Parameter Estimation for Singular Gaussian Functions

Kyle Kastner
University of Texas-San Antonio
Department of Electrical and Computer Engineering
fps041@utsa.edu

Abstract

This paper shows one technique for parameter estimation of a singular Gaussian function, given only random samples from the peak shape, and the location where these samples were taken. Using Gibbs sampling, the height, mean, and variance of a Gaussian peak can be estimated.

1. Introduction

A common technique in signal processing is to approximate a distorted peak shape with a Gaussian function of form *Eq. 1*

$$Ae^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

(*Eq. 1*)

The ability to find the height, width, and position of a gaussian function is vital in many areas of automated precessing. In this exercise, the parameters $A(height), \mu(mean), \sigma^2(variance)$ of *Eq. 1* are estimated, given a vector of sample values **y**, and a vector of sampling points **p**. Applications of this technique include biomedical measurement analysis, peak detection, and powerline communications.

2. Data Model

The data model used for this analysis data generation was *Eq. 1*, using arbitrary values. Several different values for each these parameters were used during testing, but the resultant graphs in the paper used the parameters $A = 27.50, \mu = 19.15, \sigma^2 = 4.05$.

To generate **n** points for the sampling vector **p**, the standard gaussian sampling form $\mu + \sqrt{\sigma^2} randn(n)$ was used.

For this analysis, $n = 1500$ appeared to give a good approximation of the true distribution, yielding accurate results. Applying the values in **p** to *Eq. 1* yielded the vector of sample values **y**.

3. Posterior Density Derivations

The joint conditional posterior density function for this model was approximated using Gibbs sampling. Therefore, the conditionals required are as follows.

$$p(A|\mu, \sigma^2)$$

(*Eq. 2*)

$$p(\mu|A, \sigma^2)$$

(*Eq. 3*)

$$p(\sigma^2|A, \mu)$$

(*Eq. 4*)

Using the chain rule, it is understood that

$$p(\theta|y) \propto p(y|\theta)p(\theta)$$

(*Eq. 5*)

where $p(\theta)$ is the prior distribution of θ and $p(y|\theta)$ is the likelihood of data given θ.

This means that we can derive the conditional probability of any of the parameters in question, given a likelihood $p(y|\theta)$ and a prior distribution $p(\theta)$. In each of these cases, the non-informative prior was chosen. Likelihood derivations for each parameter follow below.

3.1 Mean

Starting from *Eq. 1*, and using $\theta = \mu$, the non-informative prior $p(\theta)$is derived as 1.

$$p(\theta) \approx \sqrt{E\left[\left(\frac{x-\mu}{\sigma^2}\right)^2\right]}$$

$$= \sqrt{\frac{\sigma^2}{\sigma^4}} \propto 1$$

(*Eq.* 5)

For a vector of observations **y**, the conditional posterior $p(\theta|y)$ for $\theta = \mu$ is approximated by [1]

$$p(\theta|y) = p(\theta)\prod_{i=1}^{n} p(y_i|\theta)$$

$$\propto exp\left(-\frac{1}{2}\left[\frac{1}{\tau_0^2}(\theta-\mu_0)^2 + \frac{1}{\sigma^2}\sum_{i=1}^{n}(y_i-\theta)^2\right]\right)$$

$$p(\theta|y) \approx N(\theta|\bar{y}, \sigma^2/n)$$

(*Eq. 6*)

3.2 Variance

Once again using *Eq. 1*, this time with $\theta = \sigma^2$, the non-informative prior is derived as $\frac{1}{\sigma^2}$ [1]

$$p(\sigma^2) = \sqrt{E\left[\left(\frac{(x-\mu)^2 - \sigma^4}{\sigma^6}\right)^2\right]}$$

$$= \sqrt{\frac{2}{\sigma^2}} = \frac{1}{\sigma^2}$$

(*Eq. 7*)

Using the non-informative prior, the conditional posterior distribution is derived as follows [2]

$$p(y|\theta) = \prod_{m=1}^{M}(2\pi\sigma^2)^{-\frac{M}{2}} e^{\frac{-|y-\mu|^2}{2\sigma^2}}$$

$$\propto (\sigma^2)^{-\frac{M}{2}} e^{-\frac{|y-\mu|^2}{2\sigma^2}}$$

$$\propto IG(\frac{M}{2}, \frac{(y-\mu)^2}{2}$$

(*Eq.* 8)

It is important to note that the term $|y-\mu|^2$here represents a magnitude square, that is, subtract the mean from every value, square the result for each value, then sum the vector to form a scalar.

3.3 Height

Estimation of the height required a slightly different approach than the last two derivations. Because the form of the equation is linear with respect to A, a Bayesian regression must be performed, using the sampled vector **y**, and the unscaled vector **x,** which was generated using the values of **p** in the equation $e^{\frac{(x-\mu)^2}{2\sigma^2}}$. The linear regression appears as follows

$$\mathbf{y} = A\mathbf{x}$$

Completing the square yields

$$p(\beta|\sigma^2, y) \approx N(\beta, V_\beta \sigma^2)$$
$$V_\beta = (\mathbf{X}^T\mathbf{X})^{-1}$$
$$\beta = V_\beta \mathbf{X}^T \mathbf{y}$$

(*Eq. 9*)

Now, using the marginalization of

$$p(\sigma^2|y) = \frac{p(\beta, \sigma^2|y)}{p\beta|\sigma^2, y)}$$

a sample for σ^2 can be achieved, by drawing from

$$p(\sigma^2|y) \approx Inv - \chi^2(n-k, s^2)$$

$$n - k = degreesoffreedom = 1$$

$$s^2 = \frac{1}{n-k}(\mathbf{y} - \mathbf{X}\beta)^T(\mathbf{y} - \mathbf{X}\beta)$$

(*Eq. 10*)

4. Algorithm Explanation

The implementation was a straight forward Gibbs sampler. Before the first iteration, variables for estimated mean, estimated variance, and estimated height were initialized to rand(0,1). Next, samples for mean and variance were drawn, using the distributions found in *Eq. 6* and *Eq. 8,* respectively. Once these samples were taken, an estimate for the height was found by sampling from *Eq*. 9 using the marginal *Eq*. 10. After each iteration, the samples for each parameter were stored, and once all iterations were completed, the mean of each parameter vector was calculated and returned as the best estimate for the associated parameter.

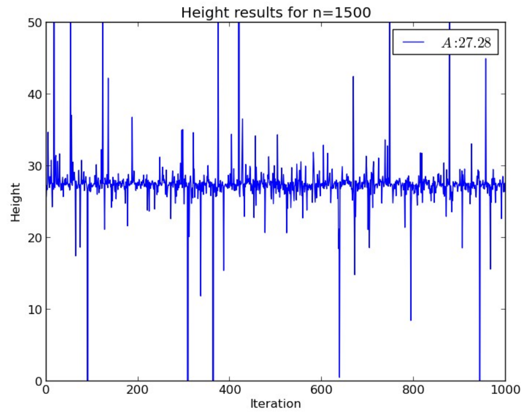

Figure 1: Point estimates for height parameter

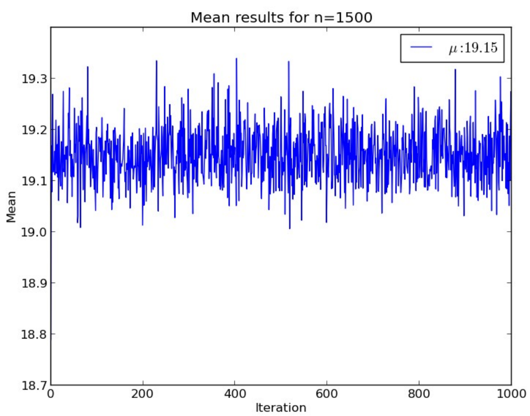

Figure 2: Point estimates for mean parameter

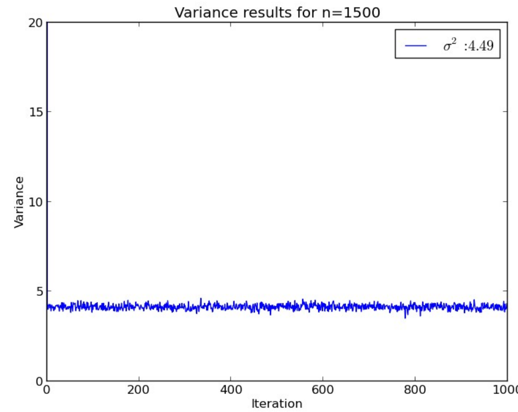

Figure 3: Point estimates for variance parameter

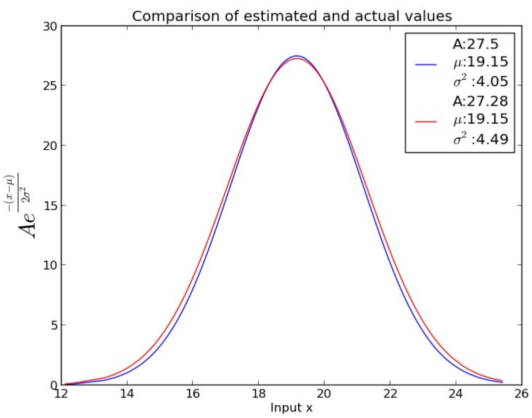

Figure 4: Comparison of final parameter estimates

5. Simulation Results

Because all of these parameters are Gibbs sampled burn-in is not required, and the value from every iteration can be used to form the parameter estimates. As seen in Figure 4., this algorithm is able to recover all three parameters for the Gaussian peak function with minimal error. Accurate estimates appear to depend largely on the number of initial samples **n**, and the variance of the peak. This is likely caused by a dependence on the sampled area. When variance is low, a relatively small number of initial gaussian samples can represent most of the area of the peak function. As the variance increases, more samples are required to describe the area of the curve, with extreme variances requiring many points to get complete coverage in order to represent the area of the peak.

6. References and Appendix

The python code used to generate the figures for this paper and verify correctness of the derivations is shown in the appended documentation. This code is also available at http://www.github.com/kastnerkyle/School/ssp/finalproj.py.

[1] A. Gelman, J.B. Carlin, H.S. Stern, D.B. Rubin, *Bayesian Data Analysis*, Chapman & Hall, 2003.

[2] J. Zhang, X. Zhou, H. Wang, A. Suffredini, L. Zhang, Y.Zhang, and S. Wong. “Bayesian Peptide Peak Detection for High Resolution TOF Mass Spectrometry”, *IEEE Transaction on Signal Processing*, Volume 58, Issue 11, pp. 5883-5894.