Proceedings of the 2013 Winter Simulation Conference
R. Pasupathy, S.-H. Kim, A. Tolk, R. Hill, and M. E. Kuhl, eds.

CRITICAL SAMPLE SIZE FOR THE L_p-NORM ESTIMATOR IN LINEAR REGRESSION MODELS

Alejandro Llorente

Instituto de Ingeniería del Conocimiento
Universidad Autónoma de Madrid
Madrid, 28049, SPAIN

Alberto Suárez

Computer Science Department
Escuela Politécnica Superior
Universidad Autónoma de Madrid
Madrid, 28049, SPAIN

ABSTRACT

In the presence of non-Gaussian noise the least squares estimator for the parameters of a regression model can be suboptimal. Therefore, it is reasonable to consider other norms. L_p-norm estimators are a useful alternative, particularly when the residuals are heavy-tailed. We analyze the convergence properties of such estimators as a function of the number samples available for estimation. An analysis based on the Random Energy Model (REM), a simplified model used to describe the thermodynamic properties of amorphous solids, shows that, in a specific limit, a second order phase transition takes place: For small sample sizes the typical and average values of the estimator are very different. For sufficiently large samples, the most probable value of the estimator is close to its expected value. The validity analysis is illustrated in the problem of predicting intervals between subsequent tweets.

1 INTRODUCTION

The goal in a regression problem is to build a model that predicts the value of a target (dependent) real-valued variable, $y \in \mathbb{R}$, from the values of a vector of attributes $x \in \mathbb{R}^D$. The model is induced from the set of training instances $\{(x_n, y_n)\}_{n=1}^{M}$, in which the value of the dependent variable is known. To address this problem one generally assumes a family of models $h(x;\theta)$ expressed in terms of the vector of parameters θ. Typically, the value of θ is determined by solving an optimization problem. One of the commonly used cost functions is the L_p-norm error

$$\theta^* = \arg\min_{\theta} \frac{1}{M} \sum_{n=1}^{M} |y_n - h(x_n;\theta)|^p . \quad (1)$$

This problem is usually referred to as L_p regression because the error function is the L_p-norm of the residuals, which are differences between the model predictions and the observed values. Least squares ($p = 2$), Minimum Absolute Deviation ($p = 1$) and Chebyshev ($p = \infty$) regression are particular instances of this class of estimators. Some questions naturally arise when this type of regression problem is addressed. Specifically, how should the order of the norm for the regression error be determined for a particular empirical sample? Some efforts in this direction have been made in the literature. In (Money, Affleck-Graves, Hart, and Barr 1982, Nyquist 1983) some heuristics are given to estimate the optimal p in terms of the kurtosis and other higher order moments of the distribution of residuals. Another issue is, given a choice of p, how large need the sample be to ensure that estimates of the regression coefficients are reliable? This is the question addressed in this research. If the residuals are i.i.d. random variables, the L_p-norm is simply an empirical estimate of the moment of order p of the probability distribution of the absolute value of the residuals. Some authors have observed that empirical estimates of moments can be rather unreliable when the data are heavy-tailed (Crovella and Lipsky 1997). The empirical estimates become more unstable for larger values of p (Angeletti, Bertin, and Abry 2012). These observations directly apply to the estimation of the L_p-norm of the residuals in regression. When these estimates are inaccurate, the value of the optimizer (1) is unreliable as well.

The effects of the finite size of the sample in the reliability of the estimators of the regression parameter is analyzed using an analogy with the computation of the partition function in the Random Energy Model (REM) (Derrida 1981). This model, which is described in Section 2, was introduced in the area of disordered systems in physics to describe the properties of amorphous solids. The analysis of the behavior of the partition function in the REM can be also applied to the characterization of the asymptotic properties of the empirical estimate of the L_p-norm error. Specifically, we consider the properties of this estimator in the limit of large samples M and large p with $2\log M/p^2$ constant. The quantity of interest is the bias of the *logarithm* of this empirical estimate. The presence of a non-zero bias in this quantity indicates that the typical and the expected values are different. This, in turn, implies that the L_p-norm estimate is unreliable. In Section 3 the analysis of the partition function in the random energy model is adapted to the L_p-norm regression problem. Assuming that the residuals are approximately lognormal i.i.d. random variables this analysis is used to derive the minimum sample size needed to obtain reliable estimates of the model parameters. Lognormal models for data are common in economics (Aitchison and Calvert Brown 1957, Crow and Shimizu 1988), insurance and finance (Frachot, Moudoulaud, and Roncalli 2003), in computer systems and networks (Downey 2001, Mitzenmacher 2004, Downey 2005a) and in social patterns of interaction (Naruse and Kubo 2006). Even though explicit expressions are give for the lognormal case only, the derivations can be readily extended to other subexponential distributions with finite moments. The validity of the analysis is illustrated in linear regression problems using simulated and real data. Finally, section 4 presents a summary of the results and conclusions of the present investigation.

2 METHODS

2.1 The Random Energy Model

The random energy model (REM) was introduced by Derrida in the area of statistical mechanics to describe the thermodynamic properties of disordered systems (Derrida 1981). Consider a system with $M = 2^K$ configurations. The energy of the ith configuration is E_i. In the REM, the energy levels $\{E_i\}_{i=1}^{M}$ are i.i.d. random variables whose probability density function is $f_E(e)$. The partition function at temperature β^{-1} for a particular realization of the system is defined as

$$Z_M(\beta) = \sum_{i=1}^{M} e^{-\beta E_i} \tag{2}$$

Thermodynamic quantities of the system such as average energy, energy fluctuations, heat capacity, etc. can be obtained from $\tilde{Z}_M(\beta)$ by taking derivatives with respect to the temperature (Mézard and Montanari 2009).

To make the connection with the L_p-norm it is preferable to work with a normalized partition function

$$\tilde{Z}_M(\beta) = \frac{1}{M}\sum_{i=1}^{M} e^{-\beta E_i}. \tag{3}$$

Note that the normalized partition function corresponds to the empirical estimate of the moment-generating function that can be made using the sample $\{E_i\}_{i=1}^{M}$. Since the residuals of the L_p regression problem can be viewed as a sequence of random i.i.d. random variables, a direct connection can be established between the REM and the L_p-norm error: The absolute value of the residuals are non-negative random variables that can be expressed in the form of Boltzmann factors $e^{-\beta E_i}$. The number of configurations M corresponds to the sample size in the estimation in the regression problem.

For the subsequent derivations, it is useful to define the following asymptotic relation between sequences of random variables, which is valid to leading order in the exponent.

Definition 1 *Let A_M and B_M be two sequences*

$$A_M \doteq B_M \quad \Leftrightarrow \quad \lim_{M\to\infty} \frac{1}{M} \log A_M = \lim_{M\to\infty} \frac{1}{M} \log B_M. \tag{4}$$

In this work if one of the sides is a sequence of random variables, the convergence is understood in probability (see (Mézard and Montanari 2009) for further details). This definition applies also when the variables are indexed by $K = \log M / \log 2$ and the limit $M \to \infty$ is taken.

The first step of the derivation of the REM is to define an entropy function in terms of $N(\varepsilon, \varepsilon+\delta)$, the number of configurations whose energies are in the interval $\mathscr{I} = [K\varepsilon, K(\varepsilon+\delta)]$, with $K = \log_2 M$. This is a binomial random variable whose first two moments are

$$\mathbb{E}\left[N(\varepsilon,\varepsilon+\delta)\right] = 2^K \mathscr{P}_{\mathscr{I}}(\varepsilon,\varepsilon+\delta) \tag{5}$$

$$\mathbb{V}\left[N(\varepsilon,\varepsilon+\delta)\right] = 2^K \mathscr{P}_{\mathscr{I}}(\varepsilon,\varepsilon+\delta)\left(1-\mathscr{P}_{\mathscr{I}}(\varepsilon,\varepsilon+\delta)\right). \tag{6}$$

In terms of the function

$$g_E(z) = \frac{1}{K} \log f_E(z), \tag{7}$$

the probability of finding a configuration whose energy is in the interval $(\varepsilon, \varepsilon+\delta)$ is

$$\mathscr{P}_{\mathscr{I}}(\varepsilon,\varepsilon+\delta) = \int_{K\varepsilon}^{K(\varepsilon+\delta)} f_E(e)de = \int_{\varepsilon}^{\varepsilon+\delta} K\exp\left[Kg_E(z)\right]dz \doteq \exp\left[K \max_{z\in[\varepsilon,\varepsilon+\delta]} \{g_E(z)\}\right] \tag{8}$$

This last expression is obtained by saddle point integration (Bender and Orszag 1978).

The *microcanonical entropy density function* (in this work, to simplify the terminology, the *entropy function*) is defined as

$$N(\varepsilon,\varepsilon+\delta) \doteq \exp\left[K \max_{y\in[\varepsilon,\varepsilon+\delta]} s_a(y)\right] \tag{9}$$

or, equivalently,

$$\lim_{K\to\infty} \frac{1}{K} \log N(\varepsilon,\varepsilon+\delta) = \max_{y\in[\varepsilon,\varepsilon+\delta]} s_a(y). \tag{10}$$

Using expression (5), to leading order in the exponent,

$$s_a(y) = \log 2 - g_E(y). \tag{11}$$

The following proposition allows us to establish a relationship between the partition functions and the entropy functions.

Proposition 1 *If $s_a(y)$ exists and the limit in (10) is uniform in y then*

$$\tilde{Z}_M(\beta) \doteq \exp\left[K \max_y \{s_a(y) - \beta y\}\right].$$

These results are a particular application of Large Deviation Theory (LDT), a branch of statistics that describes the asymptotic properties of extreme events. Further details on LDT can be found in (Touchette 2009, Dorlas and Wedagedera 2001).

The sample average (3) is an unbiased estimator of $\mathbb{E}[e^{-\beta E_i}]$ for any sample size, M. However, the *typical* value of (3) can actually be very different from the average. If the typical and the average values differ (3) is an unreliable estimate of $\mathbb{E}[e^{-\beta E_i}]$. To quantify the discrepancy between the expected value of the normalized partition function and the typical value estimated from a finite-size sample one can use the bias of the *logarithm* of the normalized partition function

$$B_M = \log \mathbb{E}[e^{-\beta E_i}] - \log \tilde{Z}_M. \tag{12}$$

The random variable B_M quantifies the discrepancy between *average* and *typical* behavior of (3). To understand why this is the case, let $f_M(z)$ be the density function of $\tilde{Z}_M$. As M increases $f_M(z)$ becomes more concentrated around its mode z_M^*. Since $\log z$ varies smoothly in the region where $f_M(z)$ is peaked, in the limit $M \to \infty$ the expectation of $\log \tilde{Z}_M$ is approximately

$$\mathbb{E}\left[\log \tilde{Z}_M\right] = \int f_M(z) \log z \, dz \approx \log z_M^*. \tag{13}$$

On the other hand

$$\mathbb{E}[e^{-\beta E_i}] = \mathbb{E}[\frac{1}{M}\sum_{i=1}^{M} e^{-\beta E_i}] = \mathbb{E}[\tilde{Z}_M]. \tag{14}$$

Therefore,

$$\mathbb{E}[B_M] = \log \mathbb{E}[e^{-\beta E_i}] - \mathbb{E}[\log \tilde{Z}_M] \approx \log \mathbb{E}[\tilde{Z}_M] - \log z_M^*,$$

which is a measure of the difference between the mode and the mean of $f_M(z)$. In the next section explicit expressions for the entropy are given for lognormal samples.

2.1.1 L_p-Norm Error for Lognormal Residuals

In this section we derive explicit expressions for the bias in the logarithm of $\tilde{Z}_M$, the normalized partition function in the random energy model, when the energy levels $\{E_i\}_{i=1}^M$ are normally distributed ($E_i \sim \mathcal{N}(\mu, \sigma)$). This means that the Boltzmann factors $\left\{e^{-\beta E_i}\right\}_{i=1}^M$, whose average is the normalized partition function, are lognormally distributed. Without loss of generality, one can that assume $\mu = 0$ because μ appears only as a multiplicative constant $e^{-\beta\mu}$ in the Boltzmann factors.

To make the connection between REM and the L_p-norm regression, the Boltzmann factors, which are non-negative, are identified with the absolute values of the residuals of the regression model

$$|r_i| \leftrightarrow e^{-\beta E_i}. \tag{15}$$

Since we are interested in the behavior of the L_p-norm error, we also need to consider the distribution of the pth power of the absolute value of residuals. However, if $|r_i| = \exp[\sigma Y], Y \sim N(0,1)$ follows a lognormal distribution whose parameters are $(\mu = 0, \sigma)$, the quantity $|r_i|^p = \exp[p\sigma Y]$ is also lognormal with parameters $(\mu = 0, p\sigma)$. For this reason, it is sufficient to analyze how the regression error varies for different values of p with $\sigma = 1$,

$$\tilde{Z}_M^{(p)} = \frac{1}{M}\sum_{i=1}^{M} |r_i|^p = \frac{1}{M}\sum_{i=1}^{M} \exp[pY_i], \quad Y_i \sim \mathcal{N}(0,1)$$

as a function of p and M.

To analyze the behavior of $\tilde{Z}_M^{(p)}$ we use the results derived in the previous section. In contrast to the Boltzmann factors that appear in the computation of the partition function of the REM (Derrida 1981), the parameters of the distribution of residuals are independent of M. Therefore one has to consider an appropriate rescaling. The asymptotic limit in which results are valid is $p \to \infty$ and $M \to \infty$ with $\log M/p^2 \to$ constant.

To carry out the analysis in this limit, one first computes the probability of finding samples in the interval $[K\varepsilon, K(\varepsilon+\delta)]$ to leading order in the exponent. From the definition of the given in (11) the entropy function for lognormal i.i.d. random variables is

$$s_a(y) = \log 2 - \frac{Ky^2}{2\ p^2} \tag{16}$$

The range of admissible values of y is restricted to the interval $|y| \leq \sqrt{\frac{2\ p^2 \log 2}{K}}$ so that the entropy function $s_a(y)$ is non negative. Depending on the location of the maximum of the exponent in Prop. 1 one can distinguish two regimes: The first one appears for small samples $M < M_c(p)$, where the critical sample size is $M_c(p) = e^{p^2/2}$. Asymptotically, the dominant contribution to the estimator (16) comes from the upper bound of the interval $y_u = \sqrt{\frac{2\ p^2 \log 2}{K}}$. The second regime corresponds to large samples $M > M_c(p)$, in which the dominant contribution comes from the local maximum of the exponent $y^* = \dfrac{p^2}{K}$. In this regime, and in the limit of large M and p with $\log M/p^2$ constant, the sample average is close to the saddle point estimate of the pth moment.

The change of behavior occurs in $y^* = y_u$ and it corresponds to a second order phase transition in the asymptotic behavior of $\tilde{Z}_M^{(p)}$

$$\tilde{Z}_M^{(p)} \doteq \begin{cases} \exp\left[\frac{p^2}{2}\right], \lambda(M,p) > 1 \\ \\ \exp\left[\sqrt{2\ p^2 \log M} - \log M\right], \lambda(M,p) < 1, \end{cases} \tag{17}$$

where

$$\lambda(M,p) = \frac{2 \log M}{p^2} \tag{18}$$

The transition is marked by a discontinuity in the second derivative of $\mathbb{E}\left[\log \tilde{Z}_M\right]$ at the transition point in the limit $p \to \infty$ and $M \to \infty$ with $\log M/p^2 \to$ constant. As in standard phase transitions, there is no mathematical discontinuity for finite M and p.

p'	$\sigma = 1$	$\sigma = 2$
1	3	8
2	8	2981
3	91	65.65×10^6
4	2981	78.96×10^{12}
5	268338	51.84×10^{20}

Figure 1: Different values of the critical sample size $M_c(p', \sigma)$ as a function of the order of the regression p' and the parameter σ such that $p = \sigma p'$.

Nevertheless the change of behavior of the sample estimates is accurate described by (17) for sufficiently large values of these parameters. In particular, for a sample of size M, there is a critical value $p_c(M) =$

$\sqrt{2\log M}$, such that, for values of $p > p_c(M)$ the sample average ceases to be an accurate approximation of the expected value of the lognormal random variable. Alternatively for a fixed value p, the critical sample size is

$$M_c(p) = e^{\frac{p^2}{2}} \tag{19}$$

From (17) one can explicitly compute B_M in the regime $\lambda(M,p) < 1$

$$B_M(p) = \frac{p^2}{2} - \sqrt{2\ p^2 \log M} + \log M. \tag{20}$$

3 EXPERIMENTS

In this section we investigate the dependence of the bias in the logarithm of the the L_p-norm error estimate as a function of sample size for a real-world problem in which the residuals of a linear model are approximately lognormal. The results of this investigation are used to illustrate the validity of the analysis of this bias using the correspondence to the random energy model carried out in the previous section. There are numerous studies in which the presence of heavy-tailed distributions in different Internet-based phenomena is discussed (Downey 2005b). To illustrate the analysis presented in the previous section we consider the distribution of inter-arrival times between subsequent Internet tweets. The dataset analyzed consists of 470K geolocalized tweets using the public Twitter API. To eliminate noisy data produced by bots and automated accounts, all tweets with only one location in our dataset and those that change their location faster than the speed of sound are removed. To extract robust information of the temporal patterns of each user, we keep the 200 most active users in our dataset. This filtering reduces the number of tweets to around 30K. We then study the times between consecutive tweeting actions. For each of user, the inter-tweet times are normalized using the mean and standard deviation of all the tweets from that user. Empirically, one observes that inter-tweet times τ_n are reasonably well approximated using a lognormal distribution. The time series of inter-tweet times exhibits low values for the autocorrelations at delay one (0.06) and even smaller for longer delays. Therefore, residuals r_n are also reasonably well approximated by a lognormal distribution.

The protocol for the experiments is as follows: Let τ_n be the time between the $(n-1)$-th and the n-th tweets. A linear autoregressive model of order one is fitted to the inter-event times

$$\tau_n = a\tau_{n-1} + \varepsilon_n, \quad n = 1,2,\ldots \tag{21}$$

The optimization problem in a is solved using Golden Section Search algorithm provided by the Python's library for scientific computation SciPy (Kiefer 1953, Jones, Oliphant, Peterson, et al. 2001). Once the optimal value of $a = \widehat{a}$ is found, the residuals $r_n = \tau_n - \widehat{a}\tau_{n-1}$ are computed. We then make estimates of the L_p-norm error from a sample of size M for different values of M.

$$\varepsilon_T(\widehat{a};p,M) = \frac{1}{M}\sum_{n=1}^{M} |r_n|^p. \tag{22}$$

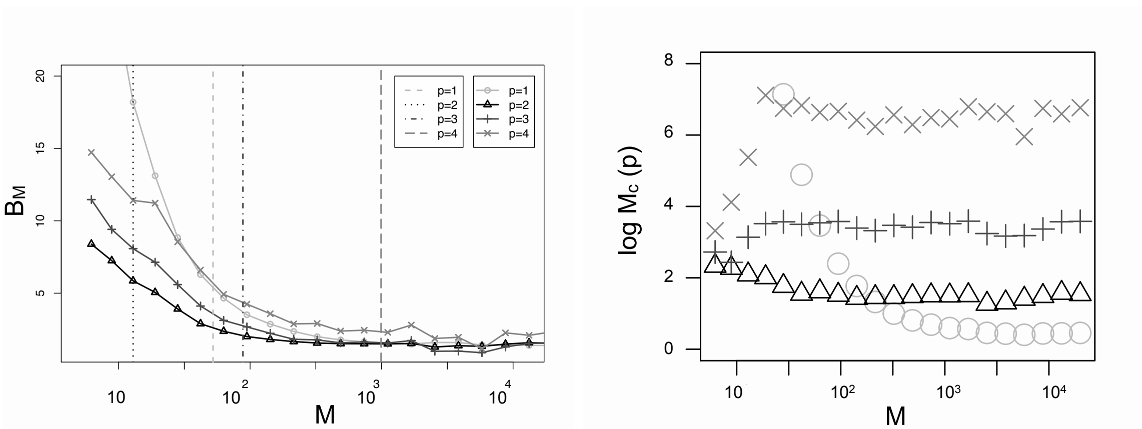
To determine the minimal sample size that should be used to compute reliable estimates of $\varepsilon_T(a;p,M)$ a lognormal distribution is fitted to the set of model residuals $\{r_n\}_{n=1}^{M}$ using Maximum Likelihood Estimation (MLE). For $M < 3000$, the estimates are repeated for different subsamples of size M and then averaged. The value of $\widehat{a}$ that minimizes the error function is influenced by both the sample size M and the order p. In consequence, different estimates of σ are obtained when different values of M and p are considered. The bias B_M also becomes a function of these parameters,

$$B_M(p,\widehat{\sigma}) = \log E[r_n](\widehat{\sigma}) - \log \varepsilon_T(\widehat{\sigma};p,M) \tag{23}$$

The behavior of $B_M(p,\widehat{\sigma})$ as a function of M is depicted in figure 2a. As expected the bias B_M decreases as M increases. This is an intuitive result because larger samples should provide more accurate estimates of the error.

The vertical lines in figure 2a mark the values estimated for the critical sample size (the size above which the bias of the logarithm L_p-norm estimator becomes negligible) for different values of p

$$M_c(p,\widehat{\sigma}) = \exp\left\{\frac{1}{2}p^2\widehat{\sigma}^2\right\}. \tag{24}$$



(a) $B_M(p,\widehat{\sigma})$ for real data.

(b) Critical size $M_c(p)$

Figure 2: (a) Solid dotted lines display the dependence B_M on the sample size for different values of p regressions whereas vertical dashed lines exhibit the transition points predicted by (17). b) Evolution of the critical sample size $M_c(p)$ as a function of M for different values of p.

Since the parameters of the lognormal fit to the residuals change with the sample size, the value of these thresholds is determined in a self-consistent manner: $M_c(p,\widehat{\sigma})$ is such that the value of σ estimated from a sample of that size coincides $\widehat{\sigma}$. This self-consistent procedure is described in Algorithm 1.

Algorithm 1 Calculate M_p in L_p regression

$P \leftarrow [1,2,...,\max_p]$
$ts \leftarrow$ load(data)
for p in P **do**
  $X \leftarrow ts[1:\text{length}(ts)-1]$
  $Y \leftarrow ts[2:\text{length}(ts)]$
  $error, residuals \leftarrow$ lp.minimize(model,X,Y)
  $\mu, \sigma \leftarrow$ lognormal.fit($residuals$)
  $min\_size \leftarrow \exp\left[(\sigma p)^2/2\right]$
  if $(\text{length}(ts)-1) \leq min\_size$ **then**
    $converged \leftarrow$ True
  else
    $converged \leftarrow$ False
  end if
end for

Figure 2b displays the dependence of the values estimated for the critical sample size as a function of M, for different values of p. The results for large M are consistent with the behavior that should be expected: The number of instances needed to obtain reliable L_p norm estimates increases in a non-linear manner for increasing p. The values of the critical sample sizes estimated for L_1 and L_2-norm regression are fairly small. This means that we are not in the asymptotic regime and that the estimate (24), which is derived in the limit $M \to \infty$ $p\sigma \to \infty$ and $\log M/(p\sigma)^2 \to$ constant, can be inaccurate. In contrast, the estimates for L_3 and L_4 are fairly accurate, as illustrated by the results in 2a. From this figure one can also observe that there is a non-zero asymptotic bias. This is due to the fact that the residuals observed actually show deviations from the lognormal distribution. To remove this effect, we generate a synthetic regression problem simulating the process

$$\tau_n = a\ \tau_{n-1} + \varepsilon_n,\ \ \tau_1 = 1, \tag{25}$$

where $\varepsilon_n = \exp[\sigma Y_n]$ and Y_n is a standard Gaussian random variable. To make the simulation closer to the inter-tweet data, we use $\sigma = 0.95$ and $a = 0.1$ in the simulation. The protocol for the estimation of B_M and $M_c(p, \widehat{\sigma})$ is the same as in the previous set of experiments. Figure 3 displays the dependence of $B_M(p, \widehat{\sigma})$ with M. The estimates of the critical sizes $M_c(p, \widehat{\sigma})$ are marked as vertical lines for $p = 2, 3, 4$. In these simulations, for a fixed M the bias becomes larger as the value of p increases. One also observes that the values of the critical sample size are ordered $M_c(p+1, \widehat{\sigma}) > M_c(p, \widehat{\sigma})$. Also in this case, the assumption that we are in the asymptotic regime holds approximately only for $p = 3, 4$. Finally, the value of the bias tends to zero as $M \to \infty$.

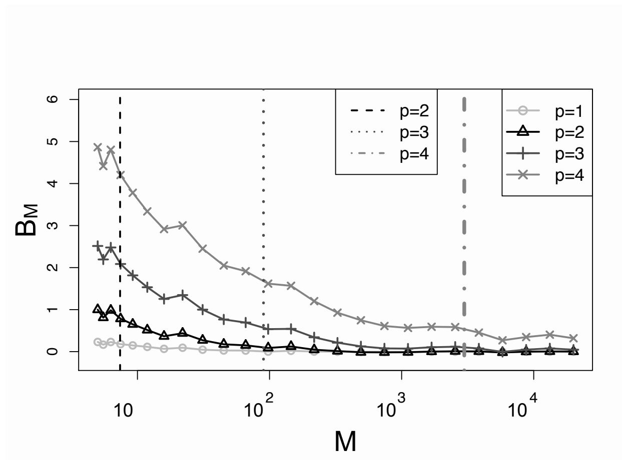


Figure 3: The solid lines display the dependence of $B_M(p, \widehat{\sigma})$ on M for different values of p. The dashed lines indicate the transition points predicted by (17)

4 CONCLUSIONS

In this work we have considered the problem of determining the minimum sample size needed to obtain robust estimates of the L_p-norm error in linear regression problems. To this end we have adapted the techniques used to analyze the random energy model (REM), a model for disordered solids introduced in (Derrida 1981). Taking advantage of the correspondence between the computation of the partition function in a REM model with M different configurations and the estimation of the L_p-norm error from an i.i.d. sample of size M, we show that two different regimes should be expected: For sample sizes smaller than a critical size $M \ll M_c(p, \sigma)$ the typical value of the L_p-norm error in a particular instance of the regression task is very different from the expected value. This difference between typical and average behavior disappears for $M \gg M_c(p, \sigma)$. In the limit $M \to \infty$ and $p \to \infty$ with $\log M/p^2 \to$ constant, a phase transition occurs between these two regimes. The transition is analyzed in terms of the bias of the *logarithm* of the empirical estimate of the L_p-norm. A small bias in this quantity signals that typical and average values of the estimated parameters are close to each other. A large bias in the logarithm indicates that the average behavior is different from the one typically observed, which makes the estimator unreliable in individual instances of the regression problem. Finally, the validity of this analysis is illustrated in experiments with data from a real-world application (the time-series of inter-tweet intervals) and in simulated data.

REFERENCES

Aitchison, J., and J. A. Calvert Brown. 1957. *The Lognormal Distribution, with Special Reference to Its Use in Economics*. Cambridge: Cambridge University Press.

Angeletti, F., E. Bertin, and P. Abry. 2012. “Critical moment definition and estimation, for finite size observation of log-exponential-power law random variables”. *Signal Processing* 92 (12): 2848 – 2865.

Bender, C., and S. Orszag. 1978. *Advanced mathematical methods for scientists and engineers: Asymptotic methods and perturbation theory*. Springer.

Bovier, A., I. Kurkova, and M. Lowe. 2002. “Fluctuations of the free energy in the REM and the p-spin SK models”. *Annals of Probability* 30 (2): 605–651.

Crovella, M., and L. Lipsky. 1997. “Long-lasting transient conditions in simulations with heavy-tailed workloads”.

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

Derrida, B. 1981, Sep. “Random-energy model: An exactly solvable model of disordered systems”. *Phys. Rev. B* 24:2613–2626.

Dorlas, T., and J. Wedagedera. 2001. “Large Deviations Theory and the random energy model”. *International Journal of Modern Physics B* 15:1–15.

Downey, A. B. 2001. “The Structural Cause of File Size Distributions”. In *MASCOTS ’01: Proceedings of the Ninth International Symposium in Modeling, Analysis and Simulation of Computer and Telecommunication Systems, 15-18 August 2001, Cincinnati, OH, USA*, 361–370. Washington, DC, USA: IEEE Computer Society.

Downey, A. B. 2005a. “Lognormal and Pareto distributions in the Internet”. *Computer Communications* 28:790–801.

Downey, A. B. 2005b, May. “Lognormal and Pareto distributions in the Internet”. *Comput. Commun.* 28 (7): 790–801.

Foss, S., D. Korshunov, and S. Zachary. 2011. *An Introduction to Heavy-Tailed and Subexponential Distributions*. Springer.

Frachot, A., O. Moudoulaud, and T. Roncalli. 2003. “Loss Distribution Approach in Practice”. In *The Basel Handbook: A Guide for Financial Practitioners*, edited by M. Ong. London: Risk Books.

E. Jones and T. Oliphant and P. Peterson and others 2001. “SciPy: Open source scientific tools for Python”.

Kagan, A., and S. Nagaev. 2001. “How many moments can be estimated from a large sample?”. *Statistics & Probability Letters* 55 (1): 99–105.

Kiefer, J. 1953. “Sequential minimax search for a maximum”. *Proceedings of the American Mathematical Society* 4 (3): 502–506.

Mézard, M., and A. Montanari. 2009. *Information, physics, and computation*. Oxford University Press.

Mitzenmacher, M. 2004. “A Brief History of Generative Models for Power Law and Lognormal Distributions”. *Internet Mathematics* 1 (2): 226–251.

Money, A., J. Affleck-Graves, M. Hart, and G. Barr. 1982. “The linear regression model: Lp norm estimation and the choice of p”. *Communications in Statistics - Simulation and Computation* 11 (1): 89–109.

Nadarajah, S. 2008. “A Review of Results on Sums of Random Variables”. *Acta Applicandae Mathematicae* 103:131–140.

Naruse, K., and M. Kubo. 2006. "Lognormal Distribution of BBS Articles and its Social and Generative Mechanism". *IEEE / WIC / ACM International Conference on Web Intelligence*:103–112.

N.B. Metha, A.F. Molisch, J. Z. 2007. "Approximating a Sum of Random Variables with a lognormal". *IEEE Transactions on Wireless Communications* 6:2690–2699.

Nyquist, H. 1983. "The optimal Lp norm estimator in linear regression models". *Communications in Statistics - Theory and Methods* 12 (21): 2511–2524.

Suárez, A., R. Silbey, and I. Oppenheim. 2012, May. "Phase transition in the Jarzynski estimator of free energy differences". *Phys. Rev. E* 85:051108.

Touchette, H. 2009, August. "The large deviation approach to statistical mechanics". *Physics Reports* 478 (1-3): 1–69.

Zhang, Q., and S. Song. 2008. "A Systematic Procedure for Accurately Approximating lognormal Sum Distributions". *IEEE Transactions on Vehicular Technology* 57:663 – 666.

Zuckerman, D., and T. Woolf. 2004. "Systematic Finite-Sampling Inaccuracy in Free Energy Differences and Other Nonlinear Quantities". *Journal of Statistical Physics* 114:1303–1323.

AUTHOR BIOGRAPHIES

ALEJANDRO LLORENTE is a Ph.D. student at Universidad Carlos III de Madrid and also holds a position as researcher at Instituto de Ingeniera del Conocimiento in Madrid (Spain). His research interests include modeling of dynamic processes on social network, analysis of human mobility, machine learning and the study of non-standard convergence of estimators. His email address is alejandro.llorente@iic.uam.es.

ALBERTO SUÁREZ is a professor in the Computer Science Department of the Universidad Autnoma de Madrid (Spain). He has worked on relaxation theory in condensed media, stochastic and thermodynamic theories of nonequilibrium systems, lattice-gas automata, and automatic induction from data. His current research interests include machine learning, quantitative and computational finance, time series analysis and information processing in the presence of noise. His email address is alberto.suarez@uam.es