Modelos jerárquicos Bayesianos

Modelos Bayesianos con aplicaciones ecológicas

Dr. Cole Monnahan
University of Concepción, Chile

Enero, 2019

Recap

- Ayer practicamos los paso para construir un modelo Bayesiano

1. Especifica las priors, estructura del modelo, y verosimilitud
2. Ajusta el modelo a los datos para obtener la posterior
3. Evalula el ajuste del modelo

- Usaron *prior and posterior predictive distributions* para chequear y DIC para seleccionar entre modelos

Un resume de modelos jerárquicos

- Un grupo de modelos con una estructura jerárquica
- Conocido por otros nombres:
  - *Random effects (mixed effects) models*
  - *State-space models*
  - *Multi-level models*
- Los jerárquicos ocurren en la naturaleza: individuos adentro sitios; subpoblaciones entre poblaciones, etc.
- **Muchas** maneras para interpretarlos, y puede ser difícil y abrumador

Royle and Dorazio 2008

Que son modelos jerárquicos?

- Los con efectos aleatorios
- Y que son esos?
  - Una fuente de varianza latente (no observable)
  - Normalmente estructurado por tiempo, espacio, sitio, región, individuo, etc.
  - En general se usa la supsicion que tienen una distribución normal, con un medio y una varianza no conocidos
  - Son estimados, y llamados: *(hyper)mean and (hyper)variance*

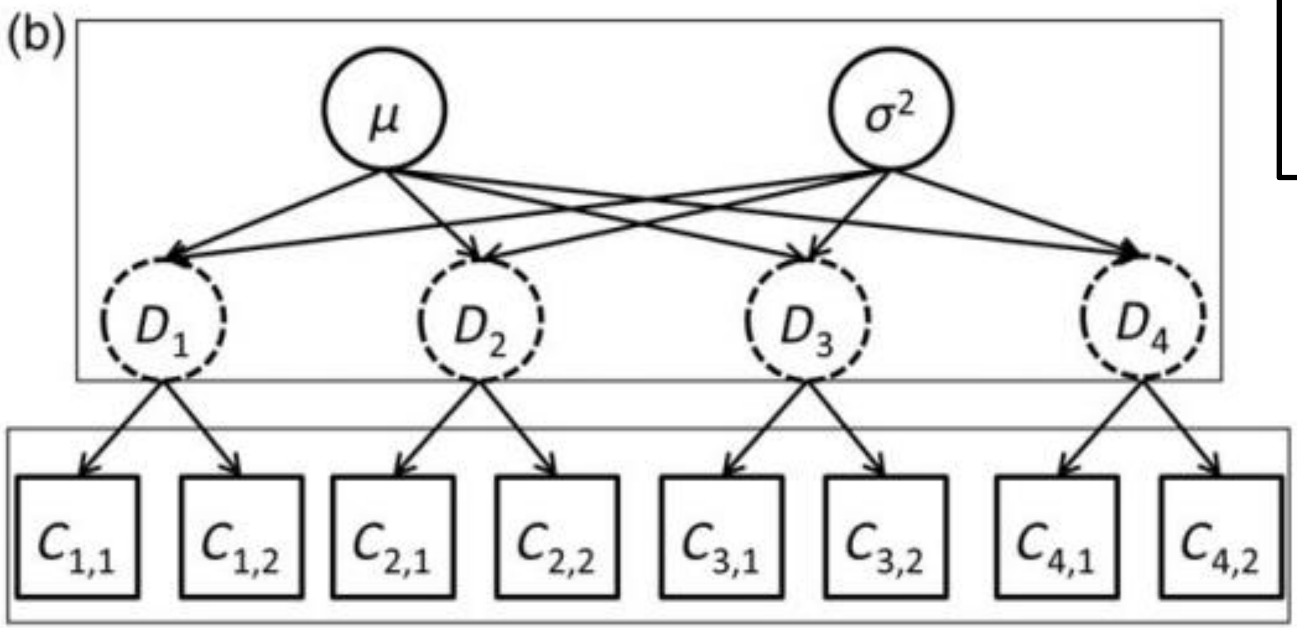
Exchangeability (intercambiables)

- Especialmente importante in la literatura Bayesiana (Gelman et al. 2004)
- *Exchangeability* sugiere que:
  - Efectos aleatorios son de un proceso común.
  - IID variables son intercambiables
  - E.g. dos poblaciones salvajes y una domesticado no seria intercambiables
  - E.g., el promedio de la densidad entre sitios

Vocabulary related to random effects

Term	Definition
Random effect	Coefficient that is “exchangeable” with one or more other coefficients
Hyperdistribution	Distribution for random effects
Exchangeable	No information is available to distinguish between residual variability in random effects
Fixed effect	Coefficient that is not exchangeable with others, and which hence is estimated without a hyperdistribution
Mixed-effect model	Model with both fixed and random effects

Motivating example



El promedio y varianza de la densidad entre sitios (*hyperdistribution*)

D= densidad (latente)

Dos observaciones de cada uno sitio (datos observados)

- Las densidades de los sitios son relacionadas
- No son observables (son latentes)

Thorson and Minto 2014

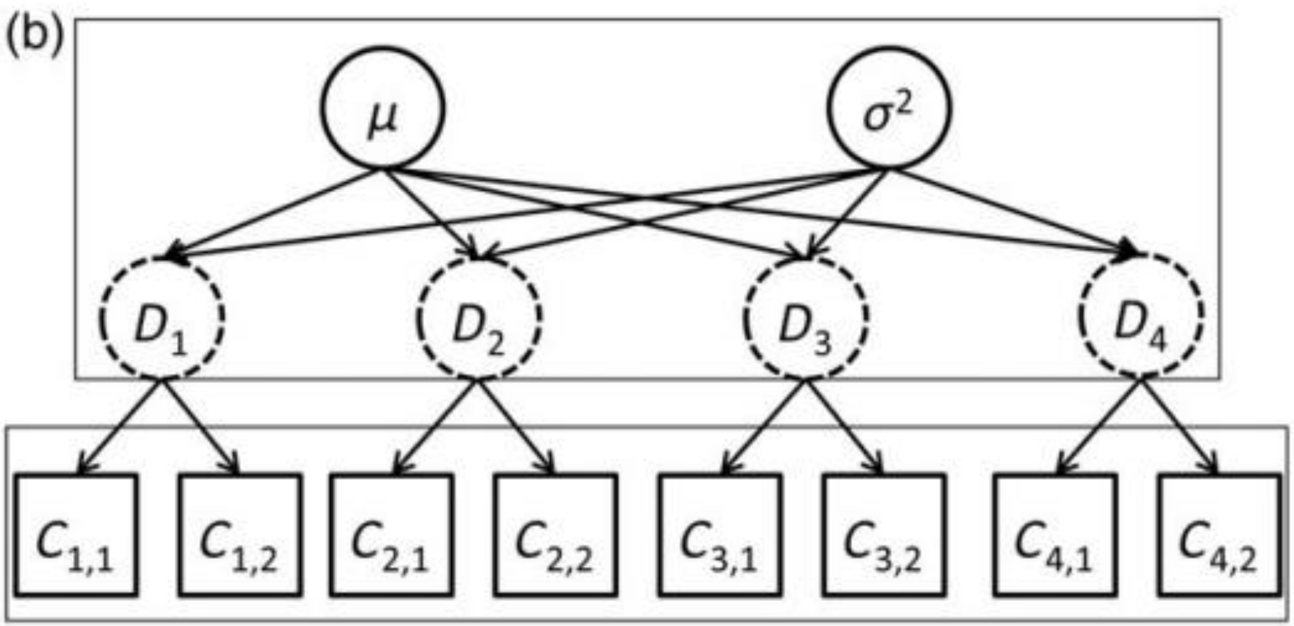
Razones para MJ

- Refleja los procesos naturales mejor
- Se puede aplicar la aleatoriedad a un sitio sin datos
- Comparta información. Los efectos no son estimados independiente sino que como un grupo asi que la información es comparticida
- Un meta-análisis de otros estudios
- Tener *process error* en el modelo

Construyendo del modelo

Paso 1: Likelihood de los datos $C_1, .. C_4$
Paso 2: Función del promedio
Paso 3: Elege la hiperdistribucion:

$$C \sim Poisson(D)$$
$$D_i = \lambda_i$$
$$\log \lambda \sim N(\mu, \sigma^2)$$

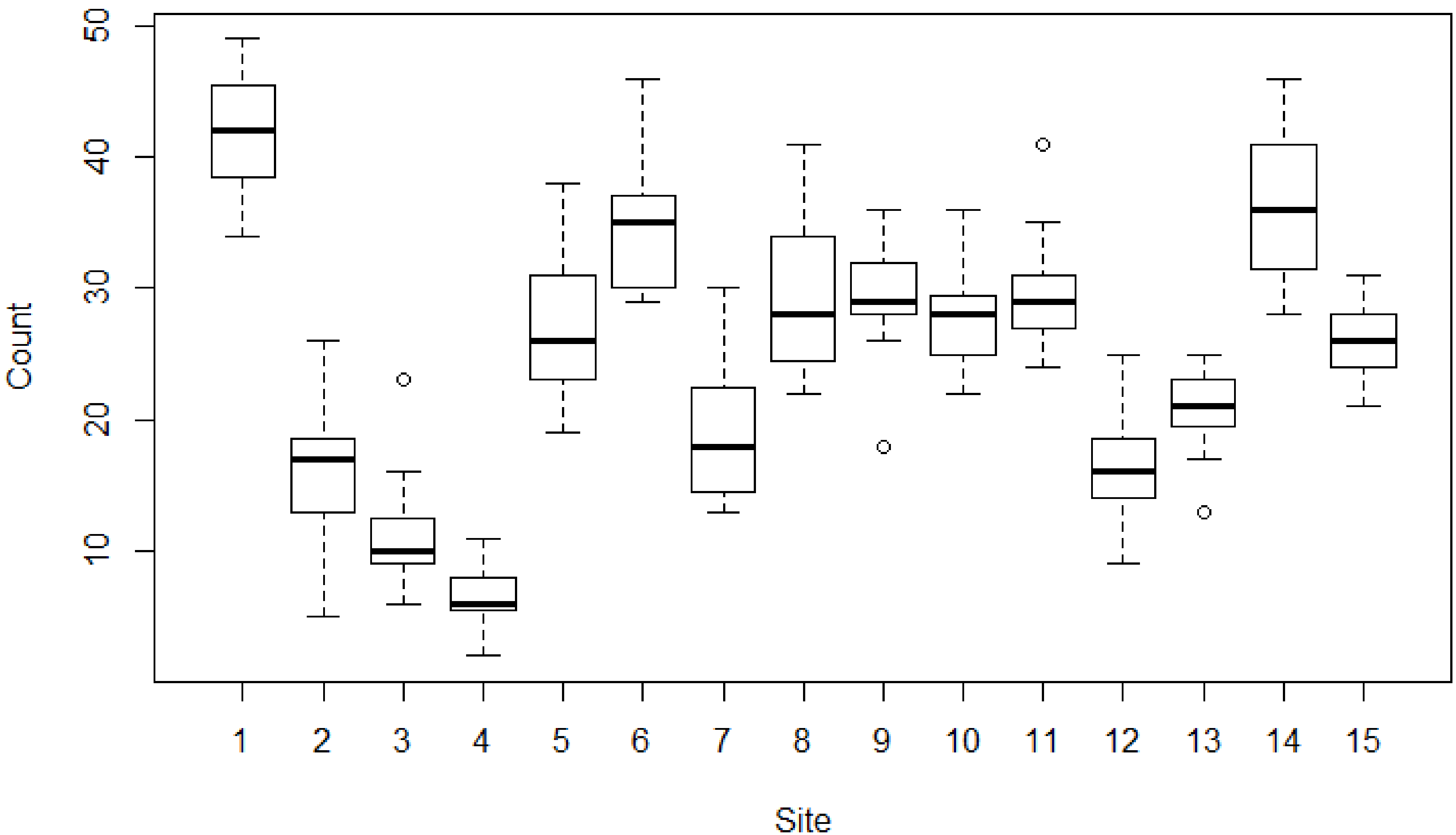


Thorson and Minto 2014

Ejercicio: Simulando datos jerárquicos

- Suponga que hay 15 IID sitios, cuyos promedios se distribuyen log-normal, i.e., loglambda~N(mu,tau)=N(3,.5)
- Y hay un proceso de los datos que es Poissoin asi que y_i~Poisson(lambda_i)
- Simula tan datos con 12 replicas por cada sitio
- Hace un *boxplot* de los datos simulados

Ejercicio: Simulando datos jerárquicos



Verosimilitudes jerárquicas

- Introduce "latent" variables into the likelihood

  $$\mathrm{L}(\theta; \varepsilon, \mathrm{y}) = \Pr(y, \varepsilon|\theta) = \Pr(y|\theta, \varepsilon)\Pr(\varepsilon|\theta)$$

  - y is data, where ε is a unobserved random variable
  - $\Pr(\varepsilon|\theta)$ is a "prior" or "hyper-distribution" for latent variables
- In our example:

  $$\Pr(y|\theta, \varepsilon)\Pr(\varepsilon) = \Pr(C|D)\Pr(D|\mu, \sigma)$$

- **This does not change our core formula:**

  $$\mathrm{P}(\theta|\mathrm{y}) \propto \Pr(y|\theta, \varepsilon)\Pr(\varepsilon|\theta)P(\theta)$$

Some slides borrowed from: https://github.com/James-Thorson/2016_Spatio-temporal_models

Comportamiento extraño de MJ

- Vamos a ignorar la verosimilitud y prior
- Que ocurre si los efectos aleatorios son del medio…
- Y la hipervarianza va a cero?

```
hyper <- function(lambda, mu, tau) {
  exp(sum(dnorm(lambda, mu, tau), log=TRUE))
}
> lambda <- rep(3,10)
> hyper(lambda, mu=3, tau=1)
[1] 146.8516
> hyper(lambda, mu=3, tau=.01)
[1] 4.928621e+173
> hyper(lambda, mu=3, tau=.0001)
[1] Inf
```

Comportamiento extraño de MJ

- La densidad de la hiperdistribucion es **infinita**!!!
- Podemos siempre encontrar una densidad mas alta como $\sigma \to 0$ y $\varepsilon \to \mu$
- Significa que la verosimilitud en un MJ **no tiene una moda valida**
- Para máxima verosimilitud eso es un problema (no hay una máxima!!)
- Métodos frequentistas tienen que **integrar** los efectos aleatorios para obtener la verosimilitud marginal

Comportamiento extraño de MJ

- Por lo tanto la inferencia frecuentista es muy difícil con MJ
- (Hasta TMB; Kristensen et al. 2016)
- **Cual es el efecto de la inferencia Bayesiana?**
  - La densidad de la posterior es infinita pero el volumen es igual de pequeño
  - Entonces esta región de la posterior no es importante[1]
  - MCMC nunca genera muestras allí porque la masa es muy baja
- No es una problema, y por eso lo métodos Bayesianos son tan común para MJ

[1] Section 1.4 of Betancourt 2017 arXiv preprint arXiv:1701.02434.

Important concepts 1

- MJ require integración para hacer inferencia
- Es difícil con máximo verosimilitud
- Pero natural con métodos Bayesianos
- Porque MCMC ya está integrando!
- MJ son herramientas muy poderosa y eran difícil de ajustar...
- Hasta software como BUGS/JAGS que son flexibles para construir modelos arbitrarios

Ejercicio

- Vamos a ajustar los datos simulados
- Primero, crea una *prior predictive distribution* en R.
- *Antes* de ver los datos, pensamos que 250 animales sería extremo (un umbral)
- La forma de la prior `tau~N(0, sigma)T(0,)` es recomendada por varianzas.
- Simula datos de un sitio y los plotea (posterior predictive distribution of new site)

References

- Kristensen, K., Nielsen, A., Berg, C. W., Skaug, H., & Bell, B. M. (2016). TMB: Automatic differentiation and Laplace approximation. *Journal of Statistical Software, 70(5), 21. doi: 10.18637/jss.v070.i05*