L3. Conjugate Models

Priors

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)} = \frac{p(y|\theta)p(\theta)}{\int_\theta p(y|\theta)p(\theta)d\theta}$$

1. Non-informative prior (objective prior, vague prior, reference prior…)
2. Weekly informative prior
3. Informative prior (subjective prior)

Conjugacy

From the Bayes rule, $p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)} = \frac{p(y|\theta)p(\theta)}{\int_\theta p(y|\theta)p(\theta)d\theta}$

$\mathcal{P}$ $\mathcal{F}$ $\mathcal{P}$

Prior + Likelihood = Posterior

Same form of dist.

$p(\theta|y) \in \mathcal{P}$ for all $p(\cdot\,|\theta) \in \mathcal{F}$ and $p(\cdot) \in \mathcal{P}$

→ $\mathcal{P}$ is conjugate for $\mathcal{F}$

If the posterior is a distribution that is of the same family as our prior
→ the prior is conjugate to the likelihood.

Advantages

- By using conjugate prior, we know the form of the resultant posterior (no math!)
  → we can easily summarize the results using mean, mode, variance, etc.
- Easy to understand the meaning of the prior used in the analysis (insight). For example, Beta prior is just adding pseudo counts to the data.
- Due to the analytical form available, we can carry out the integration

$$p(y) = \int_\theta p(y|\theta)p(\theta)d\theta$$

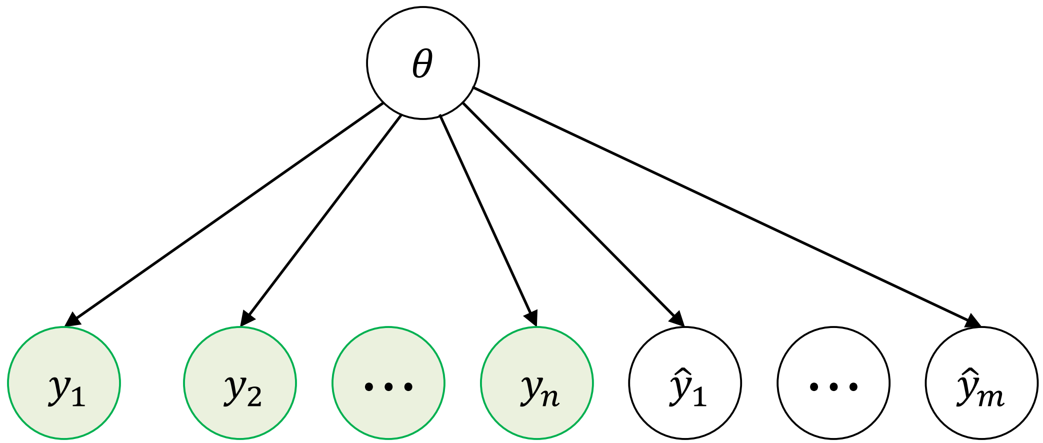
Conjugate pairs

	Likelihood	Prior	Posterior
	Binomial	Beta	Beta
	Negative Binomial	Beta	Beta
	Geometric	Beta	Beta
Single parameter model	Poisson	Gamma	Gamma
	Exponential	Gamma	Gamma
	Normal (mean unknown)	Normal	Normal
	Normal (variance unknown)	Inverse Gamma	Inverse Gamma
Multi parameters model	Normal (mean and variance unknown)	Normal-Gamma	Normal-Gamma
	Multinomial	Dirichlet	Dirichlet

Conjugate pairs

	Likelihood	Prior	Posterior
	Binomial	Beta	Beta
	Negative Binomial	Beta	Beta
	Geometric	Beta	Beta
Single parameter model	Poisson	Gamma	Gamma
	Exponential	Gamma	Gamma
	Normal (mean unknown)	Normal	Normal
	Normal (variance unknown)	Inverse Gamma	Inverse Gamma
Multi parameters model	Normal (mean and variance unknown)	Normal-Gamma	Normal-Gamma
	Multinomial	Dirichlet	Dirichlet

Bayesian Inference Problems



$y = (y_1, \dots, y_n)$ Observed

$\hat{y} = (\hat{y}_1, \dots, \hat{y}_m)$ Unobserved (future value)

Objectives

- **Prior predictive distribution**

$$p(y) = \int_\theta p(y, \theta) d\theta = \int_\theta p(y|\theta) p(\theta) d\theta$$

- **Posterior distribution**

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)} = \frac{p(y|\theta)p(\theta)}{\int_\theta p(y|\theta)p(\theta)d\theta} \propto p(y|\theta)p(\theta)$$

- **Posterior predictive distribution**

$$p(\hat{y}|y) = \int_\theta p(\hat{y}, \theta|y) d\theta = \int_\theta p(\hat{y}|\theta, y) p(\theta|y) d\theta = \int_\theta p(\hat{y}|\theta) p(\theta|y) d\theta$$

$p(\hat{y}|\theta, y) = p(\hat{y}|\theta)$ because $\hat{y} \perp y |\ \theta$

Binomial Likelihood and Beta Prior Distribution (Recap)

Binomial Likelihood and Beta Prior Distribution (Recap)

Likelihood :

$$Y \sim \mathrm{Bin}(n,\theta) \rightarrow p(y|\theta) = \binom{n}{y}\theta^{y}(1-\theta)^{n-y}$$

Prior:

$$\theta \sim \mathrm{Beta}(\alpha,\beta) \rightarrow p(\theta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}$$

Prior predictive distribution :

$$p(y) = \int_{\theta} p(y,\theta)d\theta = \int_{\infty} p(y|\theta)p(\theta)d\theta$$

$$\begin{aligned}
p(y) &= \int_0^1 p(y,\theta)d\theta \\
&= \int_0^1 p(y|\theta)p(\theta)d\theta \qquad P(y|\theta) = \mathrm{Bin}(y|n,\theta), \quad p(\theta) = \mathrm{Beta}(\alpha,\beta) \\
&= \int_0^1 \binom{n}{y}\theta^{y}(1-\theta)^{n-y}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}d\theta \qquad \binom{n}{y} = \frac{n!}{y!\,(n-y)!} = \frac{\Gamma(n+1)}{\Gamma(y+1)\Gamma(n-y+1)} \\
&= \frac{\Gamma(n+1)}{\Gamma(y+1)\Gamma(n-y+1)}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\int_0^1 \theta^{y}(1-\theta)^{n-y}\theta^{\alpha-1}(1-\theta)^{\beta-1}d\theta \\
&= \frac{\Gamma(n+1)\Gamma(\alpha+\beta)}{\Gamma(y+1)\Gamma(n-y+1)\Gamma(\alpha)\Gamma(\beta)}\int_0^1 \theta^{y+\alpha-1}(1-\theta)^{n-y+\beta-1}d\theta \\
&= \frac{\Gamma(n+1)\Gamma(\alpha+b)\Gamma(y+\alpha)\Gamma(n-y+\beta)}{\Gamma(y+1)\Gamma(n-y+1)\Gamma(\alpha)\Gamma(\beta)\Gamma(n+\alpha+\beta)}\int_0^1 \frac{\Gamma(n+\alpha+\beta)}{\Gamma(y+\alpha)\Gamma(n-y+\beta)}\theta^{y+\alpha-1}(1-\theta)^{n-y+\beta-1}d\theta \\
&= \frac{\Gamma(n+1)\Gamma(\alpha+\beta)\Gamma(y+\alpha)\Gamma(n-y+\beta)}{\Gamma(y+1)\Gamma(n-y+1)\Gamma(\alpha)\Gamma(\beta)\Gamma(n+\alpha+\beta)}\int_0^1 \mathrm{Beta}(\theta|y+\alpha,n-y+\beta)d\theta \\
&= \frac{\Gamma(n+1)\Gamma(\alpha+\beta)\Gamma(y+\alpha)\Gamma(n-y+\beta)}{\Gamma(y+1)\Gamma(n-y+1)\Gamma(\alpha)\Gamma(\beta)\Gamma(y+\alpha+\beta)}1 \\
&= \text{Beta–Binomial}(y|n,\alpha,\beta)
\end{aligned}$$

The beta-binomial distribution is the binomial distribution in *which the probability of success at each trial is not fixed but random* and follows the beta distribution

$$\mathrm{Beta-bin}(y|n,\alpha,\beta) = \int \mathrm{Bin}(y|n,\theta)\mathrm{Beta}(\theta|\alpha,\beta)d\theta$$

Binomial Likelihood and Beta Prior Distribution (Recap)

Likelihood :

$Y \sim \mathrm{Bin}(n, \theta) \rightarrow p(y|\theta) = \binom{n}{y} \theta^{y}(1-\theta)^{n-y}$ Y is the number of successes among n Bernoulli trials

Prior:

$\theta \sim \mathrm{Beta}(\alpha, \beta) \rightarrow p(\theta) = \dfrac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha-1}(1-\theta)^{\beta-1}$

Posterior :

$$P(\theta|y) \propto P(y|\theta)p(\theta)$$

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

$$\propto \theta^{y}(1-\theta)^{n-y} \theta^{\alpha-1}(1-\theta)^{\beta-1}$$

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

$$= \mathrm{Beta}(\theta|\alpha+y, \beta+n-y)$$

(α is a pseudo-count for the success, while β is a pseudo-count for the failure)

Binomial Likelihood and Beta Prior Distribution (Recap)

Likelihood :

$Y \sim \text{Bin}(n, \theta) \rightarrow p(y|\theta) = \binom{n}{y} \theta^{y}(1-\theta)^{n-y}$

Y is the number of success among n Bernoulli trial

Prior:

$\theta \sim \text{Beta}(\alpha, \beta) \rightarrow p(\theta) = \dfrac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha-1}(1-\theta)^{\beta-1}$

Posterior :

$P(\theta|y) = \text{Beta}(\theta|\alpha+y, \beta+n-y)$

Posterior predictive distribution :

$$p(\hat{y}|y) = \int_0^1 p(\hat{y}, \theta|y)d\theta = \int_0^1 p(\hat{y}|\theta)p(\theta|y)d\theta$$

$$p(y) = \int_0^1 p(y, \theta)d\theta = \int_0^1 p(y|\theta)p(\theta)d\theta$$

$$= \text{Beta−Binomial}(y|n, \alpha, \beta)$$

$p(\theta) = \text{Beta}(\alpha, \beta)$

$p(\theta|y) = \text{Beta}(\alpha+y, \beta+n-y)$

Using this result, **the posterior predictive distribution** is

$$p(\hat{y}|y) = \int_0^1 p(\hat{y}, \theta|y)d\theta = \int_0^1 p(\hat{y}|\theta)p(\theta|y)d\theta$$

$$= \text{Beta−Binomial}(\hat{y}|n, \alpha+y, \beta+n-y)$$

Poisson Likelihood-Gamma Prior

Poisson Likelihood-Gamma Prior

Likelihood :

$$Y_i \sim \text{Poisson}(\lambda) \rightarrow p(y_i|\lambda) = \frac{\lambda^{y_i} e^{-\lambda}}{y_i!}$$

Prior:

$$\lambda \sim \text{Gamma}(\alpha, \beta) \rightarrow p(\lambda) = \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda}$$

Prior predictive distribution :

$$p(y) = \int_\lambda p(y, \lambda) d\lambda = \int_\lambda p(y|\lambda) p(\lambda) d\lambda$$

$$\begin{aligned}
p(y) &= \int_0^\infty \frac{\lambda^y e^{-\lambda}}{y!} \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda} d\lambda \\
&= \frac{\beta^\alpha}{y!\,\Gamma(\alpha)} \int_0^\infty \lambda^y e^{-\lambda} \lambda^{\alpha-1} e^{-\beta\lambda}\, d\lambda \\
&= \frac{\beta^\alpha}{y!\,\Gamma(\alpha)} \int_0^\infty \lambda^{y+\alpha-1} e^{-(\beta+1)\lambda}\, d\lambda \\
&= \frac{\beta^\alpha}{y!\,\Gamma(\alpha)} \frac{\Gamma(y+\alpha)}{(\beta+1)^{y+\alpha}} \int_0^\infty \underbrace{\frac{(\beta+1)^{y+\alpha}}{\Gamma(y+\alpha)} \lambda^{y+\alpha-1} e^{-(\beta+1)\lambda}}_{\text{Gamma}(\lambda|\alpha+y, \beta+1)}\, d\lambda \\
&= \frac{\beta^\alpha}{y!\,\Gamma(\alpha)} \frac{\Gamma(y+\alpha)}{(\beta+1)^{y+\alpha}} \\
&= \binom{y+\alpha-1}{y} \left(\frac{1}{\beta+1}\right)^y \left(\frac{\beta}{\beta+1}\right)^\alpha = \text{Neg}-\text{bin}(y|\alpha, \beta)
\end{aligned}$$

Negative Binomial distribution ($\text{Neg}-\text{bin}$**) is** a discrete probability distribution of the number of successes y in a sequence of independent and identically distributed Bernoulli trials before a specified (non-random) number of failures (denoted α) occurs.

$$\text{Neg}-\text{bin}(y|\alpha, \beta) = \int \text{Poisson}(y|\lambda)\text{Gamma}(\lambda|\alpha, \beta) d\lambda$$

Poisson Likelihood-Gamma Prior

Likelihood :

$$Y_i \sim \text{Poisson}(\lambda) \rightarrow p(y_i|\lambda) = \frac{\lambda^{y_i} e^{-\lambda}}{y_i!} \qquad \text{E}(Y_i) = \lambda$$

Prior:

$$\lambda \sim \text{Gamma}(\alpha, \beta) \rightarrow p(\lambda) = \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda} \qquad \text{E}(\lambda) = \frac{\alpha}{\beta}$$

Posterior :

$P(\lambda|y) \propto P(y|\lambda)p(\lambda)$ $\quad y = (y_1, \dots, y_n)$ is a sequence of i.i.d. observation

$$= \left(\prod_{i=1}^{n} P(y_i|\lambda)\right) \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda}$$

$$= \left(\prod_{i=1}^{n} \frac{\lambda^{y_i} e^{-\lambda}}{y_i!}\right) \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda}$$

$$= \frac{\lambda^{\Sigma_i y_i} e^{-n\lambda}}{\prod_i^n y_i!} \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda}$$

$$\propto \lambda^{n\bar{y}} e^{-n\lambda} \lambda^{\alpha-1} e^{-\beta\lambda} \qquad \because \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

$$\propto \lambda^{\alpha+n\bar{y}-1} e^{-(\beta+n)\lambda}$$

$$= \text{Gamma}(\alpha + n\bar{y}, \beta + n) \qquad \text{E}(\lambda|y) = \frac{\alpha + n\bar{y}}{\beta + n} = \frac{\alpha + \Sigma_i y_i}{\beta + n}$$

(α is a pseudo-count for # of events, while β is a pseudo count for # of observations)

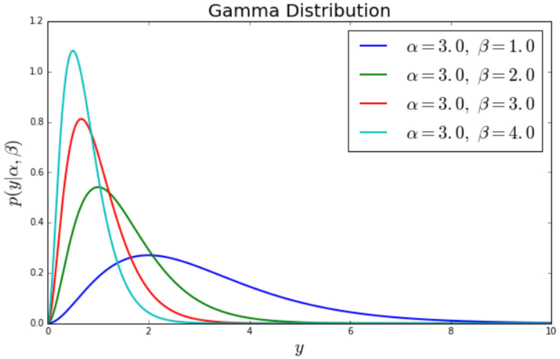
Poisson Likelihood-Gamma Prior

The posterior mean is

$$
\begin{aligned}
\mathrm{E}(\lambda|y) &= \frac{\alpha + \Sigma_i y_i}{\beta + n} \\
&= \frac{\alpha}{\beta + n} + \frac{\Sigma_i y_i}{\beta + n} \\
&= \frac{\beta}{\beta + n}\left(\frac{\alpha}{\beta}\right) + \frac{n}{\beta + n}\left(\frac{\Sigma_i y_i}{n}\right) \\
&= \frac{\beta}{\beta + n} E(\lambda) + \frac{n}{\beta + n}\hat{\theta}_{ML}
\end{aligned}
$$

Again, the data get weighted more heavily as $n \to \infty$

β control strength of prior



Example

Counts the numbers of Pokemons in each district of SF. The numbers are

14 13 7 10 15 15 2 13 13 11 10 13 5 13 9 12 9 12 8 7

- It is assumed that the numbers are independent and drawn from a Poisson distribution with mean λ

$$Y_i \sim \text{Poisson}(\lambda) \rightarrow p(y_i|\lambda) = \frac{\lambda^{y_i} e^{-\lambda}}{y_i!}$$

- The prior distribution for λ is a Gamma distribution with mean 20 and standard deviation 10

$$\lambda \sim \text{Gamma}(\alpha, \beta) \rightarrow p(\lambda) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda}$$

$$\mathrm{E}(\lambda) = \frac{\alpha}{\beta} = 20, \mathrm{Var}(\lambda) = \frac{\alpha}{\beta^2} = 10^2 \rightarrow \alpha = 4, \beta = 0.2$$

- The posterior is

$$P(\lambda|y) = \text{Gamma}(4 + 211, 0.2 + 20) \qquad P(\lambda|y) = \text{Gamma}(\alpha + n\bar{y}, \beta + n)$$

$$\mathrm{E}(\lambda|y) = \frac{215}{20.2} = 10.64, \mathrm{Var}(\lambda) = \frac{215}{20.2^2} = 0.5269$$

Poisson Likelihood-Gamma Prior

Likelihood :

$$Y_i \sim \text{Poisson}(\lambda) \rightarrow p(y_i|\lambda) = \frac{\lambda^{y_i} e^{-\lambda}}{y_i!}$$

Prior:

$$\lambda \sim \text{Gamma}(\alpha, \beta) \rightarrow p(\lambda) = \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda}$$

Posterior :

$$P(\lambda|y) = \text{Gamma}(\alpha + n\bar{y}, \beta + n)$$

$y = (y_1, \ldots, y_n)$ is a sequence of i.i.d. observations

Posterior predictive distribution (Using Bayes' rule)

$$p(\hat{y}|y) = \int_\lambda p(\hat{y}, \lambda|y) d\lambda = \int_\lambda p(\hat{y}|\lambda) p(\lambda|y) d\lambda$$

$$p(y) = \int_0^\infty p(y, \lambda) d\lambda = \int_0^\infty p(y|\lambda) p(\lambda) d\lambda$$

$$= NB(y|\alpha, \beta) = \binom{y + \alpha - 1}{y} \left(\frac{1}{\beta + 1}\right)^y \left(\frac{\beta}{\beta + 1}\right)^\alpha$$

$$p(\theta) = \text{Gamma}(\alpha, \beta)$$

$$p(\theta|y) = \text{Gamma}(\alpha + n\bar{y}, \beta + n)$$

Using this result, **the posterior predictive distribution** is

$$p(\hat{y}|y) = \int_0^\infty p(\hat{y}, \theta|y) d\lambda = \int_0^\infty p(\hat{y}|\theta) p(\theta|y) d\lambda$$

$$= NB(\hat{y}|\alpha + n\bar{y}, \beta + n) = \binom{\hat{y} + \alpha + n\bar{y} - 1}{\hat{y}} \left(\frac{1}{\beta + n + 1}\right)^{\hat{y}} \left(\frac{\beta + n}{\beta + n + 1}\right)^{\alpha + n\bar{y}}$$

Poisson model parameterized in terms of rate and exposure

Likelihood :

$$Y_i \sim \text{Poisson}(x_i\lambda) \to p(y_i|\lambda, x_i) = \frac{(x_i\lambda)^{y_i} e^{-(x_i\lambda)}}{(x_i\lambda)!}$$

where the value x_i is called the exposure of the ith unit

$$p(y|\lambda, x_i) = \prod_{i=1}^{n} \frac{(x_i\lambda)^{y_i} e^{-(x_i\lambda)}}{(x_i\lambda)!}$$

$y = (y_1, \dots, y_n)$ is a sequence of i.i.d. observations

$$\propto \lambda^{(\Sigma_i^n y_i)} e^{-(\Sigma_i^n x_i)\lambda}$$

It is more flexible model in that we can control the unit time or area

Prior:

$$\lambda \sim \text{Gamma}(\alpha, \beta) \to p(\lambda) = \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda}$$

Posterior :

$$P(\lambda|y) = \text{Gamma}\left(\alpha + \sum_{i=1}^{n} y_i\,, \beta + \sum_{i=1}^{n} x_i\right)$$

Estimating a rate from Poisson data: an idealized example

- A Poisson sampling model is often used for epidemiological data.
- 3 people, out of a population of 200,000 died of asthma
  → 1.5 cases per 100,000 people per year : exposure $x = 2.0$
- Under the Poisson model, the sampling distribution of y, the number of deaths in a city of 200,000 in one year, can be expressed as

$$y \sim \text{Poisson}(2.0\lambda)$$

Where λ represents the true underlying long-term asthma mortality rate in our city (measured in cases per 100,000 people per year)

- We can use knowledge about asthma mortality rates around the world to construct a prior distribution for λ and then combine the datum $y = 3$ to obtain a posterior distribution

Estimating a rate from Poisson data: an idealized example

Prior:

$\lambda \sim \text{Gamma}(\alpha = 3, \beta = 5)$

$\lambda = \#death\ per$ 100,000

Posterior :

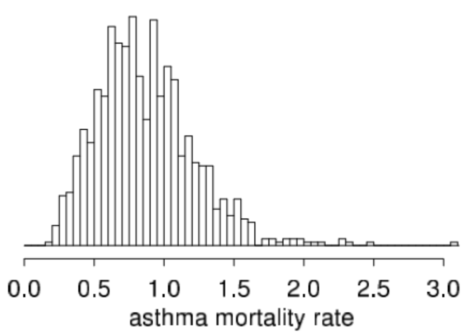
$$P(\lambda|y) = \text{Gamma}\left(\alpha + \sum_{i=1}^{n} y_i \,, \beta + \sum_{i=1}^{n} x_i\right)$$

Case 1: 3 persons, out of a population of 200,000 died of asthma for years

$\sum_{i=1}^{n} x_i = 2$: exposure

$\sum_{i=1}^{n} y_i = 3$ (average)

$P(\lambda|y) = \text{Gamma}(3 + 3, 5 + 2)$

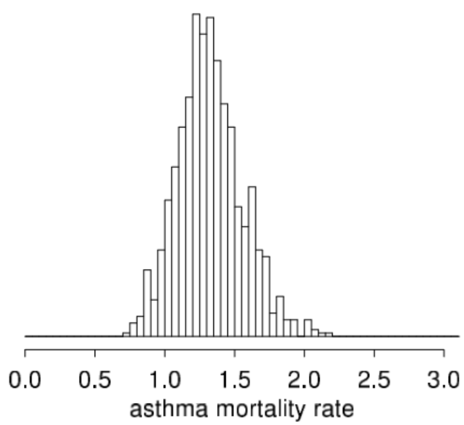


Case 1: 30 persons, out of a population of 200,000 died of asthma for 10 years

$\sum_{i=1}^{n=10} x_i = 20$: exposure

$\sum_{i=1}^{n} y_i = 30$ (average)

$P(\lambda|y) = \text{Gamma}(3 + 30, 5 + 20)$



Exponential Likelihood-Gamma Prior

Exponential Likelihood-Gamma Prior

Likelihood :

$Y_i \sim \text{Exp}(\lambda) \rightarrow p(y_i|\lambda) = \lambda e^{-\lambda y_i}$

$\text{E}(Y_i) = \frac{1}{\lambda}$

Prior:

$\lambda \sim \text{Gamma}(\alpha, \beta) \rightarrow p(\lambda) = \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda}$

Posterior :

$P(\lambda|y) \propto P(y|\lambda)p(\lambda)$ $\quad y = (y_1, \dots, y_n)$ is a sequence of i.i.d. observations

$$= \left(\prod_{i=1}^{n} P(y_i|\lambda)\right) p(\lambda)$$

$$= \left(\prod_{i=1}^{n} \lambda e^{-\lambda y_i}\right) \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda}$$

$$= \lambda^n e^{-\lambda \Sigma_i y_i} \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda}$$

$$\propto \lambda^n e^{-\lambda n \bar{y}} \lambda^{\alpha-1} e^{-\beta\lambda} \qquad \because \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

$$= \lambda^{\alpha+n-1} e^{-(\beta+n\bar{y})\lambda}$$

$$= \text{Gamma}(\alpha + n, \beta + n\bar{y})$$

Exponential Likelihood-Gamma Prior (Example)

A machine continuously produces nylon filament. From time to time the filament snaps. Suppose that the time intervals, in minutes, between snaps are random, independent and have an exponential distribution.

- Time interval between two successive failures : $Y_i \sim \mathrm{Exp}(\lambda) \rightarrow p(y_i|\lambda) = \lambda e^{-\lambda y_i}$

  The mean time = $\mathrm{E}(Y_i) = \frac{1}{\lambda}$

- Prior $\lambda \sim \mathrm{Gamma}(\alpha = 6, \beta = 1800)$

  $$\mathrm{E}(\lambda) = \frac{6}{1800} = 0.0033, \mathrm{Var}(\lambda) = \frac{6}{1800^2} = 1.85 \times 10^{-6}$$

- The mean time $\frac{1}{\lambda}$ follows the *invers-Gamma* distribution, since its inverse follows Gamma

$$\mathrm{E}\left(\frac{1}{\lambda}\right) = \int_0^\infty \frac{1}{\lambda}\frac{\beta^\alpha}{\Gamma(\alpha)}\lambda^{\alpha-1}e^{-\beta\lambda}d\lambda$$

$$\lambda \sim \mathrm{Gamma}(\alpha,\beta) \rightarrow p(\lambda) = \frac{\beta^\alpha}{\Gamma(\alpha)}\lambda^{\alpha-1}e^{-\beta\lambda}$$

$$= \frac{\beta^\alpha\Gamma(\alpha-1)}{\beta^{\alpha-1}\Gamma(\alpha)}\int_0^\infty \frac{\beta^{\alpha-1}}{\Gamma(\alpha-1)}\lambda^{\alpha-1-1}e^{-\beta\lambda}d\lambda = \frac{\beta}{\alpha-1} \qquad \alpha > 1$$

$$\mathrm{E}\left(\frac{1}{\lambda^2}\right) = \frac{\beta^2}{(\alpha-1)(\alpha-2)}$$

$$\mathrm{Var}\left(\frac{1}{\lambda}\right) = \frac{\beta^2}{(\alpha-1)(\alpha-2)} - \left(\frac{\beta}{\alpha-1}\right)^2 = \frac{\beta^2}{(\alpha-1)^2(\alpha-2)} \qquad \alpha > 2$$

Exponential Likelihood-Gamma Prior (Example)

$Y_i \sim \text{Exp}(\lambda) \rightarrow p(y_i|\lambda) = \lambda e^{-\lambda y_i}$

$\lambda \sim \text{Gamma}(\alpha = 6, \beta = 1800)$

- The prior mean for the time interval : $\text{E}\left(\frac{1}{\lambda}\right) = \frac{\beta}{\alpha-1} = \frac{1800}{6-1} = 360$(6 hours)
- The prior variance for the time interval : $\text{var}\left(\frac{1}{\lambda}\right) = \frac{\beta^2}{(\alpha-1)^2(\alpha-2)} = \frac{1800^2}{5^2 \times 4} = 32{,}400$
- The prior std. for the time interval : $\text{std}\left(\frac{1}{\lambda}\right) = \sqrt{32{,}400} = 180$(3hours)

Observations:
55 30 231 592 141 139 695 56 803 642 1890 208 246 183 38 486 264 1091 368 222 662 150 2 133 417 418 743 216 138 306 201 145 804 193 66 577 773 268 388 861

$P(\lambda|y) = \text{Gamma}(\alpha + n, \beta + \Sigma_i y_i) = \text{Gamma}(6 + 40, 1{,}800 + 15{,}841)$

- The posterior mean for the time interval : $\text{E}\left(\frac{1}{\lambda}|y\right) = \frac{\beta}{\alpha-1} = \frac{1800+15{,}841}{46-1} = 392.0(\text{minutes})$
- The posterior variance for the time interval : $\text{var}\left(\frac{1}{\lambda}|y\right) = \frac{\beta^2}{(\alpha-1)^2(\alpha-2)} = \frac{17641^2}{45^2 \times 44} = 3492.76$
- The posterior std. for the time interval : $\text{std}\left(\frac{1}{\lambda}|y\right) = \sqrt{3492.76} = 59.1$(~1hours)

Normal Likelihood-Normal Prior (unknown mean and known variance)

Normal Likelihood-Normal Prior (unknown mean and known variance)

Likelihood :

$$Y_i \sim N(\theta, \sigma_Y^2) \rightarrow p(y_i|\theta, \sigma_Y^2) = \frac{1}{\sqrt{2\pi\sigma_Y^2}} \exp\left(-\frac{(y_i - \theta)^2}{2\sigma_Y^2}\right)$$

Note that $\theta = \mu_Y$, (σ_Y^2 is known)

Prior:

$$\theta \sim N(\mu_0, \tau_0^2) \rightarrow p(\theta) = \frac{1}{\sqrt{2\pi\tau_o^2}} \exp\left(-\frac{(\theta - \mu_0)^2}{2\tau_o^2}\right)$$

Prior predictive distribution :

Before the data y are considered, the distribution of the unknown but observable y is

$$p(y) = \int_\theta p(y, \theta) d\theta = \int_{-\infty}^{\infty} p(y|\theta) p(\theta) d\theta$$

(Without integration)

$$\mathrm{E}(y) = \mathrm{E}[\mathrm{E}(y|\theta)] = E[\theta] = \mu_0$$

$\because \mathrm{E}(y|\theta) = \theta$

$$\begin{aligned} \mathrm{var}(y) &= \mathrm{E}[\mathrm{var}(y|\theta)] + \mathrm{var}\big(E(y|\theta)\big) \\ &= \mathrm{E}(\sigma_Y^2) + \mathrm{var}(\theta) \\ &= \sigma_Y^2 + \tau_0^2 \end{aligned}$$

$\because \mathrm{var}(y|\theta) = \sigma_Y^2$

$$p(y) = \mathrm{N}(y|\mu_0, \sigma_Y^2 + \tau_0^2)$$

Normal Likelihood-Normal Prior (unknown mean and known variance)

Likelihood :

$$Y_i \sim N(\theta, \sigma_Y^2) \rightarrow p(y_i|\theta, \sigma_Y^2) = \frac{1}{\sqrt{2\pi\sigma_Y^2}} \exp\left(-\frac{(y_i-\theta)^2}{2\sigma_Y^2}\right)$$

Note that $\theta = \mu_Y, (\sigma_Y^2$ is known)

Prior:

$$\theta \sim N(\mu_0, \tau_0^2) \rightarrow p(\theta) = \frac{1}{\sqrt{2\pi\tau_o^2}} \exp\left(-\frac{(\theta-\mu_0)^2}{2\tau_o^2}\right)$$

Posterior :

$y = (y_1, \dots, y_n)$ is a sequence of i.i.d. observations

$$\begin{aligned}
P(\theta|y) &\propto P(y|\theta, \sigma_Y^2)p(\theta) \\
&= \left(\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_i-\theta)^2}{2\sigma_Y^2}\right)\right) \times \frac{1}{\sqrt{2\pi\tau_o^2}} \exp\left(-\frac{(\theta-\mu_0)^2}{2\tau_o^2}\right) \\
&\propto \exp\left(-\sum_{i=1}^{n} \frac{(y_i-\theta)^2}{2\sigma_Y^2} + \frac{(\theta-\mu_0)^2}{2\tau_0^2}\right) \\
&= \exp\left[-\frac{1}{2}\left(\sum_{i=1}^{n} \frac{(y_i-\theta)^2}{\sigma_Y^2} + \frac{(\theta-\mu_0)^2}{\tau_0^2}\right)\right] \\
&= \exp\left[-\frac{1}{2\sigma_Y^2\tau_0^2}\left(\tau_0^2 \sum_{i=1}^{n} (y_i-\theta)^2 + \sigma_Y^2(\theta-\mu_0)^2\right)\right] \\
&= \exp\left[-\frac{1}{2\sigma_Y^2\tau_0^2}\left(\tau_0^2 \sum_{i=1}^{n} (y_i^2 - 2\theta y_i + \theta^2) + \sigma_Y^2(\theta^2 - 2\theta\mu_0 + \mu_0^2)\right)\right]
\end{aligned}$$

Normal Likelihood-Normal Prior (unknown mean and known variance)

Posterior :

$$P(\theta|y) \propto \exp\left[-\frac{1}{2\sigma_Y^2\tau_0^2}\left(\tau_0^2\sum_{i=1}^{n}\left(y_i^2-2\theta y_i+\theta^2\right)+\sigma_Y^2(\theta^2-2\theta\mu_0+\mu_0^2)\right)\right]$$

$$=\exp\left[-\frac{1}{2\sigma_Y^2\tau_0^2}\left(\tau_0^2\sum_{i=1}^{n}y_i^2-2\tau_0^2\theta n\bar{y}+\tau_0^2 n\theta^2+\sigma_Y^2\theta^2-2\sigma_Y^2\theta\mu_0+\sigma_Y^2\mu_0^2\right)\right]$$

$$\left(\because \bar{y}=\frac{1}{n}\sum_{i=1}^{n}y_i\right)$$

$$=\exp\left[-\frac{1}{2\sigma_Y^2\tau_0^2}\left(\tau_0^2\sum_{i=1}^{n}y_i^2-2\tau_0^2\theta n\bar{y}+\tau_0^2 n\theta^2+\sigma_Y^2\theta^2-2\sigma_Y^2\theta\mu_0+\sigma_Y^2\mu_0^2\right)\right]$$

$$\propto \exp\left[-\frac{1}{2\sigma_Y^2\tau_0^2}(\theta^2(\sigma_Y^2+n\tau_0^2)-2\theta(\mu_0\sigma_Y^2+n\bar{y}\tau_0^2)+\text{const})\right]$$

$$=\exp\left[-\frac{1}{2}\left(\theta^2\left(\frac{\sigma_Y^2+n\tau_0^2}{\sigma_Y^2\tau_0^2}\right)-2\theta\left(\frac{\mu_0\sigma_Y^2+n\bar{y}\tau_0^2}{\sigma^2\tau_0^2}\right)+\text{const}'\right)\right]$$

$$=\exp\left[-\frac{1}{2}\left(\theta^2\left(\frac{1}{\tau_0^2}+\frac{n}{\sigma_Y^2}\right)-2\theta\left(\frac{\mu_0}{\tau_0^2}+\frac{n\bar{y}}{\sigma_Y^2}\right)+\text{const}'\right)\right]$$

$$=\exp\left[-\frac{1}{2}\left(\frac{1}{\tau_0^2}+\frac{n}{\sigma_Y^2}\right)\left(\theta^2-2\theta\left(\frac{\frac{\mu_0}{\tau_0^2}+\frac{n\bar{y}}{\sigma_Y^2}}{\frac{1}{\tau_0^2}+\frac{n}{\sigma_Y^2}}\right)+\text{const}'\right)\right]$$

$$=\exp\left[-\frac{1}{2}\left(\frac{1}{\tau_0^2}+\frac{n}{\sigma_Y^2}\right)\left(\theta-\left(\frac{\frac{\mu_0}{\tau_0^2}+\frac{n\bar{y}}{\sigma_Y^2}}{\frac{1}{\tau_0^2}+\frac{n}{\sigma_Y^2}}\right)\right)^2\right]$$

Normal Likelihood-Normal Prior (unknown mean and known variance)

Likelihood :

$$Y_i \sim N(\theta, \sigma_Y^2) \rightarrow p(y_i | \theta, \sigma_Y^2) = \frac{1}{\sqrt{2\pi\sigma_Y^2}} \exp\left(-\frac{(y_i-\theta)^2}{2\sigma_Y^2}\right)$$

Prior:

$$\theta \sim N(\mu_0, \tau_0^2) \rightarrow p(\theta) = \frac{1}{\sqrt{2\pi\tau_o^2}} \exp\left(-\frac{(\theta-\mu_0)^2}{2\tau_o^2}\right)$$

Posterior on $\theta = \mu_Y$

$$P(\theta|y) = N\left(\theta \left| \frac{\frac{\mu_0}{\tau_0^2}+\frac{n\bar{y}}{\sigma_Y^2}}{\frac{1}{\tau_0^2}+\frac{n}{\sigma_Y^2}}, \left(\frac{1}{\tau_0^2}+\frac{n}{\sigma_Y^2}\right)^{-1}\right.\right)$$

- Posterior mean μ_1 :

$$\mu_1 = \frac{\frac{\mu_0}{\tau_0^2}+\frac{n\bar{y}}{\sigma_Y^2}}{\frac{1}{\tau_0^2}+\frac{n}{\sigma_Y^2}} = \frac{\sigma_Y^2\mu_0}{\sigma_Y^2+n\tau_0^2} + \frac{n\tau_0^2\bar{y}}{\sigma^2+n\tau_0^2}$$

μ_0: Prior mean

$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$: Data mean

  - $\tau_0^2 \downarrow \Rightarrow$ Prior mean μ_0 becomes accurate and influence more to μ_1
  - $\sigma_Y^2 \downarrow \Rightarrow$ the data become precise, making $\bar{y}$ stronger
  - $n \uparrow \Rightarrow \bar{y}$ stronger

- Posterior variance τ_1^2 :

$$\tau_1^2 = \left(\frac{1}{\tau_0^2}+\frac{n}{\sigma_Y^2}\right)^{-1}$$

- Posterior precision :

$$\frac{1}{\tau_1^2} = \frac{1}{\tau_0^2} + \frac{n}{\sigma_Y^2}$$

$\frac{1}{\tau_0^2}$: Prior precision

$\frac{n}{\sigma^2}$: Data precision

Normal Likelihood-Normal Prior (unknown mean and known variance)

Likelihood :

$$Y_i \sim N(\theta, \sigma_Y^2) \rightarrow p(y_i|\theta, \sigma_Y^2) = \frac{1}{\sqrt{2\pi\sigma_Y^2}} \exp\left(-\frac{(y_i-\theta)^2}{2\sigma_Y^2}\right)$$

Note that $\theta = \mu_Y$, ($\sigma_Y^2$is known)

Posterior : $\theta = \mu_Y$

$$P(\theta|y) = N(\theta|\mu_1, \tau_1^2)$$

$$\mu_1 = \frac{\frac{\mu_0}{\tau_0^2} + \frac{n\bar{y}}{\sigma_Y^2}}{\frac{1}{\tau_0^2} + \frac{n}{\sigma_Y^2}} \qquad \tau_1^2 = \left(\frac{1}{\tau_0^2} + \frac{n}{\sigma_Y^2}\right)^{-1}$$

$$E[\theta|y] = \mu_1 \qquad \mathrm{var}(\theta|y) = \tau_1^2$$

Posterior predictive distribution

$$p(\hat{y}|y) = \int_\theta p(\hat{y}, \theta|y)d\theta = \int_\theta p(\hat{y}|\theta, y)p(\theta|\mathrm{y})d\theta = \int_\theta p(\hat{y}|\theta)p(\theta|\mathrm{y})d\theta$$

(Without integration)

$$\mathrm{E}(\hat{y}|y) = \mathrm{E}[\mathrm{E}(\hat{y}|\theta, y)|y] = E[\theta|y] = \mu_1 \qquad \because \mathrm{E}(\hat{y}|\theta, y) = \mathrm{E}(\tilde{y}|\theta) = \theta$$

$$\mathrm{var}(\hat{y}|y) = \mathrm{E}[\mathrm{var}(\hat{y}|\theta, y)|y] + \mathrm{var}(E(\hat{y}|\theta, y)|y) \qquad \because \mathrm{var}(\hat{y}|\theta, y) = \sigma_Y^2$$

$$= \mathrm{E}(\sigma_Y^2|y) + \mathrm{var}(\theta|y)$$

$$= \sigma_Y^2 + \tau_1^2$$

$$p(\hat{y}|y) = \mathrm{N}(\hat{y}|\mu_1, \sigma_Y^2 + \tau_1^2)$$

Normal Likelihood-Normal Prior (unknown mean and known variance)

$$\mathrm{E}(y) = \mathrm{E}[\mathrm{E}(y|\theta)] = E[\theta] = \mu_0 \qquad \because \mathrm{E}(y|\theta) = \theta$$

$$\begin{aligned} \mathrm{var}(y) &= \mathrm{E}[\mathrm{var}(y|\theta)] + \mathrm{var}\big(E(y|\theta)\big) \qquad \because \mathrm{var}(y|\theta) = \sigma_Y^2 \\ &= \mathrm{E}(\sigma_Y^2) + \mathrm{var}(\theta) \\ &= \sigma_Y^2 + \tau_0^2 \end{aligned}$$

$$p(y) = \mathrm{N}(y|\mu_0, \sigma_Y^2 + \tau_0^2)$$

$$\mathrm{E}(\hat{y}|y) = \mathrm{E}[\mathrm{E}(\hat{y}|\theta, y)|y] = E[\theta|y] = \mu_1 \qquad \because \mathrm{E}(\hat{y}|\theta, y) = \mathrm{E}(\tilde{y}|\theta) = \theta$$

$$\begin{aligned} \mathrm{var}(\hat{y}|y) &= \mathrm{E}[\mathrm{var}(\hat{y}|\theta, y)|y] + \mathrm{var}(E(\hat{y}|\theta, y)|y) \qquad \because \mathrm{var}(\hat{y}|\theta, y) = \sigma_Y^2 \\ &= \mathrm{E}(\sigma_Y^2|y) + \mathrm{var}(\theta|y) \\ &= \sigma_Y^2 + \tau_1^2 \end{aligned}$$

$$p(\hat{y}|y) = \mathrm{N}(\hat{y}|\mu_1, \sigma_Y^2 + \tau_1^2)$$

Normal Likelihood-Inverse-Gamma Prior (known mean and unknown variance)

Normal Likelihood-Inverse-Gamma Prior (known mean and unknown variance)

Likelihood :

$$Y_i \sim N(\mu, \theta) \rightarrow p(y_i|\mu, \theta) = \frac{1}{\sqrt{2\pi\theta}} \exp\left(-\frac{(y_i - \mu)^2}{2\theta}\right)$$

Note that $\theta = \sigma^2$, (μ is known)

Prior:

$$\theta \sim \mathrm{Inv} - \mathrm{Gamma}(\theta|\alpha_0, \beta_0) \rightarrow p(\theta) = \frac{\beta_0^{\alpha_0}}{\Gamma(\alpha_0)} \theta^{-(\alpha_0+1)} \exp\left(-\frac{\beta_0}{\theta}\right)$$

Posterior :

$P(\theta|y) \propto p(y|\theta)p(\theta)$ $\quad$ $y = (y_1, \dots, y_i)$ is a sequence of i.i.d. observations

$$= \left(\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\theta}} \exp\left(-\frac{(y_i - \mu)^2}{2\theta}\right)\right) \times \frac{\beta_0^{\alpha_0}}{\Gamma(\alpha_0)} \theta^{-(\alpha_0+1)} \exp\left(-\frac{\beta_0}{\theta}\right)$$

$$\propto \left(\prod_{i=1}^{n} \theta^{-\frac{1}{2}} \exp\left(-\frac{(y_i - \mu)^2}{2\theta}\right)\right) \times \theta^{-(\alpha_0+1)} \exp\left(-\frac{\beta_0}{\theta}\right)$$

$$= \theta^{-\left(\alpha_0+1+\frac{n}{2}\right)} \exp\left(-\left(\frac{\beta_0}{\theta} + \frac{\sum_{i=1}^{n}(y_i - \mu)^2}{2\theta}\right)\right)$$

$$= \theta^{-\left(\alpha_0+\frac{n}{2}+1\right)} \exp\left(-\left(\frac{\beta_0 + \frac{1}{2}\sum_{i=1}^{n}(y_i - \mu)^2}{\theta}\right)\right)$$

$$= \mathrm{Inv} - \mathrm{Gamma}\left(\theta \middle| \alpha_0 + \frac{n}{2}, \beta_0 + \frac{1}{2}\sum_{i=1}^{n}(y_i - \mu)^2\right)$$

Multi-Parameters Model

Multinomial Likelihood-Dirichlet Prior

Likelihood :

$$p(y|\theta) = \text{Multin}(y|n, \theta_1, \dots, \theta_k) = \binom{n}{y_1\ y_2 \cdots y_k} {\theta_1}^{y_1} \cdots {\theta_k}^{y_k}$$

$$= \frac{\Gamma(n+1)}{\prod_{j=1}^{k} \Gamma(y_j+1)} \prod_{j=1}^{k} \theta_j^{y_j}$$

$y = (y_1, \dots, y_j, \dots, y_k)$

$y_j \in \{0,1, \dots, n\}, \sum_{j=1}^{k} y_j = n$

Prior:

$$\theta \sim \text{Dirichlet}(\alpha_1, \dots, \alpha_k) \rightarrow p(\theta) = \frac{\Gamma(\alpha_1 + \cdots + \alpha_k)}{\Gamma(\alpha_1) \cdots \Gamma(\alpha_k)} \theta_1^{\alpha_1 - 1} \cdots \theta_k^{\alpha_k - 1}$$

$$= \frac{\Gamma(\sum_{j=1}^{k} \alpha_j)}{\prod_{j=1}^{k} \Gamma(\alpha_k)} \theta_1^{\alpha_1 - 1} \cdots \theta_k^{\alpha_k - 1}$$

$\sum_{j=1}^{k} \theta_j = 1$

Posterior :

$$P(\theta|y) \propto P(y|\theta)p(\theta)$$

$$= \frac{\Gamma(n+1)}{\prod_{j=1}^{k} \Gamma(y_j+1)} \prod_{j=1}^{K} \theta_j^{y_j} \frac{\Gamma(\sum_{j=1}^{k} \alpha_j)}{\prod_{j=1}^{k} \Gamma(\alpha_j)} \theta_1^{\alpha_1 - 1} \cdots \theta_k^{\alpha_k - 1}$$

$$\propto \prod_{j=1}^{k} \theta_j^{y_j} \prod_{j=1}^{k} \theta_j^{\alpha_j - 1}$$

$$\propto \prod_{j=1}^{k} \theta_j^{\alpha_j + y_j - 1}$$

$$= \text{Dirichlet}(\alpha_1 + y_1, \dots, \alpha_k + y_k)$$

Multinomial Likelihood-Dirichlet Prior

Posterior predictive distribution

$$p(\hat{y}|y) = \int_\theta p(\hat{y},\theta|y)d\theta = \int_\theta p(\hat{y}|\theta,y)p(\theta|\mathrm{y})d\theta = \int_\theta p(\hat{y}|\theta)p(\theta|\mathrm{y})d\theta$$

$$\begin{aligned}
p(\hat{y}|y) &= \int_\theta p(\hat{y}|\theta_1,\dots,\theta_k)p(\theta_1,\dots,\theta_k|\mathrm{y})d\theta \\
&= \int_\theta \frac{\Gamma(n+1)}{\prod_{j=1}^{k}\Gamma(y_j+1)}\prod_{j=1}^{k}\theta_j^{y_j}\,\frac{\Gamma(\sum_{j=1}^{k}\alpha_j)}{\prod_{j=1}^{k}\Gamma(\alpha_j)}\prod_{j=1}^{k}\theta_j^{\alpha_j-1}\,d\theta \\
&= \frac{\Gamma(n+1)}{\prod_{j=1}^{K}\Gamma(y_j+1)}\frac{\Gamma(\sum_{j=1}^{K}\alpha_j)}{\prod_{j=1}^{K}\Gamma(\alpha_j)}\int_\theta \prod_{j=1}^{K}\theta_j^{y_j+\alpha_j-1}\,d\theta \\
&= \frac{\Gamma(n+1)}{\prod_{j=1}^{K}\Gamma(y_j+1)}\frac{\Gamma(\sum_{j=1}^{K}\alpha_j)}{\prod_{j=1}^{K}\Gamma(\alpha_j)}\frac{\prod_{j=1}^{K}\Gamma(y_j+\alpha_j)}{\Gamma(n+\sum_{j=1}^{K}\alpha_j)}\int_\theta \frac{\Gamma(n+\sum_{j=1}^{K}\alpha_j)}{\prod_{j=1}^{K}\Gamma(y_j+\alpha_j)}\prod_{j=1}^{K}\theta_j^{y_j+\alpha_j-1}\,d\theta \\
&= \frac{\Gamma(n+1)}{\prod_{j=1}^{K}\Gamma(y_j+1)}\frac{\Gamma(\sum_{j=1}^{K}\alpha_j)}{\prod_{j=1}^{K}\Gamma(\alpha_j)}\frac{\prod_{j=1}^{K}\Gamma(y_j+\alpha_j)}{\Gamma(n+\sum_{j=1}^{K}\alpha_j)}
\end{aligned}$$