Can Data Recognize Its Parent Distribution?

A.W. Marshall*, J.C. Meza†, I. Olkin ‡

March 2, 2001

Abstract

This study is concerned with model selection of lifetime and survival distributions arising in engineering reliability or in the medical sciences. We compare various distributions, including the gamma, Weibull and lognormal, with a new distribution called geometric extreme exponential. Except for the lognormal distribution, the other three distributions all have the exponential distribution as special cases. A Monte Carlo simulation was performed to determine sample sizes for which survival distributions can distinguish data generated by their own families. Two methods for decision are by maximum likelihood and by Kolmogorov distance. Neither method is uniformly best. The probability of correct selection with more than one alternative shows some surprising results when the choices are close to the exponential distribution.

1 INTRODUCTION

In any parametric analysis of data, the issue of model choice arises. This is particularly evident when the data are necessarily nonnegative, and the usual Gaussian model is likely to be inappropriate. Examples of nonnegative data include lifetime and other waiting time data, earthquake and flood measurements, pollutant concentrations, and material strengths.

To provide adequate selections of parametric models for various kinds of data, collections of frequency curves for data analysis were developed by Karl Pearson in 1895, Edgeworth in 1904 and Charlier in 1905. For a detailed and general discussion of frequency curves, see Elderton and Johnson (1969). Models now commonly used for nonnegative data include the exponential, gamma, Weibull, lognormal, and inverse Gaussian distributions; these distributions are discussed in Section 2.

In some fortunate circumstances, physical considerations alone can identify the appropriate family of distributions. Thus, when there is "no premium for waiting" the "lack of memory" property which characterizes the exponential distributions can identify the family of exponential distributions as the appropriate model. More commonly, no compelling physical considerations point to an appropriate model and so a choice must be made by other means. Such a choice may be based upon mathematical tractability, or on a feeling that a particular family is "rich" enough to include a good fit to the data.

*Western Washington University and the University of British Columbia

†Computational Sciences and Mathematics Research Department, MS 9217, Sandia National Laboratories, Livermore, CA 94551-0969, meza@ca.sandia.gov, supported in part by the Department of Energy under contract DE-ACO4-94AL85000

‡Statistics Department, Stanford University, supported in part by the National Science Foundation

Most parametric statistical analyses start with the assumption that a parametric family has been chosen as the model. The critical question of model choice has received relatively little attention, although the assumed model is often subjected to a goodness of fit test.

Another much less commonly used approach is to test the hypothesis that the chosen family is correct against the alternative hypothesis that a second specified family is correct. This is often referred to as "testing separate families of hypotheses"; because it is cast in the hypothesis testing framework, it treats the two families asymmetrically. Here, the work of Cox (1961, 1962) is basic, but see Pereira (1977) for further references.

There is a third approach that has received even less attention. This procedure is to choose two or more possible candidate parametric families of distributions, and then use the data to select the most appropriate candidate. Selection procedures put the alternative families on equal footings, and they were also discussed by Cox (1961, 1962).

The focus of this paper is on selection procedures for nonnegative data using the following two methods:

1. Criterion 1. **Maximum likelihood method:** For each alternative family under consideration, maximize the likelihood over the parameter values, and select the family that yields the largest maximum likelihood.

2. Criterion 2. **Minimum Kolmogorov distance method:** For each alternative family, estimate the parameters by maximum likelihood to select a specific candidate. Determine the Kolmogorov distance between the specific candidate and the empirical distribution, and select the family which yields the minimum distance.

In this paper, these methods are used only for selections where the alternative families have the same number of parameters. When the alternative families have different numbers of parameters, the appropriateness of the methods is unclear because the family with the greatest number of parameters would perhaps have an unfair advantage. The methods are certainly inappropriate when one alternative is a sub-family of the other. Such is the case, e.g., for the exponential and Weibull families, and here, the exponential family could never hope to be selected.

Some history of model selection and a discussion of various methods is given in Section 3.

How successful are methods 1 and 2 when presented with data from a known distribution and asked to choose between the actual parent parametric family and one or more alternative? How large must the sample size be to make a correct selection with specified probability? To investigate these questions, samples of various sizes from known distributions have been computer generated, and the methods for model selection have been applied. Since the parent distribution is known, the selection can be scored as correct or not; after repeating the process a large number of times, the probability of a correct selection can be accurately estimated.

Briefly, the answer to the question posed in the title of this paper is "yes," subject to the qualification that the sample size be sufficiently large — larger than many data sets encountered in practice. In general, with the understanding that there are many exceptions and qualifications, a sample size of 200 yields a reasonably high probability of correct selection. Of course, this is based upon the supposition that the correct family of distributions is amongst the possible choices.

The numerical work sheds some light on the comparative "richness" of the alternative families. For example, sometimes data from a gamma distribution is fit by a Weibull distribution even better than by its parent family, but data from a Weibull distribution is much less likely to be best fitted to a gamma distribution. This suggests that the family of Weibull distributions is "richer" than the family of gamma distributions at least in the range of some parameter values.

The importance of choosing the correct model depends upon the use to be made of the model. The correct choice is particularly critical if tail behavior is an issue, because then attempts are made to extrapolate to regions of little or no data.

Studies of the kind reported on here have already been carried out perhaps first by Bain and Engelhardt (1980), who focus on the gamma and Weibull families. This and other such studies are briefly discussed in Section 7.

2 CANDIDATE FAMILIES OF DISTRIBUTIONS

The gamma, Weibull, lognormal, and geometric extreme exponential families of distributions are included in this study. The first three are perhaps the most familiar and most widely used life distributions. The geometric extreme exponential (GE-exponential) family has been included because it was introduced only recently, and has not been widely studied. A restricted version of this distribution appears as an example of a distribution with random parameters in Gut (1995, p.117). For an extensive discussion of properties of this distribution see Marshall and Olkin (1997).

An absolutely continuous distribution can be described in terms of the distribution function F, the survival function $\bar{F} = 1 - F$, or by the density f. A distribution can also be described in terms of the hazard rate $r = f/\bar{F}$. The value $\bar{F}(x)$ of the survival function at the point x gives the probability of survival beyond x whereas $\Delta r(x)$ is the conditional probability of failure in the next increment Δ of time given survival at time x. The intuitive content of the hazard rate is often useful when model choice is to be made.

All the distributions included in this study have densities with convenient functional forms. The survival functions and hazard rates of the gamma and lognormal distributions cannot in general be written in closed form, but they can be qualitatively studied and evaluated numerically.

Weibull density. The Weibull density is given by

$$f(x) = \alpha\lambda(\lambda x)^{\alpha-1}e^{-(\lambda x)^{\alpha}}, \quad x > 0,$$

with scale parameter $\lambda > 0$ and shape parameter $\alpha > 0$. The choice $\alpha = 1$ yields the exponential density.

Gamma density. The gamma density is given by

$$f(x) = \frac{\lambda^{\nu}x^{\nu-1}e^{-\lambda x}}{\Gamma(\nu)}, \quad x > 0,$$

with scale parameter $\lambda > 0$ and shape parameter $\nu > 0$. The choice $\nu = 1$ yields the exponential density.

GE-exponential density. A new density that arises from a stability property is derived by Marshall and Olkin (1997). Its density is given by

$$f(x) = \frac{\gamma\lambda e^{\lambda x}}{[e^{\lambda x} - \bar{\gamma}]^2} = \frac{\gamma\lambda e^{-\lambda x}}{[1 - \bar{\gamma}e^{-\lambda x}]^2},$$

with scale parameter $\lambda > 0$ and shape parameter $\gamma > 0$. Here $\bar{\gamma} = 1 - \gamma$. The choice $\gamma = 1$ yields the exponential density.

Lognormal density. The lognormal density is given by

$$\begin{aligned} f(x) &= \frac{1}{\sigma\sqrt{2\pi}\,x} \exp\Big[-\frac{1}{2\sigma^2}(\log x - \mu)^2\Big], \\ &= \frac{\alpha}{\sqrt{2\pi}\,x} \exp -\tfrac{1}{2}[\log(\lambda x)^\alpha]^2, \quad x > 0, \end{aligned}$$

with scale parameter $\lambda = e^{-\mu}$ and shape parameter $\alpha = 1/\sigma > 0$.

The Weibull, gamma and GE-exponential distributions all include the exponential distribution as a special case, and they all have monotone hazard rates. Consequently, they are natural competitors when a model selection is to be made. On the other hand, the lognormal distribution does not include the exponential distribution as a special case. It does not have a monotone hazard rate, but instead has a unimodal hazard rate (increasing to a maximum, and then decreasing). Perhaps because of its substantial qualitative differences, the lognormal distribution is ordinarily not a serious competitor when data comes from a gamma, Weibull, or geometric extreme exponential distribution. On the other hand, the inverse Gaussian distribution and the F distribution, like the lognormal distribution, both have unimodal hazard rates. These three distributions may be natural competitors, and they would be interesting subjects of further simulation studies. Applications where the lognormal distribution has been used are discussed by Aitchison and Brown (1957) and Crow and Shimizu (1988); applications of the inverse Gaussian distribution are discussed by Seshadri (1993). In most of these applications, the reason for choosing the families is unclear, and alternatives are not often discussed.

3 SELECTION METHODS

The problem of choosing a model has been treated in a number of ways, some of which are described here.

Method (i): Maximum likelihood criterion

Method (i) of Section 1, i.e., choose the family which yields the largest maximum likelihood, is a modification due to Cox (1961) of a classical discrimination method. In case there are but two alternative families, $\mathcal{F} = \{f(x|\theta)\}$ and $\mathcal{G} = \{g(x|\omega)\}$, this procedure can be restated. Let

$$\hat{L} = \Sigma \log[f(x_i|\hat{\theta})/g(x_i|\hat{\omega})],$$

where $\hat{\theta}$ and $\hat{\omega}$ are maximum likelihood estimators of the parameters θ and ω. Choose the family $\mathcal{G}$ if $\hat{L} < 0$. The statistic $\hat{L}$ is sometimes called the Cox statistic. It has been shown by Cox (1962) that, properly normalized, the statistic $\log \hat{L}$ is asymptotically normal as the sample size increases without bound. White (1982) examined the regularity conditions underlying the asymptotic derivations; see also Loh (1985). The asymptotic normality of $\log \hat{L}$ has been used by Fearn and Nebenzahl (1991) to estimate sample sizes required to achieve a desired probability of correct selection.

In terms of $\hat{L}$, the described selection procedure is related to a procedure for testing the hypothesis that the sample came from a distribution in $\mathcal{F}$ versus that it came from a distribution in $\mathcal{G}$. This testing problem treats the two families asymmetrically (discussed by Cox, 1961, 1962) and so it is slightly different from the selection problem. Although a considerable amount of work has been done on the testing problem, it is not the subject of this paper and is not treated here.

Method (ii): Minimum distance criterion

With a sample in hand, it is natural to choose amongst alternative models that one which provides the distribution best fitting the empirical distribution according to some measure of goodness of fit. Here, the goodness of fit measure is taken to be the Kolmogorov distance

$$D(F, G) = \sup_{-\infty < x < \infty} |F(x) - G(x)|.$$

To implement this procedure, a candidate from each parametric family that best fits the empirical distribution must be determined, then the best fits are compared. Unfortunately, the first step of this procedure is difficult both from a theoretical and a numerical point of view. To make the method practical, the candidate from each parametric family is chosen by maximum likelihood rather than by minimizing the Kolmogorov distance. Then, the family is chosen that provides the best fit to the empirical distribution in the sense of Kolmogorov distance.

This hybrid method, mixing maximum likelihood and minimum distance methods has been utilized by Taylor and Jakeman (1985). They show (Figure 1, p. 503) that if the exponential distribution is a candidate along with the more general gamma and Weibull distributions, then with exponential data, the exponential distribution is often chosen. This happens only because mixed criteria are used, and it points to a weakness in the minimum distance method of this paper.

Both the maximum likelihood and minimum distance methods described above can be regarded as variants of the same approach. First, choose a representative from each candidate family, then choose the family providing the best candidate. Choice of candidate amounts to a choice of parameters, and in this sense, it is similar to estimation. Apart from maximum likelihood, the choice has been made by a least squares procedure (Pandy, Ferdous and Uddin, 1991). Various other discrimination procedures are discussed by Dyer (1973). McDonald, Vance and Gibbons (1995) discuss several methods in the context of lognormal and Weibull alternatives for emissions data.

Some additional methods

Several methods for choosing from alternative families have been used in the literature; some of these are briefly described here.

An approach to model selection quite different from those above can be had by utilizing Bayesian methods. As outlined by Cox (1961), the basic idea is to form the likelihoods, then instead of maximizing over the parameters, integrate the parameters out using a prior distribution, and then choose the family producing the largest integrated likelihood. Motivated by invariance considerations rather than Bayesian ideas, Siswadi and Quesenberry (1982) took this type of approach. A mixed approach is followed by Kent and Quesenberry (1982), who, again motivated by considerations of scale invariance, integrate away a scale parameter using a prior distribution, but use maximum likelihood to determine shape parameters. For scale and location parameter families, this type of procedure is discussed by Hogg, Uthoff, Randles and Davenport (1972).

Another approach essentially eliminates the issue of model choice by incorporating all candidate alternatives as special cases in a super-model. Then the super-model is used and no model choice is required.

This natural idea is proposed by Cox (1961) in the context of two competing models, say $\mathcal{F} = \{F(x|\theta)\}$ and $\mathcal{G} = \{G(x|\omega)\}$. He notes the possibility of using as a super-model densities obtained by normalizing geometric means

$$[f(x|\theta)]^{\pi}[g(x|\omega)]^{1-\pi}, \quad 0 \leq \pi \leq 1.$$

Atkinson (1970) has pointed out that with the geometric mean, the likelihood takes a more convenient form than it does when an arithmetic mean is used. However, arithmetic means were used in place of geometric means by Olkin and Spiegelman (1987) to decide between a parametric and a nonparametric model.

By introducing a third parameter, Prentice (1975) has obtained a family of distributions that includes variants of the logistic distribution as well as the Weibull and gamma distributions as limits. This specialized family can be used in some commonly encountered areas.

4 NUMERICAL PROCEDURES

The process for determining the probability of correct selection was based on a Monte Carlo procedure. For each alternative family of the study, we first fixed the scale parameter so that the expectation was equal to 1. This can be done because the results of the study do not depend upon the choice of the scale parameter (see the appendix for a proof). Next we made various choices for the shape parameter and the sample size. For each combination of these choices, we generated a random sample and computed the maximum likelihood estimate for the parameters of the parent and alternative families. The family with the largest maximum likelihood was then chosen as the *winner* for that combination of shape parameter and sample size. This procedure was then repeated $100,000$ times and we recorded the proportion of times that each alternative family yielded the best fit. This entire procedure was then repeated using the minimum Kolmogorov distance as the criterion for determining the best fit.

To compute the maximum likelihood estimate it was sometimes necessary to solve an optimization problem. The likelihood equations are provided in Appendix B, some of which appear in Lawless (1982). In these cases, we used the nonlinear optimization package NPSOL (Version 4.04) developed by Gill, Murray, Saunders, and Wright (1986). The random variables were generated using various techniques depending on the families. Random numbers from the Weibull and GE-exponential distributions were generated using inversion of uniform random variables. The lognormal random numbers were generated using the RNOR algorithm due to Marsaglia and Tsang (1984). The gamma family random numbers were generated using the GBH algorithm due to Cheng and Feast (1979).

All of the numerical experiments were run on an SGI Power Challenge using double precision arithmetic. With 100,000 trials, the variance of the estimated probability of a correct selection is theoretically at most 0.0016. In spite of using double precision arithmetic, some numerical difficulties were inevitable in instances where the likelihoods or minimum Kolmogorov distances of the competing families were nearly equal. This problem may explain the lack of complete smoothness in some of the figures.

5 SELECTION WITH GAMMA, WEIBULL AND GE-EXPONENTIAL DISTRIBUTIONS AS ALTERNATIVES

The families of Weibull, gamma and geometric extreme exponential distributions all reduce to the exponential distribution when their shape parameter is 1 and they all have monotone hazard rates, so in many applications these families are natural alternative models to consider. No matter which family is the source of the data, correct selection is relatively unlikely when the shape parameters are close to 1.

Exponential data

For data from the exponential distribution, selection of any of the three families can be regarded as "correct." With any two of the families of alternatives and the maximum likelihood selection method, the selection probabilities are close to 0.5; that is, the two alternatives are equally likely to be selected and the selection method is indifferent. It is perhaps surprising that with all three families as alternatives, the selection probabilities are not equal. For samples above 100, the selection probabilities are approximately 0.13 for the Weibull family, 0.42 for the gamma family, and 0.45 for the GE-exponential family. The fact that the Weibull family does poorly in a three-way race is reminscent of a phenomena long recognized in elections. When entering a race, a third candidate ordinarily takes more votes from one of the original candidates than the other; in this way the outcome of the election can be changed even though the third candidate does not win. The results from the numerical experiments using a Weibull distribution with parameter $\alpha = 1$ (which should be theoretically equivalent to an exponential distribution) are shown in Figure 1. Numerical experiments with the gamma and GE-exponential distributions with shape parameter equal to one resulted in similar although not identical results.

[Figure 1 about here.]

Weibull data

If the data are actually from a Weibull distribution, the probability of a correct selection is less with the Weibull and gamma families as alternatives than it is with the Weibull and GE-exponential families as alternatives. With all three families as alternatives, the GE-exponential family is least likely to be selected when the shape parameter of the parent Weibull distribution is not close to 1, but the GE-exponential family becomes more likely to be selected as the Weibull shape parameter moves closer to 1. The rate at which this change takes place depends upon the sample size. It is interesting to see that with sample sizes less than 400, and the parent Weibull with shape parameter between 0.9 and 1.1, the Weibull distribution is the *least* likely of the three alternatives to be selected.

[Figure 2 about here.]

[Figure 3 about here.]

In general, for two-way comparisons, the maximum likelihood works better than the minimum Kolmogorov distance method.

Gamma data

For gamma data with shape parameter greater than or equal to 1 and two alternative families, the Weibull family is a consistently more successful competitor to the gamma family than is the GE-exponential family. The same is true when all three families are alternatives unless the data comes from a gamma distribution with shape parameter close to 1. Then, the GE-exponential distribution begins to be a best fit more often than the Weibull distribution, first for small sample sizes, and then for all sample sizes. By the time that the shape parameter reaches 1, the GE-exponential distribution is more often a best fit than either the gamma or Weibull distribution.

[Figure 4 about here.]

[Figure 5 about here.]

[Figure 6 about here.]

For the Weibull family as the only alternative to the gamma family, the maximum likelihood method dominates the Kolmogorov distance method, which is somewhat unstable for small samples.

[Figure 7 about here.]

[Figure 8 about here.]

GE-exponential data

Suppose that data comes from a GE-exponential distribution. With one alternative family, correct selection is much more likely for the gamma as alternative than it is for the Weibull as alternative family.

[Figure 9 about here.]

[Figure 10 about here.]

For all three families as alternatives, this difference in the gamma and Weibull families is less pronounced except when the data comes from a GE-exponential distribution with a large shape parameter. When the shape parameter is close to 1, the gamma family is more likely to be selected than the Weibull family especially for small sample sizes.

[Figure 11 about here.]

[Figure 12 about here.]

If the Weibull family is the only alternative to the GE-exponential family, the minimum Kolmogorov distance method works best, although the maximum likelihood method starts to gain for samples sizes greater than 600. If the gamma family is the only alternative, then the two methods are very close except when that data comes from a distribution with shape parameter close to 1; then, the minimum Kolmogorov distance is best.

[Figure 13 about here.]

[Figure 14 about here.]

6 SELECTION WITH LOGNORMAL, GAMMA, AND WEIBULL DISTRIBUTIONS AS ALTERNATIVES

As already noted, the lognormal distribution is somewhat different in character from the other families of this study because it does not have the exponential distribution as a special case, and it does not have a monotone hazard rate. Nevertheless, the lognormal is often considered as a competitor family when nonnegative data are studied. For this reason we include it in this study.

Weibull data

With Weibull data and both the Weibull and lognormal families as alternatives, the probability of a correct selection is independent of the shape parameter. This fact, proved in an appendix, is due to the fact that the shape parameter α of the Weibull distribution and the parameter $1/\sigma$ of the lognormal distribution are both power parameters, i.e., they appear as an exponent of the argument in the distribution function.

[Figure 15 about here.]

With Weibull and lognormal alternatives and the maximum likelihood method, the probability of a correct selection is already above 0.95 for samples as small as 100. The minimum Kolmogorov distance method does even better, and is nearly foolproof with a sample of size 100.

Gamma data

With gamma data and both the gamma and lognormal families as alternatives, the probability of correct selection using maximum likelihood as the criterion is decreasing as the gamma shape parameter increases from 1. This monotonicity is not present when selection is made by minimum Kolmogorov distance. However, except for values of the shape parameter close to 1, the minimum Kolmogorov distance method gives a higher probability of a correct selection than does the maximum likelihood criterion. These comments apply only to the case of gamma shape parameter greater than or equal to 1, as simulations were not carried out with the shape parameter less than 1.

[Figure 16 about here.]

GE-exponential data

With GE-exponential data and both the GE-exponential and lognormal families as alternatives, the probability of a correct selection is quite high at least for the values 1, 2, 4, 8 and 10 of the shape parameter for which simulations were carried out. Again, the minimum Kolmogorov distance gives a higher probability of correct selection than does the maximum likelihood method.

[Figure 17 about here.]

Lognormal data

With lognormal data and both the lognormal and Weibull families as alternatives, the probability of a correct selection is independent of the shape parameter for the same reason that this independence occurs with Weibull data and the same alternatives. With samples of size at least 100, the probability of a correct selection by the maximum likelihood method exceeds 0.97, so these alternatives are relatively easily distinguished.

With the gamma family replacing the Weibull as an alternative, the situation is much different. Here, with the maximum likelihood criterion, the probability of a correct selection is rather rapidly decreasing in the shape parameter of the lognormal data. Although the probability of a correct selection is quite high for shape parameter values greater than 0.25, they approach 0.5 as the shape parameter approaches 0. The apparent reason for this is that with a fixed unit expectation, the weak limit of the lognormal distribution as the shape parameter goes to 0 is degenerate at 1, and this is also the weak limit of the gamma distribution as the shape parameter goes to ∞. So in these extreme cases, the lognormal and gamma distributions are similar and hard to distinguish.

[Figure 18 about here.]

With lognormal data and the GE-exponential as an alternative model, the maximum likelihood criterion yields a probability of correct selection that is not monotone in the shape parameter. A sample of size 30 is sufficient to achieve a 0.95 probability of correct selection for the shape parameters 1 and 5; but at the intermediate value of 1.5, a sample size of about 300 is required to achieve the same accuracy.

[Figure 19 about here.]

7 THE NUMERICAL RESULTS

Several authors have previously run simulations similar to those of this paper. Here, a brief summary of the earlier studies is offered, together with an outline of the results of this paper.

Bain and Engelhardt (1980). These authors considered the Weibull and gamma alternatives and the maximum likelihood selection method. For both families as parents, shape parameters of 0.5, 1, 2, 4, 8, 16, and sample sizes from 10 to 160, they used 4,000 replications (1000 for sample size 160) to estimate the probabilities of correct selection. Their general conclusion was that the maximum likelihood ratio provided a basis for a reasonable decision whatever the sample size and the procedure could be used to determine the sample size needed to achieve a given precision.

Kappenman (1982). This paper furthers the work of Bain and Engelhardt (1980) by considering the lognormal distribution as an alternative family along with the Weibull and gamma families. Using the maximum likelihood selection method and 1000 replications, probabilities of correct selection for the triple as well as all of the pairs of the families are tabulated. Results for 5 values of the shape parameter of the parent distribution and 5 sample sizes ($n = 10, 30, 50, 100, 200$) are given. There is good agreement with the results of Bain and Engelhardt (1980) where the studies overlap. Kappenman comes to the same conclusion as Bain and Engelhardt when their procedure is applied to a choice between Weibull and lognormal or between gamma and lognormal models. He also observes that the probability of correct selection is quite high even for moderate sample sizes when comparing the lognornmal to either the Weibull or gamma models. Kappenman's results for the three-way comparison are similar to the two-way comparisons except for the gamma model which required large sample sizes for shape parameters exceeding one to correctly select the gamma model with any substantial probability.

Taylor and Jakeman (1985). Using both the maximum likelihood method and the minimum Kolmogorov distance method (also used in this paper), these authors consider the exponential, Weibull, gamma, and lognormal families as alternatives. They made calculations with all families as competitors that pass a preliminary goodness of fit test. They estimate probabilities of correct selection for 4 or 5 shape parameter values and 6 sample sizes ranging from 10 to 365 using from 1000 to 250 replications. The results, presented graphically, show that the maximum likelihood method works better than the minimum Kolmogorov distance method especially for the gamma model, but for other parent distributions, the difference between the methods is less pronounced.

Fearn and Nebenzahl (1991). These authors use the asymptotic normality of method 1 (defined in Section 3) to estimate sample sizes required to achieve the correct selection probability of 0.8. This is done for the Weibull and gamma families as alternatives. Shape parameter values of 0.5 (0.1) 2.0 are considered. They re-computed the results from Bain and Engelhardt to verify the validity of the normal approximation and conclude that the approximation is reasonably good especially as the sample size increases. As expected the approximation deviates from the simulation results as the sample size decreased below a sample size of 80. For the Weibull, the normal approximation also deteriorated for shape parameters less than one.

Graphical Displays

The figures of this paper are organized by parent distribution with the Weibull, gamma, GE-exponential, and lognormal families abbreviated, respectively, as W, G, GE, and L. The figures are labeled first by parent distribution, followed by a slash, and then the alternative families. Finally, a letter "a" indicates that the selection method is minimum Kolmogorov distance; the absence of a letter indicates that the selection method is by maximum likelihood. Thus, Figure W/Ga is for data from a Weibull distribution with the Weibull and gamma families as alternatives and minimum Kolmogorov distance method for selection. Figure W/G, GE is for data from a Weibull distribution, with Weibull, gamma and GE-exponential families as alternatives and maximum likelihood as the selection criterion. Parameter values are indicated by keys on the graphs. For two alternatives, the graphs show the probability of correct selection for various parameter values, but for three alternatives, a separate graph is required for each parameter value so that the probability of selection can be shown for each alternative family. By comparing these various graphs, it is possible to see how the ranking of the alternatives changes with the parameter values.

Although the graphs show the probability of selection or correct selection as a function of sample size, it is easy to determine from them the sample sizes required to achieve a desired probability of correct selection.

REFERENCES

Aitchison, J. and Brown, J.A.C. (1957). *The Lognormal Distribution.* Cambridge University Press, Cambridge.

Atkinson, A.C. (1970). A method for discriminating between models (with discussion). *J. Royal Statist. Soc. B*, 32, 323–353.

Bain, Lee J. and Engelhardt, M. (1980). Probability of correct selection of Weibull versus gamma based on likelihood ratio. *Comm. Statist. Theory Methods*, A9(4), 375–381.

Cheng, R.C.H. and Feast, G.M. (1979). Some simple gamma variate generators. *Applied Statistics*, 28, 290–295.

Cox, D.R. (1961). Tests of separate families of hypotheses. In *Proceedings of the Fourth Berkeley Symposium*, Vol. 1, 105–123. Berkeley: University of California Press.

Cox, D.R. (1962). Further results on tests of separate families of hypotheses. *J. Roy. Statist. Soc.*, 24, B, 406–424.

Crow, E.L. and Shimizu, K. (eds.) (1988). *Lognormal Distributions: Theory and Applications.* Marcel Dekker: New York.

Dyer, A.R. (1973). Discrimination procedures for separate families of hypotheses. *J. Amer. Statist. Assoc.*, 68, 970–974.

Elderton, W.P. and Johnson, N.L. (1969). *Systems of Frequency Curves.* Cambridge University Press: London.

Fearn, D.H. and Nebenzahl, E. (1991). On the maximum likelihood ratio method of deciding between the Weibull and gamma distributions. *Comm. Statist. Theory Methods*, 20(2), 579–593.

Gill, P.E., Murray, W., Saunders, M.A. and Wright, M.H. (1986). User's Guide for NPSOL (Version 4.0): A FORTRAN package for nonlinear programing. *SOL* 86-2.

Gut, A (1995). An intermediate course in probability. Springer-Verlag. New York.

Hogg, R.V., Uthoff, V.A., Randles, R.H. and Davenport, A.S. (1972). On the selection of the underlying distribution and adoptive estimation. *J. Amer. Statist. Assoc.*, 67, 597–600.

Kappenman, R.F. (1982). On a method for selecting a distributional model. *Commun. Statist.*, A-11, 663–672.

Kent, Jacqueline, and Quesenberry, C.P. (1982). Selecting among probability distributions used in reliability. *Technometrics*, 24, 59–65.

Lawless, J.F. (1982) Statistical models and methods for lifetime data. Wiley, New York.

Loh, Wei-Yin. (1985). A new method for testing separate families of hypotheses. *J. Amer. Statist. Assoc.*, 80, 362–368.

Marshall, A.W. and Olkin, I. (1997). New families of life distributions with extreme stability. *Biometrika*, 84, 641–652.

Marsaglia, G. and Tsang, W.W. (1984). A fast, easily implemented method for sampling from decreasing or symmetric unimodal density functions. *SIAM J. Scient. and Stat. Computing*, 5, 349–359.

McDonald, G.C., Vance, L.C. and Gibbons, D.L. (1995). Some tests for discriminating between lognormal and Weibull distributions — an application to emissions data. Chapter 25 in *Recent Advances in Life-Testing and Reliability*, ed. by N. Balakrishnan. CRC: Boca Raton.

Olkin, I. and Spiegelman, C. (1987). A semi-parametric approach to density estimation. *J. Amer. Statist. Assoc.*, 82, 858–865.

Pandey, M., Ferdous, J. and Uddin, Md. Borhan. (1991). Selection of probability distribution for life testing data. *Comm. Statist. Theory Methods*, 20(4), 1373–1388.

Pereira, B. de B. (1977). Empirical comparisons of some tests of separate families of hypotheses. *Metrika*, 25, 219–234.

Prentice, R.L. (1975). Discrimination among some parametric models. *Biometrika*, 62, 607–614.

Seshadri, V. (1993). *The Inverse Gaussian Distribution*. Clarendon Press: Oxford.

Siswadi and Quesenberry, C.P. (1982). Selecting among Weibull, lognormal and gamma distributions using complete and censored samples. *Naval Research Logistics Quarterly*, 29, 557–569.

Taylor, J.A. and Jakeman, A.J. (1985). Identification of a distributional model. *Commun. Statist. – Simula. Computa.*, 14(2), 497–508.

White, H. (1982). Regularity conditions for Cox's test of non-nested hypotheses. *J. Econometrics*, 19, 301–318.

APPENDIX A – PARAMETER INDEPENDENCE

In the following $\{f(\cdot|\lambda,\theta)\}$ and $\{g(\cdot|\ell,\omega)\}$ are families of densities in which λ and ℓ are scale parameters, that is

$$f(x|\lambda,\theta)=\lambda f(\lambda x|1,\theta) \quad \text{and} \quad g(x|\ell,\omega)=\ell g(\ell x|1,\omega). \tag{1}$$

A.1 Theorem. If $X_1,\ldots,X_n$ is a random sample from the density $f(\cdot|\lambda_0,\theta)$ and Criterion 1 (maximum likelihood method) is used to select one of the alternative families $\{f(\cdot|\lambda,\theta)\}$ or $\{g(\cdot|\ell,\omega)\}$, then the probability of a correct selection is independent of λ_0.

Proof. In the following, $X_i \sim f(\cdot|\lambda,\theta)$ is used to mean that $X_1,\ldots,X_n$ are independent observations of random variables with density $f(\cdot|\lambda,\theta)$. It is necessary to prove that

$$P_{CS}=P\big\{\max_{\lambda,\theta}\Pi f(X_i|\lambda,\theta)>\max_{\lambda,\omega}\Pi g(X_i|\lambda,\omega)\big|X_i\sim f(\cdot|\lambda_0,\theta_0)\big\}$$

is independent of λ_0. From (1) it follows that

$$\begin{aligned}
P_{CS} &= P\big\{\max_{\lambda,\theta}\Pi f(X_i|\lambda,\theta)>\max_{\lambda,\omega}\Pi g(X_i|\lambda,\omega)\Big|X_i/\lambda_0\sim f(\cdot|1,\theta_0)\big\}\\
&= P\big\{\max_{\lambda,\theta}\Pi f(\lambda_0 Y_i|\lambda,\theta)>\max_{\lambda,\omega} g(\lambda_0 Y_i|\lambda,\omega)\Big|Y_i\sim f(\cdot|1,\theta_0)\big\}\\
&= P\big\{\max_{\lambda,\theta}\Pi f\big(Y_i\big|\frac{\lambda}{\lambda_0},\theta\big)>\max_{\lambda,\omega} g\big(Y_i\Big|\frac{\lambda}{\lambda_0},\omega\big)\big|Y_i\sim f(\cdot|1,\theta_0)\big\}\\
&= P\big\{\max_{\lambda,\theta}\Pi f(Y_i|\lambda,\theta)>\max_{\lambda,\omega} g(Y_i|\lambda,\omega)\Big|Y_i\sim f(\cdot|1,\theta_0)\big\}. \qquad \blacksquare
\end{aligned}$$

A.2 Theorem. If, in Theorem A.1, Criterion 1 is replaced by Criterion 2 (minimum Kolmogorov distance method), the conclusion remains valid.

Notation. A maximum likelihood estimator for λ computed from sample values $x_1,\ldots,x_n$ is denoted by $\hat{\lambda}(\underset{\sim}{x})$. Similarly, $\hat{\theta}(\underset{\sim}{x})$, $\hat{\ell}(\underset{\sim}{x})$ and $\hat{\omega}(x)$ are defined. The empirical distribution based on the sample values $x_1,\ldots,x_n$ is denoted by $F_n(x|\underset{\sim}{x})$.

A.3 Lemma. For any positive constant c,

$$c\hat{\lambda}(c\underset{\sim}{x})=\tilde{\lambda}(x), \quad \hat{\theta}(c\underset{\sim}{x})=\hat{\theta}(\underset{\sim}{x}).$$

If the maximum likelihood estimators are not unique, the equalities must be interpeted as saying that the left-hand side is a maximum likelihood estimator based on $\underset{\sim}{x}$.

Proof. By definition

$$\Pi f(x_i|\lambda_0,\theta_0)\le \Pi f(x_i|\hat{\lambda}(\underset{\sim}{x}),\hat{\theta}(\underset{\sim}{x})) \quad \text{for all } \lambda_0,\theta_0,\underset{\sim}{x}.$$

Thus

$$\Pi f(cx_i|\lambda_0,\theta_0)\le \Pi f\big(cx_i|\hat{\lambda}(c\underset{\sim}{x}),\hat{\theta}(c\underset{\sim}{x})\big),$$

or

$$\Pi f(x_i|c\lambda_0,\theta_0)\le \Pi f\big(x_i|c\hat{\lambda}(c\underset{\sim}{x}),\hat{\theta}(c\underset{\sim}{x})\big), \quad \text{for all } \theta_0,\lambda_0>0,$$

or

$$\Pi f(x_i|\lambda_0,\theta_0)\le \Pi f\big(x_i|c\hat{\lambda}(c\underset{\sim}{x}),\hat{\theta}(c\underset{\sim}{x})\big), \quad \text{for all } \theta_0,\lambda_0.$$

This says that $c\hat{\lambda}(c\underset{\sim}{x})$ and $\hat{\theta}(c\underset{\sim}{x})$ are also MLE estimators based on the sample $\underset{\sim}{x}$.

Proof of Theorem A.2. Note: $F_n(z|\underset{\sim}{x}) = F_n(cz|c\underset{\sim}{x})$.

$$\begin{aligned} P_{CS} &= P\Big\{ \sup_z \left|F(z|\hat{\lambda}(\underset{\sim}{X}), \hat{\theta}(\underset{\sim}{X})) - F_n(z|\underset{\sim}{X})\right| \\ &\qquad \le \sup_z \left|G(z|\hat{\ell}(\underset{\sim}{X}), \hat{\omega}(\underset{\sim}{X})) - F_n(z|\underset{\sim}{x})\right| \Big| x_i \sim f(\cdot|\lambda_0, \theta_0)\Big\}. \end{aligned} \tag{2}$$

By applying Lemma A.3 to the right-hand side of (2) it follows that

$$\begin{aligned} P_{CS} &= P\Big\{ \sup_z \left|F(z|c\hat{\lambda}(c\underset{\sim}{X}), \hat{\theta}(\underset{\sim}{X})) - F_n(cz|c\underset{\sim}{X})\right| \\ &\qquad \le \sup_z \left|G(z|c\hat{\ell}(c\underset{\sim}{X}), \hat{\omega}(c\underset{\sim}{X})) - F_n(cz|c\underset{\sim}{X})\right| \Big| X_i \sim f(\cdot|\lambda_0, \theta_0)\Big\} \\ &= P\Big\{ \sup_z \left|F(cz|\hat{\lambda}(c\underset{\sim}{X}), \hat{\theta}(c\underset{\sim}{X})) - F_n(cz|c\underset{\sim}{X})\right| \\ &\qquad \le \sup_z \left|G(cz|\hat{\ell}(c\underset{\sim}{X}), \hat{\omega}(c\underset{\sim}{X})) - F_n(cz|c\underset{\sim}{X})\right| \Big| X_i \sim f(\cdot|\lambda_0, \theta_0)\Big\}. \end{aligned} \tag{3}$$

With $Y_i = \lambda X_i$, (3) becomes

$$\begin{aligned} P_{CS} &= P\Big\{ \sup_z \left|F(cz|\hat{\lambda}(\underset{\sim}{Y}), \hat{\theta}(\underset{\sim}{Y})) - F_n(cz|\underset{\sim}{Y})\right| \\ &\qquad \le \sup_z \left|G(cz|\hat{\ell}(\underset{\sim}{Y}), \hat{\theta}(\underset{\sim}{Y})) - F_n(cz|\underset{\sim}{Y})\right| \Big| Y_i/c \sim f(\cdot|\lambda_0, \theta_0)\Big\} \\ &= P\Big\{ \sup_z \left|F(z|\hat{\lambda}(\underset{\sim}{Y}), \hat{\theta}(\underset{\sim}{Y})) - F_n(z|\underset{\sim}{Y})\right| \\ &\qquad \le \sup_z \left|G(z|\hat{\ell}(\underset{\sim}{Y}), \hat{\omega}(\underset{\sim}{Y})) - F_n(z|\underset{\sim}{Y})\right| \Big| Y_i \sim f(\cdot|c\lambda_0, \theta_0)\Big\}. \end{aligned}$$

This completes the proof because nothing has changed except the scale of the distribution from which the sample was obtained. ■

APPENDIX B – LIKELIHOOD EQUATIONS

For completeness, some of the necessary computations required for the simulations and optimizations are listed here.

Gamma density

For a data sample of size n, the log likelihood function is given by:

$$\log L = n\nu \log \lambda + (\nu - 1) \sum_{i=1}^{n} \log x_i - \lambda \sum_{i=1}^{n} x_i - n \log \Gamma(\nu). \tag{1}$$

The derivative of the log likelihood function with respect to λ yields

$$\frac{\partial \log L}{\partial \lambda} = \frac{n\nu}{\lambda} - \sum_{i=1}^{n} x_i.$$

Setting the derivative equal to zero yields the maximum likelihood estimator (MLE) for λ:

$$\hat{\lambda} = \frac{\nu}{\bar{x}}, \qquad \bar{x} = \sum_{i=1}^{n} x_i / n. \tag{2}$$

Substituting (2) into (1) and taking the derivative with respect to ν yields

$$\frac{\partial \log L}{\partial \nu} = n \log \nu - n\psi(\nu) - n \log \bar{x} + \sum_{i=1}^{n} \log x_i, \tag{3}$$

where $\psi(\nu) = \Gamma'(\nu)/\Gamma(\nu)$ is the digamma function. We can now solve for the MLE of ν by finding the unique root of equation (3); uniqueness follows from properties of the digamma function.

Weibull density

The log likelihood function is given by:

$$\log L = n \log \alpha + n\alpha \log \lambda + (\alpha - 1) \sum_{i=1}^{n} \log x_i - \lambda^{\alpha} \sum x_i^{\alpha}. \tag{4}$$

The derivatives of (4) with respect to α and λ are given by

$$\frac{\partial \log L}{\partial \alpha} = \frac{n}{\alpha} + n \log \lambda + \sum_{i=1}^{n} \log x_i - \lambda^{\alpha} \log \lambda \sum x_i^{\alpha} - \lambda^{\alpha} \sum x_i^{\alpha} \log x_i, \tag{5}$$

$$\frac{\partial \log L}{\partial \lambda} = \frac{n\alpha}{\lambda} + \alpha \lambda^{\alpha-1} \sum x_i^{\alpha}. \tag{6}$$

A straightforward calculation shows that the MLE for λ is given by

$$\hat{\lambda} = \left(\frac{n}{\sum x_i^{\alpha}} \right)^{1/\alpha}. \tag{7}$$

Substituting (7) into (5) yields

$$\frac{\partial \log L}{\partial \alpha} = \frac{n}{\alpha} + \sum \log x_i - \frac{n}{\sum x_i^{\alpha}} \sum x_i^{\alpha} \log x_i$$

which can be solved numerically for the MLE of α.

Lognormal density

The MLE are given by

$$\hat{\mu} = \frac{\sum_{i=1}^{n} \log x_i}{n}, \qquad \hat{\sigma}^2 = \frac{\sum (\log x_i - \hat{\mu})^2}{n}.$$

Geometric extreme exponential density

The log likelihood function is given by

$$\log L = n \log(\gamma\lambda) - \lambda \sum_{i=1}^{n} x_i - 2 \sum \log(1 - \bar{\gamma} e^{-\lambda x_i}). \tag{8}$$

The derivatives of (8) with respect to γ and λ are given by

$$\begin{aligned} \frac{\partial \log L}{\partial \gamma} &= \frac{n}{\gamma} - 2 \sum \frac{e^{-\lambda x_i}}{1 - \bar{\gamma} e^{-\lambda x_i}}, \\ \frac{\partial \log L}{\partial \lambda} &= \frac{n}{\lambda} - \sum_{i=1}^{n} x_i - 2 \sum \frac{\bar{\gamma} x_i e^{-\lambda x_i}}{1 - \bar{\gamma} e^{-\lambda x_i}}. \end{aligned}$$

Setting the derivatives equal to zero does not lead to closed form solutions, and the MLE must be obtained numerically.

For this density, the expectation is given by $EX = -\gamma \log \gamma / \lambda \bar{\gamma}$.

List of Figures

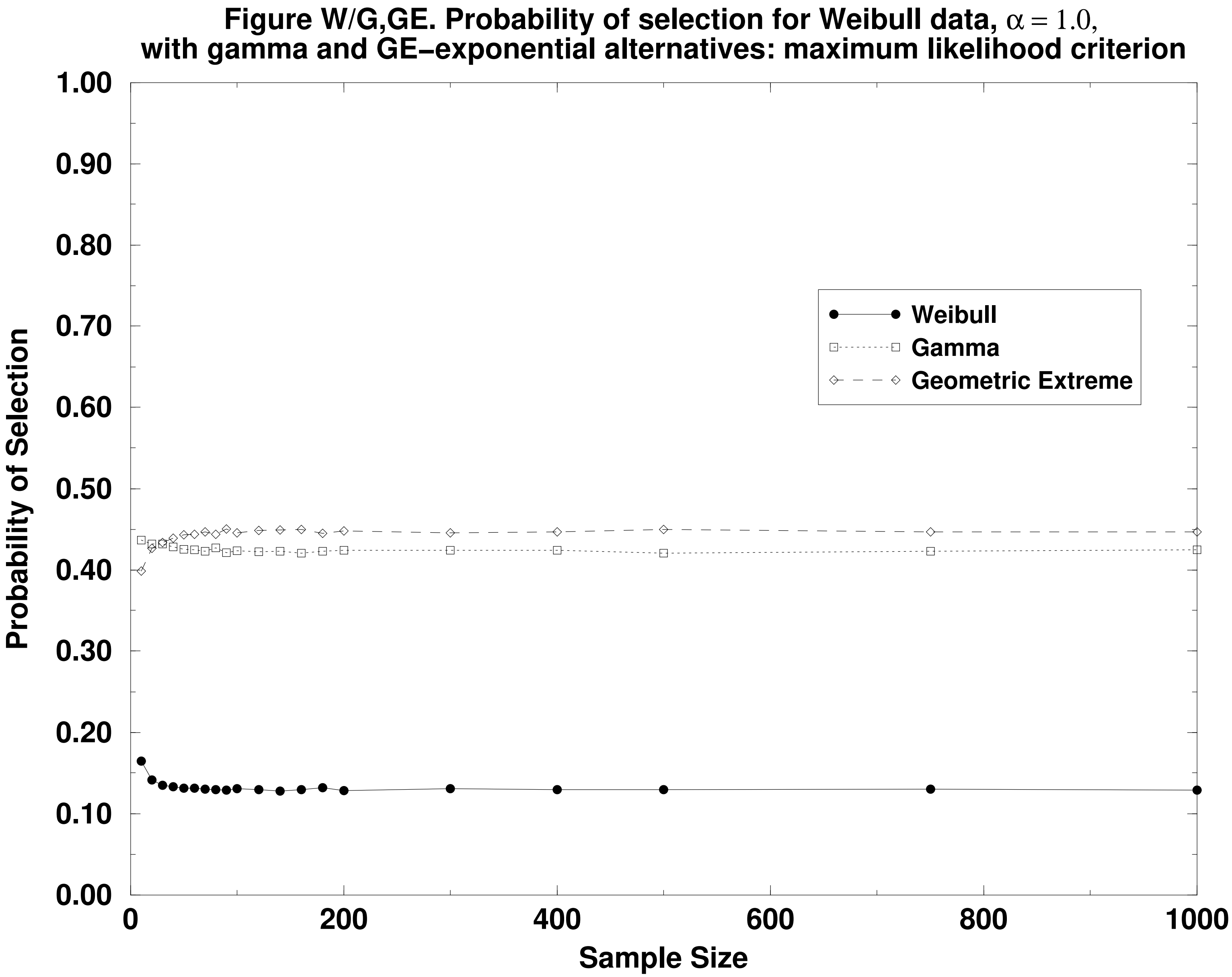

Figure 1: Weibull data with gamma, geometric extreme alternatives, $\alpha = 1.0$

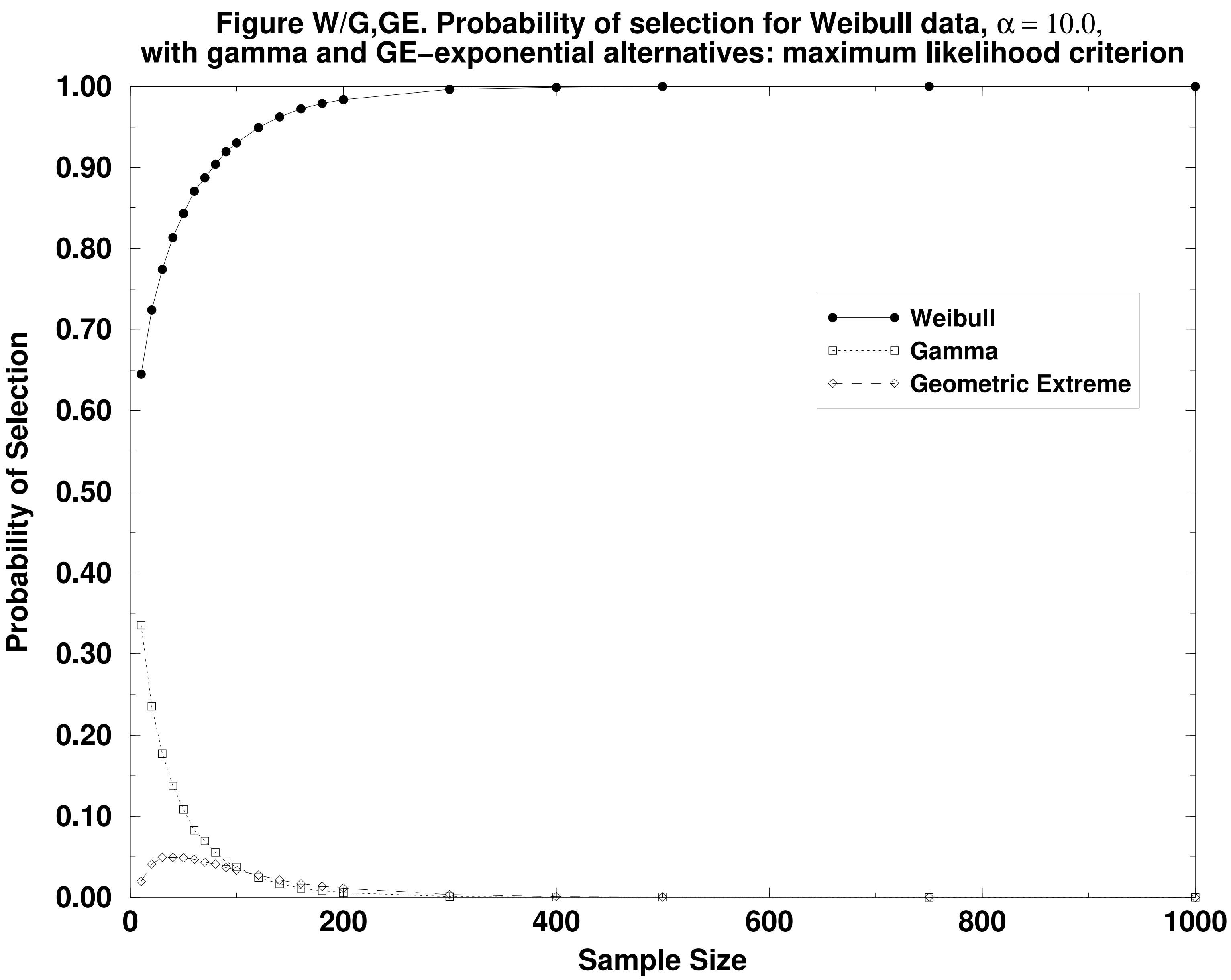

Figure 2: Weibull data with gamma, geometric extreme alternatives, $\alpha = 10.0$

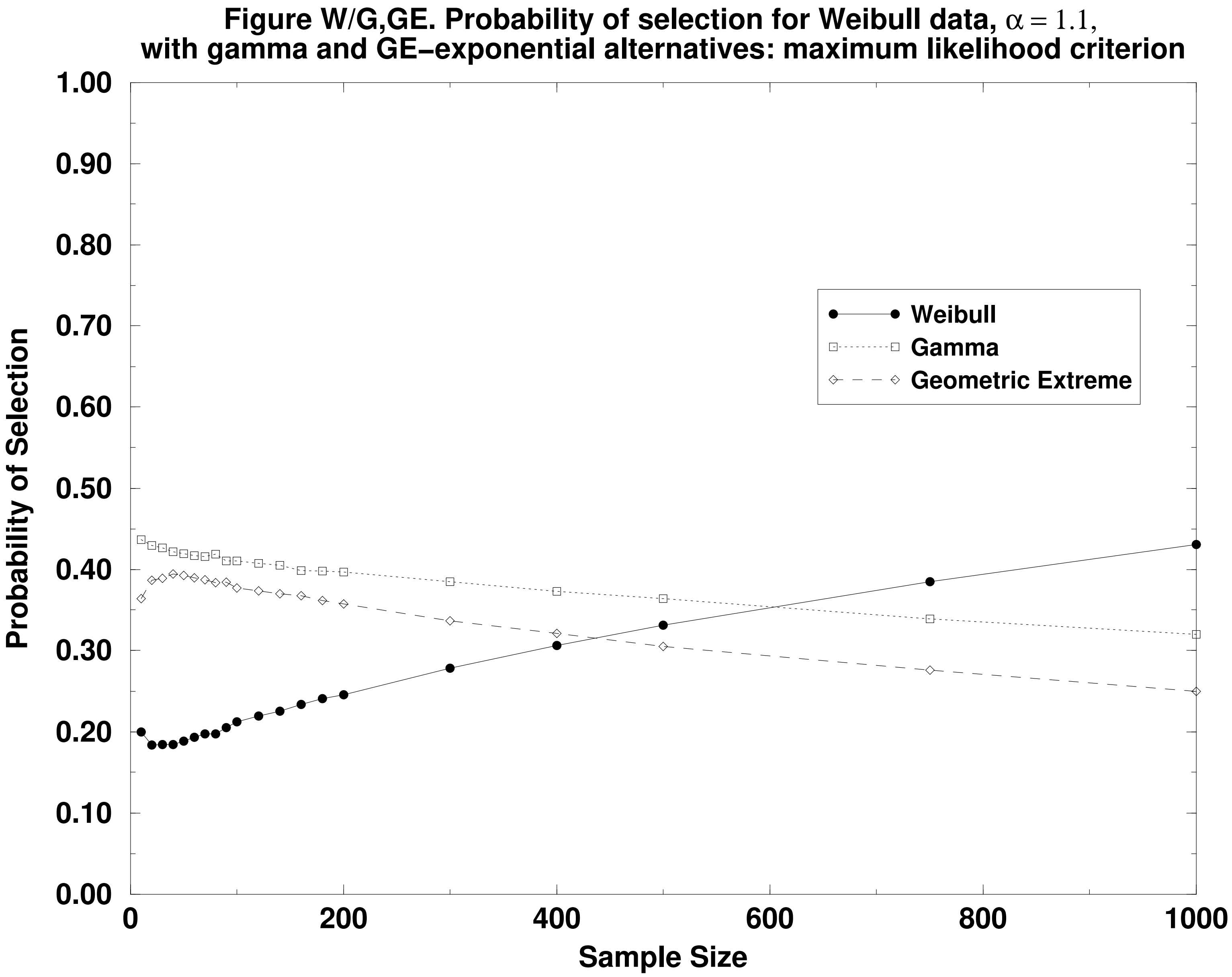

Figure 3: Weibull data with gamma, geometric extreme alternatives, $\alpha = 1.1$

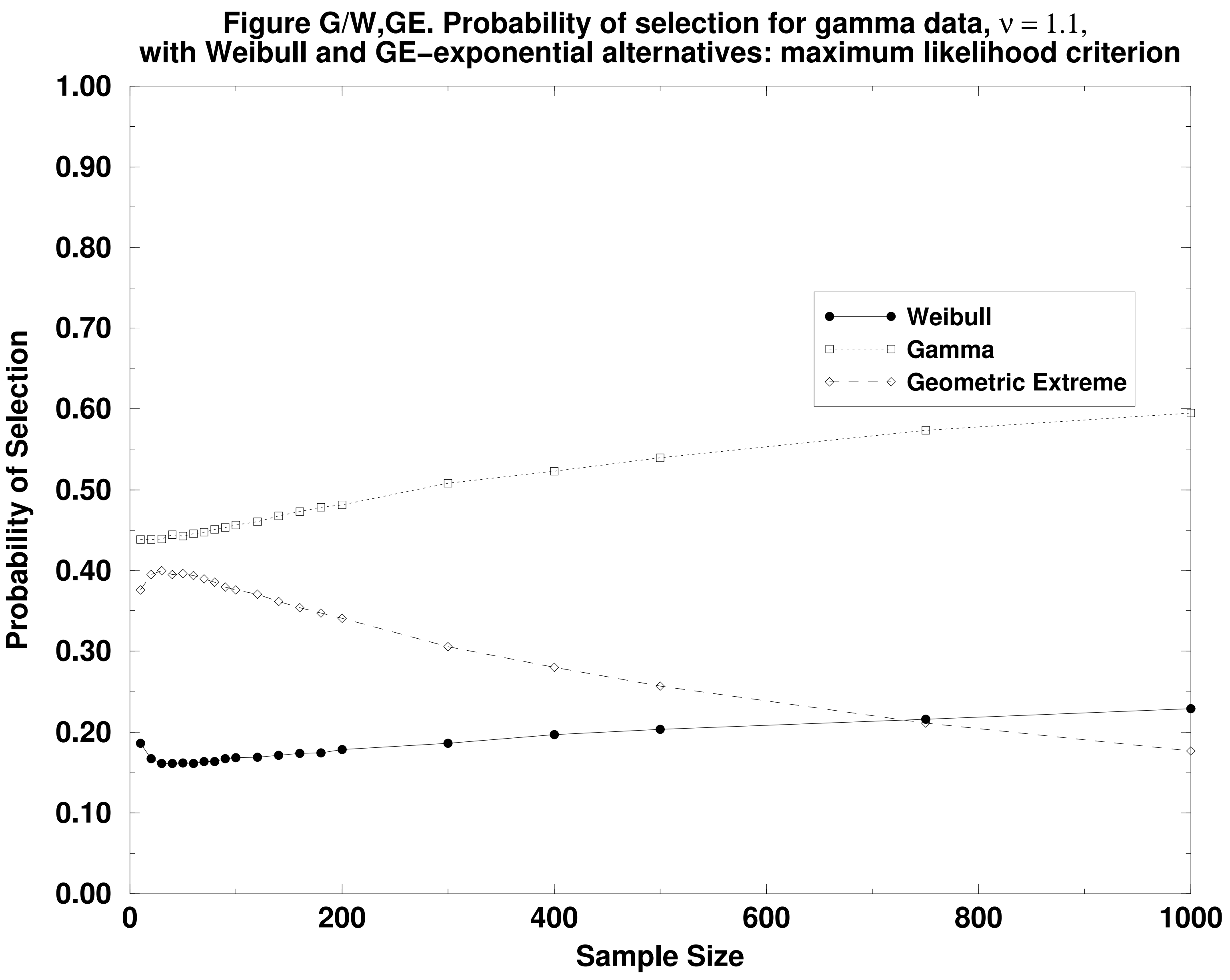

Figure 4: Gamma data with Weibull, geometric extreme alternatives, $\nu = 1.1$

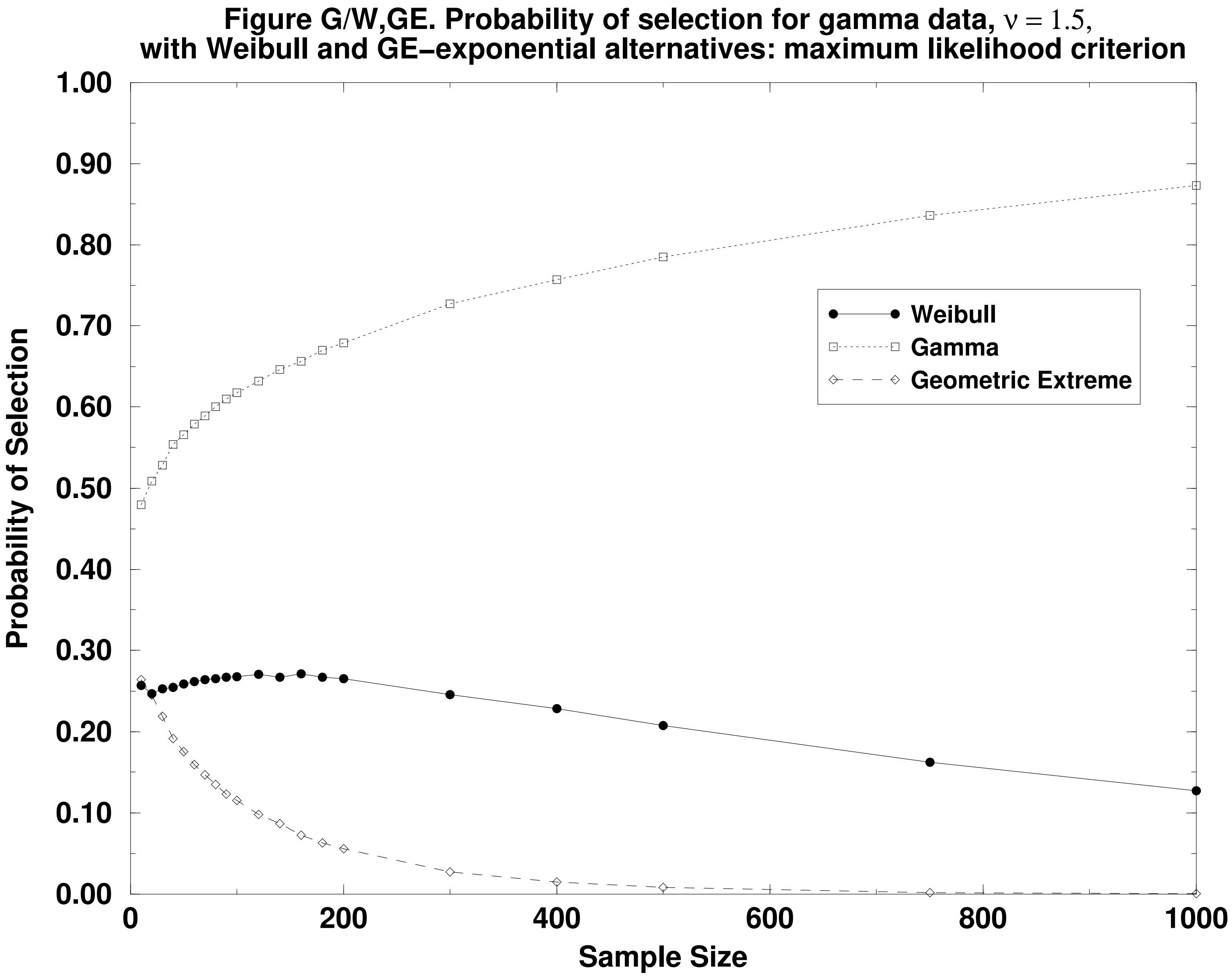

Figure 5: Gamma data with Weibull, geometric extreme alternatives, $\nu = 1.5$

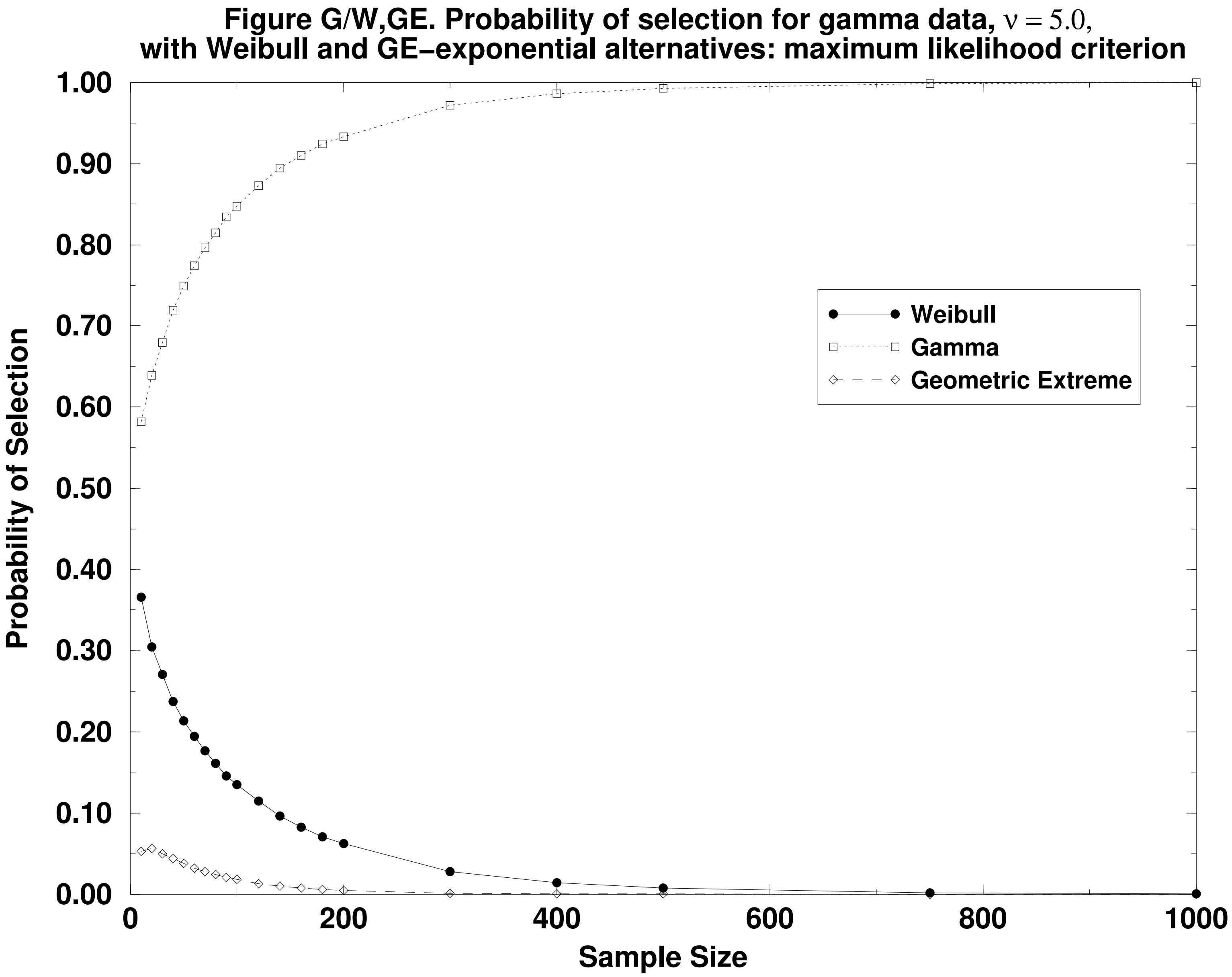

Figure 6: Gamma data with Weibull, geometric extreme alternatives, $\nu = 5.0$

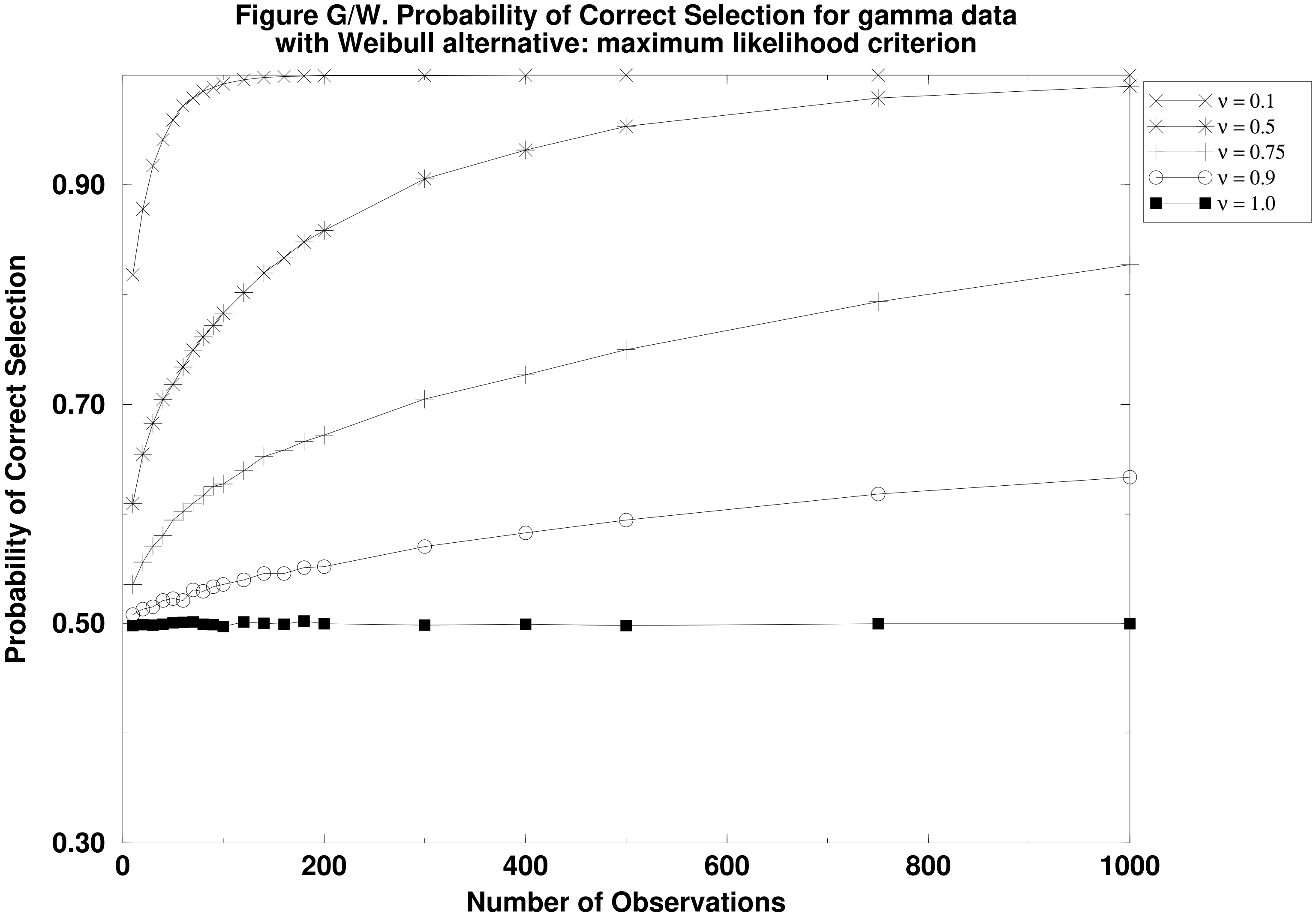

Figure 7: Gamma data with Weibull alternative, MLR criterion

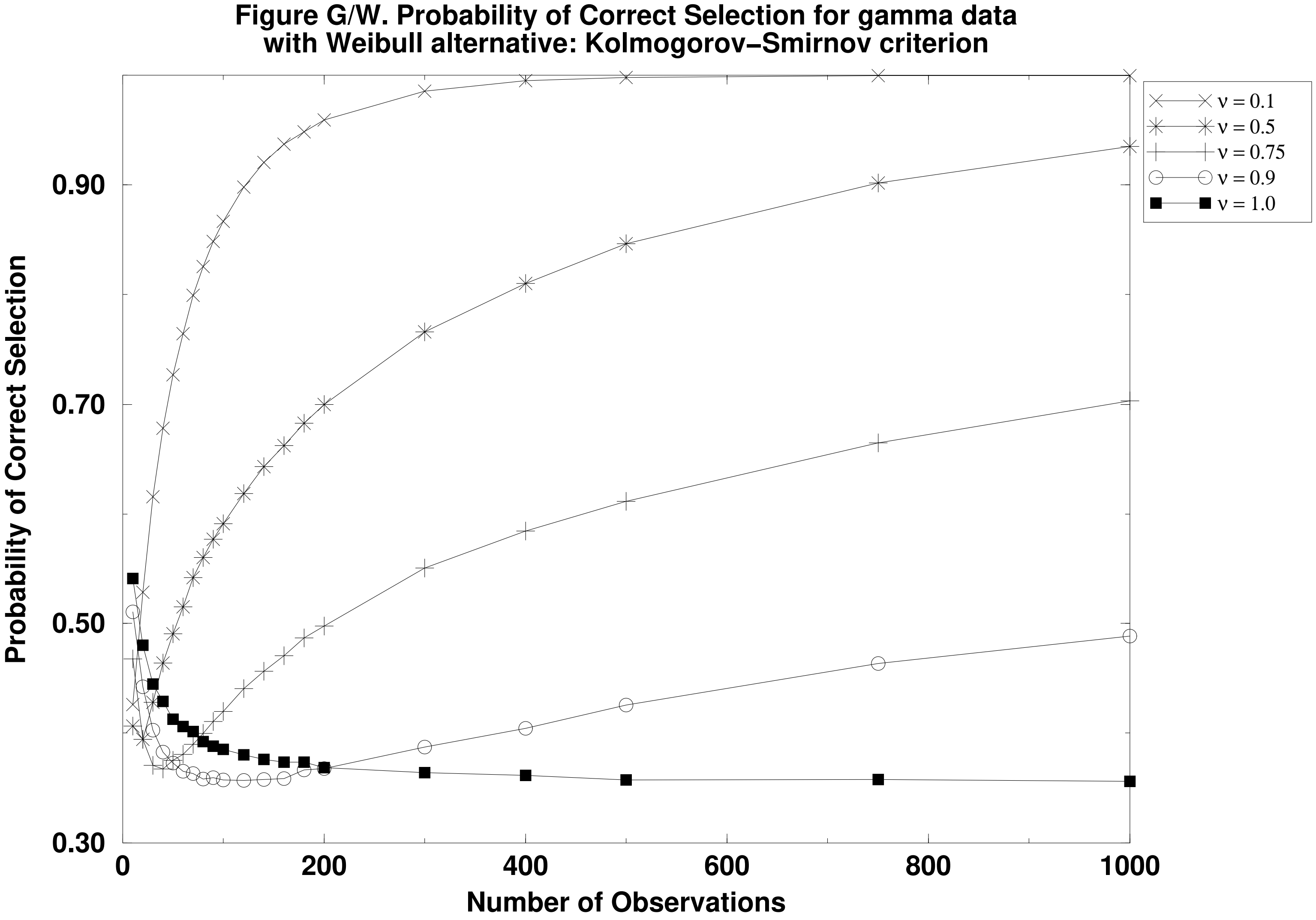

Figure 8: Gamma data with Weibull alternative, KS criterion

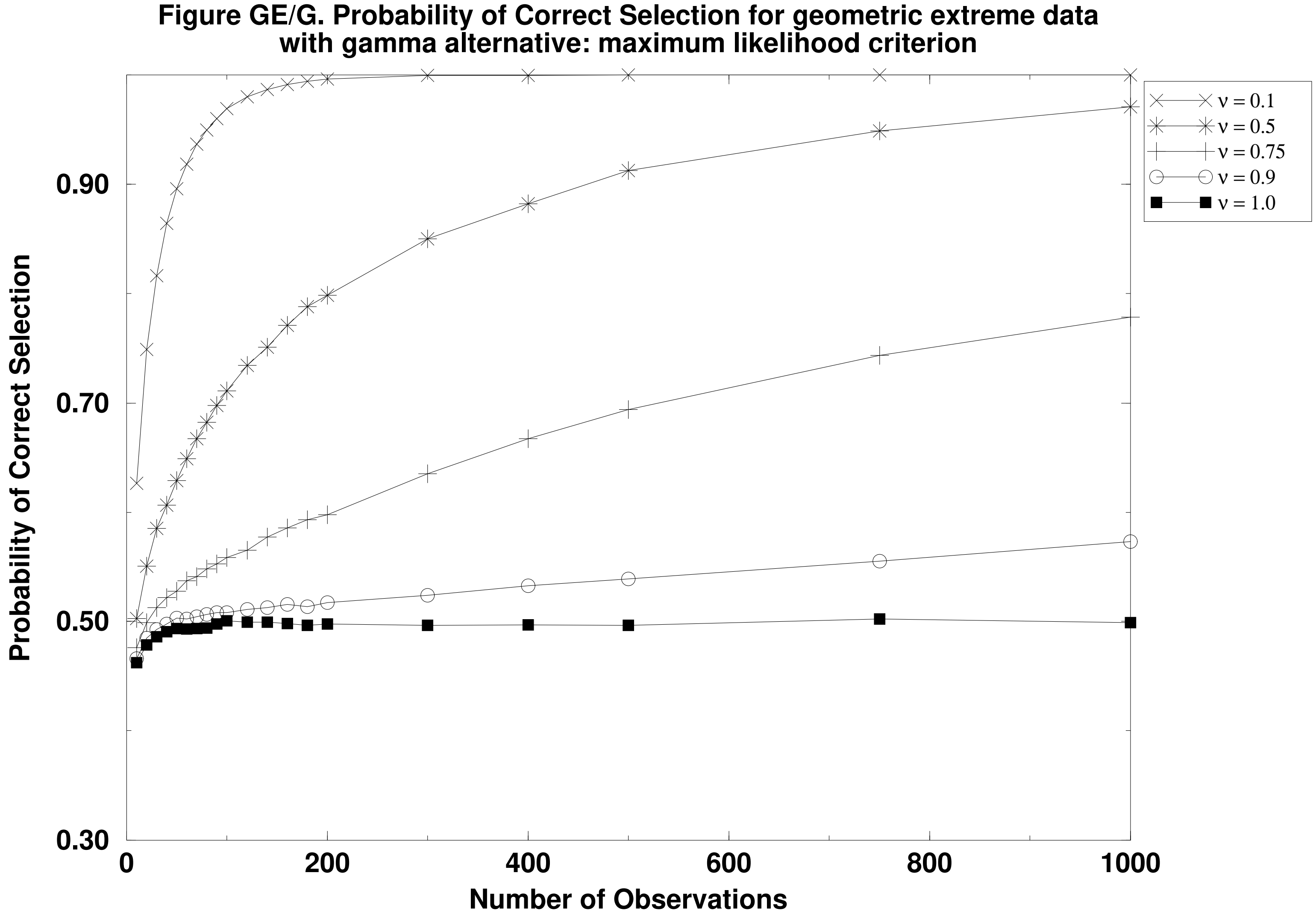

Figure 9: Geometric extreme data with gamma alternative, MLR criterion

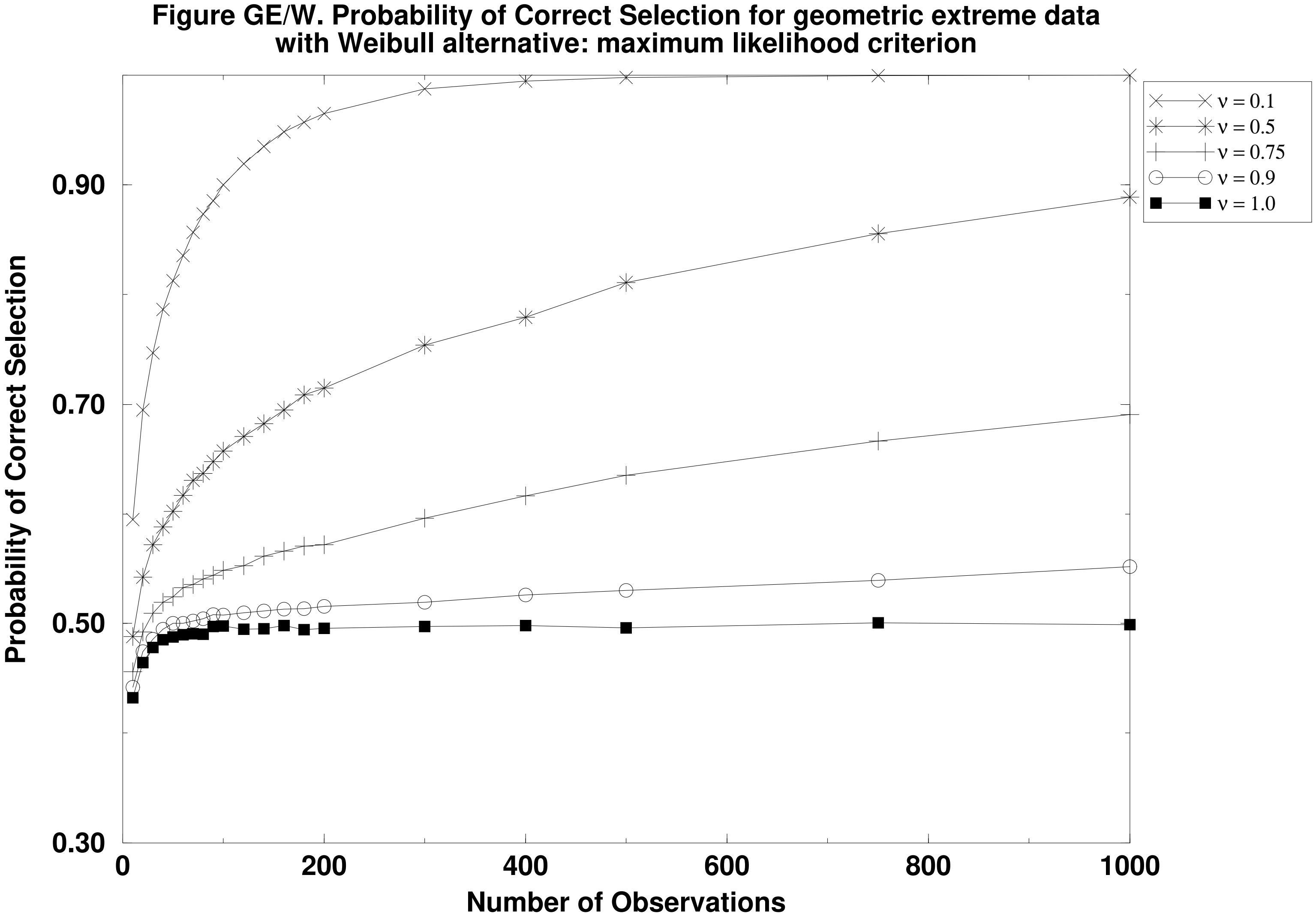

Figure 10: Geometric extreme data with Weibull alternative, MLR criterion

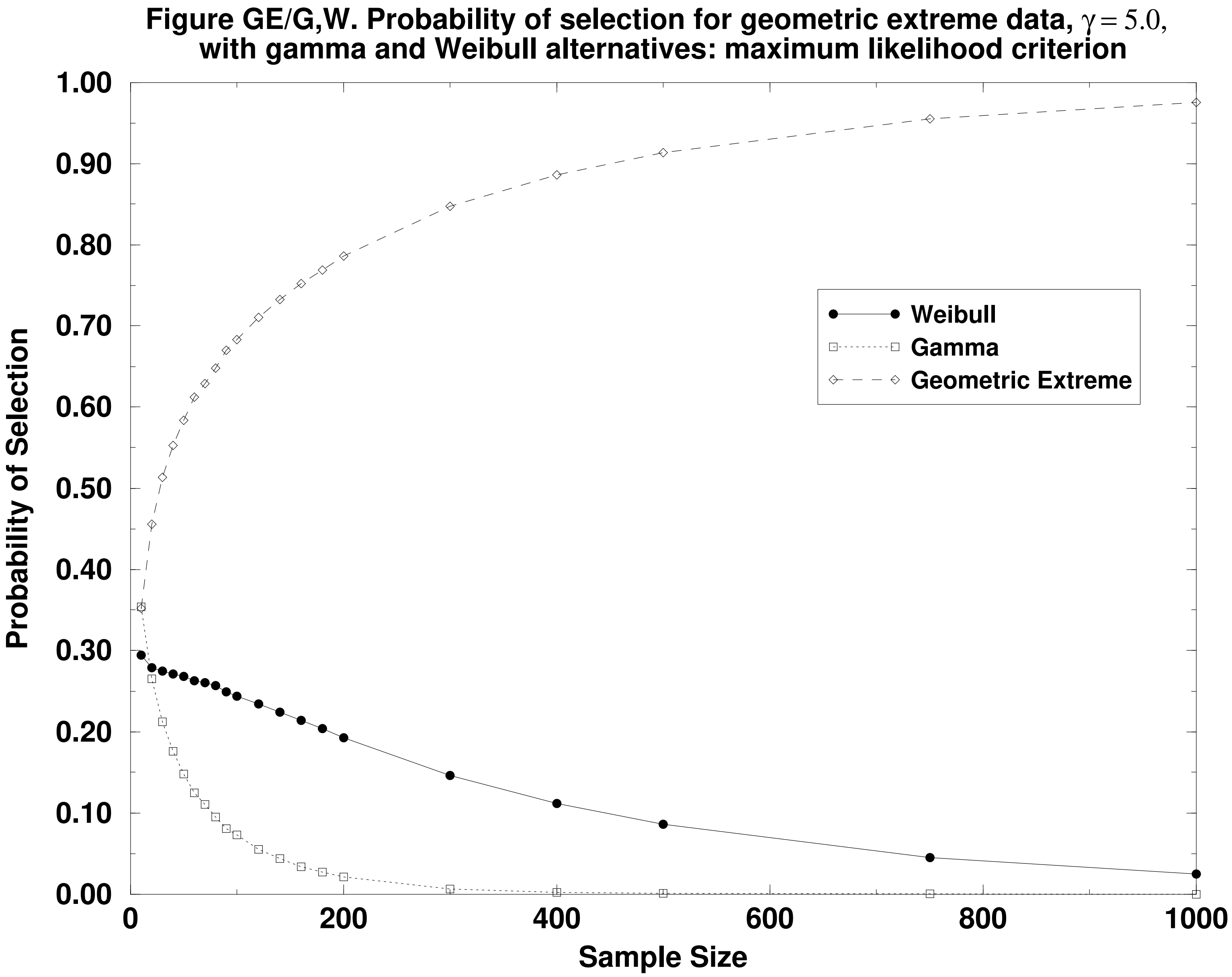

Figure 11: Geometric extreme data with gamma, Weibull alternative, $\gamma = 5.0$

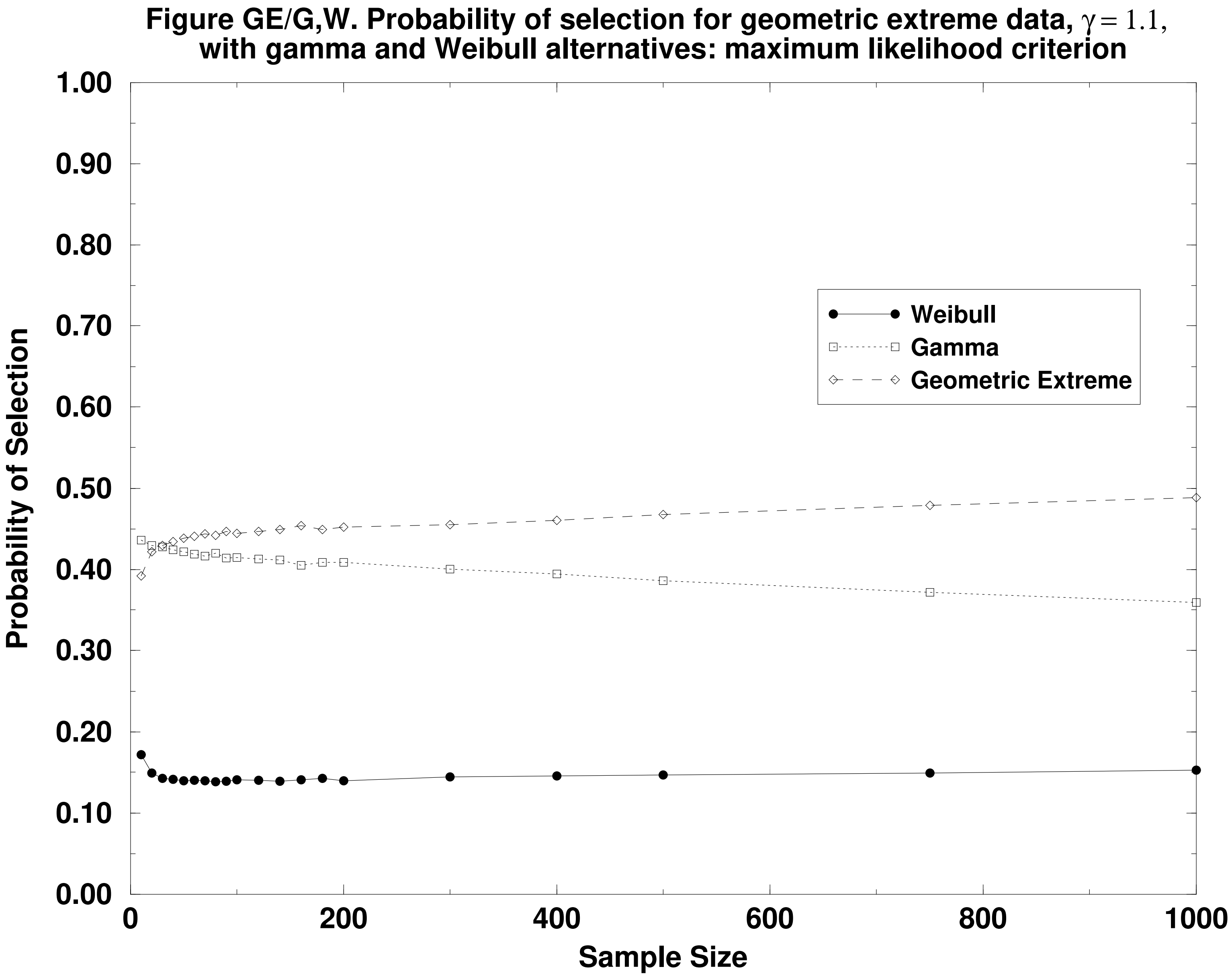

Figure 12: Geometric extreme data with gamma, Weibull alternative, $\gamma = 1.1$

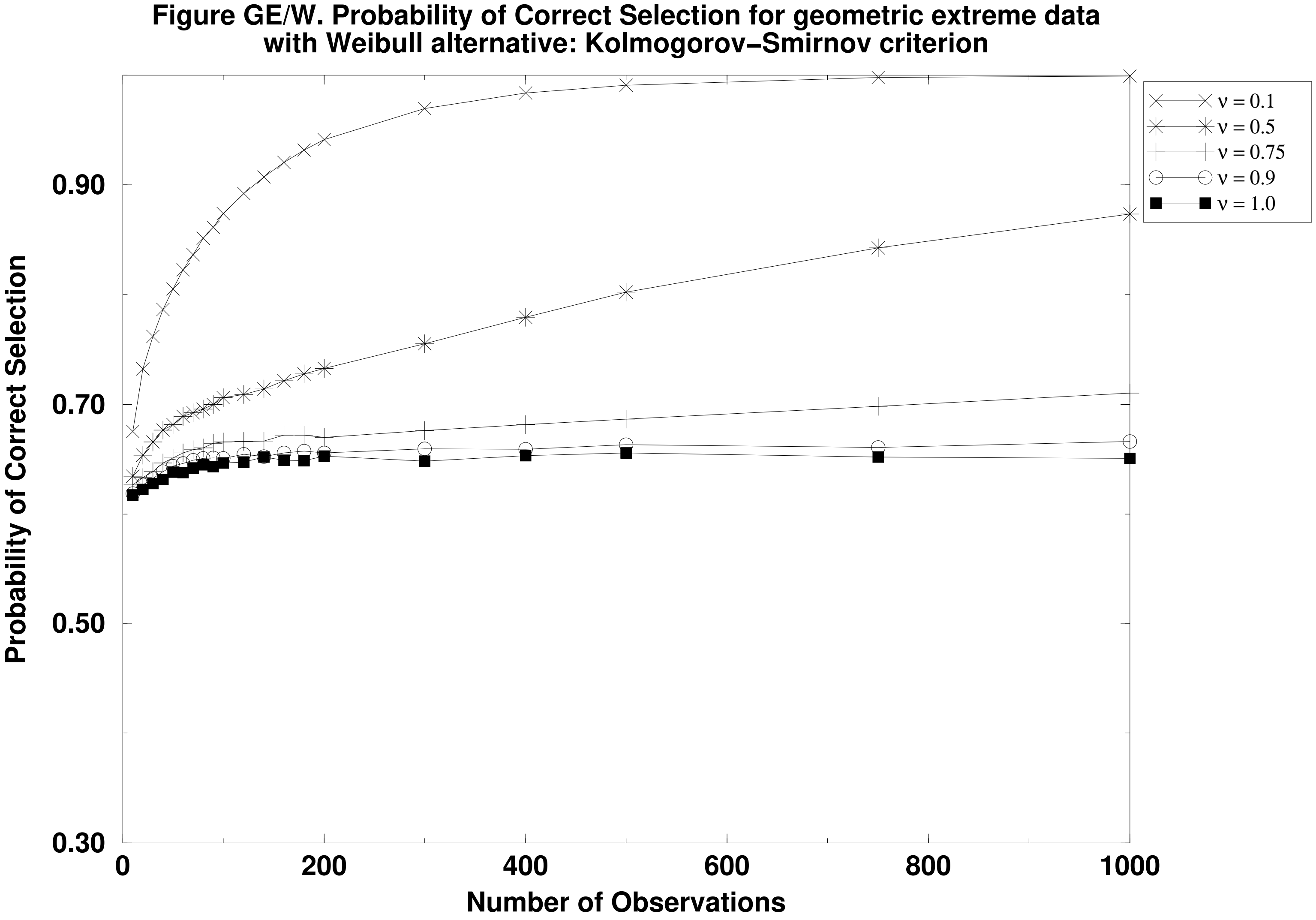

Figure 13: Geometric extreme data with Weibull alternative, KS criterion

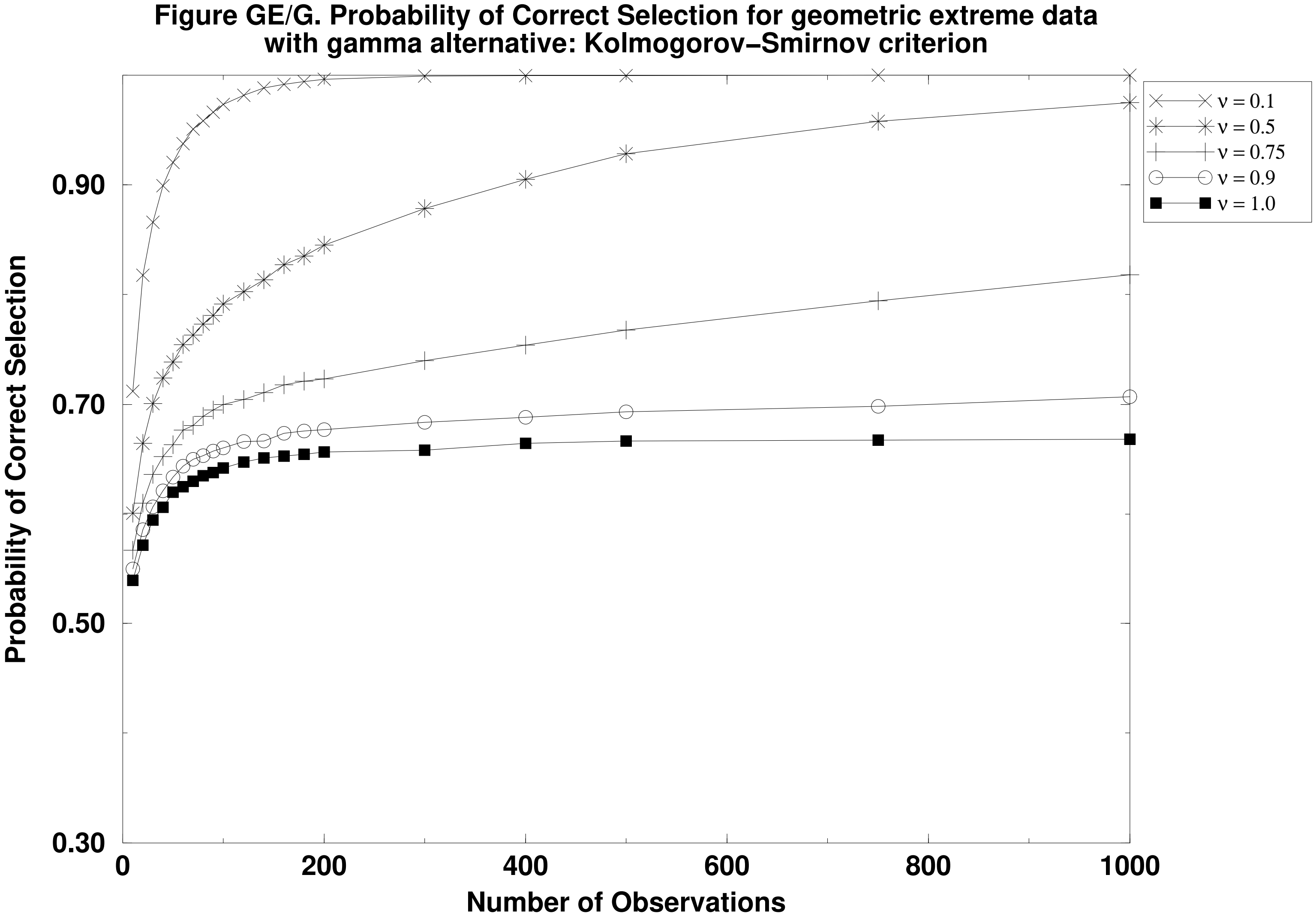

Figure 14: Geometric extreme data with gamma alternative, KS criterion

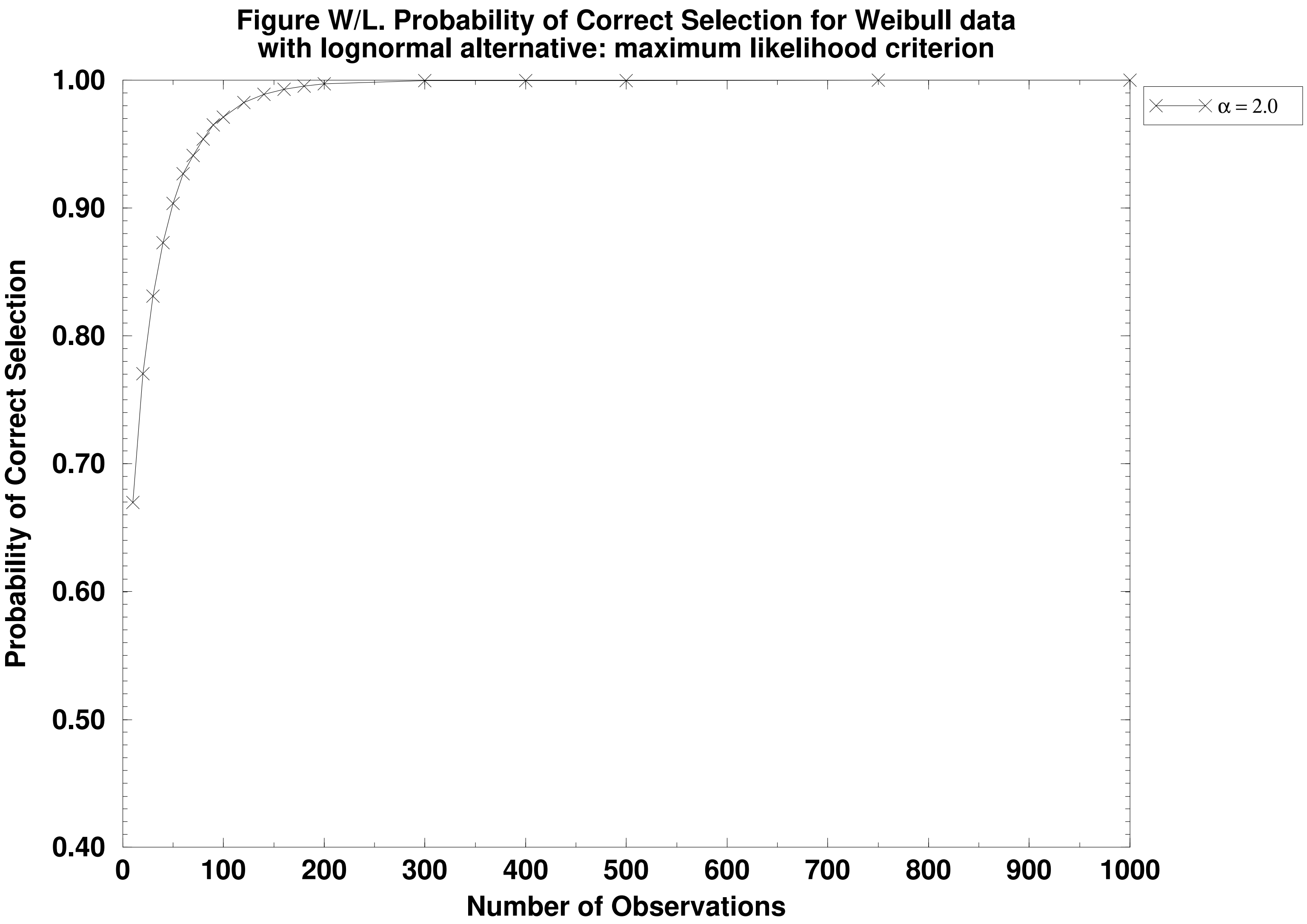

Figure 15: Weibull data with lognormal alternative, MLR criterion

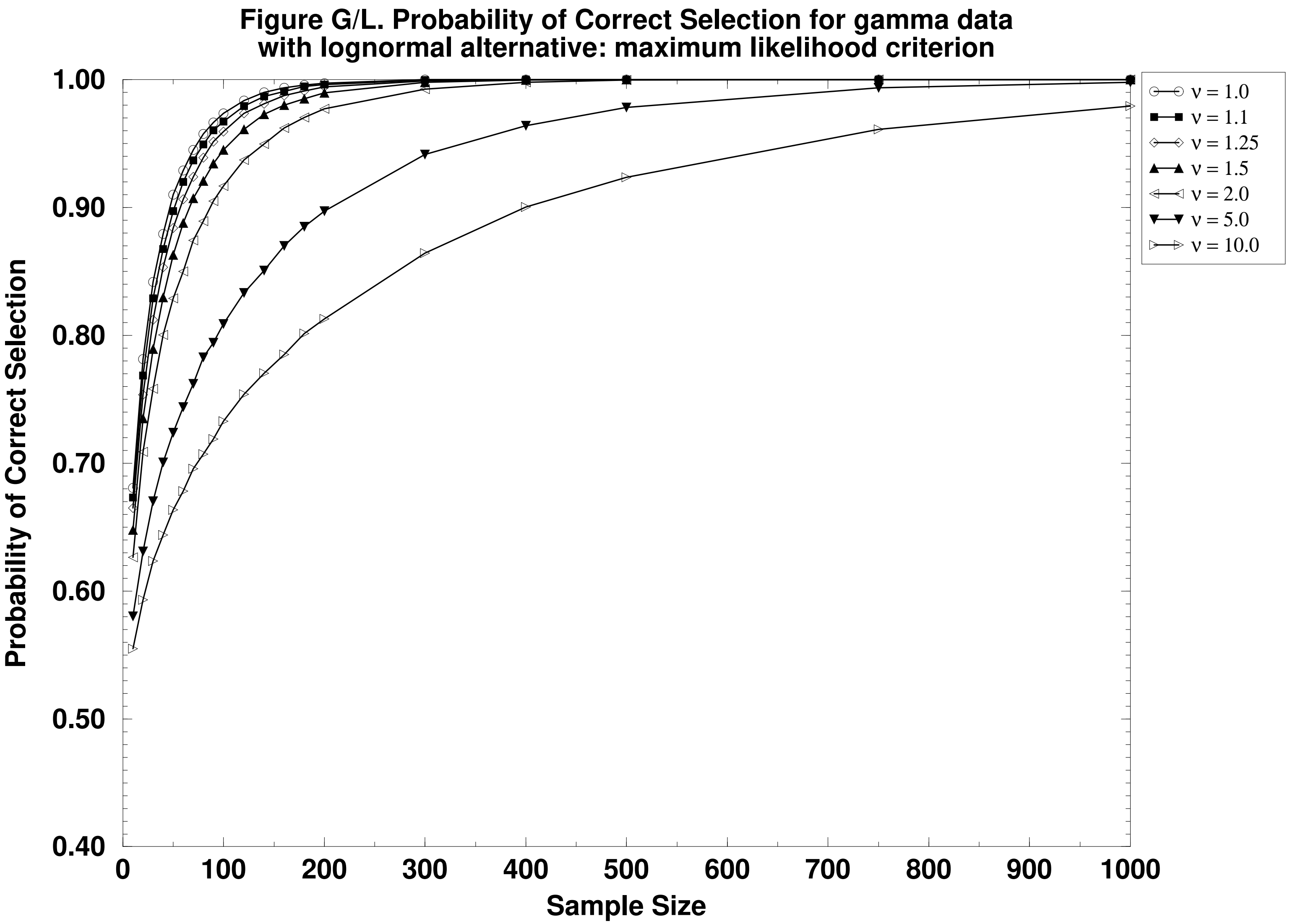

Figure 16: Gamma data with lognormal alternative, MLR criterion

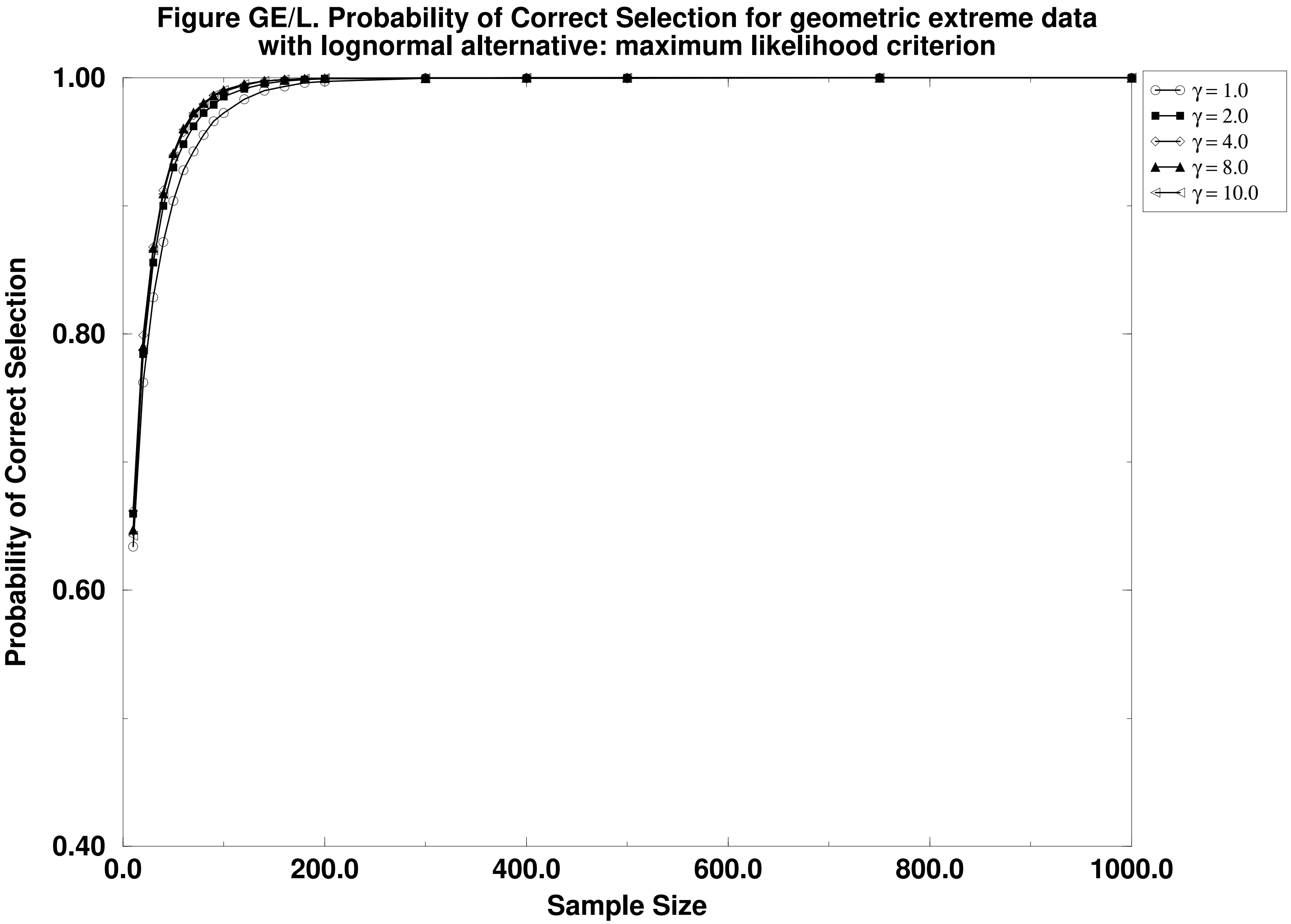

Figure 17: Geometric extreme data with lognormal alternative, MLR criterion

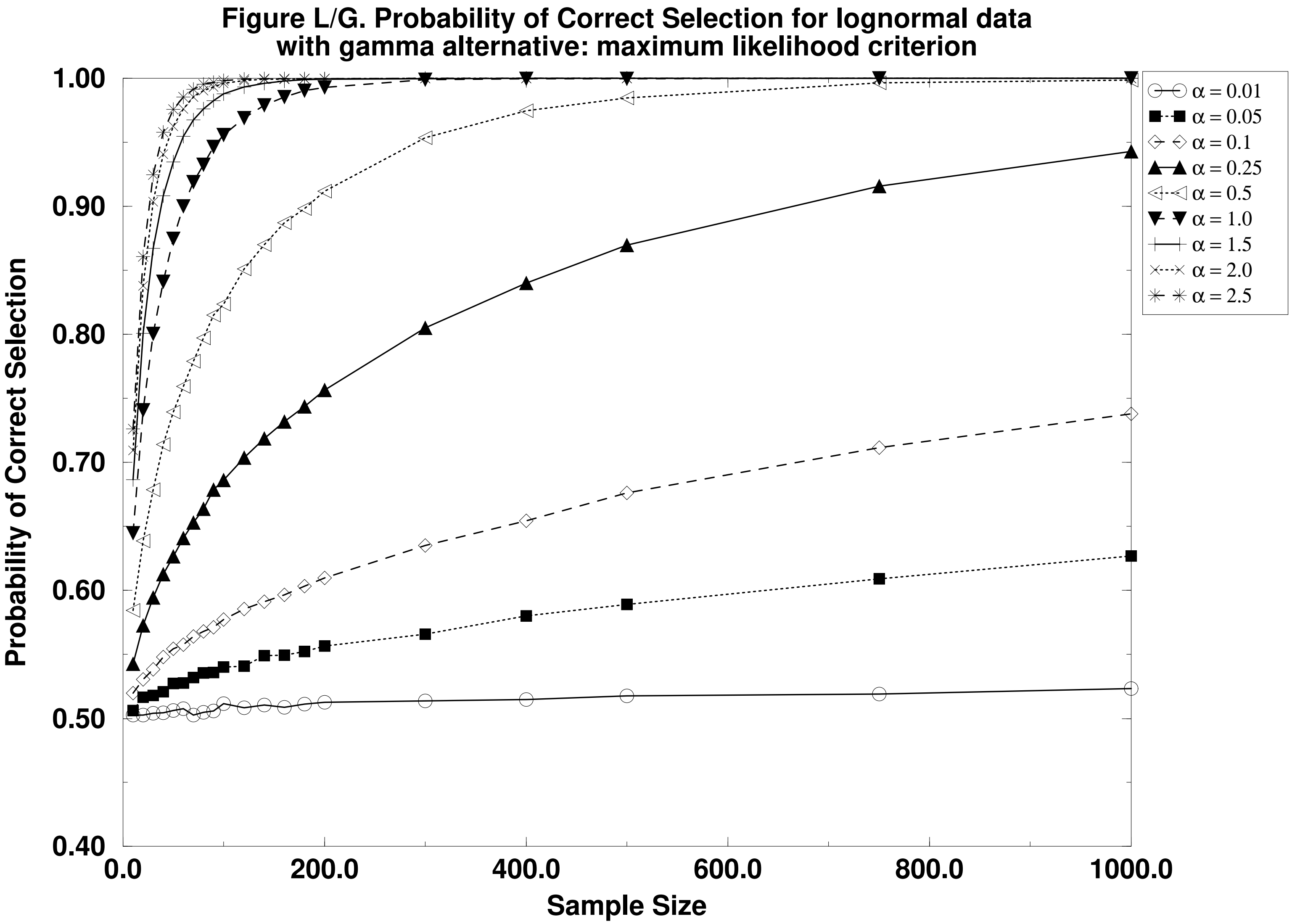

Figure 18: Lognormal data with gamma alternative, MLR criterion

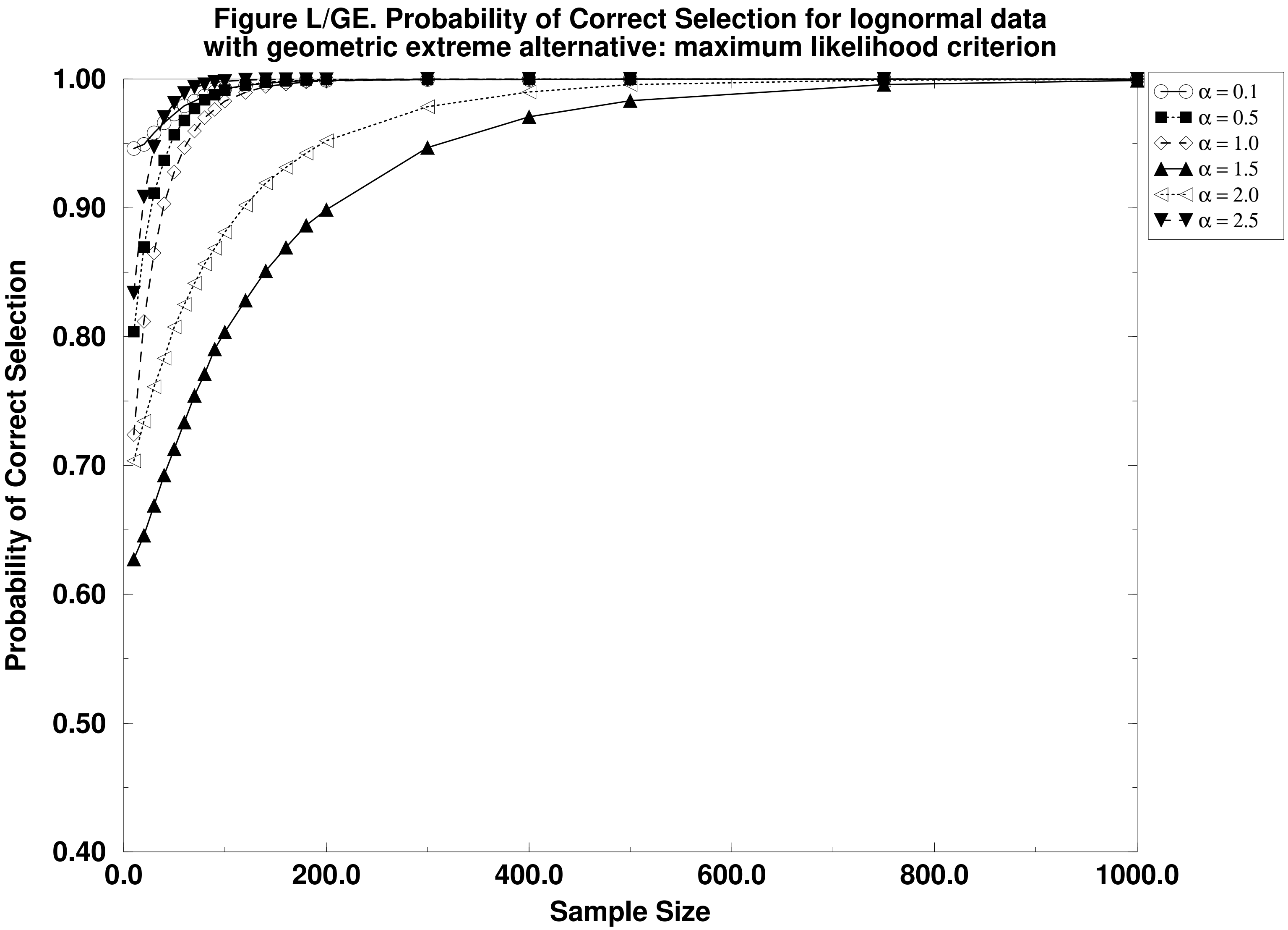

Figure 19: Lognormal data with geometric extreme alternative, MLR criterion