Bayesian BioStatistics Using BUGS

BUGS

Βιοστατιστική κατά Bayes με τη χρήση του Λογισμικού BUGS

Ioannis Ntzoufras

E-mail: ntzoufras@aegean.gr

Department of Business Administration, University of the Aegean

ΠΕΡΙΕΧΟΜΕΝΑ

- ΜΑΘΗΜΑ 1: ΕΙΣΑΓΩΓΗ ΣΤΟ BUGS
- ΜΑΘΗΜΑ 2: ΕΙΣΑΓΩΓΗ ΣΤΟ WINBUGS
- ΜΑΘΗΜΑ 3: ΣΥΝΘΕΤΑ ΠΑΡΑΔΕΙΓΜΑΤΑ
- ΜΑΘΗΜΑ 4: ΣΥΓΚΡΙΣΗ ΜΟΝΤΕΛΩΝ

Bayesian BioStatistics Using BUGS

BUGS

ΜΑΘΗΜΑ 1:
ΕΙΣΑΓΩΓΗ ΣΤΟ BUGS

Ioannis Ntzoufras

E-mail: ntzoufras@aegean.gr

Department of Business Administration, University of the Aegean

ΠΕΡΙΕΧΟΜΕΝΑ 1ου ΜΑΘΗΜΑΤΟΣ

- 1…ΕΙΣΑΓΩΓΗ ΣΤΑ ΓΕΝΙΚΕΥΜΕΝΑ ΓΡΑΜΜΙΚΑ ΜΟΝΤΕΛΑ
- 2… ΕΙΣΑΓΩΓΗ ΣΤΟΥΣ ΑΛΓΟΡΙΘΜΟΥΣ MCMC
- 3… ΜΠΕΥΖΙΑΝΗ ΣΥΜΠΕΡΑΣΜΑΤΟΛΟΓΙΑ ΜΕ ΤΗ ΧΡΗΣΗ ΤΟΥ BUGS

1… ΕΙΣΑΓΩΓΗ ΣΤΑ ΓΕΝΙΚΕΥΜΕΝΑ ΓΡΑΜΜΙΚΑ ΜΟΝΤΕΛΑ

- ΓΕΝΙΚΕΥΣΗ ΤΩΝ ΠΑΛΙΝΔΡΟΜΙΚΩΝ ΜΟΝΤΕΛΩΝ
- ΞΕΚΙΝΗΣΑΝ ΑΠΟ ΤΟΝ LEGENDRE (1805) ΚΑΙ ΤΟΝ GAUSS (1809)

1.1. ΔΕΔΟΜΕΝΑ

- RESPONSE VARIABLE (Y): εξαρτημένη ή ενδογενής μεταβλητή ή μεταβλητή απόκρισης/αντίδρασης
  - Y είναι τυχαία μεταβλητή
- EXPLANATORY VARIABLES (X_j): Ανεξάρτητες ή εξογενείς ή επεξηγηματικές μεταβλητές
  - X_j θεωρούνται συνήθως σταθερές από το σχεδιασμό του πειράματος

1.2. 3 ΚΥΡΙΕΣ ΣΥΝΙΣΤΩΣΕΣ

- ΤΥΧΑΙΑ ΣΥΝΙΣΤΩΣΑ (random component)
  - $Y_i \sim$ ΚΑΤΑΝΟΜΗ($\underline{\boldsymbol{\theta}}$)
  - $\underline{\boldsymbol{\theta}}$: διάνυσμα παραμέτρων
- ΣΥΣΤΗΜΑΤΙΚΗ ΣΥΝΙΣΤΩΣΑ (systematic component)
  - $\eta_i = \beta_0+\beta_1 X_{1i}+\ldots+ \beta_p X_{pi}$
  - η_i λέγεται γραμμικός προσδιορισμός του μοντέλου (linear predictor)

- ΣΥΝΔΕΤΙΚΗ ΣΥΝΑΡΤΗΣΗ (link function)
  - Ο τρόπος συνδεσης των παραμέτρων της τυχαίας συνιστώσας με το γραμμικό προσδιορισμό
  - $g(\underline{\boldsymbol{\theta}}) = \eta_i = \beta_0+\beta_1 X_{1i}+\ldots+ \beta_p X_{pi}$
  - συνήθως $\underline{\boldsymbol{\theta}}$ είναι ο μέσος της Υ

1.3. ΚΑΝΟΝΙΚΟ ΜΟΝΤΕΛΟ

- ΤΥΧΑΙΑ ΣΥΝΙΣΤΩΣΑ:
  - Υ ποσοτική μεταβλητή
  - $Y_i \sim$ NORMAL(μ_i , σ^2)
- ΣΥΣΤΗΜΑΤΙΚΗ ΣΥΝΙΣΤΩΣΑ:
  - X_j ποσοτικές ή κατηγορικές
- ΣΥΝΔΕΤΙΚΗ ΣΥΝΑΡΤΗΣΗ
  - $\mu_i = \eta_i = \beta_0+\beta_1 X_{1i}+\ldots+ \beta_p X_{pi}$
  - αρα συνδετική συνάρτηση: $g(\mu)=\mu$

1.4. ΜΟΝΤΕΛΑ BERNOULLI

- ΤΥΧΑΙΑ ΣΥΝΙΣΤΩΣΑ:
  - Υ δίτιμη μεταβλητή (0/1)
  - $Y_i \sim$ Bernoulli (p_i)
- ΣΥΣΤΗΜΑΤΙΚΗ ΣΥΝΙΣΤΩΣΑ:
  - X_j ποσοτικές ή κατηγορικές
- ΣΥΝΔΕΤΙΚΗ ΣΥΝΑΡΤΗΣΗ
  - $\log (p_i/(1- p_i)) = \eta_i = \beta_0+\beta_1 X_{1i}+\ldots+ \beta_p X_{pi}$
  - αρα συνδετική συνάρτηση: $g(p)=\text{logit}(p)$

1.5. ΔΙΩΝΥΜΙΚΑ ΜΟΝΤΕΛΑ

- ΤΥΧΑΙΑ ΣΥΝΙΣΤΩΣΑ:
  - Υ: # επιτυχιών σε σύνολο n επαναλήψεων
  - $Y_i \sim$ Binomial (p_i, n_i)
- ΣΥΣΤΗΜΑΤΙΚΗ ΣΥΝΙΣΤΩΣΑ:
  - X_j ποσοτικές ή κατηγορικές
- ΣΥΝΔΕΤΙΚΗ ΣΥΝΑΡΤΗΣΗ
  - $\log (p_i/(1- p_i)) = \eta_i = \beta_0+\beta_1 X_{1i}+\ldots+ \beta_p X_{pi}$
  - αρα συνδετική συνάρτηση: $g(p)=\text{logit}(p)$

1.6. ΜΟΝΤΕΛΑ POISSON

- ΤΥΧΑΙΑ ΣΥΝΙΣΤΩΣΑ:
  - Υ: # επιτυχιών σε ένα χρονικό διάστημα
  - $Y_i \sim$ Poisson (λ_i)
- ΣΥΣΤΗΜΑΤΙΚΗ ΣΥΝΙΣΤΩΣΑ:
  - X_j ποσοτικές ή κατηγορικές
- ΣΥΝΔΕΤΙΚΗ ΣΥΝΑΡΤΗΣΗ
  - $\log (\lambda_i) = \eta_i = \beta_0+\beta_1 X_{1i}+\ldots+ \beta_p X_{pi}$
  - αρα συνδετική συνάρτηση: $g(\lambda)=\log(\lambda)$

1.7. ΓΕΝΙΚΕΣ ΑΡΧΕΣ ΜΟΝΤΕΛΟΠΟΙΗΣΗΣ

- ΕΙΝΑΙ ΤΕΧΝΗ
- ΟΛΑ ΤΑ ΜΟΝΤΕΛΑ ΕΙΝΑΙ ΛΑΘΟΣ
  - ΜΕΡΙΚΑ ΠΙΟ ΧΡΗΣΙΜΑ
  - ΨΑΧΝΟΥΜΕ ΕΝΑ ΜΟΝΤΕΛΟ ΠΟΥ ΠΕΡΙΓΡΑΦΕΙ ΠΕΡΙΛΗΠΤΙΚΑ ΤΗΝ ΠΡΑΓΜΑΤΙΚΟΤΗΤΑ
- ΕΛΕΓΧΟΥΜΕ ΠΟΛΛΑ ΔΙΑΦΟΡΕΤΙΚΑ ΜΟΝΤΕΛΑ
- ΠΡΕΠΕΙ ΝΑ ΓΙΝΕΤΑΙ ΕΛΕΓΧΟΣ ΚΑΛΗΣ ΠΡΟΣΑΡΜΟΓΗΣ

1.8. ΜΕΡΙΚΑ ΙΣΤΟΡΙΚΑ ΣΤΟΙΧΕΙΑ

- **1805**: LEGENDRE:
  - ορίζει τα κατάλοιπα
  - Εκτιμάει τις παραμέτρους παλινδρόμησης μέσω των ελαχίστων τετραγώνων
- **1809**: GAUSS: Εισάγει την Κανονική κατανομή

1.8. ΜΕΡΙΚΑ ΙΣΤΟΡΙΚΑ ΣΤΟΙΧΕΙΑ

- **1805**: LEGENDRE:
  - ορίζει τα κατάλοιπα
  - Εκτιμάει τις παραμέτρους παλινδρόμησης μέσω των ελαχίστων τετραγώνων
- **1823**: GAUSS (Theoria Combinationis)
  - Εγκαταλείπει την κανονικότητα
  - Η ομοσκεδαστικοτητα διατηρεί κάποιες καλές ιδιότητες
  - Wedderburn (1974, biometrika) απέδειξε το ίδιο για τα GLM

1.8. ΜΕΡΙΚΑ ΙΣΤΟΡΙΚΑ ΣΤΟΙΧΕΙΑ

- **1919-1929**: Fisher διατυπώνει τη θεωρία "Πειραματικού σχεδιασμού" (Experimental Design).
- **1922**: Ο Έλεγχος Περιεκτικότητας του Fisher και το 1ο γενικευμένο γραμμικό μοντέλο (RSS)
- **1935**: Η Μελέτη βιολογική περιεκτικότητας του Bliss. Τα μοντέλα Probit (Ann.Appl.Biol.).
- **1944 & 1951**: Ο Berkston εισάγει τα μοντέλα logit (JASA & Biometrics)
- **1952** (biometrics) Dyke & Patterson εφαρμογή λογιστικής παλινδρόμησης σε κατηγορικά δεδομένα με 4 μεταβλητές

Bayesian Biostatistics Using BUGS

2...ΕΙΣΑΓΩΓΗ ΣΤΟΥΣ ΑΛΓΟΡΙΘΜΟΥΣ MCMC

Department of Business Administration, University of the Aegean

2...ΕΙΣΑΓΩΓΗ ΣΤΟΥΣ ΑΛΓΟΡΙΘΜΟΥΣ MCMC

- Η ΜΠΕΥΖΙΑΝΗ ΠΡΟΣΕΓΓΙΣΗ
- ΠΡΟΣΟΜΟΙΩΣΗ ΑΠΟ ΤΗΝ ΕΚ-ΤΩΝ-ΥΣΤΕΡΩΝ (posterior) ΚΑΤΑΝΟΜΗ
  - Ο Αλγόριθμος Metropolis-Hastings
  - Ο Δειγματολήπτης Gibbs
- ΠΑΡΑΔΕΙΓΜΑ ΑΠΛΗΣ ΠΑΛΙΝΔΡΟΜΙΣΗΣ

2.1. Η Μπευζιανή προσέγγιση

- Κλασσική στατιστική:
  - βασίζεται στην πιθανοφάνεια $f(\underline{\mathbf{y}}|\underline{\boldsymbol{\theta}})$
  - $\underline{\boldsymbol{\theta}}$ διάνυσμα παραμέτρων=> άγνωστες προς εκτίμηση ποσότητες
  - Εκτίμηση γίνεται μέσω εκτιμητριών συναρτήσεων με κάποιες καλές ιδιότητες (π.χ. μεροληψία)
  - Εκτιμήτριες βρίσκονται μεγιστοποιώντας την πιθανοφάνεια
  - ΠΑΡΑΔΕΙΓΜΑ: για $Y_i \sim N(\mu,\sigma^2)$
    - μ εκτιμάται από το δειγματικό μέσο $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$

- Στατιστική κατά Μπέυες:
  - Θεωρεί τις παραμέτρους τυχαίες μεταβλητές
  - Ορίζει εκ-των-προτέρων (prior) κατανομών $f(\underline{\boldsymbol{\theta}})$
  - βασίζεται στην εκ-των-υστερων (posterior) κατανομή $f(\underline{\boldsymbol{\theta}}|\underline{\mathbf{y}})$

- Πλεονεκτήματα
  - Πιο καθάρη και πιθανοθεωρητική προσέγγιση
  - Μπορεί να συμπεριλάβει πληροφορία ειδικών ή αποτελέσματα άλλων μελετών (meta-analysis)
- Μειονεκτήματα
  - Υποκειμενικότητα αποτελεσμάτων λόγω της prior
  - Δυσκολίες υπολογισμου των posterior Κατανομών

- Η Εκ-των-υστέρων κατανομή υπολογίζεται από το Θεώρημα του Bayes
- $f(\underline{\boldsymbol{\theta}}|\underline{\mathbf{y}}) = f(\underline{\boldsymbol{\theta}},\underline{\mathbf{y}}) / f(\underline{\mathbf{y}})$
- $\quad = f(\underline{\mathbf{y}}|\underline{\boldsymbol{\theta}})f(\underline{\boldsymbol{\theta}}) / f(\underline{\mathbf{y}})$
- $\quad \propto f(\underline{\mathbf{y}}|\underline{\boldsymbol{\theta}})f(\underline{\boldsymbol{\theta}})$
- Posterior = Likelihood x Prior

ΠΑΡΑΔΕΙΓΜΑ: ΕΚΤΙΜΗΣΗ ΜΕΣΟΥ

- *ΠΙΘΑΝΟΦΑΝΕΙΑ*: $Y_i \sim N(\mu, \sigma^2)$
  - σ^2 γνωστή σταθερή ποσότητα
- *PRIOR*: $\mu \sim N(\mu_0, \tau^2)$
- *POSTERIOR*:

  $f(\mu|\underline{\mathbf{y}}) = N(w\,\bar{y} + (1-w)\,\mu_0,\ w\,\sigma^2/n)$

  $w = \tau^2/(\tau^2 + \sigma^2/n)$

2.2. ΠΡΟΣΟΜΟΙΩΣΗ ΑΠΟ ΤΗΝ ΕΚ-ΤΩΝ-ΥΣΤΕΡΩΝ ΚΑΤΑΝΟΜΗ

- Υπολογισμός της posterior Κατανομής είναι δύσκολος.
  - Συζυγείς εκ-των-προτέρων κατανομές (conjugate priors, 70s)
  - Ασυμπτωτικές Προσεγγίσεις (80ς)
  - Προσομοίωση μέσω MCMC (90ς)

2.2. ΠΡΟΣΟΜΟΙΩΣΗ ΑΠΟ ΤΗΝ ΕΚ-ΤΩΝ-ΥΣΤΕΡΩΝ ΚΑΤΑΝΟΜΗ

- MCMC
  - Προϋπήρχαν σε άλλες επιστήμες
  - 1954 Metropolis etal. (Metropolis Algorithm)
  - 1970 Hastings (Metropolis-Hastings Algorithm)
  - 1984 Geman and Geman (Gibbs Sampling)
  - 1990 Smith etal (Εφαρμογή των MCMC σε Μπευζιανά Προβλήματα)
  - 1995 Green (Reversible Jump MCMC)

2.2. ΠΡΟΣΟΜΟΙΩΣΗ ΑΠΟ ΤΗΝ ΕΚ-ΤΩΝ-ΥΣΤΕΡΩΝ ΚΑΤΑΝΟΜΗ

Η ΙΔΕΑ:

ΑΦΟΥ ΔΕΝ ΜΠΟΡΟΥΜΕ ΝΑ ΥΠΟΛΟΓΙΣΟΥΜΕ ΤΗΝ ΕΚ-ΤΩΝ-ΥΣΤΕΡΩΝ ΚΑΤΑΝΟΜΗ ΤΟΤΕ ΝΑ ΒΡΟΥΜΕ ΕΝΑ ΕΝΑΝ ΑΛΓΟΡΙΘΜΟ ΚΑΙ ΝΑ ΠΡΟΣΟΜΟΙΩΣΟΥΜΕ ΤΥΧΑΙΕΣ ΤΙΜΕΣ ΑΠΟ ΑΥΤΗ

2.2. ΠΡΟΣΟΜΟΙΩΣΗ ΑΠΟ ΤΗΝ ΕΚ-ΤΩΝ-ΥΣΤΕΡΩΝ ΚΑΤΑΝΟΜΗ

Η ΛΟΓΙΚΗ:

ΦΤΙΑΧΝΟΥΜΕ ΜΙΑ ΜΑΡΚΟΒΙΑΝΗ ΑΛΥΣΙΔΑ ΤΗΣ ΟΠΟΙΑΣ Η ΣΤΑΣΙΜΗ ΚΑΤΑΝΟΜΗ ΕΙΝΑΙ ΙΣΗ ΜΕ ΤΗ ΖΗΤΟΥΜΕΝΗ (ΕΚ-ΤΩΝ-ΥΣΤΕΡΩΝ) ΚΑΤΑΝΟΜΗ.

ΚΑΘΕ ΒΗΜΑ ΕΞΑΡΤΑΤΑΙ ΜΟΝΟ ΑΠΟ ΤΟ ΠΡΟΗΓΟΥΜΕΝΟ

ΧΡΗΣΙΜΟΠΟΙΟΥΜΕ ΑΥΤΗ ΤΗΝ ΑΛΥΣΙΔΑ ΓΙΑ ΝΑ "ΚΑΤΑΣΚΕΥΑΣΟΥΜΕ" ΕΝΑ ΔΕΙΓΜΑ ΑΠΟ ΤΗ ΣΤΑΣΙΜΗ ΚΑΤΑΝΟΜΗ

2.2. ΠΡΟΣΟΜΟΙΩΣΗ ΑΠΟ ΤΗΝ ΕΚ-ΤΩΝ-ΥΣΤΕΡΩΝ ΚΑΤΑΝΟΜΗ

- **Η ΔΙΑΔΙΚΑΣΙΑ**
  - ΟΡΙΖΟΥΜΕ ΑΥΘΑΙΡΕΤΑ ΚΑΠΟΙΕΣ ΑΡΧΙΚΕΣ ΤΙΜΕΣ ΓΙΑ ΤΙΣ ΠΑΡΑΜΕΤΡΟΥΣ ΜΑΣ $\underline{\boldsymbol{\theta}}^{(0)}$
  - ΓΙΑ t=1,..., T ΓΕΝΝΑΜΕ ΤΙΜΕΣ ΣΥΜΦΩΝΑ ΜΕ ΤΟΝ ΑΛΓΟΡΙΘΜΟ ΜΑΣ
  - ΟΤΑΝ ΒΕΒΑΙΩΘΟΥΜΕ ΟΤΙ Η ΑΛΥΣΙΔΑ ΓΕΝΝΑΕΙ ΤΙΜΕΣ ΑΠΟ ΤΗ ΣΤΑΣΙΜΗ ΔΙΑΔΙΚΑΣΙΑ (ΔΗΛΑΔΗ ΕΧΕΙ **ΣΥΓΚΛΙΝΕΙ**) ΣΤΑΜΑΤΑΜΕ
  - ΠΕΤΑΜΕ ΤΙΣ ΠΡΩΤΕΣ Κ ΤΙΜΕΣ ΓΙΑ ΝΑ ΑΠΟΦΥΓΟΥΜΕ ΤΥΧΟΝ ΕΠΙΡΡΟΗ ΤΟΥ ΑΡΧΙΚΟΥ ΣΗΜΕΙΟΥ.

2.2. ΠΡΟΣΟΜΟΙΩΣΗ ΑΠΟ ΤΗΝ ΕΚ-ΤΩΝ-ΥΣΤΕΡΩΝ ΚΑΤΑΝΟΜΗ

- **ΟΡΟΛΟΓΙΑ**
  - **INITIAL VALUES (ΑΡΧΙΚΕΣ ΤΙΜΕΣ)** ΤΗΣ ΑΛΥΣΙΔΑΣ ΛΕΓΟΝΤΑΙ ΟΙ ΤΙΜΕΣ $\underline{\boldsymbol{\theta}}^{(0)}$ ΑΠΟ ΟΠΟΥ ΞΕΚΙΝΑΜΕ ΤΟΝ ΑΛΓΟΡΙΘΜΟ ΜΑΣ
  - **ITERATION (ΕΠΑΝΑΛΗΨΗ)**: ΚΑΘΕ ΠΡΟΣΟΜΟΙΩΜΕΝΗ ΠΑΡΑΤΗΡΗΣΗ ΑΝΤΙΣΤΟΙΧΕΙ ΣΕ ΜΙΑ ΕΠΑΝΑΛΗΨΗ ΤΟΥ ΑΛΓΟΡΙΘΜΟΥ
  - **BURN-IN PERIOD**: ΟΙ ΠΕΡΙΟΔΟΣ ΜΕΧΡΙ ΝΑ ΑΡΧΙΣΟΥΜΕ ΝΑ ΠΕΡΝΟΥΜΕ ΤΙΜΕΣ ΑΠΟ ΤΗ ΣΤΑΣΙΜΗ ΚΑΤΑΝΟΜΗ

2.2. ΠΡΟΣΟΜΟΙΩΣΗ ΑΠΟ ΤΗΝ ΕΚ-ΤΩΝ-ΥΣΤΕΡΩΝ ΚΑΤΑΝΟΜΗ

- **ΟΡΟΛΟΓΙΑ**
  - **CONVERGENCE (ΣΥΓΚΛΙΣΗ)**: ΟΤΑΝ Η ΔΙΑΔΙΚΑΣΙΑ ΜΑΣ ΕΧΕΙ ΔΩΣΕΙ ΤΙΜΕΣ ΑΠΟ ΤΗ ΣΤΑΣΙΜΗ ΚΑΤΑΝΟΜΗ
  - **CONVERGENCE DIAGNOSTICS:** ΔΙΑΓΝΩΣΤΙΚΟΙ ΕΛΕΓΧΟΙ ΣΥΓΛΙΣΗΣ
  - **EQUILIBRIUM**: ΣΤΑΣΙΜΗ ΚΑΤΑΝΟΜΗ
  - **MCMC OUTPUT**: ΤΟ ΠΡΟΣΟΜΟΙΩΜΕΝΟ ΔΕΙΓΜΑ

2.2. ΠΡΟΣΟΜΟΙΩΣΗ ΑΠΟ ΤΗΝ ΕΚ-ΤΩΝ-ΥΣΤΕΡΩΝ ΚΑΤΑΝΟΜΗ

- **ΟΡΟΛΟΓΙΑ**
  - ΕΠΕΙΔΗ ΟΙ ΑΛΓΟΡΙΘΜΟΙ MCMC ΒΑΣΙΖΟΝΤΑΙ ΣΕ ΜΑΡΚΟΒΙΑΝΕΣ ΑΛΥΣΙΔΕΣ, ΤΟ ΠΡΟΣΟΜΟΙΩΜΕΝΟ ΔΕΙΓΜΑ ΔΕΝ ΕΙΝΑΙ I.I.D.
  - ΕΞΑΛΕΙΨΗ ΑΥΤΟΣΥΣΧΕΤΙΣΗΣ
    - ΒΛΕΠΟΥΜΕ ΤΙΣ ΑΥΤΟΣΥΣΧΕΤΙΣΕΙΣ (AUTOCORRELATIONS)
    - ΚΡΑΤΑΜΕ 1 ΠΑΡΑΤΗΡΗΣΗ ΑΝΑ L ΕΠΑΝΑΛΗΨΕΙΣ.
  - Η ΠΟΣΟΤΗΤΑ L ΛΕΓΕΤΑΙ **THIN (ΛΕΠΤΥΝΣΗ)** ΤΗΣ ΑΛΥΣΙΔΑΣ
  - ΛΕΠΤΥΝΣΗ ΕΝΟΣ ΠΡΟΣΟΜΟΙΩΜΕΝΟΥ ΔΕΙΓΜΑΤΟΣ ΜΠΟΡΕΙ ΝΑ ΓΙΝΕΙ ΚΑΙ ΛΟΓΩ ΑΝΑΛΥΣΗΣ Η ΑΠΟΘΗΚΕΥΤΙΚΟΥ ΧΩΡΟΥ

2.2. ΠΡΟΣΟΜΟΙΩΣΗ ΑΠΟ ΤΗΝ ΕΚ-ΤΩΝ-ΥΣΤΕΡΩΝ ΚΑΤΑΝΟΜΗ

- **ΑΛΓΟΡΙΘΜΟΙ**
  - **ΑΛΓΟΡΙΘΜΟΣ METROPOLIS-HASTINGS**
  - **ΔΕΙΓΜΑΤΟΛΗΠΤΗΣ GIBBS**

2.2.1. ΑΛΓΟΡΙΘΜΟΣ METROPOLIS-HASTINGS

Έστω $\underline{\boldsymbol{\theta}}^{\mathbf{old}}$ η τρέχουσα τιμή των παραμέτρων

- Δειγματοληπτούμε $\underline{\boldsymbol{\theta}}^{\mathbf{can}}$ από μία *κατανομή εισήγησης (proposal distribution)* $q(\underline{\boldsymbol{\theta}}^{\mathbf{can}}|\underline{\boldsymbol{\theta}}^{\mathbf{old}})$.
- Υπολογίζουμε $a = \min\left\{1, \dfrac{f(\underline{\boldsymbol{\theta}}^{can}\,|\,\underline{\mathbf{y}})q(\underline{\boldsymbol{\theta}}^{old}\,|\,\underline{\boldsymbol{\theta}}^{can})}{f(\underline{\boldsymbol{\theta}}^{old}\,|\,\underline{\mathbf{y}})q(\underline{\boldsymbol{\theta}}^{can}\,|\,\underline{\boldsymbol{\theta}}^{old})}\right\}$
- Θέτουμε $\underline{\boldsymbol{\theta}}^{\mathbf{new}} = \underline{\boldsymbol{\theta}}^{\mathbf{can}}$ με πιθανότητα α και $\underline{\boldsymbol{\theta}}^{\mathbf{new}} = \underline{\boldsymbol{\theta}}^{\mathbf{old}}$ με πιθανότητα (1-α)

2.2.1. ΑΛΓΟΡΙΘΜΟΣ METROPOLIS-HASTINGS

Συνήθως

- $q(\underline{\boldsymbol{\theta}}^{\mathbf{can}}|\, \underline{\boldsymbol{\theta}}^{\mathbf{old}}) = N(\underline{\boldsymbol{\theta}}^{\mathbf{old}}, c^2)$.
- c^2 λέγεται και tuning parameter και επηρεάζει τη σύγκλιση. Επιλέγεται έτσι ωστε να δεχόμαστε 30-40%
- Η πιθανότητα αποδοχής α απλοποιείται

$$a = \min\left\{1, \frac{f(\underline{\boldsymbol{\theta}}^{can}\,|\,\underline{\mathbf{y}})}{f(\underline{\boldsymbol{\theta}}^{old}\,|\,\underline{\mathbf{y}})}\right\}$$

2.2.2. ΔΕΙΓΜΑΤΟΛΗΠΤΗΣ GIBBS

Έστω $\underline{\boldsymbol{\theta}}^{old}$ η τρέχουσα τιμή των παραμέτρων και $\underline{\boldsymbol{\theta}}^{old} = (\theta_1^{\mathbf{old}}, \ldots, \theta_p^{\mathbf{old}})^{\mathbf{T}}$

- $\theta_1^{new} \sim f(\theta_1|\theta_2^{old}, \ldots, \theta_p^{old}, \underline{\mathbf{y}})$
- …
- $\theta_j^{new} \sim f(\theta_j|\theta_1^{new}, \ldots, \theta_{j-1}^{new}, \theta_{j+1}^{old}, \ldots, \theta_p^{old}, \underline{\mathbf{y}})$
- …
- $\theta_p^{new} \sim f(\theta_j|\theta_1^{new}, \ldots, \theta_{p-1}^{new}, \underline{\mathbf{y}})$

2.2.2. ΔΕΙΓΜΑΤΟΛΗΠΤΗΣ GIBBS

- $f(\theta_j|\theta_1^{new}, \ldots, \theta_{j-1}^{new}, \theta_{j+1}^{old}, \ldots, \theta_p^{old}, \underline{\mathbf{y}})$ ονομαζέται full conditional posterior distribution και συμβολίζεται $f(\theta_j|\bullet)$

2.2.2. ΔΕΙΓΜΑΤΟΛΗΠΤΗΣ GIBBS

- Διαφορές με M-H
  - $\underline{\boldsymbol{\theta}}^{old} \neq \underline{\boldsymbol{\theta}}^{old}$ σε κάθε επανάληψη
  - Gibbs υποπεριπτωση M-H για $q()=f(\theta_j|\bullet)$
  - Κάθε φορά ανανεώνουμε μία-μία τις τιμές
  - $f(\theta_j|\bullet)$ μπορεί να είναι άγνωστη
    - => Χρήση adaptive rejection sampling για log-convave κατανομές (Gilks & Wild, 1992)
    - Στα GLM οι posterior είναι log-concave (Dellaportas & Smith, 1993)
    - Χρησιμοποιείται στο BUGS

2.3 ΠΑΡΑΔΕΙΓΜΑ: GIBBS SAMPLING ΣΤΗΝ ΑΠΛΗ ΠΑΛΙΝΔΡΟΜΗΣΗ

- $Y_i \sim N(\mu_i, \sigma^2)$ για $i=1,2,\dots,n$
- $\mu_i = \alpha + \beta X_i$
- $\underline{\boldsymbol{\theta}}=(\alpha,\beta,\sigma^2)^T$
- PRIORS: $f(\underline{\boldsymbol{\theta}})=f(\alpha,\beta,\sigma^2)= f(\alpha)f(\beta)f(\sigma^2)$
  - $f(\alpha)$ = Normal(μ_α , τ_α^2)
  - $f(\beta)$ = Normal(μ_β , τ_β^2)
  - $f(\sigma^2)$=Inverse Gamma(γ, δ)
  - αν $\tau = \sigma^{-2}$ τότε $f(\tau)$=Gamma(γ, δ)

2.3 ΠΑΡΑΔΕΙΓΜΑ: GIBBS SAMPLING ΣΤΗΝ ΑΠΛΗ ΠΑΛΙΝΔΡΟΜΗΣΗ

- Full Conditional Posteriors
  - $f(\alpha|\beta, \sigma^2,\underline{y})= N\left(w_1(\bar{y}-b\bar{x})+(1-w_1)\mu_\alpha , w_1\frac{\sigma^2}{n}\right)$
    - $w_1= \tau^2/(\tau^2+\sigma^2/n)$
  - $f(\beta|\alpha, \sigma^2,\underline{y})= N\left(w_2\frac{\sum_{i=1}^n x_i y_i - an\bar{x}}{\sum_{i=1}^n x_i^2}+(1-w_2)\mu_\beta , w_2\frac{\sigma^2}{\sum_{i=1}^n x_i^2}\right)$
    - $w_2= \tau^2/(\tau^2 +\sigma^2/ \Sigma x_i^2)$

2.3 ΠΑΡΑΔΕΙΓΜΑ: GIBBS SAMPLING ΣΤΗΝ ΑΠΛΗ ΠΑΛΙΝΔΡΟΜΗΣΗ

- Full Conditional Posteriors
  - $f(\sigma^2|\alpha, \beta,\underline{y})=IG(\gamma+n/2 , \delta+ \Sigma(y_i-\alpha-\beta x_i)^2/2)$

Bayesian Biostatistics Using BUGS

3... Μπεϋζιανή Συμπερασματολογία με τη Χρήση του BUGS

Department of Business Administration, University of the Aegean

3.1. Εισαγωγή: Τι είναι το BUGS

- BUGS: Bayesian inference Using Gibbs Sampling
- Γλώσσα προγραμματισμού που ορίζουμε μοντέλο (πιθανοφάνεια, prior)
- Υπολογίζει τις full conditional και προσομοιώνει από log-convave posterior κατανομές

3.1. Εισαγωγή: Τι είναι το BUGS

- Ξεκίνησε γύρω στο 1995
- Γύρω στο 1998 βγήκε η έκδοση για Windows (Winbugs)
- Οφείλεται σε μια ομάδα του MRC στο Cambridge (Spiegelhalter, Gilks, Best, Thomas)

3.1. Εισαγωγή: Τι είναι το BUGS

ΠΡΟΣΟΧΗ:

Ο ΔΕΙΓΜΑΤΟΛΗΠΤΗΣ GIBBS ΜΠΟΡΕΙ ΝΑ ΕΙΝΑΙ ΕΠΙΚΙΝΔΥΝΟΣ

ΓΙΑΤΙ;

3.1. Εισαγωγή: Τι είναι το BUGS

ΓΙΑΤΙ;

- ΛΑΘΟΣ ΑΠΟΤΕΛΕΣΜΑΤΑ ΛΟΓΩ ΜΗ ΣΥΓΚΛΙΣΗΣ
- ΠΟΛΥ ΑΡΓΗ ΣΥΓΚΛΙΣΗ
- ΚΑΚΕΣ ΑΡΧΙΚΕΣ ΤΙΜΕΣ
- ΚΑΚΗ ΕΠΙΛΟΓΗ ΜΟΝΤΕΛΟΥ
- ΥΠΕΡ-ΠΑΡΑΜΕΤΡΟΠΟΙΗΣΗ (over-parametrazation)
- ΣΠΑΣΙΜΟ ΝΕΥΡΩΝ

3.1. Εισαγωγή: Τι είναι το BUGS

ΓΙΑΤΙ;

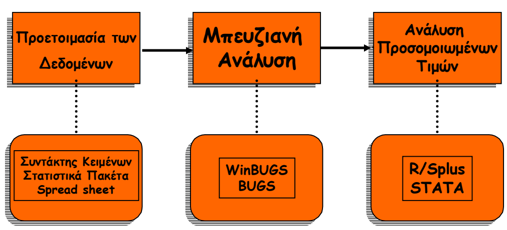
- ΛΑΘΟΣ ΑΠΟΤΕΛΕΣΜΑΤΑ ΛΟΓΩ ΜΗ ΣΥΓΚΛΙΣΗΣ
- ΠΟΛΥ ΑΡΓΗ ΣΥΓΚΛΙΣΗ
- ΚΑΚΕΣ ΑΡΧΙΚΕΣ ΤΙΜΕΣ
- ΚΑΚΗ ΕΠΙΛΟΓΗ ΜΟΝΤΕΛΟΥ
- ΥΠΕΡ-ΠΑΡΑΜΕΤΡΟΠΟΙΗΣΗ (over-parametrazation)
- ΣΠΑΣΙΜΟ ΝΕΥΡΩΝ

3.1. Εισαγωγή: Τι είναι το BUGS

ΑΡΧΕΙΑ BUGS

- ***.DAT**:Αρχείο Δεδομένων
- ***.INI**: Αρχείο Αρχικών τιμών
- ***.BUG**: Αρχείο Μοντέλου
- ***.CMD**: Αρχείο Εντολών προσομοίωσης
- ***.LOG**: Αρχείο Αποτελεσμάτων προσομοίωσης
- ***.OUT**: Αρχείο Προσομοιωμένων τιμών
- ***.IND**: Αρχείο με Περιεχόμενα του ***.OUT**

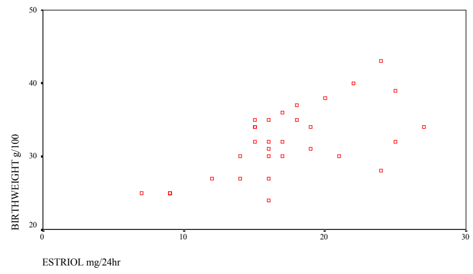
3.1. Εισαγωγή: Τι είναι το BUGS



3.2. Ένα Απλό Παράδειγμα στο BUGS

- Green & Touchston (1963, *Am.Jour. Of Obsterics & Gynecology*)
- Μελέτη σχέσης
  - Y : Βάρος γέννησης (birthweight) ενός παιδιού
  - X : Επίπεδο οιστριόλης (estriol) των εγκύων γυναικών
  - n=31
- Το γράφημα που ακολουθεί δείχνει ότι υπάρχει συσχέτιση

3.2. Ένα Απλό Παράδειγμα στο BUGS



3.2.1. Χτίζοντας το Μοντέλο

- ΤΥΧΑΙΑ ΣΥΝΙΣΤΩΣΑ: $Birth_i \sim Normal(\mu_i, \sigma^2)$
- ΣΥΣΤΗΜΑΤΙΚΗ ΣΥΝΙΣΤΩΣΑ: $\eta_i = \alpha+\beta\times Estriol_i$
- ΣΥΝΔΕΤΙΚΗ ΣΥΝΑΡΤΗΣΗ: $\mu_i=\eta_i=\alpha+\beta\times Estriol_i$
- για i=1,...,31
- PRIORS (Non-informative)
  - $f(\alpha)$=Normal (0, 10^4)
  - $f(\beta)$=Normal (0, 10^4)
  - $f(\sigma^2)$=Inverse Gamma (10^{-4}, 10^{-4}) ή
  - $f(\tau=\sigma^{-2})$=Gamma (10^{-4}, 10^{-4})

3.2.2. Δημιουργία Μοντέλων στο BUGS

- ΣΕ *.BUG Ορίζουμε το μοντέλο μας
- Εντολές => BUGS MANUAL σελ 17-18
- ΔΟΜΗ:
  - ***Προκαταρκτικό κομμάτι***: δηλώνουμε μεταβλητές, σταθερές, δεδομένα και αρχικές τιμές
  - ***Κυρίως μοντέλο***: εδώ ορίζουμε πιθανοφάνεια και priors

3.2.2. Δημιουργία Μοντέλων στο BUGS

- ΠΡΟΚΑΤΑΡΚΤΙΚΟ ΚΟΜΜΑΤΙ
  - **`;`:** τελειώνει κάθε εντολή του BUG file
  - **`model`** <*όνομα μοντέλου*>**;**
  - **`const`** <*σταθερά 1=#*>**,...,** <*σταθερά κ=#*>**;** Ορισμός σταθερών παραμέτρων
  - **`var`** <*μεταβλητή 1*>**, ... ,** <*μεταβλητή κ*>**;** Ορισμός τυχαίων μεταβλητών
  - **`data`** <*μεταβλητή 1*>**, ... ,** <*μεταβλητή κ*> **`in`** '<*ονομα αρχείου*>'**,** <*μεταβλητή κ+1*>**, ... ,** <*μεταβλητή κ+λ*> **in** '<ονομα αρχείου2>' : ορισμός αρχείων δεδομένων
  - **`inits in`** '<*ονομα αρχείου*>'**:** ορισμός αρχείων με αρχικές τιμές

3.2.2. Δημιουργία Μοντέλων στο BUGS

- ΚΥΡΙΩΣ ΜΟΝΤΕΛΟ
  - Το κυρίως μοντέλο αρχίζει και τελειώνει με **{}**
  - **~ :** ορίζουμε τις τυχαίες μεταβλητές
  - **<- :** ισότητα/ ανάθεση
  - Οι κατανομές ορίζονται στις σελίδες 17-18
  - Αν θέλουμε να περιορίσουμε μία κατανομή σε ένα διάστημα (α,β) τότε η κατανομή ακολουθείται από **I(α,β)**

3.2.2. Δημιουργία Μοντέλων στο BUGS

- ΚΥΡΙΩΣ ΜΟΝΤΕΛΟ
  - **Κανονικη Κατανομη:** `y~dnorm( mu, tau )`
    - **mu=** μέσος
    - **tau=** ακρίβεια (precision) $=1/\sigma^2$
  - **Γάμμα Κατανομη:** `y~dgamma( a, b )`
    - **μέσος** = a/b
  - `x[i]` : i στοιχείο του διανύσματος x
  - `d[i,j]` : στοιχείο της i γραμμής και j στήλης του πίνακα d

3.2.3. Εισαγωγή του Μοντέλου του Παραδείγματος στο BUGS

(1) $Birth_i \sim Normal(\mu_i, \sigma^2)$
(2) $\eta_i = \alpha+\beta \times Estriol_i$
(3) $\mu_i=\eta_i=\alpha+\beta \times Estriol_i$
για $i=1,\ldots,31$

- `for (i in 1:n) {`
- `Birth[i]~dnorm(mu[i],tau)`
- `mu[i] <-a+b*estriol[i]`
- `}`

PRIORS

- $f(\alpha)$=Normal (0, 10^4) — `a~dnorm(0.0,1.0E-04)`
- $f(\beta)$=Normal (0, 10^4) — `b~dnorm(0.0,1.0E-04)`
- $f(\tau=\sigma^{-2})$=Gamma(10^{-4}, 10^{-4}) — `tau~dgamma(1.0E-04,1.0E-04)`
- $\sigma^2=1/\tau$ — `s2<-1/tau`

3.2.3. Εισαγωγή του Μοντέλου του Παραδείγματος στο BUGS

ΠΡΟΚΑΤΑΡΚΤΙΚΟ ΚΟΜΜΑΤΙ

```
model example1;
const n=31; # n=sample size
var   estriol[n],  # estriol level of pregnant woman
      birth[n],    # birthweight
      mu[n],       # regression expected value
      a,b,tau,s2;  # model parameters,
                   # tau = precision, s2=1/tau varianc
data estriol,birth in 'estriol.dat';
inits in 'estriol.in';
```

3.2.4. Προσομοιώνοντας από την posterior

ΕΝΤΟΛΕΣ ΠΡΟΣΟΜΟΙΩΣΗΣ (σελ 31, BUGS MANUAL)

- `bugs [enter]` ΞΕΚΙΝΗΜΑ ΤΟΥ BUGS
- `compile('estriol.bug')` ΤΣΕΚΑΡΕΙ ΤΗ ΣΥΝΤΑΞΗ ΤΟΥ ΜΟΝΤΕΛΟΥ, ΒΡΙΣΚΕΙ ΤΙΣ CONDITIONALS ΚΑΙ ΤΙΣ ΑΡΧΙΚΕΣ ΤΙΜΕΣ
- `update(1000)` ΠΡΟΣΟΜΟΙΩΝΕΙ 1000 ΤΙΜΕΣ
- `monitor(a)`
- `monitor(b)` ΑΠΟ ΤΗ ΣΤΙΓΜΗ ΑΥΤΗ ΑΠΟΘΗΚΕΥΕΙ ΤΙΣ ΠΡΟΣΟΜΟΙΩΜΕΝΕΣ ΤΙΜΕΣ
- `monitor(s2)`
- `monitor(s2,10)` ΑΠΟΘΗΚΕΥΣΗ ΑΝΑ 10 (THIN=10)

3.2.4. Προσομοιώνοντας από την posterior

ΕΝΤΟΛΕΣ ΠΡΟΣΟΜΟΙΩΣΗΣ (σελ 31, BUGS MANUAL)

- `update(1000)` ΠΡΟΣΟΜΟΙΩΝΟΥΜΕ 1000 ΤΙΜΕΣ
- `stats(a)`
- `stats(b)` ΣΤΑΤΙΣΤΙΚΟΙ ΔΕΙΚΤΕΣ ΤΩΝ ΠΡΟΣΟΜΟΙΩΜΕΝΩΝ ΤΙΜΩΝ (ΤΗΣ posterior ΚΑΤΑΝΟΜΗΣ)
- `stats(s2)`
- `diag(a)`
- `diag(b)` ΔΙΑΓΝΩΣΤΙΚΟ ΤΕΣΤ ΣΥΓΚΛΙΣΗΣ
- `diag(s2)`
- `q()` ΕΞΟΔΟΣ ΑΠΟ ΤΟ BUGS

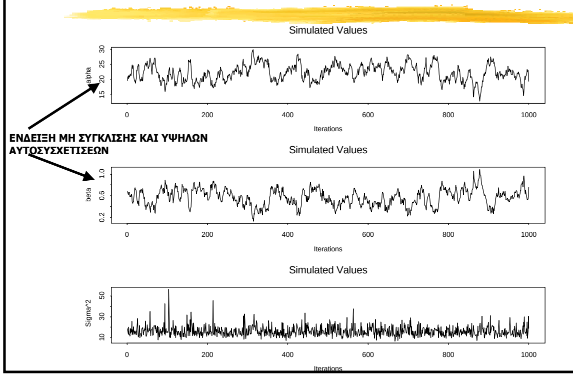
- EDIT BUGS.LOG ΒΛΕΠΟΥΜΕ ΟΛΑ ΤΑ ΑΠΟΤΕΛΕΣΜΑΤΑ
- EDIT BUGS.IND ΠΕΡΙΕΧΟΜΕΝΑ BUGS.OUT
- ΕΙΣΑΓΟΥΜΕ ΤΟ BUGS.OUT ΣΕ ΣΤΑΤΙΣΤΙΚΟ ΠΑΚΕΤΟ ΓΙΑ ΑΝΑΛΥΣΗ

3.2.4. Προσομοιώνοντας από την posterior

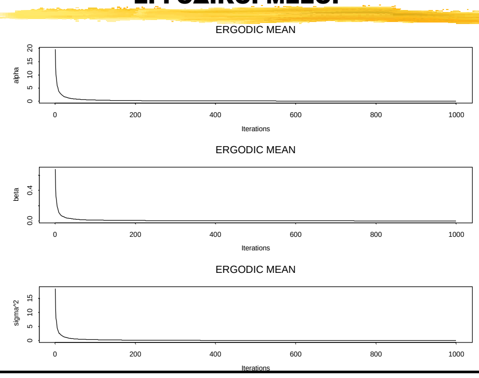
ΠΡΟΣΟΜΟΙΩΣΗ ΣΤΟ ΠΑΡΑΣΚΗΝΙΟ

- Γράφουμε τις εντολές ένα αρχείο με κατάληξη CMD πχ. **Example1.cmd**
- Τρέχουμε το μοντέλο στο παρασκήνιο με την εντολη **backbugs <όνομα αρχείου CMD>**
  π.χ. **backbugs example.cmd**
- Τα αποτελέσματα τα βλέπουμε με τον ίδιο τρόπο όπως παραπάνω δηλ.
- EDIT BUGS.LOG ΒΛΕΠΟΥΜΕ ΟΛΑ ΤΑ ΑΠΟΤΕΛΕΣΜΑΤΑ
- EDIT BUGS.IND ΠΕΡΙΕΧΟΜΕΝΑ BUGS.OUT
- ΕΙΣΑΓΟΥΜΕ ΤΟ BUGS.OUT ΣΕ ΣΤΑΤΙΣΤΙΚΟ ΠΑΚΕΤΟ ΓΙΑ ΑΝΑΛΥΣΗ

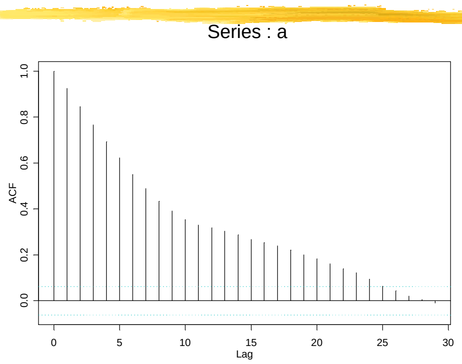
ΔΙΑΓΡΑΜΜΑ 1:
Trace of Posterior Values



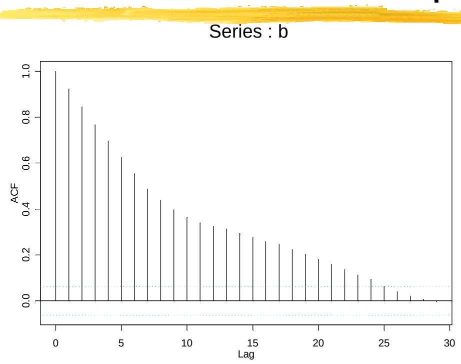
ΔΙΑΓΡΑΜΜΑ 2:
ΕΡΓΟΔΙΚΟΙ ΜΕΣΟΙ



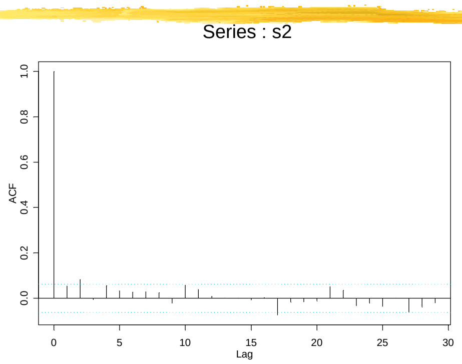
ΔΙΑΓΡΑΜΜΑ 3:
ΑΥΤΟΣΥΣΧΕΤΙΣΕΙΣ ΓΙΑ ΤΗΝ ΠΑΡΑΜΕΤΡΟ α



ΔΙΑΓΡΑΜΜΑ 4:
ΑΥΤΟΣΥΣΧΕΤΙΣΕΙΣ ΓΙΑ ΤΗΝ ΠΑΡΑΜΕΤΡΟ β



ΔΙΑΓΡΑΜΜΑ 5:
ΑΥΤΟΣΥΣΧΕΤΙΣΕΙΣ ΓΙΑ ΤΗΝ ΠΑΡΑΜΕΤΡΟ σ^2



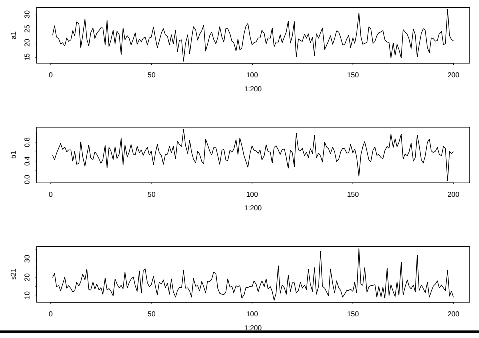
3.2.4. Προσομοιώνοντας από την posterior

ΔΙΟΡΘΩΣΗ ΤΗΣ ΑΥΤΟΣΥΣΧΕΤΙΣΗΣ

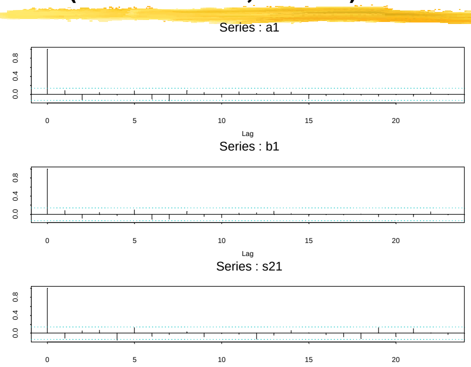
(1) ΑΥΞΑΝΟΥΜΕ ΤΟ ΔΙΑΣΤΗΜΑ ΛΕΠΤΥΝΣΗΣ (Thin=25)

- ΣΤΟ CMD ΑΡΧΕΙΟ ΤΟΥ BUGS
  - `monitor(a,25)`
  - `monitor(b,25)`

**ΔΙΑΓΡΑΜΜΑ 6:
Trace of Posterior Values
(5000 Iterations, Thin=25)**



**ΔΙΑΓΡΑΜΜΑ 7:
ΑΥΤΟΣΥΣΧΕΤΙΣΕΙΣ
(5000 Iterations, Thin=25)**



3.2.4. Προσομοιώνοντας από την posterior

ΔΙΟΡΘΩΣΗ ΤΗΣ ΑΥΤΟΣΥΣΧΕΤΙΣΗΣ

(2) ΑΛΛΑΖΟΥΜΕ ΤΗΝ ΠΑΡΑΜΕΤΡΟΠΟΙΗΣΗ ΤΟΥ ΜΟΝΤΕΛΟΥ

- ΑΦΑΙΡΕΣΗ ΤΟΥ ΜΕΣΟΥ ΑΠΟ ΚΑΘΕ ΕΠΕΞΗΓΗΜΑΤΙΚΗ ΜΕΤΑΒΛΗΤΗ ΒΟΗΘΑΕΙ ΣΤΗ ΣΥΓΚΛΙΣΗ ΑΡΑ
  - $\eta_i = \alpha^* + \beta(x_i - \bar{x})$
  - $\alpha = \alpha^* - \beta\,\bar{x}$

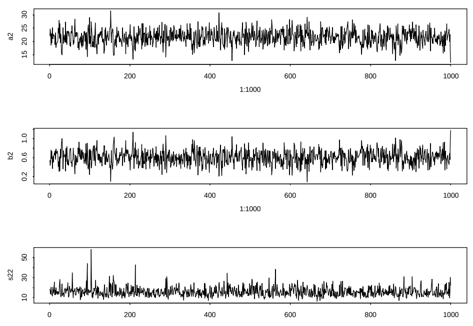
3.2.4. Προσομοιώνοντας από την posterior

ΔΙΟΡΘΩΣΗ ΤΗΣ ΑΥΤΟΣΥΣΧΕΤΙΣΗΣ

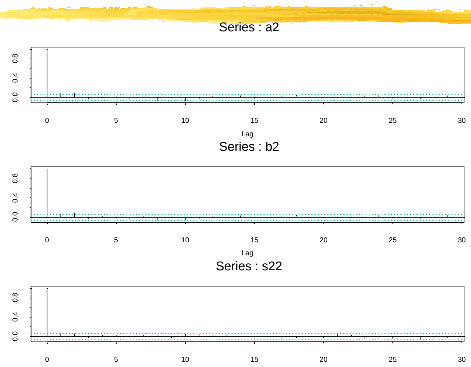
(2) ΑΛΛΑΖΟΥΜΕ ΤΗΝ ΠΑΡΑΜΕΤΡΟΠΟΙΗΣΗ ΤΟΥ ΜΟΝΤΕΛΟΥ

- ΑΦΑΙΡΕΣΗ ΤΟΥ ΜΕΣΟΥ ΑΠΟ ΚΑΘΕ ΕΠΕΞΗΓΗΜΑΤΙΚΗ ΜΕΤΑΒΛΗΤΗ ΒΟΗΘΑΕΙ ΣΤΗ ΣΥΓΚΛΙΣΗ ΑΡΑ ΣΤΟ BUGS
  - `mu[i]) <- a.star+b*(estriol[i]-mean(estriol[]));`
  - `a<-a.star-b*mean(estriol[]);`
  - `a.star~dnorm(0.0, 1.0E-04);`

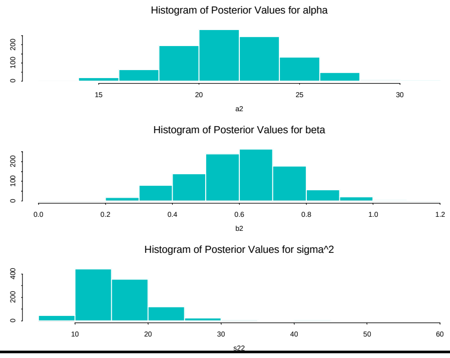
**ΔΙΑΓΡΑΜΜΑ 8:
Trace of Posterior Values**



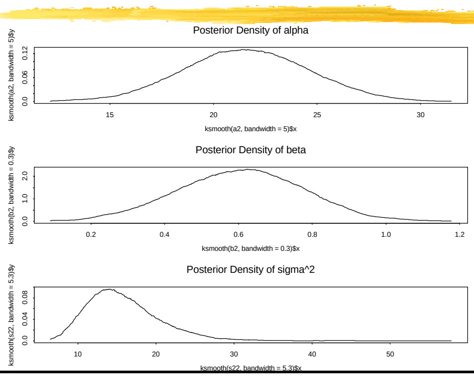
**ΔΙΑΓΡΑΜΜΑ 9:
ΑΥΤΟΣΥΣΧΕΤΙΣΕΙΣ**



ΔΙΑΓΡΑΜΜΑ 10: ΙΣΤΟΓΡΑΜΜΑΤΑ POSTERIOR ΤΙΜΩΝ



ΔΙΑΓΡΑΜΜΑ 11: ΕΚΤΙΜΩΜΕΝΕΣ POSTERIOR ΚΑΤΑΝΟΜΕΣ



3.2.5. Προσομοίωση στο Παρασκήνιο

- ΔΗΜΙΟΥΡΓΟΥΜΕ ΕΝΑ ΑΡΧΕΙΟ ΜΕ ΤΙΣ ΠΡΟΗΓΟΥΜΕΝΕΣ ΕΝΤΟΛΕΣ (ESTRIOL.CMD)
- ΠΡΟΣΟΜΟΙΩΝΟΥΜΕ ΤΙΜΕΣ ΜΕ `BACKBUGS ESTRIOL.CMD` ΣΤΟΝ ΚΑΤΑΛΟΓΟ ΤΟΥ BUGS
- ΒΛΕΠΟΥΜΕ ΤΑ ΑΠΟΤΕΛΕΣΜΑΤΑ ΣΤΟ ΑΡΧΕΙΟ BUGS.LOG

3.3. Παράδειγμα 2: Μοντέλα για Διωνυμικά Δεδομένα

- ΠΑΡΑΔΕΙΓΜΑ 12, BUGS EXAMPLES vol 2 σελ. 43
- Bliss (1935)

Εκθέτουμε 8 ομάδες εντόμων σε διαφορετικά επίπεδα Carbon dislphide και καταγράφουμε

- Συγκέντρωση (X_i)
- Συνολικός Αριθμός εντόμων στην ομάδα (n_i)
- Αριθμός εντόμων που απεβίωσαν (r_i)

3.3. Παράδειγμα 2: Μοντέλα για Διωνυμικά Δεδομένα

- (1) $r_i \sim \text{Binomial}(p_i, n_i)$
- (2) $\eta_i = \alpha^* + \beta(x_i - \bar{x})$
- (3) $\text{logit}(p_i) = \eta_i$
- για i=1,...,8
- PRIORS
  - $f(\alpha)$=Normal (0, 10^4)
  - $f(\beta)$=Normal (0, 10^4)
- $\alpha = \alpha^* - \beta\,\bar{x}$
- LINK FUNCTIONS
  - logit(p)=log{p/(1-p)}
  - probit(p)=Φ^{-1}(p)
  - cloglog(p)=log(-log(1-p))

```
for (i in 1:n) {
  r[i]~dbinom(p[i],n[i])
  logit(p[i])<-a.star+b*x[i]
}
a~dnorm(0.0,1.0E-04)
b~dnorm(0.0,1.0E-04)
a<-a.star-b*mean(x[])
```

3.3. Παράδειγμα 2: Μοντέλα για Διωνυμικά Δεδομένα

ΜΕΡΙΚΑ ΣΧΟΛΙΑ

- ΑΝΑΜΕΝΟΜΕΝΕΣ ΤΙΜΕΣ ΥΠΟΛΟΓΙΖΟΝΤΑΙ ΩΣ
  - `r.hat[i]<-n[i]*p[i];`
- ODDS RATIO ΓΙΑ LOGIT MODELS
  - `odds.ratio<-exp(b);`

3.4. Παράδειγμα 3: Μοντέλα για Δίτιμες Μεταβλητές

Ένα δείγμα ηλικιωμένων ατόμων εξετάστηκαν ψυχιατρικά αν έχουν κάποια συμπτώματα γηρατειών (senility symptoms).

Μια επεξηγηματική μεταβλητή είναι είναι το σκορ σε ένα υπο-τεστ της κλίμακας ενήλικης ευφυΐας Wechsler (Wecshler Adult Intelligence Scale - WAIS).

Να βρεθεί ποιο x αντιστοιχεί σε p=1/2 και να βρεθεί και η posterior κατανομή της πιθανότητας για κάποιον με WAIS ίσο με το μέσο όρο

3.4. Παράδειγμα 3: Μοντέλα για Δίτιμες Μεταβλητές

ΜΕΡΙΚΑ ΣΧΟΛΙΑ

- Το μοντέλο

```
for (i in 1:n) {
    symptom[i]~dbern( p[i] );
    logit( p[i] ) <- a+b*wais[i]; }
```

- Το x για p=1/2 βρίσκεται ως

```
x.half<- -a/b;
```

- Το ποσοστό για κάποιον με μέσο σκορ WAIS

```
p.mean<-
    exp(a+b*mean(x[]))/(1+exp(a+b*mean(x[])));
```

3.5. Παράδειγμα 4: Μοντέλα Poisson για 2x2 Πίνακες Συνάφειας & Η Posterior Κατανομή του Odds Ratio

Mahon *et.al.* (1970) bulletin of the world health organazation

Μελέτη για την πιθανή θετική σχέση μεταξύ ηλικίας στην 1η γέννα και καρκίνου του μαστο.

Οι περιπτώσεις (cases) προέρχονται από επιλεγμένα νοσοκομεία στις ΗΠΑ, Ελλάδα, Γιουγκοσλαβία, Βραζιλία, Ταϊβάν & Ιαπωνία

Οι Μάρτυρες (controls) επιλέχθηκαν απο γυναίκες με συγκρίσιμη ηλικία από τα ίδια νοσοκομεία

3.5. Παράδειγμα 4: Μοντέλα Poisson για 2x2 Πίνακες Συνάφειας & Η Posterior Κατανομή του Odds Ratio

	AGE AT FIRST BIRTH	
STATUS	Age>29 (1)	Age<30 (0)
Case (1)	683	2537
Control (0)	1498	8747

3.5. Παράδειγμα 4: Μοντέλα Poisson για 2x2 Πίνακες Συνάφειας & Η Posterior Κατανομή του Odds Ratio

Status	Age	Counts
1	1	683
1	0	2537
0	1	1498
0	0	8747

3.5. Παράδειγμα 4: Μοντέλα Poisson για 2x2 Πίνακες Συνάφειας & Η Posterior Κατανομή του Odds Ratio

ΤΟ ΜΟΝΤΕΛΟ

```
for (i in 1:4) {
   counts[i]~dpois(lambda[i]);
   log (lambda[i])<-mu+
     a*status[i]+b*age[i]+ab*status[i]*age[i];
   }
```

ODDS RATIO

```
odds.ratio<-exp(ab);
```

3.6. Παράδειγμα 5: Εκτίμηση κοινού Odds Ratio σε 2x2xJ Πίνακες Συνάφειας

Σε πίνακες 2x2xJ η εκτίμηση ενός κοινού OR γίνεται από το Maentel-Haenzel $OR_{MH}=(\Sigma a_i d_i/n_i)/(\Sigma b_i c_i/n_i)$

Sandler, Everson & Wilcox (1985) *Amer.Journal of Epidemiology*

Μελέτη με 518 καρκινοπαθείς με ηλικίες 15-59 και 518 μάρτυρες (ομάδα ελέγχου) ταιριασμένοι (matched) ως προς φύλο και ηλικία

Σκοπός: εκτίμηση της επίδρασης του παθητικού κανίσματος στον κίνδυνο εμφάνισης καρκίνου. Το παθητικό κάπνισμα ορίστικε θετικά αν η σύζυγος κάπνιζε τουλάχιστον 1 τσιγάρο ημερησίως τους τελευταίους 6 μήνες.

Συγχυτικός παράγοντας (confounder) αν το ίδιο άτομο καπνίζει

3.6. Παράδειγμα 5: Εκτίμηση κοινού Odds Ratio σε 2x2xJ Πίνακες Συνάφειας

	Non Smokers (0)		Smokers (1)	
	Passive Smoker (1)	Non Passive Smoker(0)	Passive Smoker (1)	Non Passive Smoker (0)
Case(1)	120	111	161	117
Control (0)	80	155	130	124

3.6. Παράδειγμα 5: Εκτίμηση κοινού Odds Ratio σε 2x2xJ Πίνακες Συνάφειας

- ΚΑΝΟΥΜΕ ΔΥΟ ΑΝΑΛΥΣΕΙΣ
  - ΑΝΑΛΥΣΗ 1:
    - ΧΕΙΡΙΖΟΜΑΣΤΕ ΞΕΧΩΡΙΣΤΑ ΤΟΝ ΚΑΘΕ ΠΙΝΑΚΑ ΚΑΙ ΕΚΤΙΜΟΥΜΕ ΕΝΑ ODDS RATIO ΓΙΑ ΚΑΘΕ ΠΙΝΑΚΑ (ΟΠΩΣ ΠΑΡΑΔΕΙΓΜΑ 4)
    - ΣΥΓΚΡΙΝΟΥΜΕ ΤΙΣ POSTERIOR ΚΑΤΑΝΟΜΕΣ ΤΟΥΣ
  - ΑΝΑΛΥΣΗ 2:
    - ΧΕΙΡΙΖΟΜΑΣΤΕ ΞΕΧΩΡΙΣΤΑ ΤΟΝ ΚΑΘΕ ΠΙΝΑΚΑ
    - ΕΚΤΙΜΟΥΜΕ ΕΝΑ ΚΟΙΝΟ ODDS RATIO ΓΙΑ ΚΑΘΕ ΠΙΝΑΚΑ

3.6. Παράδειγμα 5: Εκτίμηση κοινού Odds Ratio σε 2x2xJ Πίνακες Συνάφειας

- ΑΝΑΛΥΣΗ 1:
  - ΠΡΟΚΑΤΑΡΤΙΚΟ ΚΟΜΜΑΤΙ

```
var …, b[2,4], or[2];
```

  - ΜΟΝΤΕΛΟ

```
#model for 1st table (nonsmokers)
for (i in 1:4) {
 counts[i]~dpois( lambda[i] );
 log(lambda[i])<-b[1,1]+b[1,2]*status[i]
              +b[1,3]*passive[i]
              +b[1,4]*status[i]*passive[i];
}
```

3.6. Παράδειγμα 5: Εκτίμηση κοινού Odds Ratio σε 2x2xJ Πίνακες Συνάφειας

- ΑΝΑΛΥΣΗ 1:
  - ΠΡΟΚΑΤΑΡΤΙΚΟ ΚΟΜΜΑΤΙ

```
var …, b[2,4], or[2];
```

  - ΜΟΝΤΕΛΟ

```
#model for 2nd table (smokers)
for (i in 5:8) {
 counts[i]~dpois( lambda[i] );
 log(lambda[i])<-b[2,1]+b[2,2]*status[i]
              +b[2,3]*passive[i]
              +b[2,4]*status[i]*passive[i];
}
```

3.6. Παράδειγμα 5: Εκτίμηση κοινού Odds Ratio σε 2x2xJ Πίνακες Συνάφειας

- ΑΝΑΛΥΣΗ 1:
  - ΠΡΟΚΑΤΑΡΤΙΚΟ ΚΟΜΜΑΤΙ

```
var …, b[2,4], or[2];
```

  - ΜΟΝΤΕΛΟ

```
#priors
for (i in 1:2){
    for (j in 1:p){
        b[i,j]~dnorm(0.0, 1.0E-04)
    }
}
```

3.6. Παράδειγμα 5: Εκτίμηση κοινού Odds Ratio σε 2x2xJ Πίνακες Συνάφειας

- ΑΝΑΛΥΣΗ 2:
  - ΜΟΝΤΕΛΟ

```
#model for 1st table (nonsmokers)
for (i in 1:4) {
 counts[i]~dpois( lambda[i] );
 log(lambda[i])<-b[1,1]+b[1,2]*status[i]
              +b[1,3]*passive[i]
              +b[1,4]*status[i]*passive[i];}
#model for 2nd table (smokers)
for (i in 5:8) {
 counts[i]~dpois( lambda[i] );
 log(lambda[i])<-b[2,1]+b[2,2]*status[i]
              +b[2,3]*passive[i]
              +b[2,4]*status[i]*passive[i];  }
```

3.6. Παράδειγμα 5: Εκτίμηση κοινού Odds Ratio σε 2x2xJ Πίνακες Συνάφειας

- ΑΝΑΛΥΣΗ 2:
  - ΜΟΝΤΕΛΟ

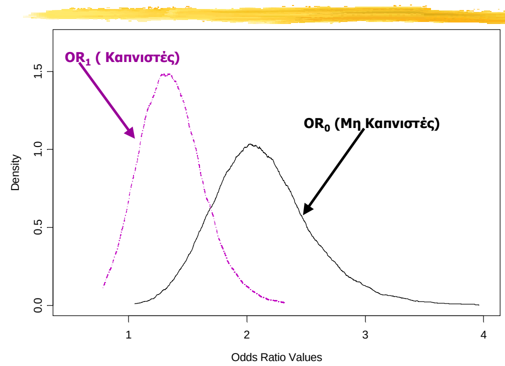
```
#model for 1st table (nonsmokers)
for (i in 1:4) {
 counts[i]~dpois( lambda[i] );
 log(lambda[i])<-b[1,1]+b[1,2]*status[i]
              +b[1,3]*passive[i]
              +  ab  *status[i]*passive[i];}
#model for 2nd table (smokers)
for (i in 5:8) {
 counts[i]~dpois( lambda[i] );
 log(lambda[i])<-b[2,1]+b[2,2]*status[i]
              +b[2,3]*passive[i]
              +  ab  *status[i]*passive[i];  }
```

3.6. Παράδειγμα 5: Εκτίμηση κοινού Odds Ratio σε 2x2xJ Πίνακες Συνάφειας

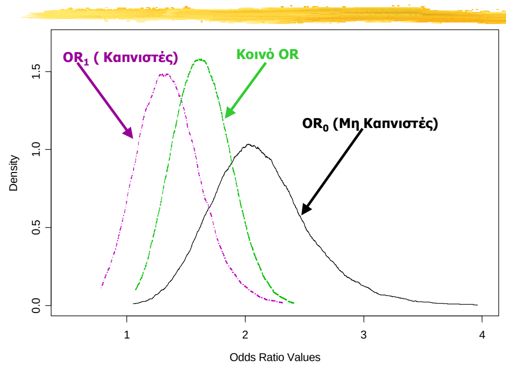
ΑΠΟΤΕΛΕΣΜΑΤΑ

	MLE	Posterior Mean	95% posterior credible interval
ΑΝΑΛΥΣΗ 1			
OR_0	2.09	2.07±0.036	1.47 - 3.09
OR_1	1.31	1.33±0.022	0.97 - 1.88
ΑΝΑΛΥΣΗ 2			
Common			
OR_{MH}	1.63	1.61 ± 0.0087	1.27 - 2.06

ΔΙΑΓΡΑΜΜΑ 12: ΕΚΤΙΜΩΜΕΝΕΣ POSTERIOR ΚΑΤΑΝΟΜΕΣ ΤΩΝ ODDS RATIOS



ΔΙΑΓΡΑΜΜΑ 13: ΕΚΤΙΜΩΜΕΝΕΣ POSTERIOR ΚΑΤΑΝΟΜΕΣ ΤΩΝ ODDS RATIOS



Bayesian Biostatistics Using BUGS

ΤΕΛΟΣ ΕΙΣΑΓΩΓΙΚΟΥ ΜΑΘΗΜΑΤΟΣ

ΕΠΟΜΕΝΟ: WINBUGS

E-mail: ntzoufras@aegean.gr

Department of
Business Administration,
University of the Aegean