

Parameter estimation of the beta-binomial distribution: an application using the SAS software

Estimação dos parâmetros da distribuição beta-binomial: uma aplicação usando o software SAS

Edson Zangiacomi Martinez^{*1}, Jorge Alberto Achcar¹, and Davi Casale Aragon²

¹Departamento de Medicina Social, Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo

²Departamento de Puericultura e Pediatria, Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo

Abstract

In this paper we describe the parameter estimation of the beta-binomial distribution using the procedure NLMIXED of the SAS software. The beta-binomial distribution is a discrete mixture distribution which can capture overdispersion in the data. The estimation of parameters of the beta-binomial distribution can lead to computational problems, since it does not belong to the exponential family and there are not explicit solutions for the maximum likelihood estimation. Using a real dataset, we show that the SAS software can be satisfactorily used for the estimation of the parameters. We also consider the possibility of including a covariate in the model. The SAS codes used in this paper are given in an Appendix.

Keywords: beta-binomial distribution, regression model, data analysis.

Resumo

Neste artigo nós descrevemos a estimação dos parâmetros da distribuição beta-binomial usando o procedimento NLMIXED do software SAS. A distribuição beta-binomial é uma distribuição de misturas discreta capaz de capturar a superdispersão dos dados. A estimação dos parâmetros de uma distribuição beta-binomial pode oferecer problemas computacionais, dado que ela não pertence a uma família exponencial e não há soluções explícitas para o método da máxima verossimilhança. Usando dados reais, nós mostramos que o software SAS pode ser satisfatoriamente usado para a estimação dos parâmetros. Nós também consideramos a possibilidade de incluir uma covariável no modelo. As linhas de comando SAS usadas neste artigo não disponibilizadas em um anexo.

Palavras-chave: distribuição beta-binomial, modelo de regressão, análise de dados.

*Corresponding author: edson@fmrp.usp.br

1 Introduction

In some applications of Bernoulli trials, the underlying success probability could change from one trial to another, that is, likelihood based techniques should be modified. This is common in situations where there are unmodeled influences that affect all the components of the binomial sum which count the number of successes in a fixed number n of Bernoulli trials. This happens, for example, in biological or agriculture applications, when we have batch effects (see for example, Kleinman (1978); Williams (1975); Crowder (1978); Haseman and Kupper (1979) or Morgan (1992)). A litter effect is related to the tendency that members of a group respond in a more similar way to some treatment than members of other groups. These random effects should be included in the modelling of the data set besides the usual covariates to account for this overdispersion.

The beta-binomial model was introduced by Pearson (1925) and more formally described by Skellam (1948). It is a popular method for explicitly account for the overdispersion. We can find several applications of this model in various areas, such as Chatfield and Goodhardt (1976), who described the buying behaviour of the consumer, and Gange et al. (1996), who studied the effect of policy changes on appropriateness of hospital admissions. In addition, Aeschbacher et al. (1977) showed that a beta-binomial distribution provided a better fit than the usual distributions in biological experiments using mice when the data used were based on a large number of counts of dead.

The parameter estimation of the beta-binomial model using the maximum likelihood method brings some challenges, since there are not explicit solutions for the maximum likelihood estimation (MLE) and it is necessary to use iteration methods. In addition, this distribution is not a member of an exponential family. Thus, the main objective of this study is present an application of the SAS software to estimate the parameters of the beta-binomial model. The SAS software is widely used in many general purpose applications in academia, industry and health literacy studies. It provides powerful tools for data analysis including linear and generalized linear models, and random and mixed effects models. We also consider the possibility of including a covariate in the model. A real dataset illustrate the methodology.

2 The beta-binomial distribution

Let y be the number of occurrences of a random variable Y in n Bernoulli trials with success probability p ($0 < p < 1$). Thus,

$$P(Y = y|n,p) = \binom{n}{y} p^y (1-p)^{n-y}, \quad y = 0, 1, \dots, n,$$

where n is known. The mean and variance of $Y|n,p$ are given, respectively, by $E(Y|n,p) = np$ and $Var(Y|n,p) = np(1-p)$. Using the principles of Bayesian inference, let us suppose that p is a random variable that follows a beta distribution with parameters a and b . Thus,

$$f(p|a,b) = \frac{1}{B(a,b)} p^{a-1} (1-p)^{b-1},$$

where $a > 0$, $b > 0$ and $B(a,b)$ is the beta function. From these expressions, we have

$$\begin{aligned} P(Y = y|n,a,b) &= \int_0^1 P(Y = y|n,p) f(p|a,b) dp \\ &= \binom{n}{y} \frac{1}{B(a,b)} \int_0^1 p^{y+a-1} (1-p)^{n-y+b-1} dp \\ &= \binom{n}{y} \frac{B(y+a, n-y+b)}{B(a,b)}. \end{aligned}$$

This expression is the probability function of a beta-binomial distribution. When a random variable Y is distributed according a beta-binomial distribution, we write $Y|n,a,b \sim \text{betabin}(n,a,b)$. The mean and the variance of this random variable are given by

$$E(Y|n,a,b) = E[E(Y|n,a,b,p)] = nE(p|a,b) = \frac{na}{a+b} \quad (1)$$

and

$$\begin{aligned} Var(Y|n,a,b) &= E[Var(Y|n,a,b,p)] + Var[E(Y|n,a,b,p)] \\ &= nE(p|a,b) - nE(p^2|a,b) + n^2 Var(p|a,b) \\ &= \frac{nab(a+b+n)}{(a+b)^2(a+b+1)}, \end{aligned} \quad (2)$$

respectively. Graphs of the probability mass function of a beta-binomial distribution for $n = 10$ and some values of a and b are shown in Figure 1. These graphs show that the probability function can assume different shapes.

A vector of covariates $\mathbf{X} = (X_1, \dots, X_k)'$ can be included in the model by replacing the parameter a by a function $a(\mathbf{x})$ given by

$$a(\mathbf{x}) = \exp(a_0 + a_1 x_1 + \dots + a_k x_k),$$

where a_0, a_1, \dots, a_k are unknown parameters.

An useful parametrization of this model considers $a = \theta\tau^{-1}$ and $b = (1-\theta)\tau^{-1}$, where $\tau > 0$ and $0 < \theta < 1$. Under this condition, the mean and the variance of $Y|n,\theta,\tau$ are given by

$$E(Y|n,\theta,\tau) = n\theta \quad (3)$$

and

$$Var(Y|n,\theta,\tau) = n\theta(1-\theta) \left[1 + (n-1) \frac{\tau}{1+\tau} \right], \quad (4)$$

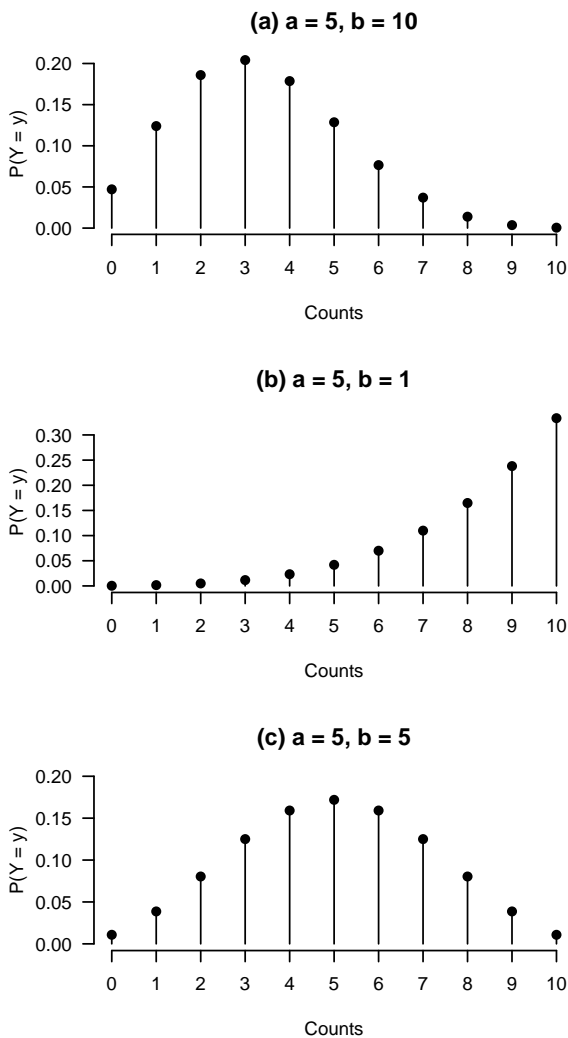


Figure 1: Graphs of the probability mass function of a beta-binomial distribution with $n = 10$ and (a) $a = 5$, $b = 10$, (b) $a = 5$, $b = 1$ and (c) $a = 5$, $b = 5$.

respectively. The parameter τ is interpreted as an overdispersion parameter, so that when $\tau = 0$ the variance (4) is equivalent to the variance of a random variable that follows a binomial distribution. The probability function is now given by

$$P(Y = y|n, \theta, \tau) = \binom{n}{y} \frac{B\left(y + \frac{\theta}{\tau}, n - y + \frac{1 - \theta}{\tau}\right)}{B\left(\frac{\theta}{\tau}, \frac{1 - \theta}{\tau}\right)}$$

$$= \binom{n}{y} \frac{\Gamma\left(y + \frac{\theta}{\tau}\right) \Gamma\left(n - y + \frac{1 - \theta}{\tau}\right) \Gamma\left(\frac{1}{\tau}\right)}{\Gamma\left(\frac{\theta}{\tau}\right) \Gamma\left(\frac{1 - \theta}{\tau}\right) \Gamma\left(\frac{1}{\tau} + n\right)} \quad (5)$$

Note that, given $c > 0$, $d > 0$ and k integer and

greater than zero, is valid the relation

$$\prod_{j=0}^k (d + jc) = c^{k+1} \frac{\Gamma\left(k + \frac{d}{c} + 1\right)}{\Gamma\left(\frac{d}{c}\right)}.$$

Thus,

$$\frac{\prod_{j=0}^{y-1} (\theta + j\tau) \prod_{j=0}^{n-y-1} (1 - \theta + j\tau)}{\prod_{j=0}^{n-1} (1 + j\tau)}$$

$$= \frac{\Gamma\left(y + \frac{\theta}{\tau}\right) \Gamma\left(n - y + \frac{1 - \theta}{\tau}\right) \Gamma\left(\frac{1}{\tau}\right)}{\Gamma\left(\frac{\theta}{\tau}\right) \Gamma\left(\frac{1 - \theta}{\tau}\right) \Gamma\left(\frac{1}{\tau} + n\right)},$$

and the expression (5) is rewritten in the form

$$P(Y = y|n, \theta, \tau) = \binom{n}{y} \frac{\prod_{j=0}^{y-1} (\theta + j\tau) \prod_{j=0}^{n-y-1} (1 - \theta + j\tau)}{\prod_{j=0}^{n-1} (1 + j\tau)}$$

(see, for example, Williams (1975) or Smith (1983)). Given N independent random variables Y_1, \dots, Y_N , the logarithm of the likelihood function for θ and τ is given by

$$l(\theta, \tau) = \sum_{i=1}^N \ln \binom{n}{y_i} + \sum_{i=1}^N \sum_{j=0}^{y_i-1} \ln(\theta + j\tau)$$

$$+ \sum_{i=1}^N \sum_{j=0}^{n-y_i-1} \ln(1 - \theta + j\tau)$$

$$- N \sum_{j=0}^{n-1} \ln(1 + j\tau).$$

The derivatives of the log likelihood function $l(\theta, \tau)$ with respect to θ and τ are given by

$$\frac{\partial}{\partial \theta} l(\theta, \tau) = \sum_{i=1}^N \sum_{j=0}^{y_i-1} \frac{1}{\theta + j\tau} - \sum_{i=1}^N \sum_{j=0}^{n-y_i-1} \frac{1}{1 - \theta + j\tau}$$

and

$$\frac{\partial}{\partial \tau} l(\theta, \tau) = \sum_{i=1}^N \sum_{j=0}^{y_i-1} \frac{j}{\theta + j\tau}$$

$$+ \sum_{i=1}^N \sum_{j=0}^{n-y_i-1} \frac{j}{1 - \theta + j\tau} - N \sum_{j=0}^{n-1} \frac{j}{1 + j\tau},$$

respectively. Maximum-likelihood estimates of θ and τ require numerical iteration and, thus, the Newton-Raphson method can be used (Griffiths, 1973; Smith, 1983). The SAS NLMIXED procedure minimizes the function $-l(\theta, \tau)$ over θ and τ numerically in order to estimate these parameters. Note that NLMIXED procedure

Table 1: Counts of points in IR subscale for female and male individuals.

Sex	Number of points												
	6	7	8	9	10	11	12	13	14	15	16	17	18
Women	3	2	3	1	3	2	5	12	21	24	55	31	40
Men	4	3	4	2	4	1	4	5	8	8	12	10	6
Total	7	5	7	3	7	3	9	17	29	32	67	41	46

is used to fit nonlinear models with random and fixed effects. However, due to its flexibility in accommodating different structures of the conditional distribution of the data and its versatility of numerical methods for estimating parameters in nonlinear models, the NLMIXED procedure can also be used to fit models without random effects when they are not specified by the programmer. The inverse Hessian matrix at the estimates is used to provide an approximate variance-covariance matrix for the estimation of the parameters. The Hessian matrix is the square matrix of second-order partial derivatives of the likelihood with respect to the parameters. Many different algorithms for optimizing general nonlinear functions are available in SAS NLMIXED procedure. In the present article, we used the dual quasi-Newton algorithm, which updates the Cholesky factor of an approximate Hessian (Littell et al., 2006).

As an alternative to the maximum likelihood method, moment estimators were introduced by Tamura and Young (1987) and Yamamoto and Yanagimoto (1992). However, a discussion of the moment-method estimation is outside the scope of this article and interested readers can refer to these authors.

A vector of covariates $\mathbf{X} = (X_1, \dots, X_k)'$ can be included in the model by replacing the parameter θ by a function $\theta(\mathbf{x})$ given by

$$\theta(\mathbf{x}) = \frac{\exp(\theta_0 + \theta_1 x_1 + \dots + \theta_k x_k)}{1 + \exp(\theta_0 + \theta_1 x_1 + \dots + \theta_k x_k)}, \quad (6)$$

where $\theta_0, \theta_1, \dots, \theta_k$ are unknown parameters (Forcina and Franconi, 1988). This link function is defined to ensure that the parameter θ remains bounded in the interval (0,1).

3 An example: the intrinsic religiosity index

To illustrate the use of the model, we consider a research conducted at the Faculty of Medicine of Ribeirão Preto (University of São Paulo) which used the Portuguese version of the Duke Religion Index (P-DUREL) in a sample of 202 female postgraduate students and 71 male students (Martinez et al., 2012). The P-DUREL is a five-item measure of religious involvement, that assesses the

three major dimensions of religiosity: organizational religious activity, non-organizational religious activity, and intrinsic religiosity (Koenig and Büssing, 2010). Intrinsic religiosity (IR) is measured by the God's presence experienced in the lives of people, the relation between religious beliefs and approach to life, and the effort to live the religion in all aspects of life (Koenig and Büssing, 2010). The greater the number of points in the IR subscale, the greater the quantification of the intrinsic religiosity of the individual. The maximum number of points in IR subscale is $n = 18$. Table 1 shows the counts of points in IR subscale for female and male individuals. Table 2 shows descriptive statistics for the intrinsic religiosity (IR) measures.

Table 2: Descriptive statistics.

Sex	Total	Mean	SD ¹	Variance
Women	202	15.42	2.523	6.364
Men	71	13.54	3.621	13.109
Both sexes	273	14.93	2.960	8.759

¹SD: Standard deviation.

The method was implemented in the SAS NLMIXED procedure and the likelihood function was maximized using a dual quasi-Newton algorithm. Akaike's information criterion (AIC) was considered for comparison between models (Burnham and Anderson, 2003). Generally, the lower the AIC value the better is the model fit. The SAS syntax is given in Appendix.

4 Results

Considering the data in Table 1, we have $n = 18$. Firstly, we consider a model that assumes that the number of points obtained by each subject in the IR dimension follows a beta-binomial distribution with parameters a and b . After this, we consider a model that assumes the parametrization $a = \theta\tau^{-1}$ and $b = (1 - \theta)\tau^{-1}$.

4.1 Model not considering the proposed parametrization

Initially, we fitted two models to the data considering the beta-binomial distribution, one which considers the number of points obtained in the IR scale considering both sexes (labeled as Model 1, without the inclusion of covariates) and another which considers a regression in the parameter a (labeled as Model 2), such that

$$a = \begin{cases} \exp(a_0), & \text{if men} \\ \exp(a_0 + a_1), & \text{if women} \end{cases} \quad (7)$$

Table 3 shows the results obtained from these models and their respective AIC values. Considering the results from Model 2, we note that the 95% confidence interval (95% CI) for a_1 does not contain the value 0, containing only positive values, which evidences a difference of IR mean scores between the sexes (we can conclude that there is evidence that women have higher IR scores). The mean and variance of the IR subscale were obtained from equations (1) and (2), respectively. The SAS NLMIXED procedure considers the mean and variance as additional parameters and their standard errors are approximated using the delta method.

Figure 2 compares the estimated frequencies obtained from Model 1 with the observed frequencies obtained from Table 1. We observe that the estimated frequencies are in good agreement with those obtained directly from the sample (Table 1).

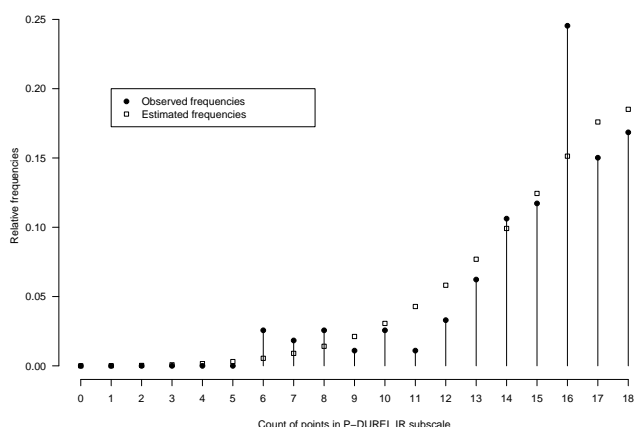


Figure 2: Observed frequencies of the IR scores compared to the predicted frequencies from the Model 1.

4.2 Model considering the proposed parametrization

Table 4 shows the maximum-likelihood results obtained from the model considering the parametrization $a =$

$\theta\tau^{-1}$ and $b = (1 - \theta)\tau^{-1}$. A third model do not considers the inclusion of covariates (labeled as Model 3) and a fourth model (Model 4) considers the variable sex, so that the link function $\theta(\mathbf{x})$ is given by

$$\theta(\mathbf{x}) = \begin{cases} \frac{\exp(\theta_0)}{1+\exp(\theta_0)}, & \text{if men} \\ \frac{\exp(\theta_0+\theta_1)}{1+\exp(\theta_0+\theta_1)}, & \text{if women} \end{cases} \quad (8)$$

We note in Table 4 that the AIC values for the Models 3 and 4 are quite similar to those from Models 1 and 2, respectively. Thus, the advantage of the parametrization used in these models is not necessarily a better fit to the data, but it is that in this case the mean of the variable of interest is a function of an unique parameter (see expression (3)), which makes the model more suitable for introducing a vector of covariates. This can be done by using the link function (6). On the other hand, when the parametrization is not considered, a model which considers a regression only in one of the parameters (a or b) can be arbitrary.

4.3 A comparison between models based on other distributions

It was fitted a model for the data of Table 1, assuming that the number of points obtained in IR subscale follows a binomial distribution with parameters $n = 18$ and p unknown. The mean obtained by this model was 14.93 points, exactly the value observed directly from the data (Table 2). However, the variance was estimated by 2.549, a value smaller than those described in Table 1. In this way we understand that the beta-binomial model is more appropriated to the data presented, as it is able to describe a dispersion higher than that of the binomial model. On the other hand, when it was fitted a model based on the negative binomial distribution, the mean was estimated by 14.93 and variance equals to 27.30. That is, the variance obtained by this model is much higher than that obtained from the beta-binomial model. In addition, the obtained AIC values from the beta-binomial model, binomial and negative binomial are given, respectively, 1231.1, 1535.9 and 1508.2, indicating a better fit of the data by a beta-binomial model. Observe that smaller values of AIC indicates better models.

5 Conclusion

Although there is no explicit solution to score functions obtained from the application of the maximum likelihood method, this study suggests that SAS program is efficient for the estimation of the model parameters, given that we have not problems of convergence for the

Table 3: Maximum likelihood estimates for the beta-binomial model.

Parameter	Estimate	Standard error	95%CI ¹	AIC
Model 1				1231.1
a	5.7759	0.1470	(5.487 , 6.064)	
b	1.2029	0.0651	(1.074 , 1.331)	
Mean	14.898	0.1723	(14.558 , 15.236)	
Variance	8.039	0.3886	(7.275 , 8.802)	
Model 2				1211.0
a_0	1.4158	0.0731	(1.2722 , 1.5594)	
a_1	0.6714	0.1137	(0.4480 , 0.8949)	
b	1.3636	0.0617	(1.2423 , 1.4848)	
Mean, women	15.396	0.1375	(15.126 , 15.666)	
Variance, women	5.859	0.3884	(5.095 , 6.622)	
Mean, men	13.524	0.3272	(12.881 , 14.167)	
Variance, men	12.181	0.9360	(10.342 , 14.020)	

¹95%CI: 95% confidence interval.

Table 4: Maximum likelihood estimates for the beta-binomial model, considering the proposed parametrization.

Parameter	Estimate	Standard error	95%CI ¹	AIC
Model 3				1231.1
τ	0.1433	0.0098	(0.1241 , 0.1625)	
θ	0.8276	0.0091	(0.8098 , 0.8455)	
Mean	14.897	0.1639	(14.575 , 15.220)	
Variance	8.039	0.5027	(7.050 , 9.026)	
Model 4				1215.1
θ_0	1.1503	0.1273	(0.9001 , 1.4004)	
θ_1	0.5942	0.1272	(0.3443 , 0.8442)	
τ	0.1283	0.0116	(0.1054 , 0.1512)	
Mean, women	15.323	0.2128	(14.904 , 15.741)	
Variance, women	6.685	0.7036	(5.302 , 8.068)	
Mean, men	13.672	0.4186	(12.849 , 14.494)	
Variance, men	9.642	0.9478	(7.780 , 11.504)	

¹95%CI: 95% confidence interval.

numerical algorithm (dual quasi-Newton) chosen to obtain the estimates and their standard errors. When it was added a covariate in the model, we observed that the same allowed us to estimate the mean and variance (see Table 3) close to those obtained directly from the data (Table 1). In addition, we note that an advantage of using the beta-binomial distribution for the data of Table 1 is in its adjustment for the dispersion of data (overdispersion), since the variance obtained from the model is closer to that obtained directly from the data than the the variances obtained from usual models based on the binomial and the negative binomial distributions. Thus, we conclude that the model based on the beta-binomial distribution can be used in the analysis of real data, whose dispersion is not satisfactorily estimated by binomial and negative binomial models, the usual

models considered in the health area.

Acknowledgments

The first and second authors are supported by a grant of the Brazilian Government (CNPq). We are also grateful to the anonymous reviewers for excellent suggestions for improving this paper.

References

Aeschbacher, H. U., Vuataz, L., Sotek, J., Stalder, R. (1977). Use of the beta-binomial distribution in dominant-lethal testing for "weak mutagenic activity"

- Part 1. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, 44(3), 369–390.
- Burnham, K. P., Anderson, D. R. (2003). *Model selection and inference: a practical information-theoretic approach*. Springer-Verlag, New York.
- Chatfield, C., Goodhardt, G. J. (1976). The beta-binomial model for consumer purchasing behaviour. In: *Mathematical Models in Marketing*, pp. 53–57. Springer Berlin Heidelberg.
- Crowder, M. J. (1978). Beta-binomial ANOVA for proportions. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 27(1), 34–37.
- Forcina, A., Franconi, L. (1988). Regression analysis with the Beta-Binomial distribution. *Rivista de Statistica Applicata*, 21(1), 7–12.
- Gange, S. J., Munoz, A., Saez, M., Alonso, J. (1996). Use of the beta-binomial distribution to model the effect of policy changes on appropriateness of hospital stays. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 45(3), 371–382.
- Griffiths, D. A. (1973). Maximum likelihood estimation for the beta-binomial distribution and an application to the house-hold distribution of the total number of cases of a disease. *Biometrics*, 29(4), 637–648.
- Haseman, J. K., Kupper, L. L. (1979). Analysis of dichotomous response data from certain toxicological experiments, *Biometrics*, 35(1), 281–293.
- Kleinman, J. C. (1978). Proportions with extraneous variance: single and independent samples. *Journal of the American Statistical Association*, 68(341), 46–53.
- Koenig, H. G., Büssing, A. (2010). The Duke University Religion Index (DUREL): a five-item measure for use in epidemiological studies. *Religions*, 1(1), 78–85.
- Littell, R. C., Milliken, G. A., Stroup, W. W., Wolfinger, R. D., Schabenberger, O. (2006). *SAS for Mixed Models*. Second Edition. Cary: SAS Institute.
- Martinez, E. Z., Santos-Almeida, R. G., Carvalho, A. C. D. (2012). Propriedades da Escala de Religiosidade de Duke em uma amostra de pós-graduandos. *Revista de Psiquiatria Clínica*, 39:(5), 180.
- Morgan, B. J. T. (1992). *Analysis of quantal response data*. London: Chapman and Hall.
- Pearson, E. S. (1925). Bayes' theorem, examined in the light of experimental sampling. *Biometrika*, 17(3/4), 388–442.
- Skellam, J. G. (1948). A probability distribution derived from the binomial distribution by regarding the probability of success as variable between the sets of trials. *Journal of the Royal Statistical Society. Series B*, 10(2), 257–261.
- Smith, D. M. (1983). Maximum likelihood estimation of the parameters of the beta binomial distribution. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 32(2), 196–204.
- Tamura, R. N., Young, S. S. (1987). A stabilized moment estimator for the beta-binomial distribution. *Biometrics*, 43, 813–824.
- Yamamoto, E., Yanagimoto, T. (1992). Moment estimators for the beta-binomial distribution. *Journal of Applied Statistics*, 19(2), 273–283.
- Williams, D. A. (1975). The analysis of binary responses from toxicological experiments involving reproduction and teratogenicity. *Biometrics*, 31(4), 949–952.

Appendix

The SAS procedure NLMIXED was designed for fitting nonlinear and generalized linear models with random effects (Littell et al., 2006). If the random effects are not reported, they are not included in the model. PROC NLMIXED specifies a conditional distribution for the dependent variable having a usual form (as normal, binomial or Poisson distributions) or specifies a general log likelihood function using SAS programming statements.

The SAS code for Model 1 is the following.

```
proc nlmixed data=example df=500;
y=IR;
parms a=6 b=1;
bounds a>0, b>0;
n = 18;
L = fact(n)/(fact(y)*fact(n-y))
*beta(y+a,n-y+b)/beta(a,b);
logL = log(L);
m=n*a/(a+b);
v=n*a*b*(a+b+n)/((a+b)*(a+b)*(a+b+1));
model y ~ general(logL);
estimate "Mean" m;
estimate "Variance" v;
run; .
```

In this SAS code, L is the likelihood function and logL is its logarithm. The option df in the first line specifies the degrees of freedom to be used in computing

asymptotic confidence intervals based on the Student t distribution. We set $df=500$ to consider confidence intervals based on the normal distribution. The PARMs statement describes the names of parameters and specifies initial values. In order to avoid problems of convergence or a Hessian matrix with negative eigenvalues, reasonable initial values should be specified for each parameter. This can be a trial-and-error process; if the initial values are far from the solution, this method can suffer from convergence problems. The ESTIMATE statement enables additional estimates that is a function of the parameter values, and their approximate standard errors are obtained from the delta method. Note that in the SAS code above, we used the ESTIMATE statement to obtain inferences about the mean and the variance of the beta-binomial distribution.

```
m=n*theta;
v=n*theta*(1-theta)*(1+(n-1)*tau/(1+tau));
model y ~ general(logL);
estimate "Mean" m;
estimate "Variance" v;
run; .
```

The SAS code for Model 2 is

```
proc nlmixed data=example df=500;
y=IR;
parms a0=1.4 a1=0.5 b=1.3;
bounds b>0;
if sex="M" then a=exp(a0);
if sex="F" then a=exp(a0+a1);
n = 18;
L = fact(n)/(fact(y)*fact(n-y))
*beta(y+a,n-y+b)/beta(a,b);
logL = log(L);
m1=n*exp(a0)/(exp(a0)+b);
v1=n*exp(a0)*b*(exp(a0)+b+n)
/((exp(a0)+b)*(exp(a0)+b)*(exp(a0)+b+1));
m2=n*exp(a0+a1)/(exp(a0+a1)+b);
v2=n*exp(a0+a1)*b*(exp(a0+a1)+b+n)
/((exp(a0+a1)+b)*(exp(a0+a1)+b)
*(exp(a0+a1)+b+1));
model y ~ general(logL);
estimate "Mean M" m1;
estimate "Variance M" v1;
estimate "Mean F" m2;
estimate "Variance F" v2;
run; .
```

The SAS code for Model 3 is

```
proc nlmixed data=example df=500;
y=IR;
parms theta=0.5 tau=0.1;
bounds theta>0, tau>0;
n = 18;
a = theta/tau;
b = (1-theta)/tau;
L = fact(n)/(fact(y)*fact(n-y))*
beta(y+a,n-y+b)/beta(a,b);
logL = log(L);
```