Robust Cardinal Interpolation

Eric C. Like and Steven C. Gustafson*

Air Force Institute of Technology, AFIT/ENG, Wright-Patterson AFB, OH, 45433, USA

Abstract: A new method for modeling functions that intersect given points is developed and demonstrated. This method yields a generally non-Gaussian probability density of y given x that has properties which are often desired in practice. It is shown that this density can have a smoother mean function and a variance which is never larger than that of a classic Gaussian process density.

INTRODUCTION

Suppose that many functions are chosen which intersect given (x, y) points and which model the process that generated the points. Also suppose that each function is a line plus Gaussian radial kernels of the same variance centered at the points, which is a classic choice. Further suppose that each function has a probability of occurrence such that the probability density of function y values upon extrapolation (i.e., at any large-magnitude x) has a mean on the least squares line of the points, which is also a classic choice. Finally suppose, as is clearly reasonable, that each function is as smooth as possible, where smoothness has the classic inverse integrated squared second derivative definition. Then it is shown here that these straightforward suppositions require the probability density of the function y values at any x to be, in general, non-Gaussian.

Thus the density of y given x cannot be produced by a Gaussian process, and an alternative to these classic processes is required. Here, for the first time to our knowledge, a practical alternative, called "robust cardinal interpolation", is developed and demonstrated. Also, and most importantly, it is shown that the robust cardinal interpolation density can have a smother mean function and a variance which is never larger than that of a classic Gaussian process density.

There is extensive literature on interpolation methods that extends over decades and even centuries. However, only in the past several decades have methods been proposed that specify a probability density of many interpolating functions rather than a single interpolating function. These methods, including spline and neural network methods, are typically special cases of Gaussain process methods [1], which, however, can only yield Gaussaian densities of y given x. Also, Gaussian radial kernels are commonly used to form basis functions for modeling [2], as is the case here for robust cardinal interpolation.

A recent paper [3] considers non-Gaussian interpolation; however, it differs from the method developed here in that it does not employ a robust optimization which requires that each function in the ensemble which forms the density be as smooth as possible. In particular, the cardinal interpolation in [3] is non-robust in that it requires each basis function to have a specified roughness that is not the least possible roughness.

DESCRIPTION OF ROBUST CARDINAL INTERPOLATION

A description of robust cardinal interpolation is as follows:

Given data points $D = \{(x_i, y_i), i = 1, 2, \ldots, n\}$ with $\Sigma x_i = 0$, $\Sigma y_i = 0$, $\Sigma x_i^2 = n$, $\Sigma y_i^2 = n$, and $\Sigma x_i y_i = 0$, the robust cardinal interpolation (RCI) probability density is $p(y \| x, D) \sim \iiint \delta[y - f(x| a, b, c)]\, w(a, b, c)\, da\, db\, dc$ with $w(a, b, c) = \exp\{-[\Sigma(y_i - a - bx_i)^2 + c^2]/2\}$ and $f(x| a, b, c) = a + bx + c + \Sigma A_i \exp[-(x - x_i)^2/(2s^2)]$, where the A_i are such that $y_i = f(x_i| a, b, c)$, s is such that $\int f''^2(x|a, b, c)\, dx$ is minimized, and all integrals are from $-\infty$ to $+\infty$.

Comments on this description are as follows:

(1) The data points are normalized to have zero mean and unit variance in both x and y and, due to $\Sigma x_i y_i = 0$, to be such that their least squares line is the x axis; this normalization can obviously be accomplished for any points that have distinct x_i. It is also apparent that the determination of s for each f(x| a, b, c) involves a nonlinear optimization which depends on the normalization.

(2) The weight w(a, b, c) is such that upon extrapolation, i.e., at large magnitude x where f(x| a, b, c) is linear, the mean of the ensemble of f(x| a, b, c) functions is the least squares line of the points (and is therefore the x axis) and the variance of the ensemble is the classic quadratic function for the Gaussian linear model density (as demonstrated below in the section on properties). Thus this choice is made to ensure a required property for the RCI density.

(3) For given a, b, and c and a specified s, the A_i are determined so that f(x| a, b, c) interpolates (i.e., intersects) the points, and this determination requires the solution of n linear equations in n unknowns (as demonstrated below in the section on properties). Here s is determined by varying its value and recalculating

*Address correspondence to this author at the Air Force Institute of Technology, AFIT/ENG, Wright-Patterson AFB, OH, 45433, USA;
E-mail: steven.gustafson@afit.edu

1874-110X/08 2008 Bentham Open

the A_i until the resulting f(x| a, b, c) minimizes $\int f''^2(x|a, b, c)\, dx$, which is a classic definition of roughness. Thus the determination of s, unlike the determination of the A_i, is nonlinear.

(4) The RCI density may be computed numerically as $p(y\| x, D) = \text{histogram}_{a,b,c}\{ f(x| a, b, c), w(a, b, c)\}$, with $\{a, b, c\} = -L, -L + \varepsilon, \ldots, L$, where $\varepsilon << 1 << L$. Here the histogram operation involves specifying an ensemble function f(x| a, b, c), assigning this function a weight w(a, b, c), repeating for many sets of a, b, and c values, and, at any x, generating a histogram of the weighted y values of the ensemble functions. This procedure is valid because in the limit of small ε and large L it evaluates the triple integral that formally describes the RCI. It implements a form of the Markov chain Monte Carlo procedures [4] that have contributed to the practical implementation of Bayesian methods.

(5) Computation of the RCI density involves, for each ensemble function, specifying a candidate s, solving n linear equations in n unknowns to obtain the A_i repeating for many s values, and determining the s that

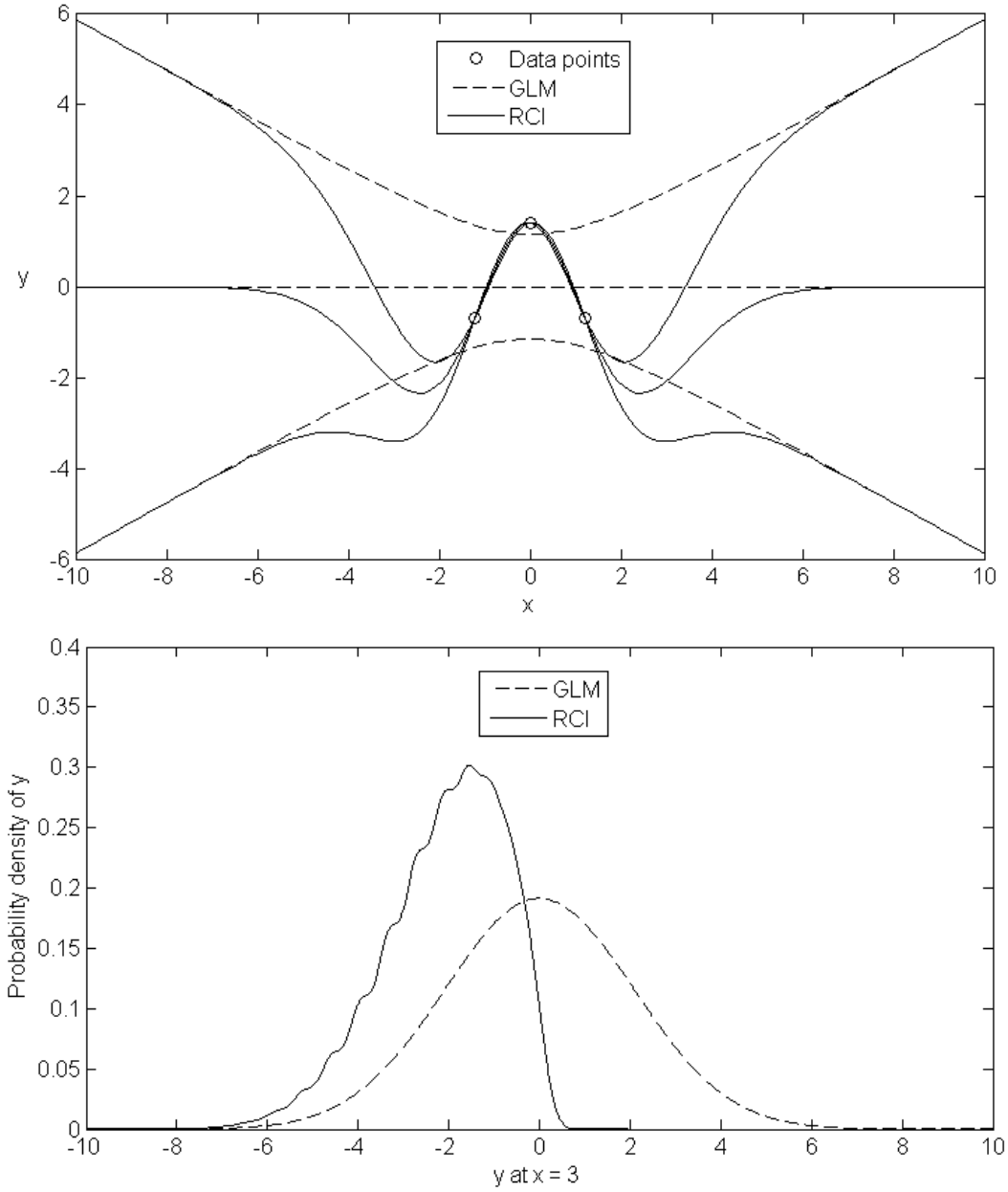

Fig. (1). The mean function and the mean function plus and minus the standard deviation function for the Gaussian linear model (GLM) density and for the robust cardinal interpolation (RCI) density for three symmetrically positioned data points; also, the GLM and RCI densities at x =3. The points are equally spaced in x and are normalized to have zero mean and unit variance in both x and y and to have the x axis as their least squares line. Note that both densities have the same mean and standard deviation for large magnitude x, thus illustrating the extrapolation property of the RCI density. Note also that the mean function of the RCI density interpolates the points and that the standard deviation function is zero at the points, thus illustrating the interpolation property of the RCI density. Finally, note that at x = 3 the RCI density is obviously asymmetric and thus non-Gaussian, as it is, in general, at all x which are not at the points or not large magnitude.

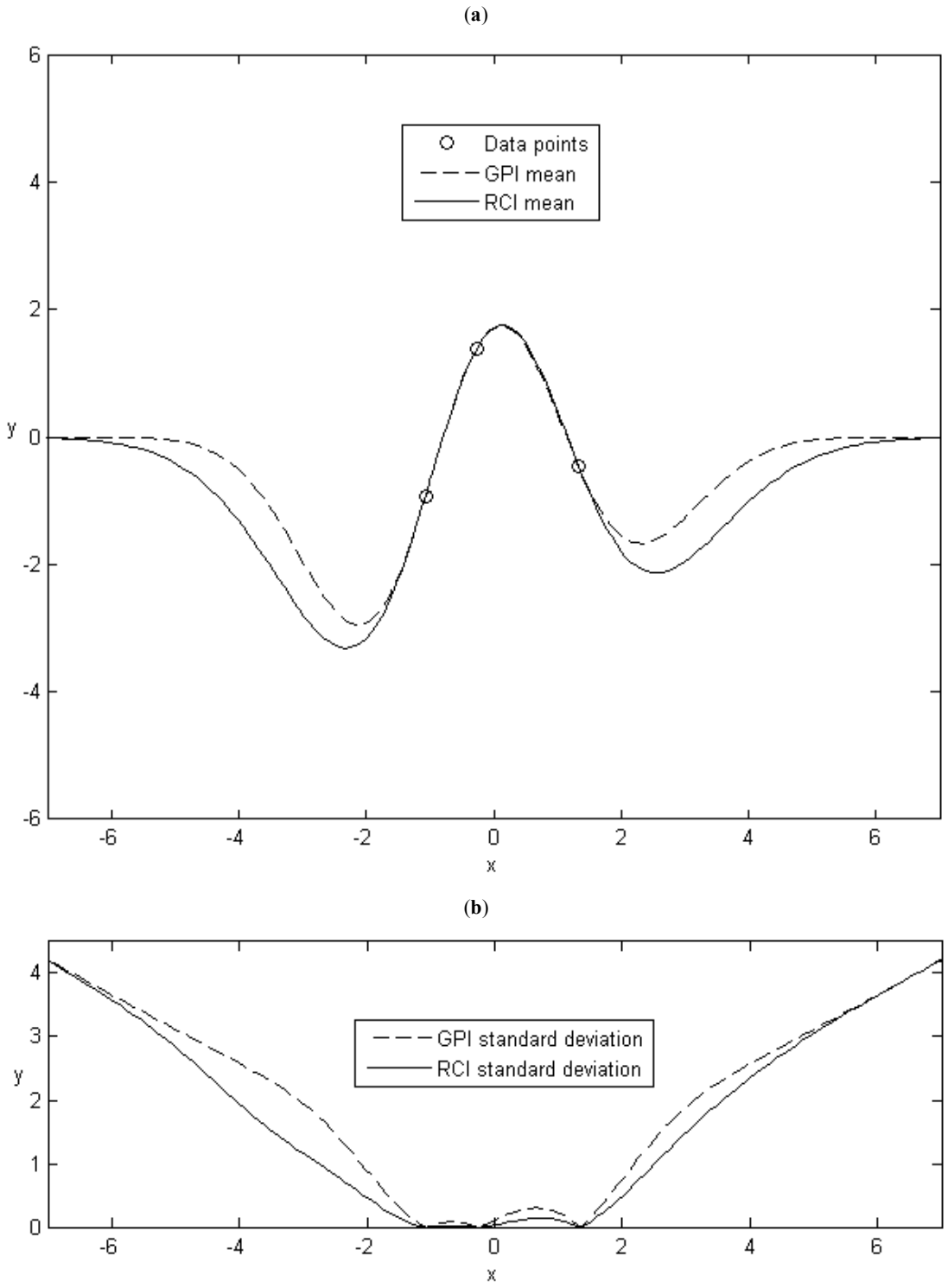

Fig. (2). The mean function and the standard deviation function for a Gaussian process interpolation (GPI) density and the RCI density for three asymmetrically positioned data points. The three points are such the x spacing of the first two points is half that of the last two points, and they are normalized to have zero mean and unit variance in both x and y and to have the x axis as their least squares line. The mean function of the RCI density is less rough than that of the GPI density, and, as shown in the figure, the standard deviation of the RCI density is not larger than that of the GPI density for any x.

minimizes roughness. In practice, linear equation solving is typically performed less than 30 times for each ensemble function, and less than 30 ensemble functions are employed to form histograms of the RCI density for any x. This "brute force" computational approach (more efficient approaches may ultimately be found) thus requires solving n linear equations in n unknowns less than 900 times. Computation time for n less than 30 typically requires a few hours is thus not excessive for non-real-time applications.

PROPERTIES OF ROBUST CARDINAL INTERPOLATION

Properties of the RCI density with comments on their origin and significance are as follows:

(1) The extrapolation property is that $\lim_{|x|\to\infty} p(y\| x, D)$ is Gaussian with a mean that is the x axis and a variance that is the quadratic $1 + 1/n + x^2/n$. This Gaussian linear model (GLM) density is the classic result obtained for a linear model, i.e., the mean is the least squares line of the points, and the variance is a quadratic function that increases with the magnitude of x. The extrapolation property results because $f(x| a, b, c) = a + bx + c$ for large magnitude x, and with this linear form the triple integral that defines the RCI density may be evaluated analytically to obtain the GLM density. Fig. (**1**) shows, for three symmetrically-placed points, the mean function and the mean function plus and minus the standard deviation function for the GLM density and for the RCI density. Note that both densities have the same mean and standard deviation for large magnitude x, thus illustrating the extrapolation property.

(2) The interpolation property is that $p(y\| x_i, D) = \delta(y - y_i)$. Thus at $x = x_i$ the RCI density has a mean that is y_i and a variance that is zero. The interpolation property results because each function $f(x| a, b, c)$ in the ensemble of functions that forms the RCI density intersects the data points, i.e., $y_i = f(x_i| a, b, c)$. Fig. (**1**) shows that the mean function of the RCI density interpolates the points and that the standard deviation function is zero at the points, thus illustrating the interpolation property.

(3) The RCI density is non-Gaussian, except for $x = x_i$ and $|x|\to\infty$, in that $R\{M\{p\}\} < R\{E\{p\}\}$, where $R\{\ \} = \int\{\ \}''^2\, dx$ is roughness, $M\{\ \}$ is mode versus x, and $E\{\ \}$ is mean versus x. This property states that $p(y\| x, D)$ is generally non-Gaussian at any x that is not at a data point or that does not have large magnitude. The property is expressed in terms of the mean versus x and the mode versus x, where the mode is the value of y at which the density has its maximum and is thus the most probable value of y. For a Gaussian density of y given x the mode equals the mean, and thus the mode function has the same roughness as the mean function. However, for the RCI density the mode function is smoother than the mean function, so the RCI density is not symmetric but is skewed, as shown in Fig. (**1**) such that the most probable values of y versus x form a less-rough function than the mean values of y versus x. This property is intuitively reasonable and results because the roughness of each function $f(x| a, b, c)$ in the ensemble is individually minimized by choosing an optimal s subject to the condition that the function interpolates the points.

(4) The RCI density can have a smoother mean function and a never-larger variance than a classic Gaussian process interpolation (GPI) density [2]. This classic GPI density $q(y\| x, D)$ has zero mean and covariance matrix elements $c_{ij} = \exp[-(x_i - x_j)^2/(2t^2)]\{4/3 + [(x_i + x_j)/2]^2/3\}$, where i, j = 1, 2,…, n. This density has properties (1) and (2) above, however, for the points in Fig. (**2**), which are asymmetrically positioned to illustrate a general case, $R\{E\{p\}\} < R\{E\{q\}$, even if, as chosen here, t is such that $R\{E\{q\}\}$ is minimized. Also, as shown in the figure, $V\{p\} \le V\{q\}$, where $V\{\ \}$ is variance versus x. Thus compared to the GPI density, the RCI density can have a less-rough mean function and a not-smaller variance at any x.

CONCLUSION

Robust cardinal interpolation is developed as a method for generating a probability density that models functions which intersect given points. This probability density (1) extrapolates to the classic Gaussian linear model density with a least squares line mean function and a quadratic variance function, (2) has a mean function that intersects the points and a variance function that is zero at the points, and (3) is non-Gaussian in that its mode function is less rough than its mean function, where roughness is integrated squared second derivative. It is shown that, compared to a classic Gaussian process interpolation density, the robust cardinal interpolation density has a less-rough mean function and a variance function that is never larger.

ACKNOWLEDGMENTS

The views expressed in this paper are those of the authors and do not reflect the official policies of position of the United States Air Force, US Department of Defense, or the US Government.

REFERENCES

[1] C. Rasmussen and C.K.I. Williams, *Gaussian Processes for Machine Learning,* MIT Press, 2006. Available at: www.gaussianprocess.org.

[2] C.M. Bishop, Pattern Recognition and Machine Learning, Springer 2006, and D.J.C. MacKay, *Information Theory, Inference, and Learning Algorithms*, Cambridge, 2003.

[3] S.C. Gustafson, D.R. Parker, and R.K. Martin, "Cardinal Interpolation", *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, pp. 1538-1545, September 2007.

[4] A.E. Gelfand and A.F.M. Smith, "Sampling-Based Approaches to Calculating Marginal Densities", *J. Am. Stat. Assoc.*, vol. 85, pp. 398-409, 1990.

Received: June 2, 2008 Revised: June 28, 2008 Accepted: July 3, 2008