

⑧ 或る領域假設の検定について

菅 原 正 巳

1. 事のおこり

真空管材料の規格に抜取り検査の條項を入れたいといふので、昭和23年の初めにその委員会が開かれ、まずニッケル條の検査がとりあげられた。真空管材料の規格は昭和23年3月中に決定することになっていたので、規格は一応抜取り検査の條項なしにきまつてしまい、抜取り検査の委員会だけが解散せず、問題を考えることになった。以下に述べるのはこの委員会で問題となつたニッケル條の厚さの検定である。ほほはだ不満足なものであるが、複合假設、領域假設の検定が、特殊な場合を除くとうまく行かない一例とはなるであらう。

2 規格の要求と、母集団パラメーターの満足する條件

ニッケル條には種々のものがあるが、最も薄く、そして実用上最も大切である約0.1mm程度の厚さのものに対する検定方式を考えることにした。例えば厚さ0.127mmの厚さのニッケル條についていえば、その厚さが (0.127 ± 0.007) mmの範囲内であれば一級品、 (0.127 ± 0.015) mmの範囲内であれば二級品と規定されている。

幾つかのニッケル條につき、その厚さを実測してみると、その長さは10mから100m以上にも及び、その厚さの分布は正規分布と考へてもよいようである。また数cm以上離れた處の厚さを測るならば、その測定値は互に独立であるとみてもよい。また厚さについての週期性や、一方に行くに従つて厚くなるということも先づないとみてよい。標準偏差は個々の條によつて異り、 $0.003\text{mm} \sim 0.008\text{mm}$ 程度である。

以上の標準偏差の値より見ると、先に述べた規格は稍、厳格すぎるようである。例えば平均値が 0.127mm 、標準偏差が 0.004mm というかなりよい製品でも、全長の約8%は一級品の規格の範囲外に出てしまうからである。そこで規格を少しくゆるめ、「厚さが $(0.127 \pm 0.007)\text{mm}$ の外に落ちる部分が全長の10%以下であれば一級品とする」というように変更することにする。即ち規格外の部分がある割合より小さければ合格と定めるのである。

かくて規格は一般に次の形式になる。

「厚さが A へられた区間 $(A-l, A+l)$ の外に落ちる部分が全長の $\alpha\%$ 以下であるならば、そのニッケル條は良品である」。ここに A は公稱の厚さ、 l は許容限度である。 $\alpha\%$ は1%程度にしたのであるが、今は10%程度に讓歩せねばなるまい。

測定箇所が数cm以上離れていれば、その厚さは独立であるということ、條の長さが10mから100m以上にも及ぶということから、厚さが A へられた区間の外に出る部分が $\alpha\%$ 以下であるということ、厚さを確率変数と考へ、それが A へられた区間外に出る確

率が $a\%$ 以下である」ことに置きかえても大きな差がないことになる。そこで問題はつぎの形になる。

「ある正規母集団からとられたサンプルが区間 $(A-l, A+l)$ の外に落ちる確率が $a\%$ 以下であれば、この母集団は規格に合格する」母集団の平均値を μ 、標準偏差を σ とすればこの条件は μ と σ との関係を示へるものである。一般性を失うことなく $A=0$ 、 $l=1$ としてもよい。そうすればこの条件は

$$\int_{-1}^1 \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \geq 1 - \frac{a}{100}$$

で与えられる。これを変形して

$$\int_{-\frac{1+\mu}{\sigma}}^{\frac{1-\mu}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt \geq 1 - \frac{a}{100}$$

a を与えたとき

$$\int_{-\xi_1}^{\xi_2} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt \geq 1 - \frac{a}{100}$$

を満足する ξ_1, ξ_2 の値の対は、 $\int_{-\infty}^t \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$ の表を用いて求めることが出来る。この ξ_1, ξ_2 の関係を数値的に求めた上で

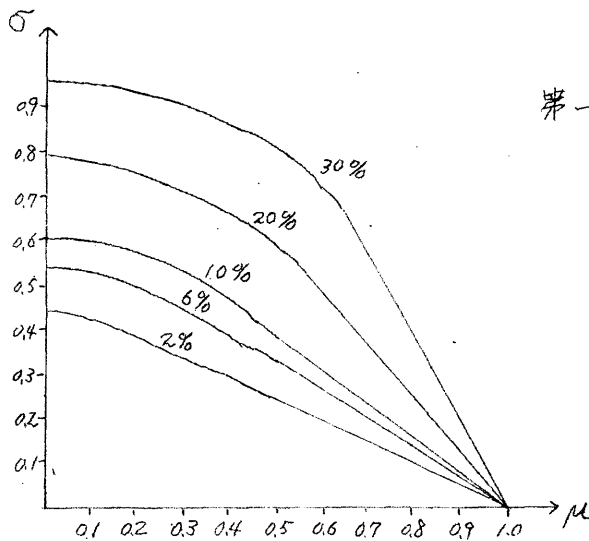
$$\xi_1 = \frac{1+\mu}{\sigma}, \quad \xi_2 = \frac{1-\mu}{\sigma}; \quad \sigma = \frac{2}{\xi_1 + \xi_2}, \quad \mu = \frac{\xi_1 - \xi_2}{\xi_1 + \xi_2}$$

の関係を使えば σ と μ との対応が数値的に求められる。この σ と μ との関係を $\sigma = \varphi_a(\mu)$ によって表わそう。表一表、表一図は

この関係を示すものである。

第一表 $\sigma = \varphi_a(\mu)$ の表

μ \ a	2%	6%	10%	20%	30%
0.0	0.430	0.532	0.6080	0.780	0.965
0.1	0.418	0.523	0.5996	0.772	0.960
0.2	0.386	0.494	0.5745	0.754	0.946
0.3	0.340	0.446	0.5294	0.717	0.916
0.4	0.292	0.385	0.4652	0.665	0.873
0.5	0.243	0.321	0.3902	0.584	0.809
0.6	0.195	0.257	0.3121	0.476	0.715
0.7	0.146	0.193	0.2341	0.357	0.569
0.8	0.097	0.129	0.1561	0.238	0.379
0.9	0.049	0.064	0.0780	0.119	0.189



第一図

$$\sigma = \varphi_a(\mu)$$

$\sigma = \varphi_a(\mu)$ の満足する微分方程式は容易に得られるが、解けそうにもなく、役に立たない。

$\sigma \leq \varphi_a(\mu)$ は母集団が規格に合格するための条件である。

3. 領域仮説に対する検定

以上の様に考へて母集団のパラメーターに対する条件

$$\sigma \leq \varphi_a(\mu)$$

が得られた。しかし母集団のパラメーターは未知であるから、標本値より上の条件が成立しているか、いないかを判定しなければならぬ。

今 n 箇の標本を x_1, x_2, \dots, x_n とする。問題は (x_1, x_2, \dots, x_n) を座標に持つ n 次元の標本空間内に領域 D を求め、標本点で D 内に落ちれば $\sigma \leq \varphi_a(\mu)$ と判定し、 D 外に落ちれば $\sigma > \varphi_a(\mu)$ と判定することで、^良 とうゆう領域のうちで最ものが求められればよいのである。この領域 D は n 次元の標本空間内に考えるべきであるが、こゝでは標本平均値 m 、標本不偏標準偏差 s の間の関係でこの領域が与へられとする。たゞし

$$m = \frac{1}{n} (x_1 + x_2 + \dots + x_n), \quad s^2 = \frac{1}{n-1} \sum (x_i - m)^2$$

m と s とは正規母集団の充足統計量であるから、この仮定はそう不都合であるまい。かくて問題は m, s 平均上に領域 D を求めることに帰する。(Wald の様に損失に対する重さ^を考へて、最良

の検定や統計量を求めると、最良の統計量が必ずしも充足統計量でないこともあるから、上の假定は考慮を要する)

Dは次の性質を満足してほしい。我々は $\sigma \leq \varphi_a(\mu)$ を検定したいのであるが、丁度 $\sigma = \varphi_a(\mu)$ を満足する母集団ならば、Dで検定したときに合格する確率がいずれも $\alpha\%$ であつてほしい。この条件を満足する領域を複台假設 $\sigma = \varphi_a(\mu)$ に対する *size* $\alpha\%$ の *similar region* と言う。そういう *similar region* の中から何事かの意味で最良のものを見つける必要があるが、今の場合は假設 $\sigma = \varphi_a(\mu)$ が簡単でないので *similar region* を求めることがうまく行かない。(*similar region* はないかも知れないが、あるとすれば一意的でない) しかし検定の問題としては *similar region* を求めることだけが大切であるとも言えない。それは以下の事情による。

規格に準へられた区間 ($A-l, A+l$) の外に落ちる部分が全体の $\alpha\%$ である様を、不良部分 $\alpha\%$ の様ということにしよう。今不良部分 10% 以下の品を良品ということにし、これを検定する場合不良部分 10% 以下の良品をすべて通し、不良部分 10% 以上 (10% のものを除く) の不良品をすべて落すと言う検定がある筈はない。検定は不良部分 15% 以上の不良品は 10% 以下しか通さないという不良部分 5% 以下の良品は 90% 以上通し 様に作るべきであらう。もちろんこのように作られた検定によれば、不良部分 10% の品は約半数検定に通り、約半数は検定に落ちるであらう。ところが不幸にして、このような検定を作ることは

望が薄いのである。即ち不良部分5%の品を90%以上通すためには、 $\sigma = \varphi_5(\mu)$ なる複合仮設に対して、size 90% の similar region を作る必要がある。その region が $\sigma = \varphi_{15}(\mu)$ なる複合仮設に対してもまた similar region になるということは、特殊な場合を除いては望み得ないからである。(正規分布に対し、 σ の値に無関係に $\mu = \mu_0$ を検定するのが有名な Student. の test であるが、 $\mu = \mu_0$, $\mu = \mu_1$ ($\mu_0 \neq \mu_1$) のそれぞれに対して σ の値に無関係な検定力を持つ test が存在しないことがわかっている)

そこで similar region を求めることを一応諦めて、次のように考えよう。

不良部分10%の品を検定することを目的とするならば、パラメータの条件は $\sigma \leq \varphi_{10}(\mu)$ であるが、

$$\Delta \leq \varphi_{15}(m)$$

により検定を行うのである。こゝに

$$m = \frac{1}{n} \sum x_i, \quad s^2 = \frac{1}{n-1} \sum (x_i - m)^2$$

であり、その分布函数は

$$\frac{\sqrt{2n}}{\sigma \sqrt{\pi}} \left(\frac{n-1}{2}\right)^{\frac{n-1}{2}} \Gamma\left(\frac{n-1}{2}\right) e^{-\frac{(n-1)s^2 + n(m-\mu)^2}{2\sigma^2}} \delta^{n-2} d\delta dm$$

で与えられる。 m, s^2 の期望値は μ, σ^2 であるから、 $\sigma = \varphi_\alpha(\mu)$ なる条件を満足する母集団を、 $\Delta \leq \varphi_\alpha(m)$ な領域で検定すれば、

略、50%合格するであらう。即ち $\Delta \leq \varphi(m)$ なる領域は假設 $\sigma = \varphi(\mu)$, $\alpha = 50\%$ に対する *similar region* に略、なつてい
と考へられる。更にこの領域は標本の個数を増すに従つて、假設
に対するよい検定領域になる。その意味でこれは *consistent* な
検定領域と言つてよからう。

さてこのように $\Delta \leq \varphi_\alpha(m)$ によつて検定すれば、ある任意の
母集団はこの検定によつてどの程度合格するかを知る必要がある。
それには種々の μ, σ を持つ母集団に対し、 Δ, m がこの領域に
含まれる確率を計算すればよい。数値積分によつてこの確率を計
算したものがオニ表とオ三表とである。計算の基礎表の手落ちか
ら、 σ が小で、 μ が大である場合の計算値には誤差が多くなつた
ので、 $|\mu| \leq 0.5$ の範囲だけをのせる。また $n=20$ の場合の基礎
表に誤りがあつたので、これにも少し誤がある。

このようにして確率を計算すれば、領域 $\Delta \leq \varphi_\alpha(m)$ で検定し
たとき、検定に通る確率が $\alpha\%$ である尖の軌跡が $\mu-\sigma$ 平面上に
描かれる。これがオニ四及びオ三四である。この曲線はオ一四の
曲線とかなり似ている。即ち $\Delta \leq \varphi_\alpha(m)$ で与えられた領域は略、
similar region と見てもよいのである。

第二表 正規母集団 $N(\mu, \sigma)$ が検定に合格する確率

($\Delta \leq \varphi_{10}(m)$ なる領域で検定した場合)

1. 標本数 5

$\sigma \backslash \mu$	0.1	± 0.1	± 0.2	± 0.3	± 0.4	± 0.5
0.1	100%	100	100	100	100	100
0.2	100	100	100	100	99	97
0.3	100	99	98	95	89	76
0.4	91	90	86	78	67	53
0.5	71	70	65	57	47	36
0.6	51	49	45	39	32	25
0.7	33	32	30	26	22	17
0.8	21	21	20	17	15	12
0.9	14	14	13	12	10	9
1.0	10	9	9	8	7	6

2. 標本数 10

$\sigma \backslash \mu$	0.0	± 0.1	± 0.2	± 0.3	± 0.4	± 0.5
0.1	100%	100	100	100	100	100
0.2	100	100	100	100	100	99
0.3	100	100	100	99	94	81
0.4	97	96	93	85	70	50
0.5	79	77	69	57	42	28
0.6	49	47	41	33	23	15
0.7	26	25	22	18	13	9
0.8	13	12	11	9	7	5

3. 標本数 20

$\sigma \backslash \mu$	0.0	± 0.1	± 0.2	± 0.3	± 0.4	± 0.5
0.3	100	100	100	100	98	87
0.4	100	99	98	93	75	47
0.5	88	86	77	61	40	15
0.6	52	49	40	28	16	7
0.7	19	18	14	9	5	2

第三表 正規母集団 $N(\mu, \sigma)$ が検定に合格する確率

($\lambda \leq \varphi_0(m)$ なる領域で検定した場合)

1. 標本数 5

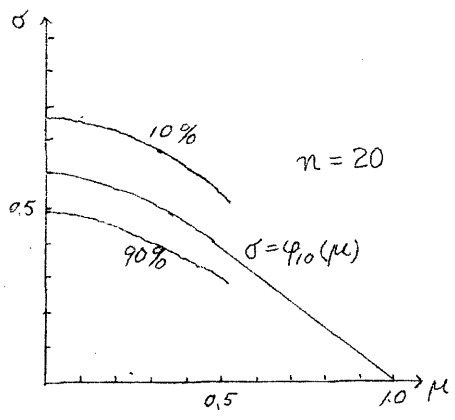
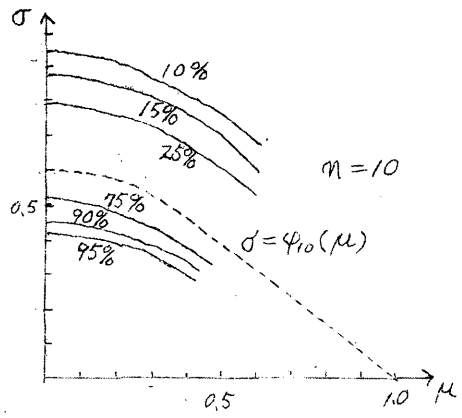
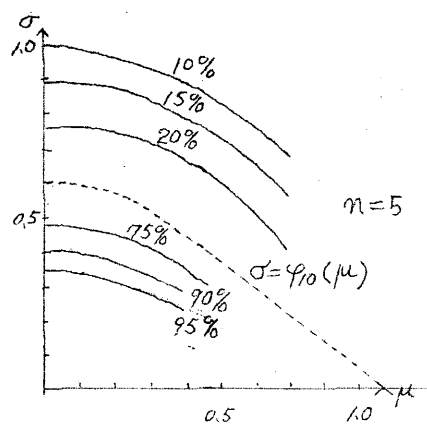
σ μ	0.0	± 0.1	± 0.2	± 0.3	± 0.4	± 0.5
0.1	100	100	100	100	100	100
0.2	100	100	100	100	98	92
0.3	98	97	94	88	78	62
0.4	81	79	75	64	52	38
0.5	56	54	49	42	34	25
0.6	36	34	31	27	22	16
0.7	22	22	20	18	14	11
0.8	13	13	12	11	9	7
0.9	9	8	8	7	7	5

2. 標本数 10

σ μ	0.0	± 0.1	± 0.2	± 0.3	± 0.4	± 0.5
0.1	100	100	100	100	100	100
0.2	100	100	100	100	100	96
0.3	100	100	99	94	83	62
0.4	90	88	81	70	54	33
0.5	63	58	48	36	24	14
0.6	28	27	22	17	11	7
0.7	13	12	10	8	5	3
0.8	5	5	4	3	3	2

3. 標本数 20

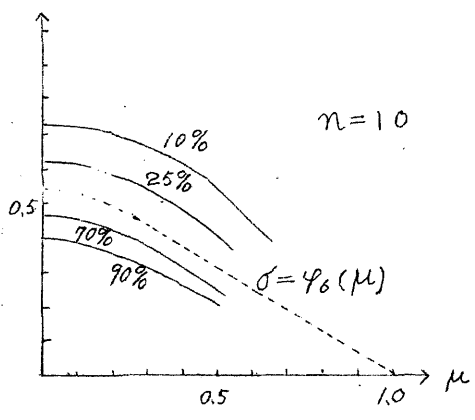
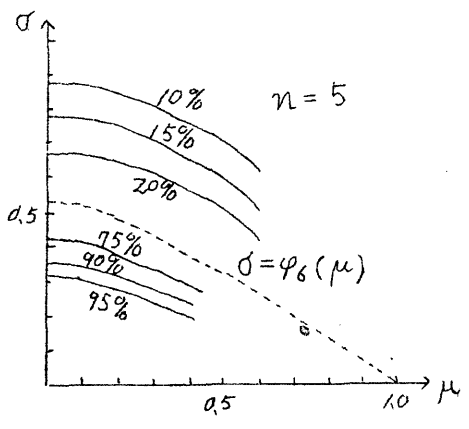
σ μ	0.0	± 0.1	± 0.2	± 0.3	± 0.4	± 0.5
0.3	100	100	100	100	98	88
0.4	96	94	87	71	47	24
0.5	62	58	46	30	16	6
0.6	23	21	16	10	5	2
0.7	6	5	4	3	1	1

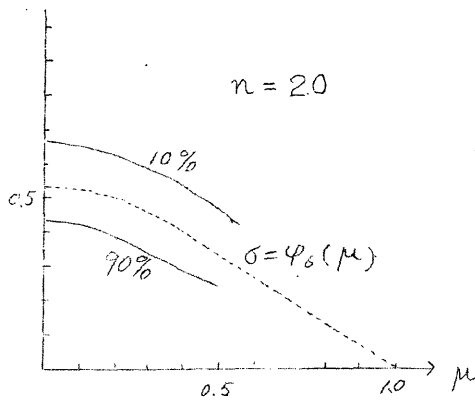


第二四

$\Delta \leq \varphi_{10}(\mu)$ により検定したとき
 検定に合格する確率が $\alpha\%$
 になる実の軌跡。 n は標本数

第三四 $\Delta \leq \varphi_6(m)$ により検定したとき合格する確率が
 $\alpha\%$ になる実の軌跡





実際の検定方式を決定する
ためにはオニ、三四を、オ一四
と見比べる必要がある。たと
えば $\sigma \leq \varphi_{10}(\mu)$ の領域で、
標本数10箇で検定するなら、
消費者としては、不良部分20
%にも及ぶ品が、確率15%強

で検定を通過する覚悟をしなければならぬ。製造者として
は不良部分が6%に過ぎない良品も約25%は検定で落ちること
を覚悟しなければならない。これを見てもわかるように標本数
が大でないとその検定はあまりよくない。特に不良品がかなり
の割合で検査に通るという欠点がある。これは標本数が小さい
ときに標準偏差の信頼限界が大であることによるものである。
そこで不良品が検査に通過するのを防ぐために検定の領域を厳
格にとれば、(例えば $\sigma \leq \varphi_2(\mu)$ で検査する)、良品を落とす確率
が増大し、製造水準が規格水準と接近しているときには、殆ん
どすべてが検定で落ちてしまう。標準偏差が未知、または変動
し、しかも製造の水準が規格の水準に近いときは、標本数を増
さねば安全な検定は不可能になるのである。

最後に一、二の注意をつけ加える。この検査では不良品がかな
りの割合で検定に通ることを述べたが、実際にさういう不良品
が少なければ、それが検定に通過する確率が大であっても、こ

の検定方式の欠点にはならない。即ち検定方式の決定には母集団パラメーターの現実における分布が必要であって、それなしには効果的な採取方式を決定できない。つぎに複合假説の検定に当って考えられることは、尤度比を使うことである。即ち μ, σ に条件をつけないときの *maximum likelihood* の比を λ とし、 $\sigma = \varphi_a(\mu)$ という条件をつけたときの *maximum likelihood* の値によつて検査するという考え方である。不幸にして、この場合は $\sigma = \varphi_a(\mu)$ が簡単な形をしていないから、 λ を μ と σ のうまい形で表わすことができません、行きつまつてしまう。

最後に、真空管材料規格の採取検査委員会の諸氏、特に委員長山内二郎先生の御指導に深く感謝する。また数値の計算に当っては、不完全 Γ 函数の数値表を借して下さった石田保士氏、計算に多大の助力を与えられた塩生コト君に感謝する。