
Bootstrap bandwidth selection in kernel hazard rate estimation

S. Jansen van Vuuren (Hons. B.Sc.)

Dissertation submitted in partial fulfillment of the requirements
for the degree Magister Scientiae in Statistics at the
North-West University

Supervisor: Prof. F. C. Van Graan

May 2011

Potchefstroom

Abstract

The purpose of this study is to thoroughly discuss kernel hazard function estimation, both in the complete sample case as well as in the presence of random right censoring. Most of the focus is on the very important task of automatic bandwidth selection. Two existing selectors, least-squares cross validation as described by Patil (1993a) and Patil (1993b), as well as the bootstrap bandwidth selector of Gonzalez-Manteiga, Cao and Marron (1996) will be discussed. The bandwidth selector of Hall and Robinson (2009), which uses bootstrap aggregation (or “bagging”), will be extended to and evaluated in the setting of kernel hazard rate estimation. We will also make a simple proposal for a bootstrap bandwidth selector. The performance of these bandwidth selectors will be compared empirically in a simulation study. The findings and conclusions of this study are reported.

Key words: Hazard function; Kernel estimation; Bandwidth selection; Bootstrap; Bagging; Cross validation; Censored data.

Uittreksel

Die doel van hierdie studie is om kernberaming van die uitvalsfunksie deeglik te bespreek, in die geval van volledige data sowel as in die geval van ewekansige regs-gesensoreerde data. Die fokus sal meestal op die belangrike taak van outomatiese seleksie van die bandwydte wees. Twee bestaande selektors, kleinste-kwadrate kruisgeldigheid soos vermeld in Patil (1993a) en Patil (1993b), sowel as die skoenlus bandwydte selektor van Gonzalez-Manteiga et al. (1996) sal bespreek word. Die bandwydte selektor van Hall and Robinson (2009), wat gebruik maak van skoenlus samevoeging (oftewel “bagging”), sal uitgebrei en ondersoek word in die konteks van kernberaming van die uitvalsfunksie. Ons sal ook ’n eenvoudige voorstel maak vir ’n bandwydte selektor, gebasseer op die skoenlusmetode. Die prestasie van hierdie metodes sal empiries vergelyk word in ’n simulasiestudie. Die bevindinge en gevolgtrekkings van die studie word bespreek.

Sleutelwoorde: Uitvalsfunksie; Kernberaming; Bandwydte seleksie; Skoenlusmetode; Skoenlus samevoeging; Kruisgeldigheid; Gesensoreerde data.

Summary

Chapter 1 provides an overview of survival analysis, the hazard function and some well-known lifetime distributions. Various types of censoring are discussed. Furthermore we introduce the empirical distribution function, and conclude the chapter by deriving the Kaplan-Meier estimator.

Chapter 2 discusses bootstrap methodology. We explain the plug-in principle, as well as bootstrap estimates of some discrepancy measures. We discuss the smooth bootstrap, and motivate why the smooth bootstrap is required in this context. Finally we discuss bootstrapping in the setting of random right censored data.

Chapter 3 covers the topic of kernel smoothing. Various discrepancy measures are introduced and discussed. We also explain the trade-off concept with regards to the choice of the bandwidth h . Kernel smoothing is then introduced. We discuss properties of kernel functions, and provide examples of kernel estimators in various contexts.

Chapter 4 provides an overview of the bandwidth selectors considered in this study. We define the kernel hazard function estimators to be considered, and we derive an explicit expression for the bootstrap mean squared error of the so-called internal estimator in the uncensored case. We discuss the least-squares cross validation selector as described by Patil (1993a) and Patil (1993b). Secondly the bootstrap bandwidth selector of Gonzalez-Manteiga et al. (1996) is discussed. A bandwidth selector which uses bagging in the context of kernel density estimation as described in Hall and Robinson (2009) is extended to the setting of kernel hazard rate estimation. We then propose a simple bootstrap bandwidth selector.

Chapter 5 deals with the simulation study. We explain the structure of the empirical study and also provide algorithms for implementing the automatic bandwidth selectors. We provide results from the empirical study in the form of tables and graphs. Finally we come to conclusions regarding the performance of the bandwidth selectors, suggest some areas of future research and state results from secondary studies.

Opsomming

Hoofstuk 1 gee 'n oorsig van oorlewingsteorie, die uitvalsfunksie asook sekere bekende leeftyd verdelings. Verskeie vorme van sensorering word bespreek. Die empiriese verdelingsfunksie word voorgestel, waarna die afleiding van die Kaplan-Meier beramer behandel word.

Hoofstuk 2 bespreek skoenusmetodologie. Ons verduidelik die inprop beginsel, sowel as skoenusberamers vir sekere verliesfunksies. Die gladde skoenus word bespreek, en die motivering vir die gebruik daarvan in hierdie konteks word gegee. Laastens word die skoenusmetode vir ewekansige regs-gesensoreerde data bespreek.

Hoofstuk 3 handel oor kernberaming. Verskeie verliesfunksies word voorgestel en bespreek. Ons verduidelik die “trade-off concept” met betrekking tot die keuse van die bandwydte h . Kernberaming word dan voorgestel. Ons bespreek eienskappe van kernfunksies en gee voorbeelde van kernberamers in verskeie kontekste.

Hoofstuk 4 verskaf 'n oorsig van die bandwydte selektors wat in die studie oorweeg word. Ons definieer relevant kernberamers van die uitvalsfunksie, en ons lei 'n eksplisiete uitdrukking af vir die skoenus gemiddelde kwadraat fout van die sogenaamde interne beramer in die geval van volledige data. Ons bespreek die kleinste-kwadrade kruisgeldigheid selektor soos beskryf in Patil (1993a) en Patil (1993b). Tweedens word die skoenus bandwydte selektor van Gonzalez-Manteiga et al. (1996) bespreek. 'n Bandwydte selektor wat gebruik maak van “bagging” in die konteks van kernberaming van die digtheidsfunksie, voorgestel deur Hall and Robinson (2009), word uitgebrei na die geval van kernberaming van die uitvalsfunksie. Laastens word 'n eenvoudige skoenus bandwydte selektor voorgestel.

Hoofstuk 5 handel oor die simulasiestudie. Ons verduidelik die struktuur van die empiriese studie en verskaf algoritmes vir die implementasie van die outomatiese bandwydte selektors wat bespreek word. Resultate van die empiriese studie word weergegee in die vorm van tabelle en grafieke. Gevolgtrekkings word gemaak rakende die prestasie van die bandwydte selektors, waarna ons sekere areas vir verdere studie voorstel en resultate van sekondere studies rapporteer.

Dankbetuigings

Ek wil hiermee graag die volgende bedankings doen:

- Vir die geleentheid en voorreg om die studie aan te pak en te voltooi, wil ek al die eer aan God bring. Dit was slegs deur U krag moontlik om hierdie studie te voltooi.
- Prof. F.C. van Graan, vir die voorstel van die onderwerp, u leiding, motivering, geduld en ondersteuning. Ek het ontsettend baie waardering vir al die laat nagte en uitdagende tye wat u saam met my deurgemaak het.
- My vrou, Leani, vir jou geduld, jou opbouende motivering en onvoorwaardelike liefde.
- Prof. J.W.H. Swanepoel en Prof. C.J. Swanepoel, vir blootstelling aan die vakgebied, inspirasie en waardevolle insette.
- Dr. L. Santana, jou hulpvaardigheid en geduld met al my vrae en versoeke word opreg waardeer.
- Dr. G. Koekemoer, Me. L. Boshoff, Dr. J.S. Allison en Mnr. W.D. Schutte. Ek het soveel by elkeen van julle geleer. Baie dankie vir jul beskikbaarheid en bereidwilligheid om te luister, te redeneer en waardevolle advies te bied.
- My familie en vriende vir volgehoue ondersteuning en belangstelling.

Symbols and Notation

Notation	Meaning
$k_h(x - y) = \frac{1}{h}k\left(\frac{x - y}{h}\right)$	Rescaled kernel k with bandwidth h in the point $(x - y)$
$I(A)$	Indicator function for the event A , where $I(A) = 1$ if the event A occurs, and $I(A) = 0$ if the event A does not occur
$F_n(x) = \frac{1}{n} \sum_{i=1}^n I(X_i \leq x)$	Empirical distribution function (EDF) in the point x
$F(x) = P(X \leq x)$	Distribution function F in the point x
$S(x) = P(X > x) = 1 - F(x)$	Survival function S in the point x
$f(x)$	Density function f in the point x
$\lambda(x) = \frac{f(x)}{S(x)} = \frac{f(x)}{1 - F(x)}$	Hazard function λ in the point x
$\hat{\lambda}_{n,h}(x)$	Kernel estimator of λ based on sample of size n and bandwidth h
$Bias(\hat{\lambda}_{n,h}(x)) = E[\hat{\lambda}_{n,h}(x)] - \lambda(x)$	Bias of the estimator $\hat{\lambda}_{n,h}$ in the point x
$Var(\hat{\lambda}_{n,h}(x)) = E\left[(\hat{\lambda}_{n,h}(x) - E[\hat{\lambda}_{n,h}(x)])^2\right]$	Variance of the estimator $\hat{\lambda}_{n,h}$ in the point x
$MSE(\hat{\lambda}_{n,h}(x)) = E\left[(\hat{\lambda}_{n,h}(x) - \lambda(x))^2\right]$	Mean squared error of the estimator $\hat{\lambda}_{n,h}$ in the point x
$w(x)$	Weight function w in the point x
$ISE_w(h) = \int_0^\infty [\hat{\lambda}_{n,h}(x) - \lambda(x)]^2 w(x) dx$	Weighted integrated squared error

Notation	Meaning
$MISE_w(h) = E[ISE_w(h)]$ $= E \left[\int_0^\infty [\hat{\lambda}_{n,h}(x) - \lambda(x)]^2 w(x) dx \right]$ argmin Q.E.D. i.i.d. iff or $\Leftrightarrow$ $\forall$ $\in$ lim $(f * g)(x)$	Weighted mean integrated squared error the argument that minimizes... “quod erat demonstrandum”; that which was to be demonstrated independent and identically distributed if and only if for all element of limit Convolution of f and g in the point x , $\int_{-\infty}^\infty f(x-y)g(y)dy$

Sequence notation
Let a_n and b_n be two real-valued deterministic sequences. $a_n = o(b_n)$ as $n \rightarrow \infty$ if and only if $\lim_{n \rightarrow \infty} a_n/b_n = 0$ $a_n = O(b_n)$ as $n \rightarrow \infty$ if and only if $\limsup_{n \rightarrow \infty} a_n/b_n < \infty$ $a_n \sim b_n$ if and only if $\lim_{n \rightarrow \infty} (a_n/b_n) = 1$ Let A_n and B_n be two real-valued random sequences. $A_n = o_p(B_n)$ if $\forall \epsilon > 0, \lim_{n \rightarrow \infty} P(A_n/B_n > \epsilon) = 0$ $A_n = O_p(B_n)$ if $\forall \epsilon > 0$, there exist $\delta > 0$ and M such that $\lim_{n \rightarrow \infty} P(A_n/B_n > \delta) < \epsilon \forall n > M$

Contents

1	Introduction	1
1.1	Survival analysis	1
1.2	Hazard rate function	2
1.3	Lifetime distributions	3
1.3.1	Constant hazard function	3
1.3.2	Monotonic hazard function	4
1.3.3	Unimodal hazard function	5
1.3.4	Bathtub hazard function	5
1.4	Censoring	7
1.4.1	Types of censoring	7
1.4.2	The empirical distribution function F_n	11
1.4.3	The Kaplan-Meier estimator	12
2	Bootstrap methodology	17
2.1	Introduction	17
2.2	The naive bootstrap procedure	18
2.2.1	The plug-in principle	18
2.2.2	Bootstrap estimate of bias	20
2.2.3	Bootstrap estimate of standard error	21
2.3	The smooth bootstrap	21
2.3.1	Motivation for using the smooth bootstrap	23
2.4	Bootstrapping censored data	25
3	Kernel smoothing	28
3.1	Discrepancy measures	28
3.1.1	Pointwise discrepancy measures	28
3.1.2	Global discrepancy measures	30

3.1.3	Integrated squared error	31
3.1.4	Mean integrated squared error	31
3.1.5	Asymptotic mean integrated squared error (AMISE)	31
3.1.6	The smoothing parameter h and the trade-off concept	33
3.2	Different estimation methods	34
3.2.1	Parametric estimation	34
3.2.2	Nonparametric estimation	35
3.3	Kernel smoothing	36
3.3.1	Introduction and motivation	36
3.3.2	Kernel functions	36
3.3.3	The concept of kernel smoothing	37
4	Smoothers and automatic bandwidth selectors	46
4.1	Introduction	46
4.2	Uncensored case	46
4.2.1	External estimator	46
4.2.2	Internal estimator	47
4.2.3	An explicit expression for bootstrap MSE in the case of $\hat{\lambda}_{n,h}^{*(2)}$	49
4.3	Random right censoring	55
4.3.1	External estimator	56
4.3.2	Internal estimator	56
4.4	Automatic bandwidth selectors	57
4.4.1	Least-squares cross validation (LSCV)	58
4.4.2	Bagged cross validation	60
4.4.3	Bootstrap AMISE_w^*	62
4.4.4	Bootstrap approximation to MISE_w	66
5	Simulation study	70
5.1	Structure of simulation study	70
5.1.1	Details of settings	70
5.1.2	Layout	72
5.1.3	Secondary studies	73
5.2	Algorithms	73
5.2.1	Notation	73

5.2.2	Random number generator	74
5.2.3	Generating censored data	75
5.2.4	Theoretical optimal bandwidth $\hat{h}_{MISE_w}$	77
5.2.5	LSCV selector $\hat{h}_{CV}$	78
5.2.6	Bagged LSCV selectors $\hat{h}_{CV,B_1}$ and $\hat{h}_{CV,B_2}$	79
5.2.7	Bootstrap $AMISE_w^*$ bandwidth selector $\hat{h}^*$	80
5.2.8	Bootstrap approximation to $MISE_w$ selector $\hat{h}_P^*$	81
5.3	Ranking methods	83
5.4	Simulation study results	85
5.4.1	Theoretical optimal h	85
5.4.2	Automatic bandwidth selectors	85
5.5	Conclusions	113
5.5.1	Main simulation study	113
5.5.2	Secondary studies	116

Bibliography	123
---------------------	------------

Chapter 1

Introduction

1.1 Survival analysis

Survival analysis, also known as Reliability analysis, is a field of Statistics that studies lifetime distributions, or in other words time-to-event data. Survival data arise in numerous fields, for example engineering, medicine, economics and so forth. An engineer might be interested in the failure times of machine parts. Manufacturing managers will study the times between breakdowns in the production line. A logistics clerk will note the times or distances between breakdowns of trucks. In medical studies the time it takes before treatment takes effect on a patient can be monitored. In an economic climate sales representatives can measure the time it takes between price decreases and significant boosts in sales, or economists can measure times between economic recessions or changes in the business cycle.

In almost any field or industry one can identify instances of where survival data can be measured. In order to understand where the hazard rate function, which is the main quantity of interest in this dissertation, fits into the framework of survival analysis, we need to define some basic concepts. The following two definitions are due to Smith (2002).

Definition 1.1. *A random variable X is called a survival random variable if an observed outcome x of X lies in the interval $[0, \infty)$.*

It is important here to note that the random variables under study are always non-negative, as a consequence of working with lifetimes.

Definition 1.2. *The survival function (or reliability function), S , is defined for all values of x by $S(x) = 1 - F(x)$, where $F(x) = P(X \leq x) = \int_0^x f(y)dy$ is the distribution function of the random variable X in the point x , and f is the probability density function of X in the point x .*

By considering the above definition we note that the survival function can be written as follows: $S(x) = P(X > x) = \int_x^\infty f(y)dy$, which clearly supports the interpretation of the survival function as the probability of surviving beyond a given time. We have now defined the necessary concepts in order to define and understand a very important quantity in survival analysis, namely the hazard rate function.

1.2 Hazard rate function

The hazard rate function (henceforth referred to as the hazard function) is defined as follows:

Definition 1.3. *Let X be a continuous survival random variable, and let Δx denote an increment in the value of x . Then*

$$\begin{aligned}\lambda(x) &= \lim_{\Delta x \rightarrow 0} \frac{P(x \leq X < x + \Delta x | X \geq x)}{\Delta x} \\ &= \lim_{\Delta x \rightarrow 0} \frac{P(x \leq X < x + \Delta x)}{\Delta x P(X \geq x)} \\ &= \frac{f(x)}{S(x)} \\ &= \frac{f(x)}{1 - F(x)}.\end{aligned}\tag{1.1}$$

The only restriction on the hazard function is that $\lambda(x) \geq 0$. The hazard function is also known as (among others) the age-specific failure rate, the force of mortality and the conditional failure rate, depending on the situation at hand.

By considering the definition, it is seen that the hazard function has a probabilistic interpretation. Note the following:

$$\begin{aligned}\lambda(x)\Delta x &\approx P(x \leq X < x + \Delta x | X \geq x) \\ &= P(\text{Failure in } [x, x + \Delta x) \mid \text{Survival past } x).\end{aligned}$$

Similarly, $\lambda(x)\Delta x$ can also be interpreted as the instantaneous failure rate (risk) for a device (individual) still functioning (alive) at time x . An associated function can be obtained by integrating $\lambda(x) = -\frac{d}{dx}(\log S(x))$ and using $S(0) = 1$ to obtain $S(x) = e^{-\int_0^x \lambda(y)dy} = e^{-\Lambda(x)}$, where

$$\Lambda(x) = \int_0^x \lambda(y)dy\tag{1.2}$$

is the cumulative hazard function. We have $f(x) = \lambda(x)e^{-\Lambda(x)}$, that is the hazard function uniquely determines the density, as well as $\int_0^\infty \lambda(y)dy = +\infty$. For additional properties of

the hazard function see Watson and Leadbetter (1964a).

The hazard function is extremely useful as it describes the way in which the likelihood of experiencing a failure changes with time. In a sense it is easier to interpret the shape of lifetime distributions by studying the hazard function (as opposed to studying the survival function or distribution function).

1.3 Lifetime distributions

As one would expect, there are many possible shapes that the hazard function may assume. Some of the common shapes include constant, monotone decreasing, monotone increasing and unimodal, as well as the so-called “bathtub” shape. This section will discuss some known distributions that assume these shapes.

1.3.1 Constant hazard function

In this section a distribution with a constant hazard function will be considered.

Exponential

Denote the exponential distribution by $[\text{Exp}(\alpha), \alpha > 0]$. The density, distribution and hazard functions are respectively given by

$$f(x) = \alpha e^{-\alpha x}, \quad (1.3)$$

$$F(x) = 1 - e^{-\alpha x}, \quad (1.4)$$

$$\text{and } \lambda(x) = \alpha. \quad (1.5)$$

An important property of the exponential distribution is the fact that the hazard function is constant, and thus does not depend on the value of x . This is referred to as the *memoryless* property of the exponential distribution, and this fact characterizes the distribution.

As a simple example, suppose the lifetime of a light bulb follows an exponential distribution. The previous property then implies that the probability of the light bulb functioning for another hour given that it has been functioning for 100 hours, is approximately the same as the probability that the light bulb will function for another hour given it has been functioning for 1 000 hours.

Clearly this is not a useful model for studying lifetimes in general, but a practical example of the use of the exponential density is the modeling of times between occurrences of earthquakes.

1.3.2 Monotonic hazard function

In this section distributions with monotonic hazard functions will be considered.

Weibull

Denote the Weibull distribution by $[\text{Wei}(\alpha, \beta), \alpha > 0, \beta > 0]$. The density, distribution and hazard functions are respectively given by

$$f(x) = \alpha\beta(\beta x)^{\alpha-1}e^{-(\beta x)^\alpha}, \quad (1.6)$$

$$F(x) = 1 - e^{-(\beta x)^\alpha}, \quad (1.7)$$

$$\text{and } \lambda(x) = \alpha\beta(\beta x)^{\alpha-1}. \quad (1.8)$$

The parameter α dictates the shape of the hazard function, whereas the parameter β dictates the location. The exponential distribution is a special case of the Weibull distribution when $\alpha = 1$. The Weibull distribution is more flexible than the exponential distribution as it can model monotone increasing hazard functions ($\alpha > 1$) as well as monotone decreasing hazard functions ($\alpha < 1$).

A typical example of a survival variable with an increasing hazard function is the lifetimes of car tires, where the tendency of failure is initially very low, but steadily increases as the tires wear out. In terms of a decreasing hazard function, a common example is the failure rate of infants of any wild animal. The hazard function of a young deer for instance is extremely high, as it cannot defend itself and evade predators, but as the young deer grows up, it becomes more resilient and the hazard function decreases.

Rayleigh

Denote the Rayleigh distribution by $[\text{Ray}(\alpha, \beta), \alpha > 0]$. The density, distribution and hazard functions are respectively given by

$$f(x) = (\alpha + \beta x)e^{-\alpha x + \frac{1}{2}\beta x^2}, \quad (1.9)$$

$$F(x) = 1 - e^{-\alpha x + \frac{1}{2}\beta x^2}, \quad (1.10)$$

$$\text{and } \lambda(x) = \alpha + \beta x. \quad (1.11)$$

The Rayleigh distribution can also model both monotone increasing ($\beta > 0$) as well as monotone decreasing ($\beta < 0$) hazard functions, and reduces to the exponential distribution when $\beta = 0$.

Gamma

Denote the gamma distribution by $[\text{Gam}(\alpha, \beta), \alpha > 0, \beta > 0]$. The density is given by

$$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}. \quad (1.12)$$

No closed form expressions exist for F and λ . Similar to the other distributions in this section, monotone increasing ($\alpha > 1$) hazard functions as well as monotone decreasing ($\alpha < 1$) hazard functions can be obtained.

1.3.3 Unimodal hazard function

In this section a distribution with a unimodal hazard function will be considered.

Lognormal

Denote the lognormal distribution by $[\text{LN}(\mu, \sigma), \sigma > 0]$. The density, distribution and hazard functions are respectively given by

$$f(x) = \frac{1}{x\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(\ln(x)-\mu)^2}, \quad (1.13)$$

$$F(x) = \Phi\left(\frac{\ln(x) - \mu}{\sigma}\right), \quad (1.14)$$

$$\text{and } \lambda(x) = \frac{\frac{1}{x\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(\ln(x)-\mu)^2}}{1 - \Phi\left(\frac{\ln(x)-\mu}{\sigma}\right)}. \quad (1.15)$$

where Φ is the standard normal distribution function, defined as

$$\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} dt, \quad -\infty < x < \infty. \quad (1.16)$$

The shape of the hazard function for this distribution is initially increasing, then it will flatten out and reach a peak, and finally it decreases monotonically in the right tail. This shape is sometimes referred to as an “upside-down bathtub” shape.

As an example one might consider the lifespan of typical household appliances. Appliances tend to have a low tendency of failure initially, but this failure rate then increases to a peak where most appliances typically wear out, after which the failure rate decreases monotonically as some appliances continue to function properly for much longer than expected.

1.3.4 Bathtub hazard function

In this section a distribution that has a bathtub hazard function will be considered.

Reduced additive Weibull

Denote the so-called reduced additive Weibull distribution, as introduced by Xie and Lai (1995) with $[\text{RAW}(\alpha, \beta), \alpha > 0, \beta > 1]$. The density, distribution and hazard functions are respectively given by

$$\begin{aligned} f(x) &= (\alpha\beta(\alpha x)^{\beta-1} + \frac{\alpha}{\beta}(\alpha x)^{\frac{1}{\beta}-1})e^{-(\alpha x)^\beta - (\alpha x)^{\frac{1}{\beta}}}, \\ F(x) &= 1 - e^{-(\alpha x)^\beta - (\alpha x)^{\frac{1}{\beta}}}, \\ \text{and } \lambda(x) &= \alpha\beta(\alpha x)^{\beta-1} + \frac{\alpha}{\beta}(\alpha x)^{\frac{1}{\beta}-1}. \end{aligned}$$

As mentioned previously, the Weibull distribution can model either increasing or decreasing hazard functions, depending on the choice of α . The idea of a distribution like this is to combine both shapes into one distribution, leading to a hazard function that initially decreases monotonically, then it will flatten out and have a turning point, after which it increases monotonically. This shape reminds one of the side-on view of a bathtub, and these distributions are commonly referred to as distributions with bathtub hazard functions.

Human lifetimes are typically modeled using distributions of this kind. In certain populations human lifetimes exhibit high infant mortality rates, after which the hazard function steadily decreases, and eventually turns upward and increases as high mortality rates occur as a result of old age.

Table 1.1 summarizes the different shapes of the hazard function and lists the distributions that represent each shape.

Constant	Increasing	Decreasing	Unimodal	Bathtub
Exp	Wei ($\alpha > 1$)	Wei ($\alpha < 1$)	LN	RAW
	Ray ($\beta > 0$)	Ray ($\beta < 0$)		
	Gam ($\alpha > 1$)	Gam ($\alpha < 1$)		

Table 1.1: *Different hazard function shapes.*

Figure 1.1 illustrates the different shapes.

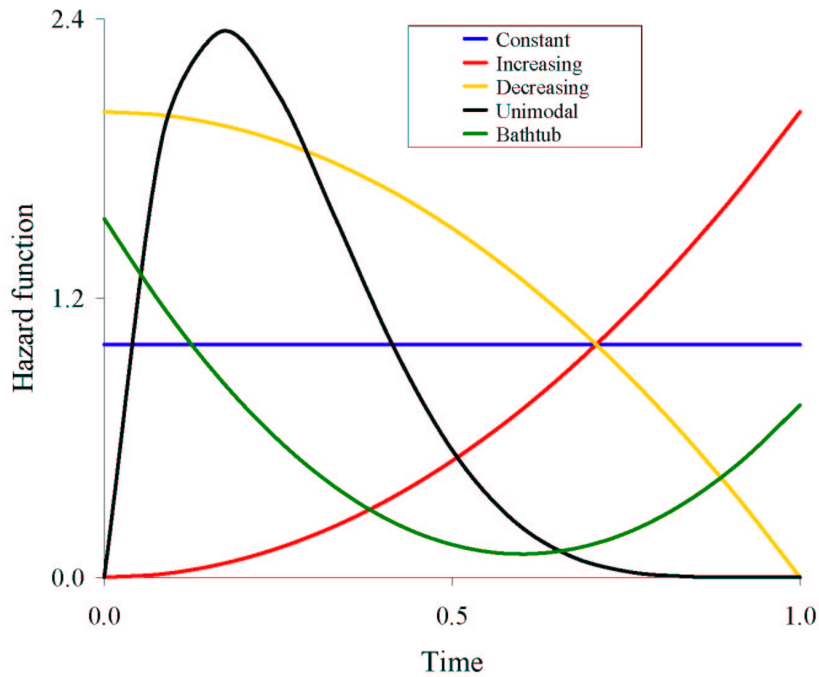


Figure 1.1: *Different shapes of the hazard function.*

1.4 Censoring

1.4.1 Types of censoring

For more details regarding the concepts defined in this section, the reader can consult Act (2005), Klein and Moeschberger (1997) and Smith (2002). Censoring is a key feature of survival analysis, and the mechanisms that give rise to censoring play an important part in statistical inference. In this section a number of different, but not mutually exclusive, types of censoring will be introduced. An important assumption to note beforehand is that the lifetimes are independent of the censoring times. Censoring occurs when a lifetime of interest is not observed completely, but for some reason the investigator only has limited information regarding the specific lifetime. Different types of censoring include the following:

Type I censoring

If the censoring time is known *in advance*, the censoring mechanism is known as Type I censoring. As an example, suppose an investigator is measuring remuneration tendencies among

professionals up to their retirement. The investigator will stop following the individuals once they reach a predetermined retirement age, for instance 65 years of age. Formally:

Definition 1.4. Denote by $X_1^0, X_2^0, \dots, X_n^0$ the actual survival times, and let t_c denote a predetermined censoring time. The observed sample $X_1, X_2, \dots, X_n$ is constructed as follows (for $i = 1, 2, \dots, n$):

$$X_i = \begin{cases} X_i^0 & \text{if } X_i^0 \leq t_c \\ t_c & \text{if } X_i^0 > t_c. \end{cases}$$

Figure 1.2 illustrates this graphically.

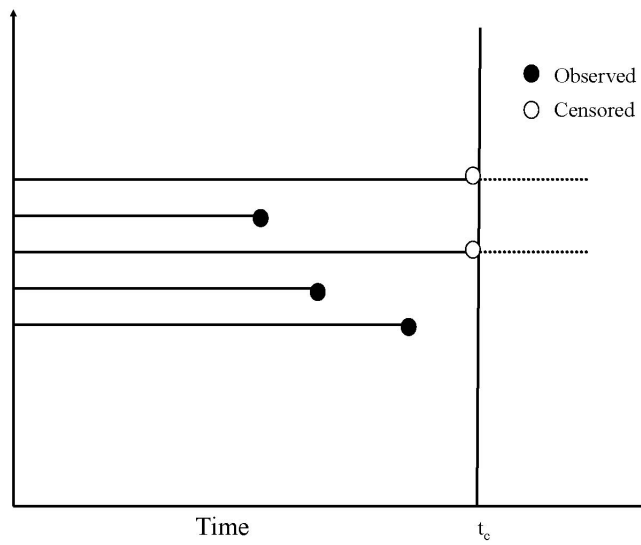


Figure 1.2: Type 1 censoring with 3 observed lifetimes and 2 censored lifetimes.

Type II censoring

In some instances an investigator will have a predetermined number of cases that need to be observed. In a medical study 1 000 rats might be subjected to a hazardous material, and the trial will end as soon as the times of the first 200 fatalities are recorded. Formally:

Definition 1.5. Let $r < n$ be a fixed number. Denote by $X_{(1)}^0, X_{(2)}^0, \dots, X_{(n)}^0$ the ordered survival variables. The observed sample $X_{(1)}, X_{(2)}, \dots, X_{(r)}$ is constructed as follows:

$$X_{(i)} = X_{(i)}^0, \quad i = 1, 2, \dots, r.$$

Note that this means that the number of observed values is non-random, since observation stops as soon as the predetermined number of observations (r) have been recorded. This also

implies that the observed data consists of the r smallest observations. Figure 1.3 illustrates this graphically.

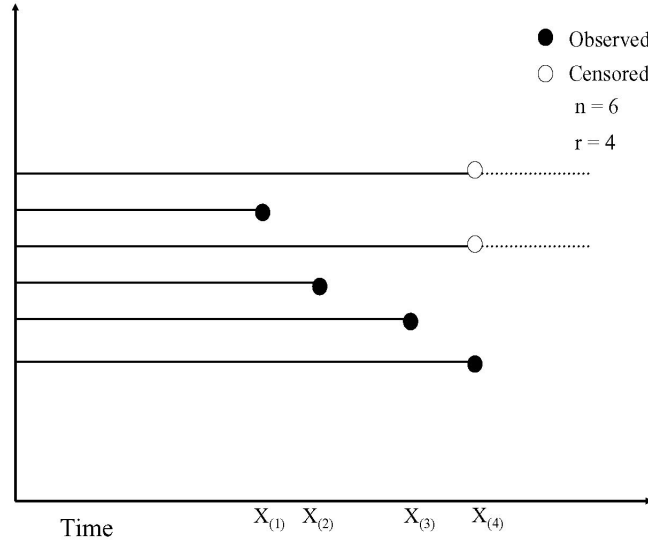


Figure 1.3: *Type II censoring with 4 observed lifetimes and 2 censored lifetimes.*

Random right censoring

Random right censoring is the censoring mechanism that has received the most attention in the literature, since it frequently occurs in practice. Data are randomly right censored if the censoring mechanism cuts short observations in progress, in other words a lifetime is only observed up to a certain point. This results in the investigator only knowing that the true, unobserved lifetime exceeds the observed censoring time. Random right censoring can be formally defined as follows:

Definition 1.6. *The survival variables $X_1^0, X_2^0, \dots, X_n^0$ are randomly right censored by the random variables $Z_1, Z_2, \dots, Z_n$, if the observed sample consists of the ordered pairs (X_i, Δ_i) , for $i = 1, 2, \dots, n$, where for each i : $X_i = \min(X_i^0, Z_i)$,*

$$\Delta_i = \begin{cases} 1 & \text{if } X_i^0 \leq Z_i \\ 0 & \text{if } X_i^0 > Z_i, \end{cases}$$

where Z_i is the random censor time, $\Delta_i = I(X_i^0 \leq Z_i)$ is the censor indicator for X_i and $i = 1, 2, \dots, n$.

Note that Type I censoring is a special (or degenerate) case of random right censoring, where $Z_i = t_c$, $i = 1, 2, \dots, n$. See Figure 1.4 for a graphical representation of random right

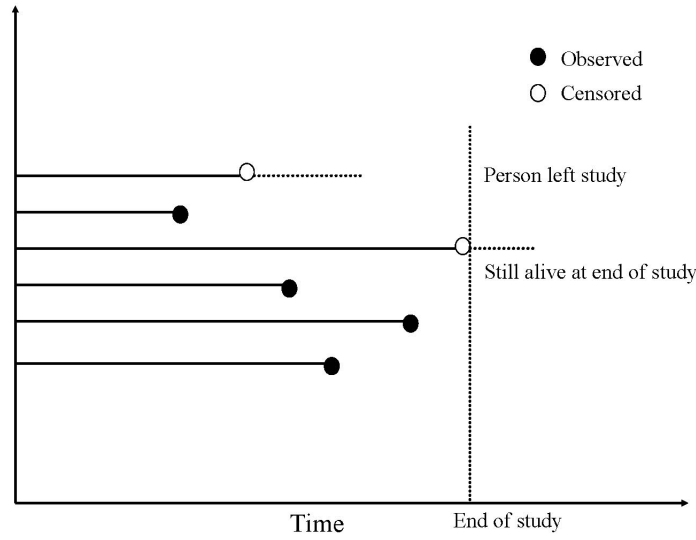


Figure 1.4: *Random right censoring with 4 observed lifetimes and 2 censored lifetimes.*

censoring. Since the emphasis will be on random right censored data in the dissertation, the following important remarks need to be noted:

Remarks:

1. $\{X_i^0\}_{i=1}^n$ denotes the true survival times of the observed sample of size n . These survival times are assumed to be independent random variables from a common (or identical), but unknown distribution function F^0 with density f^0 .
2. $\{Z_i\}_{i=1}^n$ denotes the censoring random variables of the observed sample of size n .
3. In the random censorship model $\{Z_i\}_{i=1}^n$ are assumed to be an i.i.d. random sample drawn *independently* of the $\{X_i^0\}_{i=1}^n$, from an unknown distribution H^0 with density h^0 .
4. The observed sample $\{X_i\}_{i=1}^n$ is said to be right censored, since one observes only the pair $X_i = \min(X_i^0, Z_i)$ and $\Delta_i = I(X_i = X_i^0)$, defined as

$$\Delta_i = \begin{cases} 1, & \text{if } X_i = X_i^0 \\ 0, & \text{if } X_i = Z_i, \end{cases}$$

for $i = 1, 2, \dots, n$.

5. The observed $\{X_i\}_{i=1}^n$ values constitute a random sample from the distribution function F with density f .

6. The investigator is interested in the unknown distribution of the true survival times, that is F^0 .

Random left censoring

Left censoring occurs when the censoring mechanism prevents the investigator from knowing when the state of interest was entered into. As an example, consider a doctor studying the duration of a disease among patients. When a patient is treated by the doctor for the disease, the doctor only knows that the patient has already contracted the disease, but the precise time when the disease was contracted is unknown, leading to the disease duration time being left censored. For left censored data Smith (2002) defines the observed lifetimes as $X_i = \max(X_i^0, Y_i)$, where Y_i is the (left) censor time associated with X_i^0 . It follows that

$$-X_i = \min(-X_i^0, -Y_i),$$

which illustrates that left censoring can be defined as a special case of right censoring where the time axis has been reversed. For this reason, most techniques have been developed for right censored data. See Figure 1.5 for an illustration of random left censoring.

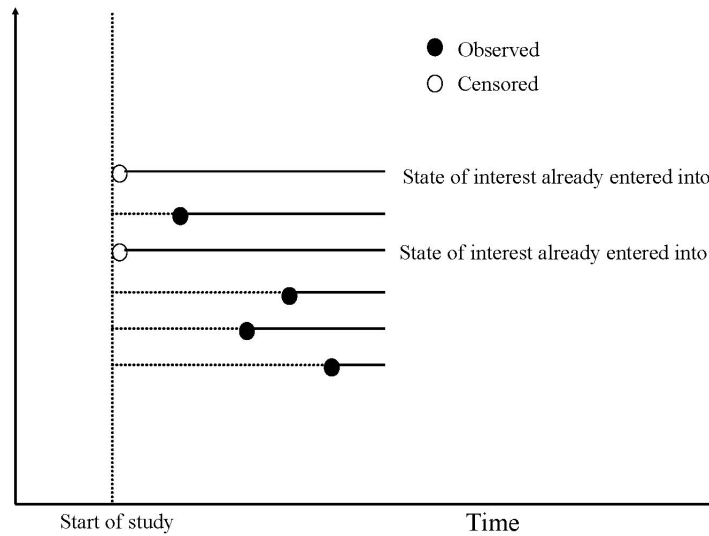


Figure 1.5: *Random left censoring with 4 observed lifetimes and 2 censored lifetimes.*

1.4.2 The empirical distribution function F_n

At this point we will introduce a function known as the empirical distribution function (or EDF). This function plays an important part in bootstrap methodology as discussed in Chap-

ter 2, and is also the basis for the Kaplan-Meier function that will be discussed in the next section. The EDF, denoted by F_n , is a discrete approximation of the distribution function F constructed entirely from the observed sample, or empirical data. Let $X_1, X_2, \dots, X_n$ be a random sample from a continuous distribution function F . The EDF is defined as

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n I(X_i \leq x), \quad (1.17)$$

where I is the indicator function and is defined as:

$$I(A) = \begin{cases} 1, & \text{if the event } A \text{ occurs} \\ 0, & \text{if the event } A \text{ does not occur.} \end{cases}$$

F_n assigns an equal probability of $\frac{1}{n}$ to each of the observed values $X_1, X_2, \dots, X_n$, and is a step function. It can easily be shown that, for any event A we have the following equality:

$$P(X \in A) = E\{I(X \in A)\}.$$

That is, the probability of an event A occurring, is equal to the expected value of the indicator function of that event.

Remarks:

1. $E\{F_n(x)\} = F(x)$, that is F_n is an *unbiased* estimator of F .
2. $Var\{F_n(x)\} = \frac{1}{n} \left[F(x)(1 - F(x)) \right]$.
3. The *Glivenko-Cantelli* Lemma states that $P\left(\limsup_{n \rightarrow \infty} |F_n(x) - F(x)| = 0\right) = 1$. Therefore, as n increases to infinity, the probability that F_n converges to F is 1. That is, F_n converges *almost-surely (a.s)* to F as n tends to infinity.

1.4.3 The Kaplan-Meier estimator

Kaplan and Meier (1958) generalized the EDF to the censored case. In the next paragraph the Kaplan-Meier estimator $\hat{F}_n^0$ of the true distribution function F^0 in the case where no ties are assumed will be derived.

Derivation of Kaplan-Meier estimator

The following derivation is in the setting of random right censoring. Before we begin the reader must note that any censoring mechanism can be classified as *informative* or *non informative*. If censoring is non informative, the censoring does not give any information

about the true survival variables X^0 . This implies that the mortality of the lives that remain in the at-risk group is the same as the mortality of the lives that have been censored. In this derivation the censoring mechanism is assumed to be non informative. Also, since the observations $\{X_i, \Delta_i\}_{i=1}^n$ are assumed to be i.i.d, it suffices to calculate the likelihood of a single point (X_i, Δ_i) .

If $\Delta = 1$, we have

$$\begin{aligned} \lim_{\epsilon \rightarrow 0} \frac{P(x - \epsilon < X_i < x + \epsilon, \Delta = 1)}{2\epsilon} &= \lim_{\epsilon \rightarrow 0} \frac{P(x - \epsilon < X_i^0 < x + \epsilon, X_i^0 \leq Z_i)}{2\epsilon} \\ &= \lim_{\epsilon \rightarrow 0} \frac{1}{2\epsilon} \int_{x-\epsilon}^{x+\epsilon} \int_y^{\infty} dH^0(z) dF^0(y) \quad (\text{by independence}) \\ &= \lim_{\epsilon \rightarrow 0} \frac{1}{2\epsilon} \int_{x-\epsilon}^{x+\epsilon} (1 - H^0(y)) dF^0(y) \\ &= (1 - H^0(x)) f^0(x). \end{aligned}$$

If $\Delta = 0$, we have

$$\begin{aligned} \lim_{\epsilon \rightarrow 0} \frac{P(x - \epsilon < X_i < x + \epsilon, \Delta = 0)}{2\epsilon} &= \lim_{\epsilon \rightarrow 0} \frac{P(x - \epsilon < Z_i < x + \epsilon, X_i^0 > Z_i)}{2\epsilon} \\ &= \lim_{\epsilon \rightarrow 0} \frac{1}{2\epsilon} \int_z^{\infty} \int_{x-\epsilon}^{x+\epsilon} dH^0(z) dF^0(y) \quad (\text{by independence}) \\ &= \lim_{\epsilon \rightarrow 0} \frac{1}{2\epsilon} \int_{x-\epsilon}^{x+\epsilon} (1 - F^0(z)) dH^0(z) \\ &= (1 - F^0(x)) h^0(x). \end{aligned}$$

Combining these results we arrive at the contribution to the likelihood of the i^{th} observation (X_i, Δ_i) , that is

$$[(1 - H^0(X_i)) f^0(X_i)]^{\Delta_i} [(1 - F^0(X_i)) h^0(X_i)]^{1-\Delta_i}.$$

Mathematically, the concept of non informative censoring means that the censoring distribution H^0 and its associated density h^0 do not depend on the parameters of interest, and can therefore be removed from the likelihood function. So by removing the factors $(1 - H^0(X_i))^{\Delta_i}$ and $h^0(X_i)^{1-\Delta_i}$ we arrive at the likelihood function

$$\prod_{i=1}^n [1 - F^0(X_i)]^{1-\Delta_i} f^0(X_i)^{\Delta_i} = \prod_{\substack{i=1 \\ \Delta_i=1}}^n f^0(X_i) \prod_{\substack{i=1 \\ \Delta_i=0}}^n [1 - F^0(X_i)], \quad (1.18)$$

which can be generalized to the following likelihood function:

$$L = \prod_{\substack{i=1 \\ \Delta_i=1}}^n [F^0(X_i) - F^0(X_i-)] \prod_{\substack{i=1 \\ \Delta_i=0}}^n [1 - F^0(X_i)], \quad (1.19)$$

where X_i- denotes a value just smaller than X_i .

Now that we have the likelihood function, we look for a discrete distribution that maximizes the likelihood function L . What we need is to find a step function $\hat{F}_n^0$ that

1. Maximizes the jump $\hat{F}_n^0(X_i) - \hat{F}_n^0(X_i-)$ at X_i where $\Delta_i = 1$
2. Minimizes $\hat{F}_n^0(X_i)$ at X_i where $\Delta_i = 0$.

This results in a function which jumps only at the X_i where $\Delta_i = 1$, that is the uncensored observations. So we know that $\hat{F}_n^0$ has the form

$$1 - \hat{F}_n^0(x) = \prod_{\substack{i=1 \\ X_{(i)} \leq x \\ \Delta_i = 1}}^n (1 - \hat{\lambda}_i), \quad (1.20)$$

for some $\hat{\lambda}_1, \hat{\lambda}_2, \dots, \hat{\lambda}_n$ that is to be determined. Let $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ denote the order statistics of $X_1, X_2, \dots, X_n$ and let $\Delta_{(1)}, \Delta_{(2)}, \dots, \Delta_{(n)}$ denote the associated indicator values. Recall the first product term in the likelihood function as given in (1.19). Introducing the notation described above, we have that

$$\begin{aligned} 1 - \hat{F}_n^0(X_i) &= \prod_{\substack{j=1 \\ \Delta_{(j)}=1}}^i (1 - \hat{\lambda}_j); \\ 1 - \hat{F}_n^0(X_i-) &= \prod_{\substack{j=1 \\ \Delta_{(j)}=1}}^{i-1} (1 - \hat{\lambda}_j) \end{aligned}$$

Note that $\hat{F}_n^0(X_i) - \hat{F}_n^0(X_i-) = 1 - \hat{F}_n^0(X_i-) - [1 - \hat{F}_n^0(X_i)]$, which implies

$$\begin{aligned} \hat{F}_n^0(X_i) - \hat{F}_n^0(X_i-) &= \prod_{\substack{j=1 \\ \Delta_{(j)}=1}}^{i-1} (1 - \hat{\lambda}_j) - \prod_{\substack{j=1 \\ \Delta_{(j)}=1}}^i (1 - \hat{\lambda}_j) \\ &= \prod_{\substack{j=1 \\ \Delta_{(j)}=1}}^{i-1} (1 - \hat{\lambda}_j) [1 - (1 - \hat{\lambda}_i)] \\ &= \prod_{\substack{j=1 \\ \Delta_{(j)}=1}}^{i-1} (1 - \hat{\lambda}_j) \hat{\lambda}_i. \end{aligned}$$

We can now replace these equalities into the likelihood function in (1.19) and proceed to develop the log-likelihood function.

$$\log L = \sum_{\substack{i=1 \\ \Delta_{(i)}=1}}^n \log \left[\prod_{\substack{j=1 \\ \Delta_{(j)}=1}}^{i-1} (1 - \hat{\lambda}_j) \hat{\lambda}_i \right] + \sum_{\substack{i=1 \\ \Delta_{(i)}=0}}^n \log \left[\prod_{\substack{j=1 \\ \Delta_{(j)}=1}}^i (1 - \hat{\lambda}_j) \right]$$

$$\begin{aligned}
&= \sum_{\substack{i=1 \\ \Delta_{(i)}=1}}^n \left[\sum_{\substack{j=1 \\ \Delta_{(j)}=1}}^{i-1} \log(1 - \hat{\lambda}_j) + \log \hat{\lambda}_i \right] + \sum_{\substack{i=1 \\ \Delta_{(i)}=0}}^n \sum_{\substack{j=1 \\ \Delta_{(j)}=1}}^i \log(1 - \hat{\lambda}_j) \\
&= \sum_{\substack{j=1 \\ \Delta_{(j)}=1}}^n [(n-j) \log(1 - \hat{\lambda}_j) + \log(\hat{\lambda}_j)].
\end{aligned}$$

Now, in order to maximise the log-likelihood we need to calculate its derivative to $\hat{\lambda}_j$, $j = 1, 2, \dots, n$ with $\Delta_{(j)} = 1$. This means we have to solve

$$\frac{\partial}{\partial \hat{\lambda}_j} \log L = 0,$$

for $j = 1, 2, \dots, n$ with $\Delta_{(j)} = 1$. This gives

$$-\frac{n-j}{1 - \hat{\lambda}_j} + \frac{1}{\hat{\lambda}_j} = 0,$$

or

$$\hat{\lambda}_j = \frac{1}{n-j+1},$$

for $j = 1, 2, \dots, n$ with $\Delta_{(j)} = 1$. Substituting this back into (1.20) results in the Kaplan-Meier estimator

$$1 - \hat{F}_n^0(x) = \prod_{\substack{i=1 \\ X_{(i)} \leq x}}^n \left(\frac{n-i}{n-i+1} \right)^{\Delta_{(i)}}, \quad (x \leq X_{(n)}).$$

From this definition we see that the Kaplan-Meier estimator for F^0 only makes jumps at the uncensored observations, and the size of the jumps is random and also increasing. Similarly, the Kaplan-Meier estimator for H^0 only makes jumps at the censored observations, and the size of the jumps is also random and increasing.

We end off this section with complete definitions of both $\hat{F}_n^0$ and $\hat{H}_n^0$, and we also show that $\hat{F}_n^0$ reduces to F_n when there is no censoring.

$$1 - \hat{F}_n^0(x) = \begin{cases} 1, & \text{if } x < X_{(1)}, \\ \prod_{j=1}^{i-1} \left(\frac{n-j}{n-j+1} \right)^{\Delta_j} & \text{if } X_{(i-1)} \leq x < X_{(i)}, \\ 0, & \text{if } x \geq X_{(n)}. \end{cases}$$

$\hat{F}_n^0$ is also a discrete approximation to F^0 (i.e. it is a step function), and is right continuous. Analogously, $\hat{H}_n^0$ is the Kaplan-Meier estimator of the censoring distribution H^0 and is defined by

$$1 - \hat{H}_n^0(x) = \begin{cases} 1, & \text{if } x < X_{(1)}, \\ \prod_{j=1}^{i-1} \left(\frac{n-j}{n-j+1} \right)^{1-\Delta_j} & \text{if } X_{(i-1)} \leq x < X_{(i)}, \\ 0, & \text{if } x \geq X_{(n)}. \end{cases}$$

Note that the only difference between $\hat{F}_n^0$ and $\hat{H}_n^0$ is in the exponent term, where the Δ_j term in $\hat{F}_n^0$ is replaced by $1 - \Delta_j$, which leads to $\hat{H}_n^0$. We will now show how $\hat{F}_n^0$ reduces to F_n when there is no censoring (that is, when given $\Delta_i = 1$ for $i = 1, 2, \dots, n$).

$$\begin{aligned}
1 - \hat{F}_n^0(x) &= \prod_{j=1}^{i-1} \left(\frac{n-j}{n-j+1} \right)^{\Delta_j} \quad \text{if } X_{(i-1)} \leq x < X_{(i)} \\
&= \prod_{j=1}^{i-1} \left(\frac{n-j}{n-j+1} \right) \quad (\text{given no censoring}) \\
&= \frac{(n-1)(n-2)}{n(n-1)} \cdots \frac{(n-i+2)(n-i+1)}{(n-i+3)(n-i+2)} \\
&= \frac{(n-i+1)}{n} \quad (\text{all the other terms cancel out}) \\
&= 1 - \frac{i-1}{n}. \\
\text{Therefore } \hat{F}_n^0(x) &= \frac{i-1}{n} \quad \text{for } X_{(i-1)} \leq x < X_{(i)} \\
&= \frac{1}{n} \sum_{i=1}^n I(X_i \leq x) \\
&= F_n(x).
\end{aligned}$$

Chapter 2

Bootstrap methodology

2.1 Introduction

The bootstrap was introduced by Bradley Efron in 1979, and is described by Efron and Tibshirani (1993) as a computer-based method for assigning measures of accuracy to statistical estimates. The bootstrap is a data-based simulation method used for statistical inference. It can be implemented to assess (among others) various discrepancy measures, like bias terms, standard error terms (and consequently variance terms) and confidence intervals. For a thorough discussion on the bootstrap and its applications, the interested reader can consult Efron and Tibshirani (1993) or Davison and Hinkley (1997).

The advantages of applying the bootstrap (and to applying nonparametric methods in general) is that no assumptions regarding the underlying distribution that generated the data at hand are required, and the method can be applied irrespective of the sample size. The method can also be applied to (almost) any statistic, no matter how complicated it might be, as long as it can be calculated from the sample data.

The scenario just mentioned is an application of the bootstrap in a nonparametric way. The bootstrap can also be applied parametrically, in which case an assumption is made regarding the distribution that generated the data.

Some statistics (for example the sample mean) have large sample theory to assist in the inferences drawn from it, but there are many statistics that do not have the same solid theoretical background, and rely more heavily on assumptions or large sample sizes. The bootstrap is a very useful tool in determining distributional characteristics of these types of statistics.

2.2 The naive bootstrap procedure

In this section, denote by F the probability distribution of a random variable X . Now, for the purposes of this dissertation, suppose that the investigator has observed a random sample of size n from this unknown probability distribution F , resulting in $\mathbf{X} = (X_1, X_2, \dots, X_n)$. Essentially, the idea of the bootstrap is to shift from an environment where there are many unknown parameters (like the distribution function F and the discrepancy measures of any estimators of the parameters) to an environment where everything is known to the investigator, or can be estimated by resampling from the observed sample.

This new environment is commonly referred to as the bootstrap world, and for our purposes we will explain the effect of moving from the “real” world to the bootstrap world. This shift takes place when one applies the plug-in principle, which will now be discussed.

2.2.1 The plug-in principle

To explain the plug-in principle, some notation needs to be introduced. Let θ denote any parameter of interest, and let $\hat{\theta}_n = \hat{\theta}(X_1, X_2, \dots, X_n)$ denote an estimator of that parameter based on a sample of size n . Suppose we express the parameter as a function that depends on the unknown distribution function F , that is

$$\theta = t(F).$$

This notation implies that the value of the parameter θ is obtained by applying some function, or a function of another function (called a functional, and denoted by $t(\cdot)$), to F . The plug-in principle then requires “plugging in” some approximation of the unknown distribution function F into the functional. We will denote the approximation of F by $\hat{F}$. This results in the bootstrap plug-in estimator of θ , given by $t(\hat{F})$. In the naive nonparametric bootstrap, F_n as defined in (1.17) is used as an approximation to F . As an example, consider the parameter μ , the population mean. Applying the plug-in principle using F_n we see that the plug-in estimator of μ is simply $\bar{X}_n$, the sample mean. In other words,

$$\mu = t(F) = \int x dF(x).$$

It follows that

$$\hat{\mu} = t(\hat{F}) = t(F_n) = \bar{X}_n.$$

The idea is to see the statistic $\hat{\theta}_n = \hat{\theta}(X_1, X_2, \dots, X_n)$ (which is calculated from the sample) as the parameter in the bootstrap world. We can then repeat the plug-in principle to estimate

the statistic. That is, using F_n as the estimate of F , define the bootstrap estimator

$$\hat{\theta}_n^* = t(\hat{F}^*) = t(F_n^*),$$

where

$$F_n^*(x) = \frac{1}{n} \sum_{i=1}^n I(X_i^* \leq x).$$

This implies applying the functional $t(\cdot)$ to the bootstrap estimate of F_n , denoted here by F_n^* . In order to create a bootstrap sample, denoted by $\mathbf{X}^* = (X_1^*, X_2^*, \dots, X_n^*)$, we simply sample with replacement from the original sample $X_1, X_2, \dots, X_n$, with equal probability of obtaining any of the original n sample values. This is obtained by executing the following steps :

1. Draw I_i uniformly and with replacement from $1, 2, \dots, n$.
2. Define the bootstrap variate $X_i^* = X_{I_i}$.

We denote this sampling structure as:

$$P_*(X_j^* = X_i) = \frac{1}{n}, \forall i, j = 1, 2, \dots, n. \quad (2.1)$$

The simulation part can be explained by the following simple algorithm:

1. Create a single bootstrap sample value using the sampling structure in (2.1).
2. Independently repeat (1) n times to create the first bootstrap sample, denoted by $\mathbf{X}^{*1} = X_1^*, X_2^*, \dots, X_n^*, (1)$.
3. Independently repeat (2) B times, where B indicates the number of bootstrap samples required.

The resulting bootstrap samples are denoted as follows:

$$\begin{aligned} \mathbf{X}^{*1} &= X_1^*, X_2^*, \dots, X_n^*, (1) \\ \mathbf{X}^{*2} &= X_1^*, X_2^*, \dots, X_n^*, (2) \\ &\vdots \\ \mathbf{X}^{*B} &= X_1^*, X_2^*, \dots, X_n^*, (B). \end{aligned}$$

For each of these B samples the bootstrap statistic $\hat{\theta}_n^* = \hat{\theta}(X_1^*, X_2^*, \dots, X_n^*)$ is calculated. Once all B replicates $\hat{\theta}_n^*(1), \hat{\theta}_n^*(2), \dots, \hat{\theta}_n^*(B)$ have been calculated, the distribution of the statistic $\hat{\theta}_n^*$ can be estimated. As $B \rightarrow \infty$, the accuracy of this distribution approximation increases. Now, the distribution of $\hat{\theta}_n^*$ can be used as an approximation of the distribution of the statistic $\hat{\theta}_n$. Table 2.1 summarizes the principle.

Real world	Bootstrap world
• F is the unknown distribution function. • θ is the unknown parameter. • $(X_1, X_2, \dots, X_n)$ is the observed sample. • $\hat{\theta}_n$ is the sample estimate of θ.	• $\hat{F}$ is an approximation to F. • $\hat{\theta}_n$ is the estimated parameter. • $(X_1^*, X_2^*, \dots, X_n^*)$ is a bootstrap sample. • $\hat{\theta}_n^*$ is a bootstrap replicate.

Table 2.1: *Effect of applying the plug-in principle.*

2.2.2 Bootstrap estimate of bias

This section describes the discussion on bias estimates taken from Efron and Tibshirani (1993). The bias of a kernel hazard function estimator as defined in (3.1) is discussed in Section 3.1.1. Since the value of the parameter θ is typically unknown, the bootstrap can be used to obtain an estimate of the bias. Denote the given observations from the unknown distribution function F by $\mathbf{X} = (X_1, X_2, \dots, X_n)$. We want to estimate the parameter $\theta = t(F)$. Let the estimator used be any statistic $\hat{\theta}_n = s(\mathbf{X})$. This estimator does not have to be the plug-in estimate $t(\hat{F})$. Define the bias of $\hat{\theta}_n = s(\mathbf{X})$ as follows:

$$bias_F = E_F[s(\mathbf{X})] - t(F). \quad (2.2)$$

The *ideal* bootstrap estimate of bias is defined as

$$bias_{\hat{F}} = E_{\hat{F}}[s(\mathbf{X}^*)] - t(\hat{F}). \quad (2.3)$$

Note that $t(\hat{F})$ may differ from $\hat{\theta}_n = s(\mathbf{X})$. We see here that $bias_{\hat{F}}$ is the plug-in estimate of $bias_F$, even if $\hat{\theta}_n = s(\mathbf{X})$ is not the plug-in estimate of θ . In some cases the *ideal* bootstrap bias as defined in (2.3) can be calculated explicitly, see Section 4.2.3 for an example. But for most statistics the ideal bootstrap estimate of bias needs to be approximated by simulation. This approximation requires the generation of B independent bootstrap samples $\mathbf{X}^{*1}, \mathbf{X}^{*2}, \dots, \mathbf{X}^{*B}$ and is consequently defined as

$$\hat{bias}_B = \hat{\theta}_n^*(\cdot) - t(\hat{F}), \quad (2.4)$$

where $\hat{\theta}_n^*(\cdot) = \frac{1}{B} \sum_{b=1}^B \hat{\theta}_n^*(b) = \frac{1}{B} \sum_{b=1}^B s(\mathbf{X}^{*b})$ and $\hat{\theta}_n^*(b) = s(\mathbf{X}^{*b})$.

2.2.3 Bootstrap estimate of standard error

In this paragraph we want to discuss the use of the bootstrap in estimating the standard error of some statistic $\hat{\theta}_n$, as an estimator of the parameter θ . Define the standard error of $\hat{\theta}_n$ by

$$\sigma_{SE} = \sqrt{\text{Var}_F(\hat{\theta}_n)}. \quad (2.5)$$

Now, by applying the plug-in principle we arrive at the bootstrap estimate of the standard error of $\hat{\theta}_n$, here denoted by $\hat{\sigma}_{SE}$. That is, we have

$$\hat{\sigma}_{SE} = \sqrt{\text{Var}_{F_n}(\hat{\theta}_n^*)}, \quad (2.6)$$

if we use F_n as approximation to the unknown F . Now, once again this expression is known as the *ideal* bootstrap estimate of standard error, and can be calculated explicitly in some situations, see for example Section 4.2.3.

In the cases where the value cannot be calculated explicitly, one has to approximate the value using Monte-Carlo (or *MC*) simulation based on a large number of bootstrap samples, denoted by B . The approximated bootstrap standard error value for $\hat{\theta}_n$ is denoted by $\hat{\sigma}_{SE_B}$ and is defined as follows:

$$\hat{\sigma}_{SE_B} = \sqrt{\frac{1}{B-1} \sum_{b=1}^B (\hat{\theta}_n^*(b) - \hat{\theta}_n^*(\cdot))^2}, \quad (2.7)$$

where $\hat{\theta}_n^*(\cdot)$ is defined as previously. Note that to obtain the *ideal* bootstrap estimate of standard error, we need $B \rightarrow \infty$. But we can only generate a finite number of bootstrap samples, and therefore the simulated standard error is just an estimate of the *ideal* standard error. It can be shown that $\hat{\sigma}_{SE_B} \rightarrow \hat{\sigma}_{SE}$ as $B \rightarrow \infty$.

2.3 The smooth bootstrap

In the preceding sections, F_n was used as an approximation to F . In other words, we were approximating a continuous function by a discrete function. In an attempt to improve the accuracy of the approximation, the smooth bootstrap can be used. The smooth bootstrap uses a kernel distribution estimator $\hat{F}_{n,h}$ as an approximation to F , thus resulting in the approximation to F also being a continuous function. Consult Section 3.3.3 for an example of a kernel distribution function estimator.

Here we will list three ways of drawing samples from $\hat{F}_{n,h}$:

$$X_i^* = X_{I_i} + hT_i, \quad (2.8)$$

$$X_i^* = \bar{X}_n + (X_{I_i} - \bar{X}_n + hT_i) \left(1 + \frac{h^2 \sigma_k^2}{S_n^2}\right)^{-\frac{1}{2}}, \quad (2.9)$$

$$X_i^* = \bar{X}_n + (X_{I_i} - \bar{X}_n + S_n T_i) (1 + \sigma_k^2)^{-\frac{1}{2}}. \quad (2.10)$$

Remarks

1. In the above list X_{I_i} denotes the i^{th} value drawn from F_n based on $X_1, X_2, \dots, X_n$ by applying the principle noted in (2.1).
2. The value T_i denotes a random value drawn from the kernel function k , and is drawn independently of X_{I_i} (see Section 3.3.3 for a discussion on kernel functions).
3. S_n^2 denotes the sample variance defined as $S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$.
4. σ_k^2 denotes the variance of the kernel function k .
5. h denotes the smoothing parameter used in the kernel distribution function estimator $\hat{F}_{n,h}$.

We will now show that using sampling scheme (2.8) is equivalent to sampling from $\hat{F}_{n,h}$. The derivation will be complete if we can show that $P(X^* \leq u) = \hat{F}_{n,h}(u)$. Note that

$$P(X_i^* \leq u) = P(X_{I_i} + hT_i \leq u).$$

Using the *Law of Total Probability*, we can deduce that

$$\begin{aligned} P(X_i^* \leq u) &= P(X_{I_i} + hT_i \leq u) \\ &= \sum_{j=1}^n P(X_{I_i} + hT_i \leq u | X_{I_i} = X_j) P(X_{I_i} = X_j) \\ &= \sum_{j=1}^n P(X_{I_i} + hT_i \leq u | X_{I_i} = X_j) \frac{1}{n} \\ &= \sum_{j=1}^n P(X_j + hT_i \leq u) \frac{1}{n} \\ &= \frac{1}{n} \sum_{j=1}^n P(T_i \leq \frac{u - X_j}{h}) \\ &= \frac{1}{n} \sum_{j=1}^n K\left(\frac{u - X_j}{h}\right) \\ &= \hat{F}_{n,h}(u), \end{aligned}$$

where K is an integrated kernel function as defined in Section 3.3.3. Recall that we aim to draw bootstrap samples that mimic the sampling structure one would expect from the unknown distribution function F . From (2.8) we have

$$\begin{aligned} E(X^*) &= E(X_{I_i} + hT) \\ &= E(X_{I_i}) + hE(T) \\ &= \bar{X}_n + h \cdot 0 \\ &= \bar{X}_n. \end{aligned}$$

From this we see that when a symmetric kernel is used, the expected values of the naive bootstrap observation X_{I_i} and the smooth bootstrap observation X^* are the same, i.e. equal to $\bar{X}_n$. This is what we expect from the sample. The shortcoming of (2.8) becomes clear when one studies the variance term of the different bootstrap observations:

$$\begin{aligned} Var_{\hat{F}_{n,h}}(X^*) &= Var_{\hat{F}_{n,h}}(X_{I_i} + hT) \\ &= Var_{F_n}(X_{I_i}) + h^2 Var(T) \\ &= S_n^2 + h^2 Var(T). \end{aligned}$$

Clearly this is not a desirable property, as smooth bootstrap observations obtained by using (2.8) lead to bootstrap samples with inflated variance (since $h^2 Var(T) > 0$). In an effort to counter this obvious problem, different smooth bootstrap sampling structures were introduced. The sampling schemes (2.9) and (2.10) are two alternatives that have the same expected value and variance as the naive bootstrap observation, therefore resulting in a better approximation of F .

2.3.1 Motivation for using the smooth bootstrap

Marron (1992) reviews results with regard to the application of the bootstrap to the problem of smoothing parameter selection. The discussion is formulated in the setting of kernel density estimation, but can be applied analogously to settings such as hazard function estimation. The kernel density estimator $\hat{f}_{n,h}$ as defined in (3.19) is introduced, as well as the corresponding MISE global discrepancy measure, defined in (3.8). The decomposition of the MISE term as the sum of the integrated variance and the squared integrated bias is shown, as well as an AMISE (see Section 3.1.5) expression illustrating the *trade-off concept* associated with h more clearly.

The author goes on to explain the idea of using the bootstrap to approximate some distributional aspect of the estimator under discussion, denoted in the article as $\mathcal{L}\{\hat{f}_{n,h}(x) - f(x)\}$. Using the plug-in principle, and F_n as approximation to the unknown distribution function F , one hopes that the following holds true:

$$\mathcal{L}\{\hat{f}_{n,h}^*(x) - \hat{f}_{n,h}(x)\} \doteq \mathcal{L}\{\hat{f}_{n,h}(x) - f(x)\}.$$

Faraway and Jhun (1990) proved that this approximation is not useful for bandwidth selection, since $E_*\{\hat{f}_{n,h}^*(x)\} = \hat{f}_{n,h}(x)$. This implies that there is *no bias* in the bootstrap world, which is very undesirable, since we will show in Section 3.1.6 that with smoothing parameter selection a value for h is sought that will balance the opposing values of bias and variance in discrepancy measures like the MISE.

If there is no bias, the applicable discrepancy measure will decrease monotonically, and the selected smoothing parameter value will always be the largest possible value. In general, the bias term is *increasing* in h . Therefore, if there is no bias, the applicable discrepancy measure will not incorporate this property, and behave just like the variance term, that is it will be strictly *decreasing* in h .

As a remedy to this flaw, the smooth bootstrap is proposed. Denote by g the bandwidth used when creating the smooth bootstrap samples. We are now in a position where we have the following approximation:

$$\mathcal{L}\{\hat{f}_{n,h}^*(x) - \hat{f}_{n,g}(x)\} \doteq \mathcal{L}\{\hat{f}_{n,h}(x) - f(x)\}.$$

One can then select a bandwidth h^* that minimizes the bootstrap approximation to the MISE, denoted by MISE^* . Decomposition of the MISE^* term is then given explicitly, such that the bootstrap approximation based on simulation is unnecessary.

The remainder of the article discusses asymptotic theory on the topic, as well as the connection with other methods and finally a section focusing on simulation and application of the method. In short, the article illustrates the problems of the naive bootstrap (resampling simply from F_n) when applied to the context of smoothing parameter selection, and also describes how the use of the smooth bootstrap can overcome those problems. Figure 2.1 shows this graphically (for the estimator $\hat{f}_{n,h}$). Note how the naive bootstrap approximation to $\text{MISE}_w(h)$ is monotone decreasing in h , whereas the smooth bootstrap provides a more accurate approximation to $\text{MISE}_w(h)$, where $\text{MISE}_w(h)$ represents a weighted MISE, as defined in (3.12).

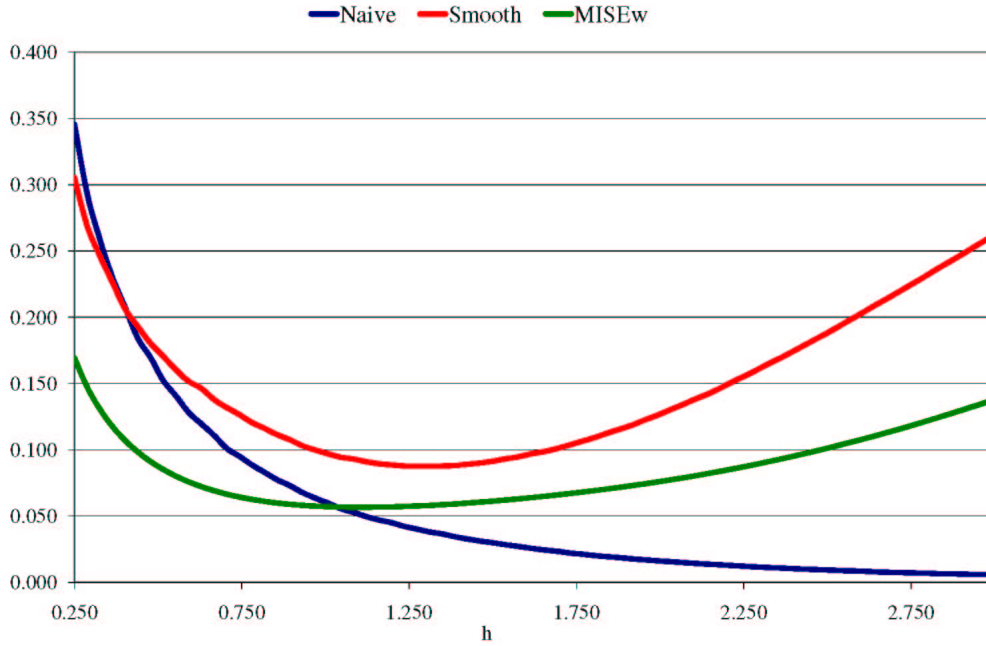


Figure 2.1: *Different bootstrap approximations to the function $MISE_w(h)$.*

2.4 Bootstrapping censored data

Recall Section 1.4.1 where the notation of random right censoring was introduced. There we defined the observed sample $\{X_i\}_{i=1}^n$ as a sample drawn from F . When we assume independence between F^0 and H^0 we have the following equality: $(1 - F) = (1 - F^0)(1 - H^0)$. This equality can be derived as follows:

$$\begin{aligned}
 1 - F(x) &= 1 - P(X_i \leq x) \\
 &= P(X_i > x) \\
 &= P(X_i^0 > x, Z_i > x) \\
 &= P(X_i^0 > x)P(Z_i > x) \text{ (by independence)} \\
 &= [1 - P(X_i^0 \leq x)][1 - P(Z_i \leq x)] \\
 &= (1 - F^0(x))(1 - H^0(x)).
 \end{aligned}$$

We proceed to discuss the two procedures introduced by Efron (1981) for bootstrapping in the case of random right censoring.

Procedure 1

The procedure resamples independently from the Kaplan-Meier estimators of F^0 and H^0 respectively, and then constructs the bootstrap sample by selecting the minima and defining indicators. This procedure could be called model based since it uses the specific structure of the random right censoring model.

1. Independently resample $\{X_i^{*0}\}_{i=1}^n \stackrel{i.i.d}{\sim} \hat{F}_n^0$ and $\{Z_i^*\}_{i=1}^n \stackrel{i.i.d}{\sim} \hat{H}_n^0$.
2. Calculate $X_i^* = \min(X_i^{*0}, Z_i^*)$ and $\Delta_i^* = I(X_i^* = X_i^{*0})$ for $i = 1, 2, \dots, n$.

Procedure 2

In this procedure one resamples from the empirical distribution function of the pairs. This procedure is more naive and could be called model free.

1. Resample (X_i^*, Δ_i^*) as a random sample with replacement from the pairs $\{(X_i, \Delta_i)\}_{i=1}^n$.
2. Independently repeat (1) n times to obtain $\{(X_i^*, \Delta_i^*)\}_{i=1}^n$.

Equality of the two procedures

Recall the definitions of $\hat{F}_n^0$ and $\hat{H}_n^0$ as given in Section 1.4.3. At the start of this section we derived the equality

$$1 - F(x) = (1 - F^0(x))(1 - H^0(x)),$$

by the assumed independence of F^0 and H^0 . Efron (1981) concludes that the nonparametric MLE for F , denoted $\hat{F}_n$ is given by

$$\begin{aligned} 1 - \hat{F}_n(x) &= [1 - \hat{F}_n^0(x)][1 - \hat{H}_n^0(x)] \\ &= \prod_{j=1}^{i-1} \left(\frac{n-j}{n-j+1} \right)^{\Delta_j} \left(\frac{n-j}{n-j+1} \right)^{1-\Delta_j} \\ &= \prod_{j=1}^{i-1} \left(\frac{n-j}{n-j+1} \right) \\ &= 1 - F_n(x). \end{aligned}$$

Remarks:

1. The above is valid for x satisfying $X_{(i-1)} \leq x < X_{(i)}$.
2. The last step follows from the derivation at the end of Section 1.4.3.

From this we see that resampling from $1 - \hat{F}_n$ is the same as resampling from the pairs $\{X_i, \Delta_i\}_{i=1}^n$ with probability mass of $1/n$ at each pair, censored or not. Because of this equality, X_i^* equals X_j with probability $1/n$, and this holds for $j = 1, 2, \dots, n$.

Furthermore, if $X_i^* = X_j$, then $\Delta_i^* = \Delta_j$, because $\hat{F}_n^0$ only puts mass on those X_j where $\Delta_j = 1$, $\hat{H}_n^0$ only puts mass on those X_j where $\Delta_j = 0$ and we have assumed no ties. This shows that the two procedures introduced in this section are equivalent.

Chapter 3

Kernel smoothing

3.1 Discrepancy measures

Before we can discuss kernel smoothing in detail, we need to understand how to go about measuring the accuracy and precision of any estimation method. All the concepts in this section will be discussed in the context of kernel estimators of the hazard function.

3.1.1 Pointwise discrepancy measures

As a starting point, discrepancy measures that quantify the accuracy and precision of an estimator in a single point x will be discussed. The values of these measures depend among others on the following factors: the point of evaluation x , the kernel function k and most importantly, the smoothing parameter h . It is the focus of this dissertation to study the effect of h on the discrepancy measures introduced below. The dependence of these measures on h will be discussed in Section 3.1.6.

Bias

The pointwise bias of a kernel hazard function estimator $\hat{\lambda}_{n,h}$ measures how close the expected estimated value is to the true hazard function λ . The bias therefore measures the accuracy of the estimator, by measuring the difference between the expected value of the estimator and the (unknown) parameter. Formally, pointwise bias is defined as

$$Bias(\hat{\lambda}_{n,h}(x)) = E[\hat{\lambda}_{n,h}(x)] - \lambda(x). \quad (3.1)$$

An estimator might overestimate the parameter value (resulting in a positive value for the bias), or it might underestimate the parameter value (resulting in a negative value for the bias). Preferably an estimator should be unbiased, a term used to indicate a bias of zero.

This means that the expected value of the estimator in the point x is exactly equal to the parameter value in the point x , i.e. $E[\hat{\lambda}_{n,h}(x)] = \lambda(x)$. Note that, in general, the squared bias term of any estimator is increasing in h , meaning that if the value of h increases, the value of the squared bias term will increase as well.

Variance

The pointwise variance of a kernel estimator $\hat{\lambda}_{n,h}$ measures the spread of the estimator around the expected value of the estimator. One can interpret the pointwise variance as measuring the precision of the estimator, as it indicates how the value of the estimator varies from the expected value of the estimator in the point x . Pointwise variance is defined as

$$Var(\hat{\lambda}_{n,h}(x)) = E \left[(\hat{\lambda}_{n,h}(x) - E[\hat{\lambda}_{n,h}(x)])^2 \right]. \quad (3.2)$$

The ideal is that the variance term (which is always positive) should be as small as possible, as this will indicate that the estimator is at least precise when estimating the hazard function. If the variance term is large, it indicates that the estimator varies significantly from sample to sample, which is an undesirable property for an estimator. Note that in general the variance term of any estimator is decreasing in h , which means that if the value of h increases, the value of the variance term will decrease.

Mean squared error

The bias quantifies the accuracy of the estimator, whereas the variance quantifies the precision of the estimator. A measure that incorporates both these discrepancy measures is the pointwise mean squared error, or MSE, which is defined as:

$$MSE(\hat{\lambda}_{n,h}(x)) = E \left[(\hat{\lambda}_{n,h}(x) - \lambda(x))^2 \right]. \quad (3.3)$$

This is quite useful, as one can measure both accuracy and precision by studying the value of the MSE. The following will show how the bias and variance combine to give the value of the MSE.

Representation in terms of bias and variance terms

$$\begin{aligned}
MSE(\hat{\lambda}_{n,h}(x)) &= E[(\hat{\lambda}_{n,h}(x) - \lambda(x))^2] \\
&= E[(\hat{\lambda}_{n,h}(x) - E(\hat{\lambda}_{n,h}(x))) + (E(\hat{\lambda}_{n,h}(x)) - \lambda(x))^2] \\
&= E[(\hat{\lambda}_{n,h}(x) - E(\hat{\lambda}_{n,h}(x)))^2 + 2[\hat{\lambda}_{n,h}(x) - E(\hat{\lambda}_{n,h}(x))][E(\hat{\lambda}_{n,h}(x)) - \lambda(x)] \\
&\quad + (E(\hat{\lambda}_{n,h}(x)) - \lambda(x))^2] \\
&= Var(\hat{\lambda}_{n,h}(x)) + 2E[(\hat{\lambda}_{n,h}(x) - E(\hat{\lambda}_{n,h}(x)))(E(\hat{\lambda}_{n,h}(x)) - \lambda(x))] \\
&\quad + E[(E(\hat{\lambda}_{n,h}(x)) - \lambda(x))^2] \\
&= Var(\hat{\lambda}_{n,h}(x)) + 2(E(\hat{\lambda}_{n,h}(x)) - \lambda(x))E[(\hat{\lambda}_{n,h}(x) - E(\hat{\lambda}_{n,h}(x)))] \\
&\quad + (E(\hat{\lambda}_{n,h}(x)) - \lambda(x))^2 \\
&= Var(\hat{\lambda}_{n,h}(x)) + 2(E(\hat{\lambda}_{n,h}(x)) - \lambda(x))[E(\hat{\lambda}_{n,h}(x)) - E(\hat{\lambda}_{n,h}(x))] \\
&\quad + [Bias(\hat{\lambda}_{n,h}(x))]^2.
\end{aligned}$$

Therefore we have the following decomposition of the MSE:

$$MSE(\hat{\lambda}_{n,h}(x)) = Var(\hat{\lambda}_{n,h}(x)) + [Bias(\hat{\lambda}_{n,h}(x))]^2. \quad (3.4)$$

3.1.2 Global discrepancy measures

When trying to quantify the quality of an estimator, one frequently desires a global discrepancy measure, in other words a measure that quantifies accuracy and precision across the entire domain of the variable under study (this will typically be the interval $[0, \infty)$ for survival random variables), as opposed to a single point x . These measures thus give an overall picture of the closeness of the estimator to the parameter.

These measures depend among others on the kernel function k and most importantly, the smoothing parameter h . The dependence of these measures on h will be discussed in Section 3.1.6, whereas kernel functions will receive more attention in Section 3.3.2. Note that in contrast to pointwise discrepancy measures, the global discrepancy measures do not depend on an evaluation point x .

Integrated bias

This measure represents the bias of the estimator across the entire domain of the variable under study, and can be intuitively interpreted as the overall bias of the estimator. It is defined by

$$IBias(h) = \int_0^\infty (E[\hat{\lambda}_{n,h}(x)] - \lambda(x)) dx. \quad (3.5)$$

Integrated variance

In the same way the integrated variance determines the variance of the estimator over the whole domain of the variable, with the intuitive interpretation of measuring the overall

variance of the estimator. It is defined as

$$IVar(h) = \int_0^\infty E \left[(\hat{\lambda}_{n,h}(x) - E[\hat{\lambda}_{n,h}(x)])^2 \right] dx. \quad (3.6)$$

3.1.3 Integrated squared error

The integrated squared error (or ISE) is an appropriate global discrepancy measure if one is only concerned with the data set at hand, since it does not take into account other possible data sets (see Wand and Jones (1995)). The ISE is defined as

$$ISE(h) = \int_0^\infty [\hat{\lambda}_{n,h}(x) - \lambda(x)]^2 dx. \quad (3.7)$$

Note that the ISE is a random quantity that varies from sample to sample.

3.1.4 Mean integrated squared error

The mean integrated squared error (or MISE) is a global discrepancy measure that calculates the expected value of the random ISE quantity, and is therefore a non-random value. The MISE averages the ISE value over all possible data sets. It can be defined as

$$MISE(h) = E[ISE(h)] = E \left\{ \int_0^\infty [\hat{\lambda}_{n,h}(x) - \lambda(x)]^2 dx \right\}. \quad (3.8)$$

By using Fubini's theorem, we note that the order of integration can be reversed (since the integrand is positive and continuous), thus resulting in

$$MISE(h) = \int_0^\infty \left\{ E[\hat{\lambda}_{n,h}(x) - \lambda(x)]^2 \right\} dx = \int_0^\infty MSE(\hat{\lambda}_{n,h}(x)) dx. \quad (3.9)$$

As a consequence, this global discrepancy measure is sometimes referred to as the integrated mean squared error (or IMSE). Also note that, analogous to the pointwise MSE, the MISE can be written as the sum of the integrated variance and squared integrated bias, that is

$$MISE(h) = IVar(h) + [IBias(h)]^2. \quad (3.10)$$

3.1.5 Asymptotic mean integrated squared error (AMISE)

As will be discussed in Section 3.1.6, we are interested in studying the behavior of these discrepancy measures as a function of the smoothing parameter h . One obstacle in this regard is that the dependence on h cannot be explicitly expressed in most situations.

It is therefore required to try and obtain a simple representation of the dependence of these measures on h . This is achieved by studying the measures in an asymptotic sense, that is in cases where the sample size n increases to infinity, or $n \rightarrow \infty$.

Taylor expansions

Theorem 1. *Suppose f is a real-valued function defined on $\mathbb{R}$ and let $x \in \mathbb{R}$. Assume that f has p continuous derivatives in an interval $(x - \Delta, x + \Delta)$ for some $\Delta > 0$. Then, for any sequence α_n converging to zero, we have $f(x + \alpha_n) = \sum_{j=0}^p \frac{\alpha_n^j}{j!} f^{(j)}(x) + o(\alpha_n^p)$.*

By using Taylor expansions one can express the influence of h on the global discrepancy measures in a much simpler way, which will simplify the selection of an appropriate h . Usually an AMISE expression can be derived with respect to h and solved to give an asymptotically optimal value for h in a closed form. But note that these asymptotic expressions are derived based on sample sizes increasing to infinity, and should be used with caution! Some AMISE expressions will be discussed in Section 3.3.3.

Weight functions

When integrating certain discrepancy measures (like bias or variance), it can sometimes happen that one or both of the integrals are undefined, or in some cases the values of the integrals tend to infinity. In order to correct for this, a so-called weight function is employed to ensure that the calculated values of the integrals are finite. The weight function usually alters the bounds over which the associated term is integrated.

In short, weight functions help to eliminate endpoint effects, resulting in integrated terms being computable and finite. For example, consider for some distribution function F the weight function:

$$w(x) = I_{F^{-1}(0.25), F^{-1}(0.75)}(x). \quad (3.11)$$

By applying this weight function to the MISE term, the resulting weighted MISE function is :

$$MISE_w(h) = \int_0^\infty MSE(\hat{\lambda}_{n,h}(x))w(x)dx = \int_{F^{-1}(0.25)}^{F^{-1}(0.75)} MSE(\hat{\lambda}_{n,h}(x))dx. \quad (3.12)$$

Analogously we can define a weighted ISE expression by

$$ISE_w(h) = \int_0^\infty [\hat{\lambda}_{n,h}(x) - \lambda(x)]^2 w(x)dx \quad (3.13)$$

Please note that, unless otherwise mentioned, the weight function defined in (3.11) was used throughout the simulation study in Chapter 5, but with $F = F^0$.

3.1.6 The smoothing parameter h and the trade-off concept

This section focuses on the dependence on the smoothing parameter h of the discrepancy measures defined in Section 3.1.1 and Section 3.1.2.

Trade-off concept

The reader is referred to the decomposition of MSE given in (3.4). This property of the MSE is very useful indeed as it allows one to study the combined effect of varying the smoothing parameter h on the performance of the estimator. Recall from Section 3.1.1 that the squared bias term increases as the value of h increases, and that the variance term decreases as the value of h increases. It is thus clear that there is a trade-off between these measures with regard to h .

Choosing h too small results in the variance term dominating the MSE term, whereas choosing h too large will result in the squared bias term dominating the MSE term. When the squared bias term dominates, the estimator is said to be *oversmoothed*, and if the variance term dominates the estimator is said to be *undersmoothed*. Ideally one would like to balance these two opposing measures (with regard to h), and find a value for h that keeps the two measures at the same order (thereby hoping to minimise the relevant discrepancy measure). This problem is known as the smoothing parameter selection problem, and is the main focus of this dissertation.

Figure 3.1 shows the smoothing parameter selection problem graphically. From this one can get an intuitive understanding of the influence of h on the performance of the estimator. One can clearly see how the blue line is undersmoothed ($h = 0.25$), since the estimator varies substantially along the interval of estimation. Also note that the red line oversmooths the data ($h = 2.5$), and results in an estimated curve that does not follow the trend of the underlying hazard. Finally, the green line is an attempt to find a balance ($h = 0.5$) between following the data pattern but not being too variable (or “spiky”).

Furthermore, the value of h depends on the sample size n , but mostly this dependence is suppressed in the notation, and therefore we shall write only h instead of h_n . Also, the following two assumptions are made regarding h :

1. $\lim_{n \rightarrow \infty} h = 0$.
2. $\lim_{n \rightarrow \infty} nh = \infty$.

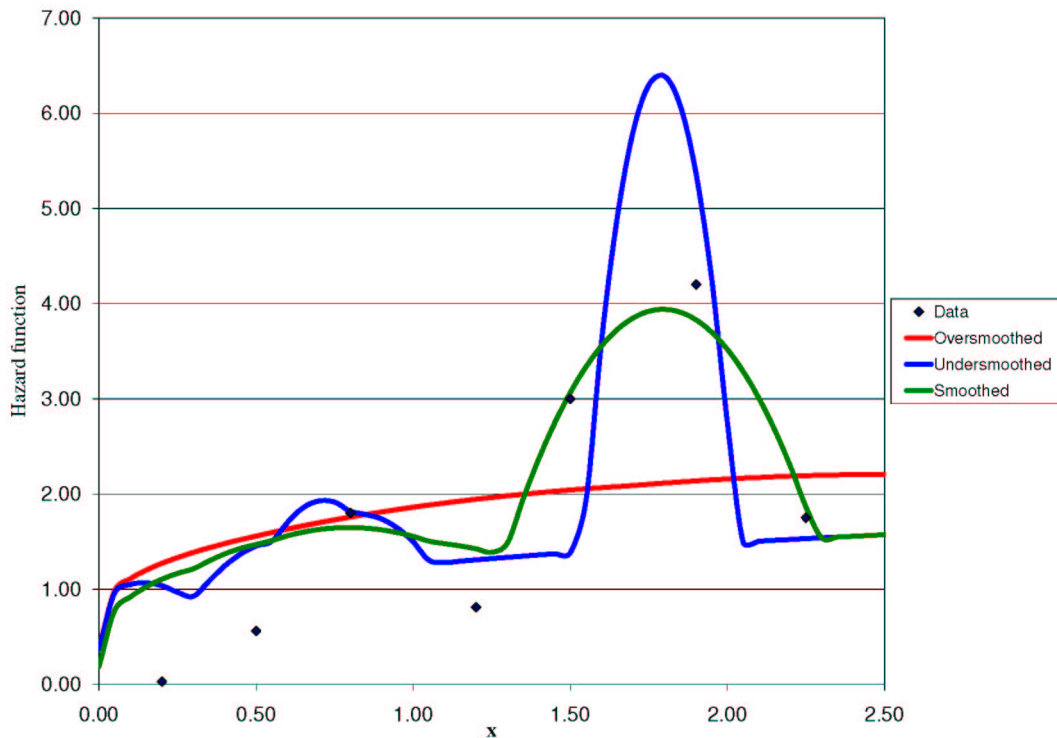


Figure 3.1: *Graphical representation of smoothing parameter selection problem*

From this we gather that h decreases as n increases, but h converges to zero at a rate slower than n tends to infinity.

3.2 Different estimation methods

Since the hazard function is unknown in practice, it is of interest to find an efficient way of estimating this function. Estimation methods can be mainly divided into two types, namely parametric estimation and nonparametric estimation methods. The following paragraphs briefly discuss the two approaches. For a more thorough discussion, the interested reader can consult Härdle (1990), Chapter 1.

3.2.1 Parametric estimation

Parametric estimation involves assuming a known functional form for the hazard function beforehand (for instance constant). This functional form can then be fully described by a finite set of parameters, and the task at hand is to estimate these parameters. The success of the estimator depends heavily on the choice of the functional form, since it is believed that the approximation bias of the best parametric fit is a negligible quantity (see

Härdle (1990)). An advantage of parametric estimation methodology lies in the inferences made from an estimation procedure, where parametric methods lead to higher power values when compared to the power values obtained from nonparametric methods, that is if the assumptions of the parametric method hold.

A disadvantage of parametric estimators is the inflexibility of the methods, in the sense that once the functional form is fixed the methods focus only on obtaining the best parameter estimates, and there is no room to adapt or modify the functional form. Also note that, in general, the usefulness of parametric methods rely heavily on the validity of the assumptions made beforehand, one possible problem for instance is the misspecification of the functional form.

3.2.2 Nonparametric estimation

Nonparametric estimation, on the other hand, avoids the assumption of a functional form beforehand. These methods are sometimes referred to as distribution-free methods, since the functional form of the estimators are not specified. The key idea of nonparametric methods is to let the data at hand dictate the shape or form of the estimator, rather than placing a constraint on the form. Nonparametric methods are often used as a starting point in parametric investigations. Once a nonparametric method has been applied and interpreted, a suitable parametric form might be suggested.

Another possible approach is the so-called semiparametric estimation methods, where the aim is to combine the strengths of both parametric and nonparametric methods. Examples of semiparametric methods include the Cox proportional hazards (CPH) model, as well as the accelerated failure time (AFT) model. Since the focus of this dissertation will be on nonparametric methods, these methods will only be mentioned.

Nonparametric estimation methods are collectively known as smoothing methods, or simply smoothers. In this dissertation kernel smoothing will receive the most attention, and the motivation for this choice will be given in Section 3.3. There are a number of other hazard function smoothers, for instance nearest neighbour smoothers (see Gefeller and Dette (1992)), penalized likelihood methods (see Antoniadis and Grégoire (1990)), and orthogonal wavelet smoothers (see Patil (1997)).

3.3 Kernel smoothing

3.3.1 Introduction and motivation

Kernel smoothing provides a simple way of finding structure in data sets without the imposition of a parametric model. Wand and Jones (1995) state that kernel smoothers have the advantage of being very intuitive and relatively simple to analyze mathematically. Before discussing some specific aspects of kernel smoothing, the basic idea of the method must be introduced.

3.3.2 Kernel functions

Kernel functions have been studied quite thoroughly, and there are numerous alterations and variations available. Usually (that is in simple cases) kernel functions have the following properties:

1. $k(x) \geq 0 \forall x \in \mathbb{R}$.
2. $\int k(x)dx = 1$.
3. $k(x) = k(-x)$.
4. $\int xk(x)dx = 0$.
5. $\int x^2k(x)dx = c^2 < \infty$.

In summary, a kernel function is usually a non-negative density function that is symmetric about 0, with a first moment of zero and a finite second moment.

Different kernels

A large number of popular kernel functions are part of the kernel family defined by

$$k(x; p) = 2^{2p+1} B(p+1, p+1)^{-1} (1-x^2)^p I(|x| < 1), \quad (3.14)$$

where $B(\cdot, \cdot)$ is the Beta function which can be defined as

$$B(x, y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}, \quad (3.15)$$

where Γ is the Gamma function defined by

$$\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} dt. \quad (3.16)$$

Note that for n a positive integer we have $\Gamma(n) = (n-1)!$ (that is $(n-1)$ factorial). Now, using $p = 0$ in equation (3.14) we obtain the uniform kernel. Substituting $p = 1$ results in the Epanechnikov kernel, whereas $p = 2$ delivers the biweight kernel and $p = 3$ the triweight kernel. The standard normal kernel can be seen as the limiting case where $p \rightarrow \infty$. The only kernel mentioned here that is not part of this kernel family is the triangular kernel. Table 3.1 summarizes the kernel functions. Note that all the kernels defined above, except one, have

Kernel	Expression	p
Triangular	$k(x) = (1 - x)I_{(x < 1)}$	-
Uniform	$k(x) = \frac{1}{2}I_{(x < 1)}$	0
Epanechnikov	$k(x) = \frac{3}{4}(1 - x^2)I_{(x < 1)}$	1
Biweight	$k(x) = \frac{15}{16}(1 - x^2)^2I_{(x < 1)}$	2
Triweight	$k(x) = \frac{35}{32}(1 - x^2)^3I_{(x < 1)}$	3
Normal	$k(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$	∞

Table 3.1: *Different kernel functions.*

bounded support in the sense that only values of x in the interval $(-1, 1)$ are assigned strictly positive kernel values. The only kernel mentioned here with unbounded support (that means all values of x in the interval $(-\infty, \infty)$ are assigned strictly positive kernel values) is the standard normal kernel. Figure 3.2 illustrates the different kernel functions noted in Table 3.1.

The kernel function K is known as an integrated kernel function, and is defined as

$$K(x) = \int_{-\infty}^x k(t)dt. \quad (3.17)$$

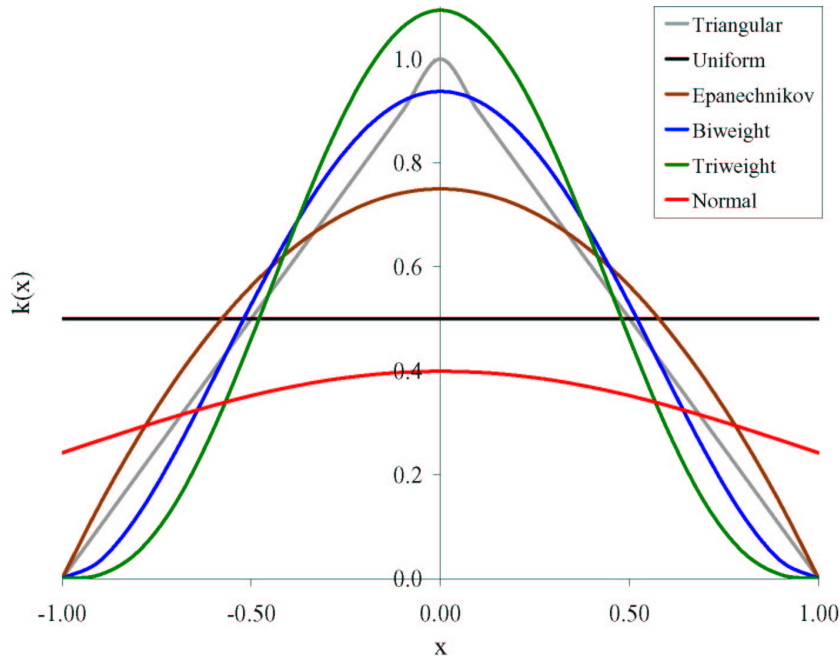
The integrated kernel function has the following properties:

1. $\lim_{t \rightarrow -\infty} K(t) = 0.$
2. $\lim_{t \rightarrow \infty} K(t) = 1.$

Therefore, an integrated kernel function is (usually) a distribution function, since it is defined as the integral of a density function.

3.3.3 The concept of kernel smoothing

To illustrate the use of kernel smoothing, we will use the setting of density estimation. This involves the estimation of a continuous probability density function f in a specific

Figure 3.2: *Different kernel functions*

point x . There are two reasons for starting with density estimation. Firstly, kernel density estimators are the most straightforward among kernel estimators, and therefore will be a sensible starting point. Secondly, the reader is reminded that the hazard function can be expressed as $\lambda(x) = f(x)/[1-F(x)]$, which emphasizes the importance of the density function when estimating the hazard function.

The kernel density estimator of f which will be used is known as the Rosenblatt-Parzen estimator. Consult Rosenblatt (1956) and Parzen (1962) for more details. It is defined as

$$\hat{f}_{n,h}(x) = \frac{1}{nh} \sum_{i=1}^n k\left(\frac{x - X_i}{h}\right). \quad (3.18)$$

In this definition k represents a weighting function, known as the kernel function, and h represents the window width (also bandwidth or smoothing parameter). Note that the rescaling function $k_h(x) = \frac{1}{h}k\left(\frac{x}{h}\right)$ is a more compact way of expressing the kernel function. We can then define the estimator concisely as

$$\hat{f}_{n,h}(x) = \frac{1}{n} \sum_{i=1}^n k_h(x - X_i). \quad (3.19)$$

The bandwidth h regulates the size of the interval over which the local average will be calculated, whereas the kernel k regulates how those observations will be weighted in the

calculation of the local average. The smoothing parameter h is discussed in Section 3.1.6.

To construct a kernel estimate, a scaled kernel (that is a kernel k_h with smoothing parameter h) is centered at each of the n observations. The kernel estimate is then calculated as the average of the n kernel values at the specific point x . Before giving an example, a kernel distribution function estimator and a kernel hazard function estimator will be discussed.

We defined the kernel density estimator earlier in (3.19). The distribution function of a continuous random variable X in the point x is defined as

$$F(x) = \int_{-\infty}^x f(y)dy. \quad (3.20)$$

The kernel distribution function estimator $\hat{F}_{n,h}$ is defined by substituting the kernel density estimator $\hat{f}_{n,h}$ for f in equation (3.20). This results in the following:

$$\hat{F}_{n,h}(x) = \int_{-\infty}^x \frac{1}{nh} \sum_{i=1}^n k\left(\frac{t - X_i}{h}\right) dt \quad (3.21)$$

$$= \frac{1}{n} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \quad (3.22)$$

$$= \frac{1}{n} \sum_{i=1}^n K_h(x - X_i), \quad (3.23)$$

where $K_h(x - X_i) = K\left(\frac{x - X_i}{h}\right)$.

The properties of the integrated kernel were noted in Section 3.3.2. Note once more that the smoothing parameter h regulates the size of the interval over which the local average will be calculated, whereas the integrated kernel K determines the weights of the observations included in the local average.

To construct a kernel distribution function estimator in the point x , center an integrated kernel K_h with smoothing parameter h on each of the n observations. This means we place the integrated kernel in such a way that $K_h(X_i) = 1 - K_h(X_i)$, or simply such that $K_h(X_i) = 0.5$. Therefore the integrated kernel is positioned (centered) such that when evaluated in the observation X_i , the function value corresponds to the median of the integrated kernel. This is done for each of the n observations. The value of the estimator is calculated as the average of the n kernel values in the point x .

Lastly we will explore the use of kernel smoothing in the setting of hazard function estimation. The estimator to be used was first introduced by Watson and Leadbetter (1964a) and will be discussed in more detail in Section 4.2. The kernel estimator is defined as follows:

$$\hat{\lambda}_{n,h}^{(2)}(x) = \sum_{i=1}^n \frac{k_h(x - X_{(i)})}{n - i + 1}, \quad (3.24)$$

where $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$ denote the order statistics. The smoothing parameter h once again controls the size of interval over which the smoothing will take place, and the kernel function k regulates the weights of the observations that are included in the smoothing interval.

The estimator is constructed as follows: Center on each of the n ordered observations a kernel k_h , and at each ordered observation point divide that kernel value by $(n - i + 1)$, where $i = 1, 2, \dots, n$. The estimated value in a point x is then calculated as the sum of the n kernel values.

Consider the following examples to establish an intuitive understanding of kernel estimation. Note beforehand that all three examples to be given below are based on very small sample sizes (for illustration purposes), and therefore may appear to be ineffective. The accuracy of these estimators increases as the sample size n increases.

Examples of kernel density estimation

To illustrate the use of a kernel density estimator, $n = 8$ values were randomly generated from the Beta(4, 4) density, symmetric about 0 with domain $(-1, 1)$. This density function (also known as the triweight density) is defined as:

$$f(x) = \frac{35}{32}(1 - x^2)^3 I_{(|x| < 1)}. \quad (3.25)$$

The sorted data set is given by the following:

-0.613	-0.563	-0.308	-0.215	-0.087	0.191	0.400	0.425
--------	--------	--------	--------	--------	-------	-------	-------

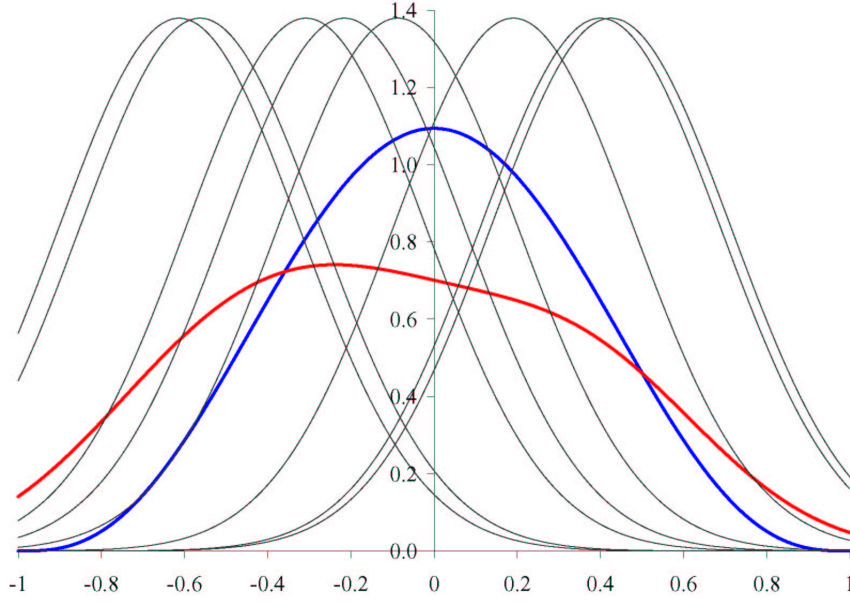
The kernel function used is the standard normal density ϕ , where

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \quad -\infty < x < +\infty. \quad (3.26)$$

The smoothing parameter is simply chosen as $h = 0.29$ (note that the choice of the smoothing parameter will be discussed at length later on). Figure 3.3 includes the target density f as the blue line, the kernel estimator $\hat{f}_{n,h}$ defined in (3.19) as the red line, as well as the rescaled kernels centered at each of the observation points. By studying this illustration, one can readily see how the kernel estimate is constructed in each point.

We will now show some asymptotic results for the accuracy and precision of the estimator $\hat{f}_{n,h}$, using results stated in Koekemoer (2004). From the definition of bias as given in (3.1) and applying Taylor expansions, the asymptotic bias of $\hat{f}_{n,h}$ is given by

$$\text{Bias}(\hat{f}_{n,h}(x)) = \frac{1}{2}h^2\mu_2(k)f''(x) + o(h^2), \quad (3.27)$$

Figure 3.3: *Kernel density estimator with $k = \phi$, and $h = 0.29$.*

where we introduce the notation

$$\mu_2(k) = \int_{-\infty}^{\infty} z^2 k(z) dz. \quad (3.28)$$

Note that the leading term of (3.27) is of order $o(h^2)$. Recall the two assumptions made regarding the value of h in Section 3.1.6 when the value of n tends to infinity. This leads us to the conclusion that $\hat{f}_{n,h}$ is asymptotically unbiased as an estimator of f .

The asymptotic variance term is given by

$$\text{Var}(\hat{f}_{n,h}(x)) = \frac{1}{nh} R(k) f(x) + o(n^{-1}h^{-1}), \quad (3.29)$$

where we introduce the notation

$$R(k) = \int_{-\infty}^{\infty} k(z)^2 dz. \quad (3.30)$$

Analogously, note that the leading term of (3.29) is $o(n^{-1}h^{-1})$. The assumptions regarding h in Section 3.1.6 now lead to the conclusion that the variance of $\hat{f}_{n,h}$ converges to zero as $n \rightarrow \infty$. Combining these results to obtain the asymptotic mean squared error and integrating that function over the entire domain of x leads to the following expression for the asymptotic mean integrated squared error:

$$\text{AMISE}(h) = \frac{1}{nh} R(k) + \frac{1}{4} h^4 \mu_2(k)^2 R(f''). \quad (3.31)$$

This function can now be evaluated as a function of h and the objective is to determine the value of h that minimizes this AMISE criteria. To do this, take the derivative of AMISE with regards to h and set the resulting function equal to zero. To obtain a minimum value for AMISE we must therefore choose the smoothing parameter as

$$h_{AMISE} = \left[\frac{R(k)}{\mu_2(k)^2 R(f'')} \right]^{\frac{1}{5}} n^{-1/5}, \quad (3.32)$$

the asymptotically optimal selection of h .

Example of kernel distribution estimation

As an example of kernel distribution estimation, the same $n = 8$ observations that were generated from (3.25) will be used to estimate the Beta(4, 4) distribution function, which we define as

$$F(x) = 140 \int_0^{\frac{x+1}{2}} t^3(1-t)^3 dt, \quad 0 \leq x \leq 1. \quad (3.33)$$

The integrated kernel used is the standard normal distribution function Φ as defined in (1.16). The smoothing parameter used in this example is $h = 0.35$ (once again, note that the choice of the smoothing parameter will be discussed in detail later on). Figure 3.4 includes the target distribution F as the blue line, the kernel estimator $\hat{F}_{n,h}$ defined in (3.21) as the red line, as well as the rescaled integrated kernels centered at each of the observation points. By studying this illustration, one can readily see how the kernel estimate is constructed in each point.

We will now briefly discuss asymptotic properties of $\hat{F}_{n,h}$ as an estimator of F , using results stated in Koekemoer (2004). The asymptotic bias term is given as

$$Bias(\hat{F}_{n,h}(x)) = \frac{1}{2} h^2 f'(x) \mu_2(k) + o(h^3). \quad (3.34)$$

Once again we see that the assumptions regarding h made in Section 3.1.6 imply that $\hat{F}_{n,h}$ is an asymptotically unbiased estimator of F .

The asymptotic variance of $\hat{F}_{n,h}$ is given by

$$Var(\hat{F}_{n,h}(x)) = \frac{1}{n} F(x)[1 - F(x)] - \frac{2h}{n} f(x) C_1 + o(n^{-1}h^{-1}), \quad (3.35)$$

where we introduce the notation

$$C_1 = \int_{-\infty}^{\infty} zk(z)K(z)dz. \quad (3.36)$$

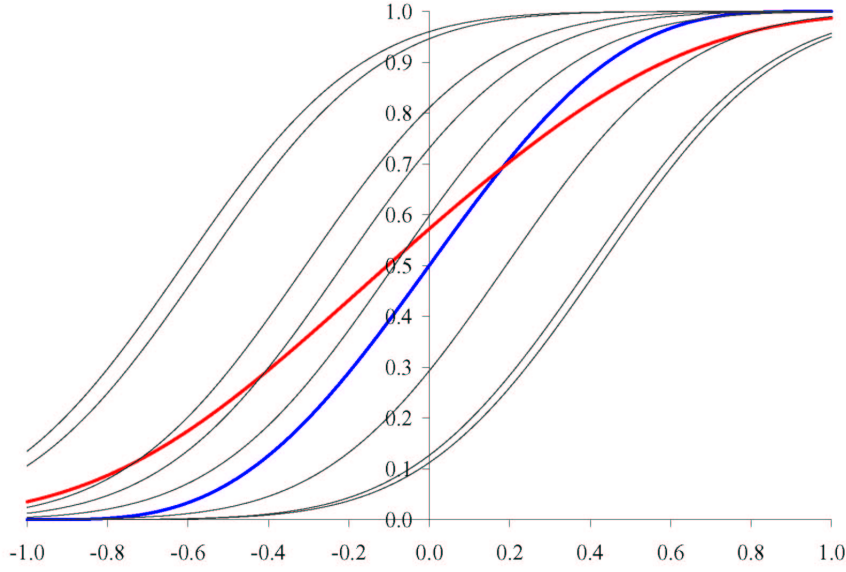


Figure 3.4: *Kernel distribution estimator with $K = \Phi$, and $h = 0.35$.*

Recall from Section 1.4.2 that the variance term of F_n is given by

$$\text{Var}\{F_n(x)\} = \frac{1}{n}F(x)[1 - F(x)].$$

By comparing this expression with (3.35) we see that $\hat{F}_{n,h}$ has a smaller variance term, since $C_1 > 0$. From this we also see that the variance term can be made smaller by selecting a larger value of h , but this will in turn inflate the bias term (which is $o(h^3)$) so we are once again faced with the trade-off concept in the smoothing parameter selection problem.

Combining these bias and variance results to obtain the asymptotic mean squared error and integrating that function over the entire domain of x leads to the following expression for the asymptotic mean integrated squared error:

$$AMISE(h) = \frac{1}{n} \int_0^\infty F(x)[1 - F(x)]dx - \frac{2h}{n}C_1 + \frac{1}{4}h^4\mu_2(k)^2R(f'). \quad (3.37)$$

Once again we need to determine the value of h that will minimize this AMISE criterion. We obtain the asymptotically optimal bandwidth by differentiating the AMISE function with regards to h and setting the resulting function equal to zero. The value of h that satisfies this equality is

$$h_{AMISE} = \left[\frac{2C_1}{\mu_2(k)^2R(f')} \right]^{\frac{1}{3}} n^{-1/3}. \quad (3.38)$$

Example of kernel hazard function estimation

For this example, $n = 8$ observations were generated from the $\text{Wei}(1.25, 1)$ distribution. The sorted generated sample is given by:

0.030	0.561	0.597	0.808	1.069	1.151	1.845	2.168
-------	-------	-------	-------	-------	-------	-------	-------

Figure 3.5 includes the target hazard function λ as the blue line, the kernel estimator $\hat{\lambda}_{n,h}$ defined in (3.24) as the red line, as well as the rescaled kernels placed on each of the ordered sample values. From this illustration one can clearly see how the kernel estimate is calculated as the sum of all n of the rescaled kernel values in each point x . We will now discuss some

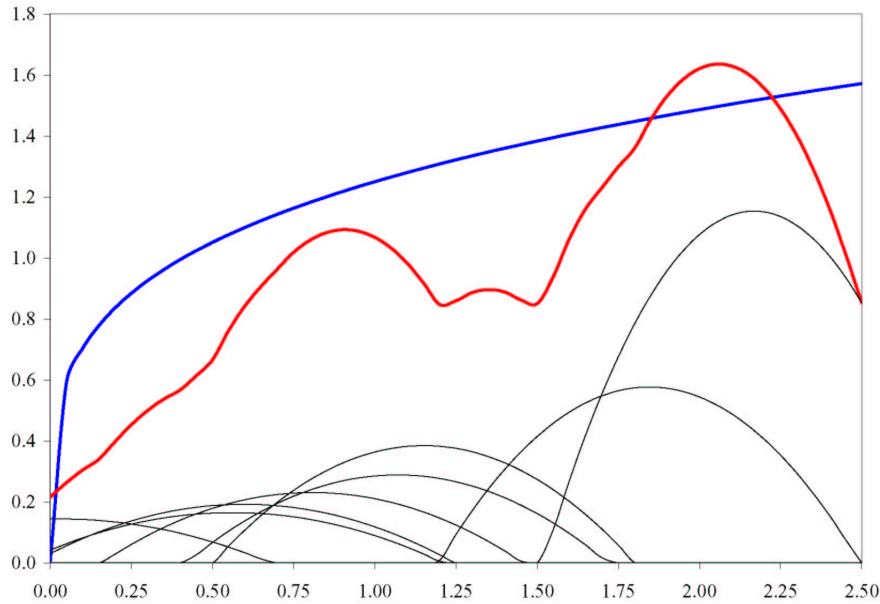


Figure 3.5: *Kernel hazard function estimator with $k = \phi$, and $h = 0.50$.*

asymptotic properties of $\hat{\lambda}_{n,h}$ as an estimator of λ , using results stated in Rice and Rosenblatt (1976). The asymptotic bias term is given by

$$\text{Bias}(\hat{\lambda}_{n,h}^{(2)}(x)) = \frac{h^2}{2} \lambda''(x) \mu_2(k) + o(h^2), \quad (3.39)$$

and

$$\text{Var}(\hat{\lambda}_{n,h}^{(2)}(x)) = \frac{1}{nh} \frac{\lambda(x)}{1 - F(x)} R(k) + o(n^{-1}h^{-1}). \quad (3.40)$$

Recall the decomposition of MSE derived in (3.4). Once again the trade-off concept (see Section 3.1.6) with regard to h can be seen more clearly. From (3.39) we see that the

asymptotic squared bias term is *increasing* in h , and is of order $o(h^4)$. In contrast, we note by inspecting (3.40) that the asymptotic variance term is *decreasing* in h , and is of order $o(n^{-1}h^{-1})$. The problem at hand is to find a value of h that balances these two opposing measures in order to minimize the MSE.

Now, since we have an expression for the asymptotic MSE given by

$$MSE(\hat{\lambda}_{n,h}^{(2)}(x)) = \{Bias(\hat{\lambda}_{n,h}^{(2)}(x))\}^2 + Var(\hat{\lambda}_{n,h}^{(2)}(x)), \quad (3.41)$$

one can integrate this expression to obtain an AMISE expression, that is an asymptotic global discrepancy measure. Define this measure as:

$$AMISE(h) = \frac{h^4}{4}\mu_2(k)^2 \int_0^\infty [\lambda''(x)]^2 dx + \frac{1}{nh}R(k) \int_0^\infty \frac{\lambda(x)}{1-F(x)} dx. \quad (3.42)$$

It is here that *weight functions* must be introduced. Upon inspection of the integrals noted in (3.42), it becomes evident that these integrals do not have finite solutions, since their values tend to infinity. Therefore, in order to obtain a finite value for the AMISE, a weight function should be introduced in such a way that the integrals deliver finite values. Thus, define the *weighted* AMISE as:

$$AMISE_w(h) = \frac{h^4}{4}\mu_2(k)^2 \int_0^\infty [\lambda''(x)]^2 w(x) dx + \frac{1}{nh}R(k) \int_0^\infty \frac{\lambda(x)}{1-F(x)} w(x) dx. \quad (3.43)$$

For an example of a typical weight function, see Section 3.1.5. Now, this $AMISE_w$ function can be used to obtain an asymptotic optimal choice of h by calculating its derivative to h and setting the resulting function equal to zero. Solving this expression for h results in the following asymptotically optimal choice of h :

$$h_{AMISE_w} = \left[\frac{R(k)B_{F,w}}{\mu_2(k)^2 A_{F,w}} \right]^{\frac{1}{5}} n^{-\frac{1}{5}}, \quad (3.44)$$

where $A_{F,w} = \int_0^\infty [\lambda''(x)]^2 w(x) dx$ and $B_{F,w} = \int_0^\infty \frac{\lambda(x)}{1-F(x)} w(x) dx$.

Chapter 4

Smoothers and automatic bandwidth selectors

4.1 Introduction

Note that the discussion below will only involve certain kernel smoothers for the hazard function. For other nonparametric estimators, the reader is reminded of the references thereof given in Chapter 3. In this chapter we will discuss two kernel hazard rate estimators as well as their extensions to the censored case. Thereafter automatic bandwidth selectors and their theoretical properties will be discussed.

4.2 Uncensored case

In this section two estimators, first introduced by Watson and Leadbetter (1964a) will be discussed. The emphasis will be on the asymptotic bias and variance terms of these estimators, more specifically their dependence on the smoothing parameter h . Two estimators will be discussed, the so-called *internal* and *external* estimators, terms introduced by Jones, Davies and Park (1994). More detail about the results discussed in this section can be found in Watson and Leadbetter (1964a), Watson and Leadbetter (1964b) and Rice and Rosenblatt (1976).

4.2.1 External estimator

This estimator estimates the two functions that make up the hazard rate as defined in (1.1) separately, and the hazard rate estimate is defined as the ratio of the two estimates. It is defined by

$$\hat{\lambda}_{n,h}^{(1)}(x) = \frac{\hat{f}_{n,h}(x)}{1 - F_n(x)}, \quad (4.1)$$

where $\hat{f}_{n,h}$ and F_n are defined in (3.19) and (1.17) respectively.

Asymptotic bias and variance

Under conditions specified in Rice and Rosenblatt (1976), the asymptotic bias of $\hat{\lambda}_{n,h}^{(1)}$ is given by

$$\text{Bias}(\hat{\lambda}_{n,h}^{(1)}(x)) = \frac{h^2}{2} \frac{f''(x)}{1 - F(x)} \mu_2(k) + o(h^2), \quad (4.2)$$

with μ_2 defined in (3.28). The asymptotic variance of $\hat{\lambda}_{n,h}^{(1)}$ is

$$\text{Var}(\hat{\lambda}_{n,h}^{(1)}(x)) = \frac{1}{nh} \frac{\lambda(x)}{1 - F(x)} R(k) + o(n^{-1}h^{-1}), \quad (4.3)$$

with R defined in (3.30). One can clearly see the conflicting effect of h on the magnitude of the bias and variance terms, as discussed in Section 3.1.6. The discussion of $\hat{\lambda}_{n,h}^{(1)}$ will end here, since most of our attention will be focused on the internal estimator.

Remarks:

1. $\hat{\lambda}_{n,h}^{(1)}$ is *asymptotically unbiased*.
2. $\hat{\lambda}_{n,h}^{(1)}$ is a *consistent* estimator of the hazard function, i.e. $\text{var}[\hat{\lambda}_{n,h}^{(1)}(x)] \rightarrow 0$ as $n \rightarrow \infty$.

4.2.2 Internal estimator

In contrast to $\hat{\lambda}_{n,h}^{(1)}$, the internal estimator attempts to estimate the ratio that defines the hazard rate as a whole, that is as one entity rather than two separate parts. We will provide the following heuristic derivation of the estimator:

Let ψ be a continuous function with a number of derivatives, and let k_h be the rescaled kernel function. It follows that

$$\begin{aligned} (k_h * \psi)(x) &= \int_{-\infty}^{\infty} k_h(x - y) \psi(y) dy \\ &= \frac{1}{h} \int_{-\infty}^{\infty} k\left(\frac{x - y}{h}\right) \psi(y) dy \\ &= \int_{-\infty}^{\infty} k(z) \psi(x - zh) dz \quad (\text{where } z = \frac{x - y}{h}) \\ &= \int_{-\infty}^{\infty} k(z) \left\{ \psi(x) - hz\psi'(x) + \frac{1}{2}h^2z^2\psi''(x) + \dots \right\} dz \\ &= \psi(x) \int_{-\infty}^{\infty} k(z) dz \\ &= \psi(x), \text{ as } h \rightarrow 0. \end{aligned}$$

Choosing $\psi(y) = \frac{f(y)}{1 - F(y)}$ we have

$$\int_{-\infty}^{\infty} k_h(x - y) \frac{f(y)}{1 - F(y)} dy = \int_{-\infty}^{\infty} k_h(x - y) \frac{dF(y)}{1 - F(y)}.$$

Plugging in the EDF as an approximation to F results in

$$\int_{-\infty}^{\infty} k_h(x - y) \frac{dF_n(y)}{1 - F_n(y)} \approx \frac{1}{n} \sum_{i=1}^n \frac{k_h(x - X_i)}{1 - F_n(X_i)},$$

which can be rewritten as

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n \frac{k_h(x - X_i)}{1 - F_n(X_i)} &= \frac{1}{n} \sum_{i=1}^n \frac{k_h(x - X_{(i)})}{1 - F_n(X_{(i)})} \\ &= \frac{1}{n} \sum_{i=1}^n \frac{k_h(x - X_{(i)})}{1 - \frac{i}{n}} \\ &= \sum_{i=1}^n \frac{k_h(x - X_{(i)})}{n - i} \\ &\approx \sum_{i=1}^n \frac{k_h(x - X_{(i)})}{n - i + 1}, \end{aligned}$$

where an adjustment is made to the denominator to avoid division by zero, and

$X_{(1)}, X_{(2)}, \dots, X_{(n)}$ are the order statistics of the sample. We can now define the estimator as

$$\hat{\lambda}_{n,h}^{(2)}(x) = \sum_{i=1}^n \frac{k_h(x - X_{(i)})}{n - i + 1}. \quad (4.4)$$

Remarks:

1. $\hat{\lambda}_{n,h}^{(2)}$ is *asymptotically unbiased*.
2. $\hat{\lambda}_{n,h}^{(2)}$ is a *consistent* estimator of the hazard function.
3. Note that $\hat{\lambda}_{n,h}^{(2)}$ has the same asymptotic variance as $\hat{\lambda}_{n,h}^{(1)}$ (see (4.3) and (3.40)), but there is a difference in the asymptotic bias expressions given by (4.2) and (3.39). Upon investigation of these expressions we see that neither estimator is uniformly superior to the other, as the performance depends on the underlying distribution as well as the point of estimation.
4. The notation $*$ is used to indicate the *convolution* of two functions, and is defined by

$$(f * g)(x) = \int_{-\infty}^{\infty} f(x - y)g(y)dy \quad (4.5)$$

for functions f and g .

4.2.3 An explicit expression for bootstrap MSE in the case of $\hat{\lambda}_{n,h}^{*(2)}$

In this section, explicit expressions for the bias and variance terms of $\hat{\lambda}_{n,h}^{*(2)}$ will be derived. Note that this derivation is based on the naive bootstrap, that is where the bootstrap samples $X_1^*, X_2^*, \dots, X_n^*$ are drawn from F_n , the EDF of the sample $X_1, X_2, \dots, X_n$ defined in (1.17). The bootstrap estimator of (4.4) is denoted by $\hat{\lambda}_{n,h}^{*(2)}$, and is defined as

$$\hat{\lambda}_{n,h}^{*(2)}(x) = \sum_{i=1}^n \frac{k_h(x - X_{(i)}^*)}{n - i + 1}, \quad (4.6)$$

where $X_{(1)}^*, X_{(2)}^*, \dots, X_{(n)}^*$ are the order statistics of the bootstrap sample $X_1^*, X_2^*, \dots, X_n^*$. The bootstrap mean squared error in a fixed point x can be defined as

$$MSE_*(\hat{\lambda}_{n,h}^{*(2)}(x)) = \left[Bias_*(\hat{\lambda}_{n,h}^{*(2)}(x)) \right]^2 + Var_*(\hat{\lambda}_{n,h}^{*(2)}(x)),$$

where

$$Bias_*(\hat{\lambda}_{n,h}^{*(2)}(x)) = E_*(\hat{\lambda}_{n,h}^{*(2)}(x)) - \hat{\lambda}_{n,h}^{(2)}(x). \quad (4.7)$$

It now follows that

$$E_*(\hat{\lambda}_{n,h}^{*(2)}(x)) = E_* \left(\sum_{i=1}^n \frac{k_h(x - X_{(i)}^*)}{n - i + 1} \right) \quad (4.8)$$

$$= \sum_{i=1}^n \frac{E_*(k_h(x - X_{(i)}^*))}{n - i + 1} \quad (4.9)$$

$$= \sum_{i=1}^n \sum_{j=1}^n \frac{k_h(x - X_{(j)})}{n - i + 1} P^*(X_{(i)}^* = X_{(j)}). \quad (4.10)$$

For the variance term we have

$$Var_*(\hat{\lambda}_{n,h}^{*(2)}(x)) = E_* \left[(\hat{\lambda}_{n,h}^{*(2)}(x))^2 \right] - \left[E_*(\hat{\lambda}_{n,h}^{*(2)}(x)) \right]^2, \quad (4.11)$$

where

$$E_* \left[(\hat{\lambda}_{n,h}^{*(2)}(x))^2 \right] = \sum_{i=1}^n \frac{E_*(k_h^2(x - X_{(i)}^*))}{(n - i + 1)^2} + 2E_* \left(\sum_{i < j} \frac{k_h(x - X_{(i)}^*)k_h(x - X_{(j)}^*)}{(n - i + 1)(n - j + 1)} \right). \quad (4.12)$$

In order to calculate the MSE expression explicitly, we must now find expressions for

$$P^*(X_{(i)}^* = X_{(j)}) \text{ in (4.10) and } 2E_* \left(\sum_{i < j} \frac{k_h(x - X_{(i)}^*)k_h(x - X_{(j)}^*)}{(n - i + 1)(n - j + 1)} \right) \text{ in (4.12).}$$

Lemma 1. Let $X_1, X_2, \dots, X_n$ be i.i.d. random variables (continuous or discrete). Let $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ denote their order statistics, and analogously let $X_{(1)}^*, X_{(2)}^*, \dots, X_{(n)}^*$ denote the order statistics of the bootstrap sample. Then

$$P^*(X_{(i)}^* = X_{(j)}) = P(\text{Bin}(n, (j-1)/n) \leq i-1) - P(\text{Bin}(n, j/n) \leq i-1),$$

where $\text{Bin}(n, p)$ denotes a binomial random variable with parameters n and p .

Proof:

$$\begin{aligned} P^*(X_{(i)}^* = X_{(j)}) &= P^*(X_{(i)}^* > X_{(j-1)}) - P^*(X_{(i)}^* > X_{(j)}) \\ &= P^*\left(\sum_{k=1}^n I(X_k^* \leq X_{(j-1)}) \leq i-1\right) - P^*\left(\sum_{k=1}^n I(X_k^* \leq X_{(j)}) \leq i-1\right) \\ &= P\left(\text{Bin}\left(n, \frac{j-1}{n}\right) \leq i-1\right) - P\left(\text{Bin}\left(n, \frac{j}{n}\right) \leq i-1\right) \end{aligned}$$

Q.E.D.

Lemma 2. Let $X_1, X_2, \dots, X_n$ be i.i.d. random variables (continuous or discrete) with distribution function F . Denote the joint distribution function of $X_{(i)}$ and $X_{(j)}$,

$1 \leq i < j \leq n$ by

$$F_{ij}(x, y) := P(X_{(i)} \leq x, X_{(j)} \leq y)$$

Then

$$F_{ij}(x, y) = \begin{cases} \sum_{k=j}^n \sum_{m=i}^k \frac{n!}{m!(k-m)!(n-k)!} [F(x)]^m [F(y) - F(x)]^{k-m} [1 - F(y)]^{n-k}, & \text{if } x < y \\ \sum_{k=j}^n \binom{n}{k} [F(y)]^k [1 - F(y)]^{n-k}, & \text{if } x \geq y \end{cases}$$

Proof:

For $x < y$ we have:

$$\begin{aligned}
 F_{ij}(x, y) &= P(X_{(i)} \leq x, X_{(j)} \leq y) \\
 &= P(\text{At least } i \text{ of the } X's \text{ are } \leq x, \text{ at least } j \text{ of the } X's \text{ are } \leq y) \\
 &= \sum_{k=j}^n \sum_{m=i}^k P(\text{Exactly } m \text{ of the } X's \text{ are } \leq x, \text{ exactly } k \text{ of the } X's \text{ are } \leq y) \\
 &= \sum_{k=j}^n \sum_{m=i}^k \frac{n!}{m!(k-m)!(n-k)!} [F(x)]^m [F(y) - F(x)]^{k-m} [1 - F(y)]^{n-k}
 \end{aligned}$$

Also, for $x \geq y$ the inequality $X_{(j)} \leq y$ implies $X_{(i)} \leq x$, so that

$$F_{ij}(x, y) = P(X_{(j)} \leq y) = \sum_{k=j}^n \binom{n}{k} [F(y)]^k [1 - F(y)]^{n-k}$$

Q.E.D.

It follows from **Lemma 1** and equation (4.10) that

$$\begin{aligned}
 E_*(\hat{\lambda}_{n,h}^{*(2)}(x)) &= \sum_{i=1}^n \sum_{j=1}^n \frac{k_h(x - X_{(j)})}{n - i + 1} \left[P\left(\text{Bin}\left(n, \frac{j-1}{n}\right) \leq i-1\right) - P\left(\text{Bin}\left(n, \frac{j}{n}\right) \leq i-1\right) \right] \\
 &= \sum_{i=1}^n \sum_{j=1}^n \frac{k_h(x - X_{(j)})}{n - i + 1} \left[\sum_{k=0}^{i-1} \binom{n}{k} \left\{ \left(\frac{j-1}{n}\right)^k \left(1 - \frac{j-1}{n}\right)^{n-k} - \left(\frac{j}{n}\right)^k \left(1 - \frac{j}{n}\right)^{n-k} \right\} \right]
 \end{aligned}$$

The bootstrap bias term of $\hat{\lambda}_{n,h}^{*(2)}$ is then given by

$$\begin{aligned}
 \text{Bias}_*(\hat{\lambda}_{n,h}^{*(2)}(x)) &= E_*(\hat{\lambda}_{n,h}^{*(2)}(x)) - \hat{\lambda}_{n,h}^{(2)}(x) \\
 &= \sum_{i=1}^n \sum_{j=1}^n \frac{k_h(x - X_{(j)})}{n - i + 1} \left[\sum_{k=0}^{i-1} \binom{n}{k} \left\{ \left(\frac{j-1}{n}\right)^k \left(1 - \frac{j-1}{n}\right)^{n-k} - \left(\frac{j}{n}\right)^k \left(1 - \frac{j}{n}\right)^{n-k} \right\} \right] \\
 &\quad - \sum_{i=1}^n \frac{k_h(x - X_{(i)})}{n - i + 1}.
 \end{aligned} \tag{4.13}$$

Lemma 3. Let $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ denote the order statistics of the sample of random variables $X_1, X_2, \dots, X_n$, and let $X_{(1)}^*, X_{(2)}^*, \dots, X_{(n)}^*$ denote the order statistics of the bootstrap sample $X_1^*, X_2^*, \dots, X_n^*$. Then

$$\begin{aligned} E_* \left(2 \sum_{i < j} \sum \frac{k_h(x - X_{(i)}^*) k_h(x - X_{(j)}^*)}{(n - i + 1)(n - j + 1)} \right) \\ = 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^n \sum_{r=1}^n \frac{k_h^2(x - X_{(r)})}{(n - i + 1)(n - j + 1)} A_{i,j,r}(n) \\ + 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^n \sum_{r=1}^{n-1} \frac{k_h(x - X_{(r)}) k_h(x - X_{(r+1)})}{(n - i + 1)(n - j + 1)} B_{i,j,r}(n) \\ + 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^n \sum_{r=1}^{n-2} \sum_{s=r+2}^n \frac{k_h(x - X_{(r)}) k_h(x - X_{(s)})}{(n - i + 1)(n - j + 1)} C_{i,j,r,s}(n), \end{aligned}$$

where

$$\begin{aligned} A_{i,j,r}(n) &:= P^*(X_{(i)}^* = X_{(r)}, X_{(j)}^* = X_{(r)}), \quad i < j, \\ B_{i,j,r}(n) &:= P^*(X_{(i)}^* = X_{(r)}, X_{(j)}^* = X_{(r+1)}), \quad i < j, \quad r < r + 1, \\ C_{i,j,r,s}(n) &:= P^*(X_{(i)}^* = X_{(r)}, X_{(j)}^* = X_{(s)}), \quad i < j, \quad r < s + 1. \end{aligned}$$

Proof:

$$\begin{aligned} E_* \left(2 \sum_{i < j} \sum \frac{k_h(x - X_{(i)}^*) k_h(x - X_{(j)}^*)}{(n - i + 1)(n - j + 1)} \right) \\ = 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^n \frac{E_* \left(k_h(x - X_{(i)}^*) k_h(x - X_{(j)}^*) \right)}{(n - i + 1)(n - j + 1)} \\ = 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^n \sum_{r=1}^n \sum_{s=r}^n \frac{k_h(x - X_{(r)}) k_h(x - X_{(s)})}{(n - i + 1)(n - j + 1)} P^*(X_{(i)}^* = X_{(r)}, X_{(j)}^* = X_{(s)}) \\ = 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^n \sum_{r=1}^n \frac{k_h^2(x - X_{(r)})}{(n - i + 1)(n - j + 1)} P^*(X_{(i)}^* = X_{(r)}, X_{(j)}^* = X_{(r)}) \\ + 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^n \sum_{r=1}^{n-1} \frac{k_h(x - X_{(r)}) k_h(x - X_{(r+1)})}{(n - i + 1)(n - j + 1)} P^*(X_{(i)}^* = X_{(r)}, X_{(j)}^* = X_{(r+1)}) \\ + 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^n \sum_{r=1}^{n-2} \sum_{s=r+2}^n \frac{k_h(x - X_{(r)}) k_h(x - X_{(s)})}{(n - i + 1)(n - j + 1)} P^*(X_{(i)}^* = X_{(r)}, X_{(j)}^* = X_{(s)}) \end{aligned} \tag{4.14}$$

Note that for $i < j$ we have:

$$\begin{aligned} P^*(X_{(i)}^* = X_{(r)}, X_{(j)}^* = X_{(r)}) \\ = P^*(X_{(i)}^* \leq X_{(r)}, X_{(j)}^* \leq X_{(r)}) - P^*(X_{(i)}^* \leq X_{(r-1)}, X_{(j)}^* \leq X_{(r)}) \\ = P^*(X_{(j)}^* \leq X_{(r)}) - P^*(X_{(i)}^* \leq X_{(r-1)}, X_{(j)}^* \leq X_{(r)}). \end{aligned}$$

Note that $P^*(X_{(i)}^* \leq X_{(r)}) = F(X_{(r)}) = F_n(X_{(r)}) = \frac{r}{n}$. By applying **Lemma 2** we obtain

$$\begin{aligned}
& P^*(X_{(j)}^* \leq X_{(r)}) - P^*(X_{(i)}^* \leq X_{(r-1)}, X_{(j)}^* \leq X_{(r)}) \\
&= \sum_{k=j}^n \binom{n}{k} [F_n(X_{(r)})]^k [1 - F_n(X_{(r)})]^{n-k} \\
&\quad - \sum_{k=j}^n \sum_{m=i}^k \frac{n!}{m!(k-m)!(n-k)!} [F_n(X_{(r-1)})]^m [F_n(X_{(r)}) - F_n(X_{(r-1)})]^{k-m} [1 - F_n(X_{(r)})]^{n-k} \\
&= \sum_{k=j}^n \binom{n}{k} \left[\frac{r}{n} \right]^k \left[1 - \frac{r}{n} \right]^{n-k} - \sum_{k=j}^n \sum_{m=i}^k \frac{n!}{m!(k-m)!(n-k)!} \left[\frac{r-1}{n} \right]^m \left[\frac{1}{n} \right]^{k-m} \left[1 - \frac{r}{n} \right]^{n-k} \\
&:= A_{i,j,r}(n)
\end{aligned} \tag{4.15}$$

Observe that for $i < j$ and $r < r+1$ we obtain, by applying **Lemma 2** :

$$\begin{aligned}
& P^*(X_{(i)}^* = X_{(r)}, X_{(j)}^* = X_{(r+1)}) \\
&= P^*(X_{(i)}^* \leq X_{(r)}, X_{(j)}^* \leq X_{(r+1)}) - P^*(X_{(i)}^* \leq X_{(r-1)}, X_{(j)}^* \leq X_{(r+1)}) \\
&\quad - P^*(X_{(i)}^* \leq X_{(r)}, X_{(j)}^* \leq X_{(r)}) + P^*(X_{(i)}^* \leq X_{(r-1)}, X_{(j)}^* \leq X_{(r)}) \\
&= \sum_{k=j}^n \sum_{m=i}^k \frac{n!}{m!(k-m)!(n-k)!} \left(\left[\frac{r}{n} \right]^m \left[\frac{1}{n} \right]^{k-m} \left[1 - \frac{r+1}{n} \right]^{n-k} \right. \\
&\quad \left. - \left[\frac{r-1}{n} \right]^m \left[\frac{2}{n} \right]^{k-m} \left[1 - \frac{r+1}{n} \right]^{n-k} + \left[\frac{r-1}{n} \right]^m \left[\frac{1}{n} \right]^{k-m} \left[1 - \frac{r}{n} \right]^{n-k} \right) \\
&\quad - \sum_{k=j}^n \binom{n}{k} \left[\frac{r}{n} \right]^k \left[1 - \frac{r}{n} \right]^{n-k} \\
&:= B_{i,j,r}(n)
\end{aligned} \tag{4.16}$$

Note that for $i < j$ and $r < s+1$ we obtain, by applying **Lemma 2** :

$$\begin{aligned}
& P^*(X_{(i)}^* = X_{(r)}, X_{(j)}^* = X_{(s)}) \\
&= P^*(X_{(i)}^* \leq X_{(r)}, X_{(j)}^* \leq X_{(s)}) - P^*(X_{(i)}^* \leq X_{(r-1)}, X_{(j)}^* \leq X_{(s)}) \\
&\quad - P^*(X_{(i)}^* \leq X_{(r)}, X_{(j)}^* \leq X_{(s-1)}) + P^*(X_{(i)}^* \leq X_{(r-1)}, X_{(j)}^* \leq X_{(s-1)}) \\
&= \sum_{k=j}^n \sum_{m=i}^k \frac{n!}{m!(k-m)!(n-k)!} \left(\left[\frac{r}{n} \right]^m \left[\frac{s-r}{n} \right]^{k-m} \left[1 - \frac{s}{n} \right]^{n-k} \right. \\
&\quad \left. - \left[\frac{r-1}{n} \right]^m \left[\frac{s-(r-1)}{n} \right]^{k-m} \left[1 - \frac{s}{n} \right]^{n-k} - \left[\frac{r}{n} \right]^m \left[\frac{(s-1)-r}{n} \right]^{k-m} \left[1 - \frac{(s-1)}{n} \right]^{n-k} \right. \\
&\quad \left. + \left[\frac{r-1}{n} \right]^m \left[\frac{(s-1)-(r-1)}{n} \right]^{k-m} \left[1 - \frac{s-1}{n} \right]^{n-k} \right) \\
&:= C_{i,j,r,s}(n)
\end{aligned} \tag{4.17}$$

Q.E.D.

Using **Lemma 1**, we can obtain an expression for the first term on the right hand side of (4.12) as

$$\begin{aligned} & \sum_{i=1}^n \frac{E_* \left(k_h^2(x - X_{(i)}^*) \right)}{(n-i+1)^2} \\ &= \sum_{i=1}^n \sum_{j=1}^n \frac{k_h^2(x - X_{(j)})}{(n-i+1)^2} \left[\sum_{k=0}^{i-1} \binom{n}{k} \left\{ \left(\frac{j-1}{n} \right)^k \left(1 - \frac{j-1}{n} \right)^{n-k} - \left(\frac{j}{n} \right)^k \left(1 - \frac{j}{n} \right)^{n-k} \right\} \right] \end{aligned}$$

The variance term follows from the above and **Lemma 3**, and is given by

$$\begin{aligned} & Var_*(\hat{\lambda}_{n,h}^{*(2)}(x)) \\ &= \sum_{i=1}^n \sum_{j=1}^n \frac{k_h^2(x - X_{(j)})}{(n-i+1)^2} \left[\sum_{k=0}^{i-1} \binom{n}{k} \left\{ \left(\frac{j-1}{n} \right)^k \left(1 - \frac{j-1}{n} \right)^{n-k} - \left(\frac{j}{n} \right)^k \left(1 - \frac{j}{n} \right)^{n-k} \right\} \right] \\ &+ 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^n \sum_{r=1}^n \frac{k_h^2(x - X_{(r)})}{(n-i+1)(n-j+1)} A_{i,j,r}(n) \\ &+ 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^n \sum_{r=1}^{n-1} \frac{k_h(x - X_{(r)})k_h(x - X_{(r+1)})}{(n-i+1)(n-j+1)} B_{i,j,r}(n) \\ &+ 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^n \sum_{r=1}^{n-2} \sum_{s=r+2}^n \frac{k_h(x - X_{(r)})k_h(x - X_{(s)})}{(n-i+1)(n-j+1)} C_{i,j,r,s}(n) \\ &- \left[E_*(\hat{\lambda}_{n,h}^{*(2)}(x)) \right]^2 \end{aligned} \tag{4.18}$$

Q.E.D.

We can now state the following:

Theorem 2. *Under the conditions of Lemma 1,2 and 3 we have that*

$$MSE_*(\hat{\lambda}_{n,h}^{*(2)}(x)) = \left[Bias_*(\hat{\lambda}_{n,h}^{*(2)}(x)) \right]^2 + Var_*(\hat{\lambda}_{n,h}^{*(2)}(x)),$$

with $Bias_*(\hat{\lambda}_{n,h}^{*(2)}(x))$ defined in (4.13) and $Var_*(\hat{\lambda}_{n,h}^{*(2)}(x))$ defined in (4.18).

Proof:

Using results from **Lemma 1**, **Lemma 2** and **Lemma 3** the proof follows.

Q.E.D.

Remarks:

1. Obtaining an explicit expression for the naive bootstrap MSE as above is very desirable, as it does not require any resampling to take place.
2. Unfortunately, as has been described in Section 2.3.1, the bootstrap bias is approximately zero if the naive bootstrap is used. The derived expression therefore has limited applicability for choosing the smoothing parameter.

3. In the special case where $j = 1$, the 0^0 term was treated as the value 1. In this specific context of a binomial probability, the value 0^0 is equal to the probability of obtaining 0 successes from n trials, where the probability of success is equal to 0, and that probability must be 1.

Tables 4.1 and 4.2 show results from a simple simulation. For this illustration, $n = 20$ values were drawn from a $\text{Exp}(1)$ distribution. Discrepancy measures for the kernel hazard function estimator $\hat{\lambda}_{n,h}^{(2)}$ were calculated in a single point. The setting details are as follows: $n = 20$, $x = 1.221$, $h = 0.4975$ and $k = \phi$. We denote the squared bias term by Bias^2 .

Measure	$B = 100$	$B = 10\ 000$	$B = 1\ 000\ 000$	$B = 10\ 000\ 000$	Explicit
Bias^2	0.006469	0.006728	0.006592	0.006547	0.006552
Variance	0.08040	0.09602	0.09529	0.09514	0.09517
MSE	0.08687	0.10275	0.10188	0.10169	0.10173

Table 4.1: *Convergence of empirical bootstrap to explicit bootstrap as $B \rightarrow \infty$.*

Measure	$\frac{\text{Explicit}}{B=100} \cdot 100\%$	$\frac{\text{Explicit}}{B=10\ 000} \cdot 100\%$	$\frac{\text{Explicit}}{B=1\ 000\ 000} \cdot 100\%$	$\frac{\text{Explicit}}{B=10\ 000\ 000} \cdot 100\%$	$\frac{\text{Explicit}}{\text{Explicit}} \cdot 100\%$
Bias^2	101.28%	97.38%	99.39%	100.09%	100.00%
Variance	118.37%	99.12%	99.87%	100.04%	100.00%
MSE	117.10%	99.00%	99.84%	100.05%	100.00%

Table 4.2: *Convergence expressed in percentages.*

From the results it is clear that the theoretical expressions are correct, as the independent empirical approximations tend to the explicit values as $B \rightarrow \infty$.

4.3 Random right censoring

The two estimators discussed in Section 4.2 can be extended to the present setting and will now be discussed. The reader is reminded of the definition of random right censoring given in Definition 1.6. Also note that the parameter of interest is the hazard function of the true underlying survival function, given by

$$\lambda^0(x) = \frac{f^0(x)}{1 - F^0(x)}. \quad (4.19)$$

4.3.1 External estimator

The external estimator in the censored setting is defined as

$$\hat{\lambda}_{n,h}^{(3)}(x) = \sum_{i=1}^n \frac{\hat{f}_{n,h}(x) \Delta_i}{1 - F_n(x)} \quad (4.20)$$

where $\hat{f}_{n,h}$ and F_n are defined in (3.19) and (1.17) respectively, and the indicator function $\{\Delta_i\}_{i=1}^n$ as defined previously. The interested reader is referred to Kim, Bae, Choi and Park (2005), who propose an external estimator that involves using the Kaplan-Meier function to create smooth estimators of f^0 , F^0 and consequently λ^0 .

4.3.2 Internal estimator

The internal estimator in the uncensored case as in (4.4) follows from the heuristic derivation given in Section 4.2.2. Gonzalez-Manteiga et al. (1996) state that this estimator can be extended to the censored setting by replacing the EDF with the Kaplan-Meier estimator as defined in Section 1.4.3. The resulting estimator is defined as

$$\hat{\lambda}_{n,h}^{(4)}(x) = \sum_{i=1}^n \frac{k_h(x - X_{(i)}) \Delta_i}{n - i + 1}. \quad (4.21)$$

Tanner and Wong (1983) generalized the results stated in Watson and Leadbetter (1964b) for the estimator $\hat{\lambda}_{n,h}^{(2)}$ in the uncensored case to the censored case (that is for $\hat{\lambda}_{n,h}^{(4)}(x)$). We define the concept of *compatibility* before important assumptions from Tanner and Wong (1983) are given:

Definition 4.1. *The kernel function k is said to be compatible with a distribution function F if for any $M > 0$, there exists a h small enough such that $k_h(x - y)[1 - F(y)]^{-1}$ is uniformly bounded for $|x - y| > M$.*

Assumption 1: The kernel k is a symmetric, nonnegative function, $k(x) = o(x^{-1})$ as $x \rightarrow \infty$, and $\int k(x)dx = 1$.

Assumption 2: The kernel k is *compatible* with both F^0 and H^0 .

Note that *Assumption 1* is often used in kernel estimation, but *Assumption 2* is needed in the censored case. By imposing this compatibility condition on the kernel function k the following theorem is proved:

Theorem 3. *Let $n \rightarrow \infty$, $h \rightarrow 0$ and $nh \rightarrow \infty$;*

(a) If k is compatible with F^0 , then $E(\hat{\lambda}_{n,h}^{(4)}(x)) \rightarrow \lambda^0(x)$;

(b) If k is compatible with both H^0 and F^0 , then

$$\text{Var}(\hat{\lambda}_{n,h}^{(4)}(x)) = \frac{1}{nh} \frac{\lambda^0(x)}{1 - F(x)} R(k) + o(n^{-1}h^{-1}),$$

with R as in (3.30).

Remarks:

1. As a consequence of this theorem, $\hat{\lambda}_{n,h}^{(4)}$ is an *asymptotically unbiased* estimator of λ^0 .
2. $\hat{\lambda}_{n,h}^{(4)}$ is a *consistent* estimator.
3. Tanner and Wong (1983) proceed to prove the *asymptotic normality* of $\hat{\lambda}_{n,h}^{(4)}$, by making use of the Hajek projection method.

Motivation for using the internal estimator

Patil, Wells and Marron (1994) compare the general form of the estimators $\hat{\lambda}_{n,h}^{(3)}$ and $\hat{\lambda}_{n,h}^{(4)}$. We have already noted that in the uncensored case the two estimators $\hat{\lambda}_{n,h}^{(1)}$ and $\hat{\lambda}_{n,h}^{(2)}$ have the same asymptotic variance, but different asymptotic bias expressions. Patil et al. (1994) generalize this property to the setting of random right censoring, and provide expressions for the asymptotic variance term as well as for the two different asymptotic bias terms.

Two arguments are put forth to show that the estimator $\hat{\lambda}_{n,h}^{(4)}$ is more natural. The first is that $\hat{\lambda}_{n,h}^{(4)}$ admits a useful and simple type of martingale representation (which $\hat{\lambda}_{n,h}^{(3)}$ does not). Secondly, the so-called “effective kernel” of $\hat{\lambda}_{n,h}^{(4)}$ is smooth whilst the “effective kernel” of $\hat{\lambda}_{n,h}^{(3)}$ has undesirable discontinuities. The interested reader is referred to Patil et al. (1994) to get a more thorough discussion of these concepts. As a result of these arguments and its properties, the estimator $\hat{\lambda}_{n,h}^{(4)}$ was used in our investigations.

4.4 Automatic bandwidth selectors

It is clear that the specification of the smoothing parameter h in any kernel estimator is crucial to the performance of the estimator. This holds (among others) in the settings of density, distribution and hazard function estimation. The choice of h is especially important when it comes to practical implementation of kernel estimators. In some cases, if the investigator can inspect the performance of the estimator visually, the choice of h can be made by eye, that is in an informal and subjective way.

Preferably, since the value of h is such an important factor, one would like the value of h to be based on some mathematical justification, and one would also desire a procedure that can determine the value of h automatically for a given set of data. Procedures that provide

a value for h given the data and the estimator used are called *automatic bandwidth selectors*, and we will refer to them simply as selectors.

In this section an overview of existing selectors in the case of right censored data will be given. Their theoretical properties will be discussed, and a new nonparametric bandwidth selector based on the bootstrap will be discussed. These selectors will be compared in the simulation study that is discussed in Chapter 5.

4.4.1 Least-squares cross validation (LSCV)

LSCV is a nonparametric procedure that aims to estimate the theoretical (and unknown in practice) ISE_w global discrepancy measure using only the sample data on hand. The concept of cross validation and especially least-squares cross validation is described in Rudemo (1982) as well as Bowman (1984) in the context of density estimation. Wand and Jones (1995) also discuss smoothing parameter selection via cross validation in the same context.

LSCV was generalized to the censored case for density estimation by Marron and Padgett (1987). The idea was applied to uncensored hazard estimation in Patil (1993a) as well as Sarda and Vieu (1991), and adapted to censored hazard rate estimation in Patil (1993a) and Patil (1993b). To illustrate the concept of least-squares cross validation, the derivation of the criterion will be shown in the case of the hazard function estimator $\hat{\lambda}_{n,h}^{(4)}$, along the same lines as discussed in Patil (1993a). Consider the following expansion of ISE_w :

$$\begin{aligned} ISE_w(h) &= \int_0^\infty [\hat{\lambda}_{n,h}^{(4)}(x) - \lambda^0(x)]^2 w(x) dx \\ &= \int_0^\infty [\hat{\lambda}_{n,h}^{(4)}(x)]^2 w(x) dx - 2 \int_0^\infty \hat{\lambda}_{n,h}^{(4)}(x) \lambda^0(x) w(x) dx + \int_0^\infty [\lambda^0(x)]^2 w(x) dx \\ &= \int_0^\infty [\hat{\lambda}_{n,h}^{(4)}(x)]^2 w(x) dx - 2 \int_0^\infty \left[\frac{\hat{\lambda}_{n,h}^{(4)}(x)}{1 - F^0(x)} \right] f^0(x) w(x) dx + \int_0^\infty [\lambda^0(x)]^2 w(x) dx. \end{aligned}$$

The third term on the right hand side of the equation is independent of h , so the minimization problem can be restricted to the first two terms. A nearly unbiased discrete estimate of the second term is given by

$$n^{-1} \sum_{i=1}^n \frac{\hat{\lambda}_{n,h}^{(4)-i}(X_i)}{[1 - F_n^0(X_i)][1 - H_n^0(X_i)]} w(X_i) \Delta_i, \quad (4.22)$$

where $\hat{\lambda}_{n,h}^{(4)-i}$ is the so-called “leave-one-out” version of $\hat{\lambda}_{n,h}^{(4)}$, and is defined as:

$$\hat{\lambda}_{n,h}^{(4)-i}(x) = (n-1)^{-1} \sum_{j=1, j \neq i}^n \frac{k_h(x - X_j)}{1 - \tilde{F}_n(X_j)} \Delta_j, \quad (4.23)$$

where

$$\tilde{F}_n(x) = \left(\frac{n}{n+1} \right) F_n(x). \quad (4.24)$$

The assumed independence between F^0 and H^0 enables (4.22) to be rewritten as

$$n^{-1} \sum_{i=1}^n \frac{\hat{\lambda}_{n,h}^{(4)-i}(X_i)}{[1 - \tilde{F}_n(X_i)]} w(X_i) \Delta_i. \quad (4.25)$$

Now we can define the LSCV criterion as follows:

$$CV(h) = \int_o^\infty \left[\hat{\lambda}_{n,h}^{(4)}(x) \right]^2 w(x) dx - 2n^{-1} \sum_{i=1}^n \frac{\hat{\lambda}_{n,h}^{(4)-i}(X_i)}{[1 - \tilde{F}_n(X_i)]} w(X_i) \Delta_i. \quad (4.26)$$

The criterion in (4.26) can now be evaluated over all possible values of h , and we can define the selected optimal smoothing parameter as

$$h_{CV} = \underset{h}{\operatorname{argmin}} \{ CV(h) \}. \quad (4.27)$$

Remarks:

1. Asymptotic optimality of h_{CV} was established in the sense that

$$\frac{ISE_w(h_{CV})}{\inf_{h \in H_n} ISE_w(h)} \rightarrow 1 \text{ a.s.},$$

where H_n is a set of possible bandwidths based on certain regularity conditions stated in Patil (1993a), and ISE_w as defined in (3.13).

2. The convergence rate is given as

$$\frac{h_{CV}}{h_{MISE_w}} - 1 = O_p(n^{-\frac{1}{10}}),$$

where h_{MISE_w} is the minimizer of $MISE_w$, and the two terms are defined as

$$MISE_w(h) = E[ISE_w(h)] = E \left\{ \int_0^\infty [\hat{\lambda}_{n,h}^{(4)}(x) - \lambda^0(x)]^2 w(x) dx \right\}, \quad (4.28)$$

and

$$h_{MISE_w} = \underset{h}{\operatorname{argmin}} \{ MISE_w(h) \}. \quad (4.29)$$

Even though the asymptotic optimality of the LSCV criteria is desirable, the rate of convergence is not particularly fast. This motivated the search for improved methods. Gonzalez-Manteiga et al. (1996) focused on the use of the smooth bootstrap in this regard. The ideas of Hall and Robinson (2009) lead to the use of bootstrap aggregation to improve the performance of least-squares cross validation in the case of density estimation.

4.4.2 Bagged cross validation

Bootstrap aggregation or “bagging” as it is commonly known, was first introduced by Breiman (1996) in the context of machine learning. The method is simple to implement, and can be used in a variety of contexts. The ideas discussed in this section were taken from Hall and Robinson (2009) and modified for the setting of hazard function estimation.

In the case where kernel methods are used to estimate a function, the theoretical optimal bandwidth h_0 which minimizes the MISE typically satisfies

$$h_0 \sim Cn^{-a}, \quad (4.30)$$

where $a > 0$ depends on the “regularity” of the problem and C is some functional of the target function (which is unknown in practice). The method of LSCV as discussed in Section 4.4.1 produces an implicit estimator $\hat{C}$ of the unknown functional C . This implies that we can express the bandwidth chosen by LSCV as $\hat{h}_{CV} = \hat{C}n^{-a}$.

By implementing bagging we essentially replace $\hat{C}$ by $E_*(\hat{C}^* | \{X_i, \Delta_i\}_{i=1}^n)$, where $\hat{C}^*$ is a bootstrap version of $\hat{C}$, and $\{X_i, \Delta_i\}_{i=1}^n$ represents the sample data. There are two approaches to implement bagging in LSCV, but both involve independently drawing B bootstrap samples of size $m < n$ without replacement from the original sample of size n . We will denote the sample by $\{(X_i^*, \Delta_i^*)\}_{i=1}^m$. Once each method has been applied, we have to rescale the smoothing parameter value chosen by that method to account for the smaller sample size. Define the following when calculating the cross validation criteria for a bootstrap sample:

$$\hat{\lambda}_{m,h}^{*(4)}(x) = \sum_{i=1}^m \frac{k_h(x - X_{(i)}^*) \Delta_i^*}{m - i + 1}, \quad (4.31)$$

$$\hat{\lambda}_{m,h}^{*(4)-i}(x) = (m - 1)^{-1} \sum_{j=1, j \neq i}^m \frac{k_h(x - X_j^*)}{1 - \tilde{F}_m(X_j^*)} \Delta_j^*, \quad (4.32)$$

and

$$CV^*(h) = \int_0^\infty \left[\hat{\lambda}_{m,h}^{*(4)}(x) \right]^2 w(x) dx - 2m^{-1} \sum_{i=1}^m \frac{\hat{\lambda}_{m,h}^{*(4)-i}(X_i^*)}{[1 - \tilde{F}_m(X_i^*)]} w(X_i^*) \Delta_i^*, \quad (4.33)$$

where $\tilde{F}_m$ is defined similar to (4.24) but with $n = m$.

Method 1

For each bootstrap sample $\{X_i^*, \Delta_i^*\}_{i=1}^m$ we calculate the LSCV criteria defined in (4.33) and denote it by $CV^*(h; b)$, where $b = 1, 2, \dots, B$. For a fixed value of h , the empirical mean of

all B values of $CV^*(h, b)$ can now be defined by

$$CV_{bagg}(h, \cdot) = \frac{1}{B} \sum_{b=1}^B CV^*(h, b). \quad (4.34)$$

This criterion can now be evaluated over all possible values of h , and we can define the selected optimal smoothing parameter as

$$h_{CV, B_1} = \left(\frac{m}{n} \right)^a \underset{h}{\operatorname{argmin}} \{ CV_{bagg}(h, \cdot) \}. \quad (4.35)$$

Method 2

For each bootstrap sample $\{X_i^*, \Delta_i^*\}_{i=1}^m$ we do the following: Calculate the LSCV criteria defined in (4.33) using the bootstrap data, and determine the LSCV optimal h using the same method as described in (4.27), where we now have

$$h^*(b) = \underset{h}{\operatorname{argmin}} \{ CV^*(h, b) \}, \quad (4.36)$$

for $b = 1, 2, \dots, B$. Define the selected optimal smoothing parameter as

$$h_{CV, B_2} = \left(\frac{m}{n} \right)^a \frac{1}{B} \sum_{b=1}^B h^*(b). \quad (4.37)$$

Remarks:

1. As mentioned in Section 4.4.1, LSCV has a very slow convergence rate of $n^{-1/10}$. Hall and Robinson (2009) state the following remark: In the case of density estimation, the reduction in asymptotic variance and mean squared error of a cross validation bandwidth selector, obtainable by applying bootstrap aggregation, is by a factor of at least $(m/n)^{4/5}$, and can be as much as $(m/n)^{9/5}$, unless (m/n) is very small.
2. By implementing half-sample bagging, where $(m/n) = 0.5$, we see that the reduction in variability is between a factor of $0.5^{4/5} = 0.57$ and $0.5^{9/5} = 0.29$, or alternatively stated this means that the variability is reduced by between 43% and 71%. We will investigate this factor empirically in the setting of hazard function estimation in the simulation study.
3. For hazard function estimation we have that $a = 1/5$, resulting in the rescaling factor $\left(\frac{m}{n} \right)^{1/5}$, see the discussion around (3.44) for an explanation.

4.4.3 Bootstrap $AMISE_w^*$

Gonzalez-Manteiga et al. (1996) suggested a selector based on the use of the bootstrap. Under certain conditions and by drawing smooth bootstrap samples in a specific way, they proceeded to derive an explicit expression for the $AMISE_w^*$ term. Their selector then minimizes this $AMISE_w^*$ term as a function of h . We will discuss the bootstrap approach put forth by Gonzalez-Manteiga et al. (1996), as well as the theory and explicit expressions they derived.

Recall the expression of $MISE_w$ given in (3.12). Gonzalez-Manteiga et al. (1996) show the following as a consequence of theorem 2 (Tanner and Wong 1983) for $MISE_w$ using estimator $\hat{\lambda}_{n,h}^{(4)}$, where we note two additional assumptions beforehand:

Assumption 3: The weight function w has compact support contained in $(0, T)$, where T satisfies $1 - F^0(T) > 0$.

Assumption 4: The variables X^0 and Z are independent.

Theorem 4. Under Assumptions 1,2,3 and 4 previously defined, when $n \rightarrow \infty$, $h \rightarrow 0$, and $nh \rightarrow \infty$, $MISE_w(h) = AMISE_w(h) + o(AMISE_w(h))$, where

$$AMISE_w(h) = \int \left[p \left((k_h * \frac{f_1^0}{1-F})(x) - \frac{f_1^0(x)}{1-F(x)} \right) \right]^2 w(x) dx + \frac{p}{nh} R(k) \int \frac{f_1^0(x)}{[1-F(x)]^2} w(x) dx,$$

f_1^0 is the conditional density of $X|\Delta = 1$, $p = P(\Delta = 1)$ is the proportion of uncensored observations, R is defined in (3.30) and $*$ denotes convolution as defined in (4.5).

Patil (1993b) states the following important result regarding the asymptotic equivalence between $AMISE_w$ and ISE_w , that is

$$\sup_{h \in H_n} \left| \frac{ISE_w(h) - AMISE_w(h)}{AMISE_w(h)} \right| \rightarrow 0 \text{ a.s.}$$

where H_n is a set of possible bandwidths based on certain regularity conditions. These results allow the investigator to use $AMISE_w$ as the target function for determining the optimal bandwidth. Now, in order to approximate the theoretical $AMISE_w$ using the smooth bootstrap, there are two major issues that need to be addressed, namely smoothing and censoring. We will describe two different approaches in the construction of the bootstrap samples.

The first approach introduces the smooth bootstrap first, and then the censoring part. Gonzalez-Manteiga et al. (1996) defines this as a smooth-censored (SC) approach. The

second approach introduces the censoring first, after which the smooth bootstrap part is done (referred to as the censored-smooth (CS) approach in Gonzalez-Manteiga et al. (1996)).

SC approach

This method is basically a smooth version of naive bootstrap resampling. F^0 and H^0 are absolutely continuous and can therefore be estimated by smoothed Kaplan-Meier estimators. The Kaplan-Meier estimators are denoted by $\hat{F}_n^0$ and $\hat{H}_n^0$, and consequently the smoothed estimators are denoted by $\hat{F}_n^0 * k_{g_1}$ and $\hat{H}_n^0 * k_{g_2}$ where g_1 and g_2 are pilot bandwidths. We can define these terms by

$$(\hat{F}_n^0 * k_{g_1})(x) = \frac{1}{\sum_{j=1}^n \Delta_j} \sum_{i=1}^n K_{g_1}(x - X_i) \Delta_i, \quad (4.38)$$

and

$$(\hat{H}_n^0 * k_{g_2})(x) = \frac{1}{\sum_{j=1}^n (1 - \Delta_j)} \sum_{i=1}^n K_{g_2}(x - X_i) (1 - \Delta_i). \quad (4.39)$$

The approach is carried out as follows:

1. Draw bootstrap sample $\{X_1^{*0}, \dots, X_n^{*0}\}$ from $\hat{F}_n^0 * k_{g_1}$.
2. Generate independent bootstrap sample $\{Z_1^*, \dots, Z_n^*\}$ from $\hat{H}_n^0 * k_{g_2}$.
3. Construct $\{(X_1^*, \Delta_1^*), \dots, (X_n^*, \Delta_n^*)\}$ where $\Delta_i^* = I(X_i^* = X_i^{*0})$ and $X_i^* = \min\{X_i^{*0}, Z_i^*\}, i = 1, 2, \dots, n$.

This leads to expression (4) of Gonzalez-Manteiga et al. (1996), which is much more complex when compared to the explicit AMISE_w^* expression obtained when using the CS approach.

CS approach

As discussed previously, in this case we try to incorporate censoring as early as possible. Given the observed data $(X_1, \Delta_1), (X_2, \Delta_2), \dots, (X_n, \Delta_n)$ we construct the bootstrap sample as follows:

1. For $i = 1, 2, \dots, n$ generate a random variate ϵ_i from k_{g_1} if $\Delta_i = 1$ or from k_{g_2} if $\Delta_i = 0$.
2. Draw a random sample $(X_1^*, \Delta_1^*), (X_2^*, \Delta_2^*), \dots, (X_n^*, \Delta_n^*)$ by sampling uniformly with replacement from the constructed set $(X_1 + \epsilon_1, \Delta_1), (X_2 + \epsilon_2, \Delta_2), \dots, (X_n + \epsilon_n, \Delta_n)$.

An equivalent way of constructing the sample:

1. Draw I_i uniformly and with replacement from $1, 2, \dots, n$.
2. Independently draw random variate T_i from the density k .
3. Construct $X_i^* = X_{I_i} + \Delta_{I_i}g_1T_i + (1 - \Delta_{I_i})g_2T_i$ and $\Delta_i^* = \Delta_{I_i}$.
4. Repeat (1) - (3) for $i = 1, 2, \dots, n$ to obtain $(X_1^*, \Delta_1^*), (X_2^*, \Delta_2^*), \dots, (X_n^*, \Delta_n^*)$

Theorem 5. *Under Assumptions 1,2,3 and 4 previously defined, when $n \rightarrow \infty$, $h \rightarrow 0$, and $nh \rightarrow \infty$, we can obtain the following explicit asymptotic representation of the bootstrap approximation to $MISE_w$ when using the CS approach:*

$$AMISE_w^*(h) = \int \left[\hat{p} \left((k_h * \frac{\hat{f}_{1,g_1}^0}{1 - \hat{F}_n})(x) - \frac{\hat{f}_{1,g_1}^0(x)}{1 - \hat{F}_n(x)} \right) \right]^2 w(x) dx + \frac{\hat{p}}{nh} R(k) \int \frac{\hat{f}_{1,g_1}^0(x)}{[1 - \hat{F}_n(x)]^2} w(x) dx, \quad (4.40)$$

where $\hat{f}_{1,g_1}^0$ is the kernel density estimator of f_1^0 constructed with the uncensored observations, we define

$$\hat{p} = \frac{1}{n} \sum_{i=1}^n \Delta_i, \quad (4.41)$$

and $\hat{F}_n = \hat{p}(k_{g_1} * F_{1,n}^0) + (1 - \hat{p})(k_{g_2} * H_{1,n}^0)$. $F_{1,n}^0$ and $H_{1,n}^0$ are the empirical distribution functions constructed with the uncensored and censored observations respectively. Note that since we have an explicit expression for $AMISE_w^*$, no bootstrap resampling is required. The criterion (4.40) can now be evaluated over all possible values of h , and we can define the selected optimal smoothing parameter as

$$h^* = \underset{h}{argmin} \{AMISE_w^*(h)\} \quad (4.42)$$

Remarks:

1. The CS approach discussed above corresponds to the CS1 approach in Gonzalez-Manteiga et al. (1996). They also propose a CS2 approach, and then proceed to show that the CS1 and CS2 approaches are in fact equivalent.
2. Gonzalez-Manteiga et al. (1996) proceeded to use WARPing (weighted average of rounded points) algorithms to evaluate (4.40) practically.

3. Note that in both bootstrap sampling approaches we required the use of two *pilot* bandwidths g_1 and g_2 , which are defined as follows:

$$g_1 = \left[\int k''^2 d_k^{-1} \hat{a}_1^{-1} n^{-1} \right]^{\frac{1}{7}}, \quad (4.43)$$

and

$$g_2 = \left[\int k''^2 d_k^{-1} \hat{a}_2^{-1} n^{-1} \right]^{\frac{1}{7}} \quad (4.44)$$

where:

- (a) $\hat{a}_1$ is an estimator of $\int f^{0(3)^2}$.
- (b) $\hat{a}_2$ is an estimator of $\int h^{0(3)^2}$.
- (c) $d_k = \int t^2 k(t) dt$.

This corresponds to the asymptotically optimal bandwidth when estimating the true curvature of the kernel estimator. See Cao, Cuevas and Gonzalez-Monteiga (1994) for details.

4. Note that the normal distribution $[N(0, \sigma^2)]$ is used as reference distribution to get exact expressions for the integrals $\int f^{0(3)^2}$ and $\int h^{0(3)^2}$. The value of σ^2 is then replaced by a certain sample variance to obtain $\hat{a}_1$ and $\hat{a}_2$. For $\hat{a}_1$ we plug in the sample standard deviation of only the observed values (denoted $\hat{\sigma}_{F^0}$), and for $\hat{a}_2$ we plug in the sample variance of the censored values (denoted $\hat{\sigma}_{H^0}$). With some effort the following expressions were derived,

$$\hat{a}_1^{-1} = \frac{16(\hat{\sigma}_{F^0})^7 \sqrt{\pi}}{15}, \quad (4.45)$$

and

$$\hat{a}_2^{-1} = \frac{16(\hat{\sigma}_{H^0})^7 \sqrt{\pi}}{15}, \quad (4.46)$$

5. As mentioned in Section 4.4.1, LSCV has a very slow convergence rate of $n^{-1/10}$. Gonzalez-Manteiga et al. (1996) suggest that their method has an increased convergence rate of approximately $n^{-5/14}$, that is

$$\frac{h^*}{h_{MISE_w}} - 1 \approx O_p(n^{-\frac{5}{14}}).$$

They also suggest that the rate of convergence could be improved to $n^{-1/2}$, and the interested reader can consult the article for more details regarding these open problems.

4.4.4 Bootstrap approximation to $MISE_w$

As previously discussed in Section 2.3.1, when one wishes to obtain a bootstrap approximation to a discrepancy measure like MISE, one has to apply the smooth bootstrap in order to accurately estimate the bias term in that specific discrepancy measure. The question now is how to choose the smoothing parameter h that is used to construct smooth bootstrap samples (see (2.8) - (2.10) for different formulas).

In the following paragraph an entirely nonparametric method for determining the smoothing parameters in the construction of smooth bootstrap samples is discussed. Note that in this setting the target function is the underlying distribution function F , and the kernel estimator is $\hat{F}_{n,h}$ as defined in (3.21).

Bandwidths used in generating smooth bootstrap samples

Van Graan (1983) defines the following global discrepancy measure for $\hat{F}_{n,h}$ as an estimator for the true distribution function F :

$$MISE(h) = E \left\{ \int_{-\infty}^{\infty} [\hat{F}_{n,h}(x) - F(x)]^2 dx \right\}. \quad (4.47)$$

This measure assigns equal weight to each point x , and a more general choice would be to weigh each point x according to the probability of occurrence for that point, that is:

$$MISE(h) = E \left\{ \int_{-\infty}^{\infty} [\hat{F}_{n,h}(x) - F(x)]^2 dF(x) \right\}. \quad (4.48)$$

The equation (4.48) is then expressed in the following form:

$$\begin{aligned} MISE(h) = & \frac{1}{n} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left[K_h(x-y) - F(x) \right]^2 dF(y) dF(x) \\ & + \left(1 - \frac{1}{n} \right) \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} \{ K_h(x-y) - F(x) \} dF(y) \right]^2 dF(x). \end{aligned} \quad (4.49)$$

In the notation of Van Graan (1983) the right hand side of (4.49) is denoted as $J(F, h)$ to illustrate the dependence on F . By applying the plug-in principle as discussed in Section 2.2.1 with F_n as an approximation to F in $J(F, h)$, the following expression for $J(F_n, h)$ can be derived:

$$\begin{aligned} J(F_n, h) = & \frac{1}{n^3} \sum_{j=1}^n \sum_{i=1}^n \left[K_h(X_{(j)} - X_{(i)}) - \frac{j}{n} \right]^2 \\ & + \frac{n-1}{n^4} \sum_{j=1}^n \left[\sum_{i=1}^n \left\{ K_h(X_{(j)} - X_{(i)}) - \frac{j}{n} \right\} \right]^2, \end{aligned} \quad (4.50)$$

where $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ denotes the order statistics and

$$F_n(X_{(j)}) = \frac{j}{n}. \quad (4.51)$$

For a specified integrated kernel K the $J(F_n, h)$ criteria can be evaluated as a function of h . In the proposed procedure we will split the data set consisting of n pairs into two subsets. The first subset contains only the uncensored observations and the second subset contains only the censored observations. Denote the number of uncensored observations by $n_0 = \sum_{i=1}^n \Delta_i$. The number of censored observations is simply computed as $n_c = n - n_0$. (When there is no censoring involved we have that $n_0 = n, n_c = 0$.)

We can now construct the function $J(F_n, h)$ for each of the two samples by defining

$$\begin{aligned} J(F_{n_0}, h) = & \frac{1}{n_0^3} \sum_{j=1}^{n_0} \sum_{i=1}^{n_0} \left[K_h(X_{(j)} - X_{(i)}) - \frac{j}{n_0} \right]^2 \\ & + \frac{n_0 - 1}{n_0^4} \sum_{j=1}^{n_0} \left[\sum_{i=1}^{n_0} \left\{ K_h(X_{(j)} - X_{(i)}) - \frac{j}{n_0} \right\} \right]^2, \end{aligned} \quad (4.52)$$

and

$$\begin{aligned} J(F_{n_c}, h) = & \frac{1}{n_c^3} \sum_{j=1}^{n_c} \sum_{i=1}^{n_c} \left[K_h(X_{(j)} - X_{(i)}) - \frac{j}{n_c} \right]^2 \\ & + \frac{n_c - 1}{n_c^4} \sum_{j=1}^{n_c} \left[\sum_{i=1}^{n_c} \left\{ K_h(X_{(j)} - X_{(i)}) - \frac{j}{n_c} \right\} \right]^2, \end{aligned} \quad (4.53)$$

Define the following two bandwidths:

$$\tilde{g}_1 = \underset{h}{\operatorname{argmin}} \{ J(F_{n_0}, h) \}. \quad (4.54)$$

$$\tilde{g}_2 = \underset{h}{\operatorname{argmin}} \{ J(F_{n_c}, h) \}. \quad (4.55)$$

These bandwidths will be used when constructing smooth bootstrap samples in the proposed procedure. Note that the number of uncensored observations n_0 is a random variable in this setting, and the above method was developed for a fixed sample size n . Since our aim was to propose a simple nonparametric selector based on the bootstrap, we did not attempt to generalize this method to incorporate random right censoring, but decided to simply split the data set.

Empirical Monte-Carlo approximation to MISE_w^*

Since we do not have an explicit expression for MISE_w^* when using the estimator $\hat{\lambda}_{n,h}^{*(4)}$, we decided to estimate the discrepancy measures using empirical Monte-Carlo approximation,

that is using bootstrap resampling. Recall the decomposition of MSE into the sum of the variance and squared bias terms in (3.4). This decomposition holds in the bootstrap world as well, so for the estimator $\hat{\lambda}_{n,h}^{(4)}$ the decomposition of the bootstrap MSE is given by

$$MSE^*(\hat{\lambda}_{n,h}^{(4)}(x)) = Var^*(\hat{\lambda}_{n,h}^{(4)}(x)) + [Bias^*(\hat{\lambda}_{n,h}^{(4)}(x))]^2. \quad (4.56)$$

Now, $Bias^*(\hat{\lambda}_{n,h}^{(4)}(x))$ does not have an explicit expression in this setting, so we need to approximate the value by independent resampling. This requires the generation of B smooth bootstrap samples. The bootstrap samples were drawn following the CS approach as discussed in Section 4.4.3, using $\tilde{g}_1$ in (4.54) and $\tilde{g}_2$ in (4.55) as the bandwidths. For each bootstrap sample (iteration b), we need to calculate $\hat{\lambda}_{n,h}^{*(4)}(x, b)$ which is defined by

$$\hat{\lambda}_{n,h}^{*(4)}(x, b) = \sum_{i=1}^n \frac{k_h(x - X_{(i)}^*) \Delta_i^*}{n - i + 1}. \quad (4.57)$$

We calculate the empirical expected value of these B iterations as

$$\hat{\lambda}_{n,h}^{*(4)}(x, \cdot) = \frac{1}{B} \sum_{b=1}^B \hat{\lambda}_{n,h}^{*(4)}(x, b). \quad (4.58)$$

Recall from (2.4) that in order to determine the approximated bootstrap bias term, we need to subtract the value $t(\hat{F})$ from $\hat{\lambda}_{n,h}^{*(4)}(x, \cdot)$. In the proposed procedure we define $t(\hat{F})$ to be the internal estimator as defined in (4.21) but with $h = \tilde{g}$. The bandwidth $\tilde{g}$ is calculated using bagged cross validation method 2 as discussed in Section 4.4.2. We define

$$\tilde{g} = \left(\frac{m}{n}\right)^a \frac{1}{B} \sum_{b=1}^B h^*(b). \quad (4.59)$$

We can now define the approximated bootstrap bias term as

$$Bias^*(\hat{\lambda}_{n,h}^{(4)}(x)) = \hat{\lambda}_{n,h}^{*(4)}(x, \cdot) - \hat{\lambda}_{n,\tilde{g}}^{(4)}(x). \quad (4.60)$$

Similarly, $Var^*(\hat{\lambda}_{n,h}^{(4)}(x))$ does not have an explicit expression in this context, so we calculate the empirical approximation by

$$Var^*(\hat{\lambda}_{n,h}^{(4)}(x)) = \frac{1}{B-1} \sum_{b=1}^B [\hat{\lambda}_{n,h}^{*(4)}(x, b) - \hat{\lambda}_{n,h}^{*(4)}(x, \cdot)]^2. \quad (4.61)$$

The equation in (4.56) can now be expressed as

$$MSE^*(\hat{\lambda}_{n,h}^{(4)}(x)) = \frac{1}{B-1} \sum_{b=1}^B [\hat{\lambda}_{n,h}^{*(4)}(x, b) - \hat{\lambda}_{n,h}^{*(4)}(x, \cdot)]^2 + [\hat{\lambda}_{n,h}^{*(4)}(x, \cdot) - \hat{\lambda}_{n,\tilde{g}}^{(4)}(x)]^2. \quad (4.62)$$

The equation (4.62) can be integrated with regards to x , using a suitable weight function w , to obtain

$$MISE_w^*(h) = \int_0^\infty MSE^*(\hat{\lambda}_{n,h}^{(4)}(x))w(x)dx. \quad (4.63)$$

This term can then be evaluated as a function of h and we can define the selected optimal smoothing parameter as follows:

$$h_P^* = \underset{h}{argmin} \{MISE_w^*(h)\}. \quad (4.64)$$

This proposed bandwidth selector was not evaluated theoretically, but rather only empirically in the simulation study.

Chapter 5

Simulation study

5.1 Structure of simulation study

5.1.1 Details of settings

Distributions

For the simulation study three variations of the $[\text{Wei}(\alpha, \beta), \alpha > 0, \beta > 0]$ distribution as well as three variations of the $[\text{LN}(\mu, \sigma), \sigma > 0]$ distribution were considered. The general form of these distributions can be found in Section 1.3. Table 5.1 lists the distributions used in the simulation study.

Distribution	F^0 expression
Wei(1,1)	$1 - e^{-x}$
Wei(2,1)	$1 - e^{-(x)^2}$
Wei(3,1)	$1 - e^{-(x)^3}$
LN(0,1)	$\Phi\left(\frac{\ln(x)}{1}\right)$
LN(0.25,0.75)	$\Phi\left(\frac{\ln(x)-0.25}{0.75}\right)$
LN(0.5,0.5)	$\Phi\left(\frac{\ln(x)-0.5}{0.5}\right)$

Table 5.1: *Different distributions used in the simulation study.*

Note that Φ is the standard normal distribution function as defined in (1.16). Figure 5.1 shows the shapes of λ^0 for each of the distributions in Table 5.1.

Censoring

Each distribution will be evaluated for the uncensored (or complete sample) case, as well as in the case where the probability of censoring is equal to 25%. The procedure used to

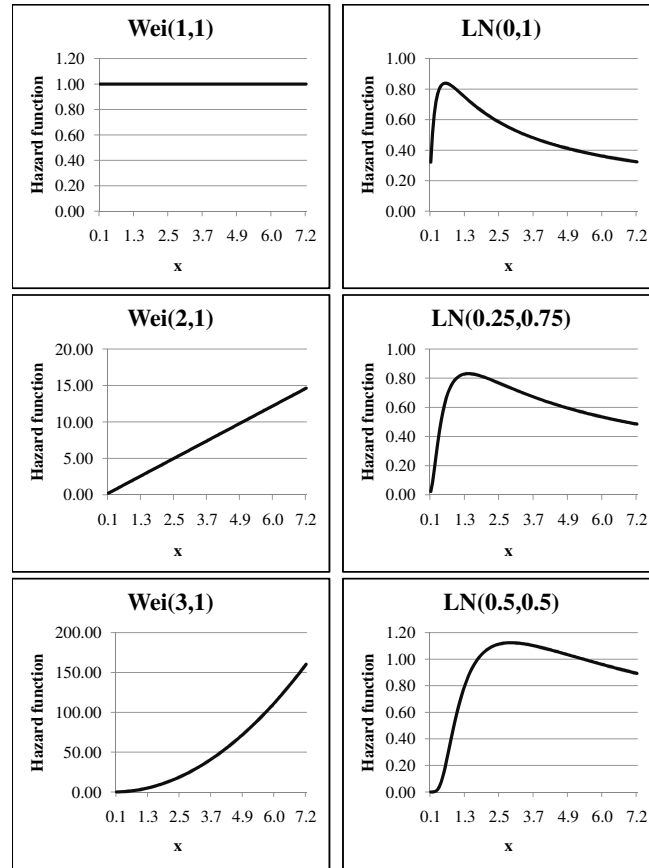


Figure 5.1: *Hazard rate functions considered in the simulation study.*

generate censored data is described in Section 5.2.3.

Sample sizes

Three different sample sizes will be considered, namely $n = 50$, $n = 100$ and $n = 200$.

Automatic bandwidth selectors

Five automatic bandwidth selectors will be compared across all the settings. The notation $\hat{h}$ will be used to refer to the selected bandwidths in general. The specific notation for each of the five selectors is given in the list below.

1. Least-squares cross validation (denoted $\hat{h}_{CV}$).
2. Least-squares cross validation with bagging method 1 (denoted $\hat{h}_{CV,B_1}$).
3. Least-squares cross validation with bagging method 2 (denoted $\hat{h}_{CV,B_2}$).
4. Bootstrap method with explicit AMISE_w^* (denoted $\hat{h}^*$).

5. Proposed bootstrap approximation of MISE_w method (denoted $\hat{h}_P^*$).

For both bagging selectors $\hat{h}_{CV,B_1}$ and $\hat{h}_{CV,B_2}$ we used $B = 100$ bootstrap iterations, as well as half-sample bagging (i.e. $m/n = 0.5$). For $\hat{h}_P^*$ we used $B = 100$ in the bagging step to determine $\tilde{g}$ in (4.59), as well as in the calculation of expressions (4.60) and (4.61).

5.1.2 Layout

Taking the breakdown of the settings described in Section 5.1.1 into account, the total number of settings across which the five selectors will be compared is equal to $(6)(2)(3) = 36$. For each setting 1 000 Monte-Carlo simulations (denoted $MC = 1\,000$) were executed in order to investigate the performance of the selectors.

Summary measures

The following summarizing statistics will be used:

1. Minimum value, denoted *Min*.
2. First quantile, denoted q_1 .
3. *Median*.
4. Third quantile, denoted q_3 .
5. Maximum value, denoted *Max*.
6. Inter quantile range, denoted *iqr*.
7. *Mean*.
8. Standard deviation, denoted *SD*.

Tables

1. For each of the five selectors, and for each of the 36 settings, we calculated the summary measures of Section 5.1.2 for the ISE_w values based on all the $MC = 1\,000$ simulations, where we multiplied the ISE_w values by a factor of 10^3 .
2. The measures defined in Section 5.1.2 will also be calculated for the selected bandwidths $\hat{h}$ for each of the five selectors, across all 36 settings, based on the $MC = 1\,000$ simulations in each case.

3. Tables summarizing the ranking results of the selectors, where the ranking methods are discussed in Section 5.3.

Graphs

1. Boxplots showing the spread of the $\hat{h}$ values along with the corresponding value of $\hat{h}_{MISE_w}$.
2. Kernel density estimators of the $\hat{h}$ values along with the corresponding value of $\hat{h}_{MISE_w}$.

5.1.3 Secondary studies

1. Empirical censoring proportion across MC versus censoring probability of 25%.
2. Reduction in variance of $\hat{h}$ as a result of bagging ($\hat{h}_{CV}$ versus $\hat{h}_{CV, B_2}$).

5.2 Algorithms

5.2.1 Notation

Since many of the quantities that we have to calculate in this section cannot be obtained exactly, we have to introduce some new notation to describe how approximate answers to the theoretical quantities are obtained. The following notation will be used to concisely define discrete sets and grid points:

$$H_p = (h_1, h_2, \dots, h_p) \tag{5.1}$$

$$G_p = (g_1, g_2, \dots, g_p) \tag{5.2}$$

$$X_p = (x_1, x_2, \dots, x_p) \tag{5.3}$$

$$Z_p = (z_1, z_2, \dots, z_p) \tag{5.4}$$

Note that H_p and G_p represent discrete sets of possible bandwidths. When we use the notation $h \in H_p$ or $g \in G_p$ we refer to the specific bandwidth values contained in these discrete sets. On the other hand, X_p and Z_p represent discrete grid points used to evaluate function values. The generic value p noted above was adapted for each setting depending on the accuracy required.

5.2.2 Random number generator

Since many different distributions were investigated during the course of this study, it was decided to develop a random number generator that is flexible and easy to implement for any distribution. Whenever random numbers were generated, the random variate drawn from the Uniform(0, 1) distribution was obtained using the routine DRNUN of the IMSL library (IMSL (1991)). The following algorithm explains the numerical search method of generating a random number from a known distribution function F^0 :

1. Generate a random number U from the Uniform(0, 1) distribution using the routine DRNUN.
2. Define the initial starting point $x_k = 0$, an initial difference measurement $dif = 1$ and an increment value $inc = 0.01$.
3. WHILE ($dif > 0$) DO the following:
 - 3.1. Set $x_k = x_k + inc$.
 - 3.2. Calculate $dif = U - F^0(x_k)$.

This means that the random number drawn from F^0 using U lies in the interval (x_{k-1}, x_k) .

Divide the interval (x_{k-1}, x_k) into 100 equally spaced grid points, where the distance between consecutive points $= \frac{inc}{100}$. Construct a matrix $difmat(100, 2)$ to store each evaluation point x_k and its corresponding absolute difference $|dif|$.

4. Set $x_k = x_k - (inc - 0.00000001)$, and define a smaller increment value $inc_2 = \frac{inc}{100}$.
5. FOR $i = 1, 100$ DO the following:
 - 5.1. Set $x_k = x_k + inc_2$.
 - 5.2. Calculate absolute value difference by $|dif| = |U - F^0(x_k)|$.
 - 5.3. Store $difmat(i, 1) = x_k$ and $difmat(i, 2) = |dif|$.
6. Rank the pairs in the matrix $difmat(100, 2)$ in ascending order based on the $|dif|$ values stored in column 2. We now have the sorted matrix $sortdif(100, 2)$.
7. Define the generated random number as $X = sortdif(1, 1)$.
8. Repeat (1) - (7) n times to construct a sample $X_1^0, X_2^0, \dots, X_n^0$ of size n from F^0 .

Remarks:

1. This method was chosen because it only requires an explicit expression for F^0 and a random number generator for $\text{Uniform}(0, 1)$.
2. Other methods like the inverse transform method require an explicit expression for the *inverse* of F^0 , which is not always available.
3. By choosing $inc = 0.01$ the accuracy of X is within $inc_2 = 0.0001$ (accurate to the 4th decimal).

5.2.3 Generating censored data

The generation of censored data was carried out using the “proportional censorship” model as discussed in Koziol and Green (1976). Define the censoring distribution H^0 as

$$1 - H^0 = (1 - F^0)^\eta \quad (5.5)$$

for $\eta > 0$. This implies a probability of censoring equal to $\frac{\eta}{1+\eta}$. We will now show this for the $\text{Wei}(\alpha, \beta)$ distribution.

Let F^0 be defined as follows:

$$F^0(x) = 1 - e^{-(\beta x)^\alpha} \quad (5.6)$$

We proceed to define H^0 in this specific context:

$$\begin{aligned} 1 - H^0(x) &= \{1 - F^0(x)\}^\eta \\ &= \{e^{-(\beta x)^\alpha}\}^\eta \\ &= e^{-\eta(\beta x)^\alpha} \end{aligned}$$

We now have $X^0 \stackrel{i.i.d}{\sim} F^0$ and $Z \stackrel{i.i.d}{\sim} H^0$. Now,

$$\begin{aligned} P(\Delta = 1) &= P(X^0 \leq Z) \\ &= \int_0^\infty \left\{ \int_0^z dF^0(x) \right\} dH^0(z) \\ &= \int_0^\infty F^0(z) dH^0(z) \\ &= \int_0^\infty \{1 - e^{-(\beta z)^\alpha}\} dH^0(z) \\ &= \int_0^\infty dH^0(z) - \int_0^\infty e^{-(\beta z)^\alpha} dH^0(z) \end{aligned}$$

$$\begin{aligned}
&= 1 - \int_0^\infty e^{-(\beta z)^\alpha} \{ \eta \alpha \beta (\beta z)^{\alpha-1} e^{-\eta(\beta z)^\alpha} \} dz \\
&= 1 - \eta \alpha \beta \int_0^\infty (\beta z)^{\alpha-1} e^{(\beta z)^\alpha(1+\eta)} dz.
\end{aligned}$$

Set $(\beta z)^\alpha(1+\eta) = y$. Then

$$\begin{aligned}
(\beta z)^{\alpha-1} &= [(1+\eta)^{-1}y]^{\frac{\alpha-1}{\alpha}} \\
&= (1+\eta)^{\frac{1}{\alpha}-1} y^{1-\frac{1}{\alpha}}.
\end{aligned}$$

Similarly,

$$\begin{aligned}
z &= \frac{1}{\beta} [(1+\eta)^{-1}y]^{\frac{1}{\alpha}} \\
&= \frac{1}{\beta} (1+\eta)^{-\frac{1}{\alpha}} y^{\frac{1}{\alpha}}.
\end{aligned}$$

This results in

$$\frac{dz}{dy} = \frac{1}{\alpha\beta} (1+\eta)^{-\frac{1}{\alpha}} y^{\frac{1}{\alpha}-1}.$$

Substituting these values into the expression for $P(\Delta = 1)$ results in:

$$\begin{aligned}
P(\Delta = 1) &= 1 - \eta \alpha \beta \int_0^\infty (1+\eta)^{\frac{1}{\alpha}-1} y^{1-\frac{1}{\alpha}} e^{-y} \frac{1}{\alpha\beta} (1+\eta)^{-\frac{1}{\alpha}} y^{\frac{1}{\alpha}-1} dy \\
&= 1 - \eta \int_0^\infty (1+\eta)^{-1} e^{-y} dy \\
&= 1 - \frac{\eta}{\eta+1}.
\end{aligned}$$

This implies $P(\Delta = 0) = 1 - P(\Delta = 1) = \frac{\eta}{\eta+1}$. For example, choosing $\eta = 1/3$ we have censoring proportion equal to 0.25, or 25%. Each generated random variable Δ_i is therefore a Bernoulli random variable, with probability of censoring $\frac{\eta}{1+\eta}$. The sum of n independent Bernoulli trials follows the binomial distribution, more specifically the binomial($n, \frac{\eta}{1+\eta}$) distribution. Recall the definition of $\hat{p}$ as the proportion of uncensored observations, given in (4.41). From this we have that

$$E(\hat{p}) = 1 - \frac{\eta}{1+\eta} \tag{5.7}$$

$$Var(\hat{p}) = \frac{1}{n} \left[\frac{\eta}{1+\eta} \left(1 - \frac{\eta}{1+\eta} \right) \right] \tag{5.8}$$

Generating a sample proceeds as follows:

1. Generate a sample $X_1^0, X_2^0, \dots, X_n^0$ from a known reference distribution F^0 using the random number generator in Section 5.2.2.
2. Independently generate a sample $Z_1, Z_2, \dots, Z_n$ from the censoring distribution H^0 using the random number generator in Section 5.2.2 and the relation in (5.5).
3. FOR $i = 1, n$ DO the following:
 - 3.1. Construct $X_i = \min \{X_i^0, Z_i\}$ and $\Delta_i = I(X_i = X_i^0)$.
4. This results in the censored data set $\{X_i, \Delta_i\}_{i=1}^n$, drawn from F .

5.2.4 Theoretical optimal bandwidth $\hat{h}_{MISE_w}$

1. Generate a sample $X_1^0, X_2^0, \dots, X_n^0$ from a known reference distribution F^0 with hazard rate λ^0 .
2. Calculate the value $SE(x, h) = [\hat{\lambda}_{n,h}^{(2)}(x) - \lambda^0(x)]^2$ for a fixed h in the point x .
3. Repeat (2) for $x \in X_p$ to obtain $SE(x_1, h), SE(x_2, h), \dots, SE(x_p, h)$
4. Numerically integrate (using a weight function) the $SE(x_p, h)$ values over the x grid to obtain an approximate value for $ISE_w(h)$.
5. Repeat (2) - (4) for $h \in H_p$ values to obtain $ISE_w(h_1), ISE_w(h_2), \dots, ISE_w(h_p)$.
6. Repeat (1) - (5) MC times to obtain the following:

$$\begin{array}{ccccccc}
 ISE_{w,1}(h_1), & ISE_{w,1}(h_2), & \dots & , & ISE_{w,1}(h_p) \\
 ISE_{w,2}(h_1), & ISE_{w,2}(h_2), & \dots & , & ISE_{w,2}(h_p) \\
 \vdots & & & & \\
 ISE_{w,MC}(h_1), & ISE_{w,MC}(h_2), & \dots & , & ISE_{w,MC}(h_p).
 \end{array}$$

7. Calculate the average $ISE(h)$ value (over MC) in each of the h grid points, to obtain estimates of $MISE_w(h_1), MISE_w(h_2), \dots, MISE_w(h_p)$ where

$$MISE_w(h) = \frac{1}{MC} \sum_{i=1}^{MC} ISE_{w,i}(h).$$

8. Calculate $\hat{h}_{MISE_w} = \underset{h \in H_p}{argmin} \{MISE_w(h_1), MISE_w(h_2), \dots, MISE_w(h_p)\}$.

5.2.5 LSCV selector $\hat{h}_{CV}$

For ease of reference, define the following terms:

$$CV_{term1}(h) = \int_o^\infty \left[\hat{\lambda}_{n,h}^{(4)}(x) \right]^2 w(x) dx, \quad (5.9)$$

$$CV_{term2}(h) = 2n^{-1} \sum_{i=1}^n \frac{\hat{\lambda}_{n,h}^{(4)-i}(X_i)}{[1 - \tilde{F}_n(X_i)]} w(X_i) \Delta_i, \quad (5.10)$$

$$CV_{term1}^*(h) = \int_o^\infty \left[\hat{\lambda}_{m,h}^{*(4)}(x) \right]^2 w(x) dx, \quad (5.11)$$

$$CV_{term2}^*(h) = 2m^{-1} \sum_{i=1}^m \frac{\hat{\lambda}_{m,h}^{*(4)-i}(X_i)}{[1 - \tilde{F}_m(X_i)]} w(X_i) \Delta_i. \quad (5.12)$$

1. Generate a paired sample $\{X_i, \Delta_i\}_{i=1}^n$ from a known reference distribution F with hazard rate λ using the algorithm described earlier in Section 5.2.3.
- 2.1. Calculate $\hat{\lambda}_{n,h}^{(4)}(x, h)$ for a fixed h in the point x .
- 2.2. Repeat (2.1) for $x \in X_p$.
- 2.3. Numerically integrate $\hat{\lambda}_{n,h}^{(4)}(x, h)$ across the x grid to obtain $CV_{term1}(h)$ (see (5.9)).
- 2.4. Calculate the value $CV_{term2}(h)$ as defined in equation (5.10).
- 2.5. Construct $CV(h) = CV_{term1}(h) - CV_{term2}(h)$.
3. Repeat (2.1) - (2.5) for $h \in H_p$ to obtain $CV(h_1), CV(h_2), \dots, CV(h_p)$.
4. Calculate $\hat{h}_{CV} = \underset{h \in H_p}{argmin} \{CV(h_1), CV(h_2), \dots, CV(h_p)\}$.

5.2.6 Bagged LSCV selectors $\hat{h}_{CV,B_1}$ and $\hat{h}_{CV,B_2}$

1. Generate a paired sample $\{X_i, \Delta_i\}_{i=1}^n$ from a known reference distribution F with hazard rate λ using the algorithm described in Section 5.2.3.
- 2.1. Draw bootstrap sample (iteration b) of size $m < n$ without replacement, and denoted by $\{X_i^*, \Delta_i^*\}_{i=1}^m$.
- 2.2. Using $\{X_i^*, \Delta_i^*\}_{i=1}^m$, calculate $\hat{\lambda}_{m,h}^{*(4)}(x, h)$ for a fixed h in the point x .
- 2.3. Repeat (2.2) for $x \in X_p$.
- 2.4. Numerically integrate $\hat{\lambda}_{m,h}^{*(4)}(x, h)$ across the x grid to obtain $CV_{term1}^*(h)$ (see (5.11)).
- 2.5. Using $\{X_i^*, \Delta_i^*\}_{i=1}^m$, calculate the value $CV_{term2}^*(h)$ as defined in equation (5.12).
- 2.6.1. Calculate $CV^*(h, b) = CV_{term1}^*(h) - CV_{term2}^*(h)$.
- 2.6.2. Calculate $CV_{bagg}(h, b) = CV_{term1}^*(h) - CV_{term2}^*(h)$.
- 2.7. Repeat (2.1) - (2.6) for $h \in H_p$ resulting in:
 $CV^*(h_1, b), CV^*(h_2, b), \dots, CV^*(h_p, b)$ and
 $CV_{bagg}(h_1, b), CV_{bagg}(h_2, b), \dots, CV_{bagg}(h_p, b)$.
- 2.8. Determine $\hat{h}^*(b) = \underset{h \in H_p}{argmin} \{CV^*(h_1, b), CV^*(h_2, b), \dots, CV^*(h_p, b)\}$.
3. Repeat (2.1) - (2.8) a large number of times (for $b = 1, 2, \dots, B$) to obtain
 $\hat{h}^*(1), \hat{h}^*(2), \dots, \hat{h}^*(B)$ and
 $CV_{bagg}(h_1, 1), CV_{bagg}(h_2, 1), \dots, CV_{bagg}(h_p, 1)$
 $CV_{bagg}(h_1, 2), CV_{bagg}(h_2, 2), \dots, CV_{bagg}(h_p, 2)$
 $\vdots$
 $CV_{bagg}(h_1, B), CV_{bagg}(h_2, B), \dots, CV_{bagg}(h_p, B)$.
4. In CV_{bagg} , determine $CV_{bagg}(h, \cdot) = \frac{1}{B} \sum_{b=1}^B CV_{bagg}(h, b)$ for $h \in H_p$ to obtain
 $CV_{bagg}(h_1, \cdot), CV_{bagg}(h_2, \cdot), \dots, CV_{bagg}(h_p, \cdot)$.
- 5.1. Calculate $\hat{h}_{CV,B_1} = \left(\frac{m}{n}\right)^{1/5} \underset{h \in H_p}{argmin} \{CV_{bagg}(h_1, \cdot), CV_{bagg}(h_2, \cdot), \dots, CV_{bagg}(h_p, \cdot)\}$.
- 5.2. Calculate $\hat{h}_{CV,B_2} = \left(\frac{m}{n}\right)^{1/5} \frac{1}{B} \sum_{b=1}^B \hat{h}^*(b)$.

5.2.7 Bootstrap $AMISE_w^*$ bandwidth selector $\hat{h}^*$

For ease of reference in following algorithm, define the following terms:

$$Convolver_x(z, h) = (k_h * \frac{\hat{f}_{1,g_1}^0}{1 - \hat{F}_n})(z), \quad (5.13)$$

$$IntConvolver(x, h) = \int Convolver_x(z, h) dz, \quad (5.14)$$

$$Bias^*(x, h) = \left[\hat{p} \left(IntConvolver(x, h) - \frac{\hat{f}_{1,g_1}^0(x)}{1 - \hat{F}_n(x)} \right) \right], \quad (5.15)$$

$$IBias_w^*(h) = \int Bias^*(x, h) w(x) dx, \quad (5.16)$$

$$Var^*(x, h) = \frac{\hat{p}}{nh} R(k) \frac{\hat{f}_{1,g_1}^0(x)}{[1 - \hat{F}_n(x)]^2}, \quad (5.17)$$

and

$$IVar_w^*(h) = \frac{\hat{p}}{nh} R(k) \int \frac{\hat{f}_{1,g_1}^0(x)}{[1 - \hat{F}_n(x)]^2} w(x) dx. \quad (5.18)$$

1. Generate a paired sample $\{X_i, \Delta_i\}_{i=1}^n$ from a known reference distribution F with hazard rate λ using the algorithm described in Section 5.2.3.
2. Using $\{X_i, \Delta_i\}_{i=1}^n$, calculate g_1 and g_2 as defined in (4.43) and (4.44) respectively.
 - 3.1. Calculate $Convolver_x(z, h)$ as defined in (5.13) for $z \in Z_p$.
 - 3.2. Numerically integrate $Convolver_x(z, h)$ over the z grid to obtain $IntConvolver(x, h)$.
as defined in (5.14) and consequently calculate $Bias^*(x, h)$ as defined in (5.15).
 - 3.3. Calculate $Bias^*(x, h)$ for $x \in X_p$ and numerically integrate over the x grid to obtain $IBias_w^*(h)$ as defined in (5.16).
 - 3.4. Calculate $Var^*(x, h)$ as defined in (5.17) for $x \in X_p$ and numerically integrate over the x grid to obtain $IVar_w^*(h)$ as defined in (5.18).
4. Construct the approximated $AMISE_w^*(h) = IBias_w^*(h)^2 + IVar_w^*(h)$.
5. Evaluate $AMISE_w^*(h)$ for $h \in H_p$ to obtain
 $AMISE_w^*(h_1), AMISE_w^*(h_2), \dots, AMISE_w^*(h_p)$.
6. Calculate $\hat{h}^* = \underset{h \in H_p}{argmin} \{AMISE_w^*(h_1), AMISE_w^*(h_2), \dots, AMISE_w^*(h_p)\}$.

5.2.8 Bootstrap approximation to $MISE_w$ selector $\hat{h}_P^*$

1. Generate a paired sample $\{X_i, \Delta_i\}_{i=1}^n$ from a known reference distribution F with hazard rate λ using the algorithm described in Section 5.2.3.

- 2.1. Split the sample into n_0 uncensored observations and n_c censored observations.

Rank the uncensored observations $\{X_i^0\}_{i=1}^{n_0}$ and the censored observations $\{Z_i\}_{i=1}^{n_c}$ to obtain the ordered samples $\{X_{(i)}^0\}_{i=1}^{n_0}$ and $\{Z_{(i)}\}_{i=1}^{n_c}$.

- 2.2. Evaluate $J(F_{n_0}, g)$ and $J(F_{n_c}, g)$ for a fixed smoothing parameter g .

- 2.3. Repeat (2.2) for $g \in G_p$ to obtain

$$J(F_{n_0}, g_1), J(F_{n_0}, g_2), \dots, J(F_{n_0}, g_p) \text{ and } J(F_{n_c}, g_1), J(F_{n_c}, g_2), \dots, J(F_{n_c}, g_p).$$

- 2.4. Obtain

$$\tilde{g}_1 = \underset{g \in G_p}{\operatorname{argmin}} \{J(F_{n_0}, g_1), J(F_{n_0}, g_2), \dots, J(F_{n_0}, g_p)\} \text{ and}$$

$$\tilde{g}_2 = \underset{g \in G_p}{\operatorname{argmin}} \{J(F_{n_c}, g_1), J(F_{n_c}, g_2), \dots, J(F_{n_c}, g_p)\}.$$

3. Generate a *smooth* bootstrap sample $\{X_i^*, \Delta_i^*\}_{i=1}^n$ using the following formulas:

$$X_i^* = X_{I_i} + \Delta_{I_i} \tilde{g}_1 T_i + (1 - \Delta_{I_i}) \tilde{g}_2 T_i, \quad i = 1, 2, \dots, n.$$

$$\Delta_i^* = \Delta_{I_i}, \quad i = 1, 2, \dots, n.$$

4. Repeat (3) B times, each time calculating $\hat{\lambda}_{n,h}^{*(4)}(x; h)$, resulting in the following:

$$\hat{\lambda}_{n,h}^{*(4)}(x; h; 1), \hat{\lambda}_{n,h}^{*(4)}(x; h; 2), \dots, \hat{\lambda}_{n,h}^{*(4)}(x; h; B)$$

- 4.1. Calculate $\tilde{g}$ as defined in (4.59) and consequently determine the following:

$$\hat{\lambda}_{n,h}^{*(4)}(x; h; \cdot) = \frac{1}{B} \sum_{b=1}^B \hat{\lambda}_{n,h}^{*(4)}(x; h; b),$$

$$\operatorname{Bias}^*(x; h) = \hat{\lambda}_{n,h}^{*(4)}(x; h; \cdot) - \hat{\lambda}_{n,h}^{(4)}(x; \tilde{g}),$$

$$Var^*(x; h) = \frac{1}{B-1} \sum_{b=1}^B (\hat{\lambda}_{n,h}^{*(4)}(x; h; b) - \hat{\lambda}_{n,h}^{*(4)}(x; h; \cdot))^2, \text{ and}$$

$$MSE^*(x; h) = (Bias^*(x; h))^2 + Var^*(x; h)$$

4.2. Repeat (4.1) for $x \in X_p$.

4.3. Using a weight function w , numerically integrate $MSE^*(x_i; h)$ over the x grid to obtain $MISE_w^*(h)$.

4.4. Repeat (4.1) - (4.3) for $h \in H_p$ to obtain $MISE_w^*(h_1), MSE_w^*(h_2), \dots, MSE_w^*(h_p)$.

4.5. Repeat (1) - (4.4) MC times, resulting in the following:

$$\begin{array}{ccccccc} MISE_{w,1}^*(h_1), & MISE_{w,1}^*(h_2), & \dots & , & MISE_{w,1}^*(h_p) \\ MISE_{w,2}^*(h_1), & MISE_{w,2}^*(h_2), & \dots & , & MISE_{w,2}^*(h_p) \\ \vdots & & & & \\ MISE_{w,MC}^*(h_1), & MISE_{w,MC}^*(h_2), & \dots & , & MISE_{w,MC}^*(h_p). \end{array}$$

4.6.1. For each MC replication (that is for $k = 1, 2, \dots, MC$), calculate

$$\hat{h}_{P,k}^* = \underset{h \in H_p}{argmin} \{MISE_{w,k}^*(h_1), MISE_{w,k}^*(h_2), \dots, MISE_{w,k}^*(h_p)\}.$$

4.6.2. Calculate $\hat{h}_P^* = \frac{1}{MC} \sum_{k=1}^{MC} \hat{h}_{P,k}^*$.

Remarks:

1. X_{I_i} is a naive bootstrap sample value generated as in Section 2.2.1.
2. A symmetric kernel is used when generating the T_i values in the *smooth* bootstrap samples.
3. The bootstrap approximation of the $MISE_w$ term, $MISE_w^*$, is generated by empirical approximation (as opposed to an explicit representation).

5.3 Ranking methods

In an attempt to rank the automatic bandwidth selectors across all the settings considered in the simulation study, we implemented the following ranking system based on approximated ISE_w values:

1. Rank the selectors based on their value for the mean of ISE_w , where the minimum value is best and receives rank 1.
2. In exactly the same manner as in (1), rank the selectors based on their value for the median of ISE_w and the standard deviation of the ISE_w values.
3. Repeat this ranking system across all settings.
4. For mean, median and standard deviation, separately calculate the average ranking across all settings each selector as follows:
Average rank = (Sum of all ranking values) / (Total number of settings).
5. Add up the mean, median and standard deviation rankings for each selector across all the settings to determine the overall ranking as:
Overall rank = (Sum of all ranking values) / (Total number of settings).
6. The overall rank will be in the interval $[1; 5]$.
7. Rank the selectors from best to worst by simply ordering them in ascending order of their average rank values.

For the selected $\hat{h}$ values, we defined the discrepancy measure $M\tilde{S}E$ as

$$M\tilde{S}E(\hat{h}) = [\hat{h}(\cdot) - h_{MISE_w}]^2 + var(\hat{h}), \quad (5.19)$$

where

$$\hat{h}(\cdot) = \frac{1}{MC} \sum_{i=1}^{MC} \hat{h}_i$$

is the mean value of the selected bandwidths across the MC iterations. The following ranking system was implemented:

1. Calculate the $M\tilde{S}E$ values across all settings for all selectors.
2. For each setting, rank the selectors based on their $M\tilde{S}E$ values, where the minimum values receives rank 1.

3. Add up the ranking values across all settings and determine the mean ranking values, which will be in the interval $[1; 5]$.
4. Rank the selectors from best to worst by simply ordering them in ascending order of their mean rank values.

Even though these are fairly simple means of grading the different selectors in terms of ISE_w and $\hat{h}$ values, it serves as a measurable and objective way of determining the performance of each selector across all the settings considered.

5.4 Simulation study results

5.4.1 Theoretical optimal h

A Monte-Carlo approximation was carried out to determine the value of $\hat{h}_{MISE_w}$ for each of the distributions considered. The algorithm in Section 5.2.4 was carried out over a grid of possible h values ranging from 0.01 to 5.01 in increments of 0.01, with $MC = 10\,000$. The selected bandwidth is denoted by h_{rough} . A finer grid of possible h values was constructed over the interval $(h_{rough} - 0.5; h_{rough} + 0.5)$, with increments of 0.002. If $(h_{rough} < 0.5)$, the lower bound of the new interval was set equal to 0.00001. The algorithm in Section 5.2.4 was applied to this finer grid with $MC = 5\,000$. The result is the selected bandwidth $\hat{h}_{MISE_w}$. Table 5.2 summarizes the results.

Distribution	$n = 50$	$n = 100$	$n = 200$
Wei(1,1)	1.0514	0.8372	0.6942
Wei(2,1)	1.3268	1.3780	0.8680
Wei(3,1)	0.8574	0.4090	0.3580
LN(0,1)	1.3776	1.2964	1.2452
LN(0.25,0.75)	1.3778	1.2654	1.1942
LN(0.5,0.5)	1.8064	1.6740	1.6022

Table 5.2: Approximate values for $\hat{h}_{MISE_w}$ for the different distributions.

5.4.2 Automatic bandwidth selectors

In each of the graphs (from pg. 104) the asymptotically optimal bandwidth is shown by a dotted red line. Figure 5.2 is a color legend that should be used to distinguish between the bandwidth selectors in the kernel density estimators based on their selected values of $\hat{h}$:

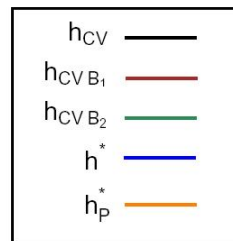


Figure 5.2: Legend to be used when interpreting graphs.

<i>Value</i>	<i>Selector</i>	<i>Min</i>	q_1	<i>Median</i>	q_3	<i>Max</i>	<i>iqr</i>	<i>Mean</i>	<i>SD</i>
ISE_w	$\hat{h}_{CV}$	2.2	27.3	57.1	133.4	6,824.0	106.1	144.8	362.7
ISE_w	$\hat{h}_{CV,B_1}$	0.7	26.4	49.9	112.7	2,503.0	86.3	106.3	185.1
ISE_w	$\hat{h}_{CV,B_2}$	1.1	22.4	44.3	85.1	844.5	62.7	68.1	77.2
ISE_w	$\hat{h}^*$	2.2	20.5	39.0	66.9	1,017.5	46.5	55.1	64.4
ISE_w	$\hat{h}_P^*$	1.4	22.8	44.8	86.7	795.3	63.9	66.4	70.3
$\hat{h}$	$\hat{h}_{CV}$	0.010	0.600	1.065	1.660	5.010	1.060	1.303	1.037
$\hat{h}$	$\hat{h}_{CV,B_1}$	0.044	0.662	0.992	1.375	4.248	0.714	1.105	0.677
$\hat{h}$	$\hat{h}_{CV,B_2}$	0.163	0.866	1.125	1.504	3.222	0.638	1.222	0.493
$\hat{h}$	$\hat{h}^*$	0.710	1.060	1.190	1.380	4.060	0.320	1.110	0.452
$\hat{h}$	$\hat{h}_P^*$	0.280	1.200	1.585	2.345	5.010	1.145	1.865	0.925

Table 5.3: *Simulation summary : Weibull(1, 1), $\eta = 0, n = 50, MC = 1\ 000$*

<i>Value</i>	<i>Selector</i>	<i>Min</i>	q_1	<i>Median</i>	q_3	<i>Max</i>	<i>iqr</i>	<i>Mean</i>	<i>SD</i>
ISE_w	$\hat{h}_{CV}$	1.3	32.7	77.4	174.1	44,369.3	141.4	253.9	1,481.0
ISE_w	$\hat{h}_{CV,B_1}$	2.0	27.9	60.2	122.1	3,625.6	94.3	123.8	242.4
ISE_w	$\hat{h}_{CV,B_2}$	1.0	26.4	54.2	102.8	2,372.1	76.4	87.6	134.1
ISE_w	$\hat{h}^*$	3.3	23.0	43.6	86.9	1,491.4	63.9	69.3	91.9
ISE_w	$\hat{h}_P^*$	1.6	28.1	56.8	110.1	1,591.6	82.0	87.2	118.8
$\hat{h}$	$\hat{h}_{CV}$	0.010	0.560	1.160	1.790	5.010	1.230	1.333	0.978
$\hat{h}$	$\hat{h}_{CV,B_1}$	0.087	0.766	1.093	1.480	3.682	0.714	1.172	0.614
$\hat{h}$	$\hat{h}_{CV,B_2}$	0.259	0.920	1.191	1.504	2.981	0.584	1.249	0.466
$\hat{h}$	$\hat{h}^*$	0.550	1.020	1.180	1.460	3.620	0.440	1.120	0.484
$\hat{h}$	$\hat{h}_P^*$	0.310	1.300	1.690	2.200	4.130	0.900	1.788	0.653

Table 5.4: *Simulation summary : Weibull(1, 1), $\eta = 1/3, n = 50, MC = 1\ 000$*

<i>Value</i>	<i>Selector</i>	<i>Min</i>	q_1	<i>Median</i>	q_3	<i>Max</i>	<i>iqr</i>	<i>Mean</i>	<i>SD</i>
ISE_w	$\hat{h}_{CV}$	0.3	27.6	72.8	204.7	16,272.0	177.1	262.4	743.8
ISE_w	$\hat{h}_{CV,B_1}$	0.0	16.8	50.6	138.6	2,876.5	121.8	151.9	304.7
ISE_w	$\hat{h}_{CV,B_2}$	0.1	13.4	37.6	85.9	1,924.0	72.5	82.4	153.8
ISE_w	$\hat{h}^*$	0.3	14.3	36.4	81.4	2,213.2	67.1	77.4	153.6
ISE_w	$\hat{h}_P^*$	0.3	16.4	42.8	91.4	704.5	75.0	72.0	89.9
$\hat{h}$	$\hat{h}_{CV}$	0.010	0.480	1.065	1.615	3.790	1.135	1.113	0.755
$\hat{h}$	$\hat{h}_{CV,B_1}$	0.026	0.548	0.844	1.167	3.238	0.618	0.879	0.479
$\hat{h}$	$\hat{h}_{CV,B_2}$	0.146	0.665	0.875	1.139	2.774	0.474	0.929	0.373
$\hat{h}$	$\hat{h}^*$	0.320	0.615	0.820	1.140	3.130	0.525	0.630	0.380
$\hat{h}$	$\hat{h}_P^*$	0.180	1.070	1.360	1.610	2.590	0.540	1.343	0.409

Table 5.5: *Simulation summary : Weibull(2, 1), $\eta = 0, n = 50, MC = 1\ 000$*

<i>Value</i>	<i>Selector</i>	<i>Min</i>	q_1	<i>Median</i>	q_3	<i>Max</i>	<i>iqr</i>	<i>Mean</i>	<i>SD</i>
ISE_w	$\hat{h}_{CV}$	0.3	36.5	109.9	267.3	14,620.0	230.8	415.6	1,176.8
ISE_w	$\hat{h}_{CV,B_1}$	0.6	24.5	72.1	182.2	12,773.7	157.8	231.8	658.0
ISE_w	$\hat{h}_{CV,B_2}$	0.6	21.2	60.6	141.4	11,996.8	120.2	139.9	442.9
ISE_w	$\hat{h}^*$	1.0	23.8	58.9	136.3	2,043.2	112.5	118.1	186.2
ISE_w	$\hat{h}_P^*$	0.5	20.8	61.0	146.4	6,158.6	125.6	134.5	341.4
$\hat{h}$	$\hat{h}_{CV}$	0.010	0.420	0.865	1.300	3.780	0.880	0.925	0.638
$\hat{h}$	$\hat{h}_{CV,B_1}$	0.035	0.514	0.731	1.001	2.594	0.488	0.778	0.412
$\hat{h}$	$\hat{h}_{CV,B_2}$	0.141	0.572	0.755	1.008	2.399	0.436	0.820	0.347
$\hat{h}$	$\hat{h}^*$	0.290	0.540	0.680	0.940	2.400	0.400	0.560	0.333
$\hat{h}$	$\hat{h}_P^*$	0.190	0.875	1.080	1.290	2.060	0.415	1.085	0.327

Table 5.6: *Simulation summary : Weibull(2, 1), $\eta = 1/3, n = 50, MC = 1\ 000$*

<i>Value</i>	<i>Selector</i>	<i>Min</i>	q_1	<i>Median</i>	q_3	<i>Max</i>	<i>iqr</i>	<i>Mean</i>	<i>SD</i>
ISE_w	$\hat{h}_{CV}$	0.2	56.9	129.6	334.3	8,469.2	277.4	346.6	722.6
ISE_w	$\hat{h}_{CV,B_1}$	0.7	31.4	75.6	199.4	6,105.6	167.9	209.3	420.5
ISE_w	$\hat{h}_{CV,B_2}$	1.0	27.7	63.0	135.3	4,689.0	107.6	138.0	276.4
ISE_w	$\hat{h}^*$	0.4	27.0	64.2	148.4	1,904.7	121.5	126.7	198.6
ISE_w	$\hat{h}_P^*$	1.5	44.8	72.9	134.6	2,533.5	89.8	127.1	197.4
$\hat{h}$	$\hat{h}_{CV}$	0.010	0.420	0.735	1.170	3.000	0.750	0.829	0.580
$\hat{h}$	$\hat{h}_{CV,B_1}$	0.017	0.418	0.601	0.814	2.333	0.396	0.640	0.345
$\hat{h}$	$\hat{h}_{CV,B_2}$	0.080	0.478	0.632	0.811	2.184	0.334	0.672	0.282
$\hat{h}$	$\hat{h}^*$	0.190	0.380	0.480	0.670	1.580	0.290	0.410	0.231
$\hat{h}$	$\hat{h}_P^*$	0.080	0.750	0.910	1.050	2.370	0.300	0.896	0.263

Table 5.7: *Simulation summary : Weibull(3, 1), $\eta = 0, n = 50, MC = 1\ 000$*

<i>Value</i>	<i>Selector</i>	<i>Min</i>	q_1	<i>Median</i>	q_3	<i>Max</i>	<i>iqr</i>	<i>Mean</i>	<i>SD</i>
ISE_w	$\hat{h}_{CV}$	1.3	79.7	179.8	434.7	13,253.2	355.0	532.4	1,186.2
ISE_w	$\hat{h}_{CV,B_1}$	2.0	46.9	111.5	267.3	18,889.0	220.4	307.7	830.1
ISE_w	$\hat{h}_{CV,B_2}$	1.8	38.4	91.0	207.7	4,198.3	169.3	196.5	354.1
ISE_w	$\hat{h}^*$	0.8	45.5	102.4	211.8	3,694.0	166.3	199.2	315.2
ISE_w	$\hat{h}_P^*$	2.4	53.4	116.6	251.3	3,476.7	197.9	201.2	268.2
$\hat{h}$	$\hat{h}_{CV}$	0.010	0.330	0.605	0.910	3.320	0.580	0.657	0.457
$\hat{h}$	$\hat{h}_{CV,B_1}$	0.017	0.348	0.505	0.696	2.168	0.348	0.543	0.285
$\hat{h}$	$\hat{h}_{CV,B_2}$	0.093	0.417	0.537	0.695	2.406	0.278	0.581	0.249
$\hat{h}$	$\hat{h}^*$	0.180	0.340	0.430	0.570	1.480	0.230	0.360	0.191
$\hat{h}$	$\hat{h}_P^*$	0.160	0.580	0.710	0.870	1.870	0.290	0.736	0.228

Table 5.8: *Simulation summary : Weibull(3, 1), $\eta = 1/3, n = 50, MC = 1\ 000$*

<i>Value</i>	<i>Selector</i>	<i>Min</i>	q_1	<i>Median</i>	q_3	<i>Max</i>	<i>iqr</i>	<i>Mean</i>	<i>SD</i>
ISE_w	$\hat{h}_{CV}$	3.5	132.8	209.7	324.5	2,949.9	191.8	273.1	250.4
ISE_w	$\hat{h}_{CV,B_1}$	7.4	126.7	196.9	285.2	2,025.5	158.5	225.7	161.2
ISE_w	$\hat{h}_{CV,B_2}$	7.5	114.8	179.4	254.2	799.2	139.4	192.9	109.3
ISE_w	$\hat{h}^*$	12.3	128.7	204.1	293.1	650.8	164.4	223.1	125.7
ISE_w	$\hat{h}_P^*$	5.3	113.5	186.3	258.0	723.8	144.5	196.8	109.5
$\hat{h}$	$\hat{h}_{CV}$	0.010	0.580	1.100	1.570	5.010	0.990	1.179	0.819
$\hat{h}$	$\hat{h}_{CV,B_1}$	0.044	0.788	1.093	1.410	4.361	0.622	1.157	0.597
$\hat{h}$	$\hat{h}_{CV,B_2}$	0.229	0.961	1.205	1.536	3.560	0.575	1.283	0.470
$\hat{h}$	$\hat{h}^*$	1.070	1.600	1.890	2.380	5.010	0.780	1.680	0.808
$\hat{h}$	$\hat{h}_P^*$	0.330	1.290	1.610	2.075	5.010	0.785	1.783	0.775

Table 5.9: *Simulation summary : lognormal(0, 1), $\eta = 0, n = 50, MC = 1\ 000$*

<i>Value</i>	<i>Selector</i>	<i>Min</i>	q_1	<i>Median</i>	q_3	<i>Max</i>	<i>iqr</i>	<i>Mean</i>	<i>SD</i>
ISE_w	$\hat{h}_{CV}$	4.5	140.3	242.2	381.8	6,776.2	241.5	343.9	472.2
ISE_w	$\hat{h}_{CV,B_1}$	2.7	113.5	202.7	308.7	7,048.6	195.2	242.7	285.3
ISE_w	$\hat{h}_{CV,B_2}$	2.6	101.6	178.7	275.7	895.3	174.2	200.3	128.2
ISE_w	$\hat{h}^*$	3.3	115.7	179.2	265.8	795.9	150.1	197.1	112.3
ISE_w	$\hat{h}_P^*$	1.5	126.6	203.7	301.8	3,575.5	175.2	227.9	171.1
$\hat{h}$	$\hat{h}_{CV}$	0.010	0.630	1.240	1.765	5.010	1.135	1.344	0.968
$\hat{h}$	$\hat{h}_{CV,B_1}$	0.017	0.918	1.275	1.628	4.361	0.709	1.338	0.670
$\hat{h}$	$\hat{h}_{CV,B_2}$	0.298	1.074	1.345	1.727	3.303	0.653	1.418	0.497
$\hat{h}$	$\hat{h}^*$	0.880	1.480	1.670	1.940	5.010	0.460	1.610	0.572
$\hat{h}$	$\hat{h}_P^*$	0.210	1.465	1.910	2.490	5.010	1.025	2.031	0.826

Table 5.10: *Simulation summary : lognormal(0, 1), $\eta = 1/3, n = 50, MC = 1\ 000$*

<i>Value</i>	<i>Selector</i>	<i>Min</i>	q_1	<i>Median</i>	q_3	<i>Max</i>	<i>iqr</i>	<i>Mean</i>	<i>SD</i>
ISE_w	$\hat{h}_{CV}$	3.3	70.9	136.7	232.0	6,910.1	161.1	208.9	374.1
ISE_w	$\hat{h}_{CV,B_1}$	2.4	63.2	113.7	196.8	1,239.5	133.6	151.6	142.7
ISE_w	$\hat{h}_{CV,B_2}$	1.0	58.6	101.8	165.5	1,183.1	106.9	120.1	88.9
ISE_w	$\hat{h}^*$	4.4	62.3	105.6	166.7	466.3	104.4	121.9	77.8
ISE_w	$\hat{h}_P^*$	8.3	64.9	112.5	169.3	499.1	104.4	125.0	77.8
$\hat{h}$	$\hat{h}_{CV}$	0.010	0.605	1.210	1.765	5.010	1.160	1.343	0.999
$\hat{h}$	$\hat{h}_{CV,B_1}$	0.052	0.810	1.201	1.541	4.361	0.731	1.274	0.706
$\hat{h}$	$\hat{h}_{CV,B_2}$	0.231	1.029	1.325	1.691	3.875	0.662	1.402	0.522
$\hat{h}$	$\hat{h}^*$	0.870	1.490	1.710	2.040	5.010	0.550	1.550	0.556
$\hat{h}$	$\hat{h}_P^*$	0.490	1.450	1.815	2.460	5.010	1.010	2.092	0.978

Table 5.11: *Simulation summary : lognormal(0.25, 0.75), $\eta = 0, n = 50, MC = 1\ 000$*

<i>Value</i>	<i>Selector</i>	<i>Min</i>	q_1	<i>Median</i>	q_3	<i>Max</i>	<i>iqr</i>	<i>Mean</i>	<i>SD</i>
ISE_w	$\hat{h}_{CV}$	1.8	79.3	155.6	281.2	4,329.4	201.9	258.9	402.9
ISE_w	$\hat{h}_{CV,B_1}$	4.1	64.2	128.9	222.2	5,516.3	157.9	188.5	262.5
ISE_w	$\hat{h}_{CV,B_2}$	1.9	58.9	114.3	187.0	1,445.5	128.1	141.4	123.6
ISE_w	$\hat{h}^*$	2.9	57.5	106.8	171.6	552.7	114.1	123.3	81.7
ISE_w	$\hat{h}_P^*$	1.6	65.3	121.1	200.0	624.1	134.7	142.6	101.2
$\hat{h}$	$\hat{h}_{CV}$	0.010	0.660	1.320	1.990	5.010	1.330	1.485	1.082
$\hat{h}$	$\hat{h}_{CV,B_1}$	0.044	0.871	1.288	1.663	4.361	0.792	1.330	0.680
$\hat{h}$	$\hat{h}_{CV,B_2}$	0.257	1.079	1.360	1.730	3.265	0.651	1.435	0.518
$\hat{h}$	$\hat{h}^*$	0.690	1.390	1.610	1.890	4.140	0.500	1.500	0.523
$\hat{h}$	$\hat{h}_P^*$	0.510	1.500	1.990	2.570	5.010	1.070	2.111	0.826

Table 5.12: *Simulation summary : lognormal(0.25, 0.75), $\eta = 1/3, n = 50, MC = 1\ 000$*

<i>Value</i>	<i>Selector</i>	<i>Min</i>	q_1	<i>Median</i>	q_3	<i>Max</i>	<i>iqr</i>	<i>Mean</i>	<i>SD</i>
ISE_w	$\hat{h}_{CV}$	1.4	32.4	65.2	155.1	5,677.2	122.6	179.4	403.9
ISE_w	$\hat{h}_{CV,B_1}$	0.9	23.3	48.2	105.2	3,564.0	81.9	105.0	221.1
ISE_w	$\hat{h}_{CV,B_2}$	1.8	21.5	40.9	78.1	1,168.7	56.6	65.2	81.6
ISE_w	$\hat{h}^*$	2.7	23.6	39.3	63.6	940.9	40.0	56.5	64.3
ISE_w	$\hat{h}_P^*$	1.9	25.0	42.5	78.1	991.0	53.2	62.3	72.8
$\hat{h}$	$\hat{h}_{CV}$	0.010	0.535	1.245	2.125	5.010	1.590	1.503	1.182
$\hat{h}$	$\hat{h}_{CV,B_1}$	0.026	0.853	1.223	1.671	4.361	0.818	1.318	0.713
$\hat{h}$	$\hat{h}_{CV,B_2}$	0.288	1.028	1.329	1.743	3.448	0.716	1.417	0.531
$\hat{h}$	$\hat{h}^*$	0.550	1.220	1.410	1.790	3.850	0.570	1.270	0.552
$\hat{h}$	$\hat{h}_P^*$	0.340	1.380	1.915	2.725	5.010	1.345	2.105	0.909

Table 5.13: *Simulation summary : lognormal(0.5, 0.5), $\eta = 0, n = 50, MC = 1\ 000$*

<i>Value</i>	<i>Selector</i>	<i>Min</i>	q_1	<i>Median</i>	q_3	<i>Max</i>	<i>iqr</i>	<i>Mean</i>	<i>SD</i>
ISE_w	$\hat{h}_{CV}$	0.8	39.9	79.6	199.8	19,557.7	159.9	282.5	993.8
ISE_w	$\hat{h}_{CV,B_1}$	1.4	30.8	61.4	131.4	2,813.1	100.5	128.2	221.8
ISE_w	$\hat{h}_{CV,B_2}$	0.6	28.0	53.9	104.8	2,758.6	76.8	90.7	144.4
ISE_w	$\hat{h}^*$	1.1	29.1	50.6	85.0	866.1	55.8	73.5	78.4
ISE_w	$\hat{h}_P^*$	1.3	30.7	55.6	106.7	3,686.2	76.0	93.6	179.8
$\hat{h}$	$\hat{h}_{CV}$	0.010	0.665	1.350	2.200	5.010	1.535	1.533	1.111
$\hat{h}$	$\hat{h}_{CV,B_1}$	0.044	0.905	1.236	1.671	4.361	0.766	1.321	0.668
$\hat{h}$	$\hat{h}_{CV,B_2}$	0.198	1.010	1.291	1.680	3.442	0.670	1.370	0.508
$\hat{h}$	$\hat{h}^*$	0.530	1.150	1.410	1.735	3.790	0.585	1.220	0.515
$\hat{h}$	$\hat{h}_P^*$	0.360	1.425	1.830	2.340	4.640	0.915	1.921	0.686

Table 5.14: *Simulation summary : lognormal(0.5, 0.5), $\eta = 1/3, n = 50, MC = 1\ 000$*

<i>Value</i>	<i>Selector</i>	<i>Min</i>	q_1	<i>Median</i>	q_3	<i>Max</i>	<i>iqr</i>	<i>Mean</i>	<i>SD</i>
ISE_w	$\hat{h}_{CV}$	0.7	17.0	32.8	73.6	1,863.0	56.6	73.0	140.2
ISE_w	$\hat{h}_{CV,B_1}$	1.5	14.9	28.9	53.4	696.6	38.5	45.6	55.7
ISE_w	$\hat{h}_{CV,B_2}$	1.6	15.6	28.8	49.0	507.1	33.3	38.7	37.8
ISE_w	$\hat{h}^*$	1.6	15.0	26.2	43.5	217.3	28.5	32.4	24.8
ISE_w	$\hat{h}_P^*$	1.3	16.0	27.9	49.8	349.0	33.9	38.5	34.7
$\hat{h}$	$\hat{h}_{CV}$	0.020	0.500	0.860	1.200	5.000	0.700	0.972	0.758
$\hat{h}$	$\hat{h}_{CV,B_1}$	0.061	0.644	0.897	1.175	3.735	0.531	0.990	0.553
$\hat{h}$	$\hat{h}_{CV,B_2}$	0.221	0.802	1.044	1.320	2.953	0.518	1.111	0.446
$\hat{h}$	$\hat{h}^*$	0.680	0.910	1.000	1.130	3.360	0.220	0.960	0.329
$\hat{h}$	$\hat{h}_P^*$	0.200	0.970	1.230	1.580	5.010	0.610	1.438	0.777

Table 5.15: *Simulation summary : Weibull(1, 1), $\eta = 0, n = 100, MC = 1\ 000$*

<i>Value</i>	<i>Selector</i>	<i>Min</i>	q_1	<i>Median</i>	q_3	<i>Max</i>	<i>iqr</i>	<i>Mean</i>	<i>SD</i>
ISE_w	$\hat{h}_{CV}$	2.3	21.6	45.2	95.7	4,322.8	74.2	99.4	231.2
ISE_w	$\hat{h}_{CV,B_1}$	1.5	19.3	37.4	74.1	1,590.2	54.8	63.0	95.9
ISE_w	$\hat{h}_{CV,B_2}$	2.2	18.2	35.1	64.0	448.8	45.8	48.6	46.9
ISE_w	$\hat{h}^*$	1.6	17.4	29.7	50.2	550.4	32.7	40.4	39.4
ISE_w	$\hat{h}_P^*$	2.4	20.9	35.3	61.6	660.6	40.6	50.2	55.0
$\hat{h}$	$\hat{h}_{CV}$	0.010	0.540	0.990	1.450	4.640	0.910	1.132	0.834
$\hat{h}$	$\hat{h}_{CV,B_1}$	0.026	0.688	0.958	1.275	3.482	0.588	1.048	0.565
$\hat{h}$	$\hat{h}_{CV,B_2}$	0.242	0.847	1.074	1.386	2.935	0.539	1.144	0.425
$\hat{h}$	$\hat{h}^*$	0.610	0.890	1.000	1.190	3.370	0.300	0.940	0.420
$\hat{h}$	$\hat{h}_P^*$	0.190	1.040	1.355	1.800	3.920	0.760	1.475	0.606

Table 5.16: *Simulation summary : Weibull(1, 1), $\eta = 1/3, n = 100, MC = 1\ 000$*

<i>Value</i>	<i>Selector</i>	<i>Min</i>	q_1	<i>Median</i>	q_3	<i>Max</i>	<i>iqr</i>	<i>Mean</i>	<i>SD</i>
ISE_w	$\hat{h}_{CV}$	0.2	15.4	40.3	97.2	2,533.1	81.8	100.9	199.7
ISE_w	$\hat{h}_{CV,B_1}$	0.0	8.2	22.4	60.5	1,876.5	52.2	62.3	127.5
ISE_w	$\hat{h}_{CV,B_2}$	0.1	6.5	17.8	44.8	1,294.3	38.3	38.1	69.6
ISE_w	$\hat{h}^*$	0.1	8.6	20.8	47.5	659.7	38.9	40.8	62.3
ISE_w	$\hat{h}_P^*$	0.1	9.8	21.6	45.0	652.6	35.1	37.1	48.7
$\hat{h}$	$\hat{h}_{CV}$	0.020	0.690	1.230	1.820	3.360	1.130	1.259	0.768
$\hat{h}$	$\hat{h}_{CV,B_1}$	0.044	0.679	1.019	1.345	2.716	0.666	1.023	0.495
$\hat{h}$	$\hat{h}_{CV,B_2}$	0.101	0.742	0.988	1.232	2.457	0.490	1.012	0.368
$\hat{h}$	$\hat{h}^*$	0.340	0.580	0.830	1.170	2.320	0.590	0.600	0.390
$\hat{h}$	$\hat{h}_P^*$	0.210	1.040	1.410	1.640	2.790	0.600	1.345	0.431

Table 5.17: *Simulation summary : Weibull(2, 1), $\eta = 0, n = 100, MC = 1\ 000$*

<i>Value</i>	<i>Selector</i>	<i>Min</i>	q_1	<i>Median</i>	q_3	<i>Max</i>	<i>iqr</i>	<i>Mean</i>	<i>SD</i>
ISE_w	$\hat{h}_{CV}$	0.4	24.6	58.6	138.7	2,800.9	114.2	149.8	279.2
ISE_w	$\hat{h}_{CV,B_1}$	0.2	13.4	35.3	89.6	1,638.4	76.2	86.0	149.6
ISE_w	$\hat{h}_{CV,B_2}$	0.3	10.8	29.0	67.7	732.9	56.9	58.8	90.0
ISE_w	$\hat{h}^*$	0.2	12.8	32.7	76.6	1,391.9	63.8	64.5	101.2
ISE_w	$\hat{h}_P^*$	0.3	11.9	34.0	86.0	1,018.8	74.2	70.2	107.7
$\hat{h}$	$\hat{h}_{CV}$	0.020	0.550	1.020	1.460	3.520	0.910	1.039	0.647
$\hat{h}$	$\hat{h}_{CV,B_1}$	0.035	0.531	0.783	1.053	2.307	0.522	0.807	0.388
$\hat{h}$	$\hat{h}_{CV,B_2}$	0.177	0.603	0.788	0.985	2.326	0.382	0.822	0.295
$\hat{h}$	$\hat{h}^*$	0.280	0.530	0.715	1.005	2.490	0.475	0.540	0.342
$\hat{h}$	$\hat{h}_P^*$	0.190	0.850	1.080	1.300	2.060	0.450	1.069	0.320

Table 5.18: *Simulation summary : Weibull(2, 1), $\eta = 1/3, n = 100, MC = 1\ 000$*

<i>Value</i>	<i>Selector</i>	<i>Min</i>	q_1	<i>Median</i>	q_3	<i>Max</i>	<i>iqr</i>	<i>Mean</i>	<i>SD</i>
ISE_w	$\hat{h}_{CV}$	0.8	35.8	99.1	196.1	3,724.6	160.3	198.8	351.2
ISE_w	$\hat{h}_{CV,B_1}$	0.6	24.4	57.2	117.4	2,387.6	93.0	111.5	178.3
ISE_w	$\hat{h}_{CV,B_2}$	0.2	19.1	41.9	91.2	868.2	72.1	74.1	89.5
ISE_w	$\hat{h}^*$	0.6	16.1	37.9	86.8	897.1	70.7	72.0	96.0
ISE_w	$\hat{h}_P^*$	0.3	36.8	63.1	106.3	831.1	69.5	93.2	102.3
$\hat{h}$	$\hat{h}_{CV}$	0.010	0.400	0.730	1.215	2.370	0.815	0.818	0.561
$\hat{h}$	$\hat{h}_{CV,B_1}$	0.026	0.427	0.636	0.923	2.142	0.496	0.686	0.368
$\hat{h}$	$\hat{h}_{CV,B_2}$	0.141	0.495	0.665	0.861	2.081	0.365	0.699	0.278
$\hat{h}$	$\hat{h}^*$	0.210	0.340	0.420	0.585	1.550	0.245	0.370	0.210
$\hat{h}$	$\hat{h}_P^*$	0.130	0.730	0.910	1.050	2.130	0.320	0.889	0.261

Table 5.19: *Simulation summary : Weibull(3, 1), $\eta = 0, n = 100, MC = 1\ 000$*

<i>Value</i>	<i>Selector</i>	<i>Min</i>	q_1	<i>Median</i>	q_3	<i>Max</i>	<i>iqr</i>	<i>Mean</i>	<i>SD</i>
ISE_w	$\hat{h}_{CV}$	0.6	49.1	107.5	229.7	11,935.1	180.6	262.4	625.1
ISE_w	$\hat{h}_{CV,B_1}$	0.8	34.6	72.6	154.5	4,312.3	119.9	153.9	279.7
ISE_w	$\hat{h}_{CV,B_2}$	0.4	26.2	58.6	121.6	1,253.6	95.4	104.8	144.1
ISE_w	$\hat{h}^*$	1.0	24.0	56.4	121.5	2,068.0	97.6	107.5	161.4
ISE_w	$\hat{h}_P^*$	0.7	26.2	61.2	124.0	1,880.2	97.9	111.2	167.5
$\hat{h}$	$\hat{h}_{CV}$	0.010	0.350	0.610	0.955	2.710	0.605	0.684	0.459
$\hat{h}$	$\hat{h}_{CV,B_1}$	0.017	0.357	0.522	0.740	1.846	0.383	0.559	0.283
$\hat{h}$	$\hat{h}_{CV,B_2}$	0.131	0.415	0.549	0.708	1.949	0.293	0.580	0.228
$\hat{h}$	$\hat{h}^*$	0.180	0.340	0.430	0.600	1.730	0.260	0.350	0.213
$\hat{h}$	$\hat{h}_P^*$	0.140	0.590	0.720	0.850	1.540	0.260	0.713	0.205

Table 5.20: *Simulation summary : Weibull(3, 1), $\eta = 1/3, n = 100, MC = 1\ 000$*

<i>Value</i>	<i>Selector</i>	<i>Min</i>	q_1	<i>Median</i>	q_3	<i>Max</i>	<i>iqr</i>	<i>Mean</i>	<i>SD</i>
ISE_w	$\hat{h}_{CV}$	18.8	122.7	183.6	258.7	1,712.5	136.1	207.1	130.5
ISE_w	$\hat{h}_{CV,B_1}$	9.4	125.3	178.3	237.0	1,506.6	111.8	190.0	100.4
ISE_w	$\hat{h}_{CV,B_2}$	10.7	113.8	169.8	225.6	506.0	111.8	176.1	81.2
ISE_w	$\hat{h}^*$	16.7	121.6	173.0	236.2	580.0	114.6	188.0	93.1
ISE_w	$\hat{h}_P^*$	19.6	117.7	167.8	220.9	642.8	103.2	174.7	80.2
$\hat{h}$	$\hat{h}_{CV}$	0.030	0.550	0.920	1.250	4.520	0.700	0.931	0.526
$\hat{h}$	$\hat{h}_{CV,B_1}$	0.035	0.705	0.940	1.175	3.395	0.470	0.964	0.419
$\hat{h}$	$\hat{h}_{CV,B_2}$	0.249	0.798	0.982	1.227	2.915	0.428	1.039	0.365
$\hat{h}$	$\hat{h}^*$	1.030	1.370	1.530	1.820	5.010	0.450	1.450	0.651
$\hat{h}$	$\hat{h}_P^*$	0.230	1.010	1.225	1.460	5.000	0.450	1.277	0.448

Table 5.21: *Simulation summary : lognormal(0, 1), $\eta = 0, n = 100, MC = 1\ 000$*

<i>Value</i>	<i>Selector</i>	<i>Min</i>	q_1	<i>Median</i>	q_3	<i>Max</i>	<i>iqr</i>	<i>Mean</i>	<i>SD</i>
ISE_w	$\hat{h}_{CV}$	13.2	125.2	194.6	284.3	3,253.7	159.0	234.5	206.5
ISE_w	$\hat{h}_{CV,B_1}$	1.0	122.4	185.0	262.7	800.4	140.3	205.3	118.4
ISE_w	$\hat{h}_{CV,B_2}$	1.1	111.5	167.0	239.8	695.5	128.3	182.7	100.2
ISE_w	$\hat{h}^*$	8.2	103.3	153.5	215.7	468.2	112.3	163.5	80.2
ISE_w	$\hat{h}_P^*$	4.7	118.0	170.9	233.9	934.2	115.9	180.8	90.0
$\hat{h}$	$\hat{h}_{CV}$	0.010	0.590	1.030	1.410	5.010	0.820	1.071	0.668
$\hat{h}$	$\hat{h}_{CV,B_1}$	0.078	0.749	1.053	1.341	3.561	0.592	1.079	0.493
$\hat{h}$	$\hat{h}_{CV,B_2}$	0.283	0.896	1.091	1.376	3.323	0.480	1.157	0.396
$\hat{h}$	$\hat{h}^*$	0.830	1.260	1.370	1.520	5.010	0.260	1.330	0.391
$\hat{h}$	$\hat{h}_P^*$	0.190	1.110	1.400	1.770	4.190	0.660	1.485	0.554

Table 5.22: *Simulation summary : lognormal(0, 1), $\eta = 1/3, n = 100, MC = 1\ 000$*

<i>Value</i>	<i>Selector</i>	<i>Min</i>	q_1	<i>Median</i>	q_3	<i>Max</i>	<i>iqr</i>	<i>Mean</i>	<i>SD</i>
ISE_w	$\hat{h}_{CV}$	3.7	64.3	101.8	156.9	1,431.1	92.6	127.4	108.4
ISE_w	$\hat{h}_{CV,B_1}$	3.9	61.2	98.5	146.8	648.6	85.6	112.3	72.6
ISE_w	$\hat{h}_{CV,B_2}$	2.6	56.9	90.6	128.1	655.2	71.2	99.1	58.2
ISE_w	$\hat{h}^*$	3.1	56.2	83.8	118.7	413.5	62.6	90.5	47.3
ISE_w	$\hat{h}_P^*$	4.8	55.8	88.2	128.7	570.2	72.9	98.5	57.5
$\hat{h}$	$\hat{h}_{CV}$	0.020	0.610	1.050	1.400	5.010	0.790	1.098	0.719
$\hat{h}$	$\hat{h}_{CV,B_1}$	0.078	0.783	1.062	1.341	4.361	0.557	1.120	0.552
$\hat{h}$	$\hat{h}_{CV,B_2}$	0.249	0.907	1.127	1.434	3.557	0.527	1.215	0.455
$\hat{h}$	$\hat{h}^*$	1.010	1.270	1.380	1.540	5.010	0.270	1.330	0.371
$\hat{h}$	$\hat{h}_P^*$	0.280	1.150	1.410	1.720	5.010	0.570	1.520	0.630

Table 5.23: *Simulation summary : lognormal(0.25, 0.75), $\eta = 0$, $n = 100$, $MC = 1\,000$*

<i>Value</i>	<i>Selector</i>	<i>Min</i>	q_1	<i>Median</i>	q_3	<i>Max</i>	<i>iqr</i>	<i>Mean</i>	<i>SD</i>
ISE_w	$\hat{h}_{CV}$	1.6	69.5	118.8	195.0	1,911.8	125.5	154.8	137.8
ISE_w	$\hat{h}_{CV,B_1}$	4.5	62.3	99.6	159.3	1,223.0	97.0	126.0	111.5
ISE_w	$\hat{h}_{CV,B_2}$	4.5	55.6	87.3	139.6	803.4	83.9	104.3	75.1
ISE_w	$\hat{h}^*$	6.4	55.4	89.3	129.9	308.5	74.5	97.2	54.9
ISE_w	$\hat{h}_P^*$	2.4	60.6	96.1	148.3	417.4	87.7	109.8	67.2
$\hat{h}$	$\hat{h}_{CV}$	0.040	0.680	1.195	1.695	5.010	1.015	1.315	0.905
$\hat{h}$	$\hat{h}_{CV,B_1}$	0.061	0.840	1.149	1.493	4.361	0.653	1.231	0.627
$\hat{h}$	$\hat{h}_{CV,B_2}$	0.190	0.986	1.259	1.572	3.423	0.586	1.314	0.466
$\hat{h}$	$\hat{h}^*$	0.770	1.210	1.350	1.560	4.350	0.350	1.300	0.441
$\hat{h}$	$\hat{h}_P^*$	0.300	1.230	1.550	1.990	4.740	0.760	1.686	0.686

Table 5.24: *Simulation summary : lognormal(0.25, 0.75), $\eta = 1/3$, $n = 100$, $MC = 1\,000$*

<i>Value</i>	<i>Selector</i>	<i>Min</i>	q_1	<i>Median</i>	q_3	<i>Max</i>	<i>iqr</i>	<i>Mean</i>	<i>SD</i>
ISE_w	$\hat{h}_{CV}$	1.2	25.8	44.7	91.5	4,348.9	65.8	91.5	194.8
ISE_w	$\hat{h}_{CV,B_1}$	2.1	20.4	37.3	71.2	645.3	50.8	60.1	69.1
ISE_w	$\hat{h}_{CV,B_2}$	0.7	18.8	31.5	54.5	391.4	35.7	45.7	45.5
ISE_w	$\hat{h}^*$	3.4	19.4	30.2	46.1	358.7	26.7	39.3	36.0
ISE_w	$\hat{h}_P^*$	1.7	20.3	32.1	48.0	400.3	27.8	40.7	35.2
$\hat{h}$	$\hat{h}_{CV}$	0.010	0.575	1.110	1.745	5.010	1.170	1.314	1.011
$\hat{h}$	$\hat{h}_{CV,B_1}$	0.078	0.805	1.123	1.523	4.361	0.718	1.248	0.683
$\hat{h}$	$\hat{h}_{CV,B_2}$	0.290	0.945	1.242	1.626	3.621	0.681	1.318	0.497
$\hat{h}$	$\hat{h}^*$	0.680	1.070	1.220	1.450	3.960	0.380	1.130	0.446
$\hat{h}$	$\hat{h}_P^*$	0.300	1.210	1.545	2.045	4.710	0.835	1.752	0.791

Table 5.25: *Simulation summary : lognormal(0.5, 0.5), $\eta = 0, n = 100, MC = 1\,000$*

<i>Value</i>	<i>Selector</i>	<i>Min</i>	q_1	<i>Median</i>	q_3	<i>Max</i>	<i>iqr</i>	<i>Mean</i>	<i>SD</i>
ISE_w	$\hat{h}_{CV}$	2.7	30.7	54.5	116.8	7,233.6	86.1	139.6	383.4
ISE_w	$\hat{h}_{CV,B_1}$	1.7	24.7	45.2	92.6	1,334.5	67.9	81.5	114.2
ISE_w	$\hat{h}_{CV,B_2}$	1.8	22.7	39.2	73.4	771.2	50.7	60.0	67.2
ISE_w	$\hat{h}^*$	3.4	22.4	35.9	59.6	499.7	37.2	48.7	47.2
ISE_w	$\hat{h}_P^*$	1.5	22.1	36.7	63.1	924.6	41.0	52.3	60.1
$\hat{h}$	$\hat{h}_{CV}$	0.010	0.590	1.220	1.955	5.010	1.365	1.385	1.000
$\hat{h}$	$\hat{h}_{CV,B_1}$	0.061	0.810	1.132	1.550	3.665	0.740	1.219	0.625
$\hat{h}$	$\hat{h}_{CV,B_2}$	0.196	0.928	1.200	1.546	3.441	0.618	1.272	0.473
$\hat{h}$	$\hat{h}^*$	0.600	1.040	1.240	1.650	3.510	0.610	1.110	0.519
$\hat{h}$	$\hat{h}_P^*$	0.350	1.220	1.610	2.135	3.720	0.915	1.702	0.633

Table 5.26: *Simulation summary : lognormal(0.5, 0.5), $\eta = 1/3, n = 100, MC = 1\,000$*

<i>Value</i>	<i>Selector</i>	<i>Min</i>	q_1	<i>Median</i>	q_3	<i>Max</i>	<i>iqr</i>	<i>Mean</i>	<i>SD</i>
ISE_w	$\hat{h}_{CV}$	0.5	10.3	20.6	43.0	841.4	32.7	39.3	64.1
ISE_w	$\hat{h}_{CV,B_1}$	0.5	9.1	17.8	32.1	419.6	23.0	26.9	32.5
ISE_w	$\hat{h}_{CV,B_2}$	1.0	8.8	16.8	29.8	236.8	20.9	22.8	21.5
ISE_w	$\hat{h}^*$	1.3	8.7	14.5	24.4	116.7	15.7	18.7	14.5
ISE_w	$\hat{h}_P^*$	1.2	10.2	17.4	29.7	181.7	19.4	22.8	18.7
$\hat{h}$	$\hat{h}_{CV}$	0.020	0.440	0.690	0.900	4.580	0.460	0.719	0.483
$\hat{h}$	$\hat{h}_{CV,B_1}$	0.044	0.548	0.740	0.923	3.369	0.374	0.762	0.372
$\hat{h}$	$\hat{h}_{CV,B_2}$	0.229	0.631	0.797	0.997	2.589	0.367	0.852	0.325
$\hat{h}$	$\hat{h}^*$	0.610	0.770	0.820	0.880	2.280	0.110	0.800	0.120
$\hat{h}$	$\hat{h}_P^*$	0.240	0.810	0.980	1.170	4.750	0.360	1.048	0.459

Table 5.27: *Simulation summary : Weibull(1, 1), $\eta = 0, n = 200, MC = 1\ 000$*

<i>Value</i>	<i>Selector</i>	<i>Min</i>	q_1	<i>Median</i>	q_3	<i>Max</i>	<i>iqr</i>	<i>Mean</i>	<i>SD</i>
ISE_w	$\hat{h}_{CV}$	1.6	14.1	25.1	49.6	699.2	35.6	46.1	64.0
ISE_w	$\hat{h}_{CV,B_1}$	1.2	11.6	22.8	40.4	457.9	28.8	35.8	46.1
ISE_w	$\hat{h}_{CV,B_2}$	1.4	11.4	21.7	37.4	357.8	26.0	29.2	28.8
ISE_w	$\hat{h}^*$	1.7	10.5	18.5	31.0	242.9	20.5	24.2	20.4
ISE_w	$\hat{h}_P^*$	1.1	13.0	23.4	39.6	192.1	26.6	30.4	24.7
$\hat{h}$	$\hat{h}_{CV}$	0.030	0.460	0.780	1.070	3.860	0.610	0.850	0.601
$\hat{h}$	$\hat{h}_{CV,B_1}$	0.052	0.575	0.797	1.036	3.360	0.461	0.871	0.486
$\hat{h}$	$\hat{h}_{CV,B_2}$	0.176	0.696	0.880	1.116	2.902	0.420	0.952	0.382
$\hat{h}$	$\hat{h}^*$	0.560	0.770	0.840	0.920	3.360	0.150	0.810	0.313
$\hat{h}$	$\hat{h}_P^*$	0.230	0.870	1.090	1.370	3.250	0.500	1.169	0.462

Table 5.28: *Simulation summary : Weibull(1, 1), $\eta = 1/3, n = 200, MC = 1\ 000$*

<i>Value</i>	<i>Selector</i>	<i>Min</i>	q_1	<i>Median</i>	q_3	<i>Max</i>	<i>iqr</i>	<i>Mean</i>	<i>SD</i>
ISE_w	$\hat{h}_{CV}$	0.2	8.7	24.5	55.0	1,245.7	46.4	55.1	104.6
ISE_w	$\hat{h}_{CV,B_1}$	0.1	5.7	15.1	33.4	498.5	27.7	33.7	58.7
ISE_w	$\hat{h}_{CV,B_2}$	0.1	3.9	10.8	24.4	375.1	20.6	20.4	31.7
ISE_w	$\hat{h}^*$	0.1	4.9	12.4	27.9	225.0	23.0	21.1	25.1
ISE_w	$\hat{h}_P^*$	0.1	5.2	11.6	24.5	391.0	19.4	21.0	32.3
$\hat{h}$	$\hat{h}_{CV}$	0.010	0.610	1.170	1.730	3.310	1.120	1.213	0.766
$\hat{h}$	$\hat{h}_{CV,B_1}$	0.035	0.692	1.053	1.484	2.603	0.792	1.071	0.528
$\hat{h}$	$\hat{h}_{CV,B_2}$	0.173	0.796	1.036	1.298	2.418	0.502	1.058	0.371
$\hat{h}$	$\hat{h}^*$	0.340	0.570	0.820	1.200	2.430	0.630	0.580	0.420
$\hat{h}$	$\hat{h}_P^*$	0.150	0.900	1.275	1.640	2.550	0.740	1.269	0.441

Table 5.29: *Simulation summary : Weibull(2, 1), $\eta = 0, n = 200, MC = 1\ 000$*

<i>Value</i>	<i>Selector</i>	<i>Min</i>	q_1	<i>Median</i>	q_3	<i>Max</i>	<i>iqr</i>	<i>Mean</i>	<i>SD</i>
ISE_w	$\hat{h}_{CV}$	0.3	11.6	31.8	85.8	1,764.9	74.3	84.9	169.5
ISE_w	$\hat{h}_{CV,B_1}$	0.1	7.6	19.8	47.8	1,009.0	40.2	47.3	83.8
ISE_w	$\hat{h}_{CV,B_2}$	0.2	5.8	15.4	35.9	543.9	30.1	31.3	52.4
ISE_w	$\hat{h}^*$	0.1	7.2	17.7	40.2	417.8	33.0	32.6	43.7
ISE_w	$\hat{h}_P^*$	0.1	6.0	16.3	38.5	1,416.9	32.5	34.2	64.2
$\hat{h}$	$\hat{h}_{CV}$	0.010	0.500	0.980	1.385	3.810	0.885	0.992	0.633
$\hat{h}$	$\hat{h}_{CV,B_1}$	0.044	0.557	0.827	1.140	2.324	0.583	0.857	0.418
$\hat{h}$	$\hat{h}_{CV,B_2}$	0.122	0.653	0.825	1.046	2.295	0.393	0.859	0.297
$\hat{h}$	$\hat{h}^*$	0.270	0.520	0.755	1.080	2.110	0.560	0.520	0.367
$\hat{h}$	$\hat{h}_P^*$	0.110	0.780	1.070	1.290	2.100	0.510	1.047	0.332

Table 5.30: *Simulation summary : Weibull(2, 1), $\eta = 1/3, n = 200, MC = 1\ 000$*

<i>Value</i>	<i>Selector</i>	<i>Min</i>	q_1	<i>Median</i>	q_3	<i>Max</i>	<i>iqr</i>	<i>Mean</i>	<i>SD</i>
ISE_w	$\hat{h}_{CV}$	0.2	19.0	61.5	133.1	2,018.0	114.1	110.8	181.5
ISE_w	$\hat{h}_{CV,B_1}$	0.4	15.4	45.3	95.9	731.3	80.5	72.0	88.2
ISE_w	$\hat{h}_{CV,B_2}$	0.5	16.5	38.2	81.4	512.3	64.9	60.2	64.3
ISE_w	$\hat{h}^*$	0.6	9.5	23.6	49.4	477.0	40.0	39.8	50.2
ISE_w	$\hat{h}_P^*$	0.6	21.1	49.6	94.4	488.1	73.3	67.8	65.3
$\hat{h}$	$\hat{h}_{CV}$	0.010	0.380	0.620	1.050	2.780	0.670	0.788	0.597
$\hat{h}$	$\hat{h}_{CV,B_1}$	0.035	0.409	0.601	0.984	1.994	0.575	0.701	0.396
$\hat{h}$	$\hat{h}_{CV,B_2}$	0.160	0.539	0.686	0.888	1.780	0.349	0.728	0.268
$\hat{h}$	$\hat{h}^*$	0.210	0.300	0.360	0.455	1.290	0.155	0.330	0.161
$\hat{h}$	$\hat{h}_P^*$	0.150	0.600	0.820	1.010	1.870	0.410	0.820	0.272

Table 5.31: *Simulation summary : Weibull(3, 1), $\eta = 0, n = 200, MC = 1\ 000$*

<i>Value</i>	<i>Selector</i>	<i>Min</i>	q_1	<i>Median</i>	q_3	<i>Max</i>	<i>iqr</i>	<i>Mean</i>	<i>SD</i>
ISE_w	$\hat{h}_{CV}$	0.7	30.3	82.6	157.7	2,356.3	127.5	146.6	235.3
ISE_w	$\hat{h}_{CV,B_1}$	0.5	19.5	45.5	94.9	2,683.8	75.4	88.0	165.3
ISE_w	$\hat{h}_{CV,B_2}$	0.3	15.6	36.3	79.9	1,828.3	64.3	64.7	97.8
ISE_w	$\hat{h}^*$	0.1	14.3	34.2	68.3	511.2	54.1	56.2	68.0
ISE_w	$\hat{h}_P^*$	0.2	20.3	43.0	87.5	682.5	67.2	70.0	83.5
$\hat{h}$	$\hat{h}_{CV}$	0.010	0.320	0.580	1.000	2.370	0.680	0.679	0.480
$\hat{h}$	$\hat{h}_{CV,B_1}$	0.017	0.361	0.531	0.766	1.671	0.405	0.575	0.294
$\hat{h}$	$\hat{h}_{CV,B_2}$	0.066	0.433	0.563	0.727	1.503	0.294	0.586	0.217
$\hat{h}$	$\hat{h}^*$	0.180	0.300	0.380	0.550	1.170	0.250	0.330	0.190
$\hat{h}$	$\hat{h}_P^*$	0.200	0.565	0.700	0.820	1.410	0.255	0.693	0.195

Table 5.32: *Simulation summary : Weibull(3, 1), $\eta = 1/3, n = 200, MC = 1\ 000$*

<i>Value</i>	<i>Selector</i>	<i>Min</i>	q_1	<i>Median</i>	q_3	<i>Max</i>	<i>iqr</i>	<i>Mean</i>	<i>SD</i>
ISE_w	$\hat{h}_{CV}$	38.6	130.2	169.8	221.2	501.7	91.0	181.1	72.2
ISE_w	$\hat{h}_{CV,B_1}$	17.1	125.5	164.5	210.1	548.3	84.6	171.4	65.0
ISE_w	$\hat{h}_{CV,B_2}$	17.8	123.0	159.8	204.6	590.9	81.5	167.7	63.6
ISE_w	$\hat{h}^*$	32.7	119.4	151.8	185.6	469.0	66.2	155.5	53.6
ISE_w	$\hat{h}_P^*$	42.3	120.9	161.0	200.5	479.6	79.6	164.8	58.6
$\hat{h}$	$\hat{h}_{CV}$	0.040	0.555	0.780	1.000	2.160	0.445	0.772	0.350
$\hat{h}$	$\hat{h}_{CV,B_1}$	0.122	0.627	0.818	0.975	2.498	0.348	0.801	0.257
$\hat{h}$	$\hat{h}_{CV,B_2}$	0.316	0.660	0.792	0.921	1.965	0.260	0.804	0.208
$\hat{h}$	$\hat{h}^*$	0.940	1.140	1.220	1.350	5.010	0.210	1.200	0.383
$\hat{h}$	$\hat{h}_P^*$	0.320	0.820	0.960	1.110	1.980	0.290	0.971	0.237

Table 5.33: *Simulation summary : $\lognormal(0, 1), \eta = 0, n = 200, MC = 1\ 000$*

<i>Value</i>	<i>Selector</i>	<i>Min</i>	q_1	<i>Median</i>	q_3	<i>Max</i>	<i>iqr</i>	<i>Mean</i>	<i>SD</i>
ISE_w	$\hat{h}_{CV}$	15.0	126.7	168.2	227.4	1, 243.6	100.6	186.9	99.0
ISE_w	$\hat{h}_{CV,B_1}$	26.3	116.7	163.0	223.7	673.8	107.0	176.3	79.8
ISE_w	$\hat{h}_{CV,B_2}$	28.6	112.4	158.3	213.7	647.8	101.3	168.4	74.8
ISE_w	$\hat{h}^*$	20.9	113.5	150.3	192.5	376.8	79.1	155.2	57.4
ISE_w	$\hat{h}_P^*$	22.3	115.8	157.6	207.3	427.9	91.4	164.8	69.5
$\hat{h}$	$\hat{h}_{CV}$	0.010	0.560	0.850	1.090	2.590	0.530	0.839	0.403
$\hat{h}$	$\hat{h}_{CV,B_1}$	0.087	0.696	0.905	1.088	3.021	0.392	0.906	0.346
$\hat{h}$	$\hat{h}_{CV,B_2}$	0.252	0.752	0.897	1.073	2.290	0.322	0.936	0.300
$\hat{h}$	$\hat{h}^*$	0.810	1.070	1.140	1.220	2.600	0.150	1.120	0.152
$\hat{h}$	$\hat{h}_P^*$	0.300	0.910	1.110	1.315	3.840	0.405	1.137	0.341

Table 5.34: *Simulation summary : $\lognormal(0, 1), \eta = 1/3, n = 200, MC = 1\ 000$*

<i>Value</i>	<i>Selector</i>	<i>Min</i>	q_1	<i>Median</i>	q_3	<i>Max</i>	<i>iqr</i>	<i>Mean</i>	<i>SD</i>
ISE_w	$\hat{h}_{CV}$	10.7	63.5	92.4	130.0	848.9	66.6	108.6	75.0
ISE_w	$\hat{h}_{CV,B_1}$	8.4	59.7	86.1	120.9	362.4	61.2	93.6	45.9
ISE_w	$\hat{h}_{CV,B_2}$	7.8	56.5	80.4	112.4	275.6	55.9	87.6	41.9
ISE_w	$\hat{h}^*$	14.3	55.5	77.3	100.6	195.4	45.1	80.1	31.7
ISE_w	$\hat{h}_P^*$	13.5	58.7	80.9	110.9	319.8	52.2	87.0	39.2
$\hat{h}$	$\hat{h}_{CV}$	0.020	0.560	0.880	1.140	3.160	0.580	0.859	0.442
$\hat{h}$	$\hat{h}_{CV,B_1}$	0.070	0.714	0.905	1.093	4.361	0.379	0.903	0.337
$\hat{h}$	$\hat{h}_{CV,B_2}$	0.305	0.765	0.916	1.100	2.593	0.335	0.949	0.294
$\hat{h}$	$\hat{h}^*$	0.880	1.070	1.150	1.250	2.820	0.180	1.120	0.156
$\hat{h}$	$\hat{h}_P^*$	0.310	0.950	1.120	1.290	4.900	0.340	1.153	0.350

Table 5.35: *Simulation summary : $\lognormal(0.25, 0.75), \eta = 0, n = 200, MC = 1\ 000$*

<i>Value</i>	<i>Selector</i>	<i>Min</i>	q_1	<i>Median</i>	q_3	<i>Max</i>	<i>iqr</i>	<i>Mean</i>	<i>SD</i>
ISE_w	$\hat{h}_{CV}$	7.1	64.9	98.2	147.5	629.8	82.6	118.4	81.4
ISE_w	$\hat{h}_{CV,B_1}$	6.8	57.3	91.3	134.0	414.9	76.7	101.8	58.3
ISE_w	$\hat{h}_{CV,B_2}$	7.6	54.0	84.2	118.7	417.3	64.7	91.5	50.2
ISE_w	$\hat{h}^*$	10.4	55.9	81.2	111.7	249.2	55.8	86.2	41.5
ISE_w	$\hat{h}_P^*$	6.6	56.4	83.1	115.6	362.0	59.2	89.8	46.7
$\hat{h}$	$\hat{h}_{CV}$	0.040	0.595	0.965	1.280	4.710	0.685	0.988	0.595
$\hat{h}$	$\hat{h}_{CV,B_1}$	0.131	0.740	0.975	1.206	3.961	0.466	1.018	0.480
$\hat{h}$	$\hat{h}_{CV,B_2}$	0.172	0.850	1.038	1.303	2.948	0.453	1.110	0.397
$\hat{h}$	$\hat{h}^*$	0.770	1.060	1.160	1.280	4.540	0.220	1.120	0.349
$\hat{h}$	$\hat{h}_P^*$	0.320	1.040	1.250	1.500	4.720	0.460	1.313	0.445

Table 5.36: *Simulation summary : $\lognormal(0.25, 0.75), \eta = 1/3, n = 200, MC = 1\ 000$*

<i>Value</i>	<i>Selector</i>	<i>Min</i>	q_1	<i>Median</i>	q_3	<i>Max</i>	<i>iqr</i>	<i>Mean</i>	<i>SD</i>
ISE_w	$\hat{h}_{CV}$	5.6	21.5	35.8	65.0	1,345.8	43.5	60.1	88.2
ISE_w	$\hat{h}_{CV,B_1}$	3.0	19.8	31.3	51.9	427.6	32.1	42.2	37.06
ISE_w	$\hat{h}_{CV,B_2}$	3.1	18.5	28.0	42.9	187.3	24.4	34.7	25.0
ISE_w	$\hat{h}^*$	3.8	18.5	26.6	40.2	143.9	21.7	31.9	19.4
ISE_w	$\hat{h}_P^*$	4.9	18.2	27.2	39.4	244.4	21.2	33.2	24.05
$\hat{h}$	$\hat{h}_{CV}$	0.010	0.515	0.965	1.350	4.890	0.835	1.088	0.827
$\hat{h}$	$\hat{h}_{CV,B_1}$	0.061	0.731	0.966	1.254	3.447	0.522	1.058	0.555
$\hat{h}$	$\hat{h}_{CV,B_2}$	0.222	0.845	1.066	1.412	3.182	0.566	1.164	0.454
$\hat{h}$	$\hat{h}^*$	0.690	0.920	1.030	1.190	3.920	0.270	0.990	0.351
$\hat{h}$	$\hat{h}_P^*$	0.280	1.010	1.240	1.570	4.890	0.560	1.363	0.594

Table 5.37: *Simulation summary : $\text{lognormal}(0.5, 0.5), \eta = 0, n = 200, MC = 1\,000$*

<i>Value</i>	<i>Selector</i>	<i>Min</i>	q_1	<i>Median</i>	q_3	<i>Max</i>	<i>iqr</i>	<i>Mean</i>	<i>SD</i>
ISE_w	$\hat{h}_{CV}$	3.8	23.7	39.4	72.6	3,480.5	48.9	75.2	145.9
ISE_w	$\hat{h}_{CV,B_1}$	3.1	20.4	33.5	59.8	897.0	39.4	52.5	62.0
ISE_w	$\hat{h}_{CV,B_2}$	2.1	18.2	29.3	47.6	317.4	29.3	38.3	31.2
ISE_w	$\hat{h}^*$	3.9	19.7	30.4	45.2	195.9	25.4	35.9	23.6
ISE_w	$\hat{h}_P^*$	2.9	18.8	28.5	45.3	416.4	26.5	37.8	33.5
$\hat{h}$	$\hat{h}_{CV}$	0.010	0.570	1.020	1.475	4.700	0.905	1.160	0.837
$\hat{h}$	$\hat{h}_{CV,B_1}$	0.061	0.757	1.019	1.410	3.700	0.653	1.120	0.598
$\hat{h}$	$\hat{h}_{CV,B_2}$	0.278	0.879	1.135	1.433	3.027	0.554	1.194	0.444
$\hat{h}$	$\hat{h}^*$	0.610	0.920	1.070	1.360	3.390	0.440	0.990	0.475
$\hat{h}$	$\hat{h}_P^*$	0.180	1.040	1.300	1.640	3.760	0.600	1.415	0.563

Table 5.38: *Simulation summary : $\text{lognormal}(0.5, 0.5), \eta = 1/3, n = 200, MC = 1\,000$*

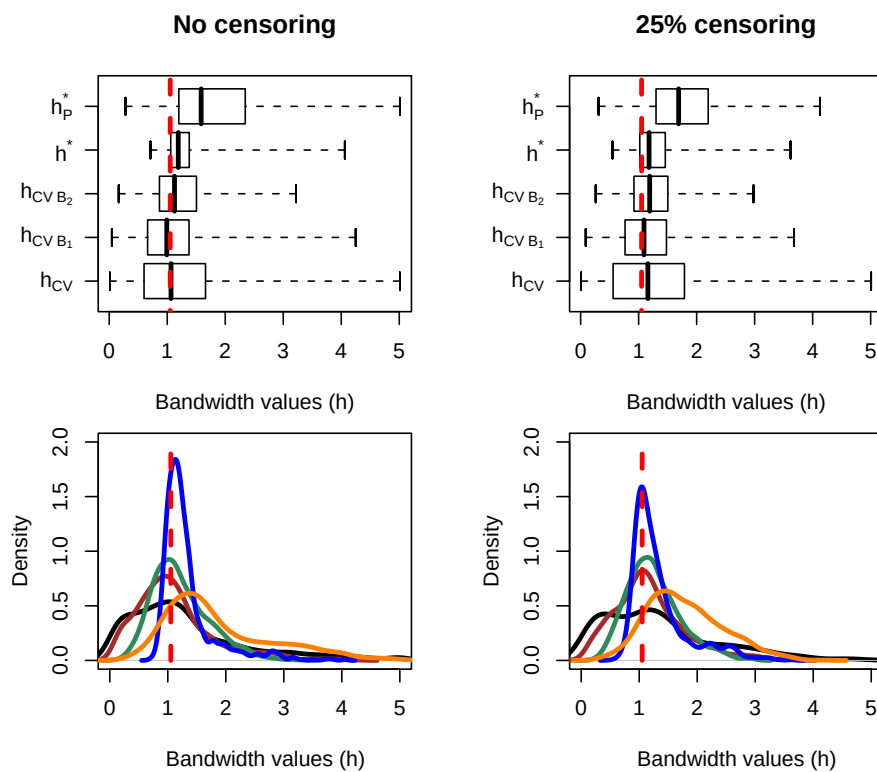


Figure 5.3: *Bandwidth distribution for Weibull(1,1), $n=50$.*

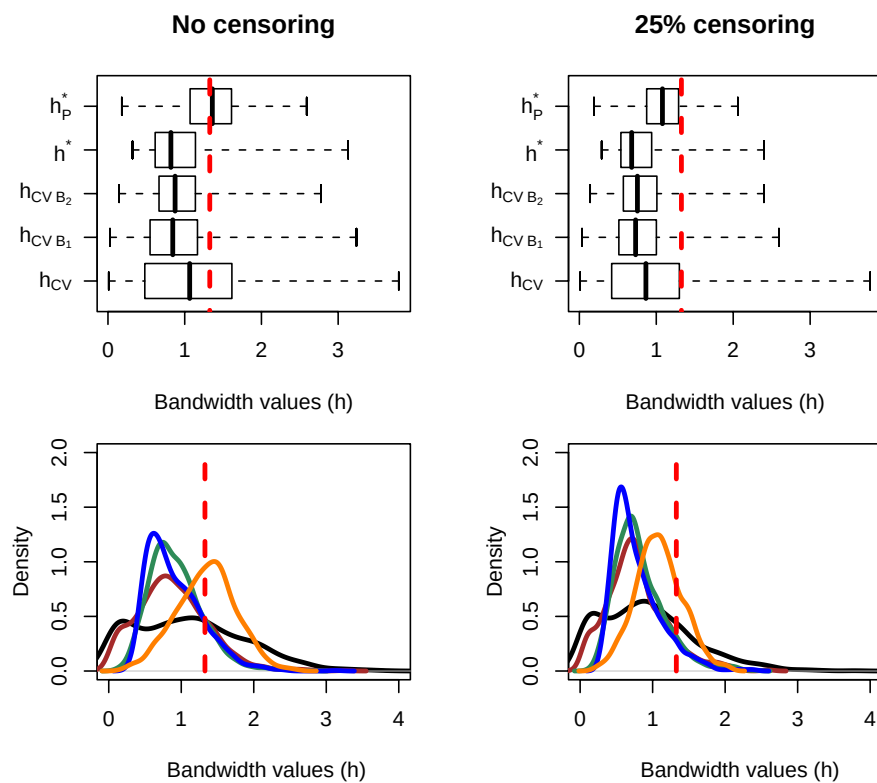


Figure 5.4: *Bandwidth distribution for Weibull(2,1), $n=50$.*

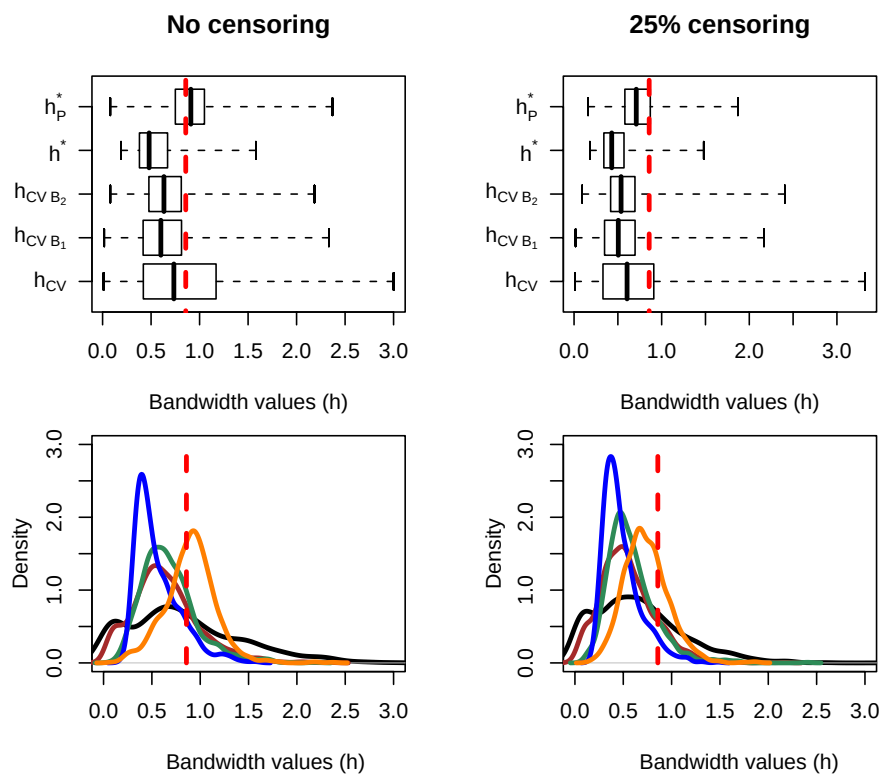


Figure 5.5: *Bandwidth distribution for Weibull(3,1), $n=50$.*

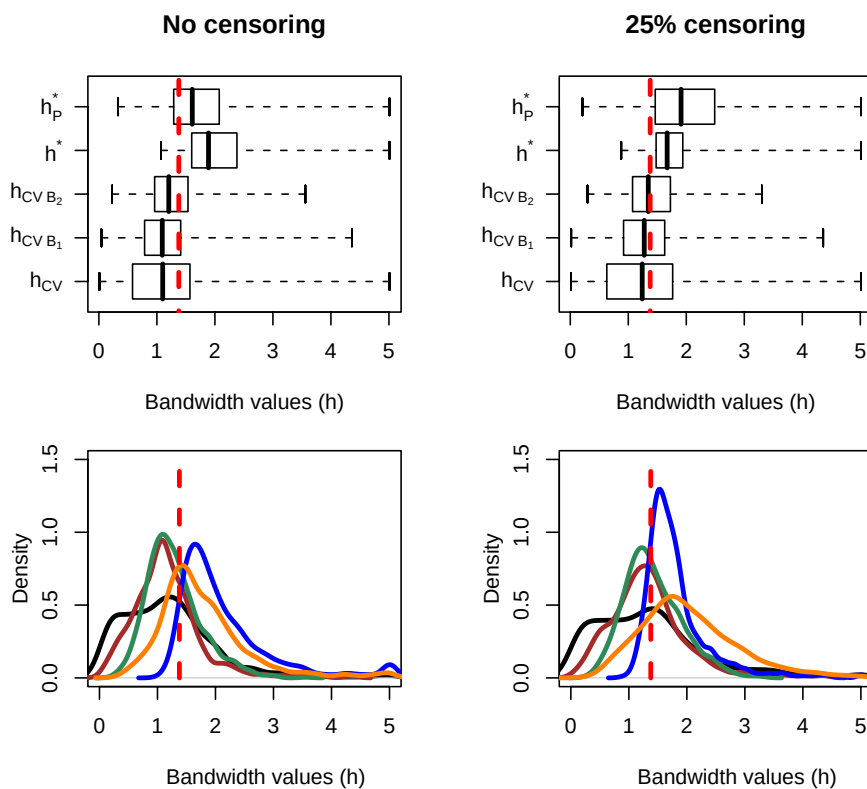
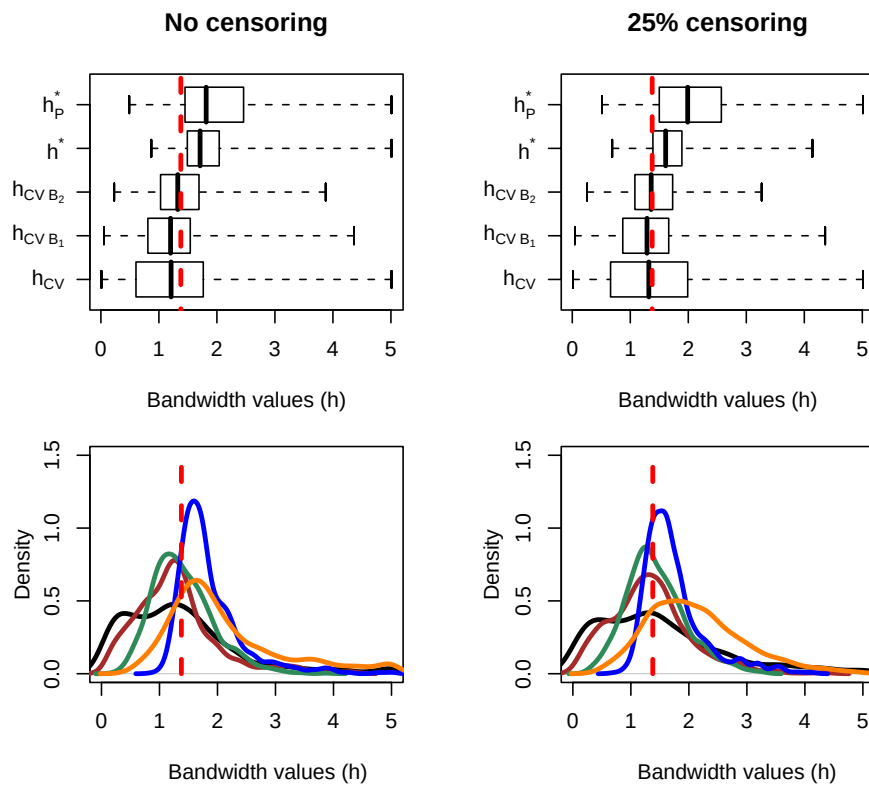
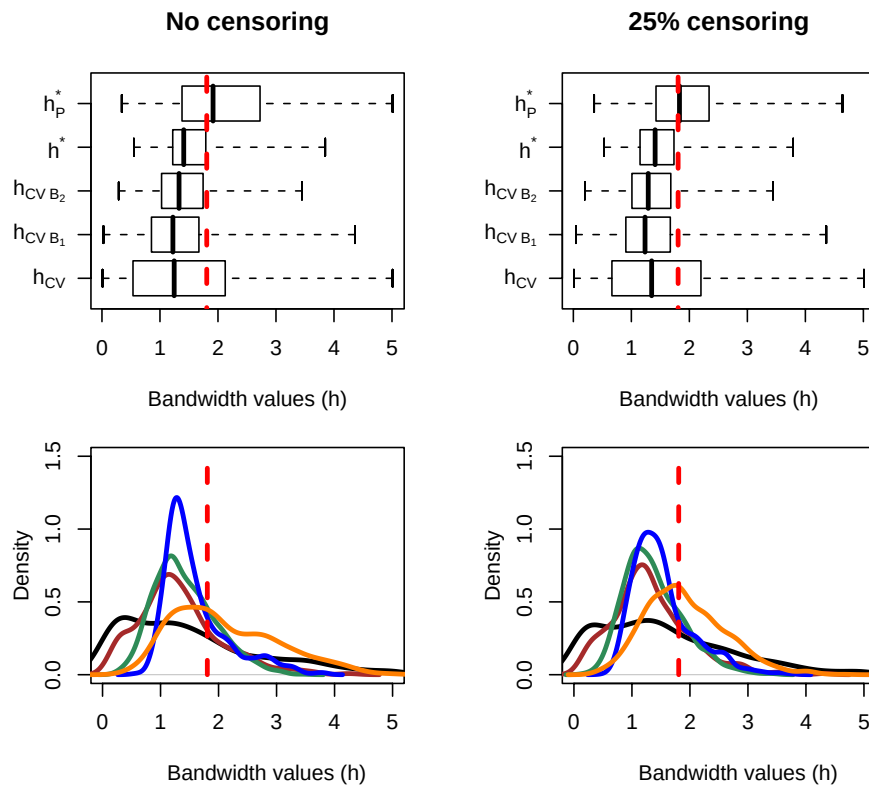
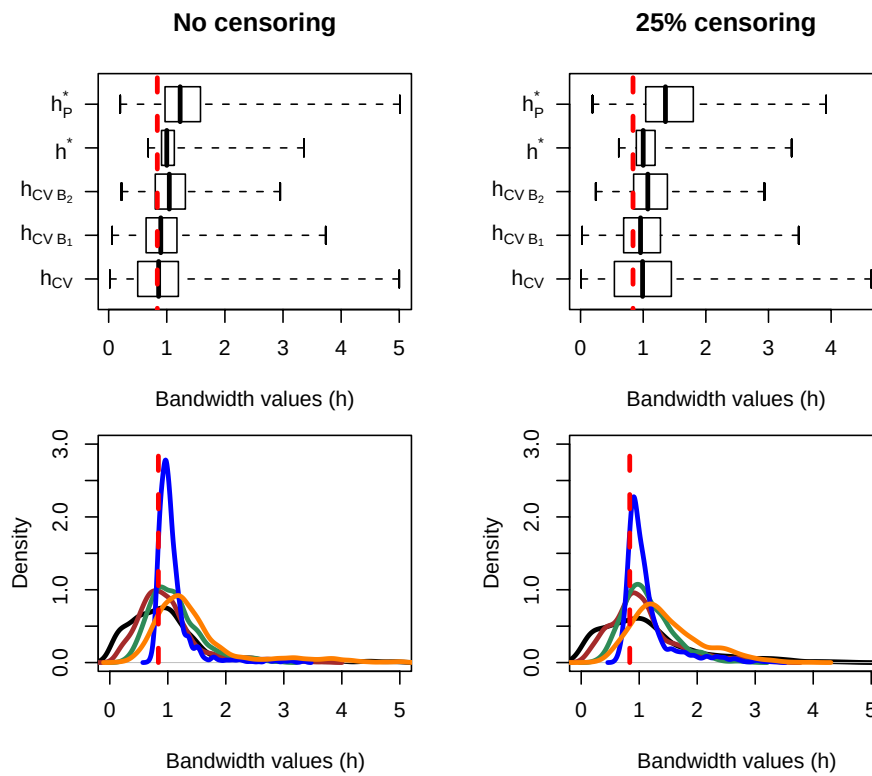
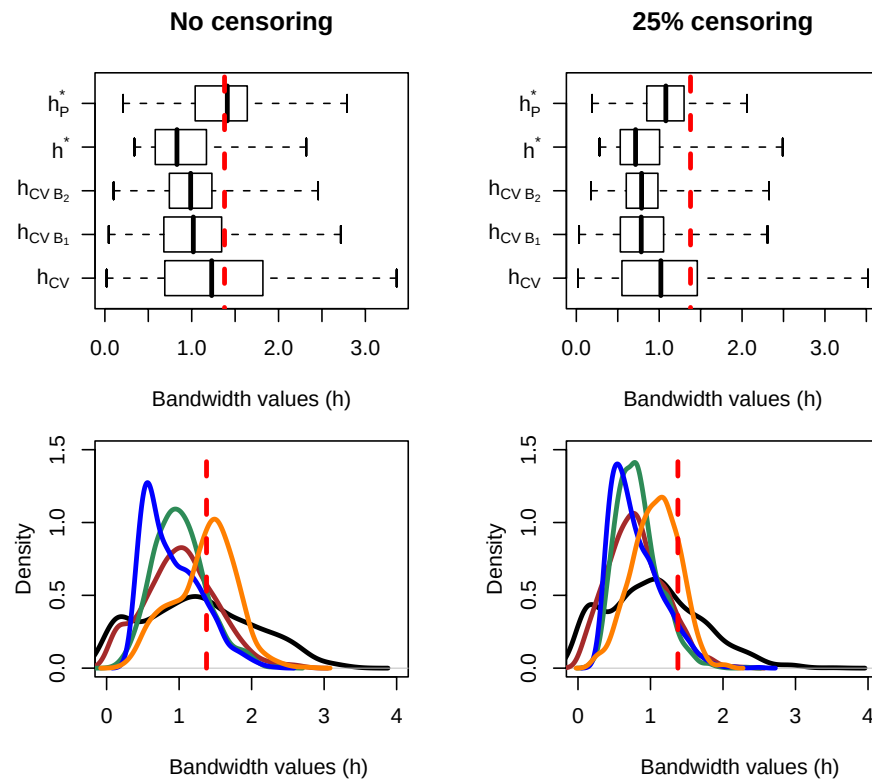


Figure 5.6: *Bandwidth distribution for lognormal(0,1), $n=50$.*

Figure 5.7: *Bandwidth distribution for $\text{lognormal}(0.25, 0.75)$, $n=50$.*Figure 5.8: *Bandwidth distribution for $\text{lognormal}(0.5, 0.5)$, $n=50$.*

Figure 5.9: *Bandwidth distribution for Weibull(1,1), $n=100$.*Figure 5.10: *Bandwidth distribution for Weibull(2,1), $n=100$.*

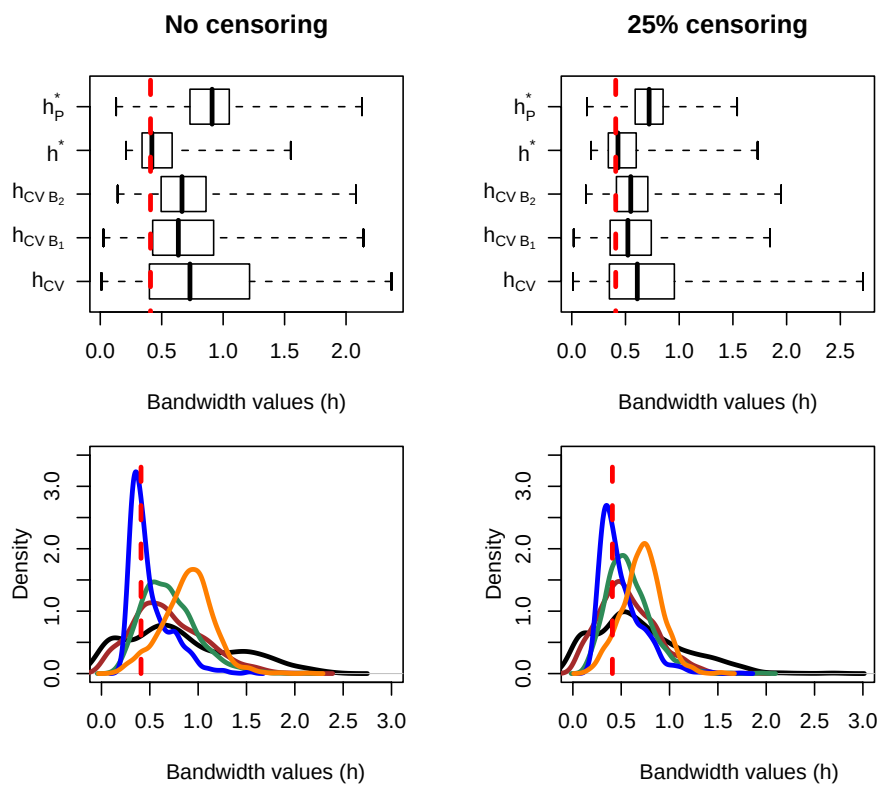


Figure 5.11: *Bandwidth distribution for Weibull(3,1), $n=100$.*

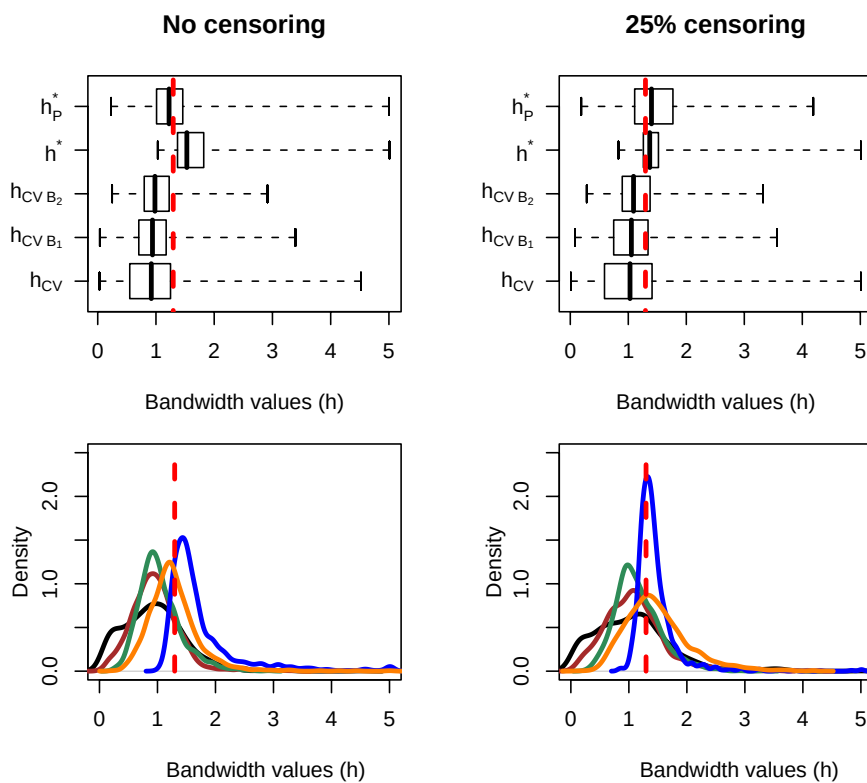


Figure 5.12: *Bandwidth distribution for lognormal(0,1), $n=100$.*

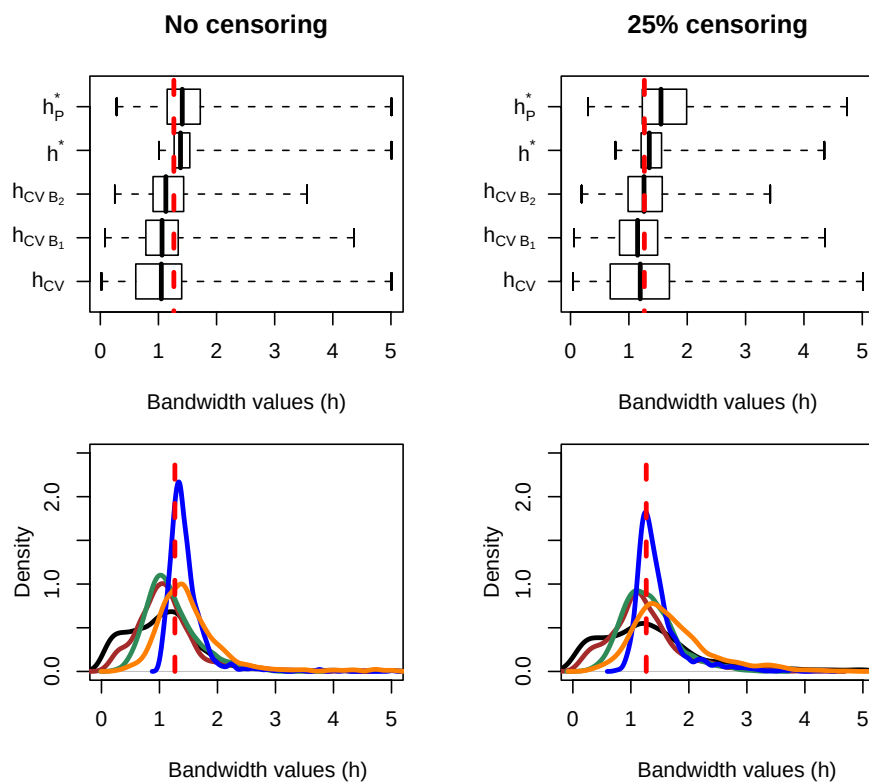


Figure 5.13: *Bandwidth distribution for $\text{lognormal}(0.25, 0.75)$, $n=100$.*

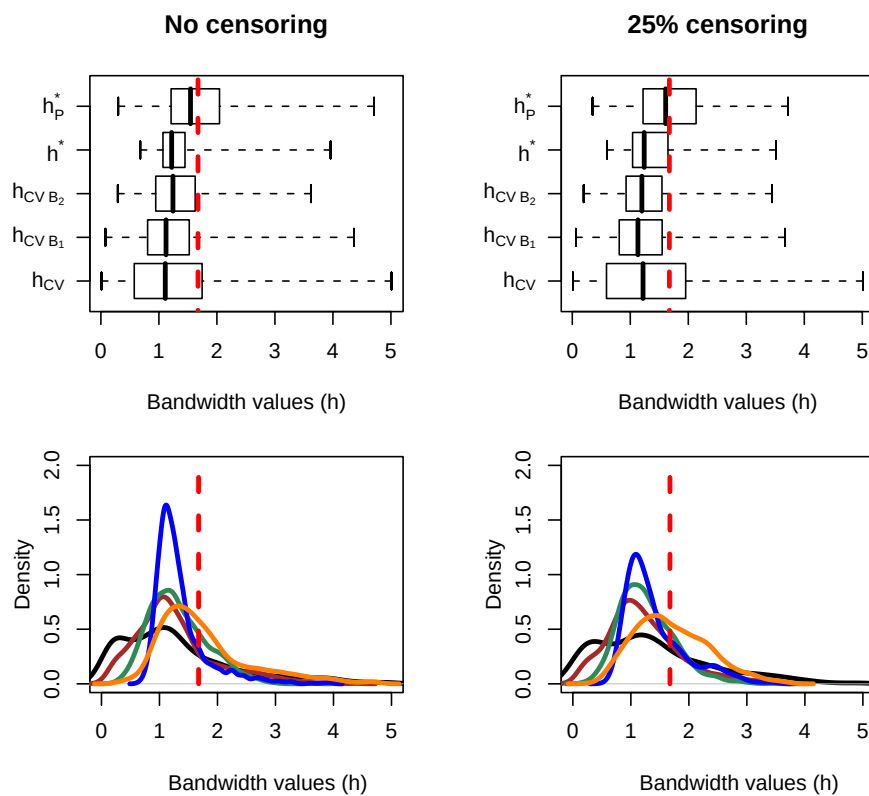


Figure 5.14: *Bandwidth distribution for $\text{lognormal}(0.5, 0.5)$, $n=100$.*

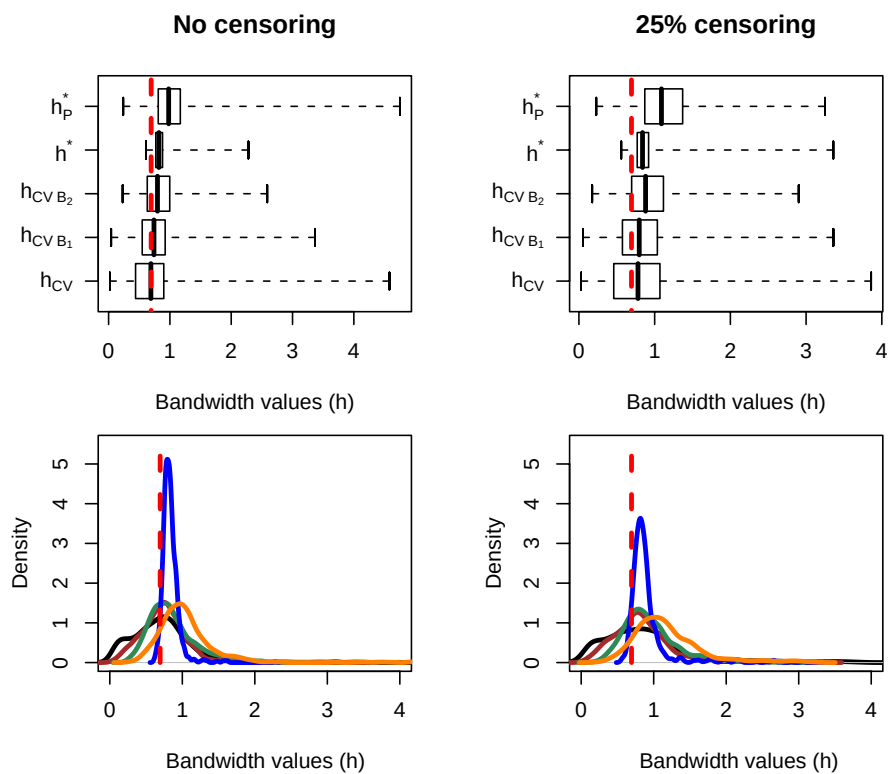


Figure 5.15: *Bandwidth distribution for Weibull(1,1), $n=200$.*

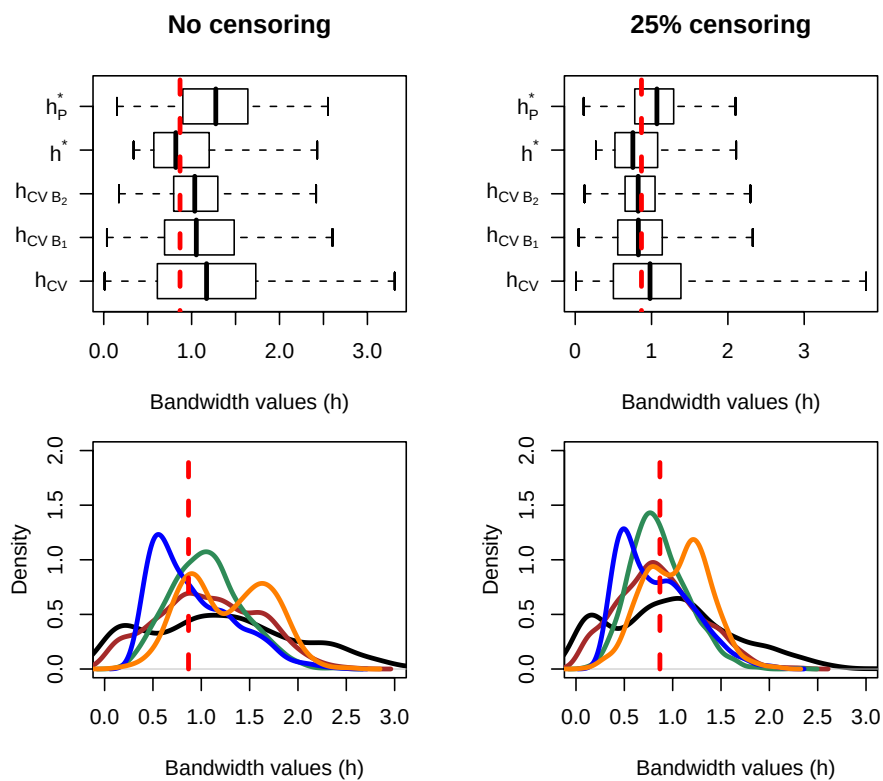


Figure 5.16: *Bandwidth distribution for Weibull(2,1), $n=200$.*

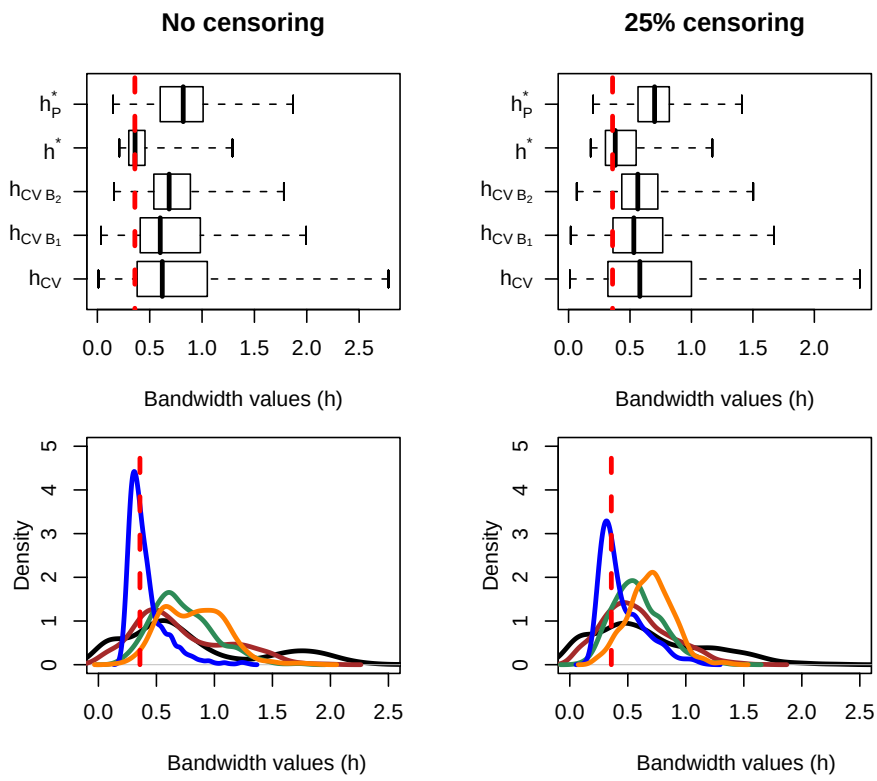


Figure 5.17: *Bandwidth distribution for Weibull(3,1), $n=200$.*

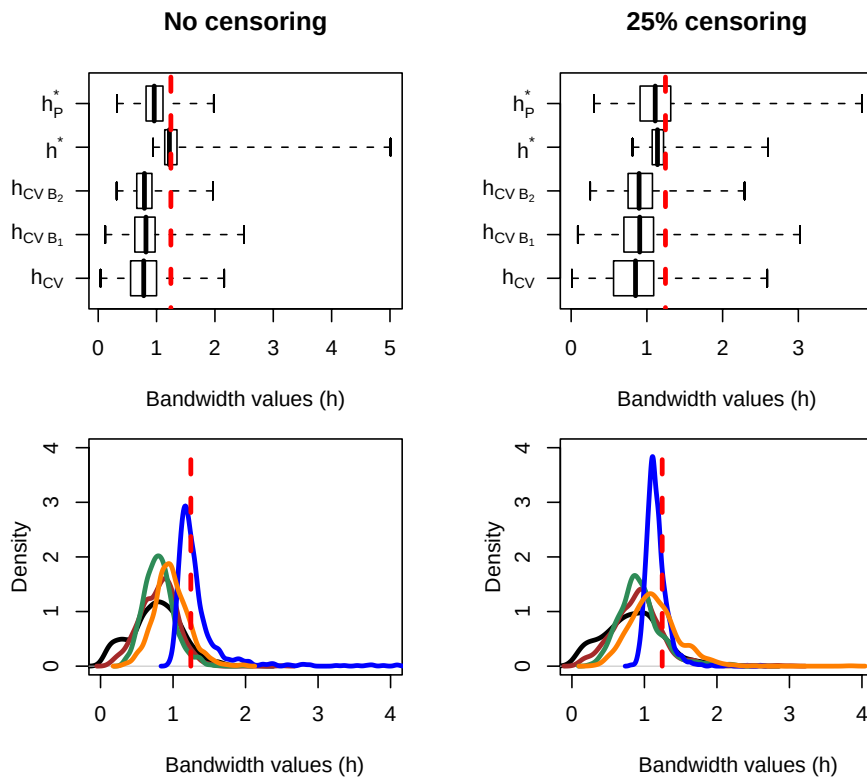
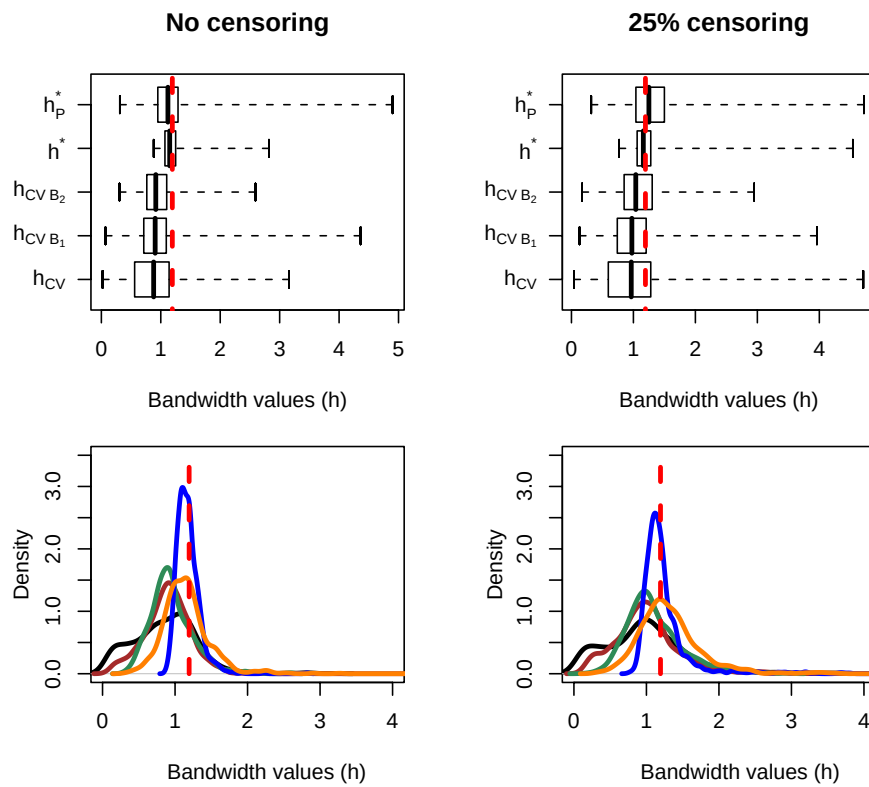
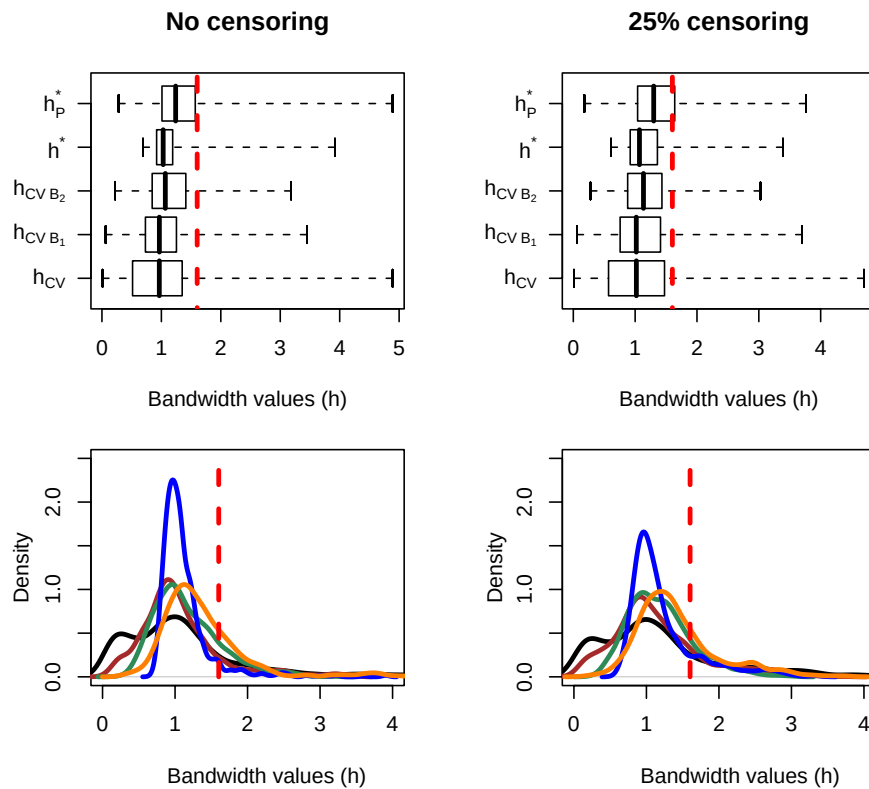


Figure 5.18: *Bandwidth distribution for lognormal(0,1), $n=200$.*

Figure 5.19: *Bandwidth distribution for $\text{lognormal}(0.25, 0.75)$, $n=200$.*Figure 5.20: *Bandwidth distribution for $\text{lognormal}(0.5, 0.5)$, $n=200$.*

5.5 Conclusions

In Section 5.5.1 we will show summarizing tables of the ranking results for both the ISE_w and $\hat{h}$ values of the five selectors. By studying the tables and graphs from Section 5.4.2 and interpreting the ranking results to be shown, the reader should have a clear understanding of the performance of each selector in each of the different settings. We will then make some remarks to conclude the comparison of the five selectors, and list some open problems for future research.

Section 5.5.2 discusses the performance of the chosen proportional censorship model, as well as findings regarding the improvement shown by the selector $\hat{h}_{CV,B_2}$ when compared to $\hat{h}_{CV}$ as a result of implementing bagging.

5.5.1 Main simulation study

Table 5.39 shows the results of the ranking procedure of the ISE_w values, Table 5.40 shows the results of the ranking procedure of the $\hat{h}$ values and Table 5.41 combines these results in order to rank the overall performance of the five selectors.

Selector	ISE_w mean rank	ISE_w median rank	ISE_w SD rank	ISE_w overall rank	Result
$\hat{h}_{CV}$	5.00	5.00	5.00	5.00	5th
$\hat{h}_{CV,B_1}$	4.00	3.83	4.00	3.94	4th
$\hat{h}_{CV,B_2}$	2.28	1.89	2.56	2.24	2nd
$\hat{h}^*$	1.39	1.50	1.33	1.41	1st
$\hat{h}_P^*$	2.31	2.78	2.08	2.39	3rd

Table 5.39: *Ranking method results for ISE_w values.*

Selector	Rank 1 %	Rank 2 %	Rank 3 %	Rank 4 %	Rank 5 %	Mean rank	Result
$\hat{h}_{CV}$	0.00%	0.00%	2.78%	36.11%	61.11%	4.58	5th
$\hat{h}_{CV,B_1}$	0.00%	2.78%	75.00%	22.22%	0.00%	3.19	4th
$\hat{h}_{CV,B_2}$	33.33%	58.33%	8.33%	0.00%	0.00%	1.75	1st
$\hat{h}^*$	44.44%	19.44%	8.33%	11.11%	16.67%	2.36	2nd
$\hat{h}_P^*$	22.22%	19.44%	5.56%	30.56%	22.22%	3.11	3rd

Table 5.40: *Relative frequency distribution and results of $\hat{h}$ $M\tilde{S}E$ rankings.*

Selector	ISE _w rank	$\hat{h}$ rank	Mean rank	Final result
$\hat{h}_{CV}$	5.00	4.58	4.79	5th
$\hat{h}_{CV,B_1}$	3.94	3.19	3.57	4th
$\hat{h}_{CV,B_2}$	2.24	1.75	2.00	2nd
$\hat{h}^*$	1.41	2.36	1.89	1st
$\hat{h}_P^*$	2.39	3.11	2.75	3rd

Table 5.41: *Final ranking of the five selectors.***Remarks:**

1. One can clearly see that LSCV (that is selector $\hat{h}_{CV}$) is the worst performing (5th) in terms of both the ISE_w and $\hat{h}$ rankings. This selector is the most basic selector considered, and as mentioned in Section 4.4.1 the convergence rate of h_{CV} to h_{MISE_w} is unacceptably slow at $n^{-1/10}$. We also noted that the variability of $\hat{h}_{CV}$ is much higher than the other four selectors, and we can therefore conclude with conviction that all the other selectors are superior to $\hat{h}_{CV}$.
2. The selector $\hat{h}_{CV,B_1}$ consistently performed better than $\hat{h}_{CV}$, but did not match up well against the remaining three selectors. This justifies the final ranking of 4th for this selector. The introduction of bagging leads to an improvement upon LSCV, but in the context of our simulation study bagging method 2 leads to an even bigger improvement upon LSCV than method 1. We therefore recommend that if bagging is considered, method 2 should be used.
3. The proposed selector $\hat{h}_P^*$ performed favorably against the other selectors. For some settings this selector outperformed the rest, but there were some settings where $\hat{h}_P^*$ performed the worst. The selector performed very well in terms of variation in the ISE_w values, and was not far behind $\hat{h}_{CV,B_2}$ in that regard. Overall we conclude that $\hat{h}_P^*$ competes fairly well with other existing methods.
4. The bagging method 2 selector $\hat{h}_{CV,B_2}$ consistently performed better than $\hat{h}_{CV}$, $\hat{h}_{CV,B_1}$ and $\hat{h}_P^*$. In terms of the $\hat{h}$ values, $\hat{h}_{CV,B_2}$ was actually the best performing selector. If we take into consideration that we used a rather conservative choice of $B = 100$ as the number of bootstrap iterations to calculate $\hat{h}_{CV,B_2}$, as well as the fact that we only implemented half-sample bagging (that is $m/n = 0.5$), one could reason that we can improve the performance of $\hat{h}_{CV,B_2}$ by increasing the values of B and m/n . But based

on these choices and the results obtained we conclude that $\hat{h}_{CV,B_2}$ is the second best selector.

5. The selector proposed by Gonzalez-Manteiga et al. (1996), $\hat{h}^*$, is the the best and most consistent of the five selectors considered. The explicit bootstrap expression for $AMISE_w^*$ is a big advantage of this selector. The proposed choices of pilot bandwidths g_1 and g_2 make use of the normal distribution as reference distribution, and is an area where the selector might be improved upon. In terms of ISE_w values, $\hat{h}^*$ performs significantly better than the other selectors. When interpreting the $\hat{h}$ values, we see that $\hat{h}^*$ is either 1st or 2nd in most cases, but there are some settings where it ranked 3rd or worse. All things considered we conclude that this selector is the recommended choice.

Future studies:

1. The effect of varying values of B and m/n on the performance of the bagging selectors $\hat{h}_{CV,B_1}$ and $\hat{h}_{CV,B_2}$ could be very informative. As mentioned, one expects the estimation to improve as we use more bootstrap samples, and include more values in those bootstrap samples.
2. As mentioned in Section 4.4.4, the current procedure of determining $\tilde{g}_1$ and $\tilde{g}_2$ for use in the selector $\hat{h}_P^*$ was not developed in the setting of random right censoring. Developing an appropriate nonparametric method of determining these pilot bandwidths can lead to an improvement in the performance of $\hat{h}_P^*$.
3. Upon inspection of generated smooth bootstrap samples, we observed that smooth bootstrap sample values could be negative. This can occur when using any of the formulas given in (2.8) - (2.10) together with a symmetric kernel with properties as described in Section 3.3.2. This is not a desirable result in the study of lifetimes, as survival variables assume strictly positive values. One method that can be used to correct this is the use of *positive* kernel functions. The effect of this choice on the theoretical properties and empirical performance of the selectors discussed can be studied.

These are but a few interesting problems that we encountered during our study, and there are certainly many more avenues of future research not mentioned here.

5.5.2 Secondary studies

Empirical censoring proportion versus censoring probability of 25%

Tables 5.42 - 5.44 show how well the generated samples adhere to this distribution for each of the six distributions used in the simulation study (for $n = 50, 100, 200$). For each distribution, 4 000 iterations were used to calculate the sample statistics. From the tables it is clear that the sample statistics tend to the theoretical values as the sample size increases. In each instance the variance values have been multiplied by a factor of 10^3 .

Distribution	$E(\hat{p})$	$Var(\hat{p})$	Sample mean	Sample variance
Wei(1,1)	0.7500	3.750	0.7495	3.658
Wei(2,1)	0.7500	3.750	0.7498	3.732
Wei(3,1)	0.7500	3.750	0.7513	3.953
LN(0,1)	0.7500	3.750	0.7490	3.826
LN(0.25,0.75)	0.7500	3.750	0.7514	3.684
LN(0.5,0.5)	0.7500	3.750	0.7494	3.708

Table 5.42: *Distribution properties of $\hat{p}$, $MC = 4\ 000, n = 50$.*

Distribution	$E(\hat{p})$	$Var(\hat{p})$	Sample mean	Sample variance
Wei(1,1)	0.7500	1.875	0.7502	1.926
Wei(2,1)	0.7500	1.875	0.7507	1.908
Wei(3,1)	0.7500	1.875	0.7499	1.904
LN(0,1)	0.7500	1.875	0.7496	1.828
LN(0.25,0.75)	0.7500	1.875	0.7512	1.921
LN(0.5,0.5)	0.7500	1.875	0.7503	1.858

Table 5.43: *Distribution properties of $\hat{p}$, $MC = 4\ 000, n = 100$.*

Distribution	$E(\hat{p})$	$Var(\hat{p})$	Sample mean	Sample variance
Wei(1,1)	0.7500	0.9375	0.7497	0.9310
Wei(2,1)	0.7500	0.9375	0.7505	0.9568
Wei(3,1)	0.7500	0.9375	0.7498	0.9376
LN(0,1)	0.7500	0.9375	0.7499	0.9506
LN(0.25,0.75)	0.7500	0.9375	0.7505	0.9548
LN(0.5,0.5)	0.7500	0.9375	0.7496	0.9518

Table 5.44: *Distribution properties of $\hat{p}$, $MC = 4\ 000, n = 200$.*

Reduction in variance of $\hat{h}$ as a result of bagging ($\hat{h}_{CV}$ versus $\hat{h}_{CV,B_2}$)

Distribution	$E(\hat{p})$	$n = 50$	$n = 100$	$n = 200$	Average
Wei(1,1)	1.00	77.4%	65.4%	54.7%	65.8%
Wei(1,1)	0.75	77.3%	74.0%	59.6%	70.3%
Wei(2,1)	1.00	75.6%	77.0%	76.5%	76.4%
Wei(2,1)	0.75	70.4%	79.2%	77.9%	75.9%
Wei(3,1)	1.00	76.4%	75.4%	79.9%	77.2%
Wei(3,1)	0.75	70.3%	75.3%	79.6%	75.1%
LN(0,1)	1.00	67.1%	51.9%	64.7%	61.2%
LN(0,1)	0.75	73.6%	64.9%	44.6%	61.0%
LN(0.25,0.75)	1.00	72.7%	59.9%	55.8%	62.8%
LN(0.25,0.75)	0.75	77.1%	73.5%	55.5%	68.7%
LN(0.5,0.5)	1.00	79.8%	75.8%	69.9%	75.2%
LN(0.5,0.5)	0.75	79.1%	77.6%	71.9%	76.2%
Average		74.7%	70.8%	65.9%	70.5%

Table 5.45: *Percentage reduction in sample variance of $\hat{h}_{CV,B_2}$ versus $\hat{h}_{CV}$, $MC = 1\,000$.*

From Table 5.45 we see that the implementation of bagging (especially method 2) reduces the sample variance of the selected bandwidths considerably, with the minimum reduction equal to 44.6%, the maximum reduction equal to 79.9%, and the average reduction across all settings equal to 70.5%. From Hall and Robinson (2009) we have that for the choice of $m/n = 0.5$ the expected reduction in variance should be between 43% and 71% for kernel density estimation. We have now shown empirically that similar results can be expected in the setting of kernel hazard function estimation.

In this dissertation we have provided the reader with sufficient background regarding the setting in which the study was conducted, and have provided various references for readers interested in more details. The automatic bandwidth selectors that we considered were explained from a theoretical as well as a practical viewpoint, that is describing the mechanics of implementing the selectors. All the selectors considered were examined and compared across various combinations of settings, and conclusions regarding their performance were made. Some problems for future studies were also mentioned. In essence we see that the application of bootstrap methodology can significantly increase the performance of automatic bandwidth selectors, and the use thereof is highly recommended.

List of Tables

1.1	<i>Different hazard function shapes.</i>	6
2.1	<i>Effect of applying the plug-in principle.</i>	20
3.1	<i>Different kernel functions.</i>	37
4.1	<i>Convergence of empirical bootstrap to explicit bootstrap as $B \rightarrow \infty$.</i>	55
4.2	<i>Convergence expressed in percentages.</i>	55
5.1	<i>Different distributions used in the simulation study.</i>	70
5.2	<i>Approximate values for $\hat{h}_{MISE_w}$ for the different distributions.</i>	85
5.3	<i>Simulation summary : Weibull(1, 1), $\eta = 0$, $n = 50$, $MC = 1\ 000$</i>	86
5.4	<i>Simulation summary : Weibull(1, 1), $\eta = 1/3$, $n = 50$, $MC = 1\ 000$</i>	86
5.5	<i>Simulation summary : Weibull(2, 1), $\eta = 0$, $n = 50$, $MC = 1\ 000$</i>	87
5.6	<i>Simulation summary : Weibull(2, 1), $\eta = 1/3$, $n = 50$, $MC = 1\ 000$</i>	87
5.7	<i>Simulation summary : Weibull(3, 1), $\eta = 0$, $n = 50$, $MC = 1\ 000$</i>	88
5.8	<i>Simulation summary : Weibull(3, 1), $\eta = 1/3$, $n = 50$, $MC = 1\ 000$</i>	88
5.9	<i>Simulation summary : lognormal(0, 1), $\eta = 0$, $n = 50$, $MC = 1\ 000$</i>	89
5.10	<i>Simulation summary : lognormal(0, 1), $\eta = 1/3$, $n = 50$, $MC = 1\ 000$</i>	89
5.11	<i>Simulation summary : lognormal(0.25, 0.75), $\eta = 0$, $n = 50$, $MC = 1\ 000$</i>	90
5.12	<i>Simulation summary : lognormal(0.25, 0.75), $\eta = 1/3$, $n = 50$, $MC = 1\ 000$</i>	90
5.13	<i>Simulation summary : lognormal(0.5, 0.5), $\eta = 0$, $n = 50$, $MC = 1\ 000$</i>	91
5.14	<i>Simulation summary : lognormal(0.5, 0.5), $\eta = 1/3$, $n = 50$, $MC = 1\ 000$</i>	91
5.15	<i>Simulation summary : Weibull(1, 1), $\eta = 0$, $n = 100$, $MC = 1\ 000$</i>	92
5.16	<i>Simulation summary : Weibull(1, 1), $\eta = 1/3$, $n = 100$, $MC = 1\ 000$</i>	92
5.17	<i>Simulation summary : Weibull(2, 1), $\eta = 0$, $n = 100$, $MC = 1\ 000$</i>	93
5.18	<i>Simulation summary : Weibull(2, 1), $\eta = 1/3$, $n = 100$, $MC = 1\ 000$</i>	93
5.19	<i>Simulation summary : Weibull(3, 1), $\eta = 0$, $n = 100$, $MC = 1\ 000$</i>	94

5.20	<i>Simulation summary : Weibull(3, 1), $\eta = 1/3, n = 100, MC = 1\ 000$</i>	94
5.21	<i>Simulation summary : lognormal(0, 1), $\eta = 0, n = 100, MC = 1\ 000$</i>	95
5.22	<i>Simulation summary : lognormal(0, 1), $\eta = 1/3, n = 100, MC = 1\ 000$</i>	95
5.23	<i>Simulation summary : lognormal(0.25, 0.75), $\eta = 0, n = 100, MC = 1\ 000$</i>	96
5.24	<i>Simulation summary : lognormal(0.25, 0.75), $\eta = 1/3, n = 100, MC = 1\ 000$</i>	96
5.25	<i>Simulation summary : lognormal(0.5, 0.5), $\eta = 0, n = 100, MC = 1\ 000$</i>	97
5.26	<i>Simulation summary : lognormal(0.5, 0.5), $\eta = 1/3, n = 100, MC = 1\ 000$</i>	97
5.27	<i>Simulation summary : Weibull(1, 1), $\eta = 0, n = 200, MC = 1\ 000$</i>	98
5.28	<i>Simulation summary : Weibull(1, 1), $\eta = 1/3, n = 200, MC = 1\ 000$</i>	98
5.29	<i>Simulation summary : Weibull(2, 1), $\eta = 0, n = 200, MC = 1\ 000$</i>	99
5.30	<i>Simulation summary : Weibull(2, 1), $\eta = 1/3, n = 200, MC = 1\ 000$</i>	99
5.31	<i>Simulation summary : Weibull(3, 1), $\eta = 0, n = 200, MC = 1\ 000$</i>	100
5.32	<i>Simulation summary : Weibull(3, 1), $\eta = 1/3, n = 200, MC = 1\ 000$</i>	100
5.33	<i>Simulation summary : lognormal(0, 1), $\eta = 0, n = 200, MC = 1\ 000$</i>	101
5.34	<i>Simulation summary : lognormal(0, 1), $\eta = 1/3, n = 200, MC = 1\ 000$</i>	101
5.35	<i>Simulation summary : lognormal(0.25, 0.75), $\eta = 0, n = 200, MC = 1\ 000$</i>	102
5.36	<i>Simulation summary : lognormal(0.25, 0.75), $\eta = 1/3, n = 200, MC = 1\ 000$</i>	102
5.37	<i>Simulation summary : lognormal(0.5, 0.5), $\eta = 0, n = 200, MC = 1\ 000$</i>	103
5.38	<i>Simulation summary : lognormal(0.5, 0.5), $\eta = 1/3, n = 200, MC = 1\ 000$</i>	103
5.39	<i>Ranking method results for ISE_w values.</i>	113
5.40	<i>Relative frequency distribution and results of $\hat{h}$ $\tilde{MSE}$ rankings.</i>	113
5.41	<i>Final ranking of the five selectors.</i>	114
5.42	<i>Distribution properties of $\hat{p}$, $MC = 4\ 000, n = 50$.</i>	117
5.43	<i>Distribution properties of $\hat{p}$, $MC = 4\ 000, n = 100$.</i>	117
5.44	<i>Distribution properties of $\hat{p}$, $MC = 4\ 000, n = 200$.</i>	117
5.45	<i>Percentage reduction in sample variance of $\hat{h}_{CV, B_2}$ versus $\hat{h}_{CV}$, $MC = 1\ 000$.</i>	118

List of Figures

1.1	<i>Different shapes of the hazard function.</i>	7
1.2	<i>Type 1 censoring with 3 observed lifetimes and 2 censored lifetimes.</i>	8
1.3	<i>Type II censoring with 4 observed lifetimes and 2 censored lifetimes.</i>	9
1.4	<i>Random right censoring with 4 observed lifetimes and 2 censored lifetimes.</i>	10
1.5	<i>Random left censoring with 4 observed lifetimes and 2 censored lifetimes.</i>	11
2.1	<i>Different bootstrap approximations to the function $MISE_w(h)$.</i>	25
3.1	<i>Graphical representation of smoothing parameter selection problem</i>	34
3.2	<i>Different kernel functions</i>	38
3.3	<i>Kernel density estimator with $k = \phi$, and $h = 0.29$.</i>	41
3.4	<i>Kernel distribution estimator with $K = \Phi$, and $h = 0.35$.</i>	43
3.5	<i>Kernel hazard function estimator with $k = \phi$, and $h = 0.50$.</i>	44
5.1	<i>Hazard rate functions considered in the simulation study.</i>	71
5.2	<i>Legend to be used when interpreting graphs.</i>	85
5.3	<i>Bandwidth distribution for Weibull(1,1), $n=50$.</i>	104
5.4	<i>Bandwidth distribution for Weibull(2,1), $n=50$.</i>	104
5.5	<i>Bandwidth distribution for Weibull(3,1), $n=50$.</i>	105
5.6	<i>Bandwidth distribution for lognormal(0,1), $n=50$.</i>	105
5.7	<i>Bandwidth distribution for lognormal(0.25,0.75), $n=50$.</i>	106
5.8	<i>Bandwidth distribution for lognormal(0.5,0.5), $n=50$.</i>	106
5.9	<i>Bandwidth distribution for Weibull(1,1), $n=100$.</i>	107
5.10	<i>Bandwidth distribution for Weibull(2,1), $n=100$.</i>	107
5.11	<i>Bandwidth distribution for Weibull(3,1), $n=100$.</i>	108
5.12	<i>Bandwidth distribution for lognormal(0,1), $n=100$.</i>	108
5.13	<i>Bandwidth distribution for lognormal(0.25,0.75), $n=100$.</i>	109
5.14	<i>Bandwidth distribution for lognormal(0.5,0.5), $n=100$.</i>	109

5.15	<i>Bandwidth distribution for Weibull(1,1), $n=200$.</i>	110
5.16	<i>Bandwidth distribution for Weibull(2,1), $n=200$.</i>	110
5.17	<i>Bandwidth distribution for Weibull(3,1), $n=200$.</i>	111
5.18	<i>Bandwidth distribution for lognormal(0,1), $n=200$.</i>	111
5.19	<i>Bandwidth distribution for lognormal(0.25,0.75), $n=200$.</i>	112
5.20	<i>Bandwidth distribution for lognormal(0.5,0.5), $n=200$.</i>	112

Bibliography

Act (2005). *CT4: Survival Analysis*.

Antoniadis, A. and Grégoire, G. (1990). Penalized likelihood estimation for rates with censored survival data, *Scandinavian Journal of Statistics* (17): 43–63.

Bowman, A. (1984). An alternative method of cross-validation for the smoothing of density estimates, *Biometrika* **71**(2): 353–360.

Breiman, L. (1996). Bagging predictors, *Machine Learning* **24**: 123–140.

Cao, R., Cuevas, A. and Gonzalez-Monteiga, W. (1994). A comparative study of several smoothing methods in density estimation, *Computational Statistics and Data Analysis* (17): 153–176.

Davison, A. and Hinkley, D. (1997). *Bootstrap Methods and their Application*, Cambridge university press.

Efron, B. (1981). Censored data and the bootstrap, *Journal of the American Statistical Association* **76**: 312–319.

Efron, B. and Tibshirani, R. J. (1993). *An Introduction to the Bootstrap*, Chapman and Hall.

Faraway, J. and Jhun, M. (1990). Bootstrap choice of bandwidth for density estimation, *Journal of the American Statistical Association* **85**: 1119–1122.

Gefeller, O. and Dette, H. (1992). Nearest neighbour kernel estimation of the hazard function from censored data, *Journal of Statistical and Computational Simulation* **43**(412): 93–101.

- Gonzalez-Manteiga, W., Cao, R. and Marron, J. (1996). Bootstrap selection of the smoothing parameter in nonparametric hazard rate estimation, *Journal of the American Statistical Association* **91**(435): 1130–1140.
- Hall, P. and Robinson, A. (2009). Reducing variability of crossvalidation for smoothing-parameter choice, *Biometrika* **96**(1): 175–186.
- Härdle, W. (1990). *Applied Nonparametric Regression*, Cambridge university press.
- IMSL (1991). *IMSL, Version 2.0*.
- Jones, M., Davies, S. and Park, B. (1994). Versions of kernel-type regression estimators, *Journal of the American Statistical Association* **89**(427): 825–832.
- Kaplan, E. and Meier, P. (1958). Nonparametric estimation from incomplete observations, *Journal of the American Statistical Association* **53**: 457–481.
- Kim, C., Bae, W., Choi, H. and Park, B. (2005). Non-parametric hazard function estimation using the kaplan-meier estimator, *Nonparametric Statistics* **17**(8): 937–948.
- Klein, J. P. and Moeschberger, M. L. (1997). *Survival analysis: Techniques for Censored and Truncated Data*, Springer.
- Koekemoer, G. (2004). *Transforming data to Normality*, PhD thesis, North-West University.
- Koziol, J. and Green, S. (1976). A cramer-von mises statistic for randomly censored data, *Biometrika* (63): 465–474.
- Marron, J. (1992). Bootstrap bandwidth selection, *Exploring the limits of Bootstrap, Wiley Series in Probability and Statistics* pp. 249–262.
- Marron, J. and Padgett, J. (1987). Asymptotically optimal bandwidth selection for a kernel density estimator, *The Annals of Statistics* **15**: 1520–1535.
- Parzen, E. (1962). On estimation of a probability density and mode, *Annals of Mathematical Statistics* (35): 1065–1076.
- Patil, P. (1993a). Bandwidth choice for nonparametric hazard rate estimation, *Journal of Statistical Planning and Inference* **35**: 15–30.

- Patil, P. (1993b). On the least-squares cross-validation bandwidth in hazard rate estimation, *The Annals of Statistics* **21**(4): 1792–1810.
- Patil, P. (1997). Nonparametric hazard rate estimation by ortogonal wavelet methods, *Journal of Statistical Planning and Inference* (60): 153–168.
- Patil, P., Wells, M. and Marron, J. (1994). Some heuristics of the kernel based estimators of ratio functions, *Nonparametric Statistics* **4**: 203–209.
- Rice, J. and Rosenblatt, M. (1976). Estimation of the log survivor function and hazard function, *Sankhya: The Indian Journal of Statistics* (38): 60–78.
- Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function., *Annals of Mathematical Statistics* (27): 642–669.
- Rudemo, M. (1982). Empirical choice of histograms and kernel density estimators, *Scandinavian Journal of Statistics* **9**: 65–78.
- Sarda, P. and Vieu, P. (1991). Smoothing parameter selection in hazard estimation, *Statistics and Probability Letters* **11**: 429–434.
- Smith, P. J. (2002). *Analysis of Failure and Survival Data*, Chapman and Hall.
- Tanner, M. and Wong, W. (1983). The estimation of the hazard function from randomly censored data by the kernel method, *The Annals of Statistics* **11**(3): 989–993.
- Van Graan, F. (1983). *Nie-parametriese beraming van verdelingsfunksies*, Master's thesis, P.U. for C.H.E, Potchefstroom. 65 p.
- Wand, M. and Jones, M. (1995). *Kernel Smoothing*, Chapman and Hall.
- Watson, G. and Leadbetter, M. (1964a). Hazard analysis i, *Biometrika* **51**: 175–184.
- Watson, G. and Leadbetter, M. (1964b). Hazard analysis ii, *Sankhya: The Indian Journal of Statistics* (26): 101–116.
- Xie, M. and Lai, C. (1995). Reliability analysis using an additive weibull model with bathtub-shaped failure rate function, *Reliability Engineering and System Safety* **52**: 87–93.