



ELSEVIER

Mathematical Social Sciences 38 (1999) 247–258

mathematical
social
sciences

A theorem on parallel processing models with a generalized stopping rule

Hans Colonius*

Institut für Kognitionsforschung, Universität Oldenburg, FB 5-A6, D-26111 Oldenburg, Germany

Accepted 1 February 1999

Abstract

Inequalities on reaction time distribution functions for parallel processing models with an unlimited capacity assumption are presented, extending previous work on first-terminating and exhaustive stopping rules to k th-terminating processes. This extension thus generates predictions for situations in which the observer's response is determined by the k th-terminating subprocess. Moreover, methods to determine the number k are discussed. © 1999 Published by Elsevier Science B.V. All rights reserved.

Keywords: Parallel process; Stopping rule; Order statistics

1. Introduction

Many models of information processing in a perceptual/cognitive task commonly assume that reaction time (RT) can be decomposed into several subprocesses comprising, among others, stimulus encoding time, response selection time, and motor execution time. In addition to processing architecture, a problem of considerable importance has been the stopping rule the information processing system employs when sufficient information has been acquired to make a correct response. Many of these results have been reviewed in (Townsend and Ashby, 1983; Townsend, 1990; Luce, 1986). The focus of the present paper are stopping rules for parallel models extending results of Colonius and Ellermeier (1997) and Colonius and Vorberg (1994). For recent results regarding stopping rules for serial models, see Townsend and Colonius (1997).

In the following, a *parallel* model is informally defined as a parallel system with n processing channels represented by a collection of nonnegative random variables

*Tel./fax: +49-411-798-5158.

E-mail address: colonius@psychologie.uni-oldenburg.de (H. Colonius)

T_1, T_2, \dots, T_n which refer to the processing durations of the channels. The term *channel* is used here in a broad sense. For example, in a redundant-signals task, where human observers monitor two or more sources of information for a target stimulus, the T_k refer to the signal-specific processing times. Channels may interact with each other; thus, processing times T_1, T_2, \dots, T_n , are allowed to be stochastically dependent. In accordance with the terminology proposed by Colonius and Vorberg (1994), a parallel system is called *first-terminating* if it initiates a response as soon as the first channel finishes processing. Thus, the contribution of the channels' processing time to overall reaction time is the time needed by the fastest channel to finish processing, i.e., the minimum of the T_k . Alternatively, a parallel system is called *exhaustive* if it cannot initiate a response unless all its channels have finished. In this case, overall reaction time is given by the time of the slowest channel, i.e., the maximum of the T_k . Here, the contributions of additional components like stimulus encoding time, response selection time, or motor execution time are all subsumed under this same random variable, for simplicity. Often, these are summarized by adding another random variable (sometimes, a constant) to the maximum/minimum random variable.

The redundant target paradigm mentioned above is a typical example of a situation where a first-terminating stopping rule has been suggested. The subject is instructed to respond to the first target detected which, in a bimodal experiment with both visual and auditory targets, could either be the visual or the auditory signal (cf. Raab, 1962; Miller, 1982; Ulrich and Giray, 1986; Diederich and Colonius, 1987). While parallel models with a first-terminating or an exhaustive stopping rule are plausible mechanisms in many situations, note that these rules only constitute two extreme cases of a more general situation where overall reaction time is determined by k channels having finished ($k = 1, \dots, n$). Depending on the specific conditions of the experiment or on instructions given to subjects, the general *kth-terminating* rule (with k different from 1 or n) may yield a more realistic description of the data.

Take, for example, the same-different paradigm (e.g., Bamber, 1969). On any trial, two letter strings are presented to the subject, e.g.

b f w s

b f r s

The subject's task is to decide as fast as possible whether the strings are the same or different. A simple model for this situation would assume channels corresponding to the columns working in parallel and trying to match the letters. Consider the condition where both strings consist of entirely different letters. A "different" response could be initiated as soon as one of the channels has registered a mismatch. However, if accuracy rather than speed is emphasized, it is not implausible to assume that a "different" response is initiated only after k ($k > 1$) of the channels indicate a mismatch.¹

In the next section, after introducing some necessary terminology and definitions,

¹No claim is made here that this type of model gives the most plausible account of all empirical results known so far. In fact, a number of different models have been discussed in the literature (see, e.g., (Townsend and Ashby, 1983), p. 153).

distribution inequalities for parallel models are derived involving different numbers of channels in generating the response time distribution. In particular, a theorem for general k th-terminating models extends a result for the second-terminating ($k = 2$) case in Colonius and Ellermeier (1997). The inequalities are illustrated by a numerical example with dependent processing times. Then, means to determine the number k of a system from the data are discussed, while proofs are deferred to the final section.

2. Distribution inequalities: k th-terminating models

Let T_1, T_2, \dots, T_n be jointly distributed random variables; the corresponding *order statistics* are the T_i 's arranged in nondecreasing order

$$T_{1:n} \leq T_{2:n} \leq \dots \leq T_{n:n}$$

Specifically, $T_{k:n}$ is called the k th *order statistic*. When actual equalities apply, one does not make any requirement about which variable should precede the other one.

This definition does not require that the T_i 's be identically distributed, or that they are independent, or that densities exist. Many classical results dealing with order statistics were originally derived in the i.i.d. case (*independent identically distributed*) with a common continuous (cumulative) distribution (see Arnold et al., 1992). In the reaction-time modelling context, existence of a multivariate density for the T_1, T_2, \dots, T_n can usually be taken for granted while neither independence nor identical distributions seem particularly plausible.

Definition 1. Let T_1, T_2, \dots, T_n denote the random channel processing times of a parallel reaction time model. The model is called *parallel k th-terminating* if its reaction time (RT) is equal to the k th order statistic of T_1, T_2, \dots, T_n , i.e.

$$\text{RT} = T_{k:n}.$$

The number k is also called the *order of the stopping rule*. In particular, the model is called *parallel first-terminating* if

$$\text{RT} = \min_{i=1, \dots, n} T_i = T_{1:n}$$

It is called *parallel exhaustive* if

$$\text{RT} = \max_{i=1, \dots, n} T_i = T_{n:n}$$

Throughout this paper, parallel models are assumed to have *unlimited processing capacity*, that is, the system allots the same amount of “capacity” to a given channel no matter how many additional channels operate at the same time. In many empirical situations, unlimited capacity can be expected to hold at most for a small number of parallel channels. The notion of unlimited capacity can be made precise by considering the joint probability distribution functions of the times T_i over experimental conditions with different sets of active channels. Suppose we have a subset of m out of the n channels of the system. Unlimited capacity stipulates that their (marginal) distributions

are the same no matter how many and which of the remaining $n - m$ channels are active. The definition below is slightly less restrictive requiring the marginal distributions to be identical only on the m -dimensional “diagonal” (t, \dots, t) rather than the full set (t_1, t_2, \dots, t_m) .

Definition 2. Let $B = \{i_1, i_2, \dots, i_m\}, m \leq n$ be the set of channels active under a given experimental condition and let P_B denote the corresponding probability measure implied by the m -channel model; an n -channel parallel model is said to have *unlimited capacity (on the diagonal)* if

$$P_B(T_{i_1} \leq t, T_{i_2} \leq t, \dots, T_{i_m} \leq t) = P(T_{i_1} \leq t, T_{i_2} \leq t, \dots, T_{i_m} \leq t)$$

for all $B \subset \{1, 2, \dots, n\}$ and all $t \in \mathfrak{R}^+$,

where P denotes the probability measure corresponding to the model where all n channels are active². Note that this definition of unlimited capacity neither implies nor is implied by stochastic independence of the processing times.

Consistent with the order statistics notation, distribution functions for the k th-terminating parallel model with n channels will be denoted by $F_{k:n}(t)$. In particular, for the first-terminating and exhaustive models with n channels,

$$F_{1:n}(t) = P(T_{1:n} \leq t) \quad (\text{first-terminating } n\text{-channel process})$$

and

$$F_{n:n}(t) = P(T_{n:n} \leq t) \quad (\text{exhaustive } n\text{-channel process}).$$

Moreover, we write $F_{k:n-1}^{(i)}(t), F_{k:n-2}^{(ij)}(t)$, etc. for the distribution function if all channels but channel i , all channels but channels i, j , etc. are active, respectively. Thus, using the notation of Definition 2 with $C = \{1, 2, \dots, n\}$ we have, for example,

$$F_{n-1:n-1}^{(i)}(t) = P_{C \setminus \{i\}}(T_1 \leq t, T_2 \leq t, \dots, T_{i-1} \leq t, T_{i+1} \leq t, \dots, T_n \leq t).$$

By the unlimited capacity assumption, this equals

$$P(T_1 \leq t, T_2 \leq t, \dots, T_{i-1} \leq t, T_{i+1} \leq t, \dots, T_n \leq t).$$

The following is the main result of this paper. It provides an upper and a lower bound for the distribution function of a parallel k th-terminating parallel model in a very general setting. Its proof is deferred to the last section.

Theorem. *For an unlimited capacity, k th-terminating n -channel parallel model ($1 \leq k \leq n - 2, n > 2$), the following holds with integer s_1, \dots, s_{k+1} and $1 \leq s_1 < \dots < s_k < s_{k+1} \leq n$*

²For simplicity, the specification “on the diagonal” will not be mentioned explicitly below.

$$\begin{aligned} \max_i F_{k:n-1}^{(i)}(t) &\leq F_{k:n}(t) \\ &\leq \min_{\mathbf{s}} \{ F_{k:n-1}^{(s_1)}(t) + \dots + F_{k:n-1}^{(s_{k+1})}(t) - F_{k:n-2}^{(s_1, s_2)}(t) - F_{k:n-2}^{(s_2, s_3)}(t) - \dots \\ &\quad F_{k:n-2}^{(s_k, s_{k+1})} \}, \end{aligned} \tag{1}$$

for any t , where the maximum is taken over all $i = 1, \dots, n - 1$ and the minimum is taken over all vectors $\mathbf{s} = (s_1, \dots, s_{k+1})$.

As mentioned above, this theorem generalizes Theorem 1 in Colonius and Vorberg (1994) and Theorem 2 in Colonius and Ellermeier (1997) where the result was given for $k = 1$ and $k = 2$, respectively.

Definition 3. If all marginal distributions $F_{k:n-1}^{(i)}, F_{k:n-2}^{(ij)}$, are independent of the choice of i and j , the parallel system is called *marginally invariant*.

As an immediate consequence of the theorem, for marginally invariant parallel systems, the inequalities take a specially simple form³

Corollary. For a marginally invariant parallel system, the inequalities of the theorem become

$$F_{k:n-1}(t) \leq F_{k:n}(t) \leq (k + 1)F_{k:n-1}(t) - kF_{k:n-2}(t) \tag{2}$$

In general, not much is known about the sharpness of the upper and the lower bounds in the above inequalities. Obviously, for t going to infinity, the inequalities become trivial. Thus, only for values of t not too large can these inequalities be expected to be useful for testing the assumption of a k th-terminating parallel stopping rule.

The above result is illustrated by a numerical example⁴ with a specified multivariate distribution function with dependent random processing times (cf. Colonius and Vorberg, 1994).

3. Example: a k th-terminating parallel model with conditionally independent exponential channel processing times

On any trial, the channel processing times are conditionally independent exponential random variables with rate α . The rates are not constant but fluctuate across trials such that $\alpha = A\beta$, where the β reflects channel-specific properties, whereas A represents the system’s momentary level of alertness which is itself exponentially distributed with parameter τ . The processing time distribution is characterized by its joint survivor-function, S_N , defined as

$$S_N(t_1, \dots, t_n) = P_N(T_1 > t_1, \dots, T_n > t_n).$$

³Note that assuming T_1, T_2, \dots, T_n to be *exchangeable random variables* implies marginal invariance (for a definition of exchangeability, see e.g. (Ross, 1983).

⁴This example was first suggested by D. Vorberg for the first-terminating/exhaustive case.

Note that the distribution is indexed by $N = \{1, 2, \dots, n\}$ to make explicit that it refers to the situation where processing is initiated on all n channels. (For simplicity, we also write $N - 1$ for $\{1, 2, \dots, n - 1\}$.) By randomizing A , S_N is obtained as

$$\begin{aligned}
 S_N(t_1, \dots, t_n) &= \int_0^\infty S_N(t_1, \dots, t_n | A = a) \tau \exp(-\tau a) da \\
 &= \int_0^\infty \prod_{i=1}^n \exp(-a\beta t_i) \tau \exp(-\tau a) da \\
 &= \int_0^\infty \exp(-a\beta \sum_{i=1}^n t_i) \tau \exp(-\tau a) da \\
 &= \frac{\tau}{\tau + \beta \sum_{i=1}^n t_i}
 \end{aligned}
 \tag{3}$$

It is straightforward to derive from Eq. (3) the joint survivor function for any subset of channels by using the fact that, e.g.,

$$\begin{aligned}
 P_{N-1}(T_1 > t_1, \dots, T_{i-1} > t_{i-1}, T_{i+1} > t_{i+1}, \dots, T_n > t_n) \\
 = S_N(t_1, \dots, t_{i-1}, 0, t_{i+1}, \dots, t_n).
 \end{aligned}$$

Obviously, the model obeys both unlimited capacity and marginal invariance. In order to derive the distribution function of the k th-order statistics, define for any $n \geq 1, 0 \leq k \leq n$

$$S_{k,n}(t) = \sum_{1 \leq i_1 < i_2 < \dots < i_k \leq n} P(T_{i_1} > t \cap T_{i_2} > t \cap \dots \cap T_{i_k} > t).
 \tag{4}$$

Moreover, let $m_n(t)$ equal the number of events among the $\{T_i > t\}_{i=1, \dots, n}$ that occur. A very useful tool in the theory of extreme values is the following identity.

Lemma. ((Galambos, 1978), Theorem 1.4.1).

$$P(m_n(t) = s) = \sum_{j=0}^{n-s} (-1)^j \binom{j+s}{s} S_{j+s,n}(t).$$

Inserting the survivor function of the model into (4) yields

$$\begin{aligned}
 S_{k,n}(t) &= \sum_{1 \leq i_1 < i_2 < \dots < i_k \leq n} \frac{\tau}{\tau + \beta kt} \\
 &= \binom{n}{k} \frac{\tau}{\tau + \beta kt}
 \end{aligned}$$

using marginal invariance and, from the lemma,

$$\begin{aligned}
 P(m_n(t) = s) &= \sum_{j=0}^{n-s} (-1)^j \binom{j+s}{s} \binom{n}{j+s} \frac{\tau}{\tau + \beta(j+s)t} \\
 &= \binom{n}{s} \sum_{j=0}^{n-s} (-1)^j \binom{n-s}{j} \frac{\tau}{\tau + \beta(j+s)t}.
 \end{aligned}
 \tag{5}$$

Recalling the definition of order statistics leads to

$$\begin{aligned}
 P(T_{k:n} \leq t) &= P(m_n(t) \leq n - k) \\
 &= \sum_{s=0}^{n-k} P(m_n(t) = s).
 \end{aligned}$$

Finally, inserting (5) into the above yields the desired distribution function of the k th order statistic

$$P(T_{k:n} \leq t) = \sum_{s=0}^{n-k} \binom{n}{s} \sum_{j=0}^{n-s} (-1)^j \binom{n-s}{j} \frac{\tau}{\tau + \beta(j+s)t}.$$

3.1. A numerical example

The figure below illustrates the distribution inequalities from the corollary for this example with values $n = 8$ and $k = 2$, where the parameters of the multivariate survivor function have been set to $\tau = 50$ and $\beta = 1$ (see Fig. 1):

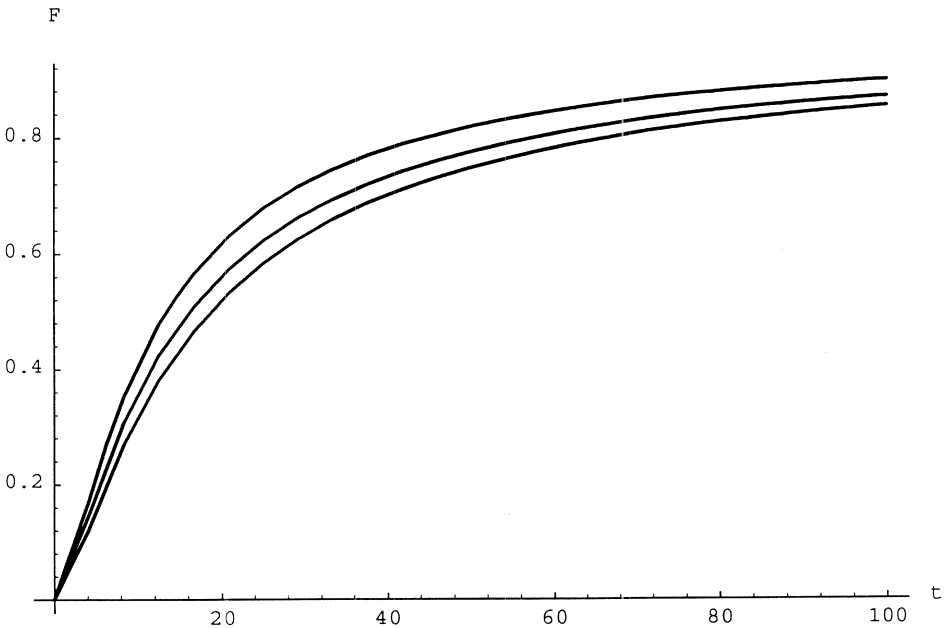


Fig. 1. Distribution functions illustrating the inequalities of the corollary for the example with conditionally independent exponential channel processing times:

$$F_{2,7}(t) \leq F_{2,8}(t) \leq 3F_{2,7}(t) - 2F_{2,6}(t)$$

$$F_{2,7}(t) \leq F_{2,8}(t) \leq 3F_{2,7}(t) - 2F_{2,6}(t)$$

4. Determining the order of the stopping rule

Consider an experimental situation with n parallel channels operating under unlimited capacity where one would like to gain information about the order of the stopping rule, i.e. the number k . Typically, k depends on the specific experimental condition under which the response times are collected. Presumably, the more the speed of the responses is emphasized in the instruction, the smaller k will be. The question is, how can k be determined quantitatively from considering appropriate distribution inequalities without making any assumptions about the parametric type of the RT distributions or the dependence structure of the channels' operating times? For simplicity, only the marginally invariant case will be discussed here.

From the definition of order statistics, for any number n of channels $F_{k:n}(t) \leq F_{k':n}(t)$ with $k > k'$. If the two distribution functions refer to two different speed conditions, this inequality is a testable hypothesis on the order of the distributions under these two conditions. However, there is no way to gain *quantitative* information about k or k' from this inequality. In fact, k and k' could be any number between 1 (first-terminating model) and n (exhaustive model).

On the other hand, the inequalities derived earlier do yield information about the order k . Let $G_n(t)$ denote the observable (rather, estimable) response time distribution in an n -channel situation. If the underlying stopping rule is of order k , then the upper bound for $G_n(t)$ according to Corollary 1 is

$$(k + 1)G_{n-1}(t) - kG_{n-2}(t).$$

This can be rewritten as

$$[kG_{n-1}(t) - (k - 1)G_{n-2}(t)] + [G_{n-1}(t) - G_{n-2}(t)],$$

where the first term in brackets is equal to the upper bound for a stopping rule of order $k - 1$ and the second term in brackets is nonnegative because of the lower bound given in Corollary 1. Thus, a first observation is that the upper bound for a stopping rule of order k cannot be stricter than the upper bound for a stopping rule of order $k - 1$. In other words, even if data are not consistent with the upper bound of the $(k - 1)$ th-terminating stopping rule, they may still be consistent with the k th-terminating rule. Second, another way to write the upper bound is

$$[2G_{n-1}(t) - G_{n-2}(t)] + (k - 1)[G_{n-1}(t) - G_{n-2}(t)].$$

Here, the first term in brackets is the upper bound for a first-terminating rule. This implies that the order can, in principle, be determined by selecting the smallest integer number k that makes the upper bound consistent with the data:

$$\text{order} \equiv \min\{k | G_n(t) \leq (k + 1)G_{n-1}(t) - kG_{n-2}(t), \text{ for all } t\}$$

5. Concluding remarks

While the theorem in this paper generalizes a previous result on first-terminating and exhaustive processing to general k th-order processing, a number of problems remain open for further investigation. First, as already mentioned above, not much is known about the sharpness of the bounds in the theorem and the corollary. Comprehensive numerical studies of the distribution inequalities for various classes of multivariate distribution functions are called for. From these one might also be able to infer multivariate dependence conditions that yield tighter bounds in the distribution inequalities. Second, the assumption of unlimited capacity, although given here in a slightly less restrictive form, can be shown not to be a necessary condition for the theorem to hold and, thus, may possibly be replaced by something more plausible from an empirical point of view. Third, no acceptable statistical test for the inequality involving the upper bound, or for the method of determining the order outlined in the last section, is known to the author. Finally, it should not escape the reader that the basic tool for deriving the inequalities, the variation of the number of channels, is a very common experimental procedure with ramifications far beyond the context mentioned here. For example, in the area of visual search, the investigation of the effects of varying display set size is a basic tool for studying visual attention (eg., (Palmer, 1995)). It is hoped that the type of results presented here will be helpful in rigorously testing parallel processing models in many other areas as well.

Acknowledgements

This research was supported in part by a grant from Deutscher Akademischer Austauschdienst (DAAD/ACLS Program). Comments by Adele Diederich and Ehtibar Dzhafarov are gratefully acknowledged. The author is also indebted to Dirk Vorberg for discussions and suggestions at an early stage of this work.

Appendix A. Proof of the Theorem

The response time distribution of a k th-terminating parallel model with n channel processing times T_1, T_2, \dots, T_n is defined by the k th order statistic, i.e.

$$\begin{aligned} P_n(T_{k:n} \leq t) &= P_n(\text{at least } k \text{ of } T_1, T_2, \dots, T_n \text{ are less than or equal to } t) \\ &= F_{k:n}(t) \end{aligned}$$

where P_n refers to the probability measure corresponding to the joint distribution of T_1, T_2, \dots, T_n . Moreover,

$$\begin{aligned} P_{n-1}(T_{k:n-1} \leq t) &= P_{n-1}(\text{at least } k \text{ of } T_1, T_2, \dots, T_{i-1}, T_{i+1}, \dots, T_n \\ &\text{are less than or equal to } t) \\ &= F_{k:n-1}^{(i)}(t) \end{aligned}$$

By unlimited capacity, the subscripts $n, n - 1$, etc. can be dropped from the probability measure P .

For arbitrary integers r_1, r_2, \dots, r_k with $1 \leq r_1 < \dots < r_k \leq n$ define

$$B_{r_1, r_2, \dots, r_k}(t) = \{T_{r_1} \leq t\} \cap \{T_{r_2} \leq t\} \cap \dots \cap \{T_{r_k} \leq t\}$$

and let

$$A(t) = \bigcup_{\substack{r_1, r_2, \dots, r_k \\ 1 \leq r_1 < \dots < r_k \leq n}} B_{r_1, r_2, \dots, r_k}(t).$$

For notational convenience, we drop the variable t in the set notation. Then, obviously,

$$F_{k:n}(t) = P(A).$$

Moreover, writing

$$A^{(i)} = \bigcup_{\substack{r_1, r_2, \dots, r_k \\ 1 \leq r_1 < \dots < r_k \leq n \\ i \notin \{r_1, \dots, r_k\}}} B_{r_1, r_2, \dots, r_k}$$

yields

$$F_{k:n-1}^{(i)}(t) = P(A^{(i)}).$$

The following technical lemma will allow us to make use of a Bonferroni-type inequality (see Appendix B) to derive the desired upper bound of the theorem.

Lemma 1. *With the notation introduced above, for arbitrary integers s_1, \dots, s_{k+1} with $1 \leq s_1 < \dots < s_{k+1} \leq n$*

$$A = A^{(s_1)} \cup A^{(s_2)} \cup \dots \cup A^{(s_{k+1})}.$$

Proof. By construction, $A^{(s_1)} \cup A^{(s_2)} \cup \dots \cup A^{(s_{k+1})}$ is contained in A . For the other direction, consider the set B_{r_1, \dots, r_k} for arbitrary r_1, \dots, r_k with $1 \leq r_1 < \dots < r_k \leq n$. To show that B_{r_1, \dots, r_k} is contained in $A^{(s_1)} \cup A^{(s_2)} \cup \dots \cup A^{(s_{k+1})}$, define the set difference of the two index sets,

$$J = \{s_1, \dots, s_{k+1}\} \setminus \{r_1, \dots, r_k\}.$$

Since $J \neq \emptyset$, by construction B_{r_1, \dots, r_k} is contained in

$$\bigcap_{p \in J} A^{(p)}.$$

This proves the lemma. \square

We now consider the case where two of the channels, i and j , say, are dropped. Extending the above notation, let us write

$$A^{(ij)} = \bigcup_{\substack{r_1, r_2, \dots, r_k \\ 1 \leq r_1 < \dots < r_k \leq n \\ i, j \notin \{r_1, \dots, r_k\}}} B_{r_1, r_2, \dots, r_k}$$

yielding

$$F_{k:n-2}^{(ij)}(t) = P(A^{(ij)}).$$

Lemma 2. *With the notation introduced above, for any i, j ($i \neq j$) from $\{1, \dots, n\}$*

$$A^{(ij)} \subset A^{(i)} \cap A^{(j)}.$$

Proof. By construction, $A^{(ij)} \subset A^{(i)}$ and $A^{(ij)} \subset A^{(j)}$. \square

We are now ready to derive the upper bound for the RT distribution function with n channels, $F_{2:n}(t)$. By Lemma 1,

$$\begin{aligned} P(A) &= P(A^{(s_1)} \cup A^{(s_2)} \cup \dots \cup A^{(s_{k+1})}) && \text{(A.1)} \\ &\leq P(A^{(s_1)}) + P(A^{(s_2)}) + \dots + P(A^{(s_{k+1})}) \\ &\quad - P(A^{(s_1)} \cap A^{(s_2)}) - P(A^{(s_2)} \cap A^{(s_3)}) - \dots - P(A^{(s_k)} \cap A^{(s_{k+1})}) \\ &\leq P(A^{(s_1)}) + P(A^{(s_2)}) + \dots + P(A^{(s_{k+1})}) \\ &\quad - P(A^{(s_1 s_2)}) - P(A^{(s_2 s_3)}) - \dots - P(A^{(s_k s_{k+1})}) \end{aligned}$$

with the first inequality being an application of the Bonferroni-type inequality (Appendix B) and the last inequality following from Lemma 2. Translating the above results from the set notation into probability distributions leads, for any t , to the upper bound for $F_{k:n}(t)$ in the theorem. A smallest upper bound is obtained, of course, by taking the minimum over all vectors s . The lower bound follows trivially from the set definitions of A and $A^{(i)}$. This completes the proof of the theorem.

Appendix B. A Bonferroni-type Inequality

The classical *Bonferroni inequalities* have been introduced by Bonferroni (1937). Let A_1, A_2, \dots, A_n be a sequence of events in a probability space. For example, the second order Bonferroni inequality is

$$\sum_{i=1}^n P(A_i) - \sum_{i < j} P(A_i \cap A_j) \leq P\left(\bigcup_{i=1}^n A_i\right) \leq \sum_{i=1}^n P(A_i)$$

These bounds can be quite inaccurate and, therefore, improved bounds, referred to as *Bonferroni-type inequalities*, have been developed (see, e.g. Glaz, 1993; Worsley, 1982). The basic idea is to express

$$A = A_1 \cup (A_2 \cap A_1^c) \cup (A_3 \cap A_2^c \cap A_1^c) \cup \dots \cup (A_n \cap A_{n-1}^c \cap \dots \cap A_1^c).$$

This immediately implies the following Bonferroni-type upper bound that is used in the proof of the theorem:

$$P\left(\bigcup_{i=1}^n A_i\right) \leq \sum_{i=1}^n P(A_i) - \sum_{i=1}^{n-1} P(A_i \cap A_{i+1}).$$

References

- Arnold, B.C., Balakrishnan, N., Nagajara, H.N., 1992. *A First Course in Order Statistics*, Wiley, New York.
- Bamber, D., 1969. Reaction times and error rates for "same"–"different" judgements of multidimensional stimuli. *Perception and Psychophysics* 6, 169–174.
- Bonferroni, C.E., 1937. Teoria statistica delle classi e calcolo delle probabilita. In: Volume in Onore di Riccardo Dalla Volta, Universita di Firenze, Firenze, pp. 1–62.
- Colonius, H., Ellermeier, W., 1997. Distribution inequalities for parallel models of reaction time, with an application to auditory profile analysis. *Journal of Mathematical Psychology* 41, 19–27.
- Colonius, H., Vorberg, D., 1994. Distribution inequalities for parallel models with unlimited capacity. *Journal of Mathematical Psychology* 38, 35–58.
- Diederich, A., Colonius, H., 1987. Intersensory facilitation in the motor component? A reaction time analysis. *Psychological Research* 49, 23–29.
- Galambo, J., 1978. *The Asymptotic Theory of Extreme Order Statistics*, John Wiley, New York.
- Luce, R.D., 1986. *Response Times: Their Role in Inferring Elementary Mental Organization*, Oxford University Press, New York.
- Miller, J.O., 1982. Divided attention: evidence for coactivation with redundant signals. *Cognitive Psychology* 14, 247–279.
- Palmer, J., 1995. Attention in visual search: Distinguishing four causes of a set-size effect. *Current Directions in Psychological Science* 4, 118–123.
- Raab, D.H., 1962. Statistical facilitation of simple reaction times. *Transactions of the New York Academy of Sciences* 24, 574–590.
- Ross, S.M., 1983. *Stochastic Processes*, John Wiley, New York.
- Townsend, J.T., Ashby, F.G., 1983. *The Stochastic Modelling of Elementary Psychological Processes*, Cambridge University Press, Cambridge.
- Townsend, J.T., 1990. Serial vs parallel processing: Sometimes they look like tweedledum and tweedledee but they can (and should) be distinguished. *Psychological Science* 1, 46–54.
- Townsend, J.T., Colonius, H., 1997. Parallel processing response times and experimental determination of the stopping rule. *Journal of Mathematical Psychology* 41, 392–397.
- Ulrich, R., Giray, M., 1986. Separate activation models with variable base-times: Testability and checking cross-channel dependency. *Perception and Psychophysics* 39, 248–254.
- Worsley, K.J., 1982. An improved Bonferroni inequality and applications. *Biometrika* 69, 297–302.
- Glaz, J. Extreme order statistics for a sequence of dependent random variables. In: Shaked, M., Tong, Y.L. (Eds.), *Stochastic Inequalities*, IMS Lecture Notes-Monograph Series, 22 (1993) 100–115