28 September 1993

Calculations of Upper Limits

Gerry Lynch
Lawrence Berkeley National Laboratory, Berkeley, CA 94720

It is the custom in high energy physics to quote a measurement either as a value with an error or as a limit. When a value is quoted, it is standard practice to quote a central value plus or minus an error of one standard deviation, or equivalently a value centered in a 68% confidence interval. For limits, not only is there no standard practice of what confidence level to use, though 90% and 95% are the most common, but also there are variations in what method to use for calculating a limit.

In this note, I consider only the case in which the measured quantity x is Gaussianly distributed:

$$f(x;\mu) = \frac{1}{\sqrt{2\pi}\sigma}\, e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

where σ is a known constant and μ is an unknown constant that we want to estimate. We may measure x many times and get an average $\overline{x}$. In this case, we still have a Gaussian distribution, but with a smaller value of σ, and the nature of the distribution is unchanged. So I will denote x as the estimator of μ.

One reason why the estimator $x \pm \sigma$ is well accepted is that it is statistically correct. By that we mean that the interval $[x-\sigma$ to $x+\sigma]$ is a 68% confidence interval no matter what the value of μ happens to be. In other words, we know (if we make no mistakes) that the interval quoted in this way has a 68% probability of containing the true value of μ. Although there are an infinite number of confidence intervals that have this property, this one stands out as the narrowest of them.

Sometimes the practice of quoting $x+\sigma$ is "derived" with a Bayesian argument, which treats μ as if it were a random variable and says that the distribution in μ with x fixed is equal to the distribution in x with μ fixed times a distribution in μ that represents our prior knowledge of μ. When there is no prior knowledge of μ, this state of ignorance is represented by a function that is flat in μ. This exercise leads to the same conclusion: that $x+\sigma$ is a 68% confidence interval for μ.

This Bayesian approach is considered unsatisfactory for two reasons. One reason is that it treats μ as a random variable, and it is not one. The second reason is that choosing a constant distribution to represent the lack of knowledge of μ is quite arbitrary. We could just as well have assumed a constant distribution in μ^2 or $\sqrt{\mu}$ and gotten a different answer. In fact, by choosing the distribution appropriately, we could get almost any answer we try to get. From this point of view, it can be argued that the Bayesian approach is useless. Nevertheless, it is worth pointing out that when we use a distribution that is constant in μ, the Bayesian approach gets the "correct answer," that is, we get the same confidence interval that is obtained by valid statistical methods. If we were to use a distribution that is flat in any non-linear function of μ, we would get a wrong answer. Thus, although the

Bayesian derivation is mathematically invalid, the Bayesian method does give the right answer for a Gaussian distribution, if and only if the a priori distribution that is assumed is one that is constant in the mean of the Gaussian distribution.

Now consider the Gaussian problem in which we still do not know the value of μ, but we know that it cannot be negative. In this case, we do have prior knowledge of the mean. Although I do not want to focus on any particular experiment, the following hypothetical experiment illustrates the type of data that I have in mind. An experiment is done to measure the rate R of some process that has never been observed, and may not even occur. This rate measurement is done in the presence of a relatively large background rate B. The experiment measures the background to be B$=5100 \pm 71$ and measures the sum R+B$=4900 \pm 70$. So the resultant measurement is R$=-200 \pm 100$. The experimenters can, and probably should, quote the result in this form. But, in addition, they feel compelled to quote an upper limit for R. They could say that the interval $[-400 < \mathrm{R} < 0]$ is a 95% confidence interval and quote a 95% upper limit of zero, or they could quote a 68% upper limit of -100. They will not do this because this is in conflict with other knowledge that they have, the knowledge that μ cannot be negative.

Of course, the above experiment (like most real experiments) does not have a distribution that is exactly Gaussian — it has the difference of two Poisson distributions. Furthermore, σ is not known exactly, but is itself an estimate. We ignore these complications partly because the differences are not large and partly because if the differences were significant, we could, with some effort, calculate the appropriate intervals anyway.

The solution to this problem that is advocated by the Particle Data Group amounts to a Bayesian method that takes the prior knowledge of the mean to be a function that is zero when μ is negative and is constant in the positive region. Then the upper limit U corresponding to a probability P is given by a renormalized form of

$$\frac{\int_0^{\mathrm{U}} f\,(x;\mu)\, d\mu}{\int_0^{\infty} f\,(x;\mu)\, d\mu} = \mathrm{P},$$

where x is set at the measured value. This approach not only suffers from the problems that we mentioned before for Bayesian methods, it also has the problem that it is not a valid confidence interval. A valid confidence interval for a probability P will have the probability P of containing the true value of μ no matter what the true value is. This upper limit does not have that property, for when the true mean is zero, the interval from zero to U is correct 100% of the time.

One can also ask whether the use of a flat distribution to represent the ignorance of the value of μ is appropriate in this case. To test this we can calculate the probability that this interval from 0 to U contains the true answer. Figures 1a and 1b show the results of such calculations for three cases, corresponding to the assumptions that the prior knowledge of μ is represented by functions that are flat in the $\sqrt{\mu}$, flat in μ and flat in μ^2. All three cases have probabilities that are one for $\mu = 0$ and approach P for large μ. Clearly, the distribution that is constant in μ stands out as the best one — it is much closer to P than are the other two. Thus, in this case, as in the case in which there are no restrictions on the value of μ, the assumption that a distribution that is flat in the mean leads to the best answer. These figures also illustrate the assertion made by the PDG that this method is conservative one, for which the probability is always greater than P. Note that if we use a distribution that is flat in the $\sqrt{\mu}$, the result loses this conservative property.

It seemed attractive to me to look for a method of calculating upper limits that produces a valid confidence interval. Clearly a scheme that calculates only an upper limit for any value of x will not satisfy this, because it will be right 100% of the time when $\mu = 0$. Then the problem is to find a prescription for defining lower and upper limits with the following properties:

1. The scheme provides a proper confidence interval for all μ;
2. When x is small, it defines an upper limit only;
3. When x is large, the result approaches the interval of $x \pm \mathrm{n}\sigma$, where n is the number of standard deviations that corresponds to P (n=1 for 68% and n=2 for 95%). One property that such a scheme must have is that the lower edge of the confidence interval be zero for small x and non-zero at large x, and the transition point is well determined. It is at the point x_{p} that has a fraction P of the area of the Gaussian to the left of it. This is at 0.47σ for p=68% and 1.64σ for P=95%.

At first thought, one might wonder whether the above criteria are so strict that no solution exists, but in fact, the real problem is that there are innumerable solutions, and there is no clear way to prefer one over another. Figure 2a shows three intervals for a 68% confidence level: the symmetric interval $x \pm \sigma$ (dot-dash lines), a simple upper limit (dashed line), and the standard Bayesian renormalized upper limit (dotted curve). One can construct a valid confidence interval by choosing an upper limit U(x) that is above (or on) the dashed line. A horizontal line through this point corresponds to a fixed value of μ, and we can always find a lower limit point L on this line such that U and L satisfy the property that for this μ, x will be between the limits 68% of the time. The locus of these lower limit points defines the curve for the lower limit L(x) that corresponds to the U(x) that was chosen. These curves also have the property that for any vertical line they define a confidence interval that has a 68% probability of containing μ. Alternatively, one could choose L(x) in the allowed region and calculate a corresponding U(x).

Figure 2b shows an example of such a construction. In this case, U(x) was set equal to the renormalized upper limit for $x < x\mathrm{p}$, and made to change smoothly to $x + \sigma$. The resultant solution for the lower limit looks rather ugly (though ugliness is in the eye of the beholder). Not shown are the first few solutions that I found, which were uglier. Figures 3a and 3b have the solution for the case in which U(x) was taken to be the renormalized upper limit displaced to the left by $\mathrm{n}\sigma - x_{\mathrm{p}}$. In this case, the curves are less ugly, though we have lost the property that at low x, the upper limit is exactly the standard one.

I considered this approach attractive because it produces a valid confidence interval with the form that we are accustomed to, namely an upper limit at low x and $x \pm \mathrm{n}\sigma$ at large x. Furthermore, it provides a way to bridge this gap between these two regimes in a continuous way. Unfortunately there exist many solutions of this type, and I am at a loss to find a reasonable criterion to use to choose one over an other. So, unless we can came up with a compelling set of criteria for choosing, I see no reason to advocate the use of such a procedure for quoting errors.

In doing this investigation I was influenced by a number of sources:

1. It was Don Groom's efforts to revise what is said on the subject in the *Review of Particle Properties* (Phys. Rev D, **45** (1992)) that motivated me to think about this topic again, and discussions with Don were quite useful.
2. I reread the treatment of confidence intervals by Cramer (Harold Cramer, Mathematical Methods of Statistics), which I regard as the gospel on this subject.
3. I read again what was written by Frank Solmitz (Analysis of Experiments in Particle Physics, Ann. Rev. Nuc. Sci., **14**, 375 (1961)), who presented this subject well from the point of view of a physicist.
4. I also read what Harold Jeffries (*Theory of Probability*, 1961) had to say on this subject, and could not make much sense out of what he recommended.
5. I read for the first time the excellent paper by Frank James and Matts Roos (Phys. Rev. D **44**, 44 (1990)). I certainly agree with them that in the interest of combining the results from different experiments, experimentalists should always publish the measured value and error, even if the value is unphysical, whenever the distribution is approximately Gaussian.

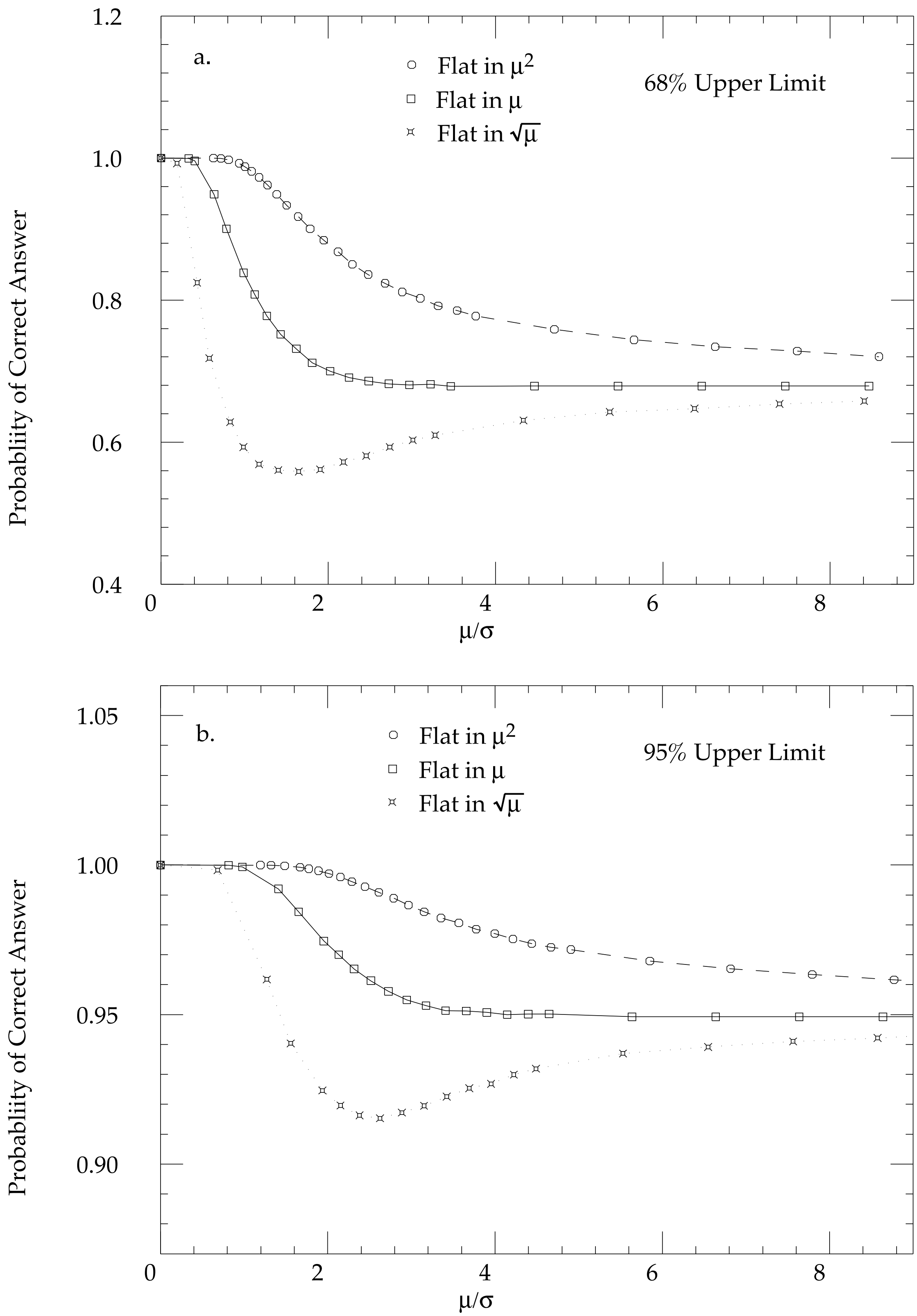
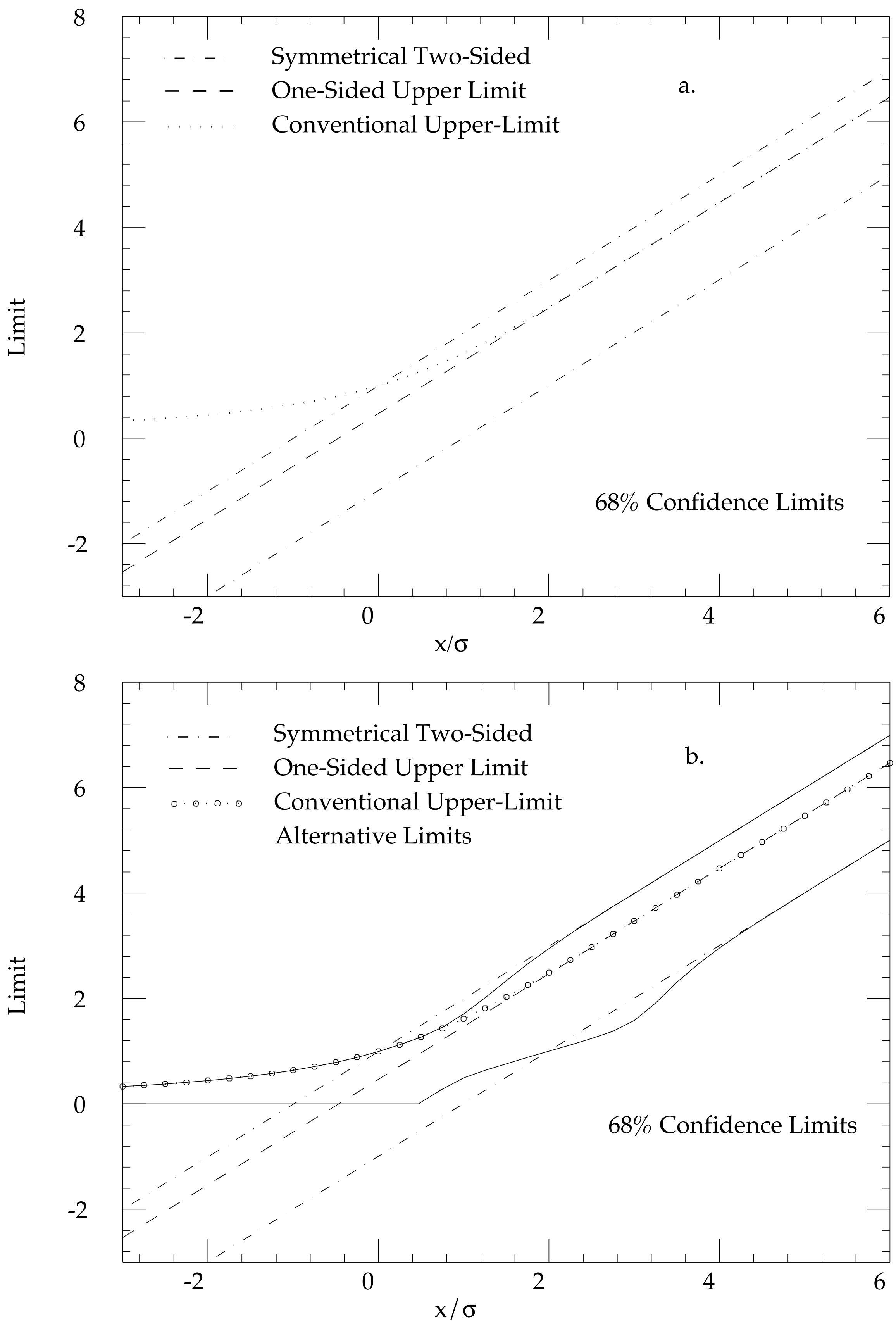

a.
Flat in μ²
Flat in μ
Flat in √μ
68% Upper Limit
1.2
1.0
0.8
0.6
0.4
0
2
4
6
8
μ/σ
Probablity of Correct Answer
b.
Flat in μ²
Flat in μ
Flat in √μ
95% Upper Limit
1.05
1.00
0.95
0.90
0
2
4
6
8
μ/σ
Probablity of Correct Answer

Figure 1



a.
Symmetrical Two-Sided
One-Sided Upper Limit
Conventional Upper-Limit
68% Confidence Limits
Limit
x/σ
b.
Symmetrical Two-Sided
One-Sided Upper Limit
Conventional Upper-Limit
Alternative Limits
68% Confidence Limits
Limit
x/σ

Figure 2

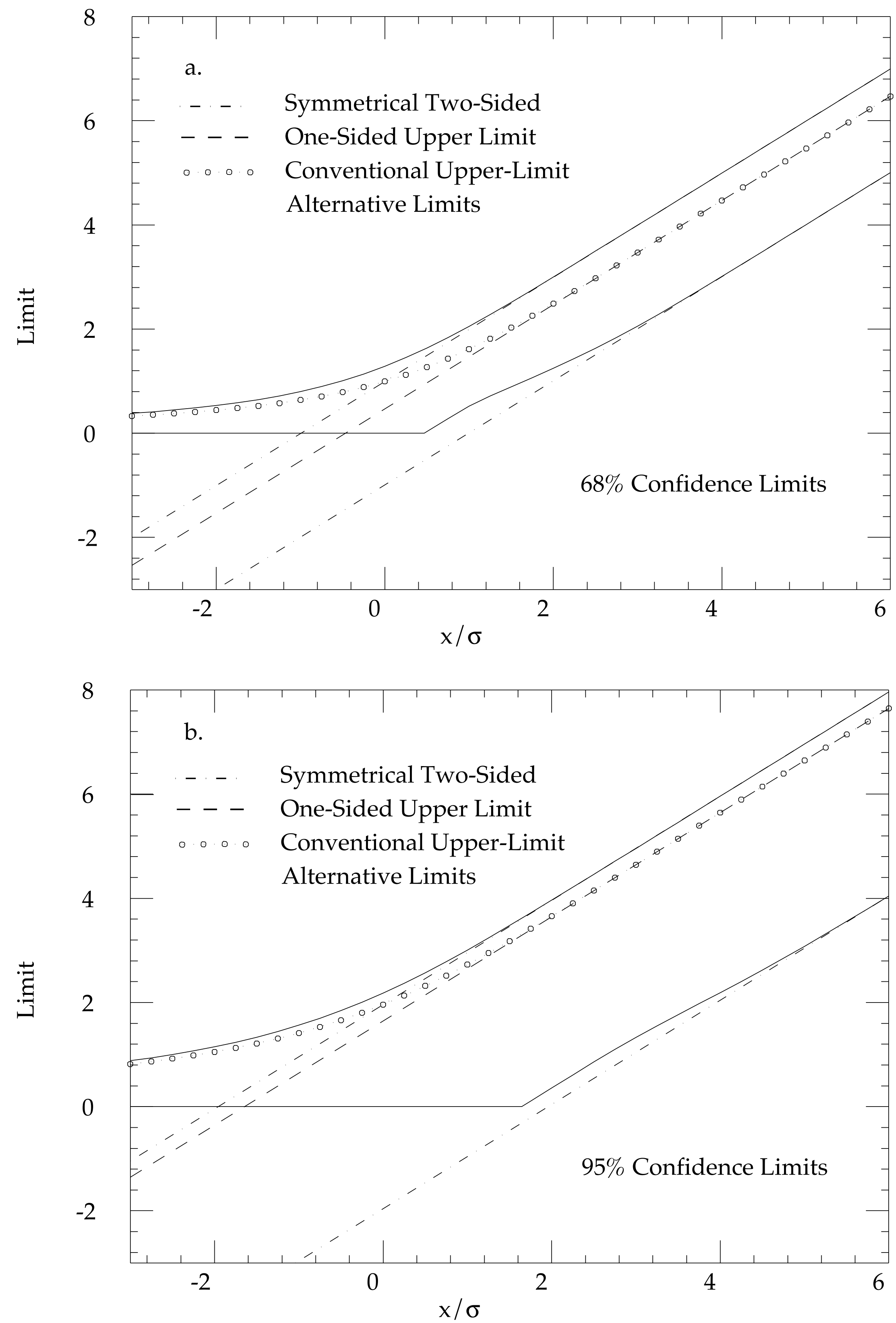

a.
Symmetrical Two-Sided
One-Sided Upper Limit
Conventional Upper-Limit
Alternative Limits
68% Confidence Limits
Limit
x/σ
b.
Symmetrical Two-Sided
One-Sided Upper Limit
Conventional Upper-Limit
Alternative Limits
95% Confidence Limits
Limit
x/σ

Figure 3