



Modelling the number of olive groves in Spanish municipalities

María-Dolores Huete¹ and Juan A. Marmolejo²

¹University of Granada, Faculty of Labour Sciences, Dept. Statistics and Operational Research. C/ Rector López Argueta s/n, 18071-Granada, Spain. ²University of Granada, Faculty of Social Science, Dept. Statistics and Operational Research. C/ Santander 1, 52071-Melilla, Spain

Abstract

The univariate generalized Waring distribution (UGWD) is presented as a new model to describe the goodness of fit, applicable in the context of agriculture. In this paper, it was used to model the number of olive groves recorded in Spain in the 8,091 municipalities recorded in the 2009 Agricultural Census, according to which the production of oil olives accounted for 94% of total output, while that of table olives represented 6% (with an average of 44.84 and 4.06 holdings per Spanish municipality, respectively). UGWD is suitable for fitting this type of discrete data, with strong left-sided asymmetry. This novel use of UGWD can provide the foundation for future research in agriculture, with the advantage over other discrete distributions that enables the analyst to split the variance. After defining the distribution, we analysed various methods for fitting the parameters associated with it, namely estimation by maximum likelihood, estimation by the method of moments and a variant of the latter, estimation by the method of frequencies and moments. For oil olives, the chi-square goodness of fit test gives *p*-values of 0.9992, 0.9967 and 0.9977, respectively. However, a poor fit was obtained for the table olive distribution. Finally, the variance was split, following Irwin, into three components related to random factors, external factors and internal differences. For the distribution of the number of olive grove holdings, this splitting showed that random and external factors only account about 0.22% and 0.05%. Therefore, internal differences within municipalities play an important role in determining total variability.

Additional key words: table olive; oil olive; agricultural holdings; Waring distribution; estimation.

Abbreviations used: L (liability); MF12 (method of one equation of moments and two relations between frequency); MF21 (method of two relations between moments and one equation of frequency); MLE (method of log-likelihood optimisation); MM3 (method of the three relations between moments); P (proneness); R (randomness); UGWD (Univariate Generalized Waring Distribution).

Citation: Huete, M. D.; Marmolejo, J. A. (2016). Modelling the number of olive groves in Spanish municipalities. Spanish Journal of Agricultural Research, Volume 14, Issue 1, e0201. <http://dx.doi.org/10.5424/sjar/2016141-7687>.

Received: 12 Mar 2015 **Accepted:** 20 Jan 2016

Copyright © 2015 INIA. This is an open access article distributed under the terms of the Creative Commons Attribution-Non Commercial (by-nc) Spain 3.0 Licence, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Funding: This research is financed by Vice-Rector's Office for Political Science and Research-University of Granada, through the project "Social-Labour Statistics and Demography" (30.BB.11.1101) at the Faculty of Labour Sciences.

Competing interests: The authors have declared that no competing interests exist.

Correspondence should be addressed to María-Dolores Huete-Morales: mdhuete@ugr.es

Introduction

The question of olive production has aroused much interest in Spain and other areas, as olive oil (especially virgin oil) is the cornerstone of the Mediterranean diet; its consumption has been associated with a lower risk of cardiovascular disease (Fernández-Jarne *et al.*, 2001), obesity, metabolic syndrome, type 2 diabetes and hypertension. Moreover, it reduces the risk of cancer (López-Miranda *et al.*, 2010) and ageing by inhibiting oxidative stress (Owen *et al.*, 2000; Gimeno *et al.*, 2002).

Spain is the world's largest producer of olive oil, and the province of Jaén (in the south of the Iberian

Peninsula) contains the country's highest concentration of olive groves, producing more oil than Italy, the second largest producer country. The most abundant variety in Jaén is Picual, although other varieties, such as Royal, Arbequina and Cornicabra, are also grown. According to the Ministry of Agriculture, Food and Environment (MAGRAMA, 2012), of the total olive production in Spain, 94% is used for olive oil and 6% for table olives. In the production of olive oil, 41% originates in the province of Jaén, 21% in Córdoba and 6% in Sevilla. Table olives are mainly grown in Sevilla (57%), followed by Córdoba and Badajoz, each with 12.5% of national production. Regarding exports, the International Olive Council reported that by the end

of the 2012/2013 season, Spain had exported 728,621 tonnes of olives/olive oil, Italy 932,000 tonnes and Greece 229,137 tonnes.

Some of the distributions used to model discrete data are well known. The most commonly used is the Poisson distribution, which is simple to use and is widely applied. However, it underestimates the variance due to the phenomenon of dispersion; to overcome this problem, mixture distributions have been proposed, such as the negative binomial distribution, derived from mixing the Poisson and Gamma distributions. Another mixture, which has been applied to address issues in the field of ecology (Katti, 1966) is the Poisson-beta distribution. In the present study, we propose to apply the Waring distribution (a mixture of the negative binomial and beta distribution) to discrete data obtained in the agricultural context.

The generalized Waring distribution (Irwin, 1968, 1975) is a discrete distribution on non-negative integers. This distribution belongs to the Kemp type 4 family of distributions, and has an analogous continuous distribution, which in general is Pearson type 6, although in special cases it may be Pearson type 3, 4 or 5. It is infinitely divisible, self-compensating as defined by Danial (1988) and is a distribution “in limit terms”, complete (Xekalaki, 1983b). This distribution can be considered a particular balance distribution (Ferrerri, 1984), and efficient algorithms have been created for it by which random variables can be generated for Sibuya’s digamma and trigamma distributions (Devroye, 1992), as have applications in stochastic aggregation models (Duerr & Dietz, 2000). Sarabia & Castillo (2003) proposed two multivariate extensions of this distribution: one by means of the Sarmanov-Lee distribution with beta marginal, and the other using the concept of the conditional specification of distributions. Finally, Rodríguez-Avi *et al.* (2003) studied different parameter estimation methods based on moments and/or frequencies for Gauss’ family of hypergeometric distributions, while they (Rodríguez-Avi *et al.*, 2007) presented an example with which they compared the results obtained applying the maximum likelihood method to the negative binomial distribution, the univariate generalized Waring distribution and the extended Waring distribution, the latter being a tetra-parametric univariate distribution generated by Gauss’ hypergeometric function. This includes the generalized Waring distribution, as a combination of the negative binomial distribution and the beta type 1 generalized distribution.

The generalized Waring distribution has been applied in many scientific fields. Newbold (1926) reported that the distribution of accidents to workers in a soap factory fitted a negative binomial distribution. This result

was later fitted by Irwin (1968) by a univariate generalized Waring distribution, improving on the results obtained by Newbold (1926). Since then, this distribution has been used, independently of the theory of accidents (Irwin, 1963; Xekalaki, 1983c), in fields such as biology (Irwin, 1968), reliability theory (Xekalaki, 1983a), library science (Boxenbaum *et al.*, 1987), computer science (Wolfran, 2003), psychiatry (Canal & Micciolo, 1999), medicine (Kemp, 2001) and linguistics and economics (Kendall & Stuart, 1969). However, it has never been applied in studies related to agriculture, as in the present paper. To the best of our knowledge, no previous application has been made of this distribution in the agricultural context, and therefore this study makes a novel contribution that may be useful for future research. Thus, we present the univariate generalized Waring distribution as a useful tool in this area of study. Let us note that the exploratory data analysis performed justifies the use of the Waring distribution with respect to the variable number of olive holdings registered in Spain. We compared parameter estimation methods and determined which methods allowed us to split the variance of this distribution into three components. This approach opened up a range of possibilities that are not possible with other distributions.

Material and methods

The micro-data used in this study were obtained from the Agricultural Census (INE, 2009). This Census provides detailed information on the crops grown on all Spanish agricultural holdings, broken down by municipalities, thus supplying the information necessary for this study. We obtained maps of shapefiles for Spain, for the corresponding municipalities, using polygons, with the ETRS89 UTM 30N coordinate system (ESRI Map Service, <http://www.arcgis.com/>). Maps, graphs and distribution fitting were obtained using R free software (<https://www.r-project.org/>) and the GWRM package (Sáez-Castillo *et al.*, 2010), together with SPSS 20.0 to adapt the data and export them to R. In the following, we define the Waring distribution and describe the fitting methods used.

Waring distribution

A random variable X follows a univariate generalized Waring distribution ($UGWD(a, k; \rho)$), with parameters a , k and ρ , when it has the following probability mass function:

$$f(r) = P\{X = r\} = \frac{\rho_{(k)}}{(a + \rho)_{(k)}} \frac{a_{(r)}k_{(r)}}{(a + k + \rho)_{(r)}} \frac{1}{r!} \quad [1]$$

where $r = 0, 1, \dots, a, k, \rho > 0$ and $x_{(r)} = \frac{\Gamma(x+r)}{\Gamma(x)}$, $x > 0$, $r \in \mathbb{R}$ being $\Gamma(\cdot)$ the gamma function. The probability generator function of X has the following expression:

$$G(s) = \frac{\rho_{(k)}}{(a + \rho)_{(k)}} {}_2F_1(a, k; a + k + \rho; s)$$

where ${}_2F_1(a, b; c; d)$ is Gauss' hypergeometric function.

Parameter estimation: method of moments

With $\alpha = a; \beta = k; \gamma = a + k + \rho; \lambda = 1$ ($a, k \in \mathbb{R}; \rho > 0$) we obtain the probability generator function of the univariate generalized Waring distribution:

$$g(t) = \frac{{}_2F_1(a, k, a + \rho + 1, t)}{{}_2F_1(a, k, a + \rho + 1, 1)}$$

Let us focus on determining the parameters of this distribution; using the method of moments, via the recurrence relation among moments with respect to the origin (Marmolejo-Martín, 2003), for the discrete distributions in the system, that is:

$$\alpha_{h+2} + (a + k + \rho - 1)\alpha_{h+1} - \sum_{m=0}^h \binom{h}{m} (\alpha_{(m+2)} + (\alpha + k)\alpha_{(m+1)} + \alpha\beta\alpha_m) = 0$$

Therefore, we have the following system of three equations and three unknowns:

$$\begin{aligned} \alpha_1(a + k + \rho - 1) - \alpha_1(a + k) - ak &= 0 \\ \alpha_2(a + k + \rho - 1) - (\alpha_2 + \alpha_1)(a + k) - (\alpha_1 + 1)ak &= \alpha_2 \quad [2] \\ \alpha_3(a + k + \rho - 1) + (\alpha_3 + 2\alpha_2 + \alpha_1)(a + k) - \\ -(\alpha_2 + 2\alpha_1 + 1)ak &= 2\alpha_3 + \alpha_2 \end{aligned}$$

After calculating the first non-centred moments ($\alpha_1, \alpha_2, \alpha_3$), \hat{a}, \hat{k} and $\hat{\rho}$ are obtained by resolving [2].

Parameter estimation: two relations between moments and an initial relation between probabilities

An alternative to the method of moments is to consider the relation between probabilities, extracted from the following equation:

$$\begin{aligned} ((r + 1)^2 + (a + k + \rho - 1)(r + 1))\beta_{r+1} - \\ - (r^2 + r(a + k) + ak)\beta_r &= 0 \end{aligned}$$

$\forall r = 0, 1, 2, \dots$ and where β_r is the r -th probability. Assuming $r = 0$, we obtain the first relation, which is added to the first two equations between moments (Eq. [2]). This approach enables us to obtain \hat{a}, \hat{k} and $\hat{\rho}$:

$$\begin{aligned} \alpha_1(a + k + \rho - 1) - \alpha_1(a + k) - ak &= 0 \\ \alpha_2(a + k + \rho - 1) - (\alpha_2 + \alpha_1)(a + k) - (\alpha_1 + 1)ak &= \alpha_2 \quad [3] \\ \beta_1(a + k + \rho - 1) - \beta_0ak &= -\beta_1 \end{aligned}$$

Another possibility is the method of one relation between moments and two between probabilities (MF12). However, careful analysis shows that this method does not provide acceptable results and does not provide a good fit; therefore, it should not be used to estimate the parameters of this distribution.

Variance decomposition

The r -th factorial moment of the generalized Waring distribution is given by the following expression:

$$\mu_r = \frac{a_{(r)}k_{(r)}}{(\rho - 1)(\rho - 2)\dots(\rho - r)}$$

It is immediately obtained that all the moments of order r (central moment about the mean) are infinite if $\rho \leq r$, in other words, the mean is finite if $\rho > 1$ and the variance is finite if $\rho > 2$. The mean and variance are expressed as follows:

$$\begin{aligned} \mu_1 &= \frac{ak}{\rho - 1}, \rho > 1 \\ \sigma^2 = \mu_2 &= \frac{ak(\rho + a - 1)(\rho + k - 1)}{(\rho - 1)^2(\rho - 2)}, \rho > 2 \end{aligned}$$

Irwin (1968) obtained the following partition of the variance, when the latter is finite ($\rho > 2$), into three components; the first of these (σ_R^2) corresponds to random factors, the second (σ_λ^2), to the variability due to external factors that affect the population (liability) and the third (σ_v^2), to the differences in the internal conditions of the individuals (prone-ness):

$$\begin{aligned} \sigma^2 = \sigma_R^2 + \sigma_\lambda^2 + \sigma_v^2 &= \frac{ak}{\rho - 1} + \\ + \frac{ak(k + 1)}{(\rho - 1)(\rho - 2)} + \frac{a^2k(\rho + k - 1)}{(\rho - 1)^2(\rho - 2)}, \rho > 2 \end{aligned}$$

Results

Exploratory analysis

The frequency observed for the discrete variable (table olive and oil olive holdings) is shown in Table 1. We analysed the 8,091 Spanish municipalities listed in the 2009 Agricultural Census on land use (with respect to crops). As shown in the frequency table, most municipalities contain very few or no agricultural holdings for the production of table olives; however, 28 Spanish municipalities contain more than 200 such holdings. The municipalities with the highest numbers of these holdings are Arahál (618), Carmona (520) and Marchena (445), all located in the province of Sevilla, in southwest Spain. Regarding the production of oil olives, approximately half of all Spanish municipalities contain at least one holding, while 472 municipalities have more than 200. The largest numbers of such holdings are found in the municipalities of Martos (2,941), Alcalá la Real (2,572) and Priego

de Córdoba (2,438), and the majority of these 472 towns are in the provinces of Jaén, Córdoba and Granada, in southern Spain. This type of olive production is more widespread in Spain than that of table olives. The maps of the spatial distribution of the two activities (Fig. 1) show that table olive-related production is located mainly in the southern half of the Iberian Peninsula.

Table 1. Observed frequency of table and oil olive holdings in Spanish municipalities

Number	Table olive	Oil olive	Number	Table olive	Oil olive
0	5304	3932	27	6	28
1	879	483	28	3	32
2	471	207	29	5	28
3	288	145	30	4	18
4	202	136	31	4	16
5	145	103	32	5	30
6	129	92	33	1	21
7	91	79	34	3	21
8	61	85	35	3	21
9	46	83	36	5	20
10	41	71	37	3	26
11	40	54	38	4	23
12	28	64	39	1	16
13	22	58	40	1	17
14	22	54	41	2	22
15	22	55	42	2	21
16	20	44	43	2	15
17	18	30	44	1	16
18	6	37	45	0	16
19	8	40	46	0	19
20	9	43	47	4	22
21	9	32	48	0	8
22	10	39	49	2	14
23	10	27	[50-100)	63	586
24	6	40	[100-200)	46	468
25	4	32	200+	28	472
26	2	30	Total	8 091	8 091

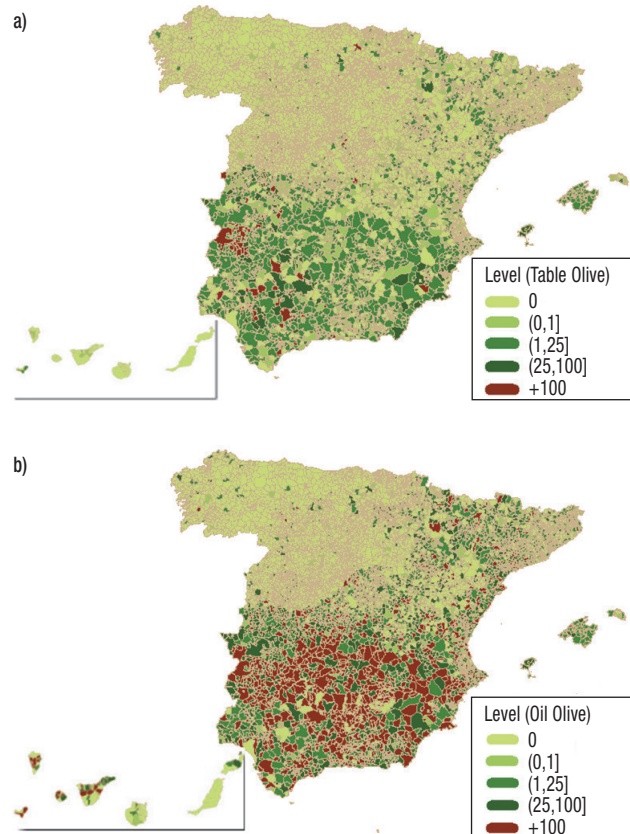


Figure 1. Distribution of the number of table olive (a) and oil olive (b) holdings in Spanish municipalities.

The type of data analysed presents very high frequencies for the first categories of the variable, but these decrease very rapidly to residual levels for the higher classes. Regarding table olives, we recorded an average of 4.06 holdings per municipality, with a coefficient of asymmetry of 12.86 (the third quartile had a value of 1), and so the distribution of this variable can be considered highly asymmetric. The same situation was found for oil olive production, with a mean of 44.84 holdings per municipality and an asymmetry of 7.94, although the value of the third quartile was 26. The frequency distribution is illustrated in Fig.2. A discrete distribution must be used to fit this type of data in order to reflect the asymmetry that is present.

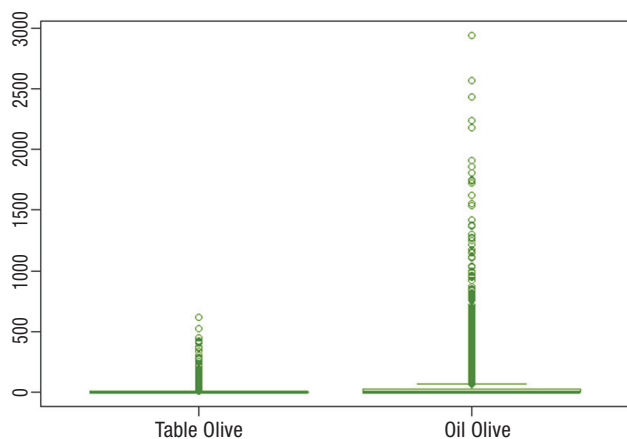


Figure 2. Box plot of the distribution of olive holdings in Spanish municipalities: both distributions are highly asymmetric, but especially that for oil olives.

Waring distribution adjustment

The system of equations [2] and the system [3] were implemented in R. Table 2 shows the results obtained after fitting the UGWD, applying the method of three relations between moments, MM3, two equations of moments and one equation of frequency, MF21, and using log-likelihood optimisation, MLE (a Newton-type algorithm), which was implemented using the GWRM package of R.

The expected values according to the different estimation methods are shown in Table 3 and Fig. 3. Due to the large number of cases, only the first 15 are shown. In cases where $\hat{\rho} > 2$, the variance can be split (following Irwin) into three factors: randomness, liability and proneness (Table 4). This is a major advantage of the Waring distribution over other discrete distributions.

Table 2. Fitting methods for olive holdings: parameters estimated using the method of three relations between moments (MM3), two equations of moments and one equation of frequency (MF21) and log-likelihood optimisation (MLE) and chi-square goodness of fit test for discrete data: statistic value and *p*-value (good fit is highlighted in bold).

Parameter	$(\hat{a}, \hat{k}, \hat{\rho})$	χ^2 statistic	χ^2 <i>p</i> -value
Table olive (MM3)	(1576.641; 0.0319; 13.3705)	5191.5	< 0.0001
Oil olive (MM3)	(2318.7957; 0.1214; 7.2804)	412.1	0.9967
Table olive (MF21)	(27.9398; 0.1801; 2.2385)	403.6	< 0.0001
Oil olive (MF21)	(2138.3544; 0.1232; 6.8775)	408.5	0.9977
Table olive (MLE)	(3.6491; 0.2402; 1.0815)	187.7	0.0071
Oil olive (MLE)	(2101.1182; 0.1267; 6.9443)	399.8	0.9992

Table 3. Observed frequencies for olive holdings in Spanish municipalities (only the first values) and estimated frequencies using the method of three relations between moments (MM3), two equations of moments and one equation of frequency (MF21) and log-likelihood optimisation (MLE).

Var. value	Table olive				Oil olive			
	Observed	MM3	MF21	MLE	Observed	MM3	MF21	MLE
0	5304	6940.0	4905.9	5297.4	3932	3986.9	3955.2	3890.7
1	879	219.3	813.0	934.3	483	482.6	485.9	491.2
2	471	112.2	442.7	451.1	207	269.8	272.0	275.8
3	288	75.4	297.7	273.0	145	190.2	191.9	194.9
4	202	56.6	219.5	184.5	136	147.9	149.3	151.8
5	145	45.3	170.6	133.4	103	121.5	122.7	124.9
6	129	37.7	137.2	101.1	92	103.4	104.5	106.3
7	91	32.2	113.1	79.2	79	90.1	91.1	92.8
8	61	28.0	94.9	63.8	85	80.0	80.8	82.4
9	46	24.8	80.8	52.5	83	72.0	72.7	74.1
10	41	22.2	69.6	43.9	71	65.4	66.1	67.4
11	40	20.1	60.6	37.2	54	60.0	60.7	61.9
12	28	18.3	53.1	32.0	64	55.4	56.0	57.2
13	22	16.8	47.0	27.8	58	51.5	52.1	53.2
14	22	15.5	41.7	24.3	54	48.1	48.7	49.7

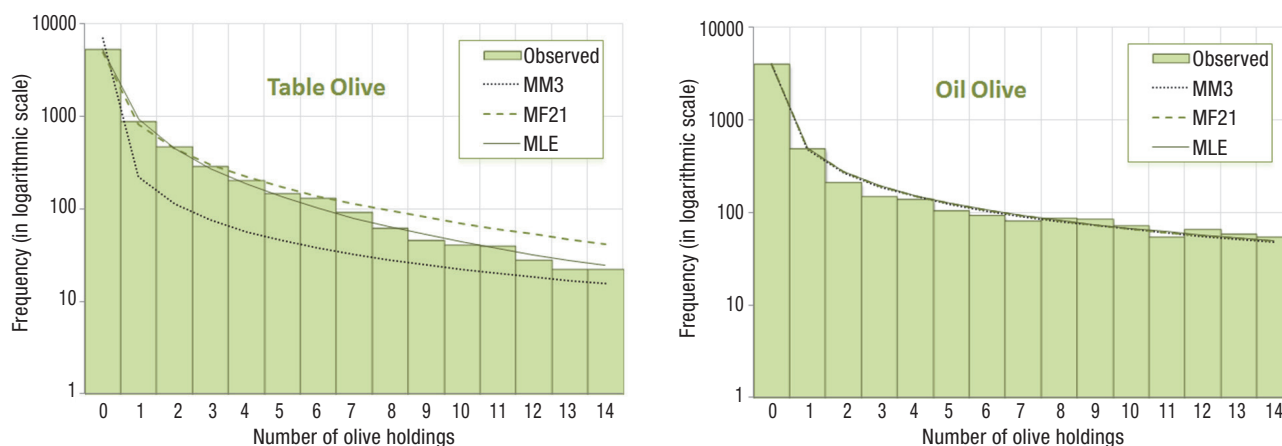


Figure 3. Observed and fitted distribution of olive holdings in Spain (for clarity, only the first values), showing the method of three relations between moments (MM3), two equations of moments and one equation of frequency (MF21) and log-likelihood optimisation (MLE).

Table 4. Breakdown of the variance according to the method used to estimate olive holdings in Spanish municipalities: method of three relations between moments (MM3), two equations of moments and one equation of frequency (MF21) and log-likelihood optimisation (MLE). Randomness (R), liability (L) and proneness (P), only for the methods in which the variance is finite, $\hat{\rho}$.

Method	Table olive				Oil olive			
	R	L	P	Total	R	L	P	Total
MM3	4.06	0.37	564.72	569.15	44.83	9.52	20,070.57	20,124.92
MF21	4.06	20.09	544.99	569.14	44.83	10.32	20,069.76	20,124.91
MLE	–	–	–	–	44.78	10.21	19,437.99	19,489.98

Discussion

The quality and health benefits of Spanish olive oil are unarguable (Barranco *et al.*, 2008) and this product is an essential element of the Mediterranean diet (Anta *et al.*, 2005). Spain is the world's largest producer of oil olives (Lambarraa *et al.*, 2007) and an extensive land area is dedicated to its cultivation; thus, olive groves form part of the landscape. From the economic standpoint, this industry is of vital importance; in agricultural production, it is second only to intensive horticulture (Sayadi *et al.*, 2012). Hence, the importance of extending the knowledge and understanding of the tools that enable in-depth analysis of this type of agricultural production.

Other types of distribution have previously been applied in agricultural research. In particular, binomial distribution, negative binomial distribution, Poisson distribution and mixture models are well known and have been used in numerous studies, in areas as diverse as counting dung patches (Monton & Baird, 1990), crop quantities and farms (Ridout *et al.*, 1998; Kim *et al.*, 2005; Bravo *et al.*, 2006; Paxton *et al.*, 2011), species (Royle, 2004; Brotons *et al.*, 2005; Kery *et al.*, 2005), and the number of food groups consumed (Hirvonen & Hoddinott, 2004), etc. Although the War-

ing distribution model is relatively unknown for studies based on discrete data, it is in fact very suitable for this application. The Waring distribution is valid when the frequency of occurrence is very low, as is the case with the distribution of olive holdings.

Several methods can be used for estimating the UGWD (a, k, ρ) parameters, including maximum likelihood, the method of moments and methods based on the relations between moments and frequencies. The results obtained by the method of moments show that the distribution is virtually biparametric, as the value of k is practically zero in most cases. Rodríguez-Avi *et al.* (2003) obtained a value of $\hat{k} < 0.067$ using the data reported by Beal & Rescia (1953) and by Katti & Gurland (1961). Canal & Micciolo (1999) also obtained values of 0.476 and 0.720 in the fits obtained \hat{k} for patients' psychiatric records.

The methods we recommend produce similar results, although the method of two relations between moments and one for frequencies (MF21) produces a good fit and good estimates of the parameters, and presents a significant advantage over the maximum likelihood method, namely its speed of calculation; numerical resolution methods are not needed, since the equations to be solved to apply this method do not require them.

Regarding the differences between the two variables analysed, a significant fit is obtained for the oil olive variable; the asymmetry of this variable (7.94), with respect to table olives (12.86), is less marked, and so the following indications are made for readers who may wish to use this type of distribution to fit discrete observations.

Observation of the parameter estimates obtained using the MLE, MM3 and MF21 methods reveals that with MM3 and MF21 the variance is finite and can be split as proposed by Irwin (see the Methodology section). However, this is not the case with the MLE method, as can be seen for the table olive variable, where $\rho < 2$. Moreover, the estimators obtained by the method based on the first relationship between moments and the first two relations between probabilities (MF12) cannot be considered, because they do not offer a good fit (for this reason, they are excluded from the present analysis and we do not recommend their use for estimating the Waring distribution). Note that the estimators obtained using the method of moments are considerably higher (Table 2), due to the multiplicative nature of their calculation.

Finally, regardless of the type of methodology used to estimate a , k , ρ the parameters, we stress that the major advantage of using the Waring distribution is that it allows us to split the variance, thus revealing the behaviour of the distribution relative to the intrinsic randomness in the observations, the external factors that may influence this behaviour (liability) and, finally, the internal differences between individuals (prone-ness). As can be seen in our split of the variance for the number of olive holdings (Table 4), random variance takes into account all the effects that cannot be explained, and this value is relatively small in both cases.

The term “Liability” refers to the variability present in each municipality, regarding parameters such as size, geographic location and local climate. Therefore, it cannot be attributed to external factors. Prone-ness, on the other hand, measures the variability between groups of municipalities, according to the number of holdings in each one. The proportion of liability to prone-ness is very low, indicating that this variable is well described. Splitting the oil olive data shows that the proportion of variability due to random factors is about 0.22% and that only 0.05% is due to external factors, for all the estimation methods used. The differences between municipalities account for 99.73% of the variability, and this plays an important role in the total variability present in the variable studied. Therefore, a large proportion of the variability corresponds to factors that are unknown and uncontrolled, such as climate and regional economic development. In the same way, splitting the table olive data shows that prone-ness is about 96%.

This split, as observed above, is a major contribution of the Waring distribution, and one that is not provided by other distributions. We show that the variability arising from external factors, among groups of municipalities, is very high. It is in this case that we might consider including explanatory variables in the model and applying a Waring regression model.

References

- Anta J, Palacios J, Guerrero F, 2005. La cultura del olivo. Ecología, economía, sociedad. University of Jaén, Spain.
- Barranco D, Fernández-Escobar R, Rallo L, 2008. El cultivo del olivo. Mundi-Prensa, Spain.
- Beal G, Rescia R, 1953. A generalization of Neyman’s contagious distribution. *Biometrics* 9: 354-386. <http://dx.doi.org/10.2307/3001711>
- Boxenbaum H, Pivinski F, Ruberg SJ, 1987. Publication rates of pharmaceutical scientists: Application of the Waring distribution. *Drug Metab Rev* 18: 553-571. <http://dx.doi.org/10.3109/03602538708994132>
- Bravo-Ureta B, Cocchi H, Solí D, 2006. Output diversification among small-scale hillside farmers in El Salvador. Office of Evaluation and Oversight. Working Paper OVE/WP-17/06.
- Brotos L, Wolff A, Paulusc G, Martin JL, 2005. Effect of adjacent agricultural habitat on the distribution of passerines in natural grasslands. *Biol Conserv* 124(3): 407-414. <http://dx.doi.org/10.1016/j.biocon.2005.01.046>
- Canal L, Micciolo R, 1999. Modelli probabilistici per l’analisi dei contatti psichiatrici. *Epidem Psich Soc* 8: 47-55. <http://dx.doi.org/10.1017/S1121189X0000751X>
- Daniail E, 1988. Generalization to the sufficient conditions for a discrete random variable to be infinitely divisible. *Stat Probabil Lett* 6: 379-382. [http://dx.doi.org/10.1016/0167-7152\(88\)90095-8](http://dx.doi.org/10.1016/0167-7152(88)90095-8)
- Devroye L, 1992. Random variate generation for the digamma and trigamma distributions. *J Stat Comput Sim* 43: 197-216. <http://dx.doi.org/10.1080/00949659208811438>
- Duerr H, Dietz K, 2000. Stochastic models for aggregation processes. *Math Biosci* 165: 135-145. [http://dx.doi.org/10.1016/S0025-5564\(00\)00014-6](http://dx.doi.org/10.1016/S0025-5564(00)00014-6)
- Fernández-Jarne J, Martínez-Losa E, Prado-Santamaría M, Brugarolas-Brufaua C, Serrano-Martínez M, Martínez-González MA, 2001. Risk of first non-fatal myocardial infarction negatively associated with olive oil consumption: a case-control study in Spain. *I J Epidemiol* 31(2): 474-480.
- Ferreri C, 1984. On the hypergeometric birth process and some implications about the gamma distribution representation. *J Roy Stat Soc B* 46:52-57.
- Gimeno E, Fitó M, Lamuela-Raventós RM, Castellote AI, Covas M, Farré M, Torre-Boronat MC, López-Sabater MC, 2002. Effect of ingestion of virgin olive oil on human low-density lipoprotein composition. *Eur J Clin Nutr* 56(2): 114-120. <http://dx.doi.org/10.1038/sj.ejcn.1601293>

- Hirvonen K, Hoddinott J, 2004. Agricultural production and children's diets: Evidence from rural Ethiopia. International Food Policy Research Institute. Working paper 69. <http://www.ifpri.org/>.
- INE, 2009. Agrarian Census. National Institute of Statistics, Spain. <http://www.ine.es>.
- Irwin J, 1963. The place of mathematics in medical and biological statistics. *J Roy Stat Soc A Sta* 126: 1-44.
- Irwin J, 1968. The generalized Waring distribution applied to accident theory. *J Roy Stat Soc A Sta* 131: 205-207.
- Irwin J, 1975. The generalized Waring distribution. *J Roy Stat Soc A Sta* 138: 18-31, 204-227, 374-378.
- Katti S, 1966. Interrelations among generalized distributions and their components. *Biometrics* 22: 44-52. <http://dx.doi.org/10.2307/2528212>
- Katti S, Gurland J, 1961. The Poisson Pascal distribution. *Biometrics* 17: 527-538. <http://dx.doi.org/10.2307/2527853>
- Kemp A, 2001. The q-beta-geometric distribution as a model for fecundability. *Commun Stat Theory* 30: 2373-2384. <http://dx.doi.org/10.1081/STA-100107692>
- Kendall M, Stuart A, 1969. The advanced theory of statistics, vol I, vol II. Griffin, London.
- Kery M, Royle A, Schmid H, 2005. Modeling avian abundance from replicated counts using binomial mixture models. *Ecol Appl* 15(4): 1450-1461. <http://dx.doi.org/10.1890/04-1120>
- Kim CS, Schluter G, Schaible G, Mishra A, Hallahan C, 2005. A decomposed negative binomial model of structural change: a theoretical and empirical application to U.S. agriculture. *Can J Agr Econ* 53: 161-176. <http://dx.doi.org/10.1111/j.1744-7976.2005.00396.x>
- Lambarraa F, Serra T, Gil JM, 2007. Technical efficiency analysis and decomposition of productivity growth of Spanish olive farms. *Span J Agric Res* 5(3): 259-270. <http://dx.doi.org/10.5424/sjar/2007053-247>
- López-Miranda J, Pérez-Jiménez F, Ros E, De Caterina R, Badimón L, Covase MI, Escrib E, Ordovás JM, Soriguerh F, Abiá R et al., 2010. Olive oil and health. *Nutr Metab Cardiovas* 20(4): 284-294. <http://dx.doi.org/10.1016/j.numecd.2009.12.007>
- MAGRAMA, 2012. Statistical yearbook, Technical report. Ministry of agriculture, food and environment. <http://www.magrama.gob.es/es/estadistica/temas/estadisticas-agrarias/>.
- Marmolejo-Martín J, 2003. Métodos de generación de distribuciones. Aplicación a la distribución de Waring. Doctoral thesis. Univ. of Granada, Spain.
- Monton J, Baird D, 1990. Spatial distribution of dung patches under sheep grazing. *New Zeal J Agr Res* 33: 285-294. <http://dx.doi.org/10.1080/00288233.1990.10428421>
- Newbold E, 1926. A contribution to the study of the human factor in the causation of accidents. Industrial Health Research Board Report, 34. London.
- Owen R, Giacosa A, Hull W, Haubner R, Wuertele G, Spiegelharder B, Bartsch H, 2000. Olive-oil consumption and health: the possible role of antioxidants. *Lancet Oncol* 1(2): 107-112. [http://dx.doi.org/10.1016/S1470-2045\(00\)00015-2](http://dx.doi.org/10.1016/S1470-2045(00)00015-2)
- Paxton K, Mishra A, Chintawar S, Roberts R, Larson J, English B, Lambert D, Marra M, Larkin S, Reeves J et al., 2011. Intensity of precision agriculture technology adoption by cotton producers. *Agr Resource Econ Rev* 40(1): 133-144.
- Ridout M, Demétrio C, Hinde J, 1998. Models for count data with many zeros. International Biometric Conference. Cape Town.
- Rodríguez-Avi J, Conde-Sánchez A, Sáez-Castillo AJ, Olmo MJ, 2003. Estimation of parameters in gaussian hypergeometric distributions. *Stat Theory* 32: 1101-1118. <http://dx.doi.org/10.1081/STA-120021322>
- Rodríguez-Avi J, Conde-Sánchez A, Sáez-Castillo AJ, Olmo MJ, 2007. A new generalization of the Waring distribution. *Comput Stat Data An* 51: 6138-6150. <http://dx.doi.org/10.1016/j.csda.2006.12.029>
- Royle A, 2004. N-mixture models for estimating population size from spatially replicated counts. *Biometrics* 60: 108-115. <http://dx.doi.org/10.1111/j.0006-341X.2004.00142.x>
- Sáez-Castillo AJ, Vilchez-López S, Olmo-Jiménez MJ, Rodríguez-Avi J, Conde-Sánchez A, Martínez-Rodríguez AM, 2010. GWRM package in "the r-project for statistical computing". Technical report. <http://cran.r-project.org/web/packages/GWRM/index.html>.
- Sarabia JM, Castillo E, 2003. Extensiones multivariantes de la distribución de Waring generalizada. Proc XVII Nat Cong Statistic and Operational Research, Lleida (Spain), Apr 8-11. pp: 344-354.
- Sayadi S, Ruiz P, Vázquez A, 2012. Prioridades de I+D en el sector agroalimentario andaluz: especial referencia a su complejo olivarero-oleícola. *Rev Esp Estudios Agrosoc Pesq* 233: 129-178.
- Wolfran D, 2003. Applying informetric characteristics of databases to ir system file desing, part I: Informetric models. *Inform Process Manag* 28: 121-133. [http://dx.doi.org/10.1016/0306-4573\(92\)90098-K](http://dx.doi.org/10.1016/0306-4573(92)90098-K)
- Xekalaki E, 1983a. Hazard functions and life distributions in discrete time. *Comm StatTheory* 12: 2503-2509. <http://dx.doi.org/10.1080/03610928308828617>
- Xekalaki E, 1983b. Infinite divisibility, completeness and regression properties of the univariate generalized waring distribution. *Ann I Stat Math* 35: 279-289. <http://dx.doi.org/10.1007/BF02480983>
- Xekalaki E, 1983c. The univariate generalized Waring distribution in relation to accident theory: proneness, spells or contagion? *Biometrics* 39: 887-895. <http://dx.doi.org/10.2307/2531324>