An introduction to Bayesian statistical inference

S. Robin

JC(2)BIM, June 2018, Fréjus

Outline

Outline

An example

Example:

- n patients: $i = 1 \dots n$
- Y_i = status (0 = healthy, 1 = sick) of patient i
- $\mathbf{x}_i = (x_{i1}, \dots x_{ip})$ = vector of gene expression for patient i (gene $j = 1 \dots p$)

Dataset: $n = 78$, $p = 15$

	AB033066	NM003056	NM000903	...	Status
1	0.178	0.116	0.22		0
2	0.065	-0.073	-0.014		0
3	-0.077	0.03	0.043		0
4	0.176	-0.041	0.362		0
5	-0.089	-0.164	-0.266		0

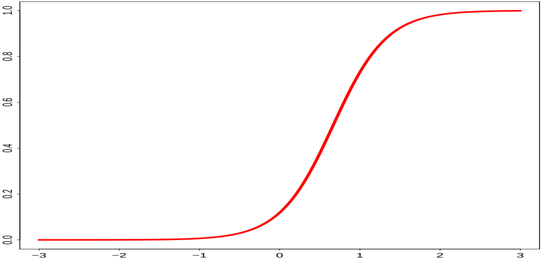
A statistical model

Logistic regression Logistic regression

- The patients are independent.
- The probability for patient i to be sick depends on $\mathbf{x}_i$:

$$\Pr\{Y_i = 1\} = \frac{e^{\mathbf{x}_i^\intercal \boldsymbol{\theta}}}{1 + e^{\mathbf{x}_i^\intercal \boldsymbol{\theta}}}, \qquad \mathbf{x}_i^\intercal \boldsymbol{\theta} = \sum_{j=1}^{p} x_{ij}\theta_j$$

- $\boldsymbol{\theta} = (\theta_1, \dots \theta_p)$: unknown parameter (regression coefficients, incl. intercept)



Frequentist inference

$\boldsymbol{\theta}$ = fixed parameter:

- Statistical model:

$$\mathbf{Y} \sim p_{\boldsymbol{\theta}}$$

- Inference: get a (point) estimate $\widehat{\boldsymbol{\theta}}$ e.g.

$$\widehat{\boldsymbol{\theta}}: \qquad \log p_{\widehat{\boldsymbol{\theta}}}(\mathbf{Y}) = \max_{\boldsymbol{\theta}} \log p_{\boldsymbol{\theta}}(\mathbf{Y})$$

- The estimate $\widehat{\boldsymbol{\theta}}$ itself is random (depends on the data) $\rightarrow$ confidence interval, tests, ...

Frequentist inference

$\boldsymbol{\theta}$ = fixed parameter:

- Statistical model:

$$\mathbf{Y} \sim p_{\boldsymbol{\theta}}$$

- Inference: get a (point) estimate $\widehat{\boldsymbol{\theta}}$ e.g.

$$\widehat{\boldsymbol{\theta}} : \qquad \log p_{\widehat{\boldsymbol{\theta}}}(\mathbf{Y}) = \max_{\boldsymbol{\theta}} \log p_{\boldsymbol{\theta}}(\mathbf{Y})$$

- The estimate $\widehat{\boldsymbol{\theta}}$ itself is random (depends on the data) $\rightarrow$ confidence interval, tests, ...

Output: `GLM = glm(Y ~ X, family=binomial)`

	Estimate	Std. Error	z value	Pr(> |z|)
(Intercept)	-0.7212697	0.6512707	-1.107481	0.2680861
XAB033066	7.23375	2.505118	2.887589	0.003882068
XNM003056	-0.6116423	1.854695	-0.3297806	0.7415658
XNM000903	1.732625	1.199888	1.443988	0.1487423

...

Bayesian inference

$\boldsymbol{\theta}$ = random parameter:

- Statistical model:

$$p(\mathbf{Y} \mid \boldsymbol{\theta}) \qquad (= \textit{likelihood})$$

- Inference: provide the conditional distribution of $\boldsymbol{\theta}$ given the observed data $\mathbf{Y}$:

$$p(\boldsymbol{\theta} \mid \mathbf{Y}) \qquad (= \textit{posterior} \text{ distribution})$$

$\rightarrow$ credibility intervals

- Requires to define a marginal distribution:

$$p(\boldsymbol{\theta}) \qquad (= \textit{prior} \text{ distribution})$$

Why 'Bayes'

Bayes formula:

$$P(A \mid B) = \frac{P(A, B)}{P(B)} = \frac{P(A)}{P(B)} P(B \mid A)$$

- $P(B)$ = marginal probability of B
- $P(A, B)$ = joint probability of A and B
- $P(A \mid B)$ = conditional probability of A given B [1]

[1] see Appendix # 35

Why 'Bayes'

Bayes formula:

$$P(A \mid B) = \frac{P(A,B)}{P(B)} = \frac{P(A)}{P(B)} P(B \mid A)$$

- $P(B)$ = marginal probability of B
- $P(A,B)$ = joint probability of A and B
- $P(A \mid B)$ = conditional probability of A given B [1]

Be careful. Many methods, e.g.

Bayesian network, Naive Bayes, ...

- use conditional probabilities
- but have nothing to do with Bayesian inference (in the statistical sense)

[1] see Appendix # 35

Bayes formula for Bayesian inference

Posterior distribution.

$$p(\boldsymbol{\theta} \mid \mathbf{Y}) = \frac{p(\mathbf{Y}, \boldsymbol{\theta})}{p(\mathbf{Y})} = \frac{\overbrace{p(\boldsymbol{\theta})}^{\textit{prior}} \, \overbrace{p(\mathbf{Y} \mid \boldsymbol{\theta})}^{\textit{likelihood}}}{p(\mathbf{Y})}$$

$\rightarrow$ Requires to evaluate the *integrated likelihood* (i.e. marginal)

$$p(\mathbf{Y}) = \int p(\boldsymbol{\theta}) p(\mathbf{Y} \mid \boldsymbol{\theta}) \, \mathrm{d}\boldsymbol{\theta},$$

which act as the normalizing constant of the posterior $p(\boldsymbol{\theta} \mid \mathbf{Y})$.

The posterior depends on the prior

Data & Model:

- $Y_i = 1$ if sick, 0 otherwise
- $n = 10$ patients
- $\vdots$: number sicks $/n$

The posterior depends on the prior

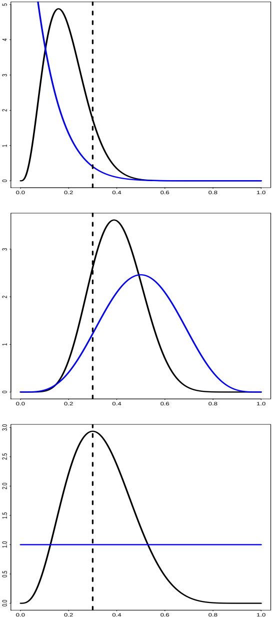
Data & Model:

- $Y_i = 1$ if sick, 0 otherwise
- $n = 10$ patients
- $\vdots$: number sicks $/n$

Param:

- $\theta =$ proba. sick
- — : prior $p(\theta)$

The posterior depends on the prior



Data & Model:

- $Y_i = 1$ if sick, 0 otherwise
- $n = 10$ patients
- $\vdots$: number sicks $/n$

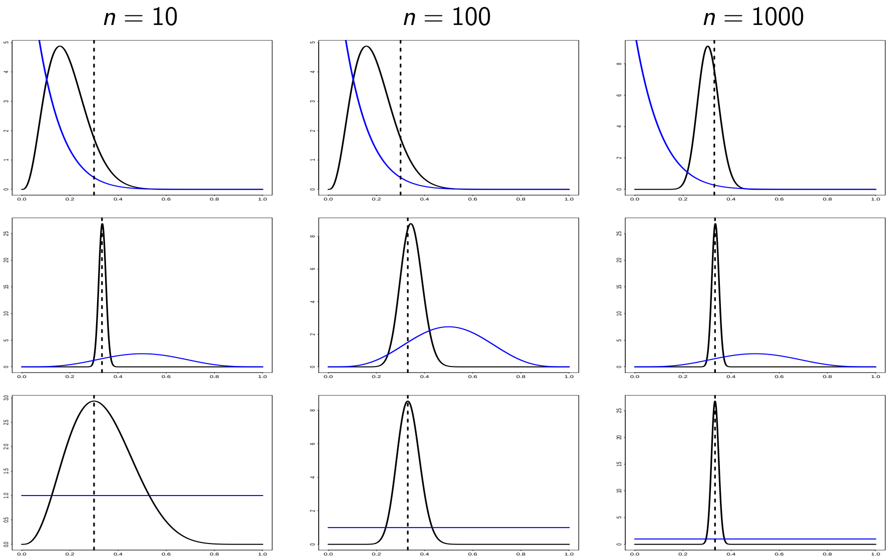
Param:

- $\theta =$ proba. sick
- — : prior $p(\theta)$

Output:

- — : posterior $p(\theta \mid \mathbf{Y})$

Dependency vanishes when n increase



Back to logistic regression

Model

- Prior: all coefficient θ_j independent:

$$\theta_j \sim \mathcal{N}(0, 100)$$

- Likelihood: all patients independent, *conditionally* on $\boldsymbol{\theta}$:

$$\Pr\{Y_i = 1 \mid \boldsymbol{\theta}\} = e^{\mathbf{x}_i^\intercal \boldsymbol{\theta}} \Big/ \left(1 + e^{\mathbf{x}_i^\intercal \boldsymbol{\theta}}\right)$$

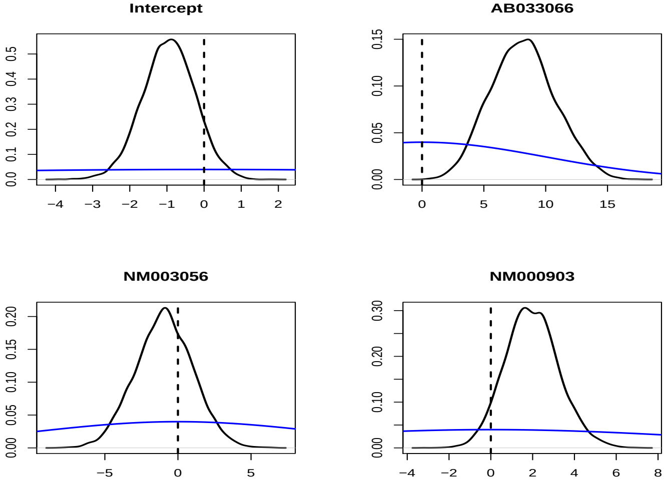
Inference:

$$\boldsymbol{\theta} \mid \mathbf{Y} \sim ?$$

(see later, but $p(\boldsymbol{\theta} \mid \mathbf{Y}) \neq \mathcal{N}(\cdot, \cdot)$, for sure.)

Bayesian inference

Output:



Posterior distribution and confidence intervals

Parameter 'estimate'.

$$\text{posterior mean:} \qquad \widehat{\theta}_j = \mathbb{E}(\theta_j \mid \mathbf{Y})$$

$$\text{posterior mode:} \qquad \widehat{\theta}_j = \arg\max_{\theta_j} \; p(\theta_j \mid \mathbf{Y})$$

Credibility interval (CI). With level $1-\alpha$ (e.g. 95%):

$$CI_{1-\alpha}(\theta_j) = [\theta_j^\ell; \theta_j^u] : \qquad \Pr\{\theta_j^\ell < \theta_j < \theta_j^u \mid \mathbf{Y}\} = 1-\alpha$$

Posterior distribution and confidence intervals

Parameter 'estimate'.

$$\text{posterior mean:} \qquad \widehat{\theta}_j = \mathbb{E}(\theta_j \mid \mathbf{Y})$$

$$\text{posterior mode:} \qquad \widehat{\theta}_j = \arg\max_{\theta_j} \; p(\theta_j \mid \mathbf{Y})$$

Credibility interval (CI). With level $1 - \alpha$ (e.g. 95%):

$$CI_{1-\alpha}(\theta_j) = [\theta_j^\ell; \theta_j^u] : \qquad \Pr\{\theta_j^\ell < \theta_j < \theta_j^u \mid \mathbf{Y}\} = 1 - \alpha$$

Example.

	post.mean	post.mode	lower.CI	upper.CI
Intercept	-0.9816181	-0.8652166	-2.41342	0.3704264
AB033066	8.395169	8.587083	3.271464	13.98373
NM003056	-1.042483	-0.9854149	-5.046639	2.819764
NM000903	1.911312	1.677234	-0.4240319	4.452512

Accounting for uncertainty

Question: What is the probability for patient 0 (with profile $\mathbf{x}_0$) to be sick?

Model answer:

$$\Pr\{Y_0 = 1 \mid \boldsymbol{\theta}\} = e^{\mathbf{x}_0^\intercal \boldsymbol{\theta}} \Big/ \left(1 + e^{\mathbf{x}_0^\intercal \boldsymbol{\theta}}\right)$$

but $\boldsymbol{\theta}$ is unknown (and random).

Bayesian answer: *posterior predictive* probability

$$\Pr\{Y_0 = 1 \mid \mathbf{Y}\} = \int \Pr\{Y_0 = 1 \mid \boldsymbol{\theta}\} p(\boldsymbol{\theta} \mid \mathbf{Y}) \, \mathrm{d}\boldsymbol{\theta}$$

Model comparison (1/2)

Problem. Which model fits the data better:

$$
\begin{aligned}
M_0 &: \text{none of the genes has an effect, i.e. } \boldsymbol{\theta} = (\theta_0, 0, \dots, 0) \\
M_1 &: \text{only the fist gene has an effect, i.e. } \boldsymbol{\theta} = (\theta_0, \theta_1, 0, \dots, 0) \\
&\dots \\
M_p &: \text{all genes have an effect, i.e. } \boldsymbol{\theta} = (\theta_0, \theta_1, \dots, \theta_p)
\end{aligned}
$$

Bayesian model comparison. For each model $M \in \mathcal{M} = \{M_0, \dots, M_p\}$, evaluate

$$p(M \mid \mathbf{Y})$$

Model comparison (2/2)

Ingredients:

- Prior on the models: $p(M)$, e.g.

$$p(M) = \text{cst} \qquad \text{(uniform prior)}$$

- Conditional prior on the parameters: $p(\boldsymbol{\theta} \mid M)$, e.g.

$$\theta_j \mid M_k \begin{cases} \sim & \mathcal{N}(0, 100) \quad \text{if } j \leq k \\ = & 0 \quad\quad\quad\quad\;\; \text{otherwise} \end{cases}$$

Model comparison (2/2)

Ingredients:

- Prior on the models: $p(M)$, e.g.

$$p(M) = \text{cst} \qquad \text{(uniform prior)}$$

- Conditional prior on the parameters: $p(\boldsymbol{\theta} \mid M)$, e.g.

$$\theta_j \mid M_k \left\{ \begin{array}{lll} \sim & \mathcal{N}(0, 100) & \text{if } j \leq k \\ = & 0 & \text{otherwise} \end{array} \right.$$

Recipe:

- Evaluate the marginal likelihood of the data for each model M:

$$p(\mathbf{Y} \mid M) = \int p(\mathbf{Y} \mid \boldsymbol{\theta}) p(\boldsymbol{\theta} \mid M) \, \mathrm{d}\boldsymbol{\theta}$$

Model comparison (2/2)

Ingredients:

- Prior on the models: $p(M)$, e.g.

$$p(M) = \text{cst} \qquad \text{(uniform prior)}$$

- Conditional prior on the parameters: $p(\boldsymbol{\theta} \mid M)$, e.g.

$$\theta_j \mid M_k \begin{cases} \sim & \mathcal{N}(0, 100) \quad \text{if } j \leq k \\ = & 0 \qquad\qquad\quad \text{otherwise} \end{cases}$$

Recipe:

- Evaluate the marginal likelihood of the data for each model M:

$$p(\mathbf{Y} \mid M) = \int p(\mathbf{Y} \mid \boldsymbol{\theta}) p(\boldsymbol{\theta} \mid M) \, \mathrm{d}\boldsymbol{\theta}$$

- Evaluate the $p(M_k \mid \mathbf{Y})$ using Bayes rule

$$p(M_k \mid \mathbf{Y}) = \frac{p(M_k) p(\mathbf{Y} \mid M_k)}{p(\mathbf{Y})} = \frac{p(M_k) p(\mathbf{Y} \mid M_k)}{\sum_{k'} p(M_{k'}) p(\mathbf{Y} \mid M_{k'})}$$

Model averaging (uncertainty on models)

Question: Probability for patient 0 to be sick?

Model averaging (uncertainty on models)

Question: Probability for patient 0 to be sick?

Model selection.

- Select the 'best' model $\widehat{M}$, i.e. with largest posterior $p(M \mid \mathbf{Y})$
- Compute

$$\Pr\{Y_0 = 1 \mid \mathbf{Y}, \widehat{M}\} = \int \Pr\{Y_0 = 1 \mid \boldsymbol{\theta}\} p(\boldsymbol{\theta} \mid \mathbf{Y}, \widehat{M}) \, \mathrm{d}\boldsymbol{\theta}$$

Model averaging (uncertainty on models)

Question: Probability for patient 0 to be sick?

Model selection.

- Select the 'best' model $\widehat{M}$, i.e. with largest posterior $p(M \mid \mathbf{Y})$
- Compute

$$\Pr\{Y_0 = 1 \mid \mathbf{Y}, \widehat{M}\} = \int \Pr\{Y_0 = 1 \mid \boldsymbol{\theta}\} p(\boldsymbol{\theta} \mid \mathbf{Y}, \widehat{M}) \, \mathrm{d}\boldsymbol{\theta}$$

Model averaging.

- Keep all models
- Compute

$$\Pr\{Y_0 = 1 \mid \mathbf{Y}\} = \sum_M \Pr\{Y_0 = 1 \mid \mathbf{Y}, M\} p(M \mid \mathbf{Y})$$

Transfer of uncertainty from one experience to another

Combining samples. Consider two independent but similar datasets $\mathbf{Y}_1$ and $\mathbf{Y}_2$.

Simple algebra gives:

$$p(\boldsymbol{\theta} \mid \mathbf{Y}_1, \mathbf{Y}_2) = \frac{p(\boldsymbol{\theta} \mid \mathbf{Y}_1) p(\mathbf{Y}_2 \mid \boldsymbol{\theta}, \mathbf{Y}_1)}{p(\mathbf{Y}_2 \mid \mathbf{Y}_1)}$$

Transfer of uncertainty from one experience to another

Combining samples. Consider two independent but similar datasets $\mathbf{Y}_1$ and $\mathbf{Y}_2$.

Simple algebra gives:

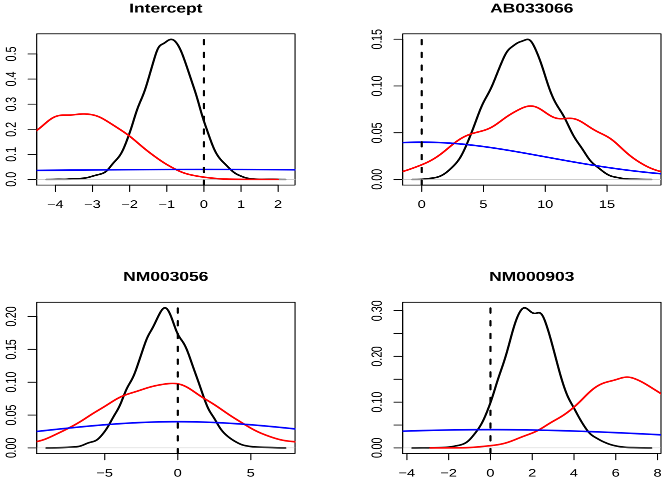
$$p(\boldsymbol{\theta} \mid \mathbf{Y}_1, \mathbf{Y}_2) = \frac{p(\boldsymbol{\theta} \mid \mathbf{Y}_1) p(\mathbf{Y}_2 \mid \boldsymbol{\theta}, \mathbf{Y}_1)}{p(\mathbf{Y}_2 \mid \mathbf{Y}_1)}$$

In practice:

1. Perform inference using $\mathbf{Y}_1$ to get $p(\boldsymbol{\theta} \mid \mathbf{Y}_1)$ from prior $p(\boldsymbol{\theta})$
2. Then perform inference using $\mathbf{Y}_2$ to get $p(\boldsymbol{\theta} \mid \mathbf{Y}_1, \mathbf{Y}_2)$ using $p(\boldsymbol{\theta} \mid \mathbf{Y}_1)$ as a prior

Combining experiments

Output: $n_1 = n_2 = 39$



Outline

Posterior distribution

Aim: Evaluate

$$E[f(\boldsymbol{\theta})|\mathbf{Y}]$$

- Posterior mean: $f(\boldsymbol{\theta}) = \theta_j$
- Credibility interval: $f(\boldsymbol{\theta}) = \mathbb{I}\{\theta_j^\ell < \theta_j < \theta_j^u\}$

Posterior distribution

Aim: Evaluate

$$E[f(\boldsymbol{\theta})|\mathbf{Y}]$$

- Posterior mean: $f(\boldsymbol{\theta}) = \theta_j$
- Credibility interval: $f(\boldsymbol{\theta}) = \mathbb{I}\{\theta_j^\ell < \theta_j < \theta_j^u\}$

Main problem: evaluate

$$p(\boldsymbol{\theta} \mid \mathbf{Y}) = \frac{p(\boldsymbol{\theta})\mathbf{p}(\mathbf{Y} \mid \boldsymbol{\theta})}{p(\mathbf{Y})}$$

which requires to evaluate

$$p(\mathbf{Y}) = \int \underbrace{p(\boldsymbol{\theta})}_{prior} \underbrace{p(\mathbf{Y}|\boldsymbol{\theta})}_{likelihood} \mathrm{d}\boldsymbol{\theta}$$

Nice case: Conjugate priors

Example: Bernoulli

Prior: $\theta =$ proba. to be sick.

$$\theta \sim \text{Beta}(a, b), \qquad p(\theta) \propto \theta^{a-1}(1-\theta)^{b-1}$$

Nice case: Conjugate priors

Example: Bernoulli

Prior: $\theta =$ proba. to be sick.

$$\theta \sim \text{Beta}(a, b), \qquad p(\theta) \propto \theta^{a-1}(1-\theta)^{b-1}$$

Likelihood: $Y_i = 1$ if sick, 0 otherwise. $S =$ number of sick

$$Y_i \mid \theta \sim \mathcal{B}(\theta), \qquad p(\mathbf{Y} \mid \theta) = \prod_i \theta^{Y_i}(1-\theta)^{1-Y_i} = \theta^S(1-\theta)^{n-S}$$

Nice case: Conjugate priors

Example: Bernoulli

Prior: $\theta =$ proba. to be sick.

$$\theta \sim \text{Beta}(a, b), \qquad p(\theta) \propto \theta^{a-1}(1-\theta)^{b-1}$$

Likelihood: $Y_i = 1$ if sick, 0 otherwise. $S =$ number of sick

$$Y_i \mid \theta \sim \mathcal{B}(\theta), \qquad p(\mathbf{Y} \mid \theta) = \prod_i \theta^{Y_i}(1-\theta)^{1-Y_i} = \theta^S(1-\theta)^{n-S}$$

Posterior:

$$p(\theta \mid \mathbf{Y}) \propto p(\theta)p(\mathbf{Y} \mid \theta) = \theta^{a+S-1}(1-\theta)^{b+n-S-1}$$

which means that

$$\theta \mid \mathbf{Y} \sim \text{Beta}(a+S, b+n-S)$$

Conjugate priors: Discrete distributions

Likelihood	Model parameters	Conjugate prior distribution	Prior hyperparameters	Posterior hyperparameters	Interpretation of hyperparameters[note 1]	Posterior predictive[note 2]
Bernoulli	p (probability)	Beta	$\alpha,\ \beta$	$\alpha+\sum_{i=1}^n x_i,\ \beta+n-\sum_{i=1}^n x_i$	$\alpha-1$ successes, $\beta-1$ failures[note 1]	$p(\tilde{x}=1)=\frac{\alpha'}{\alpha'+\beta'}$
Binomial	p (probability)	Beta	$\alpha,\ \beta$	$\alpha+\sum_{i=1}^n x_i,\ \beta+\sum_{i=1}^n N_i-\sum_{i=1}^n x_i$	$\alpha-1$ successes, $\beta-1$ failures[note 1]	$\operatorname{BetaBin}(\tilde{x}\mid\alpha',\beta')$ (beta-binomial)
Negative Binomial with known failure number r	p (probability)	Beta	$\alpha,\ \beta$	$\alpha+\sum_{i=1}^n x_i,\ \beta+rn$	$\alpha-1$ total successes, $\beta-1$ failures[note 1] (i.e. $\frac{\beta-1}{r}$ experiments, assuming r stays fixed)	
Poisson	λ (rate)	Gamma	$k,\ \theta$	$k+\sum_{i=1}^n x_i,\ \frac{\theta}{n\theta+1}$	k total occurrences in $1/\theta$ intervals	$\operatorname{NB}(\tilde{x}\mid k',\frac{\theta'}{1+\theta'})$ (negative binomial)
Poisson	λ (rate)	Gamma	$\alpha,\ \beta$ [note 3]	$\alpha+\sum_{i=1}^n x_i,\ \beta+n$	α total occurrences in β intervals	$\operatorname{NB}(\tilde{x}\mid\alpha',\frac{1}{1+\beta'})$ (negative binomial)
Categorical	$\boldsymbol{p}$ (probability vector), k (number of categories, i.e. size of $\boldsymbol{p}$)	Dirichlet	$\boldsymbol{\alpha}$	$\boldsymbol{\alpha}+(c_1,\ldots,c_k)$, where c_i is the number of observations in category i	α_i-1 occurrences of category i[note 1]	$p(\tilde{x}=i)=\frac{\alpha_i'}{\sum_i\alpha_i'}$ $=\frac{\alpha_i+c_i}{\sum_i\alpha_i+n}$
Multinomial	$\boldsymbol{p}$ (probability vector), k (number of categories, i.e. size of $\boldsymbol{p}$)	Dirichlet	$\boldsymbol{\alpha}$	$\boldsymbol{\alpha}+\sum_{i=1}^n \mathbf{x}_i$	α_i-1 occurrences of category i[note 1]	$\operatorname{DirMult}(\tilde{\mathbf{x}}\mid\boldsymbol{\alpha}')$ (Dirichlet-multinomial)
Hypergeometric with known total population size N	M (number of target members)	Beta-binomial[4]	$n=N,\alpha,\ \beta$	$\alpha+\sum_{i=1}^n x_i,\ \beta+\sum_{i=1}^n N_i-\sum_{i=1}^n x_i$	$\alpha-1$ successes, $\beta-1$ failures[note 1]	
Geometric	p_0 (probability)	Beta	$\alpha,\ \beta$	$\alpha+n,\ \beta+\sum_{i=1}^n x_i$	$\alpha-1$ experiments, $\beta-1$ total failures[note 1]	

`en.wikipedia.org/wiki/Conjugate_prior`

Conjugate priors: Continuous distributions

Likelihood	Model parameters	Conjugate prior distribution	Prior hyperparameters	Posterior hyperparameters	Interpretation of hyperparameters	Posterior predictive[note 4]
Normal with known variance σ^2	μ (mean)	Normal	$\mu_0,\ \sigma_0^2$	$\left(\frac{\mu_0}{\sigma_0^2}+\frac{\sum_{i=1}^n x_i}{\sigma^2}\right)\Big/\left(\frac{1}{\sigma_0^2}+\frac{n}{\sigma^2}\right)$, $\left(\frac{1}{\sigma_0^2}+\frac{n}{\sigma^2}\right)^{-1}$	mean was estimated from observations with total precision (sum of all individual precisions) $1/\sigma_0^2$ and with sample mean μ_0	$\mathcal{N}(\tilde{x}\mid\mu_0', \sigma_0^{2\prime}+\sigma^2)$[5]
Normal with known precision τ	μ (mean)	Normal	$\mu_0,\ \tau_0$	$\left(\tau_0\mu_0+\tau\sum_{i=1}^n x_i\right)\Big/(\tau_0+n\tau),\ \tau_0+n\tau$	mean was estimated from observations with total precision (sum of all individual precisions) τ_0 and with sample mean μ_0	$\mathcal{N}\left(\tilde{x}\mid\mu_0', \frac{1}{\tau_0'}+\frac{1}{\tau}\right)$[5]
Normal with known mean μ	σ^2 (variance)	Inverse gamma	$\alpha,\ \beta$ [note 5]	$\alpha+\frac{n}{2},\ \beta+\frac{\sum_{i=1}^n (x_i-\mu)^2}{2}$	variance was estimated from 2α observations with sample variance β/α (i.e. with sum of squared deviations 2β, where deviations are from known mean μ)	$t_{2\alpha'}(\tilde{x}\mid\mu, \sigma^2=\beta'/\alpha')$[5]
Normal with known mean μ	σ^2 (variance)	Scaled inverse chi-squared	$\nu,\ \sigma_0^2$	$\nu+n,\ \frac{\nu\sigma_0^2+\sum_{i=1}^n (x_i-\mu)^2}{\nu+n}$	variance was estimated from ν observations with sample variance σ_0^2	$t_{\nu'}(\tilde{x}\mid\mu, \sigma_0^{2\prime})$[5]
Normal with known mean μ	τ (precision)	Gamma	$\alpha,\ \beta$[note 3]	$\alpha+\frac{n}{2},\ \beta+\frac{\sum_{i=1}^n (x_i-\mu)^2}{2}$	precision was estimated from 2α observations with sample variance β/α (i.e. with sum of squared deviations 2β, where deviations are from known mean μ)	$t_{2\alpha'}(\tilde{x}\mid\mu, \sigma^2=\beta'/\alpha')$[5]
Normal[note 6]	μ and σ^2 Assuming exchangeability	Normal-inverse gamma	$\mu_0,\ \nu,\ \alpha,\ \beta$	$\frac{\nu\mu_0+n\bar{x}}{\nu+n},\ \nu+n,\ \alpha+\frac{n}{2},$ $\beta+\frac{1}{2}\sum_{i=1}^n (x_i-\bar{x})^2+\frac{n\nu}{\nu+n}\frac{(\bar{x}-\mu_0)^2}{2}$ ▪ $\bar{x}$ is the sample mean	mean was estimated from ν observations with sample mean μ_0; variance was estimated from 2α observations with sample mean μ_0 and sum of squared deviations 2β	$t_{2\alpha'}\left(\tilde{x}\mid\mu', \frac{\beta'(\nu'+1)}{\alpha'\nu'}\right)$[5]
Normal	μ and τ Assuming exchangeability	Normal-gamma	$\mu_0,\ \nu,\ \alpha,\ \beta$	$\frac{\nu\mu_0+n\bar{x}}{\nu+n},\ \nu+n,\ \alpha+\frac{n}{2},$ $\beta+\frac{1}{2}\sum_{i=1}^n (x_i-\bar{x})^2+\frac{n\nu}{\nu+n}\frac{(\bar{x}-\mu_0)^2}{2}$ ▪ $\bar{x}$ is the sample mean	mean was estimated from ν observations with sample mean μ_0, and precision was estimated from 2α observations with sample mean μ_0 and sum of squared deviations 2β	$t_{2\alpha'}\left(\tilde{x}\mid\mu', \frac{\beta'(\nu'+1)}{\alpha'\nu'}\right)$[5]
Multivariate normal with known covariance matrix $\boldsymbol{\Sigma}$	$\boldsymbol{\mu}$ (mean vector)	Multivariate normal	$\boldsymbol{\mu}_0,\ \boldsymbol{\Sigma}_0$	$\left(\boldsymbol{\Sigma}_0^{-1}+n\boldsymbol{\Sigma}^{-1}\right)^{-1}\left(\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\mu}_0+n\boldsymbol{\Sigma}^{-1}\bar{\mathbf{x}}\right)$, $\left(\boldsymbol{\Sigma}_0^{-1}+n\boldsymbol{\Sigma}^{-1}\right)^{-1}$ ▪ $\bar{\mathbf{x}}$ is the sample mean	mean was estimated from observations with total precision (sum of all individual precisions) $\boldsymbol{\Sigma}_0^{-1}$ and with sample mean $\boldsymbol{\mu}_0$	$\mathcal{N}(\tilde{\mathbf{x}}\mid\boldsymbol{\mu}_0', \boldsymbol{\Sigma}_0'+\boldsymbol{\Sigma})$[5]

en.wikipedia.org/wiki/Conjugate_prior

Computing integrals

General case: $p(\boldsymbol{\theta} \mid \mathbf{Y})$ has no close form

Goal: compute

$$\mathbb{E}(f(\boldsymbol{\theta}) \mid \mathbf{Y}) = \int f(\boldsymbol{\theta}) p(\boldsymbol{\theta} \mid \mathbf{Y}) \, \mathrm{d}\boldsymbol{\theta} = \int f(\boldsymbol{\theta}) p(\boldsymbol{\theta}) \, p(\mathbf{Y} \mid \boldsymbol{\theta}) \, \mathrm{d}\boldsymbol{\theta} \Big/ p(\mathbf{Y})$$

where

$$p(\mathbf{Y}) = \int p(\boldsymbol{\theta}) \, p(\mathbf{Y} \mid \boldsymbol{\theta}) \, \mathrm{d}\boldsymbol{\theta}$$

Monte-Carlo

Importance sampling

MCMC principle (ergodic chain with stationary distribution $p(\theta|Y)$)

Metropolis-Hastings

Gibbs

Outline

References

Outline

Joint, Marginal, Conditional

Reminder: 2 loci with 2 alleles each: (A, a), (B, b)

- Joint distribution:

	B	b	marginal
A	f_{AB}	f_{Ab}	$p_A = f_{AB} + f_{Ab}$
a	f_{aB}	f_{ab}	$p_a = f_{aB} + f_{ab}$
marginal	$q_B = f_{AB} + f_{aB}$	$q_b = f_{Ab} + f_{ab}$	1

- Marginal distribution: 'integrates out' the allele of the other locus

$$\Pr\{B\} = q_B = f_{AB} + f_{aB}$$

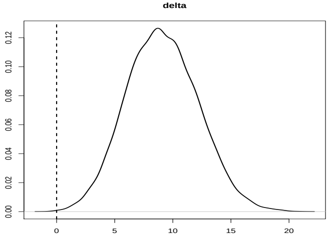
- Conditional distribution

$$\Pr\{A \mid b\} = \frac{\Pr\{A, b\}}{\Pr\{b\}} = \frac{f_{Ab}}{q_b} = \frac{f_{Ab}}{f_{Ab} + f_{ab}}$$

Posterior distribution and CI (slide # 14)

The same holds for combination of parameters, e.g.

$$\delta = \theta_2 - \theta_3$$



	post.mean	post.mode	lower.CI	upper.CI
delta	9.389825	8.824822	3.539045	15.84367