06 - **Conjugate models** - 02

Bayesian Statistics
Spring 2022-2023

Josep Fortiana

Matemàtiques - Informàtica UB

Monday, March 27, 2023

06 - Conjugate models - 02

The Poisson-Gamma model

The Dirichlet - Multinomial model

Exponential-Gamma model

Mixture priors: the spinning coin

06 - Conjugate models - 02

The Poisson-Gamma model

The Dirichlet - Multinomial model

Exponential-Gamma model

Mixture priors: the spinning coin

Likelihood

$$(y \mid \lambda) \sim \text{Poisson}(\lambda),$$

Pmf:

$$f(y \mid \lambda) = e^{-\lambda} \frac{\lambda^y}{y!}, \quad y = 0, 1, 2, \ldots, \quad \lambda > 0,$$

Conjugate prior pdf

λ's prior:

$$\text{Gamma}(\alpha, \beta), \quad \alpha, \beta > 0,$$

Pdf:

$$h(\lambda) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda}, \quad \alpha, \beta, \lambda > 0.$$

Joint pdf

Joint "density" of (y, λ):

$$f(y, \lambda) \;=\; f(y \mid \lambda) \cdot h(\lambda)$$

$$=\; \frac{\beta^{\alpha}}{\Gamma(\alpha)} \cdot \frac{\lambda^{\alpha+y-1}}{y!} \, \exp\{-(\beta + 1)\lambda\}\,,$$

$\alpha, \beta, \lambda > 0.$

y's marginal pmf (Prior predictive pmf)

To integrate with respect to λ we split $f(y, \lambda)$:

$$\frac{\beta^{\alpha}}{\Gamma(\alpha)} \cdot \frac{1}{y!} \cdot \frac{\Gamma(\alpha + y)}{(\beta + 1)^{(\alpha + y)}}$$

$$\times \frac{(\beta + 1)^{(\alpha + y)}}{\Gamma(\alpha + y)} \cdot \lambda^{\alpha + y - 1} \exp\{-(\beta + 1)\lambda\} \, ,$$

Second line is a Gamma$(\alpha + y, \beta + 1)$ pdf. Integrates to 1.

y's marginal pmf (Prior predictive pmf)

After integrating out λ we get the marginal pmf of y:

$$f(y) = \frac{\Gamma(\alpha + y)}{\Gamma(\alpha)\, y!} \frac{\beta^\alpha}{(\beta + 1)^{\alpha + y}}, \quad y = 0, 1, \ldots, \quad \alpha, \beta > 0$$

Noting that:

$$\frac{\beta^\alpha}{(\beta + 1)^{\alpha + y}} = \left(\frac{1}{\beta + 1}\right)^y \cdot \left(\frac{\beta}{\beta + 1}\right)^\alpha,$$

y's marginal pmf (Prior predictive pmf)

we see $f(y)$ is a NB(r, p) pmf, with parameters:

$$r = \alpha, \quad \text{and} \quad p = \frac{\beta}{\beta + 1}.$$

(see Section 5 in the Appendix).

This is the *prior predictive* distribution for y.

Posterior pdf

Bayes' formula, dividing $f(y, \lambda)$ by the y-marginal pmf, yields the posterior pdf of λ, given y:

$$h(\lambda \mid y) = \frac{(\beta + 1)^{\alpha + y}}{\Gamma(\alpha + y)} \, \lambda^{\alpha + y - 1} \, e^{-(\beta + 1)\lambda} \, , \quad \lambda > 0,$$

a Gamma$(\alpha + y, \beta + 1)$ pdf.

Case of an *n*-sample: posterior pdf

For $y_1, \ldots, y_n$ i.i.d. $\sim$ Poisson(λ), the sum:

$$y = \sum_{i=1}^{n} y_i \sim \text{Poisson}(n\,\lambda)$$

(Additive property of Poisson r.v.)

Thus, for a prior $\lambda \sim \text{Gamma}(\alpha, \beta)$ and n iid Poisson(λ) data, the posterior is a Gamma($\alpha + y, \beta + n$).

Case of an *n*-sample: prior predictive pmf

Similarly, for *n* observed $\text{Poisson}(\lambda)$ data, the prior predictive (marginal) distribution of the total count number $y = \sum_{i=1}^{n} y_i$ is a $\text{NB}(r, p)$ r.v.,
with parameters:

$$r = \alpha, \quad \text{and} \quad p = \frac{\beta}{\beta + n}.$$

n-sample, different *exposures*

In many applications we find data of the form:

$$y_i \sim \text{Poisson}(\lambda_i), \qquad \text{where } \lambda_i = x_i \cdot \theta, \quad 1 \le i \le n.$$

The values x_i are known positive values of an explanatory variable, usually called *exposure*, and θ is the common *rate* parameter.

n-sample, different *exposures*. Posterior pdf

With a $\theta \sim \text{Gamma}(\alpha, \beta)$ prior,

observations $y = (y_1, \ldots, y_n)$,

exposures $x = (x_1, \ldots, x_n)$,

Posterior pdf:

$$\theta \mid y \sim \text{Gamma}(\alpha + y, \beta + x),$$

where $y = \sum_{i=1}^{n} y_i$ and $x = \sum_{i=1}^{n} x_i$.

06 - Conjugate models - 02

The Poisson-Gamma model

The Dirichlet - Multinomial model

Exponential-Gamma model

Mixture priors: the spinning coin

The Dirichlet - Multinomial model

Generalizes the Beta-Binomial conjugate pair.

The Dirichlet distribution is the multivariate Beta.

Multinomial is the multivariate Binomial.

Multivariate Bernoulli distribution

A partition $\Omega = A_1 \sqcup \cdots \sqcup A_m$, where the A_j are pairwise exclusive events whose union is the total space, and:

$$\theta = (\theta_1, \ldots, \theta_m), \quad \theta_j = \mathsf{P}(A_j), \quad 1 \leq j \leq m.$$

Each indicator

$$\mathbb{1}_{A_j} \sim \mathsf{Ber}(\theta_j) \quad 1 \leq j \leq m.$$

Multivariate Bernoulli distribution

The m-dimensional vector of indicators:

$$(\mathbb{1}_{A_1}, \ldots, \mathbb{1}_{A_m}),$$

has an m-dimensional *multivariate Bernoulli* distribution with vector of probabilities:

$$\theta = (\theta_1, \ldots, \theta_m),$$

Multivariate Bernoulli distribution

The sum of the m probabilities is 1.

The sum of the m indicators is 1 **(Cannot be independent!!).**

Each j-th marginal, $\mathbb{1}_{A_j} \sim \text{Ber}(\theta_j)$.

$$\text{E}(\mathbb{1}_{A_j}) = \theta_j, \qquad\qquad \text{var}(\mathbb{1}_{A_j}) = \theta_j (1 - \theta_j),$$

$$\text{cov}(\mathbb{1}_{A_j}, \mathbb{1}_{A_k}) = -\theta_j \theta_k, \quad j \neq k.$$

The multinomial distribution

Joint pmf of an m-multinomial of size n, and probs. $\theta = (\theta_1, \ldots, \theta_m)$, $\theta_j \in [0, 1]$, $\sum_{j=1}^{m} \theta_j = 1$:

$$\frac{n!}{x_1! \cdot \cdots \cdot x_m!} \, \theta_1^{x_1} \cdot \cdots \cdot \theta_m^{x_m}.$$

for an m-dimensional vector $x = (x_1, \ldots, x_m)$, of integers $x_j \in [0, n]$ such that $\sum_{j=1}^{m} x_j = n$.

x is the sum of n m-dim. vectors i.i.d. $\sim \text{Ber}(\theta)$.

The Dirichlet distribution

θ's joint pdf, with parameters $a = (a_1, \ldots, a_m)$, $a_i > 0$:

$$h(\theta_1, \ldots, \theta_m; a_1, \ldots, a_m) = \frac{1}{\mathsf{B}(a)} \prod_{i=1}^{m} \theta_i^{a_i - 1}, \quad \text{where}$$

$$\mathsf{B}(a) = \frac{\prod_{i=1}^{m} \Gamma(a_i)}{\Gamma\left(\sum_{i=1}^{m} a_i\right)} \quad \text{is the multivariate Beta function.}$$

Multinomial likelihood with Dirichlet prior

$(x \mid \theta) \equiv (x_1, \ldots, x_m \mid \theta_1, \ldots, \theta_m) \sim \text{Multinomial}(n, \theta),$

Parameter vector:

$\theta = (\theta_1, \ldots, \theta_m), \;\; 0 < \theta_i < 1, \;\; 1 \le i \le m, \;\; \sum_{i=1}^{m} \theta_i = 1,$

and θ's joint prior is Dirichlet with parameters

$a = (a_1, \ldots, a_m)$, then θ's posterior is Dirichlet,

with parameters: $a + x = (a_1 + x_1, \ldots, a_m + x_m).$

06 - Conjugate models - 02

The Poisson-Gamma model

The Dirichlet - Multinomial model

Exponential-Gamma model

Mixture priors: the spinning coin

Exponential likelihood

Used to model *waiting times* or *insurance claims*.

The pdf of an outcome y, given θ, is:

$$f(y \mid \theta) = \theta \, \exp(-\theta \cdot y), \quad y > 0,$$

and $\theta = 1/\,\mathsf{E}(y \mid \theta)$ is called the *rate parameter.*

The exponential is a special case of a Gamma with $(\alpha, \beta) = (1, \theta)$.

Gamma prior for θ

The conjugate prior for θ is a Gamma(α, β), for $\alpha, \beta > 0$, with pdf:

$$h(\theta) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} \cdot \theta^{\alpha-1} \cdot \exp\{-\beta\,\theta\}, \quad \theta > 0.$$

Likelihood for an *n*-sample

$y = (y_1, \ldots, y_n)$ i.i.d.$\sim$ Exp(θ).

$$f(y \mid \theta) = \prod_{i=1}^{n} \left(\theta\, e^{-\theta\, y_i}\right) = \theta^n \cdot \exp\{-n\, \theta\, \bar{y}\}.$$

Joint (y, θ) pdf

$$f(y, \theta) = f(y \mid \theta) \cdot h(\theta)$$

$$= \frac{\beta^\alpha}{\Gamma(\alpha)} \cdot \theta^{\alpha+n-1} \cdot \exp\{-\theta \cdot (n\bar{y} + \beta)\}$$

To get the $y = (y_1, \ldots, y_n)$ joint marginal (prior predictive pdf) we integrate $f(y, \theta)$ with respect to θ.

(Integral is a gamma function, after a variable change $t = \theta \, (n\bar{y} + \beta)$)

Prior predictive pdf

$$f(y_1, \ldots, y_n) = \frac{\Gamma(\alpha + n)}{\Gamma(\alpha)} \cdot \frac{\beta^\alpha}{(n\,\bar{y} + \beta)^{\alpha+n}},$$

$$y_1 > 0, \ldots, y_n > 0.$$

Caution! This is the <u>joint</u> pdf of the y vector.

It is a function of the sum $n\,\bar{y}$ but NOT the pdf OF $x = n\,\bar{y}$.

Obtaining it requires $n - 1$ further integrals.

Prior predictive density of $x = n\,\bar{y}$ for small n

For $n = 1$,

$$f(x) = \frac{\alpha\,\beta^{\alpha}}{(x + \beta)^{\alpha+1}}, \quad x > 0,$$

for $n = 2$,

$$f(x) = (\alpha + 1)\,\alpha\,\beta^{\alpha}\,\frac{x}{(x + \beta)^{\alpha+2}}, \quad x > 0,$$

$\alpha > 0, \beta > 0$.

Prior predictive density of $x = n\,\bar{y}$ for small n

For $n = 1$,

$$\mathsf{E}(x) = \frac{\beta}{\alpha - 1}, \ \alpha > 1, \ \mathrm{var}(x) = \frac{\alpha\,\beta^2}{(\alpha - 2)\,(\alpha - 1)^2}, \ \alpha > 2.$$

For $n = 2$,

$$\mathsf{E}(x) = \frac{2\,\beta}{\alpha - 1}, \ \alpha > 1, \ \mathrm{var}(x) = \frac{2\,(\alpha + 1)\,\beta^2}{(\alpha - 2)\,(\alpha - 1)^2}, \ \alpha > 2.$$

$(\beta > 0)$.

Posterior pdf of θ, given y

$$(\theta \mid y) \sim \text{Gamma}(\alpha', \beta'),$$

where the updating formula is:

$$\begin{cases} \alpha' & = & \alpha + n, \\ \beta' & = & \beta + n\,\bar{y}. \end{cases}$$

06 - Conjugate models - 02

The Poisson-Gamma model

The Dirichlet - Multinomial model

Exponential-Gamma model

Mixture priors: the spinning coin

Persi Diaconis

Stanford prof. Persi Diaconis, formerly a professional magician, famously found how many shuffles a deck of cards needs to give a truly random order (seven). He's also dabbled in coin games.
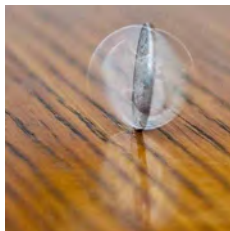

Photo Credit: L.A. Cicero

BTW: Persi Diaconis on randomness

See his 2013 video talk:

The Search for Randomness

The spinning coin

Fact: if a coin is spinned on its edge instead of being flipped, proportion of heads or tails is not around 50% but rather such values as 25% or 75% are obtained.

Persi Diaconis on the spinning coin

According to Diaconis, *"the reasons for the bias are not hard to infer. The shape of the edge will be a strong determining factor – indeed, magicians have coins that are slightly shaved; the eye cannot detect the shaving, but the spun coin always comes up heads"*.

A prior for the spinning coin problem

For *n* tosses of a spinning coin, the number *x* of heads up is a binomial Binom(n, θ), and θ's prior will typically be bimodal (pdf with two local maxima). It cannot be a Beta(α, β), which has a single mode at:

$$\frac{\alpha - 1}{\alpha + \beta - 2}$$

Diaconis, Persi and Donald Ylvisaker (1985) *Quantifying prior opinion.* In: J.M. Bernardo et al (eds), Bayesian Statistics 2, Elsevier, pp. 133-156.

A possible prior

$$0.50 \; \text{Beta}(10, 20) + 0.50 \; \text{Beta}(20, 10).$$
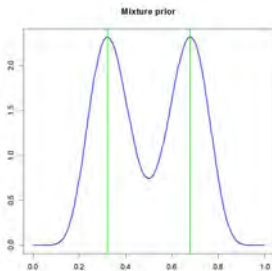
Mixture prior

Interpretation of a mixture prior

The mixture prior can be thought of as a weighted combination of "beta populations", the weights γ_i measuring the prior degree of belief that the actual coin was chosen from the i-th population.
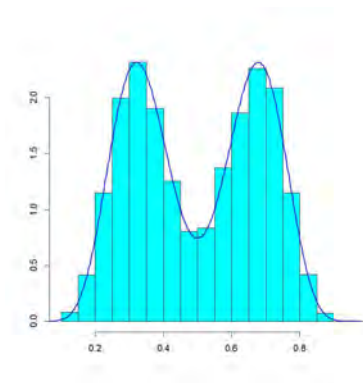
Simulation of a mixture prior

To generate random numbers from a mixture:

$$\gamma \cdot h_1 + (1 - \gamma) \cdot h_2, \quad \gamma \in (0, 1),$$

we generate a sequence of realizations of $I \sim \text{Ber}(\gamma)$ and:

▶ For each entry equal to 1, a realization of h_1,

▶ For each entry equal to 0, a realization of h_2.

Simulation of a mixture prior

Bayesian modelling with a mixture prior

Assume the prior pdf for θ is:

$$h(\theta) = \gamma \cdot h_1(\theta) + (1 - \gamma) \cdot h_2(\theta),$$

and the likelihood is: $f(x \mid \theta)$. Then, the joint pdf is:

$$f(x, \theta) = \gamma \cdot f_1(x, \theta) + (1 - \gamma) \cdot f_2(x, \theta).$$

where:

$$f_i(x, \theta) = f(x \mid \theta) \cdot h_i(\theta), \quad i = 1, 2.$$

Prior predictive pdf from a mixture prior

Integrating out θ, the marginal for x is:

$$f(x) = \gamma \cdot f_1(x) + (1 - \gamma) \cdot f_2(x),$$

where:

$$f_i(x) = \int f(x \mid \theta) \cdot h_i(\theta) \, d\theta, \quad i = 1, 2.$$

Posterior pdf from a mixture prior

From Bayes' formula:

$$h(\theta \mid x) = \frac{f(x, \theta)}{f(x)} = \frac{\gamma \cdot f_1(x, \theta) + (1 - \gamma) \cdot f_2(x, \theta)}{\gamma \cdot f_1(x) + (1 - \gamma) \cdot f_2(x)}.$$

With the obvious notation:

$$h_i(\theta \mid x) = \frac{f_i(x, \theta)}{f_i(x)} \quad i = 1, 2,$$

Posterior pdf from a mixture prior

the posterior pdf is:

$$h(\theta \mid x) = \widehat{\gamma}(x) \cdot h_1(\theta \mid x) + (1 - \widehat{\gamma}(x)) \cdot h_2(\theta \mid x),$$

where the *posterior mixture weights* are:

$$\widehat{\gamma}(x) = \frac{\gamma \cdot f_1(x)}{\gamma \cdot f_1(x) + (1 - \gamma) \cdot f_2(x)}, \text{ and } 1 - \widehat{\gamma}(x).$$

06 - Conjugate models - 02

Appendix: The negative binomial distribution

First definition

A sequence of independent binary $0/1$ experiments, with i.i.d.$\sim \text{Ber}(p)$ indicators, $p \in (0, 1)$.

Number X of realizations to obtain a number $r \in \mathbb{N}$ of successes (1's), is a *negative binomial r.v.* with *size r* and probability p.

Alternative: $Y = X - r =$ number of failures (0's) before obtaining a number r of successes.

Probability mass function

$$P(x) = \binom{x-1}{r-1} \cdot (1-p)^{x-r} \cdot p^r, \qquad x = 1, 2, \ldots,$$

$$P(y) = \binom{y+r-1}{r-1} \cdot (1-p)^y \cdot p^r, \qquad y = 0, 1, 2, \ldots,$$

The second one is more usual (e.g., `dnbinom()` in R).

Alternative (and the reason for the name)

$$P(y) = \binom{-r}{y} \cdot p^r \cdot (-q)^y, \quad \text{where } q = 1 - p.$$

Indeed:

$$\binom{-r}{y} = \frac{(-r) \cdot (-r-1) \cdots (-r-y+1)}{y!}$$

$$= (-1)^y \cdot \binom{y+r-1}{r-1}.$$

General definition

For an integer r,

$$\binom{y + r - 1}{r - 1} = \frac{(y + r - 1)!}{(r - 1)! \cdot y!} = \frac{\Gamma(y + r)}{\Gamma(r) \cdot y!}.$$

The right hand is valid for real $r > 0$. Thus the pmf:

$$P(y) = \frac{\Gamma(y + r)}{\Gamma(r) \cdot y!} \cdot (1 - p)^y \cdot p^r, \quad y = 0, 1, 2, \ldots,$$

defines the NB(r, p), for $r > 0$ and probability p.

Relation to the geometric distribution

For $r = 1$, the $\text{NB}(1, p)$ is the $\text{Geom}(p)$.

For integer r, a r.v. distributed as $\text{NB}(r, p)$ can be considered as the sum of r i.i.d. copies of a $\text{Geom}(p)$.

Expectation, variance of a negative binomial

For $Y \sim \text{NB}(r, p)$,

$$\text{E}(Y) = \mu \equiv r \cdot \frac{1-p}{p}$$

$$\text{var}(Y) = \sigma^2 \equiv r \cdot \frac{1-p}{p^2} = \frac{\mu}{p} = \mu + \frac{\mu^2}{r}.$$