

Estimation of Transformations for Microarray Data Using Maximum Likelihood and Related Methods

Blythe Durbin, Department of Statistics, UC Davis, Davis, CA 95616 *
David M. Rocke, Department of Applied Science, UC Davis, Davis, CA 95616

September 9, 2002

Abstract

Motivation and Results Durbin et al (2002), Huber et al (2002) and Munson (2001) independently introduced a family of transformations (the generalized-log family) which stabilizes the variance of microarray data up to the first order. We introduce a method for estimating the transformation parameter in tandem with a linear model based on the procedure outlined in Box and Cox (1964). We also discuss means of finding transformations within the generalized-log family which are optimal under other criteria, such as minimum residual skewness and minimum mean-variance dependency.

Contact. bpdurbin@wald.ucdavis.edu

Keywords. cDNA array, microarray, statistical analysis, transformation, ANOVA normalization.

1 Introduction

Many traditional statistical methodologies, such as regression or ANOVA, are based on the assumptions that the data are normally distributed (or at least symmetrically distributed), with constant variance not depending on the mean of the data. If these assumptions are violated, the statistician may choose either to develop some new statistical technique which accounts for the specific ways in which the data fail to comply with the assumptions, or to transform the data. Where possible, data transformation is generally the easier of these two options (see Box and Cox, 1964, and Atkinson, 1985).

*To whom correspondence should be addressed.

Data from gene-expression microarrays, which allow measurement of the expression of thousands of genes simultaneously, can yield invaluable information about biology through statistical analysis. However, microarray data fail rather dramatically to conform to the canonical assumptions required for analysis by standard techniques. Rocke and Durbin (2001) demonstrate that the measured expression levels from microarray data can be modeled as

$$y = \alpha + \mu e^\eta + \varepsilon \quad (1)$$

where y is the measured raw expression level for a single color, α is the mean background noise, μ is the true expression level, and η and ε are normally-distributed error terms with mean 0 and variance σ_η^2 and σ_ε^2 , respectively. This model also works well for Affymetrix GeneChip arrays either applied to the PM-MM data or to individual oligos.

The variance of y under this model is

$$\text{Var}(y) = \mu^2 S_\eta^2 + \sigma_\varepsilon^2, \quad (2)$$

where $S_\eta^2 = e^{\sigma_\eta^2}(e^{\sigma_\eta^2} - 1)$. In Durbin et al. (2002), Huber et al. (2002), and Munson (2001) it was shown that for a random variable z satisfying $V(z) = a^2 + b^2 \mu^2$, with $E(y) = \mu$, there is a transformation that stabilizes the variance to the first order. There are several equivalent ways of writing this transformation, but we will use

$$h_\lambda(z) = \ln(z + \sqrt{z^2 + \lambda}),$$

where $\lambda = a^2/b^2 = \sigma_\varepsilon^2/S_\eta^2$ and $z = y - \alpha$ or $y - \hat{\alpha}$. (Use of z rather than y presumes that any requisite background correction and normalization have already been applied to the data so that, to the first order, $E(z) = \mu$. The specific method normalization method used is left to the discretion of the reader.) This transformation converges to $\ln(z)$ for large z (up to an additive constant which does not affect the strength of the transformation), and is approximately linear at 0 (Durbin et al. 2002). We shall refer to this transformation as the generalized-log transformation, as in Munson (2001), as the log transformation is a special case of this family for $\lambda = 0$. The inverse transformation is

$$f_\lambda^{-1}(w) = (e^z - \lambda e^{-z})/2.$$

Both f_λ and its inverse are monotonic functions, defined for all values of z and w , with derivatives of all orders.

When transforming data from two-color arrays or from complex multi-array experiments, the closed form expression for the transformation parameter shown in (1) is less useful than in the single color, single array case. Even data from different colors on the same two-color array might have different estimated values for the model parameters σ_η^2 and σ_ε^2 , which makes it unclear exactly how we should obtain the transformation parameter. Pooling of data from different sources in order to estimate parameters could work well for some designs, but is not very flexible. An estimation method which specifically accounts for the structure of the data would be useful in these situations.

One such approach is to fit a linear model to the data while simultaneously estimating the transformation parameter via maximum likelihood, as was done in Box and Cox (1964). The linear model structure will allow us to account for the different sources of variation in the data, such as variation between arrays, between replicated spots on the same array, and between colors on the same array, in our estimation of the transformation parameter. Furthermore, the linear model, fit to appropriately transformed data, can itself be a useful analysis tool. An example of such an analysis would be the ANOVA normalization method for microarray data developed in Kerr et al (2000).

2 Maximum-Likelihood Estimation

The maximum-likelihood estimation of the linear model and transformation parameters can be conducted essentially as in Box and Cox (1964), with the key distinction being that we shall search for an optimal transformation within the family of generalized log transformations, as in (1), rather than among the power transformations.

The procedure outlined in Box and Cox (1964) is as follows: Suppose that there exists some λ such that the transformed observations $\{h_{i,\lambda}\}$ have independent normal distributions with linear mean structure and constant variance σ^2 . That is, suppose there exists lambda such that

$$\mathbf{h}_\lambda = (h_{1,\lambda}, \dots, h_{n,\lambda})^\top = \mathbf{X}\beta + \varepsilon \quad (3)$$

where n is the number of observations in the data set, \mathbf{X} is the design matrix from the linear model, β is a fixed vector of unknown linear model parameters, and $\varepsilon \sim N(\mathbf{0}, \sigma^2 \mathbf{I})$. The likelihood of the untransformed observations $\{z_i\}$ may therefore be written in terms of the transformed observations $\{h_{i,\lambda}\}$ as

$$L(\beta, \sigma^2, \lambda; \mathbf{z}) = \frac{1}{(2\pi)^{n/2} \sigma^n} \exp\left(-\frac{(\mathbf{h}_\lambda - \mathbf{X}\beta)^\top (\mathbf{h}_\lambda - \mathbf{X}\beta)}{2\sigma^2}\right) \mathbf{J}(\lambda), \quad (4)$$

where

$$\mathbf{J}(\lambda) = \prod_{i=1}^n \left| \frac{dh_{i,\lambda}}{dz_i} \right| \quad (5)$$

$$= \prod_{i=1}^n 1/\sqrt{z_i^2 + \lambda}. \quad (6)$$

Fixing λ and maximizing (4) over β and σ^2 , we arrive at the usual maximum-likelihood estimates for these parameters (presuming \mathbf{X} is of full rank):

$$\hat{\beta}(\lambda) = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{h}_\lambda$$

and

$$\hat{\sigma}^2(\lambda) = \mathbf{h}_\lambda^\top (\mathbf{I} - \mathbf{H}) \mathbf{h}_\lambda / n, \quad (7)$$

where $\mathbf{H} = \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top$ (Box and Cox (1964)).

Substituting $\hat{\beta}(\lambda)$ and $\hat{\sigma}^2(\lambda)$ into $\ln(L(\beta, \sigma^2, \lambda; \mathbf{z}))$, we obtain the partially-maximized log likelihood

$$l_{\max}(\lambda) = -\frac{n}{2} \ln \hat{\sigma}^2(\lambda) + \ln \mathbf{J}(\lambda). \quad (8)$$

For added simplicity, we may scale the transformed observations by $\mathbf{J}^{1/n}(\lambda)$, defining

$$\begin{aligned} w_\lambda &= h_\lambda / \mathbf{J}^{1/n}(\lambda) \\ &= \ln(z + \sqrt{z^2 + \lambda}) \text{gm}(\sqrt{z^2 + \lambda}), \end{aligned}$$

where

$$\text{gm}(\sqrt{z^2 + \lambda}) = \left(\prod_{i=1}^n \sqrt{z_i^2 + \lambda} \right)^{1/n}.$$

The partially-maximized log likelihood of \mathbf{z} in terms of \mathbf{w}_λ can then be closely approximated by

$$\begin{aligned} l_{\max}(\lambda) &= -\frac{n}{2} \ln \hat{\sigma}^2(\lambda) \\ &= -\frac{n}{2} \text{SSE}(\lambda)/n \end{aligned}$$

which depends on the data only through $\text{SSE}(\lambda)$, the error sum of squares from the linear model fit to the transformed data. (The approximate nature of the log likelihood arises when we choose to ignore the variability of the Jacobian, which should be quite minimal given the size of most microarray data sets. See Chapter 6 of Ferguson (1996) for further explanation.) This new, partially-maximized log likelihood is a monotone decreasing function of the error sum of squares from the linear model for fixed λ , $\text{SSE}(\lambda)$, so we may find the MLE of λ , $\hat{\lambda}$, simply by minimizing $\text{SSE}(\lambda)$ (Box and Cox (1964)). This may be accomplished by plotting the error sum of squares as a function of λ , or via numerical optimization methods. Estimates of β and σ^2 on the scale of the transformed data without the Jacobian correction may be obtained by fitting the linear model again using the MLE, $\hat{\lambda}$, as the transformation parameter or by multiplying $\hat{\beta}$ by $\mathbf{J}^{1/n}(\hat{\lambda})$ and multiplying $\hat{\sigma}^2$ by $\mathbf{J}^{2/n}(\hat{\lambda})$.

2.1 Examples

We illustrate this estimation method using two example data sets, one from a two-color cDNA-array experiment and one from an experiment conducted using Affymetrix oligonucleotide arrays. The first example comes from a toxicology experiment by Bartosiewicz et al (2002) in which male Swiss Webster mice were injected with a toxin. We shall focus on a single slide from this experiment. For this array, the treatment mouse was injected with 0.15 mg/kg of β -naphthoflavone dissolved in 10 ml/kg of corn oil, and the control mouse was injected with 10

ml/kg of corn oil. mRNA from the livers of these mice was reverse transcribed and fluor labelled, with the treatment sample labelled with Cy5 and the control sample labelled with Cy3. The samples were hybridized to a spotted cDNA array on which each of the 138 genes was replicated between 6 and 14 times, resulting in a total of 1008 spots.

For the mouse data, we will model the differences of the transformed control and treatment observations rather than the transformed observations themselves. The difference of the transformed observations from replicate j of gene i , $\Delta h_{\lambda ij}$, can be modeled as

$$\Delta h_{\lambda ij} = \mu_i + \varepsilon_{ij}, \quad (9)$$

where μ_{ij} is a gene effect and ε_{ij} is a normally distributed error term. Notice that (9) is a one-way ANOVA model.

Figure 1 shows the partially-maximized log likelihood for the mouse data as a function of the transformation parameter, λ . The likelihood is maximized at $\lambda = 1.1309 \times 10^9$. An asymptotic 95% confidence interval for the MLE, $\hat{\lambda}$, consists of those values of λ for which

$$l_{\max}(\hat{\lambda}) - l_{\max}(\lambda) < \frac{1}{2} \chi_{1,.05}^2,$$

where $\chi_{1,.05}^2$ is the upper 5% quantile of a χ_1^2 distribution (Box and Cox (1964)). This yields a confidence interval for $\hat{\lambda}$ of $(8.6827 \times 10^8, 1.4741 \times 10^9)$, which is bounded well away from 0 and thus definitively excludes the log transformation (corresponding to $\lambda = 0$).

Figure 2 shows a quantile-quantile plot of the residuals from the linear model (9) fit to the transformed data versus a standard normal distribution. The residuals appear to come from a distribution with heavier tails than a normal distribution. Although the plot appears to exhibit some skewness, this is entirely due to the four observations in the lower left-hand corner. These observations appear to be outliers resulting from phenomena such as dust on the slide, since they all occur in genes which are expressed near background in the control data, and feature a single observation which differs so hugely from the other replicates that it is unlikely to result from actual gene expression. (These observations will be excluded from the analysis of Section 3). Examination of residuals from the linear model appears to facilitate identification of outlying observations, since these outliers were much more obvious in the residuals than they would have been in the raw data.

The second example comes from an experiment using 4 Affymetrix HG_U95 arrays, which is described in Geller et al (2002). In this experiment, a lymphoblastoid cell line from a single autistic child was grown up in four separate flasks. RNA extraction, cDNA synthesis, and in-vitro labelling was conducted separately on each of the 4 samples, and each of the samples was hybridized to a separate array.

For the autism data, we model the transformed perfect match minus mismatch observation from gene j on array i , $h_{\lambda ij}$, as

$$h_{\lambda ij} = \mu_i + \eta_j + \varepsilon_{ij}, \quad (10)$$

where μ_i is a fixed array effect, η_j is a fixed gene effect, and ε_{ij} is a normally distributed error term. Notice that our model is a two-factor ANOVA model without an interaction term. (We cannot fit the interaction term due to the absence of replicated genes, but we would not expect a gene-array interaction effect anyway.)

Figure 3 shows the partially-maximized log likelihood for the autism data as a function of the transformation parameter. The likelihood is maximized at $\lambda = 3873$, and a 95% confidence interval for $\hat{\lambda}$ is (3751, 4000), which, again, excludes the log transformation. Figure 4 shows a quantile-quantile plot of the residuals from the linear model (10). The residuals, again, appear to come from a symmetric distribution with tails heavier than a normal distribution.

3 Other Methods of Estimating the Transformation Parameter

Maximum-likelihood estimation of the transformation parameter in the manner described above in essence simultaneously optimizes constancy of variance, the fit of the transformed residuals to a normal distribution, and the fit to the linear model. In some applications, some of these criteria may be more important than others. For example, for many traditional statistical techniques data that are symmetric are almost as good as data that are normally distributed, and by trying to force the transformed data to fit all of the moments of a normal distribution we may inadvertently compromise those characteristics in which we are most interested. In such cases, we may search within the family of generalized-log transformations for a transformation optimizing the quantity of interest, simply by minimizing the appropriate statistic.

For example, to find a transformation minimizing the skewness of residuals from the linear model, we would look for a transformation for which the skewness coefficient of the residuals is equal to 0. To find a transformation for which the fixed effects in an ANOVA model are the most linear, we would look for a transformation minimizing the F-statistic for the interaction term in the model. (Notice that the two estimation procedures just mentioned both incorporate the linear model structure used in the maximum-likelihood estimation.) To find a transformation minimizing the dependency of the replicate mean on the replicate variance, we would regress the replicate standard deviation of the transformed data on the replicate mean and look for the transformation minimizing the t-statistic for the significance of the slope parameter. These other optimal transformations also provide a means of assessing the quality of the maximum-likelihood estimate of the transformation parameter. If the MLE differs too greatly from the optimal transformation parameter under another criterion, this might be cause for concern.

We illustrate the skewness-minimizing transformation and the transformation minimizing dependency of the replicate mean and variance using the mouse data. Figure 5 shows a plot of the residual skewness coefficient as a function

of the transformation parameter for the mouse data, using residuals from the linear model in (9). For these data, the skewness coefficient is non-monotonic in the transformation parameter, so there are two values of λ for which the skewness coefficient is equal to 0, 2.3655×10^7 and 1.7272×10^8 . A asymptotic 95% confidence interval for the skewness-minimizing transformation consists of those values of λ for which the absolute skewness coefficient is not significant at the 5% level. For a sample of size 1008 the absolute skewness is not statistically significant if it is less than 0.1512, which yields the confidence interval $(1.1621 \times 10^6, 1.6940 \times 10^9)$. The resulting confidence interval is quite large, but notice that it excludes the log transformation (for $\lambda = 0$), indicating that the log transformation significantly skews the residuals. It does, however, include the maximum likelihood transformation, ($\lambda = 1.1309 \times 10^9$), implying the the MLE does provide sufficient symmetry.

Figure 6 shows the t-statistic for the significance of the slope parameter from the regression of the replicate standard deviation on the replicate mean as a function of the transformation parameter. For the mouse data, the t-statistic is equal to 0 at $\lambda = 4.0329 \times 10^9$. An asymptotic 95% confidence interval for the t-minimizing transformation consists of those values of λ for which the t-statistic is not significant at the 5% level. For 136 degrees of freedom (since we have 138 genes and lose 2 degrees of freedom from fitting the regression parameters) the cutoff for significance of the t-statistic is ± 1.9776 , which yields the confidence interval $(2.0172 \times 10^9, 8.1005 \times 10^9)$. This confidence interval excludes both the log transformation and the maximum-likelihood transformation. However, a Wald test to determine if the MLE and the t-minimizing parameters differed significantly (using 1/4th the length of the respective confidence intervals as a crude estimate of the standard deviation of the parameters) yields the test statistic -1.9270 , which is not significant at the 5% level.

Figure 7 shows the replicate standard deviation plotted against the replicate mean for the mouse data transformed using the t-minimizing transformation, $\lambda = 4.0329 \times 10^9$. The plot does not display any obvious mean-variance dependency, indicating that the transformation has effectively stabilized the variance of these data. For purposes of comparison, Figure 8 shows the replicate standard deviation and mean for data transformed using the maximum-likelihood transformation, $\lambda = 1.1309 \times 10^9$. Again, there does not appear to be any obvious mean-variance dependency, indicating that the MLE also provides sufficient variance stabilization. The MLE provides surprisingly good variance-stabilization and symmetrization, especially in light of the fact that the normal likelihood is almost certainly the "wrong" likelihood for the transformed data, given the apparent heavy-tailed distribution of the transformed residuals.

4 Conclusions

The generalized-log transformation of Durbin et al (2002), Huber et al (2002) and Munson (2001) with parameter $\lambda = a^2/b^2$ stabilizes the variance of data where $\text{Var}(z) = a^2 + b^2 E^2(z)$. Maximum-likelihood estimation in the manner of

Box and Cox (1964) can be used to estimate a transformation parameter for data where observations have different values of a and b . This procedure estimates the transformation parameter while simultaneously fitting a linear model to the data.

The maximum-likelihood estimate can be found by finding the transformation minimizing the error sum of squares from the linear model fit to the transformed data. Transformations minimizing residual skewness, mean-variance dependency, and other criteria may be found by minimizing the appropriate statistic. The maximum likelihood estimate appears to perform well compared to transformations specifically minimizing residual skewness and mean-variance dependency, especially in light of the fact that the normal likelihood is a first approximation to the "true" distribution of the transformed data and was chosen primarily for computational convenience. The residuals from linear model fit to the transformed data appear, in fact, to come from a distribution with heavier tails than a normal distribution. Finally, the log transformation was excluded from the confidence intervals for each of the 3 optimality criteria tested (maximum likelihood, minimum residual skewness, and minimum mean-variance dependency). The generalized-log transformation provides a good alternative to the log transformation for use with gene-expression microarray data.

A Numerical Optimization Via Newton's Method

Newton's method provides a means of finding a root of a smooth function. We use this technique to perform numerical minimization of the error sum of squares by finding a root of the first derivative of $SSE(\lambda)$. Plots of the likelihood may be used to confirm that the root found does indeed constitute a global maximum. Denote $\frac{\partial}{\partial \lambda} SSE(\lambda)$ by $SSE'(\lambda)$ and $\frac{\partial^2}{\partial \lambda^2} SSE(\lambda)$ by $SSE''(\lambda)$. A new estimate of λ , $\lambda_{(n+1)}$, may be obtained from the previous estimate, $\lambda_{(n)}$ using

$$\lambda_{(n+1)} = \lambda_{(n)} - \frac{SSE'(\lambda_{(n)})}{SSE''(\lambda_{(n)})}. \quad (11)$$

Convergence is achieved when $|SSE'(\lambda)|$ is less than the predetermined application tolerance.

For the generalized log transformation with parameter λ ,

$$SSE'(\lambda) = 2 \sum_{i=1}^n (w_{\lambda i} - \hat{w}_{\lambda i})(w'_{\lambda i} - \hat{w}'_{\lambda i}), \quad (12)$$

where

$$\hat{w}_{\lambda i} = \mathbf{x}_i^\top \hat{\beta}(\lambda),$$

\mathbf{x}_i^\top is the i^{th} row of the design matrix,

$$\begin{aligned} w'_{\lambda i} &= \frac{\partial}{\partial \lambda} w_{\lambda i} \\ &= [2\{z_i + \sqrt{z_i^2 + \lambda}\}\sqrt{z_i^2 + \lambda}]^{-1} \mathbf{J}^{-1/n}(\lambda) \\ &\quad + \ln(z_i + \sqrt{z_i^2 + \lambda}) \frac{\partial}{\partial \lambda} \mathbf{J}^{-1/n}(\lambda), \end{aligned}$$

where

$$\frac{\partial}{\partial \lambda} \mathbf{J}^{-1/n}(\lambda) = \frac{1}{n} \sum_{i=1}^n \mathbf{J}^{-1/n}(\lambda) / \{2(z_i^2 + \lambda)\},$$

and

$$\begin{aligned} \hat{w}'_{\lambda i} &= \frac{\partial}{\partial \lambda} \hat{w}_{\lambda i} \\ &= \mathbf{x}_i^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{w}'_{\lambda}, \end{aligned}$$

The second derivative of the error sum of squares is

$$\begin{aligned} \text{SSE}''(\lambda) &= 2 \sum_{i=1}^n (w'_{\lambda i} - \hat{w}'_{\lambda i})^2 \\ &\quad + 2 \sum_{i=1}^n (w_{\lambda i} - \hat{w}_{\lambda i})(w''_{\lambda i} - \hat{w}''_{\lambda i}), \end{aligned}$$

where

$$\begin{aligned} w''_{\lambda i} &= \frac{\partial^2}{\partial \lambda^2} w_{\lambda i} \\ &= -\frac{1}{4} \mathbf{J}^{-1/n}(\lambda) (z_i + \sqrt{z_i^2 + \lambda})^{-1} \{z_i^2 + \lambda\}^{-\frac{3}{2}} \\ &\quad - \frac{1}{4} \mathbf{J}^{-1/n}(\lambda) (z_i + \sqrt{z_i^2 + \lambda})^{-2} \{z_i^2 + \lambda\}^{-1} \\ &\quad + (z_i + \sqrt{z_i^2 + \lambda})^{-1} \{z_i^2 + \lambda\}^{-\frac{1}{2}} \frac{\partial}{\partial \lambda} \mathbf{J}^{-1/n} \\ &\quad + \ln(z_i + \sqrt{z_i^2 + \lambda}) \frac{\partial^2}{\partial \lambda^2} \mathbf{J}^{-1/n} \end{aligned}$$

and

$$\begin{aligned} \frac{\partial^2}{\partial \lambda^2} \mathbf{J}^{-1/n}(\lambda) &= \frac{1}{2n} \sum_{i=1}^n \{z_i^2 + \lambda\}^{-1} \frac{\partial}{\partial \lambda} \mathbf{J}^{-1/n}(\lambda) \\ &\quad - \frac{1}{2n} \sum_{i=1}^n \{z_i^2 + \lambda\}^{-2} \mathbf{J}^{-1/n}(\lambda) \end{aligned}$$

and

$$\begin{aligned} \hat{w}''_{\lambda i} &= \frac{\partial^2}{\partial \lambda^2} \hat{w}_{\lambda i} \\ &= \mathbf{x}_i^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{w}''_{\lambda}. \end{aligned}$$

Acknowledgements

The research reported in this paper was supported by grants from the National Science Foundation (ACI 96-19020, and DMS 98-70172) and the National Institute of Environmental Health Sciences, National Institutes of Health (P43 ES04699).

References

- Atkinson, A.C. (1985) *Plots, Transformations, and Regression: An Introduction to Graphical Methods of Diagnostic Regression Analysis* Clarendon Press, Oxford.
- Bartosiewicz, M., Trounstein, M., Barker, D., Johnston, R., and Buckpitt, A. (2000) Development of a toxicological gene array and quantitative assessment of this technology. *Archives of Biochemistry and Biophysics.*, 376, 66–73.
- Box, G.E.P., and Cox, D.R. (1964) An analysis of transformations. *Journal of the Royal Statistical Society, Series B (Methodological)*, 26, 211–252.
- Durbin, B., Hardin, J., Hawkins, D.M., and Rocke, D.M. (2002) A variance-stabilizing transformation for gene-expression microarray data. *Bioinformatics*, 18, 105S–110S.
- Ferguson, T.S. (1996) *A Course in Large Sample Theory* Chapman & Hall, London.
- Geller, S.C., Gregg, J.P., Hagermann, P., and Rocke, D.M. (2002) Transformation and normalization of oligonucleotide microarray data. (manuscript)
- Huber, W., von Heydebreck, A., Sülthmann, H., Poustka, A., and Vingron, M. (2002) Variance stabilization applied to microarray data calibration and to the quantification of differential expression. *Bioinformatics*, 18, 96S–104S.
- Kerr, K., Martin, M., and Churchill, G. (2000) Analysis of variance for gene expression microarray data. *Journal of Computational Biology*, 7, 819–837.
- Munson, P. (2001) A “consistency” test for determining the significance of gene expression changes on replicate samples and two convenient variance-stabilizing transformations. GeneLogic Workshop of Low Level Analysis of Affymetrix GeneChip Data.
- Rocke, D.M., and Durbin, B. (2001) A model for measurement error for gene expression arrays. *Journal of Computational Biology.*, 8, 557–569.

Figure 1: Log Likelihood by Transformation Parameter, Mouse Data

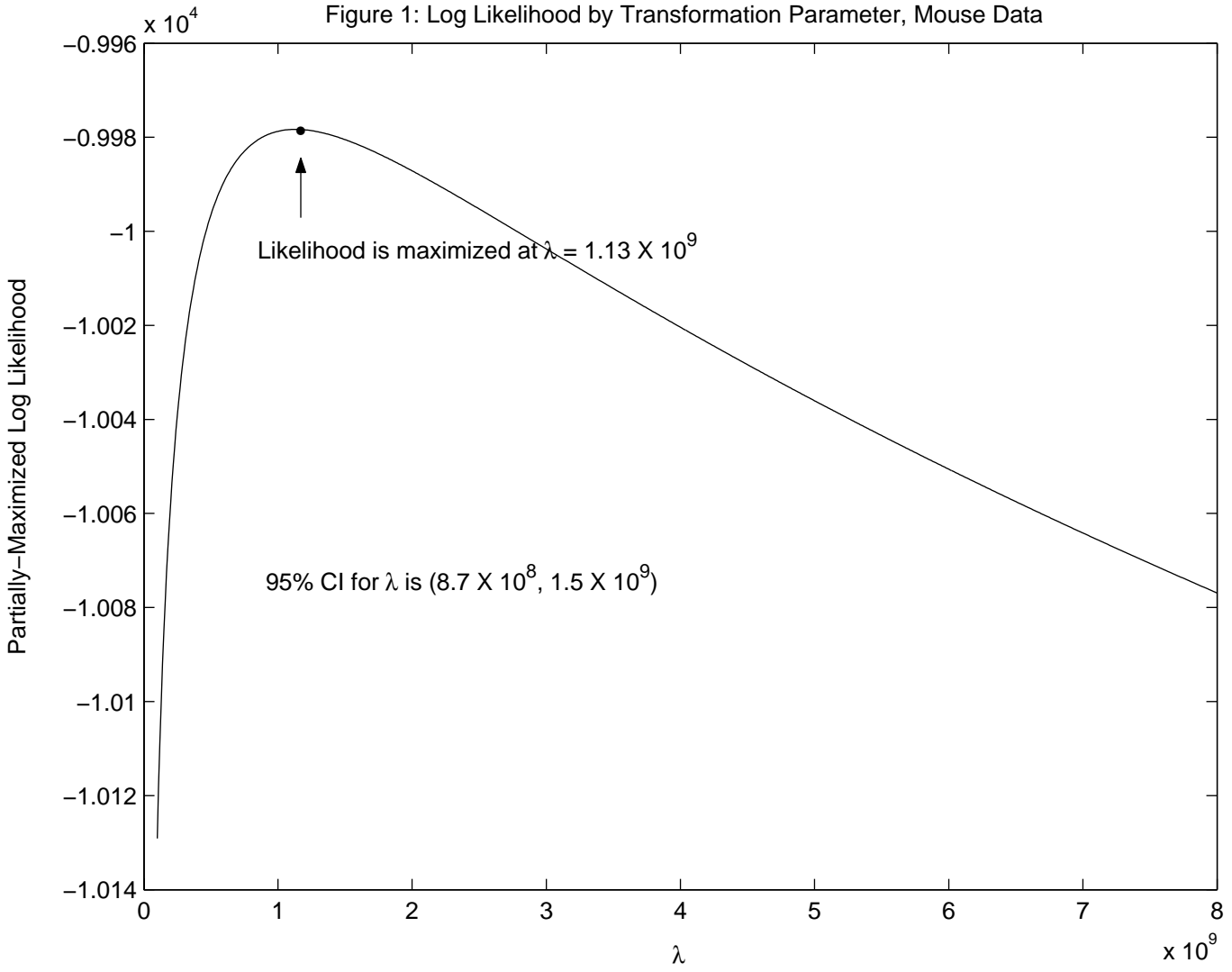


Figure 2: QQ Plot of Residuals vs. Standard Normal, Maximum-Likelihood Transformation, Mouse Data

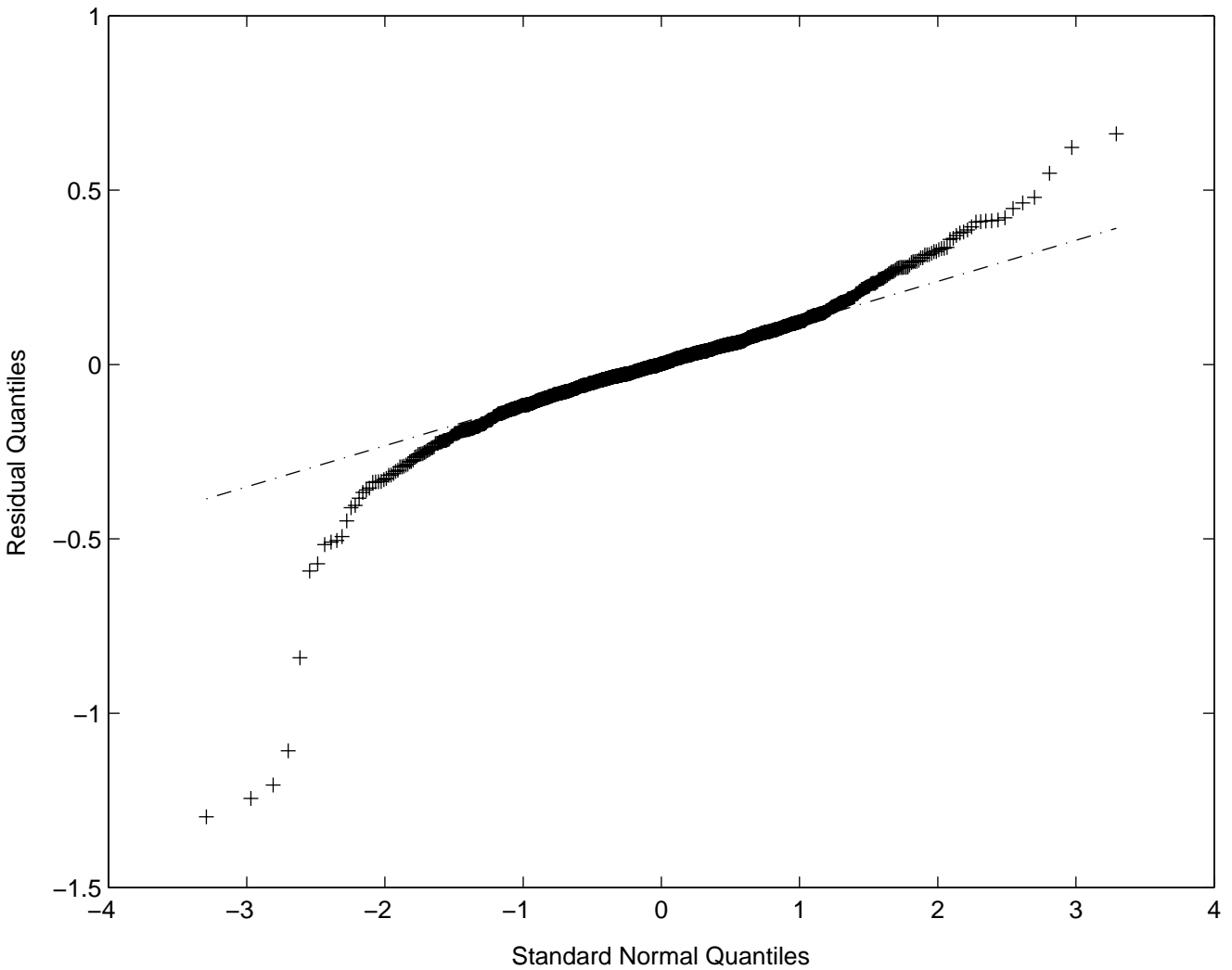


Figure 3: Log Likelihood by Transformation Parameter, Autism Data

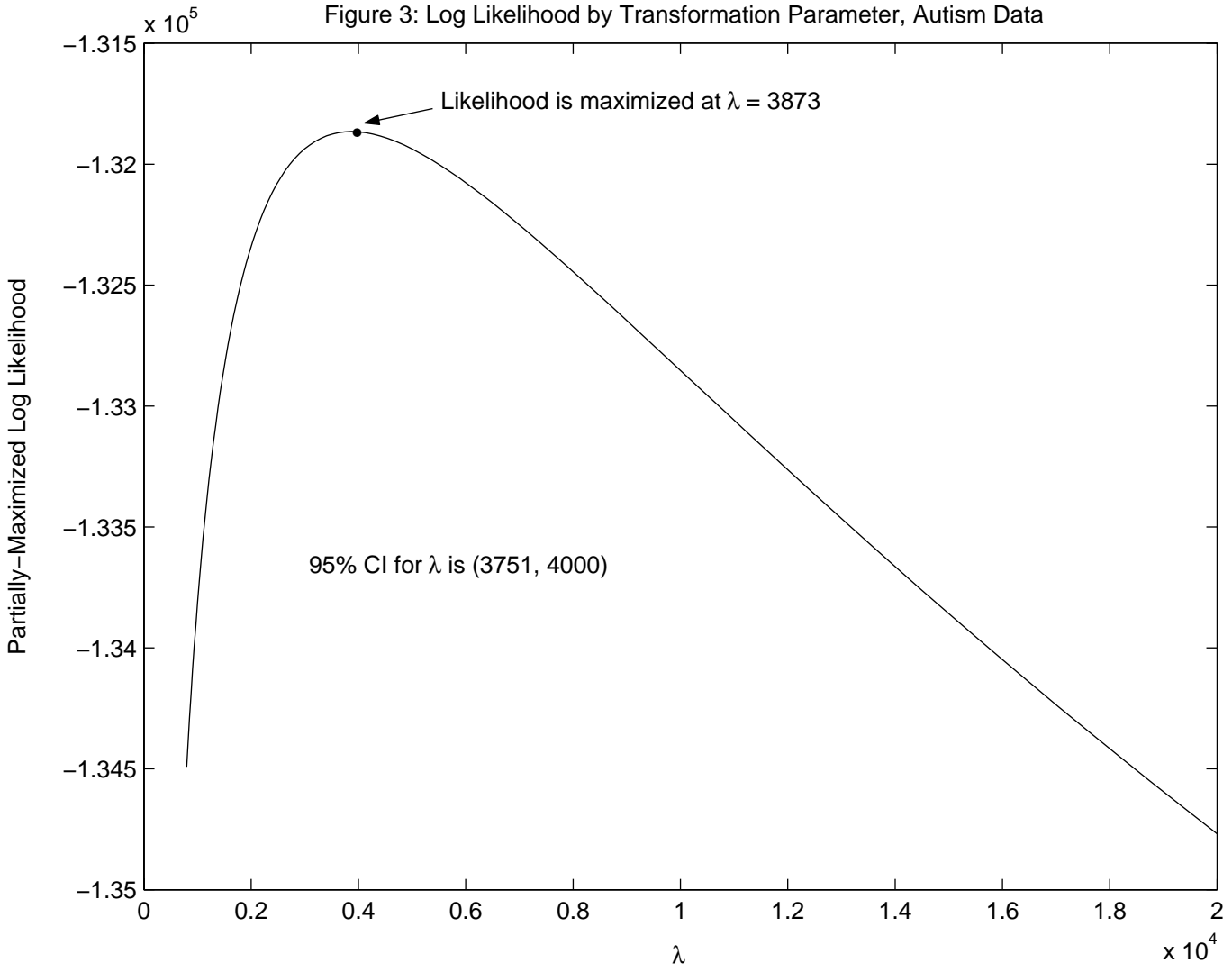


Figure 4: QQ Plot of Residuals from Linear Model vs. Standard Normal, Mouse Data

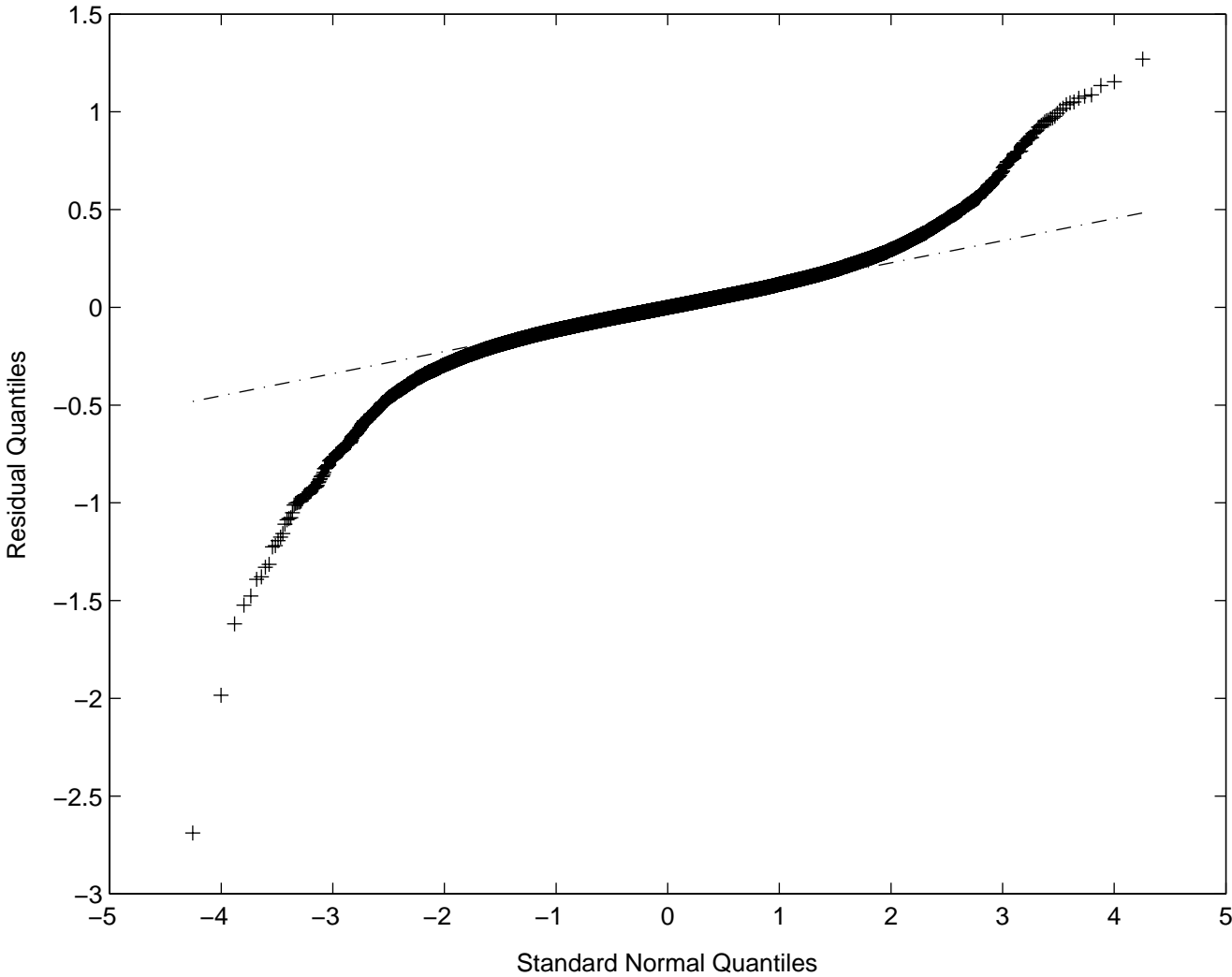


Figure 5: Residual Skewness by Transformation Parameter, Mouse Data

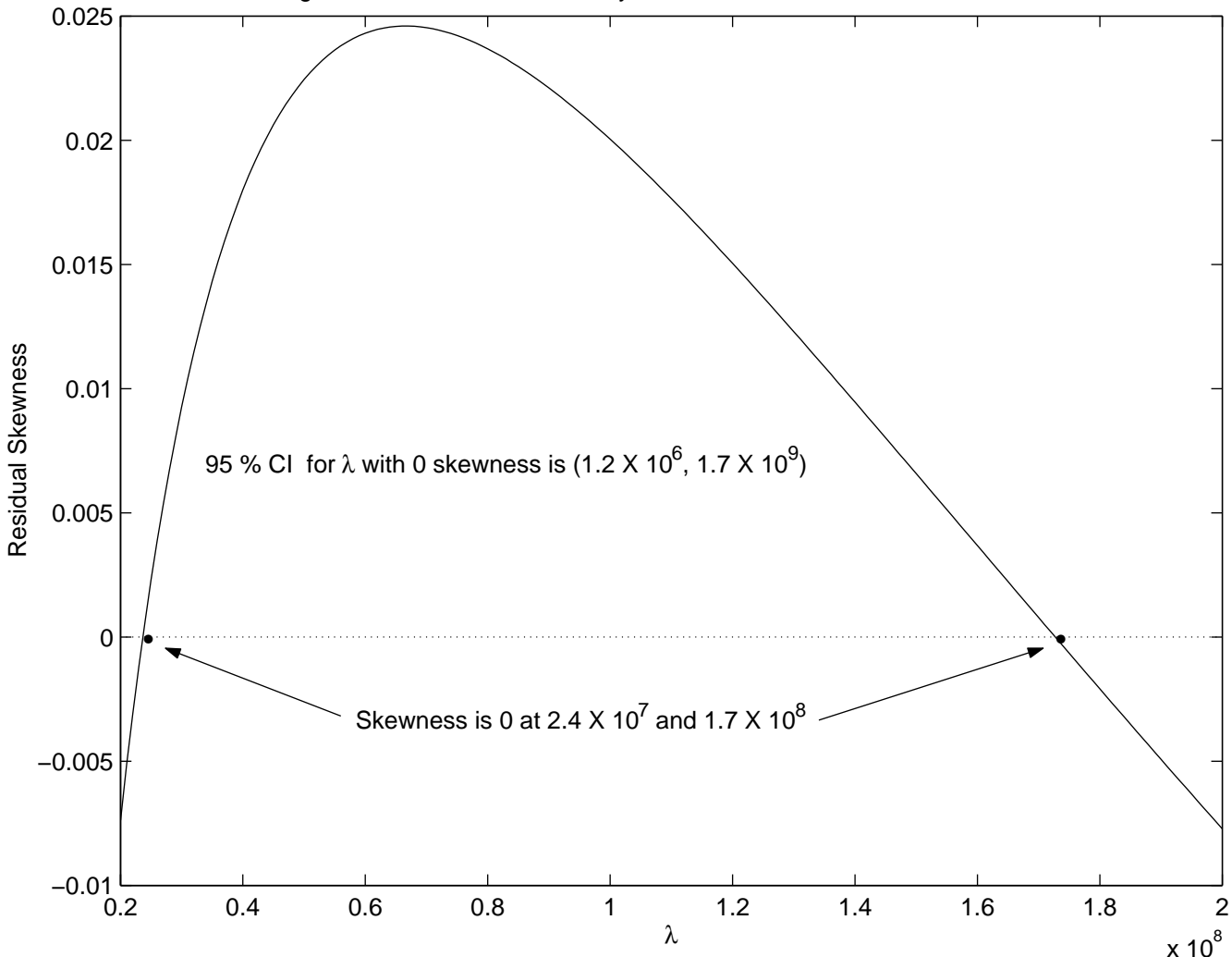


Figure 6: t Statistic for Mean–Variance Dependency by Transformation Parameter, Mouse Data

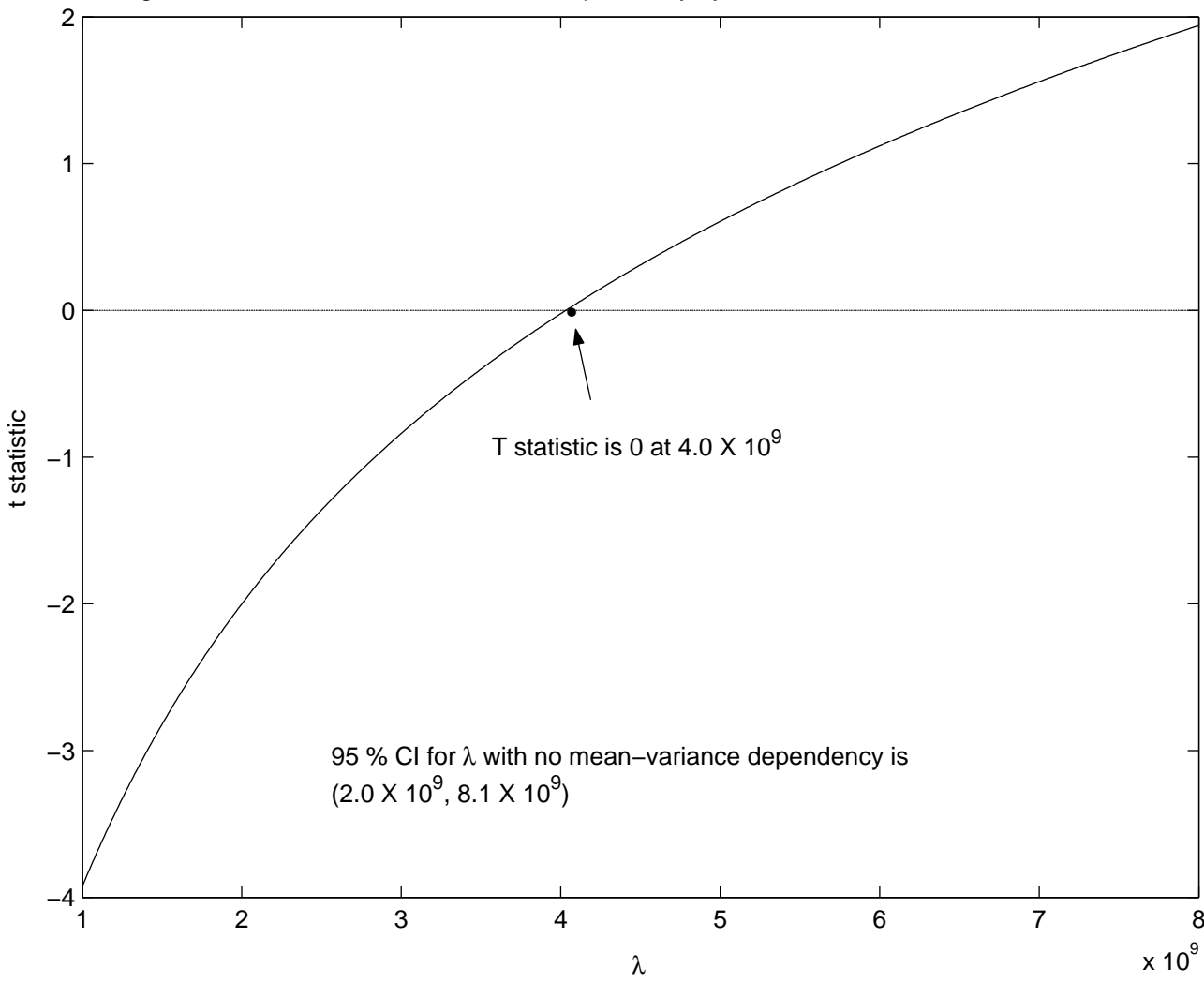


Figure 7: Replicate Mean and Standard Deviation, t-Minimizing Transformation, Mouse Data

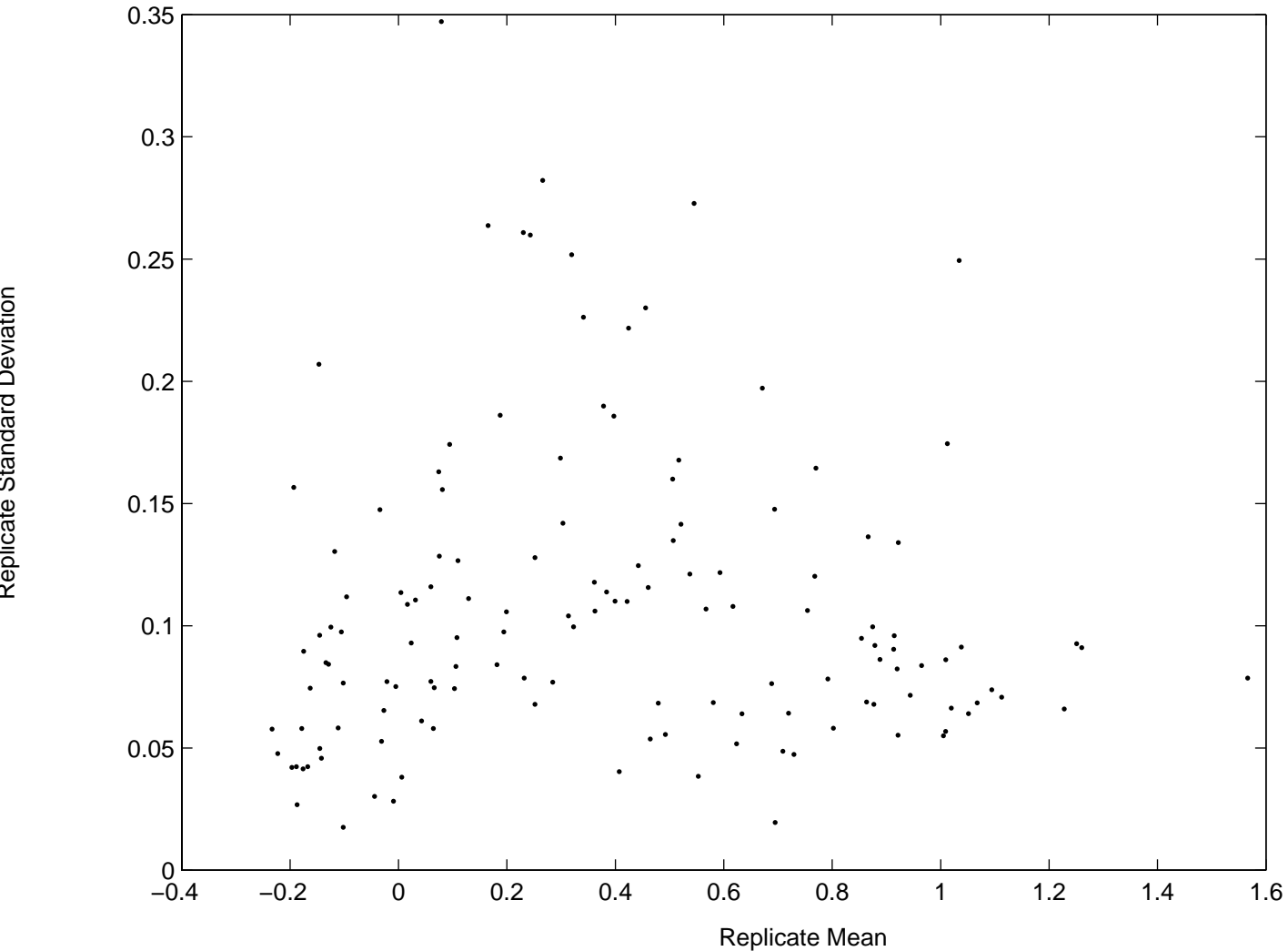


Figure 8: Replicate Mean and Standard Deviation, Maximum-Likelihood Transformation, Mouse Data

