



Published in final edited form as:

Adv Genet. 2010 ; 72: 47–71. doi:10.1016/B978-0-12-380862-2.00003-5.

Complex System Approaches to Genetic Analysis: Bayesian Approaches

Melanie A. Wilson, James W. Baurley, Duncan C. Thomas, and David V. Conti

Department of Preventive Medicine, University of Southern California, Los Angeles, California, USA

Abstract

Genetic epidemiology is increasingly focused on complex diseases involving multiple genes and environmental factors, often interacting in complex ways. Although standard frequentist methods still have a role in hypothesis generation and testing for discovery of novel main effects and interactions, Bayesian methods are particularly well suited to modeling the relationships in an integrated “systems biology” manner. In this chapter, we provide an overview of the principles of Bayesian analysis and their advantages in this context and describe various approaches to applying them for both model building and discovery in a genome-wide setting. In particular, we highlight the ability of Bayesian methods to construct complex probability models via a hierarchical structure and to account for uncertainty in model specification by averaging over large spaces of alternative models.

I. INTRODUCTION

Bayesian approaches have gained a tremendous amount of popularity in a diverse set of applications including but not limited to: economics, environmental science, bioinformatics, epidemiology, genetics, computer science, political science, and public policy. Within these fields, the Bayesian framework can be applied to a wide range of statistical model classes such as linear regression, generalized linear models, survival analysis, tree models, graphical models, and spatial analyses. The growth in popularity of Bayesian approaches is due in most part to the intuitive nature of inference within the framework, the extreme flexibility of the models, and the computational developments helping to facilitate practical analyses. While this chapter focuses more specifically on the use of Bayesian approaches for complex genetics applications, we begin with a general introduction to the fundamentals of any Bayesian analysis.

A. Fundamentals of a Bayesian approach

The fundamentals of a Bayesian approach lie in Bayes Rule, which is the tool that allows us to revise our current set of beliefs about unknown parameters given a set of observed data Y via conditional probabilities:

$$p(\theta|Y) = \frac{p(\theta, Y)}{p(Y)} = \frac{p(Y|\theta)p(\theta)}{\int p(Y|\theta)p(\theta)d\theta}$$

where the integral in the denominator can be replaced by a summation if the probability distribution of θ is discrete. Thus, any Bayesian approach has two major components: (1) defining the joint probability model $p(\theta, Y)$ and (2) computing conditional probabilities $p(\theta | Y)$. In defining the joint probability model, we must specify the likelihood of the observed data given the parameters of interest, $p(Y | \theta)$. This specification is common to both a frequentist approach and a Bayesian approach. However, instead of assuming that the parameters of the model are fixed and their true value is unknown, the Bayesian framework assumes that the parameters themselves are random variables. Thus, to define the joint probability model, we must also specify the prior distribution of the parameters, $p(\theta)$, in addition to the likelihood of the observed data.

The above framework assumes that we are interested in making inference on all of the parameters in the probability model. However, in many applications this is not the case. If there is some subset of parameters, θ_1 , that we are interested in and the remaining, θ_N , are nuisance parameters we can rewrite the conditional probability for the parameters of interest as:

$$p(\theta_1 | Y) = \frac{p(Y | \theta_1) p(\theta_1)}{p(Y)},$$

where $p(Y | \theta_1)$ is the marginal likelihood of the data given the parameters of interest for making inference and can be calculated by integrating out the nuisance parameters (or summation for discrete measures):

$$\int p(Y | \theta_1, \theta_N) p(\theta_N) d\theta_N.$$

Thus, computing conditional probabilities in the Bayesian framework often requires computing high-dimensional integrals (or summations) for both the marginal likelihoods for the parameters of interest, $p(Y | \theta_1)$, and the normalizing constant, $p(Y)$.

Finally, in many analyses, there may exist several alternative specifications for the joint model. In a Bayesian analysis, we can incorporate the uncertainty of the specification of the joint model by indexing a specific model by an indicator vector γ . Since we are now considering multiple models, $M_\gamma \in M$, we can rewrite the joint probability statement as:

$$p(Y, \theta_\gamma, M_\gamma) = p(Y | \theta_\gamma, M_\gamma) p(\theta_\gamma | M_\gamma) p(M_\gamma)$$

where θ_γ are the parameters specific to model M_γ . We can therefore make inference on the models themselves by calculating the conditional probabilities:

$$p(M_\gamma | Y) = \frac{p(Y | M_\gamma) p(M_\gamma)}{\sum_{M_\gamma \in M} p(Y | M_\gamma) p(M_\gamma)}$$

where $p(Y | M_\gamma) = \int p(Y | \theta_\gamma, M_\gamma) p(\theta_\gamma) d\theta_\gamma$ is the marginal likelihood of the observed data given any of the models of interest.

B. Bayesian advantages

Bayesian approaches come with many advantages. First and foremost, by specifying a probability distribution on the parameters we directly quantify the uncertainty in those parameters given the observed data and achieve statistical conclusions with common sense interpretations. These probability statements allow for a conceptually straightforward approach to inference, in that our prior beliefs specified in $p(\theta)$ are updated based on the observed data, Y , via conditional probabilities $p(\theta | Y)$. This is quite different from a frequentist approach. A simple example is the Bayesian credible interval versus the frequentist confidence interval. A frequentist 95% confidence interval must be interpreted based on hypothetical repeated sampling in which S repeated samples are taken from the population and, subsequently, $S \times 0.95$ of the estimated parameters would fall within the confidence interval. In contrast, the Bayesian 95% credible interval is interpreted potentially more intuitively as a 95% probability that the true value of the parameter lies within the calculated credible interval.

Another advantage of the Bayesian framework is that it provides a very natural setting for incorporating complex structures, multiple parameters, and procedures for dealing with nuisance parameters (parameters that we do not wish to make inference about). The only restriction within a Bayesian approach is that one must be able to specify a joint probability model for the observed data and parameters of interest. We can therefore include as many parameters to our models as needed and simply marginalize across (i.e., integrate out or sum over) the ones that we are not interested in making inference on. We are also able to incorporate external information in the analysis in an explicit manner by specifying prior probability distributions for the parameters of interest. This is particularly useful in the biological setting where there is often a great deal of external information and incorporating this information can potentially help the practitioner narrow the focus of an otherwise overly complex model. Finally, the Bayesian framework provides a natural setting for incorporating model uncertainty into any analysis by extending the hierarchy and viewing the model itself as a random variable with its own prior distribution.

C. Limitations

Many of the main advantages of a Bayesian approach lie in the specification of prior distributions on the model parameters and in some cases on the models themselves. However, this can also be one of the main limitations of the Bayesian framework. The prior distributions can be specified in either a subjective or objective manner depending about the amount of prior knowledge or external information one has for the parameters or the acceptable degree to which the posterior results are sensitive to the prior specification. However, even if there is a large amount of prior information regarding the parameters of interest or the models themselves, it is not always straightforward to quantify a practitioner's prior knowledge and elicit prior distributions. Also, if a limited amount of external knowledge exists about the parameters, the question remains about how to specify the priors in an objective ("non-informative") manner (Thomas *et al.*, 2007a). Advances have been

made in both the elicitation of prior distributions and in developing and investigating the asymptotic characteristics of objective priors on both the parameters of interest and the models themselves (discussed in more detail in Sections III and IV).

Another limitation to the Bayesian framework is the complexity of computing conditional probabilities. This complexity lies in the difficulty in performing potentially high-dimensional integrals (or summations) in both the normalizing constant $p(Y)$ and in the marginal likelihood of the data $p(Y | \theta)$ or $p(Y | M_\gamma)$. Because of the computational constraints on Bayesian approaches, practitioners were historically limited to choosing only conjugate likelihoods and priors for which conditional probabilities could be calculated in closed form. However, with recent advances in estimating high-dimensional integrals with stable numerical approximations or by simulation via Markov Chain Monte Carlo (MCMC), Metropolis–Hastings (MH), and stochastic model search algorithms, one is much less limited when choosing a joint probability model. These technical advances, as well as software to more easily implement Bayesian models (such as WinBUGS (Spiegelhalter *et al.*, 2003) and JAGS (<http://sourceforge.net/projects/mcmc-jags>)), have led to an increased popularity of Bayesian approaches (discussed in more detail in Sections III and IV).

D. General model specification

The remainder of the chapter utilizes a generalized linear model framework in which most any type of outcome variable may be analyzed with an appropriate link function. However, for simplicity, we will focus on a binary outcome variable. Let Y be a vector of length n comprised of some binary outcome variable for individual i with expected value

$$E(Y) = \mu$$

with mean vector $\mu = (\mu_1, \dots, \mu_n)^T$. Also, for each individual, we assume that p covariates are measured, x_1, \dots, x_p . We then use generalized regression models to relate the binary outcome variable to a subset of predictor variables. We denote the collection of all possible models by M . An individual model, denoted by M_γ , is specified by an indicator vector γ . Then under each model M_γ , μ is of the general form:

$$M_\gamma; \mu = g^{-1}[f(X_\gamma)\beta_\gamma]$$

where g is the link function (usually logit whenever Y is binary), $f(X_\gamma)$ is some general structure or parameterization of a set of covariates, and β_γ are the effects of $f(X_\gamma)$ on the outcome of interest Y (are parameters of interest θ_γ). A very general structure for the models can be seen as a directed graph in Fig. 3.1. Specifically in this example, the γ indicator allows both the covariates x_1 and x_3 to inform the model and design matrix X_γ , whereas all other covariates (e.g., x_2 and x_p) are excluded. Then $f(X_\gamma)$ defines the structure or parameterization of these variables in the regression. Finally, the link function, $g^{-1}(\cdot)$, as well as the regression coefficients, β_γ , relate the structure built in $f(X_\gamma)$ to disease.

For a simple illustration of the general model structure, we consider a simple logistic regression analysis where the link function is defined as:

$$\mu = \frac{\exp\{f(X_\gamma)\beta_\gamma\}}{1 + \exp\{f(X_\gamma)\beta_\gamma\}}$$

Each model is defined by the indicator γ denoting which of the p possible covariates will be included in the marginal model. In particular, we can have γ be an indicator vector with $\gamma_j = 1$ if covariate j is the one included in the marginal model and $\gamma_j = 0$ for every $i \neq j$. Then the model space M is made up of p single variable models M_γ . Under each model $M_\gamma \in M$ the mean vector has the form:

$$M_\gamma : \ln\left(\frac{\mu}{1-\mu}\right) = \beta_0 + X_\gamma\beta_\gamma$$

where X_γ is of dimension $n \times 1$ and β_γ is the logarithm of the odds ratio for the single covariate in model M_γ . In the simple marginal case, it is plausible to perform an enumeration of the entire model space M , so typically no stochastic model search algorithm is needed.

Given the general structure that we have developed above, the models that we describe herein can be much more complex than this simple marginal example. In Section II, we review strategies for defining the structure or parameterization of the covariates that is specified in $f(X_\gamma)$. These strategies can range from the simple marginal analyses to the more complex networks and mechanistic models. In Section III, we tackle the problem of (1) what priors should be placed on the parameters of interest, β_γ , and (2) how to compute the conditional probabilities and estimate β_γ . In Section IV, we incorporate an additional layer into the hierarchy by investigating multiple models $M_\gamma \in M$. Section IV mainly focuses on (1) defining prior distributions on the model space $p(M_\gamma)$, (2) how to approximate high-dimensional integrals for marginal likelihoods for the models of interest, $p(Y | M_\gamma)$, and (3) approximating high-dimensional sums for normalizing constants. Finally, in Section V, we describe methods for determining noteworthiness of the structures defined in Section II and marginally of the covariates themselves.

II. STRUCTURE OF COVARIATES

We begin with the simplest case of a single polymorphism in a single gene, with no genetic or environmental modifiers. In this case, X is comprised of genotypes at a single polymorphism for n individuals and the only consideration in specifying the covariate function $f(X)$ concerns issues of dominance. If the locus is diallelic, with alleles denoted a and A , then we might wish to consider any of the following codings for the genotype X_i :

$$\begin{aligned} f(X_i) &= 1 \text{ if } X=AA \text{ or } Aa, 0 \text{ if } X=aa && \text{Dominant} \\ f(X_i) &= 2 \text{ if } X=AA, 1 \text{ if } X=Aa, 0 \text{ if } X=aa && \text{Additive} \\ f(X_i) &= 1 \text{ if } X=AA, 0 \text{ if } X=Aa \text{ or } aa && \text{Recessive} \\ f(X_i) &= 1 \text{ if } X=AA, \theta \text{ if } X=Aa, 0 \text{ if } X=aa && \text{Codominant} \end{aligned}$$

(Note that the codominant model entails an additional parameter θ that would be estimated along with the other regression coefficients in β . This is most easily accomplished by fitting

a 2-degree of freedom model with two dummy variables for each genotype). In some instances, the choice of coding might be determined by prior biological knowledge, but in the absence of relevant knowledge one would consider each of these possibilities within the general model selection or model uncertainty framework developed in greater detail in Section III.

Most genes have numerous polymorphisms and any candidate pathway study might consider a broad range of genes, so multivariate models are called for. An obvious choice for a binary disease trait might be a logistic regression model

$$\text{logit}(p(Y=1|X))=\beta_0+\sum\beta_jf_j(X_j)$$

Here, we could incorporate different genetic codings for each variable and therefore we allow the function $f_j(X_j)$ to differ across the covariates. It might also be helpful to consider a hierarchical structure for multiple polymorphisms within genes within pathways. Letting $j = 1, \dots, J$ index the pathways, and $k = 1, \dots, K_j$ index the genes within pathway j , and $l = 1, \dots, L_{jk}$ polymorphisms within genes, one might extend the logistic model as

$$\text{logit}(p(Y=1|X))=\beta_0+\sum_j\beta_j\sum_k\beta_{jk}\sum_l\beta_{jkl}f_{jkl}(X_{jkl})$$

with higher level models for the β s constructed in a manner to ensure identifiability. Again, working within the general modeling framework discussed below, one might wish to consider different strategies for selection, shrinkage, or averaging at the polymorphism, gene, or pathway levels.

Gene–environment ($G \times E$) and gene–gene ($G \times G$) interactions introduce a further level of complexity, but the same general framework can be applied. Consider, for example, a model for the joint effect of two polymorphisms X_1 and X_2 . The logistic model could then be written as

$$\text{logit}(p(y=1|X))=\beta_0+\beta_1f_1(X_1)+\beta_2f_2(X_2)+\beta_{12}f_{12}(X_{12})$$

where $f_1(X_1)$ and $f_2(X_2)$ denote any of the genetic codings discussed above and $f_{12}(X_{12})$ could be a simple product X_1X_2 or a term constructed to capture phase information for two polymorphisms in linkage disequilibrium (Conti and Gauderman, 2004), or it could be some more complex epistasis model, such as 1 for $X_{12}=(aaBB, aAbB, AAbb)$, 0 otherwise. See Li and Reich (2000), Moore (2003), Moore *et al.* (2007), Zhang and Liu (2007), Cordell (2009), Moore and Williams (2009), and Tang *et al.* (2009) for further discussion of various possibilities for modeling epistasis. Similar codings for $G \times E$ interactions are possible (Thomas, 2010a,b). For multiple polymorphisms within the same gene, one might use a haplotype-based model, $\text{logit}(p(Y=1|X))=\beta_0+\beta_1h_1(X)+\beta_2h_2(X)$, where h_1 and h_2 represent the two haplotypes carried by an individual comprising the set of alleles at the different loci carried on the same chromosome. In the absence of phase information, this would require forming a likelihood by summing over all possible arrangements of the alleles

into haplotypes, weighted by their probabilities based on population linkage disequilibrium patterns (Stram *et al.*, 2003).

If specific prior knowledge of the relevant biological process is known, mechanistic approaches may be used to construct more complex models. These models will generally be quite specific to a particular pathway and their mathematical form will depend upon the nature of the pathway. Metabolic pathways, for example, might be modeled in terms of a series of latent variables $L(X; m)$ representing concentrations of intermediate metabolite m . These concentrations could be given by a system of differential equations based on the known pharmacokinetics with rates that depend in some manner on the genotypes encoding the relevant enzymes and their environmental substrates or cofactors (Thomas *et al.*, 2010). The solution to this system then provides a mathematical expression for the covariate function $f(X, \theta)$ where X now represents all the genetic and environmental inputs to the system and θ a vector of additional parameters to be estimated for a particular model. For example, Cortessis and Thomas (2003) described a metabolic model for the metabolism of polycyclic aromatic hydrocarbons and heterocyclic amines derived from tobacco smoke and well done red meat as risk factors for colorectal polyps. Each pathway involves several intermediate steps, metabolized by several genes. Using linear kinetics, the metabolic rate parameters were assumed to be lognormally distributed around population means specific to the relevant genotypes. The expected concentration of the final metabolites from each pathway was computed specific to each person's exposure and genotype, and used as covariates in a logistic model for disease. The model was fitted using MCMC methods, sampling individual metabolic rates and model parameters (regression coefficients, genotype-specific mean rates, and variances) in turn. Further elaboration of this model and comparisons with a BMA analysis using hierarchical models are provided in Conti *et al.* (2003). As before, one might wish to consider a range of alternative models, such as submodels of a general model including only some subset of the inputs or different codings of dominance or even different mathematical models (e.g., linear vs. Michaelis–Menten kinetics). These approaches have been most extensively developed in the context of population pharmacokinetic models (Best *et al.*, 1995; Bois, 2001; Clewell *et al.*, 2002; Davidian and Gallant, 1992; Gelman *et al.*, 1996; Lunn *et al.*, 2009; Racine-Poon and Wakefield, 1998; Wakefield, 1996), although so far, there has been relatively little attention to genetic variation in metabolic parameters.

III. ESTIMATION

So far, in this chapter, we have focused on how to link and combined the observed data into a probability model. However, the ability to measure numerous factors present many difficulties for analysis (Greenland, 1993). Conventional approaches have relied on either fitting a full model with all the factors included in the probability model or fitting a reduced model determined by some eliminating algorithm. Including all the factors in one model can lead to biased and unreliable estimates due to sparse data when the number of parameters approaches the number of individuals in the sample (Greenland, 2000a,b). Reduced models may avoid this complication, but they fail to account for the correlations that exist between all the factors and can lead to underestimated variance (Robins and Greenland, 1986). Additionally, when statistical tests are used for the numerous exposures, issues of multiple

comparisons arise (Thomas *et al.*, 1985). However the model γ is determined, there will be corresponding parameters β_γ that describe the effect on the outcome of interest. For this part of the chapter, we focus on how prior specification on these parameters via hierarchical models can be used to construct complex models for analysis. For example, in a genome-wide association study (GWAS), one might have little or no external information about most SNPs, so such analyses are commonly treated as exploratory; indeed, their “agnostic” or “hypothesis free” nature is commonly touted as one of their advantages. Here, treating each polymorphism as independent may be appropriate. A pathway-driven study, on the other hand, may be able to exploit extensive knowledge about the pharmacokinetics or pharmacodynamics of the pathway. Recently, there has been an intriguing convergence of the two philosophies, with external pathway information being exploited to mine GWAS data for gene sets whose components may not separately attain genome-wide levels of significance but combination implicates particular pathways (Wang *et al.*, 2007; Zhong *et al.*, 2010), or by exploring GWAS data to discover hitherto unsuspected sets of genes that may share a common biological function (Sebastiani *et al.*, 2005).

Rather than treating each factor independently, specification of the relations among the observed data with two or more stages can be used to create an intricate joint probability model. While each stage may be relatively simple and easy to understand, the entire model may be much more sophisticated, with the aim to more accurately model the underlying complex processes. Additionally, by providing a joint probability model for all exposures, hierarchical modeling offers a potential solution to problems of multiple comparisons (Greenland and Robins, 1991; Thomas *et al.*, 1985). As one of the first examples in genetic epidemiology, Thomas *et al.* (1992) used hierarchical modeling to jointly evaluate numerous human leukocyte antigen (HLA) alleles and their association to insulin-dependent diabetes mellitus (IDDM), while incorporating several environmental risk factors.

The addition of higher stage information can substantially improve the accuracy and stability of effect estimates (Greenland, 2000a,b; Greenland and Poole, 1994; Morris, 1983). However, to achieve this improvement, the model hierarchy must be specified in a manner that efficiently uses the data and is scientifically plausible (Rothman and Greenland, 1998).

While a conventional model fits a single probability level to describe the relation between the multiple factors and the outcome, hierarchical models incorporate higher level prior distributions to explain the relations among the parameters. Although multiple levels of prior structure can be constructed, in practice there is a limit to the number of levels that can be feasibly estimated from the data without relying too much on a strongly specified prior. Consider the general scenario with multiple factors, X , and outcome, Y . The first-probability model at the individual-level can be specified as:

$$Y|X, \beta \sim p(Y|X, \beta)$$

As before, Y and X are observed data and β are the corresponding coefficients of risk associated with a one-unit increase for each factor. Subsequently, probability models are

placed on the parameters, β , assuming that they come from a common probability distribution.

$$\begin{aligned}\beta|\theta &\sim p(\beta|\theta) \\ \theta &\sim p(\theta)\end{aligned}$$

Such a hierarchy can apply dependencies to the estimates of the parameters β through the structure of $p(\beta|\theta)$. These dependencies are often modeled conditionally on certain parameters, θ , called hyperparameters and these hyperparameters can also be assigned probability distributions, $p(\theta)$. An important assumption here is that the parameters are exchangeable. That is, the parameters $(\beta_1, \dots, \beta_m)$ are exchangeable in their joint distribution if $p(\beta_1, \dots, \beta_m)$ is invariant to permutations of the indexes $(1, \dots, m)$ (Gelman, 1995). Consequently, given no other information, other than the data itself, there is no prior ordering or grouping of the parameters $(\beta_1, \dots, \beta_m)$. If this assumption holds, we may assume that the corresponding coefficients for the exposures, β , are drawn from the same common population distribution.

The general hierarchical modeling framework presented in the above equations provides a quite flexible framework to model complex systems. For example, $p(\beta|\theta)$ specified as a normal distribution centered at zero is akin to ridge regression, a frequentist robust regression technique (Sorenson and Gianola, 2002). Similarly, a double exponential for $p(\beta|\theta)$ is equivalent to the Lasso procedure (Park and Casella, 2008; Chen *et al.*, 2010). To demonstrate the flexibility and nuances of specifying such a hierarchy, consider a specific analysis of case-control data with numerous genetic and environmental factors. A specific hierarchical model can be specified as:

$$\begin{aligned}\text{logit}(p(Y=1|X)) &= \beta_0 + \sum \beta_j f_j(X_j) \\ \beta &\sim MVN(\mathbf{Z}\pi, \Sigma)\end{aligned}$$

where X is a design matrix of genetic and environmental factors for the individuals within the study, Z contains second-stage covariates for each of the environmental and genetic factors reflecting higher level, often prespecified relations, π is a column vector of coefficients corresponding to these higher level effects on disease, and Σ is a matrix specifying the residual covariance of the second-stage covariates. There are many possible types of information that could be used for defining Z . In genetics, this could define simple indicator variables for which pathway(s) each gene is thought to be active (Hung *et al.*, 2004), information extracted systematically from genomic or pathway ontologies (Conti *et al.*, 2009; Thomas *et al.*, 2007b), experimental information such as eQTLs from cell cultures (Zhong *et al.*, 2010), *in silico* predictions based on evolutionary conservation or predicted effect on protein conformation (Rebeck *et al.*, 2004), or predictions from simulations of the pathway (Thomas *et al.*, 2010). Of course, multiple sources could, in principle, be combined and additional levels can easily be incorporated, such as separate models for SNPs within genes and genes within pathways (Conti and Gauderman, 2004) or for main effects and interactions (Conti *et al.*, 2003).

Overall this hierarchy will result in posterior estimates for the association parameters that are an inverse-variance weighted average between the conventional estimates from the logistic regression only and the estimated conditional second-stage means, $Z\pi$. To see this more clearly, consider a simple weighted regression approach for estimation. Here, the second-stage estimated prior means, $Z\tilde{\pi}$, and corresponding estimated covariance matrix, $(Z^T WZ)^{-1}$, can be obtained from a weighted least squares regression, $\tilde{\pi} = (Z^T WZ)^{-1} Z^T W \hat{\beta}$, where $W = (V + \Sigma)^{-1}$, and V is a diagonal matrix with elements equal to the square of the estimated standard errors for $\hat{\beta}$ (Morris, 1983). Averaging the first and second-stage estimates yields posterior estimates

$$\tilde{\beta} = BZ\tilde{\pi} + (I - B)\hat{\beta}.$$

The estimate of the covariance matrix is given by $\tilde{C} = V(\hat{I} - (I - H)^T B)$, where $H = Z(Z^T WZ)^{-1} Z^T W$ and B is the estimated shrinkage matrix

$$B = WV = (V + \Sigma)^{-1} V$$

From these equations, we can see that if the maximum likelihood first-stage estimates, $\hat{\beta}$, have large variance, V , relative to the prior variance, Σ , then B will also be large. As a result, the hierarchical analysis has several important differences from the conventional, single-stage logistic regression analysis. First, the conventional analysis has no constraints on the first-stage regression coefficients, $\hat{\beta}$. For example, if we model the effects of smoking on lung cancer, we would expect the effect estimate, $\hat{\beta}$, to be quite high. But even in this extreme case, we would not expect outrageously high odds ratios. It makes sense to incorporate this expectation into our analysis by including a probability distribution for the effect estimates that weights more probable estimates higher. This may be done by setting the elements in the Z matrix to zero and allowing Σ to reflect our prior beliefs about the extent of the probability distribution of the first-stage coefficients, $\hat{\beta}$. However, if we have information regarding the relations between the factors, we may incorporate that information into the Z design matrix and account for the effects due to these dependencies. In this case, Σ reflects residual covariance or associations between the first-stage parameters, $\hat{\beta}$, after accounting for the relations defined in Z . If we allow the elements of Σ to go to infinity—that is, we have no prior belief regarding the distribution of the residual effects—the final estimates of effect will disregard any second-stage information and be equal to the first-stage conventional maximum likelihood estimates. If we set the elements of Σ to zero, believing that there are no residual effects beyond the relationships defined in Z , then the final estimates will be equal to the estimated second-stage conditional means, $Z\pi$. Thus, Σ acts as a smoothing parameter, controlling the amount of shrinkage from the conventional first-stage maximum likelihood estimates toward the second-stage conditional means. Within this context, the specification of Σ is very flexible. Joint modeling is often represented by setting $\Sigma = \tau^2 D$, where $\tau^2 > 0$ represents the overall variance of the effect estimates and D is a positive definite correlation matrix describing the correlation between the effect estimates for any two pairs of polymorphisms. Conti and Witte (2003) utilized this specification to set D to the expected decay of linkage equilibrium. Closely related to this is

a Gaussian conditional autoregression specification where $D = (I - \rho A)^{-1}$ (Wakefield *et al.*, 2000). Here, the elements within \mathbf{A} describe the spatial weights between any two polymorphisms and ρ is interpreted as the strength of the spatial dependence. Thomas *et al.* (2010) recently investigated the use of this model in the genetics context using prior gene-gene connection information such as from gene coexpression experiments to define \mathbf{A} . The normal g-prior defines the covariance matrix as the inverse of the Fisher information matrix multiplied by a constant g , $\Sigma = g(X^T X)^{-1}$ (Zellner, 1986). This ensures that the correlation structure in the prior distribution matches the structure in the likelihood with either g being determined via information criterion (such as AIC) or estimated with a prior specified for g .

The general hierarchical model specified above often leads to analytically intractable computations of posterior distributions $p(\beta | Y)$ when estimating β . Approximate methods can be used for parameter estimation and can include a semi-Bayes, empirical Bayes, or fully Bayesian approaches. In a semi-Bayes approach, the value for τ^2 is prespecified as opposed to estimating the value from the data. This may be advantageous if the estimate of τ^2 is itself highly unstable (Greenland, 1993, 1997; Greenland and Poole, 1994). However, because the most appropriate value for τ^2 is unknown it is standard to perform a sensitivity analysis to evaluate how dependent the posterior estimates are to the choice of τ^2 . In contrast, an empirical Bayes approach uses the marginal distribution of τ^2 to obtain a point estimate that is then used to evaluate the joint posterior distribution for the log odds ratios, β (Efron and Morris, 1975; Greenland and Poole, 1994; Morris, 1983; Searle *et al.*, 1992). Semi-Bayes and empirical Bayes methods, such as two-stage weighted least squares, joint iterative weighted least squares and penalized quasi-likelihood (Breslow and Clayton, 1993; Greenland, 1997), suffer from the inability to reflect our uncertainty for τ^2 in the final posterior estimates for the log of the odds ratio. A fully Bayesian approach using MCMC methods avoids this and incorporates the uncertainty about τ^2 into the analysis by evaluating the full joint posterior distribution through simulation. However, these methods contain many potential difficulties and care must be taken when implementing and interpreting the final results (Gelman, 1995; Gilks *et al.*, 1996).

Within the hierarchical modeling framework, connections to model selection procedures can be easily made by assuming a mixture model for $p(\beta | \theta)$. George and McCulloch (1993) introduced a Stochastic Search Variable Selection (SSVS) algorithm by introducing a latent variable, $W_j = 0$ or 1, indicating whether term j is included in the model with a mixture prior for the coefficients:

$$p(\beta_j | \tau, \psi) = \begin{cases} N(0, \tau^2), & \text{if } W_j=0 \\ N(0, (\psi\tau)^2), & \text{if } W_j=1 \end{cases}$$

Here, ψ is a variance inflation factor defining the separation between two normal distributions centered at zero. Since τ^2 functions as a smoothing parameter, a specification of $\tau^2 = 0$ defines a point mass at zero for those terms not included in the model (see “spike and slab” discussion in Section IV).

The benefits of a Bayesian approach to parameter specification need not be limited to the case of numerous factors. Consider a simple scenario in which the aim is to estimate a statistical or multiplicative interaction between two factors, a genetic factor G and a dichotomous environmental factor E . In the conventional single-level case-control analysis, the departure from a multiplicative interaction model can be estimated as β_{GE}^{CC} from the following logistic model:

$$\text{logit}(p(Y=1|G, E)) = \beta_0^{CC} + \beta_G^{CC} G + \beta_E^{CC} E + \beta_{GE}^{CC} GE$$

Under the assumption of a rare disease and independence of the two factors in the source population, a case-only analysis may be used to estimate an equivalent interaction β_{GE}^{CO} , as well:

$$\text{logit}(p(E=1|Y=1, G)) = \beta_0^{CO} + \beta_{GE}^{CO} G$$

Leveraging these two approaches, Mukherjee and Chatterjee (2008) constructed a hierarchical prior for the case-control estimate as a function of the case-only estimate,

$\beta_{GE}^{CC} \sim N(\beta_{GE}^{CO}, \tau^2)$. The result is a shrinkage estimator similar to that outlined above but one that is a weighted average between the case-only and case-control estimates:

$$\tilde{\beta}_{GE} = B\beta_{GE}^{CO} + (1-B)\beta_{GE}^{CC}$$

The weight B is defined as:

$$B = \frac{\hat{V}_{CC}}{\hat{V}_{CC} + \theta_{GE}^2}$$

The amount of shrinkage is controlled by θ_{GE}^2 , the maximum likelihood estimate of the log of the G–E odds in controls relative to the estimated variance (\hat{V}_{CC}) of the case-control estimator β_{GE}^{CC} . Thus, the final estimate $\tilde{\beta}_{GE}$ shares the efficiency of the case-only estimate when the two factors are estimated to be independent within the controls and the robustness of the case-control estimate otherwise.

Further demonstrating the close connection between hierarchical mixture models and Bayesian model averaging, Li and Conti (2009) proposed a similar weighted average. However, in their Bayesian model averaging approach, the weight is defined as $B = p(M_{CO} | Y)$, the posterior probability of the case-only model relative to the case-control model. This is calculated as:

$$p(M_{CO} | Y) = \frac{p(Y | M_{CO}) p(M_{CO})}{p(Y | M_{CO}) p(M_{CO}) + p(Y | M_{CC}) p(M_{CC})}$$

where $p(M_\gamma)$ is the prespecified or semi-Bayes prior probability for model γ (either the case-only or case-control models) and $p(Y | M_\gamma)$ is the integration of the likelihood of model M_γ , estimated through a Laplace transformation. For a comparable likelihood between the two models they frame both models in a log-linear framework. Similar to the approach of Mukherjee and Chatterjee (2008), since the likelihood of the case-control model is a function of the estimated G–E association in the controls, θ_{GE} plays an important role in determining the weight.

IV. MODEL UNCERTAINTY

Model selection is the process of combining data and prior information to select among a group of statistical models $M_\gamma \in M$. In building a model, decisions to include or exclude covariates as well as uncertainty in how to code the covariates in the design matrix X_γ for any given model M_γ are based both on the prior hypotheses and the data. With many potential covariates, these decisions become difficult. Some algorithms will select variables to be included in the model, but only return a single “best” model (e.g., stepwise regression). These methods fail to account for model uncertainty, that is, a number of models may fit the data equally well. In a Bayesian framework, model uncertainty can be addressed by basing inference on the posterior distribution of models.

One simple example of the model uncertainty framework is in variable selection, where each model is defined by a distinct subset of p covariates and is specified by the indicator vector γ which is comprised of a set of p 0s or 1s indicating the inclusion or exclusion of each of the covariates in model M_γ . Then in the logistic regression framework each model is defined as:

$$\text{logit}(p(Y=1|X, M_\gamma)) = \beta_0 + X_\gamma \beta_\gamma$$

where X_γ is the design matrix that is made up of the p_γ covariates in model M_γ , and β_γ is a p_γ dimensional vector of model-specific regression coefficients. Here, the model space $M_\gamma \in M$ is made up of 2^p possible models.

In general, the main quantity of interest that needs to be calculated is the posterior probability for any model M_γ given by the equation:

$$p(M_\gamma|Y) = \frac{p(Y|M_\gamma)p(M_\gamma)}{\sum p(Y|M_\gamma)p(M_\gamma)}$$

where $p(Y | M_\gamma)$ is the marginal likelihood, $p(M_\gamma)$ is the prior probability for a particular model, and the denominator is a constant found by summing over the entire model space. When p is small, one can exhaustively visit all the possible models. For p covariates, there are $t = 2^p - 1$ possible terms (including main effect and interaction terms), and 2^t possible models. As p increases the model space M quickly outgrows what is computationally feasible, and an approximation of the posterior distribution of models by MCMC methods is required. When carefully designed, these approaches can efficiently search through the model space.

Bayesian approaches introduce an additional layer of uncertainty to the model by specifying priors on the model space itself $p(M_j)$. By Ockham's razor, the simplest models that explain our observations are preferred. This guidance can be formalized as a prior (Jefferys and Berger, 1991). Many approaches adopt a "spike and slab" prior distribution, where most regression coefficients are zero (the spike) and a few coefficients have some effect (slab) (Mitchell and Beauchamp, 1988). Others incorporate some form of lasso (least absolute shrinkage and selection operator), which either shrinks coefficients or sets them to zero (Tibshirani, 1996). Some approaches directly penalize model complexity. Chen and Chen (2008) introduced a penalty term to the Bayesian information criterion (BIC) based on the size of the model space with the same number of variables as the current model. Wilson *et al.* (2010) introduced a Beta-Binomial prior on model size that holds the prior odds of any association constant as p increases, thus limiting false discoveries.

Biological knowledge can also be incorporated into a prior on the model space. This knowledge can be used as a prior on the probability that a coefficient is involved ($p(\gamma_j = 1)$) and the effect given that it is involved. As discussed earlier, SSVS achieves this using a mixture prior for variable j , such that:

$$\beta_j = \gamma_j N(\mu_j, \tau_j) + (1 - \gamma_j) \delta(0)$$

where $\delta(0)$ is the spike, and $N(\mu_j, \tau_j)$ is the prior on the mean and variance of β_j given that it is not zero (Conti *et al.*, 2003, 2009). Biological knowledge can also form the basis for model priors. For example, a hypothesized biological pathway can be used as a reference "prior topology," and structures closer to this reference have greater prior probability (Baurley *et al.*, in press).

MCMC methods are extremely flexible. Here, we describe the general design of the MH algorithm. A random walk MH algorithm explores models in the neighborhood of the current model (Robert and Casella, 2004). A random change to the current model M_{t-1} is performed to create the proposed model M' . The changes allowed are often specific to the model structure. For instance, a change to a logistic regression model could be the addition or removal of a regression term. A new model is accepted as M_t with probability,

$$\alpha = \min \left(1, \frac{p(Y|M') p(M') q(M_{t-1}|M')}{p(Y|M_{t-1}) p(M_{t-1}) q(M'|M_{t-1})} \right)$$

where $p(Y|M')$ is the marginal likelihood of the model, $p(M')$ is the model prior, and $q(\cdot)$ is the proposal density. There are many choices for the model prior and proposal density, and these influence the performance and behavior of the algorithm. After convergence, models are sampled from the posterior distribution of models.

For model selection applications where p is large, MCMC algorithms may have difficulties traversing a multimodal posterior distribution and poor efficiency arising from the scale of the model space. Recently, new samplers have been introduced that overcome some of these issues by exploiting multiple chains running in parallel. Evolutionary stochastic search

(ESS) utilizes a “population” of MCMC chains operating at different temperatures, with hotter chains moving about the model space quicker than cooler chains. The chains are updated with local and global moves. Local moves explore models in the neighborhood of the current model (i.e., random walk) whereas global moves allow a chain to move to a new area of the model space by swapping states with another chain. This algorithm has been incorporated into the MISA (Multilevel Inference of SNP Associations) framework that computes posterior probabilities and Bayes factors (BFs) at the SNP, gene, and global level (Wilson *et al.*, 2010).

In a different strategy, parallel chains are utilized to tune the MCMC proposal density to better approximate the posterior density. This improves efficiency because less time is spent proposing models with little evidence. The methodology (known as PEAK) organizes the model space into subspaces linked through a graph (Baurley *et al.*, 2009). This graph can be informative, meaning it is derived from an ontology or domain expert or simply symmetric (a divide and conquer approach to the model space). The chains running on smaller model spaces tune the proposal densities for chains operating on larger spaces. The method has been applied to a childhood asthma case-control dataset and discovered several oxidative stress genes and gene–gene interactions for further investigation.

We can also allow for model uncertainty when incorporating interactions of covariates in the structure of the interactions and the covariates being included in the interaction. For instance, in logic regression, the model is of the form,

$$\text{logit}(p(Y=1|X)) = \beta_0 + \sum \beta_j f_j(X_j)$$

where $f_j(X_j)$ is a Boolean combination of the risk factors (called a logic tree) and β_j is the net effect of that Boolean combination (Ruczinski *et al.*, 2003). The model allows for L logic trees, each containing observed variables and logical operators (AND, OR, and NOT) that can represent different types of interactions. Baurley *et al.* (2010) extended this framework to continuous variables where the function $f_j(X_j)$ became the net effect of a pathway structure (called a topology) such that,

$$f_j(X_j) = \theta_{n,1} \text{par}_1(X_n) + \theta_{n,2} \text{par}_2(X_n) + (1 - \theta_{n,1} - \theta_{n,2}) \text{par}_1(X_n) \text{par}_2(X_n)$$

where $\text{par}_1(X_n)$ and $\text{par}_2(X_n)$ return the values of the parents of X_n in the topology (either an observed variable Z or a latent variable X) and $\theta_{n,1}$ and $\theta_{n,2}$ are parameters that can represent a range of interaction types.

V. DETERMINING NOTEWORTHINESS

Given the above defined models, we are interested in addressing two questions: (1) globally, is there an association between any of our covariates of interest and the outcome? and (2) if there is a global association, which individual covariates or structures of covariates are most likely driving the association?. Both of the questions can be answered via multilevel posterior probabilities (or conditional probabilities) and Bayes factors.

A. Global posterior quantities

We are first interested in addressing the following global hypotheses:

H_A : At least one covariate is associated with the outcome of interest.

H_0 : There is no association between the covariates of interest and the outcome.

The extent to which the data supports each of the hypotheses is calculated via posterior probabilities $p(H_A | Y)$:

$$p(H_A | Y) = \frac{p(Y | H_A)p(H_A)}{p(Y | H_A)p(H_A) + p(Y | H_0)p(H_0)},$$

and $p(H_0 | Y) = 1 - p(H_A | Y)$. These quantities are a function of both the marginal likelihood of the hypotheses ($p(Y | H_A)$ and $p(Y | H_0)$) and the prior distributions placed on the hypotheses ($p(H_A)$ and $p(H_0)$). In particular, in the Bayesian variable selection framework the posterior probability of the alternative hypothesis that at least one covariate is associated takes on the form:

$$p(H_A | Y) = \sum_{M_\gamma \neq M_0} p(M_\gamma | Y),$$

where we simply sum up the posterior probability for all non-null models ($M_\gamma \neq M_0$). Also, the posterior probability of the null hypothesis takes on the form:

$$p(H_0 | Y) = p(M_0 | Y).$$

Decisions based on which hypothesis is more likely can then be made based on the posterior probabilities. These posteriors have an intuitive interpretation of the probability of the hypothesis conditional upon seeing the data.

Given that there is posterior evidence of the global hypothesis that at least one of the covariates of interest is associated with the outcome, we are further interested in answering the question of which covariate is most likely driving the association. This question can be answered based on marginal posterior probabilities. In particular, in the Bayesian variable selection framework, the posterior probability of any covariate X_i being associated can be calculated as:

$$p(\gamma_i = 1) = \sum_{M_\gamma \in M: \gamma_i = 1} p(M_\gamma | Y),$$

which is simply the sum of the posterior probability for every model that includes the covariate X_i .

One should note that in case of calculating inclusion probabilities for highly correlated covariates (i.e., SNPs in LD) there is an expected dilution in the corresponding posterior

probabilities due to the covariates providing competing evidence for an association and therefore the posterior probability of an association will be diluted or distributed across several correlated covariates. We therefore extend this notion of marginal covariate inclusion probabilities to group inclusion probabilities (where we assume less correlation will exist across the groups) and achieve multilevel posterior probabilities (Wilson *et al.*, 2010) by considering the posterior probability that at least one covariate within a given group is associated. One example would be in genetic association studies to group SNPs according to their corresponding gene and calculate gene inclusion probabilities that are simply the sum of the posterior probability of all models that include at least one of the SNPs within the given gene.

In the Bayesian framework, we can also calculate the ratio of the weight of evidence for any two hypotheses (H_A vs. H_0) based on BFs:

$$BF[H_A:H_0] = \frac{p(H_A|Y)}{p(H_0|Y)} \bigg/ \frac{p(H_A)}{p(H_0)} .$$

A BF (Kass and Raftery, 1995) compares the posterior odds of any two hypotheses to the prior odds and measures the change of evidence provided by data for one hypothesis to the other. Goodman (1999) and Stephens and Balding (2009) provide a discussion of the usefulness of BFs in the medical context and Wakefield (2007) and Whittemore (2007) illustrate their use in controlling false discoveries in genetic epidemiology studies. Jeffreys (1961) presents a descriptive classification of BFs into “grades of evidence” (reproduced in Table 3.1) to assist in their interpretation, which is also reproduced in the work of Kass and Raftery (1995). Thus, decisions about which hypothesis are more likely can be made based on these grades of evidence.

Jeffreys (1961) was well aware of the issues that arise with testing several simple alternative hypotheses against a null hypothesis, noting that if one were to test several hypotheses separately, then by chance one might find one of the BFs to be less than one even if all null hypotheses were true. He suggested that, in this context, the BFs needed to be “corrected for selection of hypotheses.” However, it was not clear what Jeffreys meant explicitly by this correction. Experience with genetic studies shown that detectable SNP associations are relatively infrequent. For this reason, Stephens and Balding (2009) suggest that marginal BFs calculated assuming equal prior odds should alternatively be interpreted in light of prior odds more appropriate to the study at hand (leading to a much greater significance threshold than Jeffreys suggests). Another approach to the problem of exploring multiple hypotheses is to embed each of the potential submodels (corresponding to a subset of SNPs) into a single hierarchical model. Unlike the marginal (one-at-a-time) BFs indicated by Stephens and Balding (2009) that are independent of the prior odds on the hypotheses, SNP BFs computed in the Bayesian variable selection framework are based on comparing composite hypotheses and hence do depend on the prior distribution over models. Thus, it is important to select priors on the model space that have an implicit multiplicity correction. One example of this is the Beta-Binomial prior with hyperparameters $a=1$ and $b=p$ (the number of SNPs in the study) suggested by Wilson *et al.* (2010). By diluting the prior marginal

inclusion probability, this maintains constant global prior odds of an association even as the number of SNPs in the analysis increases.

VI. CONCLUSIONS

Bayesian approaches to complex analysis are becoming more and more popular. In this chapter, we have attempted to demonstrate that a Bayesian perspective provides a flexible framework for complex genetic analyses by breaking the problem into several components: (1) specification of a model structure via the covariates or combination of covariates; (2) estimation and prior specification of the corresponding parameters; and finally, (3) incorporation of the uncertainty of the specified model structure. The themes presented here focus on a generalization of Bayesian hierarchical models and their flexibility to allow the analyst to easily incorporate complex structures, multiple parameters, deal with nuisance parameters, and have common sense interpretations of the parameters of interest. Unfortunately, this added flexibility over more commonly used frequentist methods comes with the added complexity of computing conditional probabilities (high-dimensional integrals or summations) and eliciting subjective or, when there is a lack of prior knowledge, developing objective prior distributions. However, in this chapter, we discuss advances that have been made in both estimating high-dimensional integrals and in the elicitation of prior distributions.

The most notable advantage of a hierarchical perspective to data analysis is that each stage becomes relatively easy to construct, understand, and interpret. Upon aggregating these stages, the overall model can be quite complex. However, even in the face of these complexities, inference is feasible by leveraging the specified hierarchy. The potential advantages of hierarchical models are often contingent upon the ability of the model, both the individual stages and the overall probability model, to provide an accurate representation of the true data generating mechanism. The potential fear that the prior will overwhelm the data has potentially lead many to shun Bayesian approaches. While all analytic models make some level of assumptions, it is important to understand that the specification of priors is not solely subjective. A prior on parameters simply specifies or structures an exchangeable class of parameters. It does not prespecify the degree to which those parameters may differ. Further uncertainty is incorporated by also searching over alternative models. Sensitivity analysis should be performed to gauge the dependence of final inference upon the prior structure. However, the goal of such sensitivity analysis should be to better understand the balance between the data and the prior. It should not be done to ensure that final inference is reflective of the data only and not the prior. After all, our goal is to leverage the hierarchy and the external information to gain potential advantages with inference. In complex systems analysis of biological processes, we believe that a Bayesian perspective via hierarchical models is appropriate since these processes often follow a conceptual hierarchy, that is, SNPs within genes, genes within biochemical pathways, pathways within physiological processes, physiological processes within social networks. Furthermore, recent advances in technology now make it feasible to generate massive amounts of data aimed at measuring the elements in this hierarchy (i.e., genomics, proteomics, metabolomics). To integrate such data will require complex models that are reflective of the overall process, but simplistic and intuitive at each stage of construction.

References

- Baurley JW, Conti DV, Gauderman WJ, Thomas DC. Discovery of complex pathways from observational data. *Stat Med.* 2010; 29(19):1998–2011. [PubMed: 20683892]
- Best NG, Tan KK, et al. Estimation of population pharmacokinetics using the Gibbs sampler. *J Pharmacokinet Biopharm.* 1995; 23(4):407–435. [PubMed: 8882748]
- Bois FY. Applications of population approaches in toxicology. *Toxicol Lett.* 2001; 120(1–3):385–394. [PubMed: 11323198]
- Breslow NE, Clayton DG. Approximate inference in generalized linear mixed models. *J Am Stat Assoc.* 1993; 88:9–25.
- Chen JH, Chen ZH. Extended Bayesian information criteria for model selection with large model spaces. *Biometrika.* 2008; 95(3):759–771.
- Chen LS, Hutter CM, Potter JD, Liu Y, Prentice RL, Peters U, Hsu L. Insights into colon cancer etiology via a regularized approach to gene set analysis of GWAS data. *Am J Hum Genet.* 2010; 86(6):860–71. [PubMed: 20560206]
- Clewell HJ, Andersen ME, et al. A consistent approach for the application of pharmacokinetic modeling in cancer and noncancer risk assessment. *Environ Health Perspect.* 2002; 110:85–93. [PubMed: 11781169]
- Conti D, Gauderman W. SNPs, haplotypes, and model selection in a candidate gene region: The SIMPlE analysis of multilocus data. *Genet Epidemiol.* 2004; 27:429–441. [PubMed: 15543635]
- Conti DV, Witte JS. Hierarchical modeling of linkage disequilibrium: Genetic structure and spatial relations. *Am J Hum Genet.* 2003; 72(2):351–363. [PubMed: 12525994]
- Conti DV, Cortessis V, et al. Bayesian modeling of complex metabolic pathways. *Hum Hered.* 2003; 56(1–3):83–93. [PubMed: 14614242]
- Conti, DV.; Lewinger, JP., et al. Using ontologies in hierarchical modeling of genes and exposures in biologic pathways. In: Swan, GE., editor. *Phenotypes and Endophenotypes: Foundations for Genetic Studies of Nicotine Use and Dependence.* Vol. 20. NCI Tobacco Control; Bethesda, MD: 2009. p. 539-584.
- Cordell HJ. Detecting gene–gene interactions that underlie human diseases. *Nat Genet.* 2009; 10:392–404.
- Cortessis, V.; Thomas, DC. Toxicokinetic genetics: An approach to gene–environment and gene-gene interactions in complex metabolic pathways. In: Bird, P.; Boffetta, P.; Buffler, P.; Rice, J., editors. *Mechanistic Considerations in the Molecular Epidemiology of Cancer.* IARC Scientific Publications; Lyon, France: 2003.
- Davidian M, Gallant AR. Smooth nonparametric maximum likelihood estimation for population pharmacokinetics, with application to quinidine. *J Pharmacokinet Biopharm.* 1992; 20(5):529–556. [PubMed: 1287201]
- Efron B, Morris C. Data analysis using Stein’s estimator and its generalizations. *J Am Stat Assoc.* 1975; 70:311–319.
- Gelman, A. *Bayesian Data Analysis.* Chapman & Hall; London/New York: 1995.
- Gelman A, Bois F, et al. Physiological pharmacokinetic analysis using population modeling and informative prior distributions. *J Am Stat Assoc.* 1996; 91:1400–1412.
- George EI, McCulloch RE. Variable selection via Gibbs sampling. *JASA.* 1993; 88:881–889.
- Gilks, W.; Richardson, S., editors. *Markov Chain Monte Carlo in Practice.* Chapman & Hall; London: 1996.
- Goodman SN. Toward evidence-based medical statistics. 2: The Bayes factor. *Am Soc Intern Med.* 1999; 130:1005–1013.
- Greenland S. Methods for epidemiologic analyses of multiple exposures: A review and comparative study of maximum-likelihood, preliminary testing, and empirical-Bayes regression. *Stat Med.* 1993; 12:717–736. [PubMed: 8516590]
- Greenland S. Second-stage least squares versus penalized quasi-likelihood for fitting hierarchical models in epidemiologic analyses. *Stat Med.* 1997; 16:515–526. [PubMed: 9089960]

- Greenland S. Principles of multilevel modelling. *Int J Epidemiol.* 2000a; 29(1):158–167. [PubMed: 10750618]
- Greenland S. When should epidemiologic regressions use random coefficients? *Biometrics.* 2000b; 56(3):915–921. [PubMed: 10985237]
- Greenland S, Poole C. Empirical-Bayes and semi-Bayes approaches to occupational and environmental hazard surveillance. *Arch Environ Health.* 1994; 49(1):9–16. [PubMed: 8117153]
- Greenland S, Robins JM. Empirical-Bayes adjustments for multiple comparisons are sometimes useful. *Epidemiology.* 1991; 2:244–251. [PubMed: 1912039]
- Hung RJ, Brennan P, et al. Using hierarchical modeling in genetic association studies with multiple markers: Application to a case-control study of bladder cancer. *Cancer Epidemiol Biomarkers Prev.* 2004; 13(6):1013–1021. [PubMed: 15184258]
- Jefferys, WH.; Berger, JO. Technical Report #91-44C. Department of Statistics, Purdue University; 1991. Sharpening Ockham's Razor on a Bayesian Strop.
- Jeffreys, H. *Theory of Probability.* Oxford University Press; Oxford: 1961.
- Kass RE, Raftery AE. Bayes factors. *J Am Stat Assoc.* 1995; 90:773–795.
- Li D, Conti DV. Detecting gene–environment interactions using a combined case-only and case-control approach. *Am J Epidemiol.* 2009; 169(4):497–504. [PubMed: 19074774]
- Li W, Reich J. A complete enumeration and classification of two-locus disease models. *Hum Hered.* 2000; 50:334. [PubMed: 10899752]
- Lunn D, Best N, et al. Combining MCMC with 'sequential' PKPD modelling. *J Pharmacokinet Pharmacodyn.* 2009; 36(1):19–38. [PubMed: 19132515]
- Mitchell TJ, Beauchamp JJ. Bayesian variable selection in linear regression. *J Am Stat Assoc.* 1988; 83(404):1023–1032.
- Moore JH. The ubiquitous nature of epistasis in determining susceptibility to common human diseases. *Hum Hered.* 2003; 56(1–3):73–82. [PubMed: 14614241]
- Moore JH, Williams SM. Epistasis and its implications for personal genetics. *Am J Hum Genet.* 2009; 85(3):309–320. [PubMed: 19733727]
- Moore JH, Barney N, et al. Symbolic modeling of epistasis. *Hum Hered.* 2007; 63(2):120–133. [PubMed: 17283441]
- Morris C. Parametric empirical Bayes inference: Theory and applications (with discussion). *J Am Stat Assoc.* 1983; 78:47–65.
- Mukherjee B, Chatterjee N. Exploiting gene-environment independence for analysis of case-control studies: An empirical Bayes-type shrinkage estimator to trade-off between bias and efficiency. *Biometrics.* 2008; 64(3):685–694. [PubMed: 18162111]
- Park T, Casella G. The Bayesian Lasso. *J Am Stat Assoc.* 2008; 103(482):681–686.
- Racine-Poon A, Wakefield J. Statistical methods for population pharmacokinetic modelling. *Stat Methods Med Res.* 1998; 7(1):63–84. [PubMed: 9533262]
- Rebbeck TR, Martinez ME, et al. Genetic variation and cancer: Improving the environment for publication of association studies. *Cancer Epidemiol Biomarkers Prev.* 2004; 13(12):1985–1986. [PubMed: 15598750]
- Robert, CP.; Casella, G. *Monte Carlo Statistical Methods.* Springer; New York: 2004.
- Robins JM, Greenland S. The role of model selection in causal inference from nonexperimental data. *Am J Epidemiol.* 1986; 123(3):392–402. [PubMed: 3946386]
- Rothman, KJ.; Greenland, S. *Modern Epidemiology.* Lippencott-Raven; Philadelphia: 1998.
- Ruczinski I, Kooperberg C, et al. Logic regression. *J Comput Graph Stat.* 2003; 12:475–511.
- Searle, SR.; Casella, G., et al. *Variance Components.* Wiley; New York: 1992.
- Sebastiani P, Ramoni MF, et al. Genetic dissection and prognostic modeling of overt stroke in sickle cell anemia. *Nat Genet.* 2005; 37(4):435–440. [PubMed: 15778708]
- Sorenson, D.; Gianola, D. *Likelihood, Bayesian and MCMC Methods in Quantitative Genetics.* Springer; New York: 2002.
- Speigelhalter, DJ.; Thomas, A., et al. *WinBUGS Version 1.4 User Manual.* Medical Research Council Biostatistics Unit; Cambridge: 2003.

- Stephens M, Balding DJ. Bayesian statistical methods for genetic association studies. *Nat Genet.* 2009; 10:681–690.
- Stram DO, Pearce CL, et al. Modeling and E-M estimation of haplotype-specific relative risks from genotype data for a case-control study of unrelated individuals. *Hum Hered.* 2003; 55(4):179–190. [PubMed: 14566096]
- Tang W, Wu X, et al. Epistatic module detection for case-control studies: A Bayesian model with a Gibbs sampling strategy. *PLoS Genet.* 2009; 5(5):1–17.
- Thomas D. Methods for investigating gene-environment interactions in candidate pathway and genome-wide association studies. *Annu Rev Public Health.* 2010b; 31:21–36. [PubMed: 20070199]
- Thomas D. Gene-environment-wide association studies: emerging approaches. *Nat Rev Genet.* 2010a; 11(4):259–272. [PubMed: 20212493]
- Thomas DC, Siemiatycki J, et al. The problem of multiple inference in studies designed to generate hypotheses. *Am J Epidemiol.* 1985; 122:1080–1095. [PubMed: 4061442]
- Thomas D, Langholz B, Clayton D, Pitkaniemi J, Tuomilehto-Wolf E, Tuomilehto J. Empirical Bayes methods for testing associations with large numbers of candidate genes in the presence of environmental risk factors, with applications to HLA associations in IDDM. *Ann Med.* 1992; 24:387–392. [PubMed: 1418924]
- Thomas DC, Witte JS, et al. Dissecting effects of complex mixtures: Who's afraid of informative priors? *Epidemiology.* 2007a; 18(2):186–190. [PubMed: 17301703]
- Thomas PD, Mi H, et al. Ontology annotation: Mapping genomic regions to biological function. *Curr Opin Chem Biol.* 2007b; 11(1):4–11. [PubMed: 17208035]
- Thomas DC, Conti DV, et al. Use of pathway information in molecular epidemiology. *Hum Genomics.* 2010; 4(1):21–42. [PubMed: 21072972]
- Tibshirani R. Regression shrinkage and selection via the Lasso. *J R Stat Soc.* 1996; 58(1):267–288.
- Wakefield J. The Bayesian analysis of population pharmacokinetic models. *JASA.* 1996; 91:62–75.
- Wakefield J. A Bayesian measure of the probability of false discovery in genetic epidemiology studies. *Am J Hum Genet.* 2007; 81:208–227. [PubMed: 17668372]
- Wakefield, J.C.; Best, N.G., et al. Bayesian approaches to disease mapping. In: Elliot, P.; Wakefield, J.; Best, N.; Briggs, D., editors. *Spatial Epidemiology: Methods and Applications.* Oxford University Press; New York: 2000. p. 104-127.
- Wang K, Li M, et al. Pathway-based approaches for analysis of genomewide association studies. *Am J Hum Genet.* 2007; 81(6)
- Whittemore AS. A Bayesian false discovery rate for multiple testing. *J Appl Stat.* 2007; 34(1):1–9.
- Wilson MA, Iversen ES, et al. Bayesian model search and multilevel inference for SNP association studies. *Ann Appl Stat.* 2010
- Zellner, A. On assessing prior distributions and Bayesian regression analysis with g-prior distributions. In: BGoel, P.K.; Zellner, A., editors. *Bayesian Inference and Decision Techniques: Essays in Honor of Bruno de Finetti.* Elsevier; North-Holland: 1986. p. 233-243.
- Zhang Y, Liu JS. Bayesian inference of epistatic interactions in case-control studies. *Nat Genet.* 2007; 39(9):1167–1172. [PubMed: 17721534]
- Zhong H, Yang X, et al. Integrating pathway analysis and genetics of gene expression for genome-wide association studies. *Am J Hum Genet.* 2010; 86(4):581–591. [PubMed: 20346437]

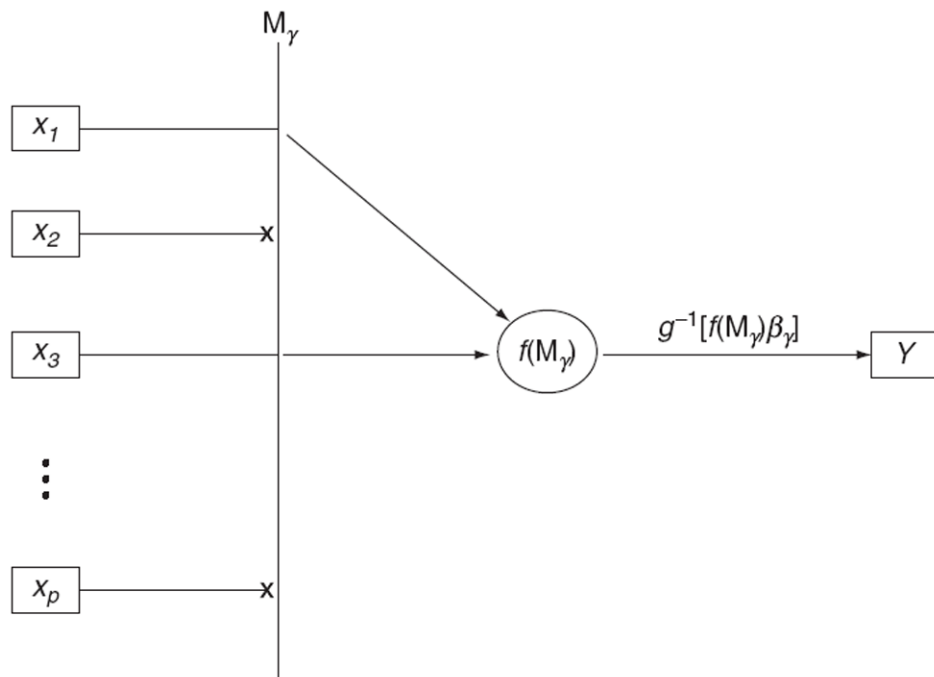


Figure 3.1.
Directed graph summarizing general model structure.

Table 3.1

Jeffreys Grades of Evidence (Jeffreys, 1961)

Grade	$\text{BF}(H_A:H_0)$	Evidence against H_0
1	1–3.2	Indeterminate
2	3.2–10	Positive
3	10–31.6	Strong
4	31.6–100	Very strong
5	>100	Decisive