BRIEF REPORT

Simplifying jackknifing of ERPs and getting more out of it: Retrieving estimates of participants' latencies

FREN T. Y. SMULDERS
Department of Psychology, Maastricht University, Maastricht, The Netherlands

Abstract

Research has demonstrated that the jackknifing procedure for estimating ERP latencies (J. Miller, T. Patterson, & R. Ulrich, 1998) yields more accurate estimates of differences between experimental conditions in ERP latency than other methods. However, the scores resulting from this procedure require special adjustments for further analyses and do not directly reflect each participant's latency. Here, a simple transform is proposed that retrieves estimates of each participant's latency from the subaverage scores, rendering further adjustments superfluous. Other advantages of working with participants' latencies are discussed. Results of simulations support the validity of jackknifing and the retrieval transform.

Descriptors: Jackknifing, Event-related brain potential, LRP, N2pc, ERP latency

Event-related brain potential (ERP) research has made important contributions to the development of models of cognitive processes (Luck, 2005). Especially good temporal precision has contributed to this success. For instance, researchers have used the lateralized readiness potential (LRP; De Jong, Wierda, Mulder, & Mulder, 1988; Gratton, Coles, Sirevaag, Eriksen, & Donchin, 1988; Smid, Mulder, & Mulder, 1987) in reaction time studies as a time marker of the moment at which one response hand is more activated than the other. Also, the N2pc (Luck & Hillyard, 1994) has been used in visual search tasks to mark the moment at which a target has been localized (Hackley, Schankin, Wohlschlaeger, & Wascher, 2007). In such studies, the accuracy of a component's latency, or at least of differences between experimental conditions in latency, is crucially important. Ideally, an ERP component's latency would be obtained for each individual trial, but a low signal-to-noise ratio generally precludes this. For LRP latency estimation, several methods have been proposed in the literature, and they were compared by Mordkoff and Gianaros (2000). The method based on jackknifing (Miller, Patterson, & Ulrich, 1998) has been used by the majority of LRP studies published since its introduction (Smulders & Miller, 2009).

The jackknifing method was originally based on the realization that LRP onset latency in a grand average waveform (i.e., averaged across *n* participants) is probably quite reliable, because the grand average will be relatively free of electroencephalogram

(EEG) noise. However, to test hypotheses about the true difference in latency between conditions, an estimate of variability among participants is required (Miller et al., 1998). With jackknifing, within each experimental condition, every participant's average waveform is replaced by the average across the other $n - 1$ participants, called the "subaverage." Obviously, LRP latency will be more reliably determined in the subaverage waveforms than in the individual participants' waveforms, because the subaverages are based on $n - 1$ participants and thus have a substantially larger signal-to-noise ratio. In each subaverage, the moment at which a criterion is exceeded by the waveform can then be used as the subaverage score; let us call it j_i for participant *i*. The set of subaverage scores is used to estimate the standard error of the mean difference in onset latency (Miller et al., 1998) or, more generally, the mean onset latency within every cell of a factorial design and the statistical significance of experimental effects (Ulrich & Miller, 2001).[1] Extensive simulation studies have shown that this jackknifing procedure yields relatively precise estimates of mean latency differences between conditions and has greater statistical power for detecting such differences than other methods while keeping Type I errors at an acceptable level (Kiesel, Miller, Jolicœur, & Brisson, 2008; Miller et al., 1998; Stahl & Gibbons, 2004; Ulrich & Miller, 2001). The jackknifing method has been further developed and validated for correlational designs (Stahl & Gibbons, 2004) and for ERP components other than LRP (Kiesel et al., 2008), and it has also been used for N2pc (Hackley et al., 2007).

[1]Note that the aim of the jackknifing method for ERP latencies deviates from the typical aim of resampling methods, including jackknifing and bootstrapping. The aim is not to estimate the sampling distribution of a statistic, but rather to estimate the variability (and mean; Ulrich & Miller, 2001, p. 818) within the sample at hand.

Considering the set of subaverage scores $j_1, \ldots, j_n$, two observations can be made. First, being based on averaged data, the variance of the set of subaverage scores is smaller than the variance of the original scores would be (Ulrich & Miller, 2001). Therefore, adjustments to the computations of standard statistical tests have been developed: Miller et al. (1998) provided an adjusted equation for the standard error, Ulrich and Miller (2001) derived an adjustment of F values in ANOVA for factorial designs, and Stahl and Gibbons (2004) provided equations for an adjusted standard deviation and correlation coefficient. Second, it is obvious that the individual latency of participant i (call it o_i) is not represented in his or her subaverage score j_i, simply because j_i is based on an average that excludes participant i's waveform. Still, it can be seen intuitively that if a participant has a relatively late LRP onset compared to the other participants, LRP onset in his or her subaverage should be relatively early, because the subaverage does not include his own (late) LRP. As a consequence, in isolation, each subaverage score j_i surely cannot reflect the individual latency of participant i. Rather, an individual participant latency o_i is effectively hidden within the pattern of subaverage scores across all participants.

Here, we propose a simple new method to retrieve estimates of the individual participants' latencies $o_1, \ldots, o_n$ from the total set of subaverage scores $j_1, \ldots, j_n$ of n participants that result from the jackknifing procedure. In the Appendix, it is shown that estimates of each individual latency o_i from the set $o_1, \ldots, o_n$ can be retrieved from the subaverage scores $j_1, \ldots, j_n$ using the following equation:

$$o_i = n\overline{J} - (n-1)j_i, \qquad (1)$$

where $\overline{J}$ is the mean of subaverage scores across n participants. The values of o_i will be referred to as the "retrieved individual" latencies.

Working with the set of individual latencies $o_1, \ldots, o_n$ has several advantages over working with the set of subaverage scores $j_1, \ldots, j_n$. First, each individual latency o_i directly reflects the latency of participant i instead of a composite of the latencies of all the other participants. Second, o_i can be interpreted on the same time scale as other variables that are not the result of jackknifing (e.g., mean reaction time [RT]). These features allow for a direct comparison of variables within each participant, instead of only on the level of the mean across participants. Third, the transform is computationally simpler and/or more convenient than some of the adjustments that are required for $j_1, \ldots, j_n$. The reason is that it may either precede the transfer of data to standard statistical software packages or be carried out within such packages at an early stage instead of at a late stage, as is necessary with the adjustments of F values (if the n is equal across groups) or of the sum of squared errors within each cell of the design (if n is unequal across groups; Ulrich & Miller, 2001). Finally, any statistical tests for which the required adjustments have not yet been worked out (e.g., multivariate analyses of variance) will likely be safely conducted on the set of o_i latencies without any further adjustments.

Below, a number of published adjustments of statistics for subaverage scores are considered. In each case, it will be demonstrated that uncorrected statistics computed with the retrieved individual participants' estimates equal the corrected statistics that are based on subaverage scores. Furthermore, simulations are presented in which the retrieved estimates equal or approach the original latencies of components that entered the jackknifing procedure.

Numerical Verification

I illustrate how the individual participant's estimates are retrieved from a set of subaverage scores using an example data set and

Table 1. *Hypothetical Subaverage Scores and Retrieved Individual Participants' Latencies, Using Equation (1), as a Function of Age Group and Sleep Deprivation*

| | | Level of sleep deprivation | | | | | |
| | | Subaverage scores | | | Retrieved latencies | | |
Age group	Omitted participant	None	Mild	Heavy	None	Mild	Heavy
Young	1	417.75	486.75	560.25	479	503	659
	2	419.75	467.5	576.75	471	580	593
	3	415	493.5	561.5	490	476	654
	4	427.5	504	570	440	434	620
	5	470	498.25	631.5	270	457	374
	Mean	430	490	580	430	490	580
	SD	91.35	56.32	118.24	91.35	56.32	118.24
Middle-aged	1	546	472.75	534.5	616	459	512
	2	566.25	493.25	535	535	377	510
	3	552.75	474.75	536.25	589	451	505
	4	571.5	471.75	531.75	514	463	523
	5	563.5	437.5	512.5	546	600	600
	Mean	560	470	530	560	470	530
	SD	41.58	80.75	39.68	41.58	80.75	39.68
Old	1	563.5	643	685.75	596	578	707
	2	573.5	614	689	556	694	694
	3	564	631	695.75	594	626	667
	4	561.75	640	675.5	603	590	748
	5	587.25	622	704	501	662	634
	Mean	570	630	690	570	630	690
	SD	42.71	48.58	42.82	42.71	48.58	42.82

Note: Subaverage scores were taken from Ulrich and Miller (2001, Table 1). All units are in milliseconds. *SD*s of subaverage scores have been adjusted in accordance with Stahl and Gibbons (2004, Equation 25); *SD*s of retrieved latencies have been computed in the conventional way.
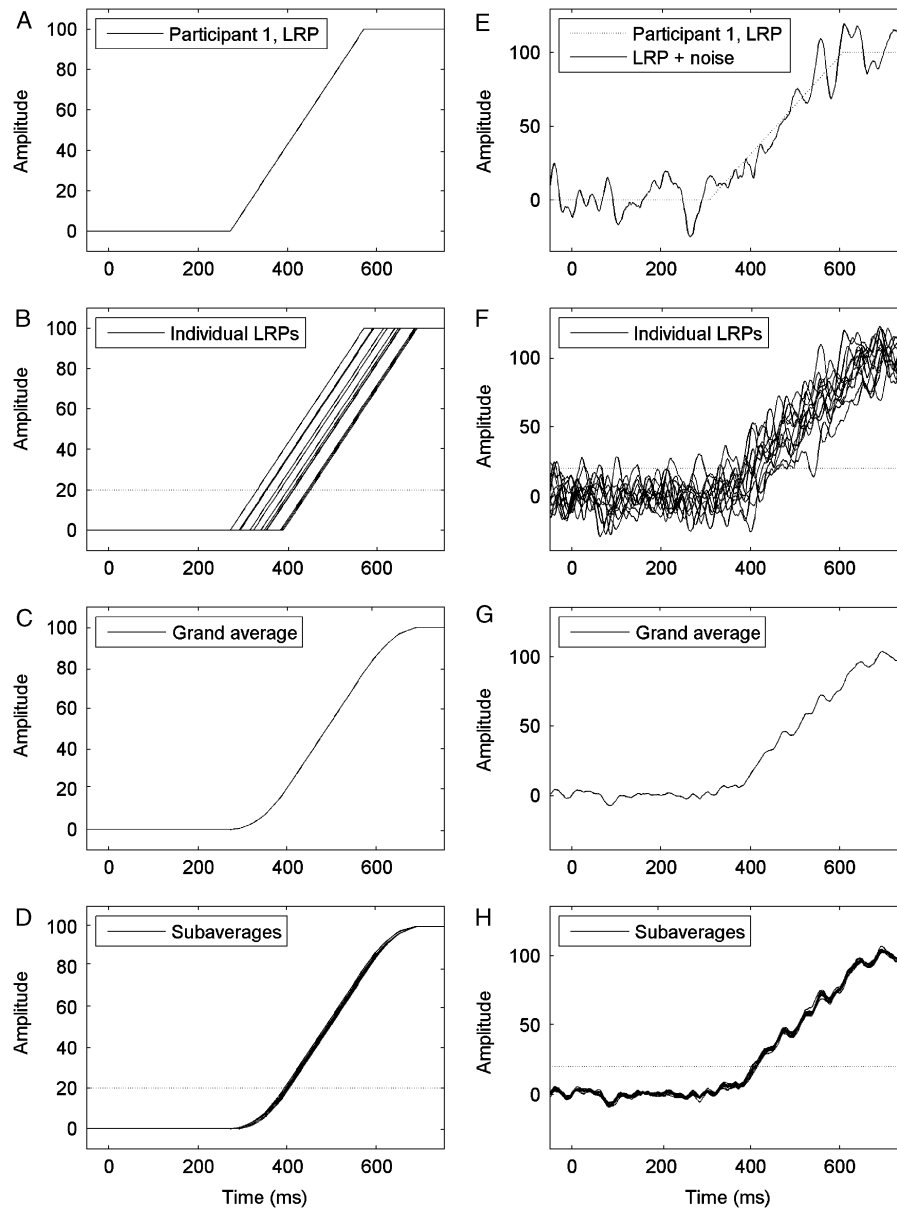
Figure 1. A–D: The jackknifing procedure applied to artificial LRPs with variable latency across participants (Simulation 1). E–H: The same, but with noise added to individual LRPs (Simulation 2). The criterion used to mark the start of the LRP is indicated by the dashed lines in panels B, D, F, and H. In panels D and H, the lines largely overlap, demonstrating the small variability among the subaverages.

show that it yields the same outcomes as more elaborate adjustments. A suitable data set involving a factorial design with two factors is provided by Ulrich and Miller (2001, their Table 1) and is copied in the first three columns of Table 1. The data are subaverage scores from a data set with one between-subjects factor (Age) and one within-subjects factor (Sleep Deprivation), both with three levels. For the first young participant, without sleep deprivation, the retrieved individual participant's estimate, using Equation (1) is $5 \times [(417.75+419.75+415+427.5+470)/5] - [(5-1) \times 417.75] = 2150-1671 = 479$ ms. In Table 1, Columns 4–6 provide the retrieved individual participants' estimates using Equation (1) within the other cells of the design.

First, the mean latency across participants within each cell of the 3×3 design was computed for the data in Columns 1–3 and Columns 4–6. For instance, for young participants in the sleep deprivation-none condition, the mean of the subaverage scores

was 430 ms, and this value was identical to mean of the retrieved estimates. Similarly, there were no differences between the means of the subaverage scores and of the retrieved estimates in any other cell. The standard deviation (*SD*) across participants within each cell was adjusted according to the method of Stahl and Gibbons (2004, Equation 25) for the subaverage scores in Columns 1–3, and an unadjusted *SD* was computed for the retrieved scores in Columns 4–6. For young participants in the sleep deprivation-none condition, these *SD* values were equal ($SD = SD_{adj} = 91.35$ ms), and they were also equal in every other cell. Next, for the young participants, the standard error of the difference between no and mild sleep deprivation was computed for the subaverage scores using the adjusted computation provided by Miller et al. (1998, their Equation 2). Again, its value ($SE_{D,adj} = 38.51$) was equal to the conventional standard error of the retrieved estimates. Finally, the data from Columns 1–3 were

Table 2. *True LRP Latencies, Latencies in LRPs Affected by Noise, Latencies in Subaverage Waveforms, and Retrieved Individual Participant's Latencies, Using Equation (1), in Two Simulations of 12 Participants*

Participant	Simulation 1: Noise-free LRPs			Simulation 2: LRPs with added noise			
	True latency	Subaverage score	Retrieved latency	True latency	Individual latency	Subaverage score	Retrieved latency
1	331.0	403.3	331.0	367.0	363.6	409.3	407.0
2	351.0	401.5	351.0	368.0	332.1	412.6	370.4
3	353.0	401.3	353.0	375.0	381.0	413.3	362.6
4	375.0	399.3	375.0	382.0	57.2	414.7	347.0
5	383.0	398.5	383.0	387.0	273.3	411.0	387.9
6	400.0	397.0	400.0	407.0	137.7	409.5	404.6
7	408.0	396.3	408.0	409.0	23.0	409.9	400.8
8	412.0	395.9	412.0	426.0	93.7	408.1	419.5
9	413.0	395.8	413.0	434.0	15.6	407.5	426.6
10	444.0	393.0	444.0	444.0	425.8	400.7	501.9
11	447.0	392.7	447.0	467.0	327.0	403.8	467.6
12	450.0	392.5	450.0	475.0	458.9	408.7	413.2
Mean	397.3	397.3	397.3	411.8	240.7	409.1	409.1
SD	37.9	3.4	37.9	36.0	157.4	3.8	41.4

Note: All units are in milliseconds.

submitted to a standard ANOVA with Age as a between-subjects factor and Sleep Deprivation as a within-subjects factor. After adjustment of the F values according to Ulrich and Miller (2001, their Equation 1), there were effects of Sleep Deprivation, $F_{adj}(2,24) = 9.8$, $p_{adj} = .001$, Age, $F_{adj}(2,12) = 9.2$, $p_{adj} = .004$, and an interaction between these factors, $F_{adj}(4,24) = 4.9$, $p_{adj} = .005$. Then, the retrieved estimates from Columns 4–6 were submitted to the same ANOVA, but F values were not adjusted. All F and p values were identical to the adjusted ones derived from the subaverage scores. In sum, there appear to be no differences between the onerously derived adjustments of standard statistics for subaverage scores and the conventional statistics computed from the retrieved individual participant's estimates.

Simulated LRPs

Above, it was shown for an example set of subaverage scores that the transformation to retrieved estimates using Equation (1) makes adjustments of a number of statistics superfluous. It was not demonstrated, however, that the retrieved estimates are actually estimates of each individual participant's true latency. To illustrate the latter, two simulations were performed.[2] In the first simulation (Figure 1A–D), a noise-free artificial LRP was generated for each of 12 participants with a variable latency. In a first analysis, the true latency was determined for each participant as the time point at which the waveform exceeded a criterion amplitude (Table 2, Column 1). Then, in a second analysis, the jackknifing procedure was applied and followed by the same latency estimation procedure, yielding the subaverage scores in Column 2.

As expected, their mean was correct, but their standard deviation was reduced. Then, the individual participants' estimates were retrieved from the subaverage scores using Equation (1), and they are given in Column 3. They were all identical to the true latencies. Under these ideal circumstances, when LRP is linear and free of EEG noise, the combination of the jackknifing procedure and retrieval yields an accurate outcome. In reality, because the number of trials included in the average ERP is finite, residual EEG noise will contaminate each individual participant's average LRP waveform. The effect of this contamination is illustrated next.

In the second simulation, simulated EEG noise was added to mimic more realistic conditions (Figure 1E–H). Table 2, Column 4, lists the true latencies of the LRP within the noisy EEG. Column 5 lists the latencies at which the noisy individual waveforms exceeded the criterion. It can be seen that these latencies were strongly affected by the noise: Their mean was decreased, and their standard deviation was increased. Figure 1F shows how the level of noise was sufficient to exceed the criterion well before LRP onset in some participants, contributing to the observed biases in mean and standard deviation. In this case, the jackknifing procedure yields more accurate estimates. The mean of the subaverage scores in Column 6 is quite close to mean true latency, but the standard deviation is, of course, artificially reduced. The retrieved estimates in Column 7 are still affected by the noise, but both their mean and standard deviation are fairly close to the true parameters. Correlations across participants with the true latencies were .13 for the latencies of the noisy individual waveforms and .72 for the retrieved individual latencies. In panel H, it can be seen why the jackknifing procedure yields better estimates than the individual scoring: With jackknifing, the residual noise in the subaverages never exceeds the criterion excessively early.[3] It should be emphasized again that

---

[2]Noise-free LRPs were generated at a sampling rate of 1000 Hz, showing a linear increase from 0 to 100 μV during 300 ms. The onset of each linear increase was adjusted so that the LRP would exceed an (arbitrary) amplitude of 20 μV at a latency drawn from a Gaussian distribution with mean = 400 ms, $SD = 45$ ms. In the second simulation, EEG noise was generated using a Gaussian distribution with mean = 0 μV, $SD = 50$ μV. This noise was low-pass filtered (17-point boxcar, 2-pass, −3dB at 25.9 Hz) and added to the noise-free LRP waveforms. The onset criterion was set to 20 μV in all cases. The parameters for each simulation were selected for illustrative purposes, not to mimic completely realistic values. For more realistic simulations, see, for example, Miller et al. (1998).

[3]Simulation 2 represented only a single realization of the random process, so the results were to some extent determined by chance. Therefore, it was repeated 1,000 times. Across the repetitions, for the true latency, the mean (SD) of the group means was 399.9 (13) ms, the mean of the group SDs was 43.3 (9.6) ms. For the latency at which the noisy individual waveforms exceeded the criterion, these values were 260.7 (40.7) ms and 138.5 (21.5) ms, respectively. Finally, for the mean retrieved latency these values were 397.1 (15.6) ms and 57.1 (41.6) ms, respectively. In sum, these results confirm the pattern that is visible in Table 2.

adjusted analyses of the subaverage scores would give the same statistical values as conventional analyses of the retrieved scores.

Discussion

Prior research has demonstrated that in many cases jackknifing yields more accurate estimates of differences between experimental conditions in ERP latency than other methods. However, the subaverage scores resulting from the jackknifing procedure do not directly reflect each individual participant's latency and require specific adjustments to further statistical analyses. It was shown above that individual participants' latencies can be retrie-

ved from the subaverage scores with a simple transformation. The combination of jackknifing and retrieval was shown to work very well if latencies are defined as the moment in time when a fixed criterion is exceeded along the rising slope of the ERP component, but it can be done in exactly the same manner with different measures (e.g., latencies derived from a segmented regression method; Mordkoff & Gianaros, 2000). The retrieval of individual latencies from the set of subaverage scores using Equation (1) yields estimates of each individual participant's latency that can be directly compared to their RT data and used in further statistical analyses without any adjustments.

REFERENCES

De Jong, R., Wierda, M., Mulder, G., & Mulder, L. J. M. (1988). Use of partial stimulus information in response processing. *Journal of Experimental Psychology: Human Perception and Performance*, *14*, 682–692.

Gratton, G., Coles, M. G. H., Sirevaag, E., Eriksen, C. W., & Donchin, E. (1988). Pre- and post-stimulus activation of response channels: A psychophysiological analysis. *Journal of Experimental Psychology: Human Perception and Performance*, *14*, 331–344.

Hackley, S. A., Schankin, A., Wohlschlaeger, A., & Wascher, E. (2007). Localization of temporal preparation effects via trisected reaction time. *Psychophysiology*, *44*, 334–338.

Kiesel, A., Miller, J., Jolicœur, P., & Brisson, B. (2008). Measurement of ERP latency differences: A comparison of single-participant and jackknife-based scoring methods. *Psychophysiology*, *45*, 250–274.

Luck, S. J., & Hillyard, S. A. (1994). Spatial filtering during visual search: Evidence from human electrophysiology. *Journal of Experimental Psychology: Human Perception & Performance*, *20*, 1000–1014.

Luck, S. J. (2005). *An introduction to the event-related potential technique*. Cambridge, MA: MIT Press.

Miller, J., Patterson, T., & Ulrich, R. (1998). Jackknife-based method for measuring LRP onset latency differences. *Psychophysiology*, *35*, 99–115.

Mordkoff, J. T., & Gianaros, P. J. (2000). Detecting the onset of the lateralized readiness potential: A comparison of available methods and procedures. *Psychophysiology*, *37*, 347–360.

Smid, H. G. O. M., Mulder, G., & Mulder, L. J. M. (1987). The continuous flow model revisited: Perceptual and central motor aspects. In J. J. R. Parasuraman & J. W. Rohrbaugh (Eds.), *Current trends in event- related potential research (EEG Suppl. 40)* (pp. 270–278). New York: Elsevier.

Smulders, F. T. Y., & Miller, J. O. (2009). The lateralized readiness potential. In S. J. Luck & E. S. Kappenman (Eds.), *Oxford handbook of event-related potential components*. New York: Oxford University Press (in press).

Stahl, J., & Gibbons, H. (2004). The application of jackknife-based onset detection of lateralized readiness potential in correlative approaches. *Psychophysiology*, *41*, 845–860.

Ulrich, R., & Miller, J. (2001). Using the jackknife-based scoring method for measuring LRP onset effects in factorial designs. *Psychophysiology*, *38*, 816–827.

APPENDIX: THE RETRIEVAL OF INDIVIDUAL PARTICIPANTS' LATENCIES FROM A SET OF SUBAVERAGE SCORES

It is shown how the individual latencies $o_1, \ldots, o_n$ of n participants can be retrieved from a set of n subaverage scores $j_1, \ldots, j_n$ that result from the jackknifing procedure for estimating ERP latencies (Miller et al., 1998).

Ulrich and Miller (2001, footnote 1) proposed that the mean subaverage score in each condition might be conceived of as an estimate of the mean of the individual participants' latencies in that condition. A demonstration given by Stahl and Gibbons (2004, Proof A, their Equation 9) is repeated here with slightly different notation:

$$\overline{O} = \overline{J}. \tag{A1}$$

The subaverage score for each participant i is defined as

$$j_i = \frac{\sum\limits_{v=1}^{n} o_v}{n-1}, \text{with } v \neq i, \tag{A2}$$

which can also be written as

$$j_i = \frac{\left(\sum\limits_{v=1}^{n} o_v\right) - o_i}{n-1}. \tag{A3}$$

The mean across all n subaverage scores is

$$\overline{J} = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{\left(\sum\limits_{v=1}^{n} o_v\right) - o_i}{n-1} \right) \tag{A4}$$

$$= \frac{1}{n(n-1)} \sum_{i=1}^{n} \left(\left(\sum_{v=1}^{n} o_v\right) - o_i \right) \tag{A5}$$

$$= \frac{1}{n(n-1)} \left(n \sum_{v=1}^{n} o_v - \sum_{i=1}^{n} o_i \right) \tag{A6}$$

$$= \frac{1}{n(n-1)} \left((n-1) \sum_{i=1}^{n} o_i \right) \tag{A7}$$

$$= \frac{\sum\limits_{i=1}^{n} o_i}{n} = \overline{O}. \tag{A8}$$

Next, we show how each original latency o_i can be retrieved from the jackknife subaverage scores j_i. As in (A3), we write the

subaverage score as

$$j_i = \frac{\left(\sum\limits_{v=1}^{n} o_v\right) - o_i}{n - 1},$$

(*A*9)

which can also be written as

$$= \frac{n\overline{O} - o_i}{n - 1}.$$

(*A*10)

Substitution of $\overline{O}$ by $\overline{J}$ (A1) yields

$$j_i = \frac{n\overline{J} - o_i}{n - 1}.$$

(*A*11)

Equation (A11) can be rearranged to obtain

$$o_i = n\overline{J} - (n - 1)j_i.$$

(*A*12)