Scientific
Research
Publishing

Techniques of Discrete Function Transfers into Continuous Function in Practice

Ming-Jong Lin

Department of Electronic Engineering, Southern Taiwan University of Science and Technology, Tainan
Email: l430107@yahoo.com.tw

Abstract

One confusing question over a long period of time is how transfer the discrete function transfers into continuous function. Recently the issue has been resolved but some details of the transformation process will be introduced in the paper. The correlation coefficients of 100,000 values are established from the two groups of data with the range between −1 and 1, creating a histogram from these correlation coefficient values known as "the probability mass function." The coefficient values are brought into the discrete distribution function, so that transfers into the discrete cumulative function, next converted into a continuous cumulative function, next which is differentiated to get the density function, so that it is easy to being research analysis. A model will be established during the process of the conversion what the medium is "the least squares algorithm." Finally, when the integral of the area within the range of the density function equals to 1, this implies that the transformation complete succeeds from the discrete function to the continuous function.

Keywords

Probability Mass Function, Discrete Cumulative Function, Continuous Cumulative Function, Probability Density Function, The Least Squares Algorithm

1. Introduction

The author has spent some time investigating the reason for the discrete function transferring into continuous function. Due to the method that is what discrete functions convert into continuous functions which are not easy to find in literature [1] [2] [3]. All studies of data start with discrete functions and finally analyze continuous functions. So that it is especially important for converting from discrete functions into continuous functions. To avoid making the same

mistakes that has been made regarding the conversion process, the following describes one case through illustration. Many of research papers comprise historical data and require certain number of data for analysis. However, the graph or the curve in the analytical process is represented by a discrete function a number of studies document, and the data of research have been represented with the discrete functions. Nonetheless, those functions must be converted to a continuous function, so that it can be included in the theoretical derivation for analyze. Therefore, the processing procedures of the two functions convert from discrete into continuous as specifically described herein. This article only narrates the steps during the conversion, as shown in the follows. For subsequently research analysis, the discrete cumulative function will be converted into a continuous cumulative function the former function is yielded from two sets of historical data by the application of correlation coefficient theory.

2. The Correlation Coefficient

The data of two groups are composed of normal 100 sets and normal 1000 sets from historical data, each of which consists of 6 elements (H_2, C_2H_4, C_2H_2, CH_4, C_2H_6, CO). The correlation coefficient of the 100,000 cases have values fallen between −1 and +1 according to the correlation coefficient theory [4]. Their distribution will be presented by the probability histogram, as shown in Figure 1. Figure 2 shows the curve of the cumulative discrete function that been transformed from Figure 1.

For the method of generating of the correlation coefficient value, please refer to the literature [5].
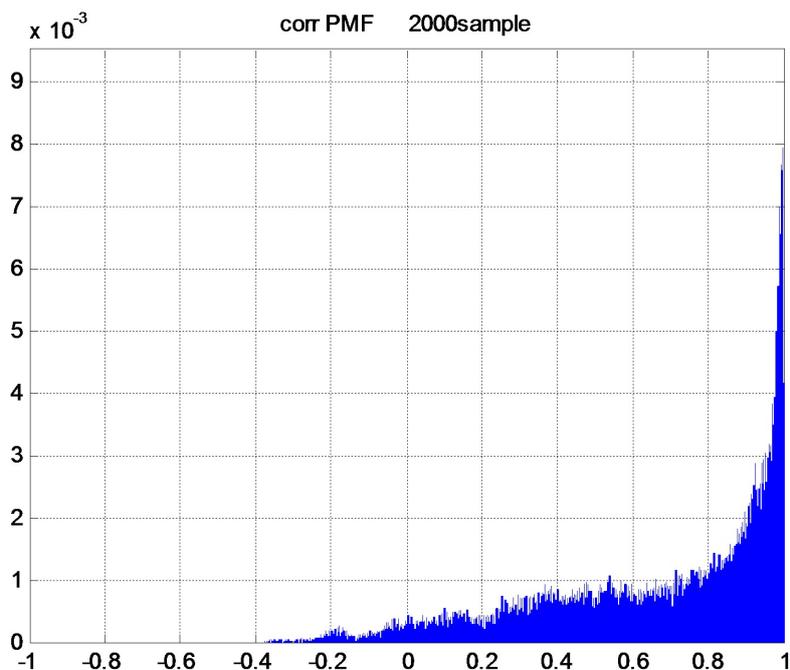
Figure 1. The probability histogram of data groups.
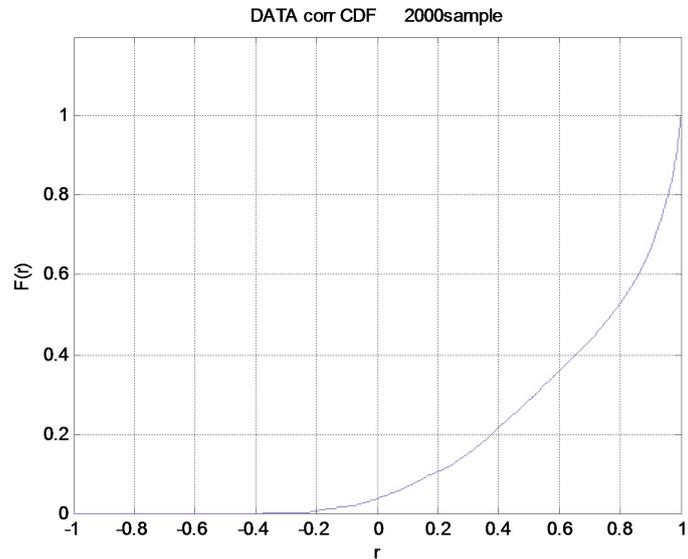
Figure 2. The curve of the discrete cumulative function for data groups.

3. The Discrete Transfers into Continuous Function

In order to find the most similar continuous accumulation function of the above discrete cumulative function, this article refers to Gaussian, Weibull, and others distrbution. The figure of the cumulative continuious function of Weibull distribution and which is widely employed as a model in testing. Maximum likelihood equations are derived for estimating the distribution parameters from complete samples, singly censored samples and progressively (multiple) censored samples. Asymptotic variance-covariance matrices are given for each of these sample types. An illustrative example is included [6]. The maximum-likelihood parameter estimation problem and how the Expectation Maximization (EM) algorithm is can be used for its solution. The form of the EM algorithm as it is often given in the literature. We then develop the EM parameter estimation procedure for two finding the parameters of a mixture of Gaussian densities, and the parameters of a hidden Markov model for both discrete and Gaussian mixture observation models [7]. Eventually, Exponential distribution function is adopted because the figure of the cumulative discrete function was known previously; this paper firstly looks into a similar figure of the cumulative continuous function from the literature. The figure of the cumulative continuous exponential function is adopted as the module conversion in this paper, as shown in Figure 3. The mathematical formula is as described in Equation (1). The symbols of B (or λ) and r that are represented what those rate parameter and variable in Equation (1). The values of the variable are between from −1 to 1.

$$F(r;\lambda) = \left\{ \begin{matrix} \left(e^{B*r} - 1 \right) \\ -1 \leq r \leq 1 \end{matrix} \right\} \tag{1}$$

The curves of cumulative for the actual data (red line) and exponential of

standard (blue line) are shown in **Figure 4**. This paper will use the method of least squares to convert the data curve (discrete) into a continuous function curve. There will be several coefficients generated in the process of conversion, such as A and B in Equation (2).

$$F(r;A;B) = \left\{ \begin{array}{c} A*\left(e^{B*r}-1\right) \\ -1 \leq r \leq 1 \end{array} \right\} \tag{2}$$
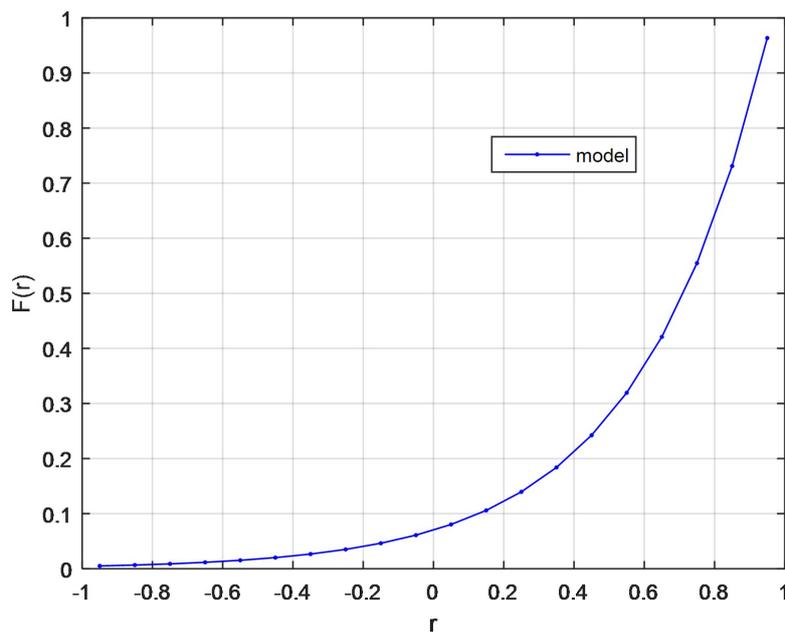
Figure 3. The curve of the cumulative continuous exponential function.
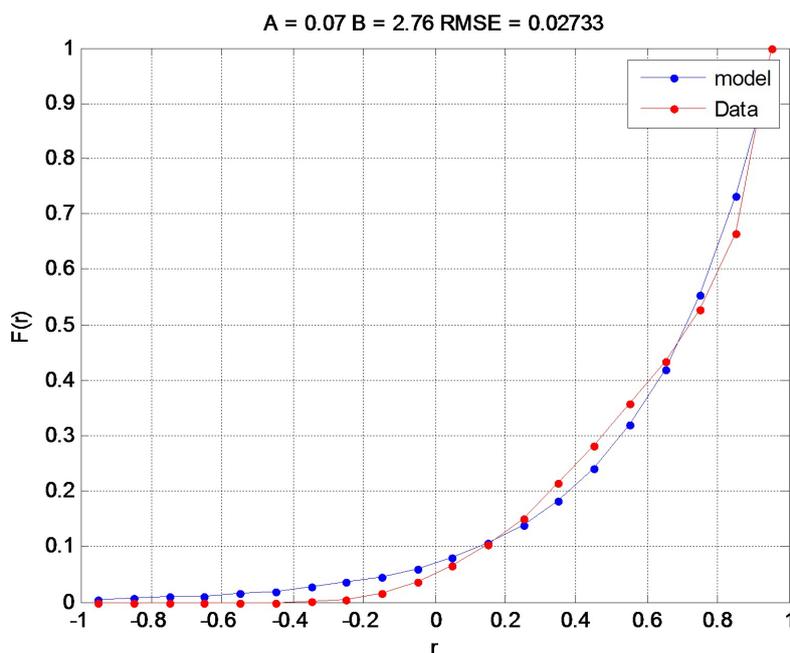
Figure 4. The curve of the actual data and exponential of standard.

After the method of the least squaresalgorithm [8], A and B are respectively calculated 0.07, 2.78 and the error rate (RMSE) of the conversion is 2.7%. The root mean square error (RMSE) is regularly employed in model evaluation studies. The RMSE measures could be more beneficial. In this technical note, we demonstrate that the RMSE is not ambiguous in its meaning, contrary to what was claimed by Willmott *et al.* (2009). The RMSE is more appropriate to represent model performance when the error distribution is expected to be Gaussian. In addition, we show that the RMSE satisfies the triangle inequality requirement for a distance metric, whereas Willmott *et al.* (2009) indicated that the sums-of-squares-based statistics do not satisfy this rule. In the end, we discussed some circumstances where using the RMSE will be more beneficial [9].

4. Verification

After the function was converted from discrete function to continuous, the function still needs verification by the integration of continuous function. First, the continuous accumulation function is differentiated into density function with adjacent re-integration from −1 to 1 of variables the value of the area equals to 1 (or approaching 1) and has been verified correct in the process, which is known as a successful conversion. The formula of the mathematical was been differentiated from continuous function, as shown the formula 3. The Y (A) stands for the area of the integration of $f(r)$ that was shown below.

$$f(r) = 0.195 * e^{2.78*r} \tag{3}$$

$$Y(A) = \int_{-1}^{1} f(r) \, dr = \int_{-1}^{1} 0.195 * e^{2.69*r} \, dr \cong 1$$

5. Conclusion

The discrete function transferring into continuous function is an important work in the process of analyzing of data and investigation in all fields. There are variously of ways for transfer but the paper only introduces which is called "the least square algorithm". This method yields the corresponding coefficient of the Equation (2), while the conversion work is performed by density function. In this paper, the area value of the coverted probability density functin by integration from −1 to 1 is not meet with theoretical value 1, but approaching, found in the process of calculation. Nonetheless, the result of the transformation was not influential. To conclude briefly, the paper will provide the graphical of the program for reference below [10].

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

[1] Chow, C. and Lin, C. (1968) Approximating Discrete Probability Distributions with

Dependence Trees. *IEEE Transactions on Information Theory*, **14**, 462-467. https://doi.org/10.1109/TIT.1968.1054142

[2] Akaike, H. (1959) On a Successive Transformation of Probability Distribution and Its Application to the Analysis of the Optimum Gradient Method. *Annals of the Institute of Statistical Mathematics*, **11**, 1-16. https://doi.org/10.1007/BF01831719

[3] Walpole, R.E., Myers, R.H., Myers, S.L. and Ye, K.Y. (2016) Probability & Statistics for Engineers and Scientists. 9th Edition, Pearson, London.

[4] Correlation Coefficient. https://en.wikipedia.org/wiki/Correlation_coefficient

[5] Lin, M.-J. (2014) A Heuristic Approach to the Diagnosis of Transformer's Insulating Oil. *Journal of Power and Energy Engineering*, **2**, 509-517. https://doi.org/10.4236/jpee.2014.24069

[6] Rockette, H., Antle, C. and Klimko, L.A. (2012) Maximum Likelihood Estimation in the Weibull Distribution Based on Complete and on Censored Samples. *Technometrics*, **7**, 579-588.

[7] Bilmes, J.A. (1998) A Gentle Tutorial of the EM Algorithm and Its Application to Parameter Estimation for Gaussian Mixture and Hidden Markov Models. International Computer Science Institute.

[8] Marquardt, D.W. (1963) An Algorithm for Least-Squares Estimation of Nonlinear Parameters. *Journal of the Society for Industrial and Applied Mathematics*, **11**, 431-441. https://doi.org/10.1137/0111030

[9] Mansard, E.P.D. and Funke, E.R. (1980) The Measurement of Incident and Reflected Spectra Using a Least Squares Method. *Conference Information* 17*th International Conference on Coastal Engineering*, Sydney, Australia, 23-28 March 1980, 154-172.

[10] Program of MATLAB Application Software.

```
clear
clear all
clc
n= 6; % number of gass types
m= 1100; % number of data sets
data = importdata('3HA1.txt');
format short
step = 0.1 ;
check = -1 : step : 1;
check_len = length( check );
pdf=zeros( 1, check_len -1 );
r=zeros( 1, check_len -1 );
corr = 0;
count = 0;
for j=1:100;
    x=data(j, :);
    for i=101:m;
        y=data(i, :);
        R = corrcoef(x,y);
        corr =R(1,2);
        for ii=1:check_len - 1
            if corr >= check( ii ) && corr < check( ii+1 )
                pdf( ii ) = pdf( ii ) + 1;
                count = count + 1;
            end
            r(ii) = ( check( ii ) + check( ii+1 ) )/2;
        end
    end
end
pdf=pdf/count;
sum(pdf)
figure(1),bar( r , pdf, 1, 'b' );
axis([ -1 1 -inf max(pdf)*1.2 ]);
title([ 'corr cdf      '  num2str(check_len-1)   'sample'    ]);
grid on;
% -----------error rms-----------
A = 0 : 0.0001 : 0.01     ;
B = 0 : 0.01 : 5    ;
A_len = length( A );
B_len = length( B );
pdf_len = length( pdf );
a = 1.1;
error_sum_min = 10^7;
error_sum_min_index = 0;
```

```
B_min_error = zeros(1, A_len);
A_min = 0;
B_min = 0;
for kk=1:A_len
    error_sum = zeros(1, B_len);
    RMSE = zeros(1, B_len);
    for jj=1:B_len
        y = A(kk)*exp( B(jj) *r );
        error = zeros(1, pdf_len);
        for ii=1:pdf_len
            error(ii) = ( pdf(ii) - y(ii) )^2;
            error_sum(jj) = error_sum(jj) + error(ii);
        end
    end
        RMSE = sqrt(error_sum/pdf_len);
        figure(2),plot( B, RMSE,'b.-'),grid on;
        title([ 'A = ' num2str( A(kk) )    '    RMSE   B = ' num2str( min(B) )
' ~ '   num2str( max(B)) ]);
        xlabel('B'),ylabel('RMSE');
        [ current_sum_min, current_sum_min_index ] = min(RMSE);
        [ 'B = ' num2str( B( current_sum_min_index ) )   ' error  min = '
num2str(current_sum_min) ]
        if current_sum_min < error_sum_min
            error_sum_min = current_sum_min;
            A_min = A( kk );
            B_min = B( current_sum_min_index );
            y = A_min*exp( B_min *r );
            figure(3),plot(r, y,'.-b' , r, pdf,'.-r'),grid on;
            title([ 'A = ' num2str(A_min) ' B = ' num2str(B_min) ' error = '
num2str( error_sum_min ) ]);
            xlabel('r'),ylabel('RMSE');
            legend('model','Data');
        end
end
```