BayesML
1. Intro 1.1
Learning proportions

DataLab CSIC

Brief description

Basic Bayesian inference concepts through simple, fundamental models

- Proportions. Beta-binomial model
- Means. Normal-normal model
- Linear regression+intro to hierarchical models
- Logistic regression (simple approach)
- Dynamic linear models

Conceptual recap: Bayesian inference, prediction and decision making

Motivating computational issues

Today props. BDA3, Hoff, French and DRI

Schedule

- Week 0. Props,
- Week 1. Means, Regression
- Week 2. DLMs, Recap.

Lab1-1

Expo-Gamma (Reliability)

Beta-binomial (Fraud detection)

Normal-normal (Ecology)

Ultrabasic concepts!!!

- Inference/Learning: Point Estimation, Interval estimation, Hypothesis testing
- Prediction
- Decision Support

- Uncertainty almost ubiquitous
  - Inherente to system
  - Incomplete observability
  - Incomplete modelling

- Probability as mesure of degree of uncertainty with certain mathematical properties

- Interpretations
  - Classical
  - Frequentist
  - Subjective

https://www.youtube.com/watch?v=KxV5kckOVeA

https://www.youtube.com/watch?v=L1Q7w3ch3

https://www.youtube.com/watch?v=OWjWYyG4Oys

Ultrabasic concepts!!!!

- Conditional probability

$$P(\mathrm{y} = y \mid \mathrm{x} = x) = \frac{P(\mathrm{y} = y, \mathrm{x} = x)}{P(\mathrm{x} = x)}$$

$$\mathrm{x} \perp \mathrm{y}$$

- Independence

$$\forall x \in \mathrm{x}, y \in \mathrm{y},\ p(\mathrm{x} = x, \mathrm{y} = y) = p(\mathrm{x} = x)p(\mathrm{y} = y)$$

$$\mathrm{x} \perp \mathrm{y} \mid \mathrm{z}$$

- Conditional independence

$$\forall x \in \mathrm{x}, y \in \mathrm{y}, z \in \mathrm{z},\ p(\mathrm{x} = x, \mathrm{y} = y \mid \mathrm{z} = z) = p(\mathrm{x} = x \mid \mathrm{z} = z)p(\mathrm{y} = y \mid \mathrm{z} = z)$$

Ultrabasic concepts!!!!

- Marginal distribution

$$P(\mathrm{x}, \mathrm{y})$$

$$\forall x \in \mathrm{x}, P(\mathrm{x} = x) = \sum_y P(\mathrm{x} = x, \mathrm{y} = y)$$

$$p(x) = \int p(x, y) dy$$

- Bayes rule

$$P(\mathrm{x} \mid \mathrm{y}) = \frac{P(\mathrm{x}) P(\mathrm{y} \mid \mathrm{x})}{P(\mathrm{y})}$$

$$P(\mathrm{y}) = \sum_x P(\mathrm{y} \mid x) P(x)$$

Beta-binomial model: A typical example

Consider recovery protocols for an SME computer service after a cyber attack. We introduce one protocol and wish to assess it, e.g. to be compared with another one.

Protocol tested in 12 attacks. Effective in 9 (e.g. attack duration was less than one hour)

Let's start with the model

A typical example

- n trials (identical, independent). Two results: success, failure
- Number X of successes in n trials
- Success probability in a trial θ_1
- Distribution of number of successes in n trials $X|\theta_1 \sim Bin(12, \theta_1)$
- For X=9,

$$Pr(X = 9|\theta_1) \propto \theta_1^9 (1 - \theta_1)^3, \ \theta_1 \in [0, 1]$$

A typical example

Likelihood

$$Pr(X = 9|\theta_1) \propto \theta_1^9(1-\theta_1)^3,\ \theta_1 \in [0,1]$$

First approach: Maximise likelihood --→ Maximum likelihood estimator MLE

Compute it!!

The MLE is 9/12

But MLE has several defects...

A typical example

We may use another source of information about the parameter. The prior distribution, e.g.

$$p(\theta_1) = 1, \theta_1 \in [0, 1].$$

Update it through Bayes formula, to get the posterior

$$p(\theta_1|x = 9) \propto p(\theta_1) \times Pr(X = 9|\theta_1) \propto \theta_1^9(1 - \theta_1)^3, \ \theta_1 \in [0, 1]$$

which summarises all the info available about the parameter in a distribution

Beta (10,4)

Check

http://en.wikipedia.org/wiki/Beta_distribution

A typical example

The posterior serves as prior for subsequent studies. E.g., if in the following 5 applications there are 3 successes the new posterior is

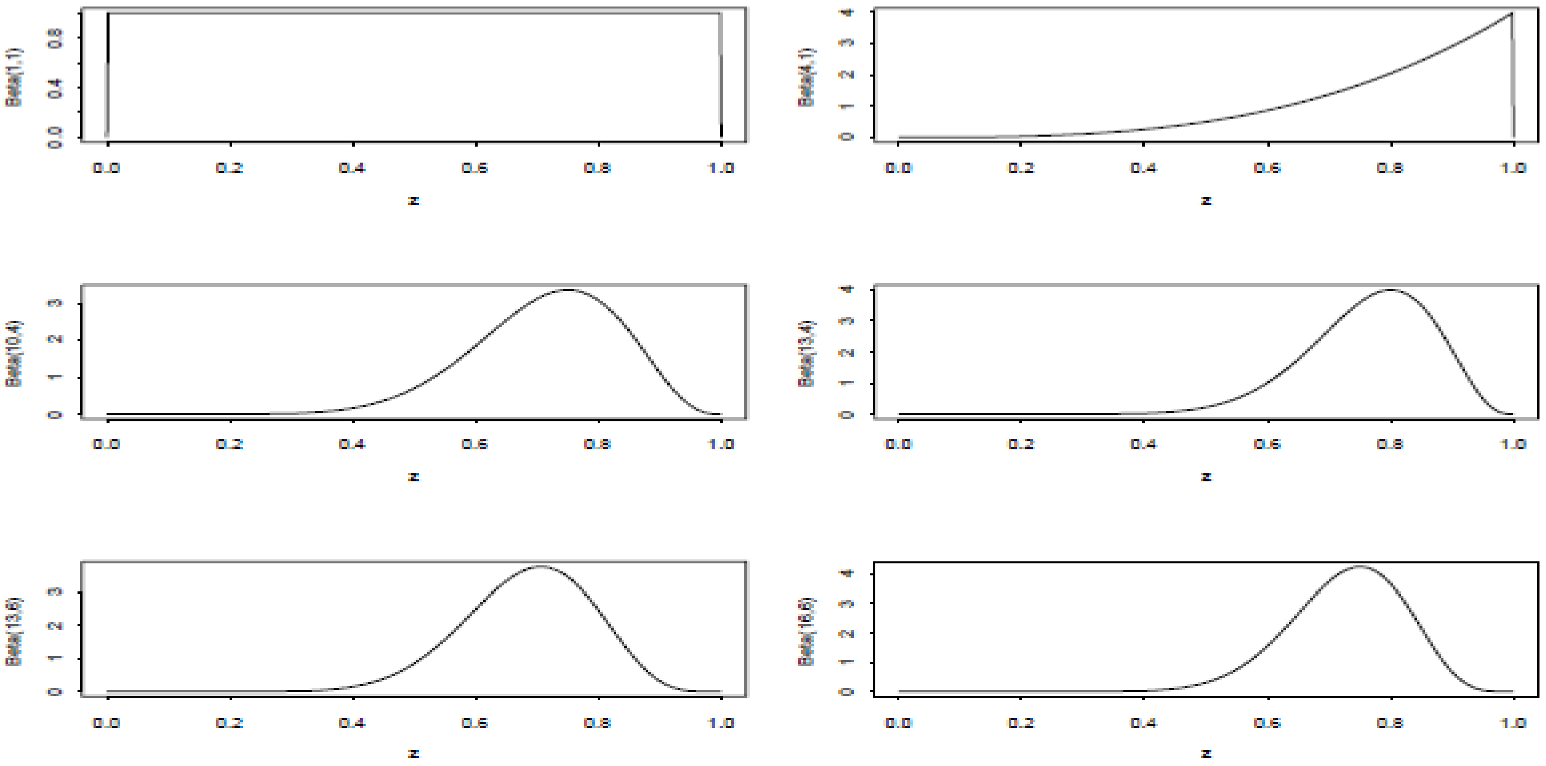
$$p(\theta_1|x=3) \propto [\theta_1^9(1-\theta_1)^3] \times [\theta_1^3(1-\theta_1)^2] = \theta_1^{12}(1-\theta_1)^5, \ \theta_1 \in [0,1] \qquad \text{Beta } (13{,}6)$$

Suppose that a priori, the probability is around 80% and bigger values are more likely, the learning goes through

$$\text{Beta } (4{,}1) \longrightarrow \text{Beta}(13{,}4) \longrightarrow \text{Beta}(16{,}6),$$

Sequential nature of Bayes rule

A typical example



Convergence in learning, consensus, asymptotic behavior

A typical example

- Focus on Beta (10,4). Try to use simulation for all computations also!!!

Point estimate. Summarise in a value, e.g. the posterior mean

$$\frac{10}{10+4} = 0.72$$

Why not the posterior median? Or the posterior mode (MAP)!!!

Interval estimate. Summarise interval with high probability e.g. 0.9.

- Symmetric probability wise

$$[0.505, .887]$$

- Highest posterior density interval. HDI

A typical example

- Focus on Beta (10,4)

Hypothesis testing. E.g Is the protocol effective? Null: Is the proportion bigger than 0.5

1-pbeta(0.5,10,4)=0.953

Forecasts Probability of more than 4 successes in 7 trials

$$
\begin{aligned}
Pr(X = k|x = 9) &= \int Pr(X = k|\theta_1)p(\theta_1|x = 9)d\theta_1 = \\
&= \int \binom{7}{k} \theta_1^k(1-\theta_1)^{7-k} \binom{13}{3} \theta_1^9(1-\theta_1)^3 d\theta_1 = \\
&= \frac{\binom{7}{k}\binom{13}{3}}{\binom{20}{9+k}}.
\end{aligned}
$$

$$Pr(X \geq 5|x = 9) = \sum_{k=5}^{7} Pr(X = k|x = 9) = 0.6641.$$

A typical example

Consider a second protocol. 10 opportunities, successful in 6. θ_2

Model

$$X|\theta_1 \sim Bin(12, \theta_1)$$

$$Y|\theta_2 \sim Bin(10, \theta_2)$$

$$\theta_1, \theta_2 \sim Unif[0,1]$$

independent

Want

$$r = Pr(\theta_1 \geq \theta_2 | x = 9, y = 6)$$

A typical example

$$\theta_1 \sim Beta(10, 4),\ \theta_2 \sim Beta(7, 5)$$

- Distribution of $\theta_1 - \theta_2$????

- Through simulation. E.g 1000 observations, compute differences, count those bigger than 0, divide by 1000.

- Which protocol is better?

$$r \approx 0.772.$$

A typical example

- Utility structure

	succeeds	does not succeed
Plan A	0.8	0
Plan B	1	0.2

- Expected utilities given probabilities

$0.8\theta_1 + 0(1-\theta_1) = 0.8\theta_1$

$\theta_2 + 0.2(1-\theta_2) = 0.2 + 0.8\theta_2$

- Epected utilities

$0.8E(\theta_1|x=9) = 0.8 \times \frac{10}{14} = \frac{4}{7}$

$0.2 + 0.8E(\theta_2|y=6) = 0.2 + 0.8 \times \frac{7}{12} = \frac{2}{3}.$

Recap: Bayesian inference with the beta-binomial model

Parameter

$$\theta$$

Model

$$\Pr(X = k \mid \theta) = \binom{n}{k} \theta^k (1-\theta)^{n-k}, \quad k = 0, 1, \dots, n$$

Data

$$x$$

Likelihood

$$\ell(\theta \mid x) \propto \theta^x (1-\theta)^{n-x}$$

$$h(\theta) = \log \ell(\theta \mid x) = x \log \theta + (n-x) \log(1-\theta)$$

(MLE)

$$h'(\theta) = 0 \Rightarrow \frac{x}{\theta} - \frac{n-x}{1-\theta} = 0 \Rightarrow \hat{\theta} = x/n$$

Recap: Bayesian inference with the beta-binomial model

Likelihood

$$\ell(\theta|x) \propto \theta^{x}(1-\theta)^{n-x}$$

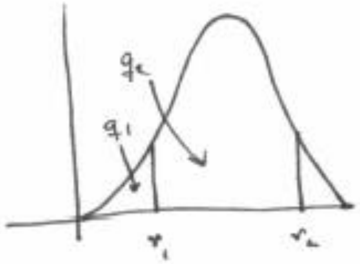
Prior

$$\pi(\theta) \propto \theta^{\alpha-1}(1-\theta)^{\beta-1} \qquad \theta \sim \mathcal{B}e(\alpha,\beta)$$

Noninformative prior

$$\pi(\theta) = I_{[0,1]}(\theta)$$

Eliciting the prior



$$\left.\begin{matrix} r_1 \\ r_2 \end{matrix}\right| \longrightarrow \alpha, \beta$$

Posterior

Sequential update

Likelihood

Prior

$$\pi(\theta|x) = \frac{\pi(\theta)\, p(\theta|x)}{p(x)} \propto \pi(\theta)\, p(\theta|x) \propto \theta^{x}(1-\theta)^{n-x}\,\theta^{\alpha-1}(1-\theta)^{\beta-1}$$

$$= \theta^{x+\alpha-1}(1-\theta)^{n-x+\beta-1}$$

$$\mathcal{B}e(\alpha,\beta) \longrightarrow \mathcal{B}e(\alpha+x,\ \beta+(n-x))$$

Recap: Bayesian inference with the beta-binomial model
(in parallel think of simulation based solutions)

Point estimation

Posterior mean

$$E(\theta|x) = \frac{\alpha+x}{\alpha+\beta+n}$$

Mix of prior and data

$$\frac{n}{\alpha+\beta+n}\,\frac{x}{n} + \frac{\alpha+\beta}{\alpha+\beta+n}\,\frac{\alpha}{\alpha+\beta}$$

What if n grows??

$$\xrightarrow[n]{} \approx \frac{x}{n}$$

$$Var(\theta|x) = \frac{(\alpha+x)(\beta+n-x)}{(\alpha+\beta+n)^2(\alpha+\beta+n+1)} \xrightarrow[n]{} 0$$

Posterior median

$$med(\theta|x) \approx \frac{\alpha+x-\frac{1}{3}}{\alpha+\beta+n-\frac{2}{3}}$$

$qbeta\,(0.5, \alpha+x, \alpha+\beta+n-x)$

Posterior mode

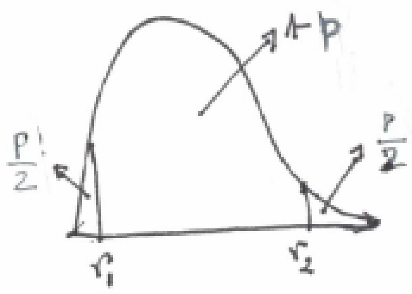
$$mode(\theta|x) = \frac{\alpha+x-1}{\alpha+\beta+n-2}$$

Recap: Bayesian inference with the beta-binomial model

Credible interval

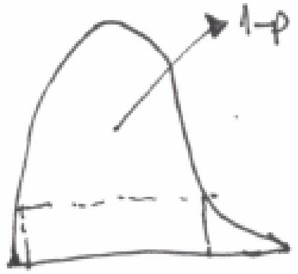
$\pi(\theta|x)$

Symmetric interval



$$\left[qbeta\left(p/2, \alpha+x, \beta+n-x\right), qbeta\left(1-\frac{p}{2}, \alpha+x, \beta+n-x\right) \right]$$

HPD



Recap: Bayesian inference with the beta-binomial model

$\pi(\theta|x)$

Hypothesis testing

Testing three hypothesis

$H_1: \theta \in \Theta_1$ $H_2: \theta \in \Theta_2$ $H_3: \theta \in \Theta_3$

$Pr(\theta \in \Theta_1 | x)$ $Pr(\theta \in \Theta_2 | x)$ $Pr(\theta \in \Theta_3 | x)$

0-1 Loss !!!

Choose Θ_i : $\max Pr(\theta \in \Theta_i | x)$

Point nulls

??

$H_0: \theta = \theta_0$ vs $H_1: \theta \neq \theta_1$.

Credible interval R for θ.

Accept H_0 if $\theta_0 \in R$.

Evidence supports H_0

Recap: Bayesian inference with the beta-binomial model

Forecasting. The predictive distribution

m future trials

$$Pr(Y=k \mid x) = \int Pr(Y=k \mid \theta)\, \pi(\theta \mid x)\, d\theta \quad k=0,\dots,m$$

$$= \int \binom{m}{k} \theta^{k} (1-\theta)^{m-k} \beta(\cdot,\cdot)\, \theta^{\alpha+x-1} (1-\theta)^{\beta+(n-x)-1}\, d\theta$$

$$= \frac{\binom{m}{k} \beta(\alpha+x, \beta+(n-x))}{\beta(k+\alpha+x, (m-k)+\beta+(n-x))}$$

Summarising the predictive distribution

$$E(Y \mid x) = \sum y\, Pr(Y=k \mid x)$$

$$= \int \pi(\theta \mid x) \left(\sum y\, Pr(Y=k \mid \theta)\right) d\theta$$

$$= \int m\theta\, \pi(\theta \mid x)\, d\theta = m \frac{\alpha+x}{\alpha+\beta+n}$$

...

See you next week

introml@icmat.es __for questions

Stuff at

https://datalab-icmat.github.io/courses_stats.html