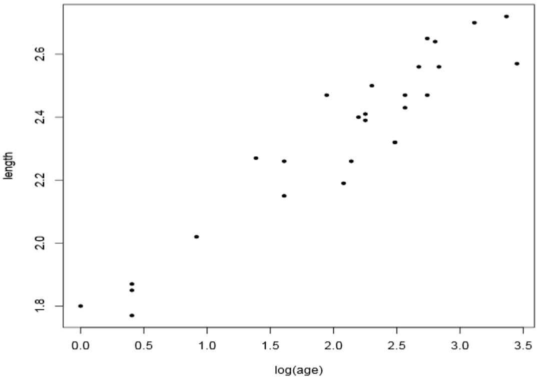
Bayesian Statistics
Note 6
BUGS Examples

Keunbaik Lee

Sungkyunkwan University

Example: Linear Regression I



Example: Linear Regression II

- For $n = 27$ captured samples of the sirenian species *dugong* (sea cow), relate an animal's length in meters, Y_i, to its age in years, x_i.
- To avoid a nonlinear model for now, transform x_i to the log scale.
- Simple linear regression in WinBUGS:

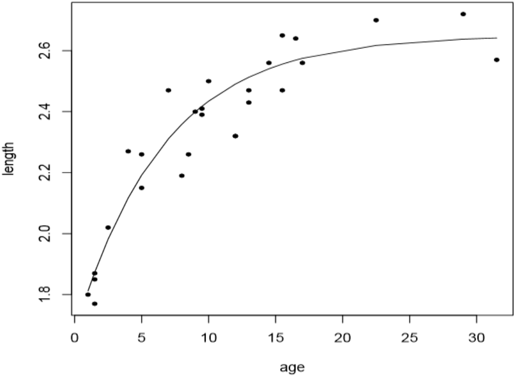
$$Y_i = \beta_0 + \beta_1 \log(x_i) + \epsilon_i, \quad i = 1, \cdots, n$$

where $\epsilon_i \sim^{iid} N(0, \sigma^2)$ and $\tau = 1/\sigma^2$ is the precision in the data.

Example: Linear Regression III

- Prior distributions:
  flat or β_0, β_1; vague gamma on τ (say, $Gamma(0.1, 0.1)$, which has mean 1 and variance 10) is traditional
- posterior correlation is reduced by centering the $\log(x_i)$ around their own mean
- Andrew Gelman suggests placing a uniform prior on σ, bounding the prior away from 0 and $\infty => U(.01, 100)$?
- WinBUGS Code: dugongs_BUGS.txt

Example: Nonlinear Regression I



Example: Nonlinear Regression II

- Model the untransformed dugong data as

$$Y_i = \alpha - \beta\gamma^{x_i} + \epsilon_i, \quad i = 1, \cdots, n$$

where $\alpha > 0$, $\beta > 0$, $0 \leq \gamma \leq 1$, and as usual $\epsilon_i \sim^{iid} N(0, \sigma^2)$ with $\tau = 1/\sigma^2 > 0$.

- In this model,
  - α corresponds to the average length of a fully grown dugong $(x \to \infty)$
  - $(\alpha - \beta)$ is the length of a dugong at birth $(x = 0)$
  - γ determines the growth rate: lower values produce an initially steep growth curve while higher values lead to gradual, almost linear growth.

Example: Nonlinear Regression III

- Prior distributions:
  flat for α and β, $U(.01, 100)$ for σ, and $U(0.5, 1.0)$ for γ (hard to estimate)
- Code: dugongsNL_BUGS.txt
- Obtain posterior density estimates and autocorrelation plots for α, β, γ, and σ, and investigate the bivariate posterior of (α, β) using the *Correlation* tool on the *inference* menu

Example: One-way ANOVA example I

- Wish to employ a new mathematics tutor.
- The ability of four candidates is examined using a small study.
- A group of 25 students was randomly divided into four classes.

Candidate	Students' grades
1	84 58 100 51 28 89
2	97 50 76 83 45 42 83
3	64 47 83 81 83 34 61
4	77 69 94 80 55 79

- For $i = 1, \cdots, 4$ and $j = 1, \cdots, n_i$,

$$y_{ij} = \mu + \alpha_i + \epsilon_{ij}$$

  where $\epsilon_{ij} \sim^{iid} N(0, \sigma^2)$ and $\sum_{i=1}^{a} \alpha_i = 0$.
- See WinBUGS code: 'chap05_ex2_tutors_evaluation.txt'.

Deviance Information Criterion I

- DIC scores for the three models:

$$DIC = \bar{D}(\theta) + p_D,$$

where $\bar{D}(\theta) = E\left[-2\log(p(y|\theta))|y\right]$, $D(\tilde{\theta}) = -2\log(p(y|\tilde{\theta}))$, and $p_D = \bar{D}(\theta) - D(\tilde{\theta})$ with $\tilde{\theta} =$ posterior estimate of θ.

Deviance Information Criterion II

- Properties of DIC
  - DIC is intended as a generalization of Akaike Information Criterion (AIC).
  - p_D is effective number of parameters.
  - Small DIC is better.
  - To compare two models, difference of DIC>10, definitely better; $5 < Diff. < 10$, substantially better; $Diff < 5$, no difference.

Example: Logistic Regression I

- Consider a binary version of the dugong data,

$$Z_i = \begin{cases} 1, & \text{if } Y_i > 2.4; \text{ (i.e., the dugong is ``full grown'')} \\ 0, & \text{otherwise.} \end{cases}$$

- A *logistic regression* model for $p_i = p(Z_i = 1)$ is then

$$\text{logit}(p_i) = \log\left[p_i/(1-p_i)\right] = \beta_0 + \beta_1 \log(x_i).$$

- Two other commonly used link functions are the *probit*,

$$\text{probit}(p_i) = \Phi^{-1}(p_i) = \beta_0 + \beta_1 \log(x_i),$$

and the *complementary log-log* (cloglog),

$$\text{cloglog}(p_i) = \log\left[-\log(1-p_i)\right] = \beta_0 + \beta_1 \log(x_i).$$

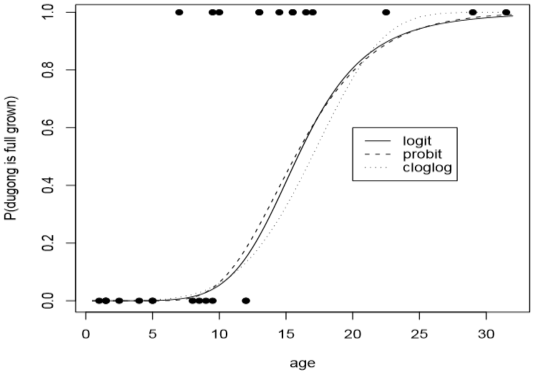
Example: Logistic Regression II

- Code: dugongsBin_BUGS.txt
- Code uses flat priors for β_0 and β_1, and the *phi* function, instead of the less stable *probit* function.

model	$\bar{D}$	p_D	DIC
logit	19.62	1.85	21.47
probit	19.30	1.87	21.17
cloglog	18.77	1.84	20.61

Example: Logistic Regression III

Figure: Fitted binary regression models



Example: Logistic Regression IV

In fact, these scores can be obtained from a single run; see the "trick version" at the bottom of the BUGS file.

- Use the *Comparison* tool to compare the posteriors of β_1 across models, and the Correlation tool to check the bivariate posteriors of (β_0, β_1) across models.
- The logit and probit fits appear very similar, but the cloglog fitted curve is slightly different.
- You can also compare p_i posterior boxplots (induced by the link function and the β_0 and β_1 posteriors) using the *Comparison* tool.

Example: Poisson hierarchical model I

- Consider 10 power plant pumps. The number of failures X_i is assumed to follow a Poisson distribution

$$X_i \sim Poisson(\theta_i t_i), \quad i = 1, \ldots, 10,$$

where θ_i is the failure rate for pump i and t_i is the length of operation time of the pump (in 1000s of hours).

Pump	1	2	3	4	5	6	7	8	9	10
t_i	94.3	15.7	62.9	126	5.24	31.4	1.05	1.05	2.1	10.5
x_i	5	1	5	14	3	19	1	1	4	22

Example: Poisson hierarchical model II

- A conjugate gamma prior distribution is adopted for the failure rates:

$$\theta_i \sim Gamma(\alpha, \beta), i = 1, \ldots, 10,$$
$$\alpha \sim Exponential(1.0),$$
$$\beta \sim Gamma(0.1, 1.0).$$

- See WinBUGS code 'Pump_Poisson.txt'.

Example: Multinomial model I

- Political Party data: Data takes from Agresti (2002), page 303, Table 7.15, problem 7.3.
- Table refers to the effect on political party identification of gender and race.

		Party Identification		
Gender	Race	Democrat	Republican	Independent
Male	White	132	176	127
	Black	42	6	12
Female	White	172	129	130
	Black	56	4	15

Example: Multinomial model II

- Let $Y_i = (Y_{i1}, \cdots, Y_{iK})$ be response with $K = 3$ levels where Y_{ik} denotes the frequency of the kth level. The multinomial logistic regression model is

$$
\begin{aligned}
&Y_i \sim \mathit{multinomial}(\pi_i, N_i), \\
&\log \frac{\pi_{ik}}{\pi_{i1}} = \beta_{0k} + \beta_{1k}\text{gender}_i + \beta_{2k}\text{race}_i
\end{aligned}
$$

- See WinBUGS code '01_multi.txt'.

Example: Economic data I

- The data come from the U.S. Department of Commerce, Survey of Current Business, and describe activity from the first quarter of 1979 to fourth quarter 1989.
- Six economic indicators are measured at 44 time points $x_1, \cdots, x_{44}$ (labeled $1, 2, \cdots, 44$).
- We model each indicator Y_{ij}, $i = 1, \cdots, 6$ and $j = 1, \cdots, 44$ as a function of (centered) time as follows:

$$
\begin{aligned}
& Y_{ij} \sim N(\beta_{0i} + \beta_{1i} x_j, \tau) \\
& \beta_{0i} \sim N(\mu_{\beta_0}, \tau_{\beta_0}), \quad \beta_{1i} \sim N(\mu_{\beta_1}, \tau_{\beta_1}), \quad \tau \sim gamma(0.01, 0.01) \\
& \mu_{\beta_0} \sim N(0, 0.01), \quad \mu_{\beta_1} \sim N(0, 0.01) \\
& \tau_{\beta_0} \sim gamma(0.01, 0.01), \quad \tau_{\beta_1} \sim gamma(0.01, 0.01)
\end{aligned}
$$

- See WinBUGS code 'Economic_BUGS.txt' for inference on β_{0i} and β_{1i}.

Example: Hierarchical model I

- Kobe Bryant's field goals in NBA

$$
\begin{aligned}
& Y_t \sim \mathit{binomial}(N_t, \pi_t), \\
& \text{logit}(\pi_t) = \log\left(\frac{\pi_t}{1-\pi_t}\right) = \theta_t, \\
& \theta_t \sim N(\mu_\theta, \sigma_\theta^2) \text{ for } t \in \{1999, 2000, \cdots, 2006\}, \\
& \mu_\theta \sim N(0, 100) \text{ and } \sigma_\theta^2 \sim \mathit{IGamma}(0.01, 0.01).
\end{aligned}
$$

- See WinBUGS code 'chap08_ex2_kobes1_hierarchical.txt'.