Supplement of "The Empirical Bayes Estimators of the Parameter of the Poisson Distribution with a Conjugate Gamma Prior under Stein's Loss Function"

Abstract

This is the supplemental file of the paper. The supplement is written for the authors, the reviewers, and the readers to check the correctness of the derivations and to replicate the results.

1. The proof of Theorem 1

Suppose that we observe $X_1, X_2, \ldots, X_n$ from the hierarchical Poisson and gamma model:

$$\begin{cases} X_i|\theta \overset{\text{iid}}{\sim} P(\theta), \ i = 1, 2, \ldots, n, \\ \theta \sim G(\alpha, \beta), \end{cases} \tag{1}$$

where $\alpha > 0$ and $\beta > 0$ are hyperparameters to be determined, $P(\theta)$ is the Poisson distribution with an unknown mean $\theta > 0$, and $G(\alpha, \beta)$ is the gamma distribution with an unknown shape parameter α and an unknown rate parameter β.

By the Bayes Theorem, the posterior distribution of θ is

$$\pi(\theta|\boldsymbol{x}) \propto f(\boldsymbol{x}|\theta)\,\pi(\theta).$$

It is easy to see that

$$\begin{aligned} f(\boldsymbol{x}|\theta) &= \prod_{i=1}^{n} f(x_i|\theta) = \prod_{i=1}^{n} \frac{\theta^{x_i}}{x_i!} \exp(-\theta) \\ &= \frac{\theta^{\sum_{i=1}^{n} x_i}}{\prod\limits_{i=1}^{n} x_i!} \exp(-n\theta) \\ &\propto \theta^{\sum_{i=1}^{n} x_i} \exp(-n\theta) \end{aligned}$$

and

$$\pi(\theta) = \frac{\beta^{\alpha}}{\Gamma(\alpha)}\theta^{\alpha-1}\exp(-\theta\beta)$$
$$\propto \theta^{\alpha-1}\exp(-\theta\beta).$$

Therefore,

$$\begin{aligned}\pi(\theta|\boldsymbol{x}) &\propto \theta^{\sum_{i=1}^{n} x_i}\exp(-n\theta)\,\theta^{\alpha-1}\exp(-\theta\beta)\\ &= \theta^{\left(\alpha+\sum_{i=1}^{n} x_i\right)-1}\exp\left[-\theta(\beta+n)\right]\\ &\sim G(\alpha^*, \beta^*),\end{aligned}$$

where

$$\alpha^* = \alpha + \sum_{i=1}^{n} x_i \text{ and } \beta^* = \beta + n.$$

Now let us calculate the marginal probability mass function (pmf) of x. We have, for $x = 0, 1, 2, \ldots,\ \alpha > 0,\ \beta > 0$,

$$\begin{aligned}m(x|\alpha,\beta) &= \int_0^{\infty} f(x|\theta)\,\pi(\theta)\,d\theta\\ &= \int_0^{\infty}\frac{\theta^x}{x!}\exp(-\theta)\frac{\beta^{\alpha}}{\Gamma(\alpha)}\theta^{\alpha-1}\exp(-\theta\beta)\,d\theta\\ &= \frac{\beta^{\alpha}}{x!\Gamma(\alpha)}\int_0^{\infty}\theta^{(x+\alpha)-1}\exp\left[-\theta(1+\beta)\right]d\theta\\ &= \frac{\beta^{\alpha}}{x!\Gamma(\alpha)}\frac{\Gamma(x+\alpha)}{(1+\beta)^{x+\alpha}}\\ &= \frac{\Gamma(x+\alpha)}{\Gamma(x+1)\Gamma(\alpha)}\frac{\beta^{\alpha}}{(1+\beta)^{x+\alpha}}. \qquad (2)\end{aligned}$$

In particular, when α is a positive integer, the marginal distribution of x is

$$\begin{aligned}m(x|\alpha,\beta) &= \frac{\Gamma(\alpha+x)}{\Gamma(x+1)\Gamma(\alpha)}\left(\frac{\beta}{1+\beta}\right)^{\alpha}\left(\frac{1}{1+\beta}\right)^{x}\\ &= \frac{\Gamma(r+x)}{\Gamma(x+1)\Gamma(r)}p^r q^x\\ &= \frac{(r+x-1)!}{x!\,(r-1)!}p^r q^x\\ &= \binom{r+x-1}{x}p^r q^x\\ &\sim NB(r,p),\end{aligned}$$

which is a negative binomial distribution, where

$$r = \alpha,\ p = \frac{\beta}{1+\beta},\ \text{and}\ q = 1 - p = \frac{1}{1+\beta}.$$

The proof of the theorem is complete. $\square$

2. The proof of Theorem 2

The hyperparameters of the model (1) are $\alpha > 0$ and $\beta > 0$. To obtain the moment estimators of the hyperparameters of the model (1), we need to calculate the first two moments of X, $\mathrm{E}X$ and $\mathrm{E}X^2$. It is easy to show that

$$\mathrm{E}\theta = \frac{\alpha}{\beta} \text{ and } \mathrm{Var}\,(\theta) = \frac{\alpha}{\beta^2}.$$

Therefore,

$$\mathrm{E}X = \mathrm{E}\left[\mathrm{E}\left(X|\theta\right)\right] = \mathrm{E}\left[\theta\right] = \frac{\alpha}{\beta}$$

and

$$\begin{aligned}
\mathrm{E}X^2 &= \mathrm{E}\left[\mathrm{E}\left(X^2|\theta\right)\right] \\
&= \mathrm{E}\left\{\mathrm{Var}\left(X|\theta\right) + \left[\mathrm{E}\left(X|\theta\right)\right]^2\right\} \\
&= \mathrm{E}\left[\theta + \theta^2\right] \\
&= \mathrm{E}\theta + \mathrm{Var}\,(\theta) + \left(\mathrm{E}\theta\right)^2 \\
&= \frac{\alpha}{\beta} + \frac{\alpha}{\beta^2} + \left(\frac{\alpha}{\beta}\right)^2 \\
&= \frac{\alpha\beta + \alpha + \alpha^2}{\beta^2} \\
&= \frac{\alpha\left(\alpha + \beta + 1\right)}{\beta^2}.
\end{aligned}$$

Furthermore, letting the population moments be equal to the sample moments, we obtain

$$\mathrm{E}X = \frac{\alpha}{\beta} = A_1, \tag{3}$$

$$\mathrm{E}X^2 = \frac{\alpha\left(\alpha + \beta + 1\right)}{\beta^2} = A_2, \tag{4}$$

where

$$A_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k,\ k = 1, 2,$$

is the sample kth moment of X. Substituting (3) into (4), we obtain

$$A_1 \frac{\alpha+\beta+1}{\beta} = A_2$$

$$\Leftrightarrow\ \frac{\alpha}{\beta} + 1 + \frac{1}{\beta} = \frac{A_2}{A_1}. \tag{5}$$

Substituting (3) into (5) and simplifying, we have

$$\beta = \frac{A_1}{A_2 - A_1 - A_1^2} = \beta_1(n). \tag{6}$$

From (3) and (6), we can solve

$$\alpha = \frac{A_1^2}{A_2 - A_1 - A_1^2} = \alpha_1(n). \tag{7}$$

Consequently, the moment estimators of the hyperparameters of the model (1) are given by (7) and (6).

The proof of the theorem is complete. □

3. The proof of Theorem 3

Now we derive the MLEs of α and β. The hyperparameters of the model are $\alpha > 0$ and $\beta > 0$. By Theorem 1, we know that the marginal density of X of the model (1) is

$$m(x|\alpha,\beta) = \frac{\Gamma(x+\alpha)}{\Gamma(x+1)\Gamma(\alpha)} \frac{\beta^{\alpha}}{(1+\beta)^{x+\alpha}},\ x = 0, 1, 2, \ldots,\ \alpha > 0,\ \beta > 0.$$

Then the likelihood function of α and β is

$$L(\alpha,\beta|\boldsymbol{x}) = m(\boldsymbol{x}|\alpha,\beta) = \prod_{i=1}^{n} m(x_i|\alpha,\beta) = \prod_{i=1}^{n} \frac{\beta^{\alpha}\Gamma(x_i+\alpha)}{x_i!\Gamma(\alpha)(1+\beta)^{x_i+\alpha}}.$$

Consequently, the log-likelihood function of α and β is

$$\begin{aligned}
&\log L(\alpha,\beta|\boldsymbol{x}) \\
&= \sum_{i=1}^{n} \left[\alpha\log\beta + \log\Gamma(x_i+\alpha) - \log(x_i!) - \log\Gamma(\alpha) - (x_i+\alpha)\log(1+\beta)\right] \\
&= n\alpha\log\beta + \sum_{i=1}^{n}\log\Gamma(x_i+\alpha) - \sum_{i=1}^{n}\log(x_i!) - n\log\Gamma(\alpha) - \left(\sum_{i=1}^{n} x_i + n\alpha\right)\log(1+\beta).
\end{aligned}$$

Taking partial derivatives with respect to α and β and setting them to zeros, we obtain

$$
\begin{aligned}
\frac{\partial}{\partial \alpha} \log L &= n \log \beta + \sum_{i=1}^{n} \frac{\Gamma'(x_i+\alpha)}{\Gamma(x_i+\alpha)} - n \frac{\Gamma'(\alpha)}{\Gamma(\alpha)} - n \log(1+\beta) = 0, \\
\frac{\partial}{\partial \beta} \log L &= n\alpha \frac{1}{\beta} - \left(\sum_{i=1}^{n} x_i + n\alpha \right) \frac{1}{1+\beta} = 0.
\end{aligned}
$$

Since

$$\psi(x) = \frac{\Gamma'(x)}{\Gamma(x)} = \frac{d}{dx} \log \Gamma(x) = \text{digamma}(x)$$

which can be directly calculated in R software by digamma(x) ([1]), after some algebra, the above equations reduce to

$$f_1(\alpha, \beta) = \log \beta + \frac{1}{n} \sum_{i=1}^{n} \psi(x_i+\alpha) - \psi(\alpha) - \log(1+\beta) = 0, \tag{8}$$

$$f_2(\alpha, \beta) = \frac{\alpha}{\beta} - \frac{\bar{x}+\alpha}{1+\beta} = 0. \tag{9}$$

The MLEs of the hyperparameters α and β are the solutions to the equations (8) and (9). The analytical calculations of the MLEs of α and β by solving the equations (8) and (9) are impossible, and thus we have to resort to numerical solutions. We can exploit Newton's method to solve the equations (8) and (9) and to obtain the MLEs of α and β. The iterative scheme of Newton's method is

$$\boldsymbol{p}^{(k+1)} = \boldsymbol{p}^{(k)} - \left[\boldsymbol{J}\left(\boldsymbol{p}^{(k)} \right) \right]^{-1} \boldsymbol{f}\left(\boldsymbol{p}^{(k)} \right), \ k = 0, 1, \ldots,$$

where $\boldsymbol{J}(\boldsymbol{p})$ is the Jacobian matrix of $\boldsymbol{f}(\boldsymbol{p}) = (f_1(\boldsymbol{p}), f_2(\boldsymbol{p}))'$ and $\boldsymbol{p} = (\alpha, \beta)'$. Note that the MLEs of α and β are very sensitive to the initial estimators, and the moment estimators are usually proved to be good initial estimators. The Jacobian matrix of α and β is given by

$$J = \boldsymbol{J}(\boldsymbol{p}) = \begin{pmatrix} J_{11} & J_{12} \\ J_{21} & J_{22} \end{pmatrix} = \begin{pmatrix} \frac{\partial f_1}{\partial \alpha} & \frac{\partial f_1}{\partial \beta} \\ \frac{\partial f_2}{\partial \alpha} & \frac{\partial f_2}{\partial \beta} \end{pmatrix},$$

where

$$
\begin{aligned}
J_{11} &= \frac{\partial f_1}{\partial \alpha} = \frac{1}{n}\sum_{i=1}^{n} \psi'\left(x_i+\alpha\right) - \psi'\left(\alpha\right), \\
J_{12} &= \frac{\partial f_1}{\partial \beta} = \frac{1}{\beta} - \frac{1}{1+\beta} = \frac{1}{\beta\left(1+\beta\right)}, \\
J_{21} &= \frac{\partial f_2}{\partial \alpha} = \frac{1}{\beta} - \frac{1}{1+\beta} = \frac{1}{\beta\left(1+\beta\right)} = J_{12}, \\
J_{22} &= \frac{\partial f_2}{\partial \beta} = \alpha\left(-\frac{1}{\beta^2}\right) - \left(\bar{x}+\alpha\right)\left[-\frac{1}{\left(1+\beta\right)^2}\right] \\
&= -\frac{\alpha}{\beta^2} + \frac{\bar{x}+\alpha}{\left(1+\beta\right)^2}.
\end{aligned}
$$

Note that

$$
\psi'\left(x\right) = \frac{d^2}{dx^2}\log\Gamma\left(x\right) = \mathrm{trigamma}\left(x\right)
$$

which can be directly calculated in R software by trigamma(x) ([1]).

The proof of the theorem is complete. $\square$

4. The two cases of the goodness-of-fit of the model to the simulated data

Let us consider the hierarchical Poisson and gamma model (1).

Case 1. The hyperparameters α and β are assumed known.

In this case, the hyperparameters α and β are assumed known. For example, $\alpha = 1$ and $\beta = 2$. Let the null hypothesis be

$$
H_0: \ X \sim P - G\left(\alpha, \beta\right),
$$

where $P - G\left(\alpha, \beta\right)$ is the marginal distribution of the hierarchical Poisson and gamma model (1) with the marginal pmf $m\left(x|\alpha, \beta\right)$ given by (2).

The chi-square goodness-of-fit is performed as follows. We first divide the domain of X, $\mathcal{X} = \{0, 1, 2, \ldots, \}$, into m groups:

$$
I_1 = \{0\}, \ I_2 = \{1\}, \ \cdots, \ I_{m-1} = \{m-2\}, \ I_m = \{k : k \geq m-1\}.
$$

Let the theoretical probabilities under H_0 on these subintervals be

$$
p_1, p_2, \ldots, p_{m-1}, p_m,
$$

where

$$\begin{aligned}
p_i &= P_{H_0}\left(X \in I_i\right),\ i = 1, \ldots, m-1, \\
&= P_{H_0}\left(X = i-1\right) \\
&= \frac{\Gamma\left(i-1+\alpha\right)}{\Gamma\left(i\right)\Gamma\left(\alpha\right)} \frac{\beta^{\alpha}}{\left(1+\beta\right)^{i-1+\alpha}},
\end{aligned}$$

$$p_m = 1 - \sum_{i=1}^{m-1} p_i,$$

and P_{H_0} is the probability when X is distributed under H_0. Let $n_i,\ i = 1, \ldots, m$, denote the number of $X_1, X_2, \ldots, X_n$ that lie in the ith subinterval I_i. Then the chi-square statistics ([2, 3])

$$K = \sum_{i=1}^{m} \frac{\left(n_i - np_i\right)^2}{np_i} \xrightarrow{d} \chi^2\left(m-1\right), \text{ as } n \to \infty,$$

where $\xrightarrow{d}$ is convergence in distribution. Moreover, we can compute the p-value which gives the probability that a value of K as large as the one observed would have occurred if the null hypothesis were true. Hence,

$$\begin{aligned}
\text{p-value} &= P\left(\chi^2\left(m-1\right) > K\right) \\
&= 1 - P\left(\chi^2\left(m-1\right) \le K\right) \\
&= \text{1 - pchisq(K, df = m-1)},
\end{aligned}$$

where pchisq(), which calculates the cdf of a chi-square random variable, is an R built-in function ([1]). Note that a large p-value (> 0.05 in the usual case) indicates that the model specified by H_0 fits the (simulated) data well, while a small p-value (≤ 0.05 in the usual case) indicates that the model specified by H_0 does not fit the (simulated) data well. The larger the p-value, the better the model specified by H_0 fits the (simulated) data.

Case 2. The hyperparameters α and β are unknown.

Let the null hypothesis be

$$H_0 :\ X \sim P - G\left(\alpha, \beta\right),$$

where α and β are unknown. First, the hyperparameters α and β need to be estimated by the sample. The estimators could be the moment estimators or the Maximum Likelihood Estimators (MLEs). Let I_i and n_i be given in Case 1. The theoretical probabilities under H_0 on the subintervals are calculated by

$$\hat{p}_i = P_{H_0}\left(X \in I_i | \alpha = \hat{\alpha}(n), \beta = \hat{\beta}(n)\right),\ i = 1, \ldots, m,$$

that is, the unknown hyperparameters α and β are estimated by their estimators $\hat{\alpha}(n)$ and $\hat{\beta}(n)$ based on the sample. Then the chi-square statistics ([2, 3])

$$\hat{K} = \sum_{i=1}^{m} \frac{(n_i - n\hat{p}_i)^2}{n\hat{p}_i} \xrightarrow{d} \chi^2(m-1-2), \text{ as } n \to \infty.$$

Note that the degree of freedom now is lost by 2, since two unknown parameters are estimated by the sample. Moreover, the p-value is given by

$$\begin{aligned} \text{p-value} &= P\left(\chi^2(m-3) > \hat{K}\right) \\ &= 1 - P\left(\chi^2(m-3) \leq \hat{K}\right) \\ &= 1 \text{ - pchisq}(\hat{K}, \text{ df = m-3}). \end{aligned}$$

The derivations are complete. □

References

[1] R Core Team. *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria 2018. URL `http://www.R-project.org/`.

[2] Xue Y, Chen LP. *Statistical Modeling and R Software.* Tsinghua University Press: Beijing, 2007.

[3] Ross S. *Simulation.* Mechanical Industry Press: Beijing, 2013. 5th edition.