Bayesian Inference in Processing Experimental Data: Conjugate Priors

Pedro Fernández Ramírez, Cora Fischer

July 17, 2013

Hadron Collider Summer School 2013, Göttingen

Outline

1. The Bayesian approach
   - Introduction
   - Bayes' Theorem
   - Importance of the Prior

2. Conjugate Priors
   - Motivation
   - The Beta distribution: A flexible prior
   - Table of Conjugate Priors

Introduction

The Bayesian approach is based on two crucial aspects:

- Definition of probability as the degree of believe that an event will occur
- Probability is unavoidable subjective

The main thrust of statistics theory in 20th century has been based on a different concept of probability: the limit of the long-term relative frequency of the outcome of the events. **This is wrong!** It revolves around the theoretical notion of infinite ensembles of "identical experiments".

Bayes' Theorem

Derive from probabilistic theory:

$$P(A|B)P(B) = P(B|A)P(A)$$

$$\Rightarrow$$

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

$P(A|B)$: posterior

$P(B|A)$: likelihood

$P(A)$: prior

$\rightarrow$ scheme for updating knowledge on basis of new conditions

Bayes' Theorem

Rewriting the theorem:

$$P(H_j|E_i, I) = \frac{P(E_i|H_j, I)P(H_j|I)}{\sum_j P(E_i|H_j, I)P(H_j|I)}$$

posterior $\propto$ likelihood $\times$ prior

with: $\bigcup_j H_j = \Omega$; $H_j \cap H_k = \varnothing$; $\sum_j P(H_j) = 1$

for continuous variables ($\rightarrow x$):

$$p(x|d, I) = \frac{p(d|x, I)p(x|I)}{\int p(d|x, I)p(x|I)\mathrm{d}x}$$

if prior chosen to be uniform:

$\rightarrow$ analysis equals maximum likelihood analysis, posterior=likelihood

Importance of the Prior

- In Bayesian theory: prior plays major role
- Advantage: include prior information in the analysis explicitly
- Prior assumption about parameter distribution, subjective, 'degree-of-belief'
- 'Nice feature': ability to transform vague, fuzzy priors into solid estimates, if sufficient amount of good quality data available

Motivation: Conjugate Priors

- calculations including priors can be complicated

  $\hookrightarrow$ modelling priors has been compromise between relatistic assessment of beliefs and simplified mathematical functions

- strategy: choose prior such that posterior belongs to same functional family
  $\hookrightarrow$ prior and posterior are **conjugate**

Example: Gaussian likelihood × Gaussian prior

$$K \exp\left[-\frac{(x_1-\mu)^2}{2\sigma_1^2} - \frac{(x_2-\mu)^2}{2\sigma_2^2}\right]$$
$$\Rightarrow \quad K' \exp\left[-\frac{(x'-\mu)^2}{2\sigma'^2}\right]$$

$\rightarrow$ Gaussian is 'auto-conjugate'

$\rightarrow$ not flexible: only one shape possible

Binomial Likelihood

Example: Binomial likelihood:

$$p(n|x, N) \propto x^n (1-x)^{N-n}$$

Then: using the **Beta distribution** as prior:

binomial likelihood $\times$ Beta prior $\rightarrow$ Beta posterior

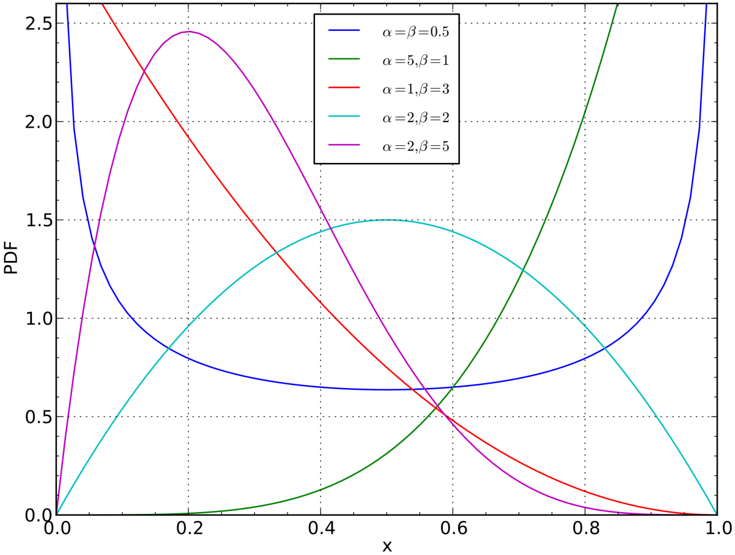
$\Rightarrow$ the Beta distribution is more flexible:

$$\text{Beta}(x|\alpha, \beta) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{\int_0^1 x^{\alpha-1}(1-x)^{\beta-1}\text{d}x} \qquad \begin{cases} 0 \leq x \leq 1 \\ \alpha, \beta > 0 \end{cases}$$

$\rightarrow$ large variety of shapes, depending on chosen α and β parameters

$\rightarrow$ α and β reflect existing belief or information (uniform prior if $\alpha = \beta = 1$)

Beta distribution: Different Parameters



Posterior Beta Distribution

$$p(x|n, N, \alpha, \beta) \propto \left[x^n(1-x)^{N-n}\right]\left[x^{\alpha-1}(1-x)^{\beta-1}\right]$$
$$\propto x^{n+\alpha-1}(1-x)^{N-n+\beta-1}$$

$\rightarrow$ posterior again Beta distribution with $\alpha' = \alpha + n$ and $\beta' = \beta + N - n$

expectation value and variance of posterior:

$$E(x) = \frac{\alpha'}{\alpha' + \beta'}$$
$$\sigma^2(x) = \frac{\alpha'\beta'}{(\alpha' + \beta' + 1)(\alpha' + \beta')^2}$$

for $n \gg \alpha$ and $N - n \gg \beta$:
posterior becomes progressively independent of prior
(same result as uniform prior)

More Conjugate Priors

likelihood	conjugate prior	posterior
Gauss(x, σ)	Gauss(μ_0, σ_0)	Gauss(μ_1, σ_1)
Binomial(N, x)	Beta(α, β)	Beta$(\alpha + n, \beta + N - n)$
Poisson(x)	Gamma(α, β)	Gamma$(\alpha + n, \beta + 1)$
Multinomial	Dirichlet$(\alpha_1, \ldots, \alpha_k)$	Dirichlet$(\alpha_1 + n_1, \ldots, \alpha_k + n_k)$

Binomial $\rightarrow$ Multinomial, Beta $\rightarrow$ Dirichlet:

multivariate extension, generalization

More Conjugate Priors

likelihood	conjugate prior	posterior
Gauss(x, σ)	Gauss(μ_0, σ_0)	Gauss(μ_1, σ_1)
Binomial(N, x)	Beta(α, β)	Beta($\alpha + n, \beta + N - n$)
Poisson(x)	Gamma(α, β)	Gamma($\alpha + n, \beta + 1$)
Multinomial	Dirichlet($\alpha_1, \ldots, \alpha_k$)	Dirichlet($\alpha_1 + n_1, \ldots, \alpha_k + n_k$)

Binomial $\rightarrow$ Multinomial, Beta $\rightarrow$ Dirichlet:

multivariate extension, generalization

Summary

- shortly introduced Bayesian reasoning
- emphasis on priors
- introduced idea of conjugate priors
- shortly discussed importance and influence of priors