**Far East Journal of Theoretical Statistics**
Published Online: October 2013
Available online at http://pphmj.com/journals/fjts.htm
Volume 44, Number 2, 2013, Pages 145-164

# ESTIMATORS COMPARISON OF THE POISSON CANONICAL PARAMETER AND MEAN DUAL DISTRIBUTIONS

**Gélin Chedly Louzayadio, Dominique Mizere and Roger Armand Makany**

Faculty of Sciences and Technics
Marien Ngouabi University
BP. 69, Brazzaville
Congo

## Abstract

The Poisson distribution and mean dual distributions have the same canonical parameter θ. We wonder the following question: which one of the used distributions allows to better estimate this parameter. By doing so, in this paper, we will examine the different estimators mean squared error of this parameter issued from each of these distributions to compare their performances. The mean squared errors of different estimators of the canonical parameter are functions of θ, we will proceed by a graphic resolution for comparison through some illustrative examples. Quality criteria estimators of θ through the amount of Fisher information of each random variable relative to the notion of mean duality between one variable overdispersed and underdispersed are proposed.

Received: August 20, 2013; Accepted: September 23, 2013

2010 Mathematics Subject Classification: 62F10.

Keywords and phrases: exponential family, Poisson distribution, weighted Poisson distribution, canonical parameter, Fisher information, overdispersion, underdispersion, mean duality, quality of an estimator, mean squared error.

## 1. Introduction

Consider $Y$ is a Poisson random variable with mass function

$$f_\theta(y) = \frac{\theta^y}{y!} e^{-\theta}, \quad y \in \mathbb{N}, \quad \theta \in \mathbb{R}_+^*, \tag{1}$$

depending on the canonical parameter $\theta$. When the Poisson model is not appropriate because the results from the processing of data do not allow to validate it, the question is to find an alternative model to describe the variations in the data studied.

Let $Z$ be a positive integer random variable, it is convenient to look for alternative families of Poisson distribution by referring to the dispersion Fisher index, which is the ratio of variance to the mean

$$I(z) = \frac{\operatorname{var}(Z)}{E(Z)}, \tag{2}$$

relatively to 1.

When $I(Z) > 1$ $(I(Z) < 1)$, we say that the variable $Z$ is *overdispersion* (*underdispersion*) (see Mizère et al. [10]). For the Poisson distribution $I(Z) = 1$; we say that in this case, Poisson is *equidispersed*. One of the alternative families is the weighted Poisson distribution (WPD) with probability mass function (pmf):

$$f_{\theta, w}(y) = \frac{w(y)}{E_\theta[w(y)]} \frac{\theta^y}{y!} e^{-\theta}, \quad \theta > 0, \quad y \in \mathbb{N}, \tag{3}$$

where $w(y)$ is called the *weight function*, a positive function and $E_\theta[w(Y)]$ denotes the mean value depending on $\theta$ such that $0 < E_\theta[w(Y)] < \infty$ (see Mizère et al. [9]). The weight function $w(y) = w(y, \phi)$ can depend on a parameter $\phi$ representing the data recording mechanism, and it may also be connected to the canonical Poisson parameter $\theta$. When the weight function $w(y)$ is a constant, we find the Poisson distribution.

For observed count data, it is of interest to use a statistical test for detecting the overdispersion or underdispersion (see Mizère et al. [9]), and it is therefore useful to have a family of count distributions possessing both overdispersion and underdispersion properties with respect to the parameter. In this case, the parameter estimation would lead to an appropriate model within the family for overdispersion or underdispersion count data. Such a process would automatically lead to an appropriate model depending on the type of observed data (see Shmueli et al. [16]).

For that in this paper, we use the notion of "duality" proposed by Mizère (see Mizère [8]) for such a family of WPD, which will be related to the combination of overdispersion and underdispersion, however, we only use the notion of mean duality. A practical meaning of mean duality for WPD is that this distribution provides the opposite dispersion (i.e., overdispersion from underdispersion and conversely).

Estimators of the parameter $\theta$ that we determine here generally are unbiased for the Poisson distribution and biased with respect to $\theta$ for the mean dual distributions. In other words, they do not have the same mean; so we cannot compare them through their variances. Otherwise, to make this comparison, we use the mean squared error of different estimators of the parameter $\theta$ with respect to $\theta$, however, it is useful to compare several estimators, especially when one of them is biased or when both are biased with respect to $\theta$.

If the estimators of $\theta$ estimate the same quantity, then this comparison is also possible when the difference in quantities of Fisher information on each variable is positive or negative for an equal number of observations. This means that the results of inference on the population parameter $\theta$ are more reliable when the amount of Fisher information contained in the sample on the parameter $\theta$ is sufficient. Subsequently, we link this comparison with the mean dual distributions.

The motivations for this work are those of Alavi [14] who considers a general exponential family having the same canonical parameter such as a weighted exponential family; he compares their quantities Fisher information to determine, which two families will be more informative for this parameter.

In this paper, we first consider, in particular, the mean dual distributions having the same Poisson canonical parameter, and we assume that these laws belong to the exponential family. Then we compare the different estimators for this parameter by using the mean square error method accompanied by illustrative examples and Fisher information method in order to assess their qualities.

## 2. Definitions

**Definition 1** (See Mizère et al. [10]). Let $w_1$ and $w_2$ be two positive weight functions generating two WPD. These two WPD are said to be (*punctually*) *dual* if their weight functions satisfy

$$w_1(y) \times w_2(y) = 1, \quad \forall y \in \mathbb{N} \tag{4}$$

with $w_i$ nonconstant.

**Remark 1.** (1) The positivity of $w_i$ $(w_i) > 0$ implies that the support of the corresponding WPD is the entire nonnegative set $\mathbb{N}$.

(2) As the weight functions cannot be constant, the Poisson distribution cannot be a point dual distribution (see Mizère [8]).

**Definition 2** (See Mizère et al. [10]). Let $w_1$ and $w_2$ be two positive weight functions generating two WPD. These two WPD are said to be *mean dual* if we have

$$E_\theta[w_1(Y)] \times E_\theta[w_2(Y)] = 1, \quad \forall y \in \mathbb{N} \tag{5}$$

with $w_i$ nonconstant.

**Remark 2.** The Poisson distribution cannot be a mean dual distribution. Thus, the Poisson weighted whose support is not equal to the entire set $\mathbb{N}$ can be for the mean dual distributions.

**Theorem 1** (See Mizère et al. [10]). *Let $Y$ be the Poisson random variable with mean $\theta > 0$ and let*

$$w(y) = w(y; \phi)$$

*be a non-null weight function. Consider* $Y^w$ *is the corresponding weighted Poisson random variable. The two following assertions are equivalent*:

(i) *The function* $\theta \mapsto E_\theta[w(Y)]$ *is logconvex* (*logconcave*).

(ii) $Y^w$ *is overdispersed* (*underdispersed*) *with respect to Y*.

**Theorem 2.** *Let* $f_{\theta, w}(y)$ *be a WPD with weight function* $w(y) = w(y, \theta, \phi)$ *depending on* $\theta$ *such that* $\ln[w(y, \theta, \phi)] = y\varphi(\theta) + \psi(\theta, \phi) + \eta(y, \phi)$ *with* $\varphi, \psi, \eta$ *functions*. *Then* $f_{\theta, w}(y)$ *is an exponential family*.

**Proof.** We know that

$$f_{\theta, w}(y) = \frac{w(y)}{E_\theta[w(Y)]} \frac{\theta^y}{y!} e^{-\theta}, \quad \theta > 0, \quad y \in \mathbb{N},$$

$$\begin{aligned} \ln[f_{\theta, w}(y)] &= \ln[w(y, \theta, \phi)] - \ln[E_\theta(w(Y))] + y\ln(\theta) - \theta - \ln(y!) \\ &= y\varphi(\theta) + \psi(\theta, \phi) + \eta(y, \phi) - \ln[E_\theta(w(Y))] + y\ln(\theta) - \theta - \ln(y!) \\ &= y(\varphi(\theta) + \ln(\theta)) + \psi(\theta, \phi) - \ln[E_\theta(w(Y))] - \theta + \eta(y, \phi) - \ln(y!) \\ &= \alpha(\theta)a(y) + \beta(\theta) + b(y) \end{aligned}$$

with $\alpha(\theta) = \varphi(\theta) + \ln(\theta)$, $a(y) = y$, $\beta(\theta) = \psi(\theta, \phi) - \ln[E_\theta(w(Y))] - \theta$ and $b(y) = \eta(y, \phi) - \ln(y!)$. □

**Remark 3.** Especially, when $\varphi(\theta) = \psi(\theta, \phi) = 0$, $\forall\theta$, then $w(y)$ being independent on $\theta$, then the distribution remains the exponential family.

## 3. Estimators Canonical Parameter

Let $(Y^{w_1}, Y^{w_2})$ be a mean dual pair of weighted Poisson random variables, of respective mass functions $f_{\theta, w_1}(y)$ and $f_{\theta, w_2}(y)$. These mean dual distributions and Poisson distribution depend on the same canonical parameter $\theta$. In this section, we will determine the characteristics estimators of $\theta$.

It is convenient also to note that the weight function that generates a Poisson weighted overdispersed (or underdispersed) used in this paper does not depend on $\theta$. However, for a mean dual distribution, the weight function of a weighted Poisson distribution overdispersed (underdispersed) depends on $\theta$ (see Mizère et al. [10]).

In other words, the weight function $w_1(y)$ does not depend on $\theta$. However, the weight function $w_2(y)$ depends on $\theta$ because of the mean duality in their mass functions $f_{\theta, w_1}(y)$ and $f_{\theta, w_2}(y)$. The expressions of $f_{\theta, w_1}(y)$ and $f_{\theta, w_2}(y)$ are obtained by subscripting the function $w(y)$ of expression (3) by 1 or 2.

Note by $\hat{\theta}_0$, $\hat{\theta}_1$ and $\hat{\theta}_2$, the parameter estimators of $\theta$ obtained by the distributions $f_\theta(y)$, $f_{\theta, w_1}(y)$ and $f_{\theta, w_2}(y)$.

Let $\tilde{y} = (y_1, y_2, \ldots, y_n)$ be a sample of size $n$ of values for the Poisson random variable $Y$ with parameter $\theta$, the estimator by the method of maximum likelihood is equal to

$$\hat{\theta}_0 = \frac{1}{n}\sum_{i=1}^{n} y_i, \tag{6}$$

which is an unbiased estimator and efficient of $\theta$, its variance is equal to

$$\operatorname{var}(\hat{\theta}_0) = \frac{\theta}{n}. \tag{7}$$

**Proposition 1.** *Let* $\tilde{y}^{w_1} = (y_1^{w_1}, y_2^{w_1}, \ldots, y_n^{w_1})$ *be a sample of size* $n$ *values of* $Y^{w_1}$ *a random variable. The domain of definition of the probability density* $f_{\theta, w_1}(y)$ *being independent of the parameter* $\theta$, *the logarithm of the probability density is as follows*:

$$\ln[f_{\theta, w_1}(y)] = \alpha_1(\theta)a_1(y) + \beta_1(\theta) + b_1(y).$$

**Proof.** Indeed,

$$f_{\theta, w_1}(y) = \frac{w_1(y)}{E_\theta[w_1(Y)]}\frac{\theta^y}{y!}e^{-\theta}, \quad \theta > 0, \quad y \in \mathbb{N}.$$

Taking the logarithm, we obtain

$$\ln[f_{\theta, w_1}(y)] = \ln[w_1(y)] - \ln[E_\theta(w_1(y))] + y\ln(\theta) - \theta - \ln(y!)$$

$$= \alpha_1(\theta)a_1(y) + \beta_1(\theta) + b_1(y)$$

with

$$\alpha_1(\theta) = \ln(\theta);\ a_1(y) = y;\ \beta_1(\theta) = -\theta - \ln[E_\theta(w_1(y))] \text{ and}$$

$$b_1(y) = \ln[w_1(y)]. \qquad \square$$

The sufficient statistic (see Saporta [15])

$$\hat{\theta}_1 = \frac{1}{n}\sum_{i=1}^{n} a_1(y_i^{w_1}) = \frac{1}{n}\sum_{i=1}^{n} y_i^{w_1} \tag{8}$$

is an unbiased estimator and efficient of function

$$k_1(\theta) = -\frac{\beta_1'(\theta)}{\alpha_1'(\theta)} = \theta + \theta\frac{d}{d\theta}\ln[E_\theta(w_1(Y))] \tag{9}$$

with $\beta_1'(\theta)$ the derivative of $\beta_1(\theta)$ with respect to $\theta$.

This leads to

$$E_\theta(\hat{\theta}_1) = k_1(\theta) = \theta + \theta\frac{d}{d\theta}\ln[E_\theta(w_1(Y))].$$

We deduce that $\hat{\theta}_1$ is a biased estimator of $\theta$ and the bias is equal to

$$Bias(\hat{\theta}_1/\theta) = \theta\frac{d}{d\theta}\ln[E_\theta(w_1(Y))] \tag{10}$$

and variance

$$\operatorname{var}(\hat{\theta}_1) = \frac{k_1'(\theta)}{n\alpha_1'(\theta)} = \frac{\theta}{n} + \frac{\theta}{n}\frac{d}{d\theta}\ln[E_\theta(w_1(Y))] + \frac{\theta^2}{n}\frac{d^2}{d\theta^2}\ln[E_\theta(w_1(Y))]. \tag{11}$$

**Proposition 2.** *Let* $\tilde{y}^{w_2} = (y_1^{w_2}, y_2^{w_2}, \ldots, y_n^{w_2})$ *be a sample of size* $n$ *values of* $Y^{w_2}$, *a mean dual random variable of* $Y^{w_1}$ *with* $w_2(y) = w_2(y, \theta)$. *The domain of definition of the probability density* $f_{\theta, w_2}(y)$ *being* *independent of the parameter* $\theta$, *the logarithm of the probability density is as follows*:

$$\ln[f_{\theta, w_2}(y)] = \alpha_2(\theta) a_2(y) + \beta_2(\theta) + b_2(y).$$

**Proof.** Indeed,

$$f_{\theta, w_2}(y) = \frac{w_2(y, \theta)}{E_\theta[w_2(Y, \theta)]} \frac{\theta^y}{y!} e^{-\theta}, \quad \theta > 0, \quad y \in \mathbb{N}.$$

Taking the logarithm, and taking into account Theorem 2, we obtain

$$\ln[f_{\theta, w_2}(y)] = \alpha_2(\theta) a_2(y) + \beta_2(\theta) + b_2(y)$$

with $\alpha_2(\theta) = \varphi(\theta) + \ln(\theta)$, $a_2(y) = y$, $\beta_2(\theta) = \psi(\theta) - \ln[E_\theta(w_2(Y, \theta))] - \theta$ and $b_2(y) = \eta(\theta) - \ln(y!)$. □

The sufficient statistic

$$\hat{\theta}_2 = \frac{1}{n} \sum_{i=1}^{n} a_2(y_i^{w_2}) = \frac{1}{n} \sum_{i=1}^{n} y_i^{w_2} \tag{12}$$

is an unbiased estimator and efficient of the function

$$k_2(\theta) = -\frac{\beta_2'(\theta)}{\alpha_2'(\theta)} = \frac{\theta + \theta \frac{d}{d\theta} \ln[E_\theta(w_2(Y))] - \theta\psi'(\theta)}{1 + \theta\varphi'(\theta)} \tag{13}$$

with $\beta_2'(\theta)$ the derivative of $\beta_2(\theta)$ with respect to $\theta$.

We therefore have

$$E_\theta(\hat{\theta}_2) = k_2(\theta) = \frac{\theta + \theta \frac{d}{d\theta} \ln[E_\theta(w_2(Y))] - \theta\psi'(\theta)}{1 + \theta\varphi'(\theta)}$$

$$= \theta + \left( \frac{\theta + \theta \frac{d}{d\theta} \ln[E_\theta(w_2(Y))] - \theta\psi'(\theta)}{1 + \theta\varphi'(\theta)} - \theta \right).$$

We deduce that $\hat{\theta}_2$ is a biased estimator of $\theta$ and the bias is equal to

$$Bias(\hat{\theta}_2/\theta) = \frac{\theta + \theta \frac{d}{d\theta} \ln[E_\theta(w_2(Y))] - \theta\psi'(\theta)}{1 + \theta\varphi'(\theta)} - \theta \tag{14}$$

and variance

$$\mathrm{var}(\hat{\theta}_2) = \frac{k_2'(\theta)}{n\alpha_2'(\theta)},$$

which gives

$$\mathrm{var}(\hat{\theta}_2) = \frac{\left(\frac{d^2}{d\theta^2} \ln[E_\theta(w_2(Y))] - \psi''(\theta, \phi)\right)\left(\varphi'(\theta) + \frac{1}{\theta}\right)}{n\left(\varphi'(\theta) + \frac{1}{\theta}\right)^3}$$

$$- \frac{\left(\varphi''(\theta) - \frac{1}{\theta^2}\right)\left(1 + \frac{d}{d\theta} \ln[E_\theta(w_2(Y))] - \psi'(\theta, \phi)\right)}{n\left(\varphi'(\theta) + \frac{1}{\theta}\right)^3}. \tag{15}$$

## 4. Estimators Comparison of the Canonical Parameter

### 4.1. Mean squared error method

The estimators $\hat{\theta}_1$ and $\hat{\theta}_2$ functions of $k_1(\theta)$ and $k_2(\theta)$, respectively, well being unbiased (see expressions (8) and (12)) do not feel the same function $(k_1(\theta) \neq k_2(\theta))$, so we cannot compare them through their variances. However, they are biased with respect to θ, which allows us to judge their quality by using the mean squared error as we note from MSE.

**Proposition 3** (See Fourdrinier [4])**.** *We have*

$$MSE(\hat{\theta}/\theta) = Bias^2(\hat{\theta}/\theta) + \mathrm{var}(\hat{\theta}). \tag{16}$$

Let $MSE(\hat{\theta}_0/\theta)$, $MSE(\hat{\theta}_1/\theta)$, $MSE(\hat{\theta}_2/\theta)$ be the mean squared errors of the estimators $\hat{\theta}_0$, $\hat{\theta}_1$, $\hat{\theta}_2$, respectively, with respect to the parameter θ.

The estimator with the value of the mean squared error is smallest is the best.

We have the following relationship (16):

$$MSE(\hat{\theta}_1/\theta) = Bias^2(\hat{\theta}_1/\theta) + \text{var}(\hat{\theta}_1),$$

$$MSE(\hat{\theta}_1/\theta) = \left(\theta \frac{d}{d\theta} \ln[E_\theta(w_1(Y))]\right)^2 + \frac{\theta}{n} + \frac{\theta}{n}\frac{d}{d\theta} \ln[E_\theta(w_1(Y))]$$
$$+ \frac{\theta^2}{n}\frac{d^2}{d\theta^2} \ln[E_\theta(w_1(Y))] \tag{17}$$

and

$$MSE(\hat{\theta}_2/\theta) = Bias^2(\hat{\theta}_2/\theta) + \text{var}(\hat{\theta}_2),$$

$$MSE(\hat{\theta}_0/\theta) = \text{var}(\hat{\theta}_0) = \frac{\theta}{n}. \tag{18}$$

**Remark 4.** To calculate the mean squared errors, it would require that the expressions of normalization constants $E_\theta[w_1(Y)]$ and $E_\theta[w_2(Y)]$ are explicit.

**Example 1.** We consider $f_{\theta, w_1}(y)$, size-biased Poisson distribution, which is an underdispersed distribution (see Mizère et al. [10]) with weight function $w_1(y) = y$, $y \in \mathbf{N}$ and with normalization constant $E_\theta[w_1(Y)] = \theta$.

Its mean dual distribution $f_{\theta, w_2}(y)$ (see Mizère et al. [10]) is the geometric distribution with success parameter $\theta \in \, ]0, 1[$ and with mass function

$$f_{\theta, w_2}(y) = \theta(1-\theta)^y, \quad y \in \mathbb{N}, \tag{19}$$

which is a family exponential distribution. The WPD corresponding is overdispersed with weight function

$$w_2(y) = \frac{(1-\theta)^y}{\theta^y} y! e^\theta$$

and with normalization constant

$$E_\theta[w_2(Y)] = \frac{1}{\theta}.$$

From expressions (11), (14), (17), (18), we establish the following table of calculations:

**Table 1**

| $f_\theta(y)$ | $f_{\theta, w_1}(y)$ | $f_{\theta, w_2}(y)$ |
|---|---|---|
| $Bias(\hat{\theta}_0/\theta) = 0$ | $Bias(\hat{\theta}_1/\theta) = 1$ | $Bias(\hat{\theta}_2/\theta) = \dfrac{1-\theta}{\theta}$ |
| $\text{var}(\hat{\theta}_0) = \dfrac{\theta}{n}$ | $\text{var}(\hat{\theta}_1) = \dfrac{\theta}{n}$ | $\text{var}(\hat{\theta}_2) = \dfrac{1-\theta}{n\theta^2}$ |
| $MSE(\hat{\theta}_0/\theta) = \dfrac{\theta}{n}$ | $MSE(\hat{\theta}_1/\theta) = 1 + \dfrac{\theta}{n}$ | $MSE(\hat{\theta}_2/\theta) = \left(\dfrac{1-\theta}{\theta}\right)^2 + \dfrac{1-\theta}{n\theta^2}$ |

We will proceed with a graphic resolution to compare the mean squared errors estimators of $\theta$ calculated in the table above.

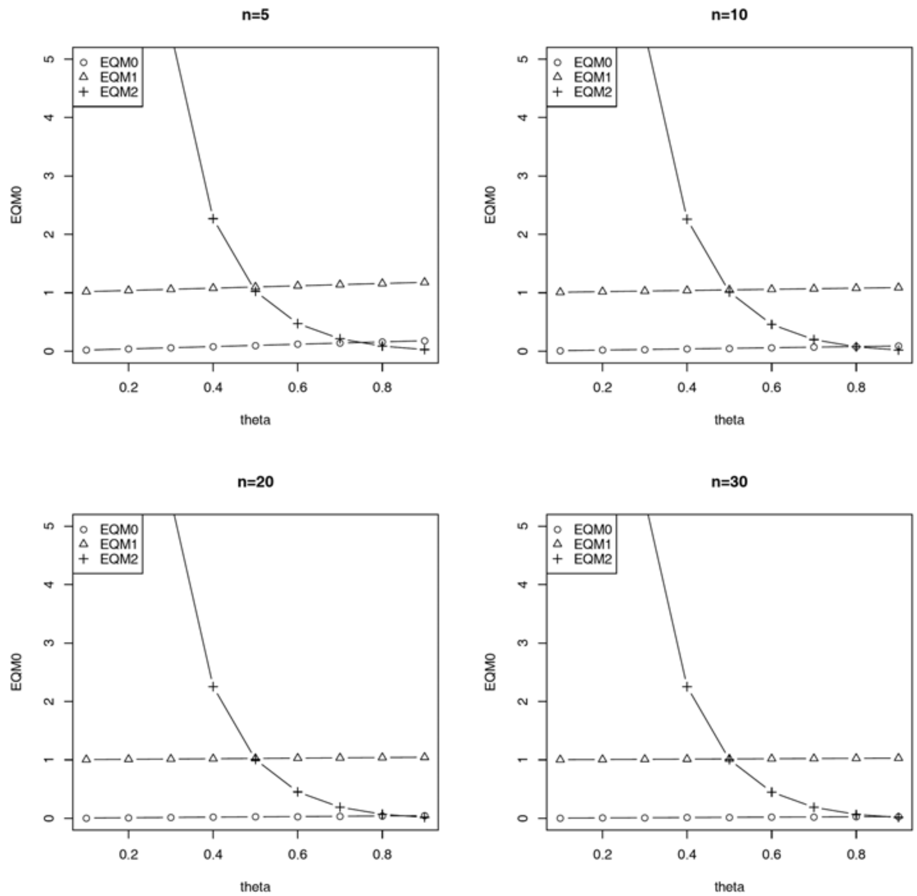


**Figure 1.** Graphical representation of $MSE(\hat{\theta}_0/\theta)$, $MSE(\hat{\theta}_1/\theta)$ and $MSE(\hat{\theta}_2/\theta)$, for $n = 5$, $n = 10$, $n = 20$ and $n = 30$.

In view of these graphs, we can conclude about the quality of the estimators: In the first graph, for $\theta \in \left]0, 0.7\right[$, $\hat{\theta}_0$ is better than $\hat{\theta}_1$ and $\hat{\theta}_2$. For $\theta \in \left]0.7, 1\right[$, $\hat{\theta}_2$ is better than $\hat{\theta}_0$ and $\hat{\theta}_1$.

In the second graph, for $\theta \in \left]0, 0.8\right[$, $\hat{\theta}_0$ is better than $\hat{\theta}_1$ and $\hat{\theta}_2$. For $\theta \in \left]0.8, 1\right[$, $\hat{\theta}_2$ is better than $\hat{\theta}_0$ and $\hat{\theta}_1$.

In the third and fourth graphs, for $\theta \in \left]0, 0.8\right[$, $\hat{\theta}_0$ is better than $\hat{\theta}_1$ and $\hat{\theta}_2$. For $\theta \in \left]0.8, 1\right[$, $\hat{\theta}_0$ and $\hat{\theta}_2$ have the same quality, but both are better than $\hat{\theta}_1$.

However, $\hat{\theta}_0$ is the best.

**Example 2.** We consider $f_{\theta, w_1}(y)$ the binomial distribution, which is an underdispersed distribution (see Mizère et al. [10]) with weight function

$$w_1(y) = \begin{cases} \dfrac{n!}{(n-y)!}, & y = 0, 1, \ldots, n, \\ 0, & y = n+1, n+2, \ldots \end{cases}$$

and with normalization constant

$$E_\theta[w_1(Y)] = (1+\theta)^n e^{-\theta}, \quad \theta \in \left]0, 1\right[.$$

Its mean dual distribution constructed by using the method of model selection (cf. Mizère et al. [10]) is an overdispersed distribution with weight function

$$w_2(y) = \frac{\Gamma(\phi + y)}{\Gamma(\phi)} \frac{(1-\theta)^\phi}{1+\theta} e^{2\theta}, \quad \phi \in \mathbb{R}_+^*,$$

and with normalization constant

$$E_\theta[w_2(Y)] = \frac{1}{(1+\theta)^n e^{-\theta}}.$$

The WPD corresponding is the negative binomial distribution with mass function

$$f_{\theta, w_2}(y) = \frac{\Gamma(\phi + y)}{y!\Gamma(\phi)}(1-\theta)^{\phi}\theta^{y}, \quad y \in \mathbb{N}, \quad \theta \in ]0, 1[, \quad \phi \in \mathbb{R}_{+}^{*}, \qquad (20)$$

which is an exponential distribution.

From expressions (11), (14), (17), (18), we have the results of the following calculations:

**Table 2**

| $f_{\theta, w_1}(y)$ | $f_{\theta, w_2}(y)$ |
|---|---|
| $Bias(\hat{\theta}_1/\theta) = \theta\left(\frac{n}{\theta+1} - 1\right)$ | $Bias(\hat{\theta}_2/\theta) = \theta\left(\frac{1-n}{\theta+1} + \frac{\phi}{1-\theta}\right)$ |
| $\mathrm{var}(\hat{\theta}_1) = \frac{\theta}{(\theta+1)^2}$ | $\mathrm{var}(\hat{\theta}_2) = \frac{\theta}{n}\left(\frac{1-n}{(1+\theta)^2} + \frac{\phi}{(1-\theta)^2}\right)$ |
| $MSE(\hat{\theta}_1/\theta) = \theta^2\left(\frac{n}{\theta+1} - 1\right)^2 + \frac{\theta}{(\theta+1)^2}$ | $MSE(\hat{\theta}_2/\theta) = \theta^2\left(\frac{1-n}{\theta+1} + \frac{\phi}{1-\theta}\right)^2 + \frac{\theta}{n}\left(\frac{1-n}{(1+\theta)^2} + \frac{\phi}{(1-\theta)^2}\right)$ |

By graphically representing those mean squared errors estimators of $\theta$, we obtain

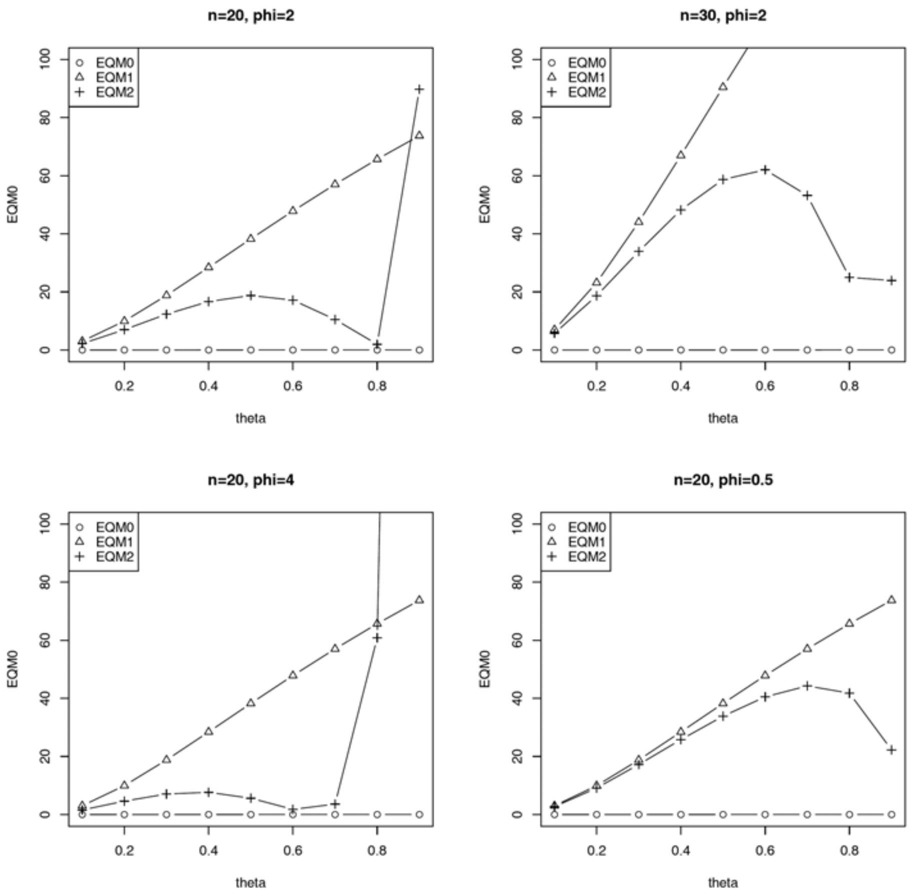


**Figure 2.** Graphical representation of $MSE(\hat{\theta}_0/\theta)$, $MSE(\hat{\theta}_1/\theta)$ and $MSE(\hat{\theta}_2/\theta)$.

In view of these graphs, we conclude the order of performance of each estimator: In the first graph, for $\theta \in ]0, 1[$, $\hat{\theta}_0$ is better than $\hat{\theta}_1$ and $\hat{\theta}_2$. And for $\theta \in ]0, 0.8[$, $\hat{\theta}_2$ is better than $\hat{\theta}_1$ and for $\theta \in ]0.8, 1[$, $\hat{\theta}_1$ is better than $\hat{\theta}_2$.

In the second and fourth graphs, for $\theta \in ]0, 1[$, $\hat{\theta}_0$ is better than $\hat{\theta}_2$, which is better than $\hat{\theta}_1$.

In the third graph, for $\theta \in ]0, 1[$, $\hat{\theta}_0$ is better than $\hat{\theta}_1$ and $\hat{\theta}_2$. However, for $\theta \in ]0, 0.7[$, $\hat{\theta}_2$ is better than $\hat{\theta}_1$ and for $\theta \in ]0.7, 1[$, $\hat{\theta}_1$ is better than $\hat{\theta}_2$.

However, $\hat{\theta}_0$ is the best.

### 4.2. Fisher information method

In this subsection, we assume that $k_1(\theta) = k_2(\theta) = k(\theta)$. This allows us to compare the estimators of θ defined in Section 3 through their variances and therefore also through their quantities of Fisher information. And we conclude that the variable, which has a much larger Fisher information in terms of θ will be more informative about the true value of θ.

Moreover, since we work with the distributions of the exponential family, the variances of different estimators of θ defined in Section 3 will be redefined as follows (Monfort [11]):

$$\operatorname{var}(\hat{\theta}_0) = \frac{1}{nI_Y(\theta)} \quad \text{and} \quad \operatorname{var}(\hat{\theta}_i) = \frac{(k'(\theta))^2}{nI_{Y^{w_i}}(\theta)} \quad (i = 1, 2) \tag{21}$$

with

$$I_Y(\theta) = E_\theta\left\{-\frac{d^2}{d\theta^2}\ln[f_\theta(y)]\right\} = \frac{1}{\theta} \tag{22}$$

is the quantity of Fisher information of the Poisson random variable $Y$.

And

$$I_{Y^{w_i}}(\theta) = E_\theta\left\{-\frac{d^2}{d\theta^2}\ln[f_{\theta, w_i}(y)]\right\} \tag{23}$$

are quantities of Fisher information of $Y^{w_i}$ $(i = 1, 2)$, of weighted Poisson random variables.

**Theorem 3.** *Let* $Y^{w_i}$ $(i = 1, 2)$ *be the weighted versions of the Poisson random variable* $Y$ *with parameter* $\theta$ *and* $w_i(y)$ $(i = 1, 2)$, *the weight functions*. *We get the following relations*:

$$I_{Y^{w_i}}(\theta) = I_Y(\theta) + \frac{d^2}{d\theta^2}\ln[E_\theta(w_i(Y))], \text{ if } w_i(y) \text{ do not depend on } \theta, \quad (24)$$

$$I_{Y^{w_i}}(\theta) = I_Y(\theta) + \frac{d^2}{d\theta^2}\ln[E_\theta(w_i(Y))] - E_\theta\left\{\frac{d^2}{d\theta^2}\ln[w_i(Y, \theta)]\right\},$$

$$\text{if } w_i(y) \text{ depend on } \theta. \quad (25)$$

**Proof.** Indeed, if weight functions $w_i(y)$ do not depend on $\theta$, we have

$$f_{\theta, w_i}(y) = \frac{w_i(y)}{E_\theta[w_i(Y)]} f_\theta(y), \quad \theta > 0, \quad y \in \mathbb{N},$$

then

$$\ln[f_{\theta, w_i}(y)] = \ln[f_\theta(y)] + \ln[w_i(y)] - \ln E_\theta[w_i(Y)].$$

Differentiating this expression twice, we obtain

$$\frac{d^2}{d\theta^2}\ln[f_{\theta, w_i}(y)] = \frac{d^2}{d\theta^2}\ln[f_\theta(y)] - \frac{d^2}{d\theta^2}\ln[E_\theta(w_i(y))].$$

However, from expressions (22) and (23), we have the result.

If weight functions $w_i(y)$ depend on $\theta$, then its first and second derivatives with respect to $\theta$ are nonzero. Therefore, we obtain the result by analogy to the foregoing. □

**Corollary 1.** *When the variables* $Y^{w_i}$ $(i = 1, 2)$ *are overdispersed* (*resp. underdispersed*) *and* $w_i$ *does not depend on* $\theta$, *then we have*

$$I_{Y^{w_i}}(\theta) > I_Y(\theta) \; (\textit{resp. } I_{Y^{w_i}}(\theta) < I_Y(\theta)). \quad (26)$$

In other words, the weighted Poisson overdispersed is more informative about the true value of the Poisson parameter θ (resp. Poisson is more informative about the true value of the WPD underdispersed parameter θ).

**Proof.** Indeed, when the variables $Y^{w_i}$ are overdispersed (resp. underdispersed), by Theorem 1, the logconvexity (resp. logconcave) of the functions $\theta \mapsto E_\theta[w_i(Y)]$ leads to $\frac{d^2}{d\theta^2}\ln[E_\theta(w_i(Y))] > 0\ (< 0)$. From expression (24), we deduce the result. CQFD. □

**Corollary 2.** *Let* $w_2(y) = w_2(y, \theta)$ *be the weight function of a random variable* $Y^{w_2}$ *mean dual of* $Y^{w_1}$ *with weight function* $w_1(y)$ *not depending on* θ. *When* $Y^{w_1}$ *is overdispersed and that the function* $\theta \mapsto w_2(y, \theta)$ *is logconvex, then* $I_{Y^{w_1}}(\theta) > I_{Y^{w_2}}(\theta)$.

**Proof.** Indeed,

$$I_{Y^{w_1}}(\theta) = I_Y(\theta) + \frac{d^2}{d\theta^2}\ln[E_\theta(w_1(Y))] \tag{27}$$

and

$$I_{Y^{w_2}}(\theta) = I_Y(\theta) + \frac{d^2}{d\theta^2}\ln[E_\theta(w_2(Y, \theta))] - E_\theta\left\{\frac{d^2}{d\theta^2}\ln[w_2(Y, \theta)]\right\}. \tag{28}$$

The difference member to member of expressions (27) and (28) gives

$$I_{Y^{w_1}}(\theta) - I_{Y^{w_2}}(\theta) = \frac{d^2}{d\theta^2}\ln[E_\theta(w_1(Y, \theta))] - \frac{d^2}{d\theta^2}\ln[E_\theta(w_2(Y, \theta))] + E_\theta\left\{\frac{d^2}{d\theta^2}\ln[w_2(Y, \theta)]\right\}. \tag{29}$$

As the variable $Y^{w_1}$ is overdispersed; then $Y^{w_2}$ is underdispersed. Moreover, as the function $\theta \mapsto w_2(y, \theta)$ is logconvex, we have

$$\frac{d^2}{d\theta^2}\ln[E_\theta(w_1(Y))] > 0, \quad \frac{d^2}{d\theta^2}\ln[E_\theta(w_2(Y, \theta))] < 0$$

and

$$\frac{d^2}{d\theta^2} \ln[(w_2(Y, \theta))] > 0.$$

And therefore, we obtain

$$I_{Y^{w_1}}(\theta) - I_{Y^{w_2}}(\theta) > 0. \tag{30}$$

Thus

$$I_{Y^{w_1}}(\theta) > I_{Y^{w_2}}(\theta) \ \mathit{CQFD}. \tag{31}$$

□

**Proposition 4.** *Let* $Y^{w_1}$ *be an overdispersed variable, and* $Y^{w_2}$ *be its mean dual version. Then* $\hat{\theta}_1$ *is better than* $\hat{\theta}_2$.

**Proof.** By Theorem 3, we have

$$I_{Y^{w_1}}(\theta) > I_{Y^{w_2}}(\theta),$$

which implies

$$\frac{1}{I_{Y^{w_1}}(\theta)} < \frac{1}{I_{Y^{w_2}}(\theta)}.$$

Taking into account expressions (21), we obtain

$$\mathrm{var}(\hat{\theta}_1) < \mathrm{var}(\hat{\theta}_2).$$

Therefore, $\hat{\theta}_1$ is better than $\hat{\theta}_2$ CQFD. □

### 4.3. Conclusion

In sum, when the estimators of θ do not estimate the same function, the estimators obtained by the Poisson distribution are better than those obtained by the mean dual distributions. But when these estimators estimate the same function, the choice is focused on the overdispersed mean dual distributions to estimate θ.

## References

[1] J. D. Castillo and M. Pérez-Casany, Weighted Poisson distributions for overdispersion and underdispersion situations, Ann. Inst. Statist. Math. 50(3) (1998), 567-585.

[2] J. D. Castillo and M. Pérez-Casany, Overdispersed and underdispersed Poisson generalizations, J. Statist. Plann. Inference 134 (2005), 486-500.

[3] E. Dietz and D. Böhning, On estimation of the Poisson parameter in zero-modified Poisson models, Comput. Statist. Data Anal. 34 (1999), 441-459.

[4] D. Fourdrinier, Inferential Statistics, Dunod Edition, 2002.

[5] A. E. Gelfand and S. R. A. Dalal, A note on overdispersed exponential families, Biometrika 77 (1990), 55-64.

[6] P. Hougaard, M.-L. Lee and G. A. Whitmore, Analysis of overdispersed count data by mixtures of Poisson variable and Poisson processes, Biometrics 53 (1997), 1225-1238.

[7] C. C. Kokonendji, D. Mizère and N. Balakrishnan, Connections of the Poisson weight function to overdispersion and underdispersion, J. Statist. Plann. Inference 138 (2008), 1287-1296.

[8] D. Mizère, Contribution to the modeling and statistical analysis of census data, Thesis presented at the University of Pau and the Adour countries, 2006.

[9] D. Mizère, C. C. Kokonendji and S. Dossou-Gbété, Some tests Poisson against general alternatives based on the dispersion index Fisher, Rev. Statist. Appl. 54(4) (2006), 61-84.

[10] D. Mizère, R. Makany, G. Kissita and Y. Nkasa, Determination of the mean dual distribution through the Poisson weight function, Far East J. Theo. Stat. 29(1) (2009), 53-63.

[11] A. Monfort, Course of Mathematical Statistics, Editions Economica, Paris, France, 1982.

[12] G. P. Patil, Weighted distributions, Encyclopedia of Environmetrics, A. H. El-Shaarawi and W. W. Piegorrsch, eds., John Wiley and Sons, Chichester, Vol. 4, 2002, pp. 2369-2377.

[13] G. P. Patil and C. R. Rao, Weighted distributions and size-biased sampling with applications to wildlife populations and human families, Biometrics 34 (1978), 179-189.

[14] S. M. R. Alavi, Fisher information in weighted distributions, Far East J. Theo. Stat. 29(1) (2009), 81-91.

[15] G. Saporta, Theory and Methods of Statistics, Editions Technip, Paris, 1978.

[16] G. Shmueli, T. P. Minka, J. B. Kadane, S. Borle and P. Boatwright, A useful distribution for fitting discrete data: revival of the Conway-Maxwell-Poisson distribution, J. Roy. Statist. Soc. Ser. C 54 (2005), 127-142.