Evaluating normality of procedure up–down trasformed response by Kolmogorov–Smirnov test applied to difference thresholds for seat vibration

Progress in Vibration and Acoustics
ISSN : 2282 − 7668

Systematic Review

Graziella Aghilone[1] Massimo Cavacece[2]

[1] *Pharmacy and Medicine Faculty, University La Sapienza of Rome, Via Regina Margherita n.273, 00198 Rome (RM) Italy, e–mail: graziella.aghilone@uniroma1.it*

[2] *ASME Member, Department of Civil and Mechanical Engineering, University of Cassino and Lazio Meridionale, Via G. Di Biasio n.43, 03043 Cassino (FR) Italy, e–mail: cavacece@unicas.it*

Abstract

Drivers and passengers appreciate the vibration of vehicles. Their subjective rating of discomfort may affect their opinion of the vehicle. This study investigated the difference threshold required for a change in vehicle ride to be perceived. We propose the Kolmogorov–Smirnov (K–S) test to decide if the sample comes from a population with a specific distribution and to evaluate the variability of the data. [DOI:10.12866/J.PIVAA.2015.09.01] [1]

Keywords: Psychophysical threshold , Discomfort, Statistical approach, Binary responses

1 Introduction

A difference threshold is defined as the difference in value of two stimuli. The difference is just sufficient to be detected. In this research we consider vibration on a seat for vehicle [Aghilone and Cavacece, 2015]. The difference threshold is the minimum change in the magnitude of the whole–body vibration required for the seat occupant to perceive the change in magnitude [Mansfield and Griffin, 2000]. The magnitude of the test stimulus was adjusted according to the subject responses using the up–and–down transformed response (UDTR) method. Experimental investigations on relative difference thresholds, measured using many subjects, offer a distribution of data. The term transformed response has meaning in a statistical sense [Zwislocki and Relkin, 2001]. The

[1]Contributed by Technical Committee for publication in the Progress in Vibration and Acoustics. Manuscript receveid July 13, 2015; final manuscript revised July 20, 2015; pubblished online September 1, 2015.

Response sequences				
Case	Up	Down	Probability	p
1	++-	+++		
2	+-			
3	-			
			$(1-2\cdot p^3)$	0.794

Table 1: Strategies for estimating midpoint

UDTR method is developed by the response sequences, classified in two mutually exclusive and exhaustive groups: an increase–S/N group and a decrease–S/N group. The UDTR method provides an efficient investigation of the difference threshold. The UDTR method can produce the spread of results [Kaernbach, 1991].

In this research we propose the Kolmogorov–Smirnov (K–S) test to decide if the sample comes from a population with a specific distribution and to evaluate the variability of the data [Treutwein, 1995]. An attractive feature is that the distribution of the K–S test statistic itself does not depend on the underlying cumulative distribution function being tested. Another important advantage is that it is an exact test. The main objective of experimental investigations on relative difference thresholds is to obtain accurate estimates [Leek, 2001]. If the distribution of the subject's responses has high variance, more trials are necessary. The examination of the the spread of results is an indication to decide the minimum number of trials necessary to obtain accurate experimental investigation [Vaerenberg et al., 2013].

2 Up–Down Trasformed Response Rules

The transformed up–down method converges to the stimulus level where the probability of obtaining an up sequence equals the probability of a down sequence [Levitt, 1971]. To find the point of convergence for a given rule, we set the probability of an up sequence equal to the probability of a down sequence. UDTR method is related to the probability of obtaining an up response at convergence. The p term represents the correct response $(+)$ (Table 1). The term $(1-p)$ indicates the incorrect response. We can articulate the following cases:

1. the expression $[p\cdot p\cdot\,(1-p)]$ describes $(++-)$;
2. the relation $[p\cdot\,(1-p)]$ corresponds to $(+-)$;
3. the term $(1-p)$ denotes $(-)$.

Setting $p_{up}=p_{down}$ we obtain the following equation

$$\underbrace{p\cdot p\cdot\,(1-p)}_{1}+\underbrace{p\,(1-p)}_{2}+\underbrace{(1-p)}_{3}=\underbrace{p}_{1}\cdot\underbrace{p}_{2}\cdot\underbrace{p}_{3}$$
$$(1-2p^3)=0\ \Rightarrow\ \boxed{p\simeq 0.794}\,. \tag{1}$$

The size of the increases and decreases in signal intensity remains constant over the experimental test of the UDTR method. The choice of step size and initial signal level is determined by the best step size acquired in the past. If the chosen starting point is above or below threshold,

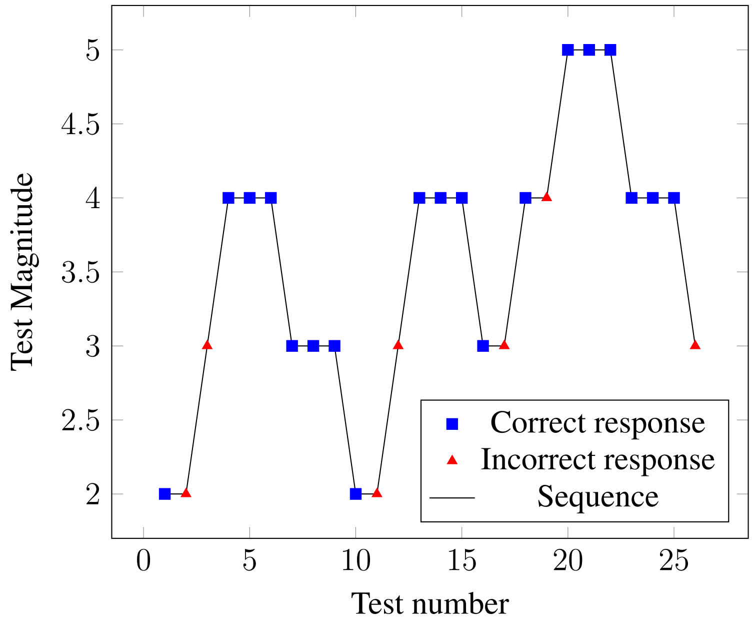

Figure 1: Stategies of the test protocol

there can be problems with this procedure. In such case the method can become inefficient because a great number of adjustments requiring a large number of responses will be needed to bring the signal intensity into the intensity region of the threshold.

If the chosen step size in signal intensity is too small or too large, there can be other aspects. If the signal intensity is changed by one small step, the step size may be so small as to produce no substantial difference in the observer's performance. In this case, many more steps and consequently, many more responses would be required to arrive at threshold. If the step size is too large, correct and incorrect response can cause swings in signal intensity from levels below to levels above tresholds. Under these conditions, the accuracy of threshold would be compromised.

The stimulus is presented to the subject at a given signal–to noise ratio. If a correct response is obtained, it means that the subject identifies correctly the stimulus. The signal–to–noise (S/N) ratio is decreased by a fixed amount. If an incorrect response is obtained, the S/N ratio is increased by the step size. The choice of S/N ratio converges on that value at which 50% of the responses are correct. The magnitude of stimulus increases, decreases or is constant with respect to the correct or incorrect answers. The stategies of the test protocol is proposed in Fig.1.

The test magnitude did not change after the correct response of the test 1. An incorrect response was given after test 2. So the test magnitude increased. After test 3 it was acquired an incorrect response. Three consecutive correct responses were given after tests 4–6. After three correct answers Mansfield and Griffin change stimulus [Mansfield and Griffin, 2000]. After tests 4–6 the magnitude of the test vibration was reduced.

Similarly, tests 7–9 offered correct responses. It follows that the magnitude of the test stimulus decreases. Also in this case, the sign of stimulus changed in magnitude.

The magnitudes of tests 4–6, tests 13–15, and tests 20–22 were obtained after having been previously increased. The incorrect responses obtained after tests 11, 17 and 26 caused the stimulus magnitude to increase after previously decreasing, and so the magnitudes of these stimuli were also used.

The UDTR method estimates the test magnitude in which there is an equal probability of obtaining a number n of consecutive and correct responses. The test protocol, developed by Mansfield and Griffin, selects $n = 3$ as consecutive correct responses. A, C and E represent correct responses. B, D and F represent single and incorrect responses. The probability of the correct responses occurring over an equal number of sets of correct and incorrect responses is equal to $P_{correct} = 0.50$.

Using three consecutive correct responses in the algorithm, the method gives the magnitude, (x), at which the probability of three consecutive correct responses, $F^{(3)}(x)$, is 0.50. The probability of a single correct response, $F^{(1)}(x)$ is 0.794. The probability of obtaining a single correct response is 79.4% at a magnitude (x).

The difference threshold for a subject is quantified by the mean of the average measured r.m.s. magnitude of the three test stimuli, giving consecutive correct responses, and the r.m.s. magnitude of the test stimulus, giving the subsequent incorrect response. In the example, the difference threshold would be calculated using the measured vibration magnitudes during tests 4–5–6 and 11 (M_4, M_5, M_6 and M_{11}). The difference threshold 1 is obtained by following relation

$$\text{difference threshold 1} = 0.5 \cdot \left(\frac{M_4 + M_5 + M_6}{3} + M_{11} \right) . \tag{2}$$

Similarly, difference threshold 2

$$\text{difference threshold 2} = 0.5 \cdot \left(\frac{M_{13} + M_{14} + M_{15}}{3} + M_{17} \right) , \tag{3}$$

and difference threshold 3

$$\text{difference threshold 3} = 0.5 \cdot \left(\frac{M_{20} + M_{21} + M_{22}}{3} + M_{26} \right) . \tag{4}$$

Precautions were taken to ensure that a true threshold was measured. If the mean magnitude of the three test stimuli for the first measured set of correct responses was lower than that for the second set of correct responses, the data, obtained from the first, were rejected and considered a false–positive response.

3 Experimental design

The subjects were exposed to the stimulus twice:

- a 10 s period of a reference motion;
- a 10 s period of a test motion, separated by a 2 s pause.

The reference and test motions had the same waveform, but the magnitude of the test motion was always slightly greater than that of the reference motion. The order of presentation of reference and test motions was assigned randomly. After each pair of stimuli, the subjects responded to the question: *Did you feel more discomfort during the first or second stimulus?*

If a subject identified the test motion as being the more uncomfortable, a correct response was registered. If a subject identified the reference motion as more uncomfortable, an incorrect response was recorded.

4 Kolmogorov-Smirnov test for normality

Let $x_1, \ldots, x_n$ be an ordered sample with $x_1 \leq \ldots\ldots \leq x_n$. We define $S_n(x)$ as follows:

$$S_n(x) = \begin{cases} 0 & x < x_1 \\ \dfrac{k}{n} & x_k \leq x \leq x_{k+1} \\ 1 & x \geq x_n \; . \end{cases} \tag{5}$$

We propose that the sample is a population with cumulative distribution function $F(x)$ and we define D_n as follows:

$$D_n = \max_x |F(x) - S_n(x)| \; . \tag{6}$$

The random variable D_n doesn't depend on $F(x)$. $S_n(x)$ depends on the sample chosen. D_n allows us to estimate $F(x)$.

The distribution of D_n can be calculated by Kolmogorov Distribution. We evaluate the critical values by the Kolmogorov–Smirnov Table [Zaiontz, 2015].

If $D_{n,\alpha}$ is the critical value from the table, then $P(D_n \leq D_{n,\alpha}) = 1 - \alpha$. $D_{n,\alpha}$ can be used to test the hypothesis that a random sample came from a population with a specific distribution function $F(x)$. If

$$\max_x |F(x) - S_n(x)| \; \leq D_{n,\alpha}. \tag{7}$$

the sample data offers a good fit with $F(x)$. It follows that

$$\begin{aligned} 1 - \alpha = & P(D_n \leq D_{n,\alpha}) = P\left(\max_x |F(x) - S_n(x)|\right) \\ = & P\left[S_n(x) - D_{n,\alpha} \leq F(x) \leq S_n(x) + D_{n,\alpha}\right] \\ = & P\left(|F(x) - S_n(x)| \leq D_{n,\alpha}\right) \; . \end{aligned} \tag{8}$$

The relation $S_n(x) \pm D_{n,\alpha}$ provides a confidence interval for $F(x)$.

5 Results

In this research we consider the data of relative difference thresholds measured using 10 subjects and Tarmac 0.4 stimulus, obtained by Mansfield and Griffin. In the Table 2 we summarize the results of Kolmogorov–Smirnov test.

Since $D_n = 0.64 > 0.41 = D_{n,\alpha}$, we conclude that the data is not a reasonably good fit with the normal distribution (Table 3). There is a significant difference between the data and data which is normally distributed. Note that is the same conclusion we reached from looking at the histogram and QQ plot (Fig.2).

x	Freq	Cumul	$S_n(x)$	Z–Score	$F(x)$	Difference
22.9	1	1	0.1	0.44	0.67	0.57
19.4	1	2	0.2	0.08	0.53	0.33
13.3	1	3	0.3	−0.56	0.29	0.01
14.4	1	4	0.4	−0.44	0.33	0.07
9.5	1	5	0.5	−0.95	0.17	0.33
36.1	1	6	0.6	1.82	0.96	0.36
7.1	1	7	0.7	−1.20	0.11	0.58
33.4	1	8	0.8	1.54	0.94	0.14
15.1	1	9	0.9	−0.37	0.35	0.54
15.3	1	10	1.0	−0.35	0.36	0.64

Table 2: KS test for data

mean	Standard Deviation	Count	D_n	$D_{n,\alpha}$
18,65	9.59	10	0.64	0.41

Table 3: Critical value of KS test

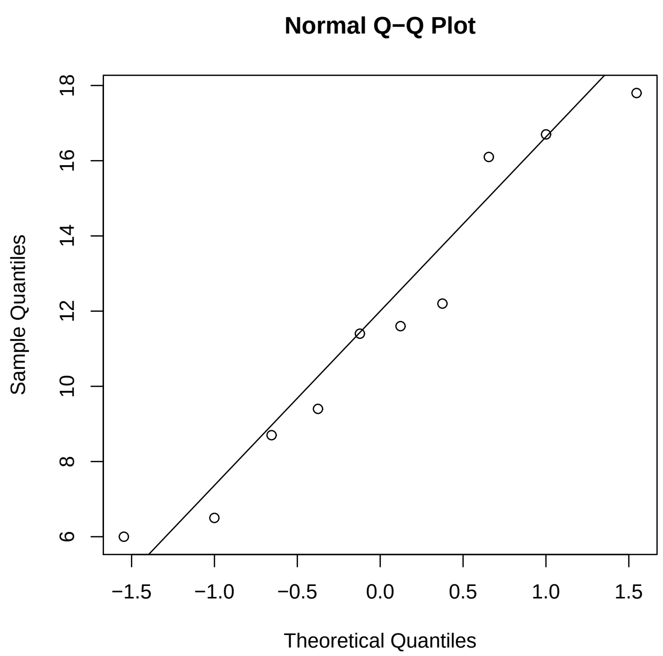

Figure 2: QQplot

6 Discussion

The response sequences are classified in two mutually exclusive and exhaustive groups: an increase–S/N group and a decrease–S/N group. If a response sequence belonging to one of the two groups is obtained, the S/N ratio is changed.

The term transformed response is used in a statistical sense. Value p is the probability of a correct response. In the UDTR procedure, the probability of decreasing S/N ratio at the next step is a function of p. The probability of decreasing S/N ratio is dependent on the grouping of the response sequences.

The disadvantages of the transformed–response procedure are the numerous observations required per run and the cost of discarding early or incomplete runs.

The rule for the intensity variation are the following ones:

- ✓ after every incorrect response, the stimulus intensity is increased by a constant step;
- ✓ after three correct responses, not necessarily consecutive, the intensity is decreased by a step of the same size.

It is important to note that the detection probability of the correct responses may not be constant in a sequence of three stimuli. The probability p depends on stimulus intensity.

At stimulus levels far higher than the threshold, subjects can exhibit a tendency not to notice the stimulus. Subjects can have lapses. We can use the term rate of false negative errors. Similar behaviour also occurs below the threshold when subjects give a yes response. The probability of such responses is termed the guessing rate or the rate of false positive errors. In a yes–no design this behaviour probably reflects noise in the sensory system.

The adaptive procedures of correct/incorrect or yes–no designs offer two aspects:

- ◦ The response domain is limited to experiments which have binary outcomes;
- ◦ The stimulus domain has to be represented by a one–dimensional continuum.

The responses are binomially distributed at any fixed stimulus level. The variability of the percentage correct measures, and the precision with which percentage correct can be measured, depends on two aspects: the number of trials and the unknown true percentage correct at stimulus level. The threshold is usually chosen to be the 50% probability for yes–no designs. The threshold corresponds to the point where same and different responses are equally likely. This type of yes–no threshold defines the point of subjective equivalence.

Psychophysical procedures should be evaluated in terms of cost and benefits. The time of patients and of the experimenter, the number of trials are relevant aspects to achieve a certain accuracy. An empirical threshold is an estimate of a theoretical parameter. The threshold is a function of the experimental data. The threshold is a concise measurement that depends on the results of a set of trials. The relevance of this function is assessed by:

- Bias or systematic error: is the estimated threshold on average equal to the true threshold ?
- Precision or the random error: If the threshold is measured repeatedly, how much variation is to be expected?
- Efficiency: how many trials are required to achieve a certain precision?
- Distribution: Does the data follow a specified distribution?

To obtain accurate estimates, we emphasize an important remark about the minimum number of trials necessary to obtain accurate estimates. If the statistical distribution of the subject's responses has high variance, more trials are necessary. An assessment of the data distribution is necessary.

7 Conclusion

An assessment of the data distribution is necessary. There are three probable reasons for the failure. If an observer commits an error in a UDTR procedure, the stimulus level is increased by a constant step. The stimulus level is decreased by a step of the same magnitude after three correct responses, not necessarily consecutive. The admission of nonconsecutive responses introduces a statistical problem. Statistical aspect depends on ratio between incorrect and correct responses. The probabilities of correct responses in a set of three non consecutive ones can be not constant. The 75% level of correct responses can be derived theoretically without referring to individual probabilities of correct responses. The derivation is based on proportions of correct and incorrect responses. The test period has to be sufficiently long to produce an approximately constant stimulus level. In this way we obtain an average probabilities. The probabilities are not constant because of natural fluctuations in an observer's sensitivity.

References

G. Aghilone and M. Cavacece. Evaluation of psycometric function of vertical whole body vibration on a rigid seat by statistical distribution functions. *Progress in Vibration and Acoustics*, 3(2):1–11, 2015. doi: 10.12866/J.PIVAA.2015.06.01. URL www.mechanical-engineering.name.

C. Kaernbach. Simple adaptive testing with the weighted up-down method. *Perception & Psychophysics*, 49:227–229, 1991.

M.R. Leek. Adaptive procedures in psychophysical research. *Perception & Psychophysics*, 63(8):1279–1292, 2001.

H. Levitt. Trsformed up-down methods in psychoacoustics. *The Journal of the Acoustical Society of America*, 49(2):467–477, 1971.

N.J. Mansfield and M.J. Griffin. Difference thresholds for automobile seat vibration. *Applied Ergonomics*, 31:255–261, 2000.

B. Treutwein. Adaptive psychophysical. *Vision Research*, 35(17):2503–2522, September 1995.

B. Vaerenberg, W. Heeren, and P.J. Govaerts. Managed estimation of ppsychological threshold. *Journal of Hearing Science*, 2013.

C. Zaiontz. Studentized range q table. http://www.real-statistics.com, 2015. URL http://www.real-statistics.com/statistics-tables/studentized-range-q-table/. Available on line.

J.J. Zwislocki and E.M. Relkin. On a psychophysical transformed-rule up and down method converging on a 75*Proceedings of the National Academy of Sciences of the United States of America*, 98(8):4811–4814, 2001. doi: 10.1073/pnas.081082598.