TESTS OF STATISTICAL HYPOTHESES WHICH ARE UNBIASED IN THE LIMIT

By J. Neyman

1. **Introduction.** The idea of unbiased tests of statistical hypotheses has been put forward and discussed in two recent papers.[1] Recently also a particular problem was solved introducing a test which has the property of being unbiased in the limit.[2] The purpose of the present note is to discuss this conception in its general form and to indicate methods of determining the tests unbiased in the limit of a broad class of simple statistical hypotheses. The notation and the terminology employed below are explained in the papers quoted.

2. **Notation and definitions.** Consider a set of n random variables

$$X_1, X_2, \cdots X_n \tag{1}$$

the particular values of which

$$x_1, x_2, \cdots x_n \tag{2}$$

can be given by observation and denote by Ω the set of hypotheses concerning the probability law of (1) which are regarded as admissible. We shall assume that all the hypotheses included in Ω specify the probability law of the X's having the same analytical form but differing among them in the value of just one parameter, θ. Thus, if E_n denotes the point (the "event point") in the space W_n of n dimensions with its coordinates equal to the values of (1) and w_n any region in W_n, then the probability of E_n falling within w_n, as determined by any of the hypotheses forming the set Ω will be denoted by

$$P\{E_n \epsilon w_n \mid \theta\} \tag{3}$$

and will be a function of the parameter θ. The probability (3) with fixed θ considered as a function of varying w_n is called the integral probability law of the X's. Frequently (3) is equal to the integral of a certain non-negative function of E_n over the region w_n. This function will always be denoted by $p(E_n \mid \theta)$ and called the elementary probability law of (1).

---

[1] J. Neyman and E. S. Pearson: Contributions to the Theory of Testing Statistical Hypotheses. Part I. Stat. Res. Memoirs, Vol. 1, (1936) pp. 1–37. Part II, ibid., Vol. II (1938).

J. Neyman: Sur la vérification des hypothèses statistiques composées. Bull. Soc. Math. de France, Vol. 63 (1935), pp. 246–266.

[2] J. Neyman: "Smooth" Test for Goodness of Fit. Skandinavisk Aktuarietidskrift, (1937), pp. 149–199.

Denote by H_0 some particular hypothesis of the set Ω and by θ_0 the value that it ascribes to the parameter θ.

A test of the statistical hypothesis H_0 consists in a rule of rejecting H_0 whenever E_n falls within a specified region w_n and in not doing so in other cases. The region w_n used for this purpose is called the critical region. It follows that to choose a test means to choose a critical region.

We shall consider below only cases such that for any region w_n the probability (3) considered as a function of θ possesses two successive derivatives.

DEFINITION 1. *If a critical region* $\bar{w}_n$ *has the property that,* α *being a fixed positive number.*

$$\text{(4)} \qquad \text{(a)} \quad P\{E_n \epsilon \bar{w}_n \mid \theta_0\} = \alpha$$

$$\text{(5)} \qquad \text{(b)} \quad \frac{d}{d\theta} P\{E_n \epsilon \bar{w}_n \mid \theta\} \Big|_{\theta=\theta_0} = 0$$

$$\text{(6)} \qquad \text{(c)} \quad \frac{d^2}{d\theta^2} P\{E_n \epsilon \bar{w}_n \mid \theta\} \Big|_{\theta=\theta_0} \geq \frac{d^2}{d\theta^2} P\{E_n \epsilon w_n \mid \theta\} \Big|_{\theta=\theta_0}$$

where w_n *is any region satisfying* (a) *and* (b), *then the region* $\bar{w}_n$ *is called the unbiased critical region of type A corresponding to the level of significance* α, *and the test of the hypothesis* H_0 *based on* $\bar{w}_n$, *the unbiased test of type A.*

This is the definition given in the first of the earlier papers quoted. Now we shall define the test which is unbiased in the limit. For this purpose we shall have to consider the situation where n is indefinitely increased and consequently we have a sequence of probability laws (3), a sequence of spaces W_n where they are defined and a sequence of regions $\bar{w}_n$, each $\bar{w}_n$ being a part of the corresponding W_n.

We must also introduce a varying scale with which to measure the differences $\theta - \theta_0$. This is due to the fact that, if the choice of the sequence of regions $\bar{w}_n$ is not very unlucky and $\theta \neq \theta_0$, then we shall frequently have

$$\text{(7)} \qquad \lim_{n \to \infty} P\{E_n \epsilon \bar{w}_n \mid \theta\} = 1$$

Comparing this with condition (4), we see that in general the limit of

$$P\{E_n \epsilon \bar{w}_n \mid \theta\}$$

for $n \to \infty$ will be discontinuous at $\theta = \theta_0$. To avoid this we shall measure $\theta - \theta_0$ in terms of $n^{-\frac{1}{2}}$ introducing instead of θ a new parameter ϑ connected with the former by means of the equality

$$\text{(8)} \qquad \theta = \theta_0 + \frac{\vartheta}{\sqrt{n}}$$

For the hypothesis tested H_0 we shall have $\vartheta = 0$ and $\vartheta \neq 0$ for any other hypothesis in Ω. The new parameter ϑ thus introduced will be called the standardized error in H_0. It will be frequently convenient to use θ but occasionally we shall use ϑ as well, for example writing $P\{E_n \epsilon w_n \mid \vartheta\}$ instead of (3) etc., and it is necessary to remember the connection (8) existing between θ and ϑ. It may be useful to notice at once that $df/d\theta = \sqrt{n}\, df/d\vartheta$.

DEFINITION 2. *We shall say that the sequence of regions*

$$\bar{w}_1, \bar{w}_2, \cdots, \bar{w}_n, \cdots \tag{9}$$

determines a test of the hypothesis H_0 which is unbiased in the limit and corresponds (in the limit) to the level of significance α, if for any n

$$\text{(d)} \quad \frac{d^2}{d\vartheta^2} P\{E_n \epsilon \bar{w}_n \mid \vartheta\}\Big|_{\vartheta=0} \geq \frac{d^2}{d\vartheta^2} P\{E_n \epsilon w_n \mid \vartheta\}\Big|_{\vartheta=0} \tag{10}$$

where w_n is any region such that

$$P\{E_n \epsilon w_n \mid \vartheta = 0\} = P\{E_n \epsilon \bar{w}_n \mid \vartheta = 0\} \tag{11}$$

and

$$\frac{d}{d\vartheta} P\{E_n \epsilon w_n \mid \vartheta\}\Big|_{\vartheta=0} = \frac{d}{d\vartheta} P\{E_n \epsilon \bar{w}_n \mid \vartheta\}\Big|_{\vartheta=0} \tag{12}$$

and if

$$\text{(e)} \quad \lim_{n\to\infty} P\{E_n \epsilon \bar{w}_n \mid \vartheta = 0\} = \alpha \tag{13}$$

$$\text{(f)} \quad \lim_{n\to\infty} \frac{d}{d\vartheta} P\{E_n \epsilon \bar{w}_n \mid \vartheta\}_{\vartheta=0} = 0 \tag{14}$$

The practical application of the test determined by the sequence of regions (7) consists in observing as large a number n of the X's of (1) and in rejecting the hypothesis H_0 whenever E_n falls within $\bar{w}_n$. If n is sufficiently large, then this rule will have about the same advantages as the application of the unbiased test of type A. In fact, allowing for the circumstance that the values of (11) and (12) will be only approximately equal to the limits (13) and (14), the properties of the test satisfying the Definition 2 will be as follows: If the hypothesis tested be true, it will be wrongly rejected with a relative frequency approximately equal to α fixed in advance. If H_0 is false and the true value say ϑ' of ϑ is not very different from zero, then the frequency of rejecting H_0 will be greater than α and could not be increased by applying some other similar test.

It may be useful to notice that in general there may be more than one test of the same hypothesis which is unbiased in the limit and corresponds to a fixed level of significance. Consequently there is a possibility of choosing between such tests, but it seems to the author that such a choice would require a previous strengthening of the theorem of S. Bernstein on which the present work is based.

3. **Theorem of S. Bernstein.** In the following, we shall have to use the following particular case of a theorem due to S. Bernstein.[3] Denote by $\mathcal{E}(x)$ the mathematical expectation of any variate x and by

$$
\begin{aligned}
&X_1, X_2, \cdots, X_n, \cdots \\
&Y_1, Y_2, \cdots, Y_n, \cdots
\end{aligned} \tag{15}
$$

two unlimited sequences of random variables.

We shall assume that

(1) X_i is independent of X_j and Y_j for any $i \neq j$.

(2) The following mathematical expectations exist and are independent of i:

$$
\begin{aligned}
&\mathcal{E}(X_i) = a \qquad \mathcal{E}(Y_i) = b \\
&\mathcal{E}(X_i - a)^2 = \sigma_1^2 \qquad \mathcal{E}(Y_i - b)^2 = \sigma_2^2 \\
&\mathcal{E}[(X_i - a)(Y_i - b)] = r\sigma_1\sigma_2 \\
&\mathcal{E}(| X_i - a |^3) = \mu \quad \mathcal{E}(| Y_i - b |^3) = \nu
\end{aligned} \tag{16}
$$

Consider now the space of $2n$ dimensions W_n and denote by E_n a point in it as determined by the values of X_i, Y_i for $i = 1, 2, \cdots n$ considered as its coordinates. Let u_n and v_n denote the sums

$$
u_n = \sum_{i=1}^{n} X_i, \qquad v_n = \sum_{i=1}^{n} Y_i \tag{17}
$$

and denote by D_n the point on a plane S with its orthogonal coordinates equal to u_n and v_n. If s is any region in S then let $P\{D_n \epsilon s\}$ be the probability of D_n falling within s.

Theorem of S. Bernstein. *If the variates* (15) *satisfy the conditions* (1) *and* (2) *then, for any* $\epsilon > 0$, *there exists a number* N_ϵ, *such that the inequality* $n > N_\epsilon$ *implies*

$$
\left| P\{D_n \epsilon s\} - \frac{1}{2\pi n\sigma_1\sigma_2\sqrt{1 - r^2}} \iint_s e^{-\frac{1}{2n(1-r^2)}\left(\frac{(u-na)^2}{\sigma_1^2} - 2r\frac{u-na}{\sigma_1}\frac{v-nb}{\sigma_2} + \frac{(v-nb)^2}{\sigma_2^2}\right)} du\, dv \right| < \epsilon, \tag{18}
$$

whatever the region s in S may be.

4. **Tests unbiased in the limit.** We shall consider the problem of determining the tests satisfying Definition 2, in the case where the following hypotheses are fulfilled.

---

[3] S. Bernstein: Sur un théorème limite du calcul des probabilités. Math. Ann., Bd. 97 (1926) p. 44.

See also V. Romanovskij, Bull. de l'Académie des Sciences de l'U. R. S. S., 1929, p. 209 and W. Kozakiewicz, Ann. Soc. Polonaise Math., t. XIII (1934), pp. 24–43.

(i) All the random variables (1) are mutually independent and each of them follows the same elementary probability law which we shall denote by $p(x_i \mid \theta)$.

(ii) The elementary probability law $p(x_i \mid \theta)$ admits three differentiations and two consecutive differentiations with respect to θ under the integral taken over any fixed finite or infinite interval, so that

$$\frac{d^k}{d\theta^k} \int_a^b p(x_i \mid \theta)\, dx_i = \int_a^b \frac{d^k}{d\theta^k} p(x_i \mid \theta)\, dx_i \tag{19}$$

for $k = 1, 2$.

(iii) If

$$\varphi_i = \left. \frac{\partial \log p(x_i \mid \theta)}{\partial \theta} \right|_{\theta=\theta_0} \quad \text{and} \quad \Psi_i = \left. \frac{\partial^2 \log p(x_i \mid \theta)}{\partial \theta^2} \right|_{\theta=\theta_0} \tag{20}$$

then we shall assume the existence of the following integrals all taken from $-\infty$ to $+\infty$

$$\sigma_1^2 = \int \varphi_i^2 p(x_i \mid \theta_0)\, dx_i \tag{21}$$

$$\sigma_2^2 = \int (\Psi_i + \sigma_1^2)^2 p(x_i \mid \theta_0)\, dx_i \tag{22}$$

$$r\sigma_1\sigma_2 = \int \varphi_i \Psi_i p(x_i \mid \theta_0)\, dx \tag{23}$$

$$\mu = \int \mid \varphi_i^3 \mid p(x_i \mid \theta_0)\, dx_i \tag{24}$$

$$\nu = \int \mid \Psi_i + \sigma_1^2 \mid^3 p(x_i \mid \theta_0)\, dx_i \tag{25}$$

Proposition I. *If the above conditions* (i), (ii) *and* (iii) *are satisfied,* Ψ_i *being a function of* x_i *and* $\mid r \mid < 1$, *then the sequence of regions* $\bar{w}_n$ *including all the points of* W_n *where* $p(E_n \mid \theta_0) = 0$ *and also those of the remaining ones which satisfy the inequality*

$$\sum_{i=1}^n \Psi_i + \left(\sum_{i=1}^n \varphi_i\right)^2 \geq M\sigma_2\sqrt{n(1-r^2)} - n\sigma_1^2 + r\frac{\sigma_2}{\sigma_1}\sum_{i=1}^n \varphi_i \tag{26}$$

where the coefficient M *is to be found from the equation*

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \left\{ e^{-\frac{1}{2}x^2} \frac{1}{\sqrt{2\pi}} \int_{M-Nx^2}^{\infty} e^{-\frac{1}{2}y^2}\, dy \right\} dx = \alpha \tag{27}$$

with

$$N = \frac{\sigma_1^2\sqrt{n}}{\sigma_2\sqrt{1-r^2}}, \tag{28}$$

defines a test of the hypothesis H_0, which is unbiased in the limit and corresponds (in the limit) to the level of significance α.

REMARK. The calculation of M satisfying the equation (27) is, of course, laborious. But a table of values of M corresponding to varying values of N is being constructed by N. L. Johnson at the Department of Statistics, University College, London, and it is hoped that it will soon be published.

To prove Proposition I, we must first prove (a) that whatever n, the region $\bar{w}_n$ determined by the inequality (26) satisfies the condition (d) in the definition 2. The proof is based on the following Lemma.[4]

LEMMA. *If $F_0, F_1, \cdots F_m$ are functions of $x_1, \cdots x_n$ integrable over any region in W_n and w_0 a region in W_n such that within w_0*

$$F_0 \geq \sum_{i=1}^{m} a_i F_i \tag{29}$$

while outside of w_0

$$F_0 \leq \sum_{i=1}^{m} a_i F_i \tag{30}$$

$a_1, a_2 \cdots a_m$ being some constant coefficients, then, whatever may be any other region w in W_n, such that

$$\int \cdots \int_w F_i dx_1 \cdots dx_n = \int \cdots \int_{w_0} F_i dx_1 \cdots dx_n, \quad \text{for } i = 1, 2, \cdots m, \tag{31}$$

we shall have

$$\int \cdots \int_{w_0} F_0\, dx_1 \cdots dx_n \geq \int \cdots \int_w F_0\, dx_1 \cdots dx_n. \tag{32}$$

PROOF OF PROPOSITION I. Denote, for simplicity, by $p(E_n)$ the elementary probability law of the X's as determined by the hypothesis tested. Comparing the statement of the Lemma with the definition (26) of $\bar{w}_n$, we immediately see that this region has the following property: whatever may be any other region w in W_n such that

$$\int\int_w p(E_n)\, dx_1 \cdots dx_n = \int \cdots \int_{\bar{w}_n} p(E_n)\, dx_1 \cdots dx_n \tag{33}$$

and

$$\frac{1}{\sqrt{n}} \int \cdots \int_w \sum_{i=1}^{n} \varphi_i p(E_n)\, dx_1 \cdots dx_n = \frac{1}{\sqrt{n}} \int \cdots \int_{\bar{w}_n} \sum_{i=1}^{n} \varphi_i p(E_n)\, dx_1 \cdots dx_n \tag{34}$$

[4] J. Neyman and E. S. Pearson: loc. cit., pp. 10–11.

we shall have

$$(35)\quad \begin{aligned} &\frac{1}{n}\int\cdots\int_{\bar{w}_n}\left(\sum_{i=1}^{n}\Psi_i+\left(\sum_{i=1}^{n}\varphi_i\right)^2\right)p(E_n)\,dx_1,\cdots dx_n \\ &\qquad\ge\frac{1}{n}\int\cdots\int_{w}\left(\sum_{i=1}^{n}\Psi_i+\left(\sum_{i=1}^{n}\varphi_i\right)^2\right)p(E_n)\,dx_1\cdots dx_n \end{aligned}$$

But under the conditions (i) and (ii)

$$(36)\quad p(E_n\mid\vartheta)=\prod_{i=1}^{n}p(x_i\mid\theta_0+\vartheta/\sqrt{n})$$

$$(37)\quad \left.\frac{\partial p(E_n\mid\vartheta)}{\partial\vartheta}\right|_{\vartheta=0}=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\varphi_i p(E_n)$$

$$(38)\quad \left.\frac{\partial^2 p(E_n\mid\vartheta)}{\partial\vartheta^2}\right|_{\vartheta=0}=\frac{1}{n}\left(\sum_{i=1}^{n}\Psi_i+\left(\sum_{i=1}^{n}\varphi_i\right)^2\right)p(E_n)$$

and it is easily seen that the relations (33), (34) and (35) are identical with (11), (12) and (10) respectively and that therefore the region $\bar{w}_n$ satisfies the condition (d) of definition 2. It remains to prove that $\bar{w}_n$ satisfies also the conditions (e) and (f), that is to say that, for $n\to\infty$, the formulas in the right hand sides of (33) and (34) tend to the prescribed values α and zero respectively. This conclusion concerning (33) is a consequence of the theorem of S. Bernstein, quoted above. To see this, write

$$(39)\quad u_n=\sum_{i=1}^{n}\varphi_i,\qquad v_n=\sum_{i=1}^{n}\Psi_i$$

and denote by s_0 the region in the plane S of (u, v) defined by the inequality

$$(40)\quad v+u^2\ge M\sigma_2\sqrt{n(1-r^2)}-n\sigma_1^2+r\frac{\sigma_2}{\sigma_1}u$$

obtained from (26) by means of (39). The right hand side of (33) represents the probability determined by the hypothesis tested of the X's satisfying the inequality (26). But this is satisfied simultaneously with the variates u_n and v_n satisfying (40). Therefore, if we denote by D_n the point in S with its coordinates equal to (39), then the right hand side of (33) may be interpreted as the probability $P\{D_n\epsilon s_0\}$ of D_n falling within s_0. Comparing (21)–(25) with (16), it is easily seen that, according to the Theorem of S. Bernstein, whatever may be $\epsilon>0$, if n is sufficiently large, then

$$(41)\quad \left|P\{D_n\epsilon s_0\}-\int\int_{s_0}G_n\,dv\right|<\epsilon$$

where

$$(42)\quad G_n=\frac{1}{2\pi n\sigma_1\sigma_2\sqrt{1-r^2}}e^{-\frac{1}{2n(1-r^2)}\left\{\frac{u^2}{\sigma_1^2}-2r\frac{u}{\sigma_1}\frac{v+n\sigma_1^2}{\sigma_2}+\frac{(v+n\sigma_1^2)^2}{\sigma_2^2}\right\}}$$

In fact, to what is given explicitly, we must only add that as

$$\int_{-\infty}^{+\infty} p(x_i \mid \theta)\, dx_i = 1 \tag{43}$$

the derivative with respect to θ of the left hand side must be identically equal to zero. Therefore

$$\frac{d}{d\theta} \int p(x_i \mid \theta)\, dx_i \Big|_{\theta=\theta_0} = \int \varphi_i p(x_i \mid \theta_0)\, dx_i = \mathcal{E}(\varphi_i) = 0 \tag{44}$$

where again the integrals are taken from $-\infty$ to $+\infty$. It follows further that the second derivative with respect to θ of (43) must be again identically equal to zero. Therefore, keeping in mind the definitions of φ_i and Ψ_i, we may write

$$\frac{\partial^2}{\partial\theta^2} \int p(x_i \mid \theta)\, dx_i \Big|_{\theta=\theta_0} = \int (\Psi_i + \varphi_i^2) p(x_i \mid \theta_0)\, dx_i = 0 \tag{45}$$

and thus

$$\mathcal{E}(\Psi_i) = -\mathcal{E}(\varphi_i^2) = -\sigma_1^2 \tag{46}$$

The proof that the right hand side of (24) tends to α with $n \to \infty$ will be completed if we manage to reduce the integral of (42) over the region s_0 to the integral (27). This is easily done by substituting

$$x = \frac{u}{\sigma_1\sqrt{n}} \tag{47}$$

$$y = \frac{v + n\sigma_1^2 - r\sigma_2 u/\sigma_1}{\sigma_2\sqrt{n(1 - r^2)}} \tag{48}$$

Thus, if the coefficient M in (26) and (40) satisfies the condition (27), then the value of the integral of G_n in (41) is permanently equal to α and this means that the right hand side of (33) tends to α as $n \to \infty$.

Denote by $p_n(u, v)$ the elementary probability law of u_n and v_n. It will be noticed that, whatever s in S

$$P\{D_n \in s\} = \iint_s p_n(u, v)\, du\, dv \tag{49}$$

and that consequently in the course of the above discussion we have proved that, whatever $\epsilon > 0$, there exists a sufficiently large number N_ϵ such that $n > N_\epsilon$ implies

$$\left| \iint_s (p_n(u, v) - G_n)\, du\, dv \right| < \epsilon \tag{50}$$

whatever may be the region s in S. We shall now use this circumstance to prove that, when $n \to \infty$, the right hand side of (34) tends to zero. It will be noticed first that

$$\int \cdots \int_{\bar{w}_n} \left(\sum \varphi_i\right)^k p(E_n)\, dx_1 \cdots dx_n = \iint_{s_0} u^k p_n(u, v)\, du\, dv \tag{51}$$

for $k = 1, 2$. Further

$$\iint_{s_0} u^2 p_n(u, v)\, du\, dv \leq \iint_S u^2 p_n(u, v)\, du\, dv = n\sigma_1^2 \tag{52}$$

Using the inequality of Schwartz,[5] we may write

$$\begin{aligned} &\left| \frac{1}{\sqrt{n}} \iint_{s_0} u(p_n(u, v) - G_n)\, du\, dx \right| \\ &\qquad \leq \frac{1}{\sqrt{n}} \left(\iint_{s_0} u^2 \left| p_n(u, v) - G_n \right| du\, dv \iint_{s_0} \left| p_n(u, v) - G_n \right| du\, dv \right)^{\frac{1}{2}} \end{aligned} \tag{53}$$

Now, it is easy to calculate that

$$\iint_{s_0} u^2 \left| p_n(u, v) - G_n \right| du\, dv \leq 2n\sigma_1^2 \tag{54}$$

On the other hand, if n is so large that (50) holds good for any region s in S and s_+ and s_- denote the two parts of s_0 where $p_n(u, v) - G_n$ is respectively positive and negative, then

$$\begin{aligned} 0 \leq \iint_{s_0} \left| p_n(u, v) - G_n \right| du\, dv &= \iint_{s_+} (p_n(u, v) - G_n)\, du\, dv \\ &\quad - \iint_{s_-} (p_n(u, v) - G_n)\, du\, dv < 2\epsilon \end{aligned} \tag{55}$$

and it follows that, for such large values of n,

$$\left| \frac{1}{\sqrt{n}} \iint_{s_0} u(p_n(u, v) - G_n)\, du\, dv \right| \leq 2\sigma_1 \sqrt{\epsilon} \tag{56}$$

On the other hand, using the transformation (47) and (48), we find that

$$\frac{1}{\sigma_1 \sqrt{n}} \iint_{s_0} u G_n\, du\, dv = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \left\{ x e^{-\frac{1}{2}x^2} \frac{1}{\sqrt{2\pi}} \int_{M - Nx^2}^{\infty} e^{-\frac{1}{2}y^2}\, dy \right\} dx \tag{57}$$

and consequently is permanently equal to zero. As ϵ is an arbitrarily small number, it follows that

$$\lim_{n \to \infty} \frac{1}{\sqrt{n}} \iint_{s_0} u p_n(u, v)\, du\, dv = \lim_{n \to \infty} \frac{1}{\sqrt{n}} \iint_{\bar{w}_n} \sum_{i=1}^{n} \varphi_i p(E_n)\, dx_1 \cdots dx_n = 0 \tag{58}$$

which fulfills the proof of Proposition I.

[5] See for example: S. Kaczmarz and H. Steinhaus, Theorie der Orthogonalreihen, Warsaw, 1935, p. 10.

PROPOSITION II. *If the conditions of Proposition I are satisfied but either* $|r| = 1$ *or* Ψ_i *is independent of* x_i, *then the test of the hypothesis* H_0 *which is unbiased in the limit and which corresponds, to the level of significance* α, *is determined by the sequence of critical regions* $\bar{w}_n$, *defined by the inequality*

$$\left|\sum_{i=1}^{n} \varphi_i\right| \geq \lambda \sigma_1 \sqrt{n} \tag{59}$$

where λ *satisfies the equation*

$$\frac{1}{\sqrt{2\pi}} \int_{-\lambda}^{+\lambda} e^{-\frac{1}{2}x^2}\, dx = 1 - \alpha \tag{60}$$

PROOF. We notice first that the condition $|r| = 1$ and the equation (44) imply

$$\left(\int \varphi_i(\Psi_i + \sigma_1^2) p(x_i \mid \theta_0)\, dx_i\right)^2 = \int \varphi_i^2 p(x_i \mid \theta_0)\, dx_i \int (\Psi_i + \sigma_1^2)^2 p(x_i \mid \theta_0)\, dx_i \neq 0 \tag{61}$$

or

$$\frac{\int \varphi_i(\Psi_i + \sigma_1^2) p(x_i \mid \theta_0)\, dx_i}{\int \varphi_i^2 p(x_i \mid \theta_0)\, dx_i} = \frac{\int (\Psi_i + \sigma_1^2)^2 p(x_i \mid \theta_0)\, dx_i}{\int \varphi_i(\Psi_i + \sigma_1^2) p(x_i \mid \theta_0)\, dx_i} = A \neq 0 \tag{62}$$

and therefore

$$\int \{(\Psi_i + \sigma_1^2)^2 - A\varphi_i(\Psi_i + \sigma_1^2)\} p(x_i \mid \theta_0)\, dx_i = 0 \tag{63}$$

$$\int \{\varphi_i(\Psi_i + \sigma_1^2) - A\varphi_i^2\} p(x_i \mid \theta_0)\, dx_i = 0 \tag{64}$$

and finally

$$\int (\Psi + \sigma_1^2 - A\varphi_i)^2 p(x_i \mid \theta_0)\, dx_i = 0 \tag{65}$$

which means that at almost every value of x_i for which $p(x_i \mid \theta_0) \neq 0$,

$$\Psi_i + \sigma_1^2 = A\varphi_i \tag{66}$$

It follows that the inequality (10) in the definition 2 of the test which is unbiased in the limit reduces to the following

$$\frac{1}{n} \int \cdots \int_{\bar{w}_n} \left(\sum \varphi_i\right)^2 p(E_n \mid \theta_0)\, dx_1 \cdots dx_n \geq \frac{1}{n} \int \cdots \int_{w} \left(\sum \varphi_i\right)^2 p(E_n \mid \theta_0)\, dx_1 \cdots dx_n \tag{67}$$

owing to (11), (12), (37) and (38). On the other hand, the inequality (59) is equivalent to

$$(68) \qquad (\sum \varphi_i)^2 \geq a + b \sum \varphi_i$$

with $a = \lambda^2\sigma_1^2 n$ and $b = 0$. Referring to the Lemma, we conclude that the regions $\bar{w}_n$ satisfy the condition (d) of the Definition 2. It remains to show that they satisfy also the conditions (e) and (f). This immediately follows from the theorem of Liapounoff[6] and the reasoning which we used above in order to prove (58).

If Ψ_i does not depend on x_i then, owing to (38) and (11), the inequality (10) immediately reduces to (67) and the proof of Proposition II follows exactly the same lines as before.

5. Limiting power function. To know the properties of a test undoubtedly means to know (i) how frequently this particular test will reject the hypothesis tested when it is in fact true and (ii) how frequently will it detect its falsehood when it is wrong. The information of this kind is provided by the properties of the so called power function of the test. This has been defined[7] as follows. Let w_n be any critical region and, as formerly, $P\{E_n \epsilon w_n \mid \theta\}$ the probability of E_n falling within w_n as determined by a specified value of θ. If w_n is fixed, then $P\{E_n \epsilon w_n \mid \theta\}$ will be a function of θ only. To emphasize this circumstance we may introduce a new symbol, writing

$$(69) \qquad P\{E_n \epsilon w_n \mid \theta\} = \beta(\theta \mid w_n\}$$

which will mean that in the above formula w_n is kept constant and θ varied. The function $\beta(\theta \mid w_n)$ thus defined is called the power function of the critical region w_n or that of the test based on w_n. If w_n corresponds to the level of significance α and θ_0 is the value of θ specified by the hypothesis tested H_0, then

$$(70) \qquad \beta(\theta_0 \mid w_n) = \alpha$$

and it will be noticed that this is the probability of rejecting H_0 when it is in fact true. As we reject H_0 only in such cases when $E_n \epsilon w_n$, the values of $\beta(\theta \mid w_n)$ corresponding to other values of $\theta \neq \theta_0$ are equal to the probability of detecting the falsehood of the hypothesis H_0 when θ has any specified value different from θ_0. The larger the value of $\beta(\theta \mid w_n)$ at a given θ, the greater will be the "detecting power" of the test, which justifies the name attached to the function $\beta(\theta \mid w_n)$. Until the present time the power function of only a few tests has been studied and it follows that we know comparatively little of the properties of the tests even if they are in frequent use. The first study of this kind was concerned with the power function of the "Student's" test as applied to the problem of one sample and there are three publications giving various

[6] See for example Paul Lévy: Théorie de l'addition des variables aléatoires. Paris, 1937. Pp. 101–107.

[7] J. Neyman and E. S. Pearson: loc. cit., p. 9.

numerical tables.[8] However, in these publications the term "power function" does not appear yet. Apart from the joint paper already referred to where the term "power function" was first defined, we may mention a few papers in *Biometrika*, the most important of which seems to be that by S. S. Wilks and Catherine M. Thompson.[9] The purpose of studying the power function of any test is to be able to answer the following three questions:

(a) What should be the size of a sample in order to have a reasonable chance of detecting the falsehood of the hypothesis tested, when the error in the parameters that it specifies has some stated value?

(b) If in some particular case a test failed to reject the hypothesis tested (which, of course, does not mean that it is necessarily true), is it likely that the error in θ_0 does not exceed some specified limit Δ?

(c) Two different tests corresponding to the same level of significance are suggested for the same hypothesis H_0, which shall we use?

In this last case the answer is obvious—the one which gives the greater chance of detecting the falsehood of the hypothesis tested in cases when it is wrong. But to know this we must know the power functions of both tests.

For the above reasons it seems to be important to study the power function of the test unbiased in the limit as defined above. It is obvious that, as in this case the elementary probability laws are not specified, it is impossible to find the actual explicit formula giving the power function. Therefore we shall endeavour to find its limiting form. This will be done by means of the two following theorems.

Consider an infinite sequence of situations

$$S_1, S_2, \cdots S_m, \cdots \tag{71}$$

In each of these situations we shall have to test the same hypothesis H_0 concerning the probability law $p(x \mid \theta)$ and specifying the value θ_0 of θ. The situations differ among themselves by the number of the X's and by the hypotheses, alternative to H_0, which are considered. For the situation S_m we shall denote them by n_m and H_m respectively. We shall assume that $\lim n_m = +\infty$ when $m \to \infty$. As to the hypothesis H_m, we shall assume that the value θ_m which it ascribes to the parameter θ is

$$\theta_m = \theta_0 + \frac{\vartheta}{\sqrt{n_m}} \tag{72}$$

---

[8] (1) S. Kołodziejczyk: Sur l'erreur de la seconde catégorie dans le problème de "Student." C. R. Academie des Sciences, Paris, t. 197 (1933) p. 814.

(2) J. Neyman with co-operation of K. Iwaszkiewicz and S. Kołodziejczyk: Statistical Problems in Agricultural Experimentation. Suppl. Journ. Roy. Stat. Soc. Vol. II (1935) pp. 107–180.

(3) J. Neyman and B. Tokarska: Errors of Second Kind in Testing "Student's" Hypothesis. J. A. S. A., Vol. 31 (1936) pp. 320–334.

[9] S. S. Wilks and Catherine M. Thompson: The Sampling Distribution of the Criterion λ_H, when the Hypothesis Tested is not true. Biometrika, Vol. XXIX (1937), pp. 124–132.

where ϑ, the standardized error in θ_0, is kept constant. We shall assume that in each situation S_m we test the hypothesis H_0 by means of the test unbiased in the limit and corresponding to the level of significance α. The power function of this test should be denoted by $\beta(\theta \mid \bar{w}_{n_m})$, but to simplify the notation we will write simply $\beta_m(\theta)$. We shall be concerned with the value of this function $\beta_m(\theta_m)$ at the point $\theta = \theta_m$ and we shall prove the following proposition.

PROPOSITION III. *If the third logarithmic derivative of $p(x_i \mid \theta)$ with respect to θ is bounded*

$$\left| \frac{\partial^3 \log p(x_i \mid \theta)}{\partial \theta^3} \right| < C = \text{constant}, \tag{73}$$

and $| r | < 1$, then

$$\lim_{m \to \infty} \beta_m(\theta_m) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \left\{ e^{-\frac{1}{2}(x - \vartheta\sigma_1)^2} \frac{1}{\sqrt{2\pi}} \int_{M - Nx^2}^{\infty} e^{-\frac{1}{2}v^2} \, dy \right\} dx \tag{74}$$

This proposition is analogous to that[10] concerning the "smooth" test for goodness of fit. It could be used in the following manner.

When testing the hypothesis H_0 and using for the purpose a certain number n of observations, we find ourselves in a situation which might be considered as one of the sequence (71). If n is large, we may hope that the right hand side of (74) will give a reasonable approximation to the actual value of the power function corresponding to the value of θ to be calculated from (72) by substituting in it $n_m = n$.

PROOF. Denote

$$\frac{\partial^3 \log p(x_i \mid \theta)}{\partial \theta^3} = \chi_i(\theta) \tag{75}$$

We may write

$$p(x_i \mid \theta_m) = p(x_i \mid \theta_0) e^{\frac{\vartheta \varphi_i}{\sqrt{n_m}} + \frac{\vartheta^2 \Psi_i}{2 n_m} + \frac{\vartheta^3 \chi(\theta_i')}{6 n_m^{3/2}}} \tag{76}$$

where θ_i' denotes some value intermediate between θ_0 and θ_m. Consequently, taking into account (39), (47) and (48), we have

$$p(E_{n_m} \mid \theta_m) = \prod_{i=1}^{n_m} p(x_i \mid \theta_m) = p(E_{n_m} \mid \theta_0)(1 + \epsilon_m) e^{x\vartheta\sigma_1 - \frac{1}{2}\vartheta^2\sigma_1^2} \tag{77}$$

where

$$\log(1 + \epsilon_m) = \frac{1}{\sqrt{n_m}} \left(\frac{1}{2} \vartheta^2 \sigma_2 (y\sqrt{1 - r^2} + xr) + \vartheta^3 \frac{\sum_{i=1}^{n_m} \chi(\theta_i)}{n_m} \right) \tag{78}$$

---

[10] J. Neyman: "Smooth" Test for Goodness of Fit. Skandinavisk Aktuarietidskrift, (1937), p. 186.

It is seen that, if $m \to \infty$ then ϵ_m tends to zero, uniformly in every bounded region of the plane, S, of x and y. Denote by s any bounded region in S and by $W_m(s)$ a region in W_{n_m} of which s is a transformation by means of the formulae (39), (47) and (48). The probability of E_{n_m} falling within $W_m(s)$ is equal to that of the point with coordinates x and y falling within s. The former of these probabilities is represented by the integral of (77) over $W_m(s)$ and the latter by the integral taken over s of the elementary probability law $p_m(x, y \mid \theta_m)$ of x and y, corresponding to the value θ_m of θ. Owing to the formula (77) we may write

$$p_m(x, y \mid \theta_m) = p_m(x, y \mid \theta_0)(1 + \eta_m)e^{x\vartheta\sigma_1 - \frac{1}{2}\vartheta^2\sigma_1^2} \tag{79}$$

where, owing to (78), η_m tends uniformly to zero in s as $m \to \infty$. Remembering the connection between u_n, v_n and x, y and also the inequality (56), which is valid for sufficiently large values of n, we conclude that

$$p_m(x, y \mid \theta_0) = \frac{1}{2\pi} e^{-\frac{1}{2}(x^2+y^2)} + Q_m \tag{80}$$

where Q_m has the property that, whatever be $\epsilon > 0$, for sufficiently large values of m

$$\left| \int\int_s Q_m \, dx \, dy \right| < \epsilon \tag{81}$$

where s is any bounded region in S. It follows that

$$p_m(x, y \mid \theta_m) = \frac{1 + \eta_m}{2\pi} e^{-\frac{1}{2}\{(x-\vartheta\sigma_1)^2+y^2\}} + Q_m e^{x\vartheta\sigma_1 - \frac{1}{2}\vartheta^2\sigma_1^2}(1 + \eta_m) \tag{82}$$

and that therefore, whatever be the bounded region s

$$\lim_{m\to\infty} \int\int_s p_m(x, y \mid \theta_m) \, dx \, dy = \frac{1}{2\pi} \int\int_s e^{-\frac{1}{2}\{(x-\vartheta\sigma_1)^2+y^2\}} \, dx \, dy \tag{83}$$

It is known however, that whenever an integral probability law tends to a fixed limit uniformly within any bounded region, then it must do so within the whole space. It follows therefore that the formula (83) is valid for any region s whether bounded or not. But

$$\beta_m(\theta_m) = \int\int_{y > M - Nx^2} p_m(x, y \mid \theta_m) \, dx \, dy \tag{84}$$

and it follows that

$$\lim_{m\to\infty} \beta_m(\theta_m) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \left\{ e^{-\frac{1}{2}(x-\vartheta\sigma_1)^2} \frac{1}{\sqrt{2\pi}} \int_{M-Nx^2}^{+\infty} e^{-\frac{1}{2}y^2} \, dy \right\} dx, \tag{85}$$

which completes the proof of Proposition III.

It is important to be clear about the exact meaning of the Proposition III. Suppose for example that in a particular case $\vartheta = \sigma_1 = 1$ and consider a sequence of situations in which

$$(86)\qquad \begin{cases} n_1 = 100, & n_2 = 100^2, \quad \cdots n_m = 100^m, \cdots \\ \theta_1 = \theta_0 + \cdot 1, & \theta_2 = \theta_0 + \cdot 01, \cdots \theta_m = \theta_0 + (\cdot 1)^m, \cdots \end{cases}$$

If this were the case, then the Proposition III would be applicable and we could affirm that the sequence of the power functions $\beta_m(\theta)$, each considered at the appropriate point θ_m, has a limit, represented by the double integral in the right hand side of (85) with $\vartheta\sigma_1 = 1$. Accordingly, if we were interested in the value of the power function at $\theta' = \theta_0 + \cdot 02$ with $n = 10000$ and $\theta_1 = 1$, then we could hope to obtain its approximate value calculating the double integral in (85) with

$$(87)\qquad \vartheta = (\theta' - \theta_0)\sqrt{n} = 2$$

These are legitimate conclusions. However, it would be wrong to consider as proved that, if in the same example we increase the size of n to $n' = 40000$, then the value of the power function at $\theta = \theta'$ will be represented by its limit (85) with $\vartheta = 4$ and with about the same accuracy as previously. It is just possible that to attain the same accuracy at $\vartheta = 4$ a value of n greater than n' will be needed. This of course would imply a corresponding change in θ'.

Proposition IV. *If the conditions of Proposition III are satisfied but either* $|r| = 1$ *or* Ψ_i *is independent of* x_i, *then*

$$(88)\qquad \lim_{m\to\infty} \beta_m(\theta_m) = 1 - \frac{1}{\sqrt{2\pi}} \int_{-\lambda}^{+\lambda} e^{-\frac{1}{2}(x-\vartheta\sigma_1)^2}\, dx$$

The proof of this proposition is quite analogous to that of Proposition III.

6. Examples.

Example 1. Consider the case where it is known for certain that

$$(89)\qquad p(x_i|\theta) = \frac{1}{\pi}\,\frac{1}{1 + (x - \theta)^2}, \quad \text{for } -\infty < x < \infty$$

but where the actual value of θ is doubtful and it is desired to test the hypothesis H_0 that $\theta = \theta_0 = 0$, the alternative possibilities being both $\theta < 0$ and $0 < \theta$. Before applying the test unbiased in the limit it is natural to try the unbiased test of type A. The critical region w_0 of this test is defined by the inequality

$$(90)\qquad \sum_{i=1}^{n} \Psi_i + \left(\sum_{i=1}^{n} \varphi_i\right)^2 \geq a + b \sum_{i=1}^{n} \varphi_i$$

where the constants a and b must be found so as to satisfy the conditions

$$(91)\qquad \int \cdots \int_{w_0} p(E_n|\theta_0)\, dx_1 \cdots dx_n = \alpha$$

$$(92)\qquad \int \cdots \int_{w_0} \sum_{i=1}^{n} \varphi_i p(E_n|\theta_0)\, dx_1 \cdots dx_n = 0$$

The technical difficulties involved in this problem are considerable and this may induce us to apply the test unbiased in the limit. Following the above theory we have

$$\varphi_i = \frac{2x_i}{1 + x_i^2} \tag{93}$$

$$\Psi_i = \frac{4x_i^2}{(1 + x_i^2)^2} - \frac{2}{1 + x_i^2} \tag{94}$$

$$\chi_i(\theta) = \frac{16(x_i - \theta)^3}{(1 + (x_i - \theta)^2)^3} - \frac{12(x_i - \theta)}{(1 + (x_i - \theta)^2)^2} \tag{95}$$

It is easily seen that all the limiting conditions of the theory are fulfilled and that, in particular $|\ \chi_i(\theta)\ |$ cannot exceed a fixed limit, approximately equal to 3. We have further

$$\mathcal{E}(\varphi_i^2) = -\mathcal{E}(\Psi_i) = \frac{4}{\pi}\int_{-\infty}^{+\infty} \frac{x_i^2}{(1 + x_i^2)^3}\, dx_i = \tfrac{1}{2} = \sigma_1^2 \tag{96}$$

Similarly

$$\mathcal{E}\{\Psi_i + \sigma_1^2)^2\} = \tfrac{5}{8} = \sigma_2^2 \tag{97}$$

$$\mathcal{E}(\varphi_i \Psi_i) = 0 = r \tag{98}$$

It follows that the regions $\bar{w}_n$, the sequence of which determines the test which is unbiased in the limit, are defined by the inequality

$$4\sum_{i=1}^{n} \frac{x_i^2}{(1 + x_i^2)^2} - 2\sum_{i=1}^{n} \frac{1}{1 + x_i^2} + 4\left(\sum_{i=1}^{n} \frac{x_i}{1 + x_i^2}\right)^2 \geq M\sqrt{\frac{5n}{8}} - \frac{n}{2} \tag{99}$$

where M should be calculated so as to satisfy (27) with

$$N = \sqrt{\frac{2n}{5}} \tag{100}$$

In order to test the hypothesis H_0 we have therefore to observe the values $x_1, x_2, \cdots x_n$ and to substitute them into the left hand side of (99). If the inequality is satisfied then the hypothesis should be rejected.

Approximate values of the power function could be obtained from the right hand side of (85) with

$$\vartheta\sigma_1 = \theta\sqrt{\frac{n}{2}}. \tag{101}$$

EXAMPLE II. Let us assume as given that

$$\begin{cases} p(x_i \mid \theta) = \theta e^{-\theta x_i} & \text{for } 0 < x_i \\ \qquad\qquad = 0 & \text{elsewhere} \end{cases} \tag{102}$$

with $\theta > 0$, the hypothesis to test being that $\theta = \theta_0 = 1$, with the alternatives both $\theta < 1$ and $\theta > 1$.

In this particular example the unbiased test of type A is easily found[11] and moreover[12] it has also the property of being of type A_1. But this circumstance does not diminish the illustrative character of the example. We have

$$\varphi_i = 1 - x_i \tag{103}$$

$$\Psi_i = -1 = \text{constant} \tag{104}$$

It follows that the regions forming the test which is unbiased in the limit are determined by the inequality (59). We have further

$$\sigma_1^2 = \mathcal{E}(\varphi_1^2) = \int_0^\infty (1 - x^2)e^{-x}\,dx = 1 \tag{105}$$

and the inequality (59) reduces to

$$\left| \sum_{i=1}^{n} (1 - x_i) \right| \geq \lambda\sqrt{n} \tag{106}$$

with λ taken from the tables of the normal integral according to (60) and to the chosen value α. Approximate values of the power function can be calculated from, say

$$\beta_\infty(\vartheta) = 1 - \frac{1}{\sqrt{2\pi}} \int_{-\lambda}^{+\lambda} e^{-\frac{1}{2}(x-\vartheta)^2}\,dx \tag{107}$$

with

$$\vartheta = (\theta - 1)\sqrt{n} \tag{108}$$

The simplicity of the example considered permits to calculate the exact power function of the test and it may be interesting to obtain its limit $\dot{\beta}_\infty(\vartheta)$ in another and a more direct way. Write

$$\sum_{i=1}^{n} x_i = y \tag{109}$$

It is known that, if the probability law of each of the X's is given by (102) then the probability law of y is

$$p(y \mid \theta) = \begin{cases} \dfrac{\theta^n y^{n-1}}{(n-1)!} e^{-\theta y} & \text{for } 0 < y \\ = 0 & \text{otherwise} \end{cases} \tag{110}$$

---

[11] J. Neyman and E. S. Pearson, loc. cit. p. 18 et seq.

[12] J. Neyman: Estimation statistique traitée comme un problème de probabilité classique. Series Actualités scientifiques et industrielles. Paris, (1938). (In the press.)

It follows that the exact form of the power function corresponding to the test (106) is

$$\beta(\theta \mid \overline{w}_n) = 1 - \frac{\theta^n}{(n-1)!} \int_{n-\lambda\sqrt{n}}^{n+\lambda\sqrt{n}} y^{n-1} e^{-\theta y}\, dy \tag{111}$$

For values of n about 100 or more and for the values of θ close to unity the distribution of say

$$z = \frac{\theta \sum x_i - n}{\sqrt{n}} = \frac{\theta y - n}{\sqrt{n}} \tag{112}$$

is practically normal with mean equal to zero and S.D. equal to unity. It follows that the integral in the right hand side of (111) is practically equal to the normal integral taken within the limits which are obtained by substituting in (112) the limits of y in (111). After some easy transformation we have, with a considerable accuracy

$$\beta(\theta \mid \overline{w}_n) = 1 - \frac{1}{\sqrt{2\pi}} \int_{(\theta-1)\sqrt{n}-\lambda\theta}^{(\theta-1)\sqrt{n}+\lambda\theta} e^{-\frac{1}{2}z^2}\, dz \tag{113}$$

or, after some further transformations and taking into account (108)

$$\beta(\theta \mid \overline{w}_n) = 1 - \frac{1}{\sqrt{2\pi}} \int_{-\lambda(1+\vartheta/\sqrt{n})}^{+\lambda(1+\vartheta/\sqrt{n})} e^{-\frac{1}{2}(u-\vartheta)^2}\, du \tag{114}$$

and it is seen that, when ϑ is fixed and n indefinitely increases, then $\beta(\theta \mid \overline{w}_n)$ does tend to $\beta_\infty(\vartheta)$.

UNIVERSITY COLLEGE, LONDON.