

Dynamic Discretization of Continuous Values from Time Series^{*}

Llanos Mora López¹, Inmaculada Fortes Ruiz², Rafael Morales Bueno¹, and
Francisco Triguero Ruiz¹

¹ Dpto. Lenguajes y C.Computacion. E.T.S.I.Informtica.
Campus de Teatinos. 29071 Málaga, Spain
{llanos,morales}@lcc.uma.es
triguero@ctima.uma.es

² Dpto. Matematica aplicada. E.T.S.I.Inf.
C. Teatinos. 29071 Málaga, Spain
ifortes@ctima.uma.es

Abstract. Two methods to assign discrete values to continuous values from time series, using dynamic information about the series, are proposed. The first method is based on a particular statistic which allows us to select a discrete value for a new continuous value from the series. The second one is based on a concept of significant distance between consecutive values from time series which is defined. This definition is based on qualitative changes in the time series values. In both methods, the conversion process of continuous values into discrete values is dynamic in opposition to static classical methods used in machine learning. Finally, we use the proposed methods in a practical case. We transform the daily clearness index time series into discrete values. The results display that the series with discrete values obtained from the dynamic process captures better the sequential properties of the original continuous series.

1 Introduction

The goal of data analysis by time series is to find models which are able to reproduce the statistical characteristics of the series. Moreover, these models allow us to predict next values of the series from its predecessors.

One of the most detailed analysis of statistic methods for the research of time series has been done by Box & Jenkins [2]. The mathematical model for a time series is the concept of discrete-time stochastic process. It is supposed that the observed value of the series at time t is a random sample of size one from a random variable X_t , for $t \in \{1, \dots, n\}$. A time series of length n is a random sample of a random vector like this (X_1, \dots, X_n) . The random vector is considered as part of a discrete-time stochastic process, and observed values of the random variables are considered as the evolution of the process. The process

^{*} This work has been partially supported by project FACA number PB98-0937-C04-01 of the CICYT, Spain. FACA is a part of the FRESCO project

is completely known if the joint probability distribution function of each random vector is known; when the series is Gaussian, the process is completely known when all first and second order moments are known.

The following steps are pursued in the analysis of data using time series theory: identification of the model, estimation of parameters, diagnosis of the model and prediction of new values. The identification of the model can be achieved either in the time domain, using the sample and partial autocorrelation functions, or in the frequency domain, using spectral analysis. In both cases, a previous selection of the possible models which can be used to fit the data must be done. This can be a restriction in the final results. Another important restriction is the following: once the model has been identified and the parameters have been estimated it is supposed that the relation between the parameters is constant along the time. However, in many time series this can not be true.

On the other hand, the analysis of time series and stochastic process has also been analyzed from machine learning techniques. Some of these techniques have solved successfully the restrictions noted above. Two of the most important works in this line have been developed by D.Ron et al., [10,11], where the use of probabilistic finite automata is proposed. In [11], a subclass of probabilistic finite automata has been used for modeling distributions on short sequences that correspond to objects such as single handwritten letters, spoken words, or short protein sequences. In [10], another subclass of probabilistic finite automata, called probabilistic suffix automata, has been used to describe variable memory length Markov processes. Other works arise from the work developed by Dagum [3], based on belief network models; in [3], it is proposed the use of dynamic network models, which are a compromise between belief network models and classical models of time series. They are based on the integration of fundamental methods of Bayesian analysis of time series. However, almost all models used for time series from machine learning are restricted to input features with known discrete values, not allowing continuous valued features as input. For this reason, before any of this method is used, it is necessary to transform the observed continuous values into discrete values. Any method to obtain discrete values must have the following two features: first, it must be known how many different discrete values can appear in the series; second, it must be able to quantify how different two or more consecutive values of the series are.

Let us consider, as an example, the following time series: $\{ \dots, 0.80, 0.82, 0.95, 0.94, 0.96, 0.94, 0.96, \dots \}$, which corresponds to measures of cloudiness index (fraction of overcast sky) for consecutive days. The possible values of this index range from 0 to 1 (clear sky and completely overcast, respectively). With this parameter precision 100 different values can be obtained in the series. However, in most applications, a few different values will suffice to characterize this index and then obtain significant information about the cloudiness of the sky. For instance, the former series can be described as: $\{ \text{overcast, overcast, completely overcast, completely overcast, completely overcast, completely overcast, completely overcast} \}$; that is, the qualitative values $\{ \text{clear sky, } \dots, \text{almost completely overcast, completely overcast} \}$ -corresponding to $\{ \text{from 0.0 to 0.1, from}$

0.1 to 0.2, ..., from 0.9 to 1.0}- give us all the information we need. Therefore, a possible way to transform into discrete values consists of using fixed-size intervals. We refer to this transformation as static discrete conversion. In this paper we propose a transformation which we will refer to as dynamic discrete conversion. The static discrete conversion methods group a set of items into a hierarchy of subsets whose items are related in some meaningful way. Typically, these algorithms perform the conversion into discrete values according to statistical stationary properties of the values, not taking into account the evolution of these values. This procedure has various problems. Consider, for example, the following cloudiness index series: {0.71, 0.89, 0.89, 0.91, 0.89}. Using the proposed static discrete conversion, we obtain the series: {half overcast, almost completely overcast, almost completely overcast, overcast, almost completely overcast}. However, if we observe the series -or this situation in the real world-, we will probably not consider as different situations those when the cloudiness index take value 0.89 or 0.91. To circumvent this problem, a new approach is developed in this paper to transform continuous values into discrete ones. We refer to it as dynamic qualitative discrete conversion: dynamic because it takes into account the evolution of the series; and qualitative because the selection of the discrete value is based on a significant distance which is defined below.

With this type of discretization, we come closer to the form of knowledge of nature phenomena we have in our mind, and we overcome the limitation of a static arithmetic concept. Machine learning techniques and qualitative reasoning allow us to overcome the rigid arithmetic concepts underlying in any equation and to come closer to the language of the brain which, as Neumann said, “is not the mathematical language”.

In Section 2 we develop our two approaches to obtain discrete values. We explain the static conversion and the two dynamic conversion that we propose. In Section 3 a practical case is described using the algorithms developed in this paper. Finally, conclusions and possible extensions are summarized in Section 4.

2 Dynamic Qualitative Discretization

In this section we explain the basic idea of this work: the development of an alternative dynamic discrete conversion method. The goal is to develop an effective and efficient method to transform continuous values into discrete ones using the overall information included in the series and, when possible, feedback with the learning system. To do this, the discrete value which corresponds to a continuous value is calculated using qualitative reasoning, taking into account the evolution of the series.

Qualitative models have been used in different areas in order to get a representation of the domain based on properties (qualities) of the systems which, additionally, allows us to avoid the use of complex mathematical models, [5], [6]. One of the objectives which has been pursued is to develop an alternative physics in which the concepts are derived from a far simpler, but nevertheless formal,

qualitative basis. Qualitative reasoning can also be used to predict the behavior of systems, [7].

On the other hand, any process of discretization has some psychological plausibility since in many cases humans apparently perform a similar preprocessing step representing temperature, weather, speed, etc., as nominal (discrete) values. Following [13], the desirable attributes for a discretization method are:

- Measure of classification “goodness”
- No specific closeness measure
- No parameters
- Globality rather than locality
- Simplicity
- Use of feedback
- Use of *a priori* knowledge
- Higher order correlations
- Fast

The more typical way to deal with numeric variables is to create a partition \mathcal{P} in the range of possible values of the variable, and treat each subset of the partition as a single discrete value. If \mathcal{P} is chosen too thick then important distinctions are missed; if \mathcal{P} is chosen too fine, the data are over-partitioned and the probability estimates may become unreliable. On the other hand, the best partition \mathcal{P} depends on the size of the series which is to be partitioned. A partition with subsets of fixed size is a static partition; if a discrete value is assigned take into account precedent values then we obtain a dynamic partition.

When using a static discrete conversion method, the continuous values are transformed into s discrete values through s intervals of same length. Specifically, the width w_X of a discretized interval is given by:

$$w_X = \frac{\max\{X_t\} - \min\{X_t\}}{s}, \tag{1}$$

where, hereafter, max and min are always considered for $t \in \{1, \dots, n\}$. The discrete value v_i corresponding to a continuous value X_i of the series is an integer from 1 to s which is given by:

$$v_i = discretize(X_i) = \begin{cases} s & \text{if } X_i = \max\{X_t\} \\ [(X_i - \min\{X_t\})/w_X] + 1 & \text{otherwise} \end{cases}, \tag{2}$$

where $[A]$ means the integer part of A . After deciding upon s and finding w_X , it is straightforward to transform the continuous values into discrete ones using this expression.

In this paper we propose the use of a qualitative dynamic discrete conversion method. It is dynamic because the discrete value associated to a particular continuous value can change along the time: that is, the same continuous value can be discretized into different values, depending on the previous values observed in the series. It is qualitative because only those changes which are qualitatively

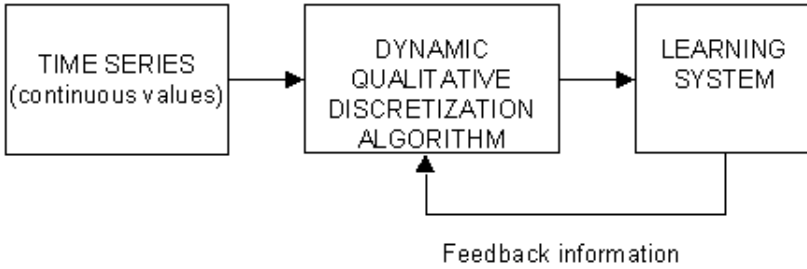


Fig. 1. Dynamic qualitative discretization model

significant appear in the discretized series. Moreover, with this dynamic method the information from the learning model can be take into account.

The process to generate the discrete values is described in Figure 1.

The dynamic qualitative discretization algorithm (DA) is connected to some learning system (LS). Information is fed forward from the DA to the LS. The LS generates feedback for the DA in order to improve the discretization of the continuous inputs. For example, when predicting using variable memory order Markov models, we can decide what discrete value is associated to a particular continuous value either using the probabilities estimated in the model or asking information to the LS about the number of states used to construct the model. Based on this feedback, the DA may perform the adjust of the discrete values corresponding to continuous ones.

With these ideas we propose two procedures to obtain time series with discrete values taking into account the preceding values for the discretization of each value. For both methods, we first justify its use and then propose an algorithm to implement it.

2.1 Using a *t* Statistic

The idea behind this method is to use statistical information about the preceding values observed from the series to select the discrete value which corresponds to a new continuous value of the series. A new continuous value will be associated to the same discrete value as its preceding values if the continuous value belongs to the same population. Otherwise, the static discrete conversion method will assign a new discrete value to this new continuous value. To decide if a new continuous value belongs to the same population as the previous ones, a statistic with Student’s *t* distribution is computed. The method is formally described below.

Given a set of observations, X_1, \dots, X_n, X_{n+1} , it is possible to examine whether X_{n+1} belongs to the same population as the previous values using the statistic:

$$t_{observed} = \frac{X_{n+1} - \bar{X}}{\sqrt{\hat{\sigma}^2(1 + 1/n)}}, \tag{3}$$

where $\bar{X} = n^{-1} \sum_{i=1}^n X_i$, and $\hat{\sigma}^2 = (n - 1)^{-1} \sum_{i=1}^n (X_i - \bar{X})^2$. As it is proved in the Appendix, if certain statistical conditions are met, when X_{n+1} comes from the same population as the previous values the statistic $t_{observed}$ has Student's t distribution with $n - 1$ degrees of freedom. This property suggests the following algorithm:

Algorithm to discretize continuous values from time series using the t statistic

Input:

- continuous time series $\{X_t\}$
- α : significance level, t_α
- s : number of intervals

Method:

```

v1 ← discretize(X1)
ini = 1
for i = 2, ..., n do
  if ((i - ini) > 1) then
     $\bar{X} \leftarrow \left( \sum_{j=ini}^{i-1} X_j \right) / (i - ini)$ 
     $\hat{\sigma} = \sqrt{\frac{\sum_{j=ini}^{i-1} (X_j - \bar{X})^2}{(i - ini) - 1}}$ 
     $t_{observed} \leftarrow |X_i - \bar{X}| / \hat{\sigma}$ 
     $v_i = \begin{cases} discretize(X_i) & t_{observed} > t_{\alpha/2} \\ v_{i-1} & otherwise \end{cases}$ 
  else
     $v_i = discretize(X_i)$ 
  if  $v_{i-1} \neq v_i$  then
     $ini = i$ 
end

```

Output:

Discrete time series: $\{v_i\} = \{v_1, \dots, v_n\}$

2.2 Using Qualitative Reasoning

This method is based on the ideas of qualitative reasoning. In order to characterize the evolution of the system and select discrete values, we propose to use distance functions. These distance functions measure the relationship between consecutive values. They have been used in Instance-Based learning, to determine how close a new input vector is to each stored instance, and use the nearest instance or instances to predict the output class. Therefore, distances are often normalized by dividing the distance for each attribute by the range (i.e. the difference between maximum and minimum) of that attribute, so that the distance for each attribute is in the approximate range $[0, 1]$. It is also common to use

standard deviation instead of range in the denominator. Domain knowledge can often be used to decide which method is most appropriate.

Using some of these ideas we have defined the concept of significant distance between values of the series: two consecutive continuous values correspond to the same discrete value when the distance between them is smaller than a threshold significant distance. This significant distance can be absolute (ASD) -the same for all the sequence- or relative (RSD) to the values which are being compared. We propose the use of the following expressions for these two distance functions:

$$ASD = \frac{|X_i - X_j|}{range\{X_t\}}, \quad (4)$$

$$range\{X_t\} = \max\{X_t\} - \min\{X_t\}, \quad (5)$$

$$RSD = \frac{|X_i - X_j|}{|X_i|}. \quad (6)$$

The proposed expression for the ASD is based on the euclidean metric distance, [14]. The new discrete value is determined depending on how far it is from the preceding values. Changes above the threshold involve changes in the discrete value. When this procedure is used, smooth changes may not be detected, especially if the time series evolves slowly but always in an increasing or decreasing way. For instance, in the time series $\{0.87, 0.88, 0.89, 0.90, 0.91, 0.92, 0.93, 0.94, 0.95, 0.96\}$ all continuous values would be assigned to the same discrete value. To solve this problem we propose to consider only the most recent values of the series to estimate the significant distance.

The first continuous value of the time series is used as reference value. The next values in the series are compared with this reference. When the distance between the reference and a specific value is greater than the threshold (there is a significant difference between them), the comparison process stops. For each value between the reference and the last value which has been compared, the following distances are computed: distance between the value and the first value of the interval, and distance between the value and the last value of the interval. If the former one is lower than the latter one, the discrete value assigned is the one corresponding to the first value; otherwise, the discrete value assigned is the one corresponding to the last value. We now formally describe the algorithm which implements this dynamic qualitative discrete conversion process.

Algorithm to discretize continuous values from time series using flags

Input:

- continuous time series $\{X_t\}$,
- absolute (relative) significant distance, ADF (RSD)
- s : number of intervals

Method:

$$v_1 \leftarrow discretize(X_1)$$

```

ref ←  $X_1$ 
marked ← 1
for  $i = 2, \dots, n$  do
  dist ←  $|X_i - \text{ref}|$ 
   $v_i = \begin{cases} v_{i-1} & \text{if } \text{dist} < \text{ADS (or RDS)} \\ \text{discretize}(X_i) & \text{otherwise} \end{cases}$ 
  if  $v_{i-1} \neq v_i$  then
    ref ←  $X_i$ 
     $j \leftarrow i$ 
     $k \leftarrow 1$ 
    while  $|z_i - z_{i-k}| < |z_{i-k} - z_{j-\text{marked}}|$ 
       $v_{j-k} \leftarrow v_j$ 
       $k \leftarrow k + 1$ 
    end
    marked ← 1
  else
    marked ← marked + 1
  end
end
Output:
Discrete time series:  $\{v_n\}$ 

```

3 Experimental Results

We have used the proposed dynamic discretization algorithms to obtain discrete series from continuous ones. The input data we have used are daily clearness index time series recorded in 10 Spanish stations. The daily clearness index is a climatic parameter obtained from the normalization of daily global radiation received in the surface of the earth. The normalization factor is the extraterrestrial radiation. The values of clearness index range between 0 and 1. In Figure 2 we can observe a fragment of these series, corresponding to data of Málaga in May, 1993 (the total number of observations from Málaga is approximately 2500).

For this series we have used the following method of discretization:

1. Static discretization
2. Dynamic qualitative discretization using the t statistic
3. Dynamic qualitative discretization using flags.

With the first method, the resulting discretized series has jumps which do not correspond to significant changes in the continuous values. The other two methods show more accurately the changes which are observed in the continuous time series.

To analyze whether the continuous series $\{X_t\}$ and the discrete series $\{v_t\}$ -obtained using any of the tree discretization methods- are similar, statistical tests were used to compare their means, variances and cumulative probability distribution functions (cpdf). In all cases, we can accept the hypothesis that the discretized series has the same mean, variance and pdf as the original one

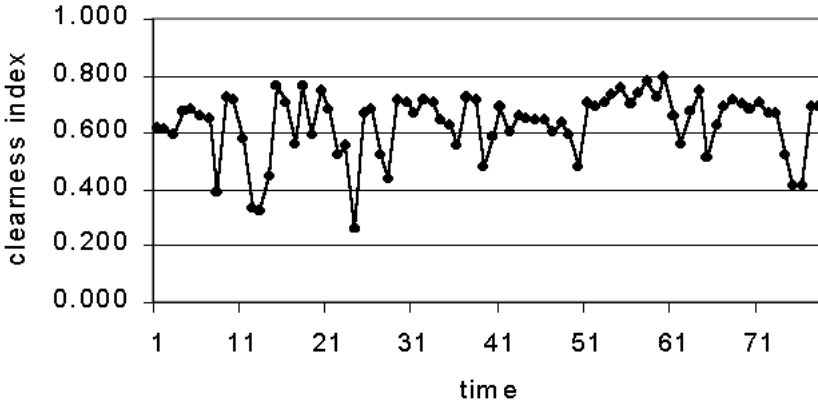


Fig. 2. Fragment of a series of daily cleanness index

Table 1. Number of different sequences which appear in discretized series

Method	Order(length)			
	2	3	4	5
Static discret.	91	596	2627	6373
Dynamic disc. (t-statistic)	84	520	2157	4699
Dynamic disc.(qualitative reasoning)	86	571	2230	4875

(tests have been carried out with 0,95 as significance level; the pdf's have been compared using the Kolmogorov-Smirnov two-sample statistic -see [12], pp.401-403, for a description of this statistic). Thus, the original time series and the discretized one seem to have similar statistical characteristics. In Figure 3, it is depicted the cpdf of the original series from Málaga and the cpdf of the three discretized series which have been obtained from the original one.

On the other hand, cleanness index series have been studied using time series models, such as Markov models, [1,8]. To use Markov models, the input data are sequences built with a set of terminal symbols. In this case, the terminal symbols are the different discrete values. Basically, Markov models of order m analyze the probability of appearance of sequence with length m . Using this information, a model is built for the series (for instance, probabilistic finite automata). With this idea in mind, we have analyzed the number of different sequence which appear in our discretized series for order $m = 2, 3, 4$ and 5 . In Table 1 we show the results which have been obtained:

We observe that the number of different sequences which appear is lower when dynamic models are used to discretize the series (compare rows two and three in Table 1 with row one). This is a logical result because when the evolution of the series is taking into account, the real changes which take place in the original series are detected more precisely.

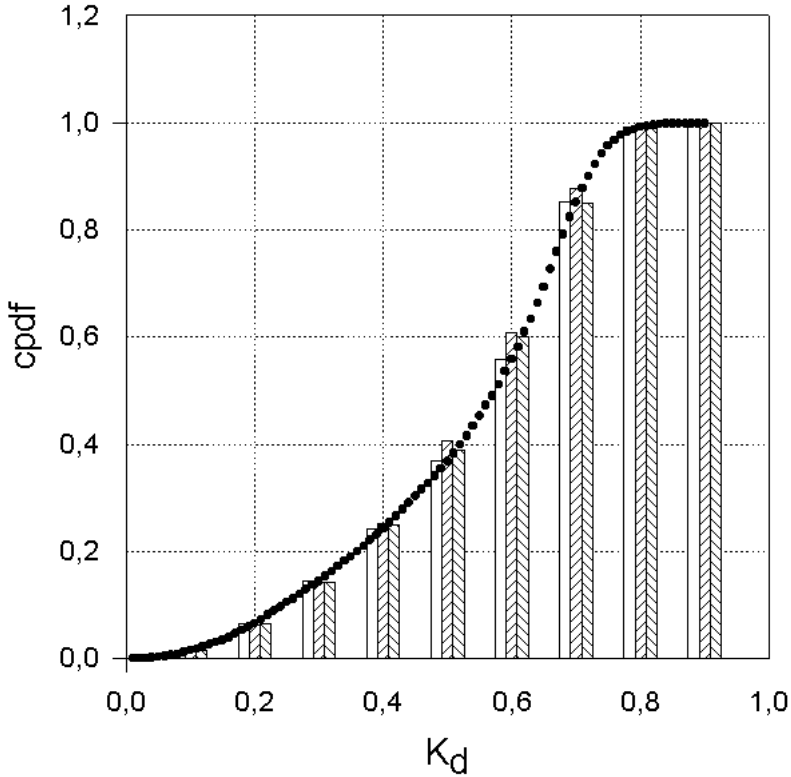


Fig. 3. cpdf for continuous (*points*) and discretized series (*rectangles; from left to right: static discretization, dynamic discretization using the t-statistic and dynamic qualitative discretization using flags*)

4 Conclusions and Future Work

Most of the machine learning models used for time series only work with discrete values. For this reason, before using any of these methods, it is necessary to convert them into discrete values. We have developed two dynamic methods to discretize continuous input values from time series. The main contribution of the work is that, with these methods, the evolution of the time series can be taken into account in the discretization process.

The algorithms proposed have been used in a practical case: the discretization of a climatic parameter -namely, the daily clearness index. Our results show that if a dynamic method is used to discretize continuous values from a time series, then the resultant series captures more accurately the true evolution of the series.

A fixed number of intervals has been used in the discretization process. It would be interesting to design a discretization process which decides, for each

specific case, what is the maximum and minimum number of intervals which can be used to obtain a discretized series which behaves similarly as the original one. We will evaluate our methods with another data sets in a immediately future.

References

1. AGUIAR, R.J., COLLARES-PEREIRA, M., CONDE J.P. Simple procedure for generating sequences of daily radiation values using a library of Markov Transition Matrix. *Solar Energy*, 40, 269-279, 1988. 288
2. BOX, G.E.P., JENKINS, G.M. *Time Series Analysis forecasting and control*. USA: Prentice Hall, 1976. 280
3. DAGUM, P., GALPER, A. Time series prediction using belief network models. *Int. Journal Human-Computer Studies*, 42, 617-632, 1995. 281
4. FELLER, W. *An Introduction to Probability Theory and Its Applications*. Vol. 1. Chapman & Hall, London; John Wiley & Sons, New York, 3rd ed., 1968. 290
5. FORBUS, K.D. Qualitative Process Theory, *Artif. Intell.* 24: 85-168, 1984. 282
6. KLEER, J. BROWN, J.S. A qualitative physics based on confluences, *Artif. Intell.* 24: 7-83, 1984. 282
7. KUIPERS, B. Commonsense reasoning about causality: deriving behavior from structure, *Artif. Intell.* 24: 169-203, 1984. 283
8. MORA-LÓPEZ, L., SIDRACH-DE-CARDONA, M. Characterization and simulation of hourly exposure series of global radiation. *Solar Energy*, 60, No.5, 257-270, 1997. 288
9. RABINER, L.R. A tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. *Proceedings of the Seventh Annual Workshop on Computational Learning Theory*, 1994.
10. RON, D., SINGER, Y., TISHBY, N. Learning Probabilistic Automata with Variable Memory Length. *Proceedings of the Seventh Annual Workshop on Computational Learning Theory*, 1994. 281
11. RON, D., SINGER, Y., TISHBY, N. On the Learnability and Usage of Acyclic Probabilistic Finite Automata. *Journal of Computer and System Sciences*, 56, 133-152, 1998. 281
12. SHORACK, G.R., WELLNER, J.A. *Empirical Processes with Applications to Statistics*. Wiley Series, New York, 1986. 288
13. VENTURA, D., MARTINEZ, T.R. BRACE: A Paradigm for the Discretization of Continuously Valued Data. *Proceedings of the Seventh Florida Artificial Intelligence Research Symposium*, 117-21, 1994. 283
14. WILSON, D.R., MARTINEZ, T.R. Improved Heterogeneous Distance Functions. *Journal of Artificial Intelligence Research*, 6, 1-34, 1997. 286

Appendix

Assume that X_1, \dots, X_n are independent and identically distributed observations from a Gaussian population. It is well-known (see [4]) that:

$$\bar{X} \sim N(\mu, \sigma^2/n); \quad \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\sigma^2} \sim \chi_{n-1}^2, \quad (7)$$

where $\bar{X} = n^{-1} \sum_{i=1}^n X_i$, $\mu = E[X_i]$, $\sigma^2 = var(X_i)$ and χ_{n-1} denotes the chi-square distribution with $n - 1$ degrees of freedom; moreover these two distributions are independent. Hence, if X_{n+1} is another observation from the same Gaussian population, and independent of the previous ones, then:

$$\frac{X_{n+1} - \bar{X}}{\sqrt{\sigma^2(1 + 1/n)}} \sim N(0, 1); \quad \frac{(n - 1)}{\sigma^2} \hat{\sigma}^2 \sim \chi_{n-1}^2, \tag{8}$$

where $\hat{\sigma}^2 = (n - 1)^{-1} \sum_{i=1}^n (X_i - \bar{X})^2$, and these statistics are also independent. Therefore:

$$\frac{\frac{X_{n+1} - \bar{X}}{\sqrt{\sigma^2(1 + 1/n)}}}{\sqrt{\frac{(n - 1)}{\sigma^2} \cdot \frac{\hat{\sigma}^2}{(n - 1)}}} \sim t_{n-1}, \tag{9}$$

where t_{n-1} denotes Student's t distribution with $n - 1$ degrees of freedom. The left-hand member of this last expression is precisely our statistic $t_{observed}$.