Estimating the mixing distribution of a mixed Poisson distribution

Matheus Silva[a]

[a]New York University. Email: matheus@nyu.edu

ARTICLE HISTORY

Compiled November 15, 2022

ABSTRACT

I propose a computationally simple estimator for the mixing distribution of a mixed Poisson distribution. Despite being asymptotically equivalent to the estimators studied in Talenti (1987) Hengartner (1997), Monte Carlo simulations show this new estimator has considerably better finite sample properties.

1. Introduction

Say an absolutely continuous distribution $f(\cdot)$ on the interval $[0,1]$ is the mixing distribution of a random variable X that follows a mixed Poisson process,

$$\mathbb{P}(X = x) = \int_0^\infty p_x(\theta) f(\theta) d\theta \tag{1}$$

Where $p_x(\theta) \equiv \mathbb{P}(X = x|\theta) = exp(-\theta)\theta^x/x!$ is the probability mass function of a Poisson distribution. In this case, I say X follows a mixed Poisson with mixing distribution f, denoted $X \sim MP(f)$.

Estimating the distribution f is a well-studied problem. Hengartner (1997) proposes and studies a series estimator for this problem and shows its best minmax convergence rate is $log(N)/log(log(N))$, where N is the sample size. Even though series estimators are well understood theoretically, their implementation may yield less than satisfactory numerical results due to their slow asymptotic convergence rates. In practice, this may leave the researcher with estimates for the mixing distribution that do not respect probability axioms (for example, they estimate points where the p.d.f. is negative).

In this paper, I propose a generalized method of moments approach to this problem that restricts the estimated mixing distribution f to be, in fact, a probability density function. This estimator is asymptotically equivalent to Hengartner's (1997), but Monte Carlo simulations show it has better finite-sample properties.

2. The estimator

The goal of the researcher is to estimate the probability density function f having access to an i.i.d. sample of size N, $\{X_1, X_2, \ldots, X_N\}$, generated according to a $MP(f)$. The probability mass function (1) implies that the m^{th} noncentered moment of X can be obtained with the recursion,

$$\mathbb{E}(X^m) = \sum_{j=1}^{r} S(r,j)\mathbb{E}(\theta^j),\ \forall r \geq 1 \tag{2}$$

Where $S(r,j)$ is the (r,j) Stirling number of the second kind. One can also solve this recursion to find that:

$$\mathbb{E}(\theta^m) = \mathbb{E}[X(X-1)\ldots(X-m+1)],\ \forall m \geq 1 \tag{3}$$

The researcher can then estimate the moments of θ directly with the sample analogue of (3),

$$\overline{(\theta^m)} = \frac{\sum_{i=1}^{N} X_i(X_i-1)\ldots(X_i-m+1)}{N}$$

The procedure I propose finds the polynomial that best approximates f when minimizing the distance between the estimated moments of θ (as calculated above) under the restriction that the resulting estimate $\hat{f}$ is indeed a probability distribution function.

Define:

$$d_m(f) \equiv \int_0^1 \theta^m f(\theta)d\theta - \overline{(\theta^m)}$$

The estimator of the mixing density is the solution of the following problem:

$$\hat{f} = arg\min_{f} \begin{pmatrix} d_1(f) & d_2(f) & \ldots & d_M(f) \end{pmatrix}' \Sigma^{-1} \begin{pmatrix} d_1(f) & d_2(f) & \ldots & d_M(f) \end{pmatrix} \tag{4}$$

$$s.t. \begin{cases} f(\theta) \geq 0 \\ \int_0^1 f(\theta)d(\theta) = 1 \end{cases} \tag{5}$$

Where M is the number of moments included in the estimation and Σ is a weighting matrix. In this general form, solving this problem is computationally infeasible because one would need to investigate all possible functions f that are probability distribution functions and pick the best among them (according to this criterion). Instead, I approximate f with a $J-th$ degree polynomial, $f_J^P(\theta)^1$,

$$f_J^P(\theta) \equiv \delta_0 + \delta_1\theta + \cdots + \delta_J\theta^J$$

This approximation makes $d_m(f_K^P)$ easy to compute since,

$$\int_0^1 \theta^m f_J^P(\theta)d\theta = \int_0^1 \theta^m \left[\delta_0 + \delta_1\theta + \cdots + \delta_J\theta^J\right] d\theta = \sum_{j=0}^{J} \frac{\delta_j}{m+j+1}$$

Assumption 2.1. The mixing distribution has an absolutely continuous probability density function in the interval $[0, 1]$.

This approximation allows the researcher to find an approximate solution to the problem posed in (4) by finding the vector $\hat{\delta} \equiv (\hat{\delta}_0, \hat{\delta}_1, \ldots, \hat{\delta}_J)$ that minimizes the loss function in (4). For the weighting matrix, I pick the diagonal matrix with the inverse of the variance of each moment estimator. I calculate it by bootstrapping the dataset, computing the moments used in the estimation, and calculating their variance. Finally, I approximate the restriction that $f_J^P(\theta;\delta) \geq 0$ by checking that it is indeed positive in a fine grid between 0 and 1.

With these steps, the feasible optimization problem is:

$$\hat{\delta} = arg\min_{\delta} \begin{pmatrix} d_1(f_J^P) & d_2(f_J^P) & \ldots & d_M(f_J^P) \end{pmatrix}' \hat{\Sigma}^{-1} \begin{pmatrix} d_1(f_J^P) & d_2(f_J^P) & \ldots & d_M(f_J^P) \end{pmatrix} \tag{6}$$

subject to the following restrictions:

$$\begin{cases} f_J^P(\vartheta;\delta) \geq 0, \ \vartheta \in \Theta \\ \int_0^1 f_J^P(\theta;\delta)d(\theta) = 1 \end{cases}$$

Where Θ is a fine grid on which I verify that the estimated polynomial is positive. Notice that this set of restrictions allow vectors δ that generate probability density functions to be candidates. This optimization is easily solvable with off-the-shelf software[2].

[1]Even though I use polynomials of this form, Legendre or Chebyshev polynomials may be used as well.

[2]Please find MATLAB code available on the website or contact the author.

3. Monte Carlo simulations

To show the performance of this GMM estimator in finite samples, this section provide simulations for the case where the mixing distribution is a Beta distribution. This distribution is flexible in terms of its shape and its support is $[0, 1]$, making it unlikely to observe high values of X. In the following simulations, I consider a sample of size $50,000$ and pick the number of moments and the degree of the polynomial as in Hengartner (1997), that is $M = J = \lfloor log(N)/log(log(N)) \rfloor$. I repeat the experiment $50,000$ times. $\hat{\Sigma}^{-1}$ is the inverse of the covariance of the moments, obtained by bootstraping the researcher's dataset. The figures also present the point-wise 90% confidence intervals.

3.1. *Setup #1*

$$X \sim MP(\beta(1, 3))$$

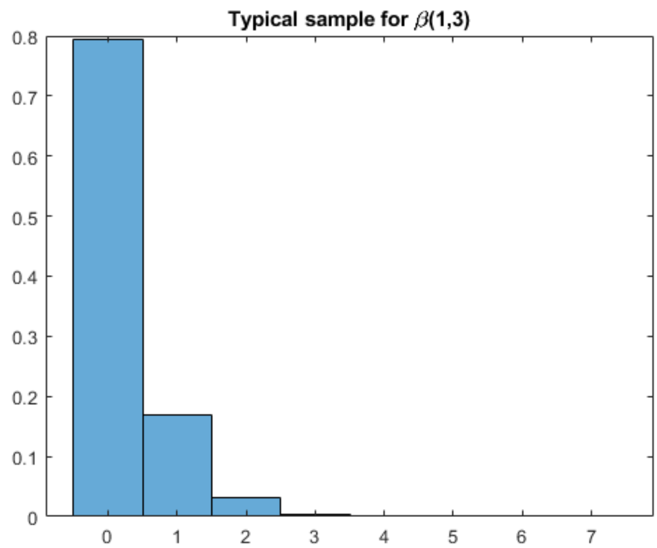

Figure 1. Histogram of a typical sample observed by the researcher, $N = 50,000$.

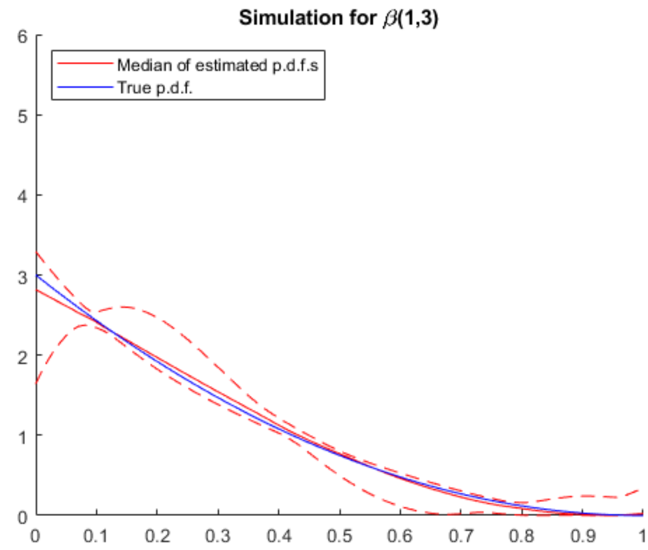

Figure 2. Estimates of the pdf of θ.

3.2. *Setup #2*

$$X \sim MP(\beta(3,3))$$

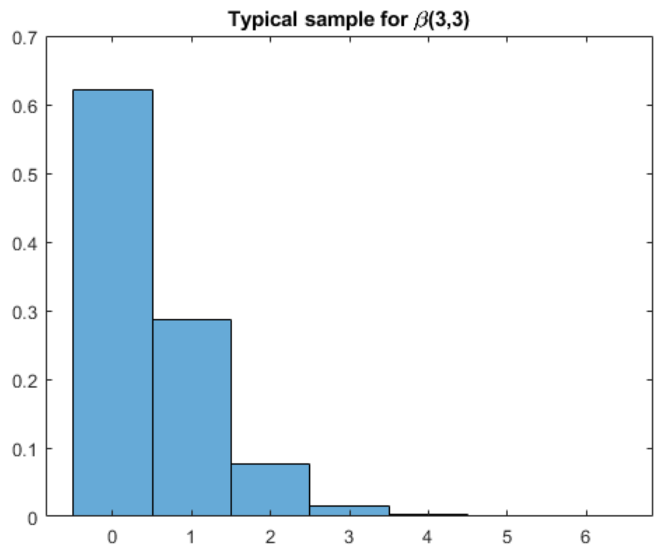

Figure 3. Histogram of a typical sample observed by the researcher, $N = 50,000$.

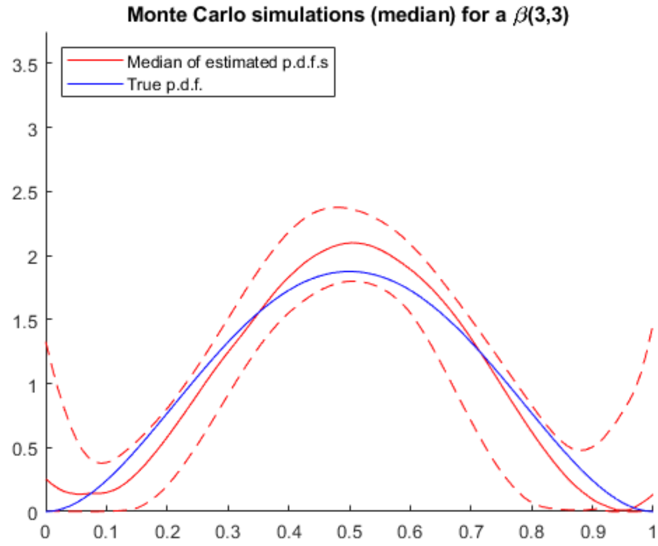

Figure 4. Estimates of the pdf of θ.

3.3. *Setup #3*

$$X \sim MP(\beta(3,1))$$

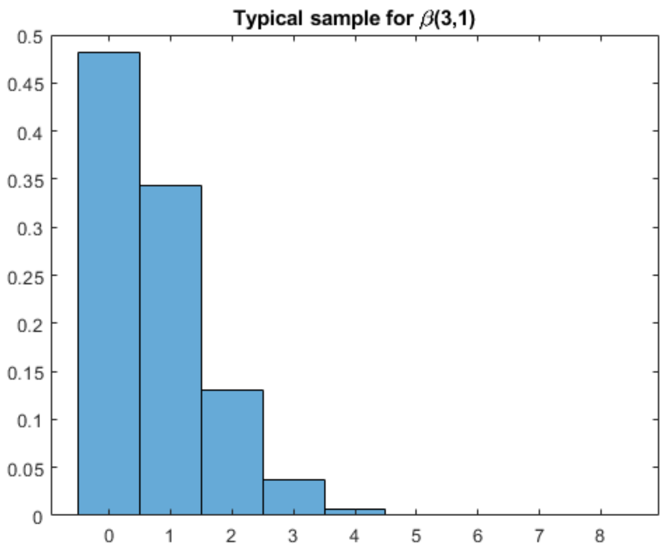

Figure 5. Histogram of a typical sample observed by the researcher, $N = 50,000$.

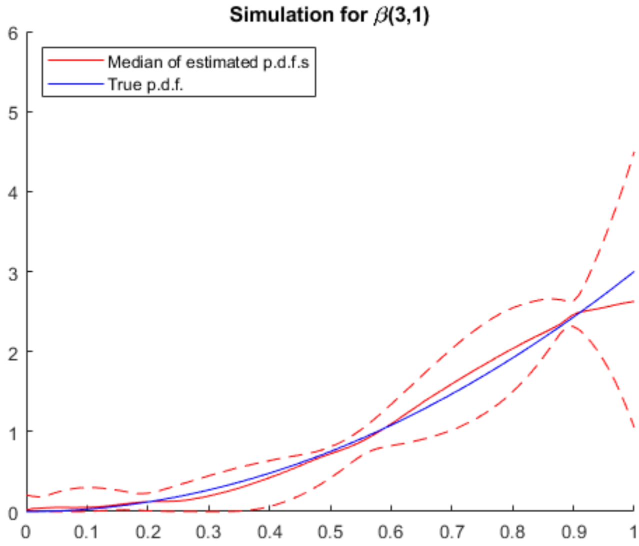

Figure 6. Estimates of the pdf of θ.

3.4. *Comparison to Hengartner's (1997) estimator*

The Talenti-Hengartner estimator is a special case of this optimization problem. If one drops the restriction that $f(\theta)$ has to be positive, the problem boils down to finding the polynomial that makes $d_j(f_J^P) = 0,\ j \in \{1, \ldots, J\}$. By setting $M = J = \lfloor log(N)/log(log(N)) \rfloor$[3], the solution yields the estimator studied in Hengartner (1997).

Figure 7 shows typical realizations of both estimators in the Talenti-Hengartner case where $N = 1,000$, $M = J = \lfloor log(1,000)/log(log(1,000)) \rfloor = 3$. This figure provides evidence that the restricted GMM estimator greatly outperforms the Talenti-Hengartner estimator in finite samples. This problem persists even when the sample size is larger. Figure 8 shows the typical realizations of both estimators when $N = 200,000$ ($M = J = 4$). The larger sample size shows that the restricted GMM estimator approximates the true beta mixing distribution well.

Despite its good finite-sample behavior, there is no asymptotic gain from using the restricted GMM estimator[4] if the researcher sets $M = J = \lfloor log(N)/log(log(N)) \rfloor$. Since the Talenti-Hengartner estimator is consistent, it will eventually estimate positive functions with high probability. When this happens, the positivity constraint is not binding and both problems are equivalent as described above.

Figures 9, 10, and 11 show the median realizations of both estimators for three different beta distributions when the for $N = 10,000$. The Talenti-Hengartner estimator still yields very poor estimates near the boundaries.

[3] Where $\lfloor x \rfloor$ is the floor of x, the largest integer number smaller than x.
[4] As expected from Hengartner's (1997) proof.

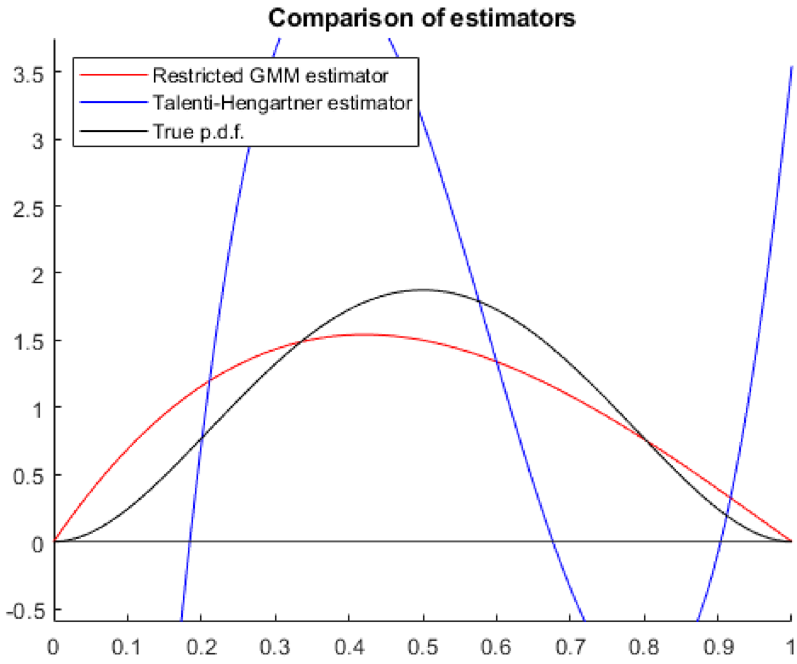

Figure 7. Typical realizations of the estimators with $N = 1,000$

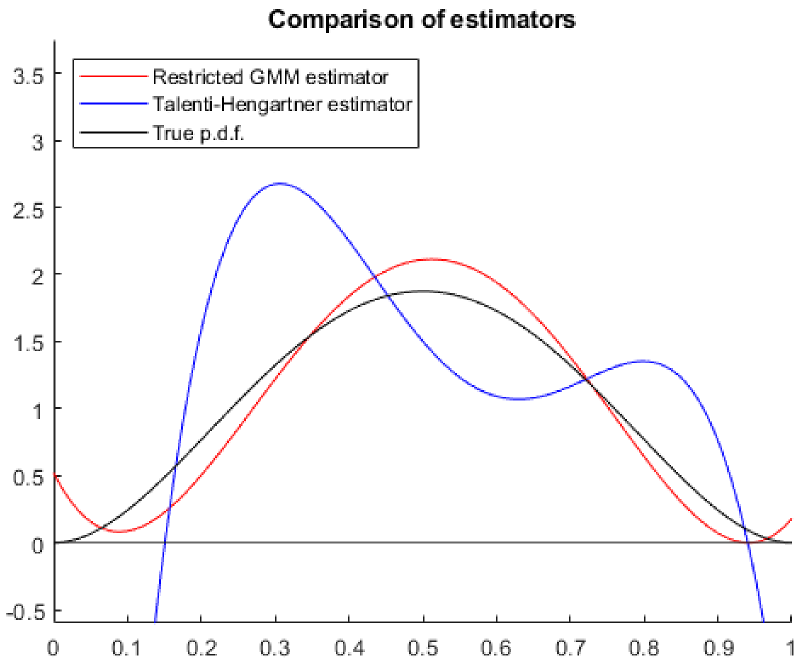

Figure 8. Typical realizations of the estimators with $N = 200,000$

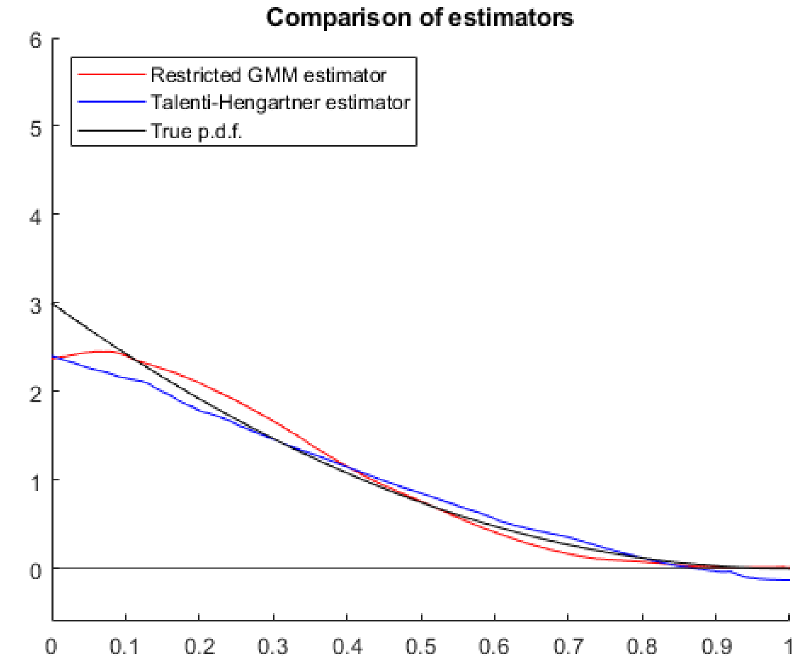

Figure 9. Median of the realizations of the estimators with $N = 10,000$ and $X \sim MP(\beta(1,3))$.

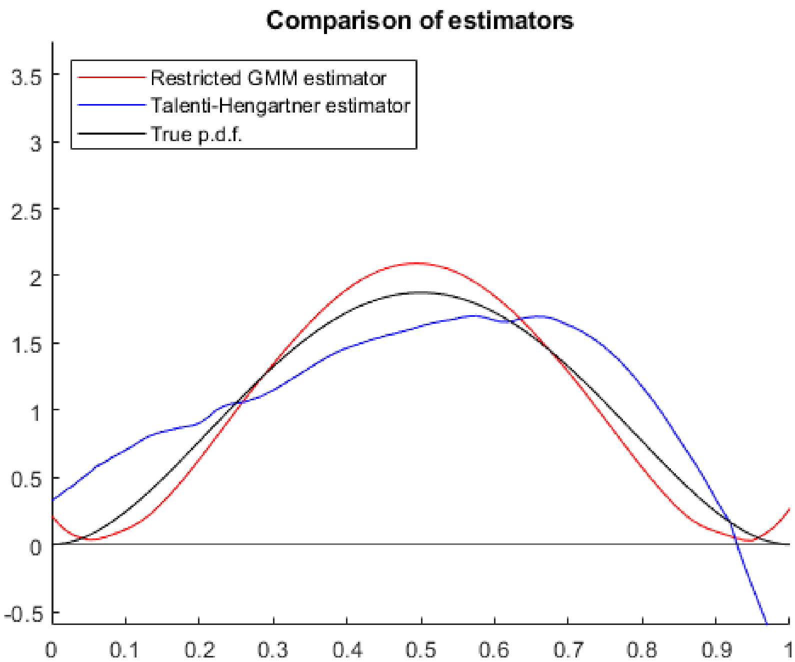

Figure 10. Median of the realizations of the estimators with $N = 10,000$ and $X \sim MP(\beta(3,3))$.

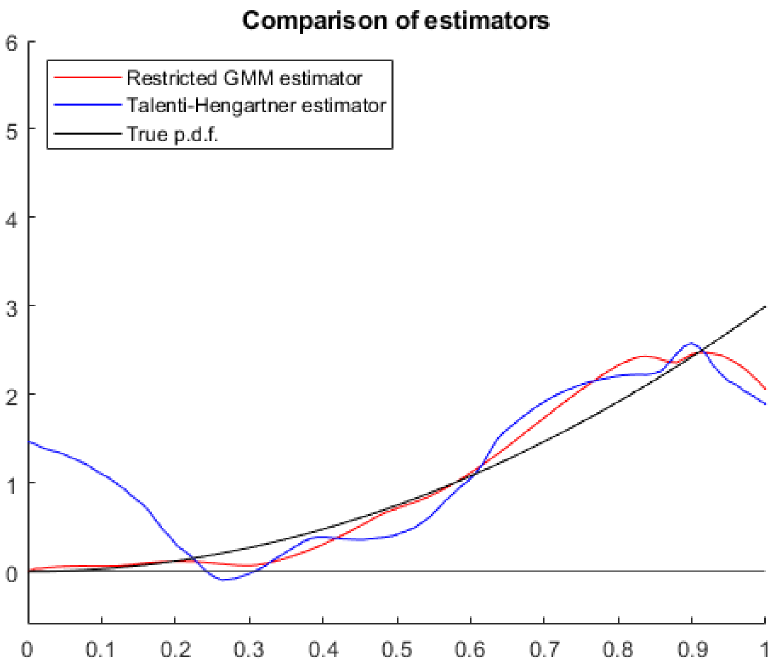

Figure 11. Median of the realizations of the estimators with $N = 10,000$ and $X \sim MP(\beta(3,1))$.

3.5. *Degree of the polynomial approximation*

The constrained GMM estimator allows some flexibility in choosing the degree of the polynomial approximation used in the estimation. This is not possible in the Hengartner's (1997) approach. The next figures compare different number of moments and different polynomial degrees for typical realizations. Unsurprisingly, setting $J = M$ generates better results.

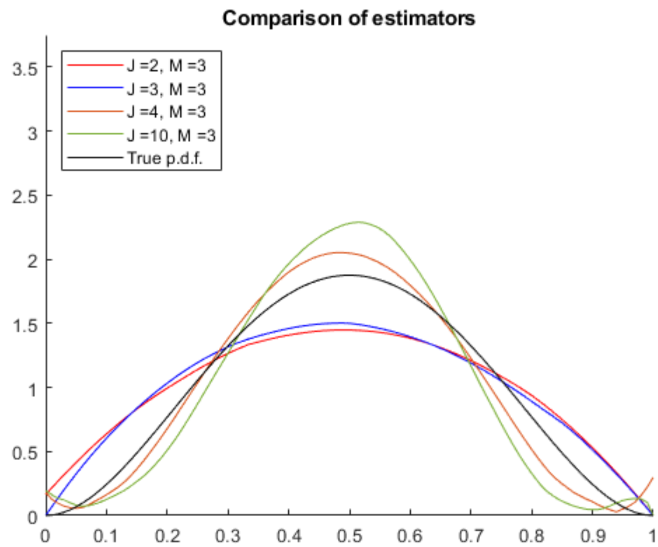

Figure 12. Comparison of the median of the GMM estimator with different polynomials ($N = 2,000$ and $10,000$ repetitions).

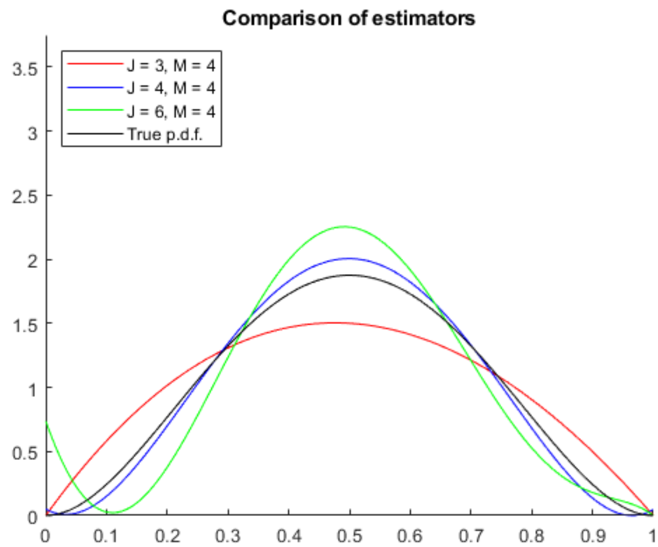

Figure 13. Comparison between the estimators with $N = 200,000$

4. Conclusion

In this paper, I present a generalized method of moments estimator for the mixing distribution of a mixed Poisson process. Even though this estimator is asymptotically equivalent to that studied in Talenti (1987) and Hengartner (1997), Monte Carlo experiments prove it has better finite sample properties. This is achieved by taking into account that the estimate must be a probability density function, which is not imposed in previous work.

References

[1] Nicolas W. Hengartner. Adaptive demixing in poisson mixture models. *The Annals of Statistics*, 25(1):917–928, 1997.

[2] G Talenti. Recovering a function from a finite number of moments. *Inverse Problems*, 3(3):501–517, 1987.