STT 465
Single-parameter models: Beta-Binomial & Poisson

(G. de los Campos)

Beta-Binomial

OUTLINE:

⟹ Elements of the models:

(1) Sampling model $p(Y|\theta)$

(2) Prior distribution $p(\theta)$

(3) From (1) and (2) and using Bayes Rule, we derive the posterior distribution of the parameter given the data, $p(\theta|Y)$

(4) In this case the posterior distribution has a recognizable form.

⟹ Inference

- Posterior mean and posterior variance
- Is the Bayesian estimator unbiased?
- What happens as sample size increases?
  (consider both the effects on bias and variance)
- Posterior credibility regions (interpretation, types,…)

.

Beta-Binomial

⇒ Sampling model

$$p(y_i \mid \theta) = \theta^{y_i} (1-\theta)^{1-y_i}$$

[Assuming IID]

$$p(y_1, \ldots, y_n \mid \theta) = \prod_{i=1}^{n} \theta^{y_i} (1-\theta)^{1-y_i} = \theta^{n_1} (1-\theta)^{n_0} \qquad n_1 = \sum_{i=1}^{n} y_i \;\; ; \;\; n_0 = n - n_1$$

⇒ Prior: we will consider a Beta distribution (the uniform is a special case)

$$p(\theta \mid \alpha, \beta) = \frac{\theta^{\alpha-1} (1-\theta)^{\beta-1}}{B(\alpha, \beta)} \quad ; \; B(\alpha, \beta) = \int_0^1 x^{\alpha-1} (1-x)^{\beta-1} dx = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

$$p(\theta \mid \alpha, \beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha-1} (1-\theta)^{\beta-1}$$

⇒ Discuss: Kernel vs Integrating constant.

Posterior Distribution

$\Rightarrow$ According to Bayes' rule

$$p(\theta \mid y_1,\ldots,y_n) = \frac{p(y_1,\ldots,y_n \mid \theta)\,p(\theta)}{p(y_1,\ldots,y_n)} \propto p(y_1,\ldots,y_n \mid \theta)\,p(\theta)$$

$$p(\theta \mid y_1,\ldots,y_n) \propto \theta^{n_1}(1-\theta)^{n_0}\,\frac{\theta^{\alpha-1}(1-\theta)^{\beta-1}}{B(\alpha,\beta)} \propto \theta^{\alpha+n_1-1}(1-\theta)^{\beta+n_0-1}$$

$\Rightarrow$ Finding the integrating constant

$$\int_0^1 \frac{\theta^{\alpha+n_1-1}(1-\theta)^{\beta+n_0-1}}{c(y)}\,d\theta = 1$$

$$\frac{1}{c(y)}\int_0^1 \theta^{\alpha+n_1-1}(1-\theta)^{\beta+n_0-1}\,d\theta = 1$$

$$\int_0^1 \theta^{\alpha+n_1-1}(1-\theta)^{\beta+n_0-1}\,d\theta = c(y) = B(\alpha+n_1,\beta+n_0,)$$

Posterior Distribution

$\Rightarrow$ Therefore

$$p\left(\theta \mid y_1,\ldots,y_n\right)=\frac{p\left(y_1,\ldots,y_n \mid \theta\right)p\left(\theta\right)}{p\left(y_1,\ldots,y_n\right)}\propto p\left(y_1,\ldots,y_n \mid \theta\right)p\left(\theta\right)$$

$$p\left(\theta \mid y_1,\ldots,y_n\right)=\frac{\theta^{\alpha+n_1-1}\left(1-\theta\right)^{\beta+n_0-1}}{B(\alpha+n_1,\beta+n_0)}=Beta\left(\theta \mid \alpha+n_1,\beta+n_0\right)$$

$\Rightarrow$ Because the posterior distribution has the same form as that of the prior, we say that the Beta prior is Conjugate to the Binomial likelihood.

$\Rightarrow$ Discuss:

- Posterior Vs. Prior Mean
- Posterior variance Vs sample size
- Posterior credibility regions (definition & interpretation).

Poisson Model: Likelihood Analyses

⇒ Data: count (y_i=0,1,2,....)

⇒ Probability model $p(y_i \mid \theta) = \dfrac{\theta^{y_i} e^{-\theta}}{y_i!} \quad > 0$

[Question: What part is the kernel and what part is the integrating constant?]
[Question: How would you find the integrating constant if are give the kernel?]

⇒ Expected value and variance: $E(y_i) = Var(y_i) = \theta$
[Question: how would you determine the E[] and Var[]?]
[Question: what is the coef. of variation sd(y)/|E(y)|?]

⇒ Joint distribution of n iid draws from a Poisson model (Sampling Model)

$$p(y_1,\dots,y_n \mid \theta) = \prod_{i=1}^{n} \frac{\theta^{y_i} e^{-\theta}}{y_i!} = \left[e^{-n\theta} \theta^{\sum_{i=1}^{n} y_i} \right] \prod_{i=1}^{n} \frac{1}{y_i!} = e^{-n\theta} \theta^{n\bar{y}} c(y)$$

[Discuss: sufficient statistic]
[Discuss: how to get MLE of theta?; Does it have a closed form?]
[Discuss: what is the sampling variance of the MLE estimator?]
[Discuss: how would you provide an approximate 95% CI for the MLE estimate?]

Poisson Model: In search for a conjugate prior

$\Rightarrow$ Likelihood

$$p(y_1,\ldots,y_n \mid \theta) \propto e^{-n\theta}\theta^{n\bar{y}}$$

$\Rightarrow$ We may guess that a conjugate prior may have this form:

$$p(\theta) \propto e^{-\tau_1\theta}\theta^{\tau_2}$$

$\Rightarrow$ If we have something like that, the posterior will have the following kernel

$$\begin{aligned} p(\theta \mid y_1,\ldots,y_n) &\propto p(y_1,\ldots,y_n \mid \theta)\, p(\theta) \\ &\propto e^{-n\theta}\theta^{n\bar{y}} \times e^{-\tau_1\theta}\theta^{\tau_2} \\ &\propto e^{-\theta(\tau_1+n)}\theta^{\tau_2+n\bar{y}} \end{aligned}$$

$\Rightarrow$ We have such a form in the Gamma distribution (a=shape; b=rate)

$$p(\theta) = \frac{b^a}{\Gamma(a)}\theta^{(a-1)}e^{-\theta b}$$

[Discuss: (i) alternative parameterizations, (ii) mean and variance]

Poisson Model: Posterior Density

⟹ According to Bayes Rule

$$p(\theta \mid y_1,\ldots,y_n) = \frac{p(y_1,\ldots,y_n \mid \theta)\, p(\theta)}{p(y_1,\ldots,y_n)}$$

$$\propto p(y_1,\ldots,y_n \mid \theta)\, p(\theta)$$

$$p(\theta) = \frac{b^a}{\Gamma(a)} e^{-\theta(a-1)} \theta^{b}$$

$$p(y_1,\ldots,y_n \mid \theta) \propto e^{-n\theta} \theta^{n\bar{y}}$$

⟹ In the Gamma-Poisson Model we have

$$p(\theta \mid y_1,\ldots,y_n) \propto \left[e^{-n\theta} \theta^{n\bar{y}} \right] \times \left[\theta^{(a-1)} e^{-\theta b} \right]$$

$$\propto e^{-\theta(n+b)}\, \theta^{n\bar{y}+a-1}$$

⟹ This can be recognized as the kernel of a Gamma distribution with the following parameters

rate: $\tilde{a} = n\bar{y} + a$

shape: $\tilde{b} = n + b$

Poisson Model: Posterior Distribution

⟹ Posterior Mean

$$E(\theta \mid y_1,\ldots,y_n) = \frac{n\bar{y}+a}{n+b}$$

⟹ Posterior Variance

$$V(\theta \mid y_1,\ldots,y_n) = \frac{n\bar{y}+a}{(n+b)^2}$$

[Discuss: what happens as n➔∞?]

Predictive Distribution

$\Rightarrow$ We have assumed that draws from the population, before we observe any data, are IID Poisson.

$\Rightarrow$ What is the distribution of a draw from the population after we have observed our sample, that is

$$p\left(y_f \mid y_1,\dots,y_n\right)?$$

$\Rightarrow$ Naïve approach (we can also think of this as a plug-in approach): estimate theta from the posterior distribution and plug that estimate in the Poisson model, that yields

$$p\left(y_f \mid y_1,\dots,y_n\right) = Poissson\left(\hat{\theta}_B\right)$$

$$\text{where: } \hat{\theta}_B = \frac{n\bar{y}+a}{n+b}$$

$\Rightarrow$ What is wrong about this?: the ‘plug-in approach does not account for uncertainty about Θ as described by our posterior distribution’.

$\Rightarrow$ The Bayesian framework allows us to arrive at the correct answer, and we will see how we can do this analytically and using MC methods.

Predictive Distribution (Poisson-Gamma model)

⇒ Using standard probability rules

$$
\begin{aligned}
p\left(y_f \mid y_1,\dots,y_n\right) &= \int p\left(y_f,\theta \mid y_1,\dots,y_n\right)d\theta \\
&= \int p\left(y_f \mid \theta, y_1,\dots,y_n\right)p\left(\theta \mid y_1,\dots,y_n\right)d\theta \\
&= \int Poisson\left(y_f \mid \theta\right)Gamma\left(\theta \mid y_1,\dots,y_n\right)d\theta
\end{aligned}
$$

⇒ The above is called the predictive distribution. Let's discuss it conceptually:

- For each possible value of Θ, the conditional likelihood is Poisson,
- The predictive distribution is an average (this is what the integral is) of all possible Poisson distributions, where the average is taken with respect to the posterior distribution of Θ given the data.

Predictive Distribution (Poisson-Gamma model)

⇒ Analytical derivation

$$p\left(y_f \mid y_1,\dots,y_n\right) = \int Poisson\left(y_f \mid \theta\right) Gamma\left(\theta \mid y_1,\dots,y_n\right) d\theta$$

[remove anything that does not involve θ or y_f]

$$\propto \int \frac{e^{-\theta}\theta^{y_f}}{y_f!} \times e^{-\theta(n+b)}\theta^{(n\bar{y}+a)-1}\, d\theta$$

[take y_f outside of the integral a

$$\propto \frac{1}{y_f!}\int e^{-\theta}\theta^{y_f} \times e^{-\theta(n+b)}\theta^{(n\bar{y}+a)-1}\, d\theta$$

[combine terms]

$$\propto \frac{1}{y_f!}\int e^{-\theta(n+1+b)}\theta^{(n\bar{y}+y_f+a)}\, d\theta$$

[the expression iside the integral is the kernel of the Gamma density, we know the integrating constatn]

$$\propto \frac{\Gamma\left(n\bar{y}+y_f+a\right)}{\Gamma\left(y_f+1\right)}\left(n+1+b\right)^{-(n\bar{y}+y_f+a)}$$

[remove elements that are part of the integrating constant, and recall $\Gamma\left(t\right)=(t-1)!$]

$$\propto \frac{\left(n\bar{y}+a+1+y_f\right)!}{y_f!}\left(n+1+\beta\right)^{-y_f}$$

[compare with the kernel of the negative binomial]

$\beta = n+b \quad ; \quad \alpha = a+n\bar{y}$

Some Questions for HW 2

- What are the mean and variances of the predictive distribution?

- Compare that with the mean and variance of the ‘plug-in distribution’

- Plot the predictive distribution and the naïve predictive distribution?

Data (use 1st column): http://www.jstatsoft.org/v16/i09/