

Analyzing the Number of Samples Required for an Approximate Monte-Carlo LMS Line Estimator

David M. Mount, Nathan S. Netanyahu, and Eliyahu Zuck

Abstract. This paper analyzes the number of samples required for an approximate Monte-Carlo least median of squares (LMS) line estimator. We provide a general computational framework, followed by detailed derivations for several point distributions and subsequent numerical results for the required number of samples.

1. Introduction and Problem Definition

We are given a set of n observations in the plane (x_i, y_i) , for $i = 1, \dots, n$, and consider the linear regression model (with intercept)

$$y_i = (ax_i + b) + r_i,$$

where a and b are the slope and y -intercept of the regression line and r_i is the residual associated with the observation (x_i, y_i) . We consider robust regression, in which the goal is to estimate the line parameters (a, b) that fit the bulk of the data, even when outliers are present. We assume that an upper bound on the fraction of *outliers* q , $0 \leq q < 0.5$, is known, so that the number of outliers is at most qn . We make no assumptions about the distribution of these outliers. The remaining points, of which there are at least $(1 - q)n$ are called *inliers*, and are assumed to lie near the line (see further specifications below).

A popular approach for robust estimation of this type is Rousseeuw's *least median of squares* (LMS) line estimator [9]. The estimator computes the line of fit that minimizes the median squared residual (or any order statistic). Define a *strip* to be the region bounded between two nonvertical, parallel lines, and define the *vertical width* of a strip to be the vertical distance between these lines. LMS

Received by the editors May, 2003.

1991 *Mathematics Subject Classification.* Primary 62J05; Secondary 68W40.

Key words and phrases. Least median of squares (LMS) line estimator, Monte-Carlo LMS, approximation algorithms.

This material is based upon work supported by the National Science Foundation under Grant No. 0098151.

is computationally equivalent to determining a strip of minimum vertical width that contains at least 50% of the points. Thus, LMS estimation problem can be reduced into a geometric optimization problem.

A number of algorithms have been derived for the computation of the LMS estimator. Stromberg [12] gave an exact algorithm for computing the LMS hyperstrip in d -dimensional space. His algorithm runs in $O(n^4 \log n)$ time for simple regression. A slightly improved algorithm which runs in $O(n^4)$ time is due to Agulló [1]. The best algorithm known for finding the LMS strip (in the plane) is the topological plane-sweep algorithm due to Edelsbrunner and Souvaine [3]. It runs in $O(n^2)$ time and requires $O(n)$ space. However, even quadratic running time is unacceptably high for many applications involving large data sets.

In practice, simple Monte-Carlo approximation algorithms are often used. The *feasible solution algorithm* due to Hawkins [5] is one such algorithm. Additional Monte-Carlo-based algorithms are described in Rousseeuw and Leroy [11] and Rousseeuw and Hubert [10]. Some number of pairs of distinct points are sampled at random. Each pair of points determines the slope of a line. The optimal intercept is then computed by a process called *intercept adjustment*, where the optimal intercept is determined by reducing the simple regression problem to an LMS location estimation problem. This can be done in $O(n \log n)$ time (see [11]). Thus for a fixed number of subsamples, it is possible to determine the best LMS line approximation in $O(n \log n)$ time. This is equivalent to finding the narrowest strip containing (at least) half of the points over all the strips of slopes defined by the subsamples. The midline of the narrowest strip is the LMS line approximation.

Although inliers are usually assumed to lie exactly on the line, this is rarely the case in practice. We assume that the distribution of residuals is known and is independent of the x -coordinate. The main issue considered is how many pairs of points should be sampled to guarantee that, with sufficiently high probability, this Monte-Carlo estimate returns a sufficiently close approximation to the LMS line of fit? The meaning of “sufficiently close” is based on the formulation given by Mount *et al.* [6]. Let w^* denote the vertical width of the narrowest strip containing (at least) 50% of the points. Given an approximation error bound, $\epsilon \geq 0$, we say that any strip that contains (at least) 50% of the points and is of vertical width at most $(1 + \epsilon)w^*$ is an ϵ -*approximate solution* to the LMS problem. Given a *confidence probability* p_c , our goal is to compute the minimum number of random samples needed, so that the Monte-Carlo algorithm returns an ϵ -approximate solution to the LMS problem with probability at least p_c . This analysis is based on some assumptions about the distribution of inliers, which are given in the next section.

2. Assumptions on the Point Distribution

Suppose that the line in question is given by $y = ax + b$. As previously noted, we make some concrete assumptions about the distribution of the inliers. Specifically, we assume that the x -coordinates of the points are uniformly distributed over an interval $[x_0, x_1]$, and that the y -coordinate for a point with x -coordinate x is $ax + b + z$, where the deviate z is a random variable having a certain probability density

function (pdf). In Section 4 we consider specifically Gaussian and exponential pdf's.¹ In the full paper we also consider the uniform pdf [7].

Because the approximation bound is based on the ratio of vertical widths of two parallel strips, we may apply any affine transformation to the plane that preserves such ratios. In particular, in order to simplify the analysis, we transform the x -coordinates to lie within the interval $[0, 1]$, and we transform the y -coordinates by subtracting $ax - b$, so that the line of fit is mapped to the x -axis, and finally we scale the y -coordinates so that the standard deviation of residuals is some desired value. After this transformation, the optimal slope is zero. It is easy to verify that all these transformations preserve the ratio of vertical widths of parallel strips. Our analysis holds in the limit as n tends to infinity.

3. Sketch of Computational Procedure

Our analyses for the various residual distributions are all based on a common approach. Any sampled pair of points defines a line with some slope s . Based on our knowledge of the distributions of the x - and y -coordinates of the inliers, it will be possible to derive the pdf of s , for pairs of inliers. In later sections we provide this derivation, but for now, let $f_S(s)$ denote this slope pdf.

Consider a parallel strip whose sides are of slope s , whose central line has y -intercept b , and whose vertical width is w . Let $A_b(s, w)$ denote the probability mass of this strip relative to the given point distribution. We will omit b when it is clear from context. The algorithm computes a strip of slope s that contains at least half of the points. Since we know nothing about outliers, if we want to guarantee that at least half of all the n points lie within a given strip, the strip should contain a fraction of at least $1/(2(1-q))$ of the $(1-q)n$ inliers. Throughout, we let $\mathcal{F}_I = 1/(2(1-q))$ denote this fraction of inliers. (Recall that $q < 0.5$.) Define the *slope-width function*, $w(s)$, to be the vertical width of the narrowest strip that has this fraction of the inlier probability mass. That is,

$$(3.1) \quad w(s) = \operatorname{argmin}_w \exists b (A_b(s, w) \geq \mathcal{F}_I).$$

For most reasonable residual distributions, we would expect that $w(s)$ increases monotonically relative to its optimum value of $w(0)$, denoted w^* . (See Figure 1.) This can be shown for the distributions presented here, by a straightforward perturbation argument, which will be presented in the full version of the paper [7]. Let $[s_{\min}, s_{\max}]$ denote the largest interval such that, for all $s \in [s_{\min}, s_{\max}]$, $w(s) \leq (1+\epsilon)w^*$. A slope s is ϵ -good if it lies within this interval. A sampled point pair is ϵ -good if the line passing through this pair has an ϵ -good slope. Recalling the $f_S(s)$ is the slope pdf, the probability of randomly sampling an ϵ -good point pair, denoted p_g , is

$$(3.2) \quad p_g = \int_{s_{\min}}^{s_{\max}} f_S(s) ds.$$

¹Assuming normally distributed residuals is fairly common. The exponential pdf was chosen as a case study for a one-sided distribution.

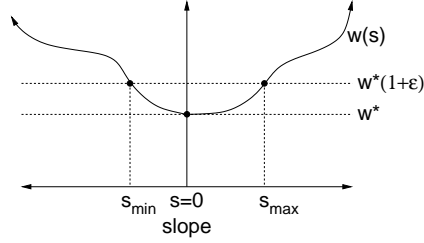


FIGURE 1. An illustration of the slope-width function $w(s)$ and minimum and maximum allowable slopes.

Finally, given that with probability p_c at least one good sample should be chosen, it follows that if N is the number of point pairs to be sampled, the probability of generating no good pair in N trials should be at most $1 - p_c$. Since we assume that sampling is performed independently, and since the probability of selecting a pair of inliers is $(1 - q)^2$, N should be selected such that

$$(1 - (1 - q)^2 p_g)^N \leq 1 - p_c.$$

Solving for N , we obtain the desired bound on the number of samples.

$$N \geq \frac{\log(1 - p_c)}{\log(1 - (1 - q)^2 p_g)}.$$

The remainder of the analysis reduces to computing the slope density function $f_S(s)$, computing the slope-width function $w(s)$, inverting this function to compute s_{\min} and s_{\max} , and finally solving the integral of Eq. (3.2). In the rest of the paper we will provide detailed derivations of these results for two common residual pdf's.

4. Derivation of Slope Distribution

Let $p_i = (x_i, y_i)$, $p_j = (x_j, y_j)$ denote two distinct inlying points sampled at random. According to the point distribution model described in Section 2, x_i , x_j are independent, uniformly distributed in the interval $[0, 1]$, that is, $x_i, x_j \sim \mathcal{U}(0, 1)$ and y_i, y_j are independent, identically distributed (i.i.d.) according to some specified pdf. We make the general position assumption that the x -coordinates of the points are distinct. The slope of the line joining p_i and p_j is

$$(4.1) \quad s = \frac{y_j - y_i}{x_j - x_i}.$$

Thus to find $f_S(s)$, it suffices to find the pdf's of $(y_j - y_i)$ and $(x_j - x_i)$, and then find the pdf of their quotient. The latter can be found, using the following result from basic probability theory [4, 8]. Let $S = V/U$ denote the quotient of two random variables V and U . The pdf of S is given by $f_S(s) = \int_{-\infty}^{\infty} f_{VU}(us, u)|u|du$,

where $f_{VU}(\cdot)$ is the joint pdf of V and U . Thus if V and U are independent,

$$(4.2) \quad f_S(s) = \int_{-\infty}^{\infty} f_V(us) f_U(u) |u| du.$$

Let U and V denote, respectively, the random variables $x_j - x_i$ and $y_j - y_i$. The pdf of U is given by the convolution of the pdf's corresponding to $\mathcal{U}(0, 1)$ and $\mathcal{U}(-1, 0)$. It is easy to show that

$$(4.3) \quad f_U(u) = \begin{cases} 1+u & -1 \leq u \leq 0 \\ 1-u & 0 \leq u \leq 1 \\ 0 & \text{otherwise.} \end{cases}$$

In the following subsections we derive the pdf of V for two common distributions. Also, we derive for each case the resulting pdf of the slope, using (4.2).

4.1. The Gaussian Case

We assume, without loss of generality, that $y_i, y_j \sim \mathcal{N}(0, 1)$. Thus $-y_i \sim \mathcal{N}(0, 1)$, as well. Based on the fact that the sum of (two) independent, normally distributed random variables has a normal pdf whose expected value is the sum of the expected values and whose variance is the sum of the variances, we have $y_j - y_i \sim \mathcal{N}(0, \sqrt{2})$. In other words, $f_V(v) = (1/(2\sqrt{\pi})) \exp(-v^2/4)$. Employing (4.2) we thus obtain

$$(4.4) \quad f_S(s) = \frac{1}{2\sqrt{\pi}} \int_{-1}^1 e^{-k(s)u^2} f_U(u) |u| du,$$

where $k(s) \equiv s^2/4$. Since in view of (4.3) the integrand is an even function of u , we may rewrite the above equation as:

$$(4.5) \quad f_S(s) = \frac{1}{\sqrt{\pi}} \int_0^1 e^{-k(s)u^2} (1-u) u du.$$

Solving the definite integral yields the following bottom-line expression:

$$(4.6) \quad f_S(s) = \frac{1}{2\sqrt{\pi k(s)}} \left(1 - 0.5\sqrt{\pi/k(s)} \operatorname{erf}(\sqrt{k(s)}) \right),$$

where $\operatorname{erf}(x) = (2/\sqrt{\pi}) \int_0^x e^{-t^2} dt$. (See [7], for a detailed derivation.) Figure 2(a) gives a plot of $f_S(s)$ for the Gaussian case. Note that although the denominator of $f_S(s)$ approaches zero as $s \rightarrow 0$, the function is well defined. Specifically, it can be shown that the limit of $f_S(s)$ as $s \rightarrow 0$ is equal to $1/(6\pi)$.

4.2. The Exponential Case

We now consider the exponential case. Without loss of generality, we assume that $y_i, y_j \sim \mathcal{E}(1)$, that is, y_i, y_j are independent, exponentially distributed with $\lambda = 1$. (This one-sided distribution is a special case of the gamma and Weibull distributions.) In other words, $y_j \sim e^{-y} H(y)$ and $-y_i \sim e^y H(-y)$, where $H(y)$ denotes a step function, such that $H(y) = 1$ for $y \geq 0$ and $H(y) = 0$ otherwise. It can be shown that the convolution of these pdf's results in a Laplace pdf (see [7]):

$$(4.7) \quad y_j - y_i \sim \frac{1}{2} e^{-|v|}.$$

Employing (4.2) we now obtain

$$(4.8) \quad f_S(s) = \frac{1}{2} \int_{-1}^1 e^{-|us|} f_U(u) |u| du.$$

As before, since the integrand is an even function of u , we may rewrite this as:

$$(4.9) \quad f_S(s) = \int_0^1 e^{-|s|u} (1-u) u du.$$

Solving the definite integral yields the desired slope pdf:

$$(4.10) \quad f_S(s) = \frac{1}{|s|^3} \left((|s| + 2)e^{-|s|} + |s| - 2 \right).$$

(See [7], for a detailed derivation.) Although the denominator is equal to zero at $s = 0$, $f_S(s)$ is well-defined at the origin. In particular, it can be shown (by a Taylor expansion of the numerator up to terms of degree 4) that the limit of $f_S(s)$ as $s \rightarrow 0$ is equal to $1/6$. Figure 2(b) gives a plot of $f_S(s)$ for the exponential case.

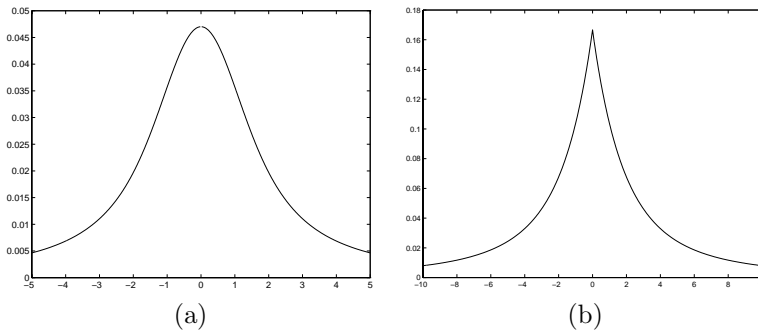


FIGURE 2. Slope pdf: (a) Gaussian case, (b) exponential case.

5. Derivation of the Slope-Width Function

In this section we compute the slope-width function $w(s)$, defined in Section 3.

5.1. The Gaussian Case

We consider first the special case $s = 0$ for which $w(s) = w(0) = w^*$. We make the observation that the desired parallelogram (a horizontal rectangle of width w^* in this case) must be symmetric with respect to the x -axis. By definition of the pdf of the y -coordinates, it is easy to show that the “area” (that is, probability mass) of any other horizontal rectangle of width w^* will be smaller than $A(0, w^*) = A_0(0, w^*)$. Thus

$$(5.1) \quad A(0, w^*) = \int_{-w^*/2}^{w^*/2} \int_0^1 f_Y(y) dx dy = \frac{1}{\sqrt{2\pi}} \int_{-w^*/2}^{w^*/2} e^{-y^2/2} dy,$$

that is,

$$(5.2) \quad A(0, w^*) = \sqrt{\frac{2}{\pi}} \int_0^{w^*/2} e^{-y^2/2} dy.$$

Given \mathcal{F}_I , we can find w^* (the solution to the implicit equation $A(0, w^*) = \mathcal{F}_I$), by using standard numerical analysis techniques [2], such as the Gaussian quadrature (to approximate the various definite integrals) and the iterative bisection method for root-finding of an equation. For $\mathcal{F}_I = 0.5$, for example, we obtain $w^* \approx 1.349$.

We now assume, without loss of generality, that $s > 0$, and consider a parallelogram of slope s and width w . It can be shown that among all such parallelograms, the one that is symmetric with respect to the line $y = sx - s/2$ has maximum area. Specifically this implies that $y = sx - s/2$ is also the midline of the narrowest strip containing half of the data points.² Henceforth, let $A(s, w) = A_{-s/2}(s, w)$. The non-vertical sides of the desired parallelogram are thus given by the line equations: $y_I = sx - (s - w)/2$ and $y_{II} = sx - (s + w)/2$. We distinguish between two cases: (1) $s \leq w$ and (2) $s > w$ (see Figures 3(a) and 3(b), respectively). Due to space limitations, we only discuss here the first case. (See [7] for further details.)

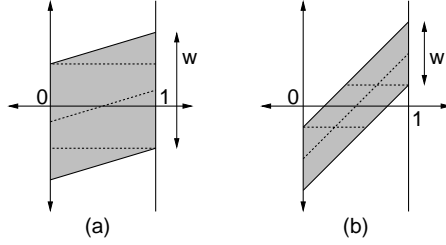


FIGURE 3. The low and high slope cases for the Gaussian distribution.

We decompose the desired parallelogram into the horizontal rectangle whose sides are $y = \pm(w - s)/2$, and the triangles above and below this rectangle. The area (i.e., probability mass) of the two triangles is clearly the same. Thus the total probability mass of the parallelogram is given by

$$(5.3) \quad A(s, w) = \int_{-(w-s)/2}^{(w-s)/2} f_Y(y) dy + 2 \int_{-(w+s)/2}^{-(w-s)/2} x_{II}(y) f_Y(y) dy,$$

where $x_{II}(y) = (1/s)y + (1 + w/s)/2$, which yields

$$\begin{aligned} A(s, w) &= \int_{-(w-s)/2}^{(w-s)/2} f_Y(y) dy + \frac{2}{s} \int_{-(w+s)/2}^{-(w-s)/2} y f_Y(y) dy + \\ &\quad \left(1 + \frac{w}{s}\right) \int_{-(w+s)/2}^{-(w-s)/2} f_Y(y) dy. \end{aligned}$$

²Moving any other strip in the direction of the above midline increases the probability mass; thus the width of any other strip is not a local minimum, in the sense that if the probability mass exceeds, say, 0.5, it is possible to locally shrink that strip so that it contains half of the data.

Solving the second definite integral leads to the following expression for $A(s, w)$:

$$A(s, w) = \sqrt{\frac{2}{\pi}} \int_0^{(w-s)/2} e^{-y^2/2} dy - \sqrt{\frac{8}{\pi}} \left(\frac{1}{s}\right) e^{-(w^2+s^2)/8} \sinh\left(\frac{ws}{4}\right) + \left(1 + \frac{w}{s}\right) \frac{1}{\sqrt{2\pi}} \int_{-(w+s)/2}^{-(w-s)/2} e^{-y^2/2} dy.$$

Using standard numerical analysis techniques, as before, we can obtain numerical solutions for the implicit equation $A(s, w(s)) = \mathcal{F}_I$. That is, for a given \mathcal{F}_I and s , we can look up $w(s)$ and vice versa.

Figures 4(a), 4(b) show plots of the strip width as a function of slope, for the Gaussian case and the exponential case, respectively, for various values of \mathcal{F}_I .

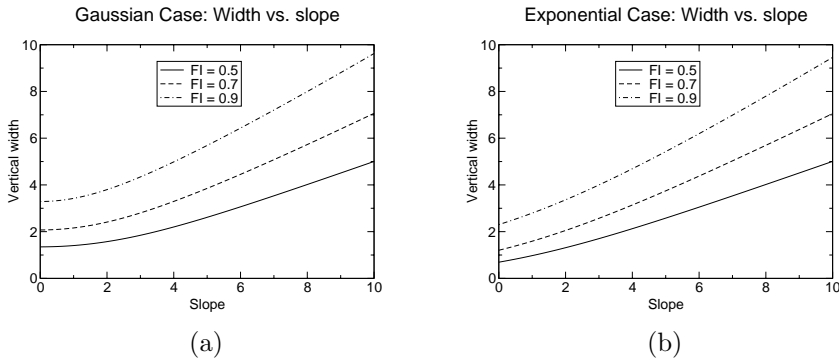


FIGURE 4. Strip width as a function of slope for various \mathcal{F}_I values: (a) Gaussian case, (b) exponential case.

5.2. The Exponential Case

We consider first the special case $s = 0$. Given the one-sided pdf of the y -coordinates, it is obvious that the lower edge of the desired strip (of width w^*) coincides in this case with the x -axis. Thus letting $A(0, w^*) = A_0(0, w^*)$ we have

$$(5.4) \quad A(0, w^*) = \int_0^{w^*} e^{-y} dy = 1 - e^{-w^*},$$

that is, for a given \mathcal{F}_I , we have

$$(5.5) \quad w^* = \ln\left(\frac{1}{1 - \mathcal{F}_I}\right).$$

Because of the lack of symmetry, determining the optimal strip for $s \neq 0$ is much less obvious than the Gaussian case. To do so, we will derive the probability mass associated with a differential strip of slope s , that is, the probability density of a line of slope s as a function of its y -intercept. Consider the family of lines

$y = sx + y_{\text{int}}$, where the intercept y_{int} is a random variable. (Assume that $s > 0$. An analogous analysis to the one below applies to $s < 0$.) To obtain the distribution of y_{int} , that is, $F(z) = \mathcal{P}(y_{\text{int}} < z)$, we distinguish between the two cases: (1) $z \geq 0$ and (2) $-s < z < 0$. The distribution is 0 for $z < -s$, due to the underlying exponential pdf of the y -coordinates, and is essentially equal to the probability mass of the strip defined by $y = sx + z$ and $y = sx - s$ (see Figure 5). We have for $z > 0$

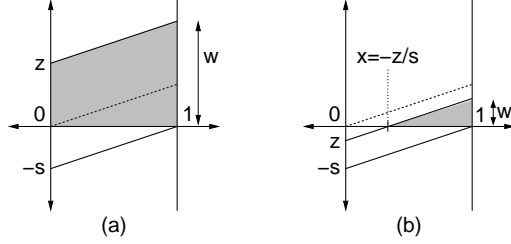


FIGURE 5. Regions for probability mass computation due to (a) (5.6) and (b) (5.7).

$$F(z) = \mathcal{P}(y_{\text{int}} < z) = \int_0^1 \left(\int_0^{sx+z} e^{-y} dy \right) dx = \int_0^1 (1 - e^{-sx-z}) dx,$$

which yields

$$(5.6) \quad F(z) = \mathcal{P}(y_{\text{int}} < z) = 1 + \frac{1}{s} (e^{-z-s} - e^{-z}).$$

For $-s < z \leq 0$ we have

$$F(z) = \int_{-z/s}^1 \left(\int_0^{sx+z} e^{-y} dy \right) dx = \int_{-z/s}^1 (1 - e^{-sx-z}) dx,$$

which yields

$$(5.7) \quad F(z) = 1 + \frac{z}{s} + \frac{1}{s} (e^{-z-s} - 1).$$

Differentiating the above distribution with respect to z provides the probability density of a line having a slope s and y -intercept z . We obtain

$$(5.8) \quad f_Z(z) = \begin{cases} \frac{1}{s} (1 - e^{-z-s}) & -s \leq z \leq 0 \\ \frac{1}{s} (e^{-z} - e^{-z-s}) & z \geq 0. \end{cases}$$

It can easily be seen that the above pdf is unimodal. The mode occurs at $z = 0$ and its value is $f_Z(0) = m$, where $m = (1/s) (1 - e^{-s})$.

Having established (for a given slope $s > 0$) the pdf of a line's intercept and its unimodality, we make the observation that the width of the desired strip is given by $w(s) = y_1 - y_0$, for some $y_0 < 0$, $y_1 > 0$, such that:

- (1) $f_Z(y_0) = f_Z(y_1)$, and
- (2) $F(y_1) - F(y_0) = A(s, w(s)) = \mathcal{F}_I$.

In this context, it is understood that $A(s, w(s))$ applies to the strip bounded by the y -intercepts y_0 and y_1 . The first observation arises as a local-minimality constraint on the strip lying between the intercepts $[y_0, y_1]$. (If this did not hold, a differential argument shows that an infinitesimal shift of the strip, either up or down, would increase the probability mass without increasing the strip's width, contradicting the strip's minimality.) The second fact simply states that the strip must contain a fraction \mathcal{F}_I of the probability mass. From (5.8) and the first constraint we obtain

$$(5.9) \quad 1 - e^{-y_0-s} = e^{-y_1} - e^{-y_1-s},$$

and substituting $(y_0 + w)$ for y_1 yields

$$(5.10) \quad e^{-y_0} (e^{-s} + e^{-w} - e^{-w-s}) = 1.$$

From (5.6), (5.7), we have

$$F(y_0) = 1 - \frac{1}{s} + \frac{y_0}{s} + \frac{1}{s} e^{-y_0-s}$$

and

$$F(y_1) = 1 + \frac{1}{s} (e^{-s} - 1) e^{-y_1}.$$

From the last two equations and the second constraint we obtain

$$A(s, w(s)) = \frac{1}{s} ((e^{-s} - 1)e^{-y_1} + 1 - y_0 - e^{-s}e^{-y_0}) = \mathcal{F}_I.$$

Combining the latter equation with (5.9) and (5.10) yields

$$A(s, w(s)) s = -y_0,$$

or

$$A(s, w(s)) s = -\ln(e^{-s} + e^{-w} - e^{-w-s}) = -\mathcal{F}_I s.$$

Hence we get the following closed-form expression

$$(5.11) \quad w(s) = \ln \left(\frac{1 + e^{-s}}{e^{-\mathcal{F}_I s} - e^{-s}} \right).$$

Finding s from a given w does not yield a closed-form solution. As in the previous subsection, a standard iterative root-finding technique was used.

6. Numerical Results

We show the concrete results of the analysis for the Gaussian and exponential distributions. Figure 6 shows the probability that single random sampled pair is *good* as a function of ϵ . Recall that a sampled pair is good if there is a strip whose slope is determined by the pair, that contains a fraction of \mathcal{F}_I of the inliers, and whose width is at most $(1 + \epsilon)$ times the optimum width. Recall that $\mathcal{F}_I = 1/(2(1 - q))$, where q is the expected fraction of outliers, and hence the values $\mathcal{F}_I = \{0.5, 0.7, 0.9\}$ shown in the plots correspond to outlier fractions of $q =$

$\{0, 0.29, 0.44\}$, respectively. Figure 7 shows the required number of samples as a function of ϵ for various confidence levels p_c , and for the fixed value $\mathcal{F}_I = 0.5$.

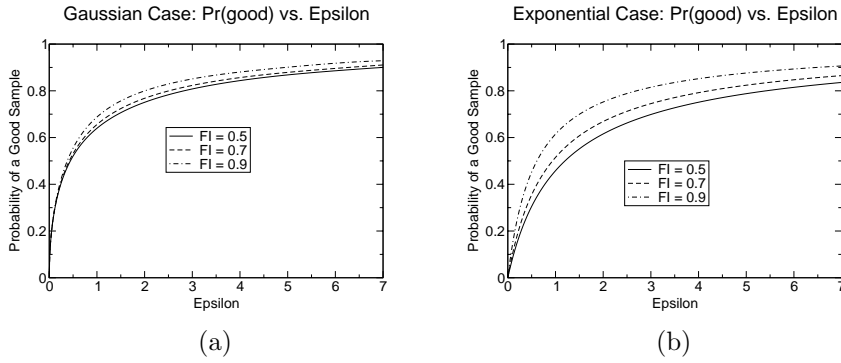


FIGURE 6. Probability that a single random sample is good as a function of ϵ for various \mathcal{F}_I values: (a) Gaussian case, (b) exponential case.

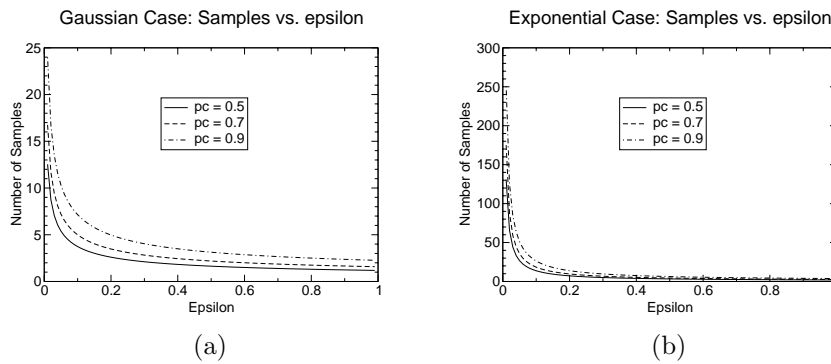


FIGURE 7. Required number of samples as a function of ϵ for $\mathcal{F}_I = 0.5$ and for various confidence levels p_c : (a) Gaussian case, (b) exponential case.

References

- [1] J. Agulló. Exact algorithms for computing the least median of squares estimate in multiple linear regression algorithm. In Y. Dodge, editor, *L₁-Statistical Procedures and Related Topics*, volume 31 of *IMS Lecture Notes Monograph Series*, pages 133–146. Institute of Mathematical Statistics, 1997.

- [2] K.E. Atkinson. *An Introduction to Numerical Analysis*. John Wiley, New York, 1978.
- [3] H. Edelsbrunner and D.L. Souvaine. Computing median-of-squares regression lines and guided topological sweep. *Journal of the American Statistical Association*, 85:115–119, 1990.
- [4] M. Fisz. *Probability Theory and Mathematical Statistics*. John Wiley, New York, 1963. 3rd edition.
- [5] D.M. Hawkins. The feasible set algorithm for least median of squares regression. *Computational Statistics and Data Analysis*, 16:81–101, 1993.
- [6] D.M. Mount, N.S. Netanyahu, K. Romanik, R. Silverman, and A.Y. Wu. Practical approximation algorithm for the LMS line estimator. In *Proceedings of the Eighth ACM-SIAM Symposium on Discrete Algorithms*, pages 464–472, New Orleans, Louisiana, January 1997.
- [7] D.M. Mount, N.S. Netanyahu, and E. Zuck. Analyzing the number of samples required for an approximate Monte-Carlo LMS line estimator. Technical Report in preparation, 2003.
- [8] A. Papoulis. *Probability, Random Variables, and Stochastic Processes*. McGraw-Hill, Tokyo, 1965.
- [9] P.J. Rousseeuw. Least median of squares regression. *Journal of The American Statistical Association*, 79:871–880, 1984.
- [10] P.J. Rousseeuw and M. Hubert. Recent developments in PROGRESS. In Y. Dodge, editor, *L₁-Statistical Procedures and Related Topics*, volume 31 of *IMS Lecture Notes Monograph Series*, pages 201–214. Institute of Mathematical Statistics, 1997.
- [11] P.J. Rousseeuw and A.M. Leroy. *Robust Regression and Outlier Detection*. Wiley & Sons, New York, 1987.
- [12] A.J. Stromberg. Computing the exact least median of squares estimate and stability diagnostics in multiple linear regression. *SIAM Journal on Scientific Computing*, 14:1289–1299, 1993.

Acknowledgments

We thank Jeremy Schiff for several insightful discussions and Adi Sweete for his help in producing the numerical results contained in this paper.

Department of Computer Science, University of Maryland, College Park, MD 20742, U.S.A.

E-mail address: `mount@cs.umd.edu`

Department of Computer Science, Bar-Ilan University, Ramat-Gan 52900, Israel, and Center for Automation Research, University of Maryland, College Park, MD 20742, U.S.A.

E-mail address: `nathan@macs.biu.ac.il` or `nathan@cfar.umd.edu`

Department of Computer Science, Bar-Ilan University, Ramat-Gan 52900, Israel,

E-mail address: `zuckel@macs.biu.ac.il`