On a Formulation of Classical Problems of Statistics[1]

Hirokichi Kudō (工藤弘吉)
Department of Mathematics, Faculty of Science,
Ochanomizu University.

In this paper we shall discuss the properties of a 'separation' function of hypotheses, which measures the degree of separating hypotheses by test, and then by using it a formulation of the classical problem of estimation is tried from the standpoint where a one-parameter group of transformations can be defined in a universe.

1. Notations. Let Ω be a space, whose elements will be denoted by $\omega, \omega_0, \omega_1, \ldots$, $\mathfrak{B}$ a Borel field of subsets of Ω, and $P(E)$, $P_0(E)$, $P_1(E)$, etc. probability measures on it. The direct product of n such measure-spaces $(\Omega, \mathfrak{B}, P)$ will be called an n-dimentional (measure-) space, and denoted by $(\Omega^{(n)}, \mathfrak{B}^{(n)}, P^{(n)})$. Thus the notations with a superscript (n) mean the corresponding ones on an n-dimensional space $(\Omega^{(n)}, \mathfrak{B}^{(n)}, P^{(n)})$. In this paper let Ω and $\mathfrak{B}$ be fixed, and on each measure P on it the following assumption will be assumed.

Assumption (1.1). *For any set $A \in \mathfrak{B}$ and any positive number p $(p < P(A))$, there exists a subset $B(\in \mathfrak{B})$ of A with P-measure p.*

2. The separation function of simple hypotheses. For any two measures P_0, and P_1 there exist, by Radon-Nikodym's theorem, a measurable function $f(\omega)$ and a measurable set S of P_0-measure zero, such that for each set $E(\in \mathfrak{B})$

$$P_1(E) = P_1(E \frown S) + \int_E f(\omega) dP_0(\omega).$$

Making use of the assumption (1.1) we can now define the family $\mathfrak{R}$ of measurable sets R_α so that *(1)* to each α $(0 \leqq \alpha \leqq 1)$ there corresponds one and only one R_α with P_0-measure α, *(2)* $\mathfrak{R}$ contains, for each α, just one set R_α of the form $\{\omega \mid f(\omega) \geqq k\} \smile S$ among those (if such sets exist) having equal P_0-measure α, *(3)* $\mathfrak{R}$ contains $R_1 = \Omega$ and the empty set R_0, *(4)* $R_\alpha \supset R_{\alpha'}$ if $\alpha > \alpha'$. This family will be called *a system of the best critical regions of P_0 against P_1.*

Consider the function $\gamma(\alpha ; P_1, P_0) = P_1(R_\alpha)^2$ of $\alpha(0 \leqq \alpha \leqq 1)$, which will be called a *separation function* of P_0 against P_1.[3] Then we have

Theorem (2.1). $\gamma(0 ; P_1, P_0) = 0$, $\gamma(1 ; P_1, P_0) = 1$, *and* $\gamma(\alpha ; P_1, P_0)$

---

[1] Contributions from Department of Mathematics, Faculty of Science, Ochanomizu University, No. 3.

[2] $1-\gamma(\alpha ; P_1, P_0)$ is the probability of an error of the second kind of the most powerful test of P_0 against P_1 whose error of the first kind is of probability α.

[3] Though $\mathfrak{R}$ is not unique, $\gamma(\alpha ; P_1, P_0)$ is unique for a pair P_0, P_1.

is a convex, non-decreasing and continuous function of α *in the interval* $(0<\alpha<1)$. *If, moreover,* P_1 *is absolutely continuous with respect to* P_0, *then* γ *is also continuous at* $\alpha=0$. *The graph of* $y=\gamma(x\,;P_0, P_1)$ *is always symmetrical with the graph of* $y=\gamma(x\,;P_1, P_0)$ *with respect to the line* $x+y=1$.

Further if the separation function $\gamma^{(n)}(\alpha; P_1, P_0)$ of $P_0^{(n)}$ against $P_1^{(n)}$ on the n-dimentional space is defined as above, then we have

Theorem (2.2). *If* $P_1(E)\neq P_0(E)$, *then*

$$\gamma^{(n)}(\alpha\,; P_1, P_0)\leqq\gamma^{(n+1)}(\alpha; P_1, P_0)\ \textit{for all}\ \alpha\ (0<\alpha\leqq1),$$

and

$$\lim_{n\to\infty}\gamma^{(n)}(\alpha; P_1; P_0)=1\ \textit{for all}\ \alpha\ (0<\alpha\leqq1).$$[4]

Remark. Let us write

$$\delta^{(n)}(P_1, P_0)=\max_{0\leqq\alpha\leqq1}\{\gamma^{(n)}(\alpha; P_1, P_0)-\alpha\},$$

then $\delta^{(n)}(P_1, P_0)$ satisfies the three conditions of distance, and enables us to introduce a metric, as a measure of separating hypotheses by test, into the set of hypotheses. But this distance is equivalent to the distance by the absolute variation of $P_0(E)-P_1(E)$.

3. **Dissipative group of transformations.** Let σ^s be a measurable one-one transformation on Ω whose inverse transformation is also measurable, $\mathfrak{G}$ a one-parameter group of σ^s (s being a parameter), and P a fixed probability measure on $(\Omega, \mathfrak{B})$. The transformation $\omega\to\sigma^s\omega=(\sigma^s\omega_1, \sigma^s\omega_2, \ldots, \sigma^s\omega_n)$ on an n-dimensional space $\Omega^{(n)}$ will be denoted by the same letter σ^s, where $\omega=(\omega_1, \omega_2, \ldots, \omega_n)$.

In this and the next sections, let us consider a set $\mathfrak{H}$ of probability measures $\sigma^sP(E)$ satisfying

$$\sigma^sP(E)=P(\sigma^{-s}E)\qquad\text{for every}\quad E\in\mathfrak{B},$$

or

$$\sigma^sP^{(n)}(E)=P^{(n)}(\sigma^{-s}E)^{5}\quad\text{for every}\quad E\in\mathfrak{B}^{(n)}$$

in the case of the n-dimensional space, and assume that the measures of $\mathfrak{H}$ are absolutely continuous each other. Then, from this assumption, follows

$$\sigma^sP(E)=\int_E f(\omega\,|\,s)\,dP(\omega)=\int_E\frac{f(\omega|s)}{f(\omega|s')}d\sigma^{s'}P(E)$$

where $f(\omega\,|\,s)$ is a non-negative measurable function of ω with a parameter s. By this function we can define the system $\mathfrak{R}(s, s')$ of the best critical regions of $\sigma^{s'}P$ against σ^sP for each pair s, s' $(s>s')$.

---

[4] S. Kakutani 1) has proved, in our notation, that $\gamma^{(\infty)}(\alpha; P_1^{(\infty)}, P_0^{(\infty)})\equiv1$, but this equation is slightly different from ours.

[5] It is clear that $\sigma^sP^{(n)}=(\sigma^sP)^{(n)}$.

Definition. If $\mathfrak{R}(s, s')$ is an invariant family for each choice of s and s', then the group $\mathfrak{G}$ is called *dissipative with respect to P.*

In this case $\mathfrak{R}(s, s')$ will be denoted simply by $\mathfrak{R}$ and the element of $\mathfrak{R}$ with P-measure α by $R_\alpha{}^*$.

Furthermore, consider the function $\gamma(\alpha|s)=\sigma^s P(R_\alpha{}^*)$ of α $(0\leqq\alpha\leqq 1)$ with a real parameter s. Then we have

Theorem (3.1).[6] *If $\mathfrak{G}$ is dissipative with respect to P, then $\gamma(\alpha|s)$ has the following properties: (1) if $s>0$, then $\gamma(\alpha|s)=\gamma(\alpha; \sigma^s P, P)$, if $s<0$, then $\gamma(\alpha|s)$ is the inverse function of $\gamma(\alpha; P, \sigma^s P)$, and $\gamma(\alpha|0)=\alpha$; (2) $\gamma(0|s)\equiv 0$, $\gamma(1|s)\equiv 1$, and it is a continuous, monotone increasing function of α $(0\leqq\alpha\leqq 1)$, and it is convex if $s>0$, and concave if $s<0$; (3) $\gamma(\gamma(\alpha|s)|s')=\gamma(\alpha|s+s')$, therefore $\gamma(\alpha|-s)$ is the inverse function of $\gamma(\alpha|s)$; (4) for each α $(0<\alpha<1)$, it is continuous and monotone increasing with respect to s, $\lim_{s\to\infty}\gamma(\alpha|s)=1$, and $\lim_{s\to-\infty}\gamma(\alpha|s)=0$; (5) it is differentiable with respect to α for each s, and its derivative will be denoted by $k(\alpha|s)$; (6) it has the derivative, denoted as $\varphi_\alpha(-s)$, with respect to s, and we have $\varphi_\alpha(s)=k(\alpha|-s)\varphi_\alpha(0)$; (7) for each α $(0<\alpha<1)$, $\varphi_\alpha(0)$ is a convex, positive function of α; (8) there exists a monotone function $\Phi(x)$ such that $\gamma(\alpha|s)=\Phi(\Phi^{-1}(\alpha)+s)$. (9)[7] let us put $\gamma^{(n)}(\alpha|s)=\sigma^s P^{(n)}(R_\alpha{}^{*(n)})$, then $\gamma^{(n)}(\alpha|s)$ is non-decreasing with n for $s>0$, and non-increasing with n for $s<0$, and moreover $\lim_{n\to\infty}\gamma^{(n)}(\alpha|s)=1\ (s>0)$, $\lim_{n\to\infty}\gamma^{(n)}(\alpha|s)=0\ (s<0)$, provided that $0<\alpha<1$.*

4. **α-estimate.** In this paragraph, let us suppose that $\mathfrak{G}$ is dissipative with respect to P, and consider the random variable $\hat{s}_\alpha(\omega)$, called an *α-estimate of s*, defined by

$$\hat{s}_\alpha(\omega)=\sup\{s\,|\,\sigma^s R_\alpha{}^*\ni\omega\}$$

where $R_\alpha{}^*$ is an element of the system $\mathfrak{R}$ defined in the preceding paragraph. Then we have

Theorem (4.1). *The α-estimate $\hat{s}_\alpha(\omega)$ has the following properties: (1) $\hat{s}_\alpha(\sigma^s\omega)=\hat{s}_\alpha(\omega)+s$; (2) let the distribution function of $\hat{s}_\alpha(\omega)$ be $D_\alpha(s'|s)=\sigma^s P(\omega|\hat{s}_\alpha(\omega)\leqq s')$ then $D_\alpha(s'|s)=1-\gamma(\alpha|s-s')$, therefore $D_\alpha(s'|s)=D_\alpha(s'-s|0)$; (3) the derivative of $D_\alpha(s|0)$ with respect to s exists and is equal to $\varphi_\alpha(s)$; (4) if $\mathfrak{G}$ is dissipative with respect to $P^{(n)}$ for every n, $\hat{s}_\alpha{}^{(n)}(\omega)$ is a consistent and sufficient estimate of s whatever α may be; (5) let one of the values of α maximizing $\varphi_\alpha(0)$ be α_0,[8] then $\hat{s}_{\alpha_0}(\omega)$ is a maximum likelihood estimate; (6) if $x(\omega)$ is a random variable with the distribution $F(x|s)=\sigma^s P(x(\omega)<x)$ which satisfies the equation $F(x|s)=F(x-s|0)$, and let*

---

[6] The properties (4)~(8) follow from (2) and (3), and an elementary proof of this fact is given by Mr. Turane Iwamura.

[7] In this case we were to assume that $\mathfrak{G}$ is dissipative with respect to $P^{(n)}$ for all integers n.

[8] The existence of this α_0 has been assumed by the property (7) of Theorem (3.1).

us put $e(\delta \mid x) = \sigma^s P(s-\delta < x(\omega) \leqq s+\delta) = P(-\delta < x(\omega) \leqq \delta)$, *then* $e(\delta \mid x) \leqq e(\delta \mid \hat{s}_{\alpha_0}(\omega))$. From the property (6) $\hat{s}_{\alpha_0}(\omega)$ is, in a sense, an efficient estimate of s. If, in particular, $x(\omega)$ and $\hat{s}_{\alpha_0}(\omega)$ are unbiased estimates and their second moments exist, the commonly used efficiency is deduced from this property.

5. Distribution with a dissipative group of transformations. In this paragraph we shall discuss what types of distributions will be allowed for a dissipative group of transformations.

Theorem (5.1). *If there exists a one-parameter group $\mathfrak{G}$ of transformations in $(\Omega, \mathfrak{B}, P)$, then $\mathfrak{G}$ is dissipative with respect to P if and only if $f(\omega \mid s)$ satisfies the equation* $F\left(\frac{f(\omega \mid s_1)}{f(\omega \mid s_2)}; s_1, s_2, s_3, s_4\right) = \frac{f(\omega \mid s_3)}{f(\omega \mid s_4)}$ *for all $\omega \in \Omega$, where $F(k; s_1, s_2, s_3, s_4)$ is a positive function of a positive variable k with four parameters s_1, s_2, s_3, s_4, such that (1) if $s_1 > s_2$, $s_3 > s_4$ or if $s_1 < s_2$, $s_3 < s_4$, then F is monotone increasing with k. (2) if $s_1 > s_2$, $s_3 < s_4$ or if $s_1 < s_2$, $s_3 > s_4$, then F is monotone decreasing with k. (3) if $s_3 = s_4$, then $F \equiv 1$ whatever s_1 and s_2 may be, (4) if $s_1 = s_2$, $s_3 \neq s_4$, then F is not defined, (5) if $s_1 = s_3$, $s_2 = s_4$, then $F(k; s_1, s_2, s_3, s_4) = k$, (6) $F(F(k; s_1, s_2, s_3, s_4); s_3, s_4, s_5, s_6) = F(k; s_1, s_2, s_5, s_6)$.*

Theorem (5.2). *If $\mathfrak{G}$ is defined in $(\Omega, \mathfrak{B}, P)$, then $\mathfrak{G}$ is dissipative with respect to $P^{(n)}$ on each n-dimensional space $(\Omega^{(n)}, \mathfrak{B}^{(n)}, P^{(n)})$ if and only if $f(\omega \mid s)$ has the form:* $\log f(\omega \mid s) = p(\omega)q(s) + r(s)$,[9] *where $p(\omega)$ is a real-valued function on Ω, $q(s)$, $r(s)$ also real valued functions of a real variable s, and moreover $q(s)$ is monotone.*

Theorem (5.3). *If there exists a group $\mathfrak{G}$ dissipative with respcct to $P^{(n)}$ on $(\Omega^{(n)}, \mathfrak{B}^{(n)}, P^{(n)})$ for every n, and if no element of $\mathfrak{G}$ makes P invariant, then there exists a function $u(\omega)$ satisfying $u(\sigma^s\omega) = u(\omega) - s$, for which $f(\omega \mid s)$ can be written in one of the following forms:*

(I) $\log f(\omega \mid s) = -2Asu(\omega) - As^2 + Bs$, $(A > 0)$

(II) $\log f(\omega \mid s) = -B(e^{As}-1)e^{Au(\omega)} + Cs$, $(B > 0,\ AC > 0)$,

where A, B, and C are constants.

Theorem (5.4). *Let $\{P_s(E)\}$ be a set of probability measures $P_s(E)$ absolutely continuous each other (s being a parameter), and represented as $P_s = \int f(\omega \mid s) dP_0$, where $f(\omega \mid s)$ is one of the two types (I) and (II) of Theorem (5.3), and $u(\omega)$ is a measurable function which is a one-one mapping of the whole space Ω onto the set of all the real numbers, then for all integers n a dissipative group of transformations with respect to $P_0^{(n)}$ can be defined such that $\sigma^s P_0(E) = P_s(E)$ for any set $E \in \mathfrak{B}$.*

Corollary. If $(\Omega, \mathfrak{B}, P_s)$ is a direct product of two measure spaces $(\Omega^*, \mathfrak{B}^*, P_s^*)$ and $(U, \mathfrak{U}, \Phi)$, where P_s^* satisfies the conditions of Theorem (5.4), then the same result as the above theorem (5.4) holds.

---

[9] cf. Koopman 2).

6. The general form of α-estimates and examples. In this paragraph we shall give α-estimates in the general form and some examples.

The next table shows the general forms of α-estimates.[10]

Type of hypothesis	Type (I)	Type (II)
$\log f(\omega \mid s)$	$-2Asu(\omega)-As^2+Bs,\ (A>0)$	$-B(e^{As}-1)e^{Au(\omega)}+Cs,\ (B,\ AC>0)$
$\hat{s}_\alpha(\omega)$	$-\frac{1}{N}\{\sum u(\omega_i)+\Phi^{-1}(\alpha)\}$	$-\frac{1}{A}\log\frac{B\sum e^{Au(\omega_i)}}{\Phi^{-1}(\alpha)}$
$\gamma^{(\alpha \mid s)}$	$\Phi(\Phi^{-1}(\alpha)+Ns)$	$\Phi(e^{As}\Phi^{-1}(\alpha))$
$\varphi_\alpha(s)$	$\sqrt{\frac{NA}{\pi}}\exp\left\{-NA\left(s-\frac{2A\Phi^{-1}(\alpha)-NB}{NA}\right)^2\right\}$	$\frac{\lvert A\rvert\,(\Phi^{-1}(\alpha))^{N\frac{C}{A}}}{\Gamma\left(N\frac{C}{A}\right)}\exp\{-e^{-As}\Phi^{-1}(\alpha)-NCs\}$
$\hat{s}_{\alpha_0}(\omega)$	$-\frac{1}{N}\sum u(\omega_i)+\frac{B}{2A}$	$\frac{-1}{A}\log\frac{AB}{NC}\sum e^{Au(\omega_i)}$
$\varphi_{\alpha_0}(s)$	$\sqrt{\frac{NA}{\pi}}\,e^{-NAS^2}$	$\frac{\lvert A\rvert(NC/A)^{NC/A}}{\Gamma\left(N\frac{C}{A}\right)}\exp\left\{-NCS-N\frac{C}{A}e^{-As}\right\}$
$\Phi(r)$	$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\sqrt{\frac{2A}{N}}\left(r-\frac{NB}{2A}\right)}e^{-\frac{t^2}{2}}dt$	$\frac{1}{\Gamma(NC/A)}\int_0^r t^{(NC/A)-1}e^{-t}\,dt$[11]

$\Phi^{-1}(\alpha)$ is the inverse function of $\Phi(r)$.

Examples of this estimation may be found in the problem of estimation of (a) the mean and (b) the variance of the normal distribution, (c) the mean of the exponential distribution, (d) m of the Pearson's distribution of the type III: $\frac{m^\lambda}{\Gamma(\lambda)}x^{\lambda-1}e^{-mx}$ $(x>0,\ m>0,\ \lambda>0)$, (e) of the type V: $\frac{m^\lambda}{\Gamma(\lambda)}x^{-\lambda-1}e^{-m/x}$ $(x>0,\ m>0,\ \lambda>0)$, (f) of the type VIII: $(1-m)(1+x)^{-m}$ $(-1<x<0,\ 1>m)$, (g) of the type IX: $(1+m)(1-x)^m$ $(0<x<1,\ -1>m)$, and (h) of the type XI: $(m-1)x^{-m}$ $(1<x,\ 1<m)$.

References

1) Kakutani, S. On equivalence of infinite product measures, Annals of Math., **49** (1948).

2) Koopman, B. O. On distributions admitting a sufficient statistic, Trans. Amer. Math. Soc., **39** (1936).

---

[10] The size of the sample ω is assumed to be N, that is, $\omega=(\omega_1,\dots,\omega_N)$, and $\sum$ indicates the summation over $i=1,2,\dots,N$.

[11] These limits of the integration are ones for a positive A, otherwise the integral is to be taken from r to ∞.