Journal of Physics: Conference Series

PAPER • OPEN ACCESS

Model for estimating of population abundance using line transect sampling

To cite this article: Gamil Abdulraqeb Abdullah Saeed *et al* 2017 *J. Phys.: Conf. Ser.* **890** 012153

Model for estimating of population abundance using line transect sampling

Gamil Abdulraqeb Abdullah Saeed, Noryanti Muhammad, Chuan Zun Liang, Wan Nur Syahidah Wan Yusoff, Mohd Zuki Salleh

Fakulti Sains & Teknologi Industri, Universiti Malaysia Pahang, 26300 Gambang, Kuantan, Pahang, Malaysia

E-mail: hamadijamil2010@gmail.com, noryanti@ump.edu.my, chuanzunliang@ump.edu.my, wnsyahidah@ump.edu.my, zuki@ump.edu.my

Abstract. Today, many studies use the nonparametric methods for estimating objects abundance, for the simplicity, the parametric methods are widely used by biometricians. This paper is designed to present the proposed model for estimating of population abundance using line transect technique. The proposed model is appealing because it is strictly monotonically decreasing with perpendicular distance and it satisfies the shoulder conditions. The statistical properties and inference of the proposed model are discussed. In the presented detection function, theoretically, the proposed model is satisfied the line transect assumption, that leads us to study the performance of this model. We use this model as a reference for the future research of density estimation. In this paper we also study the assumption of the detection function and introduce the corresponding model in order to apply the simulation in future work.

1. Introduction

Line transect method is an important procedure for estimating the population abundance or density (D) of objects in a study area. It is the simplest, most practical and least expensive method among other methods of population abundance estimation. The principle of this method is as follows:- An observer traverses a line and perpendicular measurements X_i are taken from the line to each observed object. Let $g(x)$ be a detection function, it is defined as

$$g(x) = \mathrm{P}(\text{an object is detected given its perpendicular distances } x \text{ from line}).$$

The logical assumption on $g(x)$ is that $g(x)$ is monotonically decreasing in X and $g(0) = 1$. These two assumptions indicate that the probability of detection decreases as the perpendicular distance increases and the probability of detecting an object on the line center (i.e. $x = 0$) will never be missed respectively. Let $f(x)$ be the probability density function (pdf) from which the perpendicular distances $X_1, ..., X_n$ are drawn. Burnham and Anderson [4] showed that

$$f(x) = \frac{g(x)}{\int_0^w g(t)dt}; 0 \leq x \leq w \tag{1}$$

where w is a truncated distance which is the perpendicular distance beyond w will not be considered. Burnham et al [5] suggested that the detection function $g(x)$ should have a shoulder at the origin, which means the detection in a very narrow distance from the line transect center should remain very close to one. This can be expressed mathematically as $g'(0) = 0$ or equivalently $f'(0) = 0$ see also, [3]. Let $\hat{f}(0)$ be the estimator of $f(0)$ which is evaluated based on n observed perpendicular distances $X_1, ..., X_n$. Burnham and Anderson [4] introduced the general estimator for D, which is given as

$$\hat{D} = \frac{n\hat{f}(0)}{2L} \tag{2}$$

This formula shows that the estimation of D is equivalent to estimate $f(0)$. The estimator D can be obtained by using a parametric or a nonparametric approach. Normally, in parametric approach, $f(x,\theta)$ will take parametric families with unknown parameter θ. However, the $f(x,\theta)$ depends on unknown parameter θ, where θ be vector of unknown parameters should be estimated based on the perpendicular distances $X_1, ..., X_n$. Gates [6] suggested the negative exponential detection function,

$$g(x;\theta) = e^{-x/\theta}; x \geq 0, \theta > 0 \tag{3}$$

with pdf,

$$f(x;\theta) = \frac{1}{\theta}e^{-x/\theta}; x \geq 0, \theta > 0 \tag{4}$$

The MLE method indicates that the MLE of $f(0)$ is $\hat{f}_{MLE}(0) = 1/\bar{X}$, where $\bar{X}$ is the sample mean. The detection function $g(x)$ does not satisfy the shoulder condition, which less importance the utilization of this model in line transect sampling.
In contrast, the half normal model, which suggested by Hemingway [7] satisfies the shoulder condition assumption. The half normal detection function is given by

$$g(x;\sigma^2) = e^{-x^2/2\sigma^2}; x \geq 0, \sigma^2 > 0 \tag{5}$$

with pdf,

$$f(x;\sigma^2) = \frac{2}{\sigma\sqrt{2\pi}}e^{-x^2/2\sigma^2}; x \geq 0, \sigma^2 > 0 \tag{6}$$

satisfies the shoulder condition because $g'(x;\sigma^2) = -\frac{x}{\sigma^2}e^{-x^2/2\sigma^2}$, which gives $g'(0;\sigma^2) = 0$.
The MLE of $f(0;\sigma^2)$ is $\hat{f}_{MLE}(0) = \sqrt{\frac{2}{\pi T}}$, where $T = \sum_{i=1}^{n} x_i^2/n$. Quinn and Gallucci [9] derived the minimum variance unbiased estimator (MVUE) for the half-normal model in equation (6) given by,

$$\hat{f}_{MV}(0) = \frac{1}{\beta(n)}\sqrt{\frac{2}{\pi T}} \text{ where, } \beta(n) = \frac{\Gamma((n-1)/2)}{\Gamma(n/2)}$$

Based on half-normal model in equation (6), Zhang [10] proposed the shrinkage estimator(SH)to estimate $f(0)$ given as

$$\hat{f}_{SH}(0) = \frac{n-2}{n}\beta(n)\sqrt{\frac{2}{\pi T}}.$$

However, the estimator $\hat{f}_{SH}(0)$ is biased for $f(0)$, it achieves the smallest mean square error. There exist a relationship between the estimators $\hat{f}_{SH}(0)$, $\hat{f}_{MV}(0)$ and $\hat{f}_{MLE}$ as

$$\hat{f}_{SH}(0) = \frac{n-2}{n}\beta^2(n)\hat{f}_{MV}(0) = \frac{n-2}{n}\beta(n)\hat{f}_{MLE}(0)$$

The function $\beta(n) \to 1$ [8] and $\frac{n-2}{n} \to 1$ as $n \to \infty$, then the three estimators are asymptotically equivalent, that means

$$\hat{f}_{SH}(0) = \hat{f}_{MV}(0) = \hat{f}_{MLE}(0).$$

Al-ababned and Eidous [1] suggested detection function defined as

$$g(x;\theta) = e^{-\theta x}\left(2 - e^{-\theta x}\right); x \geq 0, \theta > 0 \quad (7)$$

The model in equation (7) satisfies the shoulder condition. The corresponding pdf is

$$f(x;\theta) = \frac{2\theta}{3}e^{-\theta x}\left(2 - e^{-\theta x}\right); x \geq 0, \theta > 0. \quad (8)$$

Since $g(0;\theta) = 1$ then $f(0;\theta) = \frac{2\theta}{3}$. The model in equation (8) can be expressed in term of $f(0)$ as

$$f(x;\theta) = f(0)e^{-3f(0)x/2}\left(2 - e^{-3f(0)x/2}\right); x \geq 0, \theta > 0 \quad (9)$$

Based on the model in equation (9), Al-ababned and Eidous [1] derived the expected value of x which is $7/9f(0)$. This leads to $\hat{f}_{MME}(0) = 7/9\bar{X}$ as the moments estimator for $f(0)$. They also gave the result $\hat{f}_{MLE}(0) \sim N\left(f(0), \sigma^2_{f(0)}\right)$, where $\sigma^2_{f(0)} = \frac{f^2(0)}{0.80153n}$ and $\hat{f}_{MLE}(0)$ is the MLE of $f(0)$. Saeed [11] proposed the weighted exponential model for grouped and ungrouped line transect data with the detection function

$$g(x;\theta,\beta) = e^{-\theta x}\left(2 - e^{-\theta x/\beta}\right)^{\beta}; x \geq 0, \theta, \beta > 0 \quad (10)$$

The corresponding pdf given as

$$f(x;\theta,\beta) = \frac{2\theta\Gamma(\beta+1/2)}{\sqrt{\pi}\Gamma(\beta+1)+\Gamma(\beta+1/2)}e^{-\theta x}\left(2 - e^{-\theta x/\beta}\right)^{\beta}; x \geq 0, \theta, \beta > 0. \quad (11)$$

The parametric model in equation (11) is proposed to estimate the population density based on line transect data. $g(0;\theta,\beta) = 1$ and $g'(0;\theta,\beta) = 0$, so that the $g(x;\theta,\beta)$ satisfies the shoulder condition assumption for any values of θ. A simulation study is conducted to study the performances of the proposed model. The results show the good statistical properties of the model compared to some other models that exist in the literature. Based on equation (11) where the model in equation (8) is considered special case of the model in equation (11) where $\beta = 1$. Al Eibood and Eidous [2] studied the model in equation (8) and the data was assumed to be a grouped data, the MLE for $f(0)$ and the population abundance D which are given as

$$\hat{f}_G(0) = \frac{2\hat{\theta}}{3} \text{ and } \hat{D}_G(0) = \frac{n\hat{f}_G(0)}{2L}.$$

The MLE is calculated numerically with respect to θ by solving the equation

$$\sum_{i=1}^{n} n_i\left[\frac{c_i e^{-\theta c_i}\left(e^{-\theta c_i} - 2\right) - c_{i-1}e^{-\theta c_{i-1}}\left(2 - e^{-\theta c_{i-1}}\right)}{e^{-2\theta c_i} - e^{-2\theta c_{i-1}} + 4e^{-\theta c_{i-1}} - 4e^{-\theta c_i}}\right] = 0.$$

The MLE of θ which leads to estimate the $f(0)$ and D is shown in Table 1. In addition, the proposed estimator for grouped data is closed to the proposed estimator for ungrouped data [1] which gives $\hat{f}(0) = 0.12647$ and $\hat{D} = 0.0043$ stakes/m^2 = 43 stakes/hectare.

Table 1. The summary of proposed estimator for ungrouped data [2].

Interval Number	$\hat{\theta}$	$\hat{f}_G(0)$	$\hat{D}_G$
6	0.203346	0.135564	0.004609 $stakes/m^2$
8	0.189965	0.126643	0.004306 $stakes/m^2$
10	0.127708	0.127708	0.004342 $stakes/m^2$

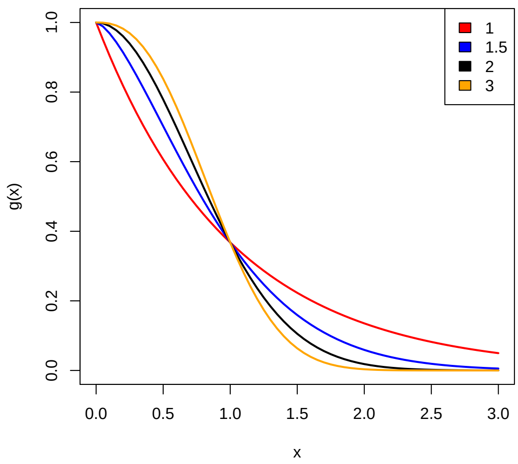

Figure 1. The shape of detection function $g(x)$ of the suggested model; $\sigma^2 = 1, 1.5, 2, 3$.

2. The Proposed Model

Let $X_1, ..., X_n$ be perpendicular distances (assumed to be independent and identically distributed), $g(x;\sigma^2)$ is the detection function in the equation (12), where σ^2 is an unknown parameter. The detection function $g(x;\sigma^2)$ is proposed to be

$$g(x;\sigma^2) = e^{-x^2/2\sigma^2}\left(2 - e^{-x^2/2\sigma^2}\right); x \geq 0, \sigma^2 > 0 \tag{12}$$

The shape of this detection function for $\sigma^2 = 1, 1.5, 2, 3$ is presented in Figure 1.
Based on this detection function, the probability of detecting an object given that its perpendicular distance on the transect line is one (i.e. $g(0;\sigma^2) = 1$), which indicates that the probability of detection on the line transect center is certain. The first derivative of the detection function $g(x;\sigma^2)$ with respect to x is,

$$\acute{g}(x;\sigma^2) = \frac{2}{\sigma^2} x e^{-x^2/2\sigma^2}\left(e^{-x^2/2\sigma^2} - 1\right),$$

which gives $g'(0;\sigma^2) = 0$ and the $g(x;\sigma^2)$ satisfies the shoulder condition assumption for any values of σ^2. The pdf of $g(x;\sigma^2)$ can be obtained by normalizing the detection function as $f(x;\sigma^2) = \frac{1}{\mu} g(x;\sigma^2)$ where $\mu = \int_0^\infty g(x;\sigma^2)dx$. Then the pdf is given by,

$$f(x;\sigma^2) = \frac{2}{(2\sqrt{2}-1)\sigma\sqrt{\pi}} e^{-x^2/2\sigma^2}\left(2 - e^{-x^2/2\sigma^2}\right); x \geq 0, \sigma^2 > 0. \tag{13}$$

Since $g(0;\sigma^2)=1$, the parameter $f(0;\sigma^2)$ is given by

$$f(0;\sigma^2)=\frac{2}{(2\sqrt{2}-1)\sigma\sqrt{\pi}}$$

The detection function $g(x;\sigma^2)$ in equation (12) is monotonically decreasing. The $g'(0;\sigma^2)=0$. Now, $g(x;\sigma^2)$ is monotonically decreasing for $x\in(0,\infty)$, if and only if,

$$g'(x;\sigma^2)=\frac{2}{\sigma^2}xe^{-x^2/2\sigma^2}\left(e^{-x^2/2\sigma^2}-1\right)<0,$$

since $0\le e^{-x^2/2\sigma^2}\le 1, \forall x\in(0,\infty)$, then $e^{-x^2/2\sigma^2}-1\le 0$, which is true for all $x>0$ and σ^2. Accordingly $f(x;\sigma^2)$ is monotonically decreasing for $x\in(0,\infty)$. We can introduce the model in equation (13) as term of $f(0)$ as

$$f(x;\sigma^2)=f(0)e^{-((2\sqrt{2}-1)\sqrt{\pi}f(0))^2x^2/8}\left(2-e^{-((2\sqrt{2}-1)\sqrt{\pi}f(0))^2x^2/8}\right); x\ge 0. \tag{14}$$

3. Statistical Properties

In this section, We obtain the 1^{st} and 2^{nd} moments for the model in equation (12) to derive some properties for the proposed model as.

$$E(x)=\frac{3\sigma}{(2\sqrt{2}-1)\sqrt{\pi}}, \text{ and } E(x^2)\left(\frac{4\sqrt{2}-1}{2(2\sqrt{2}-1)}\right)\sigma^2.$$

The variance of x is form as

$$Var(x)=\left(\frac{(17-6\sqrt{2})\pi-18}{2(2\sqrt{2}-1)^2\pi}\right)\sigma^2$$

4. Statistical Inference

In this section, the statistical inference about the parameter $f(0)$ to be considered, the method of moment estimator (MME), maximum likelihood estimator (MLE) and the asymptotic distribution are discussed. The moments method can be applied to estimate the parameter $f(0)$ as closed form while the maximum likelihood estimator for $f(0)$ must be calculated numerically.

4.1. Moments Estimators for $f(0)$

According to the model in equation (14) the first moment is given as,

$$E(x)=\frac{6}{(2\sqrt{2}-1)^2\pi f(0)}, \text{ which leads to estimate } f(0) \text{ as } \hat{f}(0)=\frac{6}{(2\sqrt{2}-1)^2\pi\bar{X}},$$

where $\hat{f}(0)$ is MME for $f(0)$.

4.2. Maximum Likelihood Estimators for $f(0)$

The MLE of $f(0)$ can be calculated using the MLE of σ^2 which in equation (13) or can be obtained directly using the log maximum likelihood function for the model in equation (14). The log maximum likelihood function for the model in equation (14) is given by

$$\ln L(f(0))=n\ln(f(0))-\frac{(2\sqrt{2}-1)^2\pi f^2(0)}{8}\sum_{i=1}^{n}x_i^2+\sum_{i=1}^{n}\ln\left(2-e^{-(2\sqrt{2}-1)^2\pi f^2(0)x_i^2/8}\right). \tag{15}$$

We obtain the MLE of $f(0)$ by solving the equation,

$$\frac{n}{f(0)} - \frac{(2\sqrt{2}-1)^2\pi f(0)}{4}\sum_{i=1}^{n} x_i^2 + \frac{(2\sqrt{2}-1)^2\pi f(0)}{4}\sum_{i=1}^{n} \frac{x_i^2}{\left(2e^{(2\sqrt{2}-1)^2\pi f^2(0)x_i^2/8}-1\right)} = 0. \quad (16)$$

We can solve equation (16) using numerical methods such as the Newton-Raphson method. The MLE of $f(0)$ is consistent estimator and asymptotically normal to $N\left(f(0), I^{-1}\left(f(0)\right)\right)$, where $I(f(0))$ is called fisher information. The fisher information can be computed and is given by

$$I(f(0)) = -E\left(\frac{\partial^2 \ln L(f)}{\partial f^2}\right) = \frac{4.88758n}{f^2(0)}$$

Therefore, the MLE for $f(0)$ is a consistent estimator and asymptotically normal with mean $f(0)$ and variance $\frac{f^2(0)}{4.88758n}$.

5. Conclusion

This paper shows a primary results for the proposed model. Under the detection function assumption, the proposed model is monotonically decreasing with the perpendicular distance X and it satisfies the shoulder condition. The primary results lead us to make the simulation to examine the performance of the proposed model . In order to use this model for estimating the population density.

References

[1] Ababneh F and Eidous O M 2012 *Journal of Modern Applied Statistical Methods* **11** 144-151
[2] Al Eibood F and Eidous O M 2017 *Mathematics and Statistics* **5** 1-4
[3] Buckland S T, Anderson D R, Burnham K P, Laake J L, Borchers D L and Thomas L 2001 *Introduction to Distance Sampling: Estimating Abundance of Biological Populations* (Oxford: Oxford University Press)
[4] Burnham K P and Anderson D R 1976 *Biometrics* **32** 325-336
[5] Burnham K P, Anderson D R and Laake, J L 1980 *Wildlife Monograph* **72** 3-202.
[6] Gates C E, Mashall W H and Olson, D P 1968 *Biometrics* **24** 135-145
[7] Hemingway P 1971 *Field trials of the line transect method of sampling large populations of herbivores: The scientific management of animal and plant communities for conservation* (Blackwell scientific. Publ. Oxford.) pp 405-411.
[8] Magnus W Oberhettinger F and Soni R P 1967 *Formulas and theorems for the special functions of mathematical physics* (New-York and Berlin: Spring-Verlag)
[9] Quinn T J and Gallucci V F 1980 *Ecology* **61** 293-302
[10] Zhang S 2009 *Environmental and Ecological Statistics* **18** 79-92
[11] Saeed G A A 2013 *New Parametric Model for Grouped and Ungrouped Line Transect Data* Master Thesis Department of statistics, Yarmouk University