

Homework 3

1. (Ex 1) The table below reports results from a toxicological experiment, including the number of beetles killed (y_i) after 5 hours exposure to gaseous carbon disulphide at various concentrations. Concentration (log dose, x_i) is given on the \log_{10} scale.

Consider a binomial response distribution, and assume that the y_i are independent

Log Dose, x_i	Number of beetles, m_i	Number killed, y_i
1.6907	59	6
1.7242	60	13
1.7552	62	18
1.7842	56	28
1.8113	63	52
1.8369	59	53
1.8610	62	61
1.8839	60	60

realizations from $\text{Bin}(m_i, \pi_i)$, $i = 1, \dots, n$. The objective is to study the effect of the choice of link function $g(\cdot)$, where $\pi_i = g^{-1}(\eta_i) = g^{-1}(\beta_1 + \beta_2 x_i)$.

- (a) Using R, fit binomial GLMs for these data corresponding to 3 link functions, logit, probit and complementary log-log. Perform residual analysis for each model, using the deviance residuals. Obtain fitted values, $\hat{\pi}_i$, under each model and compare with observed proportions, y_i/m_i . Obtain the estimated dose-response curve under each model by evaluating $\hat{\pi}(x) = g^{-1}(\hat{\beta}_1 + \hat{\beta}_2 x)$ over a grid of values x for log dose in the interval $(1.65, 1.9)$. Plot these curves and compare with the scatter plot of the observed x_i plotted against the observed proportions. Based on all the results above, discuss the fit of the different models.
- (b) One of the more general (parametric) link function for binomial GLMs that has been suggested in the literature is defined through

$$g_\alpha^{-1}(\eta_i) = \frac{\exp(\alpha \eta_i)}{\{1 + \exp(\eta_i)\}^\alpha} \text{ for } \alpha > 0. \quad (1)$$

Note that the logit link arises as a special case of 1, where $\alpha = 1$. Discuss the effect of the additional model parameter α , in particular, for values $0 < \alpha < 1$ and $\alpha > 1$. Provide the expression for the log-likelihood for β_1 , β_2 and α under the link in 1, and discuss the complications that arise for maximum likelihood estimation under this more general model compared with the logit GLM. (You do not need to fit the model, estimates are given below.)

- (c) The MLEs under model with link given in 1 are $\hat{\beta}_1 = -113.625$, $\hat{\beta}_2 = 62.5$ and $\hat{\alpha} = 0.279$. (The MLEs can be obtained using the Newton-Raphson method.) Using these estimates, obtain the fitted values $\hat{\pi}_i$ and the estimated dose-response curve

under the link 1. Compare with the corresponding results under the 3 models in (a). Obtain the deviance residuals from the model with link 1 and analyze them graphically.

- Compute the AIC and BIC for the 4 models considered above to compare them.

Sol 1 (a) The residual analysis for three models with logit, probit and complementary log-log link functions is summarized as following: Their mean function are $g^{-1}(\mathbf{X}\boldsymbol{\beta}) = \frac{1}{1+\exp(-\mathbf{X}\boldsymbol{\beta})}$, $g^{-1}(\mathbf{X}\boldsymbol{\beta}) = \Phi(\mathbf{X}\boldsymbol{\beta})$ and $g^{-1}(\mathbf{X}\boldsymbol{\beta}) = 1 - \exp[-\exp(\mathbf{X}\boldsymbol{\beta})]$. First, the deviance residuals for three models are shown in Table 1.

Table 1: Deviance residuals for four models: Model 1 refers to logit link; Model 2 refers to probit link; Model 3 refers to complementary log-log link, Model 4 refers to general link function

$\hat{r}_{\text{Model1}}^D$	1.28	1.06	-1.20	-1.59	0.61	-0.13	1.25	1.59
$\hat{r}_{\text{Model2}}^D$	1.34	0.75	-1.46	-1.57	0.75	-0.14	1.00	1.25
$\hat{r}_{\text{Model3}}^D$	0.18	0.56	-0.80	-0.63	1.29	-0.52	-0.12	0.32
$\hat{r}_{\text{Model4}}^D$	-0.17	0.42	-0.72	-0.56	1.03	-0.84	0.12	0.73

Second, the comparison between observed proportions and fitted values for different models are shown in the Table 2.

Table 2: Observed proportions and fitted values for four models: Model 1 refers to logit link; Model 2 refers to probit link; Model 3 refers to complementary log-log link, Model 4 refers to general link function

y_i/m_i	0.10	0.22	0.29	0.50	0.83	0.90	0.98	1.00
$\hat{\pi}_{\text{Model1}}$	0.06	0.16	0.36	0.61	0.80	0.90	0.96	0.98
$\hat{\pi}_{\text{Model2}}$	0.06	0.18	0.38	0.60	0.79	0.90	0.96	0.99
$\hat{\pi}_{\text{Model3}}$	0.09	0.19	0.34	0.54	0.76	0.92	0.99	1.00
$\hat{\pi}_{\text{Model4}}$	0.11	0.19	0.33	0.54	0.77	0.93	0.98	1.00

Third, estimated dose-response curves under each model are estimated in the interval (1.65, 1.9) and plotted with the observed x_i against the observed proportion in Figure 1. Those figures show the model with complementary log-log link is better than other two because the estimated curve is closer to observed points.

- (b) Figure 2 shows that as α increases the mean function is going to become a step function with the jump point 0, which suggests that the MLE is more likely close to 0 or 1. And the log-likelihood is derived as follows

$$\ell(\boldsymbol{\beta}, \alpha | \mathbf{y}) = \sum_{i=1}^n (y_i \log(\frac{\pi_i}{1 - \pi_i}) + m_i \log(1 - \pi_i) + \log \left[\frac{m_i}{y_i} \right]) \quad (2)$$

$$= \sum_{i=1}^n (y_i \log(\frac{\exp(\alpha \mathbf{X}\boldsymbol{\beta})}{(1 + \exp(\mathbf{X}\boldsymbol{\beta}))^\alpha - \exp^\alpha \mathbf{X}\boldsymbol{\beta}}) + m_i \log(\frac{(1 + \exp(\mathbf{X}\boldsymbol{\beta}))^\alpha - \exp^\alpha \mathbf{X}\boldsymbol{\beta}}{(1 + \exp(\mathbf{X}\boldsymbol{\beta}))^\alpha}) + \log \left[\frac{m_i}{y_i} \right]) \quad (3)$$

$$m_i \log(\frac{(1 + \exp(\mathbf{X}\boldsymbol{\beta}))^\alpha - \exp^\alpha \mathbf{X}\boldsymbol{\beta}}{(1 + \exp(\mathbf{X}\boldsymbol{\beta}))^\alpha}) + \log \left[\frac{m_i}{y_i} \right] \quad (4)$$

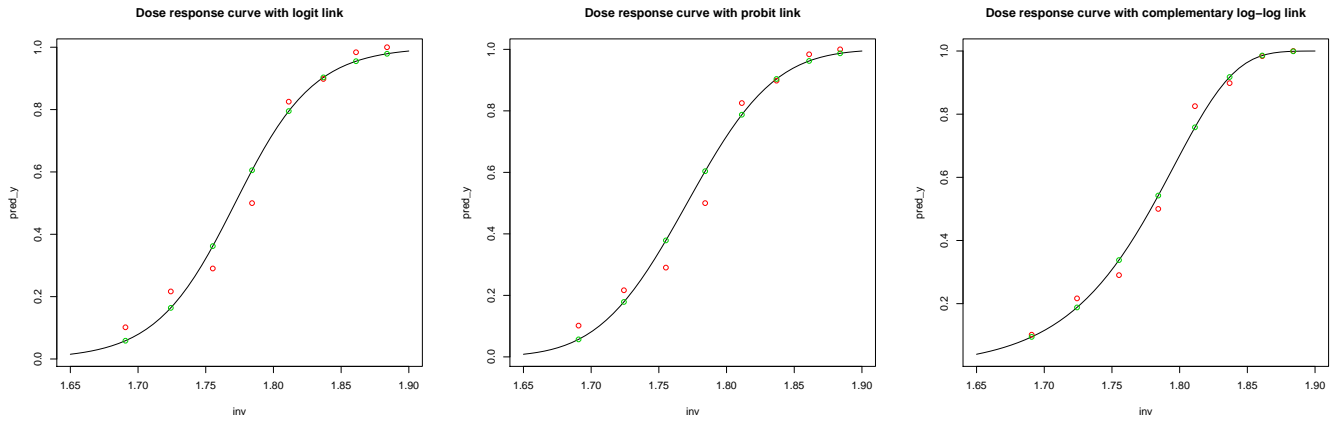


Figure 1: Three estimated dose-response curves compared with the observed x_i plotted against the observed proportions

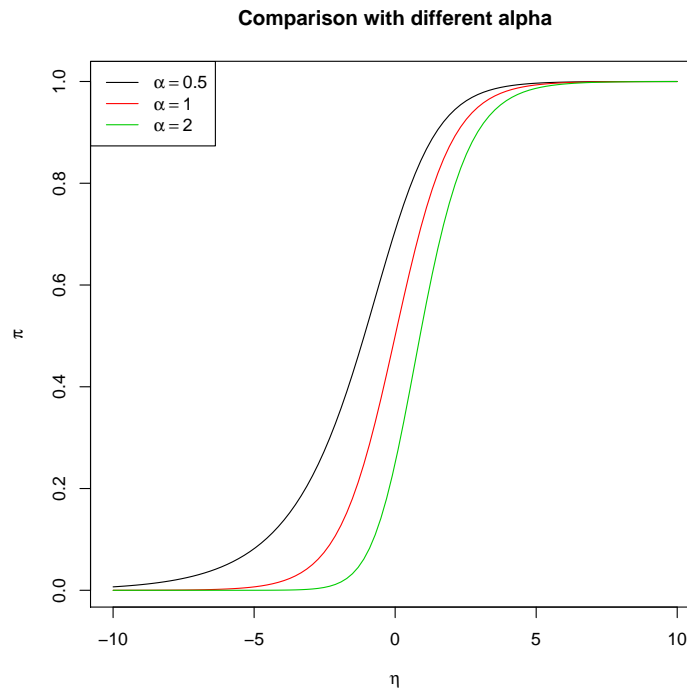


Figure 2: The link function with different α

Maximizing $\ell(\boldsymbol{\beta}, \alpha | \mathbf{y})$ is difficult, because of the complexity of $(1 + \exp(\mathbf{X}\boldsymbol{\beta}))^\alpha - \exp^{\alpha\mathbf{X}\boldsymbol{\beta}}$. Compared with the logit link function in which $\alpha = 1$, the complexity can be reduced by $(1 + \exp(\mathbf{X}\boldsymbol{\beta}))^\alpha - \exp^{\alpha\mathbf{X}\boldsymbol{\beta}} = 1$.

- (c) The fitted values are shown in Table 1 and the estimated dose-response curve are shown in Figure 3. From the deviance residuals, overall the deviance residuals are closer to 0 compared with the previous models. Also, the observed points are very close to the estimated curve, which illustrates the model with this general link function performs well.
- (d) The AIC and BIC are shown in Table 3. Table 3 shows that considering the com-

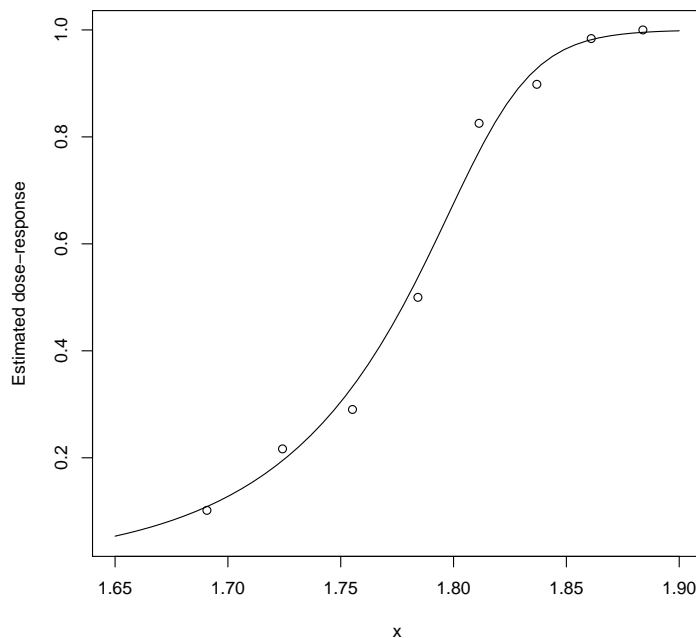


Figure 3: The dose-response curve with general link function

Table 3: AIC and BIC for four different models

Model	AIC	BIC
Model 1	41.43	41.59
Model 2	40.32	40.47
Model 3	33.64	33.80
Model 4	35.55	35.78

plexity of model, Model 3 performs better than others from both AIC and BIC respective.

2. (Ex 2) This problem involves Bayesian analysis of the beetle mortality data from the previous problem.

(a) Consider a Bayesian binomial GLM with a complementary log-log link, i.e., assume that, given β_1 and β_2 , the y_i are independent from $\text{Bin}(m_i, \pi(x_i))$, $i = 1, \dots, 8$, where

$$\pi(x) = \pi(x; \beta_1, \beta_2) = 1 - \exp\{-\exp(\beta_1 + \beta_2 x)\}.$$

Design and implement an MCMC method to sample from the posterior distribution of (β_1, β_2) . Study the effect of the prior for (β_1, β_2) , for example, consider a flat prior as well as (independent) normal priors. Under the flat prior, obtain the posterior distribution for the median lethal dose, LD_{50} , that is, the dose level at which the probability of response is 0.5. Finally, plot point and interval estimates for the dose-response curve $\pi(x)$ (over a grid of values x for log dose).

- (b) Next, consider a binomial GLM with a logit link, i.e., now the y_i are assumed independent, given β_1 and β_2 , from $\text{Bin}(m_i, \pi(x_i))$, $i = 1, \dots, 8$, where

$$\pi(x) \equiv \pi(x; \beta_1, \beta_2) = \exp(\beta_1 + \beta_2 x) / \{1 + \exp(\beta_1 + \beta_2 x)\}.$$

Working with a flat prior for (β_1, β_2) , obtain MCMC samples from the posterior distributions for β_1 , β_2 , and for LD_{50} , along with point and interval estimates for the dose-response curve $\pi(x)$.

- (c) As a third model, consider the binomial GLM with the parametric link given in 1. Develop an MCMC method to sample from the posterior distribution of $(\beta_1, \beta_2, \alpha)$, and obtain the posterior distribution for LD_{50} , and point and interval estimates for $\pi(x)$.
- (d) Use the results from parts (a), (b) and (c) for an empirical comparison of the three Bayesian binomial GLMs for the beetle mortality data. Moreover, perform a residual analysis for each model using the *Bayesian residuals*: $(y_i/m_i) - \pi(x_i; \beta_1, \beta_2)$ for the first two models, and $(y_i/m_i) - \pi(x_i; \beta_1, \beta_2, \alpha)$ for the third. Finally, use the *quadratic loss L measure* for formal comparison of the three models.

Sol 2 (a) As for a Bayesian binomial GLM with a complementary log-log link, the posterior distribution of parameters is

$$P(\beta_1, \beta_2 | \mathbf{x}, \mathbf{m}, \mathbf{y}) \propto P(\beta_1, \beta_2 | \mathbf{x}, \mathbf{m}) P(\mathbf{y} | \mathbf{x}, \mathbf{m}, \beta_1, \beta_2) \quad (5)$$

$$= P(\beta_1, \beta_2) P(\mathbf{y} | \mathbf{x}, \mathbf{m}, \beta_1, \beta_2) \quad (6)$$

$$= P(\beta_1, \beta_2) \prod_{i=1}^n \binom{m_i}{y_i} \pi(x_i)^{y_i} (1 - \pi(x_i))^{m_i - y_i} \quad (7)$$

$$= P(\beta_1, \beta_2) \prod_{i=1}^n (1 - \exp\{-\exp(\beta_1 + \beta_2 x_i)\})^{y_i} \quad (8)$$

$$(\exp\{-\exp(\beta_1 + \beta_2 x_i)\})^{m_i - y_i}. \quad (9)$$

To compare the different prior effects, we take the flat prior and normal prior for instance. Using Metropolis Hasting algorithm with both flat prior and standard normal prior, we obtain the posterior distributions of β by sampling 1000 times and choosing the burn-in $T = 100$. The posterior distributions are shown in Figure 4. Those figures show that the posterior distribution depends on the prior distributions heavily.

Because $\pi(x) = 1 - \exp\{-\exp(\beta_1 + \beta_2 x)\}$, the median lethal dose is $x_{0.5} = \frac{\log \log 2 - \beta_1}{\beta_2}$. Then thanks to calibration inference, the posterior distribution of median lethal dose with flat prior is shown in Figure 4. Using 1000 sampled β with the burn-in $T = 100$, the point and centered 95% confidence interval estimates for the dose-response curve $\pi(x)$ on the grid $[1.65, 1.9]$ are plotted in Figure 5.

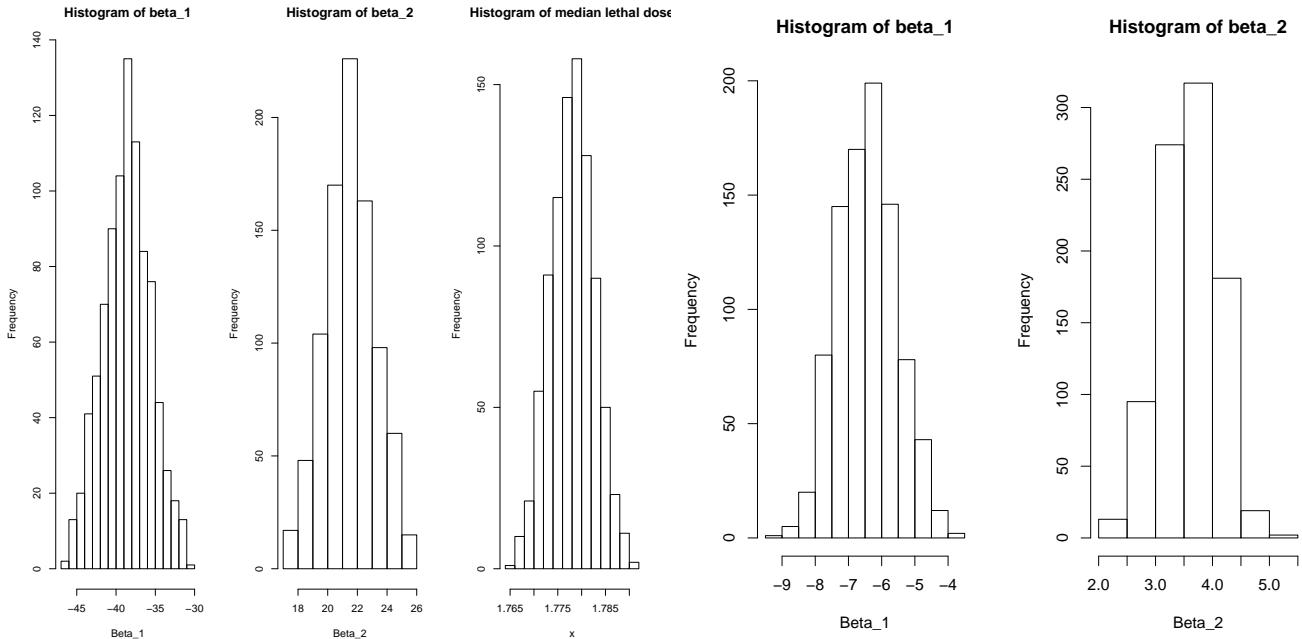


Figure 4: The posterior distributions for flat prior and standard normal prior with complementary log-log link function

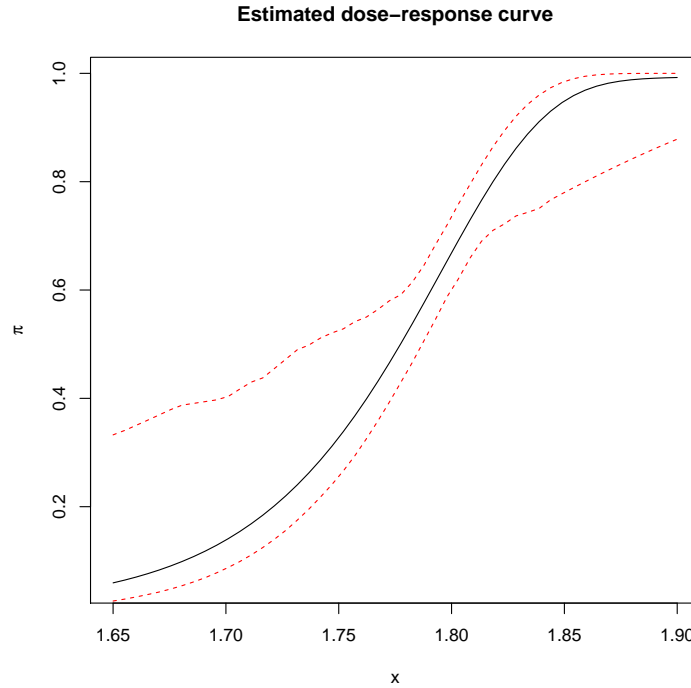


Figure 5: Estimated dose-response curve with its 95% interval

- (b) Also, with Metropolis Hasting algorithm with 1000 samples and setting burn-in as $T = 100$, the posterior distributions for β_1, β_2 and LD₅₀ and point and centered 95% interval estimates for the dose-response curve $\pi(x)$ are shown in Figure 6.
- (c) Also, with Metropolis Hasting algorithm with 1000 samples and setting burn-in as

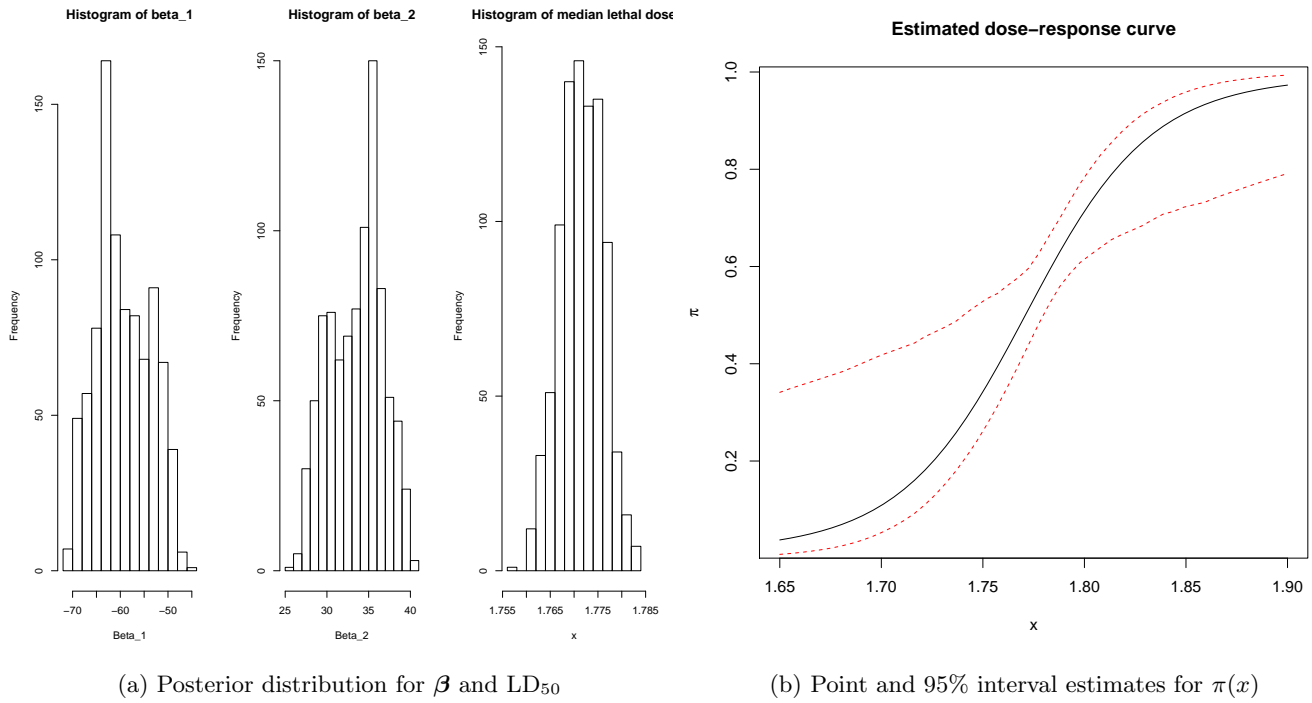


Figure 6: Summary of model with logistic link function

$T = 100$, the posterior distributions for α, β_1, β_2 and LD_{50} and point and centered 95% interval estimates for the dose-response curve $\pi(x)$ are shown in Figure 7.

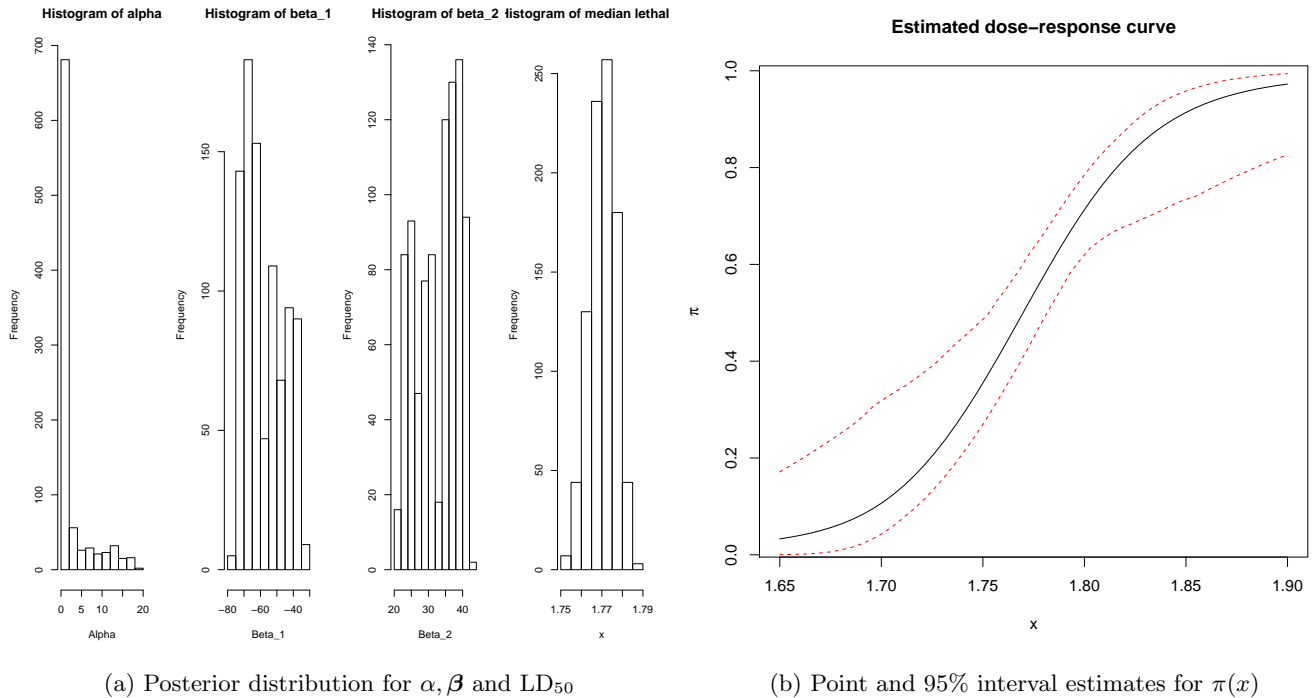


Figure 7: Summary of model with general link function

(d) From the estimated dose-response curves with respect to three models, the first and

Table 4: The Bayesian residuals and quadratic loss L measure for three models

x	1.69	1.72	1.76	1.78	1.81	1.84	1.86	1.88	QL_0	QL_1
res_1	-0.01	0.02	-0.05	-0.04	0.07	-0.01	0.01	0.01	0.022	0.032
res_2	0.02	0.03	-0.08	-0.10	0.04	0.01	0.04	0.03	0.029	0.053
res_3	0.02	0.01	-0.11	-0.12	0.05	0.03	0.06	0.04	0.024	0.060

last models have narrower interval estimates than the second one which suggests that the second model has relative higher variance in interval estimation than other two models.

From the residual aspect, we check it by two aspects. First, for each x_i , we sample the posterior $\pi(x_i; \tilde{\beta})$, where $\tilde{\beta}$ follow the posterior distribution $Pr(\beta|\mathbf{y}, \mathbf{m}, \mathbf{x})$. And taking the mean of $\pi(x_i; \tilde{\beta})$ as the estimated posterior predictive value denoted as $\hat{\pi}(x_i)$, Bayesian residual are estimated as $\hat{r}(x_i) = y_i/m_i - \hat{\pi}(x_i)$. The estimates for three model are in Table 4 Quadratic loss L measure is

$$L_q(v) = \sum_{i=1}^n var(\pi(x_i; \tilde{\beta})) + v \sum_{i=1}^n (\hat{\pi}(x_1) - y_i/m_i)^2, \quad (10)$$

where $v = k/k + 1$ is a tuning from $[0, 1)$. For each model, we show the quadratic loss L measure $v = 0$ and $v \rightarrow 1$ in Table 4, which suggests the first model has the smallest loss and so it performs better than others.