The Annals of Statistics
1983, Vol. 11, No. 2, 588–599

GALTON'S TEST AS A LINEAR RANK TEST WITH ESTIMATED SCORES AND ITS LOCAL ASYMPTOTIC EFFICIENCY

BY KONRAD BEHNEN AND GEORG NEUHAUS

University of Hamburg

In the general two-sample testing problem $F = G$ versus $F \leq G$, $F \neq G$ the shift-model score function $f'\circ F^{-1}/f\circ F^{-1}$ has to be replaced by the nonparametric score function $b = \bar{f} - \bar{g}$, where $\bar{f} = d(F\circ H^{-1})/dx$, $\bar{g} = d(G\circ H^{-1})/dx$, $H = (mF + nG)/(m + n)$, and adaption of linear rank tests should be based on rank estimators of b. We consider an easy but rough and inconsistent rank estimator of b. The resulting rank test turns out to be a generalization of Galton's test. A formula for local asymptotic power under arbitrary local alternatives is derived which allows for comparison of Galton's test with every linear rank test. For various types of alternatives the Galton test is compared with the optimal linear rank test and with the Wilcoxon test. In order to get an impression of the validity of extrapolation to finite sample sizes, we included a Monte Carlo study under the same types of fixed alternatives.

1. Introduction. Let $X_1, \cdots, X_m, Y_1, \cdots, Y_n$ be independent real valued random variables and suppose that the distribution of $X_i[Y_j]$ is given by a continuous (cumulative) distribution function $F[G]$. Let $N = m + n$ be the size of the combined sample. If a shift model

$$G(x) = F(x + \vartheta), \qquad x \in \mathbb{R}, \quad \vartheta \in \mathbb{R},$$

is assumed, it is well known that (under additional assumptions) it is possible to construct tests which are asymptotically optimum for testing $\{\vartheta \leq 0\}$ versus $\{\vartheta > 0\}$ with respect to large classes of contiguous distributions. This is done by using linear rank statistics with scores which are based on an estimator of the (unknown) score function

$$-f'\circ F^{-1}/f\circ F^{-1},$$

e.g. Hájek and Šidák (1967), VII.1.6, and many successive papers.

Behnen (1975) demonstrates a breakdown of power of such procedures in the more general model of stochastically-larger-alternatives, i.e.

$$F \geq G \quad \text{versus} \quad F \leq G, \quad F \neq G.$$

The reason for this breakdown is the asymptotic power behavior of a linear rank test with score function ψ(ψ-test) for *local nonparametric alternatives of type b* (*b-type alternative*), i.e.

$$\frac{dF}{dH} = 1 + \rho\frac{n}{N}\sqrt{\frac{N}{mn}}\,b_N\circ H, \qquad \frac{dG}{dH} = 1 - \rho\frac{m}{N}\sqrt{\frac{N}{mn}}\,b_N\circ H, \tag{1.1}$$

with $\rho > 0$, $\| b_N - b \| \to 0$ as $N \to \infty$, and H continuous d.f., the asymptotic power in this situation being

$$\beta(\psi, \alpha, b, \rho) = 1 - \Phi(u_\alpha - \rho\,\langle b, \psi\rangle / \|\psi\|), \tag{1.2}$$

where Φ, φ, and u_α denote the distribution function, the density, and the upper α-quantile of the standard normal distribution $\mathcal{N}(0, 1)$, respectively, and $\langle\cdot, \cdot\rangle$ denotes the usual

Received December 1981; revised September 1982.

AMS 1970 *subject classifications.* Primary 62G10; secondary 62G20, 62E25.

Key words and phrases. Galton test, adaptive rank tests, two-sample problem, estimated scores, power simulation and extrapolation, stochastically-larger-alternatives.

scalar product in L_2 space of Lebesgue (λ)-square integrable functions on [0, 1] with corresponding norm $\|\cdot\|$. In case of $\psi = b$ we have asymptotic optimality according to (maximum) asymptotic power

$$\beta^*(\alpha, b, \rho) = 1 - \Phi(u_\alpha - \rho \| b \|), \tag{1.3}$$

cf. Behnen (1972).

Also in this more general (local) nonparametric model (1.1), the problem is the unknown optimal score function, i.e. the underlying b. One might try to estimate this b from the data, but (without additional information) it is not possible to have a consistent estimator of the underlying b under the local model (1.1). This is true because of contiguity: Consider the asymptotic testing problem $b = 0$ (i.e. $F = G = H$) versus $b = b_1$, $\| b_1 \| > 0$ (i.e. F_N and G_N according to (1.1) with $b = b_1$). Now assume $\hat{b}_N$ to be an estimator such that $\| \hat{b}_N \| = \| \hat{b}_N - b \| \to 0$ in probability under $b = 0$. Then contiguity implies $\| \hat{b}_N \| \to 0$ in probability under $b = b_1$, therefore excluding $\| \hat{b}_N - b \| \to 0$ in probability under $b = b_1$. So again the questions are: What has to be estimated in reality? What is the right asymptotic in order to give valuable extrapolations? In order to give an answer we shall consider the case of a general fixed alternative.

2. The general model. Let us consider the *null hypothesis* $H_0{:}F = G$ versus the general "stochastically larger" *alternative* $H_1{:}F \le G,\ F \ne G$. Now consider the simple alternative $(F, G) \in H_1$ and put

$$H = H_{F,G} = (mF + nG)/N. \tag{2.1}$$

Then, obviously, the measure H dominates the measures F and G. Moreover, without further assumptions it can be shown (cf. Behnen, 1981) that $F\circ H^{-1}$ and $G\circ H^{-1}$ are distribution functions on [0, 1] with existing Lebesgue-densities (λ-densities) according to

$$\bar{f} = d(F\circ H^{-1})/d\lambda, \qquad \bar{g} = d(G\circ H^{-1})/d\lambda, \qquad (m\bar{f} + n\bar{g})/N = 1, \tag{2.2}$$

and

$$\frac{dF}{dH} = \bar{f}\circ H = 1 + \frac{n}{N}(\bar{f} - \bar{g})\circ H, \qquad \frac{dG}{dH} = \bar{g}\circ H = 1 - \frac{m}{N}(\bar{f} - \bar{g})\circ H. \tag{2.3}$$

See also Pyke and Shorack (1968) and Hájek (1974) for such representation.

The following useful properties, which reveal the function $\bar{f} - \bar{g}$ as a standardized description of the problem H_0 vs. H_1, are easy consequences of the above representation:

$$\frac{-N}{n} \le \bar{f} - \bar{g} \le \frac{N}{m}, \qquad \int_0^1 (\bar{f} - \bar{g})\, d\lambda = 0, \tag{2.4}$$

$$F = G \Leftrightarrow \bar{f} - \bar{g} = 0, \tag{2.5}$$

$$F \underset{\ne}{\le} G \Leftrightarrow \int_0^t (\bar{f} - \bar{g})\, d\lambda \begin{array}{l} \le 0\ \forall t \in (0, 1) \\ \ne 0\ \exists t \in (0, 1), \end{array}$$

$$F \underset{\ne}{\le} G \Leftrightarrow (F - G)\circ H^{-1} \underset{\ne}{\le} 0. \tag{2.6}$$

The optimal test for the simple problem (H, H) versus (F, G) is based on the log-likelihood statistic

$$\begin{aligned} L &= \textstyle\sum_{i=1}^m \log \dfrac{dF}{dH}(X_i) + \sum_{j=1}^n \log \dfrac{dG}{dH}(Y_j) \\ &= \textstyle\sum_{i=1}^m \log\left\{1 + \dfrac{n}{N}(\bar{f} - \bar{g})(H(X_i))\right\} + \sum_{j=1}^n \log\left\{1 - \dfrac{m}{N}(\bar{f} - \bar{g})(H(Y_j))\right\}. \end{aligned}$$

Since, in the general case, H and $\bar{f} - \bar{g}$ are unknown there has to be some assumption or estimation. Obviously, because of (2.1) the natural estimator of H *under* (H, H) *as well as under* (F, G) is

$$\hat{H}_N = (m\hat{F}_m + n\hat{G}_n)/N,$$

where $\hat{F}_m$ and $\hat{G}_n$ are the empirical distribution functions of $X_1, \cdots, X_m$ and $Y_1, \cdots, Y_n$, respectively. Now notice

$$N\hat{H}_N(X_i) = R_{1i} = \text{rank of } X_i \text{ in pooled sample,}$$

$$N\hat{H}_N(Y_j) = R_{2j} = \text{rank of } Y_j \text{ in pooled sample.}$$

Thus, for the problem (H, H) vs. (F, G) the natural approximation of the optimal statistic L is the "*exact rank statistic*"

$$(2.7) \qquad T_b = \textstyle\sum_{i=1}^m \log\left\{1 + \frac{n}{N}(\bar{f} - \bar{g})\left(\frac{R_{1i}}{N+1}\right)\right\} + \sum_{j=1}^n \log\left\{1 - \frac{m}{N}(\bar{f} - \bar{g})\left(\frac{R_{2j}}{N+1}\right)\right\}.$$

(Here we use $i/(N+1)$ instead of i/N in order to have the usual symmetry.) This statistic depends on the alternative (F, G) only by the "*score function*" $b = \bar{f} - \bar{g}$.

Before discussing estimates of b we establish the relationship to usual simple linear rank statistics: Local alternatives (H_0-contiguity) imply (cf. Behnen and Neuhaus, 1975)

$$\limsup_{(m,n)\to\infty}\left\{m\left(\int \left|\frac{dF}{dH} - \frac{dH}{dH}\right| dH\right)^2 + n\left(\int \left|\frac{dG}{dH} - \frac{dH}{dH}\right| dH\right)^2\right\} < \infty.$$

Because of (2.3) this means

$$\limsup_{(m,n)\to\infty}\left\{\sqrt{\frac{mn}{N}} \int_0^1 |\bar{f} - \bar{g}|\, d\lambda\right\}^2 < \infty.$$

Thus, in local situations, the statistic (2.7) may be approximated by

$$(2.8) \qquad S_b = \frac{n}{N}\textstyle\sum_{i=1}^m (\bar{f} - \bar{g})\left(\frac{R_{1i}}{N+1}\right) - \frac{m}{N}\sum_{j=1}^n (\bar{f} - \bar{g})\left(\frac{R_{2j}}{N+1}\right),$$

which is a simple linear rank statistic with score function $b = \bar{f} - \bar{g}$.

3. Estimation of scores and Galton's test. In case of the exact rank statistic (2.7) as well as in the case of the approximate rank statistic (2.8) we have to estimate the function $b = \bar{f} - \bar{g}$. Because of $b = \bar{f} - \bar{g} = dB/d\lambda$, where $B = F \circ H^{-1} - G \circ H^{-1}$ is invariant under strictly isotone transformations

$$T(X_1), \cdots, T(X_m), \qquad T(Y_1), \cdots, T(Y_n),$$

estimators of B and b should be based on the ranks only. The natural rank estimator of B is

$$\hat{F}_m \circ \hat{H}_N^{-1} - \hat{G}_n \circ \hat{H}_N^{-1}.$$

Up to a maximal difference of $m^{-1} + n^{-1}$ this is just the usual *empirical two-sample rank-process* $\hat{D}$, according to

$$(3.1) \qquad \hat{D}(t) = \frac{1}{m}\textstyle\sum_{i=1}^m 1_{(0,t]}\left(\frac{R_{1i}}{N}\right) - \frac{1}{n}\sum_{j=1}^n 1_{(0,t]}\left(\frac{R_{2j}}{N}\right).$$

Therefore, good estimators of b should be the derivatives of good smooth approximations of $\hat{D}$. On the basis of suitable modifications of kernel estimators which are applied to the rank data

$$R_{11}/N, \cdots, R_{1m}/N \quad \text{and} \quad R_{21}/N, \cdots, R_{2n}/N,$$

respectively, this concept is studied in Behnen, Neuhaus, and Ruymgaart (1982). In the present paper we study a rough and easy estimator of the scores, which surprisingly leads to Galton's test.

Because of $F \leq G$ we have $B = F \circ H^{-1} - G \circ H^{-1} \leq 0$. Therefore we should adjust the estimator $\hat{D}$ of B to this condition, i.e., we take the estimator $\hat{B}$ of B according to

$$\hat{B}(t) = \begin{cases} \hat{D}(t), & \text{if } \hat{D}(t) \leq 0, \\ 0 & \text{if } \hat{D}(t) > 0, \quad 0 \leq t \leq 1. \end{cases} \tag{3.2}$$

And the scores

$$b\left(\frac{i}{N+1}\right) = \frac{dB}{d\lambda}\left(\frac{1}{N+1}\right), \qquad i = 1, \cdots, N,$$

should be estimated by a "derivative" of B at those points. A very rough "derivative" at the point $i/(N+1)$ is

$$\hat{b}\left(\frac{i}{N+1}\right) = \frac{\hat{B}\left(\frac{i}{N}\right) - \hat{B}\left(\frac{i-1}{N}\right)}{1/N}.$$

Obviously, this "derivative" has no consistency properties for estimating the underlying b, but at least it should estimate some tendency of the underlying b.

Now, if we insert these "estimated" scores into the following representation of S_b (which is just an easy reformulation of (2.8) with respect to the definition (3.1) of $\hat{D}$),

$$S_b = \frac{mn}{N} \sum_{i=1}^{N} \left\{\hat{D}\left(\frac{i}{N}\right) - \hat{D}\left(\frac{i-1}{N}\right)\right\} b\left(\frac{i}{N+1}\right),$$

we get the statistic

$$S^* = NS_{\hat{b}} = \frac{mn}{N} \sum_{i=1}^{N} \left\{\hat{D}\left(\frac{i}{N}\right) - \hat{D}\left(\frac{i-1}{N}\right)\right\}\left\{\hat{B}\left(\frac{i}{N}\right) - \hat{B}\left(\frac{i-1}{N}\right)\right\}. \tag{3.3}$$

(Here a factor N is included in order to have a good standardization for evaluating the null distribution.) The corresponding rank test for $H_0: F = G$ versus $H_1: F \leq G$, $F \neq G$ is the upper level α test based on S^*. This test has some interesting features:

$$S^* = \text{Lebesgue measure of } \{t \in [0, 1]: \bar{D}(t) < 0\}, \tag{3.4}$$

where $\bar{D}$ is the version of $\hat{D}$ which is smoothed by linear interpolation between the points $(i/N, \hat{D}(i/N))$, $i = 0, 1, \cdots, N$. (In the case $m = n$ this is an easy consequence of the representation

$$\begin{aligned} S^* &= \frac{1}{N} \# \left\{i \in \{1, \cdots, N\} : \hat{D}\left(\frac{i-1}{N}\right) \leq 0, \hat{D}\left(\frac{i}{N}\right) \leq 0\right\} \\ &= \frac{1}{m} \# \left\{k \in \{1, \cdots, m\} : \hat{D}\left(\frac{2k-1}{N}\right) < 0\right\}. \end{aligned} \tag{3.5}$$

In case of $m \neq n$ the proof is done by summing up the fractions where the $\bar{D}$-process is below zero.) Under the null hypothesis $H_0: F = G$ we have

$$\mathscr{L}_{H_0}(S^*) \to \mathscr{U}(0, 1) \quad as \quad m, n \to \infty \tag{3.6}$$

uniform distribution on [0, 1]. This is true since the $C[0, 1]$-process $\sqrt{mn/N}(\bar{D})$ converges in distributrion to the Brownian bridge W_0, cf. Hájek and Šidák (1967, V.3.5), and since the statistic

$$\Lambda(W_0) = \lambda\{t \in [0, 1]: W_0(t) < 0\} \tag{3.7}$$

is distributed according to $\mathscr{U}(0, 1)$, cf. Billingsley (1968, pages 85–86).

In the case $m = n$ we have

$$mS^* = \# \{k \in \{1, \cdots, m\} : R_2^{(k)} < R_1^{(k)}\}, \tag{3.8}$$

where $R_1^{(1)} < \cdots < R_1^{(m)}$, $R_2^{(1)} < \cdots < R_2^{(m)}$ are the ordered ranks of the X- and Y-sample in the pooled sample, respectively. (This is an easy consequence of the representation (3.5).) This statistic is known as *Galton rank statistic*, which has exact null distribution

$$P_{H_0}(mS^* = i) = \frac{1}{m+1}, \qquad i = 0, 1, \cdots, m, \tag{3.9}$$

cf. Feller (1968), page 94. We shall use this name in the more general (m, n) case, also.

Under local nonparametric alternatives (F, G) of type b as defined in formula (1.1), we get by standard contiguity arguments

$$\mathscr{L}_{(F,G)}(S^*) \to \mathscr{L}(\lambda\{t \in [0, 1]: W_0(t) + \rho B(t) < 0\}), \qquad \text{as } m, n \to \infty. \tag{3.10}$$

where

$$B(t) = \int_0^t b \, d\lambda \le 0 \; \forall t \in [0, 1], \qquad B(0) = B(1) = 0,$$

because of $F \le G$.

An immediate consequence of (3.6) and (3.10) is the *asymptotic unbiasedness* of the Galton rank test for $H_0^*: F \ge G$ vs. $H_1: F \le G$, $F \ne G$, and also the *consistency* of the Galton rank test for a fixed alternative (F, G), $F \le G$, $F \ne G$, iff

$$\lambda\{t \in [0, 1]: B(t) = 0\} \le \alpha.$$

On the other hand the Galton test is *highly specific* for the testing problem $H_0: F = G$ vs. $H_1: F \le G$, $F \ne G$. This is true because of the following properties: If $(F, G) \notin H_1$, then $\varepsilon = \lambda\{t \in [0, 1]: B(t) > 0\} > 0$. Thus, because $\sup_{0 \le t \le 1} |\bar{D}_N(t) - B(t)| \to 0$ in (F, G)-probability as $N \to \infty$, we have in the case $\varepsilon > \alpha$

$$P_{F,G}\{S_N^* > 1 - \alpha\} = P_{F,G}\{\lambda\{t \in [0, 1]: D_N(t) < 0\} > 1 - \alpha\} \to 0 \quad \text{as} \quad n \to \infty.$$

This means that the power of the Galton test asymptotically stays below the level α for any fixed (F, G) which deviates from H_1 more than α, if the measure of deviation is

$$\lambda\{t \in [0, 1]: B(t) > 0\}, \qquad B = F \circ H^{-1} - G \circ H^{-1}.$$

4. Local asymptotic efficiency. First we show the concept of Bahadur efficiency to be inadequate in case of Galton's test. For simplicity reasons we restrict the discussion to the case $m = n$. Then we have (under the null hypothesis)

$$\alpha_N\left(\frac{i}{m}\right) = P_{H_0}\left(S^* \ge \frac{i}{m}\right) = \frac{m + 1 - i}{m + 1}, \qquad i = 0, 1, \cdots, m.$$

Therefore the level actually attained $\alpha_N(S^*)$ has the property

$$1 \ge \alpha_N(S^*) = \frac{m + 1 - mS^*}{m + 1} \ge \frac{1}{m+1}.$$

This implies (for any underlying distribution)

$$0 \ge \frac{1}{N} \log \alpha_N(S^*) \ge \frac{1}{N} \log \frac{1}{m+1} \to 0 \quad \text{as} \quad m \to \infty.$$

From this result we cannot conclude that the Galton test has bad power behavior, but it is obvious that the concept of Bahadur efficiency is inadequate for this type of test, the reason being the concentration of too much mass on the maximal value 1 of S^*.

Now let us consider the concept of *local asymptotic efficiency* as defined in Section VII.2.3 of Hájek and Šidák (1967). Because of (3.6) and (3.10) the asymptotic power of the Galton test under local alternatives (F, G) of type b as defined in formula (1.1) is

$$\beta_G(\alpha, b, \rho) = \lim_{(m,n)\to\infty} P_{(F,G)}(S^* > 1 - \alpha) \tag{4.1}$$

$$= P(\lambda\{t \in [0, 1]: W_0(t) + \rho B(t) < 0\} > 1 - \alpha).$$

Similar to Hájek and Šidák (1967), VI.4.5, we get

$$\beta_G(\alpha, b, \rho) = \alpha + \rho \langle b, \dot{F}_\alpha \rangle + o(\rho), \tag{4.2}$$

with $F_\alpha(t) = \int_A W_0(t)\, dP$ and $\dot{F}_\alpha(t) = (\partial/\partial t)F_\alpha(t)$ $0 < t < 1$, where P is the distribution of Brownian bridge W_0 on $C[0, 1]$ and A is defined by

$$A = \{f \in C[0, 1]: \lambda\{t \in [0, 1]: f(t) < 0\} > 1 - \alpha\}.$$

In order to evaluate the slope $\langle b, \dot{F}_\alpha \rangle$ of the asymptotic power function $\beta_G(\alpha, b, \rho)$ at $\rho = 0$, we have to evaluate $F_\alpha(t)$, $0 < t < 1$. In order to do this, we define functionals Λ and Λ_t, $t \in (0, 1)$, on $C[0, 1]$ according to

$$\Lambda(z) = \lambda\{s \in [0, 1]: z(s) < 0\}, \qquad z \in C[0, 1],$$

$$\Lambda_t(z) = \lambda\{s \in [0, t]: z(s) < 0\}, \qquad z \in C[0, 1].$$

With this notation we get, for each $t \in (0, 1)$,

$$F_\alpha(t) = \int_A W_0(t)\, dP = \int yP[\Lambda(W_0) > 1 - \alpha \mid W_0(t) = y]P^{W_0(t)}\,(dy)$$

and

$$\begin{aligned}\mathscr{L}(\Lambda(W_0) \mid W_0(t) = y) &= \mathscr{L}(\Lambda_t(W_0) \mid W_0(t) = y) * \mathscr{L}(\Lambda(W_0) - \Lambda_t(W_0) \mid W_0(t) = y)\\ &= \mathscr{L}(\Lambda_t(W_0) \mid W_0(t) = y) * \mathscr{L}(\Lambda_{1-t}(W_0) \mid W_0(1 - t) = y)\\ &= \mathscr{L}(\Lambda_t(W) \mid W(t) = y) * \mathscr{L}(\Lambda_{1-t}(W) \mid W(1 - t) = y)\\ &= \mathscr{L}(t\Lambda(W) \mid W(1) = y/\sqrt{t}) * \mathscr{L}(1 - t)\Lambda(W) \mid W(1) = y/\sqrt{1 - t}),\end{aligned}$$

where W denotes the Brownian motion in $C[0, 1]$. From Billingsley (1968), formula (11.24), we get a Lebesgue density of $\mathscr{L}(t\Lambda(W) \mid W(1) = y/\sqrt{t})$ in the following form $(0 \le u \le t)$:

$$g_{t,y}(u) = \frac{1}{t\varphi(y/\sqrt{t})} \begin{cases} \displaystyle\int_{u/t}^{1} g(s, y/\sqrt{t})\, ds, & \text{if } y \ge 0\\[2ex] \displaystyle\int_{1-u/t}^{1} g(s, y/\sqrt{t})\, ds, & \text{if } y < 0,\end{cases}$$

where

$$g(s, x) = \frac{1}{2\pi} \frac{|x|}{\{s(1 - s)\}^{3/2}} \exp\left(-\frac{1}{2} \frac{x^2}{1 - s}\right), \qquad 0 < s < 1, \quad x \in \mathbb{R}.$$

Therefore we get a Lebesgue density of

$$\mathscr{L}(\Lambda(W_0) \mid W_0(t) = y)$$

TABLE 1
Evaluation of $F_\alpha(t)$ according to (4.3).

t or $1-t$	$F_{0.10}(t)$	$F_{0.05}(t)$	$F_{0.15}(t)$
0	0	0	0
0.05	−0.01789	−0.01135	−0.02245
0.10	−0.03052	−0.01780	−0.03963
0.15	−0.03925	−0.02240	−0.05305
0.20	−0.04613	−0.02584	−0.06272
0.25	−0.05151	−0.02848	−0.07041
0.30	−0.05557	−0.03050	−0.07654
0.35	−0.05859	−0.03200	−0.08113
0.40	−0.06066	−0.03303	−0.08428
0.45	−0.06188	−0.03364	−0.08614
0.50	−0.06229	−0.03384	−0.08675

according to

$$h_{t,y}(v) = \int_0^v g_{t,y}(u) g_{1-t,y}(v-u)\, du, \qquad 0 < v < 1.$$

Since $\mathscr{L}(W_0(t))$ has the density

$$\varphi(y/\sqrt{t(1-t)})/\sqrt{t(1-t)}, \qquad y \in \mathbb{R},$$

we have

$$\begin{aligned} F_\alpha(t) &= \int_{-\infty}^{+\infty} y\left(\int_{1-\alpha}^1 h_{t,y}(v)\, dv\right) \frac{\varphi(y/\sqrt{t(1-t)})}{\sqrt{t(1-t)}}\, dy \\ &= \frac{(2\pi)^{-3/2}}{\{t(1-t)\}^2} \int_{-\infty}^0 dy \int_{1-\alpha}^1 dv \int_0^v du \int_{(t-u)/t}^1 dr \int_{(1-t-v+u)/(1-t)}^1 ds h_t(y, r, s) \\ &\quad + \frac{(2\pi)^{-3/2}}{\{t(1-t)\}^2} \int_0^\infty dy \int_{1-\alpha}^1 dv \int_0^v du \int_{u/t}^1 dr \int_{(v-u)/(1-t)}^1 ds h_t(y, r, s), \end{aligned}$$

where

$$h_t(y, r, s) = \frac{y^3}{\{r(1-r)s(1-s)\}^{3/2}} \exp\left\{-\frac{y^2}{2} \frac{t(1-r) + (1-t)(1-s)}{t(1-t)(1-r)(1-s)}\right\}.$$

By a straightforward but tedious calculation of the last integrals we finally get

$$F_\alpha(t) = (2\pi)^{-3/2} \int_0^t dx' \int_0^{1-t} dy B_t(x, y) A_\alpha(x, y), \tag{4.3}$$

where

$$\begin{aligned} B_t(x, y) &= [(t-x)(1-t-y)]^{1/2}(xy)^{-3/2}(1-x-y)^{-2}, \\ A_\alpha(x, y) &= [(\alpha-x)^+]^2 + [(\alpha-y)^+]^2 - [(\alpha-x-y)^+]^2 - \alpha^2 \\ &\quad - [(x-1+\alpha)^+]^2 - [(y-1+\alpha)^+]^2 + [(x+y-1+\alpha)^+]^2. \end{aligned}$$

A further evaluation of $F_\alpha(t)$ in (4.3) seems to be very difficult, but a numerical integration is possible (see Table 1). Therefore, the slopes $\langle b, \dot{F}_\alpha \rangle$ of the asymptotic power function (4.2) of the Galton test may be computed numerically for practically all b of interest. This has been done for a set of alternatives A.1 to A.7 (below). The results are contained in the first column of Table 2.

TABLE 2
Slopes of Galton's test, LARE *of Galton's test relative to Wilcoxon's test, and local asymptotic efficiencies of Galton's test and Wilcoxon's test with level* $\alpha = 0.10$ *for seven b-type alternatives.*

b-type alternative	$\langle b, \dot{F}_\alpha \rangle$	LARE $(Ga : \psi_W \mid \alpha, b)$	$e(Ga \mid \alpha, b)$	$e(W \mid b)$
A.1	0.1187	0.81	0.81	1.00
A.2	0.0997	0.67	0.50	0.75
A.3	0.0495	0.54	0.13	0.23
A.4	0.0161	0.30	0.01	0.05
A.5	0.0407	1.15	0.11	0.10
A.6	0.0450	1.99	0.12	0.06
A.7	0.0333	0.76	0.06	0.08

In case of local alternatives (1.1) of type b according to

$$b(t) = \psi_W(t) = \sqrt{3}(2t - 1), \qquad 0 \le t \le 1,$$

(*Wilcoxon-type alternative*) an explicit evaluation of $\langle \psi_W, \dot{F}_\alpha \rangle$ is possible. By partial integration and Fubini we get

$$\begin{aligned}\langle \psi_W, \dot{F}_\alpha \rangle &= \int_0^1 \sqrt{3}(2t-1)\frac{\partial}{\partial t}F_\alpha(t)\,dt = \sqrt{3}(2t-1)F_\alpha(t)\Big|_0^1 - 2\sqrt{3}\int_0^1 F_\alpha(t)\,dt \\ &= -\frac{2\sqrt{3}}{(2\pi)^{3/2}}\int_0^1 dt\int_0^t dx\int_0^{1-t} dy B_t(x,y)A_\alpha(a,y) \\ &= -\frac{\sqrt{3}}{8\sqrt{2\pi}}\int_0^1 dx\int_0^1 dy 1_{[0,1]}(x+y)(xy)^{-3/2}A_\alpha(x,y).\end{aligned}$$

By an elementary but very tedious evaluation of integrals this implies the following explicit formula, if $0 < \alpha < 1/4$:

$$\langle \psi_W, \dot{F}_\alpha \rangle = \frac{\sqrt{3}}{\sqrt{2\pi}}\left[\pi\alpha + (1-2\alpha)\left\{\frac{\pi}{2} - \arcsin(1-2\alpha)\right\} - 2\sqrt{\alpha(1-\alpha)}\right]. \tag{4.4}$$

In order to compare with linear rank tests with score function ψ, write (1.2) as

$$\beta(\psi, \alpha, b, \rho) = 1 - \Phi(u_\alpha - \rho\langle b, \psi\rangle/\|\psi\|) = \alpha + \rho\langle b, \psi\rangle\varphi(u_\alpha)/\|\psi\| + o(\rho) \tag{4.5}$$

and define *local asymptotic relative efficiency* (LARE) of the Galton test relative to the ψ-test under b-type alternatives (1.1) according to (cf. Hájek and Šidák, 1967, VIII.2.3)

$$\text{LARE}(Ga{:}\psi \mid \alpha, b) \equiv e,$$

where $e > 0$ and (em, en) are sample sizes of the ψ-test which has the same asymptotic slope as the Galton test with sample sizes (m, n). This means that e is a solution of

$$e^{1/2}\langle b, \psi\rangle\varphi(u_\alpha)/\|\psi\| = \langle b, \dot{F}_\alpha\rangle,$$

i.e.,

$$\text{LARE}(\text{Ga}{:}\psi \mid \alpha, b) = \{\langle b, \dot{F}_\alpha\rangle\,\|\psi\|/(\langle b, \psi\rangle\varphi(u_\alpha))\}^2. \tag{4.6}$$

From (4.6) the LARE of the Galton test relative to nearly arbitrary ψ-tests under nearly arbitrary b-type alternatives can be computed. In column 2 of Table 2 this is done for the ψ_W-Test (Wilcoxon) under seven b-type alternatives A.1 to A.7 (defined below) by partial integration and using the numerical values of Table 1.

For given b-type alternative (1.1), the ψ-test with $\psi = b$ is optimal. Therefore, an absolute measure of performance of the Galton test under b-type alternative is

$$e(Ga \mid \alpha, b) = \mathrm{LARE}(Ga{:}b \mid \alpha, b),$$

which is called *local asymptotic efficiency.* Obviously, we have from (4.6)

$$e(Ga \mid \alpha, b) = \langle b, \dot{F}_\alpha \rangle^2 / (\varphi(u_\alpha) \, \| b \|)^2. \tag{4.7}$$

Again by partial integration and using the values of Table 1, $e(Ga \mid \alpha, b)$ is evaluated for the b-type alternatives A.1 to A.7 in Table 2. For comparison reasons we included in Table 2 the *local asymptotic efficiency of the Wilcoxon test* under b-type alternatives (cf. (4.5))

$$e(W \mid b) = \langle b, \psi_W \rangle^2 / \| b \|^2,$$

which is independent of the level α.

Table 2 reveals that even in cases designed in favour of Galton and against Wilcoxon (A.3) the LARE(Galton:Wilcoxon) is only 0.54, whereas in cases where (under fixed alternatives of type b) the Wilcoxon test is consistent but the Galton test is not consistent, the LARE(Galton:Wilcoxon) is larger than 1 (cf. A.5). The values of the local asymptotic efficiency are rather low in case of A.3 to A.5 for both tests.

For alternatives of Wilcoxon type ($b = \psi_W$), we get from (4.4) and (4.7) the explicit formula

$$e(Ga \mid \alpha, \psi_W) = 3\left\{\pi\alpha + (1 - 2\alpha)\left(\frac{\pi}{2} - \arcsin(1 - 2\alpha)\right) - 2\sqrt{\alpha(1 - \alpha)}\right\}^2 \exp(u_\alpha^2). \tag{4.8}$$

Some values are as follows:

α	0.20,	0.10,	0.05,	0.01
$e(Ga \mid \alpha, \psi_W)$	0.902,	0.813,	0.726,	0.557

For small levels α these results do not look very good for the Galton test. But it should be kept in mind that the Galton test is compared with the optimal ψ-test (here the Wilcoxon test), that these results are local ($\rho \to 0$) under H_0-contiguous alternatives, and that under H_0-contiguity the effects of (possibly) estimating some characteristics of the alternative $\bar{f} - \bar{g}$ disappear. Therefore the Galton test might give a better performance under fixed alternatives, especially if they are different from "Wilcoxon type."

5. Power extrapolation and power simulation. From (4.5) and the definition of $\mathrm{LARE}(Ga{:}\psi \mid \alpha, b) = e$, we get an approximation of the asymptotic power of Galton's test according to

$$\beta_G(\alpha, b, \rho) \approx 1 - \Phi(u_\alpha - e^{1/2}\rho\langle b, \psi\rangle / \| \psi \|). \tag{5.1}$$

In case of some fixed alternative (F, G) and sample sizes (m, n) we have the representation (2.3), which formally may be embedded in a sequence of type (1.1) if we put

$$b = \bar{f} - \bar{g}, \qquad \rho = (mn/N)^{1/2}. \tag{5.2}$$

Therefore, by combining (5.1), (5.2), and (4.6) we obtain

$$\beta(Ga \mid \alpha, F, G, m, n) = 1 - \Phi(u_\alpha - (mn/N)^{1/2}\langle b, \dot{F}_\alpha\rangle / \varphi(u_\alpha)) \tag{5.3}$$

as an approximation of the power of Galton's test under (F, G), which we call *extrapolated power of Galton's test under* (F, G). Similarly, using (5.2) and (4.5), we get an *extrapolated power of the* ψ*-test under* (F, G),

$$\beta(\psi \mid \alpha, F, G, m, n) = 1 - \Phi(u_\alpha - (mn/N)^{1/2}\langle b, \psi\rangle / \| \psi \|). \tag{5.4}$$

Finally, from (5.2) and (1.3), we get an *extrapolated envelope power under* (F, G),

$$\beta^*(\alpha, F, G, m, n) = 1 - \Phi(u_\alpha - (mn/N)^{1/2} \, \| b \|). \tag{5.5}$$

In order to get an impression of the power behavior of Galton's test under fixed alternatives

TABLE 3

Extrapolated power according to (5.3), (5.4), *and* (5.5) *and Monte Carlo simulation of power of Galton's test, Wilcoxon's test, and optimal* ψ*-tests, respectively, under null hypothesis* H_0 *and seven types of fixed alternatives* (A.1–A.7) *for sample sizes* $m = n = 10, 20, 40$.

Type	Sample Size	Galton's Test	Wilcoxon's Test	Optimal ψ-Test
	$m = n$	Monte Carlo Power/ Extrapol. Power	Monte Carlo Power/ Extrapol. Power	Monte Carlo Power (Level)/ Extrapol. Power
H_0	10	9.2/10	11.2/10	/10
	20	9.9/10	10.7/10	/10
	40	9.9/10	10.1/10	/10
A.1	10	54.0/59	69.2/65	68(10.5)/65
	20	77.2/80	88.6/86	88(10.4)/86
	40	94.6/96	98.9/98	99(10.1)/98
A.2	10	39.9/50	63.2/61	61(8.8)/69
	20	58.4/70	83.3/82	78(6.0)/89
	40	78.7/90	97.5/97	99 (13.2)/99
A.3	10	19.9/26	35.5/34	42(11/6)/69
	20	26.5/35	47.6/47	67(11.3)/89
	40	35.6/49	67.5/67	90(9.3)/99
A.4	10	11.4/14	20.0/18	40(14.6)/66
	20	12.5/16	23.4/22	57(11.5)/87
	40	12.7/19	30.0/30	84(10.6)/98
A.5	10	16.9/22	23.1/21	31(6.2)/62
	20	19.3/29	28.1/28	73(12.9)/83
	40	21.7/40	38.2/38	88(7.9)/97
A.6	10	18.1/24	20.7/19	36(9.0)/65
	20	21.6/32	24.9/24	54(5.6)86
	40	24.0/45	32.3/32	93(13.5)/98
A.7	10	12.4/20	23.5/21	51(11.2)/67
	20	12.6/25	29.0/28	74(9.3/88
	40	10.9/33	32.2/38	95(9.60/98

(F, G) relative to the "optimal" ψ-test and the Wilcoxon test and also in order to see the quality of power extrapolations (5.3) to (5.5), a Monte Carlo study of Galton's test, Wilcoxon's test, and the optimal ψ-tests was done under the *null hypothesis* H_0 and under *seven types of non-parametric alternatives* (A.1–A.7), which are given in the form of (2.3). The *sample sizes* were $m = n = 10, 20, 40$. The Monte Carlo sample size was 10000. In case of Galton's test we used the exact critical values with natural levels $\alpha = 1/11, 2/21, 4/41$, respectively. In case of Wilcoxon's test and optimal ψ-tests, we used the 10%-critical value from normal approximation. Results are given in Table 3.

The alternatives were designed to bring out some special features of Galton's test against Wilcoxon's test. In order to get comparable envelope power for all seven types of alternatives, the distances from H_0 were adjusted by using (5.5), i.e., by using $\| b \| = \| \bar{f} - \bar{g} \|$ as a measure of distance from H_0.

Since the power of rank tests under alternatives (2.3) is independent of the special H in (2.3) and since we assume $m = n$, i.e. $m/N = n/N = ½$, the alternatives are given by Lebesgue densities on [0, 1] of the form

$$\bar{f} = 1 + b/2, \qquad \bar{g} = 1 - b/2, \tag{5.6}$$

with b according to A.1–A.7:

ALTERNATIVE 1 (A.1). $b(t) = 1.3(2t - 1)$, $0 \le t \le 1$, $\| b \|^2 = 0.5633$, $\langle b, \psi_W \rangle = 0.7506$. For this type of alternative the Wilcoxon test is asymptotically optimal.

ALTERNATIVE 2 (A.2). $(5/4)b = -1_{[0,0.5)} + 1_{[0.5,1]}$, $\| b \|^2 = 0.6400$, $\langle b, \psi_W \rangle = 0.6928$. For this type of alternative the rank median test is asymptotically optimal.

ALTERNATIVE 3 (A.3). $b = (-0.3)1_{[0,0.3)} - (1.2)1_{[0.3,0.5)} + (1.2)1_{[0.5,0.7)} + (0.3)1_{[0.7,1]}$, $\| b \|^2 = 0.6300$, $\langle b, \psi_W \rangle = 0.3845$. This type of alternative is designed against Wilcoxon and in favor of Galton.

ALTERNATIVE 4 (A.4). $(5/6)b = -1_{[0.3,0.5)} + 1_{[0.5,0.7)}$, $\| b \|^2 = 0.5760$, $\langle b, \psi_W \rangle = 0.1663$. This type of alternative is even more against Wilcoxon, but the Galton test is not even consistent.

ALTERNATIVE 5 (A.5). $b = -1_{[0,0.25)} + 1_{[0.25,0.5)}$, $\| b \|^2 = 0.5000$, $\langle b, \psi_W \rangle = 0.2165$. This unsymmetric type of alternative is easier for Wilcoxon than A.4, but again the Galton test is not consistent.

ALTERNATIVE 6 (A.6). $(4/3)b = -1_{[0,0.2)} + 1_{[0.2,0.5)} - 1_{[0.5,0.8)} + 1_{[0.8,1]}$, $\| b \|^2 = 0.5625$, $\langle b, \psi_W \rangle = 0.1819$. This type of alternative does *not* correspond to alternatives from $H_1: F \le G$, $F \ne G$ since $B \le 0$ is *not* true. But the Wilcoxon test has asymptotic power 1 (undesirable), whereas the power of the Galton test asymptotically stays below the level α, which is desirable because of the deviation from H_1.

ALTERNATIVE 7 (A.7). $b = (0.3)1_{[0,0.3)} - (0.9)1_{[0.3,0.7)} + (0.9)1_{[0.7,1]}$, $\| b \|^2 = 0.5940$, $\langle b, \psi_W \rangle = 0.2182$. Again, this type of alternative does *not* correspond to alternatives from $H_1: F \le G$, $F \ne G$ since $B \le 0$ is *not* true. The deviation from H_1 is much larger than in A.6, but again the Wilcoxon test has asymptotic power 1 (undesirable), whereas the power of the Galton test asymptotically stays below the level α, which is desirable because of the deviation from H_1.

REMARK. Because of (5.5) and $0.50 \le \| b \|^2 \le 0.64$ in all seven cases, the extrapolated envelope powers under A.1 to A.7 are of comparable order. On the other hand the extrapolated power of Galton's test and Wilcoxon's test show large variability under A.1 to A.7. The last two types of alternatives (A.6, A.7) are included in order to find out whether Galton's test is *specific for* H_1 in finite situations, too.

Discussion of the results and final remarks. The first thing to notice from Table 3 is the rather bad power behavior of the Galton test with respect to the Wilcoxon tests. Even in cases designed in favour of the Galton test (A.3) the power of the Galton test is substantially lower than the power of the Wilcoxon test. Especially, there is no indication for an adaptive behavior of the Galton test. Comparison with the power of the respective optimal tests (envelope power) reveals the bad power behavior of Wilcoxon and Galton tests for alternatives of type A.3 to A.5. On the other hand the simulation shows (at least in the substantial case A.7 of deviation from H_1) that *the Galton test is specific for* H_1, whereas the Wilcoxon power increases with sample sizes. In case of A.6 the deviation from H_1 seems to be too small to see some effect up to sample sizes (40, 40).

Second, notice the very good extrapolations in the case of the Wilcoxon test; the extrapolations are valid up to more than 90% power. In the case of the Galton test the extrapolations are surprisingly good in case of Wilcoxon alternatives (A.1). For the other alternatives (A.2–A.7) the extrapolations are rather bad; but at least they give some rough tendency of the power behavior. The results in the case of A.7 show that the extrapolation of slopes of local asymptotic power does not bring out that the Galton test is specific for H_1.

Finally, the simulation of the locally optimal rank tests with respect to alternatives of type A.1 to A.7 shows that the normal approximation under H_0 depends very much on the type of the score function, leading to substantial deviations of the empirical levels from the asymptotic level $\alpha = 0.10$. On the other hand, the corresponding empirical power shows substantial deviation from the extrapolated envelope power, also in cases where the empirical level is approximately 10% (cf. A.4). Thus, the extrapolated envelope power is only a limited measure of distance from hypothesis.

The ideas of the present paper carry over to other testing problems in a natural way. E.g. the one sample problem of testing "symmetry" versus "positive asymmetry" may be treated in an analogous way, leading to Galton's test for symmetry as defined in Bickel and Hodges (1967).

As a summary, we may conclude that some very rough estimation of scores is not sufficient in order to get an adaptive procedure. For consistent estimation of scores under mild assumptions on the underlying alternatives see Behnen, Neuhaus and Ruymgaart (1982).

Acknowledgment. The authors are very grateful to W. Cöllen and W. Schütt for writing the programs for simulation and integration.

REFERENCES

BEHNEN, K. (1972). A characterization of certain rank-order tests with bounds for the asymptotic relative efficiency. *Ann. Math. Statist.* **43** 1839–1851.

BEHNEN, K. (1975). The Randles-Hogg test and an alternative proposal. *Comm. Statist.* **4** 203–238.

BEHNEN, K. (1981). Nichtparametrische Statistik: Zweistichproben Rangtests. *Zeitschr. Angew. Math. Mech.* **61** T203–T212.

BEHNEN, K. and NEUHAUS, G. (1975). A central limit theorem under contiguous alternatives. *Ann. Statist.* **3** 1349–1353.

BEHNEN, K., NEUHAUS, G., and RUYMGAART, F. (1982). A Chernoff-Savage theorem for rank statistics with estimated scores, and rank estimators of score-functions. Preprint, Univ. of Hamburg.

BICKEL, P. J. and HODGES, J. L. (1967). The asymptotic theory of Galton's test and a related simple estimate of location. *Ann. Math. Statist.* **38** 73–89.

BILLINGSLEY, P. (1968). *Convergence of Probability Measures.* Wiley, New York.

CSÁKI, E., and VINCZE, I. (1961). On some problems connected with the Galton-test. *Publ. Math. Inst. Hung. Acad. Sci.* **6** 97–109.

CSÁKI, E., and VINCZE, I. (1963). On some distributions connected with the arcsine law. *Publ. Math. Inst. Hung. Acad. Sci.* **8** 281–290.

FELLER, W. (1968). *An Introduction to Probability Theory and its Applications. Vol.* I. Wiley, New York.

HÁJEK, J. (1974). Asymptotic sufficiency of the vector of ranks in the Bahadur sense. *Ann. Statist.* **2** 75–83.

HÁJEK, J. and ŠIDÁK, Z. (1967). *Theory of Rank Tests.* Academic, New York.

PYKE, R. and SHORACK, G. R. (1968). Weak convergence of a two-sample empirical process and a new approach to Chernoff-Savage theorems. *Ann. Math. Statist.* **39** 755–771.

INSTITUT FUR MATHEMATISCHE STOCHASTIK
UNIVERSITAT HAMBURG
BUNDESSTRASSE 55
2000 HAMBURG 13, WEST GERMANY