A Course on Bayesian Modeling Using WinBUGS

Lecture 4 (Part A): Generalized Linear Models in WinBUGS

Ioannis Ntzoufras

Associate Professor,

Department of Statistics,

Athens University of Economics and Business,

Athens, Greece;

e-mail: `ntzoufras@aueb.gr`.

Web-page: `http://www.stat-athens.aueb.gr/~jbn`

This presentation is based on Chapter 7 of Ntzoufras (2009): *Bayesian Modeling Using WinBUGS*, Wiley.

Synopsis

1. Introduction: The exponential family, Link functions, Common GLMs.
2. Prior distributions
3. Posterior inference and GLM specification in WinBUGS
4. Poisson regression models
5. Binomial response models

4.1 Introduction

- Generalized linear models (GLMs) constitute a wide class of models encompassing stochastic representations used for the analysis of both quantitative (continuous or discrete) and qualitative response variables.
- Natural extension of normal linear regression models
- They are based on the exponential family of distributions, which includes the most common distributions such as the normal, binomial and, Poisson.
- Generalized linear models have become very popular because of their generality and wide range of application.
- They can be considered as one of the most prominent and important components of modern statistical theory.
- They have provided not only a family of models that are widely used in practice but also a unified, general way of thinking concerning the formulation of statistical models.

The Components of a GLM

As we have already mentioned, three are the components of a GLM:

1. **The random/stochastic component**:

$$Y_i \sim \mathcal{D}(\boldsymbol{\theta}) \in \text{Exponential family of distributions} \; .$$

2. **The systematic component (or linear predictor)**:

   Linear function of the explanatory variables (or covariates) similarly as in normal regression models called *linear predictor*.

3. **The link function**:

   Function $g(\boldsymbol{\theta})$ which connects the parameters of the response Y with the linear predictor and the covariates. In GLM a location parameter (e.g., the mean) is usually linked with the linear predictor.

Model Specification

$$
\begin{array}{lcll}
Y_i & \sim & \mathrm{expf}\Big(\vartheta_i, \phi, a(), b(), c()\Big) & \text{(stochastic component)} \\
\eta_i & = & \boldsymbol{X}_{(i)}\boldsymbol{\beta} = \beta_0 + \sum_{j=1}^{p} x_{ij}\beta_j & \text{(systematic component)} \\
\vartheta_i & = & R(\theta_i) & \text{(canonical — distribution parameter function)} \\
g(\theta_i) & = & g(R^{-1}(\vartheta_i)) = g_\vartheta(\vartheta_i) = \eta_i & \text{(link function)} \\
\boldsymbol{\theta}_m & = & (\boldsymbol{\beta}^T, \phi)^T & \text{(model parameters)}
\end{array}
\qquad (1)
$$

4.1.1 The exponential family

$\text{expf}\Big(\vartheta, \phi, a(), b(), c()\Big)$ denotes the exponential family with density or probability function

$$f(y|\vartheta, \phi) = \exp\left(\frac{y\vartheta - b(\vartheta)}{a(\phi)} + c(y, \phi)\right). \tag{2}$$

- ϕ dispersion parameter of the exponential family
- ϑ is the canonical location parameter of the exponential family
- θ is the location parameter of the corresponding distribution
- $R(\theta)$ the function that connects the two parameters.
- $a()$, $b()$, $c()$ are functions which specify the density or probability function
- $g(\theta)$ and $g_\vartheta(\vartheta)$ are the link functions that associate the location parameter θ and the canonical parameter ϑ, respectively, with the linear predictor η $\Rightarrow g_\vartheta(\vartheta) = g(R^{-1}(\vartheta))$.

The mean and the variance of Y with distribution in the exponential family with parameters ϑ and ϕ are equal to

$$E(Y) = \frac{db(\vartheta)}{d\vartheta} = b'(\vartheta) \;\text{ and }\; V(Y) = \frac{d^2 b(\vartheta)}{d\vartheta^2} a(\phi) = b''(\vartheta) a(\phi) \;. \tag{3}$$

Indicative Bibliography

Additional details concerning generalized linear models can be found in a variety of well-written books related to the topic such as

- McCullagh and Nelder (1989),
- Lindsey (1997), and
- Fahrmeir and Tutz (2001).

A detailed illustration of Bayesian inference and analysis focusing on GLMs can be found in

- Dey et al. (2000)
- Chapter 7 of Ntzoufras (2009)

4.1.2 Common distributions as members of the exponential family

Members of the exponential family are popular distributions such as

- the normal
- the binomial
- the Poisson
- the gamma
- the inverse Gaussian

We may further include distributions that are

- Special cases of these distributions (e.g., exponential, Pareto)
- Distributions that result as transformations of the above mentioned random variables (e.g. log-normal, inverse gamma)

Details concerning the most popular distributions of the exponential family are provided in Section 7.1.2 of Ntzoufras (2009); see Table 4.1 for a tabulated summary.

Table 4.1: Details of most common members of exponential dispersion family

Distribution	Notation	Values of Y	Mean	Variance	ϑ	$b(\vartheta)$	$a(\phi)^a$	$\mu(\vartheta)$	$c(y,\phi)$
Normal	$N(\mu,\sigma^2)$	$\mathbb{R}$	μ	σ^2	μ	$\vartheta^2/2$	σ^2	ϑ	$-\frac{1}{2}\log(2\pi\sigma^2)$ $-\frac{1}{2}y^2/\sigma^2$
Binomial	binomial(π,N)	$\{0,1,\ldots,N\}$	$N\pi$	$N\pi(1-\pi)$	$\log[\pi/(1-\pi)]$	$N\log(1+e^{\vartheta})$	1	$N/(1+e^{-\vartheta})$	$\log(N!/y!)$ $-\log(N-y)!$
Negative Binomial	NB(π,k)	$\mathbf{N}=\{0,1,2,\ldots\}$	$k(1-p)/p$	$k(1-p)/p^2$	$\log(1-\pi)$	$-k\log(1-e^{\vartheta})$	1	$ke^{\vartheta}/(1-e^{\vartheta})$	$\log(y+k-1)!$ $-\log[y!(k-1)!]$
Poisson	Poisson(λ)	$\mathbf{N}=\{0,1,2,\ldots\}$	λ	λ	$\log\lambda$	e^{ϑ}	1	$\log(y!)$	e^{ϑ}
Gamma	gamma$(a,b)^b$	$(0,\infty)$	$\mu=a/b$	$\mu^2/a=a/b^2$	$-\mu^{-1}=-b/a$	$-\log(-\vartheta)$	a^{-1}	$-\vartheta^{-1}$	$\phi^{-1}\log(y\phi^{-1})$ $-\log\left[y\Gamma(1/\phi)\right]$
Inverse Gaussian	IGaussian(μ,λ)	$(0,\infty)$	μ	μ^3/λ	$-\mu^{-2}$	$-(-2\vartheta)^{1/2}$	λ^{-1}	$(-2\vartheta)^{-1/2}$	$-\frac{1}{2}(\phi y)^{-1}$ $-\frac{1}{2}\log(2\pi\phi y^3)$

[a] $a(\phi)=\phi$.

[b] $b=a/\mu$.

4.1.3 Link functions

4.1.3.1 Common link functions.

- The link function must be a monotonic and differentiable function.
- It is used to match the parameters of the response variable with the systematic component (i.e. the linear predictor) and the associated covariates.
- We focus on the mean of the distribution because the measures of central location are usually of main interest.
- GLM-based extensions: dispersion or shape parameters are linked with covariates [e.g., see Rigby and Stasinopoulos (2005)].

- **Desirable property**: the link function should map the range of values in which the parameter of interest lies with the set of real numbers $\mathbb{R}$ in which the linear predictor takes values.

  For example, in the binomial case we wish to identify link functions that map the success probability π from $[0, 1]$ to $\mathbb{R}$.

- The simplest link function : sets the linear predictor equal to the mean μ.
  - It is used in the normal models.
  - Not appropriate for other distributions such as the Poisson distribution since their mean is positive while $\eta \in \mathbb{R}$.

- **Default choice of link function**: is the *canonical link* = the canonical parameter is set equal to the linear predictor.

- The canonical link function for common distributions are summarized in Table 4.2.

Table 4.2: Canonical link functions of most common members of exponential dispersion family

	Link	Link function	
Distribution	name	$g(\mu)$	$g(\boldsymbol{\theta})$
Normal	Identity	μ	
Binomial	Logit	$\log\left[(\mu/N)/(1-\mu/N)\right]$	$\log\left[\pi/(1-\pi)\right]$
Negative binomial	Complementary log	$\log\left[\mu/(k+\mu)\right]$	$\log(1-\pi)$
Poisson	Logarithmic	$\log\lambda$	
Gamma	Reciprocal	$1/\mu$	
Inverse Gaussian	Squared reciprocal	$1/\mu^2$	

Link Functions for Binomial Models

1. The canonical link is the so-called logit link defined as $g(\pi) = \log\left(\frac{\pi}{1-\pi}\right)$.
   The corresponding model is the well known logistic regression models.
2. The probit link (frequently used in econometrics): $g(\pi) = \Phi^{-1}(\pi)$;
   where $\Phi^{-1}(\pi)$ is the inverse function of the cdf of $N(0,1)$.
3. The complementary log–log link function : $g(\pi) = \log\left\{ -\log(1-\pi)\right\}$.
4. General links: given by the inverse cdf $g(\pi) = F^{-1}(\pi; \theta)$ of a random variable $Z \sim D(\theta)$.
   - Logit cdf of logistic distribution
   - Complementary log–log cdf of extreme value distribution.

4.1.3.2 More complicated link functions for binomial data.

- A wide variety of link distributions have been proposed for binomial models.
- Details can be found in Section 7.1.3.2 of Ntzoufras (2009).

4.1.4 Common generalized linear models

- Different models of the exponential family are appropriate for different types of response variables.
- In this section, we summarize which are the most common models for each type of response variable.

Response variables defined in $\mathbb{R}$

- The normal regression model is the most popular choice.
- When the normality of error assumption is not appropriate, the normal model can be extended using errors that follow the Student's t distribution.
- Although this model cannot be considered as a member of the exponential family, it can be easily fitted using WinBUGS.

Positive continuous response variables

Initial approach

Transform the response variable using the Box–Cox transformation or the logarithm and use a common (normal) regression model for the transformed response variable.

- Can be treated as usual regression models — parameter interpretation might be difficult.
- First consider the logarithm of the original response since it usually eliminates problems related to the assumptions of the model (normality or errors, linearity of the mean, or homoscedasticity).
- Using the logarithm in a normal model is equivalent to assuming the log-normal distribution for the original response variable.

Common distributional choices

- Gamma
- Exponential
- Inverse Gaussian
- Weibull

Survival Analysis Models

A positive response variable is the survival time (or more generally the time until an event of interest occurs).

- It is of central interest in medical studies, especially in clinical trials.
- Censoring is an additional characteristic that must be considered in the model.
- A survival time is considered as censored when part of its information is not available (e.g. we may know that a patient was alive for the 50 first days of the study, but we might be ignorant of the exact time of failure).
- The Weibull distribution is commonly used for modeling such response data.

Binary (success/failure) responses

- Binary (zero–one) data (i.e., $y \in \{0, 1\}$) $\Rightarrow$ Bernoulli distribution.
- Y = number of successes after the repetition of N Bernoulli experiments $\Rightarrow$ binomial distribution with success probability π and N replications.
- $y \in \{0, 1, \ldots, n\}$.
- Bernoulli is a special case of binomial distribution with $N = 1$.
- The canonical link is the logit function $\log\big(\pi/(1-\pi)\big)$, which models the log-odds of success as linear combination of the covariates (Berkson, 1944, 1951).
- Logit models are the most popular stochastic formulations for such data and are cited as *logistic regression models.*
- We may also use Probit model (similar to logit) or model with complementary log–log link or other link function.

Counts and responses defined in $\mathbb{N}$

- Response variables defined in $\mathbf{N} = \{0, 1, 2, \ldots, \}$ frequently represent **number of events occurred within a prespecified time interval** (i.e. counts or frequencies).
- The Poisson distribution is naturally adopted $\Rightarrow$ Poisson regression models
- Also called *Poisson log-linear* or simply *log-linear*) *models* due to the canonical log-link.
- Poisson log-linear models are used for the analysis of high-dimensional contingency data (cross-classification tables of categorical variables).
- Restrictive assumption of Poisson: mean equal to variance.
- Alternative models that allow for overdispersion (larger variance than mean) or underdispersion
- An overdispersed popular distribution is the negative binomial.

Continuous Responses with a Specific Range

For variables that are defined in a range $y \in (a, b)$

- we may rescale them in the zero–one interval by setting $y^* = (y - a)/(b - a)$ and use the beta distribution for the stochastic component.
- Use a logit-like transformation by setting $y^{**} = \log\left\{(y - a)/(b - y)\right\}$ and use normal regression models.

Integer Valued Responses

For response variables $y \in \mathbf{Z} = \{\ldots, -3, -2, -1, 0, 1, 2, 3, \ldots\}$:

- Poisson difference model based on the Skellam distribution (Karlis and Ntzoufras, 2006, 2008).
- The model is based on the differences of Poisson latent variables.
- The Skellam distribution cannot be considered as a member of the exponential family but the conditional likelihood (given the latent Poisson variables) is a simple Poisson likelihood.

Categorical Responses with $k > 2$ Levels

- The multinomial distribution may be used as a natural extension of the binomial models.
- The same distribution can be used for grouped categorical variables where the frequencies of k different outcomes will be recorded as responses.
- Finally, such responses can be modeled indirectly using the Poisson log-linear models for contingency tables when all covariates are categorical [see, e.g., Fienberg (1981, chaps. 6, 7)].

Ordinal Responses

- Can be modeled using a variety of alternative approaches that have been introduced in the literature.
- A natural extension of the usual log-linear model used for contingency models can be adopted by Goodman's (1979) association models, which were originally used for two-way contingency models. Such models cannot be considered as GLMs because of the multiplicative expression between the model parameters and Poisson's expected values.
- Also a variety of logistic regression models; details can be found in Agresti (2010).

4.1.5 Interpretation of GLM coefficients

4.2 Prior distributions

Independent Priors

- Usually, a normal prior distributions is used for $\boldsymbol{\beta}$: $\boxed{\beta_j|\phi \sim N(\mu_{\beta_j}, \sigma^2_{\beta_j}\phi)}$.
- The variance of $\boldsymbol{\beta}$ depends on the dispersion parameter ϕ in order to achieve an appropriate scaling of the prior distribution.
- In the normal model: $\phi = \sigma^2 \sim \text{IG}(a, b) \Rightarrow$ conjugate prior distribution.
- When no prior information is available

  $\Rightarrow$ Prior mean = zero

  $\Rightarrow$ variance = large (to express prior ignorance).

  – A prior independent to the dispersion parameter can be also considered.

Independent Priors - Some Comments

- A priori independence between all model parameters is plausible when the design or data matrix is orthogonal:

  $\Rightarrow$ Model parameters have similar interpretation over all models.

  – We can easily incorporate such priors in ANOVA-type models with sum-to-zero constraints.

  – When we are interested in prediction we may orthogonalize the design matrix and proceed with model selection in the new orthogonal model space (Clyde et al., 1996).

  – Independent priors $\Rightarrow$ prior of Knuiman and Speed (1988) for Poisson log-linear models in contingency tables for STZ parametrization.

- In nonorthogonal cases: an independent prior $\Rightarrow$ undesirable influence on the posterior distribution and hence must be avoided.

Multivariate normal prior

$$\boldsymbol{\beta}|\phi \sim N(\boldsymbol{\mu}_\beta, \boldsymbol{\Sigma}_\beta) \ .$$

Extension of the Zellner's g-prior

$$\boldsymbol{\Sigma}_\beta = c^2 \Big(- H(\widehat{\boldsymbol{\beta}}) \Big)^{-1}, \tag{4}$$

- $\widehat{\boldsymbol{\beta}}$: maximum-likelihood estimate of $\boldsymbol{\beta}$
- $H(\boldsymbol{\beta})$ is the second derivative matrix of $\log f(\boldsymbol{y}|\boldsymbol{\beta}, \phi)$, given by

  $$-H(\boldsymbol{\beta}) = \boldsymbol{X}^T \boldsymbol{H} \boldsymbol{X}$$

  where $\boldsymbol{H}$ is a $n \times n$ diagonal matrix with elements

$$h_i = \left(\frac{\partial \mu_i}{\partial \eta_i} \right)^2 \frac{1}{a_i(\phi) b''(\vartheta)} \ . \tag{5}$$

Details concerning h_i for some popular distributions are provided in Table 4.3.

Table 4.3: Generalized linear model weights h_i

Model	Link	GLM weights h_i
Normal	Identity	σ^{-2}
Poisson	Log	λ_i
Binomial	Logit	$N_i \pi_i (1 - \pi_i)$
	Probit[a]	$N_i \left[\pi_i (1 - \pi_i) \left\{ \varphi(\pi_i) \right\}^2 \right]^{-1}$
	clog–log	$-N_i (1 - \pi_i) \left\{ \log(1 - \pi_i) \right\}^2 \pi_i^{-1}$

[a] $\varphi(z)$ is the density function of standardized normal distribution evaluated at z.

- *Unit information prior* for $c^2 = n$.
- This prior has precision ≈ the precision provided by one data point.
- More detailed discussion of this prior can be found in Spiegelhalter and Smith (1988) and Kass and Wasserman (1995).

For the normal model:

– $h_i = \sigma^{-2} \Rightarrow$ Zellner's g-prior.

For the remaining GLMS:

– h_i depends on estimated parameter values for each case; e.g.
  * In Poisson model: $h_i = \widehat{\lambda}_i = \exp\left(\boldsymbol{X}_{(i)}\widehat{\boldsymbol{\beta}}\right)$
  * In the binomial model: $h_i = N_i \widehat{\pi}_i(1-\widehat{\pi}_i)$ with $\widehat{\pi}_i = [1+\exp(-\boldsymbol{X}_{(i)}\widehat{\boldsymbol{\beta}})]^{-1}$

– Results in a data-dependent prior.

– When c^2 is large, the effect of this data dependence will be minimal since the prior will be essentially noninformative.

– To avoid this data dependence, Ntzoufras et al. (2003) proposed using the prior mean to obtain rough prior estimates of h_i.

  In binomial logistic regression models:

$\Rightarrow h_i = N_i \exp\left(\boldsymbol{X}_{(i)}\boldsymbol{\mu}_\beta\right)\left[1+\exp\left(\boldsymbol{X}_{(i)}\boldsymbol{\mu}_\beta\right)\right]^{-2}$

$\Rightarrow h_i = N_i/4$ if the prior means are zero

$\Rightarrow \boldsymbol{\Sigma}_\beta = 4N^{-1}c^2\left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}$ if $N_i = N$ for all $i = 1, 2, \ldots, n$

4.3 Posterior inference

4.3.1 The posterior distribution of a generalized linear model

Using the multivariate normal prior described in Section 4.2, we end up with the posterior

$$\begin{aligned} f(\boldsymbol{\beta}, \phi | \boldsymbol{y}) \quad &\propto \quad \exp\Bigg(\sum_{i=1}^{n} \frac{y_i g_{\vartheta}^{-1}(\boldsymbol{X}_{(i)}\boldsymbol{\beta}) - b\Big(g_{\vartheta}^{-1}(\boldsymbol{X}_{(i)}\boldsymbol{\beta})\Big)}{a(\phi)} + \sum_{i=1}^{n} c(y_i, \phi) \\ &\qquad -\frac{1}{2}\log|\boldsymbol{\Sigma}_\beta| - \frac{1}{2}(\boldsymbol{\beta} - \boldsymbol{\mu}_\beta)^T \boldsymbol{\Sigma}_\beta^{-1}(\boldsymbol{\beta} - \boldsymbol{\mu}_\beta)\Bigg) f(\phi), \end{aligned}$$

where $f(\phi)$ in the full posterior is the prior of the dispersion parameter ϕ.

- This posterior cannot be evaluated analytically except for the normal model when using the conjugate prior.
- MCMC methods are now available and widely used for the computation of the posterior distribution.
- The Gibbs sampler can be easily applied because of the result obtained by Dellaportas and Smith (1993), which allowed for implementation of the adaptive rejection method of Gilks and Wild (1992) since the posterior distributions of the parameters in specific GLMs is log-concave.
- Alternatively, Metropolis–Hastings algorithms or the slice sampler can be used.
- This method is also used in WinBUGS for the generation of random values from the posterior distribution of GLMs.

4.3.2 GLM specification in WinBUGS

- WinBUGS code for GLM: Similar to the corresponding code for normal regression models.
- Change the stochastic component (distribution of Y) and the link function.
- Details concerning the distributions of the most popular GLMs are summarized in Table 4.4.
- The inverse Gaussian distribution is not included in the standard distributions of WinBUGS; however, it can be modeled using an alternative approach (we will discuss it later in this course).

Table 4.4: WinBUGS commands for distributions within the exponential family[a]

	Distribution name	WinBUGS syntax	Probability or density function $f(x)$	Mean	Variance
1.	Normal	`y ~ dnorm(mu,tau)`	$\sqrt{\tau/(2\pi)}\exp\left[-\frac{1}{2}\tau(y-\mu)^2\right]$	μ	$1/\tau$
	(Log-normal)[b]	`y ~ dlnorm(mu,tau)`	$\sqrt{\tau/(2\pi)}y^{-1}\exp\left[-\frac{1}{2}\tau(\log y-\mu)^2\right]$	$e^{\mu+\frac{1}{2\tau}}$	$(e^{\frac{1}{\tau}}-1)e^{2\mu+\frac{1}{\tau}}$
2.	Binomial	`y ~ dbin(p,N)`	$N!p^y(1-p)^{N-y}/[y!(N-y)!]$	Np	$Np(1-p)$
	(Bernoulli)[c]	`y ~ dbern(p)`	$p^y(1-p)^{1-y}$	p	$p(1-p)$
3.	Neg. binomial	`y ~ dnegbin(p,r)`	$(y+r-1)!p^r(1-p)^y/[y!(r-1)!]$	$r\frac{1-p}{p}$	$r(1-p)p^{-2}$
4.	Poisson	`y ~ dpois(lambda)`	$e^{-\lambda}\lambda^y/y!$	λ	λ
5.	Gamma	`y ~ dgamma(a,b)`	$b^a y^{a-1}e^{-by}/\Gamma(a)$	a/b	a/b^2
	(Chi-squared)[d]	`y ~ dchisqr(k)`	see gamma$(k/2, \frac{1}{2})$	k	$2k$
	(Exponential)[e]	`y ~ dexp(lambda)`	$\lambda e^{-\lambda y}$	$1/\lambda$	$1/\lambda^2$

[a] Terms in parentheses can be considered as special cases of the distributions shown above.

[b] $\log(y)$ follows the normal distribution.

[c] Binomial with $N=1$.

[d] Gamma with $a=k/2$ and $b=\frac{1}{2}$.

[e] Gamma with $a=1$ and $b=\lambda$.

Link Functions in WinBUGS

- Four link functions are available in WinBUGS: `log`, `logit`, `probit`, and the `cloglog`.
- Can be used only in the **left part** of the definition of the linear predictor .
- The remaining link functions can be defined by setting the parameter of interest θ_i equal to $g^{-1}(\eta_i)$.

4.4 Poisson regression models

- Here we focus on Poisson regression models for response variables defined in $\mathbf{N}$.
- Such variables usually express the number of successes (visits, telephone calls, number of scored goals in football) within a fixed time interval.
- They are frequently called *Poisson log-linear models* because of the canonical log-link, which is widely used.
- The Poisson log-linear model is summarized by the following expression:

$$Y_i \sim \text{Poisson}(\lambda_i) \text{ with } \log \lambda_i = \beta_0 + \sum_{j=1}^{p} \beta_j x_{ij} = \boldsymbol{X}_{(i)} \boldsymbol{\beta} \ .$$

4.4.1 Interpretation of Poisson log-linear parameters

- $B_j = e^{\beta_j}$: is the proportional change when X_j increases by one unit.
- Details can be found in Section 7.4.1 of Ntzoufras (2009).

4.4.2 A simple Poisson regression example

Example 4.1. Aircraft damage dataset. *Here we consider the aircraft damage dataset of Montgomery et al. (2006). The dataset refers to the number of aircraft damages in 30 strike missions during the Vietnam war. Hence it consists of 30 observations and the following four variables:*

- `damage`*: the number of damaged locations of the aircraft*
- `type`*: binary variable which indicates the type of plane (0 for A4; 1 for A6)*
- `bombload`*: the aircraft bomb load in tons*
- `airexp`*: the total months of aircrew experience*

In this example we can use the Poisson distribution to monitor the number of damages after each mission.
Data of this example are available in the book's Website and are reproduced with permission of John Wiley and Sons, Inc.

4.4.2.1 Model specification in WinBUGS .

The initial model will have the following structure

$$\begin{aligned} \text{damage}_i &\sim \text{Poisson}(\lambda_i) \\ \log \lambda_i &= \beta_1 + \beta_2\,\text{type}_i + \beta_3\,\text{bombload}_i + \beta_4\,\text{airexp}_i \\ &\quad \text{for} \quad i = 1, 2, \ldots, 30\ . \end{aligned}$$

Here, the index of β_j takes values from 1 to 4 (instead from 0 to 3 as in the previous section) to be in concordance with the WinBUGS the code that follows.

We follow the same structure as in the linear regression model with the difference that the likelihood is now defined using the following syntax:

```
for (i in 1:30){
    damage[i] ~ dpois( lambda[i] )
    log(lambda[i]) <- beta[1] + beta[2] * type[i]
                    + beta[3] * bombload[i] + beta[4] * airexp[i]
}
```

Moreover, the exponentiated parameters B_j can be easily defined using the syntax

```
for (j in 1:4){ B[j] <- exp( beta[j] ) }
```

The usual independent normal prior with large variance ($\tau_{\beta_j} = \sigma_{\beta_j}^{-2} = 10^{-4}$) is considered as prior distribution for β_j .

The full code is available in this book's Webpage.

4.4.2.2 Results.

Posterior summaries of model parameters are given in Table 4.5, while 95% posterior intervals are depicted in Figure 4.1.

Table 4.5: Posterior summaries of Poisson model parameters for Example 4.1[a]

node	mean	sd	MC error	2.5%	median	97.5%	start	sample	harmonic
beta[1]	-0.766	1.089	0.1762	-3.168	-0.835	1.619	1001	1000	
beta[2]	0.580	0.466	0.0513	-0.302	0.584	1.537	1001	1000	
beta[3]	0.177	0.068	0.0099	0.040	0.177	0.308	1001	1000	
beta[4]	-0.011	0.010	0.0015	-0.033	-0.010	0.007	1001	1000	
B[1]	0.862	1.221	0.1829	0.042	0.434	5.050	1001	1000	0.465
B[2]	1.993	0.996	0.1050	0.739	1.793	4.652	1001	1000	1.786
B[3]	1.197	0.081	0.0118	1.041	1.193	1.360	1001	1000	1.194
B[4]	0.989	0.010	0.0015	0.968	0.990	1.007	1001	1000	0.989

[a] The harmonic means of B_j are calculated outside WinBUGS using the posterior means of β_j.

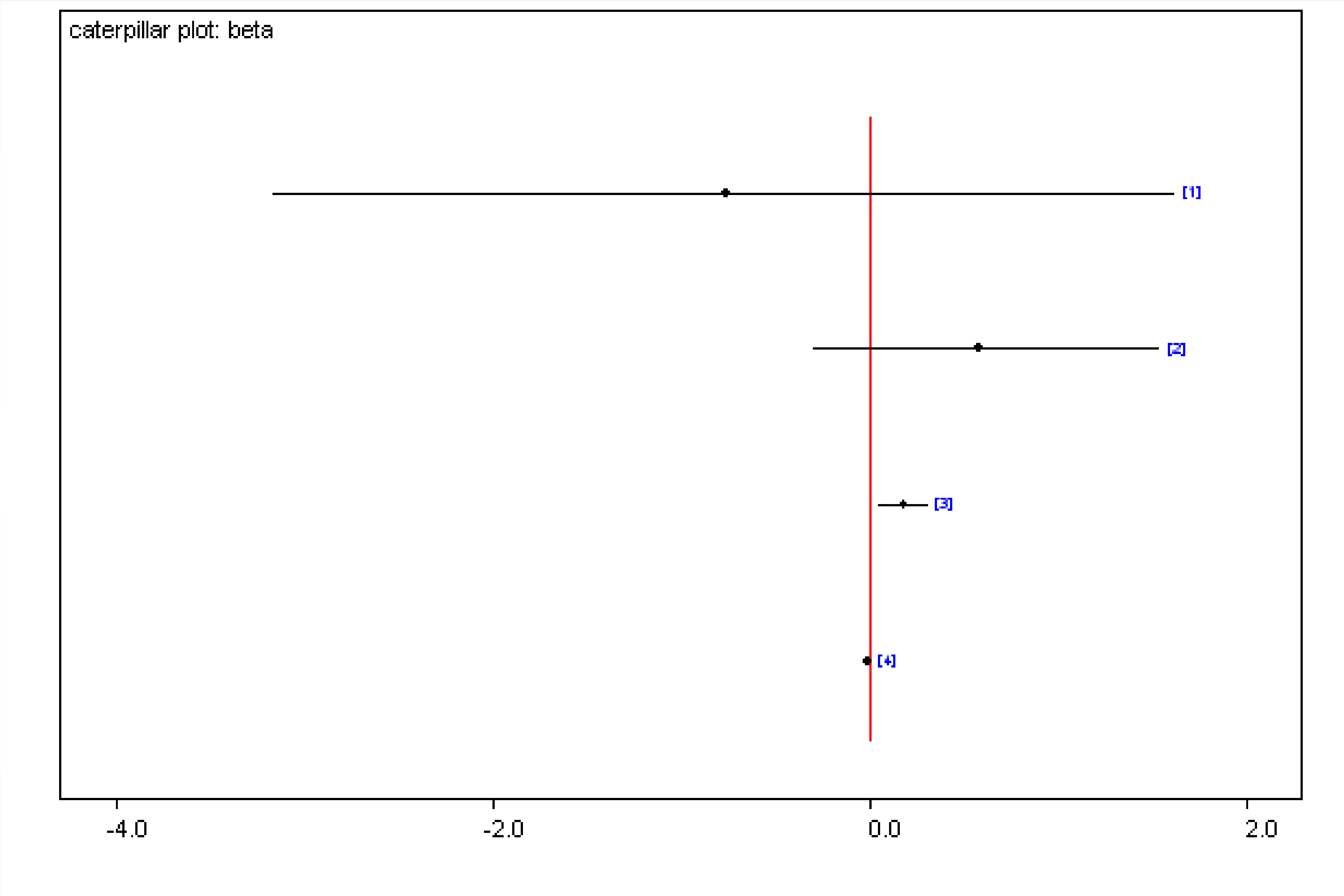


Figure 4.1: 95% posterior intervals of Poisson model parameters for Example 4.1.

A point estimate of the model can be based on the posterior means. Hence the a posteriori estimated model can be summarized by

$$\log \lambda_i = -0.77 + 0.58\,\text{type}_i + 0.18\,\text{bombload}_i - 0.011\,\text{airexp}_i \ .$$

From the 95% posterior intervals of β_j, we observe that only the posterior distribution of the bombload coefficient is away from zero, indicating a significant effect of this variable on the amount of aircraft damage.

4.4.2.3 Interpretation of the model parameters.

Interpretation can be directly based on B_j values (`B[]` in WinBUGS).

node	mean	sd	MC error	2.5%	median	97.5%	start	sample	harmonic
B[1]	0.862	1.221	0.1829	0.042	0.434	5.050	1001	1000	0.465
B[2]	1.993	0.996	0.1050	0.739	1.793	4.652	1001	1000	1.786
B[3]	1.197	0.081	0.0118	1.041	1.193	1.360	1001	1000	1.194
B[4]	0.989	0.010	0.0015	0.968	0.990	1.007	1001	1000	0.989

From Table 4.5, we may conclude the following

- The expected amount of damage for A6 (type=1) aircraft is twice as much as the corresponding damage for A4 (type=0) aircraft when the two aircraft return from missions with aircrew of the same experience and both carry the same bombload.
- Every tone of bombload increases the expected number of damaged aircraft locations by 20%.
- Every additional month of aircrew experience reduces the number of damaged aircraft locations by 1%.

4.4.2.4 Estimating specific profiles.

- The expected amount of damage for the two types of aircraft for the minimum, maximum, mean and median profiles have also been calculated.
- Minimum profile ⇒ maximum values of crew experience were considered, since these variables are negatively associated with the number of damaged locations [and maximum profile ⇒ minimum values]
- Calculation of the expected value for a profile can be easily accommodated in WinBUGS .

  e.g. a profile of an A6 aircraft is calculated by

```
a6.profile <- exp( beta[1] + beta[2] + beta[3] * bombload.profile
                                    + beta[4] * airexp.profile
```

  where `bombload.profile` and `airexp.profile` are the values of the two explanatory variables for the profile that we wish to consider.
- Substitution of these nodes by appropriate values provides the desired profiles; see Table 4.6 for the corresponding code.

- The profiles for A4 are obtained similarly by removing parameter β_2.
- Note that the minimum and maximum values of a vector $\boldsymbol{v}$ can be obtained using the commands `ranked(v[],1)` and `ranked(v[],n)`, respectively.
- Similarly, the median profile can be calculated using the command

```
ranked(v[],(n+1)/2)
```

if n is odd and by

```
0.5*(ranked(v[],n/2)+ranked(v[],n/2+1))
```

if n is even.

Table 4.6: WinBUGS syntax for calculation of expected number of damaged locations for each profile for Example 4.1

```
# profiles
# values for bombload
profiles[1,1]<- ranked( bombload[], 1 )  # minimum of bombload
profiles[2,1]<- mean(bombload[])         # mean of bombload
profiles[3,1]<- 0.5*(ranked( bombload[],15)+ranked(bombload[],16)) #median
profiles[4,1]<- ranked( bombload[], 30 ) # max
# values for airexp
profiles[1,2]<- ranked( airexp[], 1 )  # max experience
profiles[2,2]<- mean(airexp[])         # mean
profiles[3,2]<- 0.5*(ranked( airexp[], 15)+ranked(airexp[], 16))  # median
profiles[4,2]<- ranked( airexp[], 30 ) # min experience

for (k in 1:4){
a4.profile[k]<-exp(beta[1] + beta[3]*profiles[k,1] + beta[4]*profiles[k,2])
a6.profile[k]<-a4.profile[k]*exp( beta[2] )
}
```

Results

- Posterior means and the corresponding standard deviations of these profiles are provided in Table 4.7.
- For a typical mission with A4 aircraft we expect 0.8 damaged locations, while for A6 the corresponding number of damaged locations is about 1.3.
- Note that the worst-case scenario (maximum profile) where missions with 14 tons of bombload and crew with the minimum flying experience (50 months) corresponds to an expected number of 3.7 and 5.9 damaged locations for A4 and A6 aircrafts, respectively.

Table 4.7: Posterior means (standard deviations) of expected number of damaged locations for minimum, mean, median, and maximum profiles for Example 4.1

			Expected damage	
			A4	A6
Profile	Bombload	Experience	mean (SD)	mean (SD)
Minimum	4.0	120.00	0.27 (0.13)	0.50 (0.27)
Median	7.5	80.25	0.75 (0.24)	1.22 (0.94)
Mean	8.1	80.77	0.83 (0.27)	1.33 (0.40)
Maximum	14.0	50.00	3.68 (2.11)	5.90 (1.75)

4.4.2.5 Selection of variables using DIC.

- Here only three covariates are considered, resulting in eight possible models.
- We can fit each model separately to calculate DIC
- Alternatively, all models can be simultaneously fitted in WinBUGS
- Select the best model as the one with the lowest DIC value
- Code for fitting all models in a single run is provided in this book's Website (a simplified version follows)
- Results are summarized in Table 4.8 after 10,000 burnin and 10,000 additional iterations.
- Be careful to consider a sufficiently long burnin period because DIC is sensitive to initial values.

```
model{
# vectors used for the calculation of DIC
for (i in 1:n){
    y1[i] <- damage[i]
    ...................
    y8[i] <- damage[i]
    y1[i]~dpois( lambda[i,1] )
        ...................
    y8[i]~dpois( lambda[i,8] )

    log(lambda[i,1])<-beta[1,k]
    log(lambda[i,2])<-beta[1,k] + beta[2,k]*type[i]
    log(lambda[i,3])<-beta[1,k] + beta[3,k]*bombload[i]
    log(lambda[i,4])<-beta[1,k] + beta[2,k]*type[i] + beta[3,k]*bombload[i]
    log(lambda[i,5])<-beta[1,k] + beta[4,k]*airexp[i]
    log(lambda[i,6])<-beta[1,k] + beta[2,k]*type[i] + beta[4,k]*airexp[i]
    log(lambda[i,7])<-beta[1,k] + beta[3,k]*bombload[i] + beta[4,k]*airexp[i]
    log(lambda[i,8])<-beta[1,k] + beta[2,k]*type[i] + beta[3,k]*bombload[i] +
        beta[4,k]*airexp[i]
}
# prior
for (k in 1:8){ for (j in 1:4){ beta[j,k]~dnorm( 0.0, 0.001 ) } }
}
```

Table 4.8: DIC values for all eight models under consideration for Example 4.1[a]

	Dbar	Dhat	pD	DIC	Model
y1	108.6	107.6	1.01	109.6	Constant
y2	94.0	92.0	1.99	96.0	Type
y3	84.8	82.9	1.91	86.7	Bombload
y4	85.3	82.4	2.95	88.3	Type + Bombload
y5	106.2	104.3	1.97	108.2	Airexp
y6	88.9	85.9	3.01	92.0	Type + Airexp
y7	83.9	81.0	2.92	86.9	Bombload + Airexp
y8	83.7	79.7	3.98	87.7	Type + Bombload + Airexp
total	735.6	715.9	19.76	755.4	

[a]Burnin=10,000; iterations kept=10,000.

- Lowest DIC (86.7) $\Rightarrow$ only the bombload on the linear predictor.
- DIC value (86.9) of the model `bombload + Crew experience` is very close to the lowest DIC value.
- The two models have similar predictive performance.

4.4.3 A Poisson regression model for modeling football data

Example 4.2. Modeling the English premiership football data. *Modeling of football scores is becoming increasingly popular nowadays. In the present example we use the English premiership data for the season 2006–2007 to fit a simplified Poisson log-linear model for the prediction of model outcomes. Data were downloaded from the Webpage* `http://soccernet-akamai.espn.go.com`.

4.4.3.1 Background information and the model.

- Simple and basic log-linear model originally introduced by Maher (1982).
- Also used by other authors such as Lee (1997) and Karlis and Ntzoufras (2000).

We denote by

- y_{i1} = goals scored by home team (HT) in the ith game.
- y_{i2} = goals scored by away team (AT) in the ith game.

The model is expressed by

$$\begin{array}{rcll} Y_{ij} & \sim & \text{Poisson}(\lambda_{ik}) & \text{for} \quad j=1,2 \\ \log(\lambda_{i1}) & = & \mu + \text{home} + a_{\text{HT}_i} + d_{\text{AT}_i} & \\ \log(\lambda_{i2}) & = & \mu \qquad\qquad + a_{\text{AT}_i} + d_{\text{HT}_i} & \text{for} \quad i=1,2,\ldots,n, \end{array}$$

where n = number of games, μ = constant parameter; home = home effect; HT_i and AT_i = home and away teams in i game; a_k and d_k = attacking and defensive effects–abilities of k team for $k = 1, 2, \ldots, K$; and K = number of teams in the data (here $K = 20$).

Constraints

We use the sum-to-zero constraints for attacking and defensive parameters (a_k and d_k), (to make the model identifiable). Hence we set

$$\sum_{k=1}^{K} a_k = 0 \ \text{ and } \ \sum_{k=1}^{K} d_k = 0. \tag{6}$$

By this way we compare the ability of each team with an overall level of attacking and defensive abilities.

Interpretation

- μ = overall level of log–expected goals scored in away games
- $home$ = home effect (i.e. difference between the log–expected goals scored by the home and away teams of equal strength compete with each other).
- a_k = attacking ability of k team. Deviation from an overall average level.
- d_k = defensive ability of k team. Deviation from an overall average level.

Positive attacking parameter $\rightarrow$ the team under consideration has an offensive performance that is better than the average level of the teams competing in the league.

Negative defensive parameter $\rightarrow$ defensive performance better than the average level of the teams competing in the league.

Two important model assumptions

1. **Independence between home and away goals**

   Empirical evidence and exploratory analysis has shown a (relatively low) correlation between the goals in a football game.

   Possible solutions

   (a) Bivariate Poisson distribution (Karlis and Ntzoufras, 2003).

   (b) Modeling the goal differences using Skellam's distribution (Karlis and Ntzoufras, 2008).

   These models are natural extensions of the simplified model presented here.

2. **Equality between mean and the variance of goals**

   Slight overdispersion has been reported in literature; see Karlis and Ntzoufras (2000) for a discussion.

   Possible sollution: Use over-dispersed distributions such as the negative binomial distribution; see, for example, Reep and Benjamin (1968), Reep et al. (1971), and Baxter and Stevenson (1988).

Another problem: The excess of draws

Another problem appearing in football data is the excess of specific scores, especially the 0–0 and 1–1 draws.

Dixon and Coles (1997) provided an extension based on the Poisson model allowing for extra probabilities in these scores.

In a similar fashion, Karlis and Ntzoufras (2003) have proposed using a diagonal inflated bivariate Poisson model and more recently a zero inflated model for the goal differences (Karlis and Ntzoufras, 2008).

4.4.3.2 Model specification in WinBUGS .

Data

1. goals1: Goals scored by the home team.
2. goals2: Goals scored by the away team.
3. ht: Code of the home team.
4. at: Code of the away team.

WinBUGS code

```
for (i in 1:n){
    # stochastic component
    goals1[i] ~ dpois( lambda1[i] )
    goals2[i] ~ dpois( lambda2[i] )
    # linear predictor
    log(lambda1[i]) <- mu + home + a[ ht[i] ] + d[ at[i] ]
    log(lambda2[i]) <- mu        + a[ at[i] ] + d[ ht[i] ]
}
```

Imposing the constraints in WinBUGS

STZ constraints (6) can be imposed in WinBUGS by setting one set of parameters effects (e.g., here we use $k = 1$) equal to

$$a_1 = -\sum_{k=2}^{K} a_k \text{ and } d_1 = -\sum_{k=2}^{K} d_k$$

using the WinBUGS syntax

```
a[1] <- -sum( a[2:K] )
d[1] <- -sum( d[2:K] )
```

Prior distributions

Prior distributions must be defined for the remaining parameters: μ, *home* and a_k, d_k for $k = 2, \ldots, K$.

The usual normal low-information prior with zero mean and large prior variance are used here (with prior precision equal to 10^{-4}).

A large amount of historical data are available in sports. Such information can be used to form a plausible prior distribution, which will be extremely useful in the first weeks of a competition when a limited amount of data are available. Elicitation of such historical data and their potential usefulness, especially in the beginning of each season (where limited data are available), must be carefully examined.

4.4.3.3 Results.

Posterior summaries of the Poisson log-linear model parameters are provided in Table 4.9.

Posterior credible intervals of attacking and defensive parameters for each team are depicted in Figures 4.2 and 4.3.

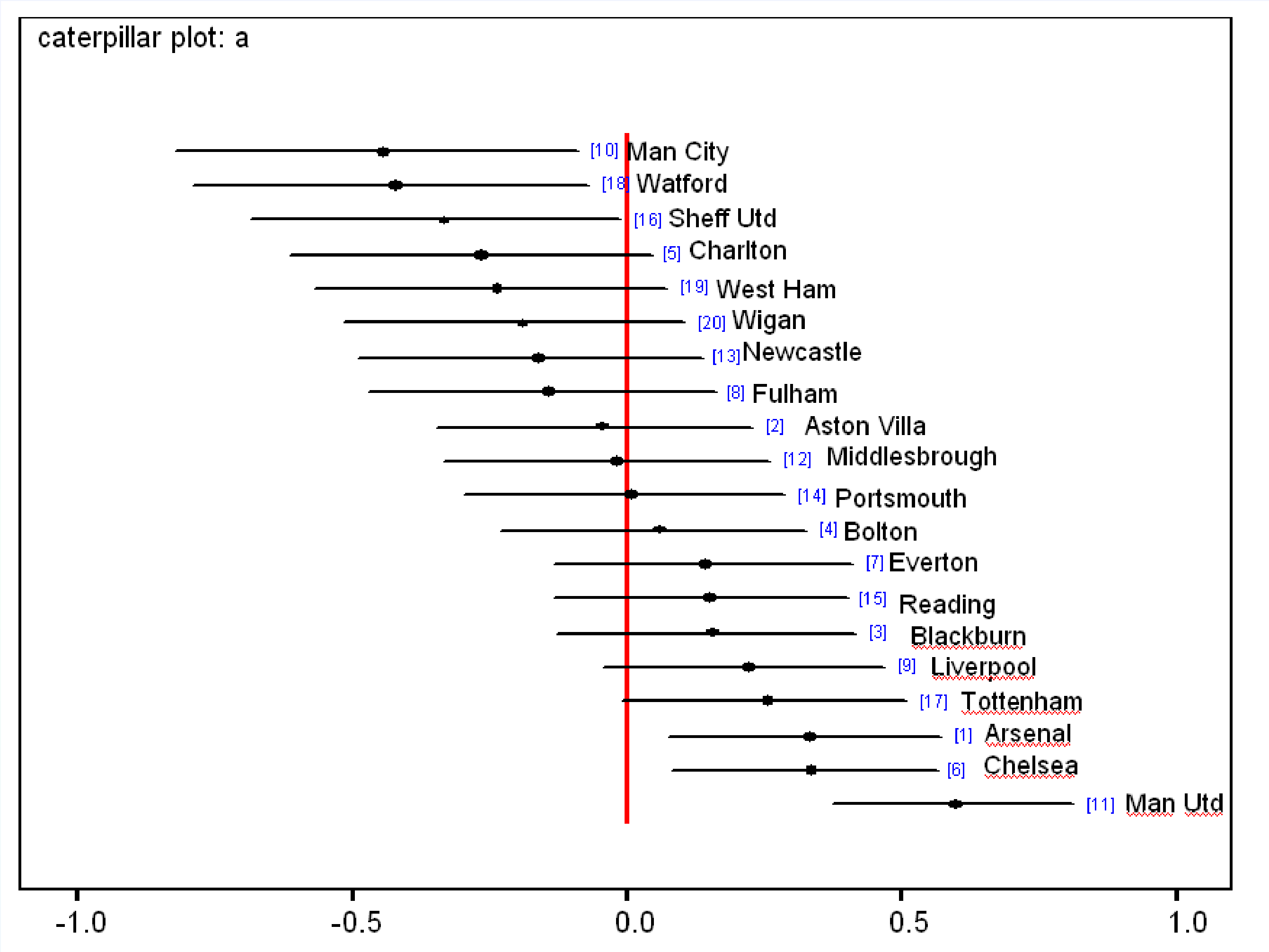
Parameter Interpretation

- Manchester United had the highest attacking parameter.
- Chelsea had the lowest (i.e., best) defensive parameter.
- In a game between two average teams: expected score = 1.3 – 0.9 .
- Home effect: Increases the expected number of goals by 46% .

Table 4.9: Posterior summaries of expected number of goals for Example 4.2

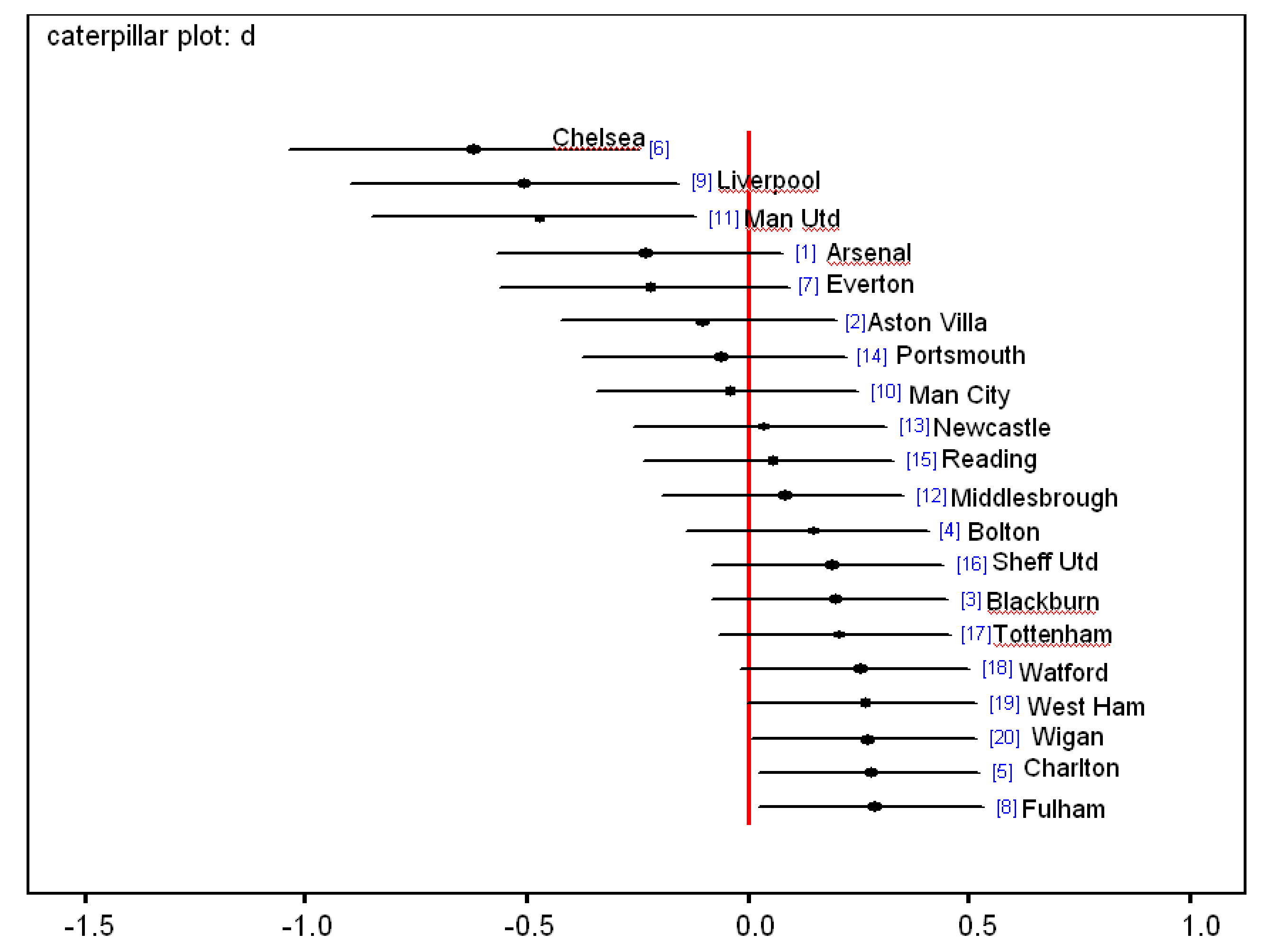
	Team[a]	Node	Posterior Mean	Posterior SD	Posterior percentiles 2.5%	Posterior percentiles 97.5%	Node	Posterior Mean	Posterior SD	Posterior percentiles 2.5%	Posterior percentiles 97.5%
1.	Arsenal	a[1]	0.33	0.12	0.08	0.57	d[1]	-0.23	0.16	-0.57	0.08
2.	Aston Villa	a[2]	-0.05	0.15	-0.34	0.23	d[2]	-0.10	0.16	-0.42	0.19
3.	Blackburn	a[3]	0.16	0.14	-0.13	0.42	d[3]	0.20	0.14	-0.08	0.45
4.	Bolton	a[4]	0.06	0.14	-0.23	0.32	d[4]	0.15	0.14	-0.13	0.41
5.	Charlton	a[5]	-0.26	0.16	-0.61	0.04	d[5]	0.28	0.13	0.03	0.52
6.	Chelsea	a[6]	0.33	0.12	0.08	0.56	d[6]	-0.62	0.20	-1.04	-0.25
7.	Everton	a[7]	0.14	0.14	-0.13	0.41	d[7]	-0.22	0.17	-0.56	0.09
8.	Fulham	a[8]	-0.14	0.16	-0.47	0.16	d[8]	0.29	0.13	0.02	0.53
9.	Liverpool	a[9]	0.22	0.13	-0.04	0.47	d[9]	-0.51	0.19	-0.90	-0.16
10.	Man City	a[10]	-0.44	0.18	-0.82	-0.09	d[10]	-0.04	0.15	-0.34	0.24
11.	Man Utd	a[11]	0.60	0.11	0.38	0.81	d[11]	-0.47	0.19	-0.85	-0.12
12.	Middlesbrough	a[12]	-0.02	0.15	-0.33	0.26	d[12]	0.09	0.14	-0.20	0.35
13.	Newcastle	a[13]	-0.16	0.16	-0.48	0.13	d[13]	0.03	0.14	-0.26	0.31
14.	Portsmouth	a[14]	0.00	0.15	-0.29	0.29	d[14]	-0.07	0.15	-0.37	0.22
15.	Reading	a[15]	0.15	0.14	-0.13	0.40	d[15]	0.05	0.15	-0.24	0.33
16.	Sheff Utd	a[16]	-0.33	0.17	-0.68	-0.01	d[16]	0.19	0.13	-0.08	0.44
17.	Tottenham	a[17]	0.26	0.13	-0.00	0.51	d[17]	0.20	0.13	-0.06	0.45
18.	Watford	a[18]	-0.42	0.18	-0.78	-0.07	d[18]	0.25	0.13	-0.01	0.50
19.	West Ham	a[19]	-0.24	0.16	-0.56	0.07	d[19]	0.27	0.13	0.01	0.51
20.	Wigan	a[20]	-0.19	0.16	-0.51	0.11	d[20]	0.27	0.13	0.01	0.51
		home	0.38	0.07	0.25	0.51	μ	-0.10	0.05	-0.20	0.003

Abbreviations: Man = Manchester; Utd = United; Sheff = Sheffield; Ham = Hampshire.



Abbreviations: Man = Manchester; Utd = United; Sheff = Sheffield; Ham = Hampshire.

Figure 4.2: 95% posterior intervals for team attacking parameters for Example 4.2.



Abbreviations: Man = Manchester; Utd = United; Sheff = Sheffield; Ham = Hampshire.

Figure 4.3: 95% posterior intervals for team defensive parameters for Example 4.2.

4.4.3.4 Prediction of future games.

Models in sports are used mainly for prediction.

Here we briefly illustrate how we can obtain predictions for two future games.

The approach can be easily generalized to additional games using the same approach.

To illustrate the implementation in WinBUGS , we have substituted the scored goals in the last two games (Tottenham – Manchester City and Watford – Newcastle) by `NA`.

WinBUGS automatically will generate values for the missing goals from the predictive distribution and will provide estimates for each score by monitoring the nodes `goals1` and `goals2`.

Posterior summaries of the predicted scores are given in Table 4.10.

Posterior means indicate that the observed actual score was expected under the fitted model.

Table 4.10: Posterior summaries of the expected scores for last two games of Example 4.2

		Actual	Posterior		Posterior summaries of goal difference		
Home team	Away team	score	Median	Mean	Mean	SD	95% CI
379. Tottenham	Manchester City	2–1	1–1	1.65 – 0.72	0.93	1.59	(-2, 4)
380. Watford	Newcastle	1–1	1–1	0.93 – 1.02	-0.09	1.43	(-3, 3)

Interest also lies in calculating in the probability of each outcome (win/draw/loss), which can be easily accommodated in WinBUGS using the following syntax:

```
# calculation of the predicted differences
pred.diff[1] <- goals1[379]-goals2[379]
pred.diff[2] <- goals1[380]-goals2[380]
#
# probability of each game outcome (win/draw/loss)
for (i in 1:2){
    outcome[i,1] <- 1 - step( -pred.diff[i] )   #home wins (diff>0)
    outcome[i,2] <- equals( pred.diff[i] , 0.0 )#draw (diff=0)
    outcome[i,3] <- 1-step( pred.diff[i] )     #home loses (diff<0)
}
```

In this syntax, the elements of `outcome` are binary indicators denoting the win, draw, and loss of the home team in each column, respectively.

Using similar syntax we can also estimate the probabilities of the expected differences. The syntax now is given by

```
# calculation of the probability of each difference
for (i in 1:2){
    pred.diff.counts[i,1]<- 1-step(pred.diff[i]+5) # less than -5
    # equal to k-7 (-5 to 5)
    for (k in 2:12){
        pred.diff.counts[i,k]<-equals(pred.diff[i],k-7)}
    pred.diff.counts[i,13]<-step(pred.diff[i]-6) # greater than 5
}
```

In this syntax `pred.diff.counts` is again a matrix with binary elements indicating which difference appears in each MCMC iteration.

Elements 2–12 denote differences from -5 to 5, while the first and last elements denote differences lower than -5 and higher than 5, respectively.

Posterior probabilities of each predicted outcome for the games Tottehnam vs. Man. City and Watford vs. Newcastle are summarized in Table 4.11.

Outcome probabilities indicate that Tottenham's probability of winning the game against Manchester City was about 60%.

The posterior model probabilities confirm that the two teams have about equal probabilities of winning the game.

Table 4.11: Posterior probabilities of each game outcome for last two games of Example 4.2

				Posterior Probability		
			Actual	Home		Away
	Home team	Away team	score	wins	Draw	wins
379.	Tottenham	Manchester City	2–1	0.59	0.24	0.17
380.	Watford	Newcastle	1–1	0.33	0.30	0.37

Posterior distribution of the goal difference for the games Tottehnam vs. Man. City and Watford vs. Newcastle are provided in Table 4.12.

Posterior mode

- Tottehnam vs. Man. City **=>** One goal difference.
- Tottehnam vs. Man. City **=>** Zero goal difference.

Table 4.12: Posterior probabilities of each game goal difference for last two games of Example 4.2

	Home Team	Away Team	Actual Score	Posterior Probability of goal difference[a] ≤ -3	-2	-1	0	1	2	3	4	≥ 5
379.	Tottenham	Man City	2–1	0.012	0.036	0.119	0.242	**0.257**	0.185	0.090	0.037	0.023
380.	Watford	Newcastle	1–1	0.042	0.108	0.218	**0.303**	0.210	0.086	0.024	0.006	0.002

[a] Boldface indicates the maximum probability and the corresponding posterior mode of the difference.

Abbreviations: Man = Manchester.

4.4.3.5 Regeneration of the full league.

Interest also lies in reconstructing the league using the predictive distribution.

Such practice is useful to evaluate whether the final observed ranking was plausible under the fitted model.

It can be interpreted as the uncertainty involved in the final ranking if the league is repeated and the model is true; see Karlis and Ntzoufras (2008).

In order to reconstruct the full table in WinBUGS , we need to replicate the full scores in a tabular $K \times K$ format and then calculate the number of points for each team.

The replicated league (in tabular $K \times K$ format) is calculated using the following syntax:

```
for (i in 1:K){ for (j in 1:K){
     # replicated league
     goals1.rep[i,j]~dpois(lambda1.rep[i,j])
     goals2.rep[i,j]~dpois(lambda2.rep[i,j])
     # link and linear predictor
     log(lambda1.rep[i,j])<-  mu + home + a[ i ] + d[ j ]
     log(lambda2.rep[i,j])<-  mu        + a[ j ] + d[ i ]
     # replicated difference
     goal.diff.rep[i,j] <- goals1.rep[i,j]-goals2.rep[i,j]
}
```

The **total number of points** is calculated using the following syntax:

```
for (i in 1:K){ for (j in 1:K){
    # points earned by each home team (i)
    points1[i,j]<-3 * (1-step(-goal.diff.rep[i,j]))
                    + equals(goal.diff.rep[i,j],0)
    # points earned by each away team (j)
    points2[i,j]<-3 * (1-step( goal.diff.rep[i,j]))
                    + equals(goal.diff.rep[i,j],0)
}}
# calculation of the total points for each team
for (i in 1:K){
    total.points[i] <- sum( points1[i,1:20] ) -  points1[i,i]
                     + sum( points2[1:20,i] ) -  points2[i,i] }
```

- `points1` and `points2` = number of points in each game for the home and the away teams, respectively (arranged in two $K \times K$ matrices).
- Total points of i team = points for team i in home games (sum of i row of node `points1`) plus points in away games (sum of j column of node `points2`).
- `points1[i,i]` and `points2[i,i]` refer to each team playing against itself.

Posterior analysis of predicted points

Posterior summaries of the total predicted earned points are obtained as usual (`sample monitor tool`), while posterior summaries of the ranks are obtained using the `rank monitor tool`.

Results are summarized in Table 4.14 after 5000 iterations kept (and an additional 1000 iterations removed).

Ranks refer to the total number of points in ascending order. Hence 20 refers to the team with the highest number of collected points (i.e., the champion), while one (1) refers to the team with the lowest number of collected points (i.e., the worst team in the league). The league was reproduced successfully. No differences are observed in the first four teams, while minor changes are observed for the remaining positions.

Further analysis for rankings

Probabilities for each ranking can be obtained by

1. Calculating the ranks using the commands

```
for (i in 1:K){ ranks[i] <- 21-rank(total.points[], i) }
```

2. Calculating binary dummies for each ranking using the commands

```
for (i in 1:K){ for (j in 1:K){
                          rank.probs[i,j]<-equals(ranks[i],j) }}
```

From the results provided in Table 4.15, we observe that Manchester United was clearly better than the other teams since its probability of winning the league was about 60% versus 25% for Chelsea, which ended up second in the league.

More detailed analysis on the topic, using the Skellam's Poisson difference distribution, can be found in Karlis and Ntzoufras (2008).

Table 4.13: Observed and predicted (by model) points and rankings for all teams

Predicted (actual) ranking[a]	Team	Actual points	Posterior summaries for total points					Posterior percentiles for points ranks[b]		
			Mean	SD	2.5%	Median	97.5%	2.5%	Median	97.5%
1 (1)	Man Utd	89	84.7	8.1	68	85	99	16	20	20
2 (2)	Chelsea	83	78.3	8.8	60	79	94	14	19	20
3 (3)	Liverpool	68	72.6	9.2	54	73	90	12	18	20
4 (3)	Arsenal	68	69.9	9.4	51	70	88	10	17	20
5 (6)	Everton	58	62.8	9.4	44	63	81	7	15	19
6 (8)	Reading	55	55.8	9.6	37	56	75	4	13	18
7 (5)	Tottenham	60	56.0	9.6	36	55	73	4	12	18
8 (9)	Portsmouth	54	54.3	9.6	36	54	73	3	12	18
9 (11)	Aston Villa	50	53.4	9.7	34	53	73	3	12	18
10 (10)	Blackburn	52	51.9	9.6	33	52	70	3	11	17

Table 4.14: Observed and predicted (by model) points and rankings for all teams

Predicted (actual) ranking[a]	Team	Actual points	Posterior summaries for total points					Posterior percentiles for points ranks[b]		
			Mean	SD	2.5%	Median	97.5%	2.5%	Median	97.5%
11 (7)	Bolton	56	49.6	9.3	32	49	69	2	10	17
12 (12)	Middlesbrough	46	49.2	9.4	32	49	68	2	10	17
13 (13)	Newcastle	43	46.0	9.4	28	46	65	1	8	16
14 (14)	Man City	42	40.8	9.0	24	40	59	1	6	14
15 (16)	Fulham	39	39.3	8.9	22	39	57	1	5	13
16 (17)	Wigan	38	38.8	9.1	22	39	57	1	5	13
17 (15)	West Ham	41	37.5	8.7	21	37	55	1	4	13
18 (18)	Sheff Utd	38	37.2	8.8	21	37	55	1	4	12
19 (19)	Charlton	34	36.4	8.9	19	36	55	1	4	12
20 (20)	Watford	28	33.2	8.6	17	33	51	1	3	11

[a] Predicted ranks are calculated using the median rank and then the mean points.
[b] *Ranks* here refers to the number of points in ascending order (e.g., 20 denotes the best team and 1, the worst team in terms of collected points).
Abbreviations: Man = Manchester; Utd = United; Sheff = Sheffield; Ham = Hampshire.

Table 4.15: Posterior mean, standard deviation of final league ranks, and posterior probabilities of each position (in %)

	Posterior		Posterior probability of each ranking[a]																			
Node	mean	SD	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Man Utd	1.7	1.1	**60**	23	10	4	1	1														
Chelsea	2.6	1.6	25	**33**	21	11	6	2	1	1												
Liverpool	3.7	2.0	10	21	**25**	18	11	7	3	3	1	1										
Arsenal	4.3	2.3	6	15	**21**	**21**	13	9	6	4	2	1	1	1								
Everton	6.2	2.9	1	5	10	16	**17**	13	10	8	6	5	3	2	1	1	1					
Reading	8.5	3.6		2	4	6	10	**12**	11	11	9	9	7	6	4	3	3	2	1	1	1	
Tottenham	8.8	3.6		1	3	6	10	11	**12**	10	10	9	7	6	4	4	3	2	1	1	1	
Portsmouth	9.0	3.7		1	2	6	8	11	10	**11**	9	9	8	6	5	4	3	2	2	1	1	
Aston Villa	9.5	3.8		1	2	5	7	9	10	**11**	10	9	8	7	6	5	4	3	2	2	1	
Blackburn	9.9	3.8		1	2	4	7	8	9	9	**10**	**10**	9	8	7	5	5	3	2	1	1	1
Bolton	10.8	3.8			1	2	4	6	7	**9**	**9**	**9**	**9**	**9**	8	7	6	4	3	2	2	1
Middlesbrough	11.0	3.9			1	2	4	6	7	8	**9**	**9**	**9**	**9**	**9**	7	5	5	4	3	2	1
Newcastle	12.3	3.9				1	2	3	5	6	8	8	9	9	**10**	8	9	6	6	4	4	2
Man City	14.4	3.7					1	1	2	3	4	6	6	7	8	9	10	**11**	10	9	8	6
Fulham	14.9	3.6						1	2	2	3	4	5	7	8	9	9	10	**11**	**11**	10	8
Wigan	15.1	3.6					1	1	2	2	4	4	5	5	8	10	9	10	**11**	10	10	10
West Ham	15.6	3.4						1	1	1	2	3	4	6	7	8	9	11	11	**12**	**12**	11
Sheff Utd	15.8	3.3							1	2	3	3	4	5	7	8	9	10	12	**13**	12	12
Charlton	16.0	3.3							1	2	2	3	3	5	6	8	8	10	12	12	14	**15**
Watford	17.1	2.9								1	1	2	2	3	4	5	8	8	10	14	18	**24**

[a] Boldface indicates highest probability for each team. Posterior percentages were rounded to the closest integer, while percentages $< 0.5\%$ were omitted. Sums of probabilities for each column are slightly higher than 100%, due to ties.
Abbreviations: Man = Manchester; Utd = United; Sheff = Sheffield; Ham = Hampshire.

4.5 Binomial response models

Binomial data are frequently encountered in modern science, especially in medical research, where the response is usually binary, indicating whether a person has a specific disease.

Most popular model: logistic regression model (binomial with logit link).

- Canonical link (i.e. default choice)
- It has a smooth interpretation based on the odds of $Y = 1$ versus $Y = 0$ $\dfrac{\pi}{1-\pi}$ (π is the probability of success for Y).

The logistic regression model model can be summarized by

$$Y_i \sim \text{binomial}(\ \pi_i, N_i), \quad \log \frac{\pi_i}{1-\pi_i} = \beta_0 + \sum_{j=1}^{p} \beta_j x_{ij} = \boldsymbol{X}_{(i)}\boldsymbol{\beta}$$

for $i = 1, 2, \ldots, n$.

- For $N_i = 1 \Rightarrow Y_i$ is Bernoulli.
- Other link functions $\Rightarrow$ probit and clog–log.

4.5.1 Interpretation of model parameters in binomial response models

4.5.1.1 Odds and odds ratios.

- Interpretation of the parameters in logistic regression models is based in the notion of odds and odds ratios.
- We define as *odds* the relative probability of two events.
- In binomial data *odds* is the relative probability of success $(Y = 1)$ compared to the probability of failure $(Y = 0)$.

$$\text{odds} = \frac{\pi}{1-\pi}$$

- The logistic model can be rewritten as

$$Y_i \sim \text{binomial}\left(\frac{\text{odds}_i}{1+\text{odds}_i}, N_i\right), \quad \log(\text{odds}_i) = \beta_0 + \sum_{j=1}^{p} \beta_j x_{ij} = \boldsymbol{X}_{(i)}\boldsymbol{\beta}$$

using the odds representation.

Interpretation of odds

- The number we multiply the probability of failure to obtain the probability of success: $\boxed{\pi = \text{odds} \times (1 - \pi)}$
- For example,
  - odds $= 2 \Rightarrow$ the success probability is twice as high as the failure probability
  - odds $= 0.6 \Rightarrow$ the success probability is equal to 60% of the failure probability.
- The value of 1 is of central interest $\Rightarrow$ probabilities of both outcomes are equal (to 0.5).
- Odds $> 1 \Rightarrow$ an increased probability of success in contrast to the failure probability ($\pi > 0.5$),
- Odds $< 1 \Rightarrow$ a probability of success lower than the probability of failure ($\pi < 0.5$).

Interpretation of odds - Percentage change of probabilities

Quantity $(\text{odds} - 1) \times 100 \Rightarrow$ the percentage increase or decrease (depending on the sign) of the success probability in comparison to the failure probability.

For example

- $\text{odds} = 1.6 \Rightarrow$ the success probability is 60% higher than the corresponding failure probability
- $\text{odds} = 0.6 \Rightarrow$ the success probability is 40% lower than the corresponding failure probability

Odds Ratios

- The ratio of two odds of two different outcomes are called *odds ratios* (OR)
- They provide the relative change of the odds under two different conditions (denoted by $X = 1, 2$ and subscripts 1 and 2):

$$\mathrm{OR}_{12} = \frac{\mathrm{odds}(X=1)}{\mathrm{odds}(X=2)} \ ,$$

  where $\mathrm{odds}(X = x)$ denotes the conditional success odds given that $X = x \mathrm{odds}(X = x) = \frac{P(Y=1|X=x)}{P(Y=0|X=x)}$.
- $\mathrm{OR}_{12} = 1 \Rightarrow$ conditional odds under comparison are equal $\Rightarrow$ no difference in the relative probabilities of Y under $X = 1$ and $X = 2$.
- Same interpretation as in *odds* but replace *probability* $\Rightarrow$ *odds*.
- $(\mathrm{OR}_{12} - 1) \times 100$ provides the percentage change of the odds for $X = 1$ compared with the corresponding odds when $X = 2$.

Table 4.16: Summary interpretation table for odds ratios

INTERPRETATION OF ODDS RATIOS

$$\mathrm{OR}_{12} = \frac{\mathrm{odds}(X=1)}{\mathrm{odds}(X=2)} = a \ .$$

- If $a = 1 \Rightarrow \mathrm{odds}(X=1) = \mathrm{odds}(X=2)$.
- If $a < 1 \Rightarrow \mathrm{odds}(X=1) < \mathrm{odds}(X=2)$.
- If $a > 1 \Rightarrow \mathrm{odds}(X=1) > \mathrm{odds}(X=2)$.
- The success odds when $X = 1$ is a times as high as the corresponding odds for $X = 2$
- If $a > 1$, then the success odds when $X = 1$ are $(a-1) \times 100\%$ times higher than the corresponding odds for $X = 2$.
- If $a < 1$, then the success odds when $X = 1$ are $(1-a) \times 100\%$ times lower than the corresponding odds for $X = 2$.

4.5.1.2 Logistic regression parameters and odds ratios.

- Interpretation is based on $B_j = e^{\beta_j}$ as in the Poisson log-linear models.
- B_j are associated with odds ratios since

$$\begin{aligned} \log\big(\text{odds}(x)\big) &= \beta_0 + \beta_1 x \Rightarrow \\ \text{odds}(x) &= B_0 B_1^x \Rightarrow \\ \text{OR}_{x+1,x} &= \frac{\text{odds}(x+1)}{\text{odds}(x)} = \frac{B_0 B_1^{x+1}}{B_0 B_1^x} = B_1 = e^{\beta_1} \end{aligned}$$

  in the simple logistic regression case with one numerical covariate.
- $B_1 = e^{\beta_1}$ denotes the relative odds magnitude when X increases by one unit.
- For $X_i = -\beta_0/\beta_1 \Rightarrow$ odds $= 1 \Rightarrow$ probabilities equal to 0.5.

  Threshold for prediction or for diagnosing future patients using the X variable directly.

Parameter Interpretation for categorical covariates

- When X is categorical variable with K levels (CR parametrization with the 1st level as baseline/reference category) then

$$
\begin{aligned}
\log\left(\text{odds}(x)\right) &= \beta_0 + \sum_{j=2}^{K} \beta_j I(x=j) = \beta_0 + \sum_{j=2}^{K} \beta_j D_j \Rightarrow \\
\text{odds}(x) &= B_0 \prod_{j=2}^{K} B_j^{D_j} \Rightarrow \\
OR_{j1} &= \frac{\text{odds}(j)}{\text{odds}(1)} = \frac{e^{\beta_0+\beta_j}}{e^{\beta_0}} = \frac{B_0 B_j}{B_0} = B_j = e^{\beta_j} \ .
\end{aligned}
$$

- B_j is the success odds ratio for the jth category of X versus the reference category of the same variable.

Parameter Interpretation in multiple logistic regression

- Extension of the interpretation above to multiple logistic regression models is straightforward.
- We only need to interpret each B_j as the change of Y when a single covariate X_j increases by one unit while the other covariates remain constant.
- Odds ratios estimated via multiple logistic regression models (i.e. B_j) are often reported
  - "Odds ratios after controlling for the effect" of the effect of the remaining covariates
  - "Odds ratios adjusted for" of the effect of the remaining covariates
  - Adjusted odds ratios
- Adjusted odds ratios estimate the joint effect of all covariates X_j ($j = 1, 2, \ldots, p$) on Y, and in this way we essentially calculate the effect of each covariate after the elimination of the effect of the other covariates.

Parameter interpretation in models with other link functions

- Use latent variable interpretation
- Use approximate linear effects (called marginal effects) via derivatives of the location parameters
- Use profiles.
- Logit and Probit links provide similar models; see Agresti (2002) for details.
- Clog-log provides different models since the link is not symmetric
- See Sections 7.5.1.3–5 of Ntzoufras (2009) for a detailed description and a concise summary Table (7.17)

4.5.2 A simple example

Example 4.3. Analysis of senility symptoms data using WinBUGS.

- *We consider the data of Agresti (1990, pp. 122–123)*
- *54 elderly people completed a subtest of the Wechsler Adult Intelligence Scale (WAIS) resulting in a discrete score with range from 0 to 20.*
- *Aim: identify people with senility symptoms (binary variable) using the WAIS score.*
- *Interest also lies in calculating WAIS scores that correspond to increased probability of senility symptoms (i.e., with $\pi > 0.5$).*
- *The data of this example can be found in the book's Website and are reproduced with the permission of John Wiley and Sons, Inc.*

4.5.2.1 Model specification in WinBUGS.

- **Response**: senility symptoms – binary $\Rightarrow$ Bernoulli or the binomial with $N=1$ distributions can be used
- **Explanatory variable** x: WAIS score (discrete quantitative).
- Specification of the likelihood in WinBUGS:

```
for (i in 1:n){
     senility[i] ~ dbin( pi[i], 1 )
     logit( pi[i] ) <- beta0 + beta1 * wais[i]
}
```

  where $n=54$.

- Alternatively, the Bernoulli distribution (`dbern(pi[i])`) can be used instead.

Other parameters of interest in WinBUGS

- $B_j = e^{\beta_j}$ can be defined directly in WinBUGS

```
odds0 <- exp( beta0 )
or    <- exp( beta1 )
```

- The threshold value $X = x(\pi = 0.5)$ can be defined in WinBUGS

```
wais.half.prob <- - beta0/beta1
```

- **`wais.half.prob`** refers individuals with disease probability equal to 0.5 (i.e. odds=0) since $0 = \beta_0 + \beta_1 X \Leftrightarrow X = -\beta_0/\beta_1$.
- Using the same approach, we may define X values for other probabilities (e.g., 0.25 or 0.01).

Models with other link functions in WinBUGS

- To define the probit and clog–log models, we only need to substitute the WinBUGS function `logit(pi[i])` by the corresponding link commands

```
        probit( pi[i] ) <- beta0 + beta1 * wais[i]
```

  and

```
        cloglog( pi[i] ) <- beta0 + beta1 * wais[i]
```

  respectively.

- Other, more complicated, link functions can be defined by expressing π_i as a function of the linear predictor η_i.

- Note that, for this example, arithmetic overflows occurred when using the probit or the clog-log link of WinBUGS.

- Arithmetic overflows can be avoided by truncating the tails of each link at $(-\xi, \xi)$, $\xi > 0$; see computational notes and related code in Ntzoufras (2009).

Other parameters of interest for the Probit Link

To facilitate parameter interpretation in probit models, we calculate the following quantities:

- Approximate OR interpretation: $\xi_2 \times \beta_j$ is given by the syntax

```
xi2 <- 1.6
approx.or <- xi2 * beta1
```

  This is based on Taylor expansion; see Agresti (2002) and Ntzoufras (2009) for details.

- The threshold value $x_c = -\beta_0/\beta_1$ for the probit link is the same as in the logit one and, therefore, can be obtained using the same syntax.

Other parameters of interest for the Clog-log Link

To facilitate parameter interpretation in probit models, we calculate the following quantities:

- Approximate OR interpretation: $\xi_2 \times \beta_j$ is given by the syntax

```
approx.or <- 1.39 * beta1
```

  This is based on Taylor expansion; see Agresti (2002) and Ntzoufras (2009) for details.

- Finally, the threshold value x_c is now given by $x_c = \big(\log(\log 2) - \beta_0\big)/\beta_1$ and is specified in WinBUGS using the syntax

```
wais.half.prob <- ( log(log(2))-beta0 )/beta1
```

4.5.2.2 Results and parameter interpretation.

The usual low information priors $\beta_j \sim N(0, 1000)$ are used in this example.

Posterior summaries of the parameters for each link are provided in Table 4.17.

Table 4.17: Posterior summaries for model parameters for each link function

	Logit		Probit		clog–log	
Node	Mean	SD	Mean	SD	Mean	SD
β_0	2.507	1.229	1.402	0.661	1.447	0.721
β_1	-0.339	0.119	-0.191	0.061	-0.260	0.076
OR[a]	0.718	0.083	0.748	0.074	0.700	0.073
WAIS($\pi = 0.5$)	6.975	2.104	6.677	3.195	6.752	1.575
DIC	55.105		54.997		54.998	

[a]Exact odds ratio in logit ($= e^{\beta_1}$); approximate for probit and clog–log.

Results - parameter interpretation

- All models $\Rightarrow$ significant negative association between WAIS and senility symptoms.
- From the logit model:
  - The odds of senility symptoms for an individual with WAIS=0 are a posteriori expected to be equal to 12.27.
  - For each additional WAIS point, a decrease in disease probability by **38%** is a posteriori expected.
- Posterior odds $\approx$ decrease of **25%** (probit) and **30%** (clog–log).
- These approximations are satisfactory summaries of the overall picture since
  - range from 0.61 to 0.74 for the probit link
  - range from 0.61 to 0.77 for $wais > 4$
- Generally, this approximation is more successful for π close to 0.5.

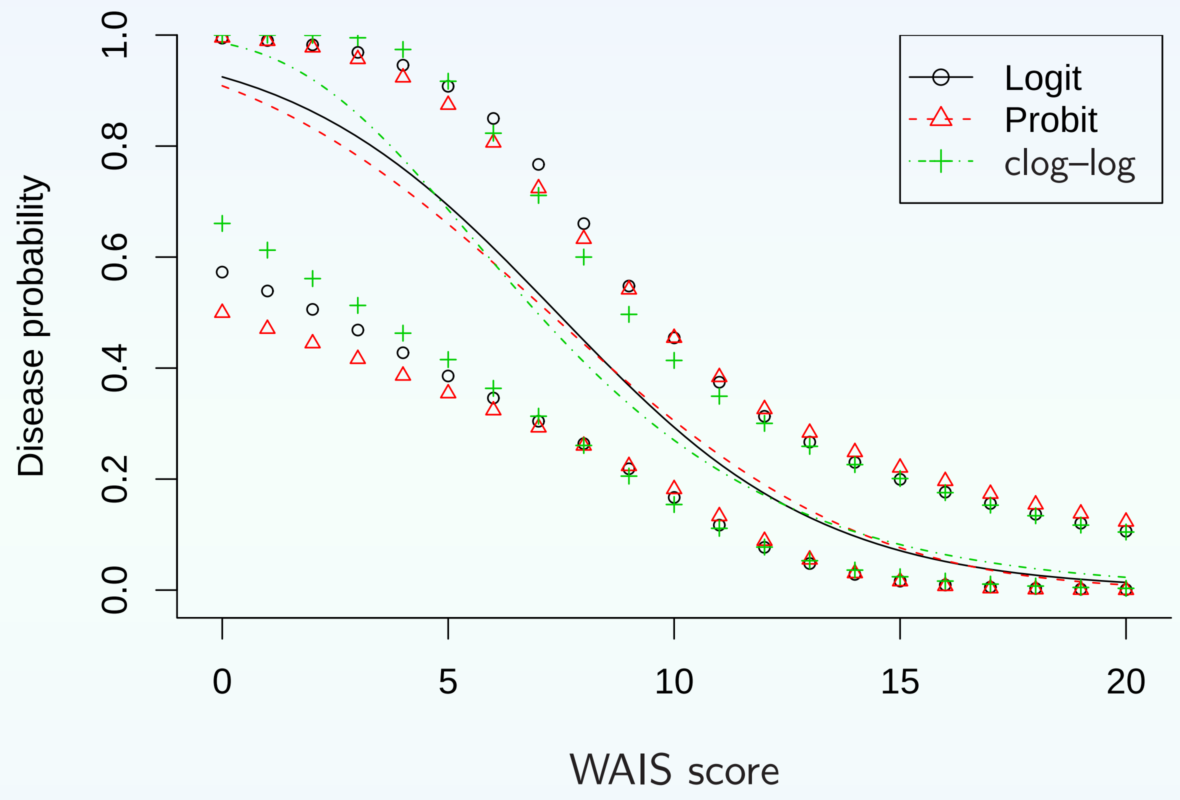


Figure 4.4: Estimated binomial models for Example 4.3. Lines represent model based on posterior means; points represent 2.5% and 97.5% posterior percentiles for disease probabilities.

- Plot was obtained in a software (`R`) outside WinBUGS using results from WinBUGS.
- The following syntax (for the probit link)

```
for (k in 1:21) {
      probit( pi.model[k] ) <- beta0 + beta1 * (k-1)
}
```

  calculates the probabilities for $x = 0, 1, 2, \ldots, 20$ used in the graph.
- For the logit and the clog–log link we only need to replace `probit` by `logit` or `cloglog` respectively in the above syntax.
- `pi.model[k]` probability of senility symptoms when $wais = k - 1$ i.e. node `pi.model` provides all possible individual "profiles" for this example.
- The central lines were obtained by the posterior means of `pi.model[k]` for $k = 1, 2, \ldots, 21$ $(x = k - 1)$ (using the monitor `tool`).
- Credible intervals were obtained by the 2.5% and 97.5% posterior percentalies of `pi.model[k]` for $k = 1, 2, \ldots, 21$ $(x = k - 1)$ (using the monitor `tool`).

Table 4.18: Posterior summaries for threshold value WAIS($\pi = 0.5$) and corresponding discrimination rules for each link

	WAIS($\pi = 0.5$)			Decision rule					
				Logit		Probit		clog–log	
Node	Logit	Probit	clog–log	Case	Healthy	Case	Healthy	Case	Healthy
Mean	6.975	6.677	6.752	$X \leq 6$	$X \geq 7$	$X \leq 6$	$X \geq 7$	$X \leq 6$	$X \geq 7$
Median	7.353	7.291	6.998	$X \leq 7$	$X \geq 8$	$X \leq 7$	$X \geq 8$	$X \leq 6$	$X \geq 7$
95% posterior interval	2.149	0.028	3.262	$X \leq 2$	$X \geq 10$	$X = 0$	$X \geq 10$	$X \leq 3$	$X \geq 9$
	9.457	9.469	8.968						

Results - threshold value

- Posterior means: All models $\Rightarrow$ cases are implied for values $X \leq 6$.
- Posterior medians:

  Logit and Probit $\Rightarrow$ cases when $X \leq 7$ (medians $= 7.35$ and 7.29).

  –log link $\Rightarrow$ cases when $X \leq 6$ (median $= 6.998$).
- Construct a more complicated decision rule based on the 95% posterior intervals:

	Logit	Probit	Clog–log
Case when WAIS	≤ 2	$= 0$	≤ 3
Healthy when WAIS	≥ 10	≥ 10	≥ 9
Cannot decide for WAIS within	3–9	1–9	4–8

- For the clog–log link, the neutral zone is narrower.

Results - DIC

- DIC for probit is the lowest with minor differences from clog–log and probit
- All differences lower than 2 $\Rightarrow$ minor differences in the fit of the three models.

References

Agresti, A. (1990), *Categorical Data Analysis*, Wiley-Interscience, New York.

Agresti, A. (2002), *Categorical Data Analysis*, 2nd ed., Wiley-Interscience, Hoboken, NJ.

Agresti, A. (2010), *Analysis of Ordinal Categorical Data*, 2nd ed., Wiley-Interscience, Hoboken, NJ.

Baxter, M. and Stevenson, R. (1988), "Discriminating between the Poisson and negative binomial distributions: An application to goal scoring in association football", *Journal of Applied Statistics* **15**, 347–438.

Berkson, J. (1944), "Application of the logistic function to bio-assay", *Journal of the American Statistical Association* **39**, 357–365.

Berkson, J. (1951), "Why I prefer logits to probits", *Biometrics* **7**, 327–339.

Clyde, M., DeSimone, H. and Parmigiani, G. (1996), "Prediction via orthogonalized model mixing", *Journal of the American Statistical Association*

91, 1197–1208.

Dellaportas, P. and Smith, A. (1993), “Bayesian inference for generalized linear and proportional hazards models via Gibbs sampling”, *Journal of the Royal Statistical Society C* **42**, 443–460.

Dey, D., Ghosh, S. and Mallick, B. (2000), *Generalized Linear Models: A Bayesian Perspective*, Marcel Dekker, New York.

Dixon, M. and Coles, S. (1997), “Modelling association football scored and inefficiencies in football betting market”, *Journal of the Royal Statistical Society C* **46**, 265–280.

Fahrmeir, L. and Tutz, G. (2001), *Multivariate Statistical Modelling Based on Generalized Linear Models*, Springer Series in Statistics, 2nd ed., Springer-Verlag, New York.

Fienberg, S. (1981), *The Analysis of Cross-Classified Categorical Data*, 2nd revised ed., The MIT Press, Cambridge, MA.

Gilks, W. and Wild, P. (1992), “Adaptive rejection sampling for Gibbs sampling”, *Journal of the Royal Statistical Society C* **41**, 337–348.

Goodman, L. (1979), "Simple models for the analysis of association in cross-classifications having ordered categories", *Journal of the American Statistical Association* **74**, 537–552.

Karlis, D. and Ntzoufras, I. (2000), "On modelling soccer data", *Student* **3**, 229–244.

Karlis, D. and Ntzoufras, I. (2003), "Analysis of sports data using bivariate Poisson models", *Journal of the Royal Statistical Society D* **52**, 381–393.

Karlis, D. and Ntzoufras, I. (2006), "Bayesian analysis of the differences of count data", *Statistics in Medicine* **25**, 1885–1905.

Karlis, D. and Ntzoufras, I. (2008), "Bayesian modelling of football outcomes: Using the Skellam's distribution for the goal difference", *IMA Journal of Management Mathematics* (to appear) **52**, 381–393.

Kass, R. and Wasserman, L. (1995), "A reference Bayesian test for nested hypotheses and its relationship to the Schwarz criterion", *Journal of the*

American Statistical Association **90**, 928–934.

Knuiman, M. and Speed, T. (1988), "Incorporating prior information into the analysis of contingency tables", *Biometrics* **44**, 1061–1071.

Lee, A. (1997), "Modeling scores in the premier league: Is Manchester United really the best?", *Chance* **10**, 15–19.

Lindsey, J. (1997), *Applying Generalized Linear Models*, Springer Texts in Statistics, Springer-Verlag, New York.

Maher, M. (1982), "Modelling association football scores", *Statistica Neerlandica* **36**, 109–118.

McCullagh, P. and Nelder, J. (1989), *Generalized Linear Models*, Monographs on Statistics and Applied Probability, Vol. 37, 2nd ed., Chapman & Hall, Cambridge, UK.

Montgomery, D., Peck, E. and Vining, G. (2006), *Introduction to Linear Regression Analysis*, 4th ed., Wiley, Hoboken, NJ.

Ntzoufras, I. (2009), *Bayesian Modeling Using WinBUGS*, Wiley Series in

Computational Statistics, Hoboken, NJ.

Ntzoufras, I., Dellaportas, P. and Forster, J. (2003), "Bayesian variable and link determination for generalized linear models", *Journal of Planning and Inference* **111**, 165–180.

Reep, C. and Benjamin, B. (1968), "Skill and chance in association football", *Journal of the Royal Statistical Society A* **131**, 581–585.

Reep, C., Pollard, R. and Benjamin, B. (1971), "Skill and chance in ball games", *Journal of the Royal Statistical Society A* **134**, 623–629.

Rigby, R. and Stasinopoulos, D. (2005), "Generalized additive models for location, scale and shape (with discussion)", *Journal of the Royal Statistical Society C* **54**, 507–554.

Spiegelhalter, D. and Smith, A. (1988), "Bayes factors for linear and log-linear models with vague prior information", *Journal of the Royal Statistical Society B* **44**, 377–387.