Statistical Papers 39, 75 – 85 (1998)

Statistical Papers

Permutation tests – a revival?!
I. Optimum properties

Jens Gebhard and Norbert Schmitz

Received: January 9, 1996; revised version: May 2, 1996

It is shown that permutation tests have optimum properties for interesting classes of continuous distributions as well as for discrete ones. General conditions sufficient for uniformly maximal power on subclasses are given. Moreover, a variety of examples is presented.

Key Words: Permutation tests, most powerful on subclasses, two-sample problem.

1. Introduction

Permutation tests play an intermediate role "between" classical parametric procedures and rank tests: On the one hand they take into account considerably larger classes of distributions than parametric models, on the other hand they make (for metric scales) full use of the data and avoid the loss of information caused by the reduction to ranks. Nevertheless, they seem to be of minor importance for practical statistics, the main objection being the computational effort needed for constructing the critical region – but in the era of computer intensive methods this argument sounds like a relic of the 60'es.

[1]Research supported by Deutsche Forschungsgemeinschaft (DFG) grant No. Schm 677/5-1

In the sequel we will show that permutation tests have convincing optimum properties for continuous distributions as well as for discrete ones. The considerations are carried out for the two-sample problem; similar results hold true for other types of permutation tests. For the comparison of two treatments, these are applied to n_1 resp. n_2 homogeneous items. It is assumed that the observational data are realizations of independent random variables $X_{1,1}, \ldots, X_{1,n_1}$ and $X_{2,1}, \ldots, X_{2,n_2}$ resp. whose distributions $P_1 = P^{X_{1,i}}$ and $P_2 = P^{X_{2,j}}$ are stochastically comparable. For the corresponding distribution functions F_1, F_2 we use the notation

$$\mathcal{F}_{\leq[\geq]} := \{(F_1, F_2) : F_1 \leq \cdot [\cdot \geq] F_2\}, \; \mathcal{F}_= := \{(F_1, F_2) : F_1 = F_2\}$$

and we consider the hypotheses

$$H : \mathcal{F}_= + \mathcal{F}_\geq, \quad K : \mathcal{F}_\leq.$$

For constructing the critical region of Pitman's two-sample permutation test one computes, for all permutations

$$\pi(x) = (\pi_1(x), \ldots, \pi_{n_1+n_2}(x)) \text{ of } x = (x_{1,1}, \ldots, x_{1,n_1}, x_{2,1}, \ldots, x_{2,n_2}),$$

the values $S(\pi) := \sum_{i=1}^{n_1} \pi_i(x)$ and compares these with $S_0 := \sum_{j=1}^{n_1} x_{1,j}$. For nearly all theoretical investigations of this problem it is additionally assumed that the F_i are continuous (then the $X_{i,j}$ as well as the $S(\pi)$ are a.s. all different); this will be marked by the notion H^c and K^c resp.. In this case the permutation test is unbiased for testing the hypotheses H^c and K^c and it maximizes the power for the class $\hat{K}$ of all normal distributions with shifted expectations and equal variances (see e.g. Witting/Nölle (1970), Ch. 3.5, or Lehmann (1986), Ch. 5.10/5.11). Our first aim is to generalize this result into two directions:

Although the assumption of (approximate) normal distributions may be justified, in many cases, by the central limit theorem, for "skew" distributions as e.g. in survival analysis other assumptions seem to be much more reasonable (e.g. exponential distributions). Hence the question arises whether the optimum property of the permutation tests can be extended to other interesting classes of distributions.

The continuity assumption has the advantage (for theoretical considerations) that all values of the $X_{i,j}$ and even of the $S(\pi)$ are a.s. different. But in most practical problems one observes, at least for larger sample sizes, ties for the $S(\pi)$ – one reason being the restricted accuracy of measurement. Hence the questions arise how to perform the permutation test in this case and whether one can prove optimum properties for this case, too (surprisingly enough, one can make use of this more complicated structure to construct efficient algorithms for computing the critical region; this will be carried out in part II).

Both problems are treated in section 2 in a unified manner. Applications to continuous distributions as well as to discrete distributions with fixed support are given in section 4.

2. Optimum properties of permutation tests

Since there does not exist, for the nonparametric hypotheses H and K, any uniformly most powerful level α-test one looks for tests which are unbiased for H against K and uniformly optimal on a suitable subclass $K_1 \subset K$.

Theorem 2.1

Let $T(x) = x_{[\,]} = t$ be the order statistic and

$$\hat{K}_\mu = \left\{ \vartheta = (F_1, F_2) \in K : \begin{array}{l} F_i \text{ dominated by } \mu; \text{ the } \mu^n\text{-density} \\ \text{is for fixed } t \text{ and } x \in T^{-1}(\{t\}) \\ \text{strictly increasing in } \sum G(x_{1,j}) \end{array} \right\},$$

where μ is a σ-finite measure on $(\mathbb{R}, \mathbb{B})$ and G is strictly increasing; let $\alpha \in (0,1)$ and

$$\phi_t^\star(t) := \begin{cases} 1 & \\ \gamma(t) \text{ for } S(x) := \sum G(x_{1,j}) & \begin{array}{c} > \\ = \\ < \end{array} \; c(t), \\ 0 & \end{cases}$$

where $c(t)$ and $\gamma(t)$ are determined such that

$$\begin{aligned} \alpha = {} & \frac{1}{n!} |\{\pi : \textstyle\sum_{i=1}^{n_1} G(\pi_i(x)) > c(t)\}| + \\ & \frac{\gamma(t)}{n!} |\{\pi : \textstyle\sum_{i=1}^{n_1} G(\pi_i(x)) = c(t)\}|. \end{aligned}$$

Then $\phi^\star(x) := \phi^\star_{T(x)}(x)$ is an unbiased level α-test for H against K and has, under the unbiasedness condition, uniformly maximal power on $\hat{K}_\mu$ (and hence on each subclass of $\hat{K}_\mu$).

Proof: We are looking for a solution of

$$\phi^\star \in \Phi_\alpha := \left\{ \phi : \mathbb{R}^n \to [0,1] : \begin{array}{ll} E_\vartheta \phi \le \alpha & \forall\, \vartheta \in H \\ & \\ E_\vartheta \phi \ge \alpha & \forall\, \vartheta \in K \end{array} \right\}$$

$$(1)\qquad E_\vartheta \phi^\star = \sup_{\phi \in \Phi_\alpha} E_\vartheta \phi \qquad \forall \vartheta \in \hat{K}_\mu .$$

Under the metric of uniform convergence one obtains that $\overline{H} \cap \overline{K} = \mathcal{F}_=$ and that the power function of each test is continuous on $\mathcal{F}_=$. Hence every unbiased level α-test is similar on $\mathcal{F}_=$. Therefore, we restrict attention on a solution of

$$(2)\qquad \begin{aligned} &\phi^\star \in \Phi'_\alpha := \{\phi : \mathrm{I\!R}^n \to [0,1] : E_\vartheta \phi = \alpha \quad \forall\, \vartheta \in \mathcal{F}_=\} \\ &E_\vartheta \phi^\star = \sup_{\phi \in \Phi'_\alpha} E_\vartheta\, \phi \qquad \forall\, \vartheta \in \hat{K}_\mu . \end{aligned}$$

The order statistic T is sufficient for $\mathcal{F}_=$ (see Witting (1985), example 3.7a) as well as complete - this follows from theorem 3.43 of Witting (1985) by replacing $\mathcal{F}^c$ by $\mathcal{F}$ (according to Witting (1985), theorem 3.42, T is complete for the class $\mathcal{P}_\mu$ of all product measures with the same ν-continuous marginal distributions; since this is true for all distributions ν the completeness of T follows).

Hence the solution of (2) is equivalent to the solution of

$$(3)\qquad \begin{aligned} &\phi^\star \in \Phi_{NS} := \{\phi : \mathrm{I\!R}^n \to [0,1] : \textstyle\int \phi \, dP_{\mathcal{F}_=}^{X|T\circ X = t} = \alpha \quad \mathcal{P}_{\mathcal{F}_=}^T \text{ a.s.}\} \\ &\textstyle\int \phi^\star \, dP_\vartheta^{X|T\circ X = t} \geq \int \phi \, dP_\vartheta^{X|T\circ X=t} \forall \phi \in \phi_N,\ \forall\, \vartheta \in \hat{K}_\mu,\ P_\vartheta^T\text{-a.s.} \end{aligned}$$

i.e. computing a uniformly optimal test with Neyman structure with respect to T. This can be done separately on the sets $T^{-1}(\{t\})$, i.e. by solving, for each t, the problem

$$(4)\qquad \begin{aligned} &\phi_t^\star \in \Phi_t := \{\phi_t : T^{-1}(\{t\}) \to [0,1] : \textstyle\int \phi_t \, dP_{\mathcal{F}_=}^{X|T\circ X=t} = \alpha\} \\ &\textstyle\int \phi_t^\star \, dP_\vartheta^{X|T\circ X=t} = \sup_{\phi_t \in \Phi_t} \int \phi_t \, dP_\vartheta^{X|T\circ X=t} \quad \forall \vartheta \in \hat{K}_\mu . \end{aligned}$$

For this aim we need the conditional distributions $P_\vartheta^{X|T\circ X=t}$, i.e. for each $B \in \mathrm{I\!B}^n$ a solution of

$$(5)\qquad \int_A P_\vartheta^{X|T\circ X=t}(B)\, dP_\vartheta^{T\circ X}(t) = P_\vartheta^X(B \cap T^{-1}(A)) \qquad \forall A \in T(\mathrm{I\!B}^n).$$

Let now Π be the set of all permutations on $\mathrm{I\!R}^n$, $\Pi_x := \{\pi \in \Pi : \pi(x) = x\}$ the set of all permutations which let x fixed, Ψ_x an arbitrary system of representants of Π/Π_x, and $f_\vartheta, f_\vartheta^T$ μ^n-densities of $P_\vartheta, P_\vartheta^T$ resp.
(i) Consider A such that $\mu^n(\{x\}) > 0\ \forall x \in A$. Since μ is σ-finite, A is countable and one obtains for the right hand side of (5)

$$\begin{aligned} &P_\vartheta^X(B \cap T^{-1}(A)) \\ &\quad = \int_{T^{-1}(A)} 1_B(x) f_\vartheta(x) d\mu^n(x) \end{aligned}$$

$$= \sum_{t \in A} \int_{T^{-1}(\{t\})} 1_B(x) f_\vartheta(x) d\mu^n(x)$$

$$= \sum_{t \in A} \sum_{\pi \in \Psi_t} \int_{\pi(t)} 1_B(x) f_\vartheta(x) d\mu^n(x)$$

$$= \sum_{t \in A} \sum_{\pi \in \Psi_t} 1_B(\pi(t)) f_\vartheta(\pi(t)) \mu^n(\{t\})$$

whereas the left hand side is

$$\int_A P_\vartheta^{X|T\circ X=t}(B) dP_\vartheta^{T\circ X}(t) = \sum_{t \in A} P_\vartheta^{X|T\circ X=t}(B) f_\vartheta^T(t) \mu^n(\{t\}).$$

Hence we may choose for $f_\vartheta^T(t) \neq 0$

$$(6) \qquad P_\vartheta^{X|T\circ X=t}(B) = \frac{\sum_{\pi \in \Psi_t} 1_B(\pi(t)) f_\vartheta(\pi(t))}{f_\vartheta^T(t)}$$

or, taking into account $f_\vartheta^T(t) = \sum_{\pi \in \Psi_t} f_\vartheta(\pi(t))$ and the multiplicities,

$$P_\vartheta^{X|T\circ X=t}(B) = \frac{\sum_{\pi \in \Pi} 1_B(\pi(t)) f_\vartheta(\pi(t))}{\sum_{\pi \in \Pi} f_\vartheta(\pi(t))}.$$

(ii) Next we assume that $\mu^n(\{x\}) = 0 \; \forall x \in A$. Then the set $T^{-1}(\{t\})$ consists of the (almost surely different) points $\pi(t), \pi \in \Pi$, and one obtains

$$\int_{T^{-1}(A)} 1_B(x) f_\vartheta(x) d\mu^n(x) = \sum_{\pi \in \Pi} \int_{\pi(A)} 1_B(x) f_\vartheta(x) d\mu^n(x)$$

$$= \sum_{\pi \in \Pi} \int_A 1_B(\pi(t)) f_\vartheta(\pi(t)) d\mu^n(t) = \int_A \sum_{\pi \in \Pi} 1_B(\pi(t)) f_\vartheta(\pi(t)) d\mu^n(t)$$

and

$$\int_A P_\vartheta^{X|T\circ X=t}(B) dP_\vartheta^{T\circ X}(t) = \int_A P_\vartheta^{X|T\circ X=t}(B) f_\vartheta^T(t) d\mu^n(t).$$

This leads, for $f_\vartheta^T(t) > 0$, to

$$P_\vartheta^{X|T\circ X=t}(B) = \frac{\sum_{\pi \in \Pi} 1_B(\pi(t)) f_\vartheta(\pi(t))}{f_\vartheta^T(t)},$$

which turns out to be the special version of (6) for the case $\Pi_t = \{id\}$.
(iii) In the general case we split A up into $A_d := \{t \in A : \mu^n(\{t\})) > 0\}$ and $A_c := A - A_d$. Then (6) yields a solution of

$$\int_{A_c} P_\vartheta^{X|T\circ X=t}(B) dP_\vartheta^{T\circ X}(t) + \int_{A_d} P_\vartheta^{X|T\circ X=t}(B) dP_\vartheta^{T\circ X}(t)$$

$$= P_\vartheta^X(B \cap T^{-1}(A_c)) + P_\vartheta^X(B \cap T^{-1}(A_d)).$$

In particular we obtain

$$P_\vartheta^{X|T\circ X=t}(\{x\}) = \frac{f_\vartheta(x)}{\sum_{\pi\in\Psi_t} f_\vartheta(\pi(t))}.$$

For $\vartheta \in \mathcal{F}_=$ it follows

$$P_\vartheta^{X|T\circ X=t}(\{x\}) = \frac{1}{|\Psi_t|} = \frac{|\Pi_t|}{n!},$$

which may also be taken for the points t with $f_\vartheta^T(t) = 0$.

To apply the Neyman-Pearson-lemma we have to consider the statistic $f_\vartheta(x)/f_\vartheta^T(t)$. Due to our assumptions this statistic is, for fixed t and $x \in T^{-1}(t)$, an increasing function of $\sum G(x_{1,j})$ which is independent of ϑ. Hence we may choose as an optimal test

$$\phi_t^\star(x) = \begin{cases} 1 & > \\ \gamma(t) : \sum G(x_{1,j}) & = \quad c(t), \\ 0 & < \end{cases}$$

where $c(t)$ and $\gamma(t)$ are determined according to

$$\begin{aligned}
\alpha &= P_{\mathcal{F}_=}^{X|T\circ X=t}(\textstyle\sum G(x_{1,j}) > c(t)) + \gamma(t) P_{\mathcal{F}_=}^{X|T\circ X=t}(\textstyle\sum G(x_{1,j}) = c(t)) \\
&= \frac{|\Pi_t|}{n!} |\{x \in \mathbb{R}^n : T(x) = t; \textstyle\sum G(x_{1,j}) > c(t)\}| \\
&\qquad + \gamma(t)\frac{|\Pi_t|}{n!} |\{x \in \mathbb{R}^n : T(x) = t; \textstyle\sum G(x_{1,j}) = c(t)\}| \\
&= \frac{1}{n!} |\{\pi \in \Pi : \textstyle\sum_{i=1}^{n_1} G(\pi_i(t)) > c(t)\}| \\
&\qquad + \frac{\gamma(t)}{n!} |\{\pi \in \Pi : \textstyle\sum_{i=1}^{n_1} G(\pi_i(t)) = c(t)\}|
\end{aligned}$$

To solve the initial problem we define $\phi^\star(x) := \phi_{T(x)}^\star(x)$. Then $\phi^\star$ is measurable; see Witting, 1985, Hilfssatz 3.61, and from $P_\vartheta(N) = 0 \ \forall \vartheta \in \mathcal{F}_=$ we obtain, for all probability measures $P_1, P_2, (P_1 + P_2, P_1 + P_2)(N) = 0$ and, therefore, $(P_1, P_2)(N) = 0$, i.e. each $P_{\mathcal{F}_=}^T$-nullset is a P_ϑ^T-nullset for $\vartheta \in \hat{K}_\mu$. Hence $\phi^\star$ is a solution of (3) and (2).

To show that $\phi^\star$ is unbiased for H against K we consider $x, \tilde{x}$ such that $\tilde{x}_{1,i} = x_{1,i} + \delta_i, \delta_i \geq 0, 1 \leq i \leq n_1, \tilde{x}_{2,j} = x_{2,j}, 1 \leq j \leq n_2$. Since the p-value of the permutation test turns out to be

$$p(x) = \frac{1}{n!} |\{\pi \in \Pi : \textstyle\sum_{i=1}^{n_1} G(\pi_i(x)) \geq \sum_{i=1}^{n_1} G(x_{1,i})\}|$$

one obtains $p(\tilde{x}) \leq p(x)$ and therefore, $\phi^\star(\tilde{x}) \geq \phi^\star(x)$. According to Witting & Nölle (1970), Satz 3.13, now the unbiasedness of $\phi^\star$ follows.

This result can, moreover, slightly be sharpened:

Corollary 2.1

Let $H_\mu := \{\vartheta \in H : P_\vartheta^{X_{i,1}} \ll \mu,\ i = 1,2\}$ *and*
$K_\mu := \{\vartheta \in K : P_\vartheta^{X_{i,1}} \ll \mu,\ i = 1,2\}$.
Then $\phi^\star$ *is an unbiased level* α*-test for* H *against* K*, and has, under all unbiased level* α*-tests for* H_μ *against* K_μ*, uniformly maximal power on each* $K_1 \subset \hat{K}_\mu$.

Proof: Let $J_\mu := \overline{H}_\mu \cap \overline{K}_\mu$. Then T is also sufficient and complete for $\mathcal{P}_{J_\mu}$ (see Witting (1985), Satz 3.42). Hence the result follows by the same arguments as before.

Applying the permutation test one will make use of the fact that all permutations $\pi \in \Pi$ which merely exchange values within the first n_1 or within the second n_2 observations lead to the same value of $\sum_{i=1}^{n_1} G(\pi_i(x))$. Using the notation

$$\Pi_{n_1,n_2} := \Pi/(\Pi_{n_1}, \Pi_{n_2})$$

one therefore obtains a simpler expression to determine $c(t)$ and $\gamma(t)$:

$$\begin{aligned} \alpha &= \frac{1}{\binom{n}{n_1}} |\{\pi \in \Pi_{n_1,n_2} : \textstyle\sum_{i=1}^{n_1} G(\pi_i(x)) > c(t)\}| \\ &+ \frac{\gamma(t)}{\binom{n}{n_1}} |\{\pi \in \Pi_{n_1,n_2} : \textstyle\sum_{i=1}^{n_1} G(\pi_i(x)) = c(t)\}|. \end{aligned}$$

3. General criteria

To apply theorem 2.1 to special classes of distributions one has to ensure that these distributions belong to $\hat{K}_\mu$, i.e. that they are stochastically comparable and that they fulfill the monotonicity assumption. For this purpose we mention:

Lemma 3.1

Let $\mathcal{P}_\Theta = \{P_\vartheta : \vartheta \in \Theta\}$ *be a class of distributions* P_ϑ *with densities* $f_\vartheta = dP_\vartheta/d\mu$ *with respect to a* σ*-finite measure* μ *and assume that* $\mathcal{P}_\Theta$ *has increasing likelihood-ratio in id. Then the elements of* $\mathcal{P}_\Theta$ *are stochastically comparable.*

For a proof see Witting (1985), Satz 2.28.

Lemma 3.2

Let $\mathcal{P}_\Theta$ be a one-parameter exponential family in $\zeta(\vartheta)$ and $G(x)$ dominated by the σ-finite measure μ, i.e. the P_ϑ have μ-densities

$$f_\vartheta(x) = C(\vartheta)\, e^{\zeta(\vartheta)G(x)} h(x),$$

and assume that ζ and G are strictly increasing. Then the elements of $\mathcal{P}_\Theta$ are stochastically comparable and for $\vartheta_1 > \vartheta_2$ holds

$$(F_{\vartheta_1}, F_{\vartheta_2}) \in \hat{K}_\mu$$

(i.e. theorem 2.1 can be applied).

Proof: Since $\mathcal{P}_\Theta$ has monotone likelihood-ratio in G and since G is strictly increasing, $\mathcal{P}_\Theta$ has monotone likelihood-ratio in *id*; Lemma 3.1 yields the stochastic comparability.

$$f(x) := \prod_{i=1}^{n_1} f_{\vartheta_1}(x_{1,i}) \prod_{i=1}^{n_2} f_{\vartheta_2}(x_{2,i})$$

is the density of the common distribution; hence we obtain for each permutation π of the $n_1 + n_2$ components

$$f(\pi(x)) = [C(\vartheta_1)]^{n_1}[C(\vartheta_2)]^{n_2} \prod_{i=1}^{n_1} h(\pi_i(x)) \prod_{j=n_1+1}^{n} h(\pi_j(x)) \times$$
$$\times \exp(\zeta(\vartheta_1)\sum_{i=1}^{n_1} G(\pi_i(x)) + \zeta(\vartheta_2) \sum_{j=n_1+1}^{n} G(\pi_j(x))).$$

For fixed x the first four factors are constant; for the remaining term we obtain

$$\exp(\zeta(\vartheta_2)(\sum_{i=1}^{n_1} G(\pi_i(x)) + \sum_{j=n_1+1}^{n} G(\pi_j(x)))) \times$$
$$\times \exp((\zeta(\vartheta_1) - \zeta(\vartheta_2))\sum_{i=1}^{n_1} G(\pi_i(x))).$$

Here again the first factor is, for $\pi \in \Pi$ and fixed x, constant whereas the second is, since ζ is strictly increasing, strictly increasing in $\sum_{i=1}^{n_1} G(\pi_i(x))$. This yields the assertion.

On the other hand, one should admit that Lemma 3.2 essentially describes the scope where Theorem 2.1 can be applied (comp. sections 1.7 and 4.5 of Pfanzagl (1994)). Hence Theorem 2.1 mainly unifies and generalizes the optimum properties of permutation tests for *parametric* subclasses $K_1 \subset K$ but is far from yielding convincing optimality results for *non-parametric* subclasses. This is underlined also by the applications given in the next session.

4. Examples

4.1 Continuous distributions

Since the family of normal distributions $\mathcal{N}(\vartheta,\sigma^2)$ with fixed $\sigma^2>0$ forms a one-parameter exponential family in $\zeta(\vartheta)=\vartheta/\sigma^2$ and $G=id$ we get back the "classical" optimum property:

Example 4.1 (Normal distributions $\mathcal{N}(\vartheta,\sigma^2)$)

$$K_1 := \{(\mathcal{N}(\vartheta_1,\sigma^2),\mathcal{N}(\vartheta_2,\sigma^2):\sigma^2>0,\vartheta_1>\vartheta_2\}\subset\hat{K}^c.$$

As further optimum properties of the permutation tests we obtain

Examples 4.2 (Gamma-distributions $\Gamma_{\kappa,\lambda}$)

$$\begin{aligned} K_2 &:= \{(\Gamma_{\kappa,1/\vartheta_1},\Gamma_{\kappa,1/\vartheta_2}):\kappa>0,\ \vartheta_1>\vartheta_2>0\}\subset\hat{K}^c,\\ K_3 &:= \{(\Gamma_{\vartheta_1,\lambda},\Gamma_{\vartheta_2,\lambda}):\lambda>0,\vartheta_1>\vartheta_2>0\}\subset\hat{K}^c.\end{aligned}$$

Proof: The Lebesgue-density $f_{\kappa,\lambda}$ of a Gamma-distribution $\Gamma_{\kappa,\lambda}$ has the form

$$f_{\kappa,\lambda}(x)=\frac{\lambda^\kappa}{\Gamma(\kappa)}x^{\kappa-1}e^{-\lambda x}1_{[0,\infty)}(x).$$

Hence the Gamma-distributions form

(i) for fixed $\kappa>0$ an exponential family in $\zeta(\vartheta)=-1/\vartheta$ and $G=id$

(ii) for fixed $\lambda>0$ an exponential family in $\zeta=id$ and $G=\ln$.

As special cases we mention

Example 4.3 (Exponential distributions $\mathrm{Exp}(\lambda)=\Gamma_{1,\lambda}$)

$$K_4=\{(Exp(\lambda_1),Exp(\lambda_2)):\lambda_2>\lambda_1>0\}\subset\hat{K}^c$$

Example 4.4 (χ^2-distributions $\chi^2_{n,\sigma}=\Gamma_{n/2,1/2\sigma^2}$)

$$\begin{aligned} K_5 &:= \{(\chi^2_{n,\sigma_1},\chi^2_{n,\sigma_2}):n\in\mathbb{N},\sigma_1>\sigma_2>0\}\subset\hat{K}^c,\\ K_6 &:= \{(\chi^2_{n_1,\sigma},\chi^2_{n_2,\sigma}):\sigma>0,n_1,n_2\in\mathbb{N},n_1>n_2\}\subset\hat{K}^c.\end{aligned}$$

Example 4.5 (Weibull distributions $\mathcal{W}_{\vartheta,\lambda}$)

$$K_7 := \{(\mathcal{W}_{\vartheta_1,\lambda},\mathcal{W}_{\vartheta_2,\lambda}):\lambda>0,\vartheta_1>\vartheta_2>0\}\subset\hat{K}^c.$$

Proof: The Lebesgue-density of $\mathcal{W}_{\vartheta,\lambda}$ is

$$f_{\vartheta,\lambda}(x) = \frac{\lambda x^{\lambda-1}}{\vartheta^{\lambda}} \exp(-(x/\vartheta)^{\lambda}) 1_{(0,\infty)}(x);$$

hence the Weibull-distributions form, for fixed λ, an exponential family in $\zeta(\vartheta) = -\frac{1}{\vartheta^{\lambda}}$ and $G = id^{\lambda}$. Since the statistic, used for the resulting permutation test, depends on λ the test can be used only for known λ.

4.2 Discrete distributions

Example 4.6 (Poisson distributions $\mathcal{P}(\vartheta)$)

$$K_8 := \{(\mathcal{P}(\vartheta_1), \mathcal{P}(\vartheta_2)) : \vartheta_1 > \vartheta_2 > 0\} \subset \hat{K}^d.$$

Proof: The Poisson distributions $\mathcal{P}(\lambda)$ form an exponential family in $\zeta = \ln$ and $G = id$.

Example 4.7 (Binomial distributions $\mathcal{B}(n, \vartheta)$)

$$K_9 := \{(\mathcal{B}(n, \vartheta_1), \mathcal{B}(n, \vartheta_2)) : n \in \mathbb{N},\ 1 > \vartheta_1 > \vartheta_2 > 0\} \subset \hat{K}^d.$$

Proof: The binomial distributions $\mathcal{B}(n,p)$ form, for fixed n, an exponential family in $\zeta(\vartheta) = \ln(\vartheta/(1-\vartheta))$, which is strictly increasing, and $G = id$.

Example 4.8 (Negative binomial distributions $\mathcal{N}b(n, \vartheta)$)

$$K_{10} := \{(\mathcal{N}b(n, \vartheta_1), \mathcal{N}b(n, \vartheta_2)) : n \in \mathbb{N}, 0 < \vartheta_1 < \vartheta_2 < 1\} \subset \hat{K}^d.$$

Proof: The negative binomial distributions form, for fixed n, an exponential family in $\zeta(\vartheta) = \ln(1-\vartheta)$ and $G = id$.

In particular, the special case of geometric distributions is covered by Example 4.8.

The results of Examples 4.6-4.8 have been obtained already by Schrage (1980).

4.3 Disturbed exponential families

Let $\mathcal{P}_{\Theta}$ be an exponential family as in Lemma 3.2 and g a non-negative function such that $f_{\vartheta}(x)g(x)$ is integrable. Then obviously there exists a norming function $\tilde{C} : \Theta \to \mathbb{R}^+$ such that the distributions $\tilde{P}_{\vartheta}$ defined by the densities $\tilde{f}_{\vartheta}(x) = \tilde{C}(\vartheta) f_{\vartheta}(x) g(x)$ form an exponential family which fulfills the conditions of Lemma 3.2.

This fact has interesting consequences:

(i) In many practical problems it is, due to the structure of the observed phenomenon, clear that the data are non-negative (e.g. for weight, age, energy) or belong to a certain interval (e.g. the values of concentrations are always between 0 % and 100 %). If nevertheless the assumption of normal distributions is taken into consideration it seems to be obvious to multiply the original densities by $1_{(0,\infty)}$ or $1_{(0,100)}$ (and to renorm). The resulting permutation test then has the same form and the same optimum properties as for the original situation.

(ii) If the data are influenced by parameter-dependent effects as well as by parameter-independent ones, with the consequence that the densities are of the form

$$f_\vartheta = h_\vartheta \cdot g,$$

then the f_ϑ fulfill the conditions of Lemma 3.2 if the h_ϑ fulfill it. This means that the permutation test is "robust" against parameter-independent disturbations.

The authors thank an anonymous referee for his careful reading and his detailed comments.

References

Good P (1995) Permutation Tests, 2nd edn. Springer, New York

Lehmann EL (1986) Testing Statistical Hypotheses, 2nd edn. Wiley, New York

Odén A, Wedel H (1975) Arguments for Fisher's permutation test. Ann. Statist. 3: 518-520

Pfanzagl J (1994) Parametric Statistical Theory, deGruyter, Berlin

Schrage C (1980) Two-sample permutation test for discrete distributions. Arbeitspapier CS 01, Inst. Math. Statistik, WWU Münster

Schrage C (1982) k-Stichprobenpermutationstests bei diskreter Verteilungsannahme. PhD thesis, Westfälische Wilhelms-Universität Münster

Witting H (1985) Mathematische Statistik I, Teubner, Stuttgart

Witting H, Nölle G (1970) Angewandte Mathematische Statistik, Teubner, Stuttgart

Jens Gebhard
Norbert Schmitz
Institut für Mathematische Statistik
Universität Münster
Einsteinstraße 62
D-48149 Münster, Germany