

DESIGN OF A ROBUST PARAMETER ESTIMATOR
FOR NOMINALLY LAPLACIAN NOISE

A Thesis

by

PANKAJ BHAGAWAT

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

August 2003

Major Subject: Electrical Engineering

DESIGN OF A ROBUST PARAMETER ESTIMATOR
FOR NOMINALLY LAPLACIAN NOISE

A Thesis

by

PANKAJ BHAGAWAT

Submitted to Texas A&M University
in partial fulfillment of the requirements
for the degree of

MASTER OF SCIENCE

Approved as to style and content by:

Don R. Halverson
(Chair of Committee)

Costas N. Georghiades
(Member)

Kai Chang
(Member)

Donald Friesen
(Member)

Channan Singh
(Head of Department)

August 2003

Major Subject: Electrical Engineering

ABSTRACT

Design of a Robust Parameter Estimator

for Nominally Laplacian Noise. (August 2003)

Pankaj Bhagawat, B.E, Regional Engineering College Trichy(India)

Chair of Advisory Committee: Dr. Don R. Halverson

In this work we have made use of a geometric approach which quantifies robustness and performance and we finally combine them using a cost function. In particular, we calculate the robustness of the estimate of standard deviation of nominally Laplacian distribution. As this distribution is imperfectly known, we employ a more general family, the generalized Gaussian; Laplacian distribution, is one of the members of this family. We compute parameter estimates and present a classical algorithm which is then analyzed for distribution from the generalized Gaussian family. We calculate the mean squared error according to the censoring height k . We measure performance as a function of $(1/\text{MSE})$ and combine it with robustness using a cost criterion and design a robust estimator which optimizes a mix of performance and robustness specified by the user.

To my Family, Friends and Teachers

ACKNOWLEDGMENTS

I am grateful to Dr.Halverson for his advice and guidance. I thank him for the support and encouragement he gave me during this research. He was always available when I sought his help. I also take the opportunity to thank my committee members Dr.Georghiades, Dr.Chang and Dr.Friesen. I would also like to express my personal gratitude to Dr.Halverson and Dr.Georghiades for their help in my academic endeavors. I am also thankful to Ms.Tammy Carda for taking care of all the administrative work. I am always indebted to my family for their encouragement and support.

TABLE OF CONTENTS

CHAPTER		Page
I	INTRODUCTION	1
	A. Parameter Estimation	1
	B. Why Study Robust Estimation Now?	2
	C. Past Work	3
	D. Overview of the Thesis	4
II	MAXIMUM LIKELIHOOD ESTIMATOR(ML)FOR LAPLA- CIAN DATA	5
	A. ML Estimator for Laplacian Data	5
	B. Efficiency Check for the Estimate	6
III	COMPARISON OF ESTIMATORS	7
	A. General Robust Parameter Estimation	7
	B. Implementation Details	11
	C. Mean Square Error Performance	12
	1. Laplacian Distribution	13
	2. Distribution of the Random Variable \mathbf{A}	15
IV	SUMMARY	17
	A. MSE as a Function of a	17
	B. Cost Function J , and Choices of α and η	17
	C. Conclusions	88
	REFERENCES	89
	VITA	90

LIST OF TABLES

TABLE		Page
I	k as a function of ϵ , which maximizes J for $\sigma^2 = 2$, and $\alpha = 0.01$. .	52
II	k as a function of ϵ , which maximizes J for $\sigma^2 = 2$, and $\alpha = 0.02$. .	53
III	k as a function of ϵ , which maximizes J for $\sigma^2 = 2$, and $\alpha = 0.03$. .	54
IV	k as a function of ϵ , which maximizes J for $\sigma^2 = 2$, and $\alpha = 0.04$. .	55
V	k as a function of ϵ , which maximizes J for $\sigma^2 = 2$, and $\alpha = 0.05$. .	56
VI	k as a function of ϵ , which maximizes J for $\sigma^2 = 2$, and $\alpha = 0.06$. .	57
VII	k as a function of ϵ , which maximizes J for $\sigma^2 = 2$, and $\alpha = 0.07$. .	58
VIII	k as a function of ϵ , which maximizes J for $\sigma^2 = 2$, and $\alpha = 0.08$. .	59
IX	k as a function of ϵ , which maximizes J for $\sigma^2 = 2$, and $\alpha = 0.09$. .	60
X	k as a function of ϵ , which maximizes J for $\sigma^2 = 2$, and $\alpha = 0.3$. . .	61
XI	k as a function of ϵ , which maximizes J for $\sigma^2 = 2$, and $\alpha = 0.5$. . .	62
XII	k as a function of ϵ , which maximizes J for $\sigma^2 = 2$, and $\alpha = 0.9$. . .	63
XIII	k as a function of ϵ , which maximizes J for $\sigma^2 = 1$, and $\alpha = 0.01$. .	64
XIV	k as a function of ϵ , which maximizes J for $\sigma^2 = 1$, and $\alpha = 0.02$. .	65
XV	k as a function of ϵ , which maximizes J for $\sigma^2 = 1$, and $\alpha = 0.03$. .	66
XVI	k as a function of ϵ , which maximizes J for $\sigma^2 = 1$, and $\alpha = 0.04$. .	67
XVII	k as a function of ϵ , which maximizes J for $\sigma^2 = 1$, and $\alpha = 0.05$. .	68
XVIII	k as a function of ϵ , which maximizes J for $\sigma^2 = 1$, and $\alpha = 0.06$. .	69
XIX	k as a function of ϵ , which maximizes J for $\sigma^2 = 1$, and $\alpha = 0.07$. .	70

TABLE	Page
XX	k as a function of ϵ , which maximizes J for $\sigma^2 = 1$, and $\alpha = 0.08$. . . 71
XXI	k as a function of ϵ , which maximizes J for $\sigma^2 = 1$, and $\alpha = 0.09$. . . 72
XXII	k as a function of ϵ , which maximizes J for $\sigma^2 = 1$, and $\alpha = 0.3$. . . 73
XXIII	k as a function of ϵ , which maximizes J for $\sigma^2 = 1$, and $\alpha = 0.5$. . . 74
XXIV	k as a function of ϵ , which maximizes J for $\sigma^2 = 1$, and $\alpha = 0.9$. . . 75
XXV	k as a function of ϵ , which maximizes J for $\sigma^2 = 0.5$, and $\alpha = 0.01$. . . 76
XXVI	k as a function of ϵ , which maximizes J for $\sigma^2 = 0.5$, and $\alpha = 0.02$. . . 77
XXVII	k as a function of ϵ , which maximizes J for $\sigma^2 = 0.5$, and $\alpha = 0.03$. . . 78
XXVIII	k as a function of ϵ , which maximizes J for $\sigma^2 = 0.5$, and $\alpha = 0.04$. . . 79
XXIX	k as a function of ϵ , which maximizes J for $\sigma^2 = 0.5$, and $\alpha = 0.05$. . . 80
XXX	k as a function of ϵ , which maximizes J for $\sigma^2 = 0.5$, and $\alpha = 0.06$. . . 81
XXXI	k as a function of ϵ , which maximizes J for $\sigma^2 = 0.5$, and $\alpha = 0.07$. . . 82
XXXII	k as a function of ϵ , which maximizes J for $\sigma^2 = 0.5$, and $\alpha = 0.08$. . . 83
XXXIII	k as a function of ϵ , which maximizes J for $\sigma^2 = 0.5$, and $\alpha = 0.09$. . . 84
XXXIV	k as a function of ϵ , which maximizes J for $\sigma^2 = 0.5$, and $\alpha = 0.3$. . . 85
XXXV	k as a function of ϵ , which maximizes J for $\sigma^2 = 0.5$, and $\alpha = 0.5$. . . 86
XXXVI	k as a function of ϵ , which maximizes J for $\sigma^2 = 0.5$, and $\alpha = 0.9$. . . 87

LIST OF FIGURES

FIGURE		Page
1	distribution function	8
2	MSE vs a	10
3	MSE vs a , for $\sigma=1.0$, $k=\infty$	14
4	$\sigma^2 = 2$, $n=2$	19
5	$\sigma^2 = 2$, $n=3$	20
6	$\sigma^2 = 2$, $n=4$	21
7	$\sigma^2 = 2$, $n=5$	22
8	$\sigma^2 = 2$, $n=6$	23
9	$\sigma^2 = 2$, $n=7$	24
10	$\sigma^2 = 2$, $n=8$	25
11	$\sigma^2 = 2$, $n=9$	26
12	$\sigma^2 = 2$, $n=10$	27
13	$\sigma^2 = 2$, $n=15$	28
14	$\sigma^2 = 2$, $n=20$	29
15	$\sigma^2 = 1$, $n=2$	30
16	$\sigma^2 = 1$, $n=3$	31
17	$\sigma^2 = 1$, $n=4$	32
18	$\sigma^2 = 1$, $n=5$	33
19	$\sigma^2 = 1$, $n=6$	34

FIGURE		Page
20	$\sigma^2 = 1, n=7$	35
21	$\sigma^2 = 1, n=8$	36
22	$\sigma^2 = 1, n=9$	37
23	$\sigma^2 = 1, n=10$	38
24	$\sigma^2 = 1, n=15$	39
25	$\sigma^2 = 1, n=20$	40
26	$\sigma^2 = 0.5, n=2$	41
27	$\sigma^2 = 0.5, n=3$	42
28	$\sigma^2 = 0.5, n=4$	43
29	$\sigma^2 = 0.5, n=5$	44
30	$\sigma^2 = 0.5, n=6$	45
31	$\sigma^2 = 0.5, n=7$	46
32	$\sigma^2 = 0.5, n=8$	47
33	$\sigma^2 = 0.5, n=9$	48
34	$\sigma^2 = 0.5, n=10$	49
35	$\sigma^2 = 0.5, n=15$	50
36	$\sigma^2 = 0.5, n=20$	51

CHAPTER I

INTRODUCTION

A. Parameter Estimation

Given X_1, X_2, \dots, X_n i.i.d. F , F belonging to a known family of distributions, the statistician is interested in estimating some descriptive measure of the distribution $\theta = T(F)$. In order to improve, if possible, on the naive choice $T(F_n)$, where F_n is the empirical distribution, he will derive the distributions of $T(F_n)$ and a number of possible competing estimators and use these distributions as a basis for his choice. Usually he brings in additional criteria such as invariance, minimax risk, and asymptotic efficiency to help simplify the selection problem, but what remains are problems often requiring great ingenuity and mathematical sophistication. Solving such problems can give the mathematical statistician a lifetime of mental pleasure and professional status, but his solutions, i.e., proposed estimators, are not of much use to the data analyst if they are extremely sensitive to slight changes in the stated assumptions. The actual observations X_1, X_2, \dots, X_n from the series of repetitions of an experiment may not be quite independent and may well involve round-off errors and occasional gross errors. Any proposed estimator should be robust in that it is insensitive to such slight changes in the underlying model. That very few people today agree in practice with the last statement is confirmed by the fact that the two most widely used descriptive statistics are the sample mean and standard deviation.

The journal model is *IEEE Transactions on Automatic Control*.

B. Why Study Robust Estimation Now?

Robust estimation has been employed for hundreds of years [1] but not seriously studied until recently. Parametric results began to be questioned in the late 1950's when asymptotic methods developed to the point where the inefficiency of the t-test to normal scores and Wilcoxon tests could be demonstrated [2]. The sensitivity of some classical estimators was documented by Tuckey [3] (1960). Other factors also contribute to the growing interest in the "robustness". A comprehensive list of them and a provocative introduction to robust methods is available.

We merely list a few main points.

1. With the coming of the computer age much data is machine recorded and passed through statistical packages: it is free of human subjectivity along the way. This means it is more likely to contain outliers and other impurities previously removed by human hands. The data is available in larger quantities, which discourages the data analyst from looking at it in too much detail.

2. While computers have changed the nature of data, they have also helped the statistician by allowing him to calculate complicated formulas in a thousandth of the time previously required. They have enabled him to conduct Monte Carlo simulations as part of an investigation.

3. More statisticians are familiar with the rudiments of functional analysis, which are helpful in defining and attacking the problems of robust estimation from a theoretical point of view.

4. Finally, the concepts introduced so far to study robustness have an intuitive appeal and can give statistical insight about an estimator whether used heuristically or as mathematical constructs.

C. Past Work

Many of the techniques used in telecommunications, image processing, speech processing, optical communications and adaptive antenna rely on various amounts of parameter estimation. To measure overall system performance, the measuring degree of robustness is very important. A considerable contribution to robustness using saddle point criteria has been made by Huber [4,5]. However, the classical saddle point criteria have a disadvantage of being non-quantitative and there is no direct way to make comparisons between various estimators. In this work we have made use of a geometric approach which quantifies robustness and performance and we finally combine them using a cost function (decided by the user). In particular, we calculate the robustness of the estimate of standard deviation of nominally Laplacian distribution, which is one of the important estimators of practical interest [4, 5, 6, 7]. Because this distribution is imperfectly known, we employ a more general family, the generalized Gaussian. The probability distribution function for this noise is given by

$$f(x) = de^{-\frac{|x|^a}{c}} \quad (1.1)$$

This yields the Laplacian for $a=1$.

We compute parameter estimates and present a classical algorithm which is then analyzed for distribution from the generalized Gaussian family. We calculate mean squared error(MSE) according to the censoring height k , including the maximum likelihood (ML) case when $k= \infty$. We measure performance as a function of $(1/\text{MSE})$ and combine it with robustness using a cost criterion and design a robust estimator which optimizes a mix of performance and robustness specified by the user. In other words, we illustrate a trade off between performance and robustness for various estimation schemes.

D. Overview of the Thesis

To evaluate robustness and performance of the estimator, we plot MSE vs. a (for a particular value of standard deviation), where a varies within a range, say from 0.1 to 2.0. The value of MSE at these points is the measure of performance, and the slope at a point is the measure of robustness of the estimator at that point. To get overall robustness and performance we need to combine the performance and robustness parameters across the whole range of a and for this we can model the value of the exponent a as a random variable \mathbf{A} . We can introduce several options for the distribution of \mathbf{A} , and from this we can define mean performance P from MSE, as well as mean robustness R by making use of slope. Finally we can combine the two parameters P and R to form a composite cost function employing a weighting factor deciding which is more important, robustness or the performance; this factor will be chosen by the user depending on his/her application. This approach can be repeated for various values of standard deviations and various values of censoring value (k) for each value of standard deviation.

CHAPTER II

MAXIMUM LIKELIHOOD ESTIMATOR(ML)FOR LAPLACIAN DATA

The probability distribution function for Laplacian noise is given by

$$f(x) = de^{-\frac{|x|^a}{c}} \quad (2.1)$$

where

$$c = \frac{\sigma}{\sqrt{2}} \text{ and } d = \frac{1}{\sqrt{2}\sigma} \quad (2.2)$$

where σ is the standard deviation. In this work we have explored how the estimator performs for various values of a .

A. ML Estimator for Laplacian Data

For Laplacian distribution, its unbiased estimator can be found as follows. Laplacian distribution can be represented as

$$f(x) = \frac{1}{\sqrt{2}\sigma} e^{-\frac{\sqrt{2}}{\sigma}|x|} \quad (2.3)$$

where σ is the standard deviation. Let f be the density of i.i.d noise and let there be n samples. Then the joint pdf will be given by

$$f(x_1, x_2, \dots, x_n) = \left[\frac{1}{\sqrt{2}\sigma}\right]^n \prod_{i=1}^n e^{-\frac{\sqrt{2}}{\sigma}|x_i|} \quad (2.4)$$

To find ML estimate of the standard deviation we solve the following equation

$$\frac{\partial \ln f(x_1, x_2, \dots, x_n)}{\partial \sigma} = 0 \quad (2.5)$$

$$\ln f(\mathbf{x}, \sigma) = \sum_{i=1}^n \ln(f(x_i, \sigma)) \quad (2.6)$$

$$= \sum_{i=1}^n \ln\left(\frac{1}{\sqrt{2}\sigma} e^{-\frac{\sqrt{2}}{\sigma}|x_i|}\right) \quad (2.7)$$

$$= n \cdot \ln\left(\frac{1}{\sqrt{2}}\right) - n \cdot \ln(\sigma) - \frac{\sqrt{2}}{\sigma} \sum_{i=1}^n |x_i| \quad (2.8)$$

Substituting (2.8) in (2.5), we get,

$$-\frac{n}{\hat{\sigma}} + \frac{\sqrt{2}}{\hat{\sigma}^2} \cdot \sum_{i=1}^n |x_i| = 0 \quad (2.9)$$

The ML estimate for the standard deviation is therefore,

$$\hat{\sigma} = \frac{\sqrt{2}}{n} \sum_{i=1}^n |x_i| \quad (2.10)$$

In order to impart robustness to above the estimator we censor it at $x_i=k$, or mathematically we have

$$|x_i| = \begin{cases} |x_i|, & \text{if } |x_i| \leq k \\ k, & \text{if } |x_i| > k \end{cases} \quad (2.11)$$

B. Efficiency Check for the Estimate

Since the estimate is obtained through the ML estimation algorithm, we can check for the efficiency of the estimate by seeing if the expectation of the estimate is unbiased.

The estimate for a single-dimensional case is,

$$\hat{\sigma} = \sqrt{2}|x| \quad (2.12)$$

Using (2.3) and (2.12), we get,

$$E[\hat{\sigma}(x)] = \int_{-\infty}^{+\infty} \sqrt{2}|x| \frac{1}{\sqrt{2}\sigma} e^{-\frac{\sqrt{2}}{\sigma}|x|} dx \quad (2.13)$$

$$= \frac{2}{\sigma} \int_0^{+\infty} x e^{-\frac{\sqrt{2}}{\sigma}x} dx \quad (2.14)$$

$$= \sigma \quad (2.15)$$

CHAPTER III

COMPARISON OF ESTIMATORS

A. General Robust Parameter Estimation

For the signal estimator under consideration, let P denote its performance measure. For example, P could be mean square error or mean absolute error function. To determine a measure of robustness for processor, it is necessary to admit perturbations in the distribution about the nominal and observe how these perturbations affect performance. First of all, we start from a distribution of the following form (see Figure 1) because it suffices whenever the performance measure involves appropriate integration:

$$F_i = \sum_{j=1}^{m_i} a_{i,j} I_{A_{i,j}}(\cdot) \quad (3.1)$$

$i=1, \dots, n$, where n is the number of independent samples, where $I_A(\cdot)$ is the indicator function of set A , where $a_{i,j}$ are positive real numbers between 0 and 1, where $A_{i,j}$ are closed intervals which generate a rectilinear partition P of R^n and where

$$I_{A_{i,j}}(x) = \begin{cases} 1, & x \in A_{i,j} \\ 0, & \text{otherwise} \end{cases} \quad (3.2)$$

Note that the value of $a_{i,j}$ is the height of $F_i(\cdot)$ within the partitioning interval $A_{i,j}$. These "step function" distributions provide a sufficient variety of perturbations from the nominal, provided that we restrict our attention to performance measures based on Steiltjes integration.

Consider parameter estimation of a parameter θ and estimator $\hat{\theta} = g(\cdot, \dots, \cdot)$ with performance criterion $E\{Q(\theta - \hat{\theta})\}$. There are n independent samples, expression of

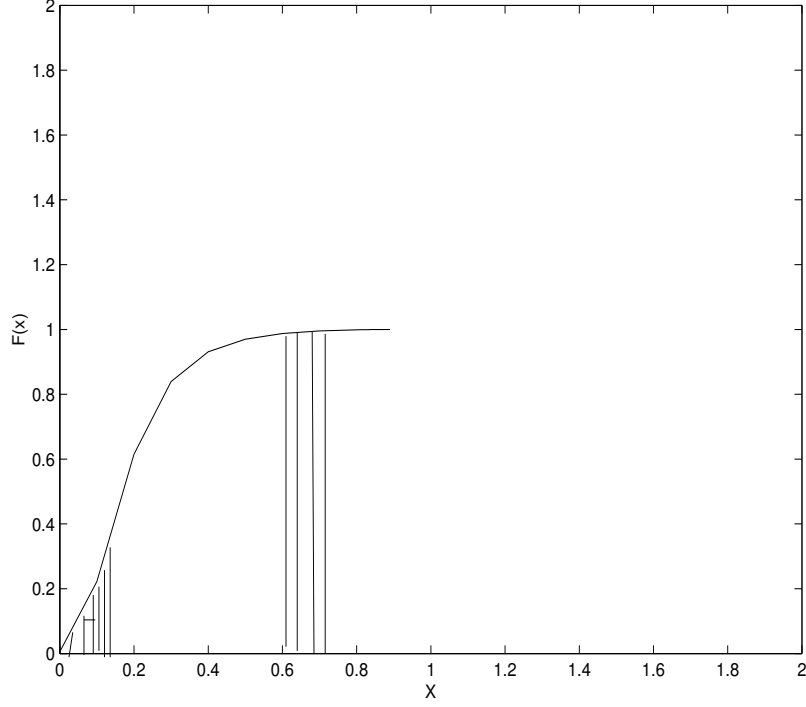


Fig. 1. distribution function

the above distribution, we can express the estimator performance criterion P by:

$$P = \sum_{j_n=1}^{m_n} \dots \sum_{j_i=1}^{m_i} \dots \sum_{j_1=1}^{m_1} \prod_{i=1}^n (a_{i,j_i} - a_{i,j_{i-1}}) \cdot Q(\theta - g(x_{1,j_1}, \dots, x_{i,j_i}, \dots, x_{n,j_n})), \quad (3.3)$$

where $Q(\cdot)$ is the error criterion, θ is the parameter and $g(\cdot, \dots, \cdot)$ is the estimator.

The variability of performance P , as the parameters $A_{i,j}$ vary about their nominal values, is reflected by $\frac{\partial P}{\partial a_{i,j}}$. The normalized first order robustness measure can also be viewed as

$$\phi = \frac{1}{1 + \Delta} \quad (3.4)$$

and Δ is

$$\Delta = \lim_{\max_i |p_i| \rightarrow 0} \sup_{P_i} \sum_{i=1}^n \left| \frac{\partial P}{\partial a_i} \right| \quad (3.5)$$

where each p_i partitions the real line It follows from [7] that we have

$$\Delta = \sum_{i=1}^n \int_{-\infty}^{\infty} \left| \frac{\partial}{\partial x_i} \int_{R^{n-1}} Q(\theta - g(x_1, \dots, x_i, \dots, x_n)) dF_i(x_i) * \dots dF_n(x_n) \right| dx_i \quad (3.6)$$

where $F(\cdot)$ is the nominal distribution for each i and the superscript* denotes omission of that term. For

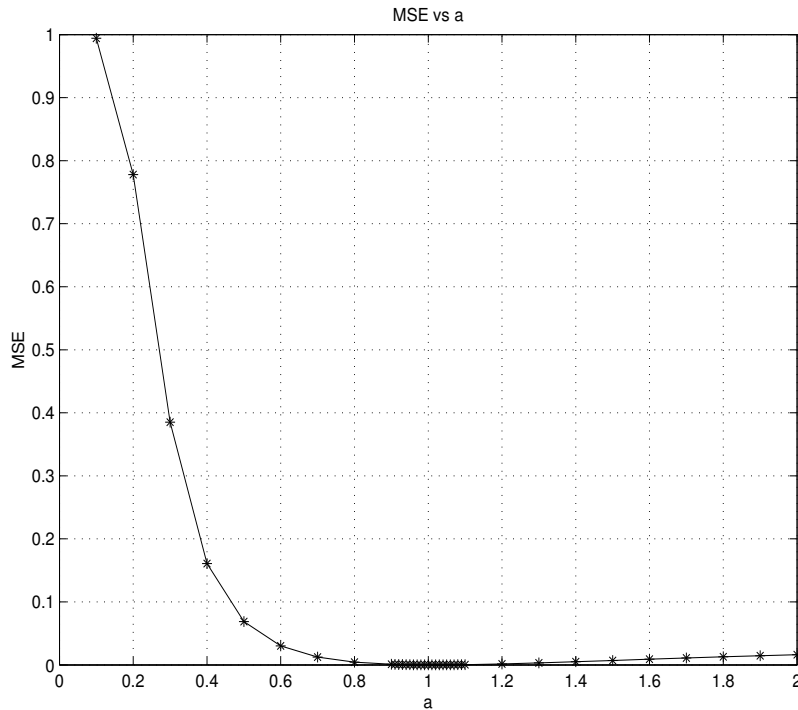
$$g(x_1, x_2, \dots, x_n) = \frac{1}{n} \sum_{j=1}^n h(x_j) \quad (3.7)$$

we obtain

$$\Delta = \sum_{i=1}^n \int_{-\infty}^{\infty} \left| \frac{\partial}{\partial x_i} \int_{R^{n-1}} \left(\theta - \frac{1}{n} \sum_{j=1}^n h(x_j) \right)^2 dF_i(x_i) * \dots dF_n(x_n) \right| dx_i \quad (3.8)$$

These new techniques can best be understood by noting that an estimator generates a multidimensional surface, as function of the underlying joint distribution, whose "shape" provides an indication of robustness. For example, like Fig. 2, an estimator whose performance surface resembles an inverted mountain, with its peak occurring at a certain distribution away from the nominal distribution, might be non-robust perturbations about the nominal. In other words, robust processor might generate a performance surface resembling an inverted plateau, or an inverted elevated plain. for an estimator of this type, a shift away from the nominal distribution will only bring about a slight change in performance.

The estimators under consideration are derived from, whenever possible, unbiased maximum likelihood estimators which may not be robust. In order to impart robustness, Huber-type data censoring is employed. In accordance with the original estimator being unbiased for the case of Gaussian, Laplacian, or Rayleigh data, in some situations variance is estimated, while in other situations it is standard deviation. As an example application of robust estimation measure ϕ previously described,

Fig. 2. MSE vs a

consider the estimation of the parameter $\theta = \frac{1}{n} \sum_{j=1}^n E\{X_j\}$, where the data may be non-stationary but independent.

The above mentioned technique can be easily applied to Laplacian distributed noise. Recall that the Laplacian noise in general can be represented as

$$f(x) = de^{-\frac{|x|^a}{c}} \quad (3.9)$$

Where $c = \frac{\sqrt{2}}{\sigma}$, $d = \frac{1}{\sqrt{2}\sigma}$, σ is the standard deviation and $a=1$. As mentioned earlier, the ML estimator for Laplacian distribution is given by:

$$\hat{\sigma} = \frac{\sqrt{2}}{n} \sum_{i=1}^n |x_i| \quad (3.10)$$

In order to impart robustness to above estimator with respect to perturbations in the

distribution we censor it at $x_i=k$, mathematically we have

$$\tilde{h}(x_i) = \begin{cases} |x_i|, & \text{if } |x_i| \leq k \\ k, & \text{if } |x_i| > k \end{cases} \quad (3.11)$$

We then let $h(x_i) = \sqrt{2\tilde{h}(x_i)}$. Doing this will ensure that the change in performance, as the distribution deviates from being Laplacian, i.e., when a perturbs from unity, will be reduced, and thus impart robustness. However, this might also lead to higher amount of MSE being introduced; this essentially means that there is a trade-off between robustness R and performance P of the estimator. The value of censor k will decide the degree of robustness and performance of the estimator. More censoring (smaller k) indicates better robustness and less censoring (larger k) indicates better performance of the estimator.

B. Implementation Details

We assume that, in general, the distribution of the received data can be represented by

$$f(x) = de^{-\frac{|x|^a}{c}} \quad (3.12)$$

To gain a complete understanding of the underlying distribution, we must have knowledge of the variables a, d , and c .

Now, since $f(x)$ is a valid probability distribution function, the following equation must hold

$$\int_{-\infty}^{\infty} de^{-\frac{|x|^a}{c}} dx = 1 \quad (3.13)$$

Also, by definition of the variance we must have

$$\int_{-\infty}^{\infty} x^2 de^{-\frac{|x|^a}{c}} dx = \sigma^2 \quad (3.14)$$

From the equations (3.13) and (3.14) we have

$$\frac{\int_{-\infty}^{\infty} x^2 e^{-\frac{|x|^a}{c}} dx}{\int_{-\infty}^{\infty} e^{-\frac{|x|^a}{c}} dx} = \sigma^2 \quad (3.15)$$

To find the values of variables a and c , we first fix the value of variable a in the vicinity of unity (for example let $a=0.9$). We then choose a value for the standard deviation σ of the distribution (for example let $\sigma = \frac{1}{\sqrt{2}}$). From equation (3.15) we find the value of c by iteration. After a and c are known, the value of variable d can be found using (3.13) as,

$$d = \frac{1}{\int_{-\infty}^{\infty} e^{-\frac{|x|^a}{c}} dx} \quad (3.16)$$

C. Mean Square Error Performance

In our work we have made use of MSE to quantify performance of the estimator. The MSE performance is expressed by the following form for stationary data:

$$E(\theta - \hat{\theta})^2 = \int_{R^n} (\theta - \frac{1}{n} \sum_{j=1}^n h(x_j))^2 f(x_1) \dots f(x_n) dx_1 \dots dx_n, \quad (3.17)$$

and then

$$= \theta^2 - \frac{2\theta}{n} \int_{R^n} h(x_j) f(x_1) \dots f(x_n) dx_1 \dots dx_n + \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \int_{R^n} h(x_i) h(x_j) f(x_1) \dots f(x_n) dx_1 \dots dx_n, \quad (3.18)$$

and this equation will be

$$= \theta^2 - 2\theta \int_{-\infty}^{\infty} h(x) f(x) dx + \frac{n^2 - n}{n^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x) h(y) f(x) f(y) dx dy + \frac{n}{n^2} \int_{-\infty}^{\infty} h^2(x) f(x) dx, \quad (3.19)$$

which is expressed by the following form

$$= \theta^2 - 2\theta \int_{-\infty}^{\infty} h(x) f(x) dx + (1 - \frac{1}{n}) (\int_{-\infty}^{\infty} h(x) f(x) dx)^2 + \frac{1}{n} \int_{-\infty}^{\infty} h^2(x) f(x) dx, \quad (3.20)$$

where $f(\cdot)$ is the univariate density of i.i.d data and n is number of samples received.

1. Laplacian Distribution

Recall that our estimator is expressed as

$$h(x) = \begin{cases} \sqrt{2}k, & x > +k \\ \sqrt{2}|x|, & |x| < +k \\ \sqrt{2}k, & x < -k \end{cases} \quad (3.21)$$

Also notice that the parameter to be estimated for Laplacian data is its standard deviation σ ; thus, θ should be replaced by σ . The specific integral parts for the Laplace distribution of the equation (3.20) can now be found as

$$\int_{-\infty}^{\infty} h(x)f(x)dx = 2\sqrt{2} \int_0^k xf(x)dx + 2\sqrt{2} \int_k^{\infty} kf(x)dx \quad (3.22)$$

$$\int_{-\infty}^{\infty} h^2(x)f(x)dx = 4 \int_0^k x^2 f(x)dx + 4 \int_k^{\infty} k^2 f(x)dx \quad (3.23)$$

$$f(x) = de^{-\frac{|x|}{c}} \quad (3.24)$$

Making use of the equations (3.20) through (3.24) we find the value of MSE for each set of variables a, d , and c . To evaluate robustness and performance of the estimator, we plot MSE vs. a (for a particular value of σ , k and n), where a varies within a range, say from 0.1 to 2.0. This can be repeated for various values of standard deviations, and various values of censoring value (k) and n , for each value of standard deviation. Figure 3 depicts a sample curve for $\sigma=1.0$, $k=\infty$ (no censoring) and $n = 20$.

The value of MSE at points a is the measure of performance, and the slopes at these points is the measure of robustness of the estimator at that point. To get

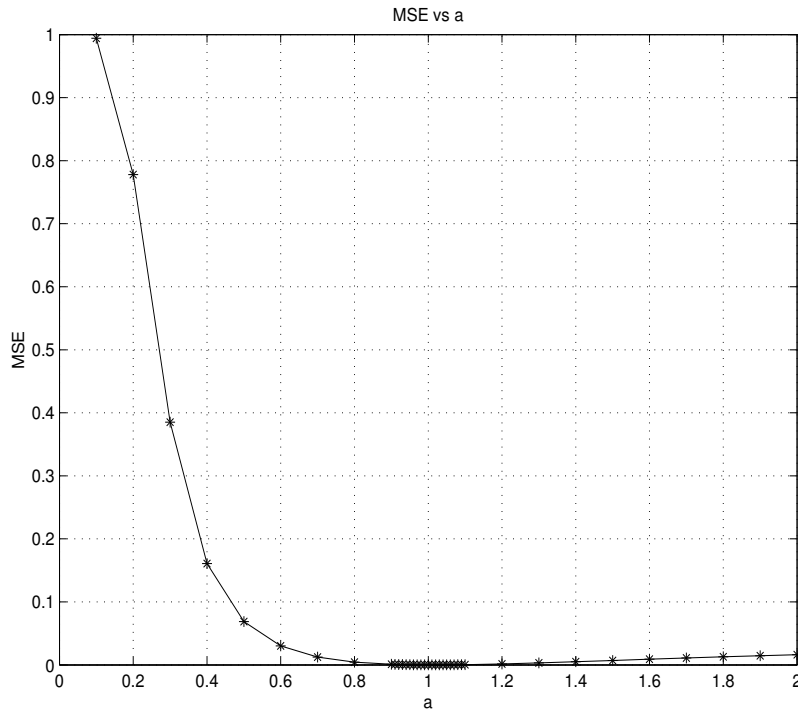


Fig. 3. MSE vs a , for $\sigma=1.0$, $k=\infty$

overall robustness R , and performance P , we need to combine the performance and robustness parameters across the whole range of a . One obvious approach is to view a as a discrete random variable \mathbf{A} taking on values $\{a_i : -\infty < i < \infty\}$ with probability $P(\mathbf{A} = a_i) = w_i$. robustness (i.e., slope) and performance (i.e., MSE) are then viewed as functions of \mathbf{A} with average values

$$\begin{aligned} \mathbf{E}\{mse\} &= \sum_{i=-\infty}^{\infty} w_i mse(a_i) \\ \mathbf{E}\{slope\} &= \sum_{i=-\infty}^{\infty} w_i slope(a_i) \end{aligned} \tag{3.25}$$

Parameters P and R can be expressed in terms of $\mathbf{E}\{mse\}$ and $\mathbf{E}\{slope\}$ as,

$$\begin{aligned} P &= \frac{1}{\mathbf{E}\{mse\}} \\ R &= \frac{1}{\mathbf{E}\{slope\}} \end{aligned} \quad (3.26)$$

We then normalize and combine these two parameters to form a composite cost function,

$$J = \epsilon P + (1 - \epsilon)R, \quad 0 \leq \epsilon \leq 1 \quad (3.27)$$

where ϵ is the weighting factor deciding how much emphasis is to be placed on robustness by the user depending on the application.

2. Distribution of the Random Variable \mathbf{A}

One of the major challenges of this work was to assign a distribution to the random variable \mathbf{A} , since there are infinitely many ways of doing this. We have chosen an approach which we consider to be both intuitive and realistic.

One of the characteristics of the employment of robustness is the concept of a nominal ($\mathbf{A} = a_0$), for example, $\mathbf{A} = 1$, the Laplace density. It would seem reasonable that as one moves in discrete intervals away from the nominal, the probability should decrease, if the meaning of the "nominal" is to be preserved. We are not describing a classical random walk, however. We remark that this discussion is directed toward a definition and should not be considered a derivation, but we will employ certain deductive tools in arriving at an appropriate definition. Suppose we consider $\mathbf{A} = 1$ as the nominal and move positively by a step size δ . Each independent step to the right has a probability β , and so with η steps to the right we have

$$P\{\mathbf{A} = 1 + \eta\delta\} = \alpha(\beta)^\eta \quad (3.28)$$

where α is the initial probability that $\mathbf{A}=1$. Similarly, we can move to the negative direction by allowing η to take on negative values, and thus

$$P\{\mathbf{A} = 1 + \eta\delta\} = \alpha(\beta)^{|\eta|} \quad (3.29)$$

we then have

$$\begin{aligned} \sum_{\eta=-\infty}^{\infty} P\{\mathbf{A} = 1 + \eta(\delta)\} &= 1, \\ \Rightarrow \sum_{\eta=-\infty}^{\infty} \alpha(\beta)^{|\eta|} &= 1 \\ \Rightarrow \alpha + 2 \sum_{\eta=1}^{\infty} \alpha(\beta)^{\eta} &= 1 \\ \Rightarrow \beta &= \frac{1 - \alpha}{1 + \alpha} \end{aligned} \quad (3.30)$$

An important thing to note here is that, even though we let the integer variable i (in equations 3.25) and η (in equations 3.30) take infinitely many values, we truncate its range in practice. This is done because we are considering only finite perturbations in a . Also notice that in Fig 3, the quality (in terms of robustness and performance) of our estimator is quite good for positive perturbations in a i.e, when a deviates to the right hand side of unity on the x-axis ($\eta > 0$).

CHAPTER IV

SUMMARY

A. MSE as a Function of a

In this section we present our results for the discussion in the previous section. We begin with plotting curves (for each value of σ , where $\sigma = \frac{1}{\sqrt{2}}, 1, \sqrt{2}$) depicting the behavior of MSE as a changes from 0.1 to 2.0, for censoring value $k = 0, 0.2, \dots, 1.6$, and for each combination of σ and k , $n = 2, 3, \dots, 10, 15, 20$. Figures 4 through 36 depict the behavior of MSE as a changes from 0.1 to 2.0.

B. Cost Function J , and Choices of α and η

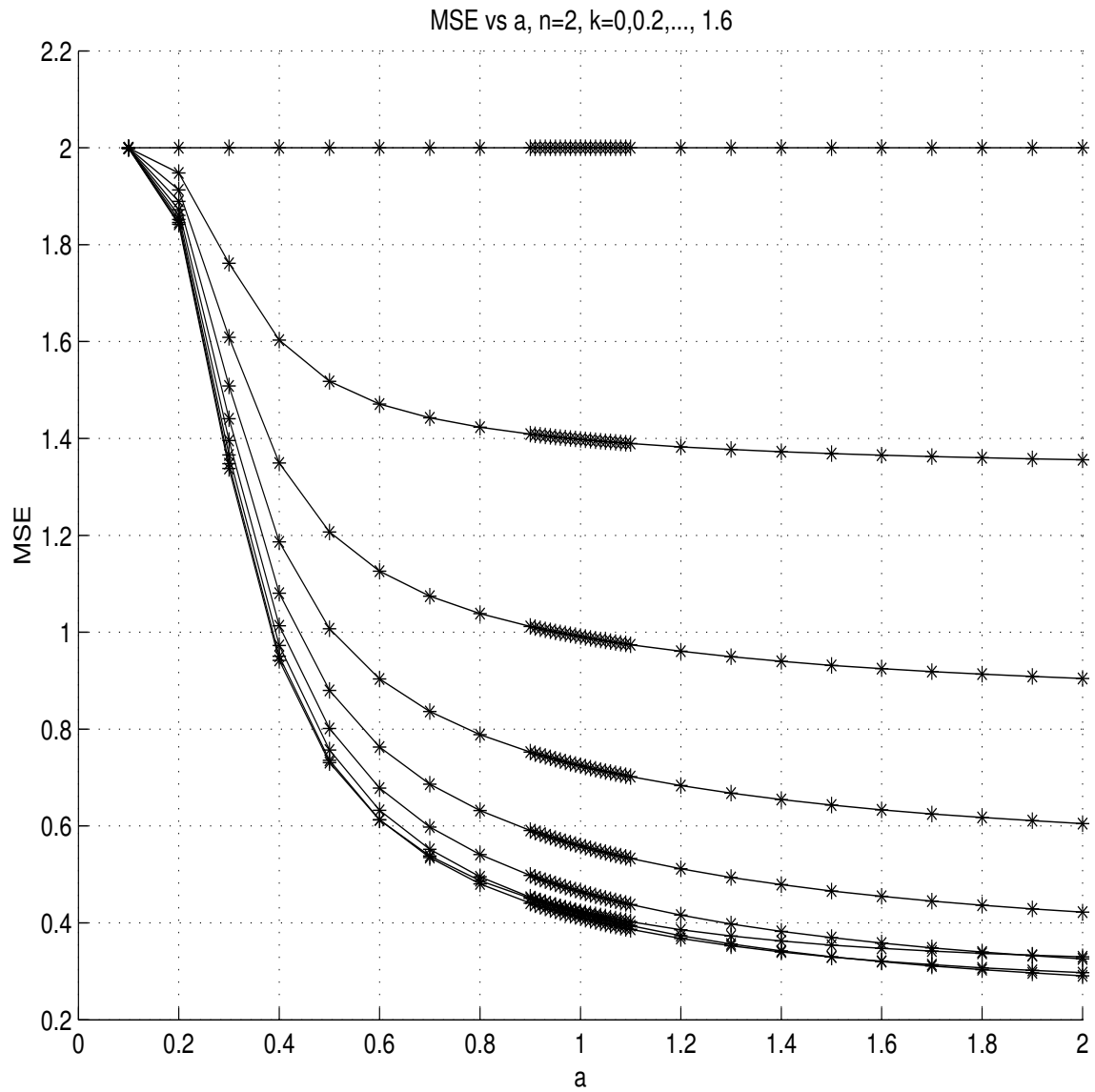
We find performance P and robustness R as explained in previous chapter. We then combine these two variables P and R using a composite cost function J as discussed in last chapter. For each combination of σ , k , and n as explained above; together with α , where α varies between 0 and 1 (to account for various distributions of the random variable \mathbf{A}), where we maintain $\delta = 0.01$ through out. We next tabulate, for various values of α and n , our results which illustrate the censoring level k that maximizes J as a function of the choice of ϵ .

We choose wide ranging values of α (from 0.01, 0.02, ..., 0.1, 0.3, 0.5, 0.9), β was then calculated using equation (3.30). A small α essentially implies that larger number of points on the curve are considered when $\mathbf{E}\{mse\}$ and $\mathbf{E}\{slope\}$ of the curve are computed. One way of interpreting this is that we have a low level of confidence about the behavior of \mathbf{A} around nominal, and thus, we would want to accommodate large perturbations in \mathbf{A} around nominal. On the other hand, large α implies that smaller number of points are considered when $\mathbf{E}\{mse\}$ and $\mathbf{E}\{slope\}$ of the curve are

computed, and this means, we have a high level of confidence about the behavior of \mathbf{A} around nominal, and thus we need to accommodate only minor perturbations in \mathbf{A} around nominal. This, precisely is the driving factor behind allowing α to take on a wide range of values.

While choosing the range of η , notice that \mathbf{A} varies from 0.1 to 1.0 (on the left hand side of the nominal); recall that we are moving away from the nominal in steps of $\delta=0.01$, thus, η 's maximum value is $1+(1.0-0.1)/0.01= 91$.

Figures 4 through 36 depict the behavior of MSE vs a for various combinations of k,n , and σ^2 .

Fig. 4. $\sigma^2 = 2$, n=2

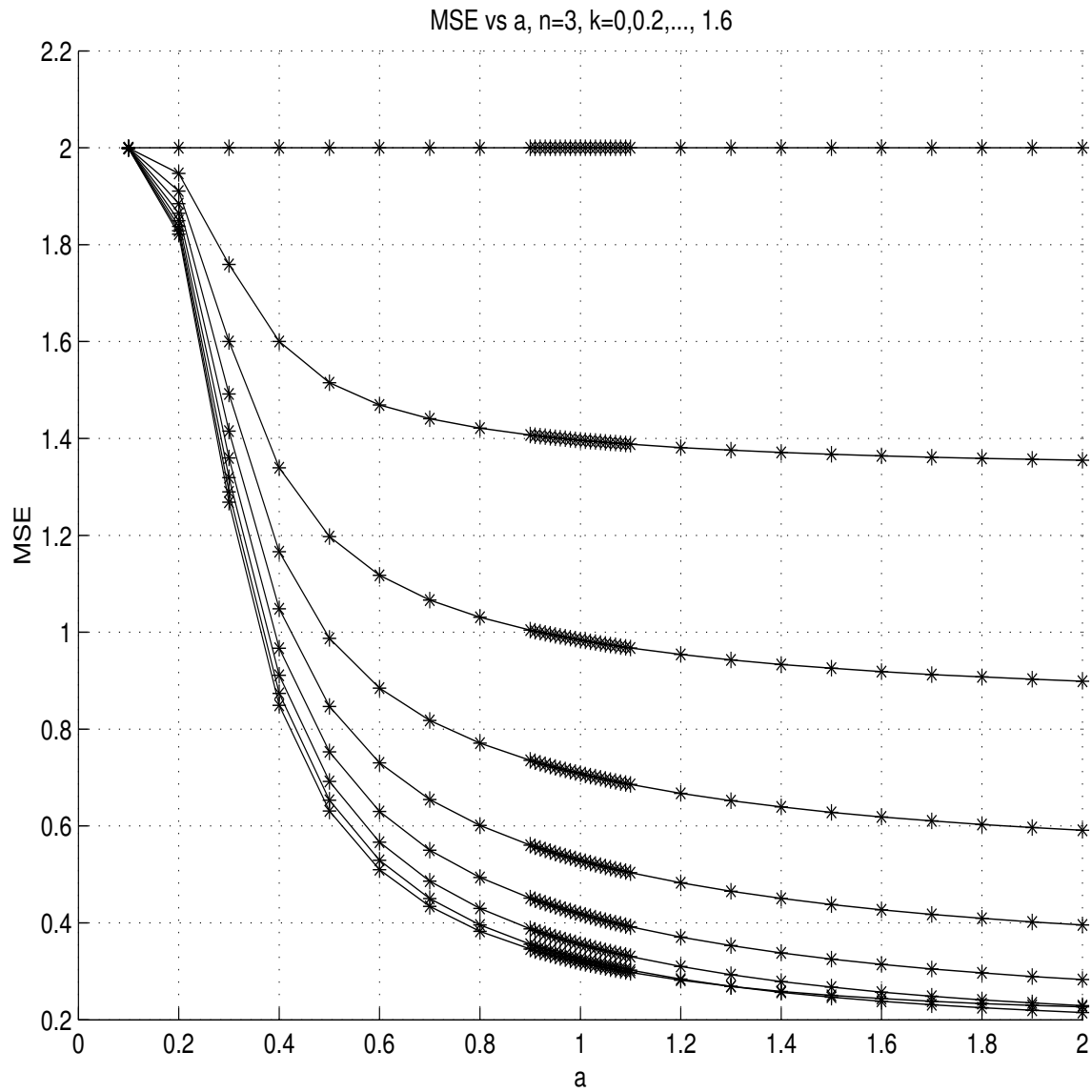
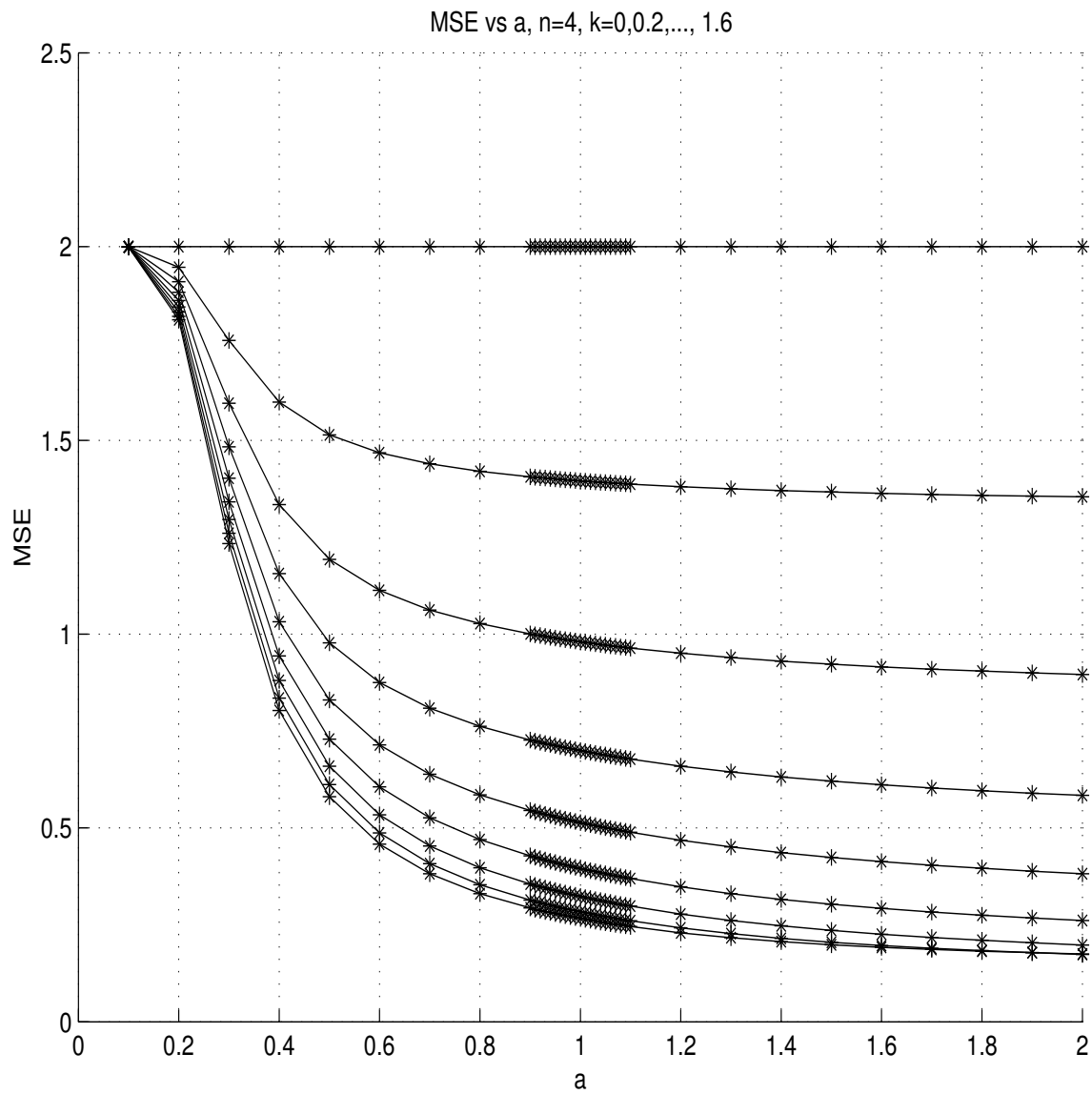


Fig. 5. $\sigma^2 = 2$, $n=3$

Fig. 6. $\sigma^2 = 2$, $n=4$

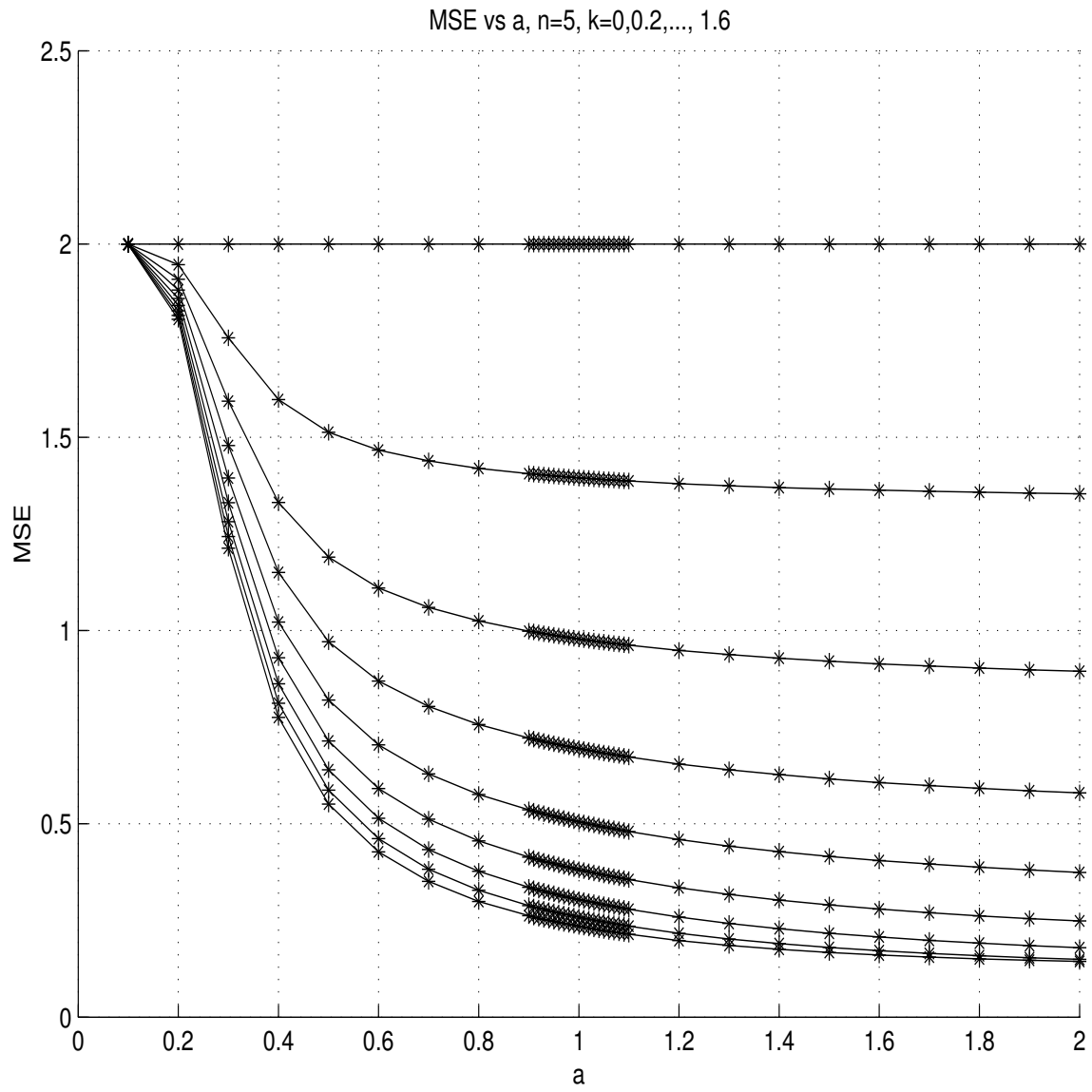


Fig. 7. $\sigma^2 = 2$, $n=5$

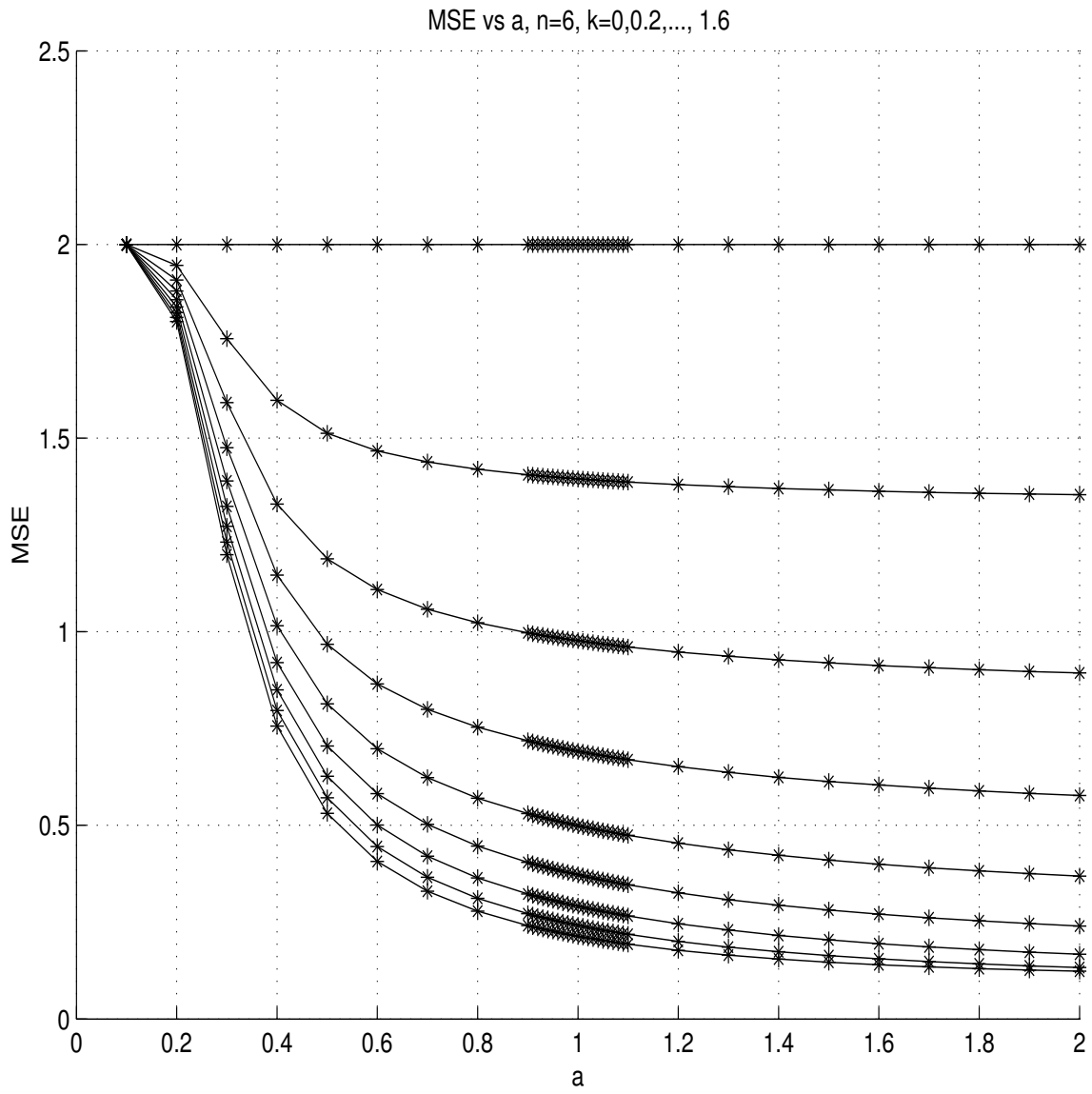


Fig. 8. $\sigma^2 = 2$, $n=6$

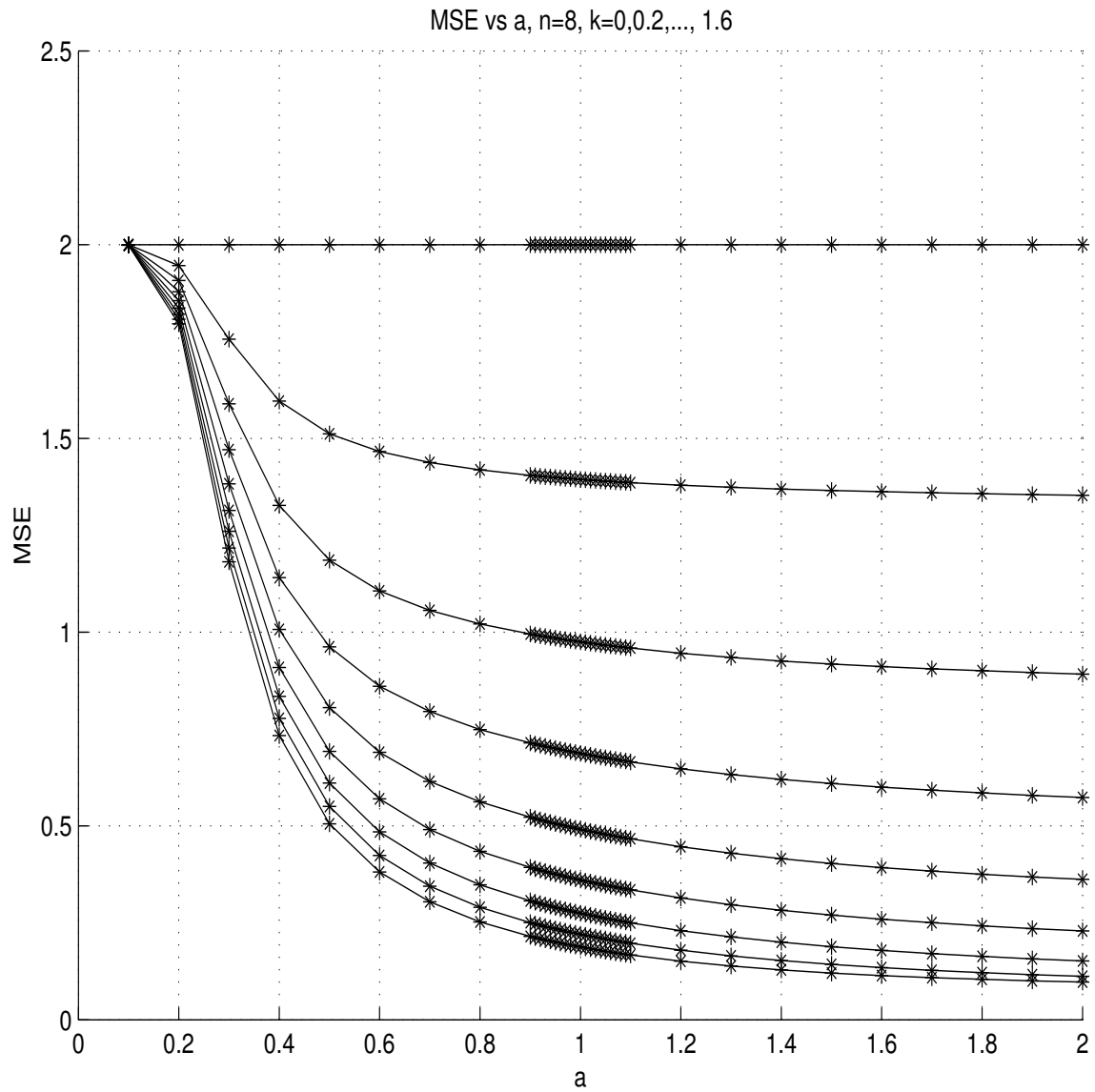
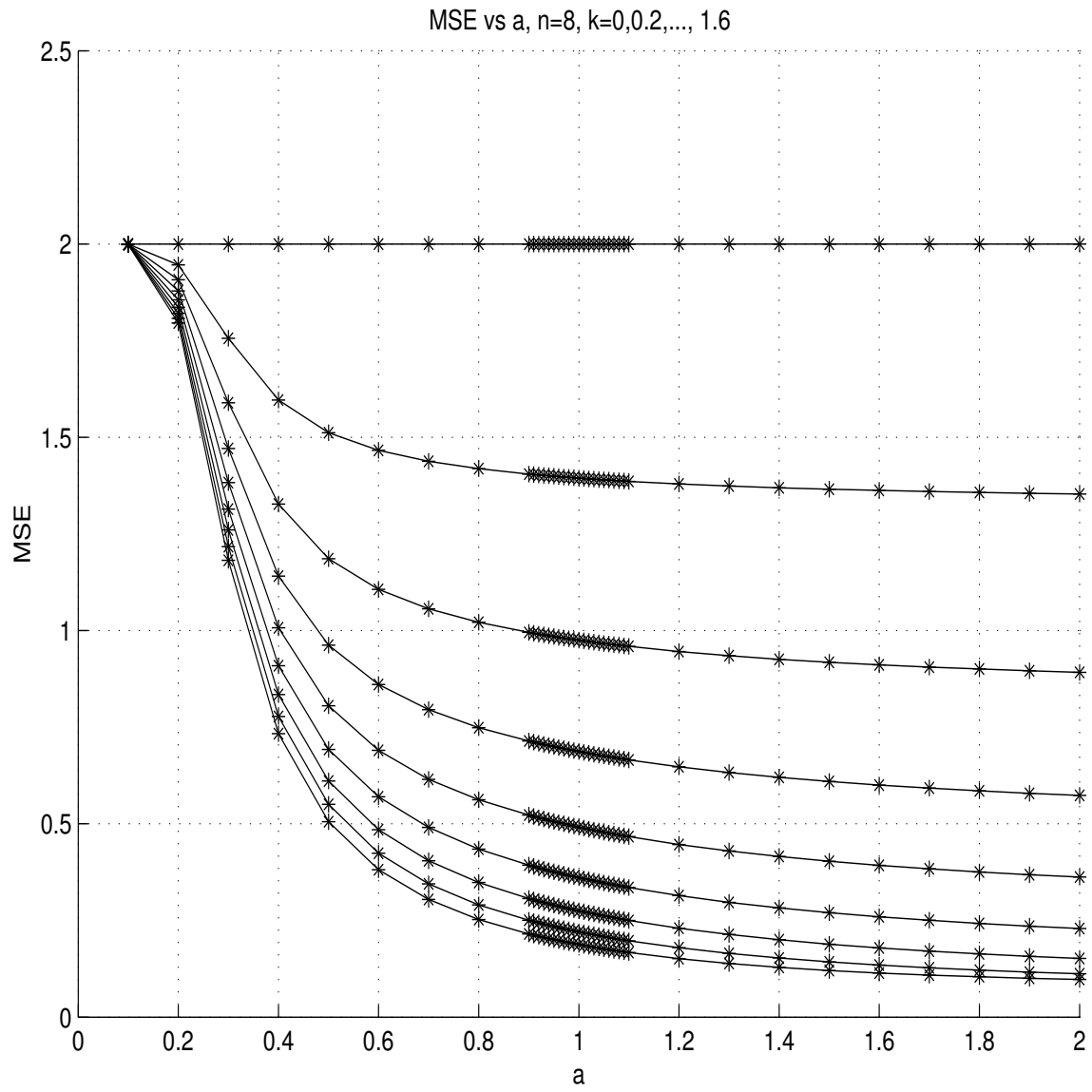


Fig. 9. $\sigma^2 = 2$, $n=7$

Fig. 10. $\sigma^2 = 2$, $n=8$

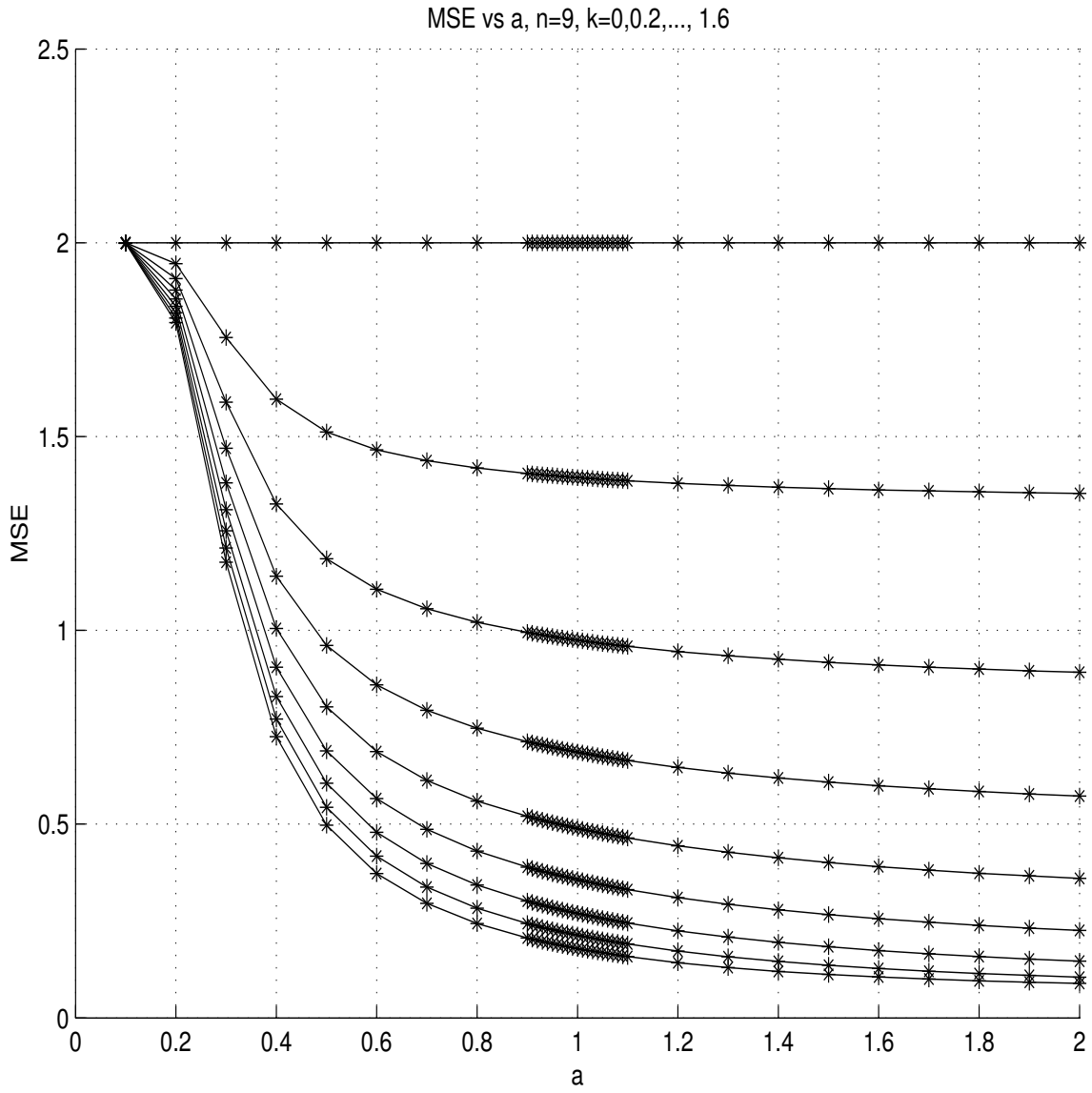
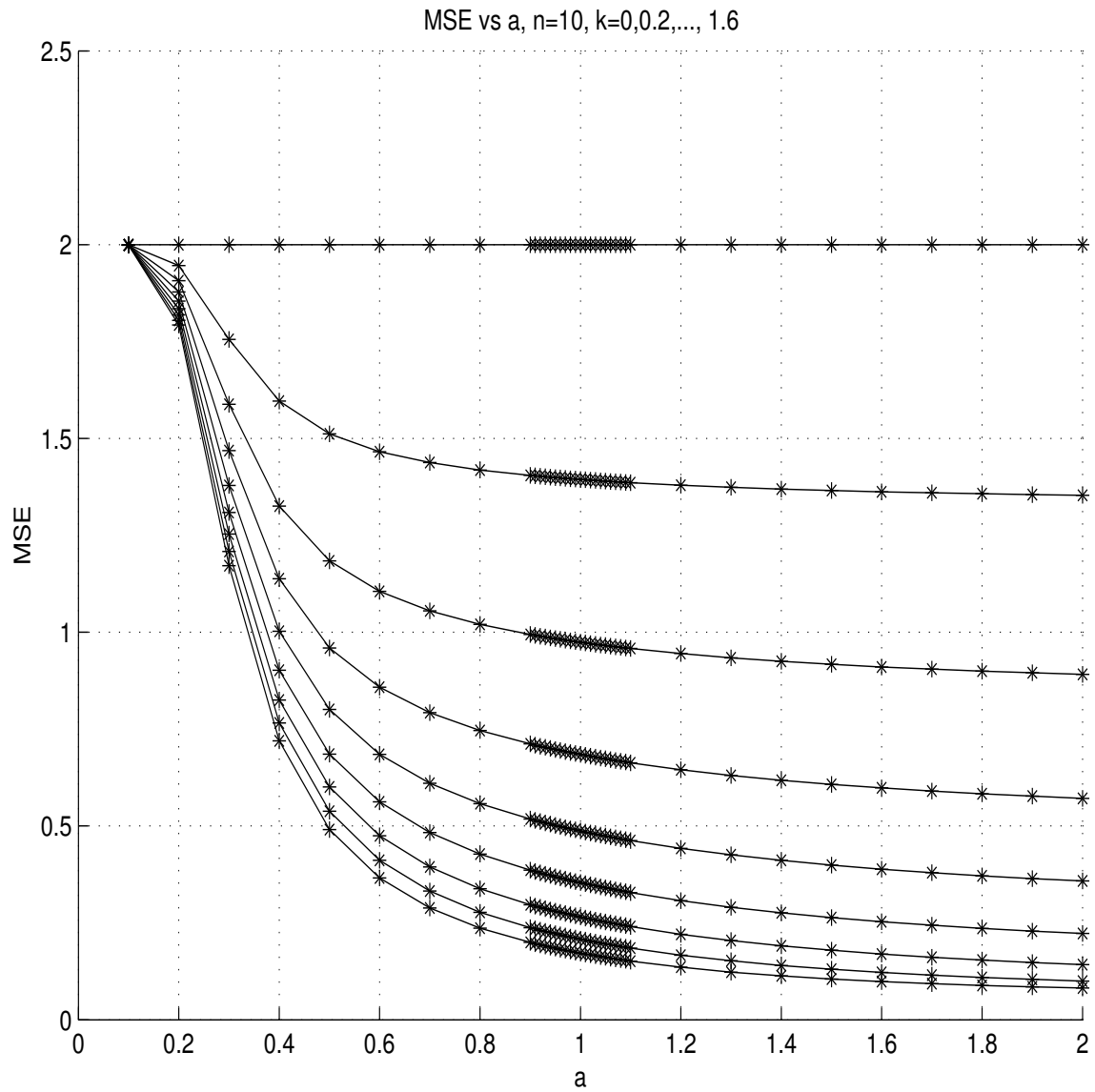
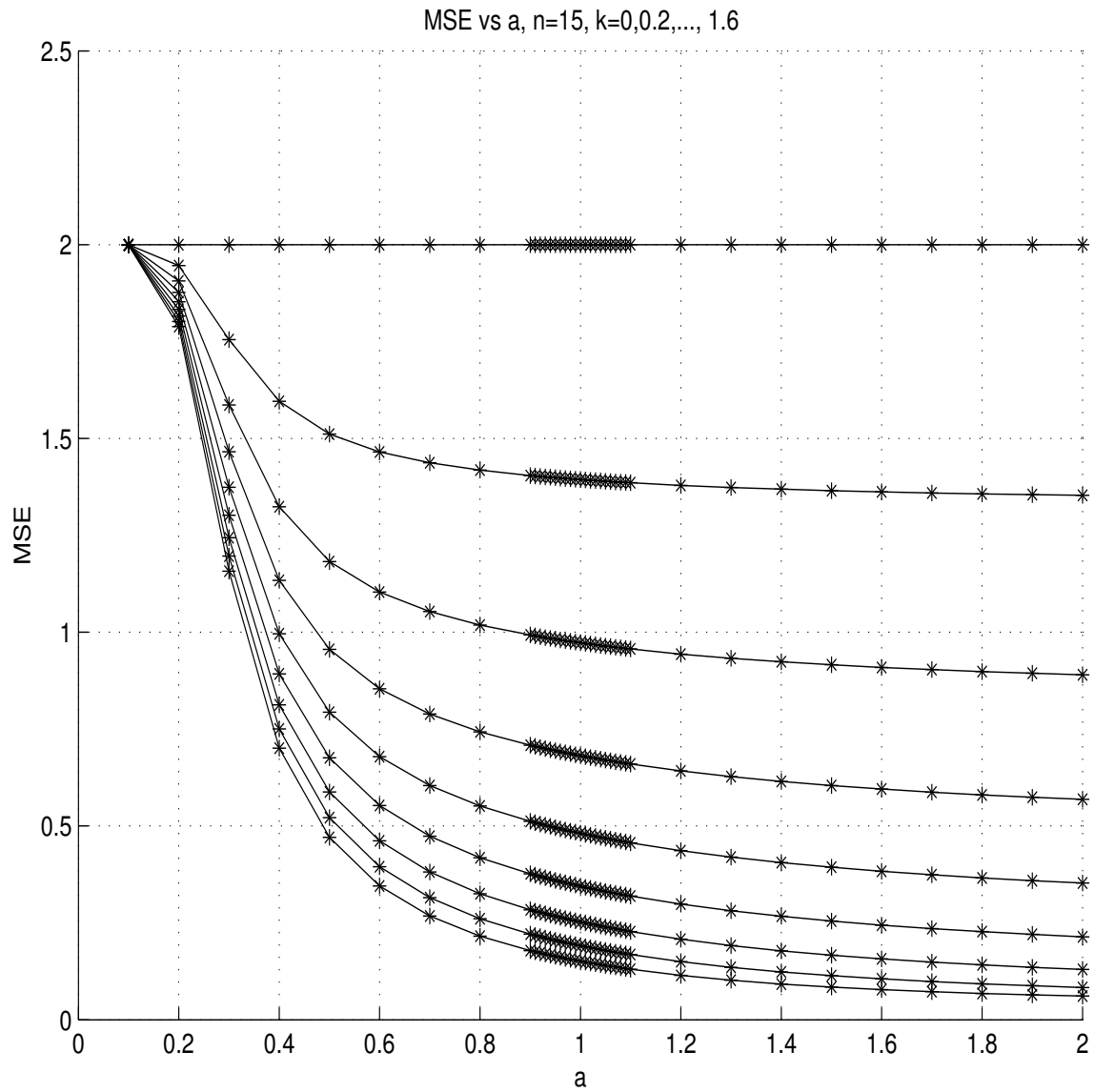
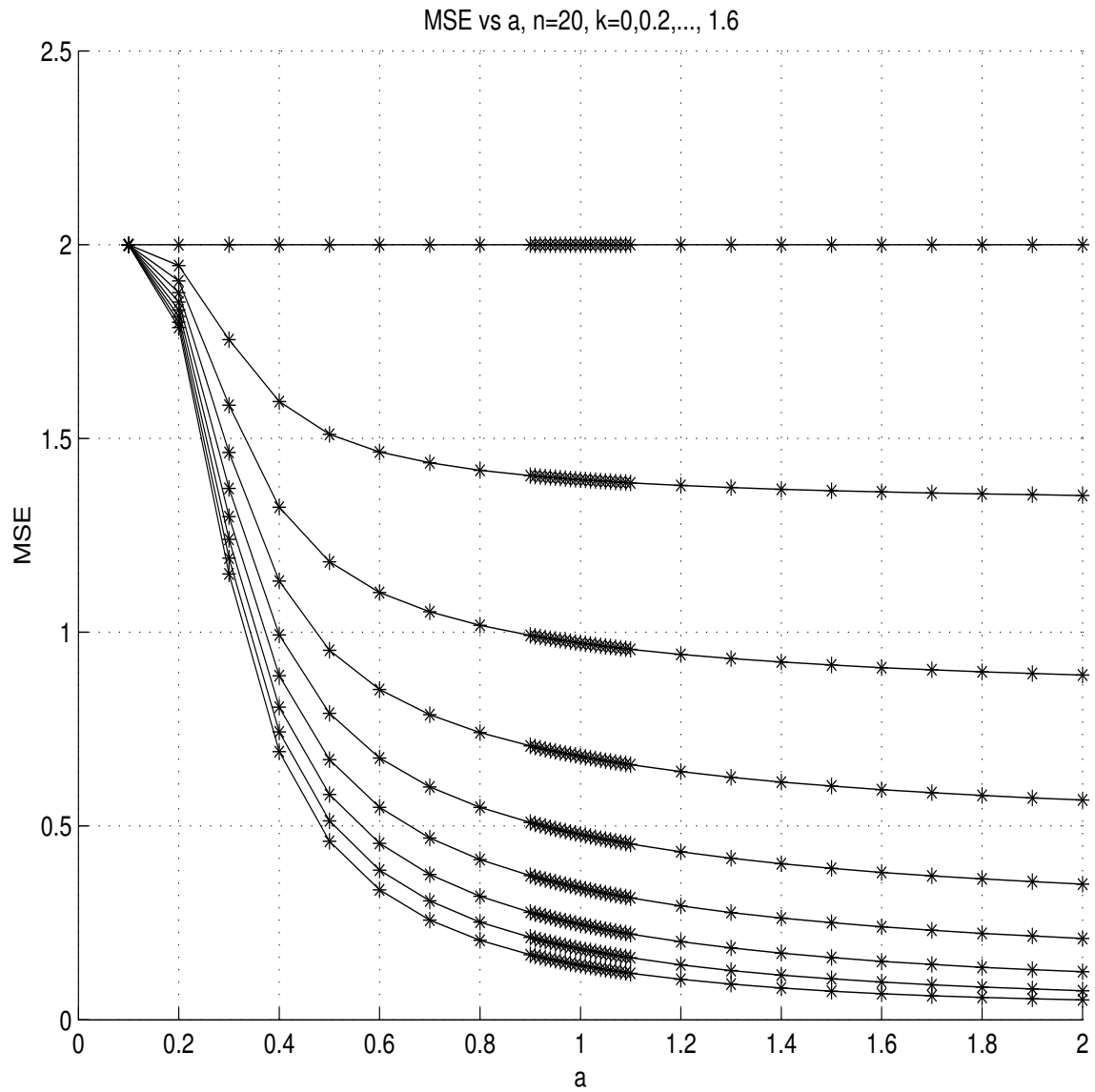


Fig. 11. $\sigma^2 = 2$, $n=9$

Fig. 12. $\sigma^2 = 2$, n=10

Fig. 13. $\sigma^2 = 2$, n=15

Fig. 14. $\sigma^2 = 2$, n=20

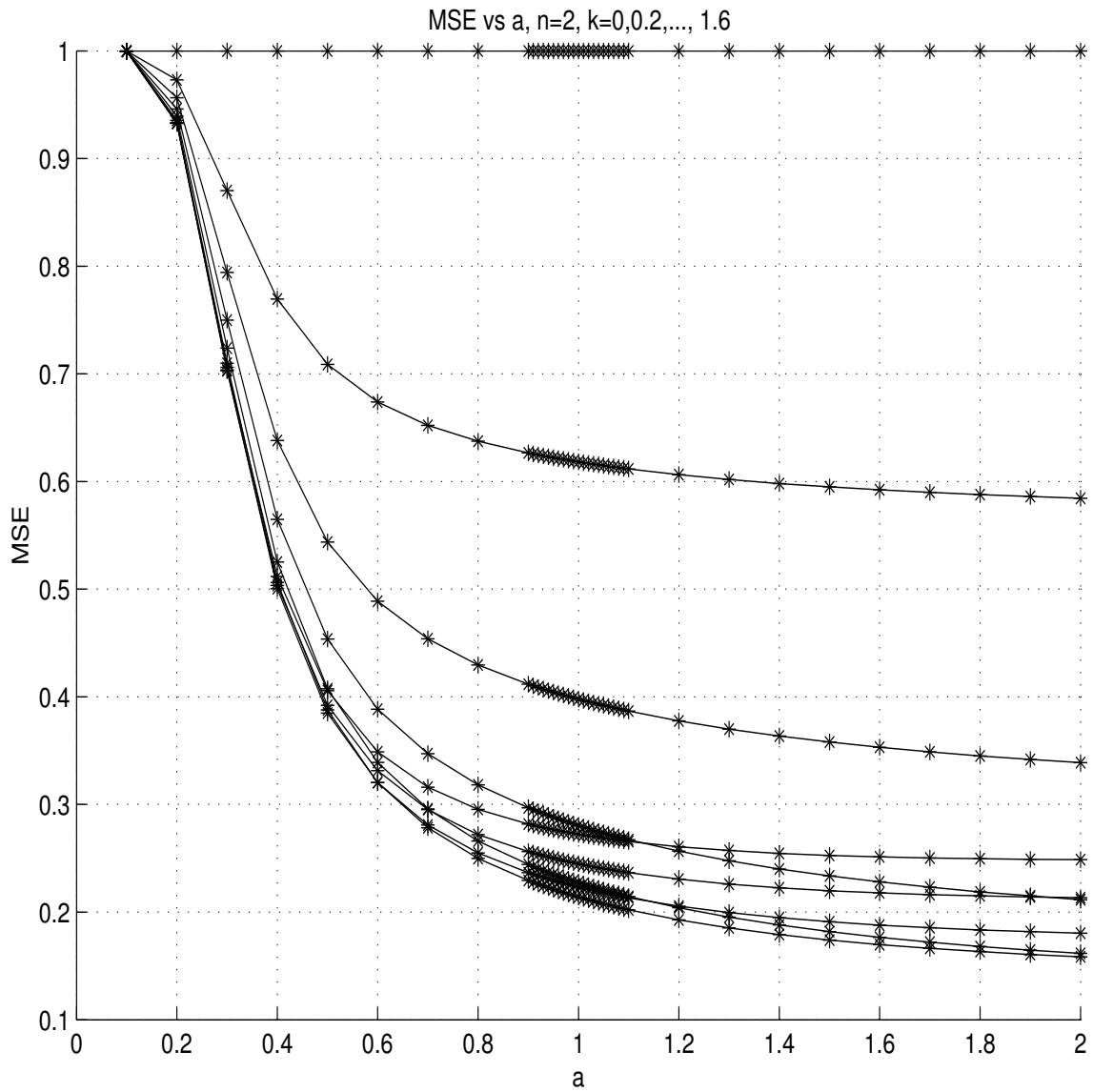


Fig. 15. $\sigma^2 = 1$, $n=2$

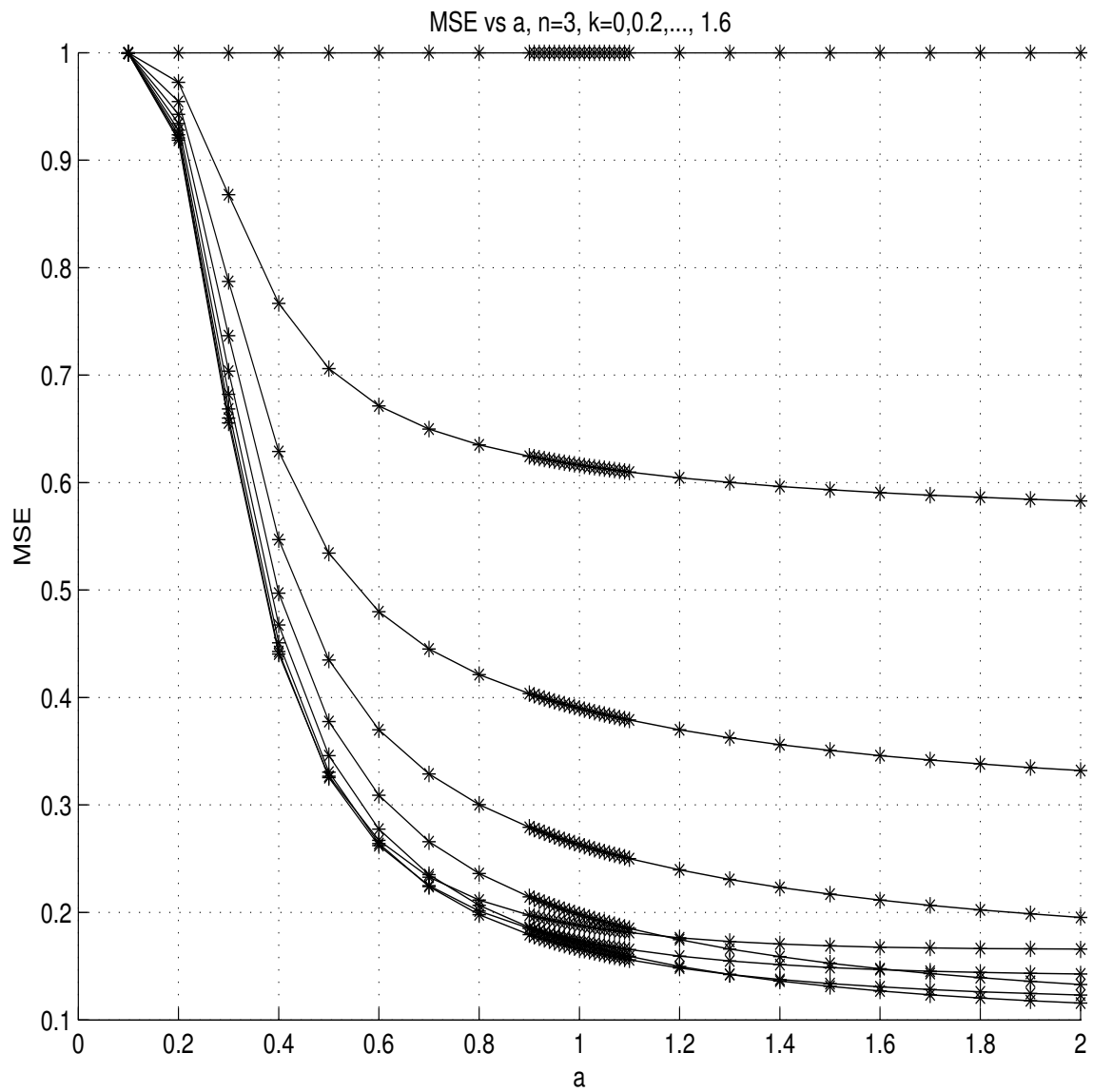


Fig. 16. $\sigma^2 = 1, n=3$

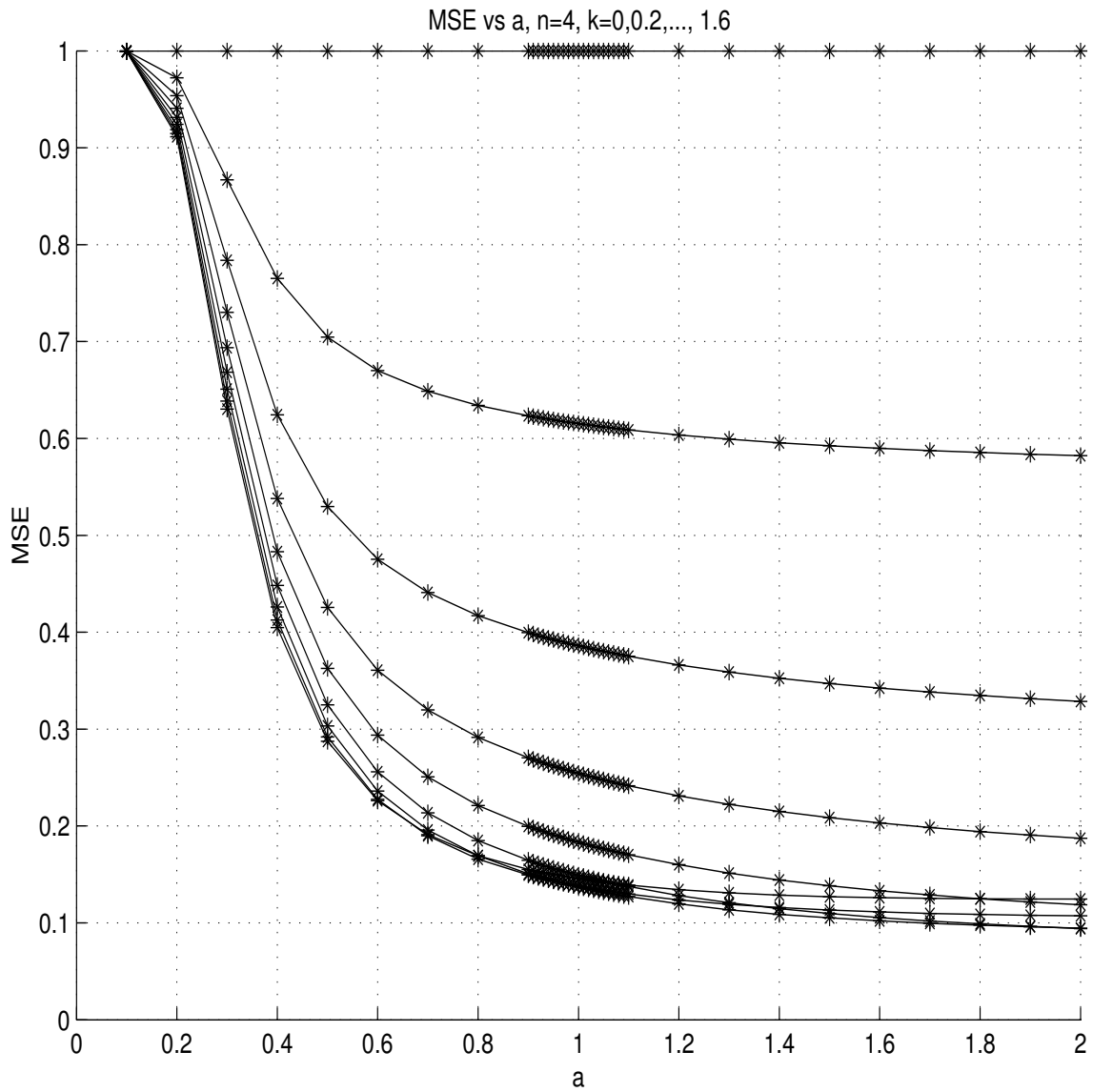


Fig. 17. $\sigma^2 = 1$, $n=4$

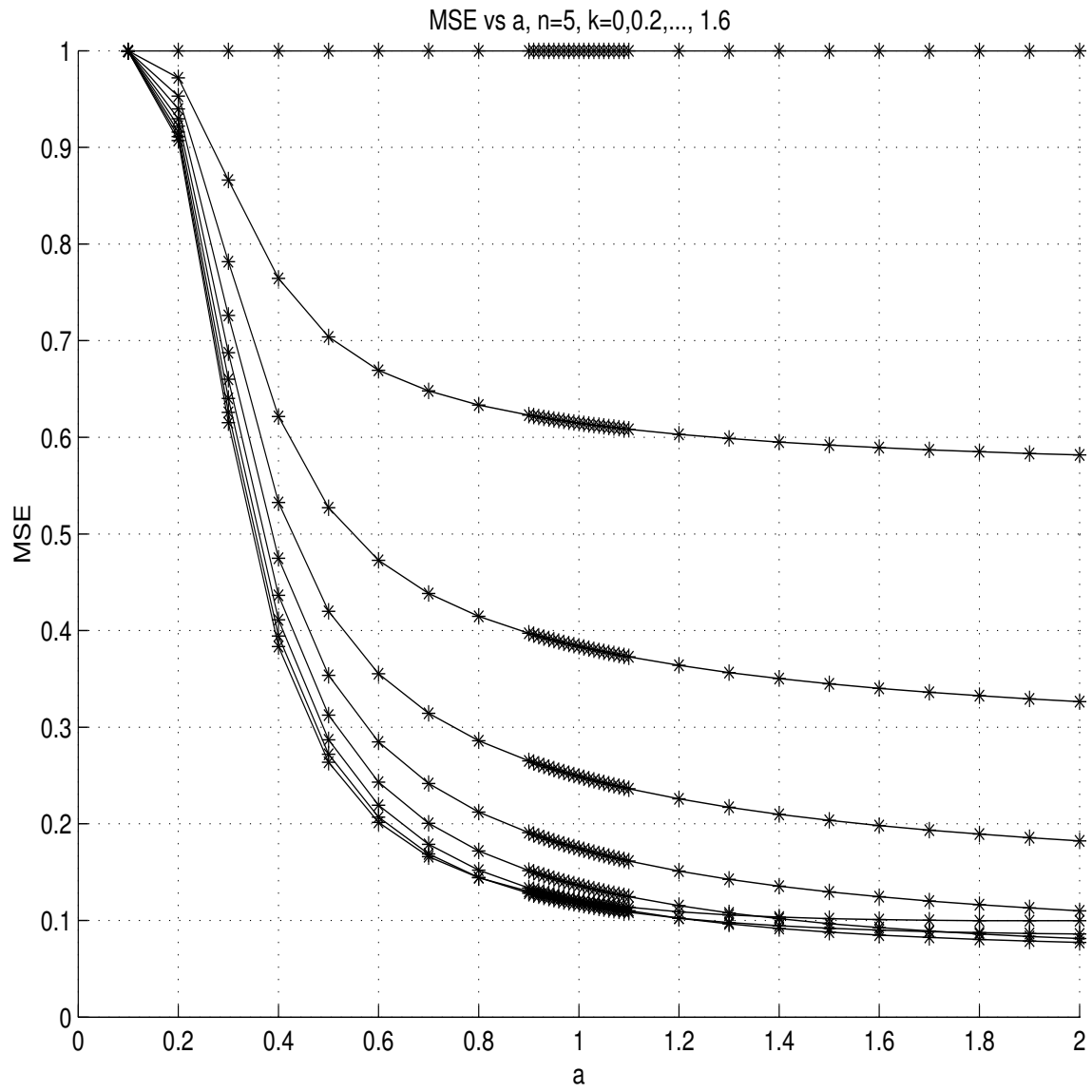


Fig. 18. $\sigma^2 = 1, n=5$

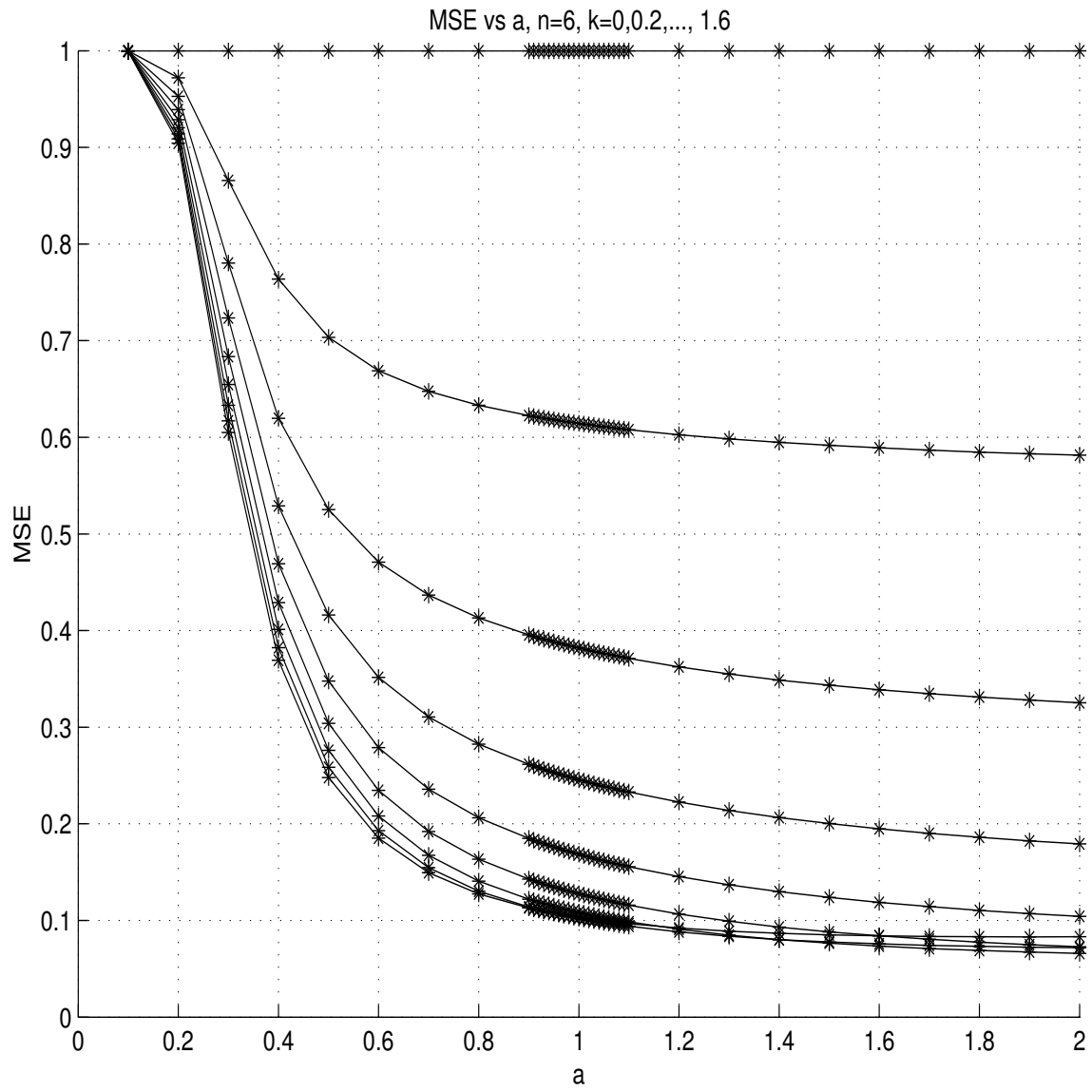
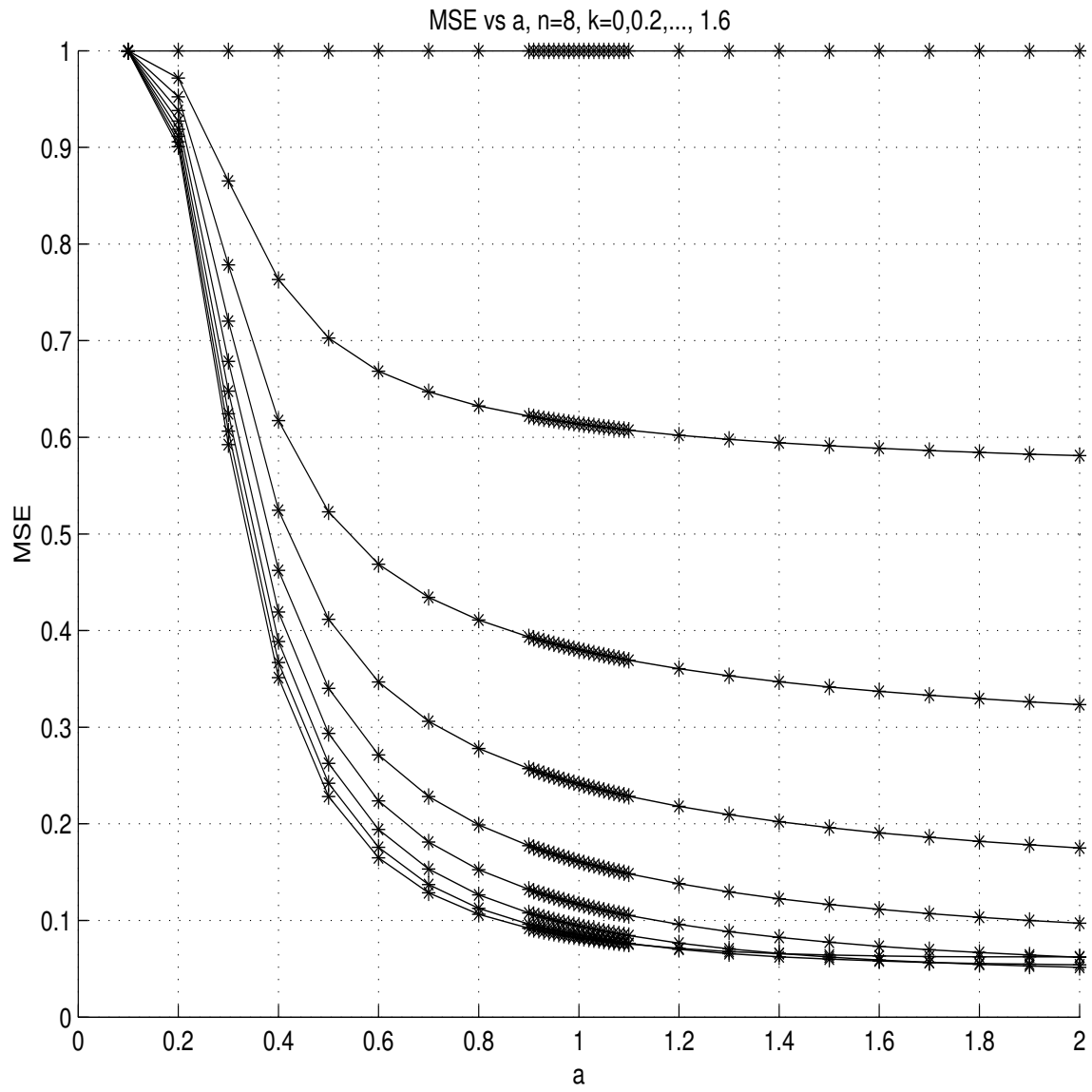
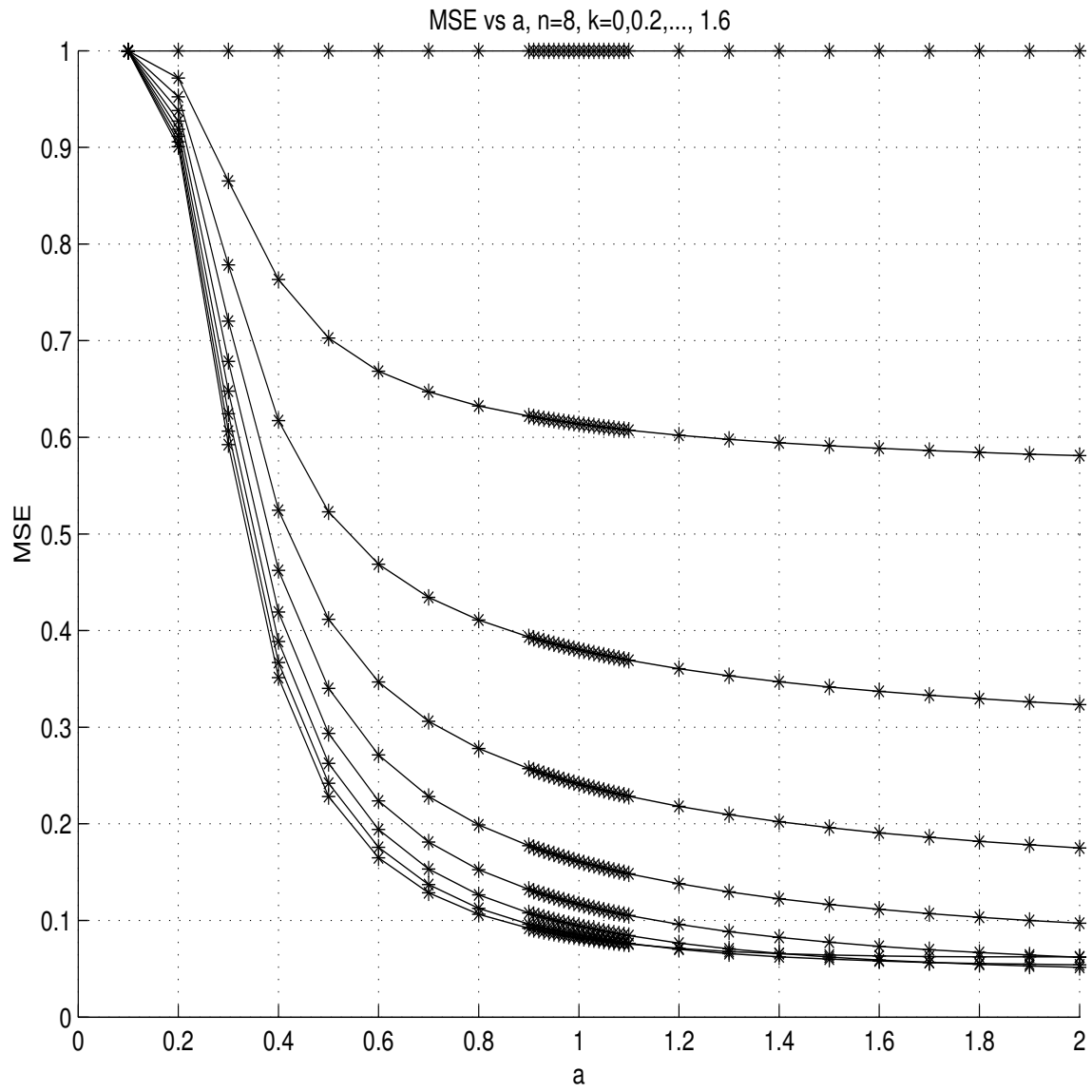


Fig. 19. $\sigma^2 = 1$, n=6

Fig. 20. $\sigma^2 = 1, n=7$

Fig. 21. $\sigma^2 = 1$, $n=8$

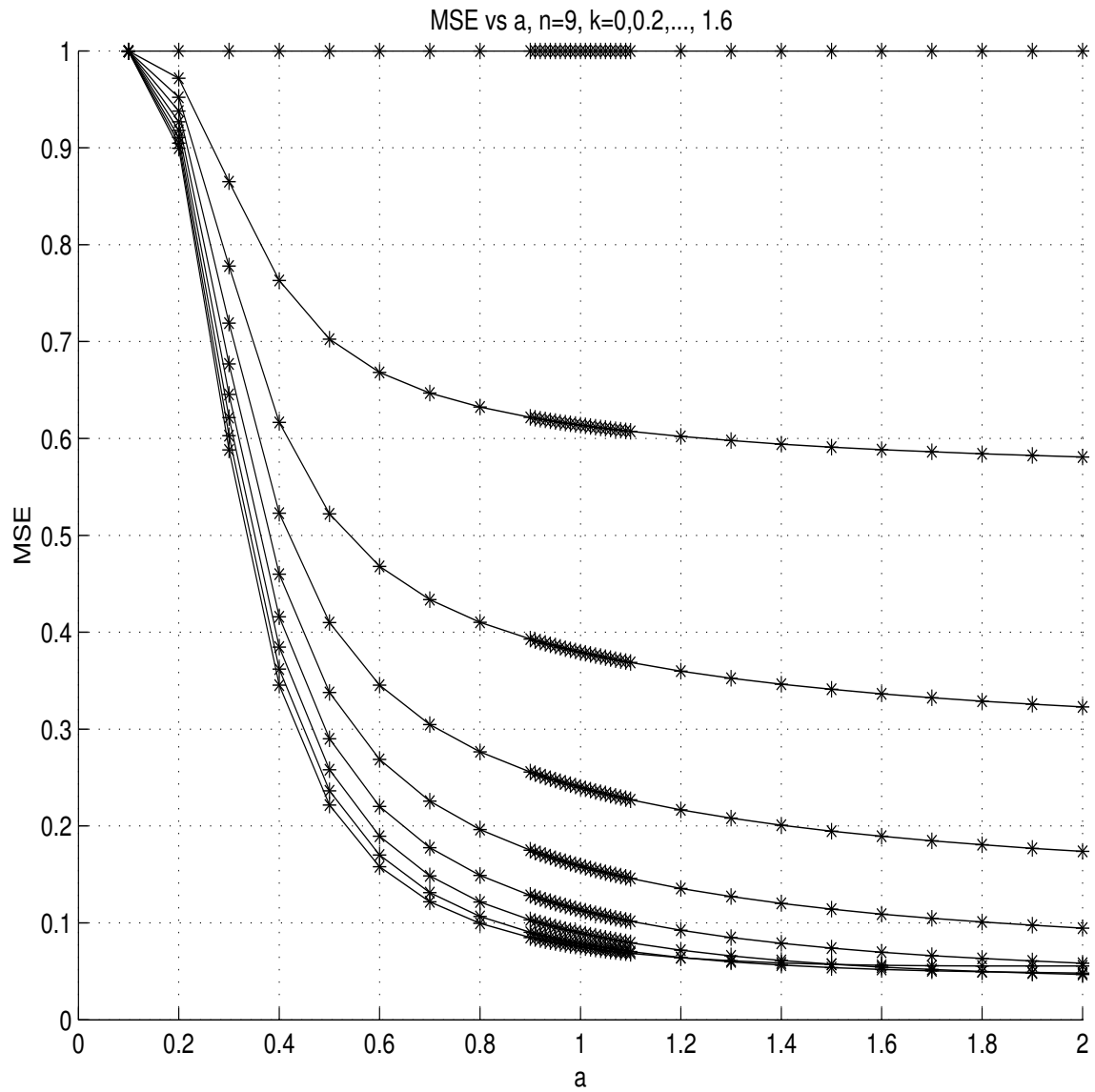
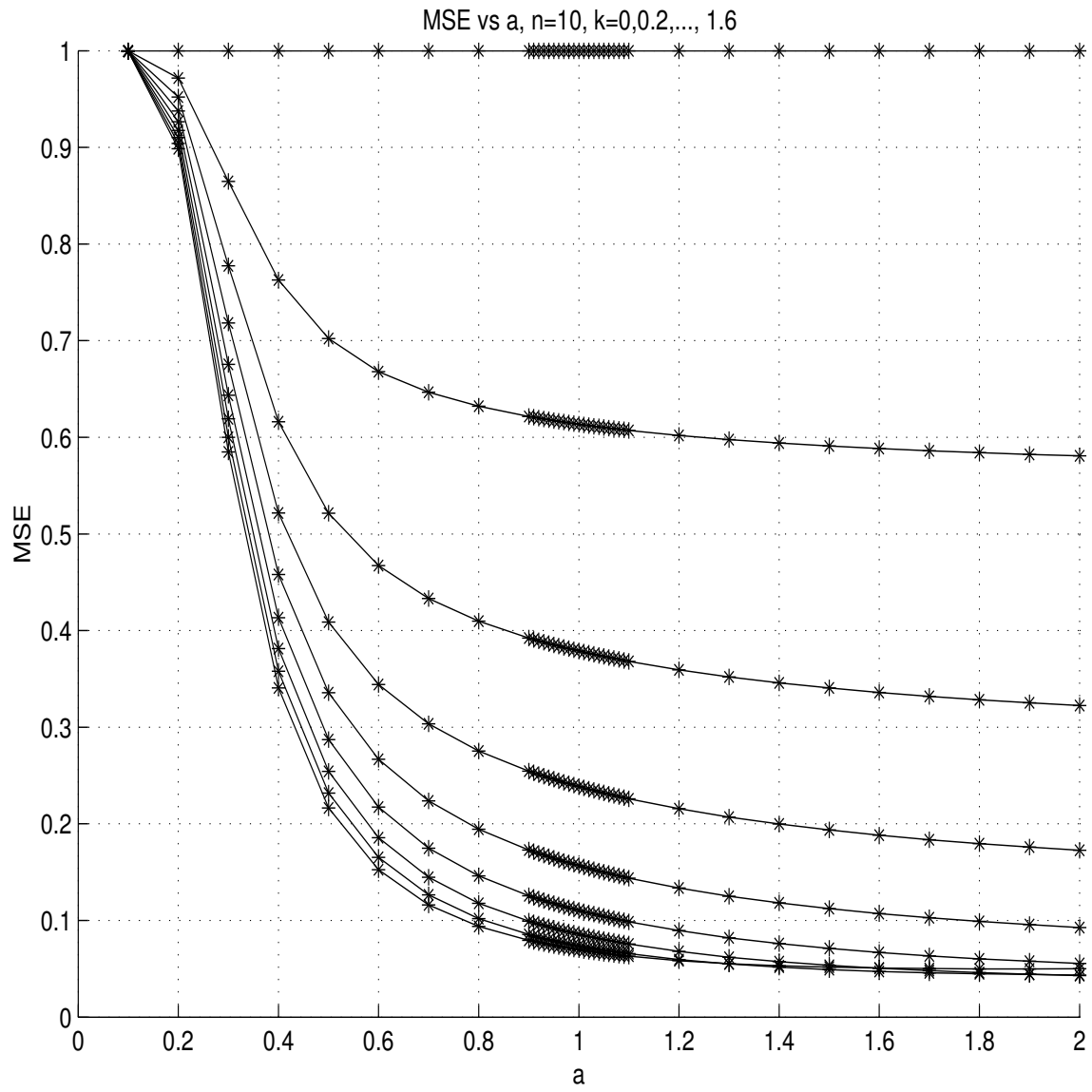


Fig. 22. $\sigma^2 = 1$, $n=9$

Fig. 23. $\sigma^2 = 1, n=10$

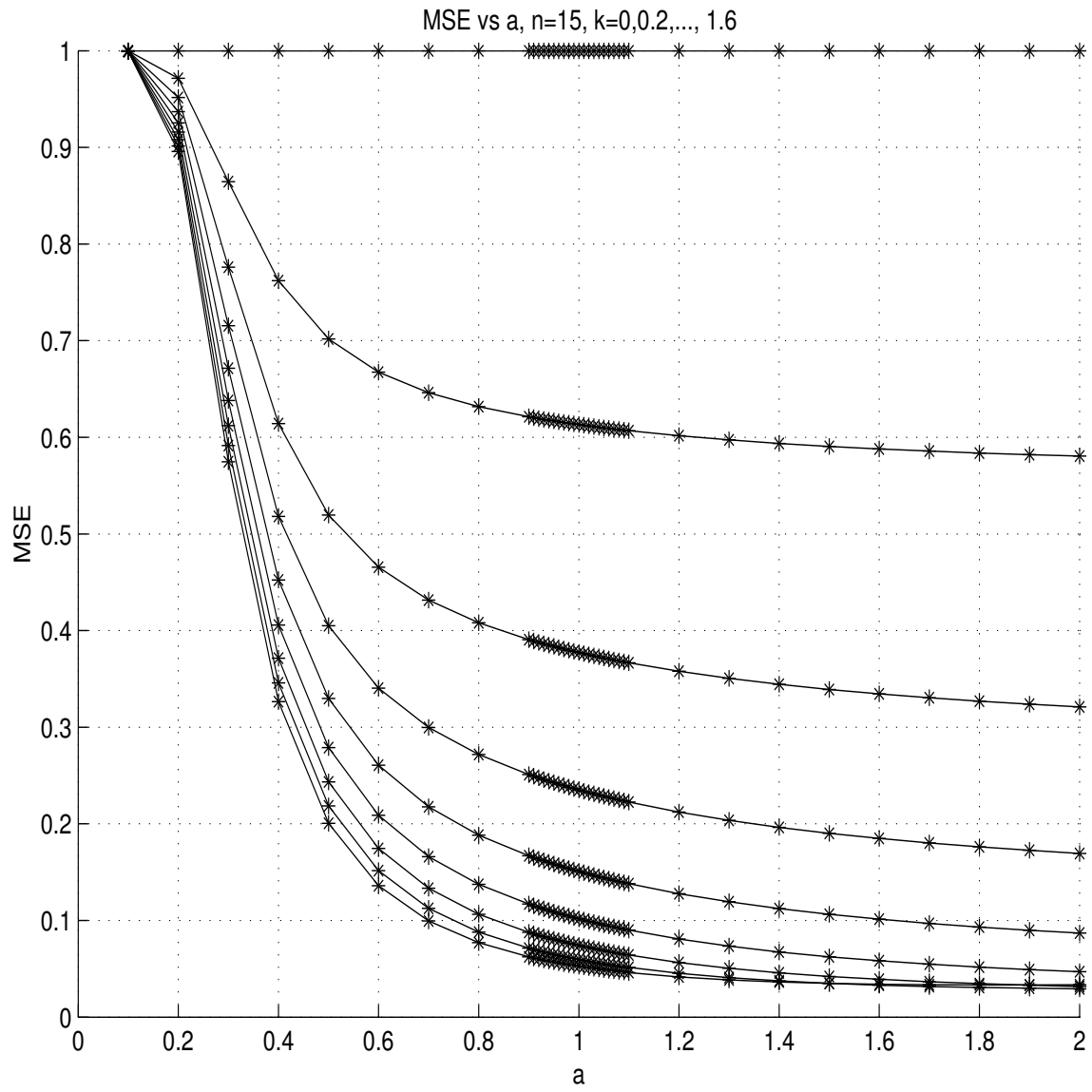
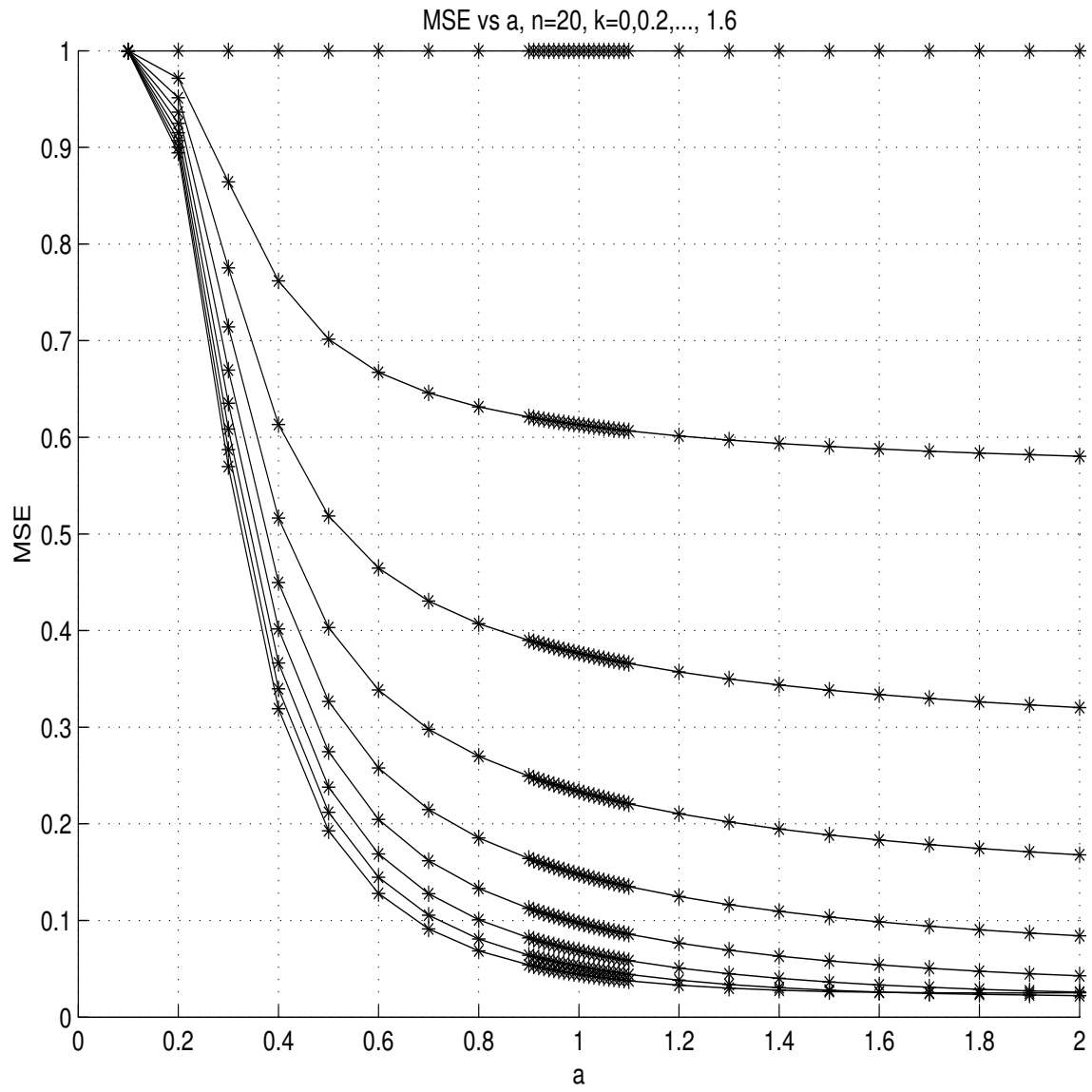
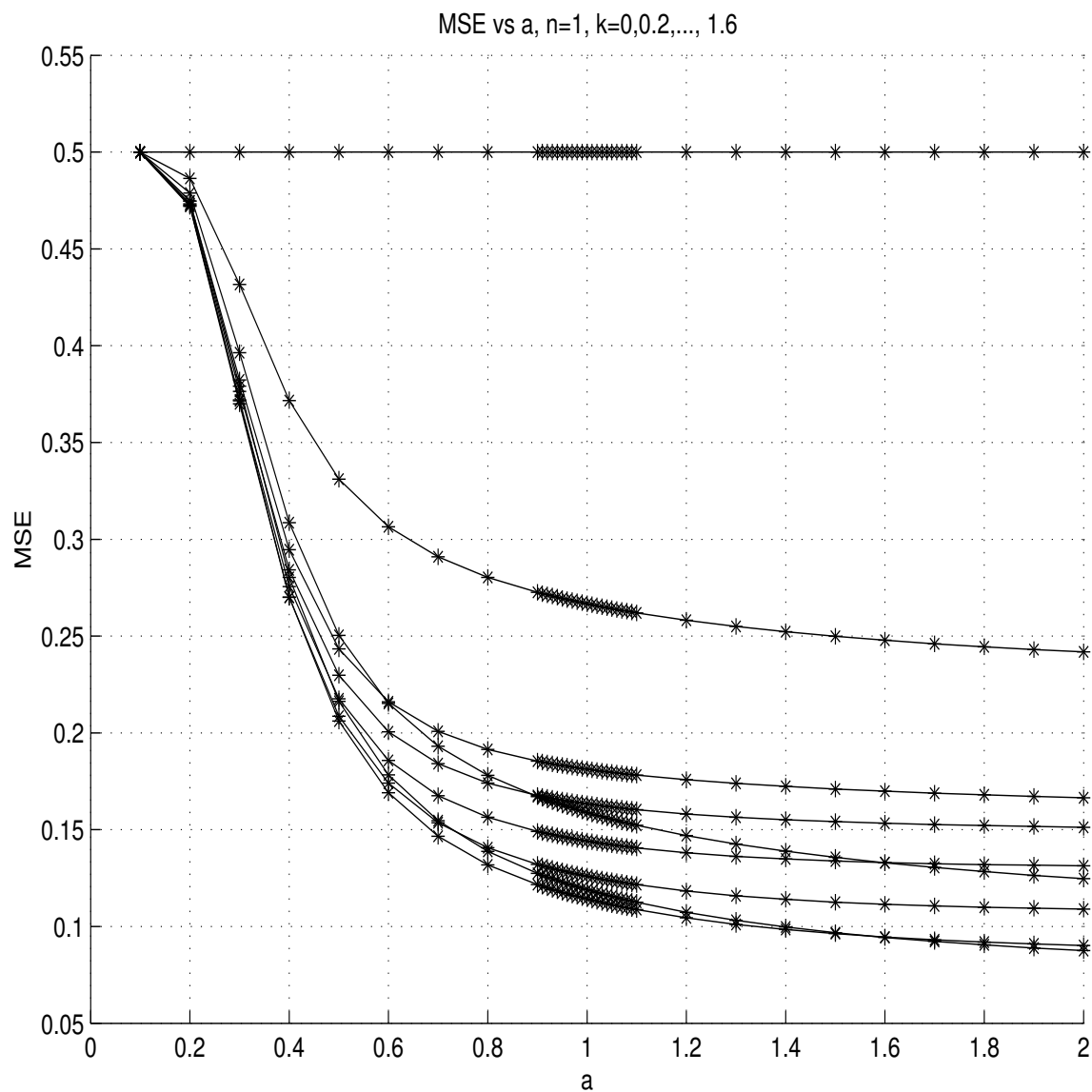
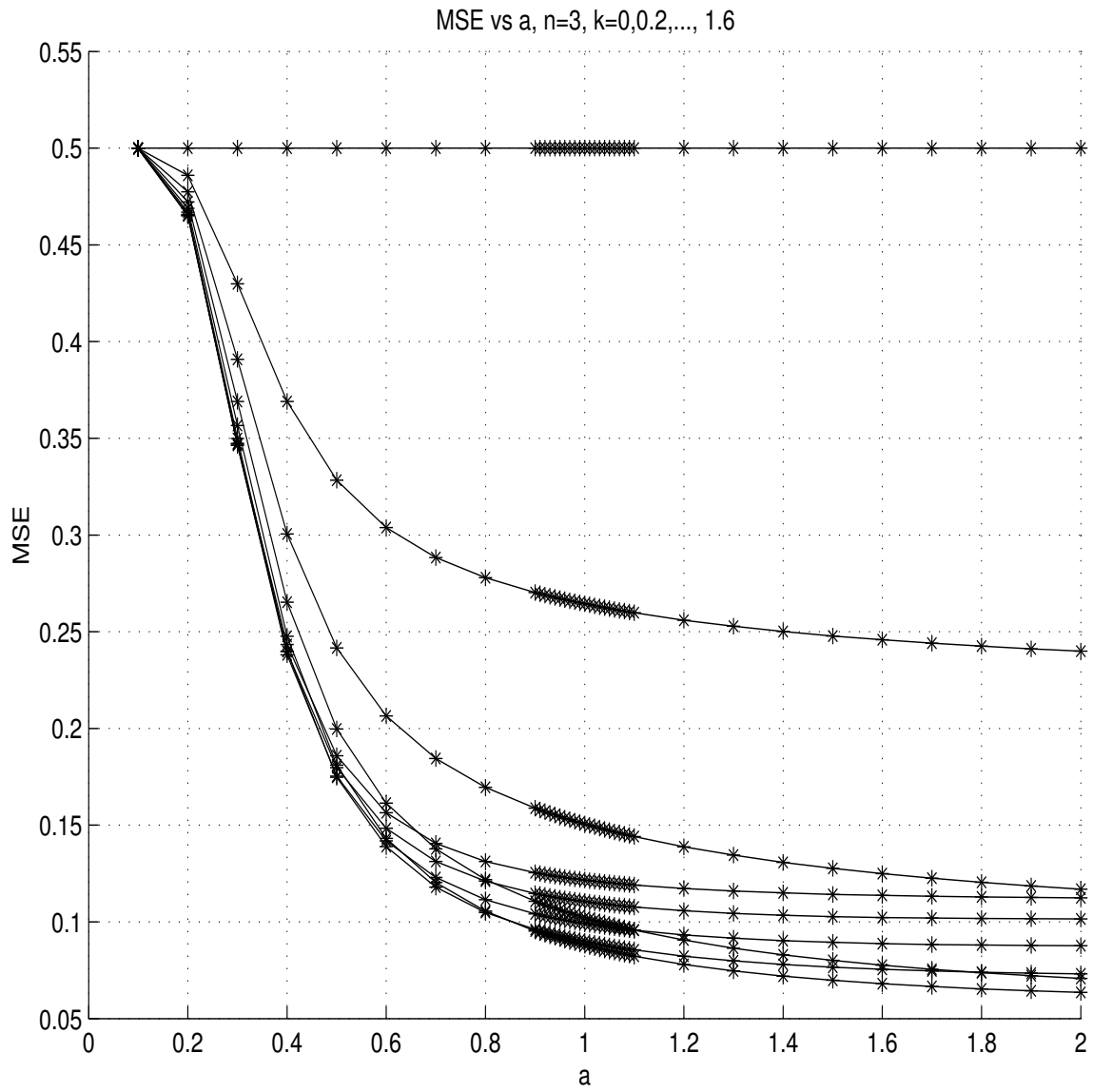
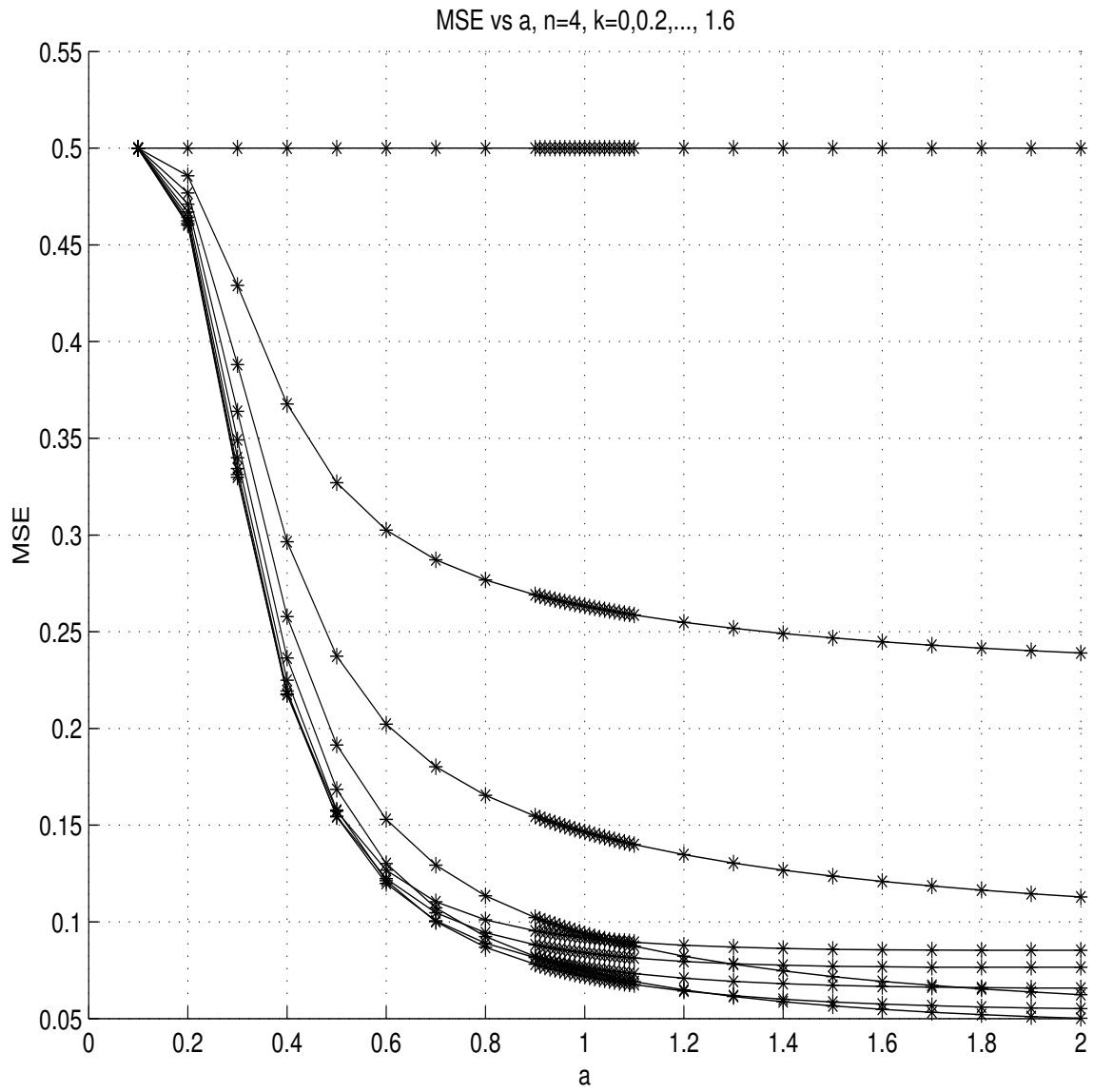


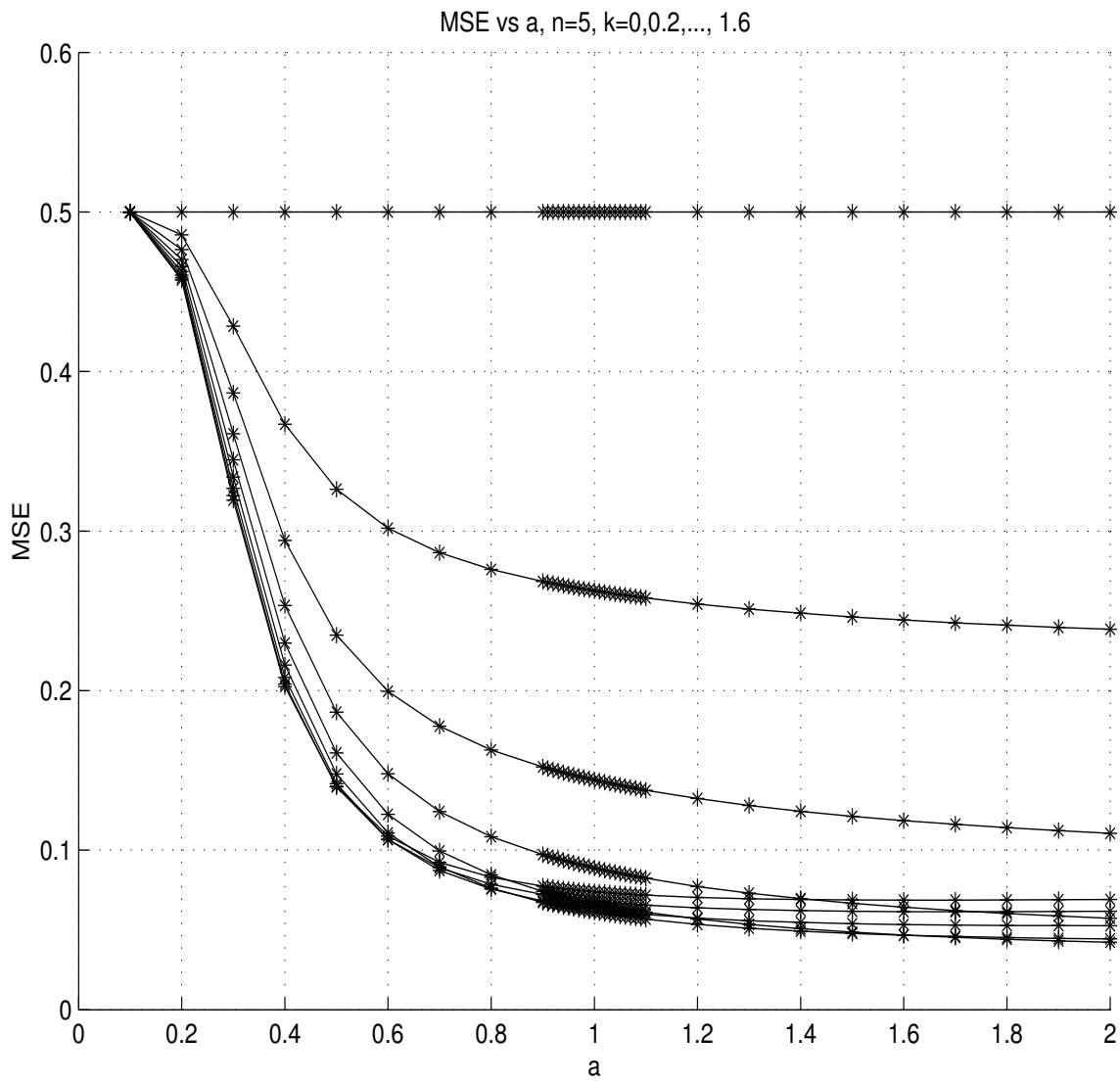
Fig. 24. $\sigma^2 = 1$, $n=15$

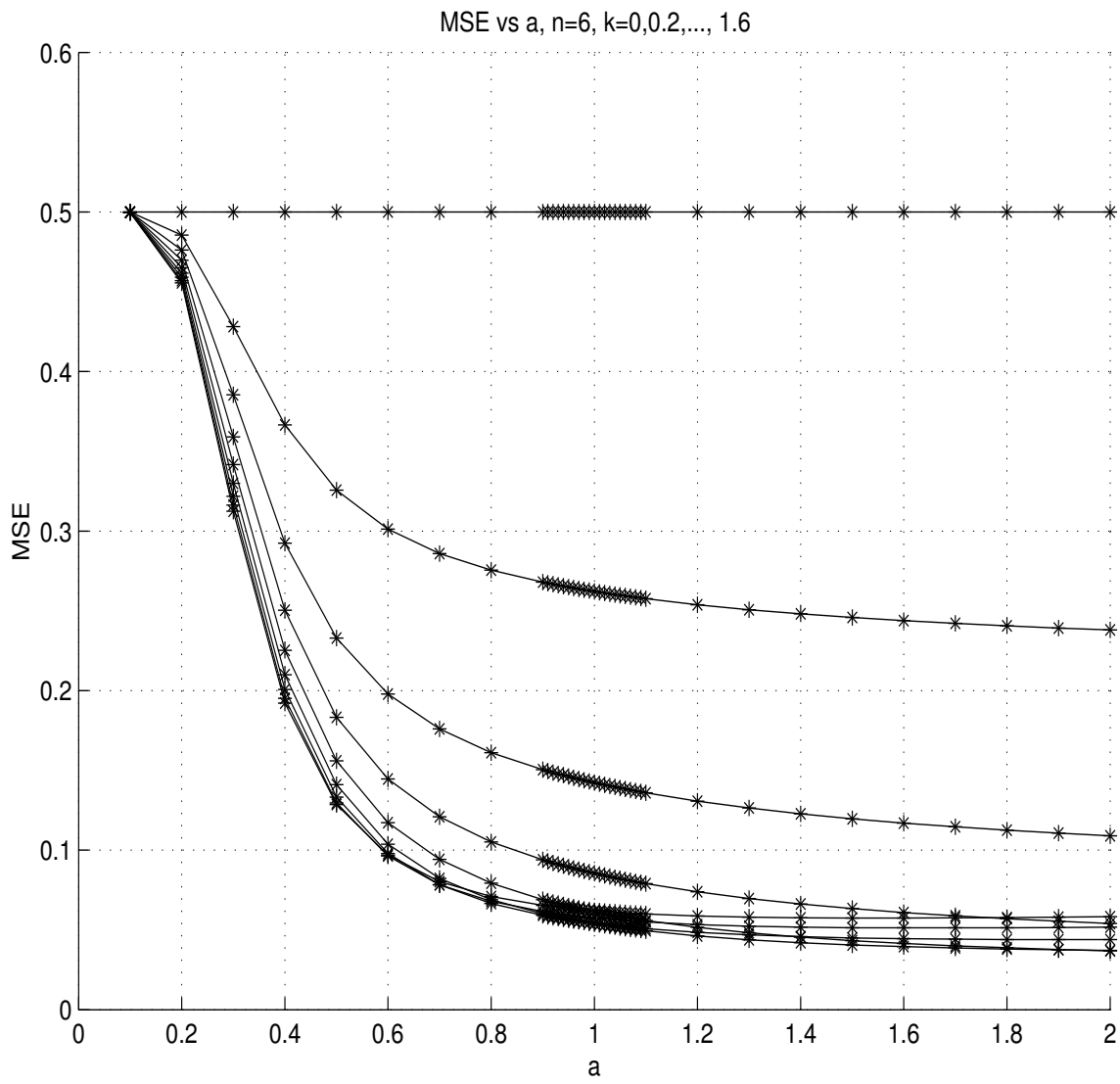
Fig. 25. $\sigma^2 = 1, n=20$

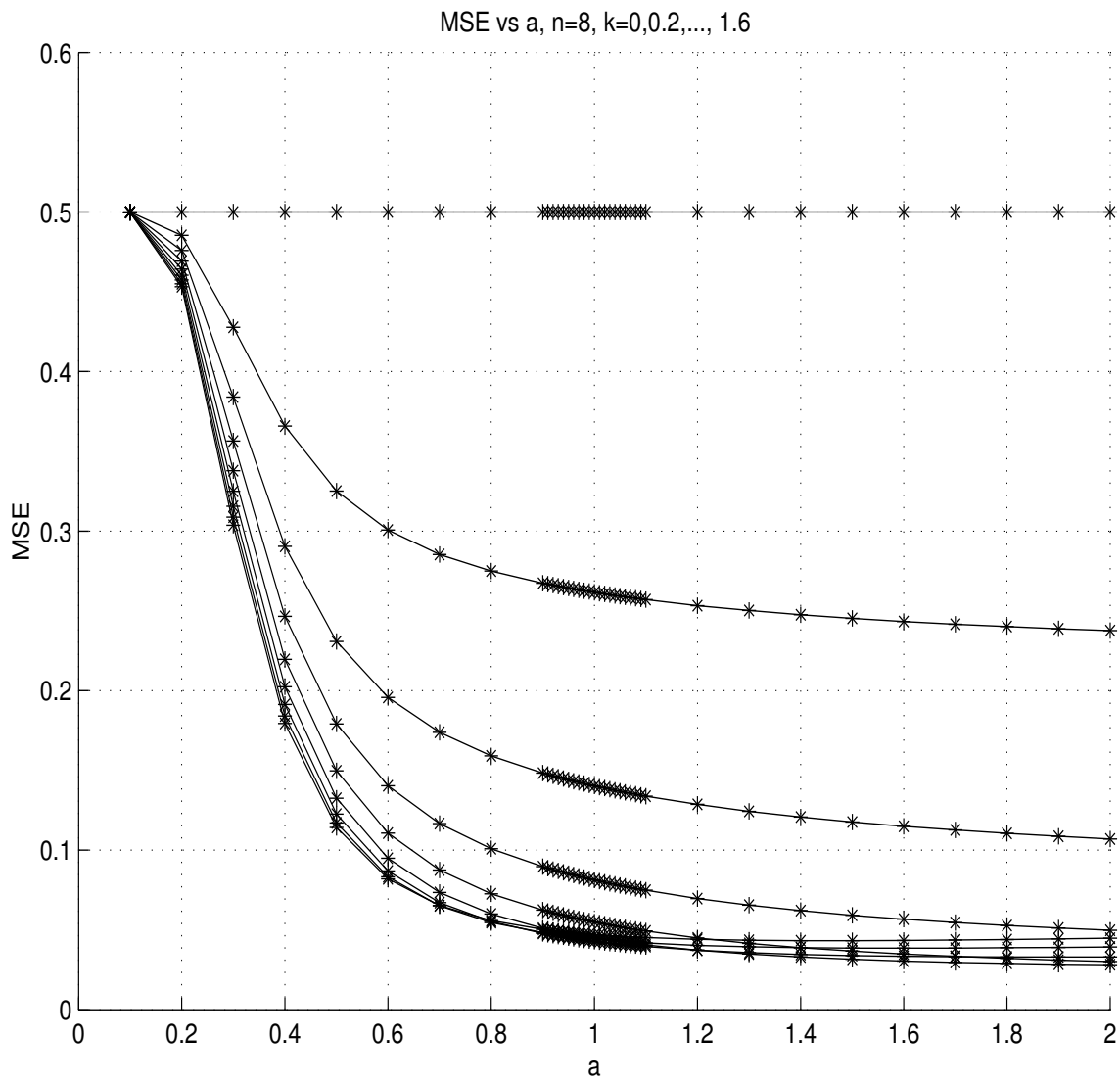
Fig. 26. $\sigma^2 = 0.5$, n=2

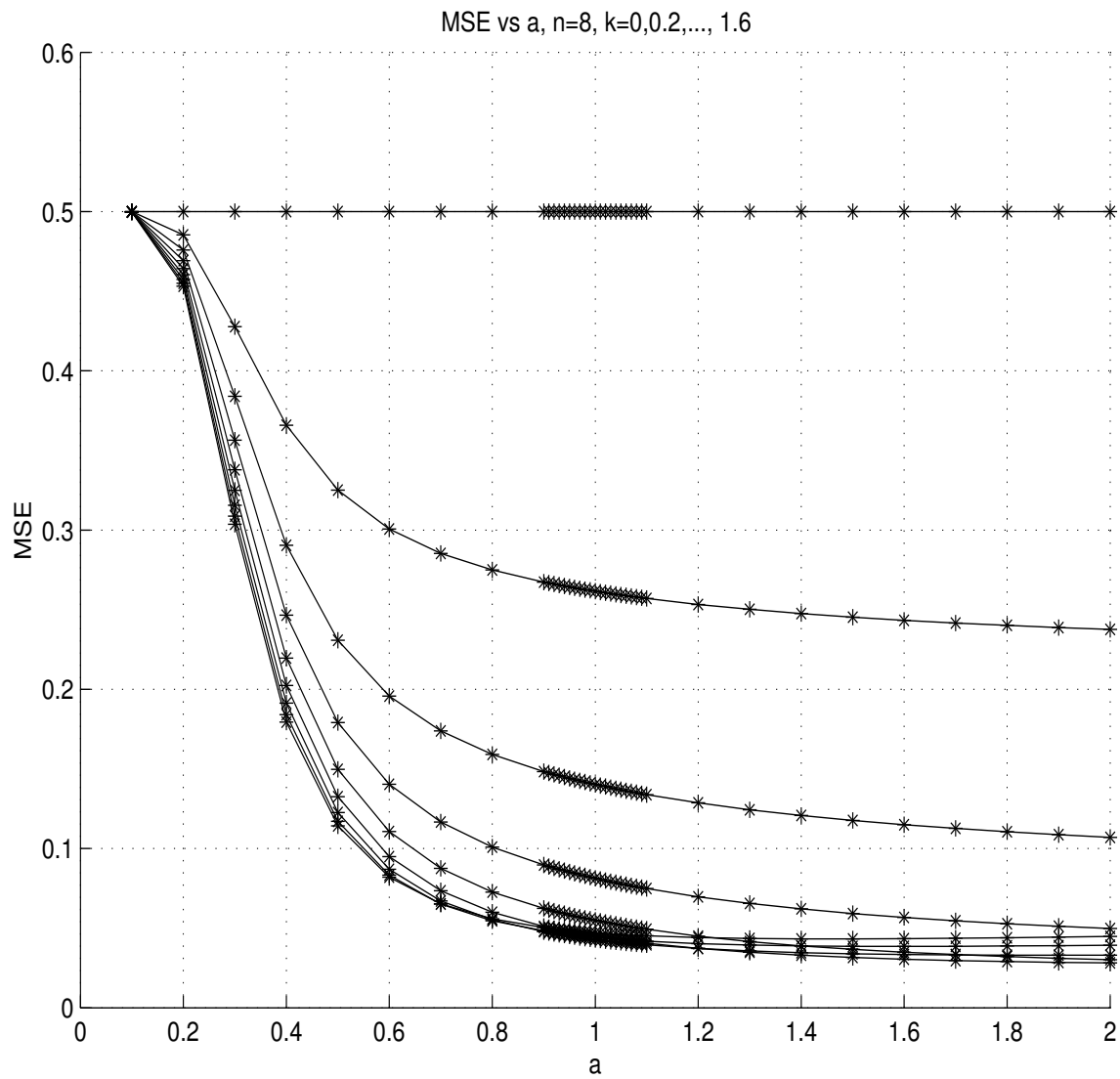
Fig. 27. $\sigma^2 = 0.5$, n=3

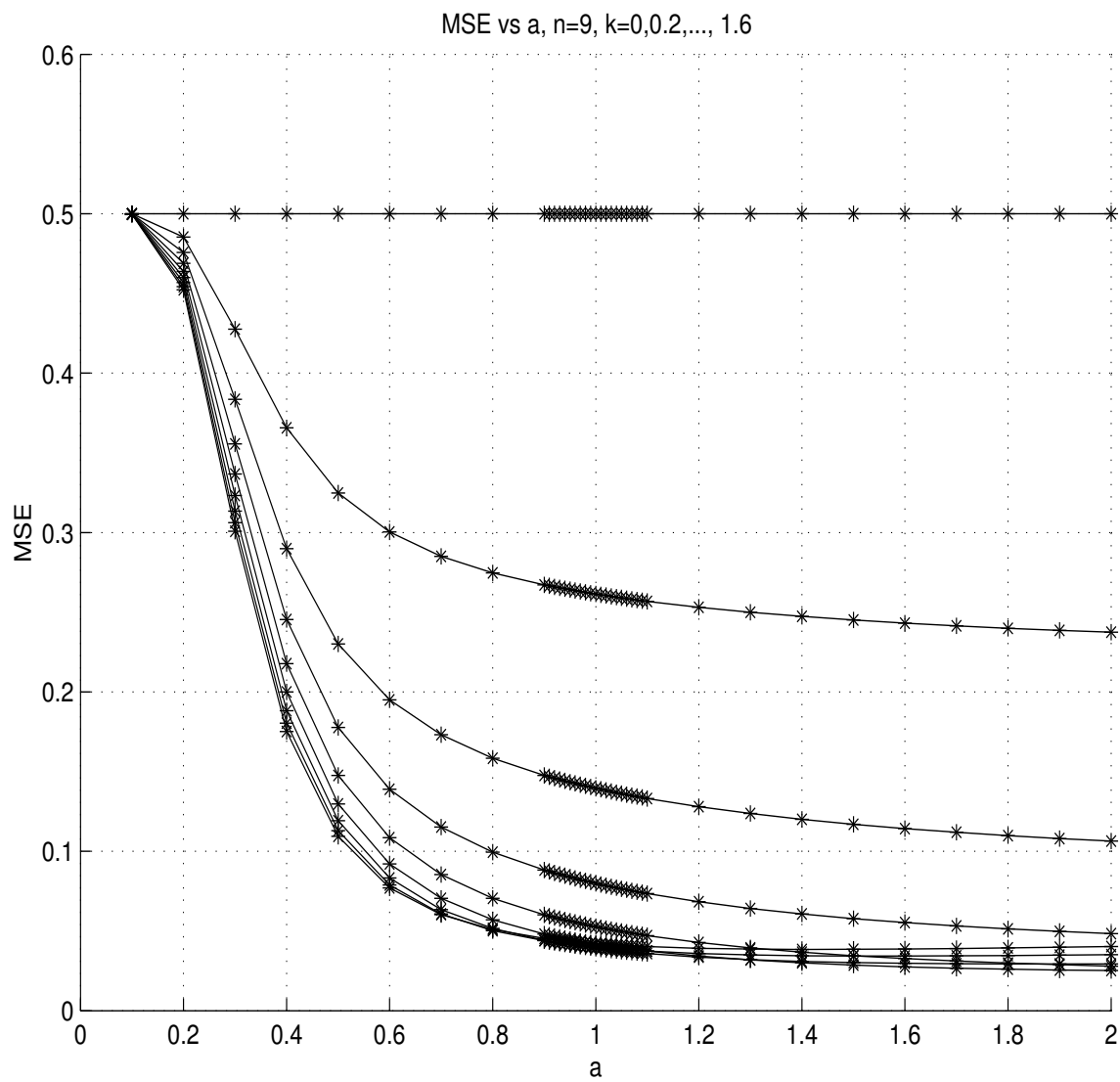
Fig. 28. $\sigma^2 = 0.5$, $n=4$

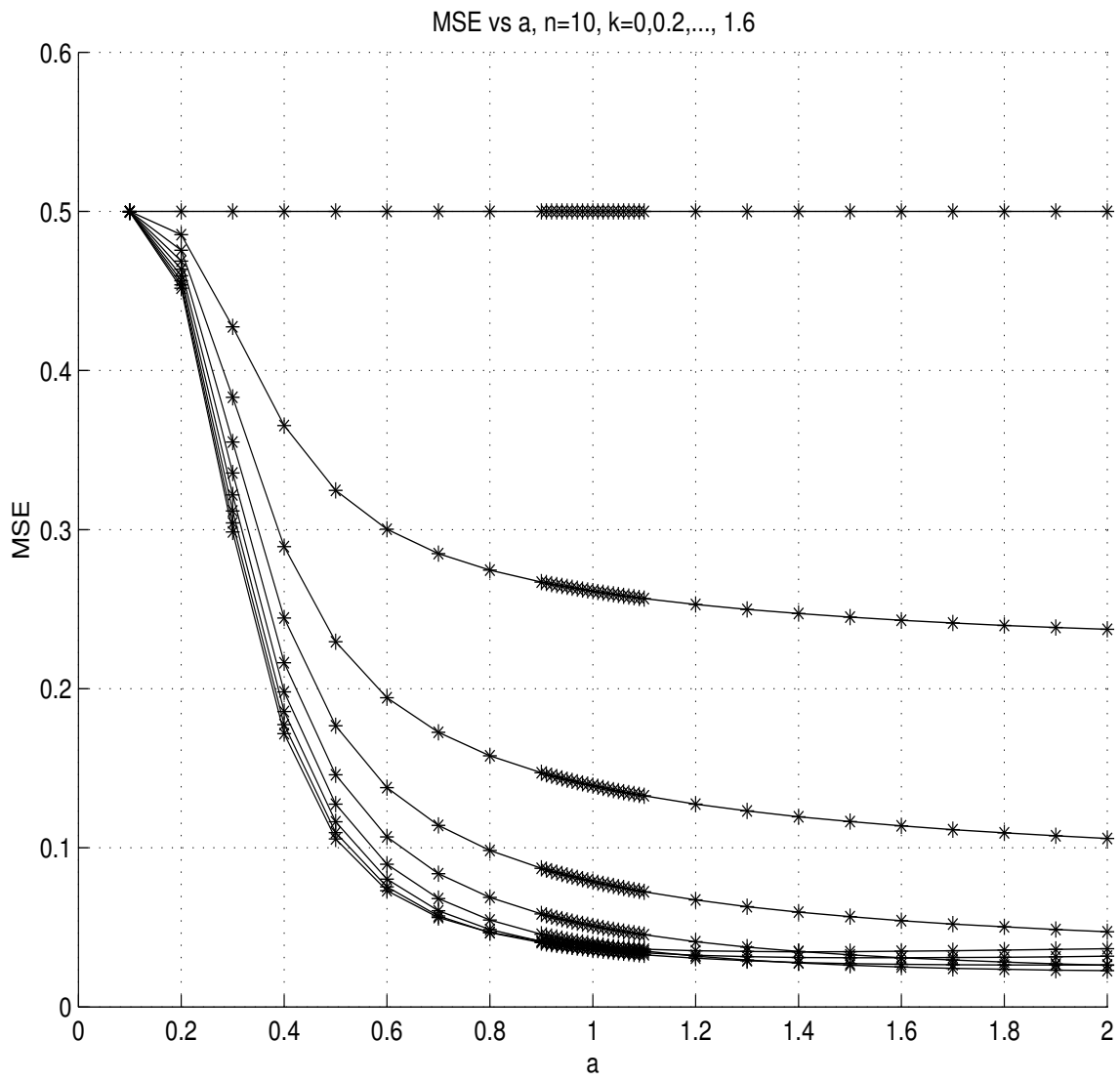
Fig. 29. $\sigma^2 = 0.5$, n=5

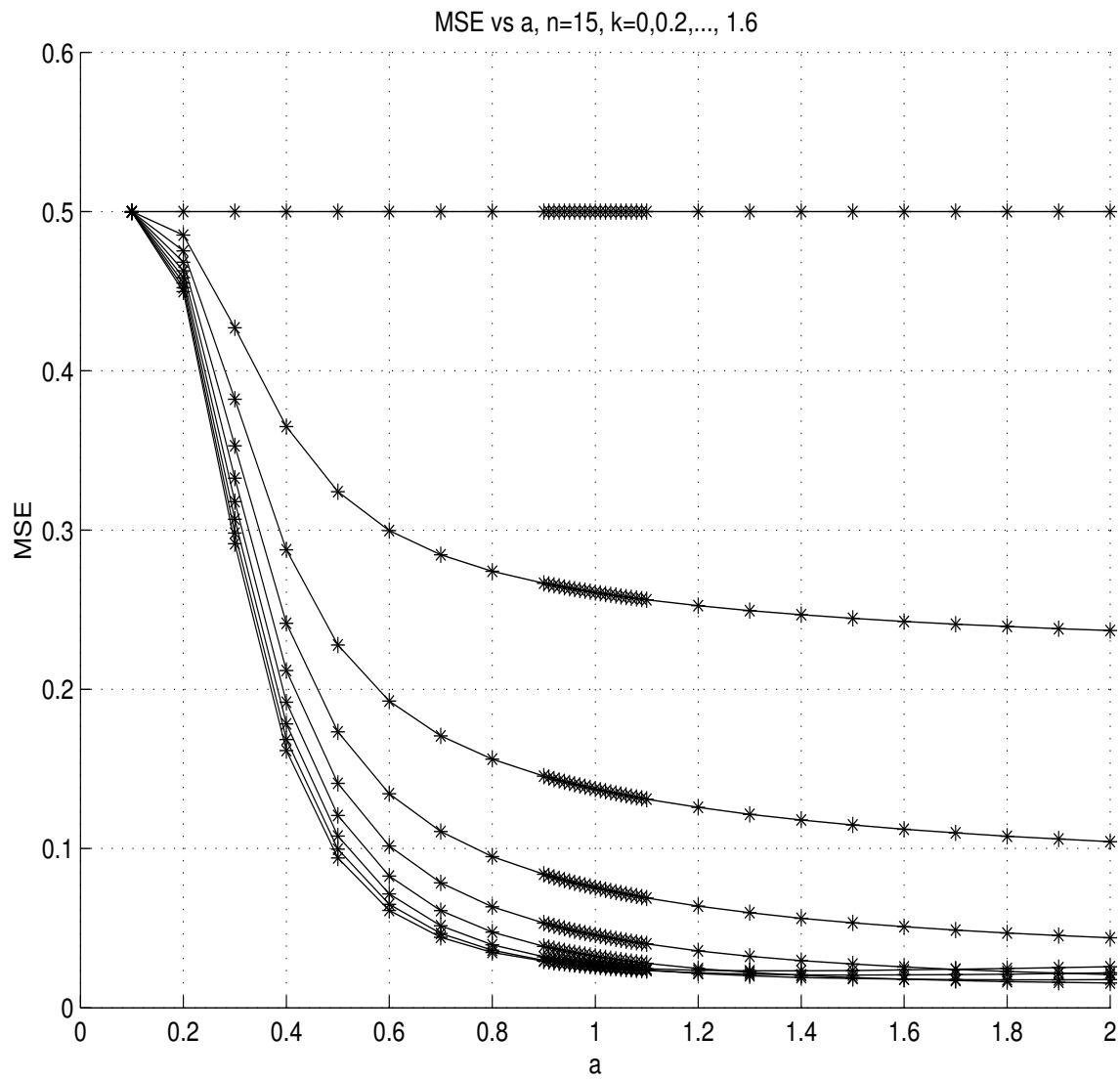
Fig. 30. $\sigma^2 = 0.5, n=6$

Fig. 31. $\sigma^2 = 0.5$, n=7

Fig. 32. $\sigma^2 = 0.5$, $n=8$

Fig. 33. $\sigma^2 = 0.5$, $n=9$

Fig. 34. $\sigma^2 = 0.5$, $n=10$

Fig. 35. $\sigma^2 = 0.5$, n=15

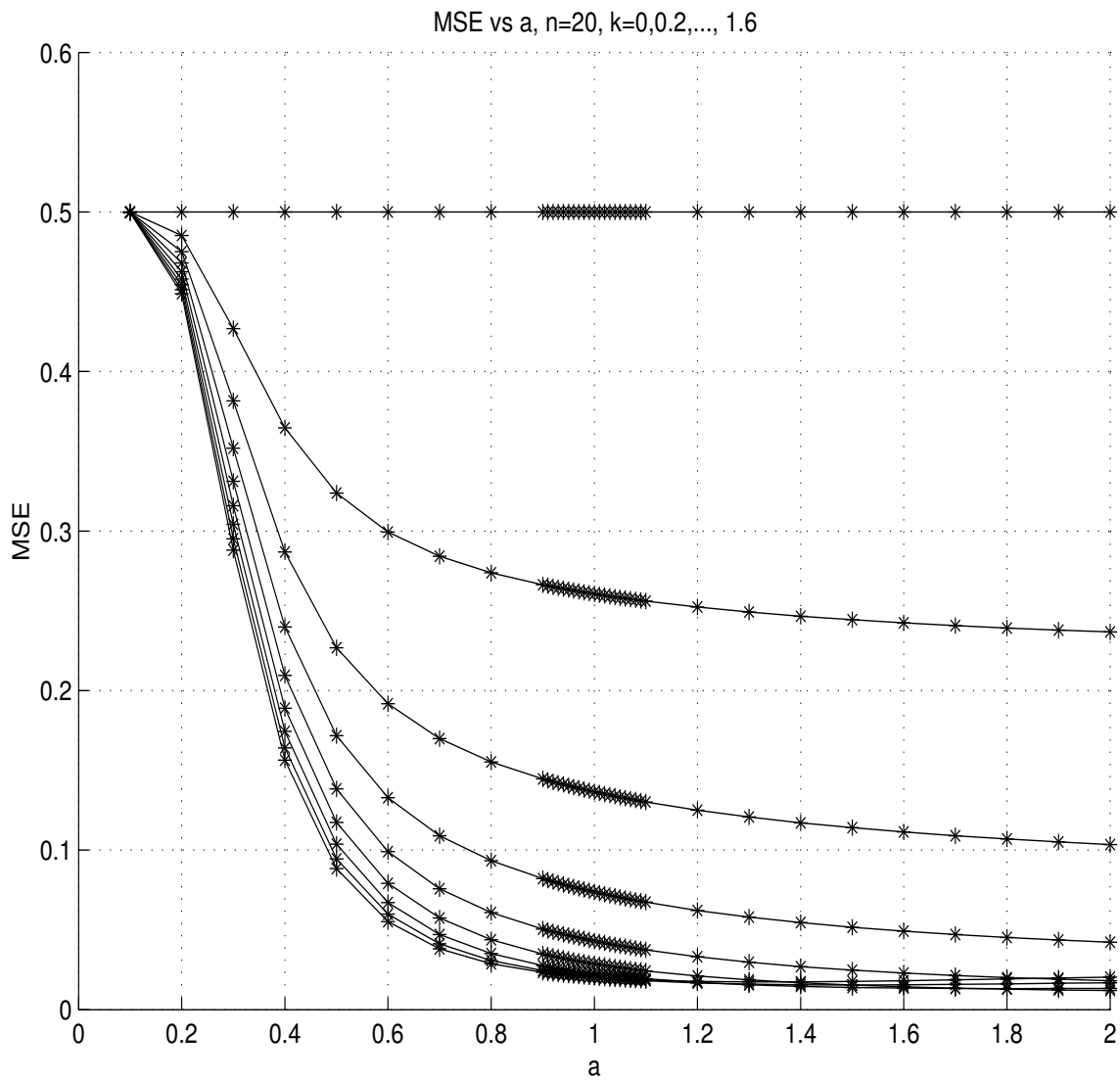
Fig. 36. $\sigma^2 = 0.5$, $n=20$

Table XXVI. k as a function of ϵ , which maximizes J for $\sigma^2 = 0.5$, and $\alpha = 0.02$

ϵ	k									
	n=2	n=3	n=5	n=6	n=7	n=8	n=9	n=10	n=15	n=20
0.00	0.70	0.90	1.00	1.10	1.20	1.20	1.30	1.30	1.50	1.60
0.10	0.80	0.90	1.10	1.20	1.20	1.30	1.40	1.40	1.60	1.60
0.20	0.80	0.90	1.20	1.30	1.50	1.60	1.60	1.60	1.60	1.60
0.30	0.80	1.00	1.60	1.60	1.60	1.60	1.60	1.60	1.60	1.60
0.40	0.90	1.50	1.60	1.60	1.60	1.60	1.60	1.60	1.60	1.60
0.50	1.40	1.60	1.60	1.60	1.60	1.60	1.60	1.60	1.60	1.60
0.60	1.50	1.60	1.60	1.60	1.60	1.60	1.60	1.60	1.60	1.60
0.70	0.10	1.60	1.60	1.60	1.60	1.60	1.60	1.60	1.60	1.60
0.80	0.10	0.10	1.60	1.60	1.60	1.60	1.60	1.60	1.60	1.60
0.90	0.10	0.10	1.60	1.60	1.60	1.60	1.60	1.60	1.60	1.60
1.00	0.10	0.10	0.10	0.10	0.10	0.10	0.10	1.60	1.60	1.60

C. Conclusions

If the user knows the number of samples received, and can pick ϵ depending on the application under consideration, then the user can easily design the estimator by looking up the value of censor k . Also, a very interesting general trend can be found in the following tables, for the cases of medium ($\sigma^2 = 1$) and large ($\sigma^2 = 2$) variances: regardless of the sample size for ($\sigma^2 = 1$) and ($\sigma^2 = 2$) we find that essentially only two values of k are required for each case as ϵ varies. For $\epsilon \leq 0.5$, k is so large that little censoring is needed (this value of k is approximately 1.5 but varies somewhat), where as for $\epsilon \geq 0.6$ a large amount is required (k on the order of 0.1). This is very fortunate because the user need therefore not be precise in specifying ϵ . On the other hand, for smaller σ^2 , e.g. $\sigma^2 = 0.5$, this is not the case and there exist examples where little censoring is required for all ϵ .

REFERENCES

- [1] S.M. Stigler, "Simon Newcomb, Percy Daniell, and the history of robust estimation 1885-1920," *J. AmeriStatist. Assoc*, vol. 68, pp.872-879, (1973).
- [2] J.L. Hodges, and E.L. Lehman, " The efficiency of some nonparametric competitors of the t-test," *Ann. Math. Statist.* 27 pp 324-338, (1956).
- [3] J.W. Tuckey, A survey of sampling from contaminated distribution, in I. Olkin, et al., (eds,) *Contribution to Probability and Statistics*. Palo Alto, Calif: Stanford University press pp. 448-485,(1960).
- [4] P.J. Huber, "Robust estimation of a local parameter," *Ann Math. Stat.*, vol.35, pp.73-101, March 1964.
- [5] P.J. Huber, *Robust Statistics*. New York: Wiley, 1981.
- [6] D.R. Halverson and H.C. Lee, "Robust detection and estimation with a differential geometric approach," *Proceedings of the Conference on Information Sciences and Systems*, Princeton University, March 2000.
- [7] M.W. Thomson, D.R. Halverson, and C. Tsai, "Robust estimation of signal parameters with dependent nonstationary and/or dependent data," *IEEE Transactions Information Theory*, vol. IT-39, pp. 617-623, March 1993.

VITA

Pankaj Bhagawat was born on November 9, 1977 in India. He obtained the Bachelor of Engineering degree in electronics and communication engineering from the Regional Engineering College, Tiruchirapalli, India in May 2000. He joined the telecommunications and signal processing group in the Electrical Engineering Department at Texas A&M University in Fall 2000 to pursue the Master of Science degree.

Permanent Address:

Bungalow no. CL VII 2/A

Ordnance Estate Ambernath

Ambernath 421502

Maharashtra

India