Foundations of Statistical Inference

Julien Berestycki

Department of Statistics
University of Oxford

MT 2015

Lecture 16 : Bayesian analysis of contingency tables. Bayesian linear regression.

Example 2×2

From Wikipedia article on contingency tables:

	Left-handed	Right-handed	Total
Male	9 (y_1)	43	52
Female	4 (y_2)	44	48
Total	13	87	100

Hypothesis: $\theta_1 =$ Proportion of left-handed men $> \theta_2 =$ proportion of left handed women.
Model: $y_1 \sim \text{Binom}(n_1, \theta_1)$, $y_2 \sim \text{Binom}(n_2, \theta_2)$.
Use uniform priors $\theta_i \sim U_{[0,1]} = \text{Beta}(1,1)$.
Posteriors

$$p(\theta_1|y_1, n_1) = \text{Beta}(y_1+1, n_1-y_1+1), p(\theta_2|y_2, n_2) = \text{Beta}(y_2+1, n_2-y_2+1).$$

Then compute posterior $P(Z_1 > Z_2)$ Either compute an integral or by simulation.

Example 2×2: simulations

See R code. Generate M sample from joint posterior

$$p(\theta_1, \theta_2|y_1, n_1, y_2, n_2) = p(\theta_1|y_1, n_1) \times p(\theta_2|y_2, n_2)$$

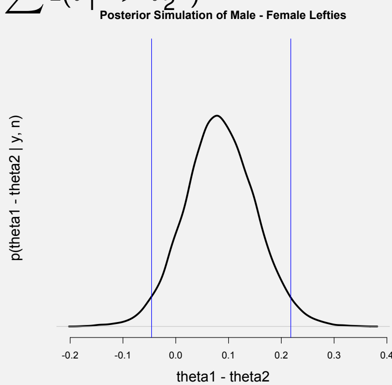
and then use Monte-Carlo approximation

$$P[\theta_1 > \theta_2] \approx \frac{1}{M} \sum \mathbb{I}(\theta_1^{(i)} > \theta_2^{(i)})$$

Outputs M=10000 :

```
  2.5%    50%   97.5%
-0.046  0.083  0.218

print(mean(theta1>theta2))
[1] 0.8997
```



Contingency table analysis

North Carolina State University data.

EC : Extra Curricular activities in hours per week.

	EC		
	< 2	2 to 12	> 12
C or better	11	68	3
D or F	9	23	5

Let $y = (y_{ij})$ be the matrix of counts.

Frequentist analysis

Usual χ^2 test from R.

Pearson's Chi-squared test

$y_{i,j}$ is cardinal of cell i, j

	< 2	2 to 12	> 12
C or better	11	68	3
D or F	9	23	5

Sum rows and columns

	< 2	2 to 12	> 12	total
C or better	11	68	3	82
D or F	9	23	5	37
total	20	91	8	119

$E_{i,j} = r_i c_j / N$

	< 2	2 to 12	> 12	total
D or F	6.22	28.3	2.49	37
total	20	91	8	119

Bayesian analysis

	EC		
	< 2	2 to 12	> 12
C or better	p_{11}	p_{12}	p_{13}
D or F	p_{21}	p_{22}	p_{23}

Let $\mathbf{p} = \{p_{11}, \dots, p_{23}\}$. The model is that $Y = (y_{1,1}, \dots, y_{2,3})$ is a multinomial $(N, \mathbf{p})$ (i.e. N trials with $P(X_k = (i,j)) = p_{i,j}$ and $y_{i,j} = \#\{X_k = (i,j)\}$.)

Bayesian method: make $\mathbf{p}$ a variable.
Consider two models

- M_I the two categorical variables are independent $(p_{11}, p_{12}, p_{13}) \propto (p_{21}, p_{22}, p_{23})$
- M_D the two categorical variables are dependent. $(p_{11}, p_{12}, p_{13}) \not\propto (p_{21}, p_{22}, p_{23})$

The Bayes factor is $BF = \frac{P(y|M_D)}{\dots}$

The Dirichlet distribution

Dirichlet integral

$$\int_{z_1+\cdots+z_k=1} z_1^{\nu_1-1} \cdots z_k^{\nu_k-1} dz_1 \cdots dz_k = \frac{\Gamma(\nu_1)\cdots\Gamma(\nu_k)}{\Gamma(\sum \nu_i)}$$

Dirichlet distribution

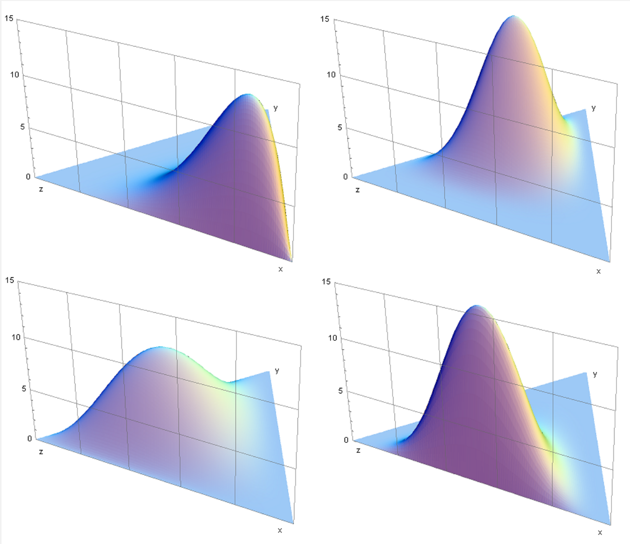
$$\frac{\Gamma(\sum \nu_i)}{\Gamma(\nu_1)\cdots\Gamma(\nu_k)} z_1^{\nu_1-1} \cdots z_k^{\nu_k-1}, \quad z_1 + \cdots + z_k = 1$$

The means are $\mathbb{E}[Z_i] = \nu_i / \sum \nu_i$, $i = 1, \dots, k$.

A representation that makes the Dirichlet easy to simulate from is the following.

Let $W_1, \dots, W_k$ be independent Gamma (ν_1, θ), ... Gamma (ν_k, θ) random variables, $W = \sum W_i$ and set $Z_i = W_i / W$, $i = 1, \dots, k$. (Does not depend on θ).

Examples of 3D Dirichlet distributions



Calculating marginal likelihoods

The model is that $f(y, \mathbf{p})$ is a

$$P(y|M_D) = \int_{\mathbf{p}} P(y|\mathbf{p})\pi(\mathbf{p})d\mathbf{p}$$
$$= \int_{p_{11}+\cdots+p_{23}=1} \binom{|y|}{y} \prod_{ij} (p_{ij})^{y_{ij}} \pi(\mathbf{p}) dp_{11}\cdots dp_{23}$$

where

$$\binom{|y|}{y} = (\sum y_{ij})!/\prod y_{ij}!$$

Under M_D $(p_{1,1}, p_{1,2}, p_{1,3}) \not\propto (p_{2,1}, p_{2,2}, p_{2,3})$ so choose a uniform distribution for $\mathbf{p}$ i.e. Dirichlet$(1, \dots, 1)$

$$\Rightarrow \pi(\mathbf{p}) = \Gamma(RC), \quad p_{11} + \cdots + p_{23} = 1$$

Calculating marginal likelihoods

$$P(y|M_D) = \binom{|y|}{y}\Gamma(RC)\int_{p_{11}+\cdots+p_{23}=1} \prod_{ij} (p_{ij})^{y_{ij}} dp_{11}\cdots dp_{23}$$
$$= \binom{|y|}{y}\Gamma(RC)\frac{\prod \Gamma(y_{ij}+1)}{\Gamma(|y|+RC)}$$
$$= \binom{|y|}{y}\frac{D(y+1)}{D(1_{RC})}$$

where

$$D(\nu) = \prod \Gamma(\nu_i)/\Gamma(\sum \nu_i)$$

and $y+1$ denotes the matrix of counts with 1 added to all entries and 1_{RC} denotes a vector of length RC with all entries equal to 1.

Calculating marginal likelihoods

Under M_I the probabilities are determined by the marginal probabilities $p_r = \{p_{1\cdot}, p_{2,\cdot}\}$ and $p_c = \{p_{\cdot 1}, p_{\cdot 2}, p_{\cdot 3}\}$

	< 2	2 to 12	> 12	
C or better	p_{11}	p_{12}	p_{13}	$p_{1\cdot}$
D or F	p_{21}	p_{22}	p_{23}	$p_{2\cdot}$
	$p_{\cdot 1}$	$p_{\cdot 2}$	$p_{\cdot 3}$	

Under M_I we have a table where $p_{ij} = p_{i\cdot}p_{\cdot j}$.
Under independence M_I the prior for the row sums and column sums are independent uniform priors: Dirichlet distribution (with R=2 and C = 3 respectively)

$$\pi(p_r) = \frac{\Gamma(R)}{\prod \Gamma(1)} p_{1\cdot}^{1-1}\cdots p_{R\cdot}^{1-1} = \Gamma(R), \; \pi(p_c) = \frac{\Gamma(C)}{\prod \Gamma(1)} p_{\cdot 1}^{1-1}\cdots p_{\cdot C}^{1-1} = \Gamma(C)$$

The marginal likelihood under M_I is therefore

$$\begin{aligned} P(y|M_I) &= \binom{|y|}{y} \int_{p_r} \int_{p_c} \prod_{ij} (p_{i\cdot} p_{\cdot j})^{y_{ij}} \pi(p_r)\pi(p_c) dp_r dp_c \\ &= \binom{|y|}{y} \Gamma(R)\Gamma(C) \int_{p_r} \prod_i (p_{i\cdot})^{y_{i\cdot}} dp_r \int_{p_c} \prod_j (p_{\cdot j})^{y_{\cdot j}} dp_c \\ &= \binom{|y|}{y} \Gamma(R)\Gamma(C) \frac{\prod \Gamma(y_{i\cdot}+1)}{\Gamma(|y|+R)} \frac{\prod \Gamma(y_{\cdot j}+1)}{\Gamma(|y|+C)} \\ &= \binom{|y|}{y} \frac{D(y_R+1)D(y_C+1)}{D(1_R)D(1_C)} \end{aligned}$$

Bayes Factor

Combining the two marginal likelihoods we get the Bayes Factor

$$BF = \frac{P(y|M_D)}{P(y|M_I)} = \frac{D(y+1)D(1_R)D(1_C)}{D(1_{RC})D(y_R+1)D(y_C+1)}$$

Our data is

11	68	3	82
9	23	5	37
20	91	8	119

The Bayes factor is

$$\frac{11!68!3!9!23!5!1!2!}{124!} \times \frac{120!121!}{5!20!91!8!82!37!} = 1.66$$

which gives modest support against independence.

Normal Linear regression model

Normal Linear regression model

Model: Response variable $n \times 1$ vector $Y = (y_1, \dots, y_n)$, predictor variables $n \times p$ matrix $X = (x_1, \dots, x_p)$.

$$Y = X\beta + \epsilon, \quad \epsilon \sim N(0, \sigma^2 I)$$

Recall that classical unbiased estimates are

$$\hat{\beta} = (X^t X)^{-1} X^T Y, \quad \hat{\sigma}^2 = (Y - X\hat{\beta})^T (Y - X\hat{\beta})$$

and predicted Y is

$$\hat{Y} = X\hat{\beta} = P_X Y, \quad P_X = X(X^T X)^{-1} X^T.$$

Normal Linear regression model

To sum up:

$$Y \mid \beta, \sigma^2, X \sim N_n(X\beta, \sigma^2 I)$$

Bayesian formulation: Assume that (β, σ^2) have a non-informative prior

$$g(\beta, \sigma^2) \propto \frac{1}{\sigma^2}$$

Posterior distribution

$$q(\beta, \sigma^2|Y) = q(\beta \mid Y, \sigma^2)q(\sigma^2 \mid Y)$$

$$q(\sigma^2 \mid Y) \propto \frac{1}{(\sigma^2)^{(n-p)/2+1}} \exp\left\{-\frac{(n-p)s^2}{2\sigma^2}\right\}$$
$$\sim IG\left(\frac{n-p}{2}, \frac{(n-p)\hat{\sigma}^2}{2}\right)$$

Recall Inverse Gamma(a, b) is $y^{-a-1}\exp\{-b/y\}$

$$q(\beta \mid Y, \sigma^2) = N(\hat{\beta}, V_\beta \sigma^2)$$

where

$$\widehat{\beta} = (X^T X)^{-1} X^T Y,\ V_\beta = (X^T X)^{-1}$$

Posterior

The posterior density comes from a classical factorization of the likelihood

$$\frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left\{-\frac{1}{2\sigma^2}(y - X\beta)^T(y - X\beta)\right\}$$

knowing that

$$(y - X\beta)^T(y - X\beta) = (y - X\widehat{\beta})'(y - X\widehat{\beta}) + (\widehat{\beta} - \beta)^T X^T X(\widehat{\beta} - \beta)$$

- $P(\beta \mid Y)$ is a non-cenral multivariate t_{n-p} distribution.
- For each j

$$\frac{\beta_j - \hat{\beta}_j}{s\sqrt{(X^T X)^{-1}_{jj}}} \sim t_{n-p}$$

Prediction

New covariate matrix $\tilde{X}$, predict $\tilde{Y}$.

$$p(\tilde{Y}|Y) = \int p(\tilde{Y}|\beta, \sigma^2)p(\beta, \sigma^2|Y)d\beta d\sigma^2$$

Simulate or $p(\tilde{Y}|Y)$ is a multivariate t distribution

$$t_{n-p}(\tilde{X}\hat{\beta}, \hat{\sigma}^2(I + \tilde{X}(X^T X)^{-1}\tilde{X}^T))$$