Predictive Inference Tools for Researchers

by

Voyze G. Harris III

A Thesis Submitted to the Faculty of the

Statistics and Data Science Graduate Interdisciplinary Program

In Partial Fulfillment of the Requirements
For the Degree of

Master of Science

In the Graduate College

The University of Arizona

2021

THE UNIVERSITY OF ARIZONA
GRADUATE COLLEGE

As members of the Master's Committee, we certify that we have read the thesis prepared by Voyze Gabriel Harris III, titled *[Enter Thesis Title]* and recommend that it be accepted as fulfilling the dissertation requirement for the Master's Degree.

____ Date: _____________
Dr. Dean Billheimer

____ Date: _____________
Dr. Edward Bedrick

____ Date: _____________
Dr. Walter Piegorsch

Final approval and acceptance of this thesis is contingent upon the candidate's submission of the final copies of the thesis to the Graduate College.

I hereby certify that I have read this thesis prepared under my direction and recommend that it be accepted as fulfilling the Master's requirement.

____ Date: _____________
Dr. Dean Billheimer
Master's Thesis Committee Chair
Biostatistics

Contents

1 Thesis Abstract

- (paragraph) Statement of the thesis topic and objectives
- (paragraph) Explanation of R package

2 Introduction: Predictive Inference

2.1 Why is predictive inference important?

2.2 Difference between parametric inference and predictive inference

2.2.1 When is predictive inference more useful?

2.2.2 When is parametric inference more useful?

[examples, comparisons]

2.3 The Bayesian Parametric Prediction Format

[Geisser p. 49]

Let

$$f\left(x^{(N)}, x_{(M)}|\theta\right) = f\left(x_{(M)}|x^{(N)}, \theta\right) f\left(x^{(N)}|\theta\right).$$

Here $x^{(N)}$ represents observed events and $x_{(M)}$ are future events. We calculate

$$f\left(x_{(M)}, x^{(N)}\right) = \int f\left(x^{(N)}, x_{(M)}|\theta\right) p(\theta) d\theta$$

where $p(\theta)$ is the prior density and

$$f\left(x_{(M)}|x^{(N)}\right) = \frac{f\left(x_{(M)}, x^{(N)}\right)}{f\left(x^{(N)}\right)} = \int f\left(x_{(M)}|\theta\right) p\left(\theta|x^{(N)}\right) d\theta$$

where

$$p\left(\theta|x^{(N)}\right) \propto f\left(x^{(N)}|\theta\right) p(\theta).$$

2.4 [Maybe] Example of Difference between results from Plug-in estimator and results using Predictive Inference

3 Chapter 1: Predictive Problems with Conjugate Priors

[Problems with closed-form solutions. These problems will be what the R package is designed for. Use problems from Geisser, Casella & Berger (Bayesian chapter), other sources. Regression problem–predictive distributions of models that include and exclude some predictor]

3.1 Prediction of Future Successes: Beta-Binomial (Geisser p. 73)

3.1.1 Derivation

Let X_i be independent binary variables with $\Pr(X_i = 1) = \theta$, and let $T = \sum X_i$. Then T has probability

$$\binom{N}{t}\theta^t(1-\theta)^{N-t}.$$

Assume $\theta \sim \text{Beta}(\alpha, \beta)$, so

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

Then

$$p\left(\theta|X^{(N)}\right) = \frac{\Gamma(N+\alpha+\beta)\theta^{t+\alpha-1}(1-\theta)^{N-t+\beta-1}}{\Gamma(t+\alpha)\Gamma(N-t+\beta)}$$

So for $R = \sum_{i=1}^{M} X_{N+i}$ we have Beta-Binomial predictive distribution

$$\Pr[R = r|t] = \int \binom{M}{r}\theta^r(1-\theta)^{M-r}p\left(\theta|X^{(N)}\right)d\theta$$

$$= \binom{M}{r}\int \theta^r(1-\theta)^{M-r}\frac{\Gamma(N+\alpha+\beta)}{\Gamma(t+\alpha)\Gamma(N-t+\beta)}\theta^{t+\alpha-1}(1-\theta)^{N-t+\beta-1}d\theta$$

$$= \frac{M!}{r!(M-r)!}\frac{\Gamma(N+\alpha+\beta)}{\Gamma(t+\alpha)\Gamma(N-t+\beta)}\int \theta^{r+t+\alpha-1}(1-\theta)^{M-r+N-t+\beta-1}d\theta$$

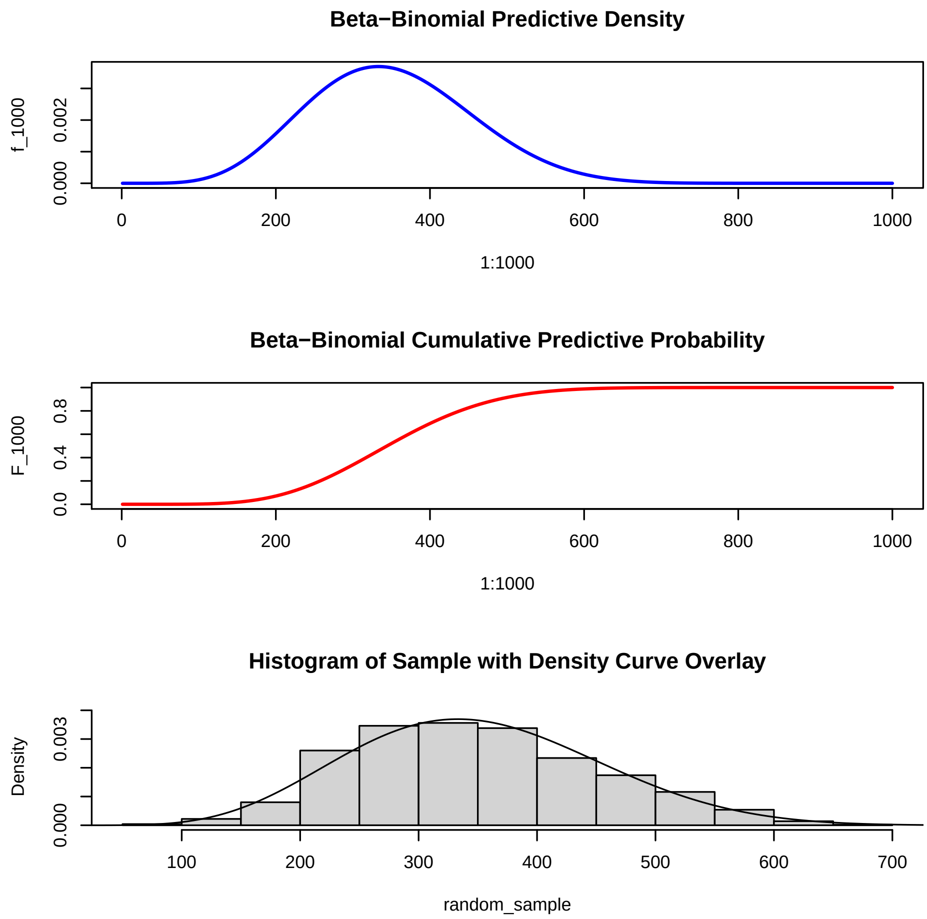
$$= \frac{\Gamma(M+1)\Gamma(N+\alpha+\beta)\Gamma(r+t+\alpha)\Gamma(M-r+N-t+\beta)}{\Gamma(r+1)\Gamma(M-r+1)\Gamma(t+\alpha)\Gamma(N-t+\beta)\Gamma(M+N+\alpha+\beta)}$$

3.1.2 R Implementation

This result has been used to create "standard" R functions dpredBB(), ppredBB(), and rpredBB() for the Beta-Binomial distribtuion for density, cumulative probability, and random sampling, respectively (see appendix). These functions are exercised in the following example.

3.1.3 Example

Suppose $t = 5$ successes have been observed out of $N = 10$ binary events, $\alpha = 2$ and $\beta = 8$. For $M = 1000$ future observations, the figures below show the predictive distribution from dpredBB(), the cumulative distribution from ppredBB(), and a histogram of random draws from rpredBB().



3.2 Survival Time: Exponential-Gamma (Geisser p. 74)

3.2.1 Derivation

Suppose $X^{(N)} = \left(X^{(d)}, X^{(N-d)}\right)$ where $X^{(d)}$ represents copies fully observed from an exponential survival time density

$$f(x|\theta) = \theta e^{-\theta x}$$

and $X^{(N-d)}$ represents copies censored at $x_{d+1}, ..., x_N$, respectively. Hence

$$L(\theta) \propto \theta^d e^{-\theta N\bar{x}}$$

when $N\bar{x} = \sum_1^N x_i$, as shown below.

The usual exponential likelihood is used for the fully observed copies, whereas for the censored copies we need $\Pr(x > \theta) = 1 - \Pr(x \leq \theta) = 1 - F(x|\theta) = 1 - (1 - e^{-\theta x}) = e^{-\theta x}$. Thus the overall likelihood is

$$L(\theta|x) = \prod_{i=1}^{d} \theta e^{-\theta x_i} \prod_{i=d+1}^{N} e^{-\theta x_i} = \theta^d e^{-\theta N\bar{x}}$$

Assuming a Gamma(δ, γ) prior for θ,

$$p(\theta) = \frac{\gamma^\delta \theta^{\delta-1} e^{-\gamma\theta}}{\Gamma(\delta)}$$

we obtain the posterior

$$p\left(\theta|X^{(N)}\right) = \frac{p\left(x^{(N)}|\theta\right) p(\theta)}{\int p\left(X^{(N)}|\theta\right) p(\theta) d\theta}$$

$$= \frac{\theta^d e^{-\theta N\bar{x}} \cdot \frac{\gamma^\delta \theta^{\delta-1} e^{-\gamma\theta}}{\Gamma(\delta)}}{\int \left(\theta^d e^{-\theta N\bar{x}} \cdot \frac{\gamma^\delta \theta^{\delta-1} e^{-\gamma\theta}}{\Gamma(\delta)}\right) d\theta}$$

$$= \frac{\cancel{\frac{\gamma^\delta}{\Gamma(\delta)}} \left(\theta^{d+\delta-1} e^{-\theta(\gamma+N\bar{x})}\right)}{\cancel{\frac{\gamma^\delta}{\Gamma(\delta)}} \int \left(\theta^{d+\delta-1} e^{-\theta(\gamma+N\bar{x})}\right) d\theta}$$

$$= \frac{\frac{(\gamma+N\bar{x})^{d+\delta}}{\Gamma(d+\delta)} \left(\theta^{d+\delta-1} e^{-\theta(\gamma+N\bar{x})}\right)}{\cancel{\frac{(\gamma+N\bar{x})^{d+\delta}}{\Gamma(d+\delta)} \int \left(\theta^{d+\delta-1} e^{-\theta(\gamma+N\bar{x})}\right) d\theta}}$$

$$= \frac{(\gamma + N\bar{x})^{d+\delta} \theta^{d+\delta-1} e^{-\theta(\gamma+N\bar{x})}}{\Gamma(d+\delta)}$$

with the Gamma$(d + \delta, \gamma + N\bar{x})$ density in the next to last step integrating to 1.

Thus the survival time predictive probability is

$$
\begin{aligned}
P\left(X=x|\theta,X^{(N)}\right) &= \int p\left(\theta|X^{(N)}\right)p(x|\theta)d\theta \\
&= \int \frac{(\gamma+N\bar{x})^{d+\delta}\theta^{d+\delta-1}e^{-\theta(\gamma+N\bar{x})}}{\Gamma(d+\delta)}\cdot\theta e^{-\theta x}d\theta \\
&= (d+\delta)(\gamma+N\bar{x})^{d+\delta}\int\frac{\theta^{(d+\delta+1)-1}e^{-\theta(\gamma+N\bar{x}+x)}}{(d+\delta)\Gamma(d+\delta)}d\theta \\
&= \frac{(d+\delta)(\gamma+N\bar{x})^{d+\delta}}{(\gamma+N\bar{x}+x)^{d+\delta+1}}\int\frac{(\gamma+N\bar{x}+x)^{d+\delta+1}\,\theta^{(d+\delta+1)-1}e^{-\theta(\gamma+N\bar{x}+x)}}{\Gamma(d+\delta+1)}d\theta \\
&= \frac{(d+\delta)(\gamma+N\bar{x})^{d+\delta}}{(\gamma+N\bar{x}+x)^{d+\delta+1}}
\end{aligned}
$$

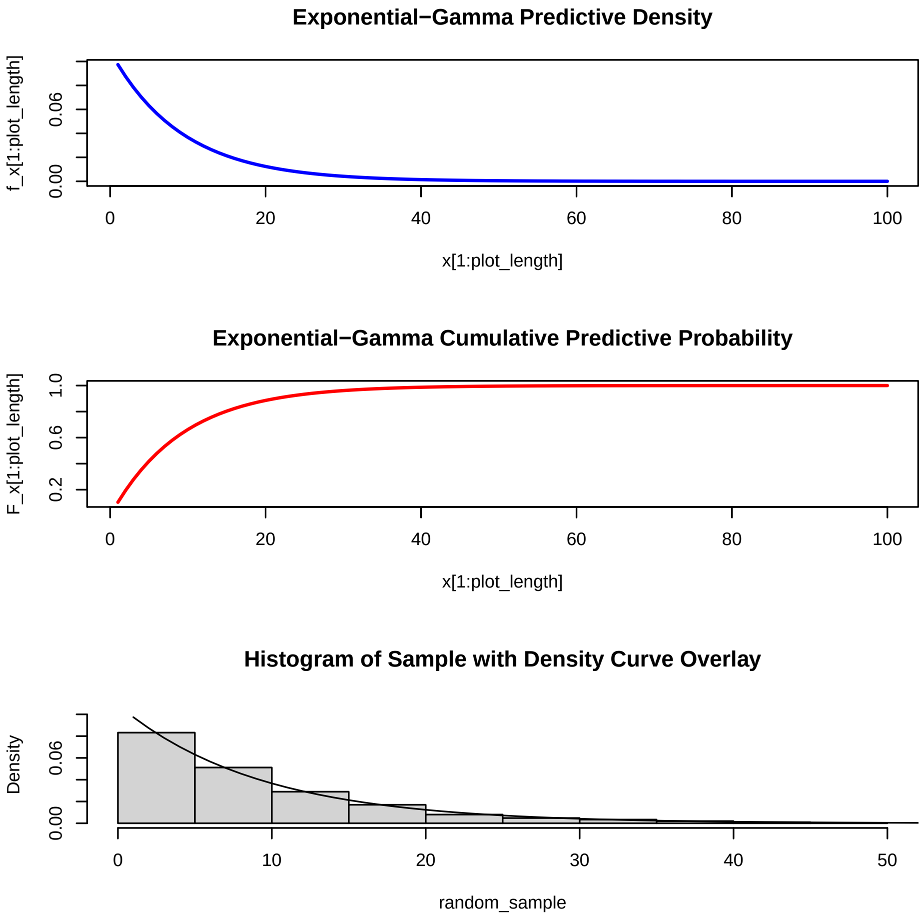
(simplifying by constructing a Gamma$(d+\delta+1,\gamma+N\bar{x}+x)$ density in the final integrand.)

3.2.2 R Implementation

This result has been used to create standard format R functions dpredEG(), ppredEG(), and rpredEG() for the Gamma-Exponential distribtuion for density, cumulative probability, and random sampling, respectively (see appendix). These functions are exercised in the following example.

3.2.3 Example

Suppose $d = 800$ out of $N = 1000$ copies have been observed, and the remaining 200 censored. Say $\delta = 20$, $\gamma = 5$, and we are interested in the number of survivors out of $M = 1000$ future observations. The figures below illustrate the predictive probability using dpredEG() and rpredEG(), along with a histogram of a random sample taken using rpredEG().


Exponential−Gamma Predictive Density
f_x[1:plot_length]
0.00
0.06
0
20
40
60
80
100
x[1:plot_length]
Exponential−Gamma Cumulative Predictive Probability
F_x[1:plot_length]
0.2
0.6
1.0
0
20
40
60
80
100
x[1:plot_length]
Histogram of Sample with Density Curve Overlay
Density
0.00
0.06
0
10
20
30
40
50
random_sample

3.3 Poisson-Gamma Model (Hoff p. 43ff)

3.3.1 Derivation

[using Hoff's notation and variable names below. Should I convert this to Geisser's $x^{(N)}, x_{(M)}$ convention for uniformity throughout my thesis?]

Suppose $Y_1, ..., Y_n|\theta \overset{i.i.d.}{\sim}$ Poisson(θ) with Gamma prior $\theta \sim$ Gamma(α, β). That is,

$$\begin{aligned} P\left(Y_1 = y_1, ..., Y_n = y_n|\theta\right) &= \prod_{i=1}^{n} p\left(y_i|\theta\right) \\ &= \prod_{i=1}^{n} \frac{1}{y!} \theta^{y_i} e^{-\theta} \\ &= \left(\prod_{i=1}^{n} \frac{1}{y!}\right) \theta^{\sum y_i} e^{-n\theta} \\ &= c\left(y_1, ..., y_n\right) \theta^{\sum y_i} e^{-n\theta} \end{aligned}$$

and

$$p(\theta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} e^{-\beta\theta}, \theta, \alpha, \beta > 0.$$

Then we have posterior distribution

$$\begin{aligned} p\left(\theta|y_1, ..., y_n\right) &= \frac{p\left(y_1, ..., y_n|\theta\right) p(\theta)}{\int_\theta p\left(y_1, ..., y_n|\theta\right) p(\theta)} \\ &= \frac{p\left(y_1, ..., y_n|\theta\right) p(\theta)}{p\left(y_1, ..., y_n\right)} \\ &= \frac{1}{p\left(y_1, ..., y_n\right)} \theta^{\sum y_i} e^{-n\theta} \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} e^{-\beta\theta} \\ &= C\left(y_1, ..., y_n, \alpha, \beta\right) \theta^{\alpha+\sum y_i - 1} e^{-(\beta+n)\theta} \\ &\sim \text{Gamma}\left(\alpha + \sum y_i, \beta + n\right). \end{aligned}$$

Here

$$C\left(y_1,...,y_n,\alpha,\beta\right) = \frac{1}{p\left(y_1,...,y_n\right)} \cdot \frac{\beta^\alpha}{\Gamma(\alpha)}$$

$$= \frac{1}{\int_\theta p\left(y_1,...,y_n|\theta\right)p(\theta)} \cdot \frac{\beta^\alpha}{\Gamma(\alpha)}$$

$$= \frac{1}{\int_\theta \left(\prod \frac{1}{y_i!}\right)\theta^{\sum y_i} e^{-n\theta}\cancel{\left(\frac{\beta^\alpha}{\Gamma(\alpha)}\right)}\theta^{\alpha-1}e^{-\beta\theta}} \cdot \cancel{\left(\frac{\beta^\alpha}{\Gamma(\alpha)}\right)}$$

$$= \frac{1}{\left(\prod \frac{1}{y_i!}\right)\frac{\Gamma(\alpha+\sum y_i)}{(\beta+n)^{\alpha+\sum y_i}}\int_\theta \frac{(\beta+n)^{\alpha+\sum y_i}}{\Gamma(\alpha+\sum y_i)}\theta^{\sum y_i+\alpha-1}e^{-(\beta+n)\theta}}$$

$$= \frac{\prod_{i=1}^n y_i!(\beta+n)^{\alpha+\sum y_i}}{\Gamma(\alpha+\sum y_i)}$$

Call this constant C_n (for n observations).

Note that an additional observation $y_{n+1} = \tilde{y}$ the constant becomes

$$C_{n+1} = \frac{\prod_{i=1}^{n+1} y_i!(\beta+n+1)^{\alpha+\sum_{i=1}^{n+1} y_i}}{\Gamma(\alpha+\sum_{i=1}^{n+1} y_i)}.$$

Also note that the marginal joint distribution of k observations is

$$p\left(\tilde{y}|y_1,...,y_k\right) = \frac{1}{C_k}\frac{\beta^\alpha}{\Gamma(\alpha)}.$$

For future observation $\tilde{y}$, then, we compute predictive distribution

$$p\left(\tilde{y}|y_1,...,y_n\right) = \frac{p\left(y_1,...,y_n,\tilde{y}\right)}{p\left(y_1,...,y_n\right)} = \frac{p\left(y_1,...,y_{n+1}\right)}{p\left(y_1,...,y_n\right)} = \frac{\frac{1}{C_{n+1}}\cancel{\frac{\beta^\alpha}{\Gamma(\alpha)}}}{\frac{1}{C_n}\cancel{\frac{\beta^\alpha}{\Gamma(\alpha)}}} = \frac{C_n}{C_{n+1}}$$

$$= \frac{\frac{\prod_{i=1}^n y_i!(\beta+n)^{\alpha+\sum_{i=1}^n y_i}}{\Gamma(\alpha+\sum_{i=1}^n y_i)}}{\frac{\prod_{i=1}^{n+1} y_i!(\beta+n+1)^{\alpha+\sum_{i=1}^{n+1} y_i}}{\Gamma(\alpha+\sum_{i=1}^{n+1} y_i)}}$$

$$= \frac{\Gamma\left(\alpha+\sum_{i=1}^{n+1} y_i\right)(\beta+n)^{\alpha+\sum_{i=1}^n y_i}}{(y_{n+1}!)\,\Gamma\left(\alpha+\sum_{i=1}^n y_i\right)(\beta+n+1)^{\alpha+\sum_{i=1}^{n+1} y_i}}$$

$$= \frac{\Gamma\left(\alpha+\sum_{i=1}^n y_i+\tilde{y}\right)(\beta+n)^{\alpha+\sum_{i=1}^n y_i}}{(\tilde{y}!)\,\Gamma\left(\alpha+\sum_{i=1}^n y_i\right)(\beta+n+1)^{\alpha+\sum_{i=1}^n y_i+\tilde{y}}}$$

$$= \frac{\Gamma\left(\alpha+\sum y_i+\tilde{y}\right)}{\Gamma(\tilde{y}+1)\Gamma(\alpha+\sum y_i)} \cdot \left(\frac{\beta+n}{\beta+n+1}\right)^{\alpha+\sum y_i} \cdot \left(\frac{1}{\beta+n+1}\right)^{\tilde{y}}$$

This is a negative binomial distribution: $\tilde{y} \sim NB\left(\alpha+\sum y_i, \beta+n\right)$.

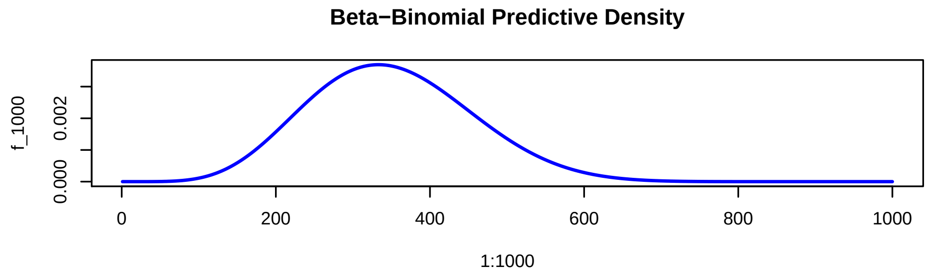
$$\theta \sim NB(\alpha,\beta) \Rightarrow p(\theta) = \binom{\theta+\alpha-1}{\alpha-1}\left(\frac{\beta}{\beta+1}\right)^{\alpha}\left(\frac{1}{\beta+1}\right)^{\theta}$$

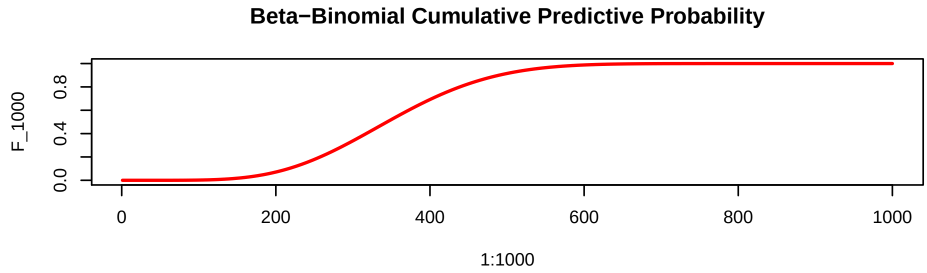
so

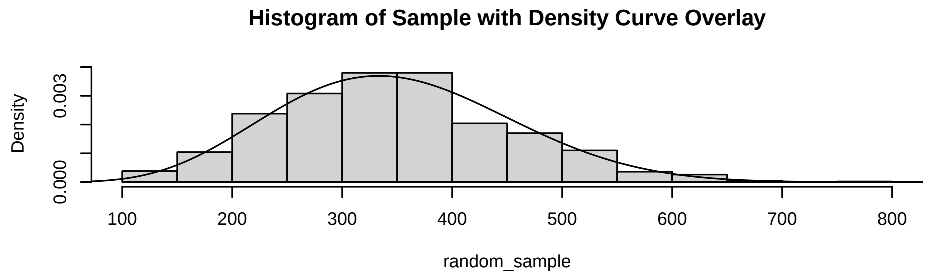
$$\tilde{y} \sim NB\left(\alpha+\sum y_i), \beta+n\right) \Rightarrow p(\tilde{y}) = \binom{\tilde{y}+\alpha+\sum y_i-1}{\alpha+\sum y_i-1}\left(\frac{\beta+n}{\beta+n+1}\right)^{\alpha+\sum y_i}\left(\frac{1}{\beta+n+1}\right)^{\tilde{y}}$$

[This is the result in Hoff. The straightforward derivation below is off by a constant multiple. Need to figure out what went awry.]

$$
\begin{aligned}
p\left(\tilde{y}|y_1,...,y_n\right) &= \int_0^\infty p\left(\tilde{y}|\theta, y_1,...,y_n\right) p\left(\theta|y_1,...,y_n\right) d\theta \\
&= \int p\left(\tilde{y}|\theta\right) p\left(\theta|y_1,...,y_n\right) d\theta \\
&= C \int \left(\frac{1}{\tilde{y}!}\theta^{\tilde{y}} e^{-\theta}\right) \theta^{\alpha+\sum y_i - 1} e^{-(\beta+n)\theta} d\theta \\
&= \frac{C}{\tilde{y}!} \int \theta^{\tilde{y}+\alpha+\sum y_i - 1} e^{-(\beta+n+1)\theta} d\theta \\
&= \frac{C\Gamma\left(\tilde{y}+\alpha+\sum y_i\right)}{\Gamma\left(\tilde{y}+1\right)\left(\beta+n+1\right)^{\tilde{y}+\alpha+\sum y_i}} \int \frac{\left(\beta+n+1\right)^{\tilde{y}+\alpha+\sum y_i}}{\Gamma\left(\tilde{y}+\alpha+\sum y_i\right)} \theta^{\tilde{y}+\alpha+\sum y_i - 1} e^{-(\beta+n+1)\theta)} d\theta \\
&= C \cdot \frac{\Gamma\left(\tilde{y}+\alpha+\sum y_i\right)}{\Gamma\left(\tilde{y}+1\right)\left(\beta+n+1\right)^{\tilde{y}+\alpha+\sum y_i}} \\
&= \frac{\prod_{i=1}^n y_i!(\beta+n)^{\alpha+\sum y_i}}{\Gamma(\alpha+\sum y_i)} \cdot \frac{\Gamma\left(\tilde{y}+\alpha+\sum y_i\right)}{\Gamma\left(\tilde{y}+1\right)\left(\beta+n+1\right)^{\tilde{y}+\alpha+\sum y_i}} \\
&= \prod_{i=1}^n y_i! \cdot \frac{\Gamma\left(\tilde{y}+\alpha+\sum y_i\right)}{\Gamma(\tilde{y}+1)\Gamma(\alpha+\sum y_i)} \cdot \left(\frac{\beta+n}{\beta+n+1}\right)^{\alpha+\sum y_i} \cdot \left(\frac{1}{\beta+n+1}\right)^{\tilde{y}}
\end{aligned}
$$







3.3.2 R Implementation

3.3.3 Example

3.4 Normal Observation with Normal-Inverse Gamma Prior

3.4.1 One sample

3.4.1.1 Derivation

3.4.1.2 R Implementation

3.4.1.3 Example

3.4.2 Two samples

3.4.2.1 Derivation

3.4.2.2 R Implementation

3.4.2.3 Example

3.4.3 k samples

3.4.3.1 Derivation

3.4.3.2 R Implementation

3.4.3.3 Example

3.4.3.4 Ranking Treatments

4 Chapter 2: Normal Regression with Zellner's g-prior

4.0.0.1 Derivation

4.0.0.2 R Implementation

4.0.0.3 Example

5 Conclusion