

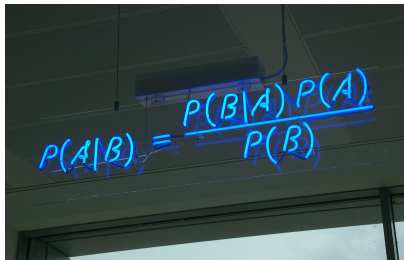


A Brief Introduction to Bayesian Statistics

Frameworks of Statistical Inference

- ▶ Frequentist Hypothesis Testing: Evaluate the probability of observing the data, or more extreme data, given that the hypothesis is true assuming that there is a single fixed True value for each parameter.
- ▶ Likelihood & Information Theoretic: Given the data at hand, compare multiple alternative hypotheses and evaluate the relative weight of evidence for each. Parameters again assumed to have True values.
- ▶ Bayesian: Using prior information and data, evaluate the degree of belief in specific hypotheses, recognizing that data is one realization of some distribution of a parameter.

Bayes Theorem



Bayes Theorem

$$p(\text{Hypothesis}|\text{Data}) = \frac{P(\text{Data}|\text{Hypothesis})p(\text{Hypothesis})}{p(\text{Data})}$$

Bayes Theorem

$$p(\theta|X) = \frac{p(X|\theta)P(\theta)}{p(X)}$$

We know/assume:

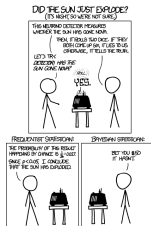
$$p(\text{Sun Explodes}) = 0.0001, P(\text{Yes} | \text{Sun Explodes}) = 35/36$$

We can calculate:

$$\begin{aligned} p(\text{Yes}) &= P(\text{Yes} | \text{Sun Explodes})p(\text{Sun Explodes}) + P(\text{Yes} | \text{Sun Doesn't Explode})p(\text{Sun Doesn't Explodes}) \\ &= 35/36 * 0.0001 + 1/36 * 0.9999 = 0.02777775 \end{aligned}$$

credit: Amelia Hoover

Bayes Theorem in Action



<http://xkcd.com/1132/>

Bayes Theorem in Action

$$p(\text{Sun Explodes}|\text{Yes}) = \frac{p(\text{Yes}|\text{Sun Explodes})p(\text{Sun Explodes})}{p(\text{Yes})}$$

We know/assume:

$$p(\text{Sun Explodes}) = 0.0001, P(\text{Yes} | \text{Sun Explodes}) = 35/36$$

We can calculate:

$$\begin{aligned} p(\text{Yes}) &= P(\text{Yes} | \text{Sun Explodes})p(\text{Sun Explodes}) + P(\text{Yes} | \text{Sun Doesn't Explode})p(\text{Sun Doesn't Explodes}) \\ &= 35/36 * 0.0001 + 1/36 * 0.9999 = 0.02777775 \end{aligned}$$

credit: Amelia Hoover

Bayes Theorem in Action

$$p(\text{Sun Explodes}|\text{Yes}) = \frac{p(\text{Yes}|\text{Sun Explodes})p(\text{Sun Explodes})}{p(\text{Yes})}$$

$$p(\text{Sun Explodes}|\text{Yes}) = \frac{0.0001 * 35/36}{0.028} = 0.0035$$

Incorporating Prior Information about the Sun Exploding gives us a very different answer

Note, we can also explicitly evaluate the probability of an alternate hypothesis - $p(\text{Sun Doesn't Explode} | \text{Yes})$

The Marginal Distribution in the Denominator

$$p(\theta|X) = \frac{p(X|\theta)P(\theta)}{\sum_{i=0}^j p(X|\theta_i)p(\theta_i)}$$

What are alternate parameter values but alternate hypotheses?

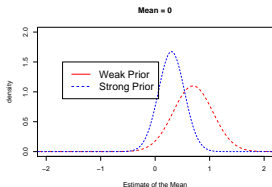
Denominator - marginal distribution - becomes an integral of likelihoods if θ is continuous. It normalizes the equation to be between 0 and 1.

How do we Choose a Prior?

- ▶ A prior is a powerful tool, but it can also influence our results of chosen poorly. This is a highly debated topic.
- ▶ Conjugate priors make some forms of Bayes Theorem analytically solvable
- ▶ If we have objective prior information - from pilot studies or the literature - we can use it to obtain a more informative posterior distribution
- ▶ If we do not, we can use a weak or flat prior (e.g., $N(0,1000)$).
Note: constraining the range of possible values can still be weakly informative - and in some cases beneficial

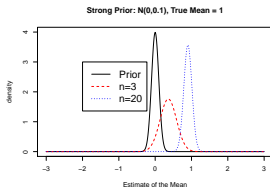
The Influence of Priors

Here's the posterior distribution drawn using the same sample - but in one case with a weak prior, and one a strong prior.



Priors and Sample Size

The influence of priors decreases with same size. A large sample size 'overwhelms' the prior.



Evaluation of a Posterior: Frequentist Confidence Intervals

In Frequentist analyses, the **95% Confidence Interval** of a parameter is the region in which, were we to repeat the experiment an infinite number of times, the *true value* would occur 95% of the time. For normal distributions of parameters:

$$\hat{\beta} - t(\alpha, df)SE_{\hat{\beta}} \leq \beta \leq \hat{\beta} + t(\alpha, df)SE_{\hat{\beta}}$$

Evaluation of a Posterior: Bayesian Credible Intervals

In Bayesian analyses, the **95% Credible Interval** is the region in which we find 95% of the possible parameter values. The observed parameter is drawn from this distribution. For normally distributed parameters:

$$\hat{\beta} - 2 * \hat{SD} \leq \beta \leq \hat{\beta} + 2 * \hat{SD}$$

where \hat{SD} is the SD of the posterior distribution of the parameter β . Note, for other types of parameters, the distribution may be different.

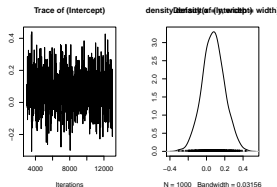
Bayes Theorem Expanded

$$p(\theta|X) = \frac{p(X|\theta)P(\theta)}{\sum_{i=0}^j p(X|\theta_i)p(\theta_i)} - \text{Algebraically Solvable}$$

$$p(\theta|X) = \frac{p(X|\theta)P(\theta)}{\int p(X|\theta)p(\theta)d\theta} - \text{Analytically Solvable for Conjugate Priors}$$

$$p(\theta|X) = \frac{\int p(X|\theta)P(\theta|\eta)p(\eta)d\eta}{\int \int p(X|\theta)p(\theta)d\theta d\eta} - \text{Hierarchical Model: need numerical integration approach with random hyperparameters}$$

Markov Chain Monte Carlo Sampling (MCMC)



Markov Chain Monte Carlo Sampling (MCMC)

If we cannot analytically solve a distribution, we can still simulate from it:

- ▶ Chose a set of starting values X at $t=0$
- ▶ Chose a random set of parameters, Y , from the distribution parameterized by X
- ▶ Select a uniform random number between 0 and 1, U
- ▶ If $U \leq f(X,Y)$, $X(t+1) = Y$. Otherwise, $X(t+1) = X$.
- ▶ Rinse and repeat

Markov Chain Monte Carlo Sampling (MCMC)

This is a time series. To use it for inference to sample from the final stationary distribution:

- ▶ Discard a 'burn in' set of samples
- ▶ 'Thin' your chain to reduce temporal autocorrelation
- ▶ Examine chain for convergence on your posterior distribution
- ▶ Evaluate multiple chains to ensure convergence to a single distribution

Many different samplers using different decision rules for f . We use the Gibbs Sampler commonly.

Software Options for MCMC

- ▶ WinBUGS <http://www.mrc-bsu.cam.ac.uk/bugs/>
- ▶ OpenBUGS <http://www.openbugs.info/w/>
- ▶ JAGS <http://mcmc-jags.sourceforge.net/>
- ▶ STAN <http://mc-stan.org/>
- ▶ MCMCglimm in R
- ▶ MCMCpack in R

BUGS code for a Simple Linear Regression

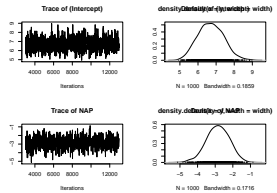
```
model {  
  # Prior  
  alpha ~ dnorm(0,0.001)  
  beta ~ dnorm(0,0.001)  
  sigma ~ dunif(0,100)  
  # Likelihood  
  for (i in 1:n){  
    y[i] ~ dnorm(mu[i],tau)  
    mu[i] <- alpha + beta*x[i]  
  }  
}
```

Example: The RIKZ Beaches and Tide Height

```
rikz <- read.csv("./data/rikz.csv")
rikz$Beach <- factor(rikz$Beach)
#
library(MCMCglmm)
NAPMod <- MCMCglmm(Richness ~ NAP, data=rikz, verbose=F)
```

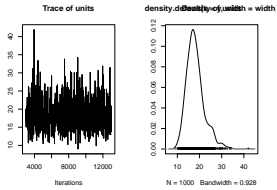
Plots of Chains

```
plot(NAPMod$Sol)
```

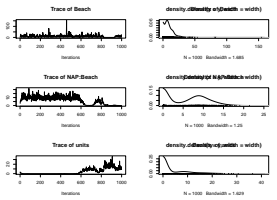


Plots of Chains

```
plot(NAPMod$VCV)
```



Sometimes Problems are Obvious



Did you Thin Enough?

```
autocorr(NAPMod$Sol)

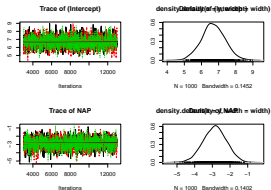
# , , (Intercept)
#
#      (Intercept)    NAP
# Lag 0      1.000000 -0.338722
# Lag 10     0.030775 -0.009913
# Lag 50     0.003102 -0.015534
# Lag 100    -0.050866  0.031194
# Lag 500    -0.077621  0.031906
#
# , , NAP
#
#      (Intercept)    NAP
# Lag 0     -0.338722  1.000000
# Lag 10    -0.008204  0.01056
# Lag 50    -0.008502  0.03985
# Lag 100   0.038907 -0.01880
# Lag 500   0.046233  0.01864
```

Did You Converge: Assessing with Multiple Chains

```
NAPMod2 <- MCMCglmm(Richness ~ NAP, data=rikz, verbose=F)
NAPMod3 <- MCMCglmm(Richness ~ NAP, data=rikz, verbose=F)
#
library(coda)
chainList <- mcmc.list(NAPMod$Sol, NAPMod2$Sol, NAPMod3$Sol)
```

Did You Converge: Assessing with Multiple Chains

```
plot(chainList)
```



The Gelman-Rubin Diagnostic

Diagnostic should be close to 1.

```
gelman.diag(chainList)

# Potential scale reduction factors:
#
#           Point est. Upper C.I.
# (Intercept)      1      1.01
# NAP              1      1.00
#
# Multivariate psrf
#
# 1
```

Evaluating Results

```
summary(NAPMod)

#
# Iterations = 3001:12991
# Thinning interval = 10
# Sample size = 1000
# ...
```

Evaluating Results

```
summary(NAPMod)

....

# R-structure: ~units
#
#      post.mean 1-95% CI u-95% CI eff.samp
# units      18.2      11.4      26.7      867
#
# ...
```

Evaluating Results

```
summary(NAPMod)

....

# Location effects: Richness ~ NAP
#
#      post.mean 1-95% CI u-95% CI eff.samp pMCMC
# (Intercept)    6.71    5.34    8.00    1000 <0.001
# NAP           -2.85   -4.02   -1.54    1000 <0.001
```

Your 95% Credible Interval

```
HPDinterval(NAPMod$Sol)

#              lower upper
# (Intercept)  5.345  7.997
# NAP         -4.019 -1.536
# attr(,"Probability")
# [1] 0.95
```


The Bayesian Approach to MMI: The DIC

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

from Spiegelhalter et al 2002

$D(\bar{\theta})$ is the average deviance and pD = Effective # of parameters

$$pD = D(\bar{\theta}) - D(\bar{\theta})$$

```
NAPMod$DIC
```

```
# [1] 260
```

Setting Priors

```
prior<-list(B=list(mu=c(0,-3),V=diag(c(1e+10, 1))))  
#  
NAPMod_Prior <- MCMCglmm(Richness ~ NAP,  
                          data=rikz, verbose=F, prior=prior)
```

Strong Priors Can Alter Parameters

```
summary(NAPMod)$solutions
```

#	post.mean	1-95% CI	u-95% CI	eff.samp	pMCMC
# (Intercept)	6.710	5.345	7.997	1000	0.001
# NAP	-2.849	-4.019	-1.536	1000	0.001

```
summary(NAPMod_Prior)$solutions
```

#	post.mean	1-95% CI	u-95% CI	eff.samp	pMCMC
# (Intercept)	6.696	5.355	7.863	1000	0.001
# NAP	-2.884	-3.838	-1.917	1000	0.001

Random Effects

MCMCglmm allows random effects & family much like nlme

```
MCMCglmm(y ~ x, random = z + x:z)
```

Implies that the intercept varies randomly by z and the slope of x varies by z. Equivalent to $(1+x | z)$

Exercise: Off the MCMC Shoreline

- ▶ Fit a model with a $\text{NAP} \times \text{angle1}$ interaction and random effect of beach
- ▶ Evaluate the model and whether it is fit well
- ▶ Compare the coefficients to a model with a strong prior that the interaction is -5.