

Application of Bayesian Linear Regression in Biomedical Research*

Chan Siew Pang *MSc(Management), MSc(Medical Statistics)*, Poh Kim Leng *PbD*¹

National Disease Registries Office, Health Promotion Board, Singapore

¹ Department of Industrial & Systems Engineering, National University of Singapore

ABSTRACT

Background. Primarily aimed at researchers with limited exposure to advanced statistics, this paper advocates the use of Bayesian linear regression analysis in biomedical research. Unlike the conventional approach, the Bayesian model allows analysis be conducted with prior information (i.e. expert opinion or published results) combined with sample information.

Methods. The regression coefficients (β) are treated as random and all statistical inferences are based on the posterior distribution. The posterior mean β^* is a vector-weighted average of the prior estimates β^0 and the conventional estimates $\hat{\beta}$ (based on maximum likelihood or ordinary least squares), with weights proportional to the precision and data matrices. In the case where there is no prior notion about β , a non-informative prior is fixed and it is not surprising that $\beta^* = \hat{\beta}$.

Results. Based on the simulated and selected real-life data sets, the properties of the Bayesian regression were illustrated. If the data shows a consistent pattern, the posterior will be dominated by the likelihood. The posterior, however, is strongly influenced by the prior if the likelihood is weak.

Discussion. Bayesian regression offers an alternative perspective to statistical modelling. When the prior information is a direct result of previous studies, or when it reflects no knowledge about the problem, Bayesian analysis becomes objective.

Conclusion. In evidenced-based biomedical research, expert opinion and published results form a rich source of information. By allowing such information be incorporated into analysis, the Bayesian model is found to be more versatile than the conventional model.

Keywords: Bayesian linear regression, conjugacy, likelihood, posterior, prior

BACKGROUND

Linear regression is one of the most widely-used statistical techniques in biomedical research. It is concerned with the study of the dependence of an outcome on one or more covariates (also known as explanatory variables or predictors). To study the pattern of dependence, one estimates the coefficients (β) which indicate the variation in the outcome with changes in the covariates. Data for regression analysis may be obtained from observational (retrospective, cross-sectional or prospective) or experimental studies.

The most distinguishing feature of conventional statistics is that researchers attempt to estimate the unknown β , considered to be non-random (constant over all observations), with no prior knowledge. The argument is that if the coefficients were known, there would be no need to make inferences about them. While it adheres to the objective principle of scientific investigation, such a framework might fail to provide a realistic and adequate approach in research. In certain applications, β is not totally unknown. For example, a large number of clinical trials, case-control analyses or cohort studies might have been conducted on a topic of similar nature. With such a large volume and preservation of published works, a well-informed

* Presented at the International Conference on Evidence-based Medicine held in Singapore on 18 to 19 October 2003.

researcher may not necessary carry out a specific investigation with a blank mind. Through literature reviews, the researcher might have acquired some useful knowledge about the topic before launching the study. Another major source of prior information comes in the form of expert opinion. Through years of professional practice, many researchers might have accumulated a large amount of knowledge useful for making intelligent guesses and providing solutions for problems that are not completely understood. Though subjective in nature, such expert opinion is helpful in planning for research studies. Statistically speaking, it is costly to ignore prior information, if available, as it may serve as a useful guide for conducting studies.

Motivated by the current need for evidence-based medicine, this paper aims to provide an introduction to Bayesian linear regression — an alternative modelling strategy — for biomedical researchers who have limited exposure to advanced statistical theories but are comfortable with running data analysis on commercial statistical software. As such, the mathematical properties of Bayesian regression are discussed with the aid of simulated data and 2 real-life examples drawn from published works.

Loosely speaking, the proposed Bayesian framework provides a mechanism for analysts to combine prior information with sample observations. Statistical analyses are based on the combined information embodied in a posterior distribution. An immediate implication of the Bayesian philosophy is that β becomes random. This idea is new to many biomedical researchers and is very much at odds with the conventional framework that β is a fixed quantity that can be accurately estimated with replicable procedures. Though widespread, this notion about β is contradictory to the way biomedical research is conducted. It is simply not possible to repeat experiments under identical conditions or re-expose clinical subjects to the same stimuli. This paper also addresses several issues related to the foundation of Bayesian analysis, including the perception that it is solely a subjective theory. Most controversies surrounding Bayesian analysis are caused by misunderstandings of the role of priors in data analysis.

METHODS

Linear Regression Model

A linear regression model with k covariates based on a sample of n observations is usually written as:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{ki} + u_i$$

where Y is the outcome, X_1, X_2, \dots, X_k the covariates, u the error term and $\beta = [\beta_0, \beta_1, \beta_2, \dots, \beta_k]'$ the set of unknown coefficients to be estimated. The subscript i refers to the i -th observation ($i=1,2,\dots,n$). There are altogether $k+1$ coefficients (β_0 : intercept; $\beta_1, \beta_2, \dots, \beta_k$: slopes) in the model. A slope coefficient, say β_1 , measures the effect of a 1 unit change of X_1 on Y , while holding other covariates constant. This is possible because the model takes the movements of X_2, \dots, X_k into account when it estimates β_1 . Note that the coefficients do not change from observation to observation, but the values of Y, X and u do.

While Y is a continuous outcome, X may be categorical, ordinal or continuous. The error term u , usually assumed to be normally distributed with constant variance σ^2 , specifies that the relationship between Y and X is not exact. The normality assumption is justified in many situations because the errors frequently represent the effects of the less important covariates omitted from the model. As such, u may also be thought of as a surrogate for all these covariates that collectively affect Y non-significantly. The term σ^2 is an important quantity in determining the precision of the estimated coefficients.

Combining all n observations, it is more illuminating to express the model in matrix form: $\mathbf{Y} = \mathbf{X}\beta + \mathbf{u}$.¹ A great advantage of such presentation is that it provides a compact and convenient way for handling regression models involving any number of covariates. In fact, all commercial statistical software read the entered data in this format.

There are 2 popular conventional methods for estimating β . To facilitate discussion, let the estimator of β be $\hat{\beta}$. In ordinary least squares (OLS), the default method employed in all commercial statistical software, $\hat{\beta}$ is obtained with the sum of squared errors minimised.² On the other hand, the maximum likelihood method (ML) generates estimates that are most likely to have caused the data to occur.³

It can be easily proved that both OLS and ML methods give $\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ in the context of normal error linear regression.⁴ Solely determined by observed data \mathbf{X} and \mathbf{Y} , this is an unbiased estimator of β , i.e., its average value is equal to the true value β . Since the OLS and ML estimators of β are identical, no explicit difference between OLS and ML is made in this paper. While more restrictive than OLS, the ML method ensures that $\hat{\beta}$ possesses some desirable properties. With a large sample, $\hat{\beta}$ approaches the true value of β and its distribution also becomes normal with the highest possible precision.^{5,6} These properties help to facilitate

inferences about β . While regression analysts are usually more familiar with OLS, the ML principle is the cornerstone of most statistical methods. A ML estimator is one that is best supported by the data embodied in a likelihood function

The likelihood is a measure of the support provided by the data for particular coefficient values of a probability model. It is useful for the development of many complicated regression models, including the proposed Bayesian model to be introduced next.

Bayesian Setup

The conventional estimation techniques treat β as a fixed quantity. The underlying principle of the proposed Bayesian framework is conceptually different. The available prior information of β is used in the estimation process and the immediate implication of such an approach is that the unknown β is viewed as the value of a random variable. The problem of estimating β is thus changed to the estimation of the value of the random variable β .⁷

Following the Bayes' theorem, the proposed Bayesian model may be implicitly formulated as:⁸

$$P[\beta, \sigma^2 \mid \text{data}] \propto L[\beta, \sigma^2 \mid \text{data}] \times g[\beta] \times g[\sigma^2]$$

where $L[\beta, \sigma^2 \mid \text{data}]$ is the likelihood of observing the data (\mathbf{X}, \mathbf{Y}) given β and σ^2 ; $g[\beta]$ and $g[\sigma^2]$ are the prior distributions concerning β and σ^2 before observing the data and $P[\beta, \sigma^2 \mid \text{data}]$ is the posterior distribution of β and σ^2 after observing the data. The post-data inference for β is a compromise between prior information and the information provided by the observed data (likelihood). In conventional statistics, β is estimated by the value that maximises $L[\beta, \sigma^2 \mid \text{data}]$. Bayesian estimation, on the other hand, is based on the posterior which includes both the prior information and the data captured in the likelihood. All information required for Bayesian analysis is described by the posterior distribution.

The basic paradigm of Bayesian statistics is straightforward. Initial information concerning the coefficient of interest (β), which could be based on objective evidence (past studies) or subjective judgement or a combination of both, are expressed as a prior distribution. Summarising what has been learnt about β , the prior carries important information about the coefficients. Evidence from further data is then summarised in the likelihood function for β . By combining the prior and the likelihood, one generates the posterior of which inference of β is drawn. Thus, one may view the Bayesian set-up as the posterior-

generating process based on a revision of prior information with observed data. Beginning with some prior information of β , one revises the knowledge of β with collected data. Intuitively, this reflects how knowledge is accumulated.

Although the prior is primarily used to combine with a likelihood to generate the posterior, it has other useful applications. Priors can be used in the design and monitoring of a clinical trial.⁹ It is also useful for reviewing the ethical aspects of a study as it documents probabilistically the current knowledge and experts' opinions concerning the proposed treatments. Priors allow available information, say in the magnitude of the effects of a treatment, be stated in a form that others can examine and compare with their own assessment.¹⁰ Details about the various methods useful for eliciting priors in the clinical setting can be found in references.¹⁰⁻¹³

In conventional statistics, the only source of uncertainty admitted to analysis is sampling uncertainty. A fundamental advantage for applying Bayesian statistics to biomedical research is that both prior and posterior estimates are described in probabilistic terms and therefore offers a more realistic procedure for dealing with the myriad sources of uncertainty faced by researchers in real applications.

Estimation of Posterior Coefficients

It follows that the choice of distributional forms for priors and likelihood is a critical feature of Bayesian analysis. In normal error linear regression, the likelihood is based on the familiar normal distribution. Describing in probabilistic terms what information gathered by the analyst before observing the data, the prior contains "known" quantities along with the normal likelihood to make inferences about the unknown coefficients.

It is well known that the posterior distribution of β might not have an analytically tractable form if the priors are freely chosen. A way to guarantee that the posterior has a calculable form is to specify a conjugate prior. When the posterior has the same distributional family as the prior, one says that the prior and the likelihood distributions are conjugate. In a normal error regression model, the most appropriate prior distributions are multivariate normal for β and inverse gamma for σ^2 . In statistical terms, $\beta \sim \text{MN}[\beta^0, \Sigma]$ and $\sigma^2 \sim \text{IG}[\alpha, \phi]$ where β^0 is the mean prior value for β , Σ the covariance matrix of β and α and ϕ are the shape and scale parameters of the inverse gamma distribution, respectively. In practice, specific prior

values are assigned once the distributions are determined. The values may be elicited from a panel of experts or pre-existing scientific knowledge (such as well-known published results and meta-analysis of similar studies). In the latter case, the following Bayesian analysis is “objectified” with the use of an objective prior.

Putting the priors and likelihood together, one obtains the Bayesian posterior estimator $\beta^* = (\Sigma^{-1} + \mathbf{X}'\mathbf{X})^{-1} (\Sigma^{-1}\beta^0 + \mathbf{X}'\mathbf{X}\hat{\beta})$, where $\mathbf{X}'\mathbf{X}$ is the data matrix and Σ^{-1} (precision) is the inverse of Σ .¹⁴ As readily seen, β^* is a weighted average of the prior coefficients and the conventional estimators, with weights provided by $\mathbf{X}'\mathbf{X}$ and Σ^{-1} . If one has reliable information concerning β , a precise prior may be fixed. In this case, the prior distribution is highly concentrated. As the prior precision increases, a greater posterior weight on prior beliefs is placed relative to the data. The derivation of the precision matrix for β^* can be found in reference.¹⁴

While adequately elegant and computationally simple for expressing an analyst’s opinion, it may be worthwhile to point out that the use of conjugate priors has no real theoretical advantage. Before 1990s, conjugacy was crucial to the ability to apply Bayesian methods because non-conjugate priors usually lead to posteriors that are not analytically tractable. With the advent of the Markov Chain Monte Carlo (MCMC) techniques, this limitation becomes greatly reduced.^{15,16} However, conjugacy offers a framework that leads to a systematic approach of prior determination. Consequently, conjugate priors can still be useful in practice and they are an excellent expository tool.¹⁷

In situations where there is no prior information about β and σ^2 , one may use constants, i.e. $g[\beta] \propto c$ and $g[\sigma^2] = 1/\sigma$. The former term, $g[\beta] \propto c$, is a flat distribution which may be thought of as a normal density with very heavy tails. Not surprisingly, the estimator of β is identical to $\hat{\beta}$, produced by the conventional linear regression model. This is because with no prior knowledge about β , the posterior is solely based on information provided by the likelihood. The idea of fitting a non-information prior is equivalent to conventional statistical analysis, although the interpretation is philosophically different. Thus, Bayesian analysis is applicable in situations where there is no available prior knowledge of the regression coefficients.

RESULTS

In fitting a Bayesian normal error linear regression model, one needs to fix 4 sets of prior quantities,

namely β^0 , Σ , α and ϕ . There are 2 prior distributions involved because both β and σ^2 are random. With conjugate normal-inverse gamma priors, the resultant posterior β^* is a vector-weighted average of the prior mean β^0 and the conventional estimator $\hat{\beta}$. The weights are proportional to the prior precision and data matrices. If the likelihood is strong, i.e. the data suggests a consistent relational pattern between the covariates (\mathbf{X}) and the outcome (\mathbf{Y}), then the posterior will be dominated by the likelihood. Similarly, if the prior is non-informative, as in the case of a flat distribution, the likelihood again dominates. On the other hand, the prior dominates if it is highly-concentrated, while the likelihood is weak.

To illustrate the properties of Bayesian linear regression, 2 simulated data sets were generated. Simulation is a very useful method for examining the properties of quantitative models. Statisticians frequently run simulated examples with desired features so that the observed results can be compared with the expected results. In addition, 2 well-known published examples based on small samples were also considered. The computations were carried out with Stata 7.0 (Stata Corporation, Texas, USA).

A Simulated Example with Strong Likelihood

With Y as the outcome, the data matrix \mathbf{X} consists of 2 covariates, namely X_1 and X_2 . While X_1 and Y are continuous (admissible range: 0 to 100), X_2 is a dummy variable with 2 categories. The simulated data with 1000 observations were constructed such that the adjusted R^2 (measuring the explanatory power of the model) is close to 0.8. In addition, there is a strong correlation between Y and X_1 , but X_1 and X_2 are mutually independent. One may imagine the proposed simulated example as a pre- and post-study design with Y and X_1 representing the post-intervention and baseline values of the outcome, respectively. The dummy variable X_2 indicates the type of treatment received by the subjects (1: Treatment A; 2: Treatment B). The objective is to compare the difference in the average values of Y for the 2 treatments, while adjusting for baseline differences. Assume here that better prognosis is characterised by higher values of Y. A coefficient is said to be precisely estimated if its standard error is small.

The results based on conventional regression model are shown in Table 1. After adjusting for X_1 , the difference in average values of Y between Treatment B and Treatment A is estimated to be about 8.8 units. As such, there is sufficient evidence suggesting that Treatment B is more effective than Treatment A. The

Table 1. Linear regression models with strong likelihood.

	Conventional Model		Bayesian Model (Non-informative)		Bayesian Model (Informative)	
	Coefficient	Standard Error	Coefficient	Standard Error	Coefficient	Standard Error
X ₁	0.77	0.01	0.77	0.01	0.88	0.01
X ₂						
Treatment A	–	–	–	–	–	–
Treatment B	8.79	0.28	8.79	0.28	6.90	0.27
Intercept	5.80	0.91	5.80	0.91	1.80	0.47

n: 1000

Table 2. Linear regression models with weak likelihood.

	Conventional Model		Bayesian Model A		Bayesian Model B	
	Coefficient	Standard Error	Coefficient	Standard Error	Coefficient	Standard Error
X ₃	-1.65	3.46	0.67	0.57	-4.19	0.67
Intercept	43.48	11.93	39.92	0.64	40.23	0.74

n: 20

errors are confirmed to be normally-distributed, so the application of normal error regression is appropriate.

The following Bayesian analysis is based on $\sigma^2 \sim \text{IG}[\alpha=3, \phi=3]$ and $\beta \sim \text{MN}[\beta^0, \Sigma]$ where $\Sigma := \text{diagonal}_{3 \times 3}(10)$ and $\beta^0 = [0, 0, 0]'$. The 3 elements in β^0 refer to the prior slope coefficients of X₁, X₂ and the intercept coefficient, respectively. It suggests that there is no difference between the 2 treatments. The prior precision Σ is a 3×3 matrix with 10's on the diagonal and zeros elsewhere. The standard errors of the prior coefficients can be computed by taking square root of the diagonal elements of Σ . With heavy tails (low precision with large standard errors), the prior of β resembles the non-informative distributions suggested earlier. From Table 1, the posterior coefficients and their respective standard errors are identical to those obtained from the conventional model.

Next, suppose again $\sigma^2 \sim \text{IG}[3, 3]$, but $\Sigma := \text{diagonal}_{3 \times 3}(0.01)$ and $\beta^0 = [0, 0, 0]$. The normal prior of β may be appropriately termed as “informative” as it is much more concentrated than that of the previous case. As shown in Table 1, the prior only exerts some slight pull on the posterior. This phenomenon can be again explained by the presence of a strong likelihood.

The above simulation confirms that when a strong likelihood (partly due to a very large sample size) exists, the prior's influence is relatively insignificant.

Moreover, there is no practical difference between the Bayesian and the conventional models if non-informative priors are specified.

A Simulated Example with Weak Likelihood

The next example differs from the previous one in 2 aspects. First, a simple regression model with one predictor is fitted to the data. The concepts developed for k-variable multiple regression can be easily modified and applied in the simple case. Second, the likelihood is relatively weak (with $R^2 < 0.01$) and the sample consists of only 20 observations. Assume again that a good prognosis is characterised by higher values of Y.

Assume that there is some prior information (possibly provided by an expert through literature review or meta-analysis) regarding the parameters, say $\beta^0 = [40, 1]'$ and $\Sigma := \text{diagonal}_{2 \times 2}(10^{-3})$. This highly-concentrated “enthusiastic” prior suggests that a 1 unit increase in the predictor would lead to a 1 unit increase in the outcome. The inverse gamma distribution for σ^2 is fixed as $\text{IG}[3, 3]$. As shown in Table 2, the posterior coefficients of Bayesian Model A are closer to the priors than the conventional estimates. Next, the analysis is repeated with a “pessimistic” prior $\beta^0 = [40, -5]'$, thus suggesting that a high value in the predictor would lead to poorer prognosis. Again, the priors dominate the posteriors (Table 2, Model B). In both cases, the posterior coefficients are more precisely

Table 3. Linear regression models on polyp size.

	Conventional Model		Bayesian Model	
	Coefficient	Standard Error	Coefficient	Standard Error
Baseline polyp size (mm)	0.21	0.24	0.60	0.16
Treatment				
0: Placebo	–	–	–	–
1: Sulindac	-1.29	0.51	-0.28	0.40
Intercept	2.42	0.88	0.31	0.44

n: 19

Table 4. Linear regression models on RMR.

	Conventional Model		Bayesian Model	
	Coefficient	Standard Error	Coefficient	Standard Error
Body Weight (kg)	6.83	1.11	6.86	0.13
Intercept	816.12	82.96	851.97	0.14

n: 32

estimated. This could be explained by the dominance of the precise priors.

Testing Sulindac Effects in Reducing Polyp Size

The third example concerns a well-known randomised, double-blind, placebo-controlled trial testing whether or not sulindac, a non-steroid anti-inflammatory agent, could reduce the size of colonic polyps in patients with familial adenomatous polyposis (FAP).¹⁸ Caused by a germ-line mutation in the adenomatous polyposis coli (APC) gene, FAP is a colon cancer predisposition syndrome characterised by the formation of a large number of precancerous colonic polyps (also called adenomas) and eventual colorectal cancer.^{18,19} The mechanism by which sulindac causes polyp regression is unknown.²⁰

Altogether, 22 patients were randomised to receive sulindac (150mg orally twice daily for 9 months) or placebo in this trial. The complete listing of data can be found in reference.²¹ In the following analysis, only the 12-month data are considered and the primary end-point is polyp size (mm). As in the simulated examples, baseline polyp size is included as one of the predictors in the analysis.

The conventional model shows that sulindac treatment was effective in reducing the polyp size at 12 months (Table 3). The average polyp size for patients receiving sulindac is 1.29mm lower than those receiving placebo, after adjusting for the baseline measurements. The

analysis is based on 19 observations (sulindac: 9; placebo: 10) as there are no data recorded for 3 patients at 12 months.

However, there are mixed findings from other studies.^{19,20,22,23} To be conservative, consider a “sceptical” prior $\beta^0 = [0, 0, 0]'$ with $\Sigma := \text{diagonal}_{3 \times 3} (10^{-1})$ for the following Bayesian analysis. This prior suggests that sulindac is not effective in reducing polyp size. Moreover, baseline polyp size is also not associated with the subsequent measurements at 12 months. The inverse gamma distribution for σ^2 is fixed as IG[3, 3]. As shown in Table 3, the posterior coefficients are closer to the priors, with slightly smaller standard errors.

Thus, the Bayesian model shows that there could be no sufficient evidence suggesting that sulindac is effective in reducing polyp size for patients with FAP. The drug may not have the colon cancer prevention properties once hoped for.

Modelling Resting Metabolic Rate

The last example attempts to relate resting metabolic rate (RMR; kcal/24 hr) with body weight (kg) in 44 healthy adult women based on published data.^{24,25} RMR is the minimum level of energy to sustain the body's vital functions, not including the energy cost of digesting and absorbing, or engaging in physical activity.

As there was only one such study based on MEDLINE search, a Bayesian analyst might model the observations with non-informative priors. However, in order to illustrate the use of published results for formulating priors, the example is modified with the original sample randomly divided into 2 sub-samples. Consisting of only 12 observations, the first sub-sample (training) is used to generate the prior parameters required for the proposed Bayesian analysis. Then, the results are combined with the remaining 32 observations in the second sub-sample (testing), which forms the likelihood. This strategy is often used in data analysis, especially for model evaluation and sample size determination.^{26,27} However, it must be emphasised that it is applied here for illustration only.

Based on the observations randomly allocated to the training sub-sample, the data-based prior β^0 turns out to be [851.97, 6.95]'. This suggests that the RMR would increase by about 6.95 kcal/24 hr with a 1kg increase in body weight. In addition, let $\Sigma := \text{diagonal}_{2 \times 2} (10^{-6})$ and $\sigma^2 \sim \text{IG}[3, 3]$. The priors are then combined with the likelihood based on observations in the testing sub-sample. As depicted in Table 4, the posterior coefficient associated with body weight is closer to the conventional estimate. However, it is more precisely estimated with the presence of a highly-concentrated prior distribution.

DISCUSSION

The Bayesian approach has begun to establish itself as a powerful and very practical tool for statistical analysis. However, it must be emphasised that the central ideas of Bayesian statistics are not new. Looking back in history, one realises that the development of Bayesian statistics actually parallels that of the conventional theories. So why do the conventional statistics dominate Bayesian usage in analysis?

There are several reasons for this. First, several prominent figures in the development of modern statistics had strong prejudices against the Bayesian ideas. This was largely caused by the misunderstanding of the nature of priors. In the case where the prior information reflects the personal opinions of individual investigators conducting the research, or possibly those of an expert who has immense knowledge of the subject matter, Bayesian statistics is subjective in nature. However, when the prior information is a direct result of previous studies, or when prior information reflects no knowledge about the problem at hand (non-informative), the Bayesian analysis becomes objective. As such, it is unfounded to allege that Bayesian theory is purely subjective. The

prior is also not deterministic in nature as it is described in probabilistic terms.

Another important reason for the dominance of conventional statistics lies with the complexity of Bayesian analysis. The development of a Bayesian model requires complicated modelling specifications and tedious computations. In the case of linear regression with conjugate priors, the estimators for regression coefficients can be derived analytically after some tedious mathematical manipulations. However, this is generally not the case for most Bayesian analyses. As described previously, the heart of Bayesian analysis lies with the posterior distribution which is derived from the likelihood and the prior. Usually, such posterior is very complicated and mathematically intractable. Thus, analysts must rely on advanced simulation techniques for finding the solutions.

As such, there is no accident that the recent rejuvenation of Bayesian statistics coincides with the development of computer-intensive statistics. With the wide applicability of high-speed computers, the new ideas offered by both branches of statistics have captured the imagination of researchers. These methods have a wide range of potential applications, especially in biomedical research, as a result of the increasing complexity of problems and data structures. Their analysis and refinement will be a formidable prospect for the statistical community in the coming years. It must be emphasised that while computers can never be as wise as people, they can explore a forest of possibilities faster than we can comprehend.

In practice, researchers may write their own computer programmes or apply user-written macros or available software (such as First Bayes and WinBUGS) for conducting Bayesian analysis.^{28,29} One may write a programme for conducting the above-mentioned Bayesian linear regression with the aid of Stata.³⁰⁻³² However, a more complete and user-friendly commercial software is clearly needed to promote the use of Bayesian statistical methods. There is an urgent need to develop a general-purpose software package that incorporates most Bayesian models useful for biomedical research. Its creation should thus be a high priority for the statistical profession.

CONCLUSION

In evidenced-based biomedical research, expert opinion and published results form a rich source of information. As such, one expects the Bayesian regression model to provide more insights to biomedical problems with reliable prior information.

As illustrated, it is very useful when prior information regarding the unknown coefficients of interest is available. However, this paper also shows that when such prior information is not available, the proposed regression model is reduced to the conventional model estimated by maximum likelihood (ML) or ordinary least squares (OLS). As such, one may view the conventional model as a special case within the broader framework of Bayesian statistics. Thus, the revolutionary Bayesian model is found to be more versatile than the conventional model.

ACKNOWLEDGEMENTS

We would like to express our appreciation to all the delegates who attended our presentation at the International Conference on Evidence-based Medicine held in Singapore on 18 October 2003. Their enlightening suggestions during the Q&A session helped to improve the draft. Special thanks also go to the anonymous reviewers for their advice and insightful critiques.

REFERENCES

- Neter J, Kutner MH, Nachtsheim CJ, Wasserman W. Applied Linear Statistical Models. Boston: McGraw-Hill, 1996;176-214.
- Gujarati DN. Basic Econometrics. Boston: McGraw-Hill, 2003;58-65.
- Wackerly DD, Mendenhall W, Scheaffer RL. Mathematical Statistics with Applications. USA: Duxbury, 2002;448-52.
- Bickel PJ, Doksum KA. Mathematical Statistics: Basic Ideas and Selected Topics. New Jersey: Prentice Hall, 2001;120-1.
- Serfling RJ. Approximation Theorems of Mathematical Statistics. Singapore: Wiley, 1980;143-9.
- Mood AM, Graybill FA, Boes DC. Introduction to the Theory of Statistics. Singapore: McGraw-Hill, 1974;358-62.
- Papoulis A, Pillai SU. Probability, Random Variables and Stochastic Processes. Singapore: McGraw-Hill, 2002;317.
- Migon HS, Gamerman D. Statistical Inference: An Integrated Approach. London: Arnold, 1999;26-7.
- Carlin B, Chaloner K, Church T, Matts JP, Louis TA. Bayesian monitoring of an AIDS clinical trial. *Stat* 1993; 42:355-67.
- Kadane JB, Wolfson LJ. Priors for the design and analysis of clinical trials. In: Berry DA, Stangl DK, editors. *Bayesian Biostatistics*. New York: Marcel Dekker, 1996;157-84.
- Freedman LS, Spiegelhalter DJ. The assessment of subjective opinion and its use in relation to stopping rules for clinical trials. *Stat* 1983; 32:153-60.
- Spiegelhalter DJ, Freedman LS. A predictive approach to selecting the size of a clinical trial, based on subjective clinical opinion. *Stat Med* 1986; 5:1-13.
- Chaloner K, Church T, Matts JP, Louis TA. Graphical elicitation of a prior distribution for an AIDS clinical trial. *Stat* 1993; 42:341-53.
- Maddala GS. *Econometrics*. Singapore: McGraw Hill, 1977;412-8.
- Casella G, George EI. Explaining the Gibbs sampler. *Am Stat* 1992; 46:167-74.
- Chib S, Greenberg E. Understanding the Metropolis-Hastings algorithm. *Am Stat* 1995; 49:327-35.
- Gill J. *Bayesian Methods: A Social and Behavioral Sciences approach*. Boca Raton: Chapman & Hall, 2002;120.
- Giardiello FM, Hamilton SR, Krush AJ, Piantadosi S, Hylind LM, Celano P, et al. Treatment of colonic and rectal adenomas with sulindac in familial adenomatous polyposis. *N Engl J Med* 1993; 328:1313-6.
- Giardiello FM, Yang VW, Hylind LM, Krush AJ, Petersen GM, Trimpath JD, et al. Primary chemoprevention of familial adenomatous polyposis with sulindac. *N Engl J Med* 2002; 346:1054-9.
- Pasricha PJ, Bedi A, O'Connor K, Rashid A, Akhtar AJ, Zahurak M, et al. The effects of sulindac on colorectal proliferation and apoptosis in familial adenomatous polyposis. *Gastroenterology* 1995; 109:994-8.
- Piantadosi S. *Clinical Trials: A Methodological Perspective*. New York: Wiley, 1997;479-81.
- Ladenheim J, Garcia G, Titzer D, Herzenburg H, Lavori P, Edson P, et al. Effect of sulindac on sporadic colonic polyps. *Gastroenterology* 1995; 108:1083-7.
- Giardiello FM, Offerhaus JA, Tersmette AC, Hylind LM, Krush AJ, Brensinger JD, et al. Sulindac induced regression of colorectal adenomas in familial adenomatous polyposis: evaluation of predictive factors. *Gut* 1996; 38:578-81.
- Owen OE, Kavle E, Owen RS, Polansky M, Caprio S, Mozzoli MA, et al. A reappraisal of caloric requirements in healthy women. *Am J Clin Nutr* 1986; 44:1-19.
- Altman DG. *Practical Statistics for Medical Research*. Boca Raton: Chapman & Hall, 2000;322-3.
- Sharma S. *Applied Multivariate Techniques*. USA: John Wiley, 1996;273-4.
- Desu MM, Raghavarao D. *Sample Size Methodology*. Boston: Academic Press, 1990;4-5.
- Albert JH. *Bayesian Computation using Minitab*. Belmont: Duxbury, 1996.
- Berger JO. Bayesian analysis: a look at today and thoughts of tomorrow. In: Raftery AE, Tanner MA, Wells MT, editors. *Statistics in the 21st Century*. Boca Raton: Chapman & Hall, 2000;275-90.
- Gould W, Sribney W. *Maximum Likelihood Estimation with STATA®*. Texas: Stata Press, 1999.
- Stata Corporation. *Stata Programming Manual (Release 7)*. Texas: Stata Press, 2001.
- Hamilton LC. *Statistics with STATA*. Toronto: Duxbury, 2003;297-328.