Journal of Statistics & Management Systems
ISSN 0972-0510 (Print), ISSN 2169-0014 (Online)
Vol. 28 (2025), No. 4, pp. 649–669
DOI : 10.47974/JSMS-1314

A notable two-parameter detection model for the line transect distance sampling

Hassan S. Bakouch [§]
Department of Mathematics
College of Sciences
Qassim University
Buraydah
Saudi Arabia

and

Department of Mathematics
Faculty of Sciences
Tanta University
Tanta
Egypt

Rawda I. Abdullaa [†]
Department of Mathematics
Faculty of Sciences
Tanta University
Tanta
Egypt

Christophe Chesneau *
Department of Mathematics
Laboratoire de Mathematiques Nicolas Oresme
University of Caen
Caen 14032
France

[§] *E-mail:* **hassan.bakouch@science.tanta.edu.eg**
[†] *E-mail:* **rawdaibrahim44@yahoo.com**
* *E-mail:* **christophe.chesneau@gmail.com (Corresponding Author)**

Abouzar Bazyari ‡
Department of Statistics
Faculty of Intelligent Systems Engineering and Data Science
Persian Gulf University
Bushehr
Iran

Abstract

Line transect sampling is a technique used to estimate the population abundance of animals, plants, birds or other objects in a given region. It is usually the most recommended method among ecologists due to its practicality, efficiency and relatively low cost. This paper contributes to the topic by proposing a new two-parameter parametric detection model designed to model line transect data. Its main properties are discussed, including a detailed analysis of the shape of the main function and the moments. Maximum likelihood estimation of the parameters and population abundance is carried out. In addition, some applications of the proposed model are investigated using two practical data sets of perpendicular distances. Based on the study of some goodness-of-fit statistics, they are compared with a set of classical and current models. The results are in favor of the new model.

Subject Classification (2010): *60E05, 62F10.*

Keywords: *Detection function, Abundance estimation, Line transect distance sampling, Shoulder condition, Statistical ecology, Data analysis.*

1. Introduction

Estimating the population abundance (density), denoted by D, of wildlife populations and the spatial distribution of objects in an area is often of biological importance and essential for the management. These objects can be animals, trees, nests or anything of interest that can be easily seen at close range. One of the most commonly used methods for determining population abundance is the line transect distance sampling method. Because of the amount of experience with this approach and its simplicity, it is the most popular among ecologists (see, for instance,[17], [6], [5] and [7]).

In line transect sampling, non-overlapping lines (at least one line) of length L are randomly placed in a region of a certain area A, independent of the population, such that the objects are equally distributed in distance from the lines. An observer traverses the lines and records the perpendicular distances $(X_i,\ i=1,\ldots,n)$ of each sighted object. Observations can be

‡ *E-mail:* `ab_bazyari@pgu.ac.ir`

either one-sided (only objects to the north or only objects to the south are observed), two-sided, or on the lines. Two-sided observations are the more common case (see [10]). The detection function (df), which expresses the probability of an object of interest as a function of its distance from the line, is the fundamental concept in assessing detectability in distance sampling. Mathematically, this function is expressed as $g(x) = g(x;\theta) =$ Probability of an object is observed at random/its perpendicular distance is x, where $0 < x < w$, θ is a vector of parameters to be estimated (see [6]), and the truncated point pointing to the largest perpendicular distance at which no objects are identified is denoted by w (this truncation may occur during the analysis or during the survey).

To calculate population abundance D, we need a conceptual model for the df that relates the data to the estimated population abundance parameters. Therefore, the df $g(x)$ is expected to have various assumptions to judge its usefulness. Typically, it's unrealistic for objects to become more detectable with increasing distance. As a result, objects are more likely to be missed as the distance from the transect line increases. So the df $g(x)$ should have a monotone, non-increasing shape with increasing distance. Mathematically, if $x_1 > x_2$ then we have $g(x_2) > g(x_1)$. Furthermore, objects on the transect path are detected with probability one (i.e., $g(0) = 1$, perfect detection along the entire path), and the detectability should be extremely close to one in a very small neighbourhood of zero (see [11]). It is also assumed that $g'(0) = 0$, where $g'(x)$ is the first derivative of g with respect to $g(x)$, which implies that the slope of the tangent at $x = 0$ should be zero. This corresponds to the shape criterion proposed in [8]. Any acceptable model for the df must include the above assumptions, often referred to as the "shoulder condition".

[23], [6] and [9] assumed that $f(x)$ is a conditional probability density function (pdf) of perpendicular distance data. The pdf is related to $g(x)$ as follows:

$$f(x) = \frac{1}{C} g(x), \tag{1}$$

for $0 < x < w$, where $C = \int_0^w g(x)\mathrm{d}x$, is the normalization constant, see [11] and [8]. They showed that $f(x)$ satisfies the shoulder condition, i.e., $f'(0) = 0$, and thus has a simple monotone non-increasing form. Also, the shapes of $f(x)$ and $g(x)$ are identical, but $f(x)$ is scaled to make the area under its curve equal to one. In this paper, for sake of simplicity, we suppose that $w = \infty$. According to [4] and [7], the following characteristics hold for line transect sampling:

1. Interested objects of total size (actual population size) N are randomly distributed over a study zone of area A. The corresponding average population abundance is given by $D = N / A$.
2. Detection is perfect on the line (at zero distance) with probability 1, i.e., $g(0) = 1$.
3. There is no movement of the objects, and they are detected at their initial sighting position.
4. There are no measurement mistakes in the distances recorded.
5. There are no objects that are identified twice.
6. Not all objects are detected; some, perhaps many, will be missed.
7. The df has a known form, but there are unknown parameters that must be estimated.

A formula is poposed in [9] and [23] to calculate D of objects in any region. It is given by

$$D = \frac{1}{2L} E(n) f(0), \tag{2}$$

where $E(n)$ denote the expected value of the detected objects with sample size n, L is the transect line length and $f(0) = 1/C$ (see [8]). The following formula can be used to calculate the estimated population abundance $\hat{D}$ of the objects:

$$\hat{D} = \frac{1}{2L} n\hat{f}(0), \tag{3}$$

where $\hat{f}(0)$ is an estimate for $f(0)$. Several parametric techniques for estimating $f(0)$ and associated parameters have been presented, including the assumption that $f(x)$ belongs to a family of known functional forms and that unknown parameters must be estimated from perpendicular distances. Some non-parametric techniques have also been presented in the literature. Hence, an estimation for $f(0)$ and, in turn, D are obtained. More information on this methodology can be found in the following papers: [4], [8], [22], [20] and [25].

In this paper, we introduce and study a remarkable two-parameter detection model (TPM) that satisfies the previously described shoulder requirement. Furthermore, the TPM has a special case detection model that has been published in the literature. In the application section, we show the ability of the TPM to use its corresponding pdf for modeling and analyzing practical data with other existing competitors.

The paper is organized as follows: In Section 2, we introduce the TPM, and its corresponding detection pdf is obtained and discussed. Moments and some related measures of the TPM are discussed in Section 3. In Section 4, a complete maximum likelihood (ML) estimation study is performed. Section 5 illustrates the performance of the TPM on two practical data sets and derives the associated metrics. The results of the paper are summarized in Section 6.

2. The proposed detection model

Suppose that we need an estimator, $\hat{D}$, of the object density based on the set, $x_1, x_2, \ldots, x_n$, of observed perpendicular distances. These distances are identically and independently distributed random variables. As we do not know the true shape of the df, we need a flexible model that can fit a variety of shapes (see [2]). Adjustment terms' aim is to give this flexibility. Furthermore, it is well known that the shape of any one-parameter model based on $g(x)$ may not be sufficiently flexible. As a result, the use of such models imposes a constraint on the shape of the detection curve, potentially leading to non-robust estimators. We therefore propose a remarkable two-parameter detection model, TPM, with a df given by

$$g(x) = g(x;\beta,\lambda) = [1+\beta(1-e^{-\lambda x})]^{1/\beta} e^{-\lambda x}, 0 \le x < \infty, \beta, \lambda > 0. \tag{4}$$

Based on the df given by Equation (4), the probability of detecting an object on the transect line path is one (i.e., $g(0;\beta,\lambda)=1$), which indicates that objects with zero distances will sure be seen. Furthermore, we have

$$g'(x) = -\frac{\lambda(1+\beta)(e^{\lambda x}-1)[1+\beta(1-e^{-\lambda x})]^{1/\beta} e^{-\lambda x}}{(1+\beta)e^{\lambda x}-\beta},$$

which gives $g'(0) = g'(0;\beta,\lambda) = 0$. This means that $g(x)$ satisfies the shoulder condition, which is seen in a variety of forms given by Figure 1. The df in Equation (4) can be rewritten by the following product:

$$g(x) = g_1(x) g_2(x),$$

where $g_1(x) = [1+\beta(1-e^{-\lambda x})]^{1/\beta}$ and $g_2(x) = e^{-\lambda x}$, and β, and λ are unknown parameters. The function $g_1(x)$ is a weighted increasing function such that $g_1(x) \ge 1$, implying the ordering $g(x) \ge g_2(x)$. It is constructed from a power transformation of the cumulative distribution function (cdf) of the exponential distribution such that, when $\beta \to 0$, $g_1(x)$ tends to $g_*(x) = e^{1-e^{-\lambda x}}$. Using a power transformation of the cdf of the exponential distribution to enhance the capabilities of the exponential model is not novel. The most popular example is the exponentiated exponential distribution (see [18]). Also, $g_2(x)=e^{-\lambda x}$ is a known df; it corresponds to the one of the negative exponential model (see [13]). By normalizing the df, the pdf associated with the TPM is as follows:

$$f(x)=f(x;\beta,\lambda)=\frac{\lambda(1+\beta)}{(1+\beta)^{1+1/\beta}-1}[1+\beta(1-e^{-\lambda x})]^{1/\beta}e^{-\lambda x}, 0\le x<\infty, \beta,\lambda>0. \tag{5}$$

The cdf related to $f(x)$ is obtained by

$$F(x)=\frac{[1+\beta(1-e^{-\lambda x})]^{1+1/\beta}-1}{(1+\beta)^{1+1/\beta}-1}. \tag{6}$$

On can remark that, when $\beta\to 0$, $F(x)$ tends to

$$F_*(x)=\frac{e^{1-e^{-\lambda x}}-1}{e-1},$$

which is the cdf of a special case of the alpha-G family (with $\alpha=e$) defined with the exponential distribution as baseline (see [16] and [19]). In this sense, the proposed TPM can be viewed as a generalized version, with the alpha-exponential distribution as a limited case. Since $g(0)=1$, we have

$$f(0)=\frac{1}{C}=\frac{\lambda(1+\beta)}{(1+\beta)^{1+1/\beta}-1}. \tag{7}$$

Then, from Equation (7), we get $\lambda=f(0)\ [(1+\beta)^{1+1/\beta}-1]/(1+\beta)$. We can introduce the pdf of the TPM in terms of β and $f(0)$ by substituting the previous value of λ as

$$f(x;\beta,f(0))=f(0)[1+\beta(1-e^{-[f(0)/(1+\beta)][(1+\beta)^{1+1/\beta}-1]x})]^{1/\beta}e^{-[f(0)/(1+\beta)][(1+\beta)^{1+1/\beta}-1]x}. \tag{8}$$

In the process of the ML estimation, the pdf in Equation (8) in terms of the parameters $f(0)$ and β will be the main base. We have

$$f'(x)=-\frac{\lambda^2(1+\beta)^2(1-e^{-\lambda x})[1+\beta(1-e^{-\lambda x})]^{1/\beta}}{[\beta-(1+\beta)e^{\lambda x}][1-(1+\beta)^{1+1/\beta}]}.$$

In the equation above, it is clear that terms in both the numerator and denominator are positive, implying that $f'(x)<0$. Hence, $f(x)$ is a monotone, non-increasing function. Also, from Equation (1), it is immediate that $f'(0)=0$ for all $\beta,\lambda>0$ and $f(x)$ has a monotone non-increasing shape. This can be observed in Figure 2.

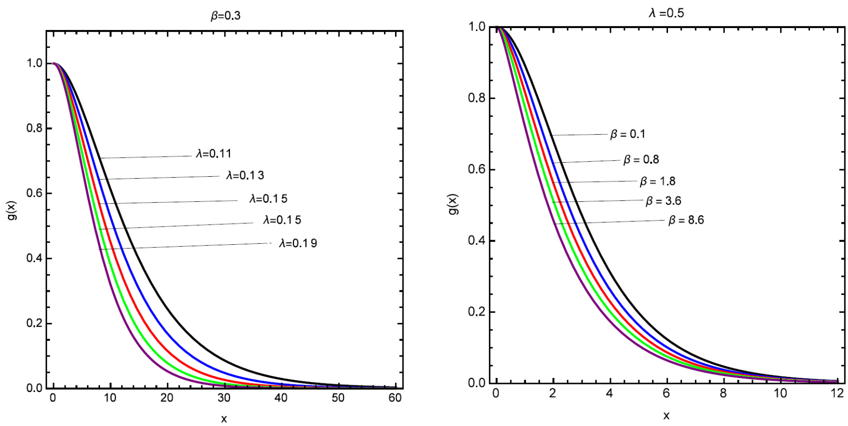

Figure 1

Plots of the df in Equation (4) for some values of parameters.

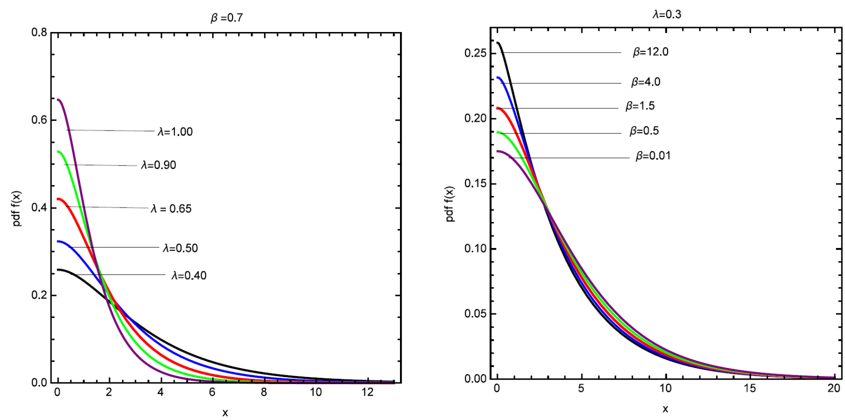

Figure 2

Plots of the pdf in Equation (5) for some values of parameters.

Figures 1 and 2 show that $g(x)$ and $f(x)$, can take a variety of shapes. The probability of detecting objects remains at or close to one at the center of the line or at small distances from it. In addition, at large distances, the curves fall off smoothly at the center of the line. The functions $g(x)$ and $f(x)$ level off close to zero as $x \to \infty$. So the proposed TPM has a proper df.

If $\hat{\beta}$ and $\hat{\lambda}$ are the estimators of the parameters β and λ, respectively, by substituting these estimators in $f(0) = f(0; \beta, \lambda)$, we obtain

$$\hat{f}(0) = f(0, \hat{\beta}, \hat{\lambda}) = \frac{\hat{\lambda}(1+\hat{\beta})}{(1+\hat{\beta})^{1+1/\hat{\beta}} - 1}. \tag{9}$$

The population abundance (density) estimator is obtained by plugging Equation (9) into Equation (3) as

$$\hat{D} = \frac{n.f(0,\hat{\beta},\hat{\lambda})}{2L} = \frac{n\hat{\lambda}(1+\hat{\beta})}{2L[(1+\hat{\beta})^{1+1/\hat{\beta}} - 1]}. \tag{10}$$

The parameters β and λ will be estimated using the ML approach. The population abundance D and $f(0)$ are thus estimated. The specifics will be covered in Section 4.

Remark: We note that the weighted exponential detection model (WEM) introduced in [1] is a sub-model. We obtain it by taking $\beta = 1$ in the df in Equation (4). The df and associated pdf of this sub-model are defined by

$$g(x) = (2 - e^{-\lambda x})e^{-\lambda x}, \tag{11}$$

and

$$f(x) = \frac{2\lambda}{3}(2 - e^{-\lambda x})e^{-\lambda x}, \lambda > 0, x \geq 0, \tag{12}$$

respectively.

3. Some statistical features

Moments, quantiles, skewness, and kurtosis are some of the basic statistical features for a distribution. Let X be a random variable with the pdf in Equation (5). It thus model the perpendicular distance with the distribution underlying the TPM. Then, the rth moment of X is

$$\mu_r' = E(X^r) = \frac{r!(\beta+1)^{1/\beta+1}}{\lambda^r[(\beta+1)^{1/\beta+1} - 1]} {}_{(r+2)}F_{(r+1)}\left(1,1,\ldots,1,-\frac{1}{\beta};2,2,\ldots,2;\frac{\beta}{\beta+1}\right), \tag{13}$$

where ${}_pF_q$ is the generalized hypergeometric function and $p > q$ are non-negative integers. This function could be expressed by

$${}_pF_q(\alpha_1, \alpha_2, \ldots, \alpha_p; \beta_1, \beta_2, \ldots, \beta_q; x) = \sum_{k=0}^{\infty} \frac{(\alpha_1)_k(\alpha_2)_k \ldots (\alpha_p)_k}{(\beta_1)_k(\beta_2)_k \ldots (\beta_q)_k} \frac{x^k}{k!},$$

such that $p = r+2$, $q = r+1$, $\alpha_1 = \alpha_2 = \ldots = \alpha_{r+1} = 1$, $\alpha_{r+2} = \alpha_p = -1/\beta$, $z = \beta / (\beta+1)$, $\beta_1 = \beta_2 = \ldots = \beta_q = 2$ and $(a)_k = \Gamma(a+k) / \Gamma(k) = a(a+1)(a+2) \ldots (a+k-1)$ is the rising factorial.

In particular, using Equation (13), the first two moments are

$$\mu = \mu_1' = \frac{(\beta+1)^{1/\beta+1}}{\lambda[(\beta+1)^{1/\beta+1}-1]}\,{}_3F_2\left(1,1,-\frac{1}{\beta};2,2;\frac{\beta}{\beta+1}\right),$$

and

$$\mu_2' = E(X^2) = \frac{2(\beta+1)^{1/\beta+1}}{\lambda^2[(\beta+1)^{1/\beta+1}-1]}\,{}_4F_3\left(1,1,1,-\frac{1}{\beta};2,2,2;\frac{\beta}{\beta+1}\right), \tag{14}$$

respectively.

Based on $F(x)$ in Equation (6), the quantile x_q of the corresponding pdf of the TPM is the real solution of $F(x_q) = q$. After some algebraic manipulations, we get

$$\ln[1+\beta(1-e^{-\lambda x_q})] = \frac{\ln(q)+\ln[(1+\beta)^{1+1/\beta}-1]}{(1+1/\beta)}. \tag{15}$$

The above equation has no closed form solution for x_q so we could use a numerical technique to get the quantiles. Setting $q = 0.5$ in Equation (15), we can get the median. Based on Equation (13), the corresponding skewness and kurtosis are obtained using the following expressions:

Table 1

Moment measures of *X* for various values of β and λ.

Parameter	$\beta = 1$			
λ	Mean	Variance	Skewness	kurtosis
1.2	0.972	0.791	1.790	7.858
1.4	0.833	0.581	1.790	7.858
1.6	0.729	0.445	1.790	7.858
1.8	0.648	0.312	1.790	7.858
2.0	0.583	0.285	1.790	7.858
2.4	0.486	0.198	1.790	7.858
Parameter	$\lambda = 1$			
β	Mean	Variance	Skewness	kurtosis
5.0	1.074	1.056	1.907	8.490
6.0	1.066	1.049	1.918	8.550
7.0	1.059	1.044	1.927	8.598
8.0	1.054	1.040	1.934	8.635
9.0	1.049	1.035	1.940	8.667
10.0	1.045	1.033	1.944	8.693

$$\text{Skewness} = \frac{\mu_3 - 3\mu_2'\mu + 2\mu^3}{\sigma^3}, \quad \text{Kurtosis} = \frac{\mu_4' - 4\mu_3'\mu + 6\mu_2'\mu^2 - 3\mu^4}{\sigma^4},$$

respectively. Table 1 indicates some values of the moment measures of X for various values of β and λ.

From this table, we see that the TPM is right skewed and leptokurtic, as anticipated in view of the graphics of the pdf.

4. Parametric estimation

In this section, we develop the method of ML in the context of the TPM.

Let $x_1, x_2, \ldots, x_n$ be realizations of a random sample of perpendicular distances of size n with pdf given by Equation (8), so $f(x) = f(x;\beta, f(0))$. The ML estimator (MLE) of λ can be computed by using the MLEs $\hat{\beta}$ and $\hat{f}(0)$ of β and $f(0)$, respectively (see [8]). The corresponding log-likelihood function is

$$\begin{aligned}\ln L = \sum_{i=1}^{n} \ln[f(x_i)] = n\ln[f(0)] - \frac{f(0)}{1+\beta}[(1+\beta)^{1+1/\beta} - 1]\sum_{i=1}^{n} x_i \\ + \frac{1}{\beta}\sum_{i=1}^{n} \ln[1+\beta(1-e^{-[f(0)/(1+\beta)][(1+\beta)^{1+1/\beta}-1]x_i})]. \end{aligned} \tag{16}$$

When we differentiate the function in Equation (16) with regard to β and $f(0)$, we obtain

$$\begin{aligned}\frac{\partial}{\partial\beta}\ln L = \left[\frac{-f(0)[\beta+(1+\beta)^{1+1/\beta}]}{\beta(1+\beta)^2} + \frac{f(0)(1+\beta)^{1/\beta}\ln[1+\beta]}{\beta^2}\right]\sum_{i=1}^{n} x_i \\ + \frac{1}{\beta}\sum_{i=1}^{n} \frac{1-e^{-\Phi x_i} - \beta e^{-\Phi x_i}\left[\frac{\Phi x_i}{1+\beta} - f(0)(1+\beta)^{1/\beta}\left[\frac{1}{\beta} - \frac{\ln(1+\beta)}{\beta^2}\right]x_i\right]}{1+\beta(1-e^{-\Phi x_i})} \\ - \frac{1}{\beta^2}\sum_{i=1}^{n} \ln[1+\beta(1-e^{-\Phi x_i})], \end{aligned} \tag{17}$$

and

$$\frac{\partial}{\partial f(0)}\ln L = \frac{n}{f(0)} - \frac{\Phi}{f(0)}\sum_{i=1}^{n} x_i + \frac{\Phi}{f(0)}\sum_{i=1}^{n} \frac{x_i e^{-\Phi x_i}}{1+\beta(1-e^{-\Phi x_i})}, \tag{18}$$

where $\Phi = f(0)\ [(1+\beta)^{1+1/\beta} - 1] \ / \ (1+\beta)$. The MLEs $\hat{\beta}$ and $\hat{f}(0)$ of β and $f(0)$, respectively, are the numerical solution of the non-linear equations $\partial \ln L \ / \ \partial\beta = 0$ and $\partial \ln L \ / \ \partial f(0) = 0$. They can be computed by using Mathematica, R, Matlab, or any other statistical software. The estimator $\hat{\lambda}$

is obtained by substituting the values of $\hat{f}(0)$ and $\hat{\beta}$ in Equation (9), then $\hat{D}$ is obtained from Equation (10). The Fisher information matrix $I(\beta, f(0)) = [I_{ij}],\ i, j = 1, 2$ is given by

$$I\left(\beta, f(0)\right) = -\begin{pmatrix} I_{11} & I_{12} \\ I_{21} & I_{22} \end{pmatrix},$$

where $I_{11} = \partial^2 \ln L / \partial\beta^2$, $I_{22} = \partial^2 \ln L / \partial(f(0))^2$, and $I_{12} = I_{21} = \partial^2 \ln L / (\partial\beta\partial f(0))$. On request, the authors can provide the elements of this matrix. As a result, the variance covariance matrix is I^{-1}. The asymptotic distributions of $\hat{\beta}$, $\hat{\lambda}$ and $\hat{f}(0)$ are $N(\beta, \widehat{\mathrm{Var}}(\hat{\beta}))$, $N(\lambda, \widehat{\mathrm{Var}}(\hat{\lambda}))$ and $N(f(0), \widehat{\mathrm{Var}}(\hat{f}(0)))$, respectively, where $\widehat{\mathrm{Var}}(\hat{\beta})$, $\widehat{\mathrm{Var}}(\hat{\lambda})$ and $\widehat{\mathrm{Var}}(\hat{f}(0))$ are the MLEs of $\mathrm{Var}(\hat{\beta})$, $\mathrm{Var}(\hat{\lambda})$ and $\mathrm{Var}(\hat{f}(0))$, respectively. Based on these results, the asymptotic $(1-\delta)\ 100\%$ confidence intervals (CIs) for β, λ, and $f(0)$ are provided by

$$\hat{\beta} \pm Z_{\delta/2}\sqrt{\widehat{\mathrm{Var}}(\hat{\beta})}, \qquad \hat{\lambda} \pm Z_{\delta/2}\sqrt{\widehat{\mathrm{Var}}(\hat{\lambda})},$$

and

$$\hat{f}(0) \pm Z_{\delta/2}\sqrt{\widehat{\mathrm{Var}}(\hat{f}(0))},$$

where $Z_{\delta/2}$ is classically the upper $\delta/2$ quantile of the standard normal distribution.

The asymptotic distributions of $\hat{D}$ is $N(D, \widehat{\mathrm{Var}}(\hat{D}))$, where $\widehat{\mathrm{Var}}(\hat{D})$ is the asymptotic sampling variance for $\hat{D}$. According to [8] and [6], the asymptotic sample variance for $\hat{D}$ is

$$\mathrm{Var}(\hat{D}) = D^2\{[Cv(n)]^2 + [Cv(\hat{f}(0))]^2\},$$

where n is the number of detected perpendicular distances. As a result, the estimated sampling variance of $\hat{D}$ is given by

$$\widehat{\mathrm{Var}}(\hat{D}) = \hat{D}^2\{[\widehat{\mathrm{Cv}}(n)]^2 + [\widehat{\mathrm{Cv}}(\hat{f}(0))]^2\}, \tag{19}$$

where $\widehat{\mathrm{Cv}}(n)$ denotes the estimated coefficient of variations that are given for n and $\hat{f}(0)$ as $[\widehat{\mathrm{Cv}}(n)]^2 = \widehat{\mathrm{Var}}(n) / n^2$ and $[\widehat{\mathrm{Cv}}(\hat{f}(0))]^2 = \widehat{\mathrm{Var}}(\hat{f}(0)) / [\hat{f}(0)]^2$, respectively. Hence, the estimation of $Var(\hat{D})$ requires the estimation of the sampling variance of both $\hat{f}(0)$ and n. Using the sample of perpendicular distances, $\hat{f}(0)$ and $\widehat{\text{Var}}(\hat{f}(0))$ are obtained. Also, because of the probabilistic nature of the detection process, n itself is a random variable (see [8]). For $\hat{D}$, the approximate $(1-\delta)100\%$ CIs is

$$\hat{D} \pm Z_{\delta/2}\sqrt{\widehat{\text{Var}}(\hat{D})}.$$

Estimating $f(0)$ and $Var(\hat{f}(0))$ can be obtained as MLEs, but the major issue is still determining $\text{Var}(n)$. As a result, the estimation of the sampling variance $\text{Var}(n)$ is based on the probability mass function (pmf) of n, say $k(n)$, which is a function of the transect lines and the research area's objects (see [22] and [8]). If the objects are dispersed randomly in the research region, then $k(n)$ is a binomial pmf, according to fundamental principles. As a result, we have $\text{Var}(n) = NP(1-P)$, where $P = n/N = 2L\hat{C}/A$ represents the probability of seeing an object and N represents the total number of objects dispersed in a particular region (see [9]). As a result, $\widehat{\text{Var}}(n) = n(1-\hat{P}) = n(1-n/N)$ (see [22]), and by employing the preceding connections, we get

$$\widehat{\text{Var}}(n) = n\left(1-\frac{2L\hat{C}}{A}\right) = n\left(1-\frac{2L}{A\hat{f}(0)}\right) = n\left(1-\frac{2L((1+\hat{\beta})^{1+1/\hat{\beta}}-1)}{A\hat{\lambda}(1+\hat{\beta})}\right). \tag{20}$$

Furthermore, if another distribution is assumed for $k(n)$ such as, for instance, the Poisson distribution (see [9]), then $\text{Var}(n) = E(n)$ and $\widehat{\text{Var}}(n) = n.$ [22], [9], [21] and [24] provide further pmfs for the distribution of $k(n)$ and various techniques for calculating $\text{Var}(n)$. The Poisson and binomial distributions are of particular relevance in this work.

5. Practical examples of two data sets

Two practical perpendicular data sets are supplied for comparing the performance of the proposed detection model to that of other detection models using a series of goodness-of-fit tests. The relevant model dfs $g(x)$ and their associated pdfs $f(x)$ are defined as follows:

1. Negative exponential model (NEM) by [13]:

$$g(x) = e^{-\lambda x},$$
$$f(x) = \lambda e^{-\lambda x}, \lambda > 0, x \geq 0.$$

2. (WEM) ([1]) given by Equations (11) and (12).
3. Model (2015) by [12]:

$$g(x) = 4e^{-\theta x}\left(1 - \frac{1}{2}e^{-\theta x/2}\right)^2,$$

$$f(x) = \frac{24\theta}{11}e^{-\theta x}\left(1 - \frac{1}{2}e^{-\theta x/2}\right)^2, \theta > 0, x \geq 0.$$

We begin by analyzing the stakes data (see [8] and [3]). This data set indicates a number of wooden stakes in the sagebrush meadow east of Logan, Utah, with a true density of $D = 0.00375$ stakes per meter. A single path of length $L = 1000$ meters is walked, recording perpendicular distances x_i's, of a sample (number) $n = 68$ stakes from a known population size of $N = 150$ stakes. Therefore, the actual value of $f(0)$ (with regard to n and D, computed from Equation (2)) is 0.11029 (see [15] and [26]), and the above mentioned region is of area of $A = 40{,}000$ m^2 (calculated from $D = N / A$, where $N = 150$) (see [14]). Table 2 shows such data collection. Table 3 provides some essential statistics of the stakes data set. The Hemingway data (see [8]) on researching numerous ungulates in Africa forms the second data set. A 60000 meters long single line transect was walked, and 73 animals were discovered. The perpendicular distances are computed using the sighted distances and sighted angles. In the Hemingway data set, the detected objects' area A and the true density D are not given (not recorded during the survey). However, an estimation procedure is produced by the TRANSECT program that adapts in turn the estimates of $f(0)$ and D, which are, respectively, 0.0065 and 0.0396 animal per hectare (or 3.096×10^{-6}) animal per meter (see [8]). Table 4 lists these perpendicular distances in ascending order.

Table 2

Wooden stakes data

2.02	2.90	11.82	4.85	3.17	15.24	1.27	9.10	1.23	4.97
0.45	8.16	14.23	1.47	7.10	3.47	13.72	3.25	1.67	3.17
10.40	6.47	2.44	18.60	10.71	3.05	6.25	8.49	4.53	7.67
3.61	5.66	1.61	0.41	3.86	7.93	3.59	6.08	3.12	18.16
0.92	2.95	31.31	0.40	6.05	18.15	9.04	0.40	3.05	4.08
1.00	3.96	6.50	0.20	6.42	10.05	7.68	9.33	6.60	-
3.40	0.09	8.27	11.59	3.79	4.41	4.89	0.53	4.40	-

Table 3

Wooden stakes data descriptive summary.

n	Skewness	Kurtosis	$\frac{Skewness}{Kurtosis}$
68	1.1083	3.8033	0.2914

On the basis of these results, the data are skewed to the right and that the TPM is, a priori, suitable for this situation.

Table 4

The Hemingway data

0	0	0	0	0	0	0	0	8.72	10.5
22.3	26.0	26.0	30.5	30.5	31.7	34.2	35.1	38.0	41.0
42.1	50.8	55.1	58.5	63.6	64.3	65.0	68.8	71.1	71.8
71.9	72.1	73.1	76.6	77.6	78.1	84.0	84.5	86.0	86.0
87.0	90.0	92.3	94.0	96.4	96.4	106.0	115.0	123.0	123.0
129.0	129.0	143.0	143.0	150.0	151.0	153.0	157.0	161.0	164.0
164.0	164.0	166.0	175.0	188.0	193.0	200.0	200.0	246.0	260.0
272.0	378.0	400.0	-	-	-	-	-	-	-

Table 5 shows some statistics of the Hemingway data.

Table 5

The Hemingway data set descriptive summary.

n	Skewness	Kurtosis	$\frac{Skewness}{Kurtosis}$
73	1.36056	5.47687	0.248418

Given the two data sets, it's clear that few objects are identified at extreme distances, so truncation may be acceptable. Table 2 shows such extreme perpendicular distances as $31.31,$ while Table 4 shows such extreme perpendicular distances as $378,$ and $400.$ These outliers cause the $f(0)$ and D estimations to be inaccurate. As a result, removing those values will enhance the accuracy of the estimators ([26] and [8]). For the sake of comparison with the considered data, we use the ML approach to estimate the unknown parameters contained in each detection model's associated pdf. We consider the Kolmogorov-Smirnov (K-S) measure, together with its corresponding p-value ($\mathrm{P-Val}$). We also consider Cramér-von Mises (W^*), and Anderson-Darling (A^*) goodness-of-fit statistics, all based on the MLEs. The collected findings for each data set are displayed in Table 6.

Table 6

The MLEs of the parameters for the considered detection models fitted to the data, as well as values of A^*, W^*, K-S and $\mathrm{P-Val}$.

Wooden stake data							
pdf of the detection models	MLEs		$-\ln L$	A^*	W^*	K-S	$\mathrm{P-Val}$
TPM (β, λ)	8.7×10^{-7}	0.2209	182.211	0.3814	0.0704	0.1132	0.356
Model (2015) (θ)	0.2369	-	182.165	0.3817	0.0707	0.1137	0.352
WEM (θ)	0.2041	-	182.603	0.5170	0.0997	0.1257	0.240
NEM (λ)	0.1745	-	183.987	1.0644	0.2127	0.1582	0.070
Hemingway data							
pdf of the detection models	MLEs		$-\ln L$	A^*	W^*	K-S	$\mathrm{P-Val}$
TPM (β, λ)	1.29×10^{-7}	0.0138	389.271	Indeter	0.2049	0.124	0.225
Model (2015) (θ)	0.0149	-	389.155	Indeter	0.2094	0.125	0.215
WEM (θ)	0.0128	-	389.957	Indeter	0.2500	0.135	0.148
NEM (λ)	0.0109	-	391.338	Indeter	0.4090	0.165	0.043

As it is clear, we can see that the smallest values indicated by the A^*, W^*, and the K-S, and the largest value of the $\mathrm{P-Val}$ are obtained for the pdf associated with the TPM model. On the other side, note that the A^* statistic in the Hemingway data is indeterminate (Indeter) because its formula contains the term $\ln(x_i)$ and some values of such data are zeros. Based on the findings of Table 6, we can conclude that the TPM is better than the WEM and NEM detection models, while it is a powerful competitor to the Model (2015) detection model. We complete this study by calculating the Akaike information criterion (AIC), corrected AIC (CAIC), Bayesian information criterion (BIC), and Hannan–Quinn information criterion (HQIC). The obtained values are contained in Table 7. Since it has model the lowest values, the TPM is the best in this setting.

Table 7

Information criteria for the wooden stakes and Hemingway data sets.

Detection Models	AIC	CAIC	BIC	HQIC
TPM (β, λ)	**357.613**	**358.812**	**360.415**	**359.754**
Model (2015) (θ)	366.330	366.392	368.535	366.392
WEM (θ)	367.052	367.113	369.256	367.924
NEM (λ)	369.723	369.785	371.928	370.596
Hemingway data				
Detection Models	**AIC**	**CAIC**	**BIC**	**HQIC**
TPM (β, λ)	**813.570**	**814.361**	**810.145**	**816.022**
Model (2015) (θ)	816.553	816.609	818.843	817.466
WEM (θ)	817.397	817.454	819.688	818.310
NEM (λ)	819.243	819.299	821.533	820.156

Consequently, the variance-covariance matrices of the MLEs for the TPM for the estimated parameters $(\beta, f(0))$ in Equation (8) and (β, λ) given in Equation (5) are given for the two considered data sets, respectively. For the wooden stakes data, the variance-covariance matrices are

$$V_1(\hat{\beta}, \hat{f}(0)) = \begin{pmatrix} 0.00058189 & 6.17291\times 10^{-6} \\ 6.17291\times 10^{-6} & 0.000187505 \end{pmatrix},$$

and

$$V_2(\hat{\beta}, \hat{\lambda}) = \begin{pmatrix} 0.00058189 & -0.0000162639 \\ -0.0000162639 & 0.000553858 \end{pmatrix}.$$

Also, the variance-covariance matrices for the Hemingway data are

$$V_1(\hat{\beta}, \hat{f}(0)) = \begin{pmatrix} 3.69729\times 10^{-7} & 2.48195\times 10^{-10} \\ 2.48195\times 10^{-10} & 6.82183\times 10^{-7} \end{pmatrix},$$

and

$$V_2(\hat{\beta}, \hat{\lambda}) = \begin{pmatrix} 3.69729\times 10^{-7} & -6.55085\times 10^{-10} \\ -6.55085\times 10^{-10} & 2.01414\times 10^{-6} \end{pmatrix}.$$

The estimated variances of the MLEs are represented by the diagonal elements of these matrices, while the estimated covariances between the MLEs are represented by the other entries. The correlation of the MLEs for the related data set may be obtained using all of these elements. Therefore, the correlations of the MLEs for the wooden stakes data are

$$\rho_{\hat{\beta},\hat{f}(0)} = 0.0186880106,$$
$$\rho_{\hat{\beta},\hat{\lambda}} = -0.0286488,$$

and, for the Hemingway data, they are

$$\rho_{\hat{\beta},\hat{f}(0)} = 4.94197\times 10^{-4},$$
$$\rho_{\hat{\beta},\hat{\lambda}} = -7.59121112\times 10^{-4}.$$

Also, 90% and 95% CIs of the parameters β, λ, and $f(0)$ for both data sets are given, respectively, in Tables 8 and 9. Also, it is noted that the interval width increases as one increases the confidence level. Furthermore, from Equation (19), 90% and 95% CIs of D with respect to $k(n)$, under the binomial and Poisson distributions, respectively, are given in Tables 10 and 11. However, the CIs of D with respect to the Hemingway data under binomial could not be obtained as the total number of objects (N the population size) in the study area of this data set is unknown.

Table 8

CIs of β, λ, and $f(0)$ of the TPM for the wooden stakes data.

CI	β	λ	$f(0)$
90%	[0,0.0309149]	[0.190769,0.25109]	[0.111027,0.146124]
95%	[0,0.0396786]	[0.182219,0.25964]	[0.106053,0.151099]

Table 9

CIs of β, λ, and $f(0)$ of the TPM for the Hemingway data.

CI	β	λ	$f(0)$
90%	[0,0.000779]	[0.119861,0.015624]	[0.0069756,0.0090926]
95%	[0,0.001000]	[0.011471,0.016139]	[0.006675,0.00939268]

Table 10

CIs of D for the wooden stakes data.

CI	binomial	Poisson
90%	[0.0035104,0.0050895]	[0.0034054,0.00519459]
95%	[0.0032865,0.005313]	[0.0031518,0.00544819]

Table 11

CIs of D for the Hemingway data.

CI	Poisson
90%	$[3.79706\times10^{-6}, 5.71014\times10^{-6}]$
95%	$[3.5259\times10^{-6}, 5.9813\times10^{-6}]$

Various descriptive statistics for the wooden stakes and Hemingway data sets with their corresponding theoretical measures of the TPM are shown in Table 12. This include the stabdard deviation (SD), the mean deviation (MD), and the standard error of means (SEM). Note that the statistics and measures are numerically closed.

For both data sets, we calculate the estimated $\hat{f}(0)$ and $\hat{D}$ the absolute difference of the standard deviation $|\widehat{SD}-SD|$, for all the compared models and using their MLEs in Table 6, where $\widehat{SD}$ is the sample standard deviation. The results of $\hat{f}(0)$, $\hat{D}$ and $|\widehat{SD}-SD|$ are summarized in Table 13.

The TPM, as shown in this table, has $\hat{f}(0)$ and $\hat{D}$ that the closest to the actual $f(0)$ and D of the wooden stakes data, which are, respectively, 0.11029 and 0.00375. Also, for the Hemingway data, $\hat{f}(0)$ and $\hat{D}$ of the TPM are the closest to the estimated $f(0)$ and D made using TRNSECT program, which are 0.0065 and 3.096×10^{-6}, respectively, among the compared model.

Table 12

Descriptive statistics for the wooden stakes and Hemingway data sets with the corresponding measures using the pdf of the TPM.

	Mean	Median	SD	SEM	MD-Mean	MD-Median
Wooden data	5.732	4.410	4.556	0.557	3.562	3.429
TPM-Wooden	5.704	4.381	5.023	0.614	3.556	3.431
Hemming data	91.086	78.100	67.151	7.969	53.876	52.889
TPM-Hemming	91.285	70.111	80.390	9.541	53.913	53.662

Table 13

Estimated $f(0)$ and D for all compared models with respect to both data sets.

Wooden stake data			
Detection models	$\hat{f}(0)$	$\hat{D}$	$\lvert \widehat{SD} - SD \rvert$
TPM (β, λ)	0.1286	0.00431	0.467
Model (2015) (θ)	0.1293	0.00433	0.416
WEM (θ)	0.1361	0.0046	0.669
NEM (λ)	0.1744	0.0058	1.176
Hemingway data			
Detection models	$\hat{f}(0)$	$\hat{D}$	$\lvert \widehat{SD} - SD \rvert$
TPM (β, λ)	0.0080	4.7356×10^{-6}	13.239
Model (2015) (θ)	0.0081	4.7962×10^{-6}	12.142
WEM (θ)	0.0085	5.0306×10^{-6}	16.526
NEM (λ)	0.0109	6.4956×10^{-6}	23.935

6. Conclusion

In this paper, we have developed a new two-parameter detection model. It is flexible in terms of the shapes of the associated functions. It also includes the weighted exponential model as a submodel. We have investigated the theoretical and practical properties of this model. A maximum likelihood estimation work was carried out. The performance of the proposed model with respect to other models was illustrated by fitting two real-world data sets of perpendicular distances and using some goodness-of-fit statistics.

References

[1] F. Ababneh and O. M. Eidous, "A weighted exponential detection function model for line transect data," *Journal of Modern Applied Statistical Methods*, vol. 11, no. 1, pp. 1–11 (2012).

[2] H. S. Bakouch, C. Chesneau, and R. I. Abdullah, "A pliant parametric detection model for line transect data sampling," *Communications in Statistics-Theory and Methods*, vol. 51, no. 21, pp. 7340–7353 (2022).

[3] L. Barabesi, "Local likelihood density estimation in line transect sampling," *Environmetrics: The Official Journal of the International Environmetrics Society*, vol. 11, no. 4, pp. 413–422 (2000).

[4] S. Buckland, "Perpendicular distance models for line transect sampling," *Biometrics,* vol. 41, no. 1, pp. 177–195 (1985).

[5] S. T. Buckland, D. R. Anderson, K. P. Burnham, J. L. Laake, D. L. Borchers, and L. Thomas, Advanced Distance Sampling, vol. 2. Oxford: Oxford University Press (2004).

[6] S. T. Buckland, D. R. Anderson, K. P. Burnham, J. L. Laake, D. L. Borchers, and L. Thomas, Introduction to Distance Sampling: Estimating Abundance of Biological Populations. United Kingdom: Oxford University Press (2001).

[7] S. T. Buckland, E. A. Rexstad, T. A. Marques, and C. S. Oedekoven, Distance Sampling Methods and Applications. The Netherlands: Springer (2015).

[8] K. P. Burnham, D. R. Anderson, and J. L. Laake, "Estimation of density from line transect sampling of biological populations," *Wildlife Monographs,* no. 72, pp. 3–202 (1980).

[9] K. Burnham and D. Anderson, "Mathematical models for nonparametric inferences from line transect data," *Biometrics,* vol. 32, no. 2, pp. 325–336 (1976).

[10] R. G. Clark, "Efficiency and robustness in distance sampling," Working paper, University of Wollongong (2015).

[11] L. Eberhardt, "Transect methods for population studies," *The Journal of Wildlife Management*, vol. 42, no. 1, pp. 1–31 (1978).

[12] O. Eidous and F. Al-Eibood, "A new parametric detection function for line transect sampling," *JP Journal of Fundamental and Applied Statistics,* vol. 8, no. 1, pp. 1–16 (2015).

[13] C. E. Gates, W. H. Marshall, and D. P. Olson, "Line transect method of estimating grouse population densities," *Biometrics,* vol. 24, no. 1, pp. 135–145 (1968).

[14] W. Jang and J. M. Loh, "Density estimation for grouped data with application to line transect sampling," *The Annals of Applied Statistics,* vol. 4, no. 2, pp. 893–915 (2010).

[15] R. J. Karunamuni and T. J. Quinn, "Bayesian estimation of animal abundance for line transect sampling," *Biometrics,* vol. 51, no. 4, pp. 1325–1337 (1995).

[16] A. Mahdavi and D. Kundu, "A new method for generating distributions with an application to exponential distribution," *Communications in Statistics-Theory and Methods,* vol. 46, no. 13, pp. 6543–6557 (2017).

[17] D. L. Miller, E. Rexstad, L. Thomas, L. Marshall, and J. L. Laake, "Distance sampling in R," *Journal of Statistical Software*, vol. 89, no. 1, pp. 1–28, (2019).

[18] S. Nadarajah, "The exponentiated exponential distribution: a survey," *AStA Advances in Statistical Analysis,* vol. 95, pp. 219–251 (2011).

[19] M. Nassar, A. Z. Afify, and M. Shakhatreh, "Estimation methods of alpha power exponential distribution with applications to engineering and medical data," *Pakistan Journal of Statistics and Operation Research*, vol. 16, no. 1, pp. 149–166 (2020).

[20] K. Pollock, "A family of density estimators for line-transect sampling," *Biometrics*, vol. 34, no. 3, pp. 475–478 (1978).

[21] T. J. Quinn, "The effects of aggregation on line transect estimators of population abundance with application to marine mammal populations," Ph.D. thesis, University of Washington (1977).

[22] T. Quinn and V. Gallucci, "Parametric models for line-transect estimators of abundance," *Ecology*, vol. 61, no. 2, pp. 293–302 (1980).

[23] G. Seber, Estimation of Animal Abundance and Related Parameters, 2nd ed. London: Griffin (1982).

[24] A. Sen, J. Tourigny, and G. Smith, "On the line transect sampling method," *Biometrics*, vol. 30, no. 2, pp. 329–340 (1974).

[25] S. Wang, W. Chen, M. Chen, and Y. Zhou, "Maximum likelihood estimation of the parameters of the inverse Gaussian distribution using maximum rank set sampling with unequal samples," *Mathematical Population Studies*, vol. 30, no. 1, pp. 1–21 (2023).

[26] S. Zhang, "Generalized likelihood ratio test for the shoulder condition in line transect sampling," *Communications in Statistics-Theory and Methods*, vol. 30, no. 11, pp. 2343–2354 (2001).

Received January, 2024