Proceedings of the 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, San Francisco, CA, Sept. 2004

# A Bayesian Approach to Biomechanical Modeling to Optimize Over Large Parameter Spaces While Considering Anatomical Variability

V. J. Santos and F. J. Valero-Cuevas
Neuromuscular Biomechanics Laboratory, Cornell University, NY, USA

*Abstract*—**We present the Markov Chain Monte Carlo (MCMC) approach in the context of a musculoskeletal model of the thumb. With special consideration for the complexities of biomechanical modeling, we present this approach as an alternative to standard parameter estimation techniques that produce a single, in some way optimal, set of parameter values. In contrast, MCMC methods are derived from a Bayesian philosophy, in which each "true" model parameter is actually a random variable with its own probability distribution. With MCMC we can (1) address challenges of model parameter estimation that are difficult for gradient-based methods to meet, (2) estimate the inherent biomechanical capabilities of a specific "model topology" for large, variable parameter spaces (e.g. 50-dimensional for the assumed thumb model), and (3) determine the functional consequences of the unavoidable anatomical variability across subjects in a population. Using the MCMC approach with a Metropolis-Hastings sampling algorithm we explored a 50-D musculoskeletal parameter space and successfully achieved convergence. We found the relatively small subspace of the expansive 50-D space that, for a hinged serial linkage model of the thumb, predicts functional outcomes that best-fit the experimental data.**

*Keywords*—**biomechanical modeling, hand, Markov Chain Monte Carlo, Metropolis-Hastings sampling, parameter estimation, stochastic simulation, thumb**

## I. Introduction

The objective of this work is to present the Markov Chain Monte Carlo (MCMC) approach as a statistics-based model parameter estimation methodology for biomechanical modeling. Modeling begins typically by selecting a model *topology* by hand (i.e., assumed biomechanical structure) and then the model *parameters* (i.e., specifics of the structure) are adjusted in an effort to explain and/or reproduce experimental data. If the unavoidable discrepancies between model predictions and experimental data are too large, one must ask whether this is due to inadequate model topology and/or because the search for satisfactory model parameter values has been unsuccessful. The challenge is to determine if additional or alternative explorations of the parameter space would improve results sufficiently, or if using an alternative model topology would be more fruitful. Improving current models necessitates that we explicitly investigate how the assumed model topology fundamentally determines and limits model behavior [1].

Parameter estimation literature has typically focused on objective functions and methods (e.g., gradient descent or simulated annealing) that produce a single, in some way optimal, set of parameters (e.g. maximum likelihood estimation). Methods such as least-squares estimation and maximum likelihood estimation, however, were derived under a Frequentist (or non-Bayesian) philosophy, which views "true" model parameters as being unknown and *fixed* (i.e. non-random). In contrast, the Bayesian philosophy views "true" model parameters as *random variables* themselves rather than fixed values [2]. This Bayesian viewpoint is perhaps more germane to biomechanics, where anatomical variability is the rule, because it acknowledges that each model parameter has its own probability distribution and is an instance of a random variable from this distribution.

In our modeling work on the human thumb, we have used a Monte Carlo approach to emphasize the need to understand the inherent biomechanical capabilities of a particular model topology by treating parameters as random variables to acknowledge the unavoidable variability and uncertainty of parameter values, and by considering numerous realistic model parameter combinations [3]. The MCMC approach is a powerful extension that can address challenges of complex biomechanical modeling that are difficult for gradient-based methods to meet. For example, MCMC can guarantee an exhaustive exploration of a complex high-dimensional parameter space of random variables (e.g., 50-dimensional for the assumed thumb model), can accommodate nonlinear and discontinuous system behavior, and yields parameter distributions (guided by experimental measurements) and their associated biomechanical output [4]. These resulting distributions can be easily compared with statistical distributions of experimental measurements. Furthermore, the MCMC approach allows us to explicitly determine the sensitivity of the model topology to parameter variability and uncertainty, resulting from anatomical variability across a population and sparseness of data, respectively.

## II. Methodology

We extend the use of stochastic methods in biomechanical modeling [3,5] by implementing Markov Chain Monte Carlo simulations using a Metropolis-Hastings sampling algorithm to investigate the effects of musculoskeletal variability on the capabilities of a thumb model to realistically predict static thumbtip force production in 3D. MCMC analysis is quite well-known in the fields of biology, chemistry, and physics, where it was first introduced [6]. To our knowledge, the MCMC technique has not yet been employed to investigate

musculoskeletal parameters and their variability in the context of biomechanical modeling.

*A. Markov Chain Monte Carlo (MCMC) simulations*

Markov Chain Monte Carlo methods can be used to estimate the parameters of complex multivariate systems that are difficult or impossible to solve in closed form by using relatively simple sampling techniques. A Markov chain is a discrete random process in which the next state of a system (e.g. parameter values) is dependent only upon the system's current state [4]. That is, the starting point and history of the system up to the current state are irrelevant. Once a sampling algorithm has been properly constructed, we can run Markov chains for a sufficiently large number of iterations and the Law of Large Numbers [2] will ensure that we converge upon the target distribution of interest $\pi(\underline{\theta})$, also known as the stationary distribution or the posterior distribution $p(\underline{\theta}|\underline{x})$, the probability of the model parameter vector $\underline{\theta}$, *given* observed data $\underline{x}$. Due to the memoryless nature of the Markov chain, independent chains started at overdispersed initial conditions should eventually converge to the target distribution, if it exists.

We performed Markov Chain Monte Carlo simulations in MATLAB™, utilizing a Metropolis-Hastings sampling algorithm [4]. At each iteration, a candidate model parameter vector $\underline{\theta}_{cand}$ is drawn from a proposal distribution $q(\underline{\theta}_{cand}|\underline{\theta}_n)$. According to the algorithm, this candidate parameter set is accepted with an acceptance probability $\alpha$ of

$$\alpha(\underline{\theta}_n, \underline{\theta}_{cand}) = \min\left\{\frac{\pi(\underline{\theta}_{cand})q(\underline{\theta}_n \mid \underline{\theta}_{cand})}{\pi(\underline{\theta}_n)q(\underline{\theta}_{cand} \mid \underline{\theta}_n)}, 1\right\} \quad (1)$$

where $\underline{\theta}_n$ is the current parameter set. At each iteration, $\alpha$ is calculated and a random uniform variable $U(0,1)$ is drawn. If the random uniform variable lies in the $[0,\alpha]$ range, $\underline{\theta}_{cand}$ is accepted and we set $\underline{\theta}_{n+1}=\underline{\theta}_{cand}$. Otherwise, we set $\underline{\theta}_{n+1}=\underline{\theta}_n$.

The formulation of the proposal distribution $q(\underline{\theta}_{cand}|\underline{\theta}_n)$ is such that the candidate parameter set depends only on the current parameter set. This is the very nature of the Markov chain. When the proposal distribution is symmetric, the Hastings ratio $q(\underline{\theta}_n|\underline{\theta}_{cand})/q(\underline{\theta}_{cand}|\underline{\theta}_n)$ equals one and (1) simplifies to the special Metropolis case [4].

$$\alpha(\underline{\theta}_n, \underline{\theta}_{cand}) = \min\left\{\frac{\pi(\underline{\theta}_{cand})}{\pi(\underline{\theta}_n)}, 1\right\} \quad (2)$$

To generate our candidate parameter sets, we perturbed elements of the most recent parameter set $\underline{\theta}_n$ by a constant drawn from a uniform proposal distribution $U(0,1)$ and scaled by a constant vector $\underline{c}$. The value of each $\underline{c}$ element was specific to each model parameter and was first arbitrarily set at 20% of the allowable range of each parameter (i.e., the bounds of the prior distribution $p(\underline{\theta})$, our assumed distribution of model parameters informed by prior knowledge). We then adjusted the individual elements of $\underline{c}$, as necessary, to achieve an acceptance rate of approximately 25% [7].

$$\underline{\theta}_{cand} = \underline{\theta}_n \pm \underline{c}*\text{Uniform}(0,1) \quad (3)$$

Due to our symmetric uniform proposal distribution, our acceptance probability equation reduced to the special Metropolis case (2).

According to Baye's Rule, the target (or posterior) distribution $\pi(\underline{\theta})$ is proportional to the product of the likelihood $p(\underline{x}|\underline{\theta})$ (also written $L(\underline{\theta})$) and the prior distribution $p(\underline{\theta})$ [2]. The likelihood characterizes the probability of observing data $\underline{x}$, *given* the model parameter vector $\underline{\theta}$.

$$\pi(\underline{\theta}) = p(\underline{\theta}|\underline{x}) \propto p(\underline{x}|\underline{\theta})*p(\underline{\theta}) \quad (4)$$

Combining (2) and (4), we get

$$\alpha(\underline{\theta}_n, \underline{\theta}_{cand}) = \min\left\{\frac{p(\underline{x} \mid \underline{\theta}_{cand}) * p(\underline{\theta}_{cand})}{p(\underline{x} \mid \underline{\theta}_n) * p(\underline{\theta}_n)}, 1\right\} \quad (5)$$

In practice, the prior distribution $p(\underline{\theta})$is chosen to reflect the amount of prior knowledge of the parameters. For example, if it is known that a particular anatomical parameter (e.g. bone length) has a limited range, the minimum and maximum values can be reflected in the boundaries of the prior distribution. Furthermore, the certainty or uncertainty in the prior knowledge can be represented in the prior. For instance, one can use a normal distribution with a small variance (high degree of certainty) or a large variance (low degree of certainty). The prior distribution can also be justifiably vague or diffuse, as in the case of a uniform distribution where no a priori knowledge about a variable's distribution can be assumed, other than its anatomical bounds.

We assumed a model topology, consisting of a hinged serial mechanism [8] based on anatomical data [9], and focused on evaluating the biomechanical capabilities (see methods in [3]) of this topology when varying its 50 parameters as per assumed prior distributions. We conservatively assumed a uniform distribution for each model parameter because experimental measurements are sparse (i.e., of unknown distribution) and we were interested in a thorough exploration of the parameter space without biasing the selection of parameters. We bounded each parameter's uniform prior based on published or measured data (mean ± std) [10]. When assuming a uniform distribution as the prior distribution for all model parameters, the $p(\underline{\theta}_{cand})$ and $p(\underline{\theta}_n)$ terms cancel in (5). As a result, the acceptance probability $\alpha$ becomes a function of a ratio of likelihood distributions.

$$\alpha(\underline{\theta}_n, \underline{\theta}_{cand}) = \min\left\{\frac{p(\underline{x} \mid \underline{\theta}_{cand})}{p(\underline{x} \mid \underline{\theta}_n)}, 1\right\} \quad (6)$$

Assuming that the experimental data vector $\underline{x}$ (n x 1) is composed of independent, identically-distributed observations, each likelihood distribution can be expanded in the following manner.

$$p(\underline{x} \mid \underline{\theta}) = \prod_{i=1}^{n} p(x_i \mid \underline{\theta}) \quad (7)$$

Our experimental data vector consisted of independent, maximal voluntary thumbtip force measurements [3]. We assumed each observation $x_i$ to be identically-distributed from a normal distribution $x_i \sim N(f(\underline{\theta}_{cand/n}), \boldsymbol{\Lambda})$ where each mean was a function of either $\underline{\theta}_{cand}$ or $\underline{\theta}_n$. As a result, our acceptance probability had the following form

$$\alpha(\underline{\theta}_n, \underline{\theta}_{cand}) = \min\left\{\frac{\prod_{i=1}^{n} \exp[-\frac{1}{2}(x_i - f(\underline{\theta}_{cand}))^T \Lambda^{-1}(x_i - f(\underline{\theta}_{cand}))]}{\prod_{i=1}^{n} \exp[-\frac{1}{2}(x_i - f(\underline{\theta}_n))^T \Lambda^{-1}(x_i - f(\underline{\theta}_n))]}, 1\right\} \quad (8)$$

Note the least-squares structure of the numerator and denominator in (8). If the thumbtip force and muscle coordination pattern predictions of a candidate parameter set $\underline{\theta}_{cand}$ do not match as well with the experimental data vector $\underline{x}$ as the current parameter set $\underline{\theta}_n$ in a least-squares sense, the argument of the exponential becomes increasingly negative, resulting in a smaller numerator. The acceptance probability $\alpha$ becomes smaller and the Markov chain is less likely to accept the candidate parameter set.

We ran MCMC simulations, using experimental data from four subjects [3] to compose $\underline{x}$ and the variance-covariance matrix $\boldsymbol{\Lambda}$. We started each Markov chain with a randomly chosen thumb model that could produce well-directed maximal static thumbtip forces in each of the five orthogonal directions for two pinch postures [3]. A total of 50 parameters were randomly selected from their respective prior distributions: 12 bone lengths/widths, 8 physiological cross-sectional areas, 4 extensor mechanism angles, 16 axis of rotation location/orientation parameters, and 10 posture joint angles (5 each for key and opposition pinch).

### *B. Convergence criteria*

We ran ten independent Markov chains for each of four subjects in order to obtain overdispersed starting points and, thus, a fair and thorough exploration of the 50-dimensional parameter space [4]. We visually checked the starting points of each chain (first randomly drawn parameter sets) to make sure that they spanned the uniform prior distribution space in an unbiased manner. However, since the Markov process is memoryless, the dispersion of the starting points was not as important as verifying that each parameter was sufficiently explored across its allowable range. We confirmed for each subject that the ten chains, collectively, explored the allowable range for each parameter.

Convergence among chains was based on the widely-used Gelman-Rubin statistic ($\sqrt{\hat{R}}$), which compares the variance between independent chains to the variance within each chain [4]. This ratio approaches 1 as the model parameters approach their target distributions. In practice, once the $\sqrt{\hat{R}}$ value falls below a threshold of 1.2 for all parameters for the remainder of the simulations, the chains are said to be "mixing well" and successful convergence to the target distributions is assumed. Once convergence is achieved, it is believed that model parameters are being randomly drawn in proportion to their true, yet unknown, distributions. The iterations preceding the crossing of the critical $\sqrt{\hat{R}}$ threshold are part of the "burn-in" period. As is typically done, we discarded iterations from the burn-in period and pooled iterations after the burn-in period for drawing inferences (conclusions).

If chains were not mixing well, elements of the scaling vector $\underline{c}$ in (3) were adjusted to increase the chain acceptance rates, thereby allowing chains to "escape" their current unacceptable locations and eventually mix well with one another, analogous to raising the temperature in simulated annealing.

Upon convergence of the ten chains, we extended chains as needed in order to adequately sample the parameter space of the converged model. We extended chains until we had obtained approximately twice as many converged iterations as it took for the ten chains to converge (e.g. extended simulations to 3,000 iterations/chain if convergence took 1,000 iterations).

## III. Results

Importantly, despite the complexities of the 50-D parameter space, the MCMC simulations converged (Fig. 1) and posterior distributions for all 50 parameters were obtained. We obtained approximately 23,000 pooled, converged data-driven iterations.

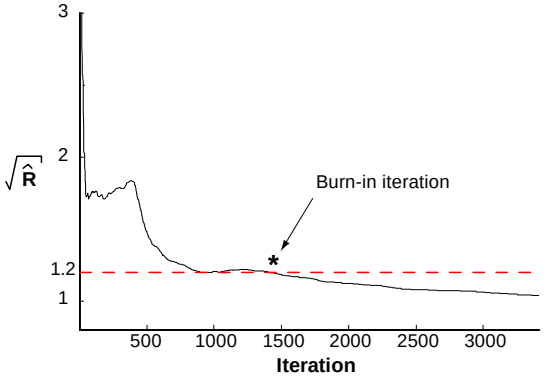


Fig. 1. The Gelman-Rubin statistic for a representative model parameter dropped below the 1.2 threshold value at the "burn-in" iteration indicated by the asterisk. Only data after the burn-in iteration were used for inference.

## IV. Discussion

The Bayesian MCMC methodology has often been criticized for its subjective selection of the prior distribution. However, proponents of the Bayesian philosophy counter that (1) prior information should be used if available and (2) if the MCMC simulations are run for a sufficiently long time (i.e. to convergence), the selection of the prior distribution becomes less important and the data, in a sense, speaks for itself. The MCMC approach essentially uses observed data to "update" the prior model parameter information.

One major advantage of the MCMC approach is that the output of the simulations consists of a posterior distribution for each parameter. These distributions can, in turn, be used as prior distributions for future simulations, instead of the diffuse priors used in this initial study. Furthermore, should one desire a single set of parameter values (for a single generic model, for example), these values can be readily obtained from the distributions (e.g., modes, means, confidence interval boundaries).

The Bayesian concept that "true" parameters are themselves random variables from statistical distributions is particularly well-suited to study biomechanical function in musculoskeletal systems that naturally vary across the population. An identical modification of two sufficiently distinct biomechanical systems can result in different functional outcomes. Thus, it is not surprising that clinicians report that subject-to-subject anatomical variability contributes to the success or failure of a clinical method, such as functional electrical stimulation, tendon transfers, arthroplasties, etc. This Bayesian approach allows us to pose questions such as: Given that anatomical variability exists from one human to the next, can the selected model topology lead to meaningful clinical predictions once we consider anatomical variability? Can one use a general model for the entire population, or are subject-specific models necessary? If so, which model features/parameters should be adjusted? We continue to address these issues in our efforts to produce clinically useful models for studying the functional consequences of orthopedic and neurological diseases, and their treatment outcomes.

## V. Conclusion

Using a Bayesian approach that considers model parameters as realizations of random variables instead of unknown constants, we found the distribution and range of all possible functional capabilities for a specific model topology resulting from measured anatomical variability of the thumb. We used a Markov Chain Monte Carlo approach with a Metropolis-Hastings sampling algorithm to explore a 50-D parameter space and successfully achieved convergence. Given 50 model parameters, each with their own variability and uncertainty, we found the relatively small subspace of the expansive 50-D space that, for the hinged serial linkage model topology, predicts functional outcomes that best-fit the experimental data. That is, there is no other subspace that would fit the experimental data better. Further details regarding the thumb model, the capabilities of the model topology, and the biomechanical consequences of the anatomical variability will be the subject of a separate publication.

## Acknowledgment

We thank Prof. Carlos Bustamante from Cornell University for his guidance in Markov Chain Monte Carlo techniques, and Prof. Mark Campbell for his simulation advice. This material is based upon work supported under a National Science Foundation (NSF) Graduate Research Fellowship (to V. J. Santos); and from ITR-0312271 grant from NSF's Robotics and Computer Vision Program, CAREER Award BES-0237258 from NSF's Biomedical Engineering/Research to Aid Persons with Disabilities Program, and Biomedical Engineering Research Grant RG-00-0397 from the Whitaker Foundation (to F. J. Valero-Cuevas).

## References

[1] F. J. Valero-Cuevas, "An integrative approach to the biomechanical function and neuromuscular control of the fingers," *J. Biomech.*, in press.

[2] J. Rice, *Mathematical Statistics and Data Analysis*. Belmont, CA: Duxbury Press, 1995, pp. 163-164, 587-589.

[3] F. J. Valero-Cuevas, M. Johanson, and J. Towles, "Towards a realistic biomechanical model of the thumb: The choice of kinematic description may be more critical than the solution method or the variability/uncertainty of musculoskeletal parameters," *J. Biomech.*, vol. 36, no. 7, pp. 1019–1030, July 2003.

[4] W. Gilks, S. Richardson, and D. Spiegelhalter, *Markov Chain Monte Carlo in Practice*. Boca Raton, FL: Chapman & Hall/CRC, 1996, pp. 1-19, 45-57, 131-143.

[5] R. E. Hughes, and K. N. An, "Monte Carlo simulation of a planar shoulder model," *Med. Biol. Eng. Comput.*, vol. 35, pp. 544-548, 1997.

[6] N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller, "Equation of state calculations by fast computing machines," *J. Chem. Phys.*, vol. 21, pp. 1087-1092, 1953.

[7] S. Chib, and E. Greenberg, "Understanding the Metropolis-Hastings Algorithm," *Amer. Stat.*, vol. 49, no. 4, pp. 327-335, Nov. 1995.

[8] V. J. Santos, and F. J. Valero-Cuevas, "Anatomical variability naturally leads to multimodal distributions of Denavit-Hartenberg parameters for the human thumb," *Proc. 25th Annual Int'l Conf. IEEE EMBS*, Cancun, Quintana Roo, Mexico, Sept. 2003.

[9] D. J. Giurintano, A. M. Hollister, W. L. Buford, D. E. Thompson, and L. M. Myers, "A virtual five-link model of the thumb," *Med. Eng. Phys.*, vol. 17, no. 4, pp. 297-303, June 1995.

[10] V. J. Santos, and F. J. Valero-Cuevas, "Investigating the interaction between variability in both musculoskeletal structure and muscle coordination for maximal voluntary static thumb forces," presented at the Annual Mtg. Society for the Neural Control of Movement, Sitges, Spain, March 2004.