An Efficient and Easy Discretizing Method for the Treatment of Noise Factors in Robust Design*

Antonio Lanzotti and Amalia Vanacore†

University of Naples "Federico II"
Department of Aerospace Engineering
E-mail: antonio.lanzotti@unina.it E-mail: vanacore@unina.it

Abstract

In this work, an efficient and easy statistical method to find an equivalent discrete distribution for a continuous random variable (r.v.) is proposed. The proposed method is illustrated by applying it to the treatment of the anthropometrical noise factors in the context of Robust Ergonomic Design (RED; Lanzotti 2006; Barone S. and Lanzotti A., 2007).

Key Words: Discrete Approximation, Noise Factor, Robust Ergonomic Design, Taguchi's Method.

1. Introduction

A critical issue for Robust Design practitioners is to take in due consideration the continuous nature of some noise factors and face, consequently, the problem of choosing and properly weighting a few representative levels for these factors to use for experiments.

Taguchi (1978) proposed a simple way to discretize a continuous distribution pointing out the relevance of the problem. D'Errico and Zaino (1988) improved the previous technique and generalise Taguchi results. Several discretization techniques have been proposed in literature (Seo and Kwak, 2002). Besides the extensions of the three-level Taguchi method, the most common approaches are numerical quadrature, matching moments, Monte Carlo simulation. Most of these methods fails to strike a balance between accuracy and computational complexity.

In this work, an efficient and easy statistical method to find an equivalent discrete distribution for a continuous random variable (r.v.) is proposed. The proposed method is illustrated by applying it to the treatment of the anthropometrical noise factors in the context of

† Corresponding Author

* The authors thank eng. Giuseppe Marino for technical support. This work was supported by the PRIN 2005 program of Italian Ministry of University and Research: Statistical design of "continuous" product innovation.

Robust Ergonomic Design (RED; Lanzotti 2006; Barone S. and Lanzotti A., 2007). In general, the anthropometrical noise factor can be modeled by a univariate/multivariate continuous variable (*e.g.* the stature or/and the weight of a class of users), or a mixture of univariate /multivariate continuous variables (*e.g.* the stature or/and the weight of two or more classes of users). In these cases an efficient way for treatment of anthropometrical continuous noise factors by a finite number of experimental levels is needed.

The article has the following structure: in Section 2, two traditional techniques commonly used in statistical tolerancing (Taguchi, 1983; D'Errico and Zaino, 1988) together with a method recently introduced for the treatment of the anthropometrical noise factor in RED (Lanzotti A., 2006; Barone S. and Lanzotti A., 2007) are briefly illustrated and reviewed. In Section 3, starting from the main criticisms of the reviewed methods, a new discretizing technique is proposed. In Section 4, the method is applied and compared with previous ones for univariate and mixture of Normal random variables (r.v.s). Section 5 provides final comments and conclusions.

2. Main Criticisms of Existing Discretizing Techniques

Among the various techniques proposed in literature to approximate a continuous r.v. through an equivalent discrete r.v., we will specifically refer to: 1) Taguchi's method (1978); 2) D'Errico and Zaino's method (1988); 3) the moments' method (Lanzotti A., 2006; Barone S. and Lanzotti A., 2007). The first two methods developed in the context of statistical tolerancing and became widely used due to their simplicity. The third technique has been recently introduced for the treatment of noise factors in the context of RED (Lanzotti A., 2006; Barone S. and Lanzotti A., 2007). It is an easy discretization method, available for non Normal r.v.s, too.

2.1 Taguchi's Method

The discretization technique proposed by Taguchi (1978) suggests to approximate a Normal distribution (with known parameters μ, σ) through a two (*i.e.* $x=\mu\pm\sigma$) or three (*i.e.* $x=\mu$, $\mu\pm\sqrt{3/2}\cdot\sigma$) points discrete r.v. whose probability masses (*i.e.* weights) are reported in Table 1.

2.2. D'Errico and Zaino's Method

The discretization technique proposed by D'Errico and Zaino (1988) suggests to approximate a Normal distribution (with known parameters μ, σ) through a three (*i.e.* $x=\mu$, $\mu\pm\sqrt{3}\cdot\sigma$), or four, (*i.e.* $x=\mu\pm\sigma(3\pm\sqrt{6})^{1/2}$), points discrete r.v. whose probability masses (*i.e.* weights) are reported in Table 1.

Table 1. Main characteristics of Taguchi's method, D'Errico-Zaino's method and moments' method

	Number of points j	*Values x_j*	*Weights w_j*	*Percentile $F(x_j)$*	*Limitations*
Taguchi 2 points	1	$\mu-\sigma$	1/2	15.87	Normality assumption Limited coverage range Fixed percentiles Equivalence only for the first two moments
	2	$\mu+\sigma$	1/2	84.13	
Taguchi 3 points	1	$\mu-\sqrt{3/2}\cdot\sigma$	1/3	11.03	
	2	μ	1/3	50.00	
	3	$\mu+\sqrt{3/2}\cdot\sigma$	1/3	88.97	
D'Errico-Zaino 3 points	1	$\mu-\sqrt{3}\cdot\sigma$	1/6	4.16	Normality assumption Fixed percentiles
	2	μ	4/6	50.00	
	3	$\mu+\sqrt{3}\cdot\sigma$	1/6	95.37	
D'Errico-Zaino 4 points	1	$\mu-\sigma(3+\sqrt{6})^{1/2}$	$(3-\sqrt{6})/12$	0.98	
	2	$\mu-\sigma(3-\sqrt{6})^{1/2}$	$(3+\sqrt{6})/12$	22.06	
	3	$\mu+\sigma(3-\sqrt{6})^{1/2}$	$(3+\sqrt{6})/12$	77.94	
	4	$\mu+\sigma(3+\sqrt{6})^{1/2}$	$(3-\sqrt{6})/12$	99.21	
Moments' method n points	No restriction	No restriction	No restriction	No restriction	Not all solutions for the weight definition are admissible

2.3 Moments' Discretization Method

The moments' discretization method (Lanzotti A., 2006; Barone S. and Lanzotti A., 2007) suggests to approximate a continuous r.v. through a discrete distribution defined in J points arbitrarily chosen. The J probability masses can be defined by solving a system of equations obtained by equalizing the $J-1$ moments of the original continuous r.v. to the $J-1$ moments of the equivalent discrete r.v. and imposing the constrain that the sum of probability masses equals the value 1:

$$\begin{cases} w_1x_1+\cdots+w_jx_j+\cdots+w_Jx_J=m_1 \\ w_1x_1^2+\cdots+w_jx_j^2+\cdots+w_Jx_J^2=m_2 \\ \vdots \\ w_1x_1^{J-1}+\cdots+w_jx_j^{J-1}+\cdots+w_Jx_J^{J-1}=m_{j-1} \\ w_1+\cdots+w_j+\cdots+w_J=1 \end{cases}$$

where $\{x_1, \cdots, x_j\}$ represent the arbitrarily chosen points; $\{m_1, \cdots, m_{J-1}\}$ are the $J-1$ moments of the original continuous r.v.; $\{w_1, \cdots, w_J\}$ are the probability masses of the equivalent discrete r.v.

The discretization techniques, briefly illustrated above, present some main limitations that compromise their applicability in many cases :

a) Normality assumption for the original continuous r.v. to be discretized (Taguchi's method, D'Errico-Zaino's method).

b) Equivalence only for the first two moments (Taguchi's method).

c) Limited coverage range for the original continuous r.v. to be discretized (Taguchi's method).

d) Constant weights assigned to fixed percentiles (Taguchi's method, D'Errico-Zaino's method).

e) Solutions are not always admissible (*i.e.* $w_j < 0$) for probability masses to assign to the equivalent discrete r.v. (Moments' method).

It is obvious that these criticisms make the three discretizing techniques not very effective to the treatment of the anthropometrical noise factors in RED context.

3. A New Efficient and Easy Discretizing Technique: SSUMM

The Square Sum Minimization method (SSUMM) proves an easy and efficient discretization technique, available for most continuous r.v.s, enabling to approximate a continuous r.v. through an equivalent discrete distribution defined in J points arbitrarily chosen.

The SSUMM method identifies the probability masses of the equivalent discrete distribution by minimizing the sum of square relative errors on the first $J+2$ moments of the original continuous r.v.:

$$Min \sum_{i=1}^{J+2} e_i^2; \quad e_i = (m_i - m_i^*)/m_i^*$$

$$\begin{cases} 0 \le w_j \le 1 & j = 1, \cdots, J \\ \sum_{j=1}^{J} w_j = 1 \end{cases}$$

where m_i represents the i-th moment of the equivalent discrete r.v.; m_i^* represents the i-th moment of the original continuous r.v.; w_j represents the probability mass associated to the j-th value of the equivalent discrete r.v.

4. Noise Factor Discretization

In Robust Design framework, the definition of the noise factor array requires the choice of the optimal levels and the relative weights for each noise factor.

With regard to the choice of the experimentation levels the results obtainable through the application of the SSUMM method are compared to the ones obtainable by applying D'Errico-Zaino's technique. Some results concerning the application of the SSUMM method to the discretization of non-symmetrical r.v.s are presented. Finally, the case of a misture of Normal r.v.s is illustrated. This latter example is of great interest for its relevance to RED context.

4.1 Univariate Normal Discretization

Consider a Normal r.v. with known parameters. Following D'Errico-Zaino's method three levels are selected corresponding to the percentiles 4.16-th, 50-th and 95.37-th, respectively. Following the SSUMM method, one can arbitrarily select n levels; in the following *three* levels are chosen corresponding to the percentiles 5-th, 50-th and 95-th, respectively. Table 2 reports the weights calculated through the SSUMM method and those one calculated through D'Errico-Zaino's method. In addition, in Table 2 the relative errors on the first six moments are reported for each method and the results show the higher accuracy of the SSUMM method.

Table 2. Application of the SSUMM method and D'Errico-Zaino method to the discretization of a Normal r.v.

	Normal Distribution **(μ=162.7, σ=6.69)**					
SSUMM 3 points	$F(x_1)=0.05$ x_1	$F(x_2)=0.50$ x_2	$F(x_3)=0.95$ x_3	e_1	e_2	e_3
	151.70	162.7	173.704	$-1.337*10^{-6}$	$-1.401*10^{-6}$	$-0.2025*10^{-6}$
	w_1	w_2	w_3	e_4	e_5	e_6
	0.1850	0.6301	0.1849	$1.405*10^{-6}$	$1.757*10^{-6}$	$-1.610*10^{-6}$
D'Errico-Zaino 3 points	$F(x_1)=0.0416$ x_1	$F(x_2)=0.50$ x_2	$F(x_3)=0.9537$ x_3	e_1	e_2	e_3
	151.11	162.7	173.945	$-353.65*10^{-6}$	$-754.9*10^{-6}$	$-1206.8*10^{-6}$
	w_1	w_2	w_3	e_4	e_5	e_6
	1/6	4/6	1/6	$-1712.15*10^{-6}$	$-2273.66*10^{-6}$	$-2894.03*10^{-6}$

4.2 Weibulll Discretization

Consider a Weibull r.v. with known parameters. D'Errico-Zaino's method cannot provide any solution for discretization of non-Normal r.v.s. Following the SSUMM method, *three* levels are chosen corresponding to the percentiles 5-th, 50-th and 95-th, respectively. In Table 3 the weight for each level and the relative errors on the first six moments are reported; the results shows the good accuracy of the proposed discretization method.

Table 3. Application of the SSUMM method to the discretization of a Weibull r.v.

	Weibull Distribution ($\alpha = 1$, $\beta = 0.6$)					
SSUMM	x_1	x_2	x_3	e_1	e_2	e_3
	0.856369	11.3993	47.8364	$-1.246*10^{-6}$	$-1.333*10^{-6}$	$-2.110*10^{-7}$
	w_1	w_2	w_3	e_4	e_5	e_6
	0.9394	0.0603	0.00030	$1.330*10^{-6}$	$1.688*10^{-6}$	$-1.537*10^{-6}$

4.3 Mixture Normal Discretization

Consider a r.v., H_{POP}, distributed as a mixture of two Normal r.v.s with known parameters. The probability density function of H_{POP} can be formulated as follows:

$$f_{H_{POP}}(h_{POP}) = p_{H_I} \cdot f_{H_I}(h_I) + p_{H_{II}} \cdot f_{H_{II}}(h_{II}) \tag{1}$$

where $f_{H_I}(h_I)$ and $f_{H_{II}}(h_{II})$ are the pdf of the two Normal r.v.s H_I and H_{II} with parameters (μ_I, σ_I) and (μ_{II}, σ_{II}), respectively; p_{H_I} and $p_{H_{II}} = (1 - p_{H_I})$ represent the population mix percentages.

Actually, the Normal r.v.s H_I and H_{II} can serve to model the distribution of the r.v. H over two (or more) specific sub-populations (*e.g.* different classes of potential users of a product).

For example, assume that H_I represents the r.v. stature over the class of female potential users of a product and H_{II} represents the r.v. stature over the class of male potential users. Assume that H_I and H_{II} are normally distributed with parameters (μ_I=162.7cm, σ_I= 6.69cm) and (μ_{II}= 175.5cm, σ_{II}= 6.78cm), respectively.

Generalizing D'Errico-Zaino's method, it is possibile to obtain solutions for the case of mixture, too. The procedure develops in two steps: first, each Normal r.v. is discretized, then the probability masses are weighted with the relative mix percentages.

Table 4 shows the probability mass function for the six points discretized mixture obtained through SSUMM method.

Table 4. Probability mass function for the six points discretized mixture obtained through SSUMM

5-th percentile (I)	*50-th percentile (I)*	*95-th percentile (I)*	*5-th percentile (II)*	*50-th percentile (II)*	*95-th percentile (II)*
$x_{I,1}$=151.70	$x_{I,2}$ = 162.72	$x_{I,3}$= 173.70	$x_{II,1}$= 164.35	$x_{II,2}$= 175.5	$x_{II,3}$= 186.65
$w_{I,1}$ = 0.185	$w_{I,2}$ = 0.630	$w_{I,3}$= 0.185	$w_{II,1}$= 0.185	$w_{II,2}$= 0.630	$w_{II,3}$= 0.185

In Figure 1 the probability masses for the discretized mixture are reported for different population mix percentages.

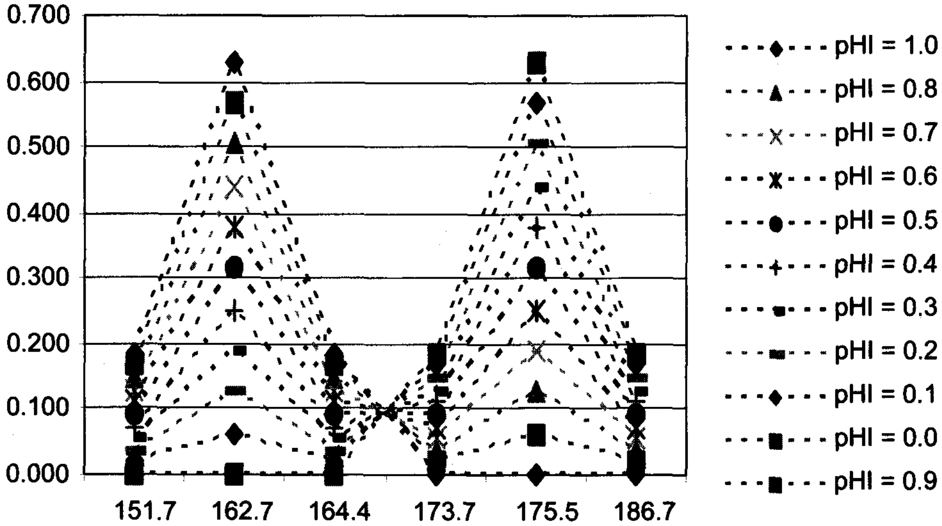

Figure 1. Relationship between probability masses of the discretized misture and p_{H_I}

Table 5 reports the relative errors on the first four moments. Following D'Errico-Zaino approach, the approximated moments are calculated as:

$$m_i = p_{H_I}(w_{I,1}\, x^i_{I,1} + w^i_{I,2}\, x^i_{I,2} + w_{I,3}\, x^i_{I,3}) + P_{H_{II}}(w_{II,1}\, x^i_{II,1} + w^i_{II,2}\, x^i_{I,2} + w_{II,3}\, x^i_{II,3}) \quad (2)$$

where for j=1, 2, 3 $x_{I,j}$ and $x_{II,j}$ are the selected values for the discretized r.v.s H_I and H_{II}; whereas, $w_{I,j}$ and $w_{II,j}$ are the associated probability masses.

The higher accuracy of the approximation obtained through the application of SSUMM method holds also in the comparison on higher order moments (5-th and 6-th moments).

Table 5. Comparison on accuracy between D'Errico-Zaino's method and SSUMM method

	Relative errors on moments D'Errico-Zaino (6 point masses)				*Relative errors on momentsSSUMM (6 point masses)*			
p_{H_I}	$e_1(\times10^{-6})$	$e_2(\times10^{-6})$	$e_3(\times10^{-6})$	$e_4(\times10^{-6})$	$e_1(\times10^{-6})$	$e_2(\times10^{-6})$	$e_3(\times10^{-6})$	$e_4(\times10^{-6})$
90%	-351.36	-749.42	-1196.90	-1696.41	-1.30566	-1.36545	-0.197336	1.36421
80%	-349.11	-744.07	-1187.45	-1681.70	-1.27467	-1.33141	-0.192371	1.32637
70%	-346.89	-738.88	-1178.45	-1667.93	-1.24415	-1.29843	-0.187641	1.29095
60%	-344.70	-733.86	-1169.86	-1655.02	-1.21410	-1.26646	-0.183129	1.25771
50%	-342.55	-728.98	-1161.66	-1642.87	-1.18450	-1.23546	0.178820	1.22646
40%	-340.43	-724.25	-1153.83	-1631.43	-1.15535	-1.20538	-0.174702	1.19703
30%	-338.34	-719.66	-1146.33	-1620.64	-1.12664	-1.17618	-0.170761	1.16927
20%	-336.29	-715.21	-1139.15	-1610.44	-1.09834	-1.14782	-0.166987	1.14303
10%	-334.26	-710.88	-1132.26	-1600.79	-1.07047	-1.12028	-0.163370	1.11819

4.4 Mixture Normal Discretization Through 4 Percentiles

When applying RED, it is necessary to reduce as much as possible the number of the levels of the noise factors in order to cut down the number of experiments required by the cross array. In particular, for RED four percentiles have been *a priori* selected: the 5-th and the 50-th percentile for the stature over the subpopulation of female potential users (*i.e.* class *I*); the 50-th and the 95-th percentile for the stature over the subpopulation of male potential users (*i.e.* class *II*).

Through the application of SSUMM method, the probability masses (or weights: $w_{I,j}$, $w_{II,j}$ for $j=1, 2$) associated to the four selected percentiles, are obtained by minimizing the square relative errors on the first six moments of the r.v. H_{POP} as follows:

$$Min\sum_{i=1}^{6}\left[\left(m_i-m_i^*\right)/m_i^*\right]^2;\begin{cases}0\leq w_{I,j},w_{II,j}\leq 1 \quad j=1,2\\ w_{I,1}+w_{I,2}+w_{II,1}+w_{II,2}=1\\ m_i=w_{I,1}\,x_{I,1}^i+w_{I,2}\,x_{I,2}^i+w_{II,1}\,x_{II,1}^i+w_{II,2}\,x_{II,2}^i\end{cases} \quad (3)$$

where $x_{I,1}=152.78$, $x_{I,2}=162.72$ represent the 5-th and the 50-th percentile for the stature of female potential users; $x_{II,1}$=175.49, $x_{II,2}$=186.65 represent the 50-th and the 95-th percentile for the stature of male potential users. The solutions to the minimization in (3) provide the four probability masses ($w_{I,j}$, $w_{II,j}$ for j=1, 2), for the equivalent discrete r.v. stature of the potential user. Table 6 reports the values and the associated probability masses of the discretized mixture for different population mix percentages, p_{H_I}.

Table 6. Probability mass function for the four points discretized mixture for different

	5-th percentile (I)	*50-th percentile (I)*	*50-th percentile (II)*	*95-th percentile (II)*
	$x_{I,1}$=152.78	$x_{I,2}$=162.72	$x_{II,1}$=175.49	$x_{II,2}$=186.65
p_{H_I}	$w_{I,1}$	$w_{I,2}$	$w_{II,1}$	$w_{II,2}$
90%	0.159	0.614	0.215	0.012
80%	0.140	0.565	0.264	0.031
70%	0.122	0.517	0.312	0.049
60%	0.103	0.468	0.361	0.068
50%	0.084	0.420	0.409	0.087
40%	0.065	0.371	0.458	0.106
30%	0.046	0.322	0.506	0.125
20%	0.028	0.274	0.555	0.144
10%	0.009	0.225	0.603	0.163

The Table 7 reports the relative errors on the first four moments. The approximated moments are calculated as in (2).

Table 7. Comparison on accuracy between D'Errico-Zaino's method (with 6 point masses) and SSUMM method (with 4 point masses)

	Relative errors on moments D'Errico-Zaino (6 point masses)				*Relative errors on moments SSUMM (4 point masses)*			
p_{H_I}	$e_1(\times10^{-6})$	$e_2(\times10^{-6})$	$e_3(\times10^{-6})$	$e_4(\times10^{-6})$	$e_1(\times10^{-6})$	$e_2(\times10^{-6})$	$e_3(\times10^{-6})$	$e_4(\times10^{-6})$
90%	-351.36	-749.42	-1196.90	-1696.41	1.1487	-0.1503	-0.9410	-0.1826
80%	-349.11	-744.07	-1187.45	-1681.70	1.1229	-0.1648	-0.9238	-0.1613
70%	-346.89	-738.88	-1178.45	-1667.93	1.0985	-0.1773	-0.9073	-0.1419
60%	-344.70	-733.86	-1169.86	-1655.02	1.0753	-0.1890	-0.8913	-0.1240
50%	-342.55	-728.98	-1161.66	-1642.87	1.0532	-0.1998	-0.8760	-0.1077
40%	-340.43	-724.25	-1153.83	-1631.43	1.0322	-0.2095	-0.8612	-0.0926
30%	-338.34	-719.66	-1146.33	-1620.64	1.0121	-0.2185	-0.8470	-0.0788
20%	-336.29	-715.21	-1139.15	-1610.44	0.9930	-0.2267	-0.8333	-0.0660
10%	-334.26	-710.88	-1132.26	-1600.79	0.9746	-0.2342	-0.8201	-0.0541

5. Conclusions

SSUMM is a discretization technique based on the constrained minimization of the relative errors on moments due to the approximation of a continuous r.v. with an equivalent discrete distribution. It proves an easy and computationally efficient discretization method. It is available for the discretization of most symmetrical distributions as well as non-symmetrical ones. It can be also directly applied to mixture of r.v.s, thus enabling to economically choose the number of levels for experimentation. The results of the comparison with D'Errico-Zaino's method highlight a higher accuracy for the approximation of the moments of the original continuous r.v.

References

1. Barone S. and Lanzotti A.(2002), "Quality Engineering approach to improve comfort of a new vehicle in virtual environment," *Proceedings of the American Statistical Association, Statistical Computing Section* [CD-ROM], Alexandria, VA: American Statistical Association.
2. Barone S. and Lanzotti A.(2007), "On the treatment of anthropometrical noise factors in Robust Ergonomic Design," *Int. Conf. ADM-Ingegraf*, Perugia (in press).
3. D'Errico J. R. and Zaino N. A. Jr.(1988), "Statistical Tolerancing Using a Modification of Taguchi's Method," *Technometrics*, Vol. 30, pp. 397-405.
4. Lanzotti A.(2006), "Robust design of minicar packaging in virtual environment," Fischer X., Coutellier D.(Eds), *Research in interactive design*, Vol. 2, Springer Paris.
5. Taguchi G.(1978), "Performance Analysis Design," *International Journal of Production Research*, Vol. 16, pp. 521-530.
6. Seo H. S. and Kwak B. M.(2002), "Efficient statistical tolerance analysis for general distributions using three-point information," *International Journal of Production Research*, Vol. 40, pp. 931-944.