Computers & Geosciences, Vol. 3, pp. 341-346. Pergamon Press, 1977. Printed in Great Britain

REGULARIZATION OF A SEMIVARIOGRAM

ISOBEL CLARK
Department of Mineral Resources Engineering, Imperial College of Science and Technology,
Prince Consort Road, London SW7 2BP, England

(Received 1 October 1976)

Abstract - The production of estimates using the technique of kriging, and the evaluation of the accuracy of these estimates depends completely on the production of a model for the semivariogram of the deposit. The process of choosing such a model can be complicated in practice by the bulk and geometry of the samples taken. Some aspects of this problem are discussed and some formulae and a FORTRAN IV subroutine are presented to ease the task of modeling.

Key Words: FORTRAN, Kriging, Regionalized variables, Semivariogram.

INTRODUCTION

It has been said that the semivariogram is the basic tool of the Theory of Regionalized Variables. The computation of an experimental semivariogram may be tedious, but rarely difficult. It is in the interpretation and fitting of a model to a practical semivariogram that the real difficulties arise. For instance, to carry out a kriging exercise of any type, we require for the deposit a theoretical model of the semivariogram. This model produces for us the relationship between two points a certain distance apart. Note, however, that we said points. Seldom in practice can we measure a value, say the grade of a mineral, at a point. At the feasibility stage of mine development we are concerned mostly with borehole cores; during development and production we may have many different types of samples, but all of them have one thing in common. They have a certain bulk, a certain geometrical shape and volume. In some instances, they may be assayed using different techniques. All of these factors combine to form the "support" of the samples. A semivariogram which has been calculated on one type of support will not be the same as that calculated on another. The process, of averaging the grade over the bulk of a sample, will influence the shape and behavior in such a manner that we must use an indirect method to derive the theoretical model for the point semivariogram necessary to further analysis and estimation.

REGULARIZATION

Figure 1. Line of point samples, spacing l.

Suppose that we had a line of n point samples at regular intervals ℓ along a line (say a drive or a raise). For example in Figure 1 points on the experimental semivariogram may be calculated with ease:

$$\gamma^*(l) = \frac{1}{2(n-1)} \sum_{i=1}^{n-1} (g_i - g_{i+1})^2$$

$$\gamma^*(2l) = \frac{1}{2(n-2)} \sum_{i=1}^{n-2} (g_i - g_{i+2})^2, \text{ etc.}$$

These points may be plotted on a graph of γ^* vs h, and a model for the point semivariogram chosen and fitted directly to the data points.

Now suppose that, instead of a line of point samples, we are concerned with a borehole which has been assayed in sections of length 1. We no longer know the assay value at some point x, but rather the average value over a length $(x-\ell/2, x+\ell/2)$. How will this affect the semivariogram? Let us consider the problem from an intuitive point of view.

Consider two segments, or "cores", of length ℓ a distance h apart as in Figure 2. Suppose that we knew the grade at every point along both of those segments. Then we could calculate $\gamma^*(h)$ for all the pairs of points, that distance h apart, by computing the difference in grade for each pair, squaring the result, and averaging through all the pairs. However, suppose that we do not know the grade at every point, but only the average grade of each segment. The semivariogram calculated then must be denoted by
$\gamma_\ell^*(h)$, and this will be computed by taking the difference in the average grades, squaring and averaging those. However, to obtain the average grade of the segment, all the variation of grades within the segment has been eliminated. Thus, the average grade of the segment will be a more regular variable than the average grade of individual points. If the variable is more regular, this implies that the differences between any two of the values will be less. That is $\gamma_\ell^*(h)$ will of necessity be less in value than the corresponding $\gamma^*(h)$, whatever the form of $\gamma^*(h)$. Now consider those models for $\gamma^*(h)$ which possess a sill, and therefore a range of influence. Two points which are farther apart than the range of influence, a, may be said to be independent. But, two segments whose centers are a distance a apart will not be independent of one another, because one-half of one segment will be within the range of influence of one-half of the other segment. It is not until $h = a + \ell$, that is the ends of the segments are a apart, that no point in one segment influences the other. Thus, for a point model $\gamma^*(h)$ with a certain range of influence, it is manifest that $\gamma_\ell^*(h)$ will have a range of influence ℓ units larger. This phenomenon of the change in character of the semivariogram as the support of the sample changes is known as *regularization,* and $\gamma_\ell^*(h)$ is termed the regularized semivariogram.

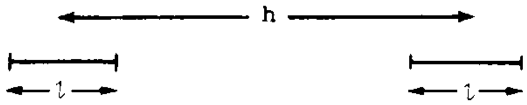

Figure 2. Two segments length l, distance h apart.

REGULARIZATION OF THE USUAL MODELS

The mathematics of the process of regularization are in Matheron (1971), and need not be discussed here. However, it may be of advantage to outline the effect of regularization on the more usual models of the semivariogram. The linear model of $\gamma^*(h)$ is the most widely used of those without a sill, and is characterized by:

$$\gamma(h) = ph \qquad h \geq o.$$

For samples of length ℓ, we determine that:

$$\gamma_l(h) = \frac{ph^2}{3l^2}(3l - h) \quad h \le l$$
$$= p(h - l/3) \quad h \geq l.$$

That is, for distances greater than ℓ - which is all we have in practice - the regularized semivariogram is a straight line with the same slope as the point model, but with an intercept on the γ axis of $-p\ell/3$ (see Fig. 3), if the semivariogram is not complicated by a nugget effect. Thus, the model for the point semivariogram may be obtained by estimating the slope for the regularized semivariogram.

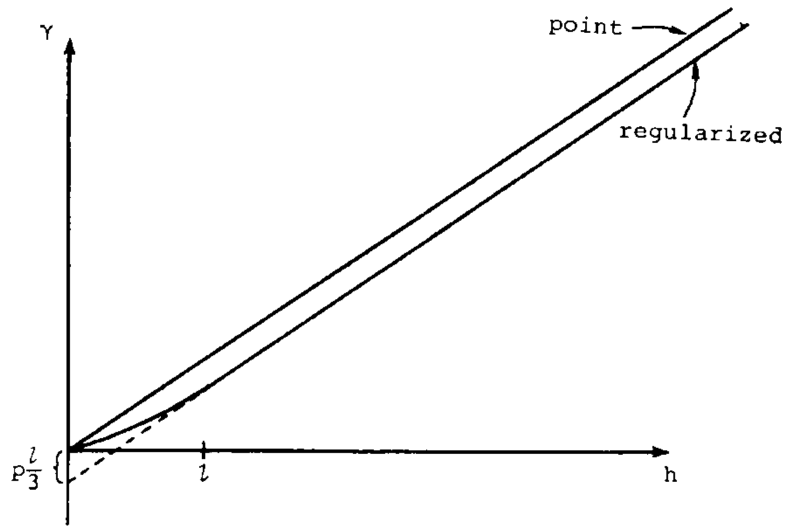

Figure 3. Linear semivariogram—point and regularized.

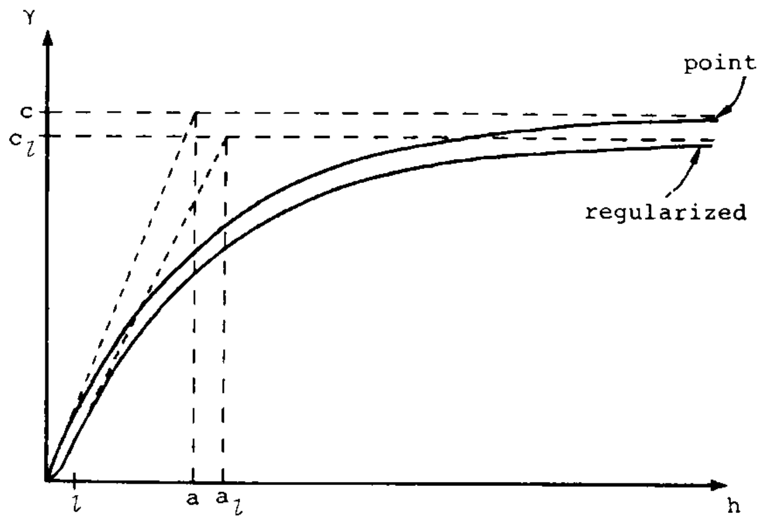

Figure 4. Exponential semivariogram—point and regularized.

There are two usual models for semivariograms with a sill. The exponential model:

$$\gamma(h) = C\{1 - e^{-h/a}\} \quad 0 \le h \le \infty.$$

When regularized by ℓ, this becomes:

$$\gamma_l(h) = C\left[2\frac{a}{l} + \frac{a^2}{l^2}(1 - e^{-l/a})\{e^{-h/a}(1 - e^{l/a}) - 2\}\right] \quad h \ge l. \tag{1}$$

With a somewhat more complex form for the parabolic part with h< ℓ. The sill of this regularized semivariogram obviously will be lower than the sill of the corresponding point model. This can be shown to be:

$$C_l = 2C\left\{\frac{a}{l} - \frac{a^2}{l^2}(1 - e^{-l/a})\right\}. \tag{2}$$

We have experimentally values of the regularized semivariogram, that is $\gamma_\ell^*(h)$ for various values of $h<\ell$. We need to be able to estimate the parameters of the point semivariogram, *a* and C, from γ_ℓ^* in some manner. Having done this, we may substitute these values into the equation (1) to give 'theoretical' points on γ_ℓ, for comparison with γ_ℓ^*.

The first step is to plot a graph of γ_ℓ^* against *h* (see Fig. 4). From this we can estimate C_ℓ immediately as the asymptote of the curve. In practice, of course, γ_ℓ^* will be more irregular. A first approximation for a, now may be determined. Draw a straight line through the first few points on γ_ℓ^* and extend it until it meets the sill level C_ℓ. The value of *h* at which the two meet should be approximately a_ℓ. We have seen that $a_\ell = a + \ell$, so an estimate for a follows immediately.

Having estimated C_ℓ and a (if only approximately) equation (2) may be used to determine an approximation for *C*. These values, *a* and *C*, then may be used in equation (1) to produce values of $\gamma_\ell(h)$. Plotting these points on the same graph as $\gamma_\ell^*(h)$ will give a visual comparison between the model and the data. The parameters can be adjusted if necessary to improve the fit, and the process repeated. For example, if all the values of γ_ℓ were slightly higher than the corresponding γ_ℓ^*, either the estimate of the range of influence is too small or the estimate of the sill is too high, or both. Visual inspection should reveal which. The process is repeated until we produce a $\gamma_\ell(h)$ which seems "close enough" to $\gamma_\ell^*(h)$ for our satisfaction. The final values of *a* and *C* then can be used to characterize the point semivariogram for the deposit.

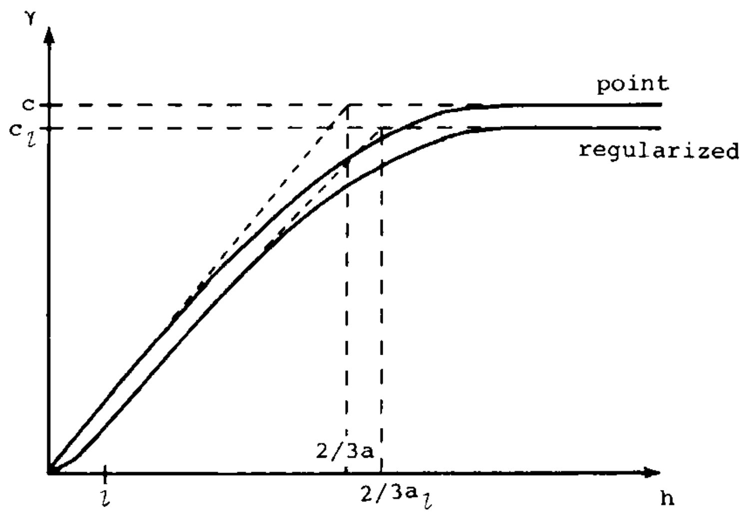

Figure 5. Spherical semivariogram—point and regularized.

The third model - and perhaps the one in widest use in mineral valuation - is the spherical or Matheron model. This semivariogram, expressed as:

$$\gamma(h) = C\left\{\frac{3h}{2a} - \frac{h^3}{2a^3}\right\} \qquad h \le a$$
$$C \qquad h \ge a$$

is the popular one which exhibits linear behavior near the origin, gradually curves off, until at h = a it reaches its sill and stays there permanently. Many mineral deposits seem to conform to this sort of behavior, but the discontinuity in the formula renders the mathematical handling more complex than the other models. This discontinuity also divides the resulting formula for $\gamma_\ell(h)$ into a series of alternatives given various boundary conditions on h and a. These formulae will not be given in the text, but are presented in Appendix I in the form of a FORTRAN IV subroutine which will be described more fully in a later section. However, we can say that the relationship between the sill of $\gamma_\ell(h)$ and that of $\gamma(h)$ is given by:

$$\frac{C}{20}\left(20 - 10\frac{\ell}{a} + \frac{\ell^3}{a^3}\right) \quad \ell \le a$$

$$C_l = \cancel{\frac{C}{20a}\left\{10 + \frac{l^2}{a^2}\right\} \quad l \le a}$$

$$= \frac{C}{20}\left\{15\frac{a}{l} - 4\frac{a^2}{l^2}\right\} \qquad l \ge a \qquad (3)$$

If the linear behavior for the first few points of $\gamma(h)$ is extended to meet the sill, the distance at which this occurs is $2a/3$ (see Fig. 5). So, we may approximate a and C as follows:

- guess C_ℓ, the sill of the regularized semivariogram;
- produce the line through the first few points up to C_ℓ, to give $h = 2a_\ell/3$, and hence a_ℓ;
- $a = a_\ell - \ell$ as before;
- calculate C from the appropriate part of equation (3);
- use attached subroutine to produce values of $\gamma_\ell(h)$ using a and C;
- plot or graph; compare with $\gamma_\ell^*(h)$;
- adjust a and C if necessary; repeat, etc.

When a satisfactory fit has been obtained, we have the point spherical model for the deposit and may proceed with estimation and kriging.

SUBROUTINE REGSPH

The subroutine included in Appendix I is a FORTRAN IV routine to evaluate points of a regularized spherical semivariogram given the following information;

NH - number of points to be evaluated.
A - range of influence of the point semivariogram.
C - sill of the point semivariogram
EL - length of samples producing regularization.
H - array containing values of h for which $\gamma_\ell(h)$ is to be calculated

The subroutine is evoked by:

CALL REGSPH (GL, H, NH, A, C, EL)

and returns in array GL the values of the semivariogram $\gamma_\ell(h)$ for the corresponding values of h in H. Note that the distances in H need not be in ascending order. To assist the user in implementing the subroutine and checking for possible mispunches, a program segment and printout are included in Appendix II.

Acknowledgments The subroutine REGSPH was developed on the CDC 6400 installation at Imperial College, London and the program and subroutine implemented on the DEC system-10 at Syracuse University, Syracuse, New York, while the author was Visiting Associate Professor in the Department of Geology.

REFERENCE

Matheron, G., 1971, The theory of regionalized variables and its applications: Cahiers du Centre de Morphologie Mathematique de Fontainebleau, no. 5, 211 p.

APPENDIX I

```
      SUBROUTINE REGSPH(GL,H,NH,A,C,EL)
      DIMENSION GL(NH),H(NH)
C
C   THIS SUBROUTINE CALCULATES THE VALUES OF A SEMIVARIOGRAM
C   REGULARISED BY LENGTH EL, WHERE THE PUNCTUAL SEMIVARIOGRAM
C   FOLLOWS A SPHERICAL MODEL WITH RANGE A AND SILL C.
C
      U=EL/A
      IF (U-1.) 16,16,17
   16 PART2=20.*U-2.*U*U*U
      GO TO 14
   17 PART2=8./(U*U)-30./U+40.
C
   14 DO 15 I=1,NH
      V=H(I)/A
      IF (V-1.) 21,22,22
   21 IF (U+V-1.) 23,23,24
   24 V2=V*V
      P1=4.-15.*(U+V)+20.*(U+V)*(U+V)-V2*V*(10.-V2)-U*V2*(30.-5.*V2)
      PART1=P1/(U*U)
      GO TO 20
   23 V2=V*V
      TUV=(U+V)*(U+V)*(U+V)
      P1=TUV*(10.-(U+V)*(U+V))-U*V2*(30.-5.*V2)-V2*V*(10.-V2)
      PART1=P1/(U*U)
      GO TO 20
   22 PART1=20.
   20 IF (V-U) 1,2,2
    1 IF (V-1.) 3,4,4
    3 IF (U-V-1.) 5,6,6
    4 IF (U-V-1.) 7,8,8
    2 IF (V-1.) 9,9,10
   10 IF (1.+U-V) 12,12,11
    5 TUV=(U-V)*(U-V)*(U-V)
      V2=V*V
      P3=TUV*(10.-(U-V)*(U-V))+5.*V2*U*(6.-V2)-V2*V*(10.-V2)
      PART3=P3/(U*U)
      GO TO 15
    7 TUV=(U-V)*(U-V)*(U-V)
      P3=TUV*(10.-(U-V)*(U-V))-15.*(U-V)-4.0+20.*V*(2.*U-V)
      PART3=P3/(U*U)
      GO TO 15
    6 V2=V*V
      UV=U-V
      P3=4.0+20.*UV*UV-15.*UV+5.*U*V2*(6.-V2)-V2*V*(10.-V2)
      PART3=P3/(U*U)
      GO TO 15
    8 PART3=30.*(V-U)/(U*U)+20.0
      GO TO 15
    9 V2=V*V
      PART3=10.*V*(3.-V2)+10.*U*(V2-1.)-U*U*(5.*V-U)
      GO TO 15
   11 TVU=(V-U)*(V-U)*(V-U)
      T2=(V-U)*(V-U)
      PART3=(15.*(V-U)+10.*TVU-T2*TVU-4.0-20.*V*(V-2.*U))/(U*U)
      GO TO 15
   12 PART3=20.
   15 GL(I)=0.025*C*(PART1-PART2+PART3)
      RETURN
      END
```

APPENDIX II

```
C   TEST PROGRAM FOR REGULARISATION SUBROUTINE
C
      DIMENSION GAMMAL(50),H(50)
      INCH=-4
      KOUCH=-1
C   INCH IS INPUT CHANNEL,  KOUCH IS OUTPUT CHANNEL
C
      NH=50
      DO 10 I=1,NH
   10 H(I)=FLOAT(I)*0.025
      A=1.00
      C=1.00
      DO 11 J=1,5
      EL=FLOAT(J)*0.1
      CALL REGSPH(GAMMAL,H,NH,A,C,EL)
      WRITE (KOUCH,100) A,C
  100 FORMAT (1H1,45HREGULARISED SEMIVARIOGRAM FOR SPHERICAL MODEL,
     1/3X,24HWITH RANGE OF INFLUENCE ,F5.2,9H AND SILL ,F5.2//)
      WRITE (KOUCH,101) EL
  101 FORMAT (29H LENGTH OF REGULARISATION IS ,F5.2///
     131H       DISTANCE        GAMMA-L(H)/)
      WRITE (KOUCH,102) (H(I),GAMMAL(I),I=1,NH)
  102 FORMAT (3X,F10.4,8X,F8.4)
   11 CONTINUE
      STOP
      END
```

REGULARISED SEMIVARIOGRAM FOR SPHERICAL MODEL
WITH RANGE OF INFLUENCE 1.00 AND SILL 1.00

LENGTH OF REGULARISATION IS 0.10

DISTANCE	GAMMA-L(H)
0.0250	0.0086
0.0500	0.0311
0.0750	0.0629
0.1000	0.0993
0.1250	0.1363
0.1500	0.1730
0.1750	0.2094
0.2000	0.2456
0.2250	0.2813
0.2500	0.3166
0.2750	0.3515
0.3000	0.3858
0.3250	0.4196
0.3500	0.4527
0.3750	0.4852
0.4000	0.5171
0.4250	0.5481
0.4500	0.5784
0.4750	0.6078
0.5000	0.6363
0.5250	0.6639
0.5500	0.6905
0.5750	0.7161
0.6000	0.7406
0.6250	0.7639
0.6500	0.7861
0.6750	0.8071
0.7000	0.8268
0.7250	0.8452
0.7500	0.8622
0.7750	0.8779
0.8000	0.8921
0.8250	0.9047
0.8500	0.9159
0.8750	0.9254
0.9000	0.9333
0.9250	0.9395
0.9500	0.9441
0.9750	0.9471
1.0000	0.9488
1.0250	0.9497
1.0500	0.9500
1.0750	0.9500
1.1000	0.9501
1.1250	0.9501
1.1500	0.9501
1.1750	0.9501
1.2000	0.9501
1.2250	0.9501
1.2500	0.9501

LENGTH OF REGULARISATION IS 0.20

DISTANCE	GAMMA-L(H)
0.0250	0.0044
0.0500	0.0169
0.0750	0.0363
0.1000	0.0614
0.1250	0.0909
0.1500	0.1238
0.1750	0.1587
0.2000	0.1944
0.2250	0.2300
0.2500	0.2651
0.2750	0.2998
0.3000	0.3339
0.3250	0.3675
0.3500	0.4005
0.3750	0.4328
0.4000	0.4644
0.4250	0.4953
0.4500	0.5253
0.4750	0.5546
0.5000	0.5829
0.5250	0.6103
0.5500	0.6367
0.5750	0.6621
0.6000	0.6864
0.6250	0.7096
0.6500	0.7316
0.6750	0.7524
0.7000	0.7719
0.7250	0.7901
0.7500	0.8070
0.7750	0.8224
0.8000	0.8364
0.8250	0.8489
0.8500	0.8599
0.8750	0.8693
0.9000	0.8772
0.9250	0.8837
0.9500	0.8888
0.9750	0.8928
1.0000	0.8956
1.0250	0.8976
1.0500	0.8989
1.0750	0.8997
1.1000	0.9001
1.1250	0.9003
1.1500	0.9004
1.1750	0.9004
1.2000	0.9004
1.2250	0.9004
1.2500	0.9004

LENGTH OF REGULARISATION IS 0.30

DISTANCE	GAMMA-L(H)
0.0250	0.0029
0.0500	0.0114
0.0750	0.0249
0.1000	0.0429
0.1250	0.0647
0.1500	0.0900
0.1750	0.1181
0.2000	0.1484
0.2250	0.1804
0.2500	0.2136
0.2750	0.2474
0.3000	0.2811
0.3250	0.3144
0.3500	0.3470
0.3750	0.3790
0.4000	0.4104
0.4250	0.4409
0.4500	0.4707
0.4750	0.4996
0.5000	0.5276
0.5250	0.5547
0.5500	0.5808
0.5750	0.6059
0.6000	0.6299
0.6250	0.6527
0.6500	0.6744
0.6750	0.6949
0.7000	0.7141
0.7250	0.7320
0.7500	0.7485

0.7750	0.7637
0.8000	0.7775
0.8250	0.7899
0.8500	0.8009
0.8750	0.8105
0.9000	0.8189
0.9250	0.8260
0.9500	0.8320
0.9750	0.8369
1.0000	0.8408
1.0250	0.8438
1.0500	0.8462
1.0750	0.8480
1.1000	0.8492
1.1250	0.8501
1.1500	0.8507
1.1750	0.8510
1.2000	0.8512
1.2250	0.8513
1.2500	0.8513

LENGTH OF REGULARISATION IS 0.40

DISTANCE	GAMMA-L(H)
0.0250	0.0022
0.0500	0.0085
0.0750	0.0186
0.1000	0.0323
0.1250	0.0492
0.1500	0.0690
0.1750	0.0914
0.2000	0.1161
0.2250	0.1427
0.2500	0.1709
0.2750	0.2004
0.3000	0.2308
0.3250	0.2619
0.3500	0.2932
0.3750	0.3244
0.4000	0.3552
0.4250	0.3853
0.4500	0.4146
0.4750	0.4431
0.5000	0.4707
0.5250	0.4973
0.5500	0.5230
0.5750	0.5476
0.6000	0.5712
0.6250	0.5936
0.6500	0.6149
0.6750	0.6350
0.7000	0.6538
0.7250	0.6714
0.7500	0.6877
0.7750	0.7027
0.8000	0.7165
0.8250	0.7290
0.8500	0.7403
0.8750	0.7505
0.9000	0.7594
0.9250	0.7673
0.9500	0.7741
0.9750	0.7799
1.0000	0.7848
1.0250	0.7889
1.0500	0.7923
1.0750	0.7951
1.1000	0.7973
1.1250	0.7990
1.1500	0.8003
1.1750	0.8013
1.2000	0.8020
1.2250	0.8025
1.2500	0.8028

LENGTH OF REGULARISATION IS 0.50

DISTANCE	GAMMA-L(H)
0.0250	0.0017
0.0500	0.0066
0.0750	0.0146
0.1000	0.0255
0.1250	0.0389
0.1500	0.0549
0.1750	0.0731
0.2000	0.0933
0.2250	0.1153
0.2500	0.1389
0.2750	0.1638
0.3000	0.1899
0.3250	0.2170
0.3500	0.2447
0.3750	0.2728
0.4000	0.3012
0.4250	0.3296
0.4500	0.3578
0.4750	0.3855
0.5000	0.4125
0.5250	0.4386
0.5500	0.4637
0.5750	0.4878
0.6000	0.5108
0.6250	0.5327
0.6500	0.5536
0.6750	0.5733
0.7000	0.5918
0.7250	0.6092
0.7500	0.6255
0.7750	0.6406
0.8000	0.6545
0.8250	0.6674
0.8500	0.6791
0.8750	0.6897
0.9000	0.6993
0.9250	0.7079
0.9500	0.7155
0.9750	0.7223
1.0000	0.7281
1.0250	0.7332
1.0500	0.7376
1.0750	0.7413
1.1000	0.7445
1.1250	0.7471
1.1500	0.7493
1.1750	0.7510
1.2000	0.7524
1.2250	0.7535
1.2500	0.7544