

③ 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日統計数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、

ある母集団(勿論これは連続な密度函数を有するとし
る)より、先づ第一の無任意標本 $\Sigma_1: X_1 < X_2 < \dots < X_n$ を抽出
し、その最小値 X_1 、最大値 X_n とする。次に別の標本
と独立に第二の無任意標本 $\Sigma_2: X'_1 < X'_2 < \dots < X'_n$ を抽出し
、その最小値 X'_1 、最大値 X'_n とすると、~~これらの標本と独立~~
~~に第二の無任意標本 $\Sigma_2: X'_1 < X'_2 < \dots < X'_n$ 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 を超える確率は如何。

又例へば工料方面の問題で、ピアノ線の強度試験等の
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、それを防ぐ側から云へば、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
が望ましいら款である。

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 の問題と考へる。

今、

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

なる正の整数列を取り、これに対応して、任意の定数列

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

を一つ決めるとき、第一の無作為標本 Σ_1 を大きさ n の順に並べて、

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

であつて、

$$(*) \quad X_{r_1} < X_{r_1} + dX_{r_1}, X_{r_2} < X_{r_2} + dX_{r_2}, \dots, X_{r_k} < X_{r_k} + dX_{r_k}$$

となる確率 P_I を求めよう。

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

$\int_{x_{r_k}}^{x_{r_k}} f(x) dx = 0$
 $\int_{x_{r_{k-1}}}^{x_{r_k}} f(x) dx = u_k$
 $\int_{x_{r_{k-2}}}^{x_{r_{k-1}}} f(x) dx = u_{k-1}$
 $\int_{x_{r_{k-3}}}^{x_{r_{k-2}}} f(x) dx = u_{k-2}$
 $\int_{x_{r_{k-4}}}^{x_{r_{k-3}}} f(x) dx = u_{k-3}$
 $\int_{x_{r_{k-5}}}^{x_{r_{k-4}}} f(x) dx = u_{k-4}$
 $\int_{x_{r_{k-6}}}^{x_{r_{k-5}}} f(x) dx = u_{k-5}$
 $\int_{x_{r_{k-7}}}^{x_{r_{k-6}}} f(x) dx = u_{k-6}$
 $\int_{x_{r_{k-8}}}^{x_{r_{k-7}}} f(x) dx = u_{k-7}$
 $\int_{x_{r_{k-9}}}^{x_{r_{k-8}}} f(x) dx = u_{k-8}$
 $\int_{x_{r_{k-10}}}^{x_{r_{k-9}}} f(x) dx = u_{k-9}$
 $\int_{x_{r_{k-11}}}^{x_{r_{k-10}}} f(x) dx = u_{k-10}$
 $\int_{x_{r_{k-12}}}^{x_{r_{k-11}}} f(x) dx = u_{k-11}$
 $\int_{x_{r_{k-13}}}^{x_{r_{k-12}}} f(x) dx = u_{k-12}$
 $\int_{x_{r_{k-14}}}^{x_{r_{k-13}}} f(x) dx = u_{k-13}$
 $\int_{x_{r_{k-15}}}^{x_{r_{k-14}}} f(x) dx = u_{k-14}$
 $\int_{x_{r_{k-16}}}^{x_{r_{k-15}}} f(x) dx = u_{k-15}$
 $\int_{x_{r_{k-17}}}^{x_{r_{k-16}}} f(x) dx = u_{k-16}$
 $\int_{x_{r_{k-18}}}^{x_{r_{k-17}}} f(x) dx = u_{k-17}$
 $\int_{x_{r_{k-19}}}^{x_{r_{k-18}}} f(x) dx = u_{k-18}$
 $\int_{x_{r_{k-20}}}^{x_{r_{k-19}}} f(x) dx = u_{k-19}$
 $\int_{x_{r_{k-21}}}^{x_{r_{k-20}}} f(x) dx = u_{k-20}$
 $\int_{x_{r_{k-22}}}^{x_{r_{k-21}}} f(x) dx = u_{k-21}$
 $\int_{x_{r_{k-23}}}^{x_{r_{k-22}}} f(x) dx = u_{k-22}$
 $\int_{x_{r_{k-24}}}^{x_{r_{k-23}}} f(x) dx = u_{k-23}$
 $\int_{x_{r_{k-25}}}^{x_{r_{k-24}}} f(x) dx = u_{k-24}$
 $\int_{x_{r_{k-26}}}^{x_{r_{k-25}}} f(x) dx = u_{k-25}$
 $\int_{x_{r_{k-27}}}^{x_{r_{k-26}}} f(x) dx = u_{k-26}$
 $\int_{x_{r_{k-28}}}^{x_{r_{k-27}}} f(x) dx = u_{k-27}$
 $\int_{x_{r_{k-29}}}^{x_{r_{k-28}}} f(x) dx = u_{k-28}$
 $\int_{x_{r_{k-30}}}^{x_{r_{k-29}}} f(x) dx = u_{k-29}$
 $\int_{x_{r_{k-31}}}^{x_{r_{k-30}}} f(x) dx = u_{k-30}$
 $\int_{x_{r_{k-32}}}^{x_{r_{k-31}}} f(x) dx = u_{k-31}$
 $\int_{x_{r_{k-33}}}^{x_{r_{k-32}}} f(x) dx = u_{k-32}$
 $\int_{x_{r_{k-34}}}^{x_{r_{k-33}}} f(x) dx = u_{k-33}$
 $\int_{x_{r_{k-35}}}^{x_{r_{k-34}}} f(x) dx = u_{k-34}$
 $\int_{x_{r_{k-36}}}^{x_{r_{k-35}}} f(x) dx = u_{k-35}$
 $\int_{x_{r_{k-37}}}^{x_{r_{k-36}}} f(x) dx = u_{k-36}$
 $\int_{x_{r_{k-38}}}^{x_{r_{k-37}}} f(x) dx = u_{k-37}$
 $\int_{x_{r_{k-39}}}^{x_{r_{k-38}}} f(x) dx = u_{k-38}$
 $\int_{x_{r_{k-40}}}^{x_{r_{k-39}}} f(x) dx = u_{k-39}$
 $\int_{x_{r_{k-41}}}^{x_{r_{k-40}}} f(x) dx = u_{k-40}$
 $\int_{x_{r_{k-42}}}^{x_{r_{k-41}}} f(x) dx = u_{k-41}$
 $\int_{x_{r_{k-43}}}^{x_{r_{k-42}}} f(x) dx = u_{k-42}$
 $\int_{x_{r_{k-44}}}^{x_{r_{k-43}}} f(x) dx = u_{k-43}$
 $\int_{x_{r_{k-45}}}^{x_{r_{k-44}}} f(x) dx = u_{k-44}$
 $\int_{x_{r_{k-46}}}^{x_{r_{k-45}}} f(x) dx = u_{k-45}$
 $\int_{x_{r_{k-47}}}^{x_{r_{k-46}}} f(x) dx = u_{k-46}$
 $\int_{x_{r_{k-48}}}^{x_{r_{k-47}}} f(x) dx = u_{k-47}$
 $\int_{x_{r_{k-49}}}^{x_{r_{k-48}}} f(x) dx = u_{k-48}$
 $\int_{x_{r_{k-50}}}^{x_{r_{k-49}}} f(x) dx = u_{k-49}$
 $\int_{x_{r_{k-51}}}^{x_{r_{k-50}}} f(x) dx = u_{k-50}$
 $\int_{x_{r_{k-52}}}^{x_{r_{k-51}}} f(x) dx = u_{k-51}$
 $\int_{x_{r_{k-53}}}^{x_{r_{k-52}}} f(x) dx = u_{k-52}$
 $\int_{x_{r_{k-54}}}^{x_{r_{k-53}}} f(x) dx = u_{k-53}$
 $\int_{x_{r_{k-55}}}^{x_{r_{k-54}}} f(x) dx = u_{k-54}$
 $\int_{x_{r_{k-56}}}^{x_{r_{k-55}}} f(x) dx = u_{k-55}$
 $\int_{x_{r_{k-57}}}^{x_{r_{k-56}}} f(x) dx = u_{k-56}$
 $\int_{x_{r_{k-58}}}^{x_{r_{k-57}}} f(x) dx = u_{k-57}$
 $\int_{x_{r_{k-59}}}^{x_{r_{k-58}}} f(x) dx = u_{k-58}$
 $\int_{x_{r_{k-60}}}^{x_{r_{k-59}}} f(x) dx = u_{k-59}$
 $\int_{x_{r_{k-61}}}^{x_{r_{k-60}}} f(x) dx = u_{k-60}$
 $\int_{x_{r_{k-62}}}^{x_{r_{k-61}}} f(x) dx = u_{k-61}$
 $\int_{x_{r_{k-63}}}^{x_{r_{k-62}}} f(x) dx = u_{k-62}$
 $\int_{x_{r_{k-64}}}^{x_{r_{k-63}}} f(x) dx = u_{k-63}$
 $\int_{x_{r_{k-65}}}^{x_{r_{k-64}}} f(x) dx = u_{k-64}$
 $\int_{x_{r_{k-66}}}^{x_{r_{k-65}}} f(x) dx = u_{k-65}$
 $\int_{x_{r_{k-67}}}^{x_{r_{k-66}}} f(x) dx = u_{k-66}$
 $\int_{x_{r_{k-68}}}^{x_{r_{k-67}}} f(x) dx = u_{k-67}$
 $\int_{x_{r_{k-69}}}^{x_{r_{k-68}}} f(x) dx = u_{k-68}$
 $\int_{x_{r_{k-70}}}^{x_{r_{k-69}}} f(x) dx = u_{k-69}$
 $\int_{x_{r_{k-71}}}^{x_{r_{k-70}}} f(x) dx = u_{k-70}$
 $\int_{x_{r_{k-72}}}^{x_{r_{k-71}}} f(x) dx = u_{k-71}$
 $\int_{x_{r_{k-73}}}^{x_{r_{k-72}}} f(x) dx = u_{k-72}$
 $\int_{x_{r_{k-74}}}^{x_{r_{k-73}}} f(x) dx = u_{k-73}$
 $\int_{x_{r_{k-75}}}^{x_{r_{k-74}}} f(x) dx = u_{k-74}$
 $\int_{x_{r_{k-76}}}^{x_{r_{k-75}}} f(x) dx = u_{k-75}$
 $\int_{x_{r_{k-77}}}^{x_{r_{k-76}}} f(x) dx = u_{k-76}$
 $\int_{x_{r_{k-78}}}^{x_{r_{k-77}}} f(x) dx = u_{k-77}$
 $\int_{x_{r_{k-79}}}^{x_{r_{k-78}}} f(x) dx = u_{k-78}$
 $\int_{x_{r_{k-80}}}^{x_{r_{k-79}}} f(x) dx = u_{k-79}$
 $\int_{x_{r_{k-81}}}^{x_{r_{k-80}}} f(x) dx = u_{k-80}$
 $\int_{x_{r_{k-82}}}^{x_{r_{k-81}}} f(x) dx = u_{k-81}$
 $\int_{x_{r_{k-83}}}^{x_{r_{k-82}}} f(x) dx = u_{k-82}$
 $\int_{x_{r_{k-84}}}^{x_{r_{k-83}}} f(x) dx = u_{k-83}$
 $\int_{x_{r_{k-85}}}^{x_{r_{k-84}}} f(x) dx = u_{k-84}$
 $\int_{x_{r_{k-86}}}^{x_{r_{k-85}}} f(x) dx = u_{k-85}$
 $\int_{x_{r_{k-87}}}^{x_{r_{k-86}}} f(x) dx = u_{k-86}$
 $\int_{x_{r_{k-88}}}^{x_{r_{k-87}}} f(x) dx = u_{k-87}$
 $\int_{x_{r_{k-89}}}^{x_{r_{k-88}}} f(x) dx = u_{k-88}$
 $\int_{x_{r_{k-90}}}^{x_{r_{k-89}}} f(x) dx = u_{k-89}$
 $\int_{x_{r_{k-91}}}^{x_{r_{k-90}}} f(x) dx = u_{k-90}$
 $\int_{x_{r_{k-92}}}^{x_{r_{k-91}}} f(x) dx = u_{k-91}$
 $\int_{x_{r_{k-93}}}^{x_{r_{k-92}}} f(x) dx = u_{k-92}$
 $\int_{x_{r_{k-94}}}^{x_{r_{k-93}}} f(x) dx = u_{k-93}$
 $\int_{x_{r_{k-95}}}^{x_{r_{k-94}}} f(x) dx = u_{k-94}$
 $\int_{x_{r_{k-96}}}^{x_{r_{k-95}}} f(x) dx = u_{k-95}$
 $\int_{x_{r_{k-97}}}^{x_{r_{k-96}}} f(x) dx = u_{k-96}$
 $\int_{x_{r_{k-98}}}^{x_{r_{k-97}}} f(x) dx = u_{k-97}$
 $\int_{x_{r_{k-99}}}^{x_{r_{k-98}}} f(x) dx = u_{k-98}$
 $\int_{x_{r_k}}^{x_{r_k}} f(x) dx = 0$

とかくと、勿論

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

であつて、且つ、

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

求める確率 P_I は多項分布に依つて

$$(1) \quad P_I = \frac{n!}{(n-1)!(n-r_1-1)!\dots(n-r_k-1)!(n-r_{k+1})!} u_1^{r_1-1}$$

$$u_2^{r_2-r_1-1} u_k^{r_k-r_k-1} u_{k+1}^{n-r_k} du_1 \dots du_k \quad (2)$$

となる。

次に、第二の無作為標本 Σ_2 について、この定められた変数列

$x_{r_1}, x_{r_2}, \dots, x_{r_k}$
 に関して、区間

$(-\infty, x_{r_1}), (x_{r_1}, x_{r_2}), \dots, (x_{r_{k-1}}, x_{r_k}), (x_{r_k}, \infty)$
 に夫々、

$N_1, N_2, \dots, N_k, N_{k+1}$ (**)
 個づつ入る確率を P_{II} とする。勿論、

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

である。

$$(2) \quad P_{II} = \frac{N!}{N_1! N_2! \dots N_{k+1}!} \mathcal{U}_1^{N_1} \mathcal{U}_2^{N_2} \dots \mathcal{U}_k^{N_k} \mathcal{U}_{k+1}^{N_{k+1}}$$

従つて、一つの定められた変数列 $x_{r_1}, x_{r_2}, \dots, x_{r_k}$ に関して、第一の無作為標本 Σ_1 が (*) なる条件を満たし、同時に第二の無作為標本 Σ_2 が (**) なる条件を満たす確率 P' は (1) と (2) の積で与えられる。

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

$$\mathcal{U}_1^{N_1+r_1-1} \mathcal{U}_2^{N_2+r_2-r_1-1} \dots \mathcal{U}_k^{N_k+r_k-r_{k-1}-1} \mathcal{U}_{k+1}^{N_{k+1}+n-r_k}$$

従つて、これを $x_{r_1}, x_{r_2}, \dots, x_{r_k}$ に関して積分すれば、 N_1, N_2, \dots, N_{k+1} ($= N - N_1 - \dots - N_k$) の同時分布が得られるのであるが、それは、 $\mathcal{U}_1, \mathcal{U}_2, \dots, \mathcal{U}_k$ に関して

$$\mathcal{U}_1 \geq 0, \mathcal{U}_2 \geq 0, \dots, \mathcal{U}_k \geq 0.$$

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

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

$$\begin{aligned} & \int \dots \int_{\mathcal{D}} u_1^{N_1+Y_1-1} \dots u_k^{N_k+Y_k-Y_{k+1}-1} (1-u_1-u_2-\dots-u_k)^{N_{k+1}+X-Y_k} du_1 \dots du_k \\ &= \frac{\Gamma(N_1+Y_1) \Gamma(N_2+Y_2-Y_1) \dots \Gamma(N_k+Y_k-Y_{k-1}) \Gamma(N_{k+1}+X-Y_k)}{\Gamma(N+X+1)} \\ &= \frac{(N_1+Y_1-1)! (N_2+Y_2-Y_1-1)! \dots (N_k+Y_k-Y_{k-1}-1)!}{(N+X)!} \end{aligned}$$

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

$$(4) P = \frac{n! N! (N_1+Y_1-1)! (N_2+Y_2-Y_1-1)! \dots (N_k+Y_k-Y_{k-1}-1)! (N_{k+1}+X-Y_k)!}{(Y_1-1)! (Y_2-Y_1-1)! \dots (Y_k-Y_{k-1}-1)! (n-Y_k)! N_1! N_2! \dots N_{k+1}! (n+N)}$$

となる。

上の結果を第一の増山氏の向題の場合に特殊化する。

るには、

$$k=1, \quad Y_1=n, \quad X_{Y_1}=X_n$$

$$N_1+N_2=N$$

として、 N_1, N_2 の同時分布を求めると、それは (4) より、

$$\begin{aligned} (5) \quad & \frac{n! N! (N_1+n+1)! N_2!}{(n-1)! N_1! N_2! (N+n)!} \\ &= \frac{n}{N+n} \frac{N C_{N_2}}{N+n-1 C_{N_2}} \end{aligned}$$

となる。従つて X'_{N_1} が X_n を越す確率 Q は次式であら

られる。

$$(6) \quad Q = \frac{n}{N+n} \sum_{v=1}^N \frac{NC_v}{N+n-1 C_v} = \frac{N}{N+n} \quad (4)$$

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) の強度の差は偶然的な変動であると考え得るとする。このやうな場合に、幾個かの試験片を強度試験せしめるとき、その最小強度を知つたとき、全ピアノ線の強度に關して吾々の備へる客観的な知識は如何と云ふ問題を考へよう。

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 個の試験片の内少なくとも N_1 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

$$n=1, \quad v=1, \quad n v = X_1, \\ N_1 = N - v, \quad N_2 = v, \quad (v \geq N_0)$$

とあけばよいから、求める確率 R は、

$$(7) \quad R = \sum_{v=N_0}^N \frac{n! N! (N-v)! (v+n-1)!}{v N_0! (n-1)! (N-v)! v! (n+N)!} \\ = \frac{n X_1 N!}{(n+N)!} \sum_{v=N_0}^N \frac{(v+n-1)!}{v!}$$

そこで、例へば

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

である如く N_0 を与へたとき、

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

となる如く n を定める、か又は、 n, N が与へられたとき

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

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。若し此處に何本かのピアノ線があつて、それか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、それから、吾々は α % 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、ピアノ線の場合にも未だ大なる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、最小値の総計値などの位を知らぬのである。勿論この爲には、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。

其の東大図書館に於て、米国奇蹟の *The Annals of Mathematical Statistics*, Vol. 13. (1942) に於

て、S. S. Wilks の *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る表現を作製してゐることを知つた。尚ほ第二の向題に關して、 N が充分大なるとき、即ち、

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

とすれば、 S' は、

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

の根であることが分つてゐる。⁽⁶⁾ この結果を用ひれば、上記第二の向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向題である。
- (2) S. S. Wilks, *Mathematical Statistics*, Princeton University Press, (1943), P. 94.
- (3) 高木貞治、群论概論、岩波書店、(昭和21年) P. 411.
- (4) この計算は、統計数理研究所、井草淳一、丸山文行両氏の御教示に依る。
- (5) $\left\{ \frac{\xi}{2} \right\}$ 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).