

TI 2024-056/III  
Tinbergen Institute Discussion Paper

# Flexible Negative Binomial Mixtures for Credible Mode Inference in Heterogeneous Count Data from Finance, Economics and Bioinformatics

*Jamie L. Cross*<sup>1</sup>  
*Lennart Hoogerheide*<sup>2</sup>  
*Paul Labonne*<sup>3</sup>  
*Herman K. van Dijk*<sup>4</sup>

<sup>1</sup> University of Melbourne and BI Norwegian Business School

<sup>2</sup> Vrije Universiteit Amsterdam and Tinbergen Institute

<sup>3</sup> BI Norwegian Business School

<sup>4</sup> Erasmus University Rotterdam, Norges Bank and Tinbergen Institute

Tinbergen Institute is the graduate school and research institute in economics of Erasmus University Rotterdam, the University of Amsterdam and Vrije Universiteit Amsterdam.

Contact: [discussionpapers@tinbergen.nl](mailto:discussionpapers@tinbergen.nl)

More TI discussion papers can be downloaded at <https://www.tinbergen.nl>

Tinbergen Institute has two locations:

Tinbergen Institute Amsterdam  
Gustav Mahlerplein 117  
1082 MS Amsterdam  
The Netherlands  
Tel.: +31(0)20 598 4580

Tinbergen Institute Rotterdam  
Burg. Oudlaan 50  
3062 PA Rotterdam  
The Netherlands  
Tel.: +31(0)10 408 8900

# Flexible Negative Binomial Mixtures for Credible Mode Inference in Heterogeneous Count Data from Finance, Economics and Bioinformatics\*

Jamie L. Cross<sup>1,2</sup>, Lennart Hoogerheide<sup>3,4</sup>, Paul Labonne<sup>2</sup>, and Herman K. van Dijk<sup>4,5</sup>

<sup>1</sup>*Melbourne Business School, University of Melbourne*

<sup>2</sup>*BI Norwegian Business School*

<sup>3</sup>*Department of Econometrics, VU University Amsterdam*

<sup>4</sup>*Tinbergen Institute*

<sup>5</sup>*Erasmus University Rotterdam and Norges Bank*

August 29, 2024

## Abstract

In several scientific fields, such as finance, economics and bioinformatics, important theoretical and practical issues exist involving multimodal and asymmetric count data distributions due to heterogeneity of the underlying population. For accurate approximation of such distributions we introduce a novel class of flexible mixtures consisting of shifted negative binomial distributions, which accommodates a wide range of shapes that are commonly seen in these data. We further introduce a convenient reparameterization which is more closely related to a moment interpretation and facilitates the specification of prior information and the Monte Carlo simulation of the posterior. This mixture process is estimated by the sparse finite mixture Markov chain Monte method since it can handle a flexible number of non-empty components. Given loan payment, inflation expectation and DNA count data, we find coherent evidence on number and location of modes, fat tails and implied uncertainty

---

\*The present paper should not be reported as representing the views of Norges Bank. The views expressed are those of the authors and do not necessarily reflect those of Norges Bank.

measures, in contrast to conflicting evidence obtained from well-known frequentist tests. The proposed methodology may lead to more accurate measures of uncertainty and risk which improves prediction and policy analysis using multimodal and asymmetric count data.

*JEL codes:* C11, C14, C63.

*Keywords:* Count data, multimodality, mixtures, shifted negative binomial, Markov chain Monte Carlo, Bayesian inference, sparse finite mixture.

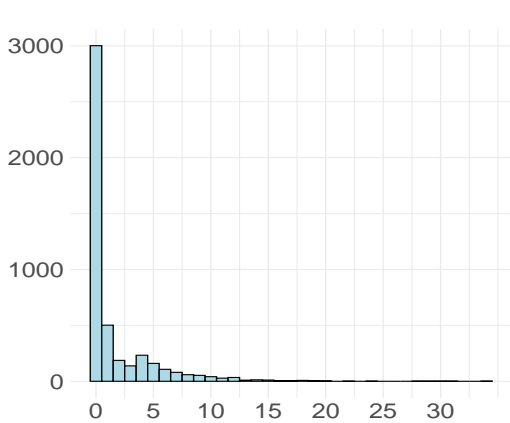
## 1 Introduction

In several scientific fields important theoretical and practical issues exist involving multimodal and/or asymmetric count data distributions due to heterogeneity of the underlying population. To determine the existence of multiple modes, statistical testing procedures are used on a regular basis. Two common examples are the dip test ([Hartigan and Hartigan, 1985](#); [Hartigan, 1985](#)) and the Silverman test and its extensions ([Silverman, 1981](#); [Fischer et al., 1994](#)). As an illustrative example we show count data in [Figure 1](#) on repayment details of clients obtaining a loan (panel (a)), inflation expectations from survey responses (panel (b)), and repetitive DNA tandem repeat sequences from human genome (panel (c)). We later show that the frequentist tests give conflicting evidence on the existence of a single mode at the 5% level of significance for two of three datasets. Moreover, it is important to note that these tests also remain silent on the mode locations and the probability that a given point is in fact a mode as opposed to noise in the data.<sup>1</sup> This is relevant since in financial economics information on multimodality may be used to categorize clients according to their defaulted installment behavior, which has direct policy relevance. Similarly, multimodality in the distribution of inflation expectations is linked with unanchored expectations, which is relevant for formulating an effective inflation policy. Deviations in the configurations in the D4Z4 count data may indicate genetic diseases; see, e.g., [Bruce et al. \(2009\)](#) and [Balog et al. \(2012\)](#) among several others.

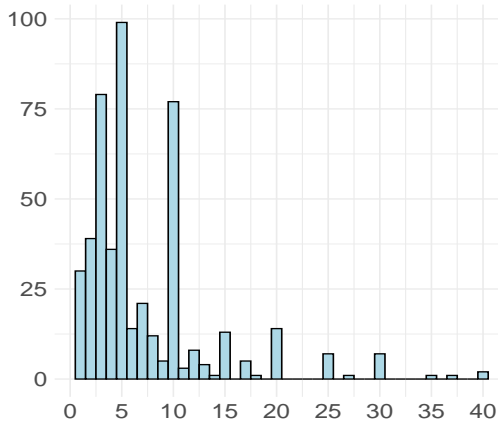
To deal with these problems, we introduce a simulation-based Bayesian approach using a

---

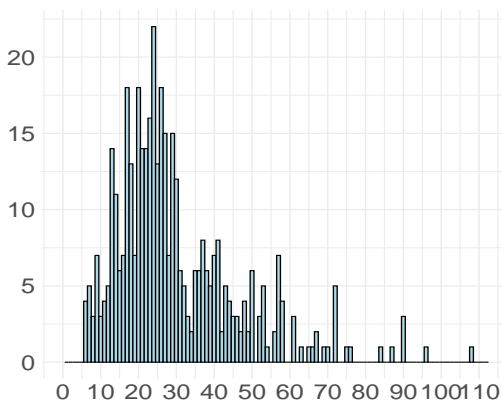
<sup>1</sup>We emphasize that when the number of modes is estimated, most existing methods also do not provide uncertainty metrics around these values ([Fischer et al., 1994](#); [Hall and Ooi, 2004](#)), with a few exceptions based on nonparametric density estimates ([Müller and Sawitzki, 1991](#); [Minnotte, 1997](#); [Chaudhuri and Marron, 1999](#)). We refer to [McLachlan and Peel \(2004\)](#) for an extensive summary of frequentist and Bayesian methods for determining the number of modes.



(a) Default data



(b) Inflation Expectations



(c) DNA tandem repeat sequence D4Z4

Figure 1: Examples of Count Data in Finance, Economics and Bioinformatics

flexible mixture of shifted negative binomial distributions. Our proposed framework provides a posterior probability of multimodality, estimates the number of modes and reveals their location in the distribution, and produces an model consistent measure of uncertainty around these values.

We start by discussing a number of features of the negative binomial distribution and consider its interpretation as the marginal of a continuous Poisson-gamma mixture. This implies adding an extra latent variable to a Poisson distribution which allows – in a Markov chain Monte Carlo (MCMC) simulation method – to make use of the well-known conditional augmented data likelihood and posterior in a Gibbs step for the mean parameter of the negative binomial distribution.

In terms of methodology we present three contributions. First, we introduce a convenient

reparameterization of the negative binomial distribution which is more closely related to a moment interpretation. It is well-known that Bayesian estimation of the dispersion parameter of the negative binomial is notoriously difficult. Consider the usual parameterization with the dispersion and probability parameters, see section 2 for details, then posterior inference of the probability parameter is straightforward due to conjugacy between the beta and negative binomial distributions. However, inference about the dispersion parameter is more difficult because there is no conjugate prior. Our experience with various simulations revealed that specifying a random walk Metropolis(-Hastings) step over this parameter is unreliable due to extremely poor mixing of the simulated draws and it leads to a lack of convergence to the true posterior. Existing literature on estimating this dispersion parameter is also surprisingly sparse. [Zhou and Carin \(2013\)](#) provide a detailed discussion of this literature and associated problems on identifying and estimating the dispersion parameter. These authors note that while the maximum likelihood estimator has been used to provide point estimates of the dispersion parameter, it is often not robust with respect to initial conditions. It can be biased or even fail to converge in small samples. The authors also note that while Bayesian approaches are able to model estimation uncertainty and incorporate prior information in small samples, the estimation is difficult as it relies on approximating the ratio of two gamma functions. In contrast, our reparameterization facilitates relatively simple and reliable estimation of these parameters.

Second, we introduce a novel shift parameter which is useful for cases where the distribution's support need not begin at zero, making it well suited for modeling data that are concentrated around certain non-zero values, such as the aforementioned DNA tandem repeat data. Moreover, allowing the shift parameter to be zero preserves all properties of the regular negative binomial distribution and allowing it, for instance, to deal with zero inflated count data. Finally, we specify a finite mixture of this general class of distributions which accommodates a wide range of distributional shapes that are commonly seen in many multimodal count data sets across the physical and social sciences, in particular, it allows for credible estimation of number of modes, their locations, and implied uncertainty of these features. Another advantage of using a negative binomial mixture, as opposed to say a mixture of Poisson distributions, is that the negative binomial distribution can account for fat tails, which are characteristic of overdispersion in count data.

In terms of computational efficiency, we apply the sparse finite mixture Markov chain Monte Carlo (SFM MCMC) method due to [Malsiner-Walli et al. \(2016\)](#), see also [Frühwirth-Schnatter and Malsiner-Walli \(2019\)](#), to our class of flexible mixtures in order to estimate the number of (non-empty) components of the mixture. This MCMC approach consists of initially overfitting the number of mixture components by selecting a larger number of components than is expected to describe the data and then using a sparse hierarchical prior on the mixture weights to shrink the model space to a credible number of components with substantial posterior probability. Estimating the number of (non-empty) components in this way has the practical benefit of reducing the computational burden when the number of alternatives is large, and reduces the probability of model misspecification when conducting inference under model specification uncertainty.

In terms of empirical analysis, the benefit of our approach is shown in the application using the three count data sets in [Figure 1](#). First, using lending data from a Spanish bank, see [Dionne et al. \(1996\)](#), the probability that a client of a bank is able to repay a loan is shown to depend on the uncertainty that exists with respect to the number of defaulted payments. Results from fitting the mixture of shifted negative binomials leads to credible evidence about a bimodal distribution with accurate location of modes including their uncertainty. Such information may be used for the institution’s risk policy of granting loans.

Using count data from survey responses to the question “By what percentage do you expect prices to go up, on average, during the next 12 months?”, from the University of Michigan’s Surveys of Consumers yields evidence about a bimodal mixture, which is suggestive of unanchored expectations, alternative learning ([Pfajfar and Santoro, 2010](#)), and possible lack of economic literacy ([Burke and Manz, 2014](#)).

The data of DNA tandem repeats from [Schaap et al. \(2013\)](#) show that applying the proposed class of flexible mixtures and the MCMC method provides substantial support with high posterior probability for a distinct unimodal distribution with a very fat right tail instead of some spurious modes.

In all three datasets, results from log-predictive likelihood analyses suggest that the mixture of shifted negative binomial distributions provides a good or better fit of the data compared to a corresponding mixture of shifted Poisson distributions. This empirical evidence reinforces the theoretical merits of our proposed shifted negative binomial mixture.

These results should be of interest to academics, policy makers and practitioners, who rely on such tests when making both inference and predictions.

The contents of this paper are organized as follows: Section 2 presents the class of mixtures of shifted negative binomials and the convenient reparameterization. Section 3 presents the SFM MCMC algorithm. Section 4 presents empirical results for three count data sets. Section 5 concludes.

## 2 A flexible mixture of shifted negative binomials

### 2.1 A class of flexible mixtures based on the negative binomial distribution

The negative binomial distribution was introduced by Pascal (1679) and subsequently studied by Montmort (1713). There exist several parameterizations of this distribution. We make use of the specification that is defined as the number of failures before the  $r$ -th success. In order to illustrate the flexibility of the negative binomial distribution Figure 2 shows examples of the shape of the Probability Mass Function (PMF) of the negative binomial distribution for a given value  $p = 0.5$  and for  $r = 1, 5, 10$ . The extreme value  $r = 1$  leads to a geometric distribution, shown in Figure 3. The other extreme case  $r \rightarrow \infty$  (while holding the mean  $\mu = \frac{r(1-p)}{p}$  constant, so that  $p = \frac{r}{r+\mu} \rightarrow 1$ ) leads to a Poisson distribution, see Figure 4. So, the interval  $1 < r < \infty$  allows for flexible shapes. For example, for the negative binomial distributions with  $p = 0.5$  in Figure 2 as  $r$  increases from 1 to 5 and 10, we observe a decreasing right skewness (with skewness equal to 2.12, 0.95 and 0.67, where the means 1, 5 and 10 are larger than the modes 0, 4 and 9) and a decreasing kurtosis (9.50, 4.30 and 3.65); see Table 1. The kurtosis of the negative binomial distribution increases for decreasing values of  $r$  and increasing values of  $p$ : it tends to  $\infty$  if  $r \rightarrow 0$  or  $p \rightarrow 1$ . On the other hand, if  $r \rightarrow \infty$ , then the skewness and kurtosis tend to well known values for the Gaussian distribution, 0 and 3, respectively. Summarizing, the class of negative binomial distributions includes both heavily right-skewed, fat-tailed distributions, and almost symmetric, thin-tailed distributions.

In Figure 5 we illustrate the difference between a Poisson distribution (with mean  $\lambda = 5$  and kurtosis 3.20) and three negative binomial distributions for  $r = 1, 5, 50$  with the same mean but different kurtosis 9.03, 4.30 and 3.30. Although the PMF and kurtosis look similar for the



Table 1: Characteristics of negative binomial, geometric and Poisson distributions

	negative binomial	geometric	Poisson
pmf	$p_{NB}(y_i p, r) = \frac{\Gamma(r+y_i)}{\Gamma(r)y_i!} (1-p)^{y_i} p^r$	$p_{Geom}(y_i p) = (1-p)^{y_i} p$	$p_{Poi}(y_i \lambda) = \frac{\lambda^{y_i} e^{-\lambda}}{y_i!}$
mean	$\frac{r(1-p)}{p}$	$\frac{1-p}{p}$	$\lambda$
mode	$\lfloor \frac{(r-1)(1-p)}{p} \rfloor$ if $r > 1$ , 0 if $r \leq 1$	0	$\lfloor \lambda \rfloor$ (& $\lambda - 1$ if $\lambda$ integer)
variance	$\frac{r(1-p)}{p^2}$	$\frac{(1-p)}{p^2}$	$\lambda$
skewness	$\frac{2-p}{\sqrt{(1-p)r}}$	$\frac{2-p}{\sqrt{(1-p)}}$	$\frac{1}{\sqrt{\lambda}}$
kurtosis	$3 + \frac{6}{r} + \frac{p^2}{(1-p)r}$	$9 + \frac{p^2}{(1-p)}$	$3 + \frac{1}{\lambda}$

Poisson distribution and the negative binomial distribution with large parameter  $r = 50$ , there is still a considerable difference in the far right tail. Figure 6 shows that the ratio of the Poisson and negative binomial PMF tends to 0, also for  $r = 50$ , so that extreme outliers are relatively much more likely for the negative binomial PMF.

### 2.1.1 A Poisson-Gamma mixture representation

The negative binomial distribution can also be specified as the marginal of a Poisson( $\lambda$ ) distribution, where  $\lambda > 0$  is a random variable having a gamma distribution with shape parameter  $r$  ( $r > 1$ ) and rate parameter  $\frac{p}{1-p}$  ( $0 < p < 1$ ):

$$p_{NB}(y_i|r, p) = \int_0^\infty p_{Poisson(\lambda_i)}(y_i) p_{Gamma}(r, \frac{p}{1-p})(\lambda_i) d\lambda_i, \quad (1)$$

see Appendix A.

So, we have as the joint probability mass/density function of  $(y_i, \lambda_i)$ :

$$p(y_i, \lambda_i|r, p) = p(y_i|\lambda_i)p(\lambda_i|r, p) = \frac{\lambda_i^{y_i} \exp(-\lambda_i)}{y_i!} \frac{\left(\frac{p}{1-p}\right)^r \lambda_i^{r-1} \exp\left(-\lambda_i \frac{p}{1-p}\right)}{\Gamma(r)}. \quad (2)$$

We note the analogue between the negative binomial and the Student-t distribution. The former is the marginal of a Poisson-gamma mixture for count data while the latter is the marginal of a normal-gamma mixture for continuous data, where the negative binomial and Student-t

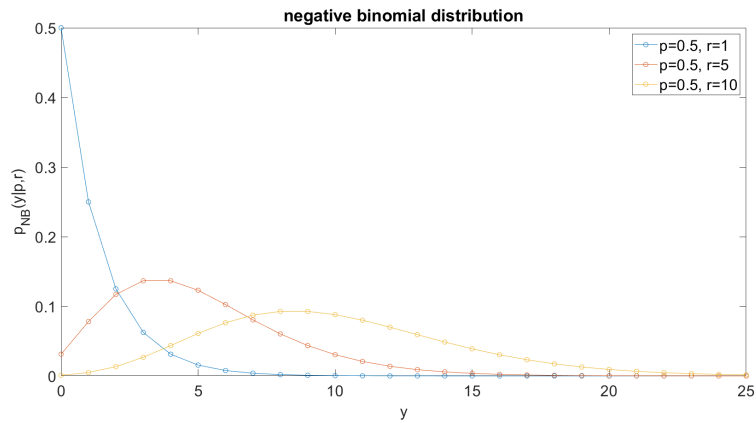


Figure 2: Probability mass function  $p_{NB(p,r)}(y_i) = \frac{\Gamma(r+y_i)}{\Gamma(r)y_i!} (1-p)^{y_i} p^r$  with  $p = 0.5$  for different values of the parameter  $r$

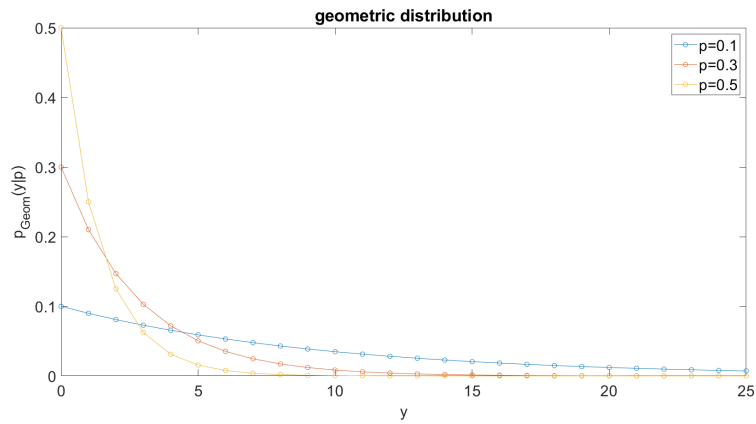


Figure 3: Probability mass function  $p_{Geom(p)}(y_i) = (1-p)^{y_i} p$  for different values of parameter  $p$

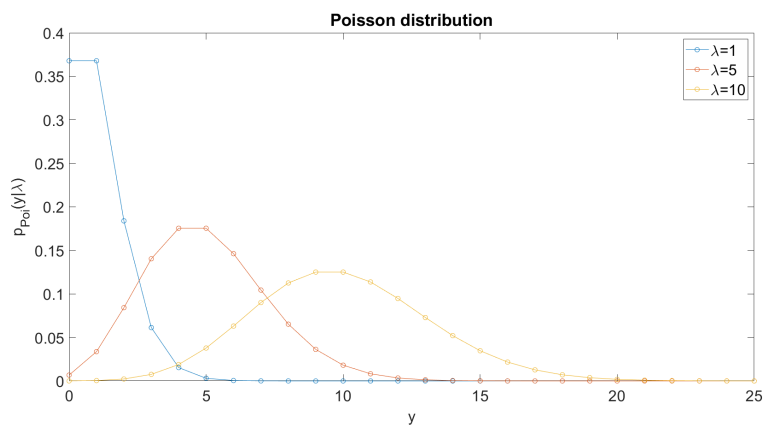


Figure 4: Probability mass function  $p_{Poisson(\lambda)}(y_i) = \frac{\lambda^{y_i} e^{-\lambda}}{y_i!}$  for different values of the mean parameter  $\lambda$

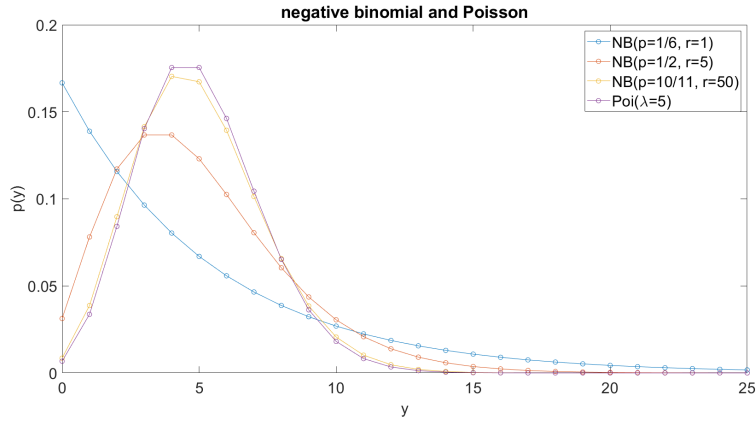


Figure 5: Comparison of probability mass functions of negative binomial distributions and Poisson distribution (all with mean 5)

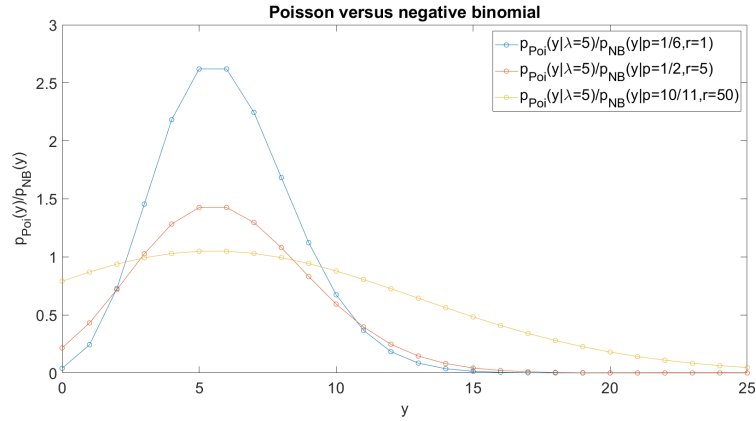


Figure 6: Ratio of probability mass functions of Poisson distribution and negative binomial distributions (all with mean 5)

distributions tend to the Poisson and normal distribution, respectively, if the shape parameter tends to infinity.

In this paper, we contribute by introducing a reparameterization that connects the parameters  $r$  and  $p$  of the negative binomial with the mean of this distribution. Moreover, this reparameterization facilitates the simulation-based Bayesian estimation of the model. It is given

as:

$$\mu = \frac{r(1-p)}{p}, \quad (3)$$

$$\tilde{r} = \frac{1}{1+r}, \quad (4)$$

with inverse transformation

$$r = \frac{1-\tilde{r}}{\tilde{r}}, \quad (5)$$

$$p = \frac{r}{r+\mu} = \frac{1-\tilde{r}}{1-\tilde{r}+\mu\tilde{r}}. \quad (6)$$

This reparametrization has multiple advantages. First, it is easily interpreted as the mean of the negative binomial distribution. Second, the conditional augmented data posterior distribution of  $\mu$  is a well-known distribution: a truncated version of the inverse gamma distribution, as discussed below. Third, the transformed shape parameter  $\tilde{r}$  takes its values in the *finite* interval  $(0, \frac{1}{2})$ , which facilitates the specification of a (non-informative) prior and the Monte Carlo simulation of the posterior. The possibly fat right tail of the distribution of the original, untransformed parameter  $r$  may harm the efficiency of the MCMC method – for example, due to long sequences of relatively high values of  $r$  and/or because it is difficult to tune an appropriate standard deviation of the candidate steps in a random walk Metropolis(-Hastings) step for  $r$ . The simulation of the transformed parameter  $\tilde{r}$  does not suffer from these problems.

Using a set of simulated data from the negative binomial distribution with  $p = 3/8$  and  $r = 1$  (or equivalently  $\mu = 5$  and  $\tilde{r} = 1/4$ ), we show in Figure 7 the shape of the likelihood for the parameters  $r$  and  $p$  and compare it with the likelihood shape using the transformed parameters  $\mu$  and  $\tilde{r}$ . Clearly the latter shape shows more curvature than the former which is an indication that the resulting shape of the posterior is more well-behaved. This usually helps the mixing properties and computational efficiency of the MCMC procedure.

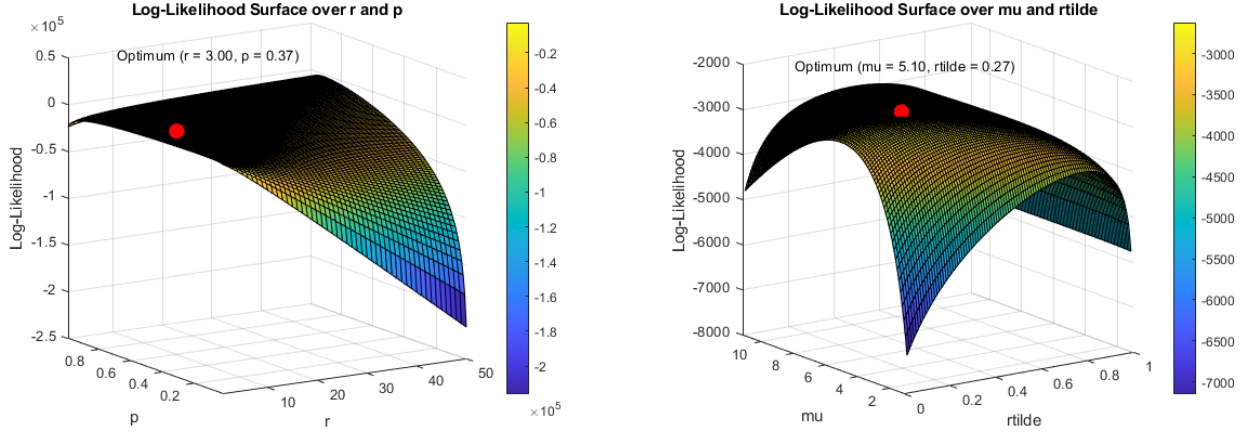


Figure 7: Log-Likelihood Surface for Negative Binomial Model

True parameter values:  $p = 0.375$ ,  $r = 3.00$ ,  $\mu = 5.00$  and  $\tilde{r} = 0.25$ .

Using this reparameterization we have

$$p_{NB}(y_i | \tilde{r}, \mu) = \frac{\Gamma(y_i + \frac{1-\tilde{r}}{\tilde{r}})}{y_i! \Gamma(\frac{1-\tilde{r}}{\tilde{r}})} \left( \frac{\mu \tilde{r}}{1 - \tilde{r} + \mu \tilde{r}} \right)^{y_i} \left( \frac{1 - \tilde{r}}{1 - \tilde{r} + \mu \tilde{r}} \right)^{\left(\frac{1-\tilde{r}}{\tilde{r}}\right)} \quad (7)$$

$$p(y_i, \lambda_i | \tilde{r}, \mu) = \frac{\lambda_i^{y_i} \exp(-\lambda_i)}{y_i!} \frac{r^r \mu^{-r} \lambda_i^{r-1} \exp\left(-\frac{\lambda_i r}{\mu}\right)}{\Gamma(r)} \quad (8)$$

with  $r = \frac{1-\tilde{r}}{\tilde{r}}$ , where (7) is the reparameterized version of the PMF in Table 1 and where (8) is the reparameterized, more convenient version of (2).

This representation implies that we can make use, in a Monte Carlo simulation procedure, of a known conditional augmented data likelihood/posterior of the mean parameter  $\mu$  instead of facing a non-standard conditional likelihood/posterior where a Metropolis step has to be used instead of a Gibbs step for the mean parameter  $\mu$ .

### 2.1.2 A mixture of shifted negative binomial distributions

We now specify the class of a finite mixture of shifted negative binomial distributions. Let  $y_i$  for  $i = 1, \dots, n$  be independent realizations from a mixture of  $J$  shifted negative binomial distributions:

$$y_i - \kappa_j \sim \text{NB}(\mu_j, \tilde{r}_j) \text{ if } z_{ij} = 1 \text{ for } i = 1, \dots, n; j = 1, \dots, J, \quad (9)$$

where  $z_{ij} = 1$  if  $y_i$  belongs to component  $j$ , and 0 otherwise and the latent variable distribution is defined as  $\Pr[z_{ij} = 1] = \pi_j$ , for  $i = 1, \dots, n; j = 1, \dots, J$ , with  $\pi_j \geq 0$  for  $j = 1, \dots, J$  and  $\sum_{j=1}^J \pi_j = 1$ , and where the shift parameter  $\kappa_j$  is a non-negative integer. That is,

$$p(y_i|\theta) = \sum_{j=1}^J \pi_j \frac{\Gamma\left(y_i - \kappa_j + \left(\frac{1-\tilde{r}_j}{\tilde{r}_j}\right)\right)}{(y_i - \kappa_j)! \Gamma\left(\frac{1-\tilde{r}_j}{\tilde{r}_j}\right)} \left(\frac{(1-\tilde{r}_j)}{(1-\tilde{r}_j) + \mu_j \tilde{r}_j}\right)^{\left(\frac{1+\tilde{r}_j(y_i-\kappa_j-1)}{\tilde{r}_j}\right)} \left(\frac{\tilde{r}_j \mu_j}{1-\tilde{r}_j}\right)^{y_i-\kappa_j} \quad (10)$$

$$p(y_i, \lambda_i|\theta) = \sum_{j=1}^J \pi_j \frac{\lambda_i^{y_i-\kappa_j} \exp(-\lambda_i)}{(y_i - \kappa_j)!} \frac{r_j^{r_j} \mu_j^{-r_j} \lambda_i^{r_j-1} \exp\left(-\frac{\lambda_i r_j}{\mu_j}\right)}{\Gamma(r_j)} \quad (11)$$

with  $\theta = \{\mu, \tilde{r}, \kappa, \pi\}$ ,  $\mu = \{\kappa_1, \dots, \kappa_J\}$ ,  $\tilde{r} = \{\tilde{r}_1, \dots, \tilde{r}_J\}$ ,  $\kappa = \{\kappa_1, \dots, \kappa_J\}$  and  $\pi = \{\pi_1, \dots, \pi_J\}$ . This class of a finite mixture of shifted negative binomial distributions is a very flexible class of distributions for non-negative integer data that allows for multimodality, underdispersion and overdispersion. In Section 4 we show the usefulness of the added flexibility in three sets of empirical data from different scientific fields.

## 2.2 Likelihood and prior

Given data  $y = \{y_1, \dots, y_n\}$  and latent variables  $z_i = \{z_{i1}, \dots, z_{iJ}\}$ ,  $z = \{z_1, \dots, z_n\}$  and  $\lambda = \{\lambda_1, \dots, \lambda_n\}$ , the augmented likelihood for the parameters  $\theta$  of model (9) (where the latent variables  $\lambda$  are integrated out) follows directly from subsection 2.1:

$$l(\theta|y, z) = \begin{cases} \prod_{i=1}^n \prod_{j=1}^J \left[ \frac{\Gamma(y_i - \kappa_j + r_j)}{(y_i - \kappa_j)! \Gamma(r_j)} (1 - p_j)^{r_j} p_j^{y_i - \kappa_j} \right]^{z_{ij}} \pi_j^{z_{ij}}, & y_i = \kappa_j, \kappa_j + 1, \dots, \forall i, j \text{ with } z_{ij} = 1 \\ 0, & \text{otherwise,} \end{cases} \quad (12)$$

and the augmented likelihood for  $\theta$  given data  $y$ , latent variables  $z$  and  $\lambda$  is:

$$l(\theta|y, z, \lambda) = \begin{cases} \prod_{i=1}^n \prod_{j=1}^J \left[ \frac{\lambda_i^{y_i - \kappa_j} \exp(-\lambda_i)}{(y_i - \kappa_j)!} \frac{r_j^{r_j} \mu_j^{-r_j} \lambda_i^{r_j-1} \exp\left(-\frac{\lambda_i r_j}{\mu_j}\right)}{\Gamma(r_j)} \right]^{z_{ij}} \pi_j^{z_{ij}}, & y_i = \kappa_j, \kappa_j + 1, \dots, \forall i, j \text{ with } z_{ij} = 1 \\ 0, & \text{otherwise,} \end{cases} \quad (13)$$

where  $r_j = \frac{1-\tilde{r}_j}{\tilde{r}_j}$  and  $p_j = \frac{r_j}{r_j + \mu_j}$ ,  $j = 1, 2, \dots, J$ .

Given the specification of the flexible class of distributions and the implied likelihood, we discuss our motivation and specification of the prior. It is relevant to distinguish between experimental sciences like medicine where data on treatment and control groups are collected and non-experimental sciences, that we deal with, where data are given and a diagnostic data step is usually taken to determine a plausible class of models and priors, implying that the latter are typically data dependent. We take the point of view that in the scientific fields where we analyze data, it is of interest to let the information in the likelihood dominate the information in the prior. Therefore, we make use of uninformative but proper priors for the parameters. Specifically, we make use of continuous and discrete uniform priors defined on bounded regions for  $\mu_j$ ,  $\tilde{r}_j$  and  $\kappa_j$  and a symmetric Dirichlet prior for the weight parameters  $\pi_j$ :

$$\mu_j \sim \text{Unif}(\mu_{\min}, \mu_{\max}), \quad (14)$$

$$\tilde{r}_j \sim \text{Unif}\left(0, \frac{1}{2}\right), \quad (15)$$

$$\kappa_j \sim \text{DiscUnif}(\kappa_{\min}, \kappa_{\max}), \quad (16)$$

$$(\pi_1, \dots, \pi_J) \sim \text{Dirichlet}(\alpha, \dots, \alpha). \quad (17)$$

where Unif and DiscUnif denote the continuous and discrete uniform distributions, respectively. We set  $[\mu_{\min}, \mu_{\max}] = [\kappa_{\min}, \kappa_{\max}] = [0, M]$  with  $M$  equal to the maximum of the data sample,  $M = \max(y_i | y_i = 1, \dots, n)$ . Similar results are obtained for different large values of  $M$ , so that the prior does not need to be data-dependent. Note that  $\tilde{r}_j < \frac{1}{2}$  and  $r_j > 1$  imply that we have  $r_j n_j - 1 > 0$  (even if  $n_j$ , the number of observations belonging to component  $j$ , is equal to 1), so that the conditional posterior distribution of  $\mu_j$  is a truncated version of an inverse gamma distribution (from which one easily simulates using inverse transform sampling). For  $0 < r_j \leq 1$  it would be possible that  $r_j n_j - 1 \leq 0$ , in which case the conditional posterior distribution of  $\mu_j$  would still be proper, but no longer a truncated version of an inverse gamma distribution. Then its properness would only stem from the fact that  $\mu_j$  takes its values on the finite interval  $[0, M]$ , since on the infinite interval  $[0, \infty)$  the right tail of the conditional posterior of  $\mu_j$  would not have a finite integral. The assumption that  $r_j > 1$  does not seem very restrictive. In literature it is sometimes assumed that the parameter  $r$  of the negative binomial distribution is a positive

integer, in which case one has  $r \geq 1$ . The geometric distribution (corresponding to  $r = 1$ ) allows for infinite right skewness and kurtosis (for  $p \rightarrow 1$ ).

### 3 Markov chain Monte Carlo method

Alternative MCMC methods exist for the estimation of the model parameters and number of components of a mixture process. We apply the SFM MCMC algorithm, see [Malsiner-Walli et al. \(2016\)](#) and [Malsiner-Walli et al. \(2017\)](#). This method consists of initially overfitting the number of mixture components by choosing a number that is larger than the number of components expected to describe the data. Next, a sparse hierarchical prior on the mixture weights is applied to shrink the model space to a credible number of components. The authors show that the SFM approach is more reliable than Dirichlet process mixtures which are known to overfit the number of components ([Miller and Harrison, 2013](#)), and that the SFM approach is conceptually simpler and substantially easier to implement than the Reversible Jump MCMC, see [Green \(1995\)](#); [Richardson and Green \(1997\)](#).

We note that, to the best of our knowledge, for the case of a set of non-experimental multimodal data there does not exist a theorem that shows which MCMC method, simulating draws from posterior distributions of mixture processes, is uniformly superior in terms of convergence properties and robustness. Different information content between model, prior information and data play an important role in choosing which MCMC method is most suitable in a certain case. The choice of the SFM MCMC method turns out to be easy to implement and seems efficient in our applications.

Two practical issues are important for the implementation of the SFM MCMC method. First, its mixing quality is an important indication for the (speed of) convergence of the algorithm, no matter the application of the model. Since data in the non-experimental sciences are usually given, an initial visual data inspection, as discussed earlier for the default data, inflation expectations data and DNA data, is a very common sense strategy in order to obtain a good guess on the prior for the parameters, see [Frühwirth-Schnatter \(2011\)](#), in particular, a plausible value and range of the shift parameter is useful. This may have a substantial effect on the speed of convergence of the SFM procedure. Second, in this study we are interested in the predictive



distribution of the observable variables of our datasets in order to credibly determine the number and location of modes. We are not interested in the identification of the mixture structure. Therefore, label switching is less of an issue here.

### 3.1 Sparse Finite Mixture Markov Chain Monte Carlo

For ease of exposition, we start to present the proposed MCMC procedure for the case of *known* number of non-empty components (with fixed Dirichlet prior for  $(\pi_1, \dots, \pi_J)$ ). For background on the case of a given number of components,  $J$ , we refer to [Frühwirth-Schnatter \(2006\)](#), Chapter 3. After that, we extend this to the case of *unknown* number of non-empty components (with hierarchical prior for  $(\pi_1, \dots, \pi_J)$ ), the case of the sparse finite mixture. We make use of the priors (14)–(17) which yield in combination with the likelihood the following conditional posteriors:

$$\lambda_i | y_i, z_i, \theta \sim \text{Gamma} \left( y_i + \sum_{j=1}^J z_{ij}(-\kappa_j + r_j), \frac{1}{\sum_{j=1}^J z_{ij} p_j} \right) \quad (18)$$

$$p(\kappa_j | y, z, \theta_{-\kappa}) \propto \prod_{i|z_{ij}=1} \frac{\Gamma(y_i - \kappa_j + r_j)}{(y_i - \kappa_j)!} \times p_j^{\sum_{i|z_{ij}=1} (y_i - \kappa_j)} \quad (19)$$

$$\frac{1}{\mu_j} \Big| y, z, \lambda, \theta_{-\mu} \sim \text{Gamma}_{[\frac{1}{M}, \infty)} \left( r_j n_j - 1, r_j \sum_{i|z_{ij}=1} \lambda_i \right) \quad (20)$$

$$p(\tilde{r}_j | y, z, \theta_{-\tilde{r}}) \propto \left( \Gamma \left( \frac{1 - \tilde{r}_j}{\tilde{r}_j} \right) \right)^{-n_j} \times \quad (21)$$

$$\prod_{i|z_{ij}=1} \Gamma \left( y_i - \sum_{j=1}^J z_{ij} \kappa_j + \left( \frac{1 - \tilde{r}_j}{\tilde{r}_j} \right) \right) \times$$

$$\left( \frac{(1 - \tilde{r}_j)}{(1 - \tilde{r}_j) + \mu_j \tilde{r}_j} \right) \left( \frac{1 + \tilde{r}_j (y_i - \sum_{j=1}^J z_{ij} \kappa_j - 1)}{\tilde{r}_j} \right) \left( \frac{\tilde{r}_j \mu_j}{1 - \tilde{r}_j} \right)^{y_i - \sum_{j=1}^J z_{ij} \kappa_j}$$

$$\pi | z \sim \text{Dirichlet} (n_1 + \alpha, \dots, n_J + \alpha), \quad (22)$$

where  $r_j = \frac{1 - \tilde{r}_j}{\tilde{r}_j}$  and  $p_j = \frac{r_j}{r_j + \mu_j}$  ( $j = 1, 2, \dots, J$ ), and where  $n_j = \sum_{i=1}^n z_{ij}$  is the number of observations in component  $j$  and  $\kappa_j$  is an integer in  $[\kappa_{\min}, \min\{\min_{i|z_{ij}=1} (y_i), \kappa_{\max}\}]$ .

That is,  $\lambda_i$  ( $i = 1, \dots, n$ ),  $\mu_j$  ( $j = 1, \dots, J$ ) and  $\pi$  are simulated from their conditional

posterior distributions – Gamma, truncated inverse Gamma and Dirichlet distributions.  $\kappa_j$  ( $j = 1, \dots, J$ ) are simulated from their conditional posterior distributions on the grids of integer values in  $[\kappa_{\min}, \min\{\min_{i|z_{ij}=1} (y_i), \kappa_{\max}\}]$ , so that in this specific case of a discrete conditional posterior distribution a Griddy-Gibbs step is equivalent with a Gibbs step. The parameters  $\tilde{r}_j$  ( $j = 1, \dots, J$ ) are simulated using a random walk Metropolis(-Hastings) step using a Gaussian candidate distribution with standard deviation 0.1.

This procedure is extended to the case of an *unknown* number of non-empty components by applying the SFM MCMC algorithm. As noted earlier, this method consists of initially overfitting the number of mixture components and next use a sparse hierarchical prior on the mixture weights is to shrink the model space to a credible number of components.

To learn from the information set (data, likelihood and prior) how much sparsity is needed, the concentration parameter in the Dirichlet distribution  $\alpha$  has a Gamma hyperprior of the form:

$$\alpha \sim \text{Gamma}(a_\alpha, b_\alpha), \quad (23)$$

where  $E(\alpha) = a_\alpha/b_\alpha = \frac{1}{200}$  strongly favors small values. By Bayes theorem the conditional posterior distribution of  $\alpha$  given the partition  $\mathcal{P}$  of components across observations,  $p(\alpha|\mathcal{P}) \propto p(\alpha)p(\mathcal{P}|\alpha)$ , is given by

$$p(\alpha) \propto \alpha^{a_\alpha-1} \exp(-b_\alpha\alpha), \quad (24)$$

$$p(\mathcal{P}|\alpha) \propto \frac{\Gamma(J\alpha)}{\Gamma(n+J\alpha)} \prod_{k=1}^K \frac{\Gamma(n_k+\alpha)}{\Gamma(\alpha)}. \quad (25)$$

Sampling from this distribution is done with a Metropolis-Hastings step.

### 3.2 Simulation-based mode inference for multimodal and asymmetric count data

The simulation-based procedure proceeds as follows. One generates draws  $\theta^{(m)} = \{\mu^{(m)}, \tilde{r}^{(m)}, \kappa^{(m)}, \pi^{(m)}\}$  ( $m = 1, \dots, M$ ), each defining a mixture of  $J^{(m)}$  non-empty components of shifted negative binomial distributions. We emphasize that draws of parameters  $r^{(m)}$  and  $p^{(m)}$  can be easily recovered from  $\mu^{(m)}$  and  $\tilde{r}^{(m)}$  using (5) and (6). The number of modes is treated as an unknown parameter

and for each draw  $\theta^{(m)}$  the number of modes  $L^{(m)}$  and their location is calculated. Each draw  $\theta^{(m)}$  leads to a probability mass function

$$p(\tilde{y}|\theta^{(m)}) = \sum_{j=1}^{J^{(m)}} \pi_j^{(m)} p_{NB(\tilde{r}_j^{(m)}, \mu_j^{(m)})}(\tilde{y} - \kappa_j^{(m)}), \quad (26)$$

where  $p_{NB(\tilde{r}_j^{(m)}, \mu_j^{(m)})}(\cdot)$  is the probability mass function of the negative binomial distribution with mean  $\mu_j^{(m)}$  and transformed shape parameter  $\tilde{r}_j^{(m)}$ .

The modes  $y_m$  must satisfy either:

1.  $p(y_m - 1) < p(y_m) > p(y_m + 1)$ , or
2.  $p(y_m - 1) < p(y_m) = p(y_m + 1) = \dots = p(y_m + l - 1) > p(y_m + l)$ .

Case 1 is a unique mode which is clearly identified. Case 2 is a mode in which  $l$  consecutive values of the posterior predictive probability mass function are of equal value. We count this as a single mode, but keep track of each location. At the leftmost boundary 0 we count a mode if  $p(0) > p(1)$ , and at the rightmost boundary  $M$  we count a mode if  $p(M - 1) < p(M)$ . We note that simulation-based mode inference can be done efficiently with parallel Monte Carlo methods, see, e.g., [Basturk et al. \(2016\)](#).

## 4 Empirical results for Financial, Economic and DNA Count Data

We report results on the accuracy of the approximation of the PMF mixture of shifted negative binomial distributions to the count data of Figure 1 including a comparison of the results from our Bayesian method with frequentist testing methods. Next, credible evidence on locations of modes with their implied uncertainty is shown.

### 4.1 Defaulted payment installments

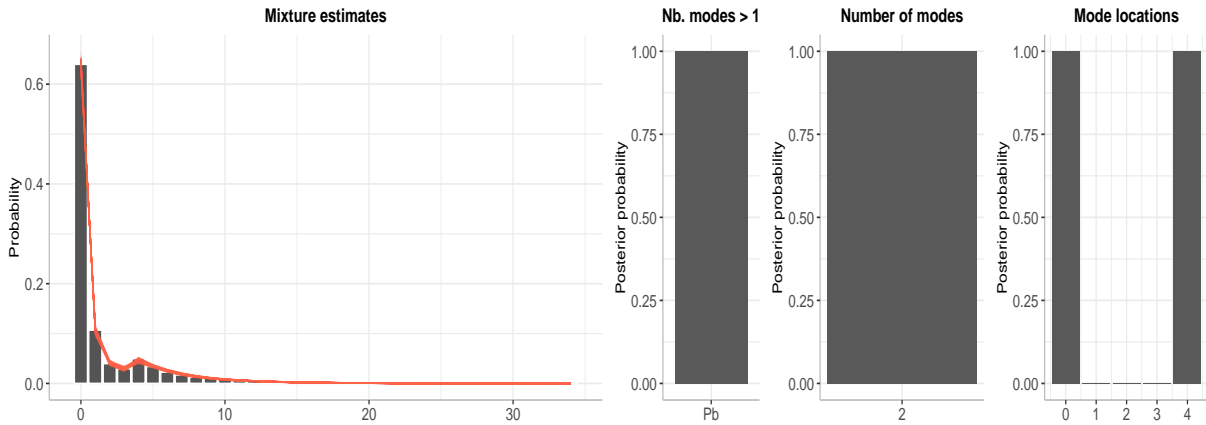
In finance, the probability that a client of a bank is able to repay a loan may depend on the number of defaulted payments and the uncertainty around it. Given the heterogeneity of the population of bank clients, analysis of this issue may involve multimodal data characteristics

of the distribution of repayments. Knowledge about these modal features is useful for risk management. Using data from [Dionne et al. \(1996\)](#) repayment details of clients obtaining a loan from a financial institution in Spain in 1990 are studied. This dataset consists of 4,691 count observations ranging from 0 to 34 defaulted loan payment installments which have been used to study the expected costs of risk misclassification in credit scoring. Since then they have been analyzed by, among others, [Karlis and Xekalaki \(2001\)](#) and [Woo and Sriram \(2007\)](#). We note that these data are a typical example of ‘zero inflated count data’ and show a high degree of ‘overdispersion’. Moreover, the possible multimodality suggests that a single negative binomial or a mixture of these distributions without the *shift* parameter may fail to approximate this data distribution given the large mass at zero and the fat tail.

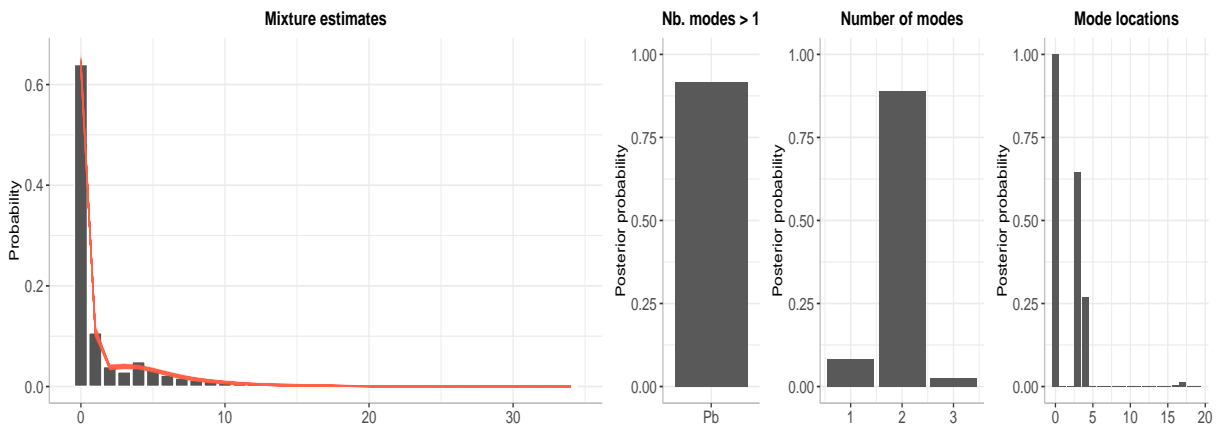
The estimated PMF of the mixture of shifted negative binomial distributions is shown in Panel (a) of Figure 8. It leads to reasonably accurate estimates from a visual diagnostic point of view. For comparison, we also provide a fitted mixture of shifted Poisson distributions in Panel (b). The shifted Poisson mixture puts quite a bit of mass at zero, around 4-5, and in the far right tail of the distribution around 10-20 while the shifted negative binomial mixture has a tighter fit around counts of 0 and 4, without categorizing the observations in the right tail as modes. This indicates a better ability of the shifted negative binomial mixture to capture the multimodality and the tail behavior. The overdispersion parameter in the shifted negative binomial distribution permits it to put relatively less weight on the observations in the right tail of the distribution, which implies that the estimation results are more robust in case of outliers. This prevents finding spurious modes. However it is still difficult to determine visually from a figure whether this is a feature of the data or the model.

In order to analyze which model is preferred by the data we consider the log-predictive likelihoods of each model in Table 3. These predictive likelihoods were computed using the cross-validation density (see [Eklund and Karlsson \(2007\)](#) and references therein). For the default data, we find that the log-predictive likelihood of the mixture of shifted negative binomial distributions is larger than the associated value for the mixture of shifted Poisson distributions. This suggests that the mixture of shifted negative binomial distributions does a better job at predicting the data.

The results so far motivate us to look at a more substantive analysis of the empirical results.



(a) Shifted Negative Binomial Mixture



(b) Shifted Poisson Mixture

Figure 8: Default Data, Fitted Mixtures and Modal Inference

The first issue is the basic question whether the PMF of the data is unimodal or multimodal which can be addressed through a straightforward hypothesis test procedure. Results from the [Silverman \(1981\)](#) test, the [Hall and York \(2001\)](#) test, the [Cheng and Hall \(1998\)](#) test and the [Hartigan and Hartigan \(1985\)](#) test are provided in the top row of [Table 2](#) together with the results using our proposed Bayesian procedure. For the frequentist test results, we find that while the Cheng and Hall test rejects the null hypothesis at the 5% level of significance, the remaining three tests all fail to reject the null. The Bayesian results from the shifted negative binomial mixture suggest that the posterior probability of uni-modality is 0%. This implies that the posterior probability of multimodality is 100%. This is further illustrated in Panel (a) of [Figure 8](#), which provides strong evidence in favor of a multimodal mixture distribution.

Table 2: Hypothesis test results

Data set	Silverman	Hall and York	Cheng and Hall	Hartigan and Hartigan	Bayesian
Default	0.51	0.09	0.01	0.48	0.00
Inflation	0.73	0.42	0.00	0.00	0.02
DNA	0.69	0.41	0.49	0.96	1.00

Hypothesis test results for  $H_0 : N_{modes} = 1$ ,  $H_1 : N_{modes} \geq 2$ , with  $N_{modes}$  the number of modes of the data distribution. Results for Frequentist tests are p-values associated with the null hypothesis. Result for the Bayesian test is the posterior probability of uni-modality.

We re-emphasize that a key advantage of our proposed Bayesian procedure is the direct obtained information about the number of modes, their locations and the degree of uncertainty surrounding them. The central column of Panel (a) in Figure 8 shows the number of modes for the shifted negative binomial mixture along with their posterior probabilities. There is credible posterior support for a total of two modes, with 100% posterior probability for two modes with a fat right tail. The right column of Panel (a) in Figure 8 shows the location of the modes along with their posterior probabilities. We observe that there are modes at values 0 and 4. The results for the shifted Poisson mixture indicate however more than two modes (albeit with low probability) where the third mode appears to be a spurious one.

Table 3: Log-predictive Likelihood Results

Data set	Shifted Negative Binomial	Shifted Poisson
Default	-6984	-7015
Inflation	-1373	-1374
DNA	-1674	-1773

Notes: Average log-predictive likelihoods of all observations, computed using the cross-validation density approach (see [Eklund and Karlsson \(2007\)](#) and references therein).

In conclusion, our analysis using a mixture of shifted negative binomial distributions with financial count data leads to credible estimates of number and locations of modes which may be used for the institution’s policy of granting loans. Our results could be used by a risk manager

to categorize clients according to their defaulted installment behavior. For each client, it is typically the case that information is provided regarding their individual characteristics, e.g., income, profession, etc, at the beginning of the contract. In further research such micro-data could be used within our methodology as explanatory variables to develop more detailed risk management profiles and quantify the expected costs of risk misclassification in credit scoring.

## 4.2 Inflation Expectations

We consider count data from survey responses to the question “By what percentage do you expect prices to go up, on average, during the next 12 months?”, from the University of Michigan’s Surveys of Consumers, which are shown in Figure 1. The right panel in Figure 1 contains response data in 2023 with a total of 480 observations. It is notable that many respondents select round numbers 10%, 20% or even 30% which suggests that such behavior is indicative of high uncertainty. Visually, there are two lower modes at 0% and 3-4%, suggesting a de-anchoring of inflation expectations away from the central bank’s target of 2% and a mode at 30%. This suggests that inflation expectations have widespread disagreement, both within the subgroup with ‘credible’ expectations and the subgroup with ‘incredible’ expectations. Heterogeneity within the joint distribution of private agents’ inflation expectations has been linked with alternative learning (Pfajfar and Santoro, 2010) and economic (il)literacy (Burke and Manz, 2014). In these studies multimodality is typically “eyeballed” from the data but our flexible class of shifted negative binomial mixtures can be used to formally detect and explore this phenomenon to which we now turn.

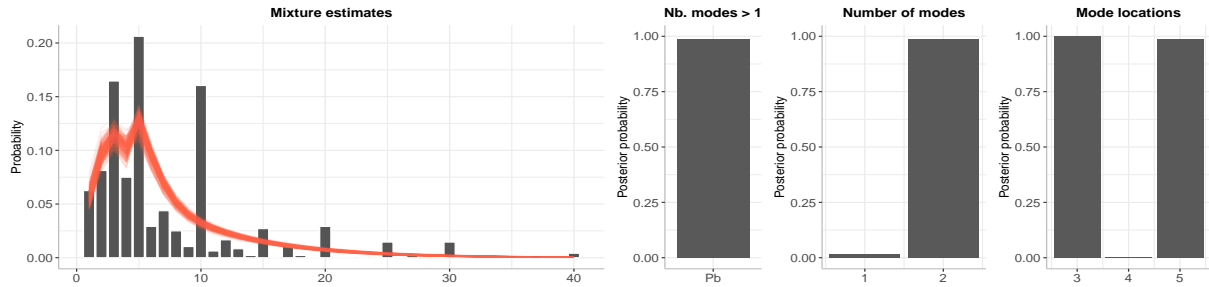
The estimated PMF of the mixture of shifted negative binomial distributions is shown in Panel (a) of Figure 9. For comparison, the fitted mixture of shifted Poisson distributions is shown in Panel (b). It is seen that both mixtures have a similar distribution of mass in the center but that the shifted Poisson mixture also puts some mass for a mode in the far right tail of the distribution while the shifted negative binomial mixture does not categorize these observations as modes. This again indicates a better ability of the shifted negative binomial mixture to capture the tail behavior. The overdispersion parameter in the shifted negative binomial distribution permits it to put relatively less weight on the observations in the right tail of the distribution, which implies that the estimation results are more robust in case of outliers.

This prevents finding spurious modes. The log-predictive likelihoods of each model are shown in the middle row of Table 3. It is seen that in terms of this measure of fit both mixtures do not differ much.

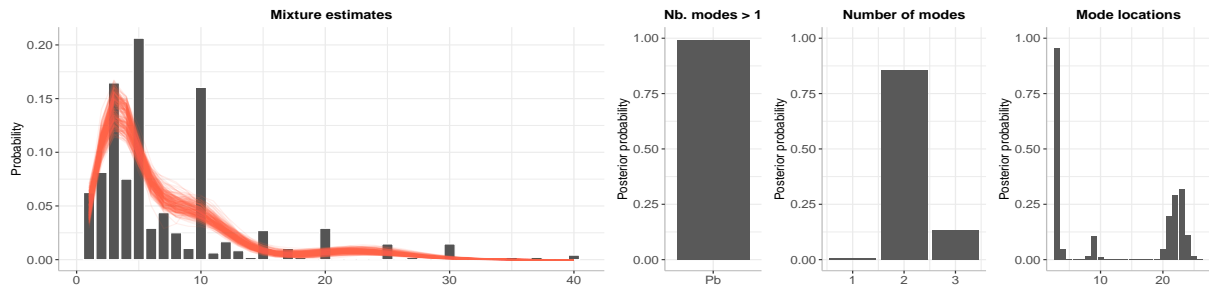
Results from the frequentist and Bayesian test procedures about the existence of multiple modes are shown in the middle row of Table 2. For the frequentist test results we find conflicting evidence on the existence of multiple mode: two test do not reject the null hypothesis of a single mode at the 5% level of significance while two tests reject the null. The Bayesian result from the shifted negative binomial mixture suggest that the posterior probability of unimodality is 2%.

The central column of Panel (a) in Figure 9 shows that both mixtures give evidence about more than one mode. However, the shifted negative binomial mixture gives substantial support for two modes while the mixture of Poissons gives some posterior probability to three modes. The right column of Panel (a) in Figure 9 shows the location of the modes along with their posterior probabilities. The third mode of the Poisson mixture seems to be a spurious one.

We conclude that the evidence about the location of modes from the shifted binomial mixture is more credible than the shifted Poisson mixture from an economic perspective.



(a) Shifted Negative Binomial Mixture



(b) Shifted Poisson Mixture

Figure 9: Inflation Expectations Data, Fitted Mixtures and Modal Inference



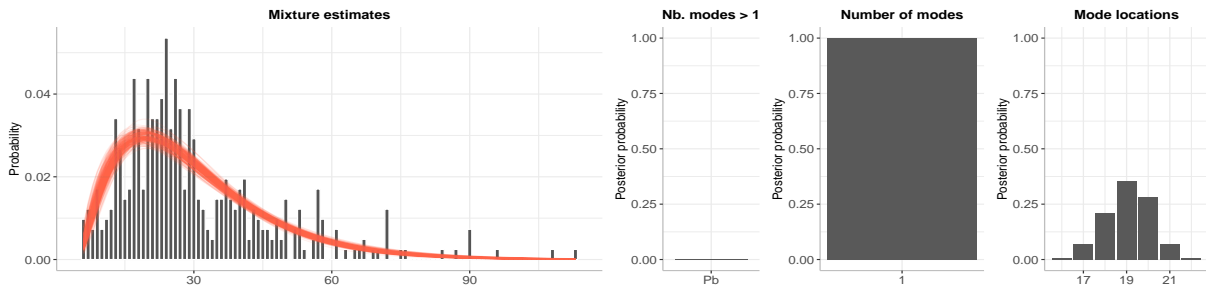
### 4.3 DNA tandem repeats data

In bioinformatics it is known that more than half of the human genome consists of repetitive DNA. Tandemly repeated DNA sequences comprise a substantial proportion thereof. In different populations one can identify commonalities and differences of so-called Macro-Satellite Repeats (MSRs) as a particular group of repetitive DNA sequences with evidence of multimodal size distributions. Estimating the number and location of modes in such distributions is important since observed deviations in the configurations of MSRs may indicate genetic diseases, see [Bruce et al. \(2009\)](#) and [Balog et al. \(2012\)](#) among several others. [Schaap et al. \(2013\)](#) provide the first comprehensive study of MSRs in different human populations, using mixtures of shifted Poisson distributions with a given number of components – based on the Bayesian Information Criterion (BIC). We construct a mixture of shifted negative binomial distributions to estimate posterior features using counts of DNA tandem repeat data for the case of D4Z4 illustrated in [Figure 10](#). These data are obtained from 270 unrelated human DNA samples from Asian, African and Caucasian origin, see [Schaap et al. \(2013\)](#). It is of substantial interest to analyze the number and location of modes in the data, since differences in these values may next be linked to e.g. genetic diseases.

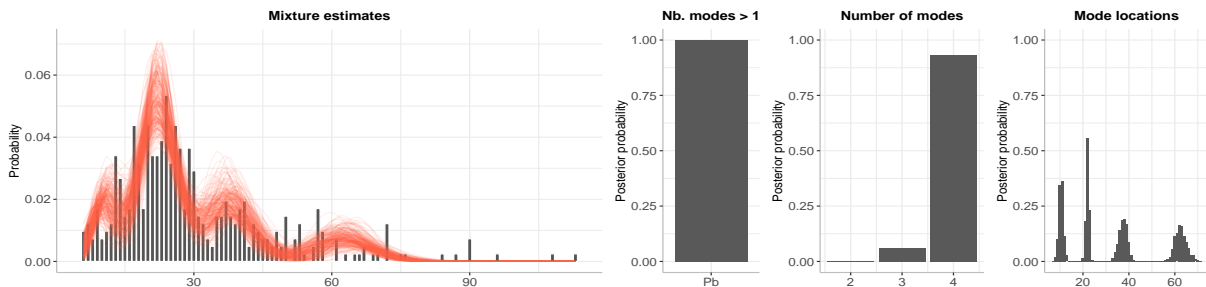
The estimated PMF for the mixture of shifted negative binomial distributions is shown in panel (a) of [Figure 10](#). For comparison in panel (b) gives the PMF of the estimated mixture of shifted Poissons. The shifted Poisson mixture indicates four modes while the shifted negative binomial mixture indicates unimodality without categorizing the observations in both tails as modes. We re-emphasize that the overdispersion parameter in the shifted negative binomial distribution makes the estimation of this distribution more robust in case of outliers. The log-predictive likelihood of the shifted binomial mixture in [Table 3](#) is substantially better than the shifted Poisson mixture.

Next, we explore the results of the frequentist tests and our Bayesian procedure about possible multimodality given in the bottom row of [Table 2](#). All four frequentist tests do not reject the null hypothesis of unimodality at the 5% level of significance and the Bayesian procedure indicates that the posterior probability of unimodality is 100%.

The second column of each sub-panel in [Figure 10](#) shows the possible number of modes along



(a) Shifted Negative Binomial Mixture



(b) Shifted Poisson Mixture

Figure 10: DNA Data, Fitted Mixtures and Modal Inference

with their posterior probabilities. There we observe that the negative binomial mixture holds on to one mode but the shifted Poisson to four modes. The third column of each sub-panel in Figure 10 shows the location of the modes along with their posterior probabilities. We conclude that the statistical evidence about the number of modes from the mixture of shifted binomials is more credible than the results of the models of the mixture of shifted Poissons, based on the predictive likelihoods.

## 5 Conclusions

We introduced a simulation-based Bayesian approach using a mixture of shifted negative binomial distributions in order to estimate multimodal and asymmetric distributions for count data sets and determine the number and location of modes with substantial posterior probability. For computational efficiency, we applied the sparse finite mixture Markov chain Monte Carlo method from [Malsiner-Walli et al. \(2016\)](#) in order to estimate the number of non-empty components of the mixture.

For empirical analysis we made use of three count data sets from finance, economics and

bioinformatics. Analysis of the financial data about repayment details of clients obtaining a loan from a financial institution in Spain in 1990 yields evidence that the data distribution is bimodal and has a fat right tail. A mixture of shifted negative binomial distributions leads to credible estimates of location of modes which may be useful information for the institution's policy of granting loans.

Using count data of inflation expectations, we again find evidence that the data distribution is bimodal and has a fat right tail, where the evidence about the location of modes from the shifted binomial mixture is more credible than the shifted Poisson mixture from an economic perspective.

Using count data of DNA tandem repeats from [Schaap et al. \(2013\)](#) we found that applying the proposed model and method provides substantial posterior probability for a unimodal distribution in agreement with the frequentist tests.

More generally, our results indicate that the proposed methodology provides effective tools for quantification of accuracy and uncertainty of the modes of multimodal data distributions which is useful for forecasting and policy analysis.

In future research, we plan to compare the proposed method with other tests to detect multimodality and to estimate in more detail quantiles of non-standard distributions for risk analysis. The method can also be extended to multivariate (e.g. panel) data.

## References

- Balog J, Miller D, Sanchez-Curtailles E, Carbo-Marques J, Block G, Potman M, De Knijff P, Lemmers RJ, Tapscott SJ, Van Der Maarel SM. 2012. Epigenetic regulation of the x-chromosomal macrosatellite repeat encoding for the cancer/testis gene ct47. *European Journal of Human Genetics* **20**: 185.
- Basturk N, Grassi S, Hoogerheide L, Van Dijk HK. 2016. Parallelization Experience with Four Canonical Econometric Models Using ParMitISEM. *Econometrics* **4**: 11.
- Bruce H, Sachs N, Rudnicki D, Lin S, Willour V, Cowell J, Conroy J, McQuaid D, Rossi M, Gaile D, Nowak N, Holmes S, Sklar P, Ross C, Delisi L, Margolis R. 2009. Long tandem repeats as a form of genomic copy number variation: structure and length polymorphism of a chromosome 5p repeat in control and schizophrenia populations. *Psychiatric Genetics* **19**: 64.
- Burke MA, Manz M. 2014. Economic literacy and inflation expectations: Evidence from a laboratory experiment. *Journal of Money, Credit and Banking* **46**: 1421–1456.
- Chaudhuri P, Marron JS. 1999. Sizer for exploration of structures in curves. *Journal of the American Statistical Association* **94**: 807–823.
- Cheng MY, Hall P. 1998. Calibrating the excess mass and dip tests of modality. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **60**: 579–589.
- Dionne G, Artís M, Guillén M. 1996. Count data models for a credit scoring system. *Journal of Empirical Finance* **3**: 303–325.
- Eklund J, Karlsson S. 2007. Forecast combination and model averaging using predictive measures. *Econometric Reviews* **26**: 329–363.
- Fischer NI, Mammen E, Marron JS. 1994. Testing for multimodality. *Computational Statistics & Data Analysis* **18**: 499–512.
- Frühwirth-Schnatter S. 2006. *Finite mixture and Markov switching models: Modeling and ap-*

- plications to random processes*. Springer Series in Statistics. New York/Berlin/Heidelberg: Springer.
- Frühwirth-Schnatter S. 2011. Dealing with label switching under model uncertainty. In Mengersen K, Robert CP, Titterton D (eds.) *Mixture estimation and applications*, chapter 10. Wiley, 193–218.
- Frühwirth-Schnatter S, Malsiner-Walli G. 2019. From here to infinity: sparse finite versus Dirichlet process mixtures in model-based clustering. *Advances in Data Analysis and Classification* **13**: 33–64.
- Green PJ. 1995. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* **82**: 711–732.
- Hall P, Ooi H. 2004. Attributing a probability to the shape of a probability density. *Annals of Statistics* **32**: 2098–2123.
- Hall P, York M. 2001. On the calibration of Silverman’s test for multimodality. *Statistica Sinica* **11**: 515–536.
- Hartigan JA, Hartigan PM. 1985. The DIP test of unimodality. *The Annals of Statistics* **13**: 70–84. ISSN 00905364.
- Hartigan PM. 1985. Computation of the DIP statistic to test for unimodality. *Journal of the Royal Statistical Society. Series C (Applied Statistics)* **34**: 320–325. ISSN 00359254.
- Karlis D, Xekalaki E. 2001. Robust inference for finite Poisson mixtures. *Journal of Statistical Planning and Inference* **93**: 93–115.
- Malsiner-Walli G, Frühwirth-Schnatter S, Grün B. 2016. Model-based clustering based on sparse finite Gaussian mixtures. *Statistics and Computing* **26**: 303–324.
- Malsiner-Walli G, Frühwirth-Schnatter S, Grün B. 2017. Identifying mixtures of mixtures using bayesian estimation. *Journal of Computational and Graphical Statistics* **26**: 285–295.
- McLachlan G, Peel D. 2004. *Finite mixture models*. John Wiley & Sons.

- Miller JW, Harrison MT. 2013. A simple example of dirichlet process mixture inconsistency for the number of components. *Advances in neural information processing systems* **26**.
- Minnotte MC. 1997. Nonparametric testing of the existence of modes. *The Annals of Statistics* **25**: 1646–1660.
- Montmort Pd. 1713. Essai d’analyse sur les jeux de hasard. 2nd ed. Quillau, Paris.
- Müller DW, Sawitzki G. 1991. Excess mass estimates and tests for multimodality. *Journal of the American Statistical Association* **86**: 738–746.
- Pascal B. 1679. *Varia opera mathematica. d. petri de fermat. tolosae.*
- Pfajfar D, Santoro E. 2010. Heterogeneity, learning and information stickiness in inflation expectations. *Journal of Economic Behavior & Organization* **75**: 426–444.
- Richardson S, Green PJ. 1997. On Bayesian analysis of mixtures with an unknown number of components. *Journal of the Royal Statistical Society. Series B (Methodological)* **59**: pp. 731–792. ISSN 00359246.
- Schaap M, Lemmers RJLF, Maassen R, van der Vliet PJ, Hoogerheide LF, Van Dijk HK, Baştürk N, de Knijff P, an der Maarel SM. 2013. Genome-wide analysis of macrosatellite repeat copy number variation in worldwide populations: evidence for differences and commonalities in size distributions and size restrictions. *BMC Genomics* **14**: 143.
- Silverman BW. 1981. Using kernel density estimates to investigate multimodality. *Journal of the Royal Statistical Society. Series B (Methodological)* **41**: 97–99.
- Woo MJ, Sriram T. 2007. Robust estimation of mixture complexity for count data. *Computational Statistics & Data Analysis* **51**: 4379 – 4392. ISSN 0167-9473.
- Zhou M, Carin L. 2013. Negative binomial process count and mixture modeling. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **37**: 307–320.

## A Specification of the negative binomial density as the marginal of a continuous mixture of Poisson densities

The negative binomial density can be specified as the marginal density of a continuous mixture of Poisson densities, where the mixing distribution of the Poisson rate is a gamma distribution. Consider the negative binomial distribution as a  $\text{Poisson}(\lambda)$  distribution, where  $\lambda > 0$  is a random variable with a gamma distribution with shape parameter  $r$  and rate parameter  $\frac{p}{1-p}$ :

$$\begin{aligned}
 p(y_i|r, p) &= \int_0^\infty p_{\text{Poisson}(\lambda_i)}(y_i) p_{\text{Gamma}(r, \frac{p}{1-p})}(\lambda_i) d\lambda_i \\
 &= \int_0^\infty \frac{\lambda_i^{y_i} \exp(-\lambda_i)}{y_i!} \frac{\left(\frac{p}{1-p}\right)^r \lambda_i^{r-1} \exp\left(-\lambda_i \frac{p}{1-p}\right)}{\Gamma(r)} d\lambda_i \\
 &= \frac{\left(\frac{p}{1-p}\right)^r}{y_i! \Gamma(r)} \int_0^\infty \lambda_i^{y_i+r-1} \exp\left(-\frac{\lambda_i}{1-p}\right) d\lambda_i \\
 &= \frac{\left(\frac{p}{1-p}\right)^r}{y_i! \Gamma(r)} (1-p)^{y_i+r} \Gamma(y_i+r) \\
 &= \frac{\Gamma(y_i+r)}{y_i! \Gamma(r)} (1-p)^{y_i} p^r, \tag{27}
 \end{aligned}$$

where the fourth equality follows from the fact that for  $a > -1$  and  $b > 0$  we have

$$\int_0^\infty \lambda^a \exp(-b\lambda) d\lambda = b^{-(a+1)} \Gamma(a+1) \tag{28}$$

with  $a = y_i + r - 1$  and  $b = \frac{1}{1-p}$ .