WATSON'S U_n^2 TEST IN THE PARAMETRIC CASE

By

Toshio HASHIMOTO[1)]

§ 1. Introduction and summary. To test whether a random sample has been drawn from a population with a specified continuous distribution, the goodness of fit problem, consisting the empirical distribution function $F_n(x)$ and hypothetical cumulative districution function $F(x)$, is treated here for the case when an auxiliary parameter is to be estimated. This extends the Watson's U_n^2 test to the parametric case, following a suggestion of Darling [1][2)].

Let $X_1, X_2, \cdots, X_n$ be independent observations (random samples) coming from a population whose absolutely continuous cumulative distrbution is $G(x)$. Let Ω be a parameter space and suppose, for each ξ contained in Ω, that $F(x;\xi)$ is a cumulative distribution function. In this paper we treat the problem of testing the null hypothesis H,

$$(1.1) \qquad H: \quad G(x)=F(x;\xi) \quad \text{for some unspecified} \quad \xi\in\Omega .$$

Well-known test statistics

$$W_n^2=n\int_{-\infty}^{\infty}\left[F(x)-F_n(x)\right]^2 dF(x) \qquad \cdots \text{Cramér-Mises-Smirnov type}$$

$$K_n^2=\sqrt{n}\sup_{-\infty<x<\infty}|F(x)-F_n(x)| \qquad \cdots \text{Kolmogorov-Smirnov type}$$

are not of this nature. These statistics are types of somewhat different hypothesis,

$$(1.2) \qquad H_0: \quad G(x)=F(x)=F(x;\xi_0) \quad \text{for a specified} \quad \xi_0\in\Omega .$$

Let $F_n(x)$ be empirical cumulative distribution function of the data; that is, $F_n(x)=k/n$ if k of the X_i, with $i=1,2,\cdots,n$ are less than x, for $-\infty<x<\infty$.

Recently G. S. Watson [7], [8] has introduced the test function

$$(1.3) \qquad U_n^2=n\int_{-\infty}^{\infty}\left\{F_n(x)-F(x)-\int_{-\infty}^{\infty}\left[F_n(y)-F(y)\right]dF(y)\right\}^2 dF(x)$$

---

1) The main result of this paper was derived in 1967 while the author was a research member, The Institute of Statistical Mathematics.

2) Numbers in brackets refer to the references at the end of the paper.

for testing the null hypothesis of the type H_0. This statistic is useful for distribution on a circle since its value does not depend on the arbitrary point chosen to begin with cumulating the probability density and the sample points. It will be noted that U_n^2 has the form of a variance while W_n^2 has its form of a second moment about the origin, that is, the modification corresponds to a 'correction for the mean'.

On the other hand, in an effort to modify the W_n^2 test to treat the hypothesis H of (1. 1), A. Darling [1] considered test function

$$C_n^2 = n\int_{-\infty}^{\infty}\left[F_n(x)-F(x;\hat{\theta}_n)\right]^2 dF(x;\hat{\theta}_n)$$

where $\hat{\theta}_n$ is a suitable estimator for the unknown parameter ξ in $F(x;\xi)$ and is a function of $X_1, X_2, \cdots, X_n$.

The suggestion of this paper is that statistic

$$A_n^2 = n\int_{-\infty}^{\infty}\Big\{F_n(x)-F(x;\hat{\theta}_n) - \int_{-\infty}^{\infty}\left[F_n(y)-F(y;\hat{\theta}_n)\right]dF(y;\hat{\theta}_n)\Big\}^2 dF(x;\hat{\theta}_n) \tag{1. 4}$$

shall be considered, with the circumference of the circle replacing the real line as the region of integration. When a measure A_n^2 is adopted, the hypothesis H is to be rejected if A_n^2 is sufficiently large.

The main purpose of this paper is to give a method for finding the asymptotic distribution of A_n^2. Clearly it turns out that matters depend crucially on the estimator chosen. There are two essentially distinct cases. By reducing the problem to straightforward considerations in the theory of continuous Gaussian stochastic process, analytical task developped by Doob and Donsker [2] [3] is used for calculating the limiting distribution of A_n^2.

In section 2, we have formulated the problem of finding the limiting distribution of A_n^2. Especially, when $\hat{\theta}_n$ is an estimator such that $nE\{(\hat{\theta}_n-\theta)^2\}\to 0$, where θ is true value of ξ, that is, $G(x)=F(x;\theta)$, then the limiting distribution of A_n^2 is same as the limiting distribution of U_n^2 given by (1.3). This is known and tabulated. In section 3, more general case when θ does not admit this so-called superefficient estimator is treated. The limiting distributions of A_n^2 and U_n^2 are not the same, in this case. Making use of the representation of the Gaussian stochastic process, we determine the characteristic function of the limiting distribution of A_n^2 explicitely in section 5. But unfortunately even in simple important cases (such as, a normal distribution with an unknown mean) the resulting characteristic function appears very difficult to invert.

It may be noted that the A_n^2 test is not distribution free unlike the U_n^2 test, and asymptotic test depends only on the structure of the family $F(x;\xi)$.

§ 2. Mathematical formulation in the superefficient case. Suppose that the hypothesis H as given by (1.1) is true, and let the true unknown value of the parameter ξ be θ, with θ are interior point of nondegenerate interval of the real axis (an element of Ω). Let the density function corresponding to $F(x;\xi)$ be $f(x;\xi)$.

If $X_{(1)}, X_{(2)}, \cdots, X_{(n)}$ is a ordered statistic of the random sample, so that $X_{(1)}<X_{(2(}<\cdots<X_{(n)}$, for computational purpose, the value of U_n^2 may be calculated from

$$(2.1)\qquad U_n^2=\sum_{i=1}^{n}\left[F(X_{(i)};\hat{\theta}_n)-\frac{2i-1}{2n}-\frac{1}{n}F(X_{(i)};\hat{\theta}_n)+\frac{1}{2}\right]^2+\frac{1}{12n}.$$

Actually it may be easier to use, instead of (1.3), the equivalent form (2.1).

When $\hat{\theta}_n$ is a superefficient estimator, we may prove that the limiting distribution of A_n^2 is the same as that of Watson's U_n^2 statistic.

Theorem 1. *Assume that the estimator $\hat{\theta}_n$ and the distribution $F(x;\xi)$ have the following properties:*

1) *$nE\{(\hat{\theta}_n-\theta)^2\}\to 0$ as $n\to\infty$, where θ is the true value of ξ, that is, $G(x)=F(x;\theta)$.*

2) *For $\xi, \xi'\in\Omega$, $F(x;\xi)$ satisfies a Lipschitz condition*

$$\left|F(x;\xi)-F(x;\xi')\right|<A(x)\left|\xi-\xi'\right|$$

where $P_r\{A^2(X_1)>A_0\}=0$ for some $A_0<\infty$, the probability according to $F(x;\theta)$.

Then we have $A_n^2=U_n^2+\delta_n$, where $\delta_n\to 0$ in probability.

Proof. From (2.1)

$$\begin{aligned}
A_n^2&=\frac{1}{12n}+\sum_{j=1}^{n}\left[F(X_{(j)};\hat{\theta}_n)-\frac{2j-1}{2n}-\frac{1}{n}\sum_{j=1}^{n}F(X_{(j)};\hat{\theta}_n)+\frac{1}{2}\right]^2\\
&=\frac{1}{12n}+\sum_{j=1}^{n}\left\{\left[F(X_{(j)};\theta)-\frac{2j-1}{2n}-\frac{1}{n}\sum_{j=1}^{n}F(X_{(j)};\hat{\theta}_n)+\frac{1}{2}\right]\right.\\
&\qquad\left.+\left[F(X_{(j)};\hat{\theta}_n)-F(X_{(j)};\theta)\right]+\left[\frac{1}{n}\sum_{j=1}^{n}F(X_{(j)};\hat{\theta}_n)-\frac{1}{n}\sum_{j=1}^{n}F(X_{(j)};\hat{\theta}_n)\right]\right\}^2\\
&=U_n^2+\sum_{j=1}^{n}\left\{F(X_{(j)};\hat{\theta}_n)-F(X_{(j)};\theta)\right\}^2+\sum_{j=1}^{n}\left\{\frac{1}{n}\sum_{j=1}^{n}\left[F(X_{(j)};\theta)-F(X_{(j)};\hat{\theta}_n)\right]\right\}^2\\
&\quad+2\sum_{j=1}^{n}\left[F(X_{(j)};\hat{\theta}_n)-\frac{2j-1}{2n}-\frac{1}{n}\sum_{j=1}^{n}F(X_{(j)};\theta)+\frac{1}{2}\right]\left[F(X_{(j)};\hat{\theta}_n)-F(X_{(j)};\theta)\right]
\end{aligned}$$

$$+2\sum_{j=1}^{n}\left[F(X_{(j)};\theta)-\frac{2j-1}{2n}-\frac{1}{n}\sum_{j=1}^{n}F(X_{(j)};\theta)+\frac{1}{2}\right]$$
$$\times\left[\frac{1}{n}\sum_{j=1}^{n}\left\{F(X_{(j)};\theta)-F(X_{(j)};\hat{\theta}_n)\right\}\right]$$
$$+2\sum_{j=1}^{n}\left[F(X_{(j)};\hat{\theta}_n)-F(X_{(j)};\theta)\right]\left[\frac{1}{n}\sum_{j=1}^{n}\left\{F(X_{(j)};\theta)-F(X_{(j)};\hat{\theta}_n)\right\}\right]$$
$$=U_n^2+\delta_1+\delta_2+2\delta_3+2\delta_4+2\delta_5\,.$$

Then $\delta_3^2\leqslant\left(U_n^2-\frac{1}{12n}\right)\delta_1$, and

$$\delta_1=\sum_{j=1}^{n}\left\{F(X_{(j)};\hat{\theta}_n)-F(X_{(j)};\ \theta)\right\}^2\leqslant n(\hat{\theta}_n-\theta)^2\max_j A^2(X_j)$$

Thus $E(\delta_3^2)\leqslant E\left\{\left(U_n^2-\frac{1}{12n}\right)\right\}^{\frac{1}{2}}E(\delta_1^2)\to 0$ as $n\to\infty$.

Similarly,

$$\delta_2\leqslant n(\hat{\theta}_n-\theta)^2B\qquad\text{where}\quad B=\left[\max_j A(x_j)\right]^2.$$

and $\delta_4^2\leqslant\left(U_n^2-\frac{1}{12n}\right)\delta_2$, hence we have $E(\delta_4^2)\leqslant E\left\{\left(U_n^2-\frac{1}{12n}\right)^2\right\}^{\frac{1}{2}}E(\delta_2^2)\to 0$. Moreover, we have $\delta_5\leqslant n(\hat{\theta}_n-\theta)B$, thus the required result is obtained. A trivial consequence of this theorem is

Corollary. *Under the conditions* 1) *and* 2) *of Theorem* 1, *the limiting distribution of A_n^2 is the same as that of the Watson's statistic U_n^2.*

The limiting distribution of U_n^2 has been tabulated [7], i.e.

$$\lim_{n\to\infty}P_r(U_n^2>z)=\sum_{m=1}^{\infty}(-1)^{m-1}2\exp(-2m^2\pi^2z)^{3)}.$$

Thus the problem in this case is completely solved. Typical example in which the estimate $\hat{\theta}_n$ is unbiased follows as already be showed by Darling [1]. When

$$\hat{\theta}_n=\frac{n+1}{n}\max\{X_1,X_2,\cdots,X_n\},\quad f(x;\xi)=\begin{cases}\frac{1}{\xi}, & (0<x<\xi)\\ 0, & (\text{otherwise})\end{cases}$$

then we have $Var(\hat{\theta}_n)=\frac{\theta^2}{n(n+1)}$, and condition 2) of the Theorem 1 is satisfied.

3) This distribution has arisen before in noise theory— See Kac-Siegert [4].

§ 3. **The regular estimation case.** In the case of regular estimation of Cramér [5], we will have

$$Var(\hat{\theta}_n) \geqslant \frac{1}{n\int_{-\infty}^{\infty}\left(\frac{\partial \log f}{\partial \theta_0}\right)^2 f(x;\theta_0)\,dx}$$

by the Cramér-Rao inequality. But it will generally happen that $n^{\frac{1}{2}-\delta}(\hat{\theta}_n-\theta)$ will converge in probability to zero for $\delta>0$. To cover this situation we have the following lemma according to Darling.

Lemma 1. *Assume that $f(x;\xi)$ and $\hat{\theta}_n$ are such that*

1) $nE\{(\hat{\theta}_n-\theta)^4\}\to 0$,

2) $\left|\frac{\partial^2 F(x;\xi)}{\partial \xi^2}\right| < g_0(x)$,

3) $\left|\frac{\partial f(x;\xi)}{\partial \xi}\right| < g_1(x)$,

for almost all x, where $g_0(x)$ and $g_1(x)$ are integrable from $-\infty$ to $+\infty$. The functions g_0 and g_1 and exceptional set do not depend upon ξ.

Then,

$$(3.1)\qquad A_n^2 = n\int_{-\infty}^{\infty}\Big\{F_n(x)-F(x)-(\hat{\theta}_n-\theta)\frac{\partial}{\partial\theta}F(x;\theta)$$
$$-\int_{-\infty}^{\infty}\Big[F_n(y)-F(y)-(\hat{\theta}_n-\theta)\frac{\partial}{\partial\theta}F(y;\theta)\Big]f(y;\theta)\Big\}^2 f(x;\theta)dx+\delta_n$$

where $\delta_n\to 0$ in probability.

Expanding $F(x;\hat{\theta}_n)$ into Taylor series, we have for almost all x,

$$F(x;\hat{\theta}_n)-F(x;\theta) = (\hat{\theta}_n-\theta)\frac{\partial}{\partial\theta}F(x;\theta)+\frac{1}{2}(\hat{\theta}_n-\theta)^2 q_0(x), \qquad |q_0|<1.$$

Putting these expansions in expression (1.4), we obtain the lemma after some calculations. Thus it is necessary for us to study only the limiting form (if it exists) of the distribution of

$$R_n^2 = n\int_{-\infty}^{\infty}\Big\{F_n(x)-F(x;\theta)-(\hat{\theta}_n-\theta)\frac{\partial}{\partial\theta}F(x;\theta)$$
$$-\int_{-\infty}^{\infty}\Big[F_n(y)-F(y;\theta)-(\hat{\theta}_n-\theta)\frac{\partial}{\partial\theta}F(y;\theta)\Big]f(y;\theta)\Big\}^2 f(x;\theta)dx.$$

A transformation which is basic in the work to follow,

$$(3.2)\qquad u = F(x;\theta),$$

defines implicitely as a function of u. Thus $x=x(u,\theta)$ except possibly for an enumerable set of u values at which x can be defined arbitrarily, except to render it monotone nondecreasing. Next we define the function $g(u)$ as

$$g(u)=\frac{\partial}{\partial\theta}F(x;\theta), \qquad 0\leqslant u\leqslant 1, \tag{3.3}$$

and note that $g(u)$ depends in general on θ. Finally, we express the empirical distribution function as a function of u. If we introduce the function $\psi_t(v)$ defined as

$$\psi_t(v)=\begin{cases} 1, & v>t \\ 0, & v\geqslant t \end{cases}$$

then we have, with probability one,

$$F_n(x)=\frac{1}{n}\sum_{j=1}^{n}\psi_x(X_j)=\frac{1}{n}\sum_{j=1}^{n}\psi_u\big[F(X_j;\theta)\big]=\begin{cases} \dfrac{1}{n}, & F(X_j;\theta)<u \\ 0, & F(X_j;\theta)\geqslant u \end{cases}$$

where u is given by $u=F(x;\theta)$. On writing

$$Z_n(u)=\sqrt{n}\left\{\frac{1}{n}\sum_{j=1}^{n}\psi_u\big[F(X_j;\theta)\big]-u\right\}, \tag{3.4}$$

$$T_n=\sqrt{n}\,(\hat{\theta}_n-\theta), \tag{3.5}$$

A_n^2 can be written, from the form of R_n^2, as

$$A_n^2=\int_0^1\left\{Z_n(u)-T_ng(u)-\int_0^1\big[Z_n(v)-T_ng(v)\big]\,dv\right\}^2du+\delta_n \tag{3.6}$$

where $\delta_n\to 0$ in probability, and $g(u)$ is given by $g(u)=\dfrac{\partial}{\partial\theta}F(x;\theta)$ for $0\leqslant u\leqslant 1$. Finally by definig the stochastic process $Y_n(u)$, $X_n(u)$ as

$$\begin{aligned} Y_n(u)&=Z_n(u)-T_ng(u) \\ X_n(u)&=Y_n(u)-\int_0^1 Y_n(v)dv \end{aligned} \tag{3.7}$$

where $Z_n(u)$ and T_n are given by (3.4) and (3.5), we have

$$A_n^2=\int_0^1 X_n^2(u)du+\delta_n \tag{3.8}$$

where $\delta_n\to 0$ in probability.

It follows that the limiting form of the stochastic process $X_n(u)$ is of central importance, and we next prove the following lemma.

Lemma 2. *Assume, in addition to condition* 1), 2), 3) *of lemma* 1, *that*

4) $nE(\hat{\theta}_n-\theta)\to 0$, *as* $n\to\infty$.

5) $T_n=\sqrt{n}\,(\hat{\theta}_n-\theta)$ *is a sum of independent identically distributed random variables having a limiting Gaussian distribution with variance* σ^2 $(\sigma^2>0)$; *and*

6) *the conditional expectation* $nuE\{\hat{\theta}_n-\theta|F(X_1;\theta)<u\}$ *converges to a certain function* $h(u)$ *with* $0<u<1$ *and* $h(0)=h(1)=0$.

Then the stochastic process $X_n(u)$ *converges in distribution to a Gaussian process* $X(u)$ *with mean* 0 *and covariance function*

$$(3.9)\quad \rho(u,v)=E\big[X(u)X(v)\big]=\min(u,v)-uv-\frac{1}{2}(u+v-u^2-v^2)$$
$$+\frac{1}{12}-h(u)g(v)-h(v)g(u)+\sigma^2g(u)g(v)$$
$$+\big[h(u)+h(v)-\sigma^2g(u)-\sigma^2g(v)\big]\int_0^1 g(x)dx+\big[g(u)+g(v)\big]\int_0^1 h(x)dx$$
$$-\int_0^1\int_0^1 h(x)g(y)dxdy-\int_0^1\int_0^1 h(y)g(x)dxdy+\Big[\sigma\int_0^1 g(x)dx\Big]^2$$

The expression "$X_n(u)$ converges to $X(u)$ in distribution" means that for every finite set $u_1,u_2,\cdots,u_k$ the joint distribution function of $X_n(u_1),X_n(u_2),\cdots,X_n(u_k)$ converges to the joint distribution of $X(u_1),X(u_2),\cdots,X(u_k)$.

Proof. The proof of this lemma is quite straightforward. The stochastic process $Z_n(u)$ defined by (3.4) is known to converge in distribution to a Gaussian process with mean 0 and covariance $k(u,v)=E[Z_n(u)Z_n(v)]$ given by

$$k(u,v)=\min(u,v)-uv$$

with $k(u,v)$ being independent of n. By assumption 5), T_n has a limiting Gaussian distribution with mean 0 and variance σ^2. It follows from the multidimensional central limit theorem that the linear combination $Y_n(u)=Z_n(u)-T_ng(u)$ converges in distribution to a Gaussian process whose mean is 0, hence $X_n(u)=Y_n(u)-\int_0^1 Y_n(v)dv$ converges in distribution to a Gaussian distribution. Thus it will be sufficient to verify that the limiting covariance $\rho_n(u,v)=E[X_n(u)X_n(v)]$ converges to $\rho(u,v)$ given by (3.9).

Denoting by $h_n(u)=E[Z_n(u)T_n]$ and using Darling's calculation [1], we have, by assumption 6)

$$h_n(u)=nuE\big\{(\hat{\theta}_n-\theta)\,|\,F(X_1;\theta)<u\big\}-nuE(\hat{\theta}_n-\theta).$$

Consequently by assumption 4) and 6), $h_n(u)$ converges to $h(u)$ for $0\leqslant u\leqslant 1$. The covariance function $\rho_n(u,v)$ is

$$\rho_n(u, v) = E\left[X_n(u)X_n(v)\right] = E\left[\left\{Z_n(u) - T_n g(u) - \int_0^1 \left[Z_n(x) - T_n g(x)\right] dx\right\}\right.$$
$$\left.\times \left\{Z_n(v) - T_n g(v) - \int_0^1 \left[Z_n(y) - T_n g(y)\right] dy\right\}\right]$$
$$= \min(u, v) - uv + \frac{1}{2}(u^2 + v^2 - u - v) + \frac{1}{12} + \sigma^2 g(u) g(v) - h_n(u) g(v)$$
$$- h_n(v) g(u) + \left[h_n(u) + h_n(v) - \sigma_n^2 g(u) - \sigma_n^2 g(v)\right] \int_0^1 g(x) dx$$
$$+ \left[g(u) + g(v)\right] \int_0^1 h_n(x) dx - \int_0^1 \int_0^1 h_n(x) g(y) dx dy$$
$$- \int_0^1 \int_0^1 h_n(y) g(x) dx dy + \left[\sigma_n \int_0^1 g(x) dx\right]^2$$

where $\sigma_n^2 = \mathrm{Var}(T_n) \to \sigma^2$. This calculation is obtained by two methods. The one is using Watson's result on the statistic U_n^2[4], the other is to calculate directly. Thus we have obtained $\rho_n(u, v) \to \rho(u, v)$ for $\rho(u, v)$ as in (3.9), and the lemma is established.

It might be concluded that the limiting distribution of A_n^2 is the distribution of $A_n^2 = \int_0^1 X^2(u) du$ where $X(u)$ is a Gaussian process with mean 0 and covariance $\rho(u, v)$ as in (3.9), following Doob-Donsker's heuristic approach. But we shall prove this fact in the next section only when the estimator $\hat{\theta}_n$ is further specialized.

Finally we remark the fact that we have $|h(u)| \leqslant \sigma\sqrt{u(1-u)}$ for the function $h(u)$ as defined by assumption 6) but $\lim_{n\to\infty} h_n'(u)$ does not always exist.

§ 4. An efficient estimator case. Thus far, we have given no special attention to the choice of the estimator $\hat{\theta}_n$. It might be thought that, paralleling the principle of minimum χ^2, we should choose $\hat{\theta}_n$ so as to make A_n^2, as given by (1.4), a minimum. As is well known, this does not lead to usable result as is often the case with minimum χ^2.

However, precisely as in the maximum likelihood principle does lead to a certain ideal properties for A_n^2, at least asymptotically.

In this section we assume that Cramér's condition [5] for a regular unbiased efficient (or minimum variance) estimate are satisfied. Following Cramér, we simplify term the estimate efficient. Then it is clear that all conditions 1) through 6) of lemma 2 are satisfied, as noted below, with the possible exception of condition 2), which we shall further presume satisfied.

4) It should be remarked that the covariance $\rho(u, v)$ for the Watson's U_n^2 is

$$\rho(u, v) = \min(u, v) - uv + \frac{1}{2}(u^2 + v^2 - u - v) + \frac{1}{12}.$$

The efficient estimator is unbiased so that condition 4) is satisfied, and implies besides that the likelihood function

$$L=\prod_{j=1}^{n} f(X_j;\xi)$$

has the property that if $\xi=\hat{\theta}_n$ is a root of a equation $\frac{\partial}{\partial\xi}\log L=0$, then, defining

$$\sigma^2=\left(\int_{-\infty}^{\infty}\left[\frac{\partial}{\partial\theta}\log f(x;\theta)\right]^2 f(x;\theta)\,dx\right)^{-1} \tag{4.1}$$

we have

$$\frac{\partial}{\partial\xi}\log L=\sum_{j=1}^{n}\frac{\partial}{\partial\xi}\log F(X_j;\xi)=\frac{n}{\sigma^2}(\hat{\theta}_n-\xi).$$

By putting $\xi=\theta$, this yields

$$\frac{n}{\sigma^2}(\hat{\theta}_n-\theta)=\sum_{j=1}^{n}\frac{\partial}{\partial\theta}\log f(X_j;\theta)=\frac{\sqrt{n}}{\sigma^2}T_n. \tag{4.2}$$

It then follows that the variance of T_n is σ^2, independent of n, given by (4.1), and that conditions 5) and 1) are satisfied. For condition 6) of lemma 2, we multiply through the last equality by σ^2 and take conditional expectations of both sides under the condition that $F(X_1;\theta)=u$. Then on using the property of the function $h_n(u)$, and the fact that $E\left[\frac{\partial}{\partial\theta}\log f(X_j;\theta)\right]=0$, we have

$$h_n'(u)=\sigma^2\frac{\partial}{\partial\theta}\log f(x;\theta)$$

using transformation $u=F(x;\theta)$. Thus $h_n'(u)$ is independent of n, and we denote it by $h(u)$, given by

$$h'(u)=\sigma^2\frac{\partial}{\partial\theta}\log f(x;\theta). \tag{4.3}$$

Because of this simple formula for $h(u)$, making it proportional to $g(u)$ as we see immediately, $\rho(u,v)$ simplifies and renders the process $X(u)$ of lemma 2 manageable.

From a definition of $g(u)$ in (3.3) we have immediately $g(0)=g(1)=0$ and

$$g'(u)=\frac{\partial}{\partial\theta}f(x;\theta)\frac{dx}{du}=\frac{\partial}{\partial\theta}\log f(x;\theta)=\frac{1}{\sigma^2}\cdot h'(u).$$

since $h(0)=1$ by the remark at the end of section 3, by integrating we obtain

$h(u)=\sigma^2 g(u)$. Thus it follows that we can define a function $\varphi(u)$

$$\varphi(u)=\frac{h(u)}{\sigma}=\sigma g(u) \tag{4.4}$$

where $h'(u)$ is given by (4.3) and $g(u)$ by (3.3), σ being given by (4.1) and x by (3.2).

Putting these values for $h(u)$ and $g(u)$ in (3.9), we obtain the following lemma.

Lemma 3. *In the case of an efficient estimator, the process $X_n(u)$ given by* (3.7) *has mean* 0 *and covariance function*

$$\begin{aligned}\rho(u, v)=\min(u, v)-uv+\frac{1}{2}(u^2+v^2-u-v)+\frac{1}{12}-\varphi(u)\varphi(v)\\ -\left[\varphi(u)+\varphi(v)\right]\int_0^1\varphi(x)dx+\left[\int_0^1\varphi(x)dx\right]^2\end{aligned} \tag{4.5}$$

independent of n, where $\varphi(x)$ is given by (4.4). *Furthermore, the function $\varphi(u)$ has the properties*

$$a)\qquad |\varphi(u)|\leqslant\sqrt{u(1-u)}\,,\qquad b)\qquad \int_0^1\varphi'^2(u)du=1\,.$$

Proof. We must show b), but this property has already established by Darling [1].

Thus we have showed that, when an efficient estimator $\hat{\theta}_n$ exists, the limiting distribution of A_n^2 is the same as the limiting distribution of $\int_0^1 X_n^2(u)du$, where $X_n(u)$ has mean 0 and covariance $\rho(u, v)$ given by (4.5), and approaches to a Gaussian process $X(u)$ in distribution. We need to show that limiting distribution is same as the distribution of $A^2=\int_0^1 X^2(u)du$ where $X(u)$ is a Gaussian process with mean 0 and covariance $\rho(u, v)$. In order to establish this conjecture, we are going to suppose that $\varphi(u)$ satisfies the following condition.

Assumption. $\varphi''(u)$ exists almost everywhere for $0\leqslant u\leqslant 1$ and

$$\int_0^1|\varphi''(u)|\,u(1-u)\log\log\frac{1}{u(1-u)}du<\infty\,. \tag{4.6}$$

Under this condition, we can prove that the process $X_n(u)$ can be expressed in terms of $Z_n(u)$. Actually we have the representation

$$\begin{aligned}X_n(u)&=Z_n(u)-T_ng(u)-\int_0^1\left[Z_n(x)-T_ng(x)\right]dx\\ &=Z_n(u)+\varphi(u)\int_0^1\varphi''(t)Z_n(t)dt-\int_0^1\left[Z_n(x)+\varphi(x)\int_0^1\varphi''(s)Z_n(s)ds\right]dx\end{aligned} \tag{4.7}$$

where $\int_0^1 \varphi''(t)Z_n(t)dt = -\frac{\sigma}{\sqrt{n}}\sum_{j=1}^{n}\frac{\partial}{\partial\theta}\log f(X_j;\theta)$.

Now the process $Z(u)$, which is the limit distribution of $Z_n(u)$ given by (3.4), is Gaussian with mean 0 and covariance $k(u, v)=\min(u, v)-uv$. By Anderson-Darling [6], the random variable $\int_0^1 \varphi''(u)Z(u)du$ exists and is Gaussian with mean 0 when the auxiliary assumption (4.6) is satisfied.

Then the process

$$(4.8)\quad X(u)=Z(u)+\varphi(u)\int_0^1\varphi''(u)Z(t)dt-\int_0^1\left[Z(x)+\varphi(x)\int_0^1\varphi''(s)Z(s)ds\right]dx$$

is Gaussian process with mean 0 and covariance

$$E\left[X(u)X(v)\right]=E\left\{\left[Y(u)-\int_0^1 Y(x)dx\right]\left[Y(v)-\int_0^1 Y(y)dy\right]\right\}$$

where $Y(u)=Z(u)+\varphi(u)\int_0^1\varphi''(t)Z(t)dt$,

using Darling's result, the covariance becomes

$$\begin{aligned}E\left[X(u)X(v)\right]&=k(u, v)-\varphi(u)\varphi(v)-\int_0^1\left[k(u, y)-\varphi(u)\varphi(y)\right]dy\\&\quad-\int_0^1\left[k(v, x)-\varphi(u)\varphi(x)\right]dx+\int_0^1\int_0^1\left[k(x, y)-\varphi(x)\varphi(y)\right]dxdy\\&=\min(u, v)-uv+\frac{1}{2}(u^2+v^2-u-v)+\frac{1}{12}-\varphi(u)\varphi(v)\\&\quad-\left[\varphi(u)+\varphi(v)\right]\int_0^1\varphi(x)dx+\left[\int_0^1\varphi(x)dx\right]^2\\&=\rho(u, v),\qquad\text{where}\quad k(u, v)-\varphi(u)\varphi(v)=E\left[Y(u)Y(v)\right].\end{aligned}$$

Hence the process

$$X(u)=Z(u)+\varphi(u)\int_0^1\varphi''(t)Z(t)dt-\left[Z(x)+\varphi(x)\int_0^1\varphi''(s)Z(s)ds\right]dx$$

is a representation in terms of the process $Z(u)$. On the other hand, since $\int_0^1 X_n^2(u)du$ is a functional of $Z_n(u)$ continuous in the uniform topology by (4.7), it follows from a theorem of Donsker [3] that its limiting distribution is the same as the distribution of $\int_0^1 X^2(u)du$ for the process (4.8). Thus we have

Theorem 2. *In the case of an efficient estimator,*

$$\lim_{n\to\infty}P_r\{A_n^2<x\}=P_r\left\{\int_0^1 X^2(u)du<x\right\},$$

where $X(u)$ is a Gaussian process with mean 0 *and covariance $\rho(u, v)$ given by* (4.5).

By virtue of this theorem, we can now concentrate on the process $X(u)$ and attempt to find the distribution of $A^2 = \int_0^1 X^2(u)\,du$.

§ 5. The limiting distribution of A_n^2.

In the preceding section we have reduced the problem of finding the limiting distribution of A_n^2 under general conditions to that of finding distribution of $\int_0^1 X^2(u)\,du$, where $X(u)$ is a Gaussian process with mean 0 and covariance $\rho(u, v)$ given by

$$\rho(u, v) = \min(u, v) - uv + \frac{1}{2}(u^2 + v^2 - u - v) + \frac{1}{12} - \varphi(u)\varphi(v) + \left[\varphi(u) + \varphi(v)\right]\int_0^1 \varphi(x)\,dx - \left[\int_0^1 \varphi(x)\,dx\right]^2, \tag{5.1}$$

$$\varphi(u) = \sigma \frac{\partial}{\partial\theta} F(x;\theta), \qquad u = F(x;\theta), \tag{5.2}$$

$$\sigma^2 = \left\{\int_{-\infty}^{\infty} \left[\frac{\partial}{\partial\theta} \log f(x;\theta)\right]^2 f(x;\theta)\,dx\right\}^{-1}.$$

The function $\varphi(u)$ can be determined from

$$\varphi'(u) = \sigma \frac{\partial}{\partial\theta} \log f(x;\theta), \qquad u = F(x;\theta).$$

In determining the distribution of $\int_0^1 X^2(u)\,du$ we shall use a basic theorem due essentially to Kac-Siegert [4]. It states as the distribution of this random variable is the same as the distribution of a sum of weighted χ^2,

$$A^2 = \int_0^1 X^2(u)\,du = \sum_{n=1}^{\infty} \frac{G_n}{\lambda_n}$$

where $G_1, G_2, \cdots$ are independent normally distributed random variables with mean 0 and variance 1, and $\lambda_1, \lambda_2, \cdots$ are eigenvalues of the kernel $\rho(u, v)$. Furthermore we have for the characteristic function

$$E\left\{\exp\left[it\int_0^1 X^2(u)\,du\right]\right\} = \prod_{j=1}^{\infty}\left(1 - \frac{2it}{\lambda_j}\right)^{-\frac{1}{2}} = \left[D(2it)\right]^{-\frac{1}{2}}$$

where $D(\lambda)$ is the Fredholm determinant associated with the kernel $\rho(u, v)$ in (4.5). To state the formally, we have

Theorem 3. *Under the conditions of theorem* 2 *and assumption* (4.6) *we have*

$$\lim_{n\to\infty} E\left\{\exp(itA_n^2)\right\} = \left[D(2it)\right]^{-\frac{1}{2}}$$

where $D(\lambda)$ is the Fredholm determinant of the integral equation

$$f(x) = \lambda\int_0^1 \rho(x, y) f(y)\, dy$$

in which $\rho(x, y)$ is given by (4.5).

Thus, for $\varphi(u) = \sigma g(u) = \sigma \dfrac{\partial}{\partial\theta} F(x;\theta)$ given, it is necessary only to determine $D(\lambda)$ and invert the characteristic function $[D(2it)]^{-\frac{1}{2}}$ in order to obtain the limiting distribution of A_n^2.

It turns out to be possible to get a fairly explicit formula for $D(\lambda)$, but resulting characteristic function seems very difficult to invert. So as to get a explicit formula for $D(\lambda)$, we shall have to use an interesting theorem of Darling [1] on the Fredholm determinant associated with a symmetric, bounded positive definite kernel over the unit square $0 \leqslant x, y \leqslant 1$. The following theorem gives us an evaluation of $D(\lambda)$.

Theorem 4. *The Fredholm determinant $D(\lambda)$ of theorem* 3 *is*

$$D(\lambda) = \frac{\sin\dfrac{\sqrt{\lambda}}{2}}{\sqrt{\lambda}}\left(1 + \lambda\sum_{j=1}^{\infty}\frac{a_j^2}{1 - 3\lambda/\pi^2 j^2}\right)$$

where $a_j = \int_0^1 \left[\varphi(x) - \int_0^1 \varphi(x)\, dx\right] \cos 2\pi j x\, dx$.

Proof. Covariance function $\rho(u, v)$ can be written as

$$\begin{aligned}\rho(u, v) &= \min(u, v) - uv + \frac{1}{2}(u^2 + v^2 - u - v)\\ &\quad - \left[\varphi(u) - \int_0^1 \varphi(x)\, dx\right]\left[\varphi(v) - \int_0^1 \varphi(x)\, dx\right] + \frac{1}{12}\\ &= \rho_1(u, v) - \varphi_1(u)\varphi_1(v).\end{aligned}$$

Then $\rho_1(u, v) = \min(u, v) - uv + \dfrac{1}{2}(u^2 + v^2 - u - v) + \dfrac{1}{12}$ is a covariance function for Watson's statistic U_n^2 and it is symmetric, bounded positive definite kernel over the unit square $0 \leqslant u, v \leqslant 1$. This Fredholm determinant $D_1(x)$ has simple zeros, $0 < \lambda_1 < \lambda_2 < \cdots$.

Let the corresponding eigenfunctions be $f_1(x), f_2(x), \cdots$ and let Fourier coefficient of $f_j(x)$ be a_j, that is,

$$a_j = c(f_j) = \int_0^1 f_j(x)\varphi_1(x)\, dx, \qquad \varphi_1(x) \in L^2(0, 1).$$

Then we have, from Watson's result [7] on eigensolutions of integral equation, $f_j(x)=\cos 2\pi jx$, $\lambda_j=\dfrac{3}{\pi^2 j^2}$, thus we have

$$a_j=\left[\varphi(x)-\int_0^1\varphi(s)\,ds\right]\cos 2\pi jx\,dx\,.$$

For the Fredholm determinant $D_1(\lambda)$ associated with this kernel $\rho_1(x, y)$ we can calculate it with tedious but well known calculation and we then have

$$D_1(\lambda)=\frac{\sin\dfrac{\sqrt{\lambda}}{2}}{\sqrt{\lambda}}\,.$$

We must consider the Fredholm determinant $D(\lambda)$, but it is easily seen that this result is immediate consequence of a theorem due to Darling [1]. Hence we have proved the required result.

From above result, the distribution of A_n^2 can be obtained by the inversion Fourier transform formula

$$p(x)=\frac{1}{2\pi}\int_{-\infty}^{\infty}e^{-itx}\left[D(2it)\right]^{-\frac{1}{2}}dt$$

$$P(x)=1-\frac{1}{\pi}\sum_{k=1}^{\infty}(-1)^{k-1}\int_{\lambda_{2k-1}}^{\lambda_{2k}}\frac{e^{-xy/2}}{\sqrt{-y^2D(y)}}\,dy$$

where $P(x)=\int_0^x p(t)dt$ is the cumulative distribution function of A_n^2, and $0<\lambda_1<\lambda_2<\cdots$, are the simple zeros of $D(\lambda)$.

§ 6. Properties of the A_n^2 test. A matter of central interest is whether the A_n^2 test has certain distribution-free properties. For examples, it is important to know if it is parameter-free, that is, if the limiting distribution of A_n^2 does not depend on the parameter θ in $F(x;\theta)$. Clearly the test will be parameter-free if and only if $\varphi(u)$ given by (6.2) does not depend on θ; only in this case do we really have a usable test. This turns out to be the case when θ is a location or scale parameter.

In section 5 we have given a method of obtainig the Fredholm determinant $D(\lambda)$ for the kernel

$$\rho(u, v)=\min(u, v)-uv+\frac{1}{2}(u^2+v^2-u-v)+\frac{1}{12}$$
$$-\left[\varphi(u)-\int_0^1\varphi(x)\,dx\right]\left[\varphi(v)-\int_0^1\varphi(x)\,dx\right]\,.$$

The eigenvalues λ_k for $k=1,2,\cdots$, of this determinant are positive since

$\rho(u, v)$ is a correlation function; For the limiting semi-invariant of A_n^2 we then have

$$\kappa_j = 2^{1-2j}\kappa_j' = 2^{j-1}(j-1)!\sum_{n=1}^{\infty}\left(\frac{1}{\pi^2 j^2}\right)^n$$
$$= 2^{j-1}(j-1)!\int_0^1 \rho_j(u, u)\,du^{5)}$$

where $\rho_j(u, v)$ is the *j-th* iterate of the kernel $\rho(u, v)$.

Thus the mean μ of A^2 is, for example,

$$\mu = \kappa_1 = \int_0^1 \rho(u, u)\,du = \Big[s(1-s) - \varphi^2(s) + s^2 - s + \frac{1}{12}$$
$$+ 2\varphi(s)\int_0^1 \varphi(x)dx - \Big\{\int_0^1 \varphi(x)dx\Big\}^2\Big]\,ds$$
$$= \frac{1}{12} - \int_0^1 \varphi^2(s)\,ds + \Big[\int_0^1 \varphi(x)\,dx\Big]^2$$
$$< \frac{1}{12}.$$

This expression may be compared with the corresponding expression for the case when there is no parameter to be estimated, that is, the U_n^2 test of the hypothesis (1.2). The mean of A^2 is less than the mean for U^2, which is $\frac{1}{12}$, by the factor $\int_0^1 \varphi(s)\,ds - \Big[\int_0^1 \varphi(s)ds\Big]^2$.

From the remark due to Darling [1], we may conclude that all the semi-invariants (in particular the mean) for A^2 are less than the corresponding semi-invariants for Watson's U^2.

§ 7. Acknowledgement. The author of the present paper wishes to express his grateful thanks to Mr. G. Suzuki for his kind advice and encouragement.

References

[1] DARLING, D. A.: The Cramér-Smirnov test in the parametric case, *Ann. Math. Stat.*, Vol. 26 (1955), pp. 1–20.

[2] DOOB, J. L.: Heuristic approach to the Kolmogorov-Smirnov theorems, *Ann. Math. Stat.*, Vol. 20 (1949), pp. 393–403.

[3] DONSKER, M. L.: Justification and extension of Doob's heuristic approach to the Kolmogorov-Smirnov theorems, *Ann. Math. Stat.*, Vol. 23 (1952), pp. 277–281.

[4] KAC. M. and SIEGERT, J. F.: An explicit representation of a stationary Gaussian

5) See E. S. Pearson and Stephens [9].

process, *Ann. Math. Stat.*, Vol. 18 (1947), pp. 438–442.

[5] CRAMÉR, H.: *Mathematical Methods of Statistics*, Princeton University Press, (1946).

[6] ANDERSON, T. W. and DARLING, D. A.: Asymptotic theory of certain goodness of fit criteria based on stochastic processes, *Ann. Math. Stat.*, Vol. 23 (1952), pp. 193–212.

[7] WATSON, G. S.: Goodness-of-fit tests on a circle, *Biometrika*, Vol. 48 (1961), pp. 109–114.

[8] WATSON, G. S.: Goodness-of-fit tests on a circle II, *Biometrika*, Vol. 49 (1962), pp. 57–63.

[9] PEARSON, E. S. and STEPHENS, M. A.: The goodness-of-fit tests based on W_N^2 and U_N^2, *Biometrika*, Vol. 49 (1962), pp. 397–402.

[10] STEPHENS, M. A.: The distribution of the goodness-of-fit statistic U_n^2 (I), *Biometrika*, Vol. 50 (1963), pp. 303–313.

Department of Mathematics,
Hokkaido University

(Received Dec. 25, 1968)