# SIGNIFICANCE LEVEL AND POWER[1]

## By E. L. Lehmann

*University of California, Berkeley*

**1. Summary and introduction.** Significance testing, as described in most textbooks, consists in fixing a standard significance level $\alpha$ such as .01 or .05 and rejecting the hypothesis $\theta = \theta_0$ if a suitable statistic $Y$ exceeds $C$ where $P_{\theta_0}\{Y > C\} = \alpha$. Such a procedure controls the probability of false rejection (error of the first kind) at the desired level $\alpha$ but leaves the power of the test and hence the probability of an error of the second kind to the mercy of the experiment. It seems more natural when deciding on a significance level (and this suggestion is certainly not new) to take into account also what power can be achieved with the given experiment. In Section 3 a specific suggestion will be made as to how to balance $\alpha$ against the power $\beta$ obtainable against the alternatives of interest.

The adoption of this or some similar rule for choosing a significance level has important consequences for the theory of testing composite hypotheses, where nuisance parameters are present. Since the quantity $\alpha$ is then potentially a function of the nuisance parameter $\vartheta$, the classical rule of a fixed significance level leads to the condition that the tests be *exact* or *similar*, that is, that $\alpha(\vartheta)$ equal the preassigned value $\alpha$ for all $\vartheta$. On the other hand, the power $\beta$ that can be attained against any alternative $\theta = \theta_1$ frequently depends on $\vartheta$. The requirement that $\alpha(\vartheta)$ and $\beta(\vartheta)$ be in a certain balance thus leads to tests which are not similar and hence do not agree with the standard solutions.

To obtain a suitable setting for this discussion, we consider first a minimal complete class of tests for testing the hypothesis $H: \theta \leq \theta_0$ in a multiparameter exponential family (Section 2). The proposed $\alpha$, $\beta$-relation is discussed in Section 3, and in Section 4 is applied to the exponential family. Section 5 gives some illustrations of the theory.

**2. A complete class theorem.** Many standard testing problems concern an exponential family of distributions, which has probability densities of the form

$$(1) \qquad p_{\theta,\vartheta}(x) = C(\theta, \vartheta) \exp\left[ \theta U(x) + \sum_{i=1}^{r} \vartheta_i T_i(x) \right] h(x)$$

with respect to a $\sigma$-finite measure $\mu$, where $\theta$, $U$, the $\vartheta_i$ and $T_i$ are real-valued and where $\vartheta = (\vartheta_1, \cdots, \vartheta_r)$. In this family, the statistics $U$ and $T = (T_1, \cdots, T_r)$ constitute a set of sufficient statistics for $(\theta, \vartheta)$.

The problem of testing the hypothesis $H: \theta \leq \theta_0$ against the one-sided al-

7742

ternatives $\theta > \theta_0$ has been treated by many authors (usually in the formulation $\theta = \theta_0$ against $\theta > \theta_0$). The solution of this testing problem according to the Neyman-Pearson theory is the uniformly most powerful unbiased test; this depends only on $U$ and $T$ and is given by the critical function[2]

$$(2) \qquad \phi(u, t) = \begin{cases} 1 & \text{if } u > C(t), \\ \gamma(t) & \text{if } u = C(t), \\ 0 & \text{if } u < C(t), \end{cases}$$

where the functions $C$ and $\gamma$ are determined by the conditions $E_{\theta_0}[\phi(U, T) \mid T = t] = \alpha$ and $E_{\theta_0}[U\phi(U, T) \mid T = t] = \alpha\, E_{\theta_0}[U \mid T = t]$ for all $t$. The condition of unbiasedness

$$E_{\theta,\vartheta}\phi(U, T) \lessgtr \alpha \qquad \text{as} \qquad \theta \lessgtr \theta_0 ,$$

and that of similarity

$$E_{\theta_0,\vartheta}\phi(U, T) = \alpha \qquad \text{for all } \vartheta$$

which it implies and which by itself is sufficient to justify the test, are not inherent in the problem but are imposed, at least in part, to facilitate the solution. Before proposing an alternative approach, it is interesting to see how far the problem can be reduced without the introduction of extraneous principles. This can be done by viewing it within the framework of decision theory.

Let $d_0$ and $d_1$ denote the decisions of accepting and rejecting the hypothesis $H$, and denote by $L_i(\theta, \vartheta)$ the loss resulting from decision $d_i$ when $(\theta, \vartheta)$ are the true parameter values. Then for fixed $\vartheta$, the function $L_0(\theta, \vartheta)$ typically will be zero for $\theta \leq \theta_0$ and increasing for $\theta \geq \theta_0$, while $L_1(\theta, \vartheta)$ will be decreasing for $\theta \leq \theta_0$ and zero for $\theta \geq \theta_0$. In particular, the difference then satisfies

$$(3) \qquad L_1(\theta, \vartheta) - L_0(\theta, \vartheta) \gtrless 0 \qquad \text{as} \qquad \theta \gtrless \theta_0 .$$

The risk function of a test $\phi$, which is the expected loss resulting from its use considered as a function of the parameters, is

$$(4) \qquad R_\varphi(\theta, \vartheta) = \int \{\varphi(U(x), T(x))L_1(\theta, \vartheta) \\ + [1 - \varphi(U(x), T(x))]L_0(\theta, \vartheta)\} p_{\theta,\vartheta}(x)\, d\mu(x).$$

Let $\mathcal{C}$ be the class of all tests satisfying (2) for some functions $C$ and $\gamma$. For all loss functions satisfying (3) it was shown by Truax [13] that $\mathcal{C}$ is essentially complete; that is, given any $\varphi$ there exists $\varphi' \ \varepsilon \ \mathcal{C}$ such that

$$(5) \qquad R_{\varphi'}(\theta, \vartheta) \leq R_\varphi(\theta, \vartheta) \qquad \text{for all } (\theta, \vartheta).$$

We shall now prove that among essentially complete classes, $\mathcal{C}$ is minimal in the sense that if (5) holds for two tests $\varphi, \varphi'$ in $\mathcal{C}$, then $\varphi = \varphi'$ a.e. $\mu$.*

---

[2] See for example [7].

* Recently I learned that this result has been obtained also by D. L. Burkholder. His results are sketched in Abstract 18, *Ann. Math. Stat.*, Vol. 29 (1958), p. 616.

7744

Let $\varphi$ and $\varphi'$ belong to $\mathcal{C}$ and let

$$(6) \qquad \alpha(\vartheta) = E_{\theta_0,\vartheta}\,\varphi(U, T), \qquad \alpha'(\vartheta) = E_{\theta_0,\vartheta}\,\varphi'(U, T).$$

(i) If the functions $\alpha$ and $\alpha'$ do not agree for all $\vartheta$, suppose without loss of generality that there exists $\vartheta_0$ such that $\alpha(\vartheta_0) < \alpha'(\vartheta_0)$. Since for $\vartheta = \vartheta_0$, the expected values of $\varphi$ and $\varphi'$ are continuous functions of $\theta$, there exist $\theta_1 < \theta_0 < \theta_2$ such that

$$(7) \qquad E_{\theta,\vartheta_0}\varphi(U, T) < E_{\theta,\vartheta_0})\varphi'(U, T) \qquad \text{for } \theta = \theta_1 \text{ and } \theta = \theta_2.$$

Then $R_\phi(\theta_1, \vartheta_0) < R_{\phi'}(\theta_1, \vartheta_0)$ and $R_\phi(\theta_2, \vartheta_0) > R_{\phi'}(\theta_2, \vartheta_0)$, and hence neither of the procedures $\phi$ and $\phi'$ is uniformly better than the other. (ii) Suppose on the other hand that $\alpha(\vartheta) \equiv \alpha'(\vartheta)$. The standard proof showing a similar test satisfying (2) to be uniformly most powerful similar also shows that a test $\phi_0$ satisfying (2) and

$$(8) \qquad E_{\theta_0,\vartheta}\,\phi_0(U, T) = \alpha(\vartheta) \qquad \text{for all } \vartheta$$

is uniformly most powerful among all tests satisfying (8). The tests $\phi$ and $\phi'$ are therefore both uniformly most powerful within this class and hence

$$E_{\theta,\vartheta}\,\phi(U, T) = E_{\theta,\vartheta}\,\phi'(U, T) \qquad \text{for all } \theta > \theta_0 \text{ and all } \vartheta.$$

Since the family of distributions of the sufficient statistics $(U, T)$ is complete, it follows that $\phi(u, t) = \phi'(u, t)$ a.e., as was to be proved.

**3. Significance level and power.** It follows from the result of the preceding section that the class $\mathcal{C}$ of tests (2) represents the maximum reduction that can be achieved by comparing only tests of which one has a uniformly better risk function than the other. The selection of a specific test from $\mathcal{C}$, involves two difficulties. It requires the adoption of some principle (Bayes, minimax, etc.) leading to a definite choice;[3] in addition, it requires knowledge of the loss functions $L_0$ and $L_1$. An alternative approach, utilizing the fortunate circumstance that the complete class is independent of the actual loss functions (subject only to their satisfying (3)), consists in making the choice by some simple rule of thumb, which does not require (the usually unavailable) knowledge of these losses.

Consider the simplest case of the family (1) with $r = 0$, which involves no nuisance parameters. The family of tests (2) is then a one-parameter family, one test corresponding to each value of

$$\alpha_0 = E_{\theta_0}\phi(X), \qquad\qquad 0 \leqq \alpha_0 \leqq 1.$$

A simple method of choice consists in specifying a value of $\alpha_0$ and selecting the test corresponding to this value. This need not be a purely formal or arbitrary

---

[3] Particular proposals of this kind that have been made in the literature include those of Jeffreys [5] involving considerations of *a priori* probabilities, and of Lindley [8] based on the concept of unlikelihood.

procedure since $\alpha_0$ as the maximum probability of false rejection is of course an important quantity in its own right.

Nevertheless, as was pointed out in Section 1, the above rule appears to neglect too many aspects of the problem. In particular, suppose that the alternatives of primary interest, for which it is important to reject the hypothesis, are those satisfying $\theta \geqq \theta_1$ ($\theta_0 < \theta_1$). Since the power function of any test (2) is increasing in $\theta$, the probability $\beta_1$ of rejection when $\theta = \theta_1$ is the minimum power against these alternatives. It seems then reasonable that the choice of test should involve at least $\beta_1$ in addition to $\alpha_0$.

The quantities $\alpha_0$ and $\alpha_1 = 1 - \beta_1$ are the error probabilities associated with the problem of testing the simple hypothesis $\theta = \theta_0$ against the simple alternative $\theta = \theta_1$. The attainable pairs ($\alpha_0$, $\alpha_1$) form a convex set, the lower boundary of which corresponds to the admissible tests (2). This lower boundary is a convex curve $S$ connecting the points $(0, 1)$ and $(1, 0)$, and what is needed is a reasonable way of selecting a point on each such curve. One possible approach to this question is in terms of indifference curves. Suppose that a system of curves could be specified in the ($\alpha_0$, $\alpha_1$)-plane such that any two points lying on the same curve are equally desirable, with the curves closer to the origin being more desirable than those further away. The optimum test would then be given by that point of $S$ lying on the indifference curve closest to the origin (Fig. 1).

It seems likely that even this approach is too complex for most applications. To obtain an even simpler formulation, consider once more the rule of fixing the significance level without regard to power. If the level is $\alpha$, this means restricting attention to the points ($\alpha_0$, $\alpha_1$) lying on the vertical line segment $L:\alpha_0 = \alpha$, $0 \leqq \alpha_1 \leqq 1 - \alpha$. The test then corresponds to the point ($\alpha_0$, $\alpha_1$), which is the
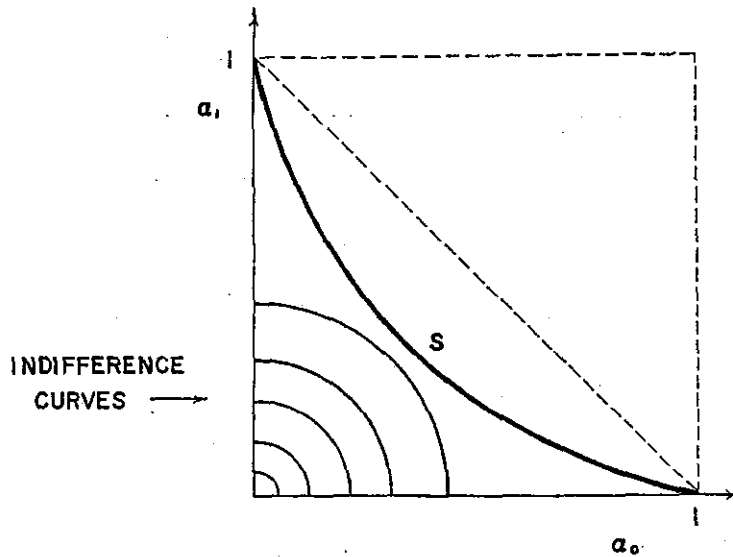


Fig. 1

intersection
that the en
therefore be
ficiently clo
is sufficientl
the assumed

A more re
$L$ by a cur
with $f(0) =$

(9)

Since $\alpha_0 \leqq$
$1/(k + 1)$ is
If $\beta_1 = 1 -$
pairs of val

with .1 bein
One woul
$\alpha_0 = .05, \beta_1$
by taking a
it can frequ
but only by

To avoid
proposed as
it seems clea

7748

Fig. 2

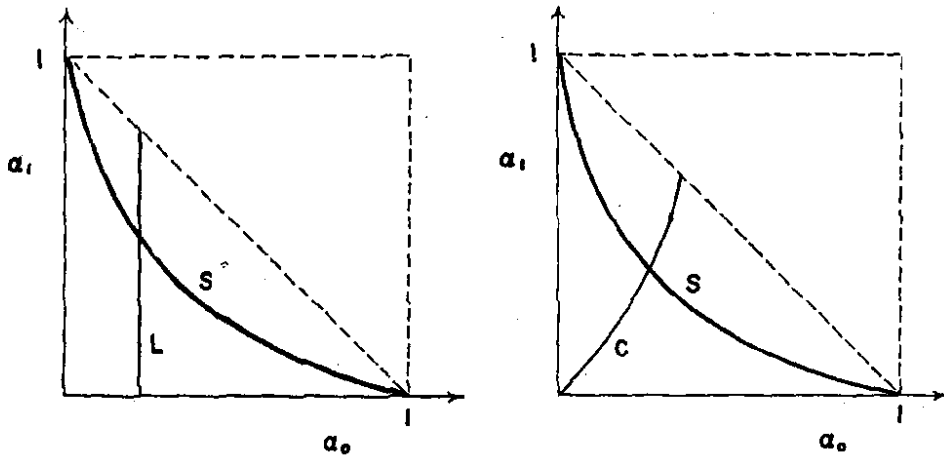intersection of $S$ and $L$. This procedure is commonly justified on the grounds that the error of the first·kind is of a higher order of importance, and should therefore be controlled at the prescribed level. However, if the curve $S$ is sufficiently close to the $\alpha_0$- and $\alpha_1$-axis, as will always be the case if the sample size is sufficiently large, then $\alpha_1$ is much smaller than $\alpha_0$, which is inconsistent with the assumed relative importance of the two errors.

A more reasonable solution is obtained if one replaces the vertical line segment $L$ by a curve $C : \alpha_1 = f(\alpha_0)$ where $f$ is a continuous strictly increasing function with $f(0) = 0$. A particularly simple choice for $f$ is a linear function

$$(9) \qquad\qquad \alpha_1 = k\alpha_0 .$$

Since $\alpha_0 \leqq 1 - \alpha_1$ for all admissible tests, one has $\alpha_0 \leqq 1/(k + 1)$ so that $1/(k + 1)$ is an upper bound.for $\alpha_0$. As an example, consider (9) with $k = 9$. If $\beta_1 = 1 - \alpha_1$ denotes the power of a test against the alternative $\theta_1$, some typical pairs of values of $(\alpha_0, \beta_1)$ are

| $\alpha_0$ | .1 | .05 | .04 | .03 | .02 | .01 | .005 |
|---|---|---|---|---|---|---|---|
| $\beta_1$ | .1 | .55 | .64 | .73 | .82 | .91 | .955 |

with .1 being an upper bound for $\alpha_0$.

One would of course hope to avoid cases such as $\alpha_0 = .1$, $\beta_1 = .1 \cdot$ or even $\alpha_0 = .05, \beta_1 = .55$. When no nuisance parameters are present, this can be achieved by taking a sample of sufficient size. In the composite case, on the other hand, it can frequently not be achieved by samples of fixed size no matter how large, but only by resorting to sequential experimentation.

To avoid misunderstandings, it should be emphasized that (9) is not being proposed as a logically convincing rule, nor as one fitting all occasions. Actually, it seems clear that no rule satisfying these requirements exists, except the Bayes

7750

solution when sufficient knowledge concerning losses and *a priori* probabilities is available. In the absence of this knowledge it may be convenient to employ a simple rule of thumb. Such a rule is in fact being used in much of present practice: It consists in choosing $\alpha$ to be .05 or .01 depending on the seriousness attached to the committing of an error of the first kind. To this, (9) is suggested as an alternative which appears to be more reasonable in many cases.

It so happens that (9) is the minimax solution if the loss for rejecting $H: \theta \leqq \theta_0$ is $a_0$ when $H$ is true, and the loss is $a_1$ for accepting $H$ when $\theta \geqq \theta_1$, where the constant $k$ of (9) is then given by $k = a_0/a_1$. However, this is not the basis for the present suggestion of (9), and the minimax property does not carry over to the application to be made in the next section to composite hypotheses.

**4. Conditional tests.** We return now to the composite case of the exponential family (1) with $r > 0$. The minimal complete class $\mathfrak{C}$ is then more complex than in the preceding section, its members being characterized by the function $\alpha(\vartheta)$ instead of the single number $\alpha_0$. Given any function $\alpha(\vartheta)$, which is the expectation of some critical function $\phi$, there exists a unique member of $\mathfrak{C}$ whose expectation function for $\theta = \theta_0$ is also $\alpha(\vartheta)$. This uniformly minimizes the risk (and maximizes the power) among all critical functions having this expectation.

If the alternatives of interest are as before those satisfying $\theta \geqq \theta_1$, let $\beta(\vartheta)$ denote the power function of a test against the alternative $(\theta_1, \vartheta)$. The proposal made in the preceding section suggests selecting that member of $\mathfrak{C}$ which satisfies

$$(10) \qquad\qquad 1 - \beta(\vartheta) = k\alpha(\vartheta) \qquad \text{for all } \vartheta.$$

However, this relationship depends on the particular parametrization chosen, and we shall not discuss it here. Instead an alternative approach will be proposed in which this difficulty does not arise.

Consider once more the case of the similar test with $\alpha(\vartheta) \equiv \alpha$. Since $T$ is a complete sufficient statistic for $\vartheta$ when $\theta = \theta_0$, the functions $C$ and $\gamma$ of (2) are determined by the requirement that the conditional probability of rejection

$$\alpha^*(t) = P_{\theta_0}\{U > C(t) \mid t\} + \gamma(t)P_{\theta_0}\{U = C(t) \mid t\}$$

be equal to $\alpha$ for all $t$.[4] However, the conditional power $\beta^*(t) = P_{\theta_1}\{\text{rejecting } H \mid t\}$ of the test against the alternative $\theta = \theta_1$, typically depends on $t$. The question then arises: Suppose that $\beta^*(t)$ is quite small for the observed $t$, or quite high; is this value not more relevant to the case in hand than the average value $\beta(\vartheta)$?

Without entering into the difficulties raised by this question, there is an alternative and simpler justification for considering $\beta^*(t)$. The actual power $\beta$ against the alternative $\theta = \theta_1$ generally depends on the nuisance parameter $\vartheta$ and is therefore unknown. It can however be estimated from the observations,

---

[4] This method of constructing exact tests was originated by Bartlett [1] and Neyman [9]. That in the present case it provides the totality of such tests has been noted by many authors. For a recent discussion and references see [7].

ori probabilities is
nient to employ a
f present practice:
:iousness attached
s suggested as an
es.
ejecting $H: \theta \leq \theta_0$
$\geq \theta_1$, where the
s not the basis for
not carry over to
ypotheses.

)f the exponential
tore complex than
the function $\alpha(\vartheta)$
ch is the expecta-
r of $\mathfrak{C}$ whose ex-
tizes the risk (and
xpectation.
$\theta \geq \theta_1$, let $\beta(\vartheta)$
$\vartheta$). The proposal
f $\mathfrak{C}$ which satisfies

itrization chosen,
t will be proposed

= $\alpha$. Since $T$ is a
$\gamma$ and $\gamma$ of (2) are
y of rejection

$t\}$

$= P_{\theta_1}\{$rejecting
pends on $t$. The
te observed $t$, or
than the average

ion, there is an
t actual power $\beta$
nce parameter $\vartheta$
the observations,

1] and Neyman [9].
en noted by many

and $\beta^*(T)$ is the unbiased estimate with (uniformly) minimum variance. That it is unbiased is clear since $\beta(\vartheta) = E_{\theta_1, \vartheta} \beta^*(T)$. The minimum variance property is an immediate consequence of the completeness of the sufficient statistic $T$ for $(\theta_1, \vartheta)$ and of Theorem 5.1 of [7].

Analogous remarks apply in the more general case, in which the tests are not required to be exact. If the relevant frame of reference is obtained by considering $t$ as fixed, the error probabilities of interest are the conditional probabilities $\alpha_0^*(t) = P_{\theta_0}$ (rejecting $H \mid t$) and $\alpha_1^*(t) = P_{\theta_1}$ (accepting $H \mid t$), and the quantities $C(t)$ and $\gamma(t)$ can therefore be determined from the relation

$$(11) \qquad \alpha_1^*(t) = k\alpha_0^*(t).$$

The resulting test will of course not be similar. However, since $\alpha_0^*(t) \leq 1/(k + 1)$ for all $t$, the quantity $1/(k + 1)$ is an upper bound also for the average probability $\alpha_0(\vartheta)$ of an error of the first kind.

The above discussion applies only to problems in which the parameter of interest is one of the "natural" parameters of the exponential distribution (1). As was pointed out in [7], any parameter of the form $\theta + \sum a_i \vartheta_i$ is natural for a suitable definition of $U$, the $T$'s and $\vartheta$'s. When the parameter of interest is not of this form, related methods may be applicable as is indicated by the following example.

If $X_1, \cdots, X_n$ are a sample from a normal distribution $N(\xi, \sigma^2)$, neither the parameter $\xi$ nor $\xi/\sigma$ are of this form. The problem of testing $\xi/\sigma \leq \delta_0$ against $\xi/\sigma \geq \delta_1$ can be reduced by invariance considerations to the statistic $\bar{X}/[\sum(X_i - \bar{X})^2]^{\frac{1}{2}}$, the distribution of which depends on the single parameter $\delta = \xi/\sigma$. If $\alpha_i = P_\delta\{\bar{X} > C[\sum(X_i - \bar{X})^2]^{\frac{1}{2}}\}$, the quantity $C$ can be determined so that $\alpha_1 = k\alpha_0$. The problem of testing $\xi \leq \xi_0$ against $\xi \geq \xi_1$ appears to be more difficult; a possible approach may be that of [4], Section 3.

**5. Examples.** We shall now briefly indicate some examples in which the natural parameter $\theta$ is the relevant one so that the method of the preceding sections is applicable. Of these, Examples 1, 2, 3 have been treated by the same method (but from a different point of view) by Tocher [12], and Examples 2, 3 by Sverdrup [11].

EXAMPLE 1. Let $X, Y$ be independent Poisson variables with $E(X) = \lambda$, $E(Y) = \mu$, and consider the problem of testing $\mu/\lambda \leq a_0$ against $\mu/\lambda \geq a_1$. The joint distribution of $X, Y$ forms an exponential family with $T = X + Y$, $U = Y$, $\theta = \log(\mu/\lambda)$ and $\vartheta = \log \lambda$. The conditional distribution of $Y$ given $X + Y = t$ is a binomial distribution corresponding to the success probability $p = \mu/(\lambda + \mu)$ and number of trials equal to $t$. In terms of $p$, the hypothesis and class of alternatives becomes $p \leq a_0/(a_0 + 1)$ and $p \geq a_1/(a_1 + 1)$ so that the test satisfying (2) and (11) can be determined from a table of the binomial distribution.

EXAMPLE 2. If $X, Y$ are independent variables with binomial distributions $b(p_1, m)$ and $b(p_2, n)$, their joint distribution has the exponential form (1) with

7754

7754

$T = X + Y$, $U = Y$, $\theta = \log(p_2/q_2 \div p_1/q_1)$ and $\vartheta = \log(p_1/q_1)$. The method is therefore applicable to the problem of testing $p_2/q_2 \leqq a_0(p_1/q_1)$, and in particular $p_2 \leqq p_1$ by letting $a_0 = 1$, against the alternatives $p_2/q_2 \geqq a_1(p_1/q_1)$. Putting $\rho = (p_2/q_2) \div (p_1/q_1)$, the conditional distribution of $Y$ given $t$ is

$$(12) \quad P_\rho\{Y = y \mid X + Y = t\} = C_t(\rho) \binom{m}{t - y} \binom{n}{y} \rho^y, \quad y = 0, 1, \cdots, t,$$

which for $\rho = 1$ reduces to the hypergeometric distribution.

EXAMPLE 3. In a $2 \times 2$ table representing the results of classifying a sample of size $s$ according to two characteristics $A$ and $B$, the joint distribution of the numbers $X$, $Y$, $Y'$ in the

|       | $A$  | $\tilde{A}$ |     |
|-------|------|-------------|-----|
| $B$   | $X$  | $X'$        | $M$ |
| $\tilde{B}$ | $Y$  | $Y'$        | $N$ |
|       | $T$  | $T'$        | $S$ |

categories $AB$, $A\tilde{B}$ and $\tilde{A}\tilde{B}$ constitute an exponential family with $U = Y$, $T_1 = X + Y$, $T_2 = Y + Y'$ and $\theta = \log(p_{AB}p_{\tilde{A}B}/p_{AB}p_{\tilde{A}B})$. Putting $\Delta = (p_{AB}p_{\tilde{A}B}/p_{AB}p_{\tilde{A}B})$ one finds

$$p_{AB} = p_A p_B + \frac{1 - \Delta}{\Delta} p_{\tilde{A}B}p_{AB}; \quad p_{\tilde{A}B} = p_{\tilde{A}} p_B + \frac{1 - \Delta}{\Delta} p_{\tilde{A}B}p_{AB}$$

$$p_{AB} = p_A p_{\tilde{B}} - \frac{1 - \Delta}{\Delta} p_{\tilde{A}B}p_{AB}; \quad p_{\tilde{A}B} = p_{\tilde{A}} p_B - \frac{1 - \Delta}{\Delta} p_{\tilde{A}B}p_{AB}$$

where $p_{AB}$ denotes the probability of having the characteristics $A$ and $B$, $p_A = p_{AB} + p_{A\tilde{B}}$ the probability of having the characteristic $A$, etc. The quantity $\Delta$ is therefore a measure of the degree of dependence,[5] $\Delta = 1$ corresponding to independence, $\Delta < 1$ to negative and $\Delta > 1$ to positive dependence. The method of the preceding section is applicable to testing $\Delta \leqq 1$ or more generally $\Delta \leqq \Delta_0$ against the alternatives $\Delta \geqq \Delta_1$. The conditional distribution of $Y$ given $X + Y = t$, $Y + Y' = n$ is given by (12) with $\Delta$ in place of $\rho$.

EXAMPLE 4. Consider a number of paired comparisons $(U_k, V_k)$ where only the sign of the differences $W_k = V_k - U_k$ are observed for each pair $k = 1, \cdots, n$. If the probability of a positive, negative and zero observation are $p_+$, $p_-$ and $p_0$ in each case and if the comparisons are independent, the joint distribution of the numbers $X$, $Y$ and $Z$ of positive, negative and zero cases is the multinomial distribution

$$\frac{n!}{x!\,y!\,z!} p_+^x p_-^y p_0^z.$$

---

[5] $\Delta$ is equivalent to Yule's measure of association, which is $Q = (1 - \Delta)/(1 + \Delta)$. For a discussion of this and related measures, see [2].

7756

7756

$q_1$) and $\vartheta = \log (p_1/q_1)$. The
f testing $p_2/q_2 \leqq a_0(p_1/q_1)$, and
inst the alternatives $p_2/q_2 \geqq$
conditional distribution of $Y$

$$\binom{}{y}\binom{n}{y}\rho^\nu, \qquad \nu = 0, 1, \cdots, t,$$

istribution.

results of classifying a sample of
}, the joint distribution of the

———

$$\frac{\begin{matrix}M\\N\end{matrix}}{S}$$

onential family with $U = Y$,
$p_{AB}p_{\bar A\bar B}$). Putting $\Delta = (p_{AB}p_{\bar A\bar B}/$

$$p_{\bar A}p_B + \frac{1-\Delta}{\Delta}\,p_{\bar A B}p_{AB}$$

$$p_{\bar A}p_B - \frac{1-\Delta}{\Delta}\,p_{\bar A B}p_{AB}$$

characteristics $A$ and $B$, $p_A =$
eristic $A$, etc. The quantity $\Delta$ is
ace,$^5$ $\Delta = 1$ corresponding to
sitive dependence. The method
$\leqq 1$ or more generally $\Delta \leqq \Delta_0$
onal distribution of $Y$ given
$\Delta$ in place of $\rho$.
nparisons $(U_k, V_k)$ where only
rved for each pair $k = 1, \cdots, n$.
) observation are $p_+$, $p_-$ and $p_0$
nt, the joint distribution of the
zero cases is the multinomial

hich is $Q = (1-\Delta)/(1+\Delta)$. For a

---

This is an exponential family with $U = Y$, $T = Z$, $\theta = \log (p_+/p_-)$ and $\vartheta = \log (p_0/p_-)$. The test of $p_+ \leqq p_-$ (or $p_+ \leqq a_0 p_-$) against $p_+ \geqq a_1 p_-$ is therefore performed conditionally given $Z = t$. Since the conditional distribution of $Y$ given $Z = t$ is the binomial distribution $b(p_+/(p_+ + p_-), n - t)$, the constants $C(t)$ and $\gamma(t)$ for which the test satisfies (2) and (11) can be obtained from the binomial tables.$^6$

EXAMPLE 5. Let $Y_1, \cdots, Y_N$ be independently distributed according to the binomial distributions $b(p_i, n_i)$ $i = 1, \cdots, N$ where

$$p_i = 1/[1 + e^{-(\alpha+\beta x_i)}]$$

This is the model frequently assumed in bioassay, where $x_i$ denotes the dose or some function of the dose such as its logarithm, of a drug given to $n_i$ experimental subjects and where $Y_i$ is the number among these subjects which respond to the drug at level $x_i$. Here the $x_i$ are known, and $\alpha$ and $\beta$ are unknown parameters. The joint distribution of the $Y$'s is

$$(13) \qquad e^{\alpha\Sigma y_i + \beta\Sigma x_i y_i} \prod_{i=1}^{N} \binom{n_i}{y_i}\left[\frac{e^{-(\alpha+\beta x_i)}}{1 + e^{-(\alpha+\beta x_i)}}\right]^{n_i},$$

which is an exponential family with the parameters $\alpha$, $\beta$ and sufficient statistics $\sum Y_i$, $\sum x_i Y_i$. The method is therefore applicable to testing $\alpha \leqq \alpha_0$ against $\alpha \geqq \alpha_1$ or $\beta \leqq \beta_0$ against $\beta \geqq \beta_1$. It is interesting to note that for the particular case $x_i = ic$ and $H:\beta \leqq 0$, the conditional test given $Y = t$ is a form of the Wilcoxon test in a setting similar to that discussed by Haldane and Smith [3].

As a last example we mention without going into details the comparison of two distributions of type (13). If the parameters in these are $\alpha$, $\beta$ and $\alpha'$, $\beta'$ the differences $\alpha' - \alpha$ and $\beta' - \beta$ are natural parameters of the resulting exponential families, and can therefore be tested by the method discussed here.

## REFERENCES

[1] M. S. BARTLETT, "Properties of sufficiency and statistical tests," Proc. Roy. Soc. London, Vol. 160 (1937), pp. 268–281.

[2] LEO A. GOODMAN AND WILLIAM H. KRUSKAL, "Measures of association for mass classifications," J. Amer. Stat. Assn., Vol. 49 (1954), pp. 732–764.

[3] J. B. S. HALDANE AND CEDRIC A. B. SMITH, "A simple exact test for birth order effect," Ann. Eugenics, Vol. 14 (1948), pp. 117–124.

[4] J. L. HODGES, JR. AND E. L. LEHMANN, "Testing the approximate validity of statistical hypotheses," J. Roy. Stat. Soc., Ser. B., Vol. 16 (1954), pp. 261–268.

[5] H. JEFFREYS, "Theory of probability (2d ed.), Oxford: Clarendon Press, 1948.

[6] SAMUEL KARLIN AND HERMAN RUBIN, "The theory of decision procedures for distribu-

---

$^6$ This problem has been considered previously in [10]. The statement made there that the test satisfying (2) and $\alpha^*(t) \equiv \alpha$ is uniformly most powerful is too strong. The test is however uniformly most powerful among all similar (and hence all unbiased) tests. An analogous remark applies to Example 2, which is among those considered in [12]. As was proved there, the test which is conditionally unbiased at fixed level $\alpha$ for each $t$, is uniformly unbiased; it is however not uniformly most powerful without this restriction as is claimed in [12] for all the cases treated there.

7758

tions with monotone likelihood ratio," *Ann. Math. Stat.*, Vol. 27 (1956), pp. 272–299.

[7] E. L. LEHMANN AND HENRY SCHEFFÉ, "Completeness, similar regions and unbiased estimation," *Sankhya*, Vol. 10 (1950), pp. 305–340 and Vol. 15 (1955), pp. 219–236.

[8] D. V. LINDLEY, "Statistical inference," *J. Roy. Stat. Soc., Ser. B.*, Vol. 15 (1953), pp. 30–76.

[9] J. NEYMAN, "Outline of a theory of statistical estimation based on the classical theory of probability," *Philos. Trans. Roy. Soc. London*, Vol. 236 (1937), pp. 333–380.

[10] JOSEPH PUTTER, "The treatment of ties in some nonparametric tests," *Ann. Math. Stat.*, Vol. 26 (1955), pp. 368–386.

[11] ERLING SVERDRUP, "Similarity, unbiasedness, minimaxibility, and admissibility of statistical test procedures," *Skand. Aktuarietids.*, Vol. 36 (1953), pp. 64–86.

[12] K. D. TOCHER, "Extension of Neyman-Pearson theory of tests to discontinuous variates," *Biometrika*, Vol. 37 (1950), pp. 130–144.

[13] DONALD R. TRUAX, "Multidecision problems for the multivariate exponential family," Stanford Technical Report No. 32 (1955) (to be published).

7760