Bayesian Analysis (2006) **1**, Number 2, pp. 00–00

Conjugate Analysis of the Conway-Maxwell-Poisson Distribution

Joseph B. Kadane*, Galit Shmueli†, Thomas P. Minka‡,
Sharad Borle§, and Peter Boatwright¶

Abstract. This article explores a Bayesian analysis of a generalization of the Poisson distribution. By choice of a second parameter ν, both under-dispersed and over-dispersed data can be modeled. The Conway-Maxwell-Poisson distribution forms an exponential family of distributions, so it has sufficient statistics of fixed dimension as the sample size varies, and a conjugate family of prior distributions. The article displays and proves a necessary and sufficient condition on the hyperparameters of the conjugate family for the prior to be proper, and it discusses methods of sampling from the conjugate distribution. An elicitation program to find the hyperparameters from the predictive distribution is also discussed.

Keywords: ba0002, convexity, exponential family, Jensen's inequality, Poisson distribution, sufficient statistics

1 Introduction

The Conway-Maxwell-Poisson (COM-Poisson) distribution (Conway and Maxwell (1961)) has been used in studies of word lengths (Wimmer et al. (1994), Wimmer and Altmann (1996)) and of marketing, in which the distribution of the times (in weeks) to next purchase had tails heavier than those of a Poisson distribution (Boatwright, Borle, and Kadane (2003)). For a general review, see Shmueli, Minka, Kadane, Borle, and Boatwright (2005). Since over-dispersion is a fairly frequent concern (Breslow (1990), Dean (1992)), it is worthwhile to explore further the properties of the COM-Poisson distribution, and in particular the conjugate family of prior distributions associated with it.

The COM-Poisson distribution is a generalization of the Poisson distribution, allowing for over- or under-dispersion. Its probability mass function is

*Department of Statistics, Carnegie Mellon University, Pittsburgh, PA, http://www.stat.cmu.edu/~kadane

†Department of Decision and Information Technologies, University of Maryland, College Park, MD, http://www.smith.umd.edu/faculty/gshmueli/

‡Microsoft Research Ltd., Cambridge, UK, mailto:minka@microsoft.com

§Jones Graduate School of Management, Rice University, Houston, TX, http://jonesgsm.rice.edu/Faculty/Sharad_Borle/

¶Tepper School of Business, Carnegie Mellon University, Pittsburgh, PA, http://business.tepper.cmu.edu/display_faculty.aspx?id=42

$$P\{X = x \mid \lambda, \nu\} = \frac{\lambda^x}{(x!)^\nu} \cdot \frac{1}{Z(\lambda, \nu)} \quad x = 0, 1, 2, \ldots \quad , \text{ where} \tag{1}$$

$$Z(\lambda, \nu) = \sum_{j=0}^{\infty} \frac{\lambda^j}{(j!)^\nu}. \tag{2}$$

The ratio of probabilities of successive integers is

$$\frac{P\{X = x - 1 \mid \lambda, \nu\}}{P\{X = x \mid \lambda, \nu\}} = \frac{x^\nu}{\lambda}. \tag{3}$$

When $\nu = 1$, a Poisson distribution results. Values of $\nu > 1$ indicate under-dispersion relative to the Poisson, while $\nu < 1$ indicates over-dispersion. As $\nu \to \infty$, $Z(\lambda, \nu) \to 1+\lambda$, and the COM-Poisson approaches a Bernoulli distribution with $P\{X = 1 \mid \lambda, \nu\} = \frac{\lambda}{\lambda+1}$.

When $\nu = 0$ and $\lambda < 1$, $Z(\lambda, \nu)$ is a geometric sum:

$$Z(\lambda, \nu) = \sum_{j=0}^{\infty} \lambda^j = \frac{1}{1 - \lambda},$$

and the distribution itself is geometric:

$$P\{X = x \mid \lambda, \nu)\} = \lambda^x (1 - \lambda) \text{ for } x = 0, 1, 2, \ldots \tag{4}$$

When $\nu = 0$ and $\lambda \geq 1$, $Z(\lambda, \nu)$ does not converge, and the distribution is undefined.

The remainder of this article is organized as follows: Section 2 reports on the sufficient statistics and conjugate family associated with the COM-Poisson distribution and the corresponding predictive distribution. Section 3 shows the use of an elicitation program (available on the web) for the hyperparameters of the COM-Poisson. Section 4 gives a numerical example of the computation of the posterior for the COM-Poisson with a conjugate prior distribution. Section 5 explores the marginal and conditional distributions arising from the conjugate distribution.

2 Sufficient Statistics, Conjugate Family and Predictive Distribution

2.1 Sufficient Statistics

The likelihood for a set of n identically distributed observations $x_1, x_2, \ldots, x_n$ is

$$\begin{aligned} L(x_1, x_2, \ldots, x_n \mid \lambda, \nu) &= \frac{\prod_{i=1}^n \lambda^{x_i}}{\left(\prod_{i=1}^n x_i!\right)^\nu} Z^{-n}(\lambda, \nu) = & (5) \\ &= \lambda^{\sum_{i=1}^n x_i} e^{-\nu \sum_{i=1}^n \log(x_i!)} Z^{-n}(\lambda, \nu) & (6) \\ &= \lambda^{S_1} e^{-\nu S_2} Z^{-n}(\lambda, \nu) & (7) \end{aligned}$$

where $S_1 = \sum_{i=1}^n x_i$ and $S_2 = \sum_{i=1}^n \log(x_i!)$. By the Factorization Theorem, (S_1, S_2) are sufficient statistics. Furthermore, (7) displays the COM-Poisson as an exponential family of distributions.

2.2 Conjugate Family

Since the COM-Poisson forms an exponential family, there is a conjugate family of priors such that, whatever the data, the posterior is of the same form. For this distribution, the conjugate prior density is of the form

$$h(\lambda, \nu) = \lambda^{a-1} e^{-\nu b} Z^{-c}(\lambda, \nu) \kappa(a, b, c) \tag{8}$$

for $\lambda > 0$ and $\nu \geq 0$, where $\kappa(a, b, c)$ is the integration constant. Then the posterior is of the same form, with $a' = a + S_1, b' = b + S_2$, and $c' = c + n$. The distribution whose density is given in (8) can be thought of as an extended bivariate Gamma distribution.

In order for equation (8) to constitute a density, it must be non-negative and integrate to one. In other words, the values of a, b, and c that lead to a finite $\kappa^{-1}(a, b, c)$, which is given by

$$\kappa^{-1}(a, b, c) = \int_0^\infty \int_0^\infty \lambda^{a-1} e^{-b\nu} Z^{-c}(\lambda, \nu) d\lambda d\nu, \tag{9}$$

will lead to a proper density.

Using Jensen's inequality and the convexity of the log-gamma function, the appendix shows that a necessary and sufficient condition for a finite $\kappa^{-1}(a, b, c)$ is:

$$b/c > \log(\lfloor a/c \rfloor !) + (a/c - \lfloor a/c \rfloor) \log(\lfloor a/c \rfloor + 1) \tag{10}$$

where $\lfloor k \rfloor$ denotes the floor of k. The set of hyperparameters (a, b, c) satisfying (10) is necessarily closed under sampling; *i.e.*, if the prior satisfies (10), so will the posterior for every possible datum x.

2.3 Predictive Distribution

Given values for the hyperparameters a, b and c, the predictive probability function is given by

$$\begin{aligned} P(X = x \mid a, b, c) &= \int_0^\infty \int_0^\infty P(X = x \mid \lambda, \nu) h(\lambda, \nu \mid a, b, c) d\lambda d\nu \\ &= \int_0^\infty \int_0^\infty \left[\frac{\lambda^x}{(x!)^\nu Z(\lambda, \nu)} \right] \left[\lambda^{a-1} e^{-b\nu} Z^{-c}(\lambda, \nu) \kappa(a, b, c) \right] d\lambda d\nu \\ &= \kappa(a, b, c) \int_0^\infty \int_0^\infty \lambda^{a+x-1} e^{-\nu(b+\log(x!))} [Z(\lambda, \nu)]^{-(c+1)} d\lambda d\nu = \\ &= \frac{\kappa(a, b, c)}{\kappa(a + x, b + \log(x!), c + 1)} \end{aligned} \tag{11}$$

where $\kappa(a,b,c)$ is defined in (9).

The calculation of the two double integrals in (11) can be done by using a non-equally spaced grid over the λ, ν space. Numerical routines that aid in the specification of this region turn out to be more robust in the transformed space

$$\lambda^* = \log(\lambda) \tag{12}$$
$$\nu^* = \log(\nu). \tag{13}$$

3 Elicitation of the Hyperparameters of the Conjugate Distribution

To facilitate practice we have created an application to help the practitioner select the hyperparameters a, b, and c or to aid the statistician in eliciting these values from the expert who understands the context of the data (see Kadane and Wolfson (1997) and Garthwaite et al. (2005) for reviews of elicitation issues). Even when a prior is of the form of a well-used Gaussian distribution, practitioners often have difficulty setting covariance parameters of the prior (Barnard, McCulloch, and Xiao-Li (2000)). Here, where the distribution is likely to be new to the practitioner, it is even more difficult to give meaningful values to a, b, and c. Because the practitioner is very likely to have some degree of knowledge about $P(X = x)$, our application calculates and plots the predictive distribution for selected values of a, b, and c. Our application aid can be found on the web at `http://www.stat.cmu.edu/COM-Poisson/`.The application has slider bars which allow the user to select and change values of a, b, and c. The application ensures that the constraints on a, b, and c are met (equation (10)). If the constraints are not met, the user is warned in a text window that they should alter their choice of the triplet (a, b, c); if the user ignores the warning, the program will give an error message. The application calculates and plots a bar chart of the predictive density for each triplet (a, b, c). The practitioner can then use knowledge of $P(X = x)$ to select reasonable values for a, b, and c.

Calculation of the predictive density involves integrals, which are calculated numerically in the elicitation application. The numerical integration requires a bounded region, specified by limits on each variable, over which the integral is to be taken. Since this approximates an open space with a closed region, it is necessary to ensure that the integrated region contains almost all of the probability mass. One way to choose such boundaries is to integrate over the rectangular region defined by

$$\widehat{\lambda^*} \pm l\sqrt{Var(\widehat{\lambda^*})} \tag{14}$$
$$\widehat{\nu^*} \pm l\sqrt{Var(\widehat{\nu^*})} \tag{15}$$

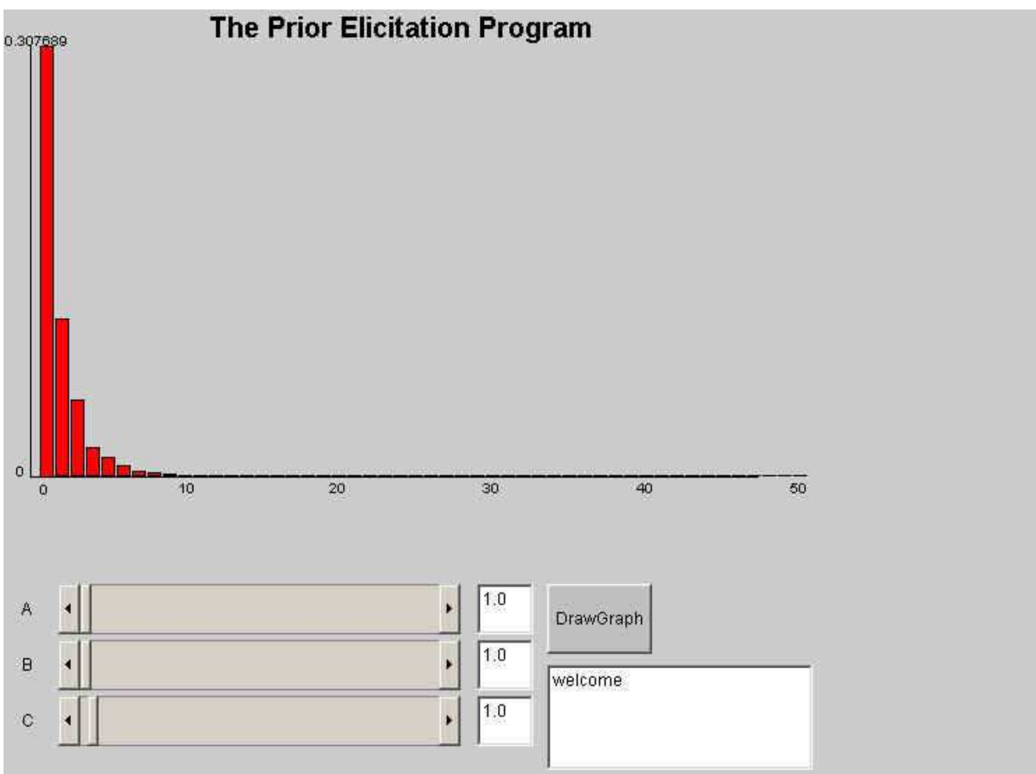
where $(\widehat{\lambda^*}, \widehat{\nu^*})$ are the points maximizing the prior density, while $Var(\widehat{\lambda^*})$ and $Var(\widehat{\nu^*})$ are the diagonal elements of the inverse of the negative Hessian, and l is a number to be chosen.

To find the point of maximizing (λ^*, ν^*) we use a globally convergent Newton Raphson routine, a double precision version of the function newt of Numerical Recipes (Flannery et al. (1994)). For the purpose of approximating the moments, a truncation of the infinite series is used that bounds the resulting error to be less than 1.0e-8.

Since λ^* and ν^* are highly correlated, integration over a rectangular space is inefficient since many calculations are made over regions with trivial probability mass. We therefore set integration boundaries using the marginal distributions $f(\nu^*|\lambda^*)$ of the bivariate normal distribution with mean $(\widehat{\lambda^*}, \widehat{\nu^*})$ and covariance matrix equal to the inverse of the negative Hessian. We retain the bounds on λ^* as given in equation (14) and use the mean and variance of $\nu^*|\lambda^*$ to replace $\widehat{\nu^*}$ and $Var(\nu^*)$ in equation (15). In our elicitation program, we set $l = 8$.

A snapshot of the predictive distribution plot, created by our web application, is given in Figure (1). The prior hyperparameters entered in this case were $a = b = c = 1$.

Figure 1: Predictive distribution for $a = b = c = 1$, using the Web application



4 A Numerical Example of Analysis with the COM-Poisson and a Conjugate Prior

Feller (1950, 1957)[p. 151, Table 5, Experiment 4; p. 152] gives the following data on chromosome interchanges induced by x-ray irradiation, from Catchside, Lea, and Thoday (1945, 1946):

0	2278
1	273
2	15
3+	0

Table 1: Numbers of cells with k interchanges.

Then

$$\begin{aligned} \sum x_i &= 273 + 15(2) = 303 \ ; \ n = 2566 \\ \sum \log x_i! &= 15 \log 2 = 10.4. \end{aligned}$$

Thus for a conjugate prior characterized by (a, b, c), the posterior is again conjugate, and characterized by $(a + 303,\ b + 10.4,\ c + 2566)$.

5 Marginal and Conditional Distribution

We investigate the marginal and conditional densities of λ and ν that arise from a bivariate distribution of the form:

$$h(\lambda, \nu) = \lambda^{a-1} e^{-b\nu} [Z(\lambda, \nu)]^{-c} \kappa(a, b, c) \tag{16}$$

where $a > 0$, $b > 0$, and $c > 0$ are hyperparameters. The marginal density of λ can then be expressed as:

$$h_1(\lambda) = \int_{\nu=0}^{\infty} h(\lambda, \nu) d\nu \quad = \quad \lambda^{a-1} \kappa(a, b, c) \int_{\nu=0}^{\infty} e^{-b\nu} [Z(\lambda, \nu)]^{-c} d\nu. \tag{17}$$

The conditional density of ν given λ is then:

$$h(\nu \mid \lambda) = \frac{h(\lambda, \nu)}{h_1(\lambda)} \propto e^{-b\nu} [Z(\lambda, \nu)]^{-c}. \tag{18}$$

Using the same logic, the marginal distribution of ν can be expressed as

$$h_2(\nu) = \int_{\lambda=0}^{\infty} h(\lambda, \nu) d\lambda = e^{-b\nu} \kappa(a, b, c) \int_{\lambda=0}^{\infty} \lambda^{a-1} [Z(\lambda, \nu)]^{-c} d\lambda. \tag{19}$$

The conditional density of λ given ν is then:

$$h(\lambda \mid \nu) = \frac{h(\lambda, \nu)}{h_2(\nu)} \propto \lambda^{a-1} [Z(\lambda, \nu)]^{-c}. \tag{20}$$

For the three well known cases ($\nu = 0, 1, \infty$) this reduces to:

$$\begin{aligned}
\lambda \mid_{\nu=0} &\sim Beta(a, c+1), && (21)\\
\lambda \mid_{\nu=1} &\sim \Gamma(a, c), && (22)\\
\lambda \mid_{\nu=\infty} &\sim F(2a, 2(c-a)) \qquad \text{if } c > a. && (23)
\end{aligned}$$

The result in (23) is obtained by noticing that

$$\frac{\lambda}{1+\lambda} \mid_{\nu=\infty} \sim Beta(a, c-a), \tag{24}$$

and using the relation between the Beta and F distributions (Mood, Graybill, and Boes (1974, p. 249)).

5.1 Generating data from $h(\lambda|\nu)$ when $c > a$

The above special cases hint that the conditional for λ always has a shorter tail than an F distribution. By constructing an F-type dominating curve, we can get a rejection sampling scheme for the conditional in cases where $c > a$. That $Z(\lambda, \nu)$ is convex in λ is a consequence of the fact that its second derivative with respect to λ is $\sum_{j=2}^{\infty} \frac{j(j-1)\lambda^{j-2}}{(j!)^{\nu}}$, which is positive. Convexity means that all tangent lines are lower bounds; thus for any λ_0 we have

$$\begin{aligned}
Z(\lambda, \nu) &\geq q(\lambda_0)(\lambda - \lambda_0) + Z(\lambda_0, \nu), && (25)\\
\text{where } q(\lambda_0) &= \left.\frac{dZ(\lambda, \nu)}{d\lambda}\right|_{\lambda=\lambda_0} = Z(\lambda_0, \nu) E[x; \lambda = \lambda_0, \nu]. && (26)
\end{aligned}$$

Substituting this bound into (20) gives an F approximation to $p(\lambda|\nu)$ as well as the minimal scale factor to make it dominate the true conditional. This gives the following rejection algorithm:

1. Choose λ_0 and compute $q(\lambda_0)$.

2. Draw λ from the F distribution proportional to $\lambda^{a-1}\left(q(\lambda_0)(\lambda - \lambda_0) + Z(\lambda_0, \nu)\right)^{-c}$. This can be done by dividing a $\Gamma(a)$ variate by a $\Gamma(c-a)$ variate and then dividing by the constant $Z(\lambda_0, \nu) - \lambda_0 q(\lambda_0)$.

3. Draw a uniform variate $u \sim U(0, 1)$ and accept λ if

$$u \leq \frac{Z(\lambda, \nu)^{-c}}{\left(q(\lambda_0)(\lambda - \lambda_0) + Z(\lambda_0, \nu)\right)^{-c}} \quad . \tag{27}$$

4. If λ is rejected, repeat from step 2.

5.2 Generating data from $h(\lambda|\nu)$, $\nu < 1$

When $\nu < 1$, an even tighter dominating curve is possible. The special cases hint that when $\nu < 1$, the conditional for λ has a shorter tail than a Gamma distribution. By constructing a Gamma-type dominating curve, we can get a more efficient rejection sampling scheme, which applies regardless of whether $c > a$. In appendix B, it is shown that $\log Z(\lambda, \nu)$ is convex in λ. Again, convexity means that all tangent lines are lower bounds, thus for any λ_0 we have

$$\begin{aligned} \log Z(\lambda,\nu) &\geq q(\lambda_0)(\lambda-\lambda_0) + \log Z(\lambda_0,\nu), & (28)\\ Z(\lambda,\nu) &\geq \exp(q(\lambda_0)(\lambda-\lambda_0))Z(\lambda_0,\nu), & (29)\\ \text{where } q(\lambda_0) &= \left.\frac{d\log Z(\lambda,\nu)}{d\lambda}\right|_{\lambda=\lambda_0} = E[x;\lambda=\lambda_0,\nu]. & (30)\end{aligned}$$

Substituting this bound into (20) gives a Gamma approximation to $p(\lambda|\nu)$ as well as the minimal scale factor to make it dominate the true conditional. This gives the following rejection algorithm:

1. Choose λ_0 and compute $q(\lambda_0)$.
2. Draw λ from the Gamma distribution $p(\lambda|\nu) \propto \lambda^{a-1}\exp(-cq(\lambda_0)\lambda)$.
3. Draw a uniform variate $u \sim U(0,1)$ and accept λ if

$$u \leq \frac{Z(\lambda,\nu)^{-c}}{\exp(-cq(\lambda_0)(\lambda-\lambda_0))Z(\lambda_0,\nu)^{-c}} \quad . \tag{31}$$

4. If λ is rejected, repeat from step 2.

Appendix

A Propriety of the conjugate density

Theorem: The conjugate density is proper if and only if

$$b/c > \log(\lfloor a/c \rfloor !) + (a/c - \lfloor a/c \rfloor)\log(\lfloor a/c \rfloor + 1). \tag{A.1}$$

Proof: The conjugate density is proper if and only if the normalizing constant $\kappa^{-1}(a,b,c)$ is finite.

A.1 Proof that (A.1) is a Necessary Condition

A lower bound on $\kappa^{-1}(a,b,c)$ is obtained via an upper bound on $Z(\lambda,\nu)$, which in turn comes from a lower bound on the factorial function. Because the log-gamma function is convex, i.e. has a positive second derivative, we can lower bound the log-factorial with a linear function. In particular, the following bound is valid for all integers j and x:

$$\log j! \geq \log x! + (j-x)\log(x+1). \tag{A.2}$$

The bound is tight at $j = x$ and $j = x+1$. This gives the following bound $\overline{Z}(\lambda,\nu)$ on $Z(\lambda,\nu)$:

$$Z(\lambda,\nu) = \sum_{j=0}^{\infty} \frac{\lambda^j}{j!^\nu} \leq \sum_{j=0}^{\infty} \frac{\lambda^j}{x!^\nu (x+1)^{(j-x)\nu}} = \overline{Z}(\lambda,\nu) \tag{A.3}$$

$$\overline{Z}(\lambda,\nu) = \frac{(x+1)^{x\nu}}{x!^\nu} \sum_{j=0}^{\infty} \left(\frac{\lambda}{(x+1)^\nu}\right)^j \tag{A.4}$$

$$= \begin{cases} \dfrac{(x+1)^{x\nu}}{x!^\nu} \dfrac{1}{1-\lambda(x+1)^{-\nu}} & \text{if } \lambda < (x+1)^\nu \\ \infty & \text{otherwise} \end{cases} . \tag{A.5}$$

From this we obtain a lower bound on the double integral:

$$\kappa^{-1}(a,b,c) \geq \int_0^\infty e^{-b\nu} \int_0^\infty \frac{\lambda^{a-1}}{\overline{Z}(\lambda,\nu)^c} d\lambda d\nu \tag{A.6}$$

$$= \int_0^\infty \frac{e^{-b\nu} x!^{c\nu}}{(x+1)^{cx\nu}} \int_0^{(x+1)^\nu} \lambda^{a-1} \left(1-\lambda(x+1)^{-\nu}\right)^c d\lambda d\nu. \tag{A.7}$$

Change from λ to $\omega = \lambda(x+1)^{-\nu}$:

$$\kappa^{-1}(a,b,c) \geq \int_0^\infty \frac{e^{-b\nu} x!^{c\nu} (x+1)^{a\nu}}{(x+1)^{cx\nu}} d\nu \int_0^1 \omega^{a-1}(1-\omega)^c d\omega. \tag{A.8}$$

The integral over ω is always finite. The integral over ν is finite only if

$$b/c > \log x! + (a/c - x)\log(x+1). \tag{A.9}$$

This condition is necessary for every x, including $x = \lfloor a/c \rfloor$, which gives condition (A.1).

A.2 Proof that (A.1) is a Sufficient Condition

An upper bound on $\kappa^{-1}(a,b,c)$ is obtained by breaking the integral into two parts and bounding each part:

$$\kappa^{-1}(a,b,c) = \int_0^\infty e^{-b\nu} \int_0^1 \frac{\lambda^{a-1}}{Z(\lambda,\nu)^c} d\lambda d\nu + \int_0^\infty e^{-b\nu} \int_1^\infty \frac{\lambda^{a-1}}{Z(\lambda,\nu)^c} d\lambda d\nu \tag{A.10}$$

$$= I_1 + I_2. \tag{A.11}$$

Since $Z(\lambda, \nu) \geq 1$, $I_1 \leq \int_0^\infty e^{-b\nu} \int_0^1 \lambda^{a-1} d\lambda d\nu$, which is finite for all $a > 0$ and $b > 0$. We use a different lower bound on $Z(\lambda, \nu)$ to handle I_2. Because $\log(x)$ is concave, i.e. the second derivative is always negative, we know from Jensen's inequality that

$$\log(\sum_{j=0}^{\infty} q_j a_j) \quad \geq \quad \sum_{j=0}^{\infty} q_j \log a_j \qquad \text{if } \sum_{j=0}^{\infty} q_j = 1. \tag{A.12}$$

Therefore by introducing variables q_j we have

$$\log Z(\lambda, \nu) = \log \sum_{j=0}^{\infty} q_j \frac{\lambda^j}{q_j (j!)^\nu} \geq \sum_{j=0}^{\infty} q_j \log \left(\frac{\lambda^j}{q_j (j!)^\nu} \right) \tag{A.13}$$

$$= \left(\sum_{j=0}^{\infty} j q_j \right) \log \lambda - \nu \left(\sum_{j=0}^{\infty} q_j \log(j!) \right) - \sum_{j=0}^{\infty} q_j \log q_j. \tag{A.14}$$

Let Q be a random variable on the non-negative integers with probability mass function $Pr(Q = j) = q_j$. Then the bound can be written succinctly as

$$\underline{Z}(\lambda, \nu) = \lambda^{E[Q]} e^{-E[\log Q!]\nu} \prod_{j=0}^{\infty} q_j^{q_j}. \tag{A.15}$$

Now we have an upper bound on the double integral:

$$I_2 \quad \leq \quad \int_0^\infty e^{-b\nu} \int_1^\infty \frac{\lambda^{a-1}}{\underline{Z}(\lambda, \nu)^c} d\lambda d\nu \tag{A.16}$$

$$= \quad \int_0^\infty e^{-b\nu} e^{cE[\log(Q!)]\nu} d\nu \int_1^\infty \lambda^{a-1} \lambda^{-cE(Q)} d\lambda \prod_{j=0}^{\infty} q_j^{-cq_j}. \tag{A.17}$$

This integral, and therefore $\kappa^{-1}(a, b, c)$, is finite if

$$E(Q) > a/c \ \text{ and } \ E(\log Q!) < b/c. \tag{A.18}$$

Given (10), we just have to show that there exists a distribution satisfying (A.18). Let

$$q_j \quad = \quad \begin{cases} 1 - (a/c - \lfloor a/c \rfloor + \epsilon) & \text{if } j = \lfloor a/c \rfloor \\ a/c - \lfloor a/c \rfloor + \epsilon & \text{if } j = \lfloor a/c \rfloor + 1 \\ 0 & \text{otherwise.} \end{cases} \tag{A.19}$$

$$\text{Then} \qquad E[Q] \quad = \quad a/c + \epsilon \tag{A.20}$$

$$\text{and } E[\log Q!] \quad = \quad \log \lfloor a/c \rfloor ! + (a/c - \lfloor a/c \rfloor + \epsilon) \log(\lfloor a/c \rfloor + 1). \tag{A.21}$$

This q_j will satisfy (A.18) if we choose $\epsilon > 0$ small enough. Therefore (A.1) is a sufficient condition. $\square$

B Proof of convexity of $\log Z(\lambda, \nu)$ when $\nu < 1$

To show convexity, we show that the second derivative of $\log Z(\lambda, \nu)$ with respect to λ is positive. First we prove that the second derivative is proportional to $\mathrm{cov}(X, (X+1)^{1-\nu})$. Then we prove that this covariance is positive.

$$\begin{aligned}
\frac{\partial^2}{\partial \lambda^2} \log Z(\lambda, \nu) &= \frac{\partial}{\partial \lambda}\left[Z^{-1}(\lambda, \nu) \sum_{x=0}^{\infty} x\lambda^{x-1}/(x!)^{\nu}\right] \\
&= \frac{\partial}{\partial \lambda}\frac{1}{\lambda}E(X) = \frac{\partial}{\partial \lambda}E(X+1)^{1-\nu} \\
&= \frac{\partial}{\partial \lambda}\sum_{x=0}^{\infty}(x+1)^{1-\nu}\lambda^{x}/(x!)^{\nu} Z^{-1}(\lambda, \nu) \\
&= \sum_{x=0}^{\infty}(x+1)^{1-\nu}/x!^{\nu}\left[x\lambda^{x-1}Z^{-1}(\lambda, \nu) - \lambda^{x-1}Z^{-1}(\lambda, \nu)E(X)\right] \\
&= \frac{1}{\lambda}\left[\sum_{x=0}^{\infty} x(x+1)^{1-\nu}\lambda^{x}/x!^{\nu}Z^{-1}(\lambda, \nu) - \right. \\
&\qquad \left. \sum_{x=0}^{\infty}(x+1)^{1-\nu}\lambda^{x}/x!^{\nu}Z^{-1}(\lambda, \nu)\right] \\
&= \frac{1}{\lambda}\left[E(X(X+1)^{1-\nu}) - E(X+1)^{1-\nu}E(X)\right] \\
&= \frac{1}{\lambda}\mathrm{cov}(X, (X+1)^{1-\nu}).
\end{aligned}$$

The less well-known Tchebychef inequality (Hardy, Littlewood, and Polya (1934, 1952, p. 43)) says that the covariance between two increasing functions of X must be positive. The function $(X+1)^{1-\nu}$ is increasing when $\nu < 1$; therefore, the covariance is positive and the proof is complete.

References

Barnard, J., McCulloch, R., and Xiao-Li, M. (2000). "Modelling covariance Matrices in Terms of Standard Deviations and Correlations with Application to Shrinkage." *Statistica Sinica*, 10(4): 1281–1311. 192

Boatwright, P., Borle, S., and Kadane, J. B. (2003). "A Model of the Joint Distribution of Purchase Quantity and Timing." *Journal of the American Statistical Association*, 98(463): 564–572. 189

Breslow, N. (1990). "Tests of Hypotheses in Overdispersed Poisson Regression and Other Quasi-Likelihood Models." *Journal of the American Statistical Association*, 85: 565–571. 189

Catchside, D. G., Lea, D. E., and Thoday, J. M. (1945, 1946). "Types of chromosome structural change induced by the irradiation of Tradescantia Microspore." *Journal of Genetics*, 47: 113–136. 194

Conway, R. W. and Maxwell, W. L. (1961). "A queuing model with state dependent service rates." *Journal of Industrial Engineering*, 12: 132–136. 189

Dean, C. B. (1992). "Testing for Overdispersion in Poisson and Binomial Regression Models." *Journal of the American Statistical Association*, 87: 451–457. 189

Feller, W. (1950, 1957). *An Introduction to Probability Theory and Its Applications*. New York: J. Wiley & Sons, 2nd edition. 194

Flannery, B. P., Press, W. H., Teukolsky, S. A., and Vetterling, W. T. (1994). *Numerical Recipes in C*. Cambridge University Press, 2nd edition. 193

Garthwaite, P. H., Kadane, J. B., and O'Hagan, A. (2005). "Statistical Methods for Eliciting Probability Distributions." *Journal of the American Statistical Association*, 100: 680–700. 192

Hardy, G., Littlewood, J. E., and Polya, G. (1934, 1952). *Inequalities*. Cambridge: Cambridge University Press, 2nd edition. 199

Kadane, J. B. and Wolfson, L. J. (1997). "Experiences in Elicitation." *The Statistician*, 47: 3–19. 192

Mood, A. M., Graybill, F. A., and Boes, D. C. (1974). *Introduction to the Theory of Statistics*. Kogakusha: McGraw Hill, 3rd edition. 195

Shmueli, G., Minka, T. P., Kadane, J. B., Borle, S., and Boatwright, P. (2005). "A Useful Distribution for Fitting Discrete Data: Revival of the COM-Poisson." *JRSS*, C(54): 127–142. 189

Wimmer, G. and Altmann, G. (1996). "The Theory of Word Length: Some Results and Generalizations." *Glottometrika*, 15: 112–133. 189

Wimmer, G., Kohler, R., Grotjahn, R., and Altmann, G. (1994). "Toward a Theory of Word Length Distributions." *Journal of Quantitative Linguistics*, 1: 98–106. 189

Acknowledgments

The first author's research is supported by National Science Foundation grants DMS-9801401 and DMS-0139911.