Predictive Inference Tools for Researchers

by

Voyze G. Harris III

Copyright © Voyze G. Harris III 2021

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

ARIZONA

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

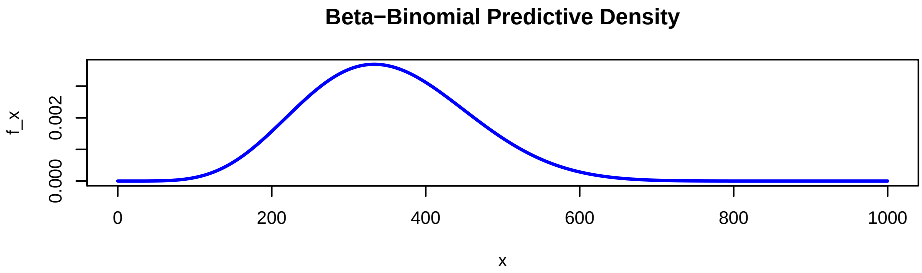
$$\begin{aligned}
\Pr[R=r|t] &= \int \binom{M}{r}\theta^r(1-\theta)^{M-r}p\left(\theta|X^{(N)}\right)d\theta \\
&= \binom{M}{r}\int \theta^r(1-\theta)^{M-r}\frac{\Gamma(N+\alpha+\beta)}{\Gamma(t+\alpha)\Gamma(N-t+\beta)}\theta^{t+\alpha-1}(1-\theta)^{N-t+\beta-1}d\theta \\
&= \frac{M!}{r!(M-r)!}\frac{\Gamma(N+\alpha+\beta)}{\Gamma(t+\alpha)\Gamma(N-t+\beta)}\int \theta^{r+t+\alpha-1}(1-\theta)^{M-r+N-t+\beta-1}d\theta \\
&= \frac{\Gamma(M+1)\Gamma(N+\alpha+\beta)\Gamma(r+t+\alpha)\Gamma(M-r+N-t+\beta)}{\Gamma(r+1)\Gamma(M-r+1)\Gamma(t+\alpha)\Gamma(N-t+\beta)\Gamma(M+N+\alpha+\beta)}
\end{aligned}$$

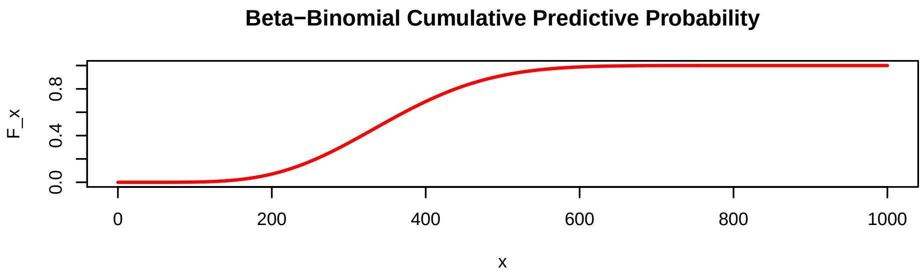
3.1.2 R Implementation

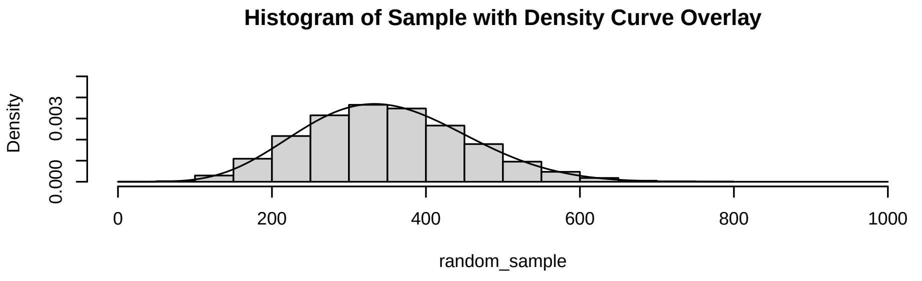
This result has been used to create "standard" R functions dpredBB(), ppredBB(), and rpredBB() for the Beta-Binomial distribtuion for density, cumulative probability, and random sampling, respectively (see appendix). These functions are exercised in the following example.

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 $\Pr(x > \theta) = 1 - \Pr(x \le \theta) = 1 - F(x|\theta) = 1 - (1 - e^{-\theta x}) = e^{-\theta x}$. 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

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

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

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

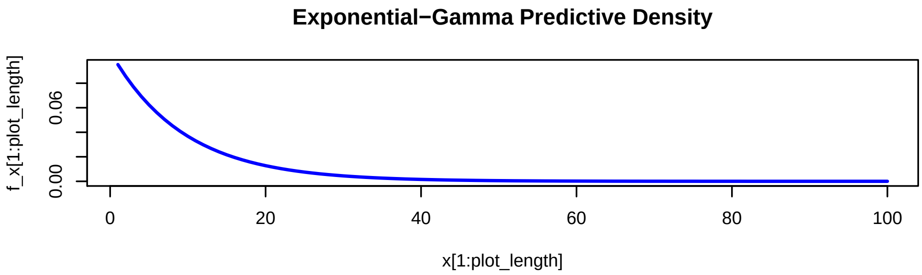
(simplifying by constructing a Gamma$(d+\delta+1, \gamma+N\bar{x}+x)$ density in the final integrand.)

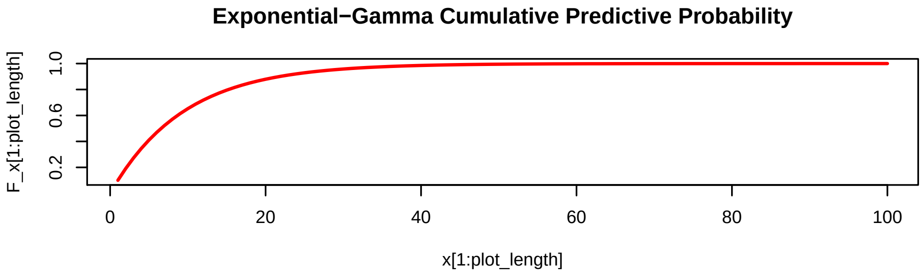
3.2.2 R Implementation

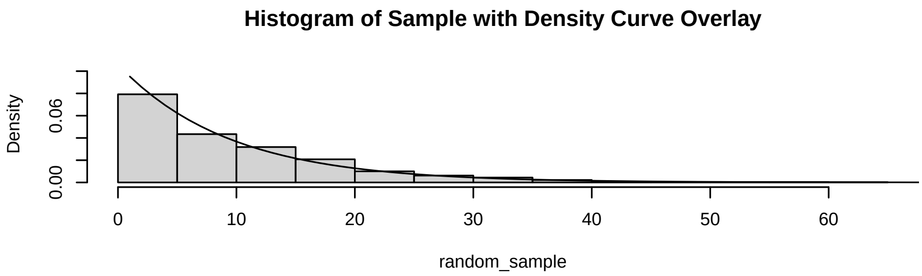
This result has been used to create standard format R functions dpredEG(), ppredEG(), and rpredEG() for the Gamma-Exponential distribtuion for density, cumulative probability, and random sampling, respectively (see appendix). These functions are exercised in the following example.

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
60
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,

$$P\left(Y_1 = y_1, ..., Y_n = y_n|\theta\right) = \prod_{i=1}^{n} p\left(y_i|\theta\right)$$

$$= \prod_{i=1}^{n} \frac{1}{y!}\theta^{y_i} e^{-\theta}$$

$$= \left(\prod_{i=1}^{n} \frac{1}{y!}\right) \theta^{\sum y_i} e^{-n\theta}$$

$$= c\left(y_1, ..., y_n\right) \theta^{\sum y_i} e^{-n\theta}$$

and

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

Then we have posterior distribution

$$p\left(\theta|y_1, ..., y_n\right) = \frac{p\left(y_1, ..., y_n|\theta\right) p(\theta)}{\int_\theta p\left(y_1, ..., y_n|\theta\right) p(\theta)}$$

$$= \frac{p\left(y_1, ..., y_n|\theta\right) p(\theta)}{p\left(y_1, ..., y_n\right)}$$

$$= \frac{1}{p\left(y_1, ..., y_n\right)} \theta^{\sum y_i} e^{-n\theta} \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} e^{-\beta\theta}$$

$$= C\left(y_1, ..., y_n, \alpha, \beta\right) \theta^{\alpha + \sum y_i - 1} e^{-(\beta+n)\theta}$$

$$\sim \text{Gamma}\left(\alpha + \sum y_i, \beta + n\right).$$

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

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

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

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

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

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

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

$$E\left[\tilde{Y}|y_1,...,y_n\right]=\frac{a+\sum y_i}{b+n}=E\left[\theta|y_1,...,y_n\right];$$

$$\begin{aligned}\mathrm{Var}\left[\tilde{Y}|y_1,...,y_n\right]&=\frac{a+\sum y_i}{b+n}\frac{b+n+1}{b+n}\\&=\mathrm{Var}\left[\theta|y_1,...,y_n\right]\times(b+n+1)\\&=E\left[\theta|y_1,...,y_n\right]\times\frac{b+n+1}{b+n}\end{aligned}$$

[Showing here that it is indeed a NB distribution]

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

so

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

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

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

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

$$p\left(\tilde{y}|y_1, ..., y_n\right) = \int_0^\infty p\left(\tilde{y}|\theta, y_1, ..., y_n\right) p\left(\theta|y_1, ..., y_n\right) d\theta$$

$$= \int p\left(\tilde{y}|\theta\right) p\left(\theta|y_1, ..., y_n\right) d\theta$$

$$= C \int \left(\frac{1}{\tilde{y}!}\theta^{\tilde{y}} e^{-\theta}\right) \theta^{\alpha + \sum y_i - 1} e^{-(\beta + n)\theta} d\theta$$

$$= \frac{C}{\tilde{y}!} \int \theta^{\tilde{y} + \alpha + \sum y_i - 1} e^{-(\beta + n + 1)\theta} d\theta$$

$$= \frac{C\Gamma\left(\tilde{y} + \alpha + \sum y_i\right)}{\Gamma\left(\tilde{y} + 1\right)\left(\beta + n + 1\right)^{\tilde{y} + \alpha + \sum y_i}} \int \frac{\left(\beta + n + 1\right)^{\tilde{y} + \alpha + \sum y_i}}{\Gamma\left(\tilde{y} + \alpha + \sum y_i\right)} \theta^{\tilde{y} + \alpha + \sum y_i - 1} e^{-(\beta + n + 1)\theta)} d\theta$$

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

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

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

Hoff p.47:

- b is interpreted as the number of prior observations
- a is interpreted as the sum of counts from b prior observations

Hoff p. 49 (Birth rate example): $a = 2, b = 1$.

3.3.2 R Implementation

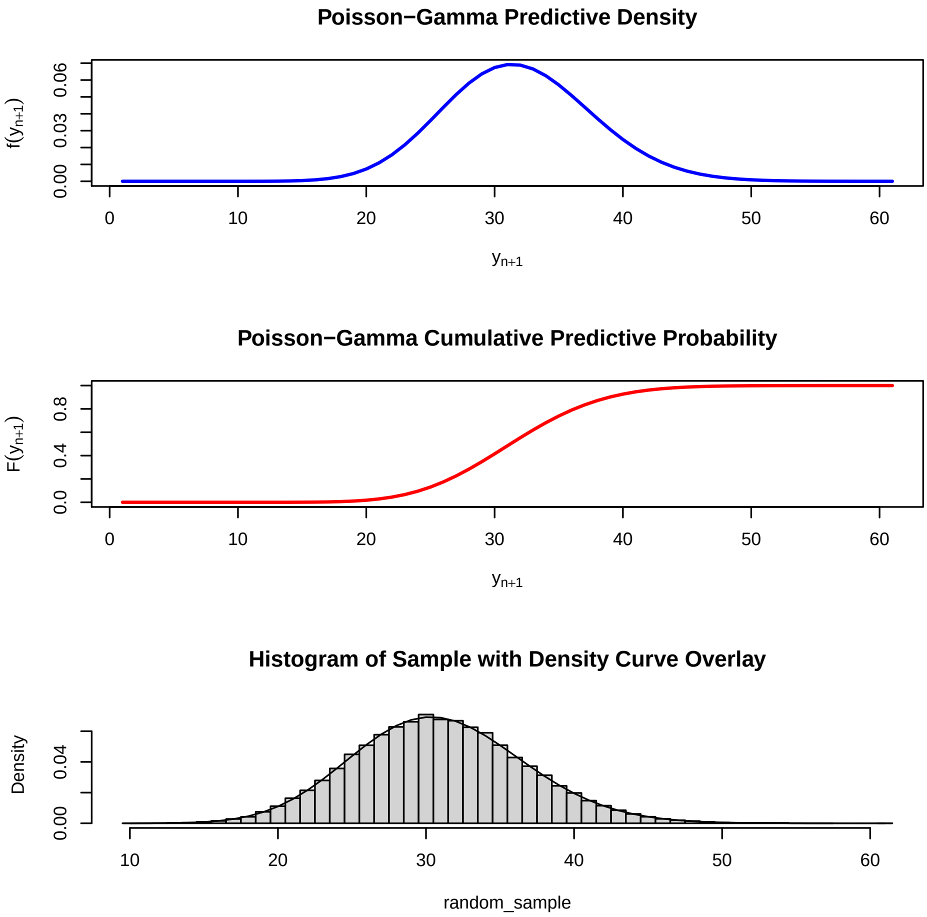
This result has been used to create standard format R functions dpredPG(), ppredPG(), and rpredPG() for the Poisson-Gamma distribution for density, cumulative probability, and random sampling, respectively (see appendix). These functions are exercised in the following example.

Developing the random sample function rpredPG(): I need to establish the support of the predictive distribution f_x from which to sample. the uniroot() function is not working because it keeps feeding non-integer values to dnbinom(). Strategy: a modified bisection method as follows:

1. set a desired tolerance ϵ.
2. Find the expected value E_x (closed formula, see above).
3. Step to the right of E_x by whole integers, in the sequence $E_x + \{1, 2, 4, ...2^n\}$, stopping at $U = f_x(E_x + 2^n) < 0$. This is the upper bound for the bisection method.
4. Bisect the interval, rounding to the nearest integer. Call the resulting mid-interval number B.
5. If B is positive, test whether $0 \leq f_x(B) \leq \epsilon$. If so, DONE. If not:
6. Establish new interval, choosing endpoints from E_x, B, and U so that the interval straddles 0, and repeat the steps until the condition in step 5 is reached.

3.3.3 Example

Suppose we have 10 prior observations with counts 27, 79, 21, 100, 8, 4, 37, 15, 3, 97. Let $\alpha = 11$ and $\beta = 3$. For $\tilde{y} = 1 : 100$ possible future occurrences, the figures below show the predictive distribution from dpredPG(), the cumulative distribution from ppredPG(), and a histogram of random draws from rpredPG().


Poisson−Gamma Predictive Density
f(yn+1)
0.00
0.03
0.06
0
10
20
30
40
50
60
yn+1
Poisson−Gamma Cumulative Predictive Probability
F(yn+1)
0.0
0.4
0.8
0
10
20
30
40
50
60
yn+1
Histogram of Sample with Density Curve Overlay
Density
0.00
0.04
10
20
30
40
50
60
random_sample

3.4 Normal Observation with Normal-Inverse Gamma Prior

3.4.1 One sample

3.4.1.1 Derivation [Hoff p. 69ff]

Let $\{Y_1, ..., Y_n|\theta, \sigma^2\} \overset{i.i.d.}{\sim} N(\theta, \sigma^2)$. Then the joint sampling density is

$$
\begin{aligned}
p\left(y_1, ..., y_n|\theta, \sigma^2\right) &= \prod_{i=1}^{n} p\left(y_i|\theta, \sigma^2\right) \\
&= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{y_i-\theta}{\sigma}\right)^2} \\
&= \left(2\pi\sigma^2\right)^{-n/2} e^{-\frac{1}{2}\sum_{i=1}^{n}\left(\frac{y_i-\theta}{\sigma}\right)^2}.
\end{aligned}
$$

It can be shown that $\{\sum y_i^2, \sum y_i\}$ and hence $\{\bar{y}, s^2\}$ are sufficient statistics, where $\bar{y} = \sum y_i/n$ and $s^2 = \sum \left(y_i - \bar{y}\right)^2/(n-1)$.

$\vdots$

Following Hoff (p. 74ff), for joint inference on both θ and σ, assume priors

$$
\begin{aligned}
\frac{1}{\sigma^2} &\sim \text{gamma}\left({}^{\nu_0}/_2, {}^{\nu_0\sigma_0^2}/_2\right) \\
\theta|\sigma^2 &\sim \text{normal}\left(\mu_0, {}^{\sigma^2}/_{\kappa_0}\right)
\end{aligned}
$$

where (σ_0^2, ν_0) are the sample variance and sample size of prior observations, and (μ_o, κ_0) are the sample mean and sample size of prior observations.

Note: μ_0, κ_0, ν_0, and σ_0^2 come from prior knowledge. [in the Hoff example (Midge Wing Length), κ_0 and ν_0 are both set to 1 so that "our prior distributions are only weakly centered around these estimates from other populations."]

From this we derive joint posterior

$$
\begin{aligned}
\left\{\theta|y_1, ..., y_n, \sigma^2\right\} &\sim \text{normal}\left(\mu_n, {}^{\sigma^2}/_{\kappa_n}\right) \\
\left\{\sigma^2|y_1, ..., y_n\right\} &\sim \text{inverse-gamma}\left({}^{\nu_n}/_2, {}^{\sigma_n^2\nu_n}/_2\right).
\end{aligned}
$$

where

$$\kappa_n = \kappa_0 + n$$

$$\mu_n = \frac{\kappa_0 \mu_0 + n\bar{y}}{\kappa_n}$$

$$\nu_n = \nu_0 + n$$

$$\sigma_n^2 = \frac{1}{\nu_n}\left[\nu_0 \sigma_0^2 + (n-1)s^2 + \frac{\kappa_0 n}{\kappa_n}\left(\bar{y} - \mu_0\right)^2\right].$$

Here $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$ is the sample mean and $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2$ is the sample variance.

From the joint posterior distribution we generate marginal samples by means of the Monte Carlo method (Hoff, p. 77):

$$\begin{array}{ll} \sigma^{2(1)} \sim \text{inverse-gamma}\left(\nu_n/2, \sigma_n^2 \nu_n/2\right), & \theta^{(1)} \sim \text{normal}\left(\mu_n, \sigma^{2(1)}/\kappa_n\right) \\ \vdots & \vdots \\ \sigma^{2(S)} \sim \text{inverse-gamma}\left(\nu_n/2, \sigma_n^2 \nu_n/2\right), & \theta^{(S)} \sim \text{normal}\left(\mu_n, \sigma^{2(S)}/\kappa_n\right) \end{array}$$

For prediction of future $\tilde{y}|y_1, ..., y_n, \theta, \sigma^2$, generate $\tilde{y}_i \sim \text{normal}\left(\theta^{(i)}, \sigma^{2(i)}\right)$.

3.4.1.2 R Implementation Standard format R functions dpredNormIG(), ppredNormIG(), and rpredNormIG() have been created for the Normal-Inverse Gamma distribution for density, cumulative probability, and random sampling, respectively (see appendix). For the random sampler rpredNormIG(), the Monte-Carlo method described above was directly employed. The predictive density and cumulative density functions depend on the random sample, and utilize Kernel Density Estimation (KDE) and R's built-in density() function. The KDE is computed by definition, using a normal kernel:

$$\hat{f}_K(x) = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{h} K\left(\frac{x - X_i}{h}\right),$$

where

X_i is the random sample generated using rpredNormIG()

K is Normal(0,1)

h is the bandwidth from R's density() function (that is, $h =$ density(X_i)$bw)

These functions are exercised in the following example.

Example (Hoff p. 72ff, using data from Grogan and Wirth (1981)): Midge wing length

Grogan and Wirth (1981) provide 9 measurements of midge wing length, in millimeters: $y = \{1.64, 1.7, 1.72, 1.74, 1.82, 1.82, 1.82, 1.90, 2.08\}$. Prior studies suggest values $\mu_0 = 1.9$ and $\sigma_0^2 = 0.01$. We choose $\kappa_0 = \nu_0 = 1$ "...so that our prior distributions are only weakly centered around these estimates from other populations" (Hoff p. 76). We compute

$$\bar{y} = 1.804$$

$$\mathrm{var}(y) = 0.0169$$

$$\kappa_n = 1 + 9 = 10$$

$$\mu_n = \frac{1 \cdot 1.9 + 9 \cdot 1.804}{10} = 1.814$$

$$\nu_n = 1 + 9 = 10$$

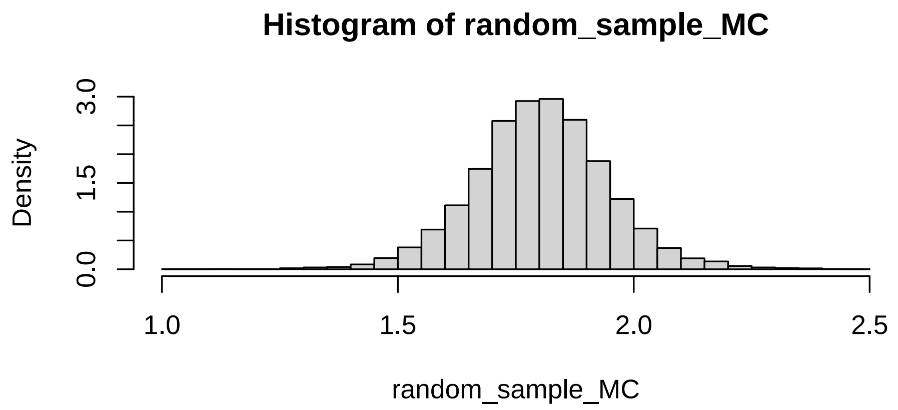
$$\sigma_n^2 = \frac{1}{10}\left[1 \cdot 0.01 + (9-1) \cdot 0.0169 + \frac{1 \cdot 9}{10}(1.804 - 1)^2\right] = 0.0153$$

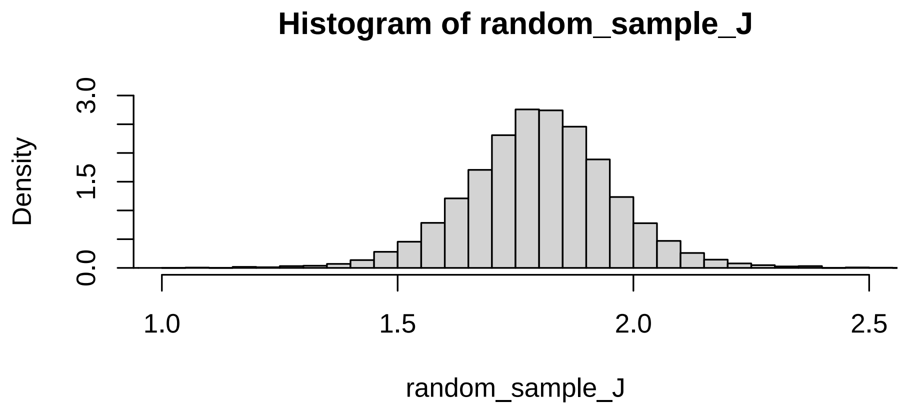
Thus $\nu_n/2 = 5$ and $\nu_n\sigma_n^2/2 = 0.7662$ and we have posteriors

$$\left\{\theta | y_1, ..., y_n, \sigma^2\right\} \sim \text{normal}\left(1.814, \sigma^2/10\right)$$

$$\left\{\sigma^2 | y_1, ..., y_n\right\} \sim \text{inverse-gamma}(5, 0.7662)$$

```
          [,1]
kn  10.0000000
mun  1.8040000
     0.1204326
nun 10.0000000
```


Histogram of random_sample_MC
Density
0.0
1.5
3.0
1.0
1.5
2.0
2.5
random_sample_MC


Histogram of random_sample_J
Density
0.0
1.5
3.0
1.0
1.5
2.0
2.5
random_sample_J

Jeffrey's Prior: $p(\theta, \sigma^2) = 1/\sigma^2$.

⋮

posteriors given Jeffrey's Prior:

$$\left\{1/\sigma^2 | y_1, ..., y_n\right\} \sim \text{gamma}\left(\frac{n-1}{2}, \frac{1}{2}\sum (y_i - \bar{y})^2\right)$$

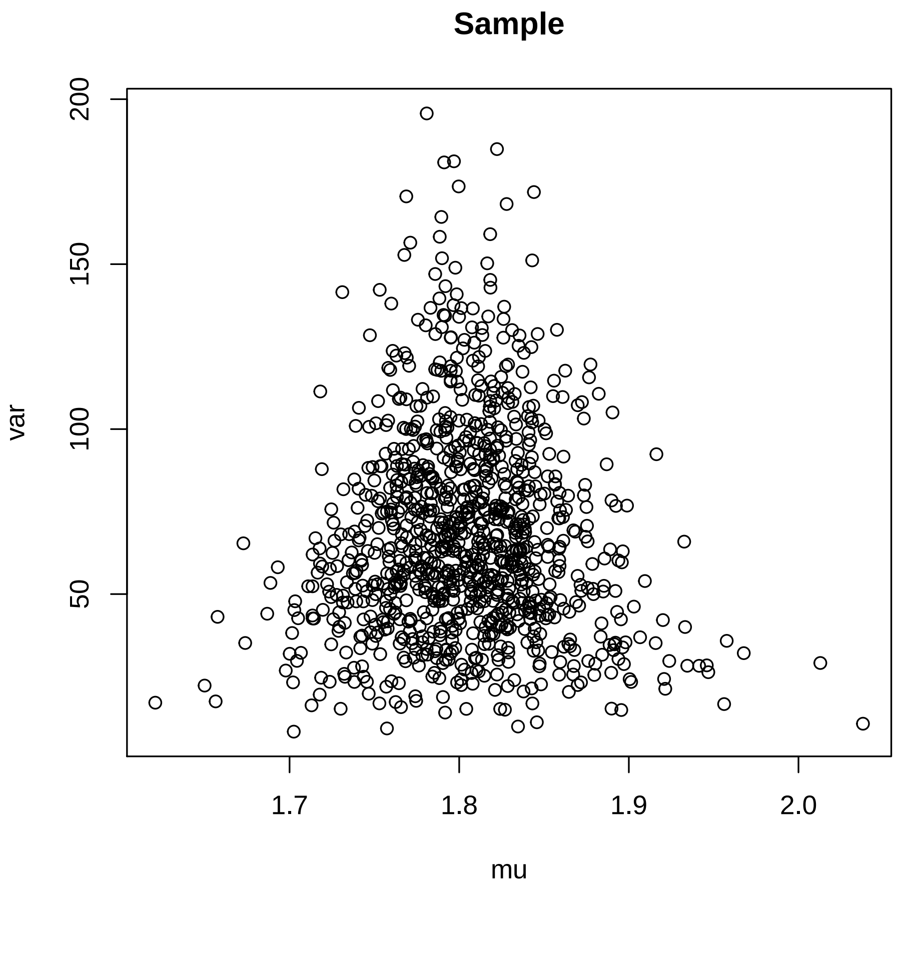
$$\left\{\theta | \sigma^2, y_1, ..., y_n\right\} \sim \text{normal}\left(\bar{y}, \frac{\sigma^2}{n}\right)$$

This leads to

$$\frac{\theta - \bar{y}}{s/\sqrt{n}} | y_1, ..., y_n \sim t_{n-1} \text{ (Hoff p. 79)}$$

So we have predictive distribution

$$\tilde{y} \sim t_{n-1} \text{ with location } \bar{y} \text{ and scale } s\sqrt{\left(1 + \frac{1}{n}\right)} \text{(Gelman et. al. p. 66)}$$



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