



FACULTAD DE MATEMÁTICAS

DEPARTAMENTO DE ESTADÍSTICA E INVESTIGACIÓN OPERATIVA

INFERENCIA EN MUESTREO POR TRANSECTOS.

Memoria realizada por José Antonio Rodríguez Jiménez

Dirigido por:

Dra. Ana María Muñoz Reyes

y

Dr. Juan Manuel Muñoz Pichardo

Resumen

En este trabajo, se desarrollará las herramientas necesarias para estimar la abundancia o densidad en un área de estudio mediante el software R y la librería *Distance* para métodos paramétricos y la librería *evmix* para métodos no paramétricos basados en la función núcleo.

Para esto se empezará describiendo los métodos de estimación de la abundancia, posteriormente un planteamiento de diseño de modelos con su desarrollo teórico.

Así como una descripción de la estimación de la densidad de probabilidad de la variable distancia y de la función de detectabilidad.

Finalmente, se describen los paquete R *Distance* y *evmix* para aplicar la metodología incluida en el trabajo.

Abstract

In this work we will develop the necessary tools to estimate abundance or density in a study area using the R software and the *Distance* library for parametric methods and the *evmix* library for non-parametric methods based on the kernel function.

This will begin by describing the methods of estimating abundance, then a design approach of models with its theoretical development.

As well as a description of the estimation of the probability density of the distance variable and the detectability function.

Finally, the R *Distance* and *evmix* packages are described to apply the methodology included in the work.

Índice general

Introducción	9
1. Planteamiento teórico	11
1.1. Muestreo por áreas	11
1.1.1. Estimación de la densidad y la abundancia	11
1.1.2. Limitaciones del muestreo por áreas	12
1.2. Métodos de muestreo por distancia.	12
1.2.1. Muestreo por transecto lineal	12
1.2.2. Muestreo por transecto puntual	13
1.3. Diseño de modelos	14
1.3.1. Diseño del modelo de transecto puntual	14
1.3.2. Diseño del modelo de transecto lineal	16
1.4. Desarrollo teórico de los modelos	17
1.4.1. Transecto lineal.	17
1.4.2. Transecto puntual.	20
1.5. Ilustración de muestreo por transecto	24
1.5.1. Datos de estacas de madera	24
2. Estimación de las funciones asociadas a la densidad.	27
2.1. Condición de la función de detectabilidad	27
2.2. Métodos paramétricos	27
2.2.1. Función de detectabilidad exponencial negativa	28
2.2.2. Función de detectabilidad semi-normal	28
2.2.3. Otras funciones de detectabilidad	29
2.2.4. Modelo propuesto	31
2.2.5. Propiedades estadísticas	32
2.2.6. Construcción de estimadores	33
2.2.6.1. Estimador por el método de los momentos para $f(0)$	33
2.2.6.2. Estimador por el método de máxima verosimilitud para $f(0)$	33
2.3. Aplicación de los métodos paramétricos.	34
2.3.1. Obtención de $g(y)$	34
2.3.2. Obtención de $f(y)$	34
2.3.3. Estimación por intervalos de la densidad.	35
2.4. Método no paramétrico basado en funciones núcleos	36

2.4.1.	El estimador	36
2.4.2.	Selección de la ventana	39
2.4.3.	Implementación práctica del estimador propuesto	39
2.4.4.	Un vistazo a las condiciones que cumple $f(x)$	40
2.5.	Método semiparamétrico para muestreo por transecto lineal.	41
2.5.1.	El estimador	41
2.5.2.	Elección de la ventana h	44
2.5.3.	Aplicación del estimador propuesto	45
2.5.3.1.	Cálculo del estimador propuesto	45
2.5.3.2.	Rendimiento del estimador propuesto mediante simulación	46
2.5.4.	Resultados y conclusiones de la simulación	50
2.6.	Mejora del método semiparamétrico mediante la cantidad k	51
2.6.1.	El estimador	51
2.6.2.	La selección de la ventana h y k	54
2.6.3.	Estudio mediante simulación	56
2.6.4.	Resultados y conclusiones	58
2.6.5.	Ejemplo	60
2.7.	Selección de estimadores paramétricos y no paramétricos	60
2.7.1.	Modelos paramétricos y AIC	61
2.7.2.	Estimadores semiparamétricos	63
3.	Aplicación en R	65
3.1.	Método paramétrico	65
3.1.1.	Ejemplo de transecto puntual	70
3.1.1.1.	Introducción del ejemplo	70
3.1.1.2.	Datos	70
3.1.1.3.	Análisis descriptivo de los datos	71
3.1.1.4.	Creación del data.frame	72
3.1.1.5.	Análisis del modelo	72
3.1.1.6.	Selección del mejor modelo	73
3.1.1.7.	Validación del modelo	73
3.1.1.8.	Función de detectabilidad	75
3.1.2.	Ejemplo de transecto lineal	75
3.1.2.1.	Introducción del ejemplo	75
3.1.2.2.	Datos	75
3.1.2.3.	Análisis descriptivo de los datos	77
3.1.2.4.	Creación del data.frame	78
3.1.2.5.	Análisis del modelo	78
3.1.2.6.	Selección del mejor modelo	78
3.1.2.7.	Validación del modelo	79
3.1.2.8.	Función de detectabilidad	81
3.2.	Métodos no paramétricos	82
3.2.1.	Estimador núcleo tradicional	83
3.2.2.	Estimador semiparamétrico	84
3.3.	Ejemplo de estacas de madera	85

3.3.1.	Datos	87
3.3.2.	Análisis descriptivo de los datos	87
3.3.3.	Creación del data.frame	88
3.3.4.	Selección del modelo paramétrico	89
3.3.5.	Validación del modelo	90
3.3.6.	Función de detectabilidad	91
3.3.7.	Estimación no paramétrica	91
3.3.8.	Método semiparamétrico	92
Bibliografía		97

Introducción

El muestreo por distancias comprende un conjunto de métodos en los cuales se recogen las distancias detectadas desde una línea o punto, con las cuales se obtiene la abundancia o la densidad de los individuos observados.

En 2001 *Buckland* y otros recogieron métodos usuales de muestreo por distancia [1] y varias extensiones de dichos métodos en 2004 [2].

El método más utilizado es el muestreo por transecto lineal. Sobre la región de interés se toma una muestra aleatoria de líneas en la región, o más comúnmente una serie de líneas paralelas espaciadas sistemáticamente con un punto de inicio aleatorio.

Un observador viaja a lo largo de cada línea, registrando cualquier individuo detectado a una distancia w de la línea. Asumimos que todos los individuos sobre la línea son detectados, y que la probabilidad de detectar a un individuo decrece con el incremento de la distancia a la línea. Por lo tanto, en el muestreo por distancias no se necesita que todos los individuos sean detectados.

La distancia de cada individuo detectado desde la línea se recoge, y usando la distribución de estas distancias, estimamos la densidad de individuos que habitan la región de interés, y con dicha proporción estimamos la abundancia.

A la probabilidad de que un individuo sea observado en el área de estudio a le llamaremos detectabilidad y será denotado por P_a .

Si podemos asegurar que la detectabilidad de todos los individuos es certera ($P_a = 1$), entonces podemos usar muestreo por áreas. Este método consiste en dividir el área de estudio en m rectángulos iguales de área x cada uno, dando un área de estudio total de $a = mx$, si además observamos n individuos, usando la definición de la densidad de población D , que no es más que el número de individuos N por unidad de área A , donde el área A es el área de estudio donde habita la población, es decir, $D = N/A$. Obtenemos una estimación de la densidad $\hat{D} = n/a$.

En el caso en el que no podamos asegurar la detectabilidad de los individuos ($0 \leq P_a \leq 1$), el método por áreas nos da una mala estimación del número de individuos de la población. Por ello se introduce el método por transectos lineales y transectos puntuales, de tal forma que a partir de la función de detectabilidad junto con las distancias perpendiculares en el caso de los transecto lineales y distancias radiales en el caso de transectos puntuales podemos obtener una aceptable estimación del número de individuos que habitan en la población.

El término "*Muestreo por distancias*" fue introducido por *Buckland* y otros en 1993 [3], así como los conceptos de muestreo por transecto lineal y puntual, los cuales estiman la densidad de animales a partir de una muestra de las distancias del observador a los individuos detectados. La primera versión del software *Distance* fue hecha por *Laake* y otros en 1993 [4], donde permitía a los usuarios implementar los métodos del libro de *Buckland* y otros de ese mismo año. Y la primera versión de *Distance* como un paquete de *R* fue realizada en 2013 por *Miller* y otros [5].

Los métodos paramétricos otorgan una mejor estimación que los métodos no paramétricos. Sin embargo, tienen el gran problema de tener que conocer la forma que tiene la función de detectabilidad, esto es el conocimiento de la familia paramétrica a la que pertenece y entonces a partir de los datos estimar el parámetro. En la práctica, hay multitud de casos en la que la forma de la función de detectabilidad es desconocida, por lo que es obligatorio el uso de otros métodos no paramétricos como los basados en la función núcleo.

Los métodos de muestreo por distancia surgen del hecho empírico de que, en un procedimiento de observación, no todos los individuos son observados. Este "*defecto*" en la observación puede deberse a distintos motivos como obstáculos que impidan la visión, movilidad de los individuos, etc. Los datos muestrales vendrán dado por las distancias a las que son observados los individuos además de otras características de interés.

Conociendo P_a conoceríamos el número de individuos fácilmente, pues si hemos observado por ejemplo solo a la mitad de individuos ($P_a = 0,5$), es suficiente con multiplicar por dos el número de individuos observados. Por tanto, es necesario estimar P_a .

Si se desea profundizar más en el conocimiento del muestreo por distancias, véase el libro de *Buckland* y otros titulado *Advance Distance Sampling* del año 2004 [2] en el que expone condiciones del problemas más complejas que van desde que los individuos no estén aislados y se encuentren agrupados, a que las distancias recogidas para evitar errores de medición sean agrupadas en intervalos.

Capítulo 1

Planteamiento teórico

En este capítulo se realizará una breve descripción sobre los métodos por distancias y la utilidad de introducir la función de detectabilidad. Para ello, se introducirá métodos de muestreo por distancias mediante transectos lineales y transectos puntuales, sus respectivos diseños de sus modelos y desarrollo teóricos de los mismos. Así mismo, para ilustrar los métodos de muestreo por distancias se describirá un ejemplo real el cual se usará a lo largo del trabajo.

1.1. Muestreo por áreas

Se empezará explicando en qué consiste el muestreo por áreas y sus limitaciones, y cómo el muestreo por distancias surge de la necesidad de suplir dichas limitaciones.

1.1.1. Estimación de la densidad y la abundancia

Se considerará primero el caso en el que todos los individuos en el área de estudio son observados ($P_a = 1$). Para transectos lineales, se tiene el caso especial de muestreo por bandas, y para transectos puntuales, se tiene el caso de muestreo por círculos. El objetivo es estimar la densidad D y la abundancia N con $N = D \cdot A$, donde A es el tamaño de la región de estudio.

Un estimador de la densidad D , vendrá dado por $\hat{D} = n/a$ siendo n el número de individuos detectados y a el área de observación. Asimismo, para la abundancia N se tiene el estimador $\hat{N} = \frac{nA}{a}$.

Si se obtiene n mediante el muestreo por bandas o por círculos, se puede obtener una estimación de la abundancia o densidad.

Para el caso de muestreo por bandas con K líneas, con $L = \sum_{k=1}^K l_k$ la longitud total de las K líneas se obtiene:

$$\hat{D} = \frac{n}{2wL} \tag{1.1.1}$$

y

$$\hat{N} = \frac{nA}{2wL} \quad (1.1.2)$$

Para muestreo por círculos con K puntos, cada punto genera un área de observación πw^2 , se obtiene un área de observación total de $a = K\pi w^2$.

$$\hat{D} = \frac{n}{K\pi w^2} \quad (1.1.3)$$

y

$$\hat{N} = \frac{nA}{K\pi w^2} \quad (1.1.4)$$

1.1.2. Limitaciones del muestreo por áreas

Debido a que se asume la certeza de detectar a todos los individuos en el área de observación, se debe realizar áreas pequeñas para que el observador pueda ver a todos los individuos. Como el área es pequeña, el observador puede detectar individuos fuera de ésta, las cuales son ignoradas bajando la precisión.

La mayor limitación del muestreo por áreas es la dificultad de asegurar por parte del observador que todos los individuos han sido observados, y más si se acepta el hecho de que los individuos puedan moverse dentro y fuera del área de observación.

Dado que la detectabilidad no es estimada, sino que es solo una suposición, no puede ser verificada como cierta. Por lo tanto, parece más razonable adoptar métodos donde la detectabilidad pueda ser estimada y controlada.

A continuación se expondrá los métodos de muestreo por distancias.

1.2. Métodos de muestreo por distancia.

Se plantean los dos métodos considerados básicos en el muestreo por distancias.

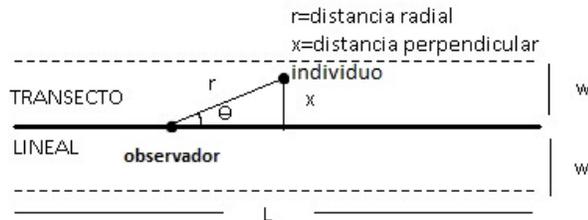
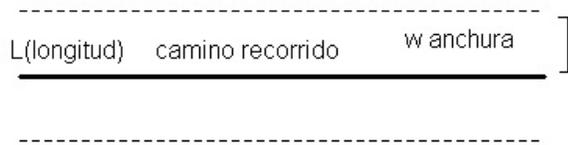
1.2.1. Muestreo por transecto lineal

En este tipo de muestreo, el observador se desplaza a lo largo de una línea anotando la localización relativa del elemento observado respecto a la línea. La posición relativa puede ser definida de dos modos:

- Distancia perpendicular desde el elemento observado al camino por donde se desplaza el observador.
- Distancia radial y ángulo de observación del observador al elemento en el momento en que es detectado.

Se denota w a la distancia máxima a la que el observador puede observar un individuo, obteniéndose una distancia mínima de $2w$ de separación entre transecto lineales en diseños de caminos paralelos para que no haya un individuo que sea observado más de una vez. Además, el área de observación es un rectángulo de área $2wL$ donde L es la longitud del transecto lineal.

Y a partir de dichas distancias junto a la función de detectabilidad podemos estimar P_a .



1.2.2. Muestreo por transecto puntual

Se establecen aleatoriamente puntos a lo largo del área de estudio, desde los cuales el observador tomará la muestra. El observador mirará a lo largo de los 360 grados y recopilará las distancias a la cual están los individuos de él, anotando las distancias r_1, r_2, r_3, \dots , y de la misma forma que pasaba en el método por transecto lineales, se denotará de la misma forma

w a la distancia máxima a la que el observador puede observar a un individuo, obteniéndose una distancia de $2w$ de separación entre transecto puntuales para que no haya individuos que sean observados más de una vez. Además el área de observación es un círculo de área πw^2 .

A partir de dichas distancias junto a la función de detectabilidad podemos estimar P_a .



1.3. Diseño de modelos

Los diseños de modelos se recogen de manera detallada en el libro de *Buckland* de 2001 [1]. Esta sección tratará sobre los diseños básicos y problemas relacionados con el análisis, comenzando con el diseño del transecto puntual y a continuación el diseño de transecto lineal.

1.3.1. Diseño del modelo de transecto puntual

Un diseño estándar de transecto puntual es extender puntos a lo largo del área de estudio, espaciados formando una malla. Un ejemplo de este diseño es el estudio de *Montrave*, el cual se puede observar en la siguiente imagen como un ejemplo de su aplicación.



Fuente: *Buckland* (2015)

El estudio de *Montrave* se ha aplicado para el cálculo de la abundancia de una zona de tamaño $33ha$ rodeada de pastos y cultivos. La zona está dividida en dos bloques. Para el diseño basado en transectos puntuales, se han generado una malla formada por 32 puntos (*indicadas por círculos*) colocados sobre el área de estudio, con una separación entre puntos vecinos de $100m$. Cada punto fue visitado dos veces. Para el diseño basado en transectos lineales hay 19 líneas (11 *en un bloque y 8 en el otro*), con una separación entre las líneas de $70m$. Cada línea fue recorrida dos veces.

Para más información sobre el estudio de *Montrave* se puede ver la sección 1.3 del libro de *Buckland* y otros de 2006 [6]

Se definen las unidades de muestreo como las intersecciones de los puntos de las líneas que generan la malla.

En algunos estudios, una unidad de muestreo aleatoria simple es preferible a una unidad de muestreo no aleatoria, para ello es conveniente definir la malla de cuadrados, de tal forma que la longitud de los lados de los cuadrados sea de al menos $2w$, donde w es la distancia a partir de la cual un individuo aun siendo observado no es considerado en la muestra.

Se coloca un punto en el centro de cada cuadrado de la malla, de tal forma que los círculos de observación no se solapen. Estos puntos serán los transectos puntuales.

Este tipo de diseños es altamente ineficiente debido a que el observador tiene que desplazarse de un punto a otro a grandes distancias. A menudo las unidades de muestreo son posicionadas a lo largo de líneas con una separación entre ellas de más de $2w$, si la separación es la misma

para todas las líneas, entonces forman una malla y los puntos pueden tratarse como una unidad de muestreo.

Cuando los transectos puntuales son localizados aleatoriamente dentro del área de estudio, algunos pueden estar a distancia menor de w de la frontera del área de estudio, por lo que el área de observación no es un círculo perfecto. En grandes estudios este tipo de puntos son ignorados, y en pequeños estudios la solución a este problema pasa por añadir el área faltante para completar el círculo al área del estudio.

Para ver más información véase *Buckland* y otros de 2006 [6].

1.3.2. Diseño del modelo de transecto lineal

El diseño estándar para transectos lineales depende en gran medida en si es práctico o eficiente recorrer la línea en el área de estudio. Por ejemplo, si en medio de la línea hay un obstáculo y este no se puede atravesar, el modelo no es práctico. En caso de que no existan obstáculos, el modelo estándar está compuesto por líneas paralelas separadas a la misma distancia una de otras.

Si es más práctico tener líneas cortas relativas a la dimensión del área de estudio, entonces se puede crear un conjunto de líneas paralelas colocadas aleatoriamente sobre el área de estudio. Es común que el coste de viajar de un transecto a otro sea elevado, por lo cual se suele usar diseños de zigzag, en los cuales el final de un transecto lineal es el inicio del siguiente transecto lineal. Esto genera diversos problemas debidos a la disponibilidad de que no existan obstáculos, ya que la longitud total de los transectos aumenta considerablemente, y además la distancia de observación de los transectos lineales van cambiando según la zona del área de estudio. Este y más problemas son resueltos en el libro *Buckland* y otros de 2004 [2].

Y de la misma forma que en el diseño de transecto puntual había ocasiones en las cuales la zona de observación no estaba totalmente dentro del área de estudio, para grandes estudios no se tienen en cuenta dichos transectos, y para estudios con áreas pequeñas se introduce el área faltante al área de estudio.

En muchas ocasiones el área de estudio no es homogénea y hay zonas donde no habitan los individuos de interés o la detectabilidad de los individuos es distinta al resto, entre otras muchas condiciones ambientales. Así que debido a que diferentes partes de un área de estudio puede tener muy diferentes características se introduce el concepto de *Estratificación*, en el cual el área de estudio se divide en áreas más pequeñas llamadas *bloques* y a cada *bloque* se le aplican los diseños de los modelos de transectos que se considere oportuno. Para más información ver el libro *Buckland* y otros de 2001 [1].

1.4. Desarrollo teórico de los modelos

El desarrollo del modelo matemático para el método de los transectos lineales y puntuales se basan en las siguientes hipótesis:

- **Hipótesis 1: Los objetos en la línea o en el punto son detectados con certeza.**

Esto es asumir que todos los objetos a distancia cero son detectados.

- **Hipótesis 2: Los objetos son detectados en su posición inicial.**

Importante para no tomar datos duplicados de un mismo individuo que ha cambiado de localización en el transcurso del experimento.

- **Hipótesis 3: Las mediciones son exactas.**

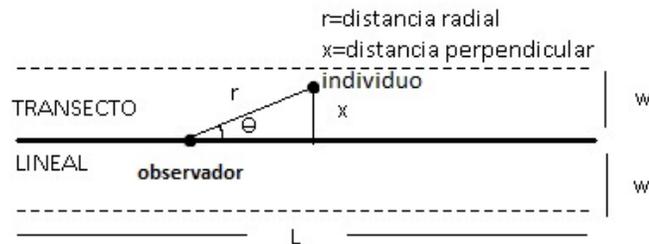
Las distancias obtenidas generalmente son exactas obviando errores de medición.

- **Hipótesis 4: Los individuos están distribuidos independientemente de las líneas o puntos.**

No hay relación entre la posición de los transectos y los individuos.

1.4.1. Transecto lineal.

Si se ha realizado la obtención de la muestra de las distancias y dichas distancias son radiales, es aconsejable pasar a distancias perpendiculares mediante el cambio $x_i = r_i \cdot \sin(\theta_i)$.



Si el observador detecta algún individuo fuera de la banda de observación, este individuo no debe ser recogido en la muestra. Por lo tanto, la banda del área observada es de anchura $2w$

y total de longitud L , obteniéndose un área de observación de $a = 2wL$. Si n individuos son detectados dentro de la banda, entonces la densidad número de objetos por unidad de área es estimada aplicando la definición de densidad por:

$$\hat{D} = n/2wL \quad (1.4.1)$$

En el muestreo por transecto lineal, solo una proporción de los objetos en el área son detectados. La proporción desconocida es P_a . Por ejemplo hemos observado 10 individuos y la proporción de detectados es $P_a = 0,5$, esto implica que el número aproximado de individuos en el área de estudio es $n/P_a = 20$.

Por lo tanto, introduciendo una constante de ajuste al número de individuos observado si P_a fuera conocido, la densidad sería estimada por:

$$\hat{D} = n/2wLP_a \quad (1.4.2)$$

Si P_a puede ser estimado de las distancias observadas, la densidad se estima por:

$$\hat{D} = n/2wL\hat{P}_a \quad (1.4.3)$$

Para la estimación de P_a se introduce el concepto de función de detectabilidad $g(x)$, que no es más que la probabilidad condicionada de que un individuo sea observado sabiendo que está a distancia x de la línea ($0 \leq x \leq w$). De la hipótesis de que todos los objetos a distancia cero son detectados, obtenemos que $g(0) = 1$, y además suponemos que la probabilidad de detectar a un individuo cuando la distancia se va haciendo grande tiende a cero, es decir $\lim_{x \rightarrow w} g(x) = 0$.

Para la obtención de P_a se usará la notación de N como el número de individuos que hay en la zona A , y se definirán las siguientes variables:

$$X_i = \text{distancia del individuo } i - \text{ésimo}, \text{ con } i = 1, \dots, N.$$

$$Z_i = \begin{cases} 1 & \text{individuo } i - \text{ésimo es observado}, \text{ con } i = 1, \dots, N. \\ 0 & \text{individuo } i - \text{ésimo no es observado}, \text{ con } i = 1, \dots, N. \end{cases}$$

Si la variable X sigue una distribución uniforme $X \sim U[0, w]$, entonces tendrá una función de densidad $f_X(x) = 1/w \cdot I_{[0, w]}(x)$ donde $I_{[0, w]}(x)$ es la función indicadora. Z_i sigue una distribución Bernoulli con probabilidad de éxito P_a .

$$P_a = P[Z_i = 1] = \int_0^w P[Z_i = 1/X_i = x] \cdot f_X(x)dx = \frac{1}{w} \cdot \int_0^w g(x)dx$$

$P[Z_i = 1/X_i = x_i]$ es la probabilidad de que un individuo sea observado suponiendo que está a distancia x_i , es decir, la función de detectabilidad en $X = x_i$, por lo que se puede sustituir la probabilidad por la función de detectabilidad $g(x)$.

Será denotada la cantidad $\int_0^w g(x)dx$ por μ .

Sustituyendo el estimador de P_a en la ecuación (1,4,2) obtenemos:

$$\hat{D} = \frac{n}{2L \int_0^w \hat{g}(x)dx} \quad (1.4.4)$$

Y si una estimación de μ es obtenida, entonces se puede escribir la ecuación como:

$$\hat{D} = n/2L\hat{\mu} \quad (1.4.5)$$

A continuación, se calculará una estimación de P_a .

Cada individuo de la población en el área de observación del transecto se caracteriza por:

$$X = \text{distancia perpendicular a la línea del transecto}$$

$$Z = \begin{cases} 1 & \text{individuo es observado} \\ 0 & \text{individuo no es observado} \end{cases}$$

Se puede suponer que $X \sim U[0, w]$, con función de densidad $f_X(x) = 1/w \cdot I_{[0,w]}(x)$.

La muestra es:

$(x_1, z_1), \dots, (x_n, z_n)$ con $z_1 = z_i = z_n = 1$, ya que las distancias x_i recogidas son aquellas que han sido observadas ($z_i = 1$). Es decir, x_1, \dots, x_n es una muestra aleatoria de la variable $X|_{Z=1}$.

La función de densidad de la variable $X|_{Z=1}$ será denotada por $f(x) = f_{X|Z=1}$.

A continuación se analiza la relación entre $f(x)$ y $g(x)$.

$$\begin{aligned} P[X \leq x | Z_i = 1] &= \frac{P[X \leq x, Z_i = 1]}{P[Z_i = 1]} = \frac{1}{P_a} P[Z_i = 1, X \leq x] = \frac{1}{P_a} \int_0^x P[Z_i = 1 | X=t] f_X(t) dt = \\ &= \frac{1}{P_a \cdot w} \int_0^x g(t) dt = \frac{1}{\mu} \int_0^x g(t) dt \end{aligned}$$

Por lo tanto la función de distribución de la variable $X|_{Z_i=1}$ es:

$$F_{X|_{Z_i=1}}(x) = \left[\frac{1}{\mu} \int_0^x g(t) dt \right] \cdot I_{[0,w]}(x) + I_{(w,+\infty)}(x)$$

Su función de densidad es

$$f(x) = f_{X|_{Z_i=1}}(x) = \frac{1}{\mu} g(x) \cdot I_{[0,w]}(x)$$

Debido a que $g(0) = 1$, la función de densidad evaluada en el cero es:

$$f(0) = \frac{1}{\mu} \tag{1.4.6}$$

De esta forma, se puede escribir de forma general el estimador de la densidad para el muestreo por transecto lineales sustituyendo el valor de P_a simplemente como :

$$\hat{D} = \frac{n \cdot \hat{f}(0)}{2L} = \frac{n}{2L\hat{\mu}} \tag{1.4.7}$$

\hat{D} es válido para w finito e infinito. La cantidad $\hat{f}(0)$ será estimada mediante los datos obtenidos por el método de transectos lineales, aplicando algún método de estimación de funciones de densidad.

1.4.2. Transecto puntual.

Se puede suponer la obtención de la muestra de las distancias, en las cuales el observador puede ver hasta distancia w y aun habiendo algún individuo más lejos, este no es recogido en la muestra.

Un transecto puntual genera un área observada πw^2 y, si se observan k transectos puntuales, se obtiene un área de observación de $a = k\pi w^2$. Si n individuos son detectados dentro del área de observación, entonces la densidad número de objetos por unidad de área es estimada aplicando la definición de densidad por :

$$\hat{D} = n/k\pi w^2 \tag{1.4.8}$$

En el muestreo por transecto puntual, solo una proporción de los objetos en el área son detectados. La proporción desconocida es P_a . Y de la misma forma que pasaba en el caso del transecto lineal se puede introducir una constante de ajuste al número de individuos observado. Si P_a fuera conocido, la densidad sería estimada por :

$$\hat{D} = n/k\pi w^2 P_a \quad (1.4.9)$$

Si P_a puede ser estimado de las distancias observadas, la densidad se estima por:

$$\hat{D} = n/k\pi w^2 \hat{P}_a \quad (1.4.10)$$

Para la estimación de P_a se usa la función de detectabilidad $g(r)$, que nos da la probabilidad condicionada de que un individuo sea detectado sabiendo que está a distancia r del punto ($0 \leq r \leq w$). De la hipótesis de que todos los individuos a distancia cero son detectados, se obtiene que $g(0) = 1$, y además se puede suponer que la probabilidad de detectar a un individuo cuando la distancia se va haciendo grande tiende a cero, es decir $\lim_{r \rightarrow w} g(r) = 0$.

A continuación, se calculará una estimación de P_a .

Cada individuo de la población en el área de observación del transecto se caracteriza por:

$R =$ *distancia radial desde el punto del transecto*

$$Z = \begin{cases} 1 & \text{individuo es observado} \\ 0 & \text{individuo no es observado} \end{cases}$$

Se puede suponer que los individuos están uniformemente distribuidos en todo el círculo de observación.

Así:

$$P[R \leq r] = \frac{r^2}{w^2} \quad 0 \leq r \leq w$$

$$P[R \leq r] = 1 \quad r \geq w$$

En consecuencia, la función de densidad de R viene dada por:

$$f_R(r) = \frac{2r}{w^2} \cdot I_{[0,w]}(r)$$

La muestra es:

$(r_1, z_1), \dots, (r_n, z_n)$ con $z_1 = z_i = z_n = 1$, ya que las distancias r_i recogidas son aquellas que han sido observadas ($z_i = 1$). Es decir, r_1, \dots, r_n es una muestra aleatoria de la variable $R|_{Z=1}$.

La función de densidad de la variable $R|_{Z=1}$ será denotada por $f(r) = f_{R|_{Z=1}}$.

La detectabilidad P_a es la probabilidad de que $Z_i = 1$.

$$P_a = P[Z_i = 1] = \int_0^w P[Z_i = 1|_{R_i=r}] \cdot f_R(r) dr = \dots = \frac{2}{w^2} \int_0^w r g(r) dr$$

donde $P[Z_i = 1|_{R_i=r}]$ es la probabilidad de que un individuo sea detectado suponiendo que está a distancia r , es decir, la detectabilidad en r , $g(r)$.

Ahora se analiza la relación entre $f(x)$ y $g(x)$.

$$\begin{aligned} P[R \leq r|_{Z_i=1}] &= \frac{P[R \leq r, Z_i = 1]}{P[Z_i = 1]} = \frac{1}{P_a} P[Z_i = 1, R \leq r] = \frac{1}{P_a} \int_0^r P[Z_i = 1|_{R=t}] f_R(t) dt = \\ &= \frac{2}{P_a \cdot w^2} \int_0^r t g(t) dt \end{aligned}$$

Por lo tanto la función de distribución de la variable $R|_{Z_i=1}$ es:

$$F_{R|_{Z_i=1}}(x) = \left[\frac{2}{P_a \cdot w^2} \int_0^r t g(t) dt \right] \cdot I_{[0,w]}(r) + I_{(w,+\infty)}(r)$$

Su función de densidad por lo tanto será:

$$f(r) = f_{R|_{Z_i=1}}(r) = \frac{2r \cdot g(r)}{w^2 \cdot P_a} \cdot I_{[0,w]}(r)$$

Así que $f(r)$ nos queda:

$$f(r) = \frac{2r \cdot g(r)}{w^2 \cdot P_a} \tag{1.4.11}$$

Ahora se denotará $\nu = 2 \int_0^w r g(r) dr$. Como $f(x)$ es una función de densidad tiene que integrar 1, obteniendo:

$$w^2 \cdot P_a = 2 \int_0^w r g(r) dr = \nu$$

De donde se obtiene que $P_a = \nu/w^2$. Sustituyendo en (1,4,11) se obtiene:

$$f(r) = \frac{2rg(r)}{2 \int_0^w rg(r)dr} \quad (1.4.12)$$

Pudiendo escribir $f(r)$ como:

$$f(r) = \frac{2rg(r)}{\nu} \quad (1.4.13)$$

Ahora la proporción P_a puede ser expresada como el área de observación entre el área total

$$P(a) = \frac{k\nu}{k\pi w^2} = \frac{\nu}{\pi w^2} \quad (1.4.14)$$

Por otro lado se tiene que

$$f(r)/r = \frac{2g(r)}{\nu} \quad (1.4.15)$$

Y además como $g(0) = 1$ se tiene

$$\lim_{r \rightarrow 0} f(r)/r = \frac{2}{\nu} \quad (1.4.16)$$

El lado izquierdo de la ecuación anterior es la pendiente de $f(r)$ en $r = 0$, a la que denominaremos $h(0)$.

$$\nu = \frac{2}{h(0)} \quad (1.4.17)$$

Además habiendo ajustado nuestro modelo por $f(r)$, podemos estimar $h(0)$ (por $\hat{h}(0)$), de lo cual se obtiene

$$P(a) = \frac{2}{\pi w^2 \hat{h}(0)} \quad (1.4.18)$$

Sustituyendo la estimación de P_a en la ecuación (1,4,10) obtenemos

$$\hat{D} = \frac{n\hat{h}(0)}{2k\pi} \quad (1.4.19)$$

La cantidad $\hat{h}(0)$ será estimada mediante los datos recogidos por el método de transecto puntual.

1.5. Ilustración de muestreo por transecto

Se recoge en este apartado un ejemplo del libro *Burnham* de 1980 [7], el cual puede ser encontrado en la página 61 de dicho libro, con motivo de ilustrar el muestreo por transecto mediante el uso de datos reales. A lo largo de este trabajo se va a utilizar el estudio que a continuación se describe, para el cálculo de los distintos estimadores que serán propuestos.

1.5.1. Datos de estacas de madera

TABLE 6.—EXAMPLE OF LINE TRANSECT DATA FROM A SAMPLE OF WOODEN STAKES OF KNOWN DENSITY; $n = 68$, $w^* = 18.60$ m, $w = 20$ m, $L = 1,000$ m, AND $D = 0.00375$ STAKES PER METER. THE DATA SHOWN ARE THE PERPENDICULAR DISTANCES IN METERS

1.	2.02	35.	3.79
2.	0.45	36.	15.24
3.	10.40	37.	3.47
4.	3.61	38.	3.05
5.	0.92	39.	7.93
6.	1.00	40.	18.15
7.	3.40	41.	10.05
8.	2.90	42.	4.41
9.	8.16	43.	1.27
10.	6.47	44.	13.72
11.	5.66	45.	6.25
12.	2.95	46.	3.59
13.	3.96	47.	9.04
14.	0.09	48.	7.68
15.	11.82	49.	4.89
16.	14.23	50.	9.10
17.	2.44	51.	3.25
18.	1.61	52.	8.49
19.	31.31	53.	6.08
20.	6.50	54.	0.40
21.	8.27	55.	9.33
22.	4.85	56.	0.53
23.	1.47	57.	1.23
24.	18.60	58.	1.67
25.	0.41	59.	4.53
26.	0.40	60.	3.12
27.	0.20	61.	3.05
28.	11.59	62.	6.60
29.	3.17	63.	4.40
30.	7.10	64.	4.97
31.	10.71	65.	3.17
32.	3.86	66.	7.67
33.	6.05	67.	18.16
34.	6.42	68.	4.08

Fuente: *Burnham* (1980)

Laake en 1979 realizó una investigación del muestreo por transectos lineales colocando un número conocido de estacas de madera en un prado al este de *Logan, Utah*. Dichos transectos fueron recorridos por estudiantes graduados en *Ciencias de Vida Salvaje* de la universidad de *Utah*.

La muestra de los datos fueron realizadas, anotando las distancias perpendicular al transecto lineal con las cuales se pretende estimar la densidad del área. Se colocaron 150 estacas de madera de forma aleatoria en un área de $L = 1000$ metros de largo, de las cuales se observaron 68 usando el muestreo por transecto lineal con una distancia de observación máxima de $w = 20$ metros. La verdadera forma de la función de densidad $f(x)$ es desconocida, pero el valor de $f(0)$ es conocido y vale $f(0) = 0,110294$. Además, la densidad de las estacas vale

0.00375 o lo que equivale a 37,5 *estacas/hectáreas*. Los datos recogido en la tabla muestran un valor $x_{19} = 31,31$, que corresponde con un valor extraño y más cuando se considera $w = 20$, por ello se va a considerar dos muestras una con $n = 68$ y otra con $n = 67$ eliminando el valor extraño $x_{19} = 31,31$, en el cual se puede considerar $w^* = 18,60$.

Capítulo 2

Estimación de las funciones asociadas a la densidad.

Para la obtención de una estimación de $\hat{f}(0)$ y $\hat{h}(0)$ se pueden usar métodos no paramétricos (usando por ejemplo estimación por funciones núcleos), métodos paramétricos en los cuales conocemos la familia paramétrica a la que pertenece la función de densidad $f(x)$ en el caso de transecto lineal y $f(r)$ en el caso de transecto puntual, y métodos semiparamétricos que combinan métodos paramétricos y no paramétricos. Se empezará introduciendo los métodos paramétricos, especialmente el caso del método paramétrico en el software R. Luego se expondrá el método no paramétrico basado en la función núcleo, y cómo este método puede ser mejorado mediante el uso de métodos paramétricos, a lo que se llamará métodos semiparamétricos.

2.1. Condición de la función de detectabilidad

Burnham [7] sugiere que la función de detectabilidad $g(x)$ debe cumplir que a distancias muy cercanas al transecto lineal, es decir, a distancias muy cercanas al origen, la función de detectabilidad tome un valor muy cercano a uno. Esto puede expresarse matemáticamente como que $g'(0) = 0$ o equivalentemente que $f'(0) = 0$. Para más información véase [1]. Esta condición es fundamental y será usada a lo largo de este trabajo.

2.2. Métodos paramétricos

Se empezará abordando el caso concreto del muestreo por transecto lineal, ya que el caso del muestreo por transecto puntual es similar. Como se recoge en el capítulo anterior, estimar D equivale a estimar $f(0)$, y en un enfoque paramétrico se tomaría $f(x; \theta)$ como una familia paramétrica con θ un vector de parámetros desconocidos. Este vector de parámetros se ha de estimar y, para ello sólo se dispone de las distancias perpendiculares observadas X_1, \dots, X_n . En este apartado se pretende introducir distintos tipos de funciones de detectabilidad, trabajar con ellas para obtener distintos tipos de propiedades estadísticas y proponer un modelo

paramétrico.

2.2.1. Función de detectabilidad exponencial negativa

La función de detectabilidad exponencial negativa propuesta por *Gates* en 1968 [8] viene dada por

$$g(x; \theta) = e^{-x/\theta}; \quad x \geq 0, \theta > 0. \quad (2.2.1)$$

Por tanto, la función de densidad de probabilidad es

$$f(x; \theta) = \frac{1}{\theta} e^{-x/\theta}; \quad x \geq 0, \theta > 0. \quad (2.2.2)$$

Entonces si se usa el método de máxima verosimilitud se obtiene un estimador para $f(0)$, $\hat{f}_{EMV}(0) = 1/\bar{X}$, donde \bar{X} es la media muestral.

Sin embargo, la función de detectabilidad debe cumplir que en el origen tome el valor 1 ($g(0) = 1$), lo que significa que a distancias cercanas del transecto lineal debe tomar valores cercanos a 1. Esto se puede expresar como que $g'(0) = 0$ o equivalentemente $f'(0) = 0$.

Propiedad que la función de detectabilidad exponencial negativa no cumple.

2.2.2. Función de detectabilidad semi-normal

La función de detectabilidad semi-normal propuesta por *Hemingway* en 1971 [9] viene dada por

$$g(x; \sigma^2) = e^{-x^2/2\sigma^2}; \quad x \geq 0, \sigma^2 > 0. \quad (2.2.3)$$

Por tanto, la función de densidad de probabilidad es

$$f(x; \sigma^2) = \frac{2}{\sigma\sqrt{2\pi}} e^{-x^2/2\sigma^2}; \quad x \geq 0, \sigma^2 > 0. \quad (2.2.4)$$

Y se satisface la propiedad ($g'(0) = 0$), ya que $g'(x; \sigma^2) = -\frac{x}{\sigma^2} e^{-x^2/2\sigma^2}$, de donde se obtiene $g'(0; \sigma^2) = 0$.

Entonces si se usa el método de máxima verosimilitud se obtiene un estimador para $f(0; \sigma^2)$, $\hat{f}_{EMV}(0) = \sqrt{\frac{2}{\pi T}}$, donde $T = \sum_{i=1}^n x_i^2/n$.

Otros posibles estimadores para $f(0; \sigma^2)$ pueden ser:

- Estimador insesgado de mínima varianza (*EIMV*). Propuesto por *Quinn y Gallucci* en 1980 [10]

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

- Estimador contraído *EC*. Propuesto por *Zhang* en 2009 [11]

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

$\hat{f}_{EC}(0)$ es un estimador sesgado de $f(0)$, pero logra un menor error cuadrático medio. Existe una relación entre los estimadores $\hat{f}_{EC}(0)$, $\hat{f}_{EIMV}(0)$ y $\hat{f}_{EMV}(0)$, que viene dada por:

$$\hat{f}_{EC}(0) = \frac{n-2}{n} \beta^2(n) \hat{f}_{EIMV}(0) = \frac{n-2}{n} \beta(n) \hat{f}_{EMV}(0)$$

Entonces los tres estimadores son asintóticamente equivalentes.

2.2.3. Otras funciones de detectabilidad

Se recogen a continuación otras funciones de detectabilidad propuestas.

En primer lugar, la función de detectabilidad propuesta por *Ababneh* y *Eidous* [12] dada por

$$g(x; \theta) = e^{-\theta x} (2 - e^{-\theta x}); \quad x \geq 0, \theta > 0. \quad (2.2.5)$$

con la función de probabilidad

$$f(x; \theta) = \frac{2\theta}{3} e^{-\theta x} (2 - e^{-\theta x}); \quad x \geq 0, \theta > 0. \quad (2.2.6)$$

La función de detectabilidad satisface que $g'(0) = 0$, y $g(0; \theta) = 1$ despejando en la función de probabilidad se obtiene que $f(0; \theta) = \frac{2\theta}{3}$.

Se puede expresar la función de probabilidad en función de $f(0)$ como

$$f(x; \theta) = f(0) e^{-3f(0)x/2} (2 - e^{-3f(0)x/2}); \quad x \geq 0, \theta > 0. \quad (2.2.7)$$

Ababneh y *Eidous* en el artículo del año 2012 [12] recogen el valor esperado $\mu = 7/[9f(0)]$. Por el cual, a partir del método de los momentos se puede obtener el estimador para $f(0)$, $\hat{f}_{EMM}(0) = 7/(9\bar{X})$. Y a partir del método de máxima verosimilitud, usando las propiedades asintóticas se puede obtener que $\hat{f}_{EMV}(0)$ sigue una $N(f(0), \sigma_{f(0)}^2)$, donde $\sigma_{f(0)}^2 = \frac{f^2(0)}{0,80153n}$. Con $\hat{f}_{EMV}(0)$ el estimador de $f(0)$ obtenido por máxima verosimilitud.

Esta función de detectabilidad se puede generalizar incluyendo un nuevo parámetro:

$$g(x; \theta, \beta) = e^{-\theta x}(2 - e^{-\theta x/\beta})^\beta; \quad x \geq 0, \theta, \beta > 0. \quad (2.2.8)$$

con la correspondiente función de probabilidad dada por

$$f(x; \theta, \beta) = \frac{2\theta\Gamma(\beta+1/2)}{\sqrt{\pi}\Gamma(\beta+1)+\Gamma(\beta+1/2)} e^{-\theta x}(2 - e^{-\theta x/\beta})^\beta; \quad x \geq 0, \theta, \beta > 0. \quad (2.2.9)$$

Si se propone el modelo paramétrico de la ecuación anterior para estimar la densidad de población basada en datos obtenidos mediante transectos lineales, se tiene que $g(0; \theta, \beta) = 1$ y $g'(0; \theta, \beta) = 0$, de tal forma que $g(x; \theta, \beta)$ cumple las condiciones esperadas para cualquier valor de θ . Según *Saeed* [13], basándose en un estudio de simulación, el modelo posee buenas propiedades estadísticas comparadas con otros modelos.

Por ejemplo, para el caso de $\beta = 1$ se tiene el caso particular de la ecuación (2.2.8), *Al Eibood y Eidous* [14] estudian este caso particular del modelo para datos agrupados, el estimador de máxima verosimilitud para $f(0)$ y la densidad de población D vienen dada por

$$\hat{f}_G(0) = \frac{2\hat{\theta}}{3} \quad y \quad \hat{D}_G = \frac{n\hat{f}_G(0)}{2L}$$

El estimador de máxima verosimilitud se puede calcular numéricamente resolviéndose la ecuación

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

Se denotará *datos agrupados* a aquellos datos que tienen como característica principal la frecuencia con que se presentan, es decir, aquellos que se encuentran contados y clasificados. Una vez anotadas las distancias a la que se han observado los individuos, si w es la distancia máxima de observación, se clasifican en m intervalos de amplitud w/m . En caso contrario, se dirá que los *datos son no agrupados*.

Con motivo de ilustrar los estimadores de máxima verosimilitud para θ , a partir del cual se puede estimar $f(0)$ y la densidad de población D . Se recogen los resultados obtenidos por *Al Eibood y Eidous* [14] para datos agrupados de los datos de estacas de madera, los cuales vienen recogida en la siguiente tabla. Donde *Interval Number* denota el número de intervalos usados en la clasificación de los datos.

Por otro lado, el estimador propuesto para datos no agrupados, proporciona las estimaciones $\hat{f}(0) = 0,12647$ y $\hat{D} = 0,0043$ *estacas/hectáreas*, resultando por lo tanto, +que el estimador propuesto para datos agrupados es muy cercano al estimador propuesto para datos no agrupados.

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

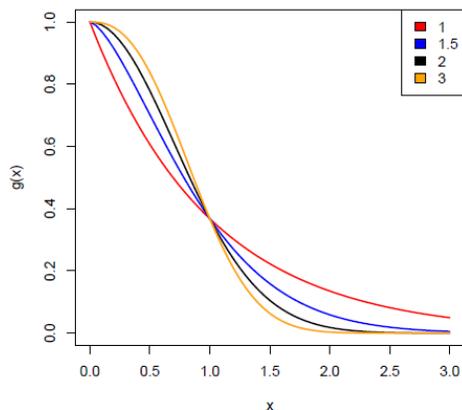
Fuente: *Al Eibood y Eidous (2017)*

2.2.4. Modelo propuesto

Supuesto que se tiene n distancias perpendiculares obtenidas a partir de transecto lineales X_1, \dots, X_n , las cuales son independientes e idénticamente distribuidas. Si se considera la función de detectabilidad $g(x; \sigma^2)$ con σ^2 un parámetro desconocido dada por

$$g(x; \sigma^2) = e^{-x^2/2\sigma^2} (2 - e^{-x^2/2\sigma^2}); \quad x \geq 0, \sigma^2 > 0. \quad (2.2.10)$$

La forma de la función de detectabilidad para los valores de $\sigma^2 = 1, 1.5, 2, 3$ vienen representada por la siguiente figura



Fuente: *Saeed (2017)*

La probabilidad de detectar a un individuo dado que su distancia perpendicular al transecto lineal es 1 ($g(0; \sigma^2) = 1$), indica certeza a la hora de detectar un individuo sobre el transecto lineal. La primera derivada de la función de detectabilidad $g(x; \sigma^2)$ con respecto a x es

$$g'(x; \sigma^2) = \frac{2}{\sigma^2} x e^{-x^2/2\sigma^2} (e^{-x^2/2\sigma^2} - 1)$$

de lo cual se obtiene que $g'(0; \sigma^2) = 0$ para cualquier valor de σ^2 .

La función de probabilidad de $g(x; \sigma^2)$ puede ser obtenida a partir de

$$f(x; \sigma^2) = \frac{g(x; \sigma^2)}{\int_0^\infty g(t; \sigma^2) dt}$$

Entonces la función de probabilidad viene dada por

$$f(x; \sigma^2) = \frac{2}{(2\sqrt{2} - 1)\sigma\sqrt{\pi}} e^{-x^2/2\sigma^2} (2 - e^{-x^2/2\sigma^2}); \quad x \geq 0, \sigma^2 > 0. \quad (2.2.11)$$

Analizando el comportamiento de la función de detectabilidad y de la función de probabilidad, se puede expresar $f(x; \sigma^2)$ en términos de $f(0)$ de la siguiente manera.

Se tiene que $g(0; \sigma^2) = 1$, por lo tanto el parámetro $f(0; \sigma^2)$ viene dado por

$$f(0; \sigma^2) = \frac{2}{(2\sqrt{2} - 1)\sigma\sqrt{\pi}} \quad (2.2.12)$$

Y $g'(0; \sigma^2) = 0$, además que la función de detectabilidad $g(x; \sigma^2)$ es una función monótona decreciente para $x \in (0, \infty)$, por lo cual esto se cumple si y solo si

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

y como se tiene que $0 \leq e^{-x^2/2\sigma^2} \leq 1$, $\forall x \in (0, \infty)$ se tiene que $e^{-x^2/2\sigma^2} - 1 < 0$, lo cual es cierto $\forall x > 0$ y para cualquier σ^2 .

De la misma forma, $f(x; \sigma^2)$ es una función monótona decreciente para $x \in (0, \infty)$. Por lo que se puede expresar $f(x; \sigma^2)$ en términos de $f(0)$ como

$$f(x; \sigma^2) = f(0) e^{-((2\sqrt{2}-1)\sqrt{\pi}f(0))^2 x^2/8} (2 - e^{-((2\sqrt{2}-1)\sqrt{\pi}f(0))^2 x^2/8}); \quad x \geq 0. \quad (2.2.13)$$

2.2.5. Propiedades estadísticas

Para el modelo de la ecuación (2.2.10), *Saeed* [15] obtiene los momentos de orden 1 y 2

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

De donde se obtiene

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

2.2.6. Construcción de estimadores

En este apartado se considera la inferencia estadística sobre el parámetro $f(0)$, centrándose en el método de los momentos y en el método de máxima verosimilitud.

2.2.6.1. Estimador por el método de los momentos para $f(0)$

Si se considera el modelo con la ecuación (2.2.13) el momento de orden 1 viene dado por

$$E(x) = \frac{6}{(2\sqrt{2} - 1)^2 \pi f(0)}$$

se puede obtener un estimador para $f(0)$, $\hat{f}(0) = \frac{6}{(2\sqrt{2}-1)^2 \pi \bar{X}}$. Donde $\hat{f}(0)$ es el estimador por el método de los momentos para $f(0)$.

2.2.6.2. Estimador por el método de máxima verosimilitud para $f(0)$

El estimador por el método de máxima verosimilitud se puede obtener usando máxima verosimilitud sobre el parámetro σ^2 en la ecuación (2.2.11) o directamente usando la función logaritmo de la función de máxima verosimilitud, para el modelo con la ecuación en función de $f(0)$ (2.2.13).

El logaritmo de la función de máxima verosimilitud para (2.2.13) viene dada por

$$\ln(L(f(0))) = n \cdot \ln(f(0)) - \frac{(2\sqrt{2}-1)^2 \pi f^2(0)}{8} \sum_{i=1}^n x_i^2 + \sum_{i=1}^n \ln(2 - e^{-(2\sqrt{2}-1)^2 \pi f^2(0) x_i^2 / 8}) \quad (2.2.14)$$

Donde $L(f(0))$ es la función de máxima verosimilitud con parámetro $f(0)$.

El estimador por el método de máxima verosimilitud para $f(0)$ se puede obtener derivando e igualando a cero

$$\frac{n}{f(0)} - \frac{(2\sqrt{2} - 1)^2 \pi f(0)}{4} \sum_{i=1}^n x_i^2 + \frac{(2\sqrt{2} - 1)^2 \pi f(0)}{4} \sum_{i=1}^n \frac{x_i^2}{(2e^{(2\sqrt{2}-1)^2 \pi f^2(0) x_i^2 / 8} - 1)} = 0 \quad (2.2.15)$$

Se puede resolver esta ecuación usando métodos numéricos como el método de *Newton – Rapshon*. El estimador por el método de máxima verosimilitud es consistente y asintóticamente normal, en concreto, con distribución asintótica a $N(f(0), I^{-1}(f(0)))$, donde

$$I(f(0)) = -E\left(\frac{\partial^2 \log(L(f))}{\partial f^2}\right) = \frac{4,88758n}{f^2(0)}.$$

2.3. Aplicación de los métodos paramétricos.

A continuación, se expondrá unos modelos especiales, centrados en el software R.

2.3.1. Obtención de $g(y)$.

El software *Distance* de R provee de tres modelos para la función de detectabilidad, que consiste en una *función clave* y en un ajuste llamado *serie de expansión*.

$$g(y) = key(y)[1 + serie(y)] \quad (2.3.1)$$

Los seis modelos se recogen en la tabla siguiente:

Función clave	Serie de expansión
<i>Uniforme</i> , $1/w$, $0 \leq y \leq w$	<i>Coseno</i> , $\sum_{j=1}^m a_j \cos(j\pi y/w)$ <i>Polinomio simple</i> , $\sum_{j=1}^m a_j (y/w)^{2j}$
<i>Semi - Normal</i> , $\exp(-y^2/2\sigma^2)$, $0 \leq y \leq w$	<i>Coseno</i> , $\sum_{j=2}^m a_j \cos(j\pi y/w)$ <i>Polinomio Hermite</i> , $\sum_{j=2}^m a_j H_{2j}(y_s)$
<i>Tasa de riesgo</i> , $1 - \exp(-(y/\sigma)^{-b})$, $0 \leq y \leq w$	<i>Coseno</i> , $\sum_{j=2}^m a_j \cos(j\pi y/w)$ <i>Polinomio simple</i> , $\sum_{j=2}^m a_j (y/w)^{2j}$

Cuadro 2.1: Los modelos más usados de $g(y)$ en el software *Distance* en R.

2.3.2. Obtención de $f(y)$.

En la práctica es razonable que la verdadera función de densidad sea cercana a una forma paramétrica conocida, por lo cual $f(y)$ se puede escribir como:

$$f(y) = \frac{\alpha(y)}{\beta} \cdot [1 + \sum_{j=1}^m a_j \cdot p_j(y_s)] \quad (2.3.2)$$

donde $\alpha(y)$ es una función clave paramétrica, conteniendo k parámetros (usualmente 0, 1 o 2)

$$p_j(y_s) = \begin{cases} y_s^j, & \text{si un polinomio simple es requerido, o} \\ H_j(y_s), & \text{el } j\text{-ésimo polinomio de Hermite, } j=1, \dots, m, \text{ o} \\ \cos(j\pi y_s), & \text{si una serie de coseno es requerida} \end{cases}$$

y_s es un valor de y estandarizado ($y_s = y/w$ con w definido como anteriormente).

$a_j = 0$ si el término j de $p_j(y_s)$ no es usado en el modelo, o es estimado por máxima verosimilitud.

β es una función normalizada de los parámetros (parámetros de la función clave y de series de coeficientes).

2.3.3. Estimación por intervalos de la densidad.

De la definición de la estimación de densidad para transecto lineales

$$\hat{D} = \frac{n \cdot \hat{f}(0)}{2L} \quad (2.3.3)$$

Una aproximación de la varianza de \hat{D} viene dada por

$$\hat{v}ar(\hat{D}) = \hat{D}^2 \cdot \left\{ \frac{\hat{v}ar(n)}{n^2} + \frac{\hat{v}ar[\hat{f}(0)]}{[\hat{f}(0)]^2} \right\} \quad (2.3.4)$$

Equivalentemente las expresiones para muestras de transecto puntual son

$$\hat{D} = \frac{n \cdot \hat{h}(0)}{2\pi k} \quad (2.3.5)$$

y

$$\hat{v}ar(\hat{D}) = \hat{D}^2 \cdot \left\{ \frac{\hat{v}ar(n)}{n^2} + \frac{\hat{v}ar[\hat{h}(0)]}{[\hat{h}(0)]^2} \right\} \quad (2.3.6)$$

Estimaciones de $f(0)$ y $h(0)$ y correspondientes a las estimaciones de las varianzas, son obtenidas a partir de la matriz de información de la función de verosimilitud.

Habiendo obtenido \hat{D} y $\hat{v}ar(\hat{D})$, se puede obtener un intervalo de confianza asintótico al $100(1 - \alpha)\%$

$$\hat{D} \pm z_\alpha \cdot \sqrt{\hat{v}ar(\hat{D})} \quad (2.3.7)$$

donde z_α es el percentil de orden $1 - \alpha$ de la distribución $N(0, 1)$.

Sin embargo, la distribución de \hat{D} es sesgada positivamente, y un intervalo mejor es obtenido

asumiendo que \hat{D} sigue una distribución logarítmica-normal. Obteniendo un intervalo de confianza $100(1 - \alpha)\%$ dado por

$$(\hat{D}/C, \hat{D} \cdot C) \quad (2.3.8)$$

donde

$$C = \exp[z_\alpha \cdot \sqrt{\hat{\text{var}}(\log_e \hat{D})}] \quad (2.3.9)$$

y

$$\hat{\text{var}}(\log_e \hat{D}) = \log_e \left[1 + \frac{\hat{\text{var}}(\hat{D})}{\hat{D}^2} \right] \quad (2.3.10)$$

Si se desea obtener más información acerca de esta sección puede verse más detallado en la subsección 3,6,1 del libro de *Buckland* y otros del año 2001 [1].

2.4. Método no paramétrico basado en funciones núcleos

Los métodos paramétricos proporcionan una muy buena estimación, incluso mejor que los métodos no paramétricos. Sin embargo, el método paramétrico es fuertemente dependiente de las especificaciones del modelo. Por ello en esta sección se desarrollará el método no paramétrico basado en las funciones núcleos.

2.4.1. El estimador

Sea X_1, \dots, X_n una muestra aleatoria de distancias perpendiculares de tamaño n , con función de detectabilidad desconocida $g(x)$ y función de probabilidad $f(x)$.

El clásico estimador núcleo $\hat{f}_k(x)$ para $f(x)$ con $x \geq 0$ propuesto por *Chen* en 1996 [16] es

$$\hat{f}_k(x) = \frac{1}{nh} \sum_{i=1}^n \left[K\left(\frac{x - X_i}{h}\right) + K\left(\frac{x + X_i}{h}\right) \right] \quad (2.4.1)$$

donde K es la función núcleo y h es la ventana. En los métodos basados en estimadores núcleos, la ventana es más importante que la elección de la función núcleo. La ventana especifica la cantidad de suavizado que se le aplica a los datos y controla el rendimiento de $\hat{f}_k(x)$.

En el muestreo por transecto lineal, el interés reside en el valor de la densidad en $x = 0$, $f(0)$, debido a que esta cantidad está relacionada a la densidad de población.

Usandose la técnica de simetrización usada por *Chen* en 1996 [16], la cual fue originalmente sugerida por *Buckland* en 1992 [17], se puede obtener un estimador núcleo para $f(0)$.

La idea de la técnica de simetrización viene dada por la reflexión de los datos con respecto el 0, es decir, duplicar los datos de la siguiente manera:

Supuesto que se tiene un valor de los datos x_i , se reemplaza por x_i y su reflexión sobre el 0 $-x_i$. Entonces los datos consiste en los valores y_1, \dots, y_{2n} donde $y_{2i-1} = x_i$ y $y_{2i} = -x_i$.

De tal forma se tiene que $X = |Y|$, y

$$f(x) = g(x) + g(-x)$$

donde g es la densidad de y . El estimador núcleo de g es

$$\hat{g}_k(x) = \frac{1}{2nh} \sum_{k=1}^{2n} K\left(\frac{x - y_k}{h}\right) = \frac{1}{2nh} \sum_{k=1}^n \left[K\left(\frac{x - x_i}{h}\right) + K\left(\frac{x + x_i}{h}\right) \right]$$

pudiendose obtener un estimador de $f(0)$,

$$\hat{f}_k(0) = 2\hat{g}_h(0) = \frac{1}{nh} \sum_{k=1}^{2n} K\left(\frac{0 - y_k}{h}\right) = \frac{2}{nh} \sum_{k=1}^n K\left(\frac{x_i}{h}\right)$$

donde la última igualdad es debido a la simetría de la función núcleo, generalmente supuesta que sigue una $N(0, 1)$.

Por lo que se obtiene que el estimador núcleo $\hat{f}_k(x)$ para $f(x)$ es

$$\hat{f}_k(0) = \frac{2}{nh} \sum_{i=1}^n K\left(\frac{X_i}{h}\right). \quad (2.4.2)$$

Si se considera las hipótesis $f'(0) = 0$ y $g'(0) = 0$, entonces *Chen* [16] demuestra que el sesgo asintótico de (2.4.2) es $h^2 f''(0) \int_0^\infty u^2 K(u) du$ bajo la condición de que $f(x)$ tiene su segunda derivada continua. Además, el estimador propuesto en este caso es

$$\hat{f}(0) = \hat{f}_k(0) - h^2 f''(0) \int_0^\infty u^2 K(u) du \quad (2.4.3)$$

donde $f''(0)$ es la derivada segunda de $f(x)$ en $x = 0$. El segundo término de (2.4.3) se puede entender como una función de ajuste y $f(x)$ es la densidad desconocida, la cual tiene derivadas continuas para todo los órdenes requeridos, entonces el valor esperado de $\hat{f}(0)$ para una muestra de tamaño n es

$$E(\hat{f}(0)) = \frac{2}{h} \int_0^\infty K\left(\frac{x_1}{h}\right) f(x_1) dx_1 - h^2 f''(0) \int_0^\infty u^2 K(u) du$$

$$= 2 \int_0^{\infty} K(t)f(ht)dt - h^2 f''(0) \int_0^{\infty} u^2 K(u)du$$

donde se ha aplicado el cambio de variable $x_1 = ht$. Usando la serie de Taylor para expandir $f(ht)$ en un entorno del cero, entonces si $h \rightarrow 0$ con $n \rightarrow \infty$ el sesgo de $\hat{f}(0)$ bajo las correctas condiciones de diferenciabilidad de f en un entorno del cero es

$$Sesgo(\hat{f}(0)) = \frac{2h^3}{3!} f''(0) \int_0^{\infty} t^3 K(t)dt + \frac{2h^4}{4!} f^{iv}(0) \int_0^{\infty} t^4 K(t)dt + O(h^5) \quad (2.4.4)$$

Chen [16] también señala de (2.4.4) que bajo las condiciones de diferenciabilidad de f en un entorno de cero, el sesgo del estimador propuesto es $O(h^3)$ frente al estimador núcleo que bajo las mismas condiciones es $O(h^2)$.

Si además, es supuesto que la tercera derivada de $f(x)$ es cero en $x = 0$ se obtiene el caso de que el sesgo es $O(h^4)$.

Si h está relacionada a n de tal manera que $h \rightarrow 0$ y $nh \rightarrow \infty$ cuando $n \rightarrow \infty$, entonces la varianza de $\hat{f}(0)$ es

$$\begin{aligned} var(\hat{f}(0)) &= \frac{4}{nh^2} Var(K(\frac{X_1}{h})) \\ &= \frac{4}{nh} f(0) \int_0^{\infty} K^2(t)dt + O(n)^{-1} \end{aligned}$$

Entonces la varianza es $O(nh)^{-1}$, lo mismo que en el caso de la varianza del estimador núcleo *Chen* [16].

El error cuadrático medio asintótico (*AMSE*) del estimador propuesto $\hat{f}(0)$ bajo la hipótesis que $f'(0) = f'''(0) = 0$ viene dada por

$$AMSE(\hat{f}(0)) = \frac{4}{nh} f(0) \int_0^{\infty} K^2(t)dt + \frac{4h^8}{576} \left[f^{iv}(0) \int_0^{\infty} t^4 K(t)dt \right]^2 \quad (2.4.5)$$

donde el primer término de (2.4.5) es la varianza y el segundo término es el sesgo al cuadrado, viene detallado en el libro de *Chen* [16].

2.4.2. Selección de la ventana

El principal problema de las estimaciones no paramétricas es la selección del valor de la ventana h . Y fijándose en el sesgo y varianza de $\hat{f}(0)$ calculado previamente, se puede observar que al aumentar h el sesgo aumenta y la varianza disminuye. Para obtener el valor de h se volverá a usar el método de minimizar el *AMSE*. En este apartado se recogerá algunos resultados de *Chen* [16] referidos a la ventana.

Considerándose la fórmula (2.4.5) como una función dependiente de h , derivándola con respecto a h e igualando a cero se obtiene

$$h = \left[288f(0) \int_0^\infty K^2(t)dt \right]^{1/9} \left[2f^{iv}(0) \int_0^\infty t^4 K(t)dt \right]^{-2/9} n^{-1/9}$$

Si $K(t)$ es la función núcleo que sigue una normal estándar ($N(0, 1)$), entonces

$$h = 1,18229 [f(0)]^{1/9} [f^{iv}(0)]^{-2/9} n^{-1/9} \quad (2.4.6)$$

Este valor de h puede ser sustituido en *AMSE* para lograr obtener el mínimo *AMSE* para $\hat{f}(0)$, cual indica que ratio de convergencia para el error cuadrático medio (*ECM*) de $\hat{f}(0)$ es de orden $n^{-8/9}$, mientras que el orden del estimador núcleo es de $n^{-4/5}$. Mas claramente, lo que indica esto es que cuando $n \rightarrow \infty$ el *ECM* de $\hat{f}(0)$ tiende más rápidamente a cero que el *ECM* del estimador núcleo tradicional.

2.4.3. Implementación práctica del estimador propuesto

El estimador propuesto (2.4.3) depende de la función desconocida $f''(0)$. Por lo tanto, la implementación práctica del estimador requiere la estimación de esta cantidad desconocida, para su sustitución en (2.4.3).

Para estimar $f''(0)$ se puede tomar la segunda derivada de $\hat{f}_k(x)$ en $x = 0$. Teniéndose en cuenta, que se puede usar $K(t)$ como la función núcleo que sigue una norma estándar, se obtiene

$$\hat{f}_k''(0) = \frac{2}{nh^5\sqrt{2\pi}} \sum_{i=1}^n (x_i^2 - h^2) e^{-x_i^2/2h^2}$$

De donde se obtiene $\hat{f}_k''(0)$ como el estimador de $f''(0)$.

Por otro lado, la formula del h óptimo prueba que este depende de una función desconocida que tiene que ser estimada. Una forma de resolver esto, es elegir h basándose en un modelo seminormal, véase *Buckland* [17]; *Chen*[16]; *Mack y Quang* [18]. Bajo esta suposición

$$f(x) = \frac{2}{\sigma} \phi(x/\sigma) I_{[0,\infty)}(x) \quad (2.4.7)$$

donde $\phi(x)$ representa la función de densidad de la normal estándar i $I_A(x)$ es la función indicadora del conjunto A .

Se tiene de (2.4.7) que el sesgo es de orden $O(h^4)$ debido a que en este caso $f'(0) = f'''(0) = 0$.

Una vez es obtenido el valor de $f(0)$ de (2.4.7) se substituye en la fórmula (2.4.6) obteniéndose h como

$$h = 0,9497\sigma n^{-1/9} \quad (2.4.8)$$

para implementar la fórmula (2.4.8), se estima σ mediante máxima verosimilitud obteniéndose el estimador $\hat{\sigma} = \sqrt{\sum_{i=1}^n x_i^2/n}$.

El valor de h también puede ser elegido usando el estimador núcleo $\hat{f}_k(0)$ como un estimador de $f(0)$. En este caso es necesario de un método numérico como el de *Newton – Raphson*.

2.4.4. Un vistazo a las condiciones que cumple $f(x)$

El estimador propuesto $\hat{f}(0)$ logra que el sesgo tenga orden $O(h^4)$ bajo la hipótesis que $f'(0) = f'''(0) = 0$.

Si $f'(0) \neq 0$ entonces $\hat{f}(0)$ logra solo orden $O(h)$ el cual es significativamente mayor que $O(h^4)$. Y hay casos que la condición no se cumple, por lo que no se puede considerar cierta siempre. Por ello se sugiere usar un nuevo estimador cuando se tengan dudas, este sería llamado $\hat{f}^*(0)$ que combina el modelo exponencial negativo el cual no satisface las condiciones esperables en el origen ($f'(0) \neq 0$ y $g'(0) \neq 0$), con la corrección del sesgo del estimador $\hat{f}(0)$.

Gates en 1968 [8] sugiere el modelo exponencial negativo para ajustar las distancias perpendiculares obtenidas en el muestreo por transecto lineal. Este modelo no satisface las condiciones esperables en el origen, y es dado por

$$f(x; \lambda) = \frac{1}{\lambda} e^{-x/\lambda}, \quad x \geq 0.$$

El parámetro $f(0; \lambda) = 1/\lambda$ puede ser estimado usando máxima verosimilitud, de donde se obtiene $f_M(0; \hat{\lambda}) = 1/\bar{X}$, o el estimador insesgado de mínima varianza $f(0; \hat{\lambda}) = (n - 1)/n\bar{X}$ cual es adoptado en el ajuste, donde \bar{X} la media de las distancias perpendiculares observadas X_1, X_2, \dots, X_n . Además, el estimador propuesto $\hat{f}^*(0)$ es de la forma

$$\hat{f}^*(0) = m f(0; \hat{\lambda}) + (1 - m) \hat{f}(0)$$

con $0 \leq m \leq 1$. El parámetro m es estimado de los datos y su estimación \hat{m} usada entonces en $\hat{f}^*(0)$ como el estimador propuesto para $f(0)$. Ahora lo que falta es la elección del valor de m , con la idea presente de que valores cercanos a ceros implica que las condiciones sobre la función de detectabilidad y de densidad se cumplen, y cercano a uno implica que no se cumplen.

Zhang en el año 2001 [19] propone un procedimiento mediante test sobre la condiciones esperables de un modelo basado en muestreo por transecto lineal. Considerando el test $H_0 : f'(0) = 0$ frente a $H_1 : f'(0) \neq 0$. Donde se rechaza la hipótesis nula H_0 para valores grande de $Z = \frac{\sqrt{\sum_{i=1}^n x_i^2}}{\sum_{i=1}^n x_i}$.

Zhang construyo una tabla con los valores críticos de la distribución de la muestra para Z con respecto a muestras de distinto tamaños mediante simulación de *Monte Carlo*. Y seleccionar el valor de m basándose en el test estadístico Z . El valor de Z el cual siempre está entre cero y uno, indica lo fuerte que H_0 puede ser aceptado. Pequeños valores de Z indica que se acepta que se cumplen las condiciones esperables y por lo tanto $\hat{f}(0)$ es cercano a $f(0)$. Y se puede usar el valor de Z para estimar m , el cual es el valor del test estadístico Z . Con estimador propuesto

$$\hat{f}^*(0) = zf(0; \hat{\lambda}) + (1 - z)\hat{f}(0)$$

2.5. Método semiparamétrico para muestreo por transecto lineal.

Se propone un método que requiere la existencia de un modelo paramétrico, y posteriormente dicho estimador mejorado mediante el uso de una función núcleo. Más concretamente, se parte del estimador de la función núcleo y se crea un nuevo estimador para $f(0)$ añadiendo la función de detectabilidad a la formula del nuevo estimador con la peculiaridad de que esta función de detectabilidad es supuesta paramétrica.

2.5.1. El estimador

Sea X_1, \dots, X_n una muestra aleatoria de distancias perpendiculares de tamaño n , con función de detectabilidad desconocida $g(x)$ y función de probabilidad $f(x)$.

El clásico estimador núcleo $\hat{f}_k(x)$ para $f(x)$ con $x \geq 0$ propuesto por *Chen* [16]

$$\hat{f}_k(x) = \frac{1}{nh} \sum_{i=1}^n \left[K\left(\frac{x - X_i}{h}\right) + K\left(\frac{x + X_i}{h}\right) \right] \quad (2.5.1)$$

donde K es la función núcleo y h es la ventana. Teniendo en cuenta la simetría de la función núcleo, el estimador núcleo $\hat{f}_k(x)$ para $f(x)$ es

$$\hat{f}_k(0) = \frac{2}{nh} \sum_{i=1}^n K\left(\frac{X_i}{h}\right). \quad (2.5.2)$$

Cuando no hay conocimiento previo acerca de $f(x)$, bajo ciertas suposiciones algunos autores indican que el estimador $\hat{f}_k(x)$ es un aceptable estimador para $f(x)$. Para más información, vease *Chen* [16], *Mack y Quang* [18].

Sin embargo, en muchos problemas en la practica el que realiza el experimento tiene conocimiento previo de la forma de $f(x)$ debido a experiencias de casos parecidos que previamente se han realizado. Por ello, para hacer valida la información de dichas experiencias, se supone que $g(x; \theta)$ es una función de detectabilidad paramétrica usada para modelar $g(x)$ (con $g(0; \theta) = 1$).

Se define $\alpha(x) = f(x)/g(x; \theta)$, y entonces el estimador núcleo de $\alpha(x)$ como

$$\hat{\alpha}_k(x) = \frac{1}{nh} \sum_{i=1}^n \left[K\left(\frac{x - X_i}{h}\right) + K\left(\frac{x + X_i}{h}\right) \right] [g(X_i; \theta)]^{-1}. \quad (2.5.3)$$

Como $g(0; \theta) = 1$ se tiene que $\alpha(0) = f(0)$, además el estimador propuesto para $f(0)$ es $\hat{\alpha}(x)$ en $x = 0$, de lo cual se obtiene

$$\hat{\alpha}_k(0) = \frac{2}{nh} \sum_{i=1}^n K\left(\frac{X_i}{h}\right) [g(X_i; \theta)]^{-1}. \quad (2.5.4)$$

Denotando $\tilde{f}(0)$ a $\hat{\alpha}_k(0)$ se introducirá un *Lema* el cual proporciona el cálculo del sesgo y la varianza.

Lema 2.5.1 *Supuesto que $g(x; \theta)$ y $f(x)$ son segundo derivables con derivadas continuas en $x = 0$ y $g'(0; \theta) = f'(0) = 0$.*

Si $h \rightarrow 0$ y $nh \rightarrow \infty$ con $n \rightarrow \infty$. Entonces

- *Sesgo($\tilde{f}(0)$) = $h^2 [f''(0) - g''(0; \theta)f(0)] \int_0^\infty t^2 K(t) dt + O(h^3)$.*
- *Var($\tilde{f}(0)$) = $\frac{4}{nh} f(0) \int_0^\infty K^2(t) dt + O(n^{-1})$.*

Prueba del Lema 2.5.1

Se define el *Sesgo*($\tilde{f}(0)$), donde $\tilde{f}(0)$ es el estimador de $f(0)$ como

$$\text{Sesgo}(\tilde{f}(0)) = E(\tilde{f}(0)) - f(0)$$

Entonces, se va a calcular $E(\tilde{f}(0))$ y usar el cambio de variable $\frac{x}{h} = t$ de donde se obtiene $x = th$.

$$\begin{aligned} E(\tilde{f}(0)) &= \frac{2}{h} E \left[K\left(\frac{X}{h}\right) [g(X; \theta)]^{-1} \right] \\ &= \frac{2}{h} \int_0^\infty K\left(\frac{X}{h}\right) [g(X; \theta)]^{-1} f(x) dx \\ &= 2 \int_0^\infty K(t) [g(ht; \theta)]^{-1} f(ht) dt \\ &= 2 \int_0^\infty K(t) \alpha(ht) dt \end{aligned}$$

Ahora, se considera la serie de expansión de Taylor de $\alpha(ht)$, y se usa que $\alpha'(0) = 0$ cuando $g'(0; \theta) = 0$ y $f'(0) = 0$.

Entonces, si $h \rightarrow 0$ cuando $n \rightarrow \infty$ se obtiene

$$\begin{aligned} E(\tilde{f}(0)) &= 2 \int_0^\infty K(t) [\alpha(0) + ht\alpha'(0) + h^2t^2\alpha''(0)/2 + \dots] dt \\ &= 2\alpha(0) \int_0^\infty K(t)dt + h^2\alpha''(0) \int_0^\infty t^2K(t)dt + O(h^3) \\ &= f(0) + h^2 [f''(0) - g''(0; \theta)f(0)] \int_0^\infty t^2K(t)dt + O(h^3) \end{aligned}$$

Y restando $f(0)$ a $E(\tilde{f}(0))$ se obtiene $Sesgo(\tilde{f}(0))$, y la primera parte del Lema.

Para la segunda parte del Lema, se realiza de forma análoga usando las propiedades de la varianza

$$\begin{aligned} Var(\tilde{f}(0)) &= \frac{4}{nh^2} E \left[K\left(\frac{X}{h}\right) [g(X; \theta)]^{-1} \right] \\ &= \frac{4}{nh^2} \int_0^\infty K^2\left(\frac{X}{h}\right) [g(X; \theta)]^{-2} f(x)dx + O(n^{-1}) \\ &= \frac{4}{nh} \int_0^\infty K^2(t) [g(ht; \theta)]^{-2} f(ht)dt + O(n^{-1}) \\ &= \frac{4}{nh} \int_0^\infty K^2(t) [g(ht; \theta)]^{-1} \alpha(ht)dt + O(n^{-1}) \end{aligned}$$

Se considera la serie de expansión de Taylor de $\alpha(ht) [g(ht; \theta)]^{-1}$ y las hipótesis que $h \rightarrow 0$ y cuando $n \rightarrow \infty$ se cumple que $nh \rightarrow \infty$. Entonces

$$\begin{aligned} Var(\tilde{f}(0)) &= \frac{4}{nh} \int_0^\infty K^2(t) [f(0) + h^2t^2(f''(0) - 2g''(0; \theta)f(0))/2 + \dots] dt + O(n^{-1}) \\ &= \frac{4f(0)}{nh} \int_0^\infty K^2(t)dt + \frac{2h}{n} (f''(0) - 2g''(0; \theta)f(0)) \int_0^\infty t^2K(t)dt + O(n^{-1}) \\ &= \frac{4f(0)}{nh} \int_0^\infty K^2(t)dt + O(n^{-1}) \end{aligned}$$

2.5.2. Elección de la ventana h

Para implementar el estimador propuesto es necesario elegir una ventana h adecuada. La ventana controla la suavidad del ajuste a la curva de densidad, por ello un h grande proporciona un estimador con pequeña varianza pero gran sesgo, por otro lado un h pequeño proporciona una gran varianza y un pequeño sesgo (es trivial viendo el comportamiento del sesgo y la varianza en el Lema con respecto a h).

Uno de los métodos más comunes de selección de la ventana es encontrar h que minimice el error cuadrático medio asintótico ($AMSE$), el cual relaciona la varianza con el sesgo.

El $AMSE$ de $f(0)$ es según *Omar y Eidous* (2006) [20]

$$AMSE = h^4 [f''(0) - g''(0; \theta)f(0)]^2 \left[\int_0^\infty t^2 K(t) dt \right]^2 + \frac{4}{nh} f(0) \int_0^\infty K^2(t) dt$$

Si se considera $AMSE$ como una función de h , se puede derivar e igualar a cero para obtener el h óptimo, denotado como h_{opt}

$$h_{opt} = \left[\frac{f(0) \int_0^\infty K^2(t) dt}{[f''(0) - g''(0; \theta)f(0)]^2 \left[\int_0^\infty t^2 K(t) dt \right]^2} \right]^{1/5} n^{-1/5} \quad (2.5.5)$$

Es importante indicar que el valor óptimo de la ventana h para el estimador $\tilde{f}(0)$, dado por (2.5.5) difiere del valor óptimo de la ventana para el estimador núcleo tradicional $\hat{f}_k(0)$. Varios estudios mediante simulación se han realizado, para el cual se han considerado la misma ventana para los estimadores $\tilde{f}(0)$ y $\hat{f}_k(0)$. Los cuales muestran una superioridad del estimador $\tilde{f}(0)$ sobre $\hat{f}_k(0)$ en la mayoría de casos.

La ecuación (2.5.5) puede considerarse algo decepcionante, al indicar que h_{opt} depende de una densidad desconocida que es necesaria determinar, el modelo paramétrico depende de un parámetro desconocido θ que debe ser estimado.

Usualmente, se suele considerar la función de detectabilidad paramétrica $g(x; \theta)$ como una seminormal, es decir, $g(x; \theta) = e^{-x^2/2\sigma^2}$, y por lo tanto $f(x; \theta) = 2e^{-x^2/2\sigma^2}/\theta\sqrt{2\pi}$. El parámetro σ es estimado mediante algún método robusto como por ejemplo mediante máxima verosimilitud, el cual proporciona el estimador $\hat{\theta} = \sqrt{\sum_{i=1}^n x_i^2/n}$. También, $f(0)$ y $f''(0)$. En este caso, para estimar $f(0)$ se ha elegido el estimador núcleo tradicional $\hat{f}_k(0)$ cual proporciona una ventana $1,06\hat{\theta}n^{-1/5}$ *Chen* [16]. Para la estimación $f''(0)$, *Mack* [31] recomienda el uso del estimador $\hat{f}''(0) = [F_n(3\delta) - 3F_n(2\delta) + 3F_n(\delta)]/\delta^3$ como un estimador para $f''(0)$. Donde $F_n(u)$ se entiende como la función de distribución empírica, con $\delta = 1,97\hat{\theta}n^{-1/9}$.

2.5.3. Aplicación del estimador propuesto

2.5.3.1. Cálculo del estimador propuesto

TABLE 6.—EXAMPLE OF LINE TRANSECT DATA FROM A SAMPLE OF WOODEN STAKES OF KNOWN DENSITY; $n = 68$, $w^* = 18.60$ m, $w = 20$ m, $L = 1,000$ m, AND $D = 0.00375$ STAKES PER METER. THE DATA SHOWN ARE THE PERPENDICULAR DISTANCES IN METERS

1.	2.02	35.	3.79
2.	0.45	36.	15.24
3.	10.40	37.	3.47
4.	3.61	38.	3.05
5.	0.92	39.	7.93
6.	1.00	40.	18.15
7.	3.40	41.	10.05
8.	2.90	42.	4.41
9.	8.16	43.	1.27
10.	6.47	44.	13.72
11.	5.66	45.	6.25
12.	2.95	46.	3.59
13.	3.96	47.	9.04
14.	0.09	48.	7.68
15.	11.82	49.	4.89
16.	14.23	50.	9.10
17.	2.44	51.	3.25
18.	1.61	52.	8.49
19.	31.31	53.	6.08
20.	6.50	54.	0.40
21.	8.27	55.	9.33
22.	4.85	56.	0.53
23.	1.47	57.	1.23
24.	18.60	58.	1.67
25.	0.41	59.	4.53
26.	0.40	60.	3.12
27.	0.20	61.	3.05
28.	11.59	62.	6.60
29.	3.17	63.	4.40
30.	7.10	64.	4.97
31.	10.71	65.	3.17
32.	3.86	66.	7.67
33.	6.05	67.	18.16
34.	6.42	68.	4.08

Fuente: *Burnham* (1980)

Omar Eidous [20] aplica el estimador propuesto al conjunto de datos de estacas de madera, recogidos por *Burnham* (1980) en la página 61.

Se computa los estimadores $\hat{f}_k(0)$ y $\tilde{f}(0)$, así como sus correspondientes sesgos (*bias*), desviaciones típicas (*SD*) y errores cuadráticos medios (*MSE*) de cada estimador, para la obtención de estos resultados se ha usado la técnica de bootstrap con 200 iteraciones.

Para el estimador $\tilde{f}(0)$, la ecuación (2.5.5) proporciona una ventana óptima $h_{opt} = 2,923$ para el caso $n = 68$ y $h_{opt} = 2,635$ para el caso $n = 67$. Resultando la estimación de $f(0)$ 0.10917 para el caso $n = 68$ y 0.11233 para el caso $n = 67$, con sus respectivas estimaciones de densidades $\hat{D} = 37,12$ y $\hat{D} = 37,63$.

Otras computaciones del problema para las dos muestras y para los dos estimadores $\hat{f}_k(0)$ y $\tilde{f}(0)$ son recogidos en la siguiente tabla.

Recordar que el verdadero valor de $f(0)$ es 0.110294 y el de D es 37.5 *estacas/hectáreas*.

Estimador	h	$f(0)$	Bias	SD	MSE	D
$n = 68$ stakes						
$\hat{f}_k(0)$	3.733	0.09889	-0.01090	0.01127	0.00025	33.6 stakes/ha
$\tilde{f}(0)$	2.923	0.10917	-0.00066	0.01344	0.00018	37.1 stakes/ha
$n = 67$ stakes						
$\hat{f}_k(0)$	3.342	0.10248	-0.00745	0.01186	0.00020	34.3 stakes/ha
$\tilde{f}(0)$	2.635	0.11233	0.00238	0.01417	0.00021	37.6 stakes/ha

Fuente: *Omar Eidous* (2006)

2.5.3.2. Rendimiento del estimador propuesto mediante simulación

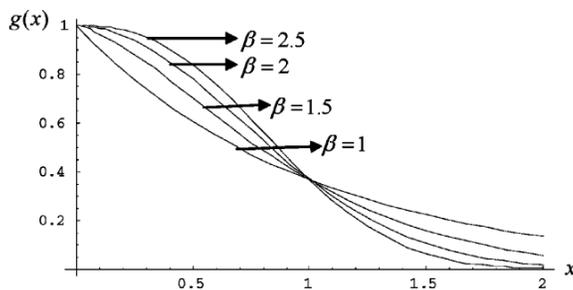
Para ver el rendimiento del estimador propuesto $\tilde{f}(0)$, se ha realizado un estudio mediante simulación. El estimador $\tilde{f}(0)$ es comparado con el estimador núcleo tradicional $\hat{f}_k(0)$ propuesto por *Chen* [16] con una seminormal y una ventana $h = 1,06\hat{\theta}n^{-1/5}$, mientras que para $\tilde{f}(0)$ se ha usado la ventana propuesta en (2.5.5).

Se van a considerar en esta simulación, tres familias de modelos, que son comúnmente usadas en el estudio de transectos lineales. Para más información véase *Barabesi* [21].

La familia exponencial que será denotada como (*EP*), *Pollock* [22]

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

Cuatro modelos fueron seleccionados para la familia *EP* con valores $\beta = 1,0, 1,5, 2,0$ y $2,5$, y sus correspondientes valores de truncación $w = 5,0, 3,0, 2,5$ y $2,0$. De donde se obtiene la siguiente función de detectabilidad para cada β

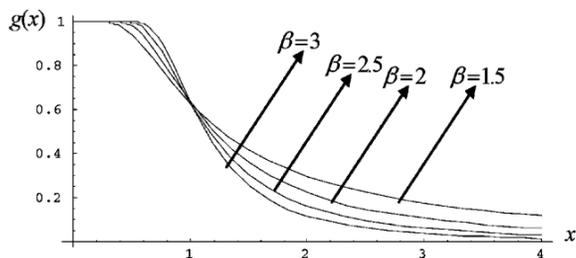


Fuente: *Omar Eidous* (2006)

La familia tasa de riesgo que será denotada como (*HR*), *Hayes y Buckland* [23]

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

Cuatro modelos fueron seleccionado para la familia *HR* con valores $\beta = 1,5, 2,0, 2,5$ y $3,0$, y sus correspondientes valores de truncación $w = 20, 12, 8$ y 6 . De donde se obtiene la siguiente función de detectabilidad para cada β

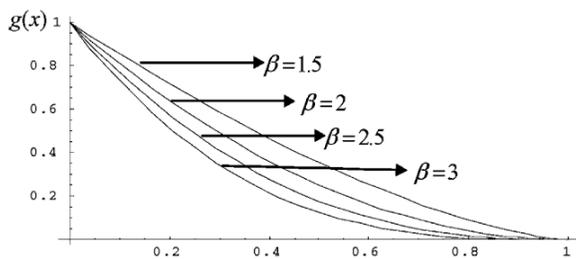


Fuente: Omar Eidous (2006)

La familia beta que será denotada como (BE) , Eberhardt [24]

$$f(x) = (1 + \beta)(1 - x)^\beta, \quad x \geq 0, \quad \beta \geq 0.$$

Cuatro modelos fueron seleccionados para la familia BE con valores $\beta = 1,5, 2,0, 2,5$ y $3,0$ y con el mismo valor de truncación en todos los casos de $w = 1$. De donde se obtiene la siguiente función de detectabilidad para cada β



Fuente: Omar Eidous (2006)

Hay que remarcar que el modelo EP para $\beta = 1$ y el modelo BE no satisfacen las condiciones que deberían cumplir, esta condición es que $g'(0) = 0$. Para cada modelo y para cada muestra de tamaño $n = 50,00$ y 200 , mil iteraciones han sido realizadas.

Para comparar ambos estimadores $\hat{f}_k(0)$ y $\tilde{f}(0)$ se emplean 3 valores para cada estimador.

El sesgo relativo (RB),

$$RB = \frac{E(\hat{f}(0)) - f(0)}{f(0)}$$

El error medio relativo (RME),

$$RME = \frac{\sqrt{MSE(\hat{f}(0))}}{f(0)}$$

La eficiencia (EFF) del estimador propuesto frente al tradicional núcleo,

$$EFF = \frac{MSE(\hat{f}_k(0))}{MSE(\tilde{f}(0))}$$

donde MSE es el error cuadrático medio.

Estos valores quedan recogidos en la tabla de la siguiente página.

Relative bias (RB), relative mean error (RME), and efficiencies (EFF) for the proposed estimator

N	β	W	$\hat{f}_k(0)$		$\hat{f}(0)$		EFF
			RB1	RME1	RB2	RME2	
Exponential power model							
50	1.0	5.0	-0.351	0.363	-0.257	0.280	1.297
100			-0.318	0.325	-0.228	0.244	1.333
200			-0.291	0.296	-0.209	0.220	1.347
50	1.5	3.0	-0.174	0.206	-0.073	0.152	1.360
100			-0.149	0.170	-0.063	0.118	1.436
200			-0.124	0.142	-0.054	0.099	1.428
50	2.0	2.5	-0.091	0.156	0.010	0.152	1.029
100			-0.076	0.120	0.002	0.110	1.092
200			-0.056	0.091	0.006	0.084	1.076
50	2.5	2.0	-0.064	0.146	0.035	0.160	0.914
100			-0.037	0.107	0.039	0.123	0.866
200			-0.031	0.081	0.028	0.090	0.894
Hazard rate model							
50	1.5	20.0	-0.421	0.435	-0.291	0.319	1.364
100			-0.375	0.384	-0.233	0.253	1.519
200			-0.326	0.331	-0.174	0.188	1.764
50	2.0	12.0	-0.265	0.292	-0.127	0.188	1.555
100			-0.223	0.240	-0.087	0.134	1.788
200			-0.178	0.190	-0.048	0.089	2.126
50	2.5	8.0	-0.130	0.179	-0.004	0.140	1.279
100			-0.094	0.130	0.014	0.106	1.226
200			-0.063	0.089	0.022	0.079	1.130
50	3.0	6.0	-0.064	0.139	0.047	0.152	0.916
100			-0.038	0.097	0.047	0.117	0.830
200			-0.017	0.070	0.044	0.095	0.737
Beta model							
50	1.5	1.0	-0.191	0.222	-0.097	0.166	1.339
100			-0.167	0.190	-0.091	0.139	1.362
200			-0.148	0.163	-0.085	0.118	1.387
50	2.0	1.0	-0.219	0.244	-0.126	0.180	1.358
100			-0.191	0.209	-0.113	0.152	1.380
200			-0.171	0.182	-0.106	0.131	1.394
50	2.5	1.0	-0.233	0.256	-0.139	0.188	1.364
100			-0.210	0.225	-0.131	0.163	1.380
200			-0.189	0.199	-0.122	0.144	1.384
50	3.0	1.0	-0.252	0.274	-0.159	0.204	1.341
100			-0.227	0.241	-0.148	0.176	1.366
200			-0.201	0.210	-0.131	0.152	1.382

Fuente: *Omar Eidous* (2006)

2.5.4. Resultados y conclusiones de la simulación

Se observa que el estimador núcleo tradicional $\hat{f}_k(0)$ tiene el mayor sesgo relativo en valor absoluto para el modelo *EP* con $\beta = 1,0, 1,5$, para el modelo *HR* con $\beta = 1,5, 2,0$ y para el modelo *BE* con diferentes valores de β .

Sin embargo, se observa que al aumentar el valor del parámetro β en los modelos *EP* y *HR* el sesgo relativo en valor absoluto disminuye. Además, se observan los casos particulares del modelo *EP* para $\beta = 2,5$ y el modelo *HR* para $\beta = 3$, para los cuales el estimador núcleo tradicional es mejor que el propuesto $\tilde{f}(0)$ debido a que la eficiencia es menor que 1, lo que indica que el error cuadrático medio del estimador núcleo tradicional es inferior al estimador propuesto. Pero estos dos casos pueden ser poco relevantes, a la hora de decir que estimador es mejor, al poder considerarse aceptable los valores de la eficiencia en ambos casos.

Los resultados de la simulación indican que $\hat{f}_k(0)$ es estimada siempre por abajo ($\hat{f}_k(0) < f(0)$), esto se debe a que siempre el sesgo es negativo. Esto indica la necesidad de usar el estimador propuesto $\tilde{f}(0)$, que proporciona al compararse los sesgos relativos en valor absoluto de los estimadores $\hat{f}_k(0)$ y $\tilde{f}(0)$, el resultado de que el sesgo relativo en valor absoluto del estimador $\hat{f}_k(0)$ es generalmente inferior al sesgo relativo en valor absoluto del estimador $\tilde{f}(0)$.

El estimador $\tilde{f}(0)$ generalmente proporciona razonables errores medio relativos, y su rendimiento es mejor que el del estimador $\hat{f}_k(0)$ claramente demostrado para el modelo *EP* con los valores $\beta = 1,0, 1,5$, el modelo *HR* con los valores $\beta = 1,5, 2,0, 2,5$ y el modelo *BE* con cualquier valor de β .

El error medio relativo en el caso del estimador $\hat{f}_k(0)$ según el tamaño de la muestra n comprende los siguientes rangos:

- Si $n = 50$, $[0,139, 0,435]$
- Si $n = 100$, $[0,097, 0,384]$
- Si $n = 200$, $[0,070, 0,331]$

El error medio relativo en el caso del estimador $\tilde{f}(0)$ según el tamaño de la muestra n comprende los siguientes rangos:

- Si $n = 50$, $[0,140, 0,319]$
- Si $n = 100$, $[0,106, 0,253]$
- Si $n = 200$, $[0,079, 0,220]$

Por lo que se observa que el estimador $\tilde{f}(0)$ proporciona una amplitud de intervalo inferior al estimador núcleo tradicional, esto implica que su rendimiento es mejor que el del estimador $\hat{f}_k(0)$.

En cuanto a la eficiencia, se observa que en 10 de los 12 casos, es mejor el estimador $\tilde{f}(0)$, y en esos 2 casos restantes la eficiencia es aceptable, por lo que puede considerar el estimador $\tilde{f}(0)$ como más eficiente que el estimador $\hat{f}_k(0)$.

Como conclusión, se ha visto que el estimador propuesto $\tilde{f}(0)$, proporciona un sesgo relativo en valor absoluto inferior, así como un inferior error medio relativo y que es más eficiente.

Hay que hacer mención, que las propiedades asintóticas del estimador $\tilde{f}(0)$ son bajo la condición de que es conocido el valor del parámetro θ de la función de detectabilidad, cosa que en la práctica suele ser desconocida y hay que estimarla. Sin embargo, el estudio mediante simulación aporta una superioridad del estimador propuesto en este caso, frente al estimador núcleo tradicional.

2.6. Mejora del método semiparamétrico mediante la cantidad k

Ya se ha visto que uno de los elementos más importante de las funciones núcleo es la selección de la ventana h , en esta sección se introducirá una nueva cantidad a determinar k que otorgará mejoras en la estimación. Más concretamente, se podría decir que es una generalización del caso del modelo semiparamétrico descrito anteriormente para $k = 1$.

2.6.1. El estimador

Sea X_1, \dots, X_n una muestra aleatoria de distancias perpendiculares de tamaño n , con función de detectabilidad desconocida $g(x)$ y función de probabilidad $f(x)$.

El clásico estimador núcleo $\hat{f}_k(x)$ para $f(x)$ con $x \geq 0$ es el propuesto por *Chen* [16]

$$\hat{f}_b(x) = \frac{1}{nb} \sum_{i=1}^n \left[K\left(\frac{x - X_i}{b}\right) + K\left(\frac{x + X_i}{b}\right) \right] \quad (2.6.1)$$

donde K es la función núcleo, asumiéndose que satisface condiciones de regularidad como que es acotada $\int_{-\infty}^{\infty} K(t)dt = 1$ y $\int_{-\infty}^{\infty} |t^p K(t)|dt < \infty$ para algun entero p tal que

$$\int_{-\infty}^{\infty} t^p K(t)dt = \begin{cases} 0, & \text{si } j = 1, 2, \dots, p-1 \\ d_p \neq 0 & \text{si } j = p \end{cases}$$

El parámetro b es la denominada ventana de la función núcleo. Teniéndose en cuenta la simetría de la función núcleo ($K(-x) = K(x)$), el estimador núcleo $\hat{f}_b(0)$ para $f(0)$ es

$$\hat{f}_b(0) = \frac{2}{nb} \sum_{i=1}^n K\left(\frac{X_i}{b}\right) \quad (2.6.2)$$

Bajo las hipótesis que $f'(0) = 0$ y que la segunda deriva de $f(x)$ existe y es continua en $x = 0$, la expresión de *AMSE* de $\hat{f}_b(0)$ es

$$AMSE = Sesgo^2 [\hat{f}_b(0)] + Var \hat{f}_b(0)$$

$$= b^2 f''(0) r_2^2 + \frac{4Q}{nb} f(0) \quad (2.6.3)$$

donde $Q = \int_0^\infty K^2(t) dt$ y $r_i = \int_0^\infty t^i K(t) dt$

Minimizando (2.6.3) con respecto a b se obtiene

$$b = [\delta(K)\beta(f)]^{1/5} n^{-1/5}$$

donde $\delta(K) = Q/r_2^2$ y $\beta(f) = f(0)/f''^2(0)$.

Si la función núcleo es elegida de tal manera que siga una normal estándar y $f(x)$ es supuesta que es una seminormal con parámetro de dispersión θ entonces *Chen* [16] propone

$$b = 1,06\hat{\theta}n^{-1/5} \quad (2.6.4)$$

donde $\hat{\theta}$ es el estimador de θ obtenido al aplicar máxima verosimilitud, cual es dado por $\hat{\theta} = \sqrt{\sum_{i=1}^n x_i^2/n}$. Supuesto que $g(x; \theta)$ es una seminormal con la forma $g(x; \theta) = \exp[-x^2/2\theta^2]$ se tiene que $g(0; \theta) = 1$ y $g'(0; \theta) = g''(0; \theta) = 0$.

Ahora se define $\alpha(x) = \frac{f(x)}{g^k(x; \theta)}$ donde $-\infty < k < \infty$, entonces el estimador núcleo de $\alpha(x)$ es propuesto como

$$\tilde{\alpha}_h(x) = \frac{1}{nh} \sum_{i=1}^n \left[K\left(\frac{x - X_i}{h}\right) + K\left(\frac{x + X_i}{h}\right) \right] [g(X_i; \theta)]^{-k}. \quad (2.6.5)$$

Como $g(0; \theta) = 1$ se tiene que $\alpha(0) = f(0)$, además el estimador propuesto para $f(0)$ es $\tilde{\alpha}(x)$ en $x = 0$, de lo cual se obtiene

$$\tilde{f}_h(0) \equiv \tilde{\alpha}_h(0) = \frac{2}{nh} \sum_{i=1}^n K\left(\frac{X_i}{h}\right) [g(X_i; \theta)]^{-k}. \quad (2.6.6)$$

La notación de h en $\tilde{f}_h(0)$ es usada para aclarar que el parámetro de $\tilde{f}_h(0)$ no tiene por qué ser el mismo que el parámetro del estimador núcleo tradicional $\hat{f}_b(0)$. *Eidous* [20] considero el caso especial de $k = 1$ y probo que el estimador propuesto es mejor que el estimador núcleo tradicional matemáticamente y numéricamente.

Lema 2.6.1 *Supuesto que $f(x)$ es cuatro veces derivable con derivadas continuas en $x = 0$ y $f'(0) = 0$.*

Si $h \rightarrow 0$ y $nh \rightarrow \infty$ con $n \rightarrow \infty$. Entonces

- $Sesgo(\tilde{f}_h(0)) = h^2\alpha''(0)r_2 + \frac{h^3}{3}\alpha'''(0)r_3 + \frac{h^4}{12}\alpha^{iv}(0)r_4 + O(h^5)$.
- $Var(\tilde{f}_h(0)) = \frac{4}{nh}Qf(0) + O(n^{-1})$.

Prueba del Lema 2.6.1

Se define el $Sesgo(\tilde{f}(0))$, donde $\tilde{f}(0)$ es el estimador de $f(0)$ como

$$Sesgo(\tilde{f}(0)) = E(\tilde{f}(0)) - f(0)$$

Entonces, se va a calcular $E(\tilde{f}(0))$ y usar el cambio de variable $\frac{x}{h} = t$ de donde se obtiene $x = th$.

$$\begin{aligned} E(\tilde{f}(0)) &= \frac{2}{h}E\left[K\left(\frac{X}{h}\right)[g(X;\theta)]^{-k}\right] \\ &= \frac{2}{h}\int_0^\infty K\left(\frac{X}{h}\right)[g(X;\theta)]^{-k}f(x)dx \\ &= 2\int_0^\infty K(t)[g(ht;\theta)]^{-k}f(ht)dt \\ &= 2\int_0^\infty K(t)\alpha(ht)dt \end{aligned}$$

Ahora, se considera la serie de expansión de Taylor de $\alpha(ht)$, y se usa que $\alpha'(0) = 0$ cuando $g'(0;\theta) = 0$ y $f'(0) = 0$.

Entonces, si $h \rightarrow 0$ cuando $n \rightarrow \infty$ se obtiene

$$\begin{aligned} E(\tilde{f}(0)) &= 2\int_0^\infty K(t)\left[\alpha(0) + ht\alpha'(0) + h^2t^2\alpha''(0)/2 + \dots\right]dt \\ &= 2\alpha(0)\int_0^\infty K(t)dt + h^2\alpha''(0)r_2 + h^3\alpha'''(0)r_3/3 + h^4\alpha^{iv}(0)r_4/12 + O(h^5) \\ &= f(0) + h^2\alpha''(0)r_2 + h^3\alpha'''(0)r_3/3 + h^4\alpha^{iv}(0)r_4/12 + O(h^5) \end{aligned}$$

Y restando $f(0)$ a $E(\tilde{f}(0))$ se obtiene $Sesgo(\tilde{f}(0))$, y la primera parte del Lema.

Para la segunda parte del Lema, se realiza de forma análoga usando las propiedades de la varianza

$$Var(\tilde{f}(0)) = \frac{4}{nh^2}E\left[K\left(\frac{X}{h}\right)[g(X;\theta)]^{-k}\right]$$

$$\begin{aligned}
 &= \frac{4}{nh^2} \int_0^\infty K^2\left(\frac{X}{h}\right) [g(X; \theta)]^{-2k} f(x) dx + O(n^{-1}) \\
 &= \frac{4}{nh} \int_0^\infty K^2(t) [g(ht; \theta)]^{-2k} f(ht) dt + O(n^{-1}) \\
 &= \frac{4}{nh} \int_0^\infty K^2(t) [g(ht; \theta)]^{-k} \alpha(ht) dt + O(n^{-1})
 \end{aligned}$$

Se considera la serie de expansión de Taylor de $C(\alpha h) = \alpha(ht) [g(ht; \theta)]^{-k}$ y las hipótesis que $h \rightarrow 0$ y cuando $n \rightarrow \infty$ se cumple que $nh \rightarrow \infty$. Entonces

$$\begin{aligned}
 Var(\tilde{f}(0)) &= \frac{4}{nh} \int_0^\infty K^2(t) \left[f(0) + h^2 t^2 C''(0)/2 + \dots \right] dt + O(n^{-1}) \\
 &= \frac{4Q}{nh} f(0) + \frac{2h}{n} C''(0) r_2 + O(n^{-1}) \\
 &= \frac{4Q}{nh} f(0) + O(n^{-1})
 \end{aligned}$$

Se obtiene que el estimador $\tilde{f}_h(0)$ tiene el mismo ratio de convergencia que el estimador núcleo tradicional y el sesgo de estos dos estimadores tienen el mismo orden, pero el estimador $\tilde{f}_h(0)$ es más proporcional a $\alpha''(0)$ que a $f''(0)$. Es decir, desde un punto de vista de la aproximación, si ambos usan el mismo valor para el parámetro de suavización (*el cual se ha denotado h*) entonces el error cuadrático medio (*ECM*) de $\tilde{f}_h(0)$ es menor que el del estimador núcleo $\hat{f}_b(0)$ si el valor absoluto de $\alpha''(0)$ es menor que el valor absoluto de $f''(0)$. En este escenario, se enfocara en el caso $k = f''(0)/g''(0; \theta)f(0)$. Esto produce una diferente expresión para el sesgo asintótico de $\tilde{f}_h(0)$, ya que este valor de k implica que $\alpha'' = 0$ por el que el primer término del sesgo desaparece. Además, el sesgo $\tilde{f}_h(0)$ es de orden $O(h^3)$ en vez de orden $O(h^2)$. Sin embargo, si $f'(0) = f'''(0) = 0$ entonces se logra un orden $O(h^4)$.

Es reseñable mencionar que si se usa el valor de k anteriormente mencionado y $g(x; \theta)$ es supuesto que tiene una distribución seminormal con parámetro de dispersión θ entonces la fórmula del estimador propuesto (2.6.3) no depende de θ . Es decir, el valor de *theta* no afecta al valor del estimador $\tilde{f}_h(0)$ si k toma el valor de $f''(0)/g''(0; \theta)f(0)$.

2.6.2. La selección de la ventana h y k

Para implementar el estimador propuesto, en la práctica es necesario elegir los valores de la ventana h y la cantidad k . La ventana puede ser elegida minimizando el *AMSE* que relaciona el sesgo y la varianza del estimador.

El *AMSE* de $\tilde{f}_h(0)$ tomando k como descrito en el apartado anterior:

$$AMSE = \frac{h^6}{9} [f'''(0)]^2 r_2^2 + \frac{4Q}{nh} f(0)$$

Considerándose *AMSE* como una función de h , derivando con respecto a h e igualando a cero, se puede obtener la elección del h óptimo,

$$h = \left[\frac{6f(0)Q}{[f'''(0)]^2 r_3^2} \right]^{1/7} n^{-1/7} \quad (2.6.7)$$

Si $f'''(0) = 0$ entonces la fórmula (2.6.2) es infinita. En tal caso, la expresión par el sesgo asintótico varía y con ello la expresión de $AMSE$, con lo que el valor óptimo de h cuando $f'''(0)$ es

$$h = \left[\frac{72f(0)Q}{[\alpha^{iv}(0)]^2 r_4^2} \right]^{1/9} n^{-1/9} \quad (2.6.8)$$

En la práctica usar (2.6.2) implica estimar las cantidades $f(0)$ y $f''(0)$, y usar (2.6.8) implica estimar las cantidades $f(0)$, $f^{iv}(0)$ y θ .

Las dos fórmulas de h , (2.6.2) y (2.6.8) son algo decepcionantes al probarse que h depende de un desconocido $f(0)$ el cual debe ser estimado. Para estimar $f(0)$, $f''(0)$, y $f^{iv}(0)$ se puede usar el estimador núcleo clásico $\hat{f}_b(x)$, en que distintos valores de b son necesarios a la hora de estimar las derivadas de $f(x)$ en $x = 0$. *Hardle, Marron y Wand* [25] condujeron una forma de estimar las derivadas de la función de densidad usando función núcleo y *Chen* [16] dio una fórmula para elegir la ventana h en el caso del muestreo por transecto lineal.

En este escenario, se adopta el método de selección del $m - vecino más cercano$, cual no usa las cantidades anteriormente dichas y tampoco usa las hipótesis de la forma de $f(x)$. *Barabesi* [26] uso el método de selección del $m - vecino más cercano$ en el muestreo por transecto lineal para mejorar la eficacia del estimador paramétrico que previamente habia desarrollado *Barabesi* en el año 2000 [21].

Loftsgaarden y Quesenberry [29] introdujeron el método de selección del $m - vecino más cercano$ dado por $h = x_{(m)}$, donde $x_{(m)}$ representa el $m - esimo$ estadístico ordenado en la muestra observada. Una común elección para la selección del m es $m = \lceil [n^\epsilon] \rceil$, donde $0 < \epsilon < 1$ y $\lceil [s] \rceil$ denota al mayor entero menor o igual que s .

En estas condiciones, se suele usar $\epsilon = 4/5$, *Barabesi* (2001).

Como se vio, el sesgo de $\tilde{f}(0)$ es de orden $O(h^3)$ en vez de $O(h^2)$ cuando k es elegido como $f''(0)/g''(0; \theta)f(0)$.

Sea $g(x; \theta)$ un modelo seminormal, entonces $g''(0, \theta) = -1/\theta^2$ y $g''(0, \hat{\theta}) = -1/\hat{\theta}^2$, donde $\hat{\theta} = \sqrt{\sum_{i=1}^n x_i^2/n}$ es el estimador de θ obtenido por máxima verosimilitud. La consistencia del estimador de la función núcleo clásica es usada para estimar $f(0)$. Para estimar $f''(0)$, *Mack* [31] recomienda usar $\hat{f}''(0) = [F_n(3\delta) - 3F_n(2\delta) + 3F_n(\delta)]/\delta^3$, donde $F_n(u)$ es la función de distribución empírica y $\delta = 1,97\hat{\theta}n^{-1/9}$. Y en consecuencia el valor de k es tomado como $\hat{k} = \hat{f}''(0)/g''(0; \hat{\theta})\hat{f}_b(0)$.

2.6.3. Estudio mediante simulación

Se recoge en este apartado el estudio de Montecarlo del autor *Omar Eidous*.

Se han usado en el método de Montecarlo tres estimadores:

- $\hat{f}_b(0)$ el estimador núcleo tradicional propuesto por *Chen* [16] con una seminormal y una ventana $b = 1,06\hat{\theta}n^{-1/5}$.
- $\tilde{f}_1(0) = \tilde{f}_h(0)$ con la ventana h como la computada mediante el $m - vecino$ más cercano.
- $\tilde{f}_2(0) = \tilde{f}_h(0)$ con la ventana h como $h = 1,06\hat{\theta}n^{-1/5}$.

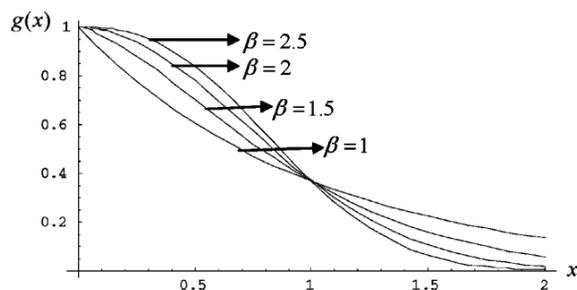
donde k es tomado como $\hat{h} = \hat{f}''(0)/g''(0; \hat{\theta})\hat{f}_b(0)$.

Entonces se ha usado tres familias de modelos, cuales son comúnmente usadas en estudios de transecto lineales, *Barabesi* (2000).

La familia exponencial que será denotada como (*EP*), *Pollock* [22]

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

Cuatro modelos fueron seleccionados para la familia *EP* con valores $\beta = 1,0, 1,5, 2,0$ y $2,5$, y sus correspondientes valores de truncación $w = 5,0, 3,0, 2,5$ y $2,0$. De donde se obtiene la siguiente función de detectabilidad para cada β

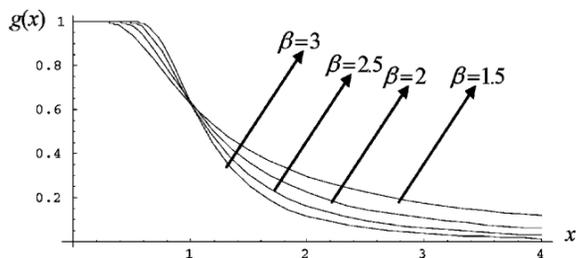


Fuente: Omar Eidous (2006)

La familia tasa de riesgo que será denotada como (*HR*), *Hayes y Buckland* [23]

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

Cuatro modelos fueron seleccionado para la familia *HR* con valores $\beta = 1,5, 2,0, 2,5$ y $3,0$, y sus correspondientes valores de truncación $w = 20, 12, 8$ y 6 . De donde se obtiene la siguiente función de detectabilidad para cada β

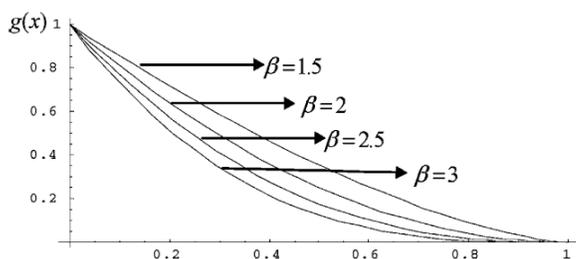


Fuente: Omar Eidous (2006)

La familia beta que será denotada como (BE) , Eberhardt [24]

$$f(x) = (1 + \beta)(1 - x)^\beta, \quad x \geq 0, \quad \beta \geq 0.$$

Cuatro modelos fueron seleccionados para la familia BE con valores $\beta = 1,5, 2,0, 2,5$ y $3,0$ y con el mismo valor de truncación en todos los casos de $w = 1$. De donde se obtiene la siguiente función de detectabilidad para cada β



Fuente: Omar Eidous (2006)

El modelo EP para $\beta = 1$ y el modelo BE no satisfacen la condición que debería cumplir la función de detectabilidad, es decir, $g'(0)$ no es cero.

Para cada modelo, se han generado muestras de tamaño $n = 50, 100, 200, 1000$ y se han realizado iteraciones con ellas.

Definiendo el error relativo (RB) como

$$RB = \frac{E(\hat{f}(0)) - f(0)}{f(0)}$$

Y el error medio relativo (RME) ,

$$RME = \frac{\sqrt{MSE(\hat{f}(0))}}{f(0)}$$

2.6.4. Resultados y conclusiones

En el estudio, para ver la influencia de k en el estimador propuesto $\tilde{f}_h(0)$, se ha consideraran distintos valores de k . Se ha estimado k por $\hat{h} = \hat{f}''(0)/g''(0; \hat{\theta})\hat{f}_b(0)$ y entonces el RB y RME de $\tilde{f}_1(0)$ para los distintos valores $k/4, \sqrt{3}, k/2, 3k/4, 1,2k, y 3k$.

Del estudio de simulación se consiguieron las siguientes conclusiones.

Se observa que el estimador núcleo tradicional $\hat{f}_b(0)$ tiene un mayor sesgo relativo en valor absoluto en el modelo EP para $\beta = 1,0, 1,5$, para el modelo HR para $\beta = 1,5, 2,0$ y para el modelo BE para los distintos valores de β .

Sin embargo, cuando el valor del parámetro β aumenta en los modelos EP y HR , estos modelos aumentan el rendimiento al disminuir el sesgo relativo en valor absoluto.

El estimador $\hat{f}_b(0)$ es muy similar al estimador $\tilde{f}_h(0)$ para el modelo EP con $\beta = 2,5$ y el modelo HR con $\beta = 3$. Pero la eficiencia en ambos estimadores es aceptable, recordándose la eficiencia como el cociente entre los RME .

Debido a que los sesgos relativos de $\hat{f}_b(0)$ son negativos, se observa que la estimaciones se realizan por debajo del valor real $f(0)$.

Esto último, motiva el uso del estimador $\tilde{f}_h(0)$, cual proporciona sesgos relativos más pequeños generalmente. Ya que si se compara los sesgos relativos en valor absoluto de los estimadores $\hat{f}_b(0)$ y $\tilde{f}_h(0)$, se observan que en este ejemplo numérico son más pequeños en el caso del estimador $\tilde{f}_h(0)$.

El estimador $\tilde{f}_h(0)$ generalmente proporciona razonables errores medio relativos, y su rendimiento es mejor que el del estimador $\hat{f}_b(0)$ si se observa los valores RME .

La mejora de rendimiento del estimador $\tilde{f}_h(0)$ sobre el estimador $\hat{f}_b(0)$ claramente está demostrado para el modelo EP con los valores $\beta = 1,0, 1,5$, el modelo HR con los valores $\beta = 1,5, 2,0, 2,5$ y el modelo BE con cualquier valor de β .

El error medio relativo RME en el caso del estimador $\hat{f}_b(0)$ según el tamaño de la muestra n comprende los siguientes rangos:

- Si $n = 50$, $[0,137, 0,436]$
- Si $n = 100$, $[0,095, 0,378]$
- Si $n = 200$, $[0,068, 0,328]$

El error medio relativo RME en el caso del estimador $\tilde{f}_1(0)$ según el tamaño de la muestra n comprende los siguientes rangos:

- Si $n = 50$, $[0,144, 0,320]$
- Si $n = 100$, $[0,107, 0,251]$
- Si $n = 200$, $[0,071, 0,240]$

El error medio relativo RME en el caso del estimador $\tilde{f}_2(0)$ según el tamaño de la muestra n comprende los siguientes rangos:

- Si $n = 50$, $[0,134, 0,403]$
- Si $n = 100$, $[0,103, 0,361]$
- Si $n = 200$, $[0,070, 0,308]$

Por lo que se observa que el estimador $\tilde{f}_1(0)$ proporciona una amplitud de intervalo inferior, y por lo tanto una mejor estabilidad, lo que implica que su rendimiento es mejor que el de los estimadores $\hat{f}_b(0)$ y $\tilde{f}_2(0)$.

El rendimiento del estimador $\tilde{f}_h(0)$ es mejor que el del $\hat{f}_b(0)$ para 11 de los 12 modelos. Para el modelo HR con $\beta = 3,0$ el estimador $\hat{f}_b(0)$ es relativamente mejor que el estimador $\tilde{f}_h(0)$, aun así la eficiencia del estimador $\tilde{f}_h(0)$ sigue siendo aceptable con respecto al estimador $\hat{f}_b(0)$.

Se observa que mientras el método elegido en la selección de k propuesto ($\hat{k} = \hat{f}''(0)/g''(0; \hat{\theta})\hat{f}_b(0)$) proporciona un rendimiento aceptable para $\tilde{f}_h(0)$, está claro que algunas mejoras pueden producirse en algunos casos.

Por ejemplo, para el modelo EP con $\beta = 1$ y para el modelo BE con los distintos valores de β , se observa una disminución del sesgo relativo y del error medio relativo de $\tilde{f}_1(0)$ cuando aumenta el valor de k hasta llegar a $3k$, por lo tanto una mejora de rendimiento.

Por otro lado, el rendimiento de $\tilde{f}_1(0)$ es relativamente mejor para k (cuando su coeficiente es 1) para los modelos EP con $\beta = 2, 2,5$ y el HR con $\beta = 2,5, 3$.

Los modelos EP con $\beta = 1$ y BE para los diferentes valores de β no satisfacen la condición $f'(0) = 0$, de la cual las propiedades del estimador considerado son obtenidas. Es decir, esta condición no se cumple entonces un método más robustos para la elección del valor k puede mejorar el rendimiento del estimador propuesto.

Como conclusión el estimador propuesto proporciona RB y RME más pequeños. Esto se deduce comparando el estimador núcleo tradicional $\hat{f}_b(0)$ con el estimador propuesto $\tilde{f}_h(0)$, este último en la mayoría de casos proporcionan un mejor rendimiento. Comparando el rendimiento de $\tilde{f}_1(0)$ con el de $\tilde{f}_2(0)$, es más eficiente en general el estimador $\tilde{f}_1(0)$, especialmente en el modelo HR con $\beta = 1,5, 2,0$ y el modelo EP con $\beta = 1,0$, donde se observan menores RB y RME .

2.6.5. Ejemplo

Usandose de nuevo los datos recogidos en *Burnham* (1980) sobre estacas de madera.

68 observaciones a lo largo de un transecto lineal de longitud 1000m.

Omar Eidous [27] computa los estimadores $\hat{f}_b(0)$, $\tilde{f}_1(0)$ y $\tilde{f}_2(0)$.

Para el estimador $\tilde{f}_1(0)$ se ha usado el método del *m - vecino* más cercano obteniendose un valor $h = 3,79$, mientras que para los estimadores $\hat{f}_b(0)$ y $\tilde{f}_2(0)$ $h = 3,73$, el proporcionado por *Chen* (1996) para el estimador núcleo tradicional.

En la siguiente tabla se recogen para cada estimador el sesgo, la desviación típica, error cuadrático medio y la estimación de la densidad.

The point estimates of $f(0)$ for stakes data and their biases, variances, and MSEs.
The true value of $f(0)$ is 0.110294 and of D is 37.5 stakes/ha

Estimator	$n = 68$ stakes					
	h	$f(0)$	<i>Bias</i>	<i>Variance</i>	<i>MSE</i>	D
$\hat{f}_b(0)$	3.733	0.09889	-0.01097	0.0001226	0.000243	33.62 stakes/ha
$\tilde{f}_1(0)$	3.733	0.10597	-0.00404	0.0001425	0.000159	36.03 stakes/ha
$\tilde{f}_2(0)$	3.790	0.10580	-0.00353	0.0001498	0.000162	35.97 stakes/ha

Fuente: Omar Eidous (2009)

2.7. Selección de estimadores paramétricos y no paramétricos

Como ya se vio estimar la densidad equivale a estimar $f(0)$, y esto puede ser abordado desde un punto de vista paramétrico si se tiene información sobre la forma de la función de densidad $f(x)$. Sin embargo, la función de densidad $f(x)$ puede depender de parámetros desconocidos θ , cual deben ser estimador usando una muestra aleatoria de las distancias perpendiculares X_1, x_2, \dots, X_n . Diferente al método paramétrico, los métodos no paramétricos no requieren información acerca de la forma de la función de densidad $f(x)$. Por las últimas dos décadas, el método no paramétrico basado en la función núcleo ha sido el más utilizado en el muestreo por transecto lineal. este método da una muy aceptable estimación de $f(0)$ y por lo tanto de la densidad D cuando no hay a priori información de la forma del modelo paramétrico.

Eidous [27] sugiere un estimador semi-paramétrico para $f(0)$ y en consecuencia para D . En este apartado se combina la estimación no paramétrica basada en función núcleo con un modelo paramétrico. Para ello se sugiere cuatro modelos, las dos primeras no satisfacen lo esperable que $g'(0; \text{parámetro}) = 0$, y las otras dos si satisfacen dicha propiedad. Y se seleccionara el mejor modelo basándose en el que mejor se ajuste a los datos mediante el *Criterio de información de Akaike (AIC)*.

2.7.1. Modelos paramétricos y AIC

Sea X_1, X_2, \dots, X_n una muestra aleatoria de n distancias perpendiculares con una función de densidad $f(x)$.

Gates [8] sugirió el modelo exponencial con función de densidad

$$f_1(x; \alpha) = \frac{1}{\alpha} e^{-x/\alpha}; \quad x \geq 0. \quad (2.7.1)$$

y su correspondiente función de detectabilidad

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

La función de detectabilidad debe cumplir que en el origen tome el valor 1 ($g(0) = 1$), lo que significa que a distancias cerca del transecto lineal debe tomar valores cercanos a 1. Esto se puede expresar como que $g'(0) = 0$ o equivalentemente que $f'(0) = 0$. Sin embargo, $g_1'(0; \alpha) = 1/\alpha \neq 0$ por lo que no cumple lo esperable que debe de cumplir una función de detectabilidad.

Usando el método de máxima verosimilitud se obtiene un estimador para $f(0)$, $f_1(0; \hat{\alpha}) = 1/\bar{X}$, donde \bar{X} es la media muestral.

Burnham [7] sugirió el modelo logístico invertido con función de densidad

$$f_2(x; \theta) = \frac{2\theta}{3\log(3)} \frac{3e^{-\theta x}}{1 + 2e^{-\theta x}}, \quad x \geq 0. \quad (2.7.2)$$

con su correspondiente función de detectabilidad

$$g_2(x; \theta) = \frac{3e^{-\theta x}}{1 + 2e^{-\theta x}}$$

Es fácil ver que $f_2(x; \theta)$ no cumple que $f'(0, \theta) = 0$. El parámetro que se necesita estimar es $f_2(0; \theta) = \frac{2\theta}{3\log(3)}$.

El estimador por el método de los momentos de $f_2(0; \theta)$ existe en forma cerrada (esto es, que se puede expresar en un número finito de operaciones), dado por $f_2(x; \hat{\theta}) = \frac{2\theta}{3\log(3)} \frac{1,3078}{\bar{X}} = \frac{0,7936}{\bar{X}}$, mientras que el estimador mediante máxima verosimilitud no existe en forma cerrada. Es necesario de un método numérico como *Newton – Raphson* para su obtención.

En contraste con estos dos modelos anteriormente descrito, se tiene estos dos modelos que si cumplen las condiciones esperadas.

Burnham [7] sugirió el modelo seminormal con función de densidad

$$f_3(x; \sigma^2) = \frac{2}{\sigma\sqrt{2\pi}} e^{-x^2/2\sigma^2}; \quad x \geq 0, \sigma^2 > 0. \quad (2.7.3)$$

y su correspondiente función de detectabilidad

$$g_3(x; \sigma^2) = e^{-x^2/2\sigma^2}; \quad x \geq 0, \sigma^2 > 0.$$

Y se satisface la propiedad ($g'(0) = 0$), ya que $g'(x; \sigma^2) = -\frac{x}{\sigma^2}e^{-x^2/2\sigma^2}$, de donde obtenemos $g'(0; \sigma^2) = 0$.

Por otro lado, usando el método de máxima verosimilitud se obtiene un estimador para $f_3(0; \sigma^2) = \frac{2}{\sigma\sqrt{2\pi}}$, que es $f_3(0; \hat{\sigma}^2) = [2n/\pi \sum_{i=1}^n X_i^2]^{1/2}$.

Ababneh y Eidous [12] sugirieron el modelo exponencial ponderado con función de densidad

$$f_4(x; \beta) = \frac{2\beta}{3}e^{-\beta x}(2 - e^{-\beta x}); \quad x \geq 0, \beta > 0. \quad (2.7.4)$$

y su correspondiente función de detectabilidad

$$g_4(x; \beta) = e^{-\beta x}(2 - e^{-\beta x}); \quad x \geq 0, \beta > 0.$$

La función de detectabilidad satisface que $g'(0) = 0$.

Ahora, el parámetro a estimar es $f_4(0; \beta)$. El método de los momentos proporciona el estimador $f_4(0; \hat{\beta}) = 7/9\bar{X}$. Y la máxima verosimilitud requiere de un método numérico para obtenerlo.

Si se asume que la muestra de las distancias perpendiculares sigue alguno de estos modelos, y entonces usando el estimador de $f(0)$ para estimar la densidad D es lo que se conoce como método paramétrico. Los métodos paramétricos proporcionan un buen ajuste cuando los datos realmente pertenecen al modelo supuesto. Si el modelo supuesto no representa a los datos, entonces su desempeño no es aceptable, *Buckland* [1].

En el caso de que la distribución de los datos no esté clara, una de las formas de seleccionar el mejor modelo es mediante el criterio de optimización de *AIC*. *Buckland* [1] define el *AIC* como

$$AIC = -2\ln(L) + p$$

donde $\log_e(L)$ es la función logarítmica de verosimilitud evaluada es la estimación de máxima verosimilitud del modelo paramétrico y p es el número de parámetros del modelo. Para un conjunto de datos dado y para seleccionar el modelo que más se ajuste a dichos datos, *AIC* es computado para cada modelo y el modelo con menor *AIC* se considera el que mejor se ajusta a los datos. El *AIC* para los diferentes modelos dados en este apartado son

- $AIC(1) = -2\ln(1/\bar{X}) + 2$ para el modelo exponencial.
- $AIC(2) = -2 \left[n \cdot \ln(2\hat{\theta}) - n \cdot \ln(\ln(3)) - \hat{\theta} \sum_{i=1}^n x_i - \sum_{i=1}^n \ln(1 + 2\exp(-\hat{\theta}x_i)) \right] + 2$ para el modelo logístico inverso, donde $\hat{\theta}$ es el estimador de θ mediante máxima verosimilitud.

- $AIC(3) = -2n \cdot \ln(2) + n \cdot \ln(2\pi\hat{\sigma}^2) + n + 2$ para el modelo seminormal, donde $\hat{\sigma}^2 = \frac{\sum_{i=1}^n x_i^2}{n}$ es el estimador de σ^2 por máxima verosimilitud.
- $AIC(4) = -2 \left[n \cdot \ln(2\hat{\beta}/3) - \hat{\beta} \sum_{i=1}^n x_i + \sum_{i=1}^n \ln(2 - \exp(-\hat{\beta}x_i)) \right] + 2$ para el modelo exponencial ponderado, donde $\hat{\beta}$ es el estimador de β por máxima verosimilitud.

2.7.2. Estimadores semiparamétricos

Sea X_1, X_2, \dots, X_n una muestra aleatoria de n distancias perpendiculares con función de densidad $f(x)$ desconocida. Es estimador no paramétrico basado en la función núcleo clásico de $f(0)$ propuesto por *Chen* [16] es

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

donde h es el parámetro de suavización y K es la función núcleo que es supuesta simétrica ($K(-x) = K(x)$). El valor óptimo de h que minimiza el error cuadrático medio asintótico de $\hat{f}_k(0)$, es dado por la siguiente formula

$$h = \left[\frac{f(0) \int_0^\infty K(t)^2 dt}{(f''(0) \int_0^\infty t^2 K(t) dt)^2} \right]^{1/5} \cdot n^{-1/5}$$

esta fórmula depende de $f(0)$ y $f''(0)$, cuales son necesarias ser estimadas.

Uno puede asumir de manera razonada la densidad $f(x)$ y proceder al cálculo de h . *Gerard* y *Schucany* [28] recomendaron el modelo seminormal como una razonable partida para $f(x)$ y da unos muy aceptable resultados para $\hat{f}_k(0)$.

Si $f(x) = f_3(x; \sigma^2)$ entonces $f(0) = \frac{2}{\sigma\sqrt{2\pi}}$ y $f''(0) = \frac{-2}{\sigma^3\sqrt{2\pi}}$, y si es asumido que la función núcleo sigue una normal estándar ($K(t) = N(0, 1)$) entonces $h = 0,933\hat{\sigma}n^{-1/5}$, donde $\hat{\sigma} = \sqrt{\frac{\sum_{i=1}^n x_i^2}{n}}$ es el estimador de σ obtenido mediante máxima verosimilitud.

De la misma forma usando distintos modelos paramétricos podemos obtener diferentes fórmulas para h , de donde se obtiene lo siguiente

- Si $f(x) = f_1(x; \alpha)$ entonces $h = 0,8918\hat{\alpha}n^{-1/5}$, donde $\hat{\alpha} = \bar{X}$ es el estimador de α obtenido mediante máxima verosimilitud.
- Si $f(x) = f_2(x; \theta)$ entonces $h = 2,3734(1/\hat{\theta})n^{-1/5}$, donde $\hat{\theta} = \frac{1,3078}{\bar{X}}$ es el estimador de θ obtenido mediante el método de los momentos.
- Si $f(x) = f_4(x; \beta)$ entonces $h = 0,7330(1/\hat{\beta})n^{-1/5}$, donde $\hat{\beta} = \frac{7}{6\bar{X}}$ es el estimador de β obtenido mediante el método de los momentos.

En otra dirección, si no se usa la fórmula de h , el método de *k vecinos más cercanos* computa h usando el muestreo por transecto lineal, dicho método fue recomendado por *Barabesi* [1].

Loftsgaarden y *Quesenberry* [29] introdujeron el método de selección del k – vecino más cercano dado por $h = x_{(k)}$, donde $x_{(k)}$ representa el k – esimo estadístico ordenado en la muestra observada. Una común elección para la selección del k es $k = \lceil n^\epsilon \rceil$, donde $0 < \epsilon < 1$ y $\lceil s \rceil$ denota al mayor entero menor o igual que s .

En estas condiciones, se suele usar $\epsilon = 4/5$ *Barabesi* [1].

Otros estimadores no paramétricos fueron propuestos y desarrollados para estimar $f(0)$. Muchos de estos estimadores consideran el estimador función núcleo como la semilla a partir de la cual se elaboran.

Supuesto que la función de detectabilidad sigue una seminormal con parámetro σ^2 , es decir, $g(x) = e^{-x^2/2\sigma^2}$, entonces *Eidous* [27] sugiere un estimador semiparamétrico para $f(0)$, cual combina el estimador clásico basado en la función núcleo y la función de detectabilidad que sigue una seminormal. De donde se obtiene el siguiente estimador

$$\hat{f}_E(0) = \frac{2}{nh} \sum_{i=1}^n K\left(\frac{x_i}{h}\right) [g(x_i; \sigma^2)]^{-1}$$

Eidous [27] deduce propiedades asintóticas de $\hat{f}_E(0)$ y prueba que este estimador es superior al estimador $\hat{f}_k(0)$ mediante técnicas de simulación. El recomienda usar el método de selección del k – vecino más cercano para computar el parámetro h en $\hat{f}_E(0)$.

La fórmula de $\hat{f}_E(0)$ depende de dos cantidades h y g las cuales están bajo el control del investigador. A pesar de poder elegir diferentes elecciones, *Eidous* [27] eligió únicamente la función de detectabilidad g como una seminormal. Sin embargo, es sugerido para una mejor estimación seleccionar, el modelo que se asemeje más a los datos mediante algún criterio como el de *AIC*, y usar este como la función de detectabilidad g .

Capítulo 3

Aplicación en R

3.1. Método paramétrico

Se empezará introduciendo el paquete de *R Distance* que implementa los métodos de muestreo por distancias para estimar la abundancia de la población. Así mismo los usuarios de *R* podrán obtener una estimación de la abundancia (así como la densidad). Además de la librería *Distance* existen otras que permiten obtener resultados similares, entre las que se puede citar a *Rdistance* de Trent McDonald, Jason Carlisle y Aidan McDonald (2019).

Librería “Distance”

Descripción

Es un paquete de *R* que da las herramientas necesarias para implementar los métodos de muestreo por distancias y obtener la abundancia (o densidad).

Las funciones principales de la librería son `ds`, `ds.gof`, `gof_ds`, `AIC.dsmodel`, `plot.dsmodel`

ds

Descripción

Este comando ajusta la función de detectabilidad para datos procedentes de transecto puntuales o lineales y entonces calcula una estimación de la abundancia y densidad.

Uso

```
ds(data,transect = c("line", "point"), key = c("hn", "hr", "unif"), adjustment = c("cos", "herm", "poly"))
```

Argumentos

data	<i>Un data.frame de la muestra con una columna llamada distance con las distancias de la muestra.</i>
transect	<i>Indica el tipo de transecto lineal (por defecto) y puntual.</i>
key	<i>La función clave que se va a utilizar en la estimación : "hn" para semi – normal (por defecto) "hr" para polinomio Hermite "unif" para uniforme</i>
adjustment	<i>La serie de expansión que se va a utilizar en la estimación : "cos" para el coseno (por defecto) "herm" para polinomio Hermite "poly" para polinomio simple</i>

Resultado

El comando devuelve una lista formada por:

ddf La función de detectabilidad del modelo.

dht Información sobre la abundancia / densidad.

Observación

El argumento “data” tiene obligatoriamente que tener una columna con el nombre “distance”, donde estarán recogidas las distancias obtenidas mediante los transectos. Y en el caso de que hubiera una columna con el nombre de “size”, “R” lo entenderá como el tamaño de las agrupaciones de individuos a la hora de observarlos, es decir, si en una observación el individuo está aislado o no.

Si se usa la función clave uniforme entonces las covariantes no pueden ser usadas en el modelo.

En la serie de expansión un valor válido es “NULL” el cual indicará que en el ajuste no será necesario introducir la serie.

ds.gof

Descripción

Si se define la bondad de ajuste de un modelo estadístico como la discrepancia entre los valores observados y los valores esperados en el modelo de estudio, entonces este comando realiza un test de bondad de ajuste sobre la función de detectabilidad del modelo obtenido previamente (por ejemplo con el comando ds). Usa los test Chi-cuadrado, Kolmogorov-Smirnov (si ks=TRUE) y Cramer-Von Mises.

Uso

`ds.gof(model,qq=TRUE,ks=TRUE)`

Argumentos

model	<i>Modelo ajustado previamente, ddf del comando ds.</i>
qq	<i>Si se desea ver el plot cuantil – cuantil.</i>
ks	<i>El tipo de test que se aplica : Chi – cuadrado Cramér – Vom Mises Kolmogorov – Smirnov que es el que usaremos, ks = TRUE.</i>

Resultado

El comando devuelve el resultado del test y el qq-plot (el plot cuantil-cuantil).

`gof_ds`

Descripción

Test de Bondad de ajuste de la función de detectabilidad usando los test Kolmogorov-Smirnov y Cramer-von Mises. Ambos test se basan en observar el plot cuantil-cuantil obtenido, y ver la desviación respecto de la recta $y = x$.

Uso

```
gof_ds(model,plot=TRUE,chisq=FALSE,nboot=100,ks=FALSE)
```

Argumentos

model	<i>Modelo ajustado previamente, ddf del comando ds.</i>
plot	<i>Si toma el valor TRUE entonces el plot es dibujado.</i>
chisq	<i>Si toma el valor TRUE entonces calcula el estadístico del Chi – cuadrado.</i>
nboot	<i>Número de réplicas que se usaran para el cálculo del p – valor para el test Kolmogorov – Smirnov.</i>
ks	<i>Si toma el valor TRUE, entonces realiza el test Kolmogorov – Smirnov.</i>

Resultado

El comando devuelve el resultado del test y un plot en caso de haber usado `plot=TRUE`.

AIC.dsmodel

Descripción

Devuelve el Criterio de Información de Akaike (AIC) de la función de detectabilidad ajustada.

Uso

`AIC(object,k=2)`

Argumentos

object	<i>Es una función de detectabilidad ajustada.</i>
k	<i>Término de penalización para el número de parámetros que por defecto es 2 (AIC usual).</i>

Resultado

Devuelve el valor AIC del modelo.

plot.dsmodel

Descripción

Dibuja la función de detectabilidad, así como el histograma de las distancias observadas.

Uso

`plot(x,pl.den=0)`

Argumentos

x	<i>Es una función de detectabilidad ajustada ddf, o el modelo ds también es válido.</i>
pl.den	<i>El sombreado del histograma, por defecto esta sin sombrear con valor 0.</i>

Resultado

Devuelve el histograma de las distancias y la gráfica de la función de detectabilidad ajustada.

3.1.1. Ejemplo de transecto puntual

3.1.1.1. Introducción del ejemplo

Este ejemplo se ha extraído del libro Buckland y otros [3] , pagina 179, ejercicio 3. El objetivo de este ejemplo es explicar como obtener la abundancia de una población mediante el paquete *Distance*

3.1.1.2. Datos

Sean 10 transectos puntuales para los cuales se han obtenido las siguientes observaciones:

- En el transecto $p1$ se han observado dos casos, a distancias 12 y 35.
- En el transecto $p2$ se han observado cuatro casos, a distancias 10, 25, 30 y 45.
- En el transecto $p3$ se han observado tres casos, a distancias 8, 15 y 50.
- En el transecto $p4$ se han observado seis casos, a distancias 20, 20, 30, 35, 45 y 60.
- En el transecto $p5$ no se han observado casos.
- En el transecto $p6$ se han observado cinco casos, a distancias 20, 25, 30, 30 y 40.
- En el transecto $p7$ se han observado cuatro casos, a distancias 30, 40, 45 y 45.

- En el transecto $p8$ se han observado un caso, a distancia 25.
- En el transecto $p9$ se han observado seis casos, a distancias 15, 25, 30, 40, 40 y 50.
- En el transecto $p10$ se han observado un caso, a distancia 35.

En total el número de casos observados es 32.

Se introduce los datos en R

```
p1=c(12,35)
p2=c(10,25,30,45)
p3=c(8,15,50)
p4=c(20,20,30,35,45,60)
p5=c()
p6=c(20,25,30,30,40)
p7=c(30,40,45,45)
p8=c(25)
p9=c(15,25,30,40,40,50)
p10=c(35)
```

Después se agrupan las distintas observaciones en una lista que llamaremos *distance*, ya que el comando `ds` necesita que una de las columnas del `data.frame` que usa como base de datos sea nombrada *distance*.

```
distance=c(p1,p2,p3,p4,p5,p6,p7,p8,p9,p10)
```

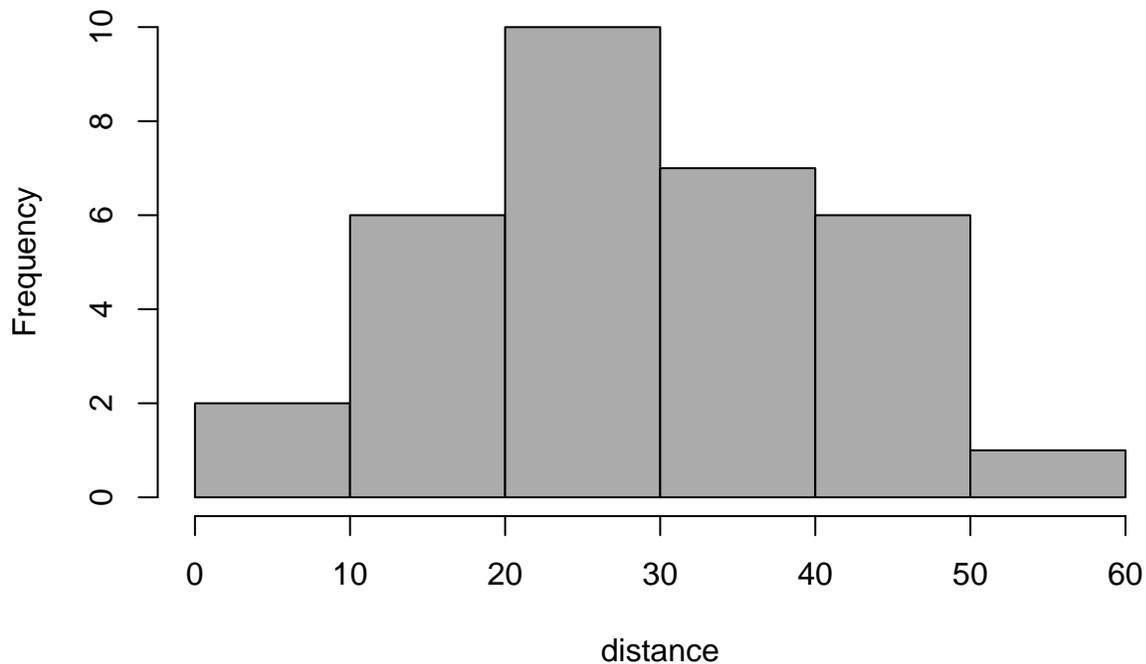
3.1.1.3. Análisis descriptivo de los datos

```
summary(distance)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      8.00  23.75   30.00   31.41  40.00   60.00
```

```
hist(distance,main="Histogramas de las distancias observadas",col="gray67")
```

Histogramas de las distancias observadas



3.1.1.4. Creación del data.frame

```
pointdata=data.frame(distance)
```

3.1.1.5. Análisis del modelo

Empezamos cargando el paquete *Distance*, después consideramos algún modelo y le aplicamos `ds`. Yo consideraré 3 modelos distintos y posteriormente veré cual es el mejor según el criterio de AIC.

```
library("Distance")

pointmodel1=ds(pointdata,transect = c("point"), key = c("hn" ))
pointmodel2=ds(pointdata,transect = c("point"), key = c( "hr"),
adjustment = c( "herm"))
pointmodel3=ds(pointdata,transect = c("point"), key = c("unif"),
adjustment = c("poly"))
```

3.1.1.6. Selección del mejor modelo

Simplemente usando el criterio de AIC, observo que modelo tiene menor AIC y me quedo con él.

```
AIC(pointmodel1)
```

```
## 'log Lik.' 254.2026 (df=1)
```

```
AIC(pointmodel2)
```

```
## 'log Lik.' 254.9039 (df=2)
```

```
AIC(pointmodel3)
```

```
## 'log Lik.' 255.0231 (df=1)
```

Como el modelo pointmodel1 es el que menos AIC tiene, me quedo con el modelo que consta de la función clave de la semi-normal y sin serie de expansión.

```
summary(pointmodel1)
```

```
##
## Summary for distance analysis
## Number of observations : 32
## Distance range       : 0 - 60
##
## Model : Half-normal key function
## AIC   : 254.2026
##
## Detection function parameters
## Scale coefficient(s):
##           estimate      se
## (Intercept) 3.31362 0.1621369
##
##           Estimate      SE      CV
## Average p    0.3809344 0.09360216 0.2457172
## N in covered region 84.0039577 23.71870209 0.2823522
```

Obteniendo una estimación de la probabilidad de detectabilidad $\hat{P}_a = 0,3809344$, y una estimación de la abundancia de $\hat{N} = 84$ ($\hat{N} = n/P_a = 32/0,3809344 = 84$).

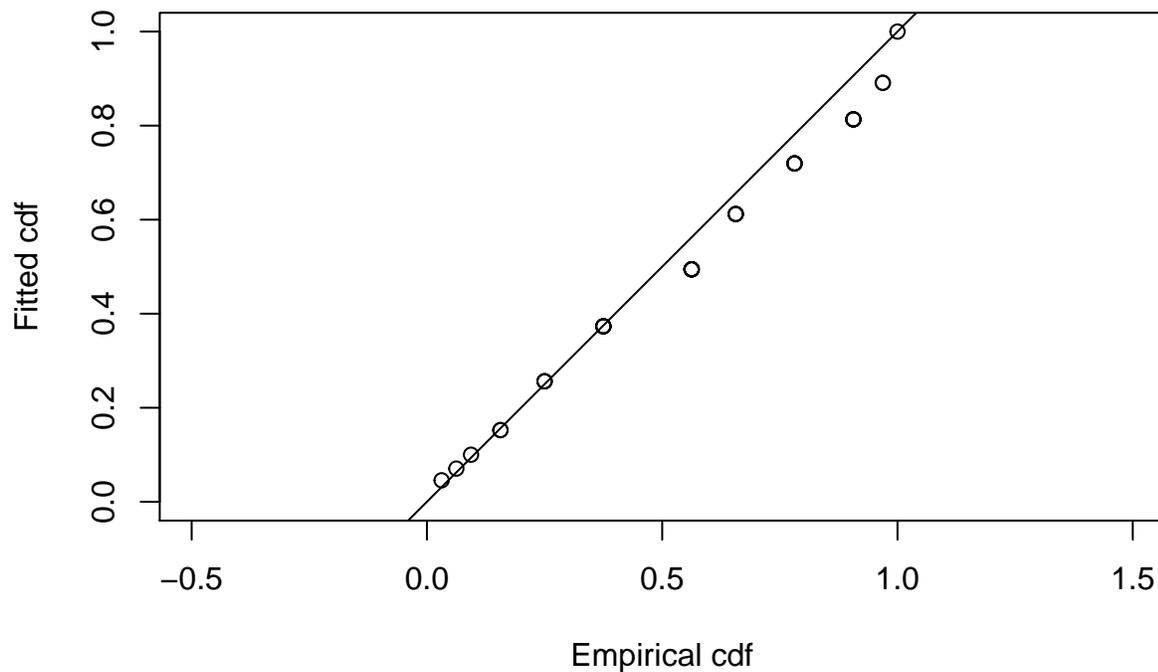
De ambas estimaciones se obtiene el error de estimación (SE) y el coeficiente de variación ($CV = SE/ESTIMATE$). Con esta estimación de la abundancia se puede obtener la correspondiente estimación de la densidad $\hat{D} = \hat{N}/\text{área observada}$

3.1.1.7. Validación del modelo

Aplico un test, por ejemplo el test de Cramer-von Mises para ver si la función de distribución empírica se ajusta lo suficiente a la función de distribución teórica. Hay dos formas de verlo:

- Visualmente, si los puntos están aproximadamente sobre la recta $y=x$
- Se realiza el contraste de hipótesis sobre la igualdad de distribuciones.

```
ds.gof(pointmodel1)
```



```
##
## Goodness of fit results for ddf object
##
## Chi-square tests
##           [0,10]      (10,20]      (20,30]      (30,40]      (40,50]
## Observed  2.00000000  6.0000000000  10.0000000  7.000000000  6.00000000
## Expected  2.25789333  5.9424601062  7.6224944  7.205191360  5.48713129
## Chisquare 0.02945621  0.0005571496  0.7415595  0.005843494  0.04793658
##           (50,60]      Total
## Observed  1.000000  32.000000
## Expected  3.484830  32.000000
## Chisquare 1.771788  2.597141
##
## P = 0.62733 with 4 degrees of freedom
##
```

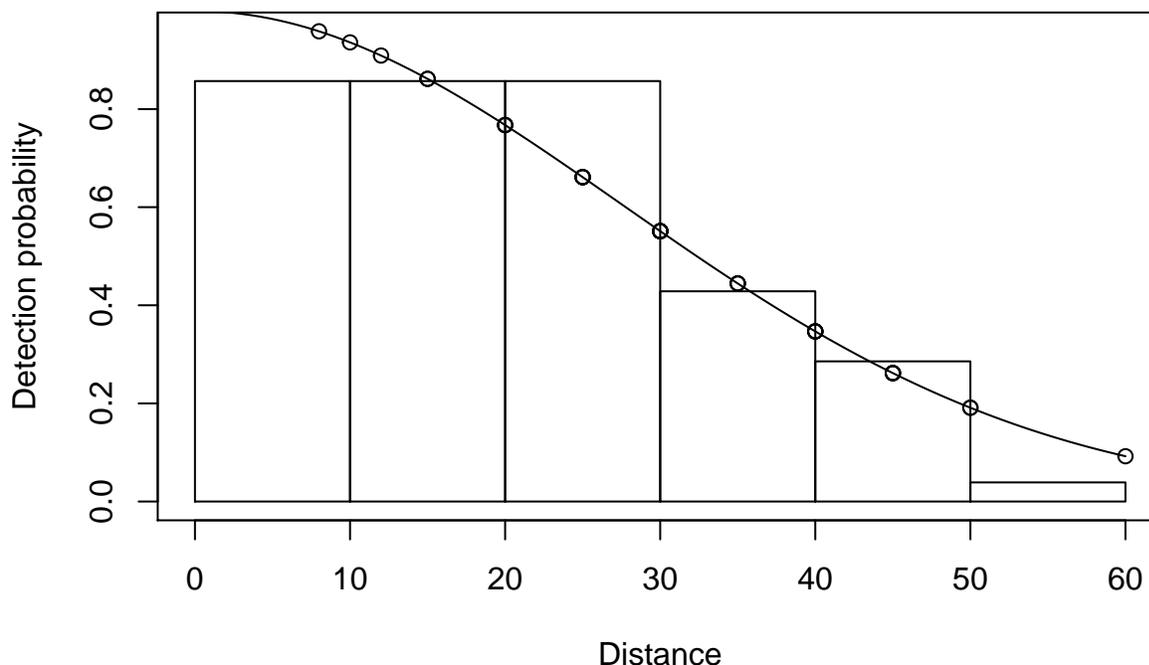
```
## Distance sampling Cramer-von Mises test (unweighted)
## Test statistic = 0.0781576 p-value = 0.702579
```

Para este modelo se obtiene un p-valor de 0.702579, por lo cual no rechazo la hipótesis nula de igualdad de las distribuciones empírica y ajustada por el modelo.

3.1.1.8. Función de detectabilidad

El dibujo de la función de detectabilidad ajustada a través de la técnica aplicada es:

```
plot(pointmodel1)
```



3.1.2. Ejemplo de transecto lineal

3.1.2.1. Introducción del ejemplo

Este ejemplo se ha extraído del libro Buckland y otros [30], página 7, tabla 1.1. El objetivo de este ejemplo es explicar cómo obtener la abundancia de una población mediante el paquete *Distance*.

3.1.2.2. Datos

Sean 19 transectos lineales para los cuales se han obtenido las siguientes distancias perpendiculares a la línea:

- En el transecto $l1$ se han observado tres casos, a distancias 12, 26 y 70.
- En el transecto $l2$ se han observado ocho casos, a distancias 5, 5, 20, 24, 30, 45, 50 y 53.
- En el transecto $l3$ se han observado once casos, a distancias 0, 5, 5, 15, 20, 22, 27, 28, 32, 40 y 60.
- En el transecto $l4$ se han observado cinco casos, a distancias 15, 35, 72, 75 y 100.
- En el transecto $l5$ se han observado ocho casos, a distancias 15, 21, 22, 28, 33, 35, 45 y 60.
- En el transecto $l6$ se han observado siete casos, a distancias 15, 15, 20, 35, 50, 65 y 70.
- En el transecto $l7$ se han observado cinco casos, a distancias 15, 30, 50, 60 y 85.
- En el transecto $l8$ se han observado siete casos, a distancias 23, 27, 38, 55, 58, 75 y 100.
- En el transecto $l9$ se han observado tres casos, a distancias 15, 45 y 80.
- En el transecto $l10$ se han observado un caso, a distancia 55.
- En el transecto $l11$ no se han observado casos.
- En el transecto $l12$ se han observado dos casos, a distancias 45 y 50.
- En el transecto $l13$ se han observado cinco casos, a distancias 5, 25, 38, 45 y 45.
- En el transecto $l14$ se han observado seis casos, a distancias 5, 10, 27, 40, 45 y 45.
- En el transecto $l15$ se han observado cuatro casos, a distancias 5, 25, 33 y 35.
- En el transecto $l16$ no se han observado casos.
- En el transecto $l17$ se han observado cuatro casos, a distancias 15, 25, 33 y 35.
- En el transecto $l18$ se han observado tres casos, a distancias 10, 25 y 70.
- En el transecto $l19$ no se han observado casos.

En total el número de casos observados es 82.

Se introduce los datos en R

```

l1=c(12,26,70)
l2=c(5,5,20,24,30,45,50,53)
l3=c(0,5,5,15,20,22,27,28,32,40,60)
l4=c(15,35,72,75,100)
l5=c(15,21,22,28,33,35,45,60)
l6=c(15,15,20,35,50,65,70)
l7=c(15,30,50,60,85)
l8=c(23,27,38,55,58,75,100)
l9=c(15,45,80)
l10=c(55)

```

```
l11=c()  
l12=c(45,50)  
l13=c(5,25,38,45,45)  
l14=c(5,10,27,40,45,45)  
l15=c(5,20,25,45)  
l16=c()  
l17=c(15,25,33,35)  
l18=c(10,25,70)  
l19=c()
```

Después se agrupan las distintas observaciones en una lista que llamaremos *distance*, como en el ejemplo anterior.

```
distance=c(l1,l2,l3,l4,l5,l6,l7,l8,l9,l10,l11,l12,l13,l14,l15,l16,l17,l18,l19)
```

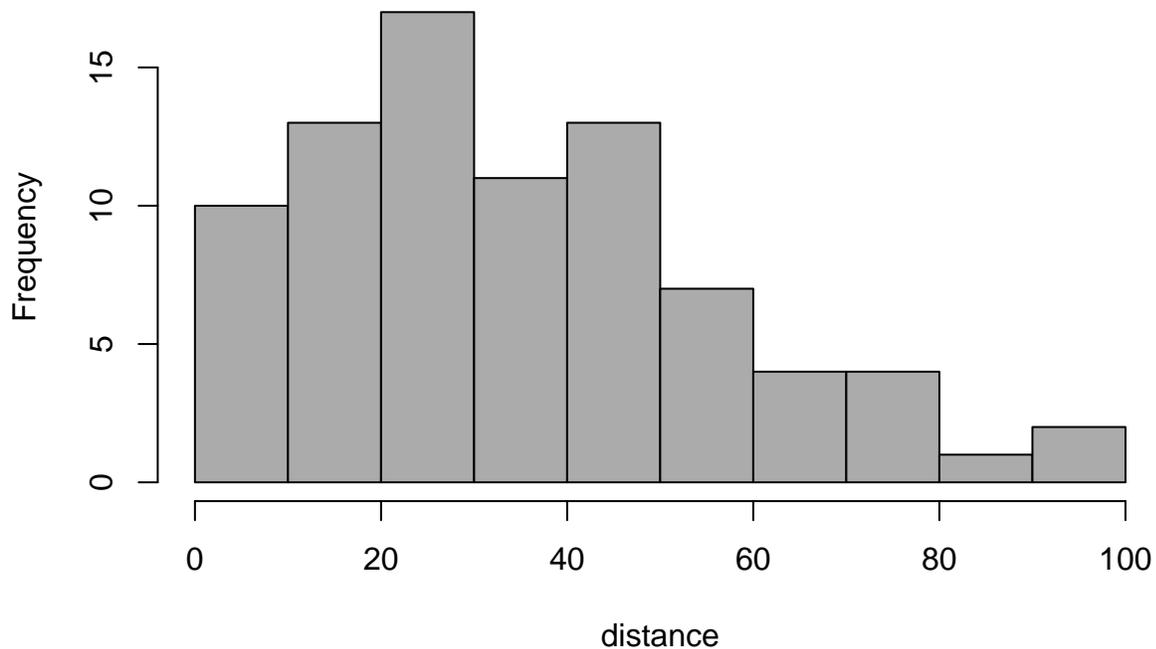
3.1.2.3. Análisis descriptivo de los datos

```
summary(distance)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   
##      0.00  20.00   32.50   36.15  50.00  100.00
```

```
hist(distance,main="Histogramas de las distancias observadas",col="gray67")
```

Histogramas de las distancias observadas



3.1.2.4. Creación del data.frame

```
linedata=data.frame(distance)
```

3.1.2.5. Análisis del modelo

Empezamos cargando el paquete *Distance*, luego aplico el comando `ds` poniendo todas las opciones de funciones clave y series de expansiones, de tal forma que el propio comando `ds` obtiene el mejor modelo.

```
library("Distance")

linemodel=ds(linedata,transect = c("line"), key = c("hn","hr","unif" ),
             adjustment = c("cos","herm","poly"))
```

3.1.2.6. Selección del mejor modelo

Con la orden anterior, se obtiene el mejor modelo.

```
## Starting AIC adjustment term selection.
## Fitting half-normal key function
```

```
## Key only model: not constraining for monotonicity.
## AIC= 732.599
## Fitting half-normal key function with cosine(2) adjustments
## AIC= 733.14
##
## Half-normal key function selected.
## No survey area information supplied, only estimating detection function.
```

Como observamos el mejor modelo según el criterio AIC, es el de la función clave de una semi-normal sin serie de expansión con un AIC=732.599.

Una visualización mas detallada de los resultados se puede obtener usando el comando *summary*

```
summary(linemodel)
```

```
##
## Summary for distance analysis
## Number of observations : 82
## Distance range       : 0 - 100
##
## Model : Half-normal key function
## AIC   : 732.5992
##
## Detection function parameters
## Scale coefficient(s):
##           estimate      se
## (Intercept) 3.858983 0.1175001
##
##           Estimate      SE      CV
## Average p      0.5735152 0.05467541 0.09533384
## N in covered region 142.9778973 17.09143666 0.11953901
```

Obteniendo una estimación de la probabilidad de detectabilidad $\hat{P}_a = 0,5735152$, y una estimación de la abundancia de $\hat{N} = 143$ ($\hat{N} = n/P_a = 82/0,5735152 = 143$).

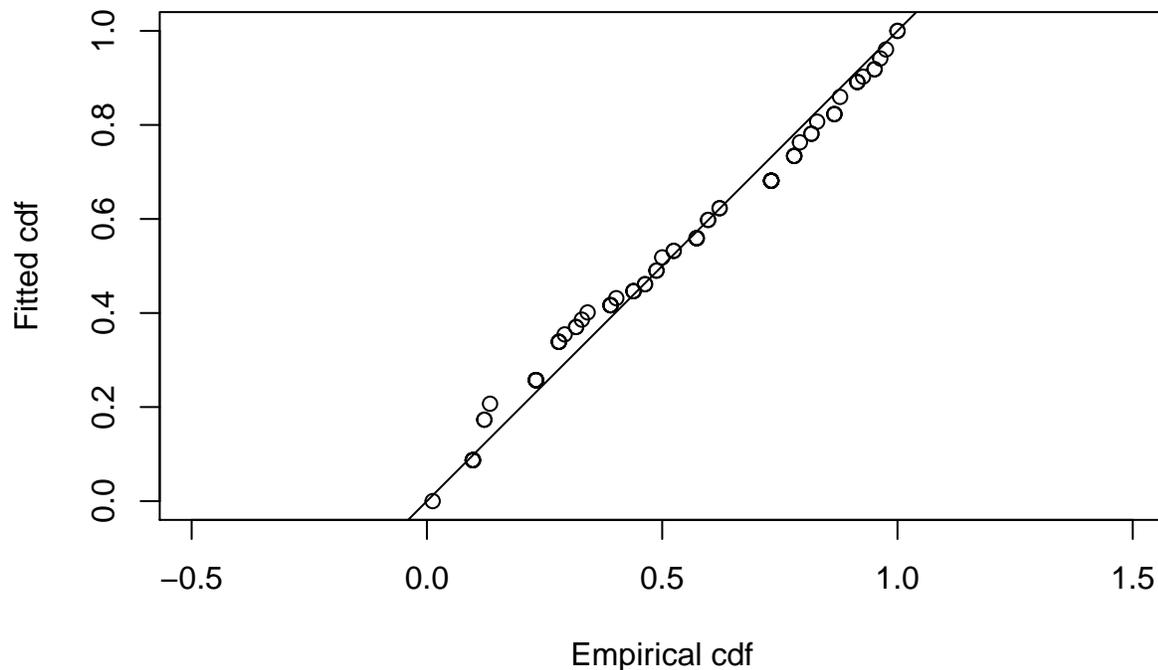
De ambas estimaciones se obtiene el error de estimación (*SE*) y el coeficiente de variación ($CV = SE/ESTIMATE$). Con esta estimación de la abundancia se puede obtener la correspondiente estimación de la densidad $\hat{D} = \hat{N}/\text{área observada}$

3.1.2.7. Validación del modelo

Aplico un test, por ejemplo el test de Cramer-von Mises para ver si las función de distribución empírica se parecen lo suficiente a la función de distribución teórica. Hay dos formas de verlo:

- Visualmente si los puntos están aproximadamente sobre la recta $y=x$.
- Observando el p-valor de la hipótesis nula, las dos distribuciones son iguales.

```
ds.gof(linemodel)
```



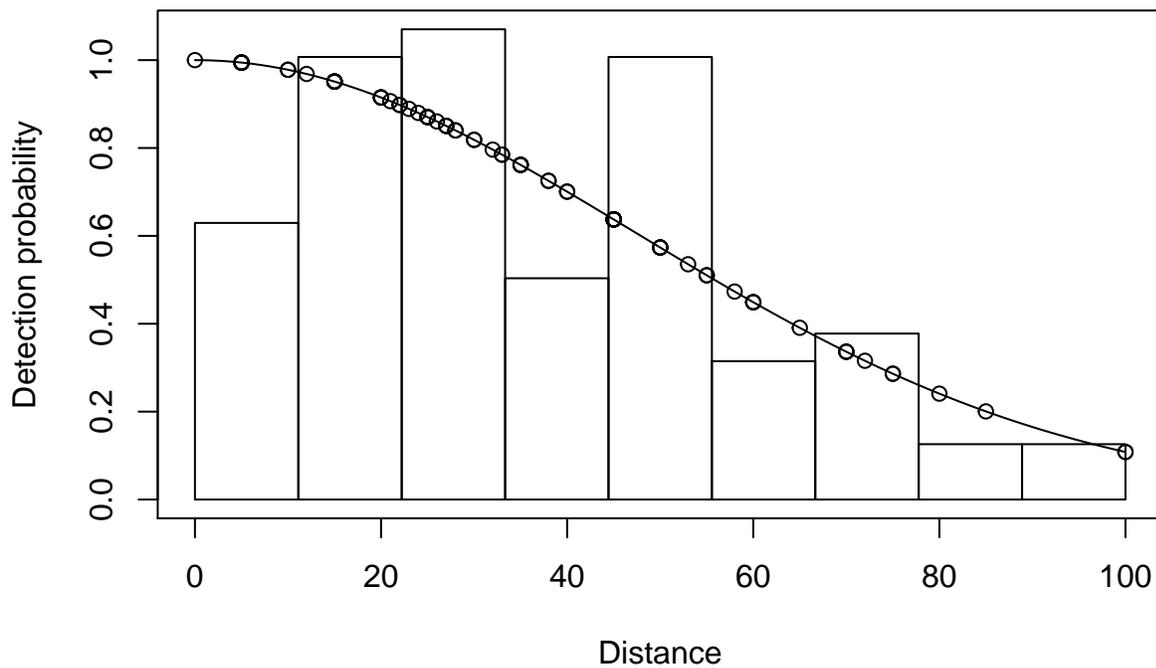
```
##
## Goodness of fit results for ddf object
##
## Chi-square tests
##           [0,11.1] (11.1,22.2] (22.2,33.3] (33.3,44.4] (44.4,55.6]
## Observed  10.000000 16.00000000 17.0000000  8.0000000  16.000000
## Expected  15.742237 14.90488422 13.3614278 11.3406822  9.113547
## Chisquare  2.094575  0.08046212  0.9908528  0.9840817  5.203598
##           (55.6,66.7] (66.7,77.8] (77.8,88.9] (88.9,100]   Total
## Observed   5.0000000  6.0000000  2.0000000 2.0000000 82.00000
## Expected   6.9342186  4.9953885  3.4072380 2.20037610 82.00000
## Chisquare  0.5395275  0.2020352  0.5812094 0.01824714 10.69459
##
## P = 0.15251 with 7 degrees of freedom
##
## Distance sampling Cramer-von Mises test (unweighted)
## Test statistic = 0.163918 p-value = 0.349818
```

Para este modelo se obtiene un p-valor de 0.349818, por lo cual no rechazo la hipótesis nula de igualdad de las distribuciones empírica y ajustada por el modelo.

3.1.2.8. Función de detectabilidad

El dibujo de la función de detectabilidad ajustada por el modelo.

```
plot(linemodel)
```



3.2. Métodos no paramétricos

Los métodos no paramétricos basados en la función núcleo como su nombre indica, es necesario el uso de funciones núcleo y por ello se empezará introduciendo el paquete *evmix* que proporciona una gran variedad de funciones núcleo.

Librería “*evmix*”

Descripción

Es un paquete de *R* que recopila las funciones núcleo más comunes y da las herramientas necesarias para evaluar una función núcleo en un punto.

Las función principal de la librería que se va usar es *kdgaussian*. Ya que uno de los requisitos de la teoría, es que la función núcleo sea simétrica y que la parte importante en el método no paramétrico es la ventana y no tanto la elección de la función núcleo.

kdgaussian

Descripción

Este comando evalúa la función núcleo en un valor *x*.

Uso

```
kdgaussian(x = 0, lambda = NULL, bw = NULL, kerncentres = 0)
```

Argumentos

Resultado

x	<i>El valor donde se quiere evaluar la función núcleo.</i>
lambda	<i>La medio – anchura de la función núcleo ó NULL.</i>
bw	<i>La desviación típica de la función núcleo ó NULL.</i>
kerncentres	<i>El centro de la función núcleo.</i>
z	<i>Estandarización de la función núcleo, $z=(x-kerncentres)/lambda$.</i>
kernel	<i>El nombre de la función núcleo, (default=gaussian).</i>

El comando devuelve el valor numérico de evaluar una función núcleo.

Observación

No siempre la función núcleo con la que se trabaja es una de las conocidas, por lo que no es posible usar un paquete de R, para estos casos la solución es definir las con el comando *function*.

3.2.1. Estimador núcleo tradicional

El estimador considerado es el propuesto por *Chen* [16]

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

con el supuesto de que K sigue una $N(0, 1)$ y $f(x)$ sigue una seminormal con parámetro θ , por lo que $h = 1,06\hat{\theta}n^{-1/5}$. Donde $\hat{\theta} = \sqrt{\sum_{i=1}^n X_i^2/n}$.

Si se llama *distance* a los datos observados, *teta* al parámetro θ y *hopt* al h obtenido para este estimador. Los cálculos en R son los siguientes:

```
teta=sqrt(sum(distance^2)/n)

hopt=1.06*teta*n^(-1/5)

Fh<-function(h){
  2*sum(kdgaussian(distance/h))/(h*n)
}

Fh(hopt)
```

3.2.2. Estimador semiparamétrico

Para implementar los estimadores semiparamétricos basados en la función núcleo en R, se definirá un estimador general que incluya a una amplia cantidad de opciones.

$$\hat{f}(0) \equiv \tilde{\alpha}_n(0) = \frac{2}{nh} \sum_{i=1}^n K\left(\frac{X_i}{h}\right) [g(X_i; \theta)]^{-k}.$$

Para el caso particular $k = 0$ se obtiene el estimador núcleo tradicional.

Se define la función α como una función que depende de h y k , y al variar dichos parámetros se adquiere una amplia variedad de distintos estimadores para $f(0)$.

```
alpha<-function(h, k){
  2*(sum((1/sqrt(2*pi))*exp(-0.5*(distance/h)^2) * (exp((-distance/h)^2/(2*teta^2))^k)))/(h*n)
}
```

Para la obtención del h por el método del m -vecino más cercano, denotado por *hquesenberry*

```
epsilon=4/5
m=trunc(n^epsilon)

estadistico=sort(distance, decreasing=FALSE)

hquesenberry=estadistico[m]
```

Para el cálculo de k se usa la estimación $\hat{k} = \hat{f}''(0)/\hat{g}''(0; \hat{\theta})\hat{f}(0)$.

Donde $g(x; \theta)$ es supuesto un modelo seminormal, entonces $g''(0, \theta) = -1/\theta^2$ y $\hat{g}''(0, \hat{\theta}) = -1/\hat{\theta}^2$, donde $\hat{\theta} = \sqrt{\sum_{i=1}^n x_i^2/n}$ es el estimador de θ obtenido por máxima verosimilitud.

$\hat{f}(0)$ es la estimación obtenida usando el estimador núcleo tradicional.

Para estimar $f''(0)$, Mack [31] recomienda usar $\hat{f}''(0) = [F_n(3\delta) - 3F_n(2\delta) + 3F_n(\delta)]/\delta^3$, donde $F_n(u)$ es la función de distribución empírica y $\delta = 1,97\hat{\theta}n^{-1/9}$.

Que se implementa en *R* usando el comando *ecdf* para el uso de la función empírica de la siguiente forma:

```
teta=sqrt(sum(distance^2) /n)

gsegunda=-1/teta^2

fcero=Fn(hopt)

delta=1.97*teta*n^(-1/9)

Fn=ecdf(distance)

fsegunda=(Fn(3*delta)- 3*Fn(2*delta) +3*Fn(delta))/delta^3

ka=fsegunda/(gsegunda*fcero)
```

Entonces los estimadores más interesantes son estos 6:

$\alpha(hopt, 0)$, $\alpha(hopt, 1)$, $\alpha(hopt, ka)$, $\alpha(hquesenberry, 0)$, $\alpha(hquesenberry, 1)$ y $\alpha(hquesenberry, ka)$.

3.3. Ejemplo de estacas de madera

Como ejemplo aplicados a datos reales, se aplicará la variedad de estimadores propuestos para $f(0)$ con los datos de estacas de madera recogidos por *Burnham* [7] de la página 61. Se colocan 150 estacas de madera de forma aleatoria a lo largo de un area de $L = 1000m$ de largo.

Tras realizar el muestreo por transecto lineal a lo largo de un transecto de longitud $L = 1000$, se han observado 68 estacas de madera y se han anotado las distancias perpendiculares al transecto de cada una de ellas. Con un truncamiento de $w = 20$.

Y se sabe que $f(0) = 0,110294$ y $D = 0,00375 = 37,5$ estacas/hectáreas.

TABLE 6.—EXAMPLE OF LINE TRANSECT DATA FROM A SAMPLE OF WOODEN STAKES OF KNOWN DENSITY; $n = 68$, $w^* = 18.60$ m, $w = 20$ m, $L = 1,000$ m, AND $D = 0.00375$ STAKES PER METER. THE DATA SHOWN ARE THE PERPENDICULAR DISTANCES IN METERS

1.	2.02	35.	3.79
2.	0.45	36.	15.24
3.	10.40	37.	3.47
4.	3.61	38.	3.05
5.	0.92	39.	7.93
6.	1.00	40.	18.15
7.	3.40	41.	10.05
8.	2.90	42.	4.41
9.	8.16	43.	1.27
10.	6.47	44.	13.72
11.	5.66	45.	6.25
12.	2.95	46.	3.59
13.	3.96	47.	9.04
14.	0.09	48.	7.68
15.	11.82	49.	4.89
16.	14.23	50.	9.10
17.	2.44	51.	3.25
18.	1.61	52.	8.49
19.	31.31	53.	6.08
20.	6.50	54.	0.40
21.	8.27	55.	9.33
22.	4.85	56.	0.53
23.	1.47	57.	1.23
24.	18.60	58.	1.67
25.	0.41	59.	4.53
26.	0.40	60.	3.12
27.	0.20	61.	3.05
28.	11.59	62.	6.60
29.	3.17	63.	4.40
30.	7.10	64.	4.97
31.	10.71	65.	3.17
32.	3.86	66.	7.67
33.	6.05	67.	18.16
34.	6.42	68.	4.08

Fuente: Burnham (1980)

3.3.1. Datos

Se introduce la muestra obtenida en R.

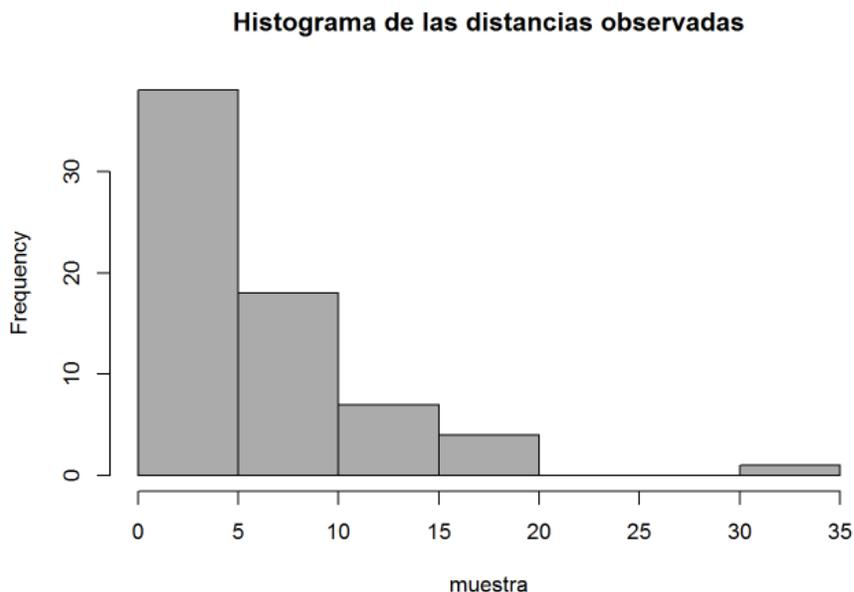
```
muestra=c(2.02,0.45,10.40,3.61,0.92,1.00,3.40,2.90,8.16,6.47,5.66,2.95,3.96,0.09,11.82,14.23,2.44,1.61,31.31,6.50,8.27,4.85,1.47,18.60,0.41,0.40,0.20,11.59,3.17,7.10,10.71,3.86,6.05,6.42,3.79,15.24,3.47,3.05,7.93,18.15,10.05,4.41,1.27,13.72,6.25,3.59,9.04,7.68,4.89,9.10,3.25,8.49,6.08,0.40,9.33,0.53,1.23,1.67,4.53,3.12,3.05,6.60,4.40,4.97,3.17,7.67,18.16,4.0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

3.3.2. Análisis descriptivo de los datos

```
summary(muestra)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.090  2.938   4.470   6.108   8.188  31.310
```

```
hist(muestra,main="Histograma de las distancias observadas",col="gray67")
```

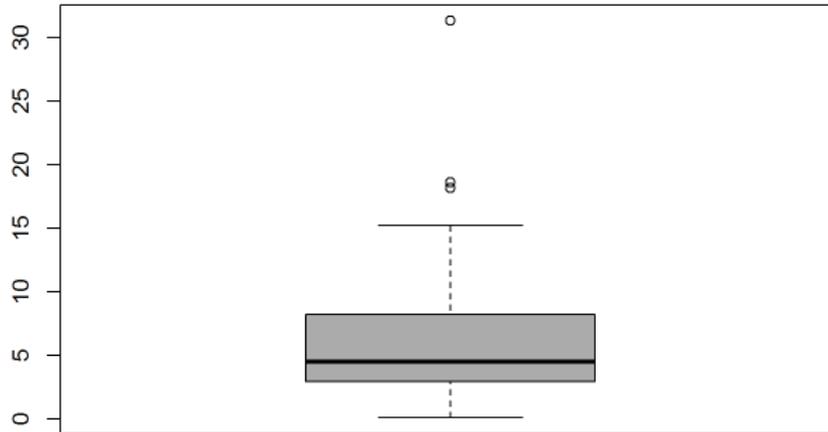


Se ve que existe un elemento, concretamente la observación 19 que toma el valor 31.31, que es un valor atípico. Por lo que tiene sentido considerar un truncamiento como se indica de $w = 20$.

Por ello a partir de ahora se considerará el conjunto de la muestra con $n = 67$ en vez de 68, eliminando esta observación atípica. En el caso del método paramétrico esto no es necesario, ya que añadiendo la opción *truncation* = 20 al comando *ds* este realiza el truncamiento de los datos.

```
boxplot(muestra,main="Boxplot de las distancias observadas",col="gray67")
```

Boxplot de las distancias observadas



Dicha muestra para diferenciarla de la anterior pasara a llamarse *distance*, lo cual es necesario si se quiere aplicar el comando `ds`, debido a que el `data.frame` necesita que una de sus columnas, concretamente las de la distancias tenga dicho nombre.

3.3.3. Creación del `data.frame`

```
distance=c(2.02,0.45,10.40,3.61,0.92,1.00,3.40,2.90,8.16,6.47,5.66,2.95,3.96,0.09,11.82,14.23,2.44,1.61,6.50,8.27
,4.85,1.47,18.60,0.41,0.40,0.20,11.59,3.17,7.10,10.71,3.86,6.05,6.42,3.79,15.24,3.47,3.05,7.93,18.15,10.05,4.41,
1.27,13.72,6.25,3.59,9.04,7.68,4.89,9.10,3.25,8.49,6.08,0.40,9.33,0.53,1.23,1.67,4.53,3.12,3.05,6.60,4.40,4.97,3.
17,7.67,18.16,4.08);
```

```
n=length(distance)
n
```

```
## [1] 67
```

```
estacadata=data.frame(distance)
```

3.3.4. Selección del modelo paramétrico

El comando `ds` elige automáticamente el mejor modelo para los datos basándose en AIC.

```

modelestaca=ds(estacadata,transect = "line",truncation = 20)

## Starting AIC adjustment term selection.

## Fitting half-normal key function

## Key only model: not constraining for monotonicity.

## AIC= 364.714

## Fitting half-normal key function with cosine(2) adjustments

## AIC= 365.664

##
## Half-normal key function selected.

## No survey area information supplied, only estimating detection function.

summary(modelestaca)

##
## Summary for distance analysis
## Number of observations : 67
## Distance range       : 0 - 20
##
## Model : Half-normal key function
## AIC   : 364.7145
##
## Detection function parameters
## Scale coefficient(s):
##           estimate      se
## (Intercept) 2.020857 0.08824599
##
##           Estimate      SE      CV
## Average p      0.4690034 0.03875852 0.08264017
## N in covered region 142.8561014 17.35258912 0.12146901

```

Obteniéndose una estimación de la probabilidad de detectabilidad de $\hat{P}_a = 0,4690034$ y una estimación del número de estacas en el área de $\hat{N} = 143$.

Y como el área de estudio es $2wL = 2 * 20 * 1000 = 40000$, la densidad estimada es $\hat{D} = \hat{N}/2wL = 143/40000 = 0,003575 = 35,75 \text{ estacas/hectáreas}$.

Y como $\hat{D} = \frac{n\hat{f}(0)}{2L}$ implica que $\hat{f}(0) = \frac{\hat{D} * 2L}{n} = \frac{0,003575 * 2000}{67} = 0,1067164$.

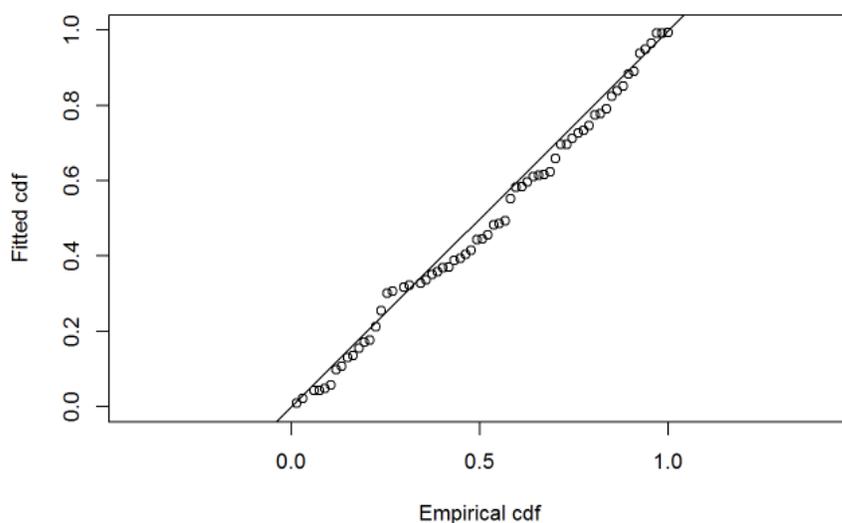
El error que $\hat{f}(0)$ comete con respecto al valor real $f(0)$ en valor absoluto es $|0,1067164 - 0,110294| = 0,0035776$

El error que \hat{D} comete con respecto al valor real D en valor absoluto es $|35,75 - 37,5| = 1,75$ estacas/hectáreas.

3.3.5. Validación del modelo

Aplico el test de Cramer-von Mises para ver si la función de distribución empírica se ajusta lo suficiente a la función de distribución teórica.

```
ds.gof(modelestaca)
```



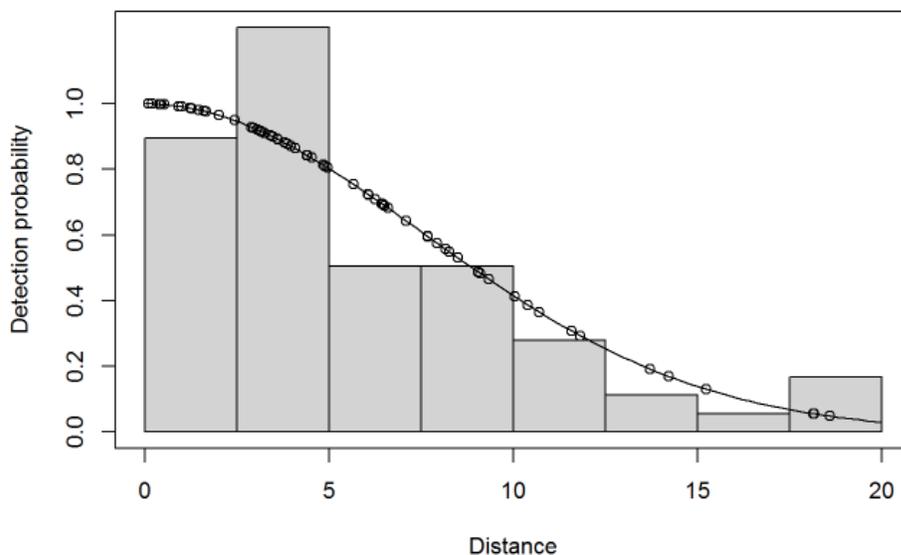
```
##
## Goodness of fit results for ddf object
##
## Chi-square tests
##           [0,2.5]  (2.5,5]  (5,7.5]  (7.5,10] (10,12.5] (12.5,15]
## Observed  16.0000000 22.000000  9.000000 9.000000000 5.0000000 2.0000000
## Expected  17.5355541 15.727887 12.652363 9.128979878 5.9077300 3.4289786
## Chisquare  0.1344655  2.501251  1.054329 0.001822308 0.1394738 0.5955067
##           (15,17.5] (17.5,20]  Total
## Observed  1.0000000 3.0000000 67.00000
## Expected  1.7850570 0.8334501 67.00000
## Chisquare 0.3452632 5.6319368 10.40405
##
## P = 0.10864 with 6 degrees of freedom
##
## Distance sampling Cramer-von Mises test (unweighted)
## Test statistic = 0.067984 p-value = 0.763993
```

Para este modelo se obtiene un p-valor de 0.763993, por lo cual no rechazo la hipótesis nula de igualdad de las distribuciones empírica y ajustada por el modelo.

3.3.6. Función de detectabilidad

El dibujo de la función de detectabilidad ajustada por el modelo

```
plot(modelestaca)
```



3.3.7. Estimación no paramétrica

El estimador núcleo tradicional propuesto por *Chen* [16] es

$$\hat{F}_h(0) = \frac{2}{nh} \sum_{i=1}^n K\left(\frac{X_i}{h}\right).$$

supuesto que K sigue una $N(0, 1)$ y el h es estimado por $\hat{h} = h_{opt} = 1,06\hat{\theta}n^{-1/5}$ con $\hat{\theta} = \sqrt{\sum_{i=1}^n x_i^2/n}$.

Usándose *library(evmix)* se obtiene

```
Fh<-function(h){
  2*sum(kdgaussian(distance/h))/(h*n)
}

n=length(distance)
n
```

```
## [1] 67
```

```
teta=sqrt(sum(distance^2) /n)
teta
```

```
## [1] 7.301217
```

```
hopt=1.06*teta*n^(-1/5)
hopt
```

```
## [1] 3.337999
```

```
Fh(hopt)
```

```
## [1] 0.1025105
```

De donde se obtiene $\hat{f}_h(0) = 0,1025105$.

Y como $\hat{D} = \frac{n\hat{f}_h(0)}{2L}$ implica que $\hat{D} = \frac{67 \cdot 0,1025105}{2000} = 0,003434102 = 34,34102 \text{ estacas/hectáreas}$.

El error que $\hat{f}_h(0)$ comete con respecto al valor real $f(0)$ en valor absoluto es $|0,1025105 - 0,110294| = 0,0077835$

El error que \hat{D} comete con respecto al valor real D en valor absoluto es $|34,34102 - 37,5| = 3,15898 \text{ estacas/hectáreas}$.

3.3.8. Método semiparamétrico

Para hacerlo lo más general posible, se plantea un estimador basado en la función núcleo multiplicado por una función paramétrica elevado a un número $-k$.

$$\tilde{f}_h(0) \equiv \tilde{\alpha}_h(0) = \frac{2}{nh} \sum_{i=1}^n K\left(\frac{X_i}{h}\right) [g(X_i; \theta)]^{-k}.$$

Ya se ha visto en teoría que podemos suponer K una $N(0, 1)$ y $g(X; \theta)$ que sigue una seminormal con $\hat{\theta} = \sqrt{\sum_{i=1}^n x_i^2/n}$.

```
epsilon=4/5
m=trunc(n^epsilon)
m
```

```
## [1] 28
```

Otro h importante es el *hquesenberry* basado en el $m - \text{vecino}$ más cercano. Para ello primero ordeno la muestra, transformándola en estadísticos ordenados y considero el elemento $m - \text{esimo}$ como h .

k es elegido como $\hat{k} = \hat{f}''(0)/g''(0; \hat{\theta})\hat{f}_h(0)$.

```
estadistico=sort(distance, decreasing=FALSE)
estadistico
```

```
## [1] 0.09 0.20 0.40 0.40 0.41 0.45 0.53 0.92 1.00 1.23 1.27 1.47
## [13] 1.61 1.67 2.02 2.44 2.90 2.95 3.05 3.05 3.12 3.17 3.17 3.25
## [25] 3.40 3.47 3.59 3.61 3.79 3.86 3.96 4.08 4.40 4.41 4.53 4.85
## [37] 4.89 4.97 5.66 6.05 6.08 6.25 6.42 6.47 6.50 6.60 7.10 7.67
## [49] 7.68 7.93 8.16 8.27 8.49 9.04 9.10 9.33 10.05 10.40 10.71 11.59
## [61] 11.82 13.72 14.23 15.24 18.15 18.16 18.60
```

```
hquesenberry=estadistico[m]
hquesenberry
```

```
## [1] 3.61
```

La consistencia del estimador de la función núcleo clásica es usada para estimar $f(0)$. Para estimar $f''(0)$, Mack [31] recomienda usar $\hat{f}''(0) = [F_n(3\delta) - 3F_n(2\delta) + 3F_n(\delta)] / \delta^3$, donde $F_n(u)$ es la función de distribución empírica y $\delta = 1,97\hat{\theta}_n^{-1/9}$.

```
teta=sqrt(sum(distance^2) /n)
teta
```

```
## [1] 7.301217
```

```
gsegunda=-1/teta^2
gsegunda
```

```
## [1] -0.01875899
```

```
fcero=Fh(hopt)
fcero
```

```
## [1] 0.1025105
```

```
delta=1.97*teta*n^(-1/9)
delta
```

```
## [1] 9.01497
```

```
Fn=ecdf(distance)
fsegunda=(Fn(3*delta)- 3*Fn(2*delta) +3*Fn(delta))/delta^3
fsegunda
```

```
## [1] 0.0006926458
```

```
ka=fsegunda/(gsegunda*fcero)
ka
```

```
## [1] -0.3601914
```

```
alpha(hquesenberry, ka)
```

```
## [1] 0.1013557
```

```
alpha(hopt, ka)
```

```
## [1] 0.1028014
```

El estimador propuesto para $f(0)$ es $\tilde{f}_h(0) = \alpha(hquesenberry, ka) = 0,1013557$.

Y como $\hat{D} = \frac{n\tilde{f}_h(0)}{2L}$ implica que $\hat{D} = \frac{67*0,1013557}{2000} = 0,003395416 = 33,95416 \text{ estacas/hectáreas}$.

El error que $\tilde{f}_h(0)$ comete con respecto al valor real $f(0)$ en valor absoluto es $|0,1013557 - 0,110294| = 0,0089383$

El error que \hat{D} comete con respecto al valor real D en valor absoluto es $|33,95416 - 37,5| = 3,54584 \text{ estacas/hectáreas}$.

Otro estimador de interés sería $\alpha(hopt, ka) = 0,1028014$. El cual proporciona una estimación $\hat{D} = 0,003443847 = 34,43847 \text{ estacas/hectáreas}$.

El error que $\alpha(hopt, ka)$ comete con respecto al valor real $f(0)$ en valor absoluto es $|0,1028014 - 0,110294| = 0,0074926$

El error que \hat{D} comete con respecto al valor real D en valor absoluto es $|34,43847 - 37,5| = 3,06153 \text{ estacas/hectáreas}$.

Otro estimador de interes sería $\alpha(hopt, 1) = 0,1017149$.

El cual proporciona una estimación $\hat{D} = 0,003407449 = 34,07449 \text{ estacas/hectáreas}$.

El error que $\alpha(hopt, 1)$ comete con respecto al valor real $f(0)$ en valor absoluto es $|0,1017149 - 0,110294| = 0,0085791$

```
alpha(hopt,1)
```

```
## [1] 0.1017149
```

El error que \hat{D} comete con respecto al valor real D en valor absoluto es $|34,07449 - 37,5| = 3,42551$ *estacas/hectáreas*.

Se obtiene el estimador núcleo tradicional en el estimador $alpha(h, k)$ para $h = hopt$ y $k = 0$.

Bibliografía

- [1] 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). Press, Oxford.
- [2] Buckland S. T., Anderson D. R., Burnham K. P., Laake J. L., Borchers D. L., and Thomas L. (2004) *Advanced Distance Sampling: Estimating Abundance of Biological Populations*. Oxford. pp 1-11.
- [3] Buckland S. T., Anderson D. R., Burnham K. P., Laake J. L. (1993) *Distance Sampling: Estimating Abundance of Biological Populations*, London: Chapman and Hall.
- [4] Laake J. L. , Buckland S. T., Anderson D. R., Burnham K. P. (1993) *DISTANCE Users Guide V2.0.*, Colorado State University.
- [5] Miller D. L., Burt M. L., Rexstad E. A., Thomas L. (2013), *Spatial models for distance sampling data: recent developments and future directions*.
- [6] Buckland S. T., Summers R. W., Borchers D. L., Thomas L. (2006) *Point transect sampling with traps or lures. Journal of Applied Ecology*.
- [7] Kenneth P., Burnham, David R., Anderson, Jeffrey L., Laake (1980) *Wildlife Monographs: Estimation of Density from Line Transect Sampling of Biological Populations*. Wildlife Monograph number 72.
- [8] Gates C. E. , Mashall W. H., Olson D. P. (1968). *Line transect method of estimating grouse population densities.. Biometrics, 24, pp 135-145*.
- [9] 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.
- [10] Quinn T. J., Gallucci V. F. (1980) *Parametric models for line transect estimators of abundance. Ecology, 61, pp 293-302*.
- [11] Zhang S. (2009) *Environmental and Ecological Statistics, 18, pp 79-92*.
- [12] Ababneh F., Eidous O. M. (2012) *Journal of Modern Applied Statistical Methods, 11, pp 144-151*.

- [13] Saeed G. A. A. (2013) *New Parametric Model for Grouped and Ungrouped Line Transect Data Master Thesis Department of statistics, Yarmouk University.*
- [14] Al Eibood F., Eidous O. M. (2017) *Mathematics and Statistics, 5, pp 1-4.*
- [15] Gamil Abdulraqueeb Abdullah Saeed (2017) *Model for estimating of population abundance using line transect sampling.*
- [16] Chen S. X. (1996) *A kernel estimate for the density of a biological population by using line transect sampling. J. Roy. Statist. Soc. Ser. 45, pp 135-150.*
- [17] Buckland S. T. (1992) *Fitting Density Functions with Polynomials., Applied Statistics, 41, pp 63-76.*
- [18] Mack Y. P., Quang P. X. (1998) *Kernel methods in line and point transect sampling. Biometrics, 54, pp 606-619.*
- [19] Zhang S. (2001) *Generalized likelihood ratio test for the shoulder condition in the line transect sampling.. 30, pp 2343-2354*
- [20] Omar Eidous (2006) *A Semiparametric Model for Line Transect Sampling. Communications in Statistics- Theory and Methods. 35, pp 1211-1222.*
- [21] Barabesi L. (2000) *Local likelihood density estimation in line transect sampling. Environmetrics, 11, pp 413-422.*
- [22] Pollock K. H. (1978) *A family of density estimators for line transect sampling. Biometrics, 34, pp 475-478.*
- [23] Hayes R. J., Buckland S. T. (1983) *Radial distance models for line transect method.*
- [24] Eberhardt L. L. (1968) *A preliminary appraisal of line transects. 32, pp 82-88*
- [25] Hardle W., Marron J., Wand M. (1990) *Bandwidth choice for density derivatives. 52, pp 223-232.*
- [26] Barabesi L. (2001) *Local parametric density estimation methods in line transect sampling. Metron, LIX, pp 21-37.*
- [27] Omar Eidous (2009) *Kernel Method Starting with Half-Normal Detection Function for Line Transect Density Estimation. Communications in Statistics- Theory and Methods*
- [28] Gerard P. D., Schucany W. R. (1999) *Local Bandwidth Selection for Kernel Estimation of Population Densities with Line Transect sampling. Biometrics, 55, pp 769-773.*
- [29] Loftsgaarden D., Quesenberry C. (1965) *A nonparametric estimate of multivariate density function. Ann. Math. Statist., 36, pp 1049-1051.*
- [30] Buckland S. T., Rexstad E. A., Marques T. A., Oedekoven C. S. (2015) *Distance sampling : methods and applications* Springer.
- [31] Mack Y. P. (2002) *Bias corrected confidence intervals for wildlife abundance estimation. Commun. Statist. Theor. Meth., 31, pp 1107-1122.*

- [32] Buckland S. T. (1985) *Perpendicular distance models for line transect sampling*. *Biometrics*, 41, pp 177-195.
- [33] Omar Eidous (2005) *On Improving Kernel Estimators Using Line Transect Sampling*. *Communications in Statistics- Theory and Methods*. 34, pp 931-941.
- [34] Stephen T., Buckland, Eric A., Rexstad, Jeff L., Laake, Samantha Strindberg, Sharon L., Hedley, Jon R.B., Bishop, Tiago A., Marques, Kenneth P., Burnham (2010) *Distance software: design and analysis of distance sampling surveys for estimating population size* *Journal of Applied Ecology*.
- [35] Woncheol Jang, Ji Meng Loh. (2010) *Density Estimation for Grouped Data with Application to Line Transect Sampling*. University of Georgia and Columbia University.
- [36] Mohammad Al-Bassam, Omar Eidous (2018) *Combination of parametric and nonparametric estimators for population abundance using line transect sampling*. *Communications in Statistics- Theory and Methods*
- [37] Francois Bolduc, Christian Roy (2018) *Analysis in R: Analysis of multi-species surveys*, Univ of St Andrews.
- [38] David L., Miller, Eric Rexstad, Len Thomas, Laura Marshall, Jeffrey L., Laake (2019) *Distance Sampling in R*.
- [39] Trent McDonald, Jason Carlisle, Aidan McDonald (2019) *Rdistance: Distance-Sampling Analyses for Density and Abundance Estimation*.