11-1-2013

Akaike Information Criterion to Select the Parametric Detection Function for Kernel Estimator Using Line Transect Data

Omar Eidous
King Abdulaziz University, Jeddah, Saudi Arabia, omarm@yu.edu.jo

Samar Al-Salman
Yarmouk University, Irbid, Jordan, alsalman85@yahoo.com

Recommended Citation

Akaike Information Criterion to Select the Parametric Detection Function for Kernel Estimator Using Line Transect Data

Omar Eidous
King Abdulaziz University
Jeddah, Saudi Arabia

Samar Al-Salman
Yarmouk University
Irbid, Jordan

Among different candidate parametric detection functions, it is suggested to use Akaike Information Criterion (*AIC*) to select the most appropriate one of them to fit line transect data. Four different detection functions are considered in this paper. Two of them are taken to satisfy the shoulder condition assumption and the other two estimators do not satisfy this condition. Once the appropriate detection function is determined, it also can be used to select the smoothing parameter of the nonparametric kernel estimator. For a wide range of target densities, a simulation results show the reasonable and good performances of the resulting estimators comparing with some existing estimator, particularly the usual kernel estimator when the half normal model is use as a reference to select the smoothing parameter.

Keywords: Line transect sampling, Akaike Information Criterion, kernel method, smoothing parameter

Introduction

Line transect sampling is one of the popular sampling method adopted by ecologists to estimate the population density D of specific objects in a given region. The estimation procedure can be performed by walking a distance L following a specific line transect, counts the number objects being investigated and records the perpendicular distance, X from the detected object to the center of the line transect. Let $g(x)$ be the detection function of observing an object at distance X, then X will tend to have a probability density function $f(x)$ of the same shape as $g(x)$ but scaled so that the area under $f(x)$ equals unity. Buckland et al. (2001) and Burnham et al. (1980) constitute the key references for this distance sampling procedure.

Dr. Eidous is in the Department of Statistics. Email him at: omarm@yu.edu.jo. Samar Al-Salman is a graduate of the Department of Statistics. Email him at alsalman85@yahoo.com.

The first logical assumption related the detection function $g(x)$ indicates that $g(x)$ is monotonically decreasing function in x. The second important assumption is that $g(0) = 1$, which indicates the objects located on the center of line will never be missed. In other words, this condition means that the probability of detected an object given that its perpendicular distance is zero equals one. In addition to the previous two assumptions, some authors (see Mack and Quang, 1998) stated that, in many practical situations the shape of the detection function of the data should have a shoulder at distance $x = 0$. If that is required then it can be translated mathematically as $g'(0) = 0$. The condition $g'(0) = 0$ is known in the literature as the shoulder condition assumption. However, Buckland et al. (2001) pointed out that the shoulder condition assumption may not be satisfied for some cases in practice, especially for the experiment with small objects or the experiment that performed with existing a fog or a tall grass etc. If $g(x)$ is monotonically decreasing and $g'(0) = 0$ then this ensures that $f(x)$ is in turn monotonically decreasing with $f'(0) = 0$.

Burnham and Anderson (1976) gave the fundamental relation for estimating the density of objects in a specific area, which can be expressed as $D = E(n) f(0) / 2L$, and the general estimate for D is given by $\hat{D} = n\hat{f}(0) / 2L$, where $E(n)$ is the expected value of the number of detected objects n, and $\hat{f}(0)$ is an approximate sample estimator of $f(0)$ based on the n observed perpendicular distances $x_1, x_2, \ldots, x_n$. Hence, the key aspect in line transects sampling can be reduced to be the modeling of $f(x)$ as well as the estimation of $f(0)$.

Let $X_1, X_2, \ldots, X_n$ be a random sample of n perpendicular distances from unknown *pdf* $f(x)$. A parametric approach would involve by assuming that $f(x)$ is a member of a family of proper *pdf* of a known functional form but depends on an unknown parameter θ, where θ may take a vector value and should be estimated by using the perpendicular distances. A variety of approaches to estimate θ will lead to $\hat{f}(0) = f(0, \hat{\theta})$. In contrast to the parametric method, the nonparametric kernel method requires no assumptions about the form of $f(x)$. This method allows the data at hand to talk about themselves.

Given that the line transect data are available and their true *pdf* is unknown, our first aim in this paper is to choose the most appropriate *pdf* for these data by considering four logical parametric models. The Akaiki Information Criterion (*AIC*) is suggested for use to select the best parametric model. The second aim is to use the *AIC* to determine the best parametric model that can be used as a reference to determine the smoothing parameter of the kernel estimator of $f(0)$.

Some Parametric Estimators

A number of parametric models have been proposed in the literature for $f(x)$. The negative exponential model and the half normal model are the most prominent models. Gates et al. (1968) suggested the negative exponential model with detection function,

$$g_1(x) = e^{-x/\alpha}, \quad x \geq 0$$

The corresponding *pdf* is,

$$f_1(x) = \frac{1}{\alpha} e^{-x/\alpha}, x \geq 0 \qquad (1)$$

The maximum likelihood (ML) method indicates that the ML estimator of $f(0)$ is $\hat{f}_1(0) = 1/\bar{X}$, where $\bar{X}$ is the sample mean. The detection function $g_1(x)$ (or the *pdf* $f_1(x)$) do not satisfy the shoulder condition, which minimizes the importance of utilizing this model in line transect sampling. In contrast to the exponential model, the half normal model (Burnham et al., 1980) satisfies the shoulder condition assumption. The half normal detection function is given by

$$g_2(x) = e^{-x^2/2\sigma^2}$$

and the *pdf* is

$$f_2(x) = \frac{2}{\sigma\sqrt{2\pi}} e^{-x^2/2\sigma^2}, x \geq 0. \qquad (2)$$

The ML estimator of $f(0)$ is $\hat{f}_2(0) = \left(\frac{2}{\pi T}\right)^{1/2}$ under the half normal model, where $T = \sum_{i=1}^{n} x_i^2 / n$ is the ML estimator of σ^2. Ababneh and Eidous (2012) suggested the weighted exponential detection function with the form,

$$g_3(x) = e^{-\theta x}\left(2 - e^{-\theta x}\right),$$

346

and the corresponding *pdf* is

$$f_3(x) = \frac{2\theta}{3} e^{-\theta x}\left(2 - e^{-\theta x}\right), x \geq 0, \theta > 0 \tag{3}$$

The parameter required to estimate is

$$f_3(0) = \frac{2\theta}{3}$$

The expected value of *X* based on Model (3) is $7/(6\theta)$, which gives $\hat{f}_3(0) = 7/9\bar{X}$ as the moment estimator for $f_3(0)$. The moment estimator for $f_3(0)$ is given in a closed form, while the maximum likelihood estimator needs a numerical method to find it. It is worthwhile to note that the Model (3) satisfies the shoulder condition assumption. That is, $f_3{}'(0) = 0$. Finally, Burnham et al. (1980) suggested the Reversed Logistic detection function, which is given by

$$g_4(x) = \frac{3e^{-0x}}{1 + 2e^{-0x}},$$

and the corresponding *pdf* is given by

$$f_4(x) = \frac{2\theta}{3\ln(3)} \frac{3e^{-\theta x}}{1 + 2e^{-\theta x}}, x \geq 0 \tag{4}$$

It is easy to verify that Model (4) does not satisfy the shoulder condition assumption. Based on Model (4), the parameter that to estimate is $f_4(0) = \frac{2\theta}{3\ln(3)}$. If one decides to use the moment estimator for $f_4(0)$, then he obtains $\hat{f}_4(0) = \frac{2}{3\ln(3)} \frac{1.3078}{\bar{X}} = \frac{0.7936}{\bar{X}}$. Again the ML estimator of θ based on Model (4) does not exist in closed form and consequently it is not exist in closed form for $f_4(0)$. Therefore, a numerical method is required to find the corresponding ML estimator.

The Nonparametric Kernel Estimator

Let X_1, X_2, …, X_n be n perpendicular distances (assumed to be independent and identically distributed) from a continuous probability density function $f(x)$. Because the perpendicular distances are nonnegative, the usual kernel estimator of $f(x)$ (Silverman, 1986 and Chen, 1996) is

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^{n} \left\{ K\left(\frac{x-x_i}{h}\right) + K\left(\frac{x+x_i}{h}\right) \right\}, x \geq 0 \qquad (5)$$

where h is called the smoothing parameter (or bandwidth) and K is a symmetric kernel function assumed to satisfy the following conditions

$$\int_{-\infty}^{\infty} K(u)du = 1$$

$$\int_{-\infty}^{\infty} uK(u)du = 0 \qquad (6)$$

$$\int_{-\infty}^{\infty} u^2 K(u)du = c \neq 0 < \infty, \text{ where } c \text{ is a constant}$$

The kernel estimator of $f(0)$ is obtained by taking $x = 0$ in Equation (5), which gives

$$\hat{f}(0) = \frac{1}{nh} \sum_{i=1}^{n} \left\{ K\left(\frac{-x_i}{h}\right) + K\left(\frac{x_i}{h}\right) \right\}.$$

Since K is a symmetric function (i.e., $K(-x) = K(x)$), then

$$\hat{f}(0) = \frac{2}{nh} \sum_{i=1}^{n} K\left(\frac{x_i}{h}\right) \qquad (7)$$

If $f(x)$ has a second continuous derivative at $x = 0$ then under the assumption that $h \rightarrow 0$ and $nh \rightarrow \infty$ when $n \rightarrow \infty$, the bias and variance of $\hat{f}(0)$ are (Chen, 1996)

348

$$bias\left(\hat{f}(0)\right)=2hf^{(1)}(0)R_1(K)+h^2 f^{(2)}(0)R_2(K)+o(h^2) \tag{8}$$

and

$$var\left(\hat{f}(0)\right)=\frac{4f(0)M_2(K)}{nh}+o(nh)^{-1} \tag{9}$$

where $f^{(1)}(0)$ is the ith derivative of $f(x)$ at $x = 0$, $R_z(K)=\int_0^\infty u^z K(u)du$ and

$M_2(K)=\int_0^\infty K^2(u)du.$ Hence, if $f'(0) = 0$, then the bias convergence rate is $O(h^2)$, if not (i.e., $f'(0) \neq 0$), the bias convergence rate is only $O(h)$, which is slower than $O(h^2)$ as $h \rightarrow 0$.

The estimator of D by using the kernel method is now obtained by substituting the estimator $\hat{f}(0)$ from (7) back into the formula of $\hat{D}$.

The Optimal Smoothing Parameter

There are many kernel functions that satisfy Condition (6). Wand and Jones (1995) pointed out that there is very little to choose between the various kernel functions on the basis of the mean square error of the estimator. In other words, given that the kernel function that satisfies (6) is selected, then the performance of the kernel estimator remains almost the same as any other kernel estimator when the kernel function is changed. However, it becomes very well know that the way to select the smoothing parameter h is very sensitive on the performance of the kernel estimator (see for example, Gerard and Schucany, 1999 and Eidous, 2005). The popular method that used to select h using line transect data is the reference method. This method can be used by adopting the half normal detection function as a reference. Gerard and Schucany (1999) pointed out that this technique is very acceptable in line transect sampling and there is no need to adopt the other computational methods such as least squares cross validation and likelihood cross validation methods.

As opposed to referring to only the half normal detection function to compute h, the other detection functions as stated in the section on Parametric Estimators are introduced as references to select h. This gives a choice to select

the most appropriate model to select the smoothing parameter and then, as expected, to improve the performances of the kernel estimator.

As stated, the smoothing parameter h has a strong effect on the accuracy of the kernel estimator (7) as illustrated by examining Formulas (8) and (9). As they demonstrated, the choice of a large value of h gives a large bias and small variance and vice versa. The logical method to determine h is to find its optimal value that minimizes the asymptotic mean square error (*MSE*) of the estimator $\hat{f}(0)$. The formula of the asymptotic *MSE* of $\hat{f}(0)$ (based on (8) and (9)) is given by

$$MSE\left(\hat{f}(0)\right) = \left[bias\left(\hat{f}(0)\right)\right]^2 + var\left(\hat{f}(0)\right)$$

$$= h^4\left(f''(0)\right)^2 R^2(K) + \frac{4}{nh}f(0)M_2(K)$$

(10)

Formula (10) is obtained by assuming that $f'(0) = 0$. By differentiating both sides of (10) with respect to h and equating the resulting equation with zero, the value of h that minimize the asymptotic *MSE* of $f(0)$ can be obtained. This value is known as the optimal smoothing parameter with respect to the asymptotic *MSE*, which is given by

$$h = \left\{\frac{f(0)M_2(K)}{R_2^2(K)\left(f''(0)\right)^2}\right\}^{1/5} n^{-1/5}$$

(11)

The smoothing parameter h is now computed by assuming a reasonable form for $f(x)$. Gerard and Schucany (1999) compared among different methods to compute h in practice. They recommended to use the half-normal *pdf* as a reference, i.e., they assumed that $f(x) = f_2(x)$ (see Formula (2)), which gives $f(0) = \frac{2}{\sigma\sqrt{2\pi}}$ and

$f''(0) = \frac{-2}{\sigma^3\sqrt{2\pi}}$, where σ is now estimated by its maximum likelihood estimator

$\hat{\sigma} = \sqrt{\frac{\sum_{i=1}^{n} x_i^2}{n}}$. Now, assume that the kernel function is the standard normal, i.e.

$K(t) = N(0,1)$, then $h = 0.933 \hat{\sigma} n^{-1/5}$. By adopting the same technique, the formulas of h for the other densities can be derived, and are stated as follows:

- If $f(x) = f_1(x)$, then $h = 0.8918 \hat{\theta} n^{-1/5}$, where $\hat{\theta} = \bar{X}$ is the maximum likelihood of θ.

- If $f(x) = f_3(x)$, then $h = 0.7330 \left(1/\hat{\theta} \right) n^{-1/5}$, where $\hat{\theta}$ is the maximum likelihood of θ under the weighted exponential pdf. However if the moments estimator of θ is required then $\hat{\theta} = \dfrac{7}{6\bar{X}}$.

- If $f(x) = f_4(x)$, then $h = 2.3734 \left(1/\hat{\theta} \right) n^{-1/5}$, where $\hat{\theta}$ is the maximum likelihood of θ under the reversed Logistic pdf. Note that the moments estimator of θ is $\hat{\theta} = \dfrac{1.3078}{\bar{X}}$.

Akaike Information Criterion (*AIC*) and the Proposed Estimators

The *AIC* (Buckland et al., 2001) is defined by

$$AIC = -2 Log_e (L) + 2p$$

Where $\log_e (L)$ is the log-likelihood function evaluated at the maximum likelihood estimates of the model parameter and p is the number of parameters in the model. The above criterion provides a method to select the best model (among a set of models) that fit the data at hand. For a given data set, *AIC* is computed for each model and the model with the smallest *AIC* is considered to be better than the others. For models (1), (2), (3), and (4), the *AIC*s are given by

- $AIC_1 = 2n + 2n Log_e (\bar{X}) + 2$ for the negative exponential (Model 1).

- $AIC_2 = -2n \log_e (2) + n \log_e (2\pi \hat{\sigma}^2) + n + 2$ for the half normal (Model 2), where $\hat{\sigma} = \sqrt{\dfrac{\sum_{i=1}^{n} x_i^2}{n}}$ is the maximum likelihood estimator for σ^2.

351

- $$AIC_3 = -2\left\{n\log_e\left(2\hat{\theta}/3\right) - \hat{\theta}\sum\nolimits_{i=1}^{n}x_i + \sum\nolimits_{i=1}^{n}\log_e\left(2-\exp\left(-\hat{\theta}x_i\right)\right)\right\} + 2$$

  for the weighted exponential (Model 3), where $\hat{\theta}$ is the maximum likelihood estimator for θ.

- $$AIC_4 = -2\left\{n\log_e\left(2\hat{\theta}\right) - n\log_e\left(\log_e(3)\right) - \hat{\theta}\sum\nolimits_{i=1}^{n}x_i - \sum\nolimits_{i=1}^{n}\log_e\left(1+2\exp\left(-\hat{\theta}x_i\right)\right)\right\} + 2$$

  for the reversed Logistic (Model 4), where $\hat{\theta}$ is the maximum likelihood estimator for θ.

Two proposed estimators will be constructed for $f(0)$ based on the *AIC*. For a random sample of n perpendicular distances X_1, X_2, …, X_n, the first proposed estimator is constructed by computing the *AIC* for each model and the model with the smallest *AIC* is selected to estimate $f(0)$. If the selected model is $f_j(x)$, $j = 1, 2, 3, 4$ then $\hat{f}_j(0)$ is the estimator of $f(0)$. For example, if $f_1(x)$ is selected based on the *AIC* then $\hat{f}(0) = 1/\bar{X}$ is the estimator of $f(0)$. The first estimator of $f(0)$ will be denoted by $\hat{f}_P(0)$, where the sub P stands for "parametric." The second estimator is the usual kernel estimator (Estimator 7) but here the smoothing parameter of the kernel estimator is computed by using the reference model that is selected based on the *AIC*. In other words, compute the *AIC* for the previous four models and then select the model that has the smallest *AIC*, then based on the selected model, use the corresponding optimal formula to compute h. For example, if $f_1(x)$ is selected based on the *AIC* then $h = 0.8918\bar{X}n^{-1/5}$. This value is substituted in Estimator (7), which enables us to compute its final value for a given data set. The second estimator of $f(0)$ is denoted as $\hat{f}_N(0)$, where the sub N stands for "non-parametric."

Simulation Study and Results

In order to assess the performances of the proposed estimators $\hat{f}_P(0)$ and $\hat{f}_N(0)$ of $f(0)$, discussed in the previous section, a simulation study is performed. For the sake of comparison, the usual kernel estimator $\hat{f}(0)$ with smoothing parameter $h = 0.933\hat{\sigma}n^{-1/5}$ (Gerard and Schucany, 1999) is also considered. Four target families were considered in the simulation. These families were chosen using the criterion that they are representative of many different shapes that might occur in

the field. The target models – not necessary the same as the four models discussed in the Introduction – that used to simulate the perpendicular distances are

1) Exponential Power (EP) family (Pollack, 1978)

$$f(x) = \frac{1}{\Gamma(1+1/\beta)} e^{-x^\beta}, \ x \geq 0, \ \beta > 1$$

2) Hazard-Rate (HR) family (Hayes and Buckland, 1983)

$$f(x) = \frac{1}{\Gamma(1-1/\beta)} \left(1 - e^{-x^{-\beta}}\right), \ x \geq 0, \ \beta > 1$$

3) Beta (BE) family (Eberhardt, 1968)

$$f(x) = (w+\beta)(1-x/w)^\beta, \ 0 \leq x \leq w, \ \beta > 0$$

4) General Reversed Logistic (GRL) family (Burnham et al., 1980)

$$f(x) = \frac{\beta b}{\ln(1+b)\left(1+be^{-\beta x}\right)}, \ x \geq 0, \ \beta, b > 0$$

Two target models with two values for parameter β are selected from each of the above families. The selected model is truncated at a distance w. The selected values for β and for w for each model are as follows: $(\beta, w) = (1,5), (2,3)$ for EP family; $(\beta, w) = (1.5,20), (2,12)$ for HR family; $(\beta, w) = (10,5), (20,9)$ for BE family; and $(\beta, b, w) = (6,10,1), (8,30,1)$ for GRL family. These models cover a wide range for the detection functions of perpendicular distances, which vary near zero from spike to flat. It is worthwhile to mention here that the Reversed Logistic model (i.e. $f_4(x)$) is a special case of the above GRL with $b = 2$. The target GRL models that selected to simulate the data are taken for $b = 10,30$, which differ in their shape for $f_4(x)$. This choice is made to avoid our knowledge of the true detection function of the perpendicular distances.

It should be remarked that the EP model with $\beta = 1$, BE family and the GRL family do not satisfy the shoulder condition assumption. These choices were made in order to assess the robustness of the considered estimators with respect to the violation of the shoulder condition assumption. Note also that the other considered models satisfy the shoulder condition assumption.

For each model and for sample sizes $n = 50,100,200$, one thousand samples of perpendicular distances were randomly drawn. For each model and for each

sample size, Tables 1 – 4 demonstrate the simulated value of the relative bias (RB); $RB = \left\{ E\left(\hat{f}(0) \right) - f(0) \right\} / f(0)$ and the relative mean error (RME); $RME = \sqrt{MSE\left(\hat{f}(0) \right)} / f(0)$ for each considered estimator.

Table 1. *RB*, *RME*, and *EFF* of the different estimators when the perpendicular distances are simulated from EP detection function

Estimator	Parameters	n = 50			n = 100			n = 200		
		RB	*RME*	*EFF*	*RB*	*RME*	*EFF*	*RB*	*RME*	*EFF*
$\hat{f}(0)$		-0.322	0.337	**1.000**	-0.288	0.298	**1.000**	-0.265	0.272	**1.000**
$\hat{f}_P(0)$	$\beta = 1$ $w = 5$	-0.043	0.209	**1.614**	-0.031	0.166	**1.798**	-0.024	0.133	**2.038**
$\hat{f}_N(0)$		-0.221	0.259	**1.302**	-0.194	0.224	**1.332**	-0.177	0.192	**1.417**
$\hat{f}(0)$		-0.083	0.156	**1.000**	-0.070	0.120	**1.000**	-0.052	0.097	**1.000**
$\hat{f}_P(0)$	$\beta = 2$ $w = 2.5$	0.043	0.172	**0.908**	0.008	0.080	**1.511**	0.006	0.055	**1.765**
$\hat{f}_N(0)$		-0.087	0.173	**0.906**	-0.082	0.124	**0.971**	-0.058	0.098	**0.994**

Table 2. *RB*, *RME*, and *EFF* of the different estimators when the perpendicular distances are simulated from HR detection function

Estimator	Parameters	n = 50			n = 100			n = 200		
		RB	*RME*	*EFF*	*RB*	*RME*	*EFF*	*RB*	*RME*	*EFF*
$\hat{f}(0)$		-0.474	0.485	**1.000**	-0.439	0.444	**1.000**	-0.398	0.401	**1.000**
$\hat{f}_P(0)$	$\beta = 1.5$ $w = 20$	-0.255	0.297	**1.633**	-0.269	0.284	**1.563**	-0.277	0.285	**1.411**
$\hat{f}_N(0)$		-0.301	0.333	**1.457**	-0.268	0.284	**1.562**	-0.227	0.238	**1.689**
$\hat{f}(0)$		-0.266	0.290	**1.000**	-0.215	0.231	**1.000**	-0.171	0.183	**1.000**
$\hat{f}_P(0)$	$\beta = 2$ $w = 12$	0.050	0.188	**1.536**	0.069	0.146	**1.581**	0.068	0.113	**1.612**
$\hat{f}_N(0)$		-0.119	0.190	**1.522**	-0.080	0.131	**1.760**	-0.050	0.094	**1.944**

Table 3. *RB*, *RME*, and *EFF* of the different estimators when the perpendicular distances are simulated from BE detection function

Estimator	Parameters	n = 50			n = 100			n = 200		
		RB	RME	EFF	RB	RME	EFF	RB	RME	EFF
$\hat{f}(0)$		-0.299	0.316	**1.000**	-0.271	0.281	**1.000**	-0.244	0.252	**1.000**
$\hat{f}_P(0)$	$\beta = 10$ $w = 5$	-0.064	0.233	**1.354**	-0.039	0.194	**1.451**	-0.014	0.147	**1.716**
$\hat{f}_N(0)$		-0.217	0.264	**1.197**	-0.184	0.217	**1.298**	-0.147	0.166	**1.516**
$\hat{f}(0)$		-0.317	0.333	**1.000**	-0.285	0.296	**1.000**	-0.257	0.264	**1.000**
$\hat{f}_P(0)$	$\beta = 20$ $w = 9$	-0.068	0.220	**1.518**	-0.036	0.168	**1.756**	-0.020	0.133	**1.984**
$\hat{f}_N(0)$		-0.229	0.266	**1.253**	-0.198	0.224	**1.318**	-0.171	0.184	**1.431**

Table 4. *RB*, *RME*, and *EFF* of the different estimators when the perpendicular distances are simulated from GRL detection function

Estimator	Parameters	n = 50			n = 100			n = 200		
		RB	RME	EFF	RB	RME	EFF	RB	RME	EFF
$\hat{f}(0)$		-0.092	0.166	**1.000**	-0.087	0.132	**1.000**	-0.070	0.107	**1.000**
$\hat{f}_P(0)$	$\beta = 6$ $b = 10$ $w = 1$	0.028	0.172	**0.968**	-0.013	0.075	**1.758**	-0.009	0.050	**2.155**
$\hat{f}_N(0)$		-0.098	0.181	**0.922**	-0.097	0.135	**0.975**	-0.073	0.107	**1.001**
$\hat{f}(0)$		-0.058	0.150	**1.000**	-0.040	0.115	**1.000**	-0.035	0.094	**1.000**
$\hat{f}_P(0)$	$\beta = 8$ $b = 30$ $w = 1$	0.101	0.168	**0.891**	0.088	0.115	**1.001**	0.080	0.095	**0.996**
$\hat{f}_N(0)$		-0.063	0.157	**0.953**	-0.045	0.114	**1.015**	-0.035	0.094	**1.000**

For simple comparison, compute the efficiency (*EFF*) of the proposed estimators $\hat{f}_P(0)$ and $\hat{f}_N(0)$ with respect to the classic kernel estimator $\hat{f}(0)$, which is given by

$$EFF = \frac{MSE\left(\hat{f}(0)\right)}{MSE\left(\hat{f}_{P \text{ or } N}(0)\right)}$$

Depending on the simulation results of Tables 1 – 4, several conclusions can be drawn by inspecting the results in regard to *RB*, *RME*, and *EFF*

- The RBs that associated with the proposed estimators $\hat{f}_P(0)$ and $\hat{f}_N(0)$ are generally smaller in their magnitude than that associated with the classic kernel estimator $\hat{f}(0)$.

- The RMEs for different estimators decrease when the sample size increases. This is a strong sign for the consistency of these estimators.

- The performance of the classical kernel estimator seems to be reasonable for EP model with $\beta = 2$ and for GRL model comparing to the proposed estimator $\hat{f}_N(0)$, at which the performances of the two estimators are similar. However, $\hat{f}_N(0)$ beats $\hat{f}(0)$ for the other cases.

- By comparing between the two proposed estimators $\hat{f}_N(0)$ and $\hat{f}_P(0)$, the performance of the former one seem to be surprisingly for most considered cases especially when the sample size increases.

Generally, Tables 1 – 4 demonstrate clearly that there is a significantly improvement by applying the estimator $\hat{f}_P(0)$ or even $\hat{f}_N(0)$ instead of the classic kernel estimator $\hat{f}(0)$.

References

Ababneh, F. & Eidous, O. (2012). A Weighted Exponential Detection Function Model for Line Transect Data. *Journal of Modern Applied Statistical Methods*, *11*(1), 144-151.

Buckland, S. T., Anderson, D. R., Burnham, K. P., Laake, J. L., Borchers, D. L., & Thomas, L. (2001). *Introduction to distance sampling*. Oxford: Oxford University Press.

Burnham, K. P., & Anderson, D. R. (1976). Mathematical models for nonparametric inferences from line transect data. *Biometrics*, *32*, 325-336.

Burnham, K. P., Anderson, D. R., & Laake, J. L. (1980). Estimation of density from line transect sampling of biological populations. *Wildlife Monograph 72*.

Chen, S. X. (1996). A kernel estimate for the density of a biological population by using line transect sampling. *Applied Statistics*, *45*, 135-150.

Eberhardt, L. L. (1968). A preliminary appraisal of line transects. *Journal of Wildlife Management, 32*, 82-88.

Eidous, O. M. (2005). Bias correction for histogram estimator using line transect sampling. *Environmetrics, 16*, 61-88.

Gates, C. E., Marshall, W. H., & Olson, D. P. (1968). Line transect method of estimating grouse population densities. *Biometrics, 24,* 135-145.

Gerard, P. D., & Schucany, W. R. (1999). Local Bandwidth Selection for Kernel Estimation of Population Densities with Line Transect sampling. *Biometrics, 55*, 769-73.

Hayes, R.J., & Buckland, S.T. (1983). Radial distance models for line-transect method. *Biometrics*, *39*, 29-42.

Mack, Y. P., & Quang, P. X. (1998). Kernel methods in line and point transect sampling. *Biometrics*, *54*, 609-619.

Pollock, K. H. (1978). A family of density estimators for line transect sampling. *Biometrics*, *34*, 475-478.

Silverman, B.W. (1986). *Density Estimation for Statistics and Data Analysis*. London: Chapman and Hall.

Wand, M.P., & Jones, M.C. (1995). *Kernel Smoothing*. London: Chapman and Hall.