DOCUMENT RESUME

ED 282 909 TM 870 321

AUTHOR	Cope, Ronald T.; Kolen, Michael J.
TITLE	A Study of Methods for Estimating Distributions of Test Scores.
PUB DATE	Apr 87
NOTE	28p.; Paper presented at the Annual Meeting of the American Educational Research Association (Washington, DC, April 20-24, 1987).
PUB TYPE	Speeches/Conference Papers (150) -- Reports - Research/Technical (143)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	College Entrance Examinations; *Estimation (Mathematics); Higher Education; Mathematical Models; *Nonparametric Statistics; *Scores; Scoring; Secondary Education; *Statistical Distributions; Statistical Studies; Test Norms; *Test Theory
IDENTIFIERS	ACT Assessment; Density; Equipercentile Equating; *Score Distribution; *Smoothing Methods

ABSTRACT

This study compared five density estimation techniques applied to samples from a population of 272,244 examinees' ACT English Usage and Mathematics Usage raw scores. Unsmoothed frequencies, kernel method, negative hypergeometric, four-parameter beta compound binomial, and Cureton-Tukey methods were applied to 500 replications of random samples of 500, 1000, 2000, and 5000 from these populations. The four-parameter beta compound binomial produced the most accurate estimates, and the kernel method yielded only slightly less accurate estimates. Cureton-Tukey ranked third in accuracy. All methods involving smoothing produced more accurate estimates than unsmoothed frequencies except the negative hypergeometric. Negative hypergeometric estimates varied erratically by test and score level. The methods studied have the potential to improve the estimation of norms and the equipercentile equating function. (Author)

A Study of Methods for Estimating Distributions of Test Scores

Ronald T. Cope

Michael J. Kolen

The American College Testing Program

Paper presented at the Annual Meeting of the American Educational Research Association

Washington, D.C., April, 1987

Abstract

This study compared five density estimation techniques applied to samples from a population of 272,244 examinees' ACT English Usage and Mathematics Usage raw scores. Unsmoothed frequencies, kernel method, negative hypergeometric, four-parameter beta compound binomial, and Cureton-Tukey methods were applied to 500 replications of random samples of 500, 1000, 2000, and 5000 from these populations. The four-parameter beta compound binomial produced the most accurate estimates, and the kernel method yielded only slightly less accurate estimates. Cureton-Tukey ranked third in accuracy. All methods involving smoothing produced more accurate estimates than unsmoothed frequencies except the negative hypergeometric. Negative hypergeometric estimates varied erratically by test and score level. The methods studied have the potential to improve the estimation of norms and the equipercentile equating function.

Key Words: Nonparametric density estimation, test score models, smoothing norms, equating

A Study of Methods for Estimating Distributions of Test Scores

Statisticians have traditionally taken a parametric approach to estimating a probability density function from sample data: assume or try to deduce the function (e.g., binomial, beta, normal, Poisson), then estimate function parameters from the sample statistics. Only recently have they actively cultivated a nonparametric approach (Silverman, 1986; Tapia & Thompson, 1978) involving few or no assumptions about the function. Yet already one finds a considerable body of theory and methods of nonparametric density estimation.

Nonparametric methods show promise for estimating test score distributions from sample data. Here we adapt one of them—the kernel method—to estimating discrete test score distributions of ACT English Usage and Mathematics Usage tests. Another nonparametric method, the Cureton-Tukey weighted moving average method (Cureton & Tukey, 1951), is also studied. We compare results by these methods to those from two parametric methods: the negative hypergeometric (Lord, 1965) and four-parameter beta compound binomial test score models (Keats & Lord, 1962; Lord & Novick, 1968, chap. 23). These methods have the potential to improve the estimation of test norms and the equipercentile equating function.

Density Estimation Techniques

Four techniques for estimating population densities are described, in addition to the sample relative frequencies. All of these techniques produce discrete density estimates.

Negative Hypergeometric Distribution (Beta Binomial)

The negative hypergeometric distribution was described by Keats and Lord (1962) and was discussed by Lord and Novick (1968). Lord and Novick (1968) present a procedure for generating the distribution given the mean and the variance for a test of a given length. One way to derive the negative hypergeometric is to assume that proportion-correct true scores have a two-parameter beta distribution, ranging from 0 to 1, and that for examinees of a given proportion-correct true score, observed scores are distributed binomial with parameters equal to the number of items and proportion-correct true score. The observed score distribution over all examinees that results from this process is the negative hypergeometric. The negative hypergeometric distribution is often said to be the observed score distribution arising from the beta binomial model.

The negative hypergeometric is a discrete unimodal distribution. If the mean proportion-correct score is below .5 then the distribution is positively skewed, and if the mean is above .5 then the distribution is negatively skewed. Keats and Lord (1962) and Lord and Novick (1968) showed that the negative hypergeometric can fit many test score distributions very well.

Four Parameter Beta Compound Binomial Method

To improve the fit to data, Lord (1965) generalized the beta binomial model. He used a four parameter beta distribution for proportion-correct true scores rather than a two parameter beta distribution. This four parameter beta distribution has parameters for the high and low proportion-correct true scores in addition to the two parameters used to describe the two parameter beta distribution. The low parameter is allowed to be greater than zero and the high parameter less than one. A lower bound for true scores that is above zero seems especially sensible for multiple choice tests, where an examinee can correctly answer a substantial proportion of items through random guessing.

In this model, Lord (1965) used a two-term approximation to the compound binomial distribution for observed scores given true score. Lord and Novick (1968, p. 525) suggested that the compound binomial may be more realistic than the binomial for this situation. Practically speaking, one major difference between the binomial and the two term approximation to the compound binomial is that the latter typically has smaller variance. The observed score distribution under this model is the four parameter beta compound binomial distribution.

The four parameter beta compound binomial distribution is unimodal. It is more general than the negative hypergeometric. For instance, it can be positively skewed even if the mean proportion-correct score is above .5.

Lord (1965) presented a method for estimating the parameters of this distribution that is based on the method of moments, and the observed score distribution is computed analytically. In implementing the method of moments, sometimes the estimate of the high parameter exceeds 1. In such cases, the high parameter is fixed at 1.0 and the remaining three parameters are estimated by the method of moments.

Cureton-Tukey Estimation

Cureton and Tukey (1951) described a method in which the estimated relative frequency for a given score is found by taking a weighted average of the relative frequencies at that score and at surrounding scores. A method using seven relative frequencies in the averaging procedure was used here. For a relative frequency at score x, $\hat{f}(x)$, the smoothed relative frequency, $\hat{f}_s^*(x)$, is taken as $[-2\hat{f}(x - 3) + 3\hat{f}(x - 2) + 6\hat{f}(x - 1) + 7\hat{f}(x) + 6\hat{f}(x + 1) + 3\hat{f}(x + 2) - 2\hat{f}(x + 3)]/21$. According to Angoff (1982), these weights were chosen to preserve "the parabolic and cubic trends within successive sets of points" (p. 68).

This procedure sometimes produces negative relative frequencies near the extremes. When negative relative frequencies occur, they are set to zero. In addition, sometimes the weights are supposed to be applied to scores outside the range of possible scores on the test. For example, on a 40-item test, f(43) would be involved in finding $f_s^*(40)$. It was assumed in this procedure that relative frequencies outside the range of possible scores were zero. The smoothing process sometimes results in $\sum \hat{f}_s^* \neq 1$. For this reason, we define $\hat{f}_s(x) = \hat{f}_s^*(x) / \sum_{x=0}^{k} \hat{f}_s^*(x)$.

<u>Kernel Estimation</u>

The kernel estimator was proposed by Rosenblatt (1956). The idea behind kernel estimation is to spread out the density of each observed score point using a probability density function. This probability density function is referred to as the kernel. The kernel estimator has been used most often with continuous data, and the normal distribution is often used as the kernel. In kernel estimation, a parameter is manipulated which controls the degree of smoothing. Silverman (1986) described in detail the use of the kernel estimator with continuous data.

In this paper, a kernel estimator is developed for discrete raw test score distributions. This estimator uses a binomial kernel to produce a discrete density estimate. The parameter H is an even integer that is the binomial "number of trials" parameter. H is set by the investigator, and larger values of H result in more smoothing. The "probability of success" binomial parameter is .5. For a test with K items and an observed relative frequency distribution $\hat{f}(x)$, x = 0, 1, ..., K, this kernel estimator is

$$\hat{f}^*_s(x) = \sum_{i=0}^{K} \text{Bin}(i - x + H/2|H, .5)\ \hat{f}(i) , \tag{1}$$

$x = 0, 1, \ldots, K$, H is an even integer, and

$$\text{Bin}(y|H,.5) = \begin{cases} \binom{H}{y} .5^y(1 - .5)^{H-y}, & y = 0, 1, \ldots, H \\ 0 & , \text{otherwise.} \end{cases} \tag{2}$$

Because $\sum_{x=0}^{K} \hat{f}^*_s(x)$ does not necessarily equal one, the estimator in Equation 1 is adjusted, and this adjusted kernel estimator is

$$\hat{f}_s(x) = \hat{f}^*_s(x) / \sum_{x=0}^{K} \hat{f}^*_s(x) . \tag{3}$$

To better understand Equations 1 and 2, first consider the special case when $H = 0$. In this case, $\text{Bin}(i - x + H/2|0, .5) = \text{Bin}(i - x|0, .5)$. By Equation 2, if $i = x$ then $\text{Bin}(0|0, .5) = 1$, and if $i \neq x$ then $\text{Bin}(0|0, .5) = 0$. Thus, for all x, when $H = 0$, $\hat{f}^*_s(x) = \hat{f}(x)$. That is, the observed relative frequency distribution is the kernel estimator when $H = 0$.

Now consider the case when $H = 2$. From Equation 2, $\text{Bin}(0|2, .5) = .25$, $\text{Bin}(1|2, .5) = .50$, and $\text{Bin}(2|2, .5) = .25$. All other values of Bin are 0 when $H = 2$. From Equation 1, if $H = 2$ and $i = x$, then $i - x + H/2 = 1$ and $\text{Bin}(1|2, .5) = .50$. Similarly, if $i - 1 = x$ or if $i + 1 = x$, then Bin = .25. For all other values of i, Bin = 0. Thus, $\hat{f}^*_s(x) = .25\ \hat{f}(x - 1) + .50\ \hat{f}(x) + .25\ \hat{f}(x + 1)$, where $\hat{f}(x)$ is defined to be zero for $x < 0$ or $x > K$. This indicates that $\hat{f}^*_s(x)$ can be written as a weighted sum of relative frequencies.

So far, we have suggested two interpretations of kernel estimators. One is that kernel estimators spread the density at each score point to other score points. The second is that the estimated density is a weighted sum of the observed densities. This second interpretation would suggest that, for discrete distributions, the Cureton-Tukey method presented earlier is similar to the kernel estimator. Actually, the only reason that the Cureton-Tukey estimator cannot qualify as a kernel estimator is because it uses negative weights. Both estimators are in the class of estimators described by Silverman (1986) as general weight function estimators.

A hypothetical example of the kernel method with H = 2 is presented in Table 1. First, assume the test has 5 items and there are 10 examinees. The third column of the table shows the computations involved to estimate each $\ddot{f}_s$ by Equations 1 and 2. Note that the weight .5 is applied to the relative frequency at the point and .25 to the two adjacent points, which suggests the weighted sum of the relative frequencies interpretation of f_s . Now focus on the .4 relative frequency at a score of 3. As can be seen, a relative frequency of .25(.4) = .1 is spread to scores of 2 and 4 and .5(.4) = .2 is kept at a score of 3, which suggests the spreading of density interpretation. The adjusted estimates if the test would have had 4 items are shown in the rightmost column in Table 1. In this case, each relative frequency in the fourth column of the table was multiplied by 1/.925.

Insert Table 1 about here

Because a binomial kernel with a parameter of .5 is used, there is an interesting relationship that involves repetition of the kernel procedure. Consider a situation where Equations 1 and 2 are applied first to the original relative frequencies and then again to the smoothed relative frequencies. The resulting distribution will be the same as that which would have been obtained by applying Equations 1 and 2 once with H equal to the sum of the two H's in the repeated application. For example, using H = 4 will result in the same smoothed distribution as applying H = 2 twice.

Illustration

To illustrate the results produced by the methods, each was applied to a frequency distribution of ACT Mathematics scores based on 3,039 examinees. This test has 40 multiple-choice items. The results are shown in Figure 1. In this figure, the observed frequency distribution is represented by a solid curve and the fitted distributions by a dotted curve.

Insert Figure 1 about here

The negative hypergeometric appears to fit poorly. The fitted frequencies are too high at the very low scores and at middle scores above 20 and too low at other score points. The observed distribution is positively skewed with a mean above .5, while the fitted distribution is nearly symmetric, which may be part of the reason for the apparent poor fit.

The four parameter beta compound binomial appears to fit this distribution very well. The Cureton-Tukey fitted distribution is close to the observed distribution. However, it is not very smooth. This is a problem we have often noted with the Cureton-Tukey method.

The kernel method is shown with H = 4, 8, 16, and 32. The distributional fit with H = 4 stays reasonably close to the observed distribution, although the fitted distribution is somewhat bumpy. As H is increased the fitted distribution becomes less bumpy, although it departs more from the observed distribution. For H = 16 and H = 32, the fitted frequencies are above the observed frequencies at the lower scores.

Overall, the negative hypergeometric appears to fit this Mathematics distribution poorly, the four parameter beta compound binomial appears to fit very well, the Cureton-Tukey method fitted distribution is not very smooth, and the kernel method seems promising.

Comparing the Methods

Mathematics and English test score distributions from a recent October administration of the ACT Assessment to 272,244 examinees were used to compare the methods. The Mathematics test contains 40 five-alternative multiple choice test questions, and the English test contains 75 four-alternative multiple choice questions.

Comparison Methodology

The relative frequency distribution for the 272,244 examinees was considered to be the population density. The following procedure was used to evaluate the methods:

1. Draw a random sample of size N from the population density $f(x)$, $x = 0, 1, \ldots, K$, and refer to this sample as replication r.
2. Construct the observed relative frequency distribution $f_r(x)$, $x = 0, 1, \ldots, K$.
3. Estimate the relative frequencies using each of the techniques described earlier, and refer to this estimated relative frequency as $f_{rs}(x)$, $x = 0, 1, \ldots, K$.
4. Repeat steps 1-3 R times.

This process was repeated for N = 500, 1000, and 5000, each with R = 500 replications. The Cureton-Tukey, negative hypergeometric, four parameter beta compound binomial (4PB), and kernel methods were used in step 3.

The following statistics were calculated at each x for each method, including the observed frequencies:

$$\text{Bias}_x^2 = \left[\sum_{r=1}^{R} \hat{f}_{rs}(x)/R - f(x) \right]^2 , \tag{4}$$

$$\text{Variance}_x = \sum_{r=1}^{R} \left[\hat{f}_{rs}(x) - \sum_{r=1}^{R} \hat{f}_{rs}(x)/R \right]^2 /R , \text{ and} \tag{5}$$

$$\text{MSE}_x = \sum_{r=1}^{R} \left[\hat{f}_{rs}(x) - f(x) \right]^2 /R = \text{Bias}_x^2 + \text{Variance}_x . \tag{6}$$

In addition, statistics over all score points were calculated as

$$\text{Bias}^2 = \sum_{x=0}^{K} \text{Bias}_x^2/(K + 1) , \tag{7}$$

$$\text{Variance} = \sum_{x=0}^{K} \text{Variance}_x/(K + 1) , \text{ and} \tag{8}$$

$$\text{MSE} = \sum_{x=0}^{K} \text{MSE}_x/(K + 1) . \tag{9}$$

The Equation 4 through 9 statistics are based on the estimation of relative frequencies, and can be viewed as adaptations, to discrete distributions, of the integrated root mean squared framework for evaluating distributional fit described by Silverman (1986).

A statistic based on relative cumulative frequencies also was used, because relative cumulative frequencies typically are the basis for calculating norms and for equipercentile equating. Only an overall statistic was calculated which is

$$\text{K-S} = \sum_{r=1}^{R} \sup_x [\hat{F}_{rs}(x) - F(x)]/R . \tag{10}$$

The K-S statistic in Equation 10 is an adaptation of the Komolgorov-Smirnov statistic. For a given replication, $\hat{F}_{rs} = \sum_{i=0}^{x} \hat{f}_{rs}(i)$, which is the relative frequency at x. $F(x) = \sum_{i=0}^{x} f(i)$, the distribution function value at x. $\sup_x$ is the supremum over x. Thus, in Equation 10 the greatest difference, over score points, between the estimated relative cumulative frequency and the population distribution function is found for each replication, and the averaged over replications.

Results

Tables 2 and 3 compare results of applying the density estimation methods to samples from distributions

---

Insert Tables 2 and 3 about here

---

of ACT English Usage and Mathematics Usage raw scores. Figures 2 and 3 plot MSE x 10,000 shown in Tables 2 and 3 against method.

Insert Figures 2 and 3 about here

The four-parameter beta-compound binomial (4PB) shows the lowest MSE and K-S statistics for both tests and all four sample sizes. The second lowest MSE and K-S statistic is associated with kernel estimates of varying degrees of smoothing. However, the Bias^2 of 4PB exceeds that of unsmoothed, Cureton-Tukey, and kernel with low H. Thus 4PB owes its low MSE to low Variance rather than low Bias^2.

Results for the kernel method show that the optimal amount of smoothing varies according to test and sample size. In addition, there is a tradeoff between Bias^2 and Variance. Namely, increased Bias^2 accompanies reduced

Variance. For English, MSE decreases up to a binomial smoothing parameter (H) of about 76 for samples of 500 and 1000. For samples of 2000 and 5000, the optimum H appears to be closer to 32. For Mathematics, the lowest MSE is obtained with H = 16 for samples of 500 and 1000, and H = 8 for samples of 2000 and 5000. Although Variance continues to decrease with increased H, Bias^2 continues to increase.

MSE for the negative hypergeometric model shows a striking difference between the two tests. Estimation by this model yields much lower MSE for English than for Mathematics.

Figures 4 and 5 plot Bias_x^2, Variance_x, and MSE_x under the different

---

Insert Figures 4 and 5 about here

---

methods for a sample size of 1000. Bias_x^2 for the negative hypergeometric is lower than Bias_x^2 for the other methods at most points of the score scale. Variance_x shows a much more even pattern for smoothed frequencies. MSE_x, being the sum of Bias_x^2 and Variance_x, retains some of the bumpiness of Bias_x^2, particularly for negative hypergeometric.

MSE_x of the kernel method tends to be slightly greater than that of 4PB except at very high English and very low Mathematics scores, where kernel shows greater Bias_x^2. This greater Bias_x^2 results from the kernel method's tendency to overestimate frequencies at the ends of the score scales. The overestimation increases as the smoothing parameter increases.

Summary and Discussion

The kernel and 4PB methods clearly do the best job of estimating the two score distributions studied. This result essentially agrees with Divgi's (1983) findings. He found a four parameter beta binomial model performed better than a smoothed cumulative distribution function, two- and three-parameter beta binomial models, and a polynomial smoothing of the distribution function. The 4PB method shows slightly lower mean squared error (MSE_x) than the kernel method over most of the score scales.

One way to compare the methods studied here is on the sample size required to achieve equal levels of estimation error. Refer to Tables 2 and 3. The MSE for the 4PB method at N = 500 is smaller than the MSE for the unsmoothed sample frequencies at N = 5000 for both English and Mathematics. Therefore, the use of the 4PB method has an effect on MSE that is similar to using the sample relative frequencies and increasing sample size tenfold. Note that the effect of the 4PB method on the K-S statistic is less drastic. From Tables 2 and 3, the 4PB method appears to be as effective in decreasing the K-S index as a two to two and one-half-fold increase in sample size. The kernel method for the H with the lowest MSE performed nearly as well as the 4PB method.

In planning a norming study a target value for estimation error often is stated and used in specifying the sample size required. The results of this study suggest that the sample size needed to meet the target estimation error may be lowered substantially by using the 4PB or kernel methods.

Kernel MSE_x tends to increase at extremely high and low scores owing to a positive bias: estimated frequencies at the ends tend to be higher as the smoothing parameter increases. This bias merits concern, especially in relation to norms estimation. The adaptive kernel method (Silverman, 1986) shows promise for reducing such bias. This method changes the kernel function according to observed relative frequencies along the score scale.

The Cureton-Tukey method failed to perform nearly as well as 4PB and kernel; nevertheless, it yielded an improvement over no smoothing, and introduced little bias. Its ease and simplicity of computation make its use still worth considering.

The erratic performance of the negative hypergeometric method prompts us to advise extreme caution in applying it. Under this method, Bias_x^2 fluctuated wildly along both score scales. Also, the Bias^2 and MSE appear to depend greatly upon the particular shape of the population distribution: MSE for English remained within reasonable limits, but for Mathematics Usage MSE often far exceeded that of unsmoothed frequencies.

In sum, all but one of the methods produced density estimates much closer on the average to population densities than did unsmoothed sample data. We expect such methods to find extensive application to future analysis of test score data.

References

Angoff, W. H. (1982). Summary and derivation of equating methods used at ETS. In P. W. Holland and D. B. Rubin (Eds.) Test Equating. New York: Academic Press.

Cureton, E. F., & Tukey, J. W. (1951). Smoothing frequency distributions, equating tests, and preparing norms. American Psychologist, 6, 404. (Abstract)

Divgi, D. R. (1983). Comparison of some methods for smoothing score distributions. Paper presented at the 1983 Annual Meeting of the American Educational Research Association, Montreal, Canada, April 1983.

Keats, J. A., & Lord, F. M. (1962) A theoretical distribution for mental test scores. Psychometrika, 27, 59-72.

Lord, F. M. (1965). A strong true-score theory with applications. Psychometrika, 30, 239-270.

Lord, F. M., & Novick, M. R. (1968). Statistical theories of mental test scores. Reading, MA: Addison-Wesley.

Rosenblatt, H. M. (1956). Remarks on some nonparametric estimates of a density function. Annals of Mathematical Statistics, 27, 832-837.

Silverman, B. W. (1986). Density estimation for statistics and data analysis. New York: Chapman and Hall.

Tapia, R. A., & Thompson, J. R. (1978). Nonparametric probability density estimation. Baltimore, MD: The Johns Hopkins University Press.

Table 1

Hypothetical Example of Kernel Estimation with H = 2

x	$\hat{f}$	$.25\hat{f}_{i-1}$	+	$.5\hat{f}_i$	+	$.25\hat{f}_{i+1}$		$\hat{f}_s(K=5)$	$\hat{f}_s(K=4)$
5	.0	.25(.0)	+	.5(.0)	+	.25(.3)	=	.075	---
4	.3	.25(.0)	+	.5(.3)	+	.25(.4)	=	.250	.270
3	.4	.25(.3)	+	.5(.4)	+	.25(.2)	=	.325	.351
2	.2	.25(.4)	+	.5(.2)	+	.25(.1)	=	.225	.243
1	.1	.25(.2)	+	.5(.1)	+	.25(.0)	=	.100	.108
0	.0	.25(.1)	+	.5(.0)	+	.25(.0)	=	.025	.027

Table 2

Fit of ACT English Usage Estimated Densities

Measure of fit x 10,000	Unsmoothed Sample Frequencies	Cureton-Tukey	Negative Hypergeometric	4 Parameter Beta Compound Binomial	Kernel H=2	Kernel H=4	Kernel H=16	Kernel H=32	Kernel H=76
Bias^2	.0005	.0005	.0200	.0013	.0003	.0003	.0006	.0011	.0038
Variance	.2590	.0828	.0058	.0094	.0941	.0670	.0312	.0206	.0118
MSE	.2595	.0833	.0258	**.0108**	.0944	.0673	.0318	.0217	**.0156**
K-S	.0353	.0313	.0293	**.0213**	.0321	.0309	.0276	.0258	**.0246**
Bias^2	.0002	.0003	.0194	.0011	.0002	.0003	.0005	.0011	.0039
Variance	.1285	.0407	.0028	.0048	.0463	.0329	.0154	.0102	.0059
MSE	.1287	.0410	.0222	**.0058**	.0465	.0332	.0160	.0114	**.0098**
K-S	.0246	.0224	.0249	**.0151**	.0224	.0216	.0194	**.0183**	.0188
Bias^2	.0001	.0003	.0193	.0010	.0002	.0002	.0005	.0012	.0041
Variance	.0645	.0203	.0015	.0024	.0231	.0165	.0078	.0052	.0030
MSE	.0646	.0206	.0207	**.0033**	.0233	.0167	.0083	**.0064**	.0071
K-S	.0174	.0158	.0226	**.0112**	.0159	.0153	.0139	**.0136**	.0156
Bias^2	.0000	.0003	.0192	.0009	.0002	.0002	.0005	.0012	.0042
Variance	.0258	.0083	.0006	.0009	.0094	.0067	.0032	.0021	.0012
MSE	.0258	.0085	.0198	**.0018**	.0095	.0069	.0037	**.0033**	.0053
K-S	.0110	.0100	.0207	**.0075**	.0100	.0097	**.0090**	.0093	.0134

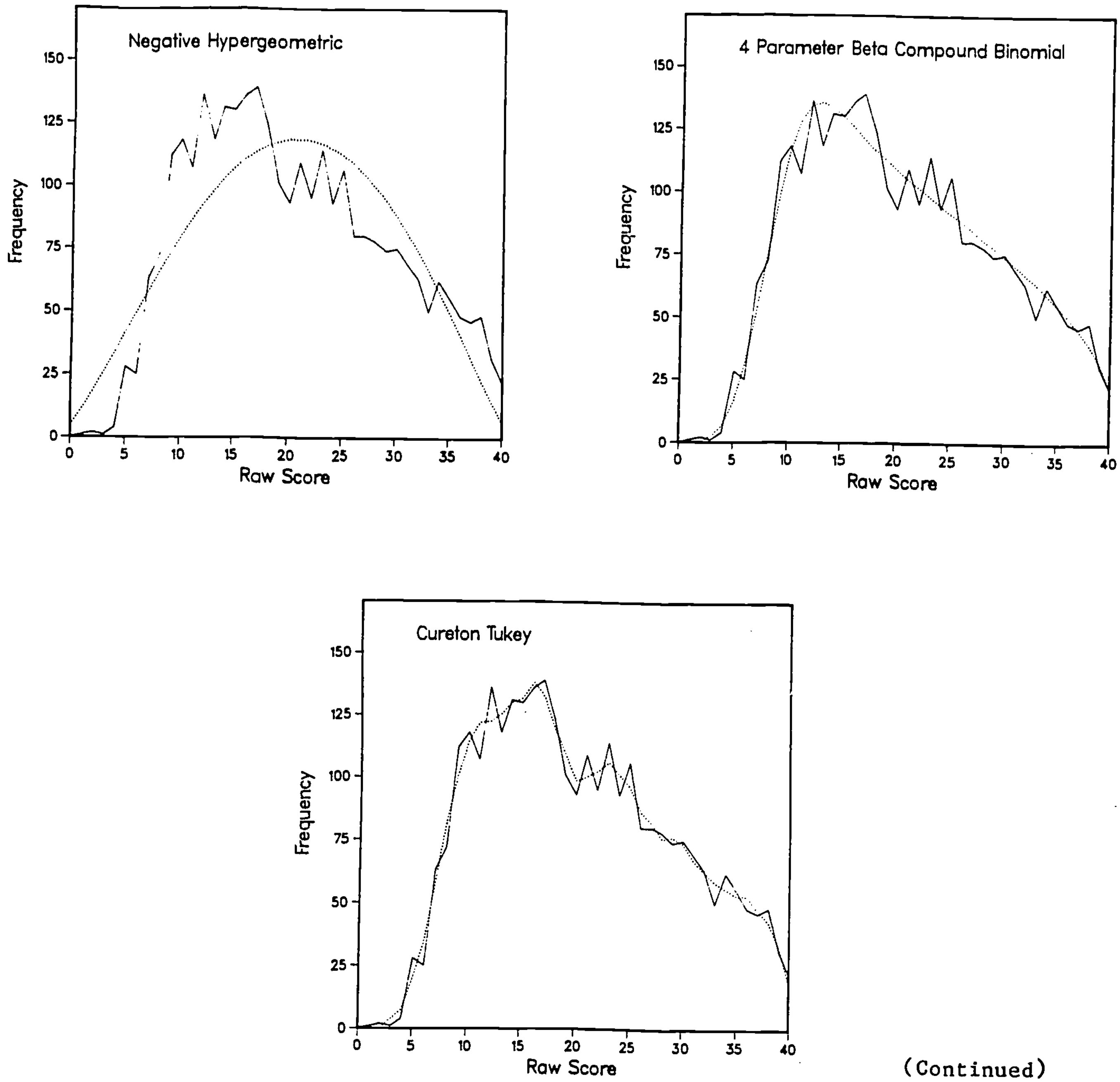
or a given sample size, the lowest two MSE and K-S appear in boldface.

Table 3

Fit of ACT Mathematics Usage Estimated Densities

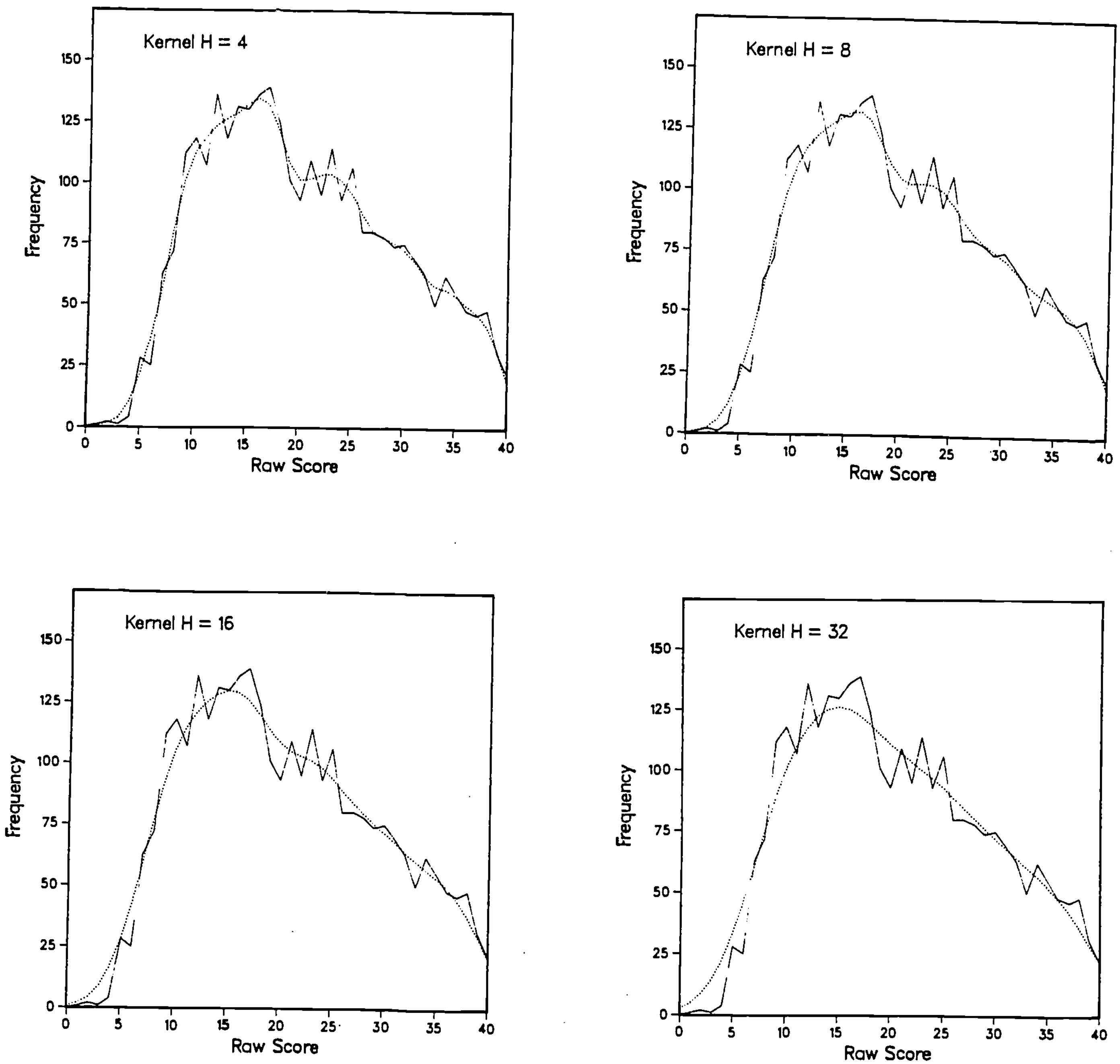
Measure of fit x 10,000	Unsmoothed Sample Frequencies	Cureton-Tukey	Negative Hypergeometric	4 Parameter Beta Compound Binomial	Kernel H=2	Kernel H=4	Kernel H=8	Kernel H=16	Kernel H
Bias^2	.0009	.0013	.2801	.0082	.0012	.0022	.0050	.0131	.
Variance	.4841	.1470	.0168	.0348	.1690	.1185	.0806	.0532	.
MSE	.4849	.1483	.2969	**.0431**	.1703	.1207	.0856	**.0663**	.0
K-S	.0344	.0303	.0466	**.0229**	.0306	.0291	.0274	**.0259**	.0
Bias^2	.0003	.0009	.2774	.0072	.0007	.0016	.0043	.0122	.0
Variance	.2392	.0724	.0083	.0178	.0835	.0584	.0398	.0265	.0
MSE	.2395	.0733	.2857	**.0251**	.0842	.0601	.0441	**.0387**	.0
K-S	.0240	.0213	.0426	**.0166**	.0214	.0204	.0193	**.0190**	.0
Bias^2	.0001	.0008	.2776	.0065	.0007	.0016	.0042	.0121	.0
Variance	.1192	.0366	.0044	.0090	.0419	.0295	.0202	.0135	.0
MSE	.1194	.0374	.2819	**.0155**	.0426	.0311	**.0245**	.0256	.0
K-S	.0170	.0150	.0408	**.0126**	.0152	.0145	**.0140**	.0151	.0
Bias^2	.0001	.0007	.2779	.0060	.0006	.0015	.0042	.0122	.0
Variance	.0474	.0149	.0016	.0035	.0170	.0120	.0082	.0054	.0
MSE	.0475	.0156	.2795	**.0095**	.0176	.0135	**.0124**	.0176	.0
K-S	.0107	.0095	.0395	**.0088**	.0096	**.0093**	.0095	.0129	.0

Within a given sample size, the lowest two MSE and K-S appear in boldface.



(Continued)

Figure 1. Fitted distributions for an ACT Mathematics form. (Observed distribution represented by solid line. Fitted distribution represented by dotted line.)



<u>Figure 1</u> (continued). Fitted distributions for an ACT Mathematics form. (Observed distribution represented by solid line. Fitted distribution represented by dotted line.)

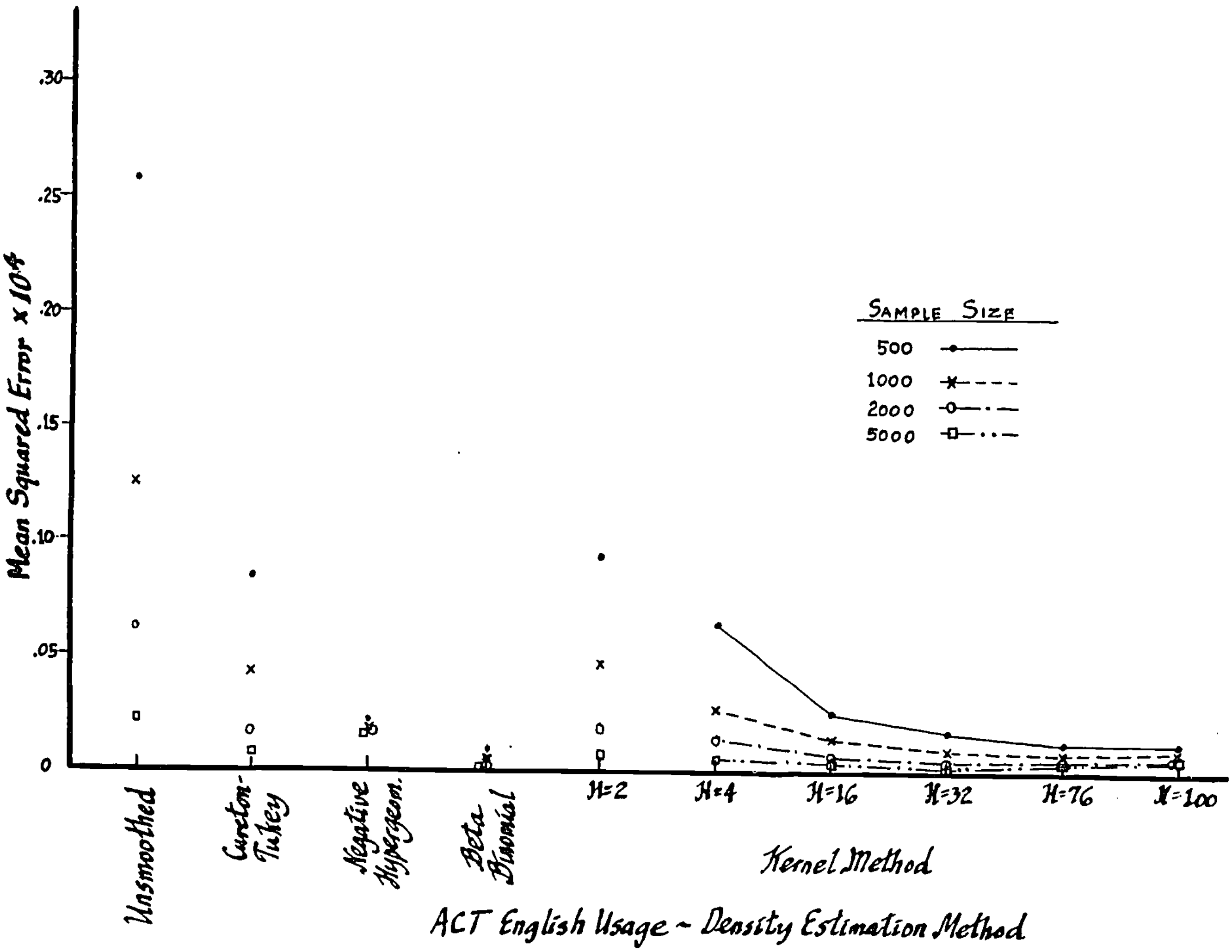


Figure 2. MSE of density estimation methods, ACT English Usage.

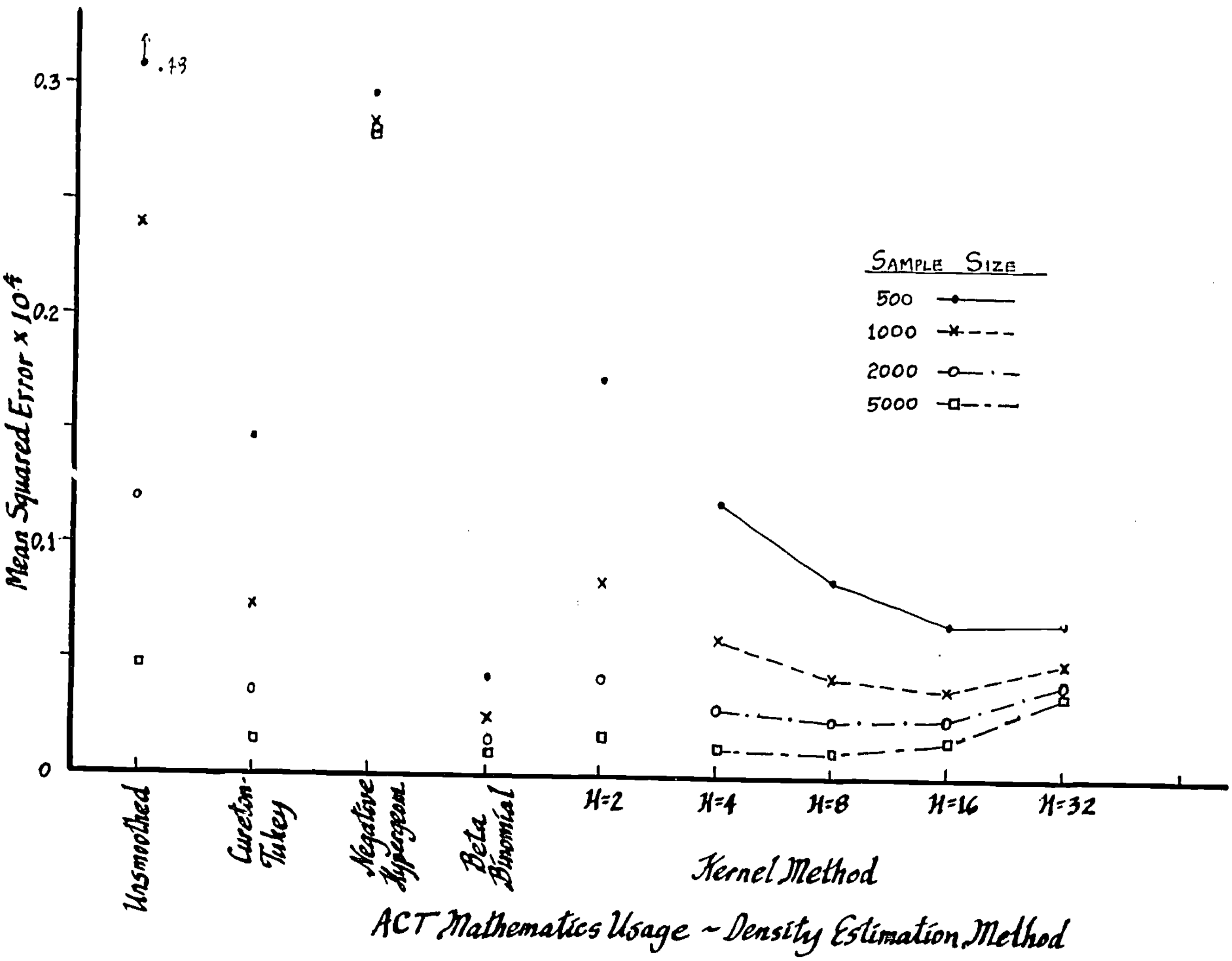


Figure 3. MSE of density estimation methods, ACT Mathematics Usage.

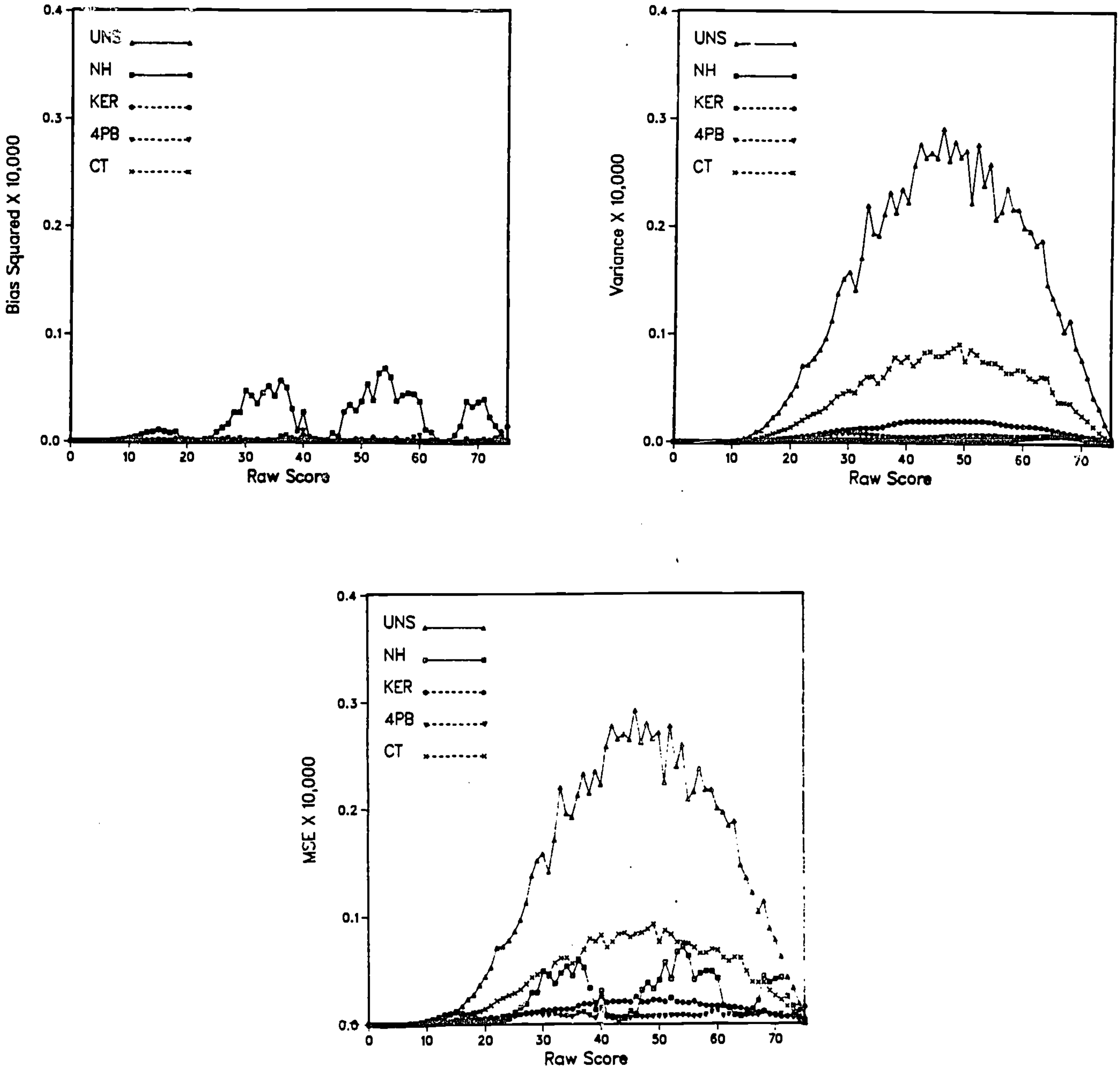
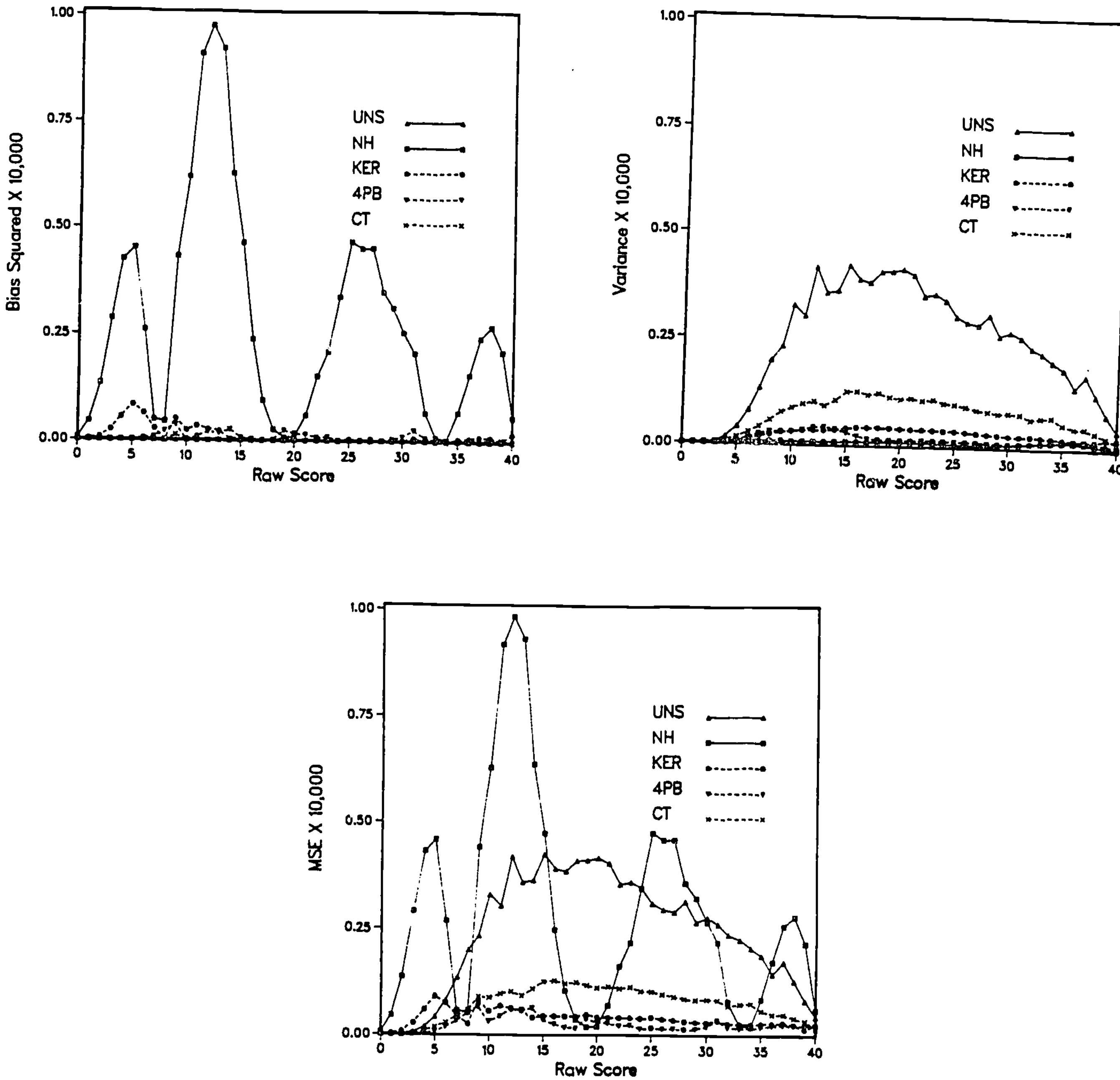


Figure 4. Bias$_x^2$, Variance$_x$, and MSE$_x$ of ACT English Usage densities estimated from samples of 1000. (UNS - unsmoothed, NH - negative hypergeometric, KER - kernel, 4PB - four-parameter beta compound binomial, CT - Cureton-Tukey.)



<u>Figure 5</u>. Bias$_x^2$, Variance$_x$, and MSE$_x$ of ACT Mathematics Usage densities estimated from samples of 1000. (UNS - unsmoothed, NH - negative hypergeometric, KER - kernal, 4PB - four-parameter beta compound binomial, CT - Cureton-Tukey.)