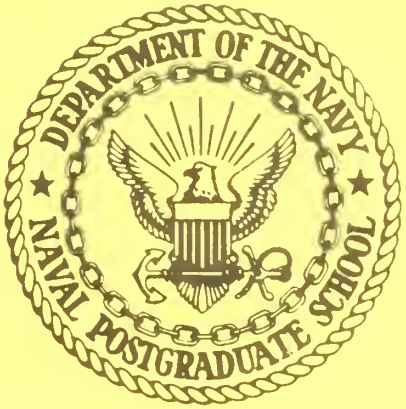
NPS52-84-006

NAVAL POSTGRADUATE SCHOOL

Monterey, California



ABSOLUTE BOUNDS ON THE MEAN AND STANDARD DEVIATION OF TRANSFORMED DATA FOR CONSTANT-DERIVATIVE TRANSFORMATIONS

Neil C. Rowe

April 1984

Prepared for:

Chief of Naval Research
Arlington, VA 22217

NAVAL POSTGRADUATE SCHOOL
Monterey, California

Commodore R. H. Shumaker
Superintendent

D. A. Schrady
Provost

The work reported herein was supported in part by the Foundation Research Program of the Naval Postgraduate School with funds provided by the Chief of Naval Research.

This report was prepared by:

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: NPS52-84-006	2. GOVT ACCESSION NO.
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE *(and Subtitle)*: Absolute Bounds on the Mean and Standard Deviation of Transformed Data for Constant-Derivative Transformations	5. TYPE OF REPORT & PERIOD COVERED
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): Neil C. Rowe	8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Naval Postgraduate School, Monterey, California 93943	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: 61152: RR000-01-10, N0014841F0001
11. CONTROLLING OFFICE NAME AND ADDRESS: Chief of Naval Research, Arlington, Virginia 22217	12. REPORT DATE: April 1984
	13. NUMBER OF PAGES: 40
14. MONITORING AGENCY NAME & ADDRESS *(if different from Controlling Office)*	15. SECURITY CLASS. *(of this report)*: Unclassified
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT *(of this Report)*

Approved for public release, distribution unlimited

17. DISTRIBUTION STATEMENT *(of the abstract entered in Block 20, if different from Report)*

18. SUPPLEMENTARY NOTES

19. KEY WORDS *(Continue on reverse side if necessary and identify by block number)*

transformations, statistical bounds, mean, standard deviation, function approximation, production systems, exploratory data analysis, estimation, nonparametric estimation, inequalities, antisampling

20. ABSTRACT *(Continue on reverse side if necessary and identify by block number)*

We investigate absolute bounds (or inequalities) on the mean and standard deviation of transformed data values, given only a few statistics on the original set of data values. Our work applies primarily to transformation functions whose derivatives are constant-sign for a positive range (e.g. logarithm, antilog, square root, and reciprocal). With such functions we can often get reasonably tight absolute bounds, so that distributional assumptions about the data needed for confidence intervals can be eliminated.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

S/N 0102-LF-014-6601

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

We investigate a variety of methods for obtaining such bounds, first examining bounding curves which are straight lines, then those that are quadratic polynomials. While the problem of finding the best quadratic bound is an optimization problem with no closed-form solution, we display a variety of closed-form quadatic bounds which can come close to the optimal solution. We emphasize what can be done with prior knowledge of the mean and standard deviation of the untransformed data values, but do address some other statistics too.

Absolute bounds on the mean and standard deviation of transformed data for constant-derivative transformations

Neil C. Rowe
Department of Computer Science
Code 52
Naval Postgraduate School
Monterey, CA 93943

Table of Contents

List of Figures

Abstract

We investigate absolute bounds (or inequalities) on the mean and standard deviation of transformed data values, given only a few statistics on the original set of data values. Our work applies primarily to transformation functions whose derivatives are constant-sign for a positive range (e.g. logarithm, antilog, square root, and reciprocal). With such functions we can often get reasonably tight absolute bounds, so that distributional assumptions about the data needed for confidence intervals can be eliminated. We investigate a variety of methods for obtaining such bounds, first examining bounding curves which are straight lines, then those that are quadratic polynomials. While the problem of finding the best quadratic bound is an optimization problem with no closed-form solution, we display a variety of closed-form quadratic bounds which can come close to the optimal solution. We emphasize what can be done with prior knowledge of the mean and standard deviation of the untransformed data values, but do address some other statistics too.

Acknowledgments

The work reported herein was suported in part by the Foundation Research Program of the Naval Postgraduate School with funds provided by the Chief of Naval Research.

keywords: transformations, statistical bounds, mean, standard deviation, function approximation, production systems, exploratory data analysis, estimation, nonparametric estimation, inequalities, antisampling

abbreviated title: Absolute Bounds on Statistics of Transformed Values

1. Introduction

Standard transformations of numeric data values such as logarithm, antilog, square root, square, cube, and reciprocal are frequently appropriate as a prelude to statistical analysis of finite data sets [7]. Sometimes, however, the data are already aggregated into counts and means, and the original data values lost. This happens when the original data is too large to handle and/or contains sensitive information, as the U. S. Census, which publishes much of its data as aggregates. We may also deliberately create "database abstracts" of aggregate statistics to facilitate quick statistical estimates by "antisampling" methods [10]. Statistics on the transformed values cannot be calculated uniquely when the original data is so preaggregated[1]. But if we are doing exploratory data analysis [13, 6], an estimate of a statistic on the transformed data may be all that we need. We address one set of methods for obtaining such estimates, by finding absolute (unconditionally guaranteed) bounds on the mean and standard deviation for data under some common transformations.

Absolute bounds are the only true "nonparametric" form of estimate, and as such have advantages. Compared to "reasonable-guess" estimates [9], biasedness of the estimator need not be dealt with, while at the same time providing numbers close to the true answer for this category of problems. As [7] discusses, confidence intervals for the mean and standard deviation of transformed data are difficult to obtain and methods are subject to exceptions, and thus absolute bounds easily obtained are appealing. Tight enough absolute bounds can be equivalent to a good estimate. An estimate of a statistic can also be logically incorrect when bounds are tight, i.e. it may not be a statistic of any possible distribution consistent with the constraints. Bounds are useful for other reasons as well. Some algorithms exploit only bounds, as the "branch and bound" methods of [4] for retrieval of information from a database. Other advantages we have investigated in previous work [10, 11, 12]. In addition, the mathematics of absolute bounds is straightforward and requires only elementary calculus.

Our approach is to give a variety of bounds formulae for the same estimation situation. In general, we do not know which of several bounding methods will be the best for a problem, and this suggests the program architecture of an artificial-intelligence "production system" [1]. We can combine results by taking the minimum of all the upper bounds, and the maximum of all the lower bounds.

---

[1] Even if the data is transformed before being aggregated, there are still many reasons to want statistics on the untransformed data. To use the example of [7], it is useful to study rainfall in the cube root of inches, but one may then be interested in statistics on the cube of that, the meaningful quantity of total volume.

2. Our approach

In this work we examine transformation functions whose derivatives have a constant sign in the interval of study. (We may be able to relax this restriction in particular cases, however; usually only a constant-sign second derivative is necessary. Chapter 3 of [5] discusses detailed restrictions, in particular the notion of function convexity, for the material we cover in section 3 below.) The so-called "power transformations" and their inverses [2] satisfy this constant-sign restriction for positive data values. Six common power transformations are log, antilog, square root, square, cube, and reciprocal, and these will be our primary examples. Logarithm is particularly important because the mean of the logs is the log of the geometric mean of a set of data values; reciprocal is also important because it provides the key to handling quotients of random variables. To summarize the six example transformations:

Function	first deriv.	second deriv.	steepest point
$\ln(x)$	+	-	left side
e^x	+	+	right side
$\sqrt{x}$	+	-	left side
x^2	+	+	right side
x^3	+	+	right side
$1/x$	-	+	left side

We shall assume the following statistics on the original (untransformed) data values are known:

- μ, the mean of the values (or equivalently, the sum of the values and the number of values)
- m, the minimum of the values
- M, the maximum of the values

Even when we do not know the minimum and maximum exactly, we can often assume extreme "safe" values which the minimum cannot be less than and the maximum cannot be greater than, and which we can use in our formulae. So it is reasonable to believe we can always come up with a minimum and maximum for a set of values.

In much of what follows we also assume the following is known:

- σ, the standard deviation of the values -- defined as $\Sigma_{1\leq i\leq n}(x_i-\mu)^2/n$, instead of the more conventional formula with a denominator of n-1

Note we use the symbols μ and σ to emphasize that we are consider finite data *populations*, which are not necessarily samples of anything.

We shall ignore linear transformations of variables as a preliminary to applying power functions, since

these can be handled trivially. For instance, f(x) = ln(ax+b) can be analyzed by defining y=ax+b and analyzing g(y)=ln(y), where $\mu_y = a\mu_x + b$ and $\sigma_y = a\sigma_x$.

Our basic idea is to find functions that are (a) entirely above, and (b) entirely below the curve of the function on the data-value interval. We shall consider two important cases: bounding curves that are straight lines (sections 3 and 4) and bounding curves that are second-degree polynomials (quadratics) (sections 5, 6, 7, and 8). Subsequent sections consider extensions to this framework: use of subset means and standard deviations in section 9, use of order statistics in section 10, use of distribution fits in 11, and adjustments for small populations in 12. We conclude with some simple test experiments in section 13.

3. Linear bounds on the mean

3.1. Overview

For straight lines, one curve can be a tangent to the curve at some point (for convenience, the mean); the other a secant of the curve through it at the minimum and the maximum. For curves with negative second derivative like logarithm and square root, the tangent is an upper bound, the secant a lower bound; for curves with positive second derivative like antilog and reciprocal, the tangent is the lower bound and the secant the upper. These bounding lines map directly into bounds on the mean and standard deviation, for note if $ax+b \geq f(x)$ for all x in a range, f some transformation functions satisfying our restrictions, and E denoting expected value, then

$E(ax+b) \geq E(f(x))$, or
$aE(x) + b \geq E(f(x))$, or
$a\mu + b \geq E(f(x))$

E(f(x)) being the quantity we are interested in bounding.

3.2. Linear bounds on the mean

Let us apply these ideas to the mean of transformed values (see figure 3-1). The tangent to f(x) at μ has equation

$$y = x * f'(\mu) + [f(\mu) - \mu * f'(\mu)]$$

This leads to a well-known bound (generalized in [5], p. 70):

$$\mu * f'(\mu) + [f(\mu) - \mu * f'(\mu)] = f(\mu)$$

On the other side of the curve, the secant through the maximum and minimum forms a bound. This line has equation

$$y = x * [(f(M)-f(m))/(M-m)] + [f(m) - m * [(f(M)-f(m))/(M-m)]]$$

which corresponds to the bound

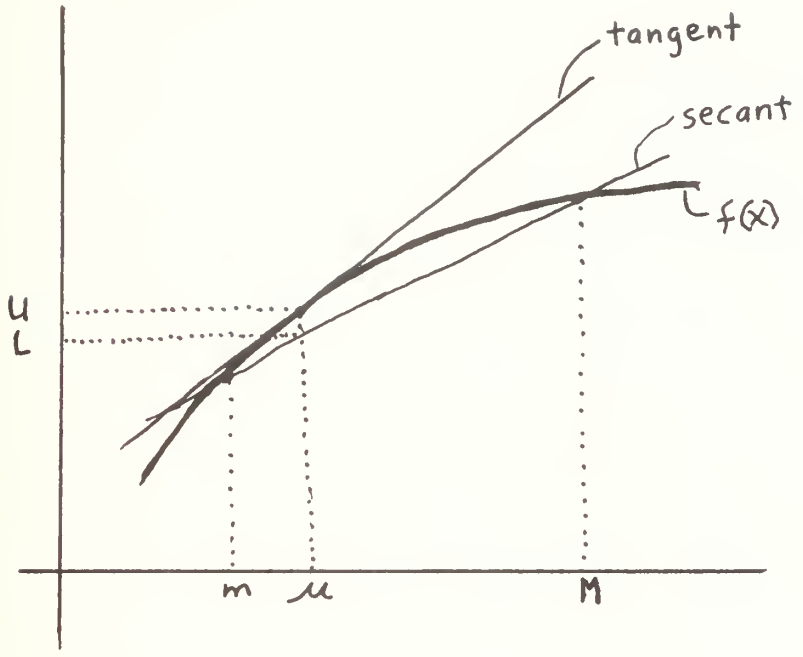


Figure 3·1: Linear bounds on the mean of transformed values

$\mu * [(f(M)-f(m))/(M-m)] + [f(m) - m * [(f(M)-f(m))/(M-m)]]$
$= (\mu-m) * [(f(M)-f(m))/(M-m)] + f(m)$
$= (1-\alpha)f(m) + \alpha f(M) = f(m) + \alpha(f(M)-f(m))$
where $\alpha = (\mu-m)/(M-m)$

To give an example, if a set of data values ranges from 10 to 100, and the mean is 23, the mean of the logarithms of the data values has

an upper bound of $\ln(23) = 3.135$
a lower bound of $77/90 \ln(10) + 13/90 \ln(100) = 2.635$

Hence the geometric mean of the original data values is between $e^{2.635} = 13.9$ and $e^{3.135} = 23$. In general from these formulae, the geometric mean is between μ and $m(M/m)^{\alpha}$; and the harmonic mean is between μ and $1/[1/m + 1/M - \mu/mM]$.

3.3. Proof that tangent at the mean is optimal

Note that the bound obtained from taking the tangent at μ is optimal for the conditions we are assuming on f. To see this, suppose we use the tangent at some other point t, i.e. the line $y = f(t) + (x-t)f'(t)$. Then the mean on this bound line is

$$E[f(t) + (x_i-t)f'(t)] = f(t) + (\mu-t)f'(t)$$

Now we want to find the maximum of this as t varies, so we take the derivative with respect to t and set it equal to zero:

$$f'(t) - f'(t) + (\mu-t)f''(t) = 0 = (\mu-t)f''(t)$$

But since we assumed that f had a constant-sign second derivative in the interval of interest, the only way this can be zero is if $\mu = t$. Hence the only extreme value for the bound will be when we take a tangent at μ -- a minimum for downwards-curving functions, and a maximum for upwards-curving.

3.4. Miscellaneous comments

In the case of a negative second derivative, the tangent bound is an upper bound, and the secant bound a lower bound; otherwise, the reverse. Note the two bounds are related, because they can be rewritten as

$f((1-\alpha)m + \alpha M)$ and
$(1-\alpha)f(m) + \alpha f(M)$, where $\alpha = (\mu-m)/(M-m)$

so they represent interchanging of a weighting and functional application.

Here is a table of the linear bounds for our six common transformations:

Function	Upper mean bound	Lower mean bound
natural log	$\ln(\mu)$	$(1-\alpha)\ln(m) + \alpha\ln(M)$
antilog	$(1-\alpha)e^m + \alpha e^M$	e^μ
square root	$\sqrt{\mu}$	$(1-\alpha)\sqrt{m} + \alpha\sqrt{M}$
square	$(1-\alpha)m^2 + \alpha M^2$	μ^2
cube	$(1-\alpha)m^3 + \alpha M^3$	μ^3
reciprocal	$\alpha/m + (1-\alpha)/M$	$1/\mu$

where $\alpha = [\mu-m]/[M-m]$

3.5. Accuracy of linear mean bounds

To illustrate effectiveness of the bounds, we tabulate the bounds for $m=10$, $M=100$, $f=\ln$, and for $\mu=19,28,37,46,55,64,73,82$, and 91. The "bounds range fraction" is the ratio of the distance between the bounds to the total range of the function on the values, the difference between f(M) and f(m); it indicates the quality of the estimate.

mean (μ)	upper bound	lower bound	bounds range fraction
19	2.944	2.533	.179
28	3.332	2.763	.247
37	3.611	2.993	.268
46	3.829	3.224	.263
55	4.007	3.454	.240
64	4.159	3.684	.206
73	4.290	3.914	.163
82	4.407	4.145	.114
91	4.511	4.374	.059

It is typical that the estimates are best for extreme μ, and the error is worst for a particular value inside the range. We can calculate this value. Assume f has negative second derivative (the other case is analogous). Then we want to find the maximum of the function representing the difference of the tangent and secant bounds, or

$$g(x) = f(\mu) - (1-\alpha)f(m) - \alpha f(M), \text{ where } \alpha = (\mu-m)/(M-m)$$

We find this by setting to zero the derivative with respect to μ, in other words

$$dg(\mu)/dx = 0 = df(\mu)/dx + f(m)/(M-m) - f(M)/(M-m)$$
$$f'(\mu) = (f(M)-f(m))/(M-m)$$

Or in other words, the maximum error occurs for any function f (that satisfies our conditions) for a mean at the point where the tangent to f is parallel to the secant through the endpoints. This makes sense because this

is the point at which f(x) stops "turning away" from the secant and begins turning back towards it. Note by Rolle's Theorem there is always one such point where the lines are parallel, and the constant sign of the second derivative ensures that there is never more than one such point.

For specific f we can tabulate the point of maximum error from this formula, as a function of m and M.

Function	Worst μ
natural log (ln x)	$(M-m)/\ln(M/m)$
antilog (e^x)	$\ln[(e^M-e^m)/(M-m)]$
square root	$(M-m)^2/4[\sqrt{M}-\sqrt{m}]^2$
square	$(M+m)/2$
cube	$\sqrt{[(m^2+mM+M^2)/3]}$
reciprocal	$\sqrt{(mM)}$

The maximum error may then be obtained as $|f(\mu) + f(M) - [(\mu_{worst}-m)(f(M)-f(m))/(M-m)]|$.

3.6. Bounds on the standard deviation, given mean

A simple application of the linear bounds on the mean of transformed values is to bounding the standard deviation of a set of values given only the maximum (M), minimum (m), and mean (μ). The variance is computed:

$$\Sigma(x-\mu)^2/n = \Sigma x_i^2/n - \mu^2$$

But since square is a continuous function with a constant-sign second derivative, we can bound the second summation, and hence the bounds on the variance are:

lower bound: $\mu^2 - \mu^2 = 0$
upper bond: $m^2 + (\mu-m)(M^2-m^2)/(M-m) - \mu^2 = \mu M + \mu m - mM - \mu^2 = (\mu-m)(M-\mu)$

And so the bounds on the standard deviation are:

lower bound: 0
upper bond: $\sqrt{[(\mu-m)(M-\mu)]}$

We will use this result frequently.

4. Linear bounds on the standard deviation

There are two methods we can use to bound the standard deviation of a set of transformed values. First, we can use the two bounds lines used previously, bound the sum of the squares, and subtract out the effect of the mean (i.e. use the formula $\Sigma x^2/n - [\Sigma x/n]^2$). Second, we can construct two new lines passing through f(x) at the mean of the transformed values.

4.1. Sum-of-squares bounds

Bound line $y = ax + b$ has second moment (sum of squares) equal to

$$E[(ax+b)^2] = E[a^2x^2 + 2abx + b^2] = a^2(\sigma^2+\mu^2) + 2ab\mu + b^2 = (a\mu+b)^2 + a^2\sigma^2$$

For our two bounds lines:

tangent: $a = f'(\mu)$, $b = [f(\mu) - \mu * f'(\mu)]$
secant: $a = (f(M)-f(m))/(M-m)$, $b = [f(m) - m * [(f(M)-f(m))/(M-m)]]$

hence the tangent bound on the sum of the squares is

$$(\sigma^2+\mu^2)[f'(\mu)^2] + 2\mu[f'(\mu)][f(\mu) - \mu * f'(\mu)] + [f(\mu) - \mu * f'(\mu)]^2$$

$$= \sigma^2[f'(\mu)]^2 + [f(\mu)]^2$$

and the secant bound is

$$\beta^2(\sigma^2+\mu^2) + 2\mu\beta[f(m) - m\beta] + [f(m) - m\beta]^2$$
where $\beta = [f(M)-f(m)]/[M-m]$

To find bounds on the variance, then, we subtract the larger of these two bounds from the square of the lower bound on the mean to get the upper bound; and subtract the smaller of these two bounds from the square of the upper bound on the mean to get the lower bound. The standard deviation then has upper bound the square root of the variance upper bound, and lower bound the square root of the variance lower bound.

To return to our previous example, suppose $f = \ln$, $m = 10$, $M = 100$, $\mu = 23$, and also suppose $\sigma = 10$. Then the bounds on the sum of squares are

tangent: $629 * (1/23)^2 + 2 * 23 * (1/23) * [\ln(23) - 23 * (1/23)]$
$+ [\ln(23) - 23 * (1/23)]^2 = 1.19 + 4.28 + 4.57 = 10.04$

secant: $\beta = \ln(100/10)/(100-10) = .02558$; hence bound is
$(.02558)^2 * 629 + 2 * 23 * .02558 * [\ln(10) - 10 * .02558] + [\ln(10) - 10 * .02558]^2$
$= .412 + 2.409 + 4.189 = 7.010$

Now since the bounds on the mean are 2.635 and 3.135 from our analysis in section 3, the bounds on the square of the mean are 6.95 and 9.82. Hence bounds on the variance are $10.04-6.95 = 3.09$ and $7.01-9.82 = -2.81$, and bounds on the standard deviation are thus $\sqrt{3.09} = 1.76$ and 0.

4.2. Special standard-deviation bounds lines

To bound the standard deviation of the transformed values we can use different bound lines than for the mean. First, let us assume we know an exact value for the mean of the transformed data values -- call it φ. Distance from φ to each transformed data value is what needs to be linearly bounded, so we use secants through $f(x)$ at φ (see figure 4-1). We assume $f(x)$ is monotonic, and hence $f^{-1}(\varphi)$ is unique, so let $f^{-1}(\varphi) = \nu$ (i.e., $\varphi = f(\nu)$). So to get an upper bound on the standard deviation of the transformed values, we use a line

below f(x) for $x<\nu$, and above for $x>\nu$; and to get a lower bound, a line above f(x) for $x<\nu$, and below for $x>\nu$. (Vice versa for a monotonically decreasing f(x).) Now since we assume f(x) has a constant-sign second derivative in the interval, the line segment from m to ν must lie constantly to one side of f(x), and similarly the line segment from ν to M. Hence choose the extensions of those two line segments into lines as our bounds lines. These lines have equations

$$y = (x-\nu)(f(\nu)-f(m))/(\nu-m) + f(\nu)$$
$$y = (x-\nu)(f(M)-f(\nu))/(M-\nu) + f(\nu)$$

Now:

$$\sigma_y^2 = E[(y-f(\nu))^2]$$

And if $y = m(x-\nu) + f(\nu)$ this is:

$$E[[m(x-\nu) + f(\nu) - f(\nu)]^2] = E[m^2(x-\nu)^2]$$
$$= m^2[\sigma^2+(\nu-\mu)^2]$$

Hence using the formula for the variance, the second moment about the mean, the variance of the transformed values is bounded by

$$[\sigma^2 + (\nu-\mu)^2]\,[(\nu-f(m))/(\mu-m)]^2 \text{ and}$$
and
$$[\sigma^2 + (\nu-\mu)^2]\,[(f(M)-\nu)/(M-\nu)]^2$$

Hence the standard deviation is bounded by

$$\sqrt{[\sigma^2+(\nu-\mu)^2]}\,[(f(\nu)-f(M))/(\nu-M)] \text{ and}$$
$$\sqrt{[\sigma^2+(\nu-\mu)^2]}\,[(f(\nu)-f(m))/(\nu-m)]$$

They are upper and lower bounds respectively for curves with positive second derivative, and vice versa for negative second derivative. Hence the bounds are just an "adjusted" standard deviation of the original values times the slopes of the lines from the mean of the transformed values to the minimum and maximum on the interval.

Note since

$\sigma f'(\nu)$ is between $\sigma[(f(\mu)-f(m))/(\mu-m)]$
and $\sigma[(f(M)-f(\mu))/(M-\mu)]$, for $f''(x)$ constant-sign

a rough *approximation* of the standard deviation of the transformed values (as opposed to *bound*) may always be obtained from $\sigma f'(\nu)$, and this will be increasingly good an approximation as σ gets smaller. Also note that for a narrow range of mean bounds, the difference between our standard deviation bounds is a rough approximation of the second derivative of f at ν:

$$\approx \sigma[(f(M)-f(\nu))/(M-\nu)] - \sigma[(f(\nu)-f(m))/(\nu-m)] \approx 2\sigma f''(\nu)$$

So the width of the bounds varies proportionately with the magnitude of the second derivative at the mean of the transformed values.

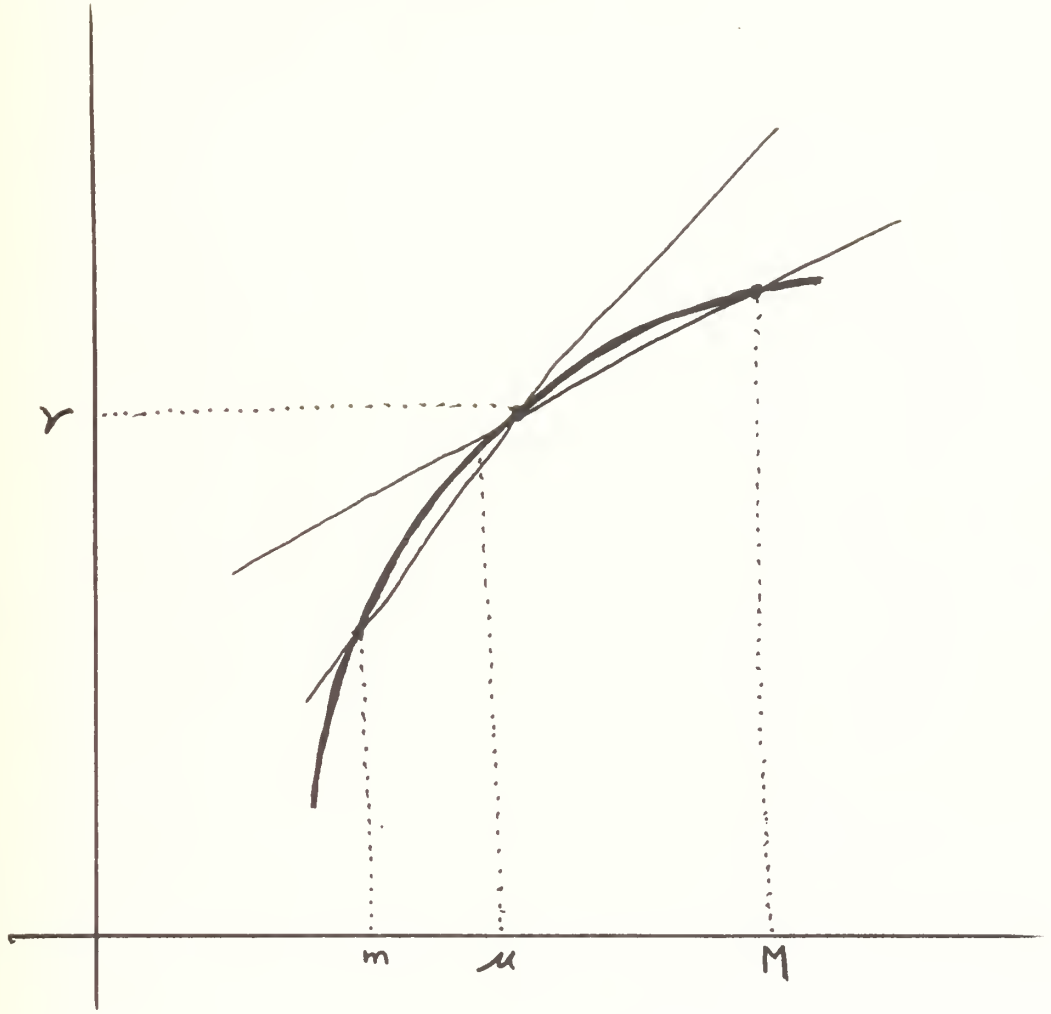


Figure 4-1: Linear bounds on the standard deviation of transformed values

4.3. Handling inexact transform means

But this assumes we know ν, the mean of the transformed values, exactly. We do for the square function, for instance. Otherwise there is an adjustment we can make. Let the upper and lower bounds on the value ν which maps to the transform mean be ν_L and ν_U. Then the bounds on the variance of the transformed values are

$$\max[\max_{\nu_L \le \nu \le \nu_U} [\sigma^2+(\mu-\nu)^2][(f(\nu)-f(m))/(\nu-m)]^2,$$

$$\max_{\nu_L \le \nu \le \nu_U} [\sigma^2+(\mu-\nu)^2][(f(\nu)-f(M))/(\nu-M)]^2]$$

$$\text{and } \min[\min_{\nu_L \le \nu \le \nu_U} [\sigma^2+(\mu-\nu)^2][(f(\nu)-f(m))/(\nu-m)]^2,$$

$$\min_{\nu_L \le \nu \le \nu_U} [\sigma^2+(\mu-\nu)^2][(f(\nu)-f(M))/(\nu-M)]^2]$$

Since $\max(\max(g(x)*s(x)),\max(h(x)*s(x))) = \max(\max(g(x)*s(x),h(x)*s(x))) = \max(\max(g(x),h(x))*s(x))$, we can simplify:

$$\max_{\nu_L \le \nu \le \nu_U} [\max[(f(\nu)-f(M))/(\nu-M)]^2,(f(\nu)-f(m))/(\nu-m)]^2]*[\sigma^2+(\mu-\nu)^2]]$$

$$\text{and } \min_{\nu_L \le \nu \le \nu_U} [\min[(f(\nu)\ f(M))/(\nu-M)]^2,(f(\nu)-f(m))/(\nu-m)]^2]*[\sigma^2+(\mu-\nu)^2]]$$

First, suppose f(x) is monotonically increasing (like all of our six important functions except 1/x). If the second derivative is positive, then the inner max is the first subexpression in the first bound above, and the inner min is the second subexpression in the second bound. We can then rewrite the formulae:

$$\max_{\nu_L \le \nu \le \nu_U} [(f(\nu)-f(M))/(\nu-M)]^2*[\sigma^2+(\mu-\nu)^2]$$

$$\text{and } \min_{\nu_L \le \nu \le \nu_U} [(f(\nu)-f(m))/(\nu-m)]^2*[\sigma^2+(\mu-\nu)^2]$$

Note that these represent the product of two functions which are both monotonically increasing with respect to ν. For a monotonically increasing f(x), μ is a lower bound on ν. The product of two monotonically increasing functions is a monotonically increasing function. The max of a monotonically increasing function is the value at the rightmost point, and the min is at the leftmost point. So the revised bounds on the variance of the transformed values, given f(x) increasing and with positive second derivative, are

upper: $[(f(\nu_U)-f(M))/(\nu_U-M)]^2*[\sigma^2+(\mu-\nu_U)^2]$

lower: $[(f(\nu_L)-f(m))/(\nu_L-m)]^2*[\sigma^2+(\mu-\nu_L)^2]$

Similarly if f(x) has a negative second derivative (again, assuming the first derivative is positive), we can show by analogous reasoning that the bounds are:

upper: $[(f(\nu_L)-f(m))/(\nu_L-m)]^2*[\sigma^2+(\mu-\nu_L)^2]$

lower: $[(f(\nu_U)-f(M))/(\nu_U-M)]^2*[\sigma^2+(\mu-\nu_U)^2]$

Using our example of f=ln, m = 10, M = 100, μ = 23, σ = 10, we use the previously found linear bounds on

the mean of the logarithms of $\nu_U = e^{3.135} = 23$ and $\nu_L = e^{2.635} = 13.9$. Hence bounds on the standard deviation of the logarithms are:

$\sqrt{[10^2+9.1^2][(2.635-\ln(10))/(13.9-10)]} = 1.16$
$\sqrt{[10^2+0^2][(\ln(100)-3.135)/(100-23)]} = .196$

both being better than the sum-of-squares bounds in section 4.1.

Unfortunately, revised formulae for monotonically decreasing functions are not as easy. The partial derivative of the bounds expressions must be set to zero and inverted. Consider the case for the upper bound for a curve with a negative second derivative (like 1/x):

$0 = \partial/\partial\nu[[(f(\nu)-f(M))/(\nu-M)]^2*[\sigma^2+(\mu-\nu)^2]]$
$0 = 2[(f(\nu)-f(M))/(\nu-M)] * [(f'(\nu)(\nu-M) - (f(\nu)-f(M)) / (\nu-M)^2] * [\sigma^2+(\mu-\nu)^2]]$
$\quad + [(f(\nu)-f(M))/(\nu-M)]^2 * -2(\mu-\nu)$
$[f'(\nu)(\nu-M) - (f(\nu)-f(M)) / (\nu-M)^2] * [\sigma^2+(\mu-\nu)^2]] = [(f(\nu)-f(M))/(\nu-M)] * (\mu-\nu)$

which is then solved for ν, and the value substituted in the function differentiated above to obtain the bound. Analogously, the other bound is found by solving

$[f'(\nu)(\nu-m) - (f(\nu)-f(m)) / (\nu-m)^2] * [\sigma^2+(\mu-\nu)^2]]$
$\quad = [(f(\nu)-f(m))/(\nu-m)] * (\mu-\nu)$

4.4. Evaluating standard-deviation bounds

The sum-of-squares bounds of section 4.1 are hard to evaluate, but we can examine the slope-based bounds of the last section, provided we assume ν is known exactly. We are interested in knowing the largest possible difference between the upper and lower bounds for an exact ν, or the maximum of

$D(\nu) = \sigma_2[[(f(M)-f(\nu))/(M-\nu)] - [(f(\nu)-f(m))/(\nu-m)]]$
where $\sigma_2^2 = \sigma^2 + (\mu-\nu)^2$

For four of our functions -- x^2, x^3, 1/x, and $\sqrt{x}$ -- this is straightforward to find:

- x^2: $D(\nu) = \sigma_2[(\nu+M)-(\nu+m)] = \sigma_2(M-m)$, so D is constant.
- x^3: $D(\nu) = \sigma_2[(\nu^2 + \nu M + M^2) - (\nu^2 + \nu m + m^2)] = \sigma_2[\nu(M-m) + (M^2-m^2)]$. This has maximum at $\nu = M$ of $\sigma_2(M-m)(2M-m)$.
- 1/x: $D(\nu) = \sigma_2[1/\nu m - 1/\nu M] = \sigma_2(1/m - 1/M)/\nu$. This has a maximum at $\nu = m$ of $\sigma_2(M-m)/m^2M$.
- $\sqrt{x}$: $D(\nu) = \sigma_2[(1/(\sqrt{\nu} + \sqrt{M})) - (1/(\sqrt{\nu} + \sqrt{m}))] = \sigma_2(\sqrt{M}-\sqrt{m})/(\nu + (\sqrt{M}-\sqrt{m})\sqrt{\nu} + \sqrt{(mM)})$. This has a maximum at $\nu = m$ of $\sigma_2(1/\sqrt{m} - 1/\sqrt{M})$.

For transcendental functions like ln(x) and e^x we can attack the problem with an infinite series obtained from the Taylor series expansion of the function about ν; when the curve is relatively flat in the interval of interest, the approximation will be good.

$$D(\nu) = \sigma_2[((f(\nu)-f(M)))/(\nu-M) - ((f(\nu)-f(m))/(\nu-m))]$$

Let us expand the first quotient in the brackets into a series.

$$\begin{aligned}(f(\nu)-f(M))/(\nu-M) &= [f(\nu) - [f(\nu) + (\nu-M)f'(\nu) + (\nu-M)^2f''(\nu)/2! + \ldots]] / (\nu-M)\\ &= -[f'(\nu) + (\nu-M)f''(\nu)/2! + (\nu-M)^2f'''(\nu)/3! + \ldots]\\ &= -\Sigma_{i=1 \text{ to } \infty}[(\nu-M)^{i-1}f^i(\nu)/i!]\end{aligned}$$

Hence

$$D(\nu) = \sigma\Sigma_{i=1 \text{ to } \infty}[[(\nu-m)^{i-1}f^i(\nu)/i!] - [(\nu-M)^{i-1}f^i(\nu)/i!]]$$

We need to take the derivative with respect to ν of this in order to see if it has a maximum in the interval. The condition for the maximum is thus:

$$0 = \Sigma_{i=1 \text{ to } \infty}[[(\nu-m)^{i-1}-(\nu-M)^{i-1}]f^{i+1}(\nu)/(i+1)*(i-1)!)]$$

To approximate this we can take the first few terms:

$$\begin{aligned}0 &= (M-m)f''(\nu)/2! + (2\nu(M-m) - (M+m)(M-m))f'''(\nu)/3!\\ 0 &= (M-m)[f''(\nu)/2 + (2\nu-m-M)f'''(\nu)/6]\end{aligned}$$

As an example, consider $f(x) = e^x$. Then:

$$0 = (M-m)[e^\nu/2 + (2\nu-m-M)e^\nu/6] = (M-m)e^\nu(1/2 + \nu/3 - m/6 - M/6)$$

which can be solved iteratively for ν.

5. Quadratic bounds on means: Taylor-series methods

5.1. The problem

A straight line is not a very good approximation to a function with a strong curvature. An obvious next step to improve our estimates of the mean is to construct quadratic bounds lines of the form $y = ax^2+bx+c$ and compute the mean along those:

$$E[ax^2+bx+c] = a(\sigma^2+\mu^2) + b\mu + c$$

However, finding quadratic bounds curves is not as easy as it might seem. We generally cannot just use the Taylor series about some point of the curve, as with the estimates (not bounds) of [9], because while such approximations may stay close to the curve of the actual function on some range, they may be above and below it at different places. For instance, take the 3-term Taylor series for $f(x) = \ln(x)$ about $x = 1$, which is

$$0 + (x-1)*(1/1) + (x-1)^2*(-1/1^2)/2 = -.5x^2 + 2x - 1.5$$

At $x=2$ this is .5, below the logarithm curve value $\ln(2) = .69$, but at $x=.5$ this is -.625, above the logarithm curve value $\ln(.5) = -.69$. Hence the approximation curve crosses $\ln(x)$, and cannot be used as a bound on the values of the latter.

5.2. Quadratic bounding by vertical shifting

There is a way we can use arbitrary polynomial approximations to get bounds: we can shift the approximation curve upwards or downwards until it no longer crosses the target curve in the interval. To put this formally for the Taylor series, we want to bound f(x) on the interval m to M by the function

$$h(x) = f(t) + (x-t)f'(t) + .5f''(t)(x-t)^2 + K$$

where t is some arbitrary point in the interval, and K is some constant. If we choose $t=\mu$ (for quadratic bounds a convenient, but not necessarily best-bound point), then the mean of the approximation function is

$$E[h(x)] = f(\mu) + (\mu-\mu)f'(t) + .5[\sigma^2 + \mu^2 - 2\mu^2 + \mu^2]\, f''(t) + K$$
$$= f(\mu) + .5\sigma^2 f''(t) + K$$

If we do not choose $t=\mu$ the formula is slightly more complicated:

$$f(t) + (\mu-t)f'(t) + .5(\sigma^2+(\mu-t)^2)f''(t) + K$$

Note for the particular function $f(x)=x^2$ the Taylor series has only three terms, and hence an exact formula for the mean of the square of a set of data values is

$$\mu^2 + .5\sigma^2(2) = \mu^2 + \sigma^2$$

The lower and upper bounds are then found from substituting K_U and K_L, which are respectively the maximum and minimum values in the interval of study of the error of the approximation e(x), defined as

$$e(x) = f(x) - f(t) - (x-t))f'(t) - .5(x-t)^2 f''(t)$$

Since the interval is finite, we cannot just find the zeros of the derivative of e(x). Zeros have to lie within the data-value interval, and they must be compared to two other points, the function values at the maximum and minimum of the range. In other words:

K_U is max[e(m),e(M), $e(z_1)$, $e(z_2)$, ...]
K_L is min[e(m),e(M), $e(z_1)$, $e(z_2)$, ...]

where the z_i are all zeros of e'(x) within the interval. To find the zeros:

$$\partial e/\partial x = f'(x) - f'(t) - (x-t)f''(t) = 0$$
$$[f'(x)-f'(t)]/(x-t) = f''(t)$$

We always know one solution of the above equation, x = t, because

$$[f'(t) - f'(t)] = (t-t)\, f''(t) = 0$$

But there are no other solutions for functions with constant-sign derivatives, implying no other local maxima or minima for a Taylor-series approximation. To see this, note the equation says the slope of f'(x) from t to some other point must be equal to the derivative of f'(x) at t. But this cannot occur if the second derivative of f'(x) (i.e., f''(x)) is constant in sign, because then each value of the first derivative (i.e., f''(x)) can occur at most once.

Hence we can write the Taylor-series quadratic bound in general as (noting $e(\mu)=0$):

upper bound: $f(t) + (\mu-t)f'(t) + .5(\sigma^2+(\mu-t)^2)f''(t) + \max(e(m),e(M),0)$
lower bound: $f(t) + (\mu-t)f'(t) + .5(\sigma^2+(\mu-t)^2)f''(t) + \min(e(m),e(M),0)$

For particular functions f we may be able to rule out some possibilities for the min and max. For instance, for $f(x)=x^3$, e(x) is just the fourth Taylor-series term, $(x-t)^3*6/6$, so $e(M) > 0$ and $e(m) < 0$, and bounds are

upper bound: $t^3 + (\mu-t)*3t^2 + .5(\sigma^2+(\mu-t)^2)*6t + (M-t)^3 = 3t^2(M-\mu) + 3t[\sigma^2+\mu^2-M^2] + M^3$
lower bound: $t^3 + (\mu-t)*3t^2 + .5(\sigma^2+(\mu-t)^2)*6t + (m-t)^3 = 3t^2(m-\mu) + 3t[\sigma^2+\mu^2-m^2] + m^3$

Similarly, $e(m)<0$ from analyzing the Taylor series for logarithm and square root; $0<e(m)$ for reciprocal; and $0<e(M)$ for antilog.

5.3. An example

To illustrate, use our previous example of f=ln, m=10, M=100, $t=\mu=23$, and $\sigma=10$. Take the Taylor series about μ. From the preceding we know that the only possible extremes occur at m, M, and μ, so note:

$e(x) = \ln(x) - [\ln(23) + (x-23)/23 - .5(x-23)^2/23^2]$
$e(m) = \ln(10) - [3.14 - .56 - .16] = 2.30 - 2.42 = -.12 = K_L$
$e(\mu) = \ln(23) - \ln(23) = 0$
$e(M) = \ln(100) - [3.14 + 3.35 - 5.6] = 4.6 - 0.9 = 3.7 = K_U$

Which are the bounds offsets we have to add to the estimate of the mean of

$\ln(23) - .5\ 10^2/23^2 = 3.06$

So we estimate the mean of the logarithms is 3.06, with an upper bound of $3.06 + \max(-.12,0,3.7) = 6.76$, and a lower bound of $3.06 + \min(-.12,0,3.7) = 2.94$. The upper bound is much worse than the linear upper bound (3.135), but the lower bound is better than the linear lower bound (2.635).

5.4. Choosing the optimal point for the Taylor series

The question arises as to the best value of t for getting an upper or lower bound. Analysis requires careful preconditions, but we can often do something like this. Suppose that e(M) is the maximum value of e(x) on the interval of study. The estimate of the transformed mean from taking the Taylor series about t is

$f(t) + (\mu-t)f'(t) + .5[\sigma^2+(\mu-t)^2]f''(t)$
$= f(t) + (\mu-t)f'(t) + .5[\sigma^2+(\mu-t)^2]f''(t)] + [f(M) - f(t) - (M-t)f'(t) - .5(M-t)^2f''(t)]$
$= f(M) + (\mu-M)f'(t) + .5[\sigma^2+\mu^2-M^2-2\mu t+2Mt)]f''(t)]$

We want to minimize this maximum error with respect to t, i.e. we want:

$0 = \partial/\partial t\ [f(M) + (\mu-M)f'(t) + .5[\sigma^2+\mu^2-M^2-2\mu t+2Mt)]f''(t)]$
$0 = (\mu-M)f''(t) + .5[\sigma^2+\mu^2-M^2-2\mu t+2Mt)]f'''(t)] + (M-\mu)f''(t)$
$0 = .5[\sigma^2+\mu^2-M^2-2\mu t+2Mt)]f'''(t)$

For a function with derivatives constant in sign, this can only be zero if the expression in brackets is zero:

$0 = \sigma^2+\mu^2-M^2-2\mu t+2Mt$

$t = [\sigma^2+\mu^2-M^2]/2(\mu-M)$
$t = [\mu + M - \delta_M] / 2$, where $\delta_M = \sigma^2/(M-\mu)$

Hence substituting back in the expression for the bound, the second derivative term must disappear, and we get

$f(M) + (\mu-M)f'((\mu+M-\delta_M)/2)$

which is an upper bound provided $e(M)>0$ and $e(M)>e(m)$.

By similar analysis we can show that

$t = [\mu + m + \delta_m] / 2$, where $\delta_m = \sigma^2/(\mu-m)$

is the best t for obtaining the other bound on the $e(x)$ on the interval of interest, leading to a lower bound of

$f(m) + (\mu-m)f'((\mu+m+\delta_m)/2)$

provided $e(m)<0$ and $e(m)<e(M)$. For $\sigma=0$ the upper and lower bounds occur at $t=(\mu+M)/2$ and $t=(\mu+m)/2$ respectively; and for σ the maximum, $\sqrt{[(M-\mu)(\mu-m)]}$, these are both $(M-m)/2$.

So for the logarithm function (where $e(m)<0$ necessarily) $\mu=23$, $m=10$, and $M=100$, and this gives for a lower bound for $t = (23 + 10 + .5*100/(23-10))/2 = 20.3$, and the bound is

$f(m) + (\mu-m)f'(20.3) = \ln(10) + 13/20.3 = 2.30 + .640 = 2.94$

which is negligibly better than for the series about μ, but may represent an improvement in other cases. In general, the Taylor series approach works well for narrow intervals of interest or intervals where $f(x)$ is rather flat. We can, however, use order statistics to improve Taylor-series bounds; see section 10.

6. Quadratic bounds on means from Lagrange interpolation

Taylor series approximations deteriorate on the edges of an approximation interval. We are more concerned with signed maximum deviation of the approximation from the function (a concept distinct from the L_∞ approximation, which minimizes the absolute value of deviations), and a better quadratic for our purposes comes from Lagrange interpolation method using the Chebyshev interpolation points. For a quadratic we need three points to fit the curve through, giving:

$h(x) = f(p)(x-q)(x-r)/(p-q)(p-r) + f(q)(x-p)(x-r)/(q-p)(q-r) + f(r)(x-p)(x-q)/(r-p)(r-q)$
$h(x) = (8/3(M-m)^2)[f(p)(x-q)(x-r) - 2f(q)(x-p)(x-r) + f(r)(x-p)(x-q)]$
where $p = m + (.5-\sqrt{3}/4)(M-m)$, $q = (M+m)/2$, and $r = m + (.5+\sqrt{3}/4)(M-m)$

Using our example of $f=\ln$, $m=10$, $M=100$, $\mu=23$, and $\sigma=10$, we have:

$p=16.029$, $q=55.0$, $r=93.971$; $\ln(p)=2.7744$, $\ln(q)=4.0073$, $\ln(r)=4.5430$
$h(x) = -.0002295x^2 + .04794x + 2.0648$

Hence an estimate of the mean of the logarithms for this example is

$-.0002295(10^2+23^2) + .04794(23) + 2.0648 = -.1444 + 1.1026 + 2.0648 = 3.0230$

This is an estimate, not a bound. Just as with Taylor-series polynomials, we can get bounds from this from knowing the extrema (maxima and minima) of the error curve on the interval of interest. For Chebyshev (as opposed to Taylor-series) approximations there are two places in the interval where $e(x)=0$, and hence one local maximum and one local minimum. We can find these by solving the error curve derivative explicitly;, for logarithm and cube this is a quadratic equation, for square root and reciprocal a cubic, and for exponential a transcendental equation. For example, for our ln(x) example:

$d/dx[\ln(x) - (-.0002295x^2 + .04794x + 2.0648)] = 1/x + .000459x - .04794 = 0$
hence $.000459x^2 - .04794x + 1 = 0$
and $x = [.4794 \pm \sqrt{(.04794^2 - .001836)}] / .000918 = 28.80$ and 75.64

So the extrema of e(x) on the interval can occur at only four points: $m=10$, $M=100$, 28.80, and 75.64. Computing e(x) there:

$e(10) = -.2187$, $e(100) = .04137$, $e(28.80) = .10526$, $e(75.64) = -.05193$

And hence the Lagrange-Chebyshev quadratic bounds on the mean of the transformed values are:

upper bound: $3.0230 + \max(-.2187, .04137, .10526, -.05193) = 3.1283$
lower bound: $3.0230 + \min(-.2187, .04137, .10526, -.05193) = 2.8043$

which are better than the linear bounds of 3.135 and 2.635 (and hence the Taylor series bounds too).

7. Quadratic bounds on means: one-sided methods

There are quadratic methods that avoid having to find the extrema of the error function in computing an approximation, by constructing approximation curves entirely above or entirely below the target function in the interval. We can do this if we can position the points of intersection of the approximation curve $ax^2 + bx + c$ with f(x) to lie either (a) outside the interval, or (b) tangent at some point. Among our six demonstration functions, reciprocal and cube lead to cubic polynomial equations.

7.1. Intersection and tangent positioning: reciprocal

Consider reciprocal first. The error curve is

$e(x) = 1/x - ax^2 - bx - c$

and it can have at most three zeros which are the solutions to

$0 = ax^3 + bx^2 + cx - 1$

To keep the approximation curve "close", we can put a point of tangency at some t inside the interval -- i.e., a double zero at t -- and another zero at M. We can write this function as $e(x) = (x/t - 1)^2(x/M - 1)$, which approaches $-\infty$ for small x, $+\infty$ for large x, reaches a local maximum at $x=t$, a local minimum at some larger x value, and then crosses zero permanently at $x=M$. Then we want

$(x/t - 1)(x/t - 1)(x/M - 1) = ax^3 + bx^2 + cx - 1$
$x^3/t^2M - x^2(2/tM + 1/t^2) + x(2/t + 1/M) - 1 = ax^3 + bx^2 + cx - 1$
$a = 1/t^2M$, $b = -(2/tM + 1/t^2)$, $c = 2/t + 1/M$

So the quadratic lower bound on the mean is

$$(\sigma^2+\mu^2)/t^2M - (2/tM + 1/t^2)\mu + 2/t + 1/M$$

We are interested in the best lower bound possible, i.e. the largest. We can find this by setting to zero the partial derivative of the preceding with respect to t:

$$\begin{aligned} 0 &= -2(\sigma^2+\mu^2)/t^3M + (2/t^2M + 2/t^3)\mu - 2/t^2 \\ 0 &= -(\sigma^2+\mu^2)/M + (t/M + 1)\mu - t \\ t &= [\mu-(\sigma^2+\mu^2)/M]/(1 - \mu/M) \\ &= \mu - \sigma^2/(M-\mu) = \mu - \delta_M \text{ where } \delta_M = \sigma^2/(M-\mu) \end{aligned}$$

So for $\sigma=0$ this is μ; for σ a maximum, namely $\sqrt{[(M-\mu)(\mu-m)]}$ (see section 3.6), this is m. We saw this δ_M term before in a different kind of quadratic approximation in section 5.4.

Substituting this t in the bound formula, we get a quadratic lower bound of

$$\begin{aligned} &[(\sigma^2+\mu^2-M\mu) + 2(\mu-\delta_M)(M-\mu) + (\mu-\delta_M)^2] / M(\mu-\delta_M)^2 \\ &= 1/M + [(\sigma^2+\mu^2-M\mu) + 2(M\mu-\mu^2-\sigma^2)] / M[(\mu - \sigma^2/(M-\mu))^2] \\ &= 1/M + [M\mu - \sigma^2 - \mu^2] / M[(M\mu - \sigma^2 - \mu^2) / (M - \mu]^2 \\ &= 1/M + [(M - \mu)^2 / M(M\mu - \sigma^2 - \mu^2)] \\ &= (1/M) [[M\mu - \sigma^2 - \mu^2 + M^2 - 2M\mu + \mu^2] / (M\mu - \sigma^2 - \mu^2)] \\ &= (1/M) [M^2 - M\mu - \sigma^2] / [M\mu - \mu^2 - \sigma^2] \\ &= (1/M) (M - \delta_M) / (\mu - \delta_M) \end{aligned}$$

Note that when $\sigma = 0$ this is equal to $1/M * M / \mu = 1/\mu$, the linear bound. Since $\mu \leq M$, a nonzero σ will cause the denominator of the fraction to decrease proportionately more than the denominator, and hence give a lower bound greater (better) than the linear lower bound. The maximum value of σ is $\sqrt{[(M-\mu)(\mu-m)]}$, whereupon $\delta_M = \mu-m$, and the lower bound is $1/M * [M - \mu + m] / m = 1/m + 1/M - \mu/mM$, exactly the upper linear bound for reciprocal (see section 3.2).

Again, let's use our standard example of $m=10$, $M=100$, $\mu=23$, $\sigma=10$, this time for the reciprocal function. Then

$$\delta_M = 10^2/(100-23) = 1.299$$

And a lower bound on the mean of the reciprocals is

$$1/100 * (100 - 1.299) / (23 - 1.299) = .04548$$

This is better than the linear lower bound, calculated as $1/\mu = .0435$.

We can get an upper quadratic bound by only minor modifications: just create a bounding curve that crosses 1/x at m instead of M, and is tangent at t in the interval. We just substitute m for M in the preceding formulae, giving

an upper bound of $(\sigma^2+\mu^2)/t^2m - (2/tm + 1/t^2)\mu + 2/t + 1/m$

taken at $t = \mu + \sigma^2/(\mu\text{-}m) = \mu + \delta_m$
which can be written as $(1/m)(m + \delta_m) / (\mu + \delta_m)$ where $\delta_m = \sigma^2/(\mu\text{-}m)$

So for our example data, $t = 23 + 10^2/(23\text{-}10) = 30.69$, and the upper bound is $1/10 - 13/10*30.69 = .0576$. This is significantly better than the linear upper bound of $(77/90)*.1 + (13/90)*.01 = .0871$. Hence by using a quadratic rather than linear bound we have narrowed the range of the answer by a factor of $(.0576\text{-}.0455)/(.0871\text{-}.0435) = .278$.

7.2. Evaluation of the quadratic reciprocal bounds

We can obtain useful approximations of the quadratic bounds by replacing the quotient with the first few terms of its binomial expansion, as here for the lower bound:

$$(\mu - \delta_M)^{-1} \approx 1/\mu + \delta_M/\mu^2 + \delta_M^2/\mu^3$$
$$\text{hence } 1/M\,(M - \delta_M)(\mu - \delta_M)^{-1} \approx 1/\mu + (1/\mu^2 - 1/M\mu)\delta_M + (1/\mu^3 - 1/M\mu^2)\delta_M^2$$
$$= 1/\mu + \delta_M(1/\mu\text{-}1/M)/\mu + \delta_M^2(1/\mu\text{-}1/M)/\mu^2$$
$$= 1/\mu + \sigma^2/M\mu^2 + \sigma^4/(M\text{-}\mu)M\mu^3$$

Hence the difference between the quadratic bounds can be approximated by

$$(1/m - 1/M)\sigma^2/\mu^2 + (1/m(m\text{-}\mu) - 1/M(M\text{-}\mu))\sigma^4/\mu^3$$
$$= [(M\text{-}m)\sigma^2/\mu^2][1/mM + (m+M\text{-}\mu)\sigma^2/\mu mM(m\text{-}\mu)(M\text{-}\mu)]$$

As suggested in the previous section, the quadratic bounds are always better than the linear bounds except at the two extreme cases of σ. We can find the μ and σ for which they are least accurate. Set the partial derivative of the difference between the quadratic bounds to 0:

$$0 = \partial/\partial\mu\,[(1/m - 1/M)\sigma^2/\mu^2 + (1/m(m\text{-}\mu) - 1/M(M\text{-}\mu))\sigma^4/\mu^3]$$
$$0 = -2(1/m - 1/M)\sigma^2/\mu^3 + [1/m(m\text{-}\mu)^2 - 1/M(M\text{-}\mu)^2]\sigma^4/\mu^3$$
$$+ -3[1/m(m\text{-}\mu) - 1/M(M\text{-}\mu)]\sigma^4/\mu^4$$
$$2(1/m - 1/M) = [1/m(m\text{-}\mu)^2 - 1/M(M\text{-}\mu)^2]\sigma^2 - 3[1/m(m\text{-}\mu) - 1/M(M\text{-}\mu)]\sigma^2/\mu$$

which can be solved iteratively.

7.3. Intersection and tangent positioning: cube

We can do something similar for the cube function:

$$c(x) = x^3 - ax^2 - bx - c$$

which is a third-degree polynomial just like the one for reciprocal. So we can position one intersection point and one tangency point. This time we can write c(x) as

$$c(x) = (x\text{-}t)^2(x\text{-}M) = x^3 - ax^2 - bx - c$$

hence

$$a = 2t+M,\ b = -(t^2+2tM),\ c = t^2M$$

so an upper bound on the mean is

$$(2t+M)(\sigma^2+\mu^2) - (t^2+2tM)\mu + t^2M$$

and this is a minimum when we choose a t such that

$$2(\sigma^2+\mu^2) - (2t+2M)\mu + 2tM = 0$$
$$[(\sigma^2+\mu^2) - M\mu] / (\mu-M) = t$$
$$t = \mu - \sigma^2/(M-\mu) = \mu - \delta_M$$

Substituting this in the equation for the bound:

$$(\mu-\delta_M)^2(M-\mu) + 2(\mu-\delta_M)(\sigma^2+\mu^2-M\mu) + M(\sigma^2+\mu^2)$$
$$= \mu^2M - \mu^3 - 2\mu\delta_M M + 2\mu^2\delta_M + \delta_M{}^2M - \delta_M{}^2\mu + 2\mu\sigma^2 + 2\mu^3 -$$
$$2\mu^2M - 2\delta_M\sigma^2 - 2\delta_M\mu^2 + 2\delta_M\mu M + \sigma^2M + \mu^2M$$
$$= \mu^3 + \delta_M{}^2(M-\mu) + (2\mu+M-2\delta_M)\sigma^2$$
$$= \mu^3 + \sigma^4/(M-\mu) + (2\mu + M - 2\sigma^2/(M-\mu))\sigma^2$$
$$= \mu^3 - \sigma^4/(M-\mu) + (2\mu+M)\sigma^2$$
$$= \mu^3 + (2\mu + M - \delta_M)\sigma^2$$

Similarly, a lower bound is

$$(2t+m)(\sigma^2+\mu^2) - (t^2+2tm)\mu + t^2m$$

and this is a maximum when we choose a t such that

$$t = \mu + \sigma^2/(\mu-m) = \mu + \delta_m$$

leading to a lower bound of

$$\mu^3 + (2\mu + m + \delta_m)\sigma^2$$

Note the quadratic lower bound is always greater than the linear lower bound, μ^3. The difference between the upper and lower bounds is

$$[M - m - \delta_M - \delta_m]\sigma^2$$

which provides a useful criterion for the effectiveness of these bounds. Note this is always nonnegative since

$$M - m - [\delta_M + \delta_m] = M - m - \sigma^2(M-m)/(M-\mu)(\mu-m)$$
$$= (M-m)[1 - \sigma^2/(M-\mu)(\mu-m)]$$

The largest possible value of σ^2 is $(M-\mu)(\mu-m)$, so the quantity in brackets is always nonnegative.

8. Optimal quadratic bounds

The problem of finding the best quadratic approximation for our bounding purposes may be viewed as an optimization problem in two variables. Since the quadratic curve $ax^2 + bx + c$ leads to a bound of

$$\text{upper bound: } a(\sigma^2+\mu^2) + b\mu + c + \max_{m\le x\le M}[f(x)-ax^2-bx-c]$$
$$\text{lower bound: } a(\sigma^2+\mu^2) + b\mu + c + \min_{m\le x\le M}[f(x)-ax^2-bx-c]$$

and the constant c can be moved out of the maximum and minimum, we can write:

$$\text{upper bound: } a(\sigma^2+\mu^2) + b\mu + \max_{m\le x\le M}[f(x)-ax^2-bx]$$
$$\text{lower bound: } a(\sigma^2+\mu^2) + b\mu + \min_{m\le x\le M}[f(x)-ax^2-bx]$$

So we have two optimization problems for real a and b: to find the values that minimize the upper bound,

and the values the maximize the lower bound. We have constructed a program that does this by estimating the gradient from exploratory steps, finding the zeros of the error function by the quadratic formula for logarithm and cube, and by iterative bisection for antilog, square root, and reciprocal. Comparison with the other obtained bounds is presented later in this paper. Unfortunately, the extrema appear to be "broad", and convergence is slow, so the other methods discussed in this paper seem clearly desirable in most cases. While these other methods cannot usually get the tightest bounds, the difference is usually not much.

A strong local maximum found by the optimization process is guaranteed to be the global maximum over all quadratic curves, because the function being optimized is convex. To see this, note for the upper bound for instance

$$(\theta a_1+(1-\theta)a_2)(\sigma^2+\mu^2) + (\theta b_1+(1-\theta)b_2)\mu$$
$$+ \max_{m\le x\le M}[f(x)-(\theta a_1+(1-\theta)a_2)x^2-(\theta b_1+(1-\theta)b_2)x]$$

$$\le a_1(\sigma^2+\mu^2) + b_1\mu + \max_{m\le x\le M}[f(x)-a_1x^2-b_1x]$$
$$+ a_2(\sigma^2+\mu^2) + b_2\mu + \max_{m\le x\le M}[f(x)-a_2x^2-b_2x]$$

since $\max(f(x)+g(x)) \le \max(f(x)) + \max(g(x))$.

For our standard example, we found the optimal quadratic bounds to be 3.00 and 3.10.

9. Improving accuracy with outliers and statistics on subsets

We can tighten bounds if we know additional information about a set of data values. We may know a few extreme values on the range (outliers), and be able to remove these points from the analysis of the rest of the points. This helps a good deal when m and/or M are unusually unrepresentative of the distribution (and notice how frequently we have used m and M in our formulas). With the outliers removed, the remaining values can have a narrower range, on which the function can be better matched by a linear or quadratic approximation. The transformed values for the known outliers can then be added to the total mean or total variance in a final step.

But we can generalize this. We can improve accuracy of bounds any time we know means and variances of arbitrary subsets of the original data values. We may then estimate statistics on the transformed values for each subset and combine them with the appropriate weighting.

9.1. An example

For instance, from [8], there were 6133 merchant ships with United States registry in 1982, of an average gross tonnage of 3120 per ship. Of these, 2941 were fishing vessels, of average tonnage 199.6 gross tons; 548 were cargo ships, of average tonnage 9790 tons; 361 were tankers, of average tonnage 2670 tons. Hence there

were 6133 - 2941 - 548 - 361 = 2283 other ships of average tonnage [(6133*3120) - (2941*200) - (548*9740) - (361*2670)]/2283 = [19,130,000 - 588,00 - 5,340,000 - 965,000] / 2283 = 5320 tons.

Now suppose we want the mean of the logarithms of the tonnage values. Consider the upper bounds on each of the four disjoint subsets. These are just the logarithms of the means, or 5.30, 9.21, 7.88, and 8.57. Hence the total upper bound is the weighted mean of these upper bounds, or [(5.30*2941)+(9.21*548)+(7.88*361)+(8.57*2283)] / 6133 = 7.018. This should be compared with the upper bound derived from the mean of the entire set, ln(3120) = 8.03, so the subdivision data gave us a significant improvement.

Unfortunately, we do not know anything about the maximum and minimum tonnage of classes of ships, so we cannot get a cumulative lower bound. However, we know m=100 for this table, and M=200,000 is a reasonable figure from knowledge of merchant shipping, so a global lower bound is found by

α = (3120-100)/(200000-100) = .0151
lower bound is ln(100) + α(ln(200000)-ln(100)) = 4.60 + .0151*7.60
= 4.60 + .115 = 4.715

9.2. Proof of desirability of subdivision for linear bounds

It can be proved that linear bounds on the mean are never worsened by using such subset statistics. This can be seen graphically in figure 9-1. We consider here the case of binary subdivision, and further subdivisions can be covered by extension. We also consider only functions concave downwards, but the other case can be handled analogously.

First consider the lower bound. If the ranges of the subdivisions are the same as the full set, then the two lower bounds must lie along the same line, and their weighted average must lie along the line too; hence the lower bound of the full set is exactly the weighted average of the two lower bounds. If one or both of the subsets has a narrower range of values than the full set, this can only increase (improve) the lower bound since a secant across a subrange lies fully above a secant across a range containing the subrange. Hence the lower bound cannot get any worse in this subdivision summation of linear lower bounds.

The upper bound also cannot be any worse. This time range reduction within a subset does not matter because the upper bound is constrained to lie along the curve of the function, which is independent of where it is sliced. The weighted average of the two subset upper bounds is a point along the line connecting two points on the function curve. But since the function is concave downwards, this point is always below the function. But since the upper bound on the full set is constrained to lie on the curve, the subdivision process always guarantees a better upper bound as long as the two subdivision means are different, and no worse if

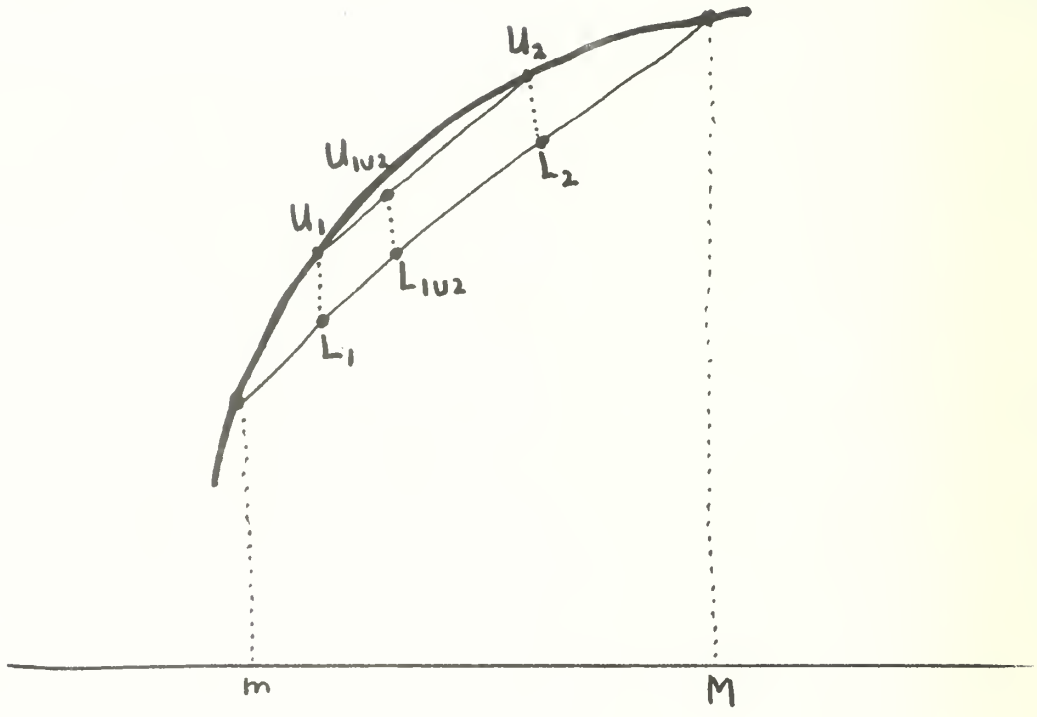


Figure 7-1: Improvements in linear bounds from combining statistics on two disjoint sets

they are not different.

10. Exploiting order statistics as well

So far we have only assumed knowledge of the maximum, minimum, mean, and (sometimes) standard deviation of sets of data values. If we have additional statistics on the data values we can do a better job of estimating statistics on the transformed values. In this section we discuss using order statistics (e.g. medians and percentiles). Order statistics have the nice property that they have one-to-one mappings from the original data values to the transformed values under the monotonic transformations we are assuming.

10.1. Using the median

First, assume we know a median in addition to the maximum, minimum, and mean. We can often get an immediate improvement in the bounds on estimates. Let the error curve (linear, quadratic, or whatever) be e(x). Then the median can be thought to partition the points into two equal-sized subranges (assume the number of points to be large enough so that even numbers of points don't bother us). Then an upper bound on the mean of the transformed values is the estimate given by the approximation curve plus one half the maximum of the error curve in the range to the left of the median plus one half the maximum of the error curve in the range to the right of the median. The lower bound on the mean is found substituting "minimum" for "maximum" in the above rule. Thus knowing the median decreases the influence of extrema of the error curve.

10.2. Other order statistics

We can generalize these ideas to the situation where we know arbitrary order statistics on the original distribution. Denote these statistics as r pairs of the form $\langle x_i, f_i \rangle$, where fraction f_i of the items in the distribution are claimed to lie to the left of value x_i. Then we can generalize the formula of section 5 as follows:

upper bound is <estimate from approximation curve> - $\Sigma_{1 \le i \le r} [f_i * \min_{x_{i-1} < x < x_i} [e(x)]$

lower bound is <estimate from approximation curve> - $\Sigma_i [f_{1 \le i \le r} * \max_{x_{i-1} < x < x_i} [e(x)]]$

where e(x) is the error curve a(x)-f(x), x_0 is defined as m, with $f_0 = 0$, and the x_r is defined as M (with corresponding f of 1). Thus the effects of the extreme points of e(x) are "diluted" by their fractional coefficients, and the more order statistics are known, the tighter the eventual bounds.

Under certain circumstances we can simplify the above formulae considerably. If we know even-subdivision order statistics (i.e., $f_i = i/r$, r the number of order statistics), and if the error curve e(x) is monotonic, then the maximum and minimum of e(x) in each subinterval between the order statistic ordinates

x_i must lie at the endpoints. So if c(x) is monotonic increasing, the upper bound is $[\Sigma_{1\le i\le m} c(x_i)]/r$ and the lower bound is $[\Sigma_{1\le i\le m} c(x_{i-1})]/r$; and vice versa if c(x) is monotonic decreasing. Hence the absolute range between the upper bound and lower bound is always the same number, $|c(x_m)-c(x_0)|/r = |c(M)-c(m)|/r$. (Note that Taylor-series quadratic approximations are monotonic if $c(m)<0<c(M)$ or $c(M)<0<c(m)$, conditions which occur frequently.)

10.3. Order statistics and the standard deviation

Order statistics are also helpful in estimating the standard deviation of the transformed values, especially order statistics for the leftmost and rightmost subranges of the interval. Recalling the bounds lines drawn through the mean of the transformed values in section 4.2, we had to draw them so they lay entirely above the curve to one side of the mean, and entirely above on the other side, and this is a highly conservative assumption. Assume ν is known precisely. We could probably get a better bound if we knew how many points lay to the left of some x_1, and the drew a secant of f(x) from the transform mean to it, rather than from the transform mean to m; or if we knew how many points lay to the right of some x_{r-1}, and drew secant from the transform mean to it instead of M. See figure 10-1.

The estimate of the standard deviation of the transformed values obtained from these lines is just their slope times the original standard deviation. But to get a bound, we need a correction for the points lying more extreme than the new point of intersection. Consider the example of curve concave downwards like logarithm, and take the upper bound line from the transform mean to some point to the left; call the point x_1, and let it be an order statistic so that fraction p of the distribution lies to the left of it. Assume the mean of the transformed values is known exactly. Then the correction for a bound corresponds to the situation where all the p points are at m, which means a difference in the variance of

$$p^*[(f(\nu)-f(m))^2 - [(\nu-m)^*(f(\nu)-f(x_1))/(\nu-x_1)]^2]$$

where ν is the number which maps functionally to the mean of the transformed values. Hence the expression for the upper bound on the standard deviation is

$$[[\sigma^2+(\mu-\nu)^2][(f(\nu)-f(x_i))/(\nu-x_i)]^2 + p^*(f(\nu)-f(m))^2 - p^*[(\nu-m)^*(f(\nu)-f(x_1))/(\nu-x_1)]^2]^{.5}$$

So using such a bounds line can give a better slope, but one pays a penalty of a correction term which subtracts from the slope improvement. An obvious question is under what conditions use of the order statistic helps. It turns out this has a surprising answer when ν is known exactly. Denote the two slopes as s_m and s_o, i.e.

$$s_m = (f(\nu)-f(m))/\nu-m),\ s_o = (f(\nu)-f(x_1))/(\nu-x_1)$$

we can rewrite our expression for the upper bound as

$$[[\sigma^2+(\mu-\nu)^2] * s_o^2 + p^*s_m^2*(\nu-m)^2 - p^*s_o^2*(\nu-m)^2]^{.5}$$

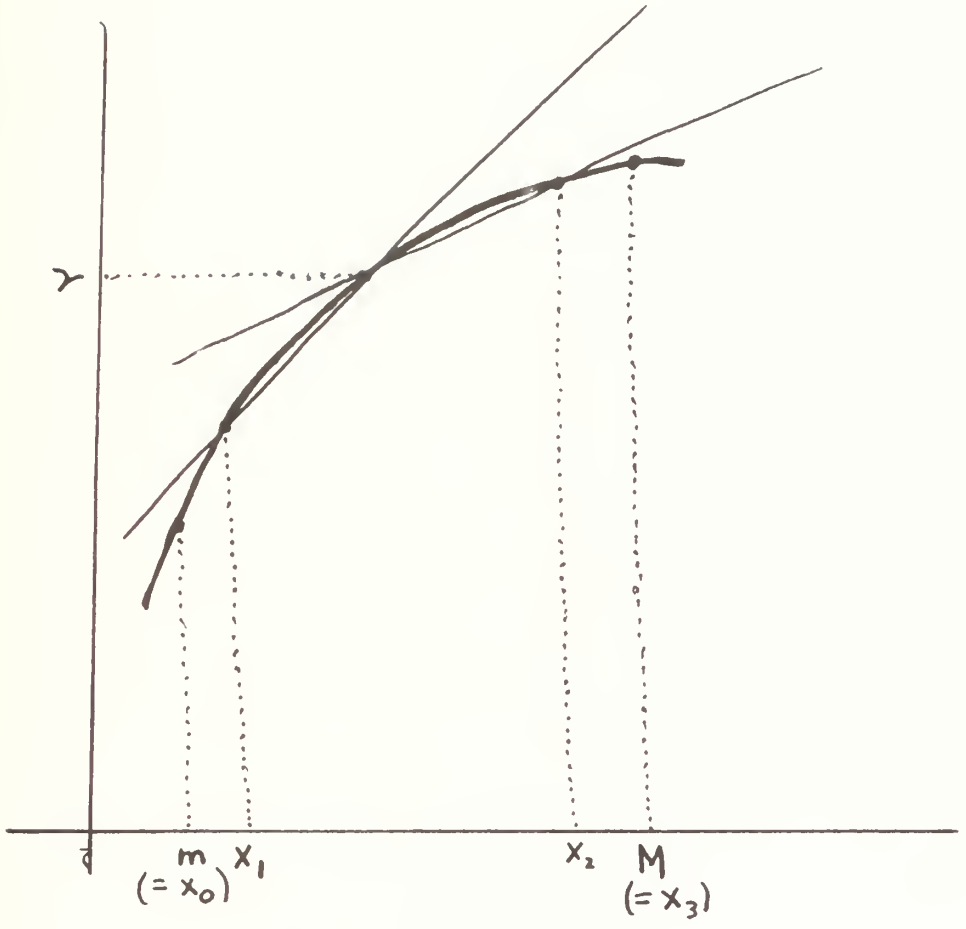


Figure 10-1: Exploiting order statistics for a better bounds on the standard deviation

This will represent an improvement on the linear upper bound $[\sigma^2+(\mu-\nu)^2]s_{in}^2$ if

$$[\sigma^2+(\mu-\nu)^2]s_m^2 > [\sigma^2+(\mu-\nu)^2] * s_o^2 + p^*s_{m_2}^2 *(\nu-m)^2 - p^*s_o^2*(\nu-m)^2$$
$$\text{or } [\sigma^2+(\mu-\nu)^2][s_m^2 - s_o^2] > p^*[s_m^2 - s_o^2]*(\nu-m)^2$$

So the slope terms cancel, and use of the order statistic $<x_1,p>$ is going to be helpful when:

$$[\sigma^2+(\mu-\nu)^2] > p * (\nu-m)^2$$
$$\text{or } p < [(\sigma^2+(\mu-\nu)^2)/(\nu-m)^2]$$

This result is independent of where the order statistic is within the distribution (x_1), and depends only on the standard deviation and minimum of the original distribution, and the mean of the transformed values. The corresponding result for the rightmost order statistic is

$$p < [(\sigma^2+(\mu-\nu)^2)/(M-\nu)^2]$$

where p is the fraction of items to the right of x_{r-1}.

If we know other order statistics than just the leftmost and rightmost (x_1 and x_{r-1}) we can get better bounds, though predicting the improvement is difficult. For instance, if we know x_2, we can take a line from ν to x_2, and estimate the contribution to the correction factor from the items between x_1 and x_2 differently than the contribution of items between m and x_1.

10.4. Adjustment of standard deviation for an inexact transform mean

If we do not know the exact mean of the transformed values, $\varphi = f(\nu)$, we must adjust these results. Let the bounds on the transform mean be ν_L and ν_U as in section 4.3. Assume f(x) has a negative second derivative. The formula for the upper bound is

$$[[\sigma^2+(\mu-\nu)^2][(f(\nu)-f(x_1))/(\nu-x_1)]^2 + p^*(f(\nu)-f(m))^2 - p^*[(\nu-m)*(f(\nu)-f(x_1))/(\nu-x_1)]^2]^{.5}$$

Since $\nu \leq \mu$, $[[\sigma^2+(\mu-\nu)^2]$ is monotonically decreasing with ν in its range. The rest of the expression is the difference of a term and the difference of two others. The first term is monotonically decreasing with increasing ν since the second derivative of the curve is negative. This represents the second moment of f_1 items grouped at m on the curve. As ν increases, the possible distance these items could be off the bound line increases, and their relative weight increases as $f(\nu)$ becomes relatively larger than f(m). Hence since this correction term is *subtracted* from the slope, the effect as ν increases will be for all the terms to decrease. Hence the adjusted value for the upper bound on the standard deviation of the transform values is just

$$[[\sigma^2+(\mu-\nu_L)^2][(f(\nu_L)-f(x_1))/(\nu_L-x_1)]^2 + p^*(f(\nu_L)-f(m))^2$$
$$- p^*[(\nu_L-m)*(f(\nu_L)-f(x_1))/(\nu_L-x_1)]^2]^{.5}$$

substituting ν_U for ν in the exact-ν formula.

Similarly, we substitute ν_L for ν to get an adjusted lower bound. Analogously, we handle curves with a positive second derivative by substituting ν_L for ν for an upper bound, ν_U for ν for a lower bound.

10.5. Quasi-order statistics from the standard deviation

If we know the mean and standard deviation of a set of data values, we can use Chebyshev's inequality to bound the number of items lying more than a certain distance from the mean. This information is like an order statistic, but since it only represents an upper bound on the number of items in a region and not an exact number of items, it must be used carefully. It can only be used for partitions of the interval of interest into two parts, the subinterval of points farther than a certain distance to the left (or right) of the mean, and a subinterval of all other points of the interval. It can also only be used for an upper bound on the mean of the transformed values, given μ, when c(x) has a maximum on the first subinterval that is more than the maximum on the second, or for a lower bound when c(x) has a minimum on the first subinterval that is less than the minimum on the second.

Actually, Chebyshev's inequality in the standard form (that only a fraction σ^2/D^2 of the points of a distribution can lie greater than distance D units from the mean) is not the best inequality we can get, since it refers to both tails of a distribution, and we are only concerned with the number of points in one tail. Only $\sigma^2/(\sigma^2+D^2)$ points can lie to the left of a point D to the left of the mean, or lie to the right of a point D to the right of the mean. To see this, note that if fraction f of the points lie to the left of a point D units to the left of the mean, then their weighted second moment about the mean is at least fD^2, which must be less than σ^2. But in order for the mean to be at the place it is, this fraction f of the points must be compensated for by (1-f) points R units to the other side of the mean. For maximal f, these other (1-f) points must all be at the same location, for otherwise they would have a nonzero variance which plus their mean would add to the variance of the whole distribution, and would require a lower maximum f. Hence we have two equations to solve simultaneously:

$$fD^2 + (1-f)R^2 = \sigma^2$$
$$fD - (1-f)R = 0$$

which imply

$$R = Df/(1-f),\ fD^2/(1-f) = \sigma^2,\ f = \sigma^2/(\sigma^2+D^2)$$

Using this result, we then can put bounds on the mean of the transformed values of

upper bound: $f(\mu) + .5\sigma^2 f''(\mu)$
$+ (\sigma^2/(\sigma^2+D^2)*\max_{m\le x\le \mu-D}(c(x))$
$+ (D^2/(\sigma^2+D^2)*\max_{\mu-D\le x\le M}(c(x))$,
provided the first max value is greater than the second

lower bound: $f(\mu) + .5\sigma^2 f''(\mu)$
$+ (\sigma^2/(\sigma^2+D^2)*\min_{m\le x\le \mu-D}(c(x))$
$+ (D^2/(\sigma^2+D^2)*\min_{\mu-D\le x\le M}(c(x))$,
provided the first min value is less than the second

These are the left-sided bounds; we can also get analogous expressions for bounds using points on the right of

a distribution. Unfortunately, we cannot find optimal values of D for these formulas because they the derivative cannot be applied.

Note that while it may be difficult to determine for an arbitrary e(x) whether the maximum in one interval is greater than in another, the Taylor-series quadratic approximation often always has this property for either the left-side or right-side rule.

10.6. Evaluation of quasi-order statistics from the standard deviation

Let us return to the analysis in section 5.3 of our standard example with the quadratic Taylor series approximation at μ. Choose as subintervals $10 \leq x \leq 33$ and $33 \leq x \leq 100$, so $D = 33-23 = 10 = \sigma$. Since the error curve is monotonically increasing ($e(m) < e(\mu) < e(M)$, and no $e'(x) = 0$ except μ) the maxima on the subintervals are at the rightmost points, and the minima at the leftmost. Hence the maxima are $e(33) = 3.50-(3.14+.435-.106) = .03$ and $e(100) = 3.7$. Similarly for the other bound, choose $D = 5$, $10 \leq x \leq 18$, and $18 \leq x \leq 100$; and the minima are $e(10) = -.12$ and $e(18) = 2.89-(3.14-.217-.023) = -.01$. The maximum fraction f for $x = 33$ is $10^2/(10^2+10^2) = .5$, and for $x = 18$ is $10^2/(10^2+5^2) = .8$. Hence the revised bounds on the mean of the transformed values are

lower: 3.06 - .5*.03 - .5*3.7 = 1.20
upper: 3.06 - .8*-.12 - .2*-.01 = 3.16

which are better than the bounds obtained in section 5.3.

D is a parameter here that can vary arbitrarily. Let us find the best value for it, for the case of a Taylor series approximation where e(x) increases with x, and a lower bound:

$$0 = \partial/\partial D\,[(\sigma^2/(\sigma^2+D^2)) * e(\mu-D) + (D^2/(\sigma^2+D^2)) * e(M)]$$
$$0 = \partial/\partial D\,[[\sigma^2 * e(\mu-D) + D^2 * e(M)] / (\sigma^2+D^2)]$$
$$\text{so } \sigma^2\partial/\partial D[f(\mu-D)-f(\mu)-Df'(\mu)-.5D^2f''(\mu)] + [2D * e(M)]$$
$$= [\sigma^2[f(\mu-D) - f(\mu) - Df'(\mu) - .5D^2f''(\mu)] + D^2e(M)] * 2D$$
$$\text{Hence } \sigma^2[-f'(\mu-D)-f'(\mu)-Df''(\mu)] + [2D * e(M)]$$
$$= [\sigma^2[f(\mu-D) - f(\mu) - Df'(\mu) - .5D^2f''(\mu)] + D^2e(M)] * 2D$$
$$\text{or } 2De(m)(1-D^2) / \sigma^2$$
$$= f'(\mu-D) + (1-2D^2)f'(\mu) + D(1-D^2)f''(\mu)) - 2Df(\mu-D) - 2Df(\mu)$$

which we can solve by iterative methods to find the best value of D.

10.7. Splines and order statistics

We have not referred to spline approximations in the preceding analysis because if an approximation curve is divided into pieces with different properties then we must know how many data points are in each to calculate means and standard deviations on the transformed values. One might think that for a given set of order statistics on a distribution we may be able to create a spline approximation broken at the points at which

the order statistics are sited, and use that for bounding. But we still need to know means of every subinterval, the knowledge discussed in section 9, which may be difficult to obtain. Thus splines may be difficult to use.

11. Using fits to known distributions

As a final kind of information which we might have about a set of values, we might know that their distribution is close to some well-known distribution, with a certain allowed tolerance. If the tolerance is small we can expect quite tight bounds on the transformed values. But estimating statistics this way requires special preparation in advance (namely, measuring fits to a predicted distribution), and is not possible with most data presented in already-aggregated units.

11.1. General formula for known distributions

A well-known result (e.g. [3], section 7.3) gives the distribution of the transform of some probability distribution p(x), under the transformation function f(x), as

$$q(y) = p(f^{-1}(y)) * |df^{-1}(y)/dy|$$

as a function of y, provided f is either monotonically increasing of decreasing in the interval.

So for instance if our p(x) approximates a uniform distribution on the interval m to M, $q(y) = (1/(M-m)) * |df^{-1}(y)/dy|$. For $f(x)=\ln(x)$, $q(y) = e^y/(M-m)$ on the interval $y=\ln(m)$ to $y=\ln(M)$; an estimate of the mean of q(y) is

$$\int yq(y)dy \,/\, \int q(y)dy = [(\ln(M)-1)M - (\ln(m)-1)m] \,/\, (M-m) = -1 + [M \ln(M) - m \ln(m)]/(M-m)$$

and an estimate of the second moment about zero is

$$\int y^2q(y)dy \,/\, \int q(y)dy = [M[\ln(M)*\ln(M) - 2\ln(M) + 2] - m[\ln(m)*\ln(m) - 2\ln(m) + 2]] \,/\, (M-m)$$

which minus the square of the estimate of the mean gives an estimate of the variance.

For p(x) uniform, $f(x)=1/x$, $q(y) = 1/y^2(M-m)$ on the interval $y=1/M$ to $y=1/m$; an estimate of the mean of q(y) is

$$[\ln(1/m)-\ln(1/M)] \,/\, (M-m) = \ln(M/m)/(M-m)$$

and an estimate of the second moment about zero is $(1/m - 1/M)/(M-m) = 1/mM$, hence an estimate of the variance is

$$1/mM - [\ln(M/m)/(M-m)]^2$$

11.2. Handling inexact fits to distributions

We have not addressed how to get bounds on means and standard deviations. We can do this by defining an "upper fit" ω_U and "lower fit" ω_L on the discrete set of n values x_i such that

$$\omega_U = \max_i [x_i - g_i], \omega_L = \min_i [x_i - g_i]$$

where $\int_{-\infty}^{g_i} p(x)dx = (i-.5)/n$, and p(x) is the distribution the x_i fit to

In other words, the fits are the maximum and minimum deviations of an x_i from its value predicted by the approximating distribution p(x).

We can exploit the assumed fact that f(x) is monotonically increasing or decreasing to say that the maximum and minimum of the mean of the transformed values occur when the x_i are all at ω_U or all at ω_L from their predicted positions, not necessarily respectively. This is because less than an extreme deviation for one point cannot improve prospects for a more extreme mean; all point deviations are independent of one another, within the tolerances. Hence to find the extreme values of the transformed mean one just calculates the means of

$$q_U(y) = p[f^{-1}(y)-\omega_U] * |df^{-1}(y)/dy| \text{ and}$$
$$q_L(y) = p[f^{-1}(y)-\omega_L] * |df^{-1}(y)/dy|$$

We can use this same approach to get bounds on the standard deviation in the manner of section 4.1. We just define a $g(x) = [f(x)]^2$ as a new transformation function, and compute the above formulae with g instead of f. We then compute bounds on the mean, square them, and subtract this interval from the interval computed on the mean of g(x).

11.3. Example of inexact distribution fit

Suppose we know the distribution of x_i fits an even distribution on the interval 10 to 100, to such an extent that a point is never further than 2 units in advance of where it would be in a perfectly even distribution, and never more than 3 units behind. Then the maximum-mean distribution is a uniform distribution from 12 to 102, and the minimum-mean distribution is a uniform distribution from 7 to 97. Suppose we want to find the mean of the logarithms of these data values. Using the formulae we obtained in section 11.1, the mean of the first distribution is [102 ln(102) - 12 ln(12) - 102 + 12] / (102-12) = (472 - 29.8)/90 - 1 = 5.02 - 1 = 4.02; and the mean of the second distribution is [97 ln(97) - 7 ln(7) - 97 + 7] / (97-7) = (443 - 13.6)/90 - 1 = 4.78 - 1 = 3.78. Hence the mean of the transformed values is between 3.78 and 4.02, corresponding to antilogs of 44 and 56. Note the mean of the original values must lie between (102 + 12)/2 = 57 and (97+7)/2 = 52.

For an estimate of the standard deviation we use the formula previously derived for an estimate of the sum of the squares, namely

$$[M[\ln(M)*\ln(M) - 2\ln(M) + 2] - m[\ln(m)*\ln(m) - 2\ln(m) + 2]] / (M-m)$$
$$= [M(\ln(M)-1)^2 - m(\ln(m)-1)^2]/(M-m) + 1$$

For the uniform distribution 12 to 102, this is

$$[102(3.62)^2 - 12(1.48)^2]/90 + 1 = (1338-26.2)/90 + 1 = 15.61$$

and for the uniform distribution 7 to 97 this is

$$[97(3.57)^2 - 7(.945)^2]/90 + 1 = (1235-6.25)/90 + 1 = 14.58$$

From the previous paragraph we know bounds on the mean of the transformed values are 3.78 and 4.02, hence bounds on the square of the mean are 14.3 and 16.2. Hence bounds on the variance are 15.61-14.3 = 1.3 and max(14.58-16.2,0) = 0. Hence bounds on the standard deviation of the transformed values are 1.14 and 0.

12. Small populations

Thusfar we have not made use of the size of the data population being analyzed. This is only significant if the population is particularly small, in which case the known maximum M and minimum m (and the median and mode too, if known) are a nonnegligible proportion of the points of the distribution. For instance, the linear bounds represent in general the two extreme cases where (a) all the points are grouped at the mean, and (b) all the points are at the maximum and the minimum. Knowledge of M and m thus decreases the distance between linear bounds by a factor of 2/n, n the size of the data population, since it represents a weighted modification of case (a) by two points from case (b).

13. Some experimental comparisons of the various bounds formulae

We have run some simple experiments of the effectiveness of our bounds formulae on the mean of the transformed values. We wrote programs in INTERLISP-VAX. We used two test functions, $f(x)=\ln(x)$ and $f(x)=1/x$. For the experiments we computed upper and lower bounds derived the following ways:

- simple linear bounds (section 3)
- Taylor-series quadratic bounds, series around the mean (section 5)
- Lagrange-Chebyshev interpolation quadratic bounds (section 6)
- For the reciprocal only, the one-sided quadratic bounds (section 7)
- Order-statistic bounds from the Chebyshev-inequality, using a Taylor series around the mean (section 10.5)
- Best quadratic bounds found by explicit optimization on quadratic coefficients a and b (section 8):

  upper bound: $a(\sigma^2+\mu^2) + b\mu + c + \max_{m\le x\le M}[f(x)-ax^2-bx-c]$
  lower bound: $a(\sigma^2+\mu^2) + b\mu + c + \min_{m\le x\le M}[f(x)-ax^2-bx-c]$

We discovered that our results for optimal bounds for the reciprocal curve were identical (except for roundoff error) to those for one-sided bounds, so we have omitted the former from the reciprocal table. Unfortunately, we have been unable to prove the connection (that is, that the one-sided bounds are indeed the optimal ones), though we strongly suspect it.

Results are contained in figures 13-1 and 13-2. Since the closed-form expressions are simple computations, in a computer implementation it is advisable to try all the different bounds methods, and take the minimum of the upper bounds to get a cumulative upper bound, and the maximum of the lower bounds to get a cumulative lower bound.

14. Application to correlated data

An application of these ideas is to estimation of statistics of one attribute from those of another if the attributes are known to have a nonlinear correlation describable by a monotonic function such as we have been analyzing. We can then bound statistics on one attribute from statistics on the other.

15. Direct optimization

We should note there is another kind of optimization that can be applied to problems of this sort. We can make the optimization variables the values themselves of an unknown distribution and perform a constrained optimization with objective function the statistic on which bounds are desired, and with constraints the values of known other statistics. Conceptually, this is a nice approach since it can be applied to arbitrary states of prior knowledge and can bound arbitrary statistics.

We have done a number of experiments which we do not have the space here to discuss, and the idea seems to work. However, we have found that this "direct optimization" is highly sensitive to optimization methods, starting points, and step sizes, and is surprisingly difficult to get convergence for; unlike quadratic optimization, the function optimized is not usually convex. But there is an even more serious problem with direct optimization, a very fundamental one: it only gives lower bounds on upper bounds, and upper bounds on lower bounds, unlike all the other bounds discused in this paper which are upper bounds on upper bounds, and lower bounds on lower bounds. For instance, for our standard example we found a lower bound on the upper bound of 3.09771 on the mean of the logarithms from direct optimization, but we have no idea how much larger a bound is possible up to the quadratic-optimization bound of 3.10383 which represents an absolute limit. Thus the utility of direct optimization is questionable in bopunded statistical estimation,, and we do not see it as a challenge to the methods developed in this paper. (It does provide a useful tool for debugging the methods, however, since for instance any supposed bound we find less than the upper bound on the lower bound is in error.)

m	M	μ	σ	linear	Taylor	LaGrange	quad.opt.	order-stat.
10	20	15	0.1	2.649	2.691	2.704	2.705	2.7
				2.708	2.718	2.712	2.708	2.718
10	20	15	1	2.649	2.689	2.701	2.703	2.693
				2.708	2.716	2.709	2.708	2.711
10	20	11	0.1	2.372	2.398	2.392	2.393	2.393
				2.398	2.512	2.4	2.398	2.455
10	20	11	1	2.372	2.396	2.39	2.394	2.393
				2.398	2.508	2.398	2.396	2.451
10	20	19	0.1	2.928	2.888	2.941	2.944	2.918
				2.944	2.944	2.949	2.944	2.944
10	20	19	1	2.926	2.887	2.939	2.942	2.915
				2.944	2.943	2.947	2.943	2.943
1	200	100	1	2.636	1.48	2.63	2.636	3.048
				4.605	4.798	4.91	4.605	4.709
1	200	100	20	2.636	1.46	2.625	2.633	3.035
				4.605	4.778	4.875	4.599	4.603
1	200	10	1	0.24	1.8	0.267	2.255	1.703
				2.303	183.798	2.687	2.302	81.915
1	200	10	20	0.24	-0.394	0.251	0.237	-0.094
				2.303	164.798	2.302	2.125	62.53
1	200	190	1	5.032	1.469	2.984	5.032	3.535
				5.247	5.247	5.534	5.247	5.248
1	200	190	20	5.032	1.464	2.948	5.032	3.535
				5.247	5.242	5.499	5.241	5.242

Figure 13-1: Some comparisons between different expressions for bounds on the mean, for $f(x) = \ln(x)$

m	M	μ	σ	linear	Taylor	LaGrange	one-sided	order-stat
10	20	15	0.1	0.0667	0.0648	0.0659	0.0637	0.0657
				0.075	0.0704	0.0677	0.0637	0.0655
10	20	15	1	0.0667	0.0651	0.0662	0.0669	0.066
				0.075	0.0707	0.068	0.0671	0.0662
10	20	11	0.1	0.0909	0.0885	0.0904	0.0909	0.0772
				0.095	0.091	0.0922	0.0909	0.091
10	20	11	1	0.0909	0.0843	0.0907	0.0918	0.075
				0.095	0.0917	0.0926	0.0917	0.0917
10	20	19	0.1	0.0526	0.0526	0.0516	0.0526	0.0526
				0.055	0.0533	0.0535	0.0526	0.0526
10	20	19	1	0.0526	0.0523	0.052	0.0526	0.0526
				0.055	0.0534	0.0538	0.0526	0.0526
1	200	100	1	0.01	0.005	-0.0116	0.01	0.0075
				0.505	0.9803	0.9882	0.0101	0.4975
1	200	100	20	0.01	0.0054	-0.0101	0.0102	0.0079
				0.505	0.9807	0.9877	0.0184	0.4975
1	200	10	1	0.1	-84.1989	0.0524	0.1001	-17.0184
				0.955	0.33	0.9903	0.1099	0.4975
1	200	10	20	0.1	-83.7949	0.0530	0.1255	-16.9784
				0.955	1.229	0.9918	0.8947	1.229
1	200	190	1	0.0053	0.0053	-0.0155	0.0053	0.0053
				0.055	0.9898	0.9923	0.0053	0.4974
1	200	190	20	0.0053	0.0055	-0.014	0.0053	0.0053
				0.055	0.9898	0.9908	0.0162	0.4975

For these results the quadratic optimum was verified to be equal to the one-sided bound when allowing for roundoff error.

Figure 13-2: Some comparisons between different expressions for bounds on the mean, for $f(x) = 1/x$

16. Conclusion

We have developed some quick closed-form expressions for bounds on the mean and standard deviation of a finite set of transformed numerical data values, where the transformation function has derivatives of constant sign in the interval of interest. In making these estimates we use only statistics on the original set of data values, and no actual values themselves. Our bounds provide a useful alternative to often difficult-to-obtain confidence intervals, requiring no distributional assumptions whatsoever. Such bounds are likely to be helpful for exploratory data analysis as an aid to getting a feel for the data, preliminary to detailed hypothesis testing.

References

[1]
R. Davis and J. King, *An Overview of Production Systems*, Machine Intelligence 8, E. W. Elcock and D. Michie, ed., Wiley, New York, 1976, pp. 300-334.

[2]
John D. Emerson, *Mathematical Aspects of Transformation*, Understanding Robust and Exploratory Data Analysis, D. Hoaglin, F. Mosteller and J. Tukey, ed., Wiley, New York, 1983, pp. 247-282.

[3]
John E. Freund and Ronald E. Walpole, *Mathematical Statistics*, Prentice-Hall, Englewood Cliffs, NJ, 1980.

[4]
J. Grant and J. Minker, *On Optimizing the Evaluation of a Set of Expressions*, International Journal of Computer and Information Science, 11 (1982), pp. 179-191.

[5]
G. H. Hardy, J. E. Littlewood, and G. Polya, *Inequalities*, Cambridge University Press, Cambridge, UK, 1952.

[6]
David C. Hoaglin, Frederick Mosteller, and John W. Tukey, editors, *Understanding Robust and Exploratory Data Analysis*, Wiley, New York, 1983.

[7]
M. H. Hoyle, *Transformations -- An Introduction and a Bibliography*, International Statistical Review, 41 (1973), pp. 203-223.

[8]
Lloyds Register, Lloyd's Register of Shipping: Statistical Tables 1982, Lloyds of London, London, U. K. 1982.

[9]
Jerzy Neyman and Elizabeth L. Scott, *Correction for Bias Introduced by a Transformation of Variables*, Annals of Mathematical Statistics, 31 (1960), pp. 643-655.

[10]
Neil C. Rowe, *Top-down Statistical Estimation on a Database*, Proceedings of the International Conference on Management of Data, ACM-SIGMOD, San Jose, CA, May, 1983, pp. 135-145.

[11]

Neil C. Rowe, *Rule-based Statistical Calculations on a Database Abstract,* Tech. Rep. STAN-CS-63-975, Stanford University, Stanford, CA, June 1983.

[12]

Neil C. Rowe, *Diophantine Compromise of a Statistical Database,* Information Processing Letters, 12 (1984), pp. 5-12.

[13]

John W. Tukey, *Exploratory Data Analysis,* Addison-Wesley, Reading, Mass., 1977.

INITIAL DISTRIBUTION LIST

Defense Technical Information Center Cameron Station Alexandria, VA 22314	2
Dudley Knox Library Code 0142 Naval Postgraduate School Monterey, CA 93943	3
Office of Research Administration Code 012A Naval Postgraduate School Monterey, CA 93943	1
Chairman, Code 52Hq Department of Computer Science Naval Postgraduate School Monterey, CA 93943	40
Associate Professor Neil C. Rowe, Code 52Rp Department of Computer Science Naval Postgraduate School Monterey, CA 93943	25
Dr. Robert Grafton Code 433 Office of Naval Research 800 N. Quincy Arlington, VA 22217	1
Dr. David W. Mizell Office of Naval Research 1030 East Green Street Pasadena, CA 91106	1