilks が Statistical Prediction with Special Reference to The Problem of Tolerance 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。尚ほ第二の問題に関して、Nが充分大なるとき、即ち、

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

とすれば、Sは

$$(9)\qquad n(n-1)\int_{S}^{1}\xi^{n-2}(1-\xi)\,d\xi=\alpha$$

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訳である。

引用文献及び註、

(1) 東大第一工学部、応用数学教室、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、統計数理研究所、井草準一、丸山文行両氏の御教示に依る。

(5) $\left[\frac{l}{2}\right]$ 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).