は適当でない。 夫で筆者は円筒を用いる二次元法を一次元法に変へることを提案した。この為へは円筒濾過面に依つて三次成長される面性連鎖状球菌，血液寒天を有 用載試験に入れることによつて成功した。菌を添入した血液寒天を細い試験管に入れ，その上に標準ペニシリン溶液を重ねるのである。この場合境界面から溶血域の境界の距離をyとすると，濃度の対数xに対し $y = L(1-e^{-\lambda(x-a)})$ がよく適合することが分つた。実験範囲は0.04〜4単位迄である。この場合~~[illegible]~~に変曲点が存在しない。常数の意味は上と似てゐる。之等を推定するにはtを一定間隔ずつとると

$$y_{t+h} = e^{-\lambda h} y_t + L(1-e^{-\lambda h})$$

であるから(y_t, y_{t+h})が同一直線上にあることを利用して，$e^{-\lambda h}$ ついで $L(1-e^{-\lambda h})$，従つてLを推定するのである。力価検定の仕方は前と殆んど同様である。
この公式は培地の量の変化や読み取り時間の変化に対して，形式を変へないことが実験的に分つてゐる。誤差論的には更に実験を重ねてから報告したい。尚この公式は前に筆者が溶血曲線に利用したことのある式であるが，他にも応用例がある（八木，小泉氏の著書参照）
尿中又は血液中のペニシリンに対しても恐らく同じ形の式が成立すると思はれるが，実験資料が未だ無い。(昭和21年12月16日)

Poisson分布に対するSequential Test

所員 小川潤次郎

A. Wald は Annals of Math. Stat. Vol 16 No 2 (1945) で Sequential Test の一般理論を展開してそれの特殊分布への適用を述べてゐるが，二項分布及び正規分布の場合には詳細に亘つて述べてゐるが，Poisson分布の場合は述べてゐないので，此処では A. Wald の方式に従つて Poisson 分布に対する Sequential Test を作つて見よう。

註(1) A. Wald Sequential Test of Statistical Hypothesis の翻訳は近くその一部が筆者に依り研究録 号に発表される予定である

例へば J. Przyborowski 及 H. Wilenski, Statistical Principles of Routine Work in Testing Clover Seed for

Dodder[2] に依れば Radiner の如き小さな種子の中に混入して居る Dodder の個数は Poisson の分布をなす由である。

而して此種子を検査するには通常100乃至200瓦の種子を抽出して其内に混入してゐる Dodder の個数を調べるのであるが此標本中に含まれる clover の種子の個数は59,000内外である。

Poisson 分布とは変量 x が $0, 1, 2, 3, \cdots$ の discrete な値を取るものとしてその $x = k$ である確率が

$$P_k = e^{-m}\frac{m^k}{k!}$$

で与へられるもので其の期望値 $E(x) = m$ 標準偏差 $\sqrt{E(x-E(x))^2} = \sqrt{m}$ である。

検査は抽出検査であるから必ず誤りを生ずる。そこで此検査の結果に対して買手の側と売手の側と二つの立場からの注文がある訳である。

例へば検査規定として per size of sample の Dodder の含有平均個数が $m_0 = 5$ 以下のとき合格 $m > m_0$ のとき不合格とすれば此の検査に依れば $m > m_0$ のものが合格品とされ又 $m \leq 5$ は $m > m_0$ のもの~~不合格品~~とされることも有る訳であるが買手の側からすれば $m > m_0$ のものが合格する確率は成る可く小さく又売手の側からすれば $m \leq m_0$ のものが不合格とされる確率が成る可く小さいことが望ましい訳である。

扨て買手の側としては $m > m_0 (= 5)$ のものは不合格となる確率が大であって欲しい訳であるが、$m = 5$ も $m = 6, 7, 8, 9$ であってもまあ気厳的には我慢出来るが $m = 10$ のものは絶対に困ると云ふ様なことがあるであらう。又売手の側とすれば $m = 4, 3, 2$ 位迄の品物が時に不合格となるのは止むを得ぬとしても $m = 1, 0$ の品物は絶対合格してくれねば困るといふ様な事情がある。

即ち $0 \leq m_1 < m_0 < m_2$ —— 此場合には $m_1 = 1$, $m_2 = 10$ である —— なる二数 m_1, m_2 を取って今

帰無仮説として H_1　$m = m_1$

対立仮説として H_2　$m = m_2$

を取り此に対して第一種の誤謬の確率 α, 第二種の誤謬の確率 β なる Sequential Probability Ratio Test を作らう。これを強さ (α, β) の Test と云ふ。

一般に A. Wald の Sequential Test とは標本空間を三つ

註(2) Biometrika, Vol. 27 (1935) 273頁参照

の部分空間に分つ方法が与へられて先ず標本を一つ抽出してそれを x_1 とするとき一次元の標本空間が R_1^1, R_1^2, R_1^3 に分割されて $x_1 \in R_1^1$ なら仮説 H_1 を棄却し $x_1 \in R_1^2$ なら仮説 H_2 を棄却し又 $x_1 \in R_1^3$ なら更に一つの標本 x_2 を取り二次元の標本空間を考へるとそれが同じ様に三部分に分割され R_2^1 R_2^2 R_2^3 となる。標本点 $P=(x_1, x_2)$ として $P \in R_2^1$ なら H_1 を棄却し $P \in R_2^2$ なら H_2 を棄却し $P \in R_2^3$ なら更に一つの標本 x_3 を抽出して三次元の標本空間を考へるとこれが三部分 R_3^1 R_3^2 R_3^3 に分割され若し標本点 $P=(x_1, x_2, x_3)$ が R_3^1 なら仮説 H_1 を棄却し $P \in R_3^2$ なら H_2 を棄却し $P \in R_3^3$ なら更に一つの標本 x_4 を抽出すると云ふ風に進んで一般に m 個の標本 $x_1, x_2, \ldots, x_m$ が抽出されたとき m 次元の標本空間が三部分空間 R_m^1 R_m^2 R_m^3 に分割され標本点 $P=(x_1, x_2, \ldots, x_m)$ が R_m^1 なら仮説 H_1 を棄却し $P \in R_m^2$ なら仮説 H_2 を棄却しこれで Test は終るのであるが $P \in R_m^3$ ならもう一つの標本 x_{m+1} を取り出して $(m+1)$ 次元の標本空間で考へる。

この場合各次元の標本空間を三部分空間に分割する一つの方法が即ち一つの Sequential Test を与へる訳である。

Single alternative H_2 を有する Simple null hypothesis H_1 の場合には事情は比較的簡単であって〜となる母集団の分布を $f(x, \theta)$ として θ は unknown parameter とする。H_1 $\theta=\theta_1$, H_2 $\theta=\theta_2$ とする。

無限次元の標本空間の内で考へて m 次元に project されたとき R_m^1 に落ちる点の全体を R_∞^1, R_m^2 に project される点の全体を R_∞^2, 残りを R_∞^3 とすれば R_∞^3 は m 次元では R_m^3 に project される。

然れば仮説 H_1 が真であるときこれが棄却される確率 α は

$$\alpha = \int \cdots \int_{R_\infty^1} dP_1$$

dP_1 は例へば m 次元なら $f(x_1, \theta_1) \cdots f(x_m, \theta_1)\, dx_1 \cdots dx_m$ と置け。又仮説 H_2 が真なるときこれが棄却される確率 β は

$$\beta = \int \int_{R_\infty^2} dP_2$$

但し dP_2 は例へば m 次元ならば $f(x_1, \theta_2) \cdots f(x_m, \theta_2)\, dx_1 \cdots dx_m$ である。

そこで標本空間の分割方法の内で特に

$$\int \int_{R_\infty^3} dP_1 = \int \int_{R_\infty^3} dP_2 = 0$$

となるものを考へれば、

$$\iint_{R_0^2} dP_1 = 1-\alpha, \quad \iint_{R_0} dP_2 = 1-\beta$$

となる。

そこで具体的に標本空間の分割を如何に定めるかと云ふことに関して Wald は Probability Ratio Test を与へてゐる。而してこれが Optimum な Sequential Test を与へることになるのである。

先ず Probability Ratio Test の発見的説明から始めよう。

$$f(x,\theta)=e^{-m}\frac{m^x}{x!} \qquad (\theta=m)$$

として母数 m が値 m_1 を取る a priori probability g_1, m_2を取る a priori probability g_2 が存在するとしよう。この母集団から n 個の標本 $x_1, x_2, \ldots, x_n$ を抽出してこれを知った後の $m=m_1$ となる a posterior probability g_{1n}, $m=m_2$ となる a posterior probability g_{2n} とする。然らば Bayes の定理に依って、

$$p(x_1, \ldots\ldots, x_n, m_i) = e^{-nm_i}\frac{m_i^{\sum_{j=1}^{n} x_j}}{x_1!\, x_2! \cdots x_n!} \qquad i=1,2$$

として

$$g_{1n} = \frac{g_1 p(x_1, \ldots x_n, m_1)}{g_1 p(x_1 \cdots x_n, m_1) + g_2 p(x_1 \cdots x_n, m_2)}, \quad g_{2n} = \frac{g_2 p(x_1, \ldots, x_n, m_2)}{g_1 p(x_1, \ldots x_n, m_1) + g_2 p(x_1, \ldots, x_n, m_2)}$$

となる。適当に二つの常数

$$0<d_1<\frac{1}{2}, \qquad 0<d_2<\frac{1}{2}$$

を取って、次の如く定めることは直観的に妥当であらう。

$g_{1n} \leq d_1$ のとき假説 H_1 を棄却する。

$g_{2n} \leq d_2$ のとき假説 H_2 を棄却する。

$g_{1n} > d_1$, $g_{2n} > d_2$ なら更にもう一つの標本を抽取って調べる。

即ち n 次元の標本空間内に $g_1 n \leqq \alpha_1$ で定められる領域 R_n^1 及 Null Hypthesis H_1 に重点を置いて云ふならば、棄却領域 (Rejitance Region)、$g_2 n \leqq \alpha_2$ で定められる領域 R_n^2 及採擇領域 (Acceptance Region) となり、$g_1 n > \alpha_1$, $g_2 n > \alpha_2$ が定める領域が R_n^3 となる。これを書改めると

$$R_n^1: \quad \frac{p_2}{p_1} \geqq \frac{g_2}{g_2}\,\frac{1-\alpha_1}{\alpha_1}$$

$$R_n^2: \quad \frac{p_2}{p_1} \leqq \frac{g_1}{g_2}\,\frac{\alpha_2}{1-\alpha_2}$$

$$R_n^3: \quad \frac{g_1}{g_2}\,\frac{\alpha_2}{1-\alpha_2} < \frac{p_2}{p_1} < \frac{g_1}{g_2}\,\frac{1-\alpha_1}{\alpha_1}$$

但し $p_i = p(x_1, \cdots\cdots, x_n, m_i)$, $i = 1.2$

斯く分布函数の ratio に対する不等式で R_n^1, R_n^2 R_n^3 が定められるから Probability Ratio Test と名づけられる所以である。

そこで a prion probability など用ひずに二つの常数 A, B を取って、

$$\frac{p_2}{p_1} = e^{-n(m_2-m_1)} \left(\frac{m_2}{m_1}\right)^{\sum_{i=1}^{n} x_i}$$

なることに注意して

$$\frac{p_2}{p_1} = e^{-n(m_2-m_1)} \left(\frac{m_2}{m_1}\right)^{\sum_{i=1}^{n} x_i} \geqq A$$

で R_n^1 を定め

$$\frac{p_2}{p_1} = e^{-n(m_2-m_1)} \left(\frac{m_2}{m_1}\right)^{\sum_{i=1}^{n} x_i} \leqq B$$

で R_n^2 を、次に R_n^3 を

$$B < e^{-n(m_2-m_1)} \left(\frac{m_2}{m_1}\right)^{\sum_{i=1}^{n} x_i} < A$$

で定めることにする。

Regaritnm を取って書改めると

$$R_n^1: \quad \sum_{i=1}^{n} x_i \leqq \frac{1}{\log m_2 - \log m_1}\left(\log A + n(m_2 - m_1)\right)$$

$$R_n^2: \quad \sum_{i=1}^{n} x_i \leqq \frac{1}{\log m_2 - \log m_1}\left(\log B + n(m_2 - m_1)\right)$$

$$R_n^3: \quad \frac{1}{\log m_2 - \log m_1}\left(\log B + n(m_2 - m_1)\right) < \sum_{i=1}^{n} x_i \cdots$$

$$\cdots < \frac{1}{\log m_2 - \log m_1}\left(\log A + n(m_2 - m_1)\right)$$

斯かる Sequential Test に対しては $\int\cdots\int_{R_n^3} dp_1 = \int\cdots\int_{R_n^3} dp_2 = 0$ であることはすぐ分るから

$$A \leqq \frac{1-\beta}{\alpha}, \quad B \geqq \frac{\beta}{1-\alpha}$$

が出る。

従って近似的には

$$A = \frac{1-\beta}{\alpha}, \quad B = \frac{\beta}{1-\alpha}$$

と取ってよいから Rejectance 及び Acceptance Region は次の如くなる。即ち

$$R_n^1 \quad \sum_{i=1}^{n} x_i \geqq \frac{\log \frac{1-\beta}{\alpha}}{\log m_2 - \log m_1} + n\,\frac{m_2 - m_1}{\log m_2 - \log m_1}$$

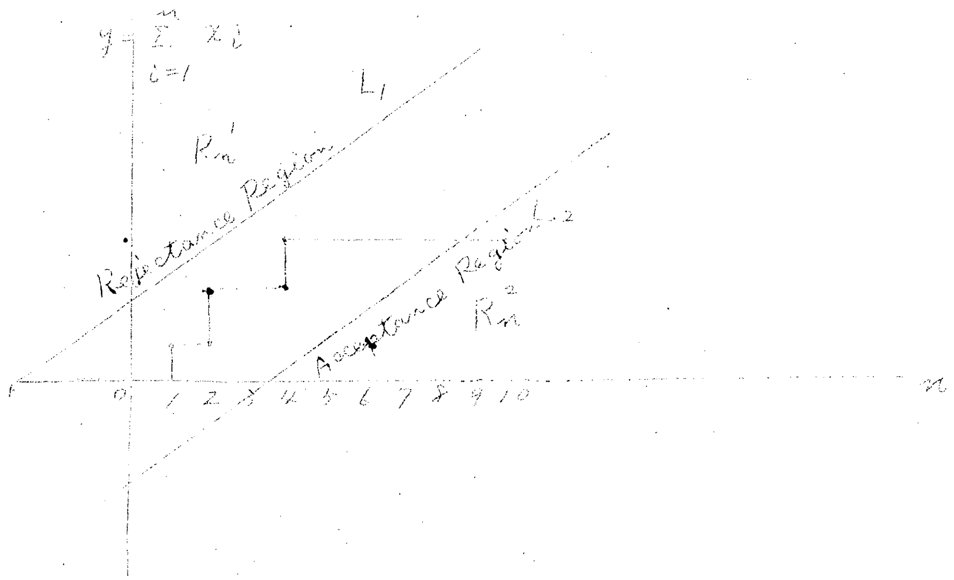
$$R_n^2 \quad \sum_{i=1}^{n} x_i \leqq \frac{\log \frac{\beta}{1-\alpha}}{\log m_2 - \log m_1} + n\,\frac{m_2 - m_1}{\log m_2 - \log m_1}$$

横軸に n を取り、縦軸に $y = \sum_{i=1}^{n} x_i$ を取ると平行線

$$L_1 \quad y = \frac{\log \frac{1-\beta}{\alpha}}{\log m_2 - \log m_1} + n\,\frac{m_2 - m_1}{\log m_2 - \log m_1}$$

$$L_2 \quad y = \frac{\log \frac{\beta}{1-\alpha}}{\log m_2 - \log m_1} + n\,\frac{m_2 - m_1}{\log m_2 - \log m_1}$$

に依って囲まれる Band が R_n^3 となる。



即ち(x, y)を plot して行って、L_1の上方に出ればH_1を reject し、L_2の下方に出ればH_1を accept するのである。

母数mが任意の値mのとき、この Sequential Test に依って、ついにH_1が reject されるという結果になる確率をγとすれば、近似的に

$$\gamma = \frac{1-B^h}{A^h-B^h}$$

但し
$$z=\log\frac{e^{-m_2}m_2^{x}/x!}{e^{-m_1}m_1^{x}/x!} = -(m_2-m_1)+x\log\left(\frac{m_2}{m_1}\right)$$

として$m=m$のときの moment generating function $\varphi(t)$を求めると、

$$\varphi(t)=Ee^{zt}=\sum_{x=1}^{\infty} e^{-(m_2-m_1)t+xt\log\left(\frac{m_2}{m_1}\right)} e^{-m}\frac{m^x}{x!}$$
$$= e^{-(m_2-m_1)t+m\left(\left(\frac{m_2}{m_1}\right)^t-1\right)}$$

となり、
$$\varphi(t)=1$$
なる方程式は0でない real root を只一つ有するから、そ

h を k とすればよいのである。従って $m = m$ のとき H_1 が accept される確率 L_m は

$$L_m = 1 - \gamma = \frac{\left(\frac{1-\beta}{\alpha}\right)^h - 1}{\left(\frac{1-\beta}{\alpha}\right)^h - \left(\frac{\beta}{1-\alpha}\right)^h}$$

となり、従って此 Test の Operating characteristic curve を描くには

$$m = \frac{(m_2 - m_1)h}{\left(\frac{m_2}{m_1}\right)^h - 1}$$

として、h に色々な値を與へて

$$(m,\ L_m)$$

を plot すれば良い。又 $m = m$ のときの Test の終る迄に要する標本数 n の平均値は

$$E_m n \sim \frac{(1-\gamma)\log B + \gamma \log A}{E z}$$

$$= \frac{L_m \log \frac{\beta}{1-\alpha} + (1 - L_m)\log \frac{1-\beta}{\alpha}}{m \log\left(\frac{m_2}{m_1}\right) - (m_2 - m_1)}$$

となる。

この方法を用ひれば前記 J. Przyborowski and H. Wilenski の論文の方法より遙かに良い Test が出来る。この論文との比較検討に関しては数値計算の完了を待って書くことにする。

(1947.1.31)