A Generalization of the Fisher Information Measure

D.E. Boekee

Delft University Press

A Generalization

of the Fisher

Information Measure

A Generalization of the Fisher Information Measure

Proefschrift ter verkrijging van
de graad van doctor in de
technische wetenschappen
aan de Technische Hogeschool Delft,
op gezag van de rector magnificus
Prof. ir. L. Huisman,
voor een commissie aangewezen
door het college van dekanen
te verdedigen op
woensdag 28 september 1977
te 14.00 uur door

Dirk Evert Boekee
elektrotechnisch ingenieur,
geboren te 's-Gravenhage

Delft University Press/1977

Dit proefschrift is goedgekeurd door de promotor
PROF. IR. IJ. BOXMA

Aan Ank,

Eddy, Wouter en Cynthia

CONTENTS

SUMMARY

The Fisher information is a well known information measure. It can be seen as a measure for the information about an unknown parameter which is contained in a set of observations. The information measure satisfies certain basic requirements which are necessary to make it a useful measure in applications. It is related to the concept of accuracy in estimation theory since it provides a lower bound to the variance of estimators of a parameter. This is the famous Cramér-Rao inequality.

In this thesis a generalization of the Fisher information measure is considered, called the Fisher information of order s. It is based on a certain convex function of the first derivative of the log-likelihood function. This measure includes the usual Fisher information measure as a special case.

In Chapter 1 some basic concepts are introduced which will be used in the thesis. Also a short introduction is given to the problems which will be studied.

In Chapter 2 the Fisher information of order s is introduced and some basic properties of this information measure are obtained. An important property is its invariance under a measurable sufficient transformation. Next an inequality for the s-th absolute central moment of unbiased parameter estimators is derived, which is based on the Fisher information of order s. We consider some consequences of this inequality, partly based on the equality condition, and also give a generalized version of the Blackwell - Rao theorem.

Then the Fisher information of order s is studied for multivariate observations and upper and lower bounds for this case are obtained.

Discrete observations are also discussed briefly. Two equivalent information measures are introduced, and their relation to the Fisher information of order s is discussed. As a final special case, random parameters are considered. For this case it is possible to define two information measures by different ways of averaging. Some basic properties for these two measures are derived.

In Chapter 3 results obtained in the previous chapter are applied to two important classes of parameters: location and scale parameters. At first expressions are derived for the Fisher information of order s for those two cases. For the class of location parameters, it is shown that there exists a family of distributions, called the exponential power distribution or gaussian family, which possesses minimal information. For estimation problems this means that observations which have this distribution provide as little information as possible and for this reason lead to the worst estimation performance (in terms of the s-th absolute central moment of estimators). It is also shown that this characterization of the exponential power distribution is stable.

For the location and scale parameters, also some multivariate aspects are discussed. Next some results for the Fisher information of order s are given for location and scale parameters by considering several examples of distributions which are used when estimating signal parameters.

Based on Shannon's measure of information, an entropy moment of order s is defined. Some results for location and scale parameters are given. It is shown that there exists an analogy between this entropy moment of order s and the inverse of the Fisher information of order s.

In Chapter 4 the relation of the Fisher information of order s to some probabilistic distance measures is considered. It is shown that there are several measures which can be related to the Fisher information of order s in a simple way. This means that they can also be used to obtain bounds on the s-th absolute moment of parameter estimators. The relations obtained also permit a geometric interpretation of the Fisher information of order s to be given.

For mixture distributions bounds on the Fisher information of order s are given. It is shown that the upper bound, which is based on the convexity of the Fisher information of order s, has some attractive properties.

Finally, the Fisher information of order s is applied to the estimation of signal parameters. Here the estimation of an unknown parameter of a furthermore completely known signal is considered, when there is non-gaussian noise. Also some models for non-gaussian signals, which were introduced in Chapter 3, are discussed.

LIST OF MAJOR SYMBOLS

a	real constant
$a(.)$	real function
A	set in $\mathcal{A}$
$A(.)$	real function
$\mathcal{A}$	Borel field
b	real constant
$b(.)$	bias function
B	set in $\mathcal{B}$
$B(.)$	beta function
$\mathcal{B}$	Borel field
c	real constant
C	real constant
$C(.)$	real function
$d(.)$	metric
$D(.)$	directed divergence
$D_f(.)$	f-divergence
$D_{f,s}(.)$	information measure
$D_M(.)$	Matusita distance
e_s	efficiency of estimator
E	expectation operator

$f(.)$	Fisher self-information
$\mathcal{f}(.)$	real convex function
$F_s(.)$	Fisher information of order s
$F_s'(.)$	information measure
$F_s''(.)$	information measure
$F_s(\xi,\phi)$	information measure for random parameter
$\overline{F}_s(\xi;\phi)$	average Fisher information of order s
${}_1F_1(.)$	hypergeometric function
$g(.)$	real function
$G(.)$	information measure
$h(.)$	real function
$H(.)$	Shannon's information measure
i	integer constant
$I_s(.)$	information measure for location parameter
$I_s^*(.)$	information measure for scale parameter
$\mathcal{I}(.)$	mutual information
j	integer constant
J_s	information function
k	real constant
$\dot{k}(.)$	real function
K	real constant
$K(.)$	real function
m	real constant
m_s	s-th absolute moment
$\mathcal{m}_s$	s-th absolute central moment
M	real constant

$M(.)$	real function
$M_s(.)$	entropy moment of order s
MSB	minimum s-th absolute central moment bound
MVB	minimum variance bound
$\mathcal{M}_\alpha$	mean of order α
n	integer constant
N	integer constant
$N_s(\mu,\beta)$	exponential power distribution
$\mathcal{O}(.)$	order sign
$p(.)$	probability density function
$\{P_\theta\}$	family of probability density functions
P	probability measure
$Pr(.)$	probability of event
$q(.)$	probability density function
Q	probability measure
r	real constant
$r(.)$	probability density function
$R(.)$	Lagrange remainder
$\mathbb{R}^n$	n-dimensional real space
s	order parameter
sgn [.]	sign function
s.a.c.m.	s-th absolute central moment
S	set {2, 3/2, 4/3, ...}
t	real constant, value of T
T	transformation, statistic
u	real variable

v	real variable
w	real variable
$W_{\lambda,\alpha}(.)$	Whittaker's function
x	element of X
X	real space
y	element of Y
Y	real space
z	element of Z
Z	real space
α	real constant
β	real constant, s-th absolute central moment of $N_s(\mu,\beta)$
γ	real parameter
$\gamma(.)$	incomplete gamma function
$\Gamma(.)$	gamma function
δ	real constant
$\Delta\theta$	change in θ
ζ	random variable
η	random variable
θ	real parameter
$\hat{\theta}$	estimator of θ
Θ	parameter space
λ	real parameter
μ	real parameter
ν	real parameter
ξ	random variable
$\underset{\sim}{\xi}$	random vector

$\pi(.)$	probability density function
ρ	real parameter
ρ_B	Bhattacharyya distance
σ	real parameter
τ	real constant
$\tau(.)$	real function
υ	curve on hypersphere
ϕ	random parameter
$\psi(.)$	psi-function

CHAPTER 1

INTRODUCTION

In this thesis we shall study a generalization of the Fisher information measure and its application to the estimation of parameters. In this first chapter we will introduce the basic concepts which will be used in this thesis.

First we will discuss some of the ideas behind this generalization in Section 1.1, in order to give a framework for the thesis. In Section 1.2 we will give a short description of the various kinds of parameters which are considered in the thesis.

Next we will discuss some measures of information. Some properties are given with respect to accuracy of estimators and invariance of transformations. We will also discuss the concept of sufficiency. These properties will be given quite briefly and are covered only to introduce the basic ideas which will be considered in more detail in the next chapters of the thesis.

Finally we will present in Section 1.4 some results for bounds on the performance of estimators. We shall discuss briefly those bounds and some related results and give a short review of the literature.

1.1 SOME GENERAL CONSIDERATIONS

When studying such fundamental areas as information theory and estimation and filtering theory, as well as related areas like pattern recognition, speech and image processing, one is often faced with two assumptions: the density functions of the signals or observations which are studied are gaussian, and the appropriate measure for distortion, accuracy or estimation error is the mean-square error.

Although at first the second assumption may seem natural, one may wonder what is natural about it. Then it is surprising to find out that it is far from simple not to use this assumption. If a different performance criterion is used, much of the mathematical framework looses its simplicity or even breaks down. A similar remark can be made about the application of information measures as a bound on the performance.

What are the advantages and shortcomings of the mean-square error criterion as a performance measure? There are several advantages from a mathematical point of view. If the mean-square error criterion is adopted, we can apply such methods as inner products, Hilbert space considerations, positive-definite forms, error spectral densities and Karhunen-Loève expansion.

Furthermore, we can embed the mean-square error criterion in an elegant way in the second order theory of stochastic processes. Several refinements, like a (frequency) weighted mean-square error criterion have been developed.

As a result of these considerations, it will be clear that it is attractive to use the mean-square error criterion. However, in many practical situations this assumption is unsatisfactory, or even inadequate. We mention the field of image processing, where it is clear that the human performance criterion is not equivalent to a mean-square error criterion. A similar remark can be made for speech processing. Although this has long been realized, it has been difficult to find other criteria which can also be handled mathematically in a satisfactory manner.

One way to approach this problem is to use test persons to give a, possibly subjective, opinion as to the performance. However, due to the difficult, psychological aspects which are involved, this has been of limited value.

A possible solution lies in the study of error criteria other than the mean-square error. In this thesis we consider the mean s-th absolute error for $s \geq 1$. For $s = 2$ this yields the mean-square error and for $s = 1$, the absolute error criterion. This mean s-th absolute error criterion is a special case of the whole class of convex error criteria. It will be shown in this thesis that the results which can be obtained can still be interpreted in a convenient manner.

The choice of s can be based on two considerations. First, we can choose s close to 2. This makes it possible to investigate the sensitivity of the model we are working with to the mean-square error criterion. Secondly, we can choose s such that small or large errors are weighted in a desired way. In Fig. 1.1.1 we have given the mean s-th absolute error criterion for some values of s.

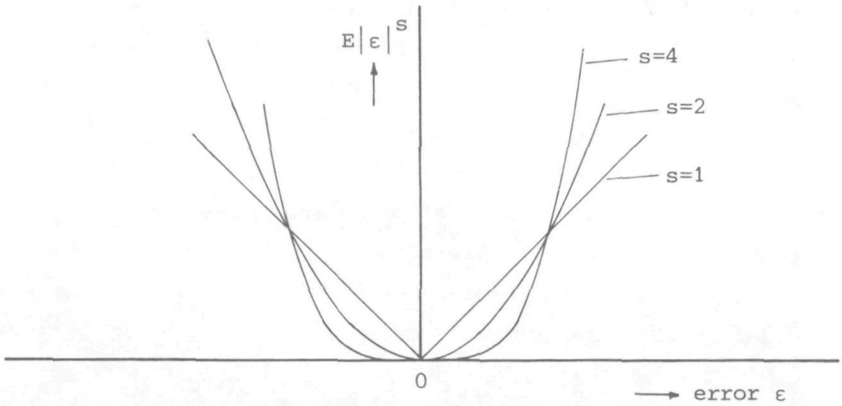

Fig. 1.1.1 s-th absolute error criterion

As we will see later on, the use of a mean s-th absolute error criterion is a natural counterpart to a density function which is called the exponential power distribution of order s. This distribution is also called the gaussian family, since it includes and is a generalization of the gaussian distribution. This relation between the mean s-th absolute error criterion and the exponential power distribution is based on information theoretic grounds.

The information theoretic approach in the thesis comes in as a tool to measure the intrinsic, or potential, information which is contained in the observations, signals or measurements, which are available. There are several ways to measure information, as we shall see in Section 1.3. Since we shall consider estimation problems where the mean s-th absolute error is used as a performance measure, we shall study an information measure which measures this information in an appropriate way.

There exists a strong relation between the performance measure which is used, the optimal density function of the observations and the information measure which is the most natural one. Although each of these three quantities can be introduced as the elementary one, from which the other two follow, it seems realistic to consider the mean s-th absolute error criterion as the most important quantity. However, the exponential power distribution is also met frequently, as a model for signals like speech and noise. In that case one might conclude that for those signals the mean s-th absolute error criterion is 'matched' to the signals which are studied.

In this thesis we shall emphasize the information measure which will be called the Fisher information of order s. We shall consider its relation to the mean s-th absolute error criterion and to the exponential power distribution of order s.

1.2 CLASSIFICATION OF PARAMETERS

We shall consider a real parameter $\theta \in \Theta$, where Θ is an open subset of the real line $\mathbb{R}^1$. We assume that the density function of the observation ξ depends on θ, and therefore we have to consider the conditional density function $p(x;\theta)$. Since $\theta \in \Theta$ we introduce a family of density functions, denoted by $\{P_\theta\} = \{p(x;\theta),\ \theta \in \Theta\}$.

Parameters can be classified in different ways. First we will consider a classification into random and non-random parameters.

The distinction between random and non-random parameters is in many cases a natural one, which follows from the practical situation from which one starts. If there is any knowledge available about the values which the parameter can take on, it is often possible to model this as a probability distribution of the parameter θ. To obtain this probability distribution in a reasonable way we can for instance use the notion of most informative prior, as discussed by Jaynes [28]. If θ is a constant, but unknown a priori, it will be called non-random.

Another way to classify parameters is to distinguish between discrete and continuous parameters.

Continuous parameters will be encountered in analog processing systems, e.g., phase, amplitude, etc., but may also be the signal value itself, e.g., in filtering or prediction problems. If θ is continuous and random, we assume that θ has a probability density function $\pi(\theta)$ such that

$$\pi(\theta) > 0 \qquad \theta \in \Theta \tag{1.2.1}$$

and

$$\int_\Theta \pi(\theta)\, d\theta = 1. \tag{1.2.2}$$

In this thesis we will only consider continuous parameters.

A third possible classification of parameters is obtained if we consider the way they influence the distribution of the random variable ξ. We will consider two important classes of parameters which are of particular importance in estimation theory: location and scale parameters. For completeness we give their definitions (see Ferguson [19]).

DEFINITION 1.2.1

A real parameter $\theta \in \Theta$ is a location parameter for the distribution of the random variable ξ iff

$$p(x;\theta) = q(x-\theta) \tag{1.2.3}$$

for some density function $q(y)$.

DEFINITION 1.2.2

A real positive parameter $\theta \in \Theta$ is a scale parameter for the distribution of the random variable ξ iff

$$p(x;\theta) = \frac{1}{\theta}\, q(\frac{x}{\theta}) \tag{1.2.4}$$

for some density function $q(y)$.

1.3 MEASURES OF INFORMATION

In this section we will briefly describe the concept of information and two measures which will be referred to in this thesis. We also give some properties which are essential for their use in estimation theory.

The concept of information may be a slightly misleading phrase, since actually there are several concepts which are involved if we speak of information. It is interesting to study those concepts in order to get a better insight in the various information measures which have been developed. The two best known information measures are Shannon's measure of information, or entropy as it is usually called, and Fisher's information measure. Several other information measures have been developed. We mention the information of order α and the information of type β. Some of the concepts which are involved in the notion of information are statistical entropy, uncertainty, accuracy, coding, questionaires, stochastic independence, probabilistic distance and discriminating ability.

Since a general discussion is beyond the scope of this section we shall consider mainly the concept of accuracy. It is this concept which relates measures of information to the performance of parameter estimators. It will be the main theme of the thesis. We may distinguish two basic problems.

The first problem is how much information is contained in observations. We can only consider this problem if we specify in what manner we intend to use the information or, in other words, what we consider to be relevant information. Since we are concerned with the performance *of parameter estimators*, we are *interested in the information which is* contained in the observations about a parameter θ. This parameter may be random or non-random.

The second problem is how to find an estimator which gives the best results, i.e., which leads to the highest accuracy of an unknown parameter. It will be clear that we would like to use all information which is relevant with respect to the estimation of θ. The concept of using all information is closely related to the concept of sufficiency. It is a generally accepted point of view that an estimator which is sufficient for a parameter θ uses all relevant information in the observations.

A measure for information which is well known is Shannon's measure of information. Its main applications can be found in communication theory. If the observation or random variable ξ is continuous with a density function $p(x;\theta)$ this information measure is given by

$$H(\xi;\theta) = - \int_X p(x;\theta) \log p(x;\theta)\, dx, \qquad (1.3.1)$$

provided the integral exists. We shall not discuss its properties, since they can be found in any text on information theory. Based on this measure, we can obtain a measure for the information which is contained on the average in an observation. To this end we assume that the parameter is a random variable, denoted by ϕ, which takes on the values $\theta \in \Theta$ and has a density function $\pi(\theta)$. Then the information provided by an observation ξ with respect to θ is given by

$$I(\xi,\phi) = H(\phi) - H(\phi/\xi). \tag{1.3.2}$$

Here $H(\phi/\xi)$ is the conditional information of ϕ given ξ and is defined by

$$H(\phi/\xi) = - \int_X \int_\Theta q(x,\theta) \log q(\theta/x)\, dx\, d\theta. \tag{1.3.3}$$

Essentially the quantity $I(\xi,\phi)$ is the mutual information between ξ and ϕ. It is a non-negative quantity which is zero iff the observation ξ and the parameter ϕ are stochastically independent. Some further results for $I(\xi,\phi)$ can be found in Fraser [20], Lindley [36], Mallows [39] and Rényi [47].

If the parameter θ is non-random we cannot follow this approach. We now introduce Fisher's information measure. This measure is defined by

$$F(\xi;\theta) = \int_X \left[\frac{\partial}{\partial\theta} \log p(x;\theta)\right]^2 p(x;\theta)\, dx, \tag{1.3.4}$$

provided some regularity conditions for the density function $p(x;\theta)$ are satisfied. An equivalent expression is

$$F(\xi;\theta) = - \int_X \frac{\partial^2}{\partial\theta^2} \log p(x;\theta)\, p(x;\theta)\, dx. \tag{1.3.5}$$

This expression is slightly less general, since some additional requirements have to be satisfied. We shall mainly use the first expression.

As will be clear from Eq. 1.3.4, the parameter θ plays a crucial role in this information measure. Though its main application is concerned with non-random parameters, the case of random parameters can be included by a suitable modification of $F(\xi;\theta)$. This can be found, for example, in Gart [22] and Van Trees [56]. The Fisher information is a non-negative quantity and is additive. Further properties can be found in Kagan et al. [30], Kendall et al. [32], Rao [46] and Stam [53].

An important property of the Fisher information is its relation to the accuracy of parameter estimators, which is given by the famous Cramér-Rao inequality. We will discuss this property in the next section. The Fisher information measure can be seen as the information contained in an observation ξ about a parameter θ. This should be interpreted as the extent to which, on the average, the accuracy of the unknown parameter can be increased as a result of the observed value x of the observation ξ. For this reason Fisher called it the intrinsic accuracy of the observation ξ.

A fundamental concept both in information theory and in statistics is sufficiency, although the background of this concept is different in those fields. We will give two definitions to demonstrate this. As we have indicated earlier in this section sufficiency is related to the idea of using all information, or losing no information. In information theory this concept is reflected in the data processing theorem in terms of the mutual information. In estimation theory sufficiency is usually considered as a property for a statistic which is used to estimate a parameter. In this application the concept of sufficiency is also to use all information contained in the observation, so as to obtain the most accurate estimate. A definition of sufficiency which is often used is the factorization theorem.

To give a more precise statement of these points of view, we will consider a random variable ξ with density function $p(x;\theta)$. Furthermore, we will consider a measurable transformation $T(\xi)$ of ξ which has an induced density function $q(T(x);\theta)$. This transformation is also called a statistic and is also a random variable. We can define sufficiency in an information context as follows.

DEFINITION 1.3.1

Let ξ and η be random variables and let $T(\xi)$ be a measurable transformation. Then the random variable $T(\xi)$ is sufficient for η if

$$I(T(\xi),\eta) = I(\xi,\eta). \tag{1.3.6}$$

This definition gives a precise version of the conventional formulation that $T(\xi)$ is sufficient for η if $T(\xi)$ contains all information in ξ with respect to η. If we replace the random variable η with ϕ, we have a formulation for sufficiency with respect to the random parameter ϕ. The use of Shannon's mutual information measure is not essential. In fact we can show that for other information measures a similar definition for sufficiency can be given. We can also give a similar result for the Fisher information, which is of importance if θ is non-random. Then a random function $T(\xi)$ is sufficient for θ if $F(T(\xi);\theta) = F(\xi;\theta)$.

If $T(\xi)$ is not sufficient this should mean that we use less information. It can be shown that in general we have an inequality of the type

$$I(T(\xi),\phi) \le I(\xi,\phi) \tag{1.3.7}$$

or $$F(T(\xi);\theta) \le F(\xi;\theta) \tag{1.3.8}$$

which is in agreement with this intuitive requirement.

In estimation theory the notion of a sufficient statistic is usually motivated in a slightly different way. We then consider a parameter θ for which several statistics can be used. We denote two of those statistics by T and T_1 and consider the joint density function. We now give the following definition of sufficiency, which is known as the factorization theorem.

DEFINITION 1.3.2

Let T and T_1 be statistics for θ with joint density function $p(t,t_1;\theta)$. Then T is a sufficient statistic for θ iff

$$p(t,t_1;\theta) = g(t;\theta)\ h(t_1;t) \tag{1.3.9}$$

where h(.) is independent of θ.

Since h(.) is independent of θ, the statistic T_1, once $T = t$ is known, cannot be used to get a better estimate of θ. For this reason it is usually said that T contains all the information about θ in the observation and is, therefore, the best statistic, if it exists.

The two definitions are equivalent in the sense that each definition implies the other. However, the concept of using all information is demonstrated more clearly by Definition 1.3.1.

1.4 *BOUNDS ON THE PERFORMANCE OF ESTIMATORS*

In this section we shall pay attention to the possible bounds on the performance of estimators. We study bounds which are depending on the observations only and are independent of actual estimation methods which are used.

As we have indicated in Section 1.1 a common choice for the performance of estimators is the mean-square error. Since in general different estimators may lead to different values of the mean-square error, we are interested in what the intrinsic mean-square error for the parameter θ is, given a set of observations. This intrinsic mean-square error then is a lower bound for the mean-square error of an arbitrary estimator for θ.

An estimator which achieves this lower bound is called an MVB (minimum variance bound) estimator or efficient estimator. Here efficiency is defined as the ratio between the actual mean-square error and the optimal mean-square error, as given by the lower bound.

We shall consider a statistic (estimator) $T(\xi)$ which is a function of the observation ξ and which is used to estimate a differentiable function $\tau(\theta)$ of the non-random parameter θ. The mean-square error, or variance, of the unbiased statistic T is denoted by $E_\theta|T - \tau(\theta)|^2$. The best-known bound for the mean-square error is the celebrated Cramér-Rao inequality. This bound is given by

$$E_\theta |T - \tau(\theta)|^2 \geq \frac{|\tau'(\theta)|^2}{F(\xi;\theta)}, \tag{1.4.1}$$

where $\tau'(\theta)$ is the derivative of $\tau(\theta)$ with respect to θ. The quantity $F(\xi;\theta)$ is Fisher's information measure. This inequality is valid if certain regularity conditions are satisfied. If $T(\xi)$ is an unbiased estimator of θ, we have the simple inequality

$$E_\theta |T - \theta|^2 \geq \frac{1}{F(\xi;\theta)}. \tag{1.4.2}$$

The Cramér-Rao inequality can be extended to vector-valued parameters and to random parameters.

An unbiased estimator is efficient if its variance is equal to the right-hand side of Eq. 1.4.1.

The equality condition in Eq. 1.4.1 is of particular interest. This condition is

$$M(\theta) \{t - \tau(\theta)\} = K(\theta) \frac{\partial}{\partial\theta} \log p(x;\theta) \quad \text{a.e. } P_\theta. \tag{1.4.3}$$

Here the two constants $M(\theta)$ and $K(\theta)$ are non-negative and not both zero. From this equality condition, we can obtain two interesting conclusions.

First of all we can obtain a family of distributions $p(x;\theta)$ which satisfies Eq. 1.4.3 and, therefore, guarantees that a statistic which has such a distribution is an MVB estimator for $\tau(\theta)$. This family is the so-called exponential family. If θ is a location parameter, we find a gaussian distribution, and if θ is a scale parameter, we find a gamma distribution (see Kagan et al. [30]).

Secondly we note that the condition of Eq. 1.4.3 is a special case of the condition for sufficiency according to Definition 1.3.2. This means that an MVB estimator for $\tau(\theta)$ can only exist if there exists a sufficient statistic for $\tau(\theta)$. We also see that sufficiency is a less restrictive condition than the attainment of the minimum variance bound.

Chapman et al. [12] have given a modification for the Cramér-Rao bound for the case that θ is a discrete parameter. Improved versions of the Cramér-Rao inequality have been given by Bhattacharyya and Barankin in the sense that their bounds are sharper. Bhattacharyya [6] has obtained a lower bound by considering higher order derivatives of the likelihood function $p(x;\theta)$. It is given by

$$E_\theta |T - \tau(\theta)|^2 \geq \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\partial^j \tau(\theta)}{\partial\theta^j} J_{ij}^{-1} \frac{\partial^i \tau(\theta)}{\partial\theta^i} \qquad (1.4.4)$$

where J_{ij}^{-1} is the ij-th element of the inverse of a matrix with coefficients

$$J_{ij} = \int_X \left[\frac{\partial^i p(x;\theta)}{\partial\theta^i} \cdot \frac{\partial^j p(x;\theta)}{\partial\theta^j} \right] \frac{1}{p(x;\theta)} \, dx. \qquad (1.4.5)$$

It is easy to see that for $n = 1$ we have the Cramér-Rao inequality. Its application has been limited, however, mainly because of its complexity.

The best possible bound has been developed by Barankin [3] for locally best estimators. It can be given by

$$E|T - \tau(\theta)|^2 \geq \text{l.u.b} \frac{\left| \sum_{i=1}^{n} a_i \, \tau(\theta_i) \right|^2}{\int_X \left| \sum_{i=1}^{n} a_i \frac{p(x;\theta_i)}{p(x;\theta)} \right|^2 p(x;\theta) \, dx} . \qquad (1.4.6)$$

The lowest upper bound is taken over any set of n parameter points $\theta_1, \ldots, \theta_n$ and any set of n real numbers $a_1, \ldots, a_n$ for every $n = 1, 2, \ldots$. Barankin has shown that his bound includes the two bounds mentioned before. The optimalization procedure which is necessary in this bound is in general quite difficult.

In his original paper [3] he used the mean s-th absolute moment as a measure for the performance of estimators. Boekee [7] [8], Mathai [40], Papantoni-Kazakos [45] and Vajda [55] have paid attention to various aspects of this performance measure. In this thesis we shall develop a unified approach, which is based on an information measure which will be called the Fisher information of order s.

CHAPTER 2

THE FISHER INFORMATION OF ORDER S

In this chapter we will introduce an information measure which is an extension of the Fisher information measure. It will be called the Fisher information of order s.

We will start with definitions of the self-information and of the information of order s. Then we will consider the major properties of this new information measure. It will be shown that it can be seen as a measure for the information about a parameter θ, contained in the observations. Using this information measure we shall obtain a lower bound on the s-th absolute central moment of parameter estimators. We also introduce the notion of an MSB estimator and give some results for such an estimator.

It will be shown that the Fisher information of order s is related to two other information measures. We shall also consider further extensions to random parameters, for which we will obtain appropriate information measures.

The main emphasis in this chapter is on continuous observations and on random and non-random parameters, but we also include some results for discrete observations.

2.1 DEFINITIONS

In this section we shall give definitions of a generalization of the Fisher information measure. Let ξ be a real-valued random variable which represents the observations and is defined over the measurable space (or sample space) $(X,\mathcal{A})$. The possible outcomes of ξ are points $x \in X$ and $\mathcal{A}$ is a Borel field. Let P be the probability measure or distribution function of ξ, defined over $\mathcal{A}$, and let the probability density function of ξ be $dP/dx = p(x)$ for every $x \in X$. We assume that P depends on a real, non-random parameter $\theta \in \Theta$, where Θ is an open subset of the real line. We also assume that X does not depend on θ. For every $\theta \in \Theta$ we consider the density function $p(x;\theta)$. This leads to a family of density functions $\{P_\theta\} = \{p(x;\theta), \theta \in \Theta\}$.

We will assume that $p(x;\theta)$ satisfies the following conditions

$$\text{(i)} \qquad p(x;\theta) > 0 \qquad x \in X,\ \theta \in \Theta \tag{2.1.1a}$$

$$\text{(ii)} \qquad \frac{\partial}{\partial\theta} p(x;\theta) = p'(x;\theta) \qquad \text{exists for every } x \in X,\ \theta \in \Theta. \tag{2.1.1b}$$

If condition (i) is not satisfied for every $x \in X$ we have to consider the support of ξ, denoted by $X_s \subset X$, with $p(x;\theta) > 0$ for $x \in X_s$.

We now introduce the Fisher self-information as the information about θ, which is contained in the observed value x of the random variable ξ.

DEFINITION 2.1.1

Let the density function $p(x;\theta)$ satisfy conditions (i) and (ii). Then the Fisher self-information of θ is defined as

$$f(x;\theta) = \left|\frac{\partial}{\partial\theta} \log p(x;\theta)\right| = \left|\frac{p'(x;\theta)}{p(x;\theta)}\right| \tag{2.1.2}$$

where p' denotes the derivative with respect to θ. Here and in the sequel log will mean the natural logarithm.

The existence of $f(x;\theta)$ follows from the conditions (i) and (ii). As we can see, the Fisher self-information is non-negative:

$$f(x;\theta) \geq 0. \tag{2.1.3}$$

Equality in Eq. 2.1.3 holds iff ξ does not depend on θ. This follows from Definition 2.1.1. The equality condition is found by noting that if $p(x;\theta) = p(x)$ then $p'(x;\theta) = 0$. The "only if" part is proved by noting that $f(x;\theta) = 0$ means that $p'(x;\theta)/p(x;\theta) = 0$. From condition (i) it follows that $p'(x;\theta) = 0$.

We now introduce a joint and a conditional Fisher self-information. We consider two random variables ξ and η with density functions $p(x;\theta)$ and $q(y;\theta)$ and with joint density function $r(x,y;\theta) = p(x;\theta).q(y/x;\theta)$.

DEFINITION 2.1.2

The joint Fisher self-information is defined as

$$f(x,y;\theta) = \left|\frac{\partial}{\partial\theta} \log r(x,y;\theta)\right| = \left|\frac{r'(x,y;\theta)}{r(x,y;\theta)}\right|, \tag{2.1.4}$$

provided $r(x,y;\theta)$ satisfies conditions similar to those of Eqs. 2.1.1a and 2.1.1b.

DEFINITION 2.1.3

The conditional Fisher self-information is defined as

$$f(y/x;\theta) = \left|\frac{\partial}{\partial\theta} \log q(y/x;\theta)\right| = \left|\frac{q'(y/x;\theta)}{q(y/x;\theta)}\right|, \tag{2.1.5}$$

provided $q(y/x;\theta)$ satisfies conditions similar to those of Eqs. 2.1.1a and 2.1.1b.

Using Definitions 2.1.2 and 2.1.3 we can prove that $f(x;\theta)$ is sub-additive.

THEOREM 2.1.1

The Fisher self-information is sub-additive in the sense that

$$f(x,y;\theta) \le f(x;\theta) + f(y/x;\theta). \tag{2.1.6}$$

Proof

From Definition 2.1.2 we have

$$\begin{aligned} f(x,y;\theta) &= \left|\frac{\partial}{\partial\theta} \log r(x,y;\theta)\right| \\ &= \left|\frac{\partial}{\partial\theta} \log p(x;\theta) + \frac{\partial}{\partial\theta} \log q(y/x;\theta)\right| \\ &\le \left|\frac{\partial}{\partial\theta} \log p(x;\theta)\right| + \left|\frac{\partial}{\partial\theta} \log q(y/x;\theta)\right| \\ &= f(x;\theta) + f(y/x;\theta). \end{aligned} \tag{2.1.7}$$

□

COROLLARY 2.1.1

If ξ and η are independent, we have

$$f(x,y;\theta) \le f(x;\theta) + f(y;\theta) \tag{2.1.8}$$

for every $\theta \in \Theta$.

Proof

This follows from Theorem 2.1.1 by noting that in this case $q(y/x;\theta) = q(y;\theta)$. □

The Fisher self-information will be the basic quantity in the following chapters, from which several information measures will be derived. It is a random variable since it depends on the value x of the random variable ξ.

The question, if it is generally true that $f(y/x;\theta) \le f(y;\theta)$, must be answered negatively. Whether $f(y/x;\theta)$ is greater or less than $f(y;\theta)$

depends on the probability density functions $q(y;\theta)$ and $q(y/x;\theta)$.

Since the self-information depends on the actual value x of ξ, we need a measure which represents an average value of the information about θ contained in ξ. To this end we define the Fisher information of order s as a function of the r-th power of $f(x;\theta)$ where $r = s/(s-1)$.

DEFINITION 2.1.4

Let the density function $p(x;\theta)$ of the random variable ξ satisfy conditions (i) and (ii) of Eqs. 2.1.1a and 2.1.1b. Then if $f(x;\theta)$ exists, we define the Fisher information of order s for $s \geq 1$ as

$$F_s(\xi;\theta) = \left[E_\theta \left\{ f(\xi;\theta) \right\}^{\frac{s}{s-1}} \right]^{s-1} \tag{2.1.9a}$$

$$= \left[E_\theta \left| \frac{\partial}{\partial\theta} \log p(\xi;\theta) \right|^{\frac{s}{s-1}} \right]^{s-1} \tag{2.1.9b}$$

$$= \left[\int_X \left| \frac{\partial}{\partial\theta} \log p(x;\theta) \right|^{\frac{s}{s-1}} p(x;\theta)\, dx \right]^{s-1} \tag{2.1.9c}$$

$$= \left[\int_X \left| \frac{p'(x;\theta)}{p(x;\theta)} \right|^{\frac{s}{s-1}} p(x;\theta)\, dx \right]^{s-1} . \tag{2.1.9d}$$

The cases $s\to\infty$ and $s\to 1$ need some special clarification and will therefore be considered in the Theorems 2.2.2 and 2.2.3.

When there is no possibility of confusion, we shall sometimes use the notation $F_s(\theta)$ instead of $F_s(\xi;\theta)$. The notation E_θ should be interpreted as an expectation with respect to $\{P_\theta\} = \{p(x;\theta),\ \theta \in \Theta\}$ as is clear from Eq. 2.1.9c. With the notation $p(\xi;\theta)$ we shall mean a random variable which can take on the values $p(x;\theta)$, $x \in X$.

Note that the well-known Fisher information measure is included in this definition if we set s=2. The Fisher information of order s can be seen as the s-th power of the s/(s-1)-norm of the function $\partial \log p(x;\theta)/\partial\theta$.

It is possible to extend the definition to the bivariate case when we consider the random variables ξ and η with density functions $p(x;\theta)$ and $q(y;\theta)$ having a joint density function $r(x,y;\theta)$. We need a joint and a conditional information measure.

DEFINITION 2.1.5

Let the joint density function $r(x,y;\theta)$ of the random variables ξ and η satisfy conditions similar to those of Eqs. 2.1.1a and 2.1.1b. Then for $s \geq 1$ the joint Fisher information of order s is defined as

$$F_s(\xi,\eta;\theta) = \left[\int_X \int_Y \left| \frac{\partial}{\partial\theta} \log r(x,y;\theta) \right|^{\frac{s}{s-1}} r(x,y;\theta) \, dx \, dy \right]^{s-1} . \tag{2.1.10}$$

DEFINITION 2.1.6

Let the conditional density function $q(y/x;\theta)$ of the random variable η, given $\xi=x$, exist for every $x \in X$, $\theta \in \Theta$ and let it satisfy conditions similar to those of Eqs. 2.1.1a and 2.1.1b. Then for $s \geq 1$ the conditional Fisher information of order s is defined as

$$F_s(\eta/\xi;\theta) = \left[\int_X \int_Y \left| \frac{\partial}{\partial\theta} \log q(y/x;\theta) \right|^{\frac{s}{s-1}} r(x,y;\theta) \, dx \, dy \right]^{s-1} . \tag{2.1.11}$$

COROLLARY 2.1.2

If ξ and η are independent random variables, then

$$F_s(\eta/\xi;\theta) = F_s(\eta;\theta). \tag{2.1.12}$$

Proof

This follows from Definition 2.1.6 by noting that in this case $q(y/x;\theta) = q(y;\theta)$ and $r(x,y;\theta) = p(x;\theta).q(y;\theta)$. □

2.2 BASIC PROPERTIES

In Section 2.1 we have given some definitions of information measures. In this section we shall consider properties of these information measures, like invariance and convexity. Also a bound on the s-th absolute central moment (s.a.c.m.) of estimators of θ, based on $F_s(\xi;\theta)$, will be derived.

First, we will show that $F_s(\xi;\theta)$ is a non-negative quantity.

THEOREM 2.2.1

For $s \geq 1$ it holds that

$$F_s(\xi;\theta) \geq 0. \tag{2.2.1}$$

Equality in Eq. 2.2.1 holds iff $p(x;\theta)$ does not depend on θ a.e. x.

Proof

The theorem follows from the property that $f(\xi;\theta)$ is non-negative and from Definition 2.1.4. □

Now we shall consider the cases $s \to \infty$ and $s \downarrow 1$.

THEOREM 2.2.2

For $s \to \infty$ the Fisher information of order s exists iff

$$\int_X \left| \frac{\partial}{\partial\theta} p(x;\theta) \right| dx < 1 \qquad \theta \in \Theta \tag{2.2.2}$$

and we have

$$F_\infty(\xi;\theta) = 0. \tag{2.2.3}$$

Proof

Using Definition 2.1.1 we find that

$$\lim_{s\to\infty} E_\theta \left\{ f(\xi;\theta) \right\}^{\frac{s}{s-1}} = E_\theta\{f(\xi;\theta)\}$$

$$= \int_X \left| \frac{\partial}{\partial\theta} p(x;\theta) \right| dx. \tag{2.2.4}$$

Therefore, $\lim_{s\to\infty} F_s(\xi;\theta)$ exists iff the expression of Eq. 2.2.4 is less than 1, and in this case Eq. 2.2.3 follows immediately from Definition 2.1.4. □

THEOREM 2.2.3

For s=1 the Fisher information of order s is

$$F_1(\xi;\theta) = \text{ess. sup.} \left| \frac{\partial}{\partial\theta} \log p(x;\theta) \right| , \tag{2.2.5}$$

provided it exists.

Proof

We have

$$\lim_{s\to 1} F_s(\xi;\theta) = \lim_{s\to 1} F_s(\xi;\theta)^{\frac{1}{s}}$$

$$= \lim_{s\to 1} \left[E_\theta \left| \frac{\partial}{\partial\theta} \log p(\xi;\theta) \right|^{\frac{s}{s-1}} \right]^{\frac{s-1}{s}}$$

$$= \lim_{r\to\infty} \left[E_\theta \left| \frac{\partial}{\partial\theta} \log p(\xi;\theta) \right|^{r} \right]^{\frac{1}{r}} \qquad (2.2.6)$$

which gives (see Hardy, et al. [25])

$$\lim_{s\to 1} F_s(\xi;\theta) = \text{ess. sup.} \left| \frac{\partial}{\partial\theta} \log p(x;\theta) \right| . \qquad (2.2.7)$$

□

This theorem can be seen as a definition of $F_s(\xi;\theta)$ for s=1.

For several reasons it is of interest to consider the behaviour of $F_s(\xi;\theta)$ as a function of $p(x;\theta)$.

THEOREM 2.2.4

For $s \geq 2$ the Fisher information of order s is a convex function of $p(x;\theta)$.

Proof

The proof for the case s=2 has been given by Stam [53] and Cohen [15]. Here we give a more general proof for the case $s \geq 2$.

Let the random variables ξ_1 and ξ_2 with density functions $p_1(x;\theta)$ and $p_2(x;\theta)$ satisfy the conditions of Eqs. 2.1.1a and 2.1.1b. We first note that from the convexity of $|w|^a$, $a \geq 1$ we have with

$$w = F_s(\xi;\theta)^{\frac{1}{s-1}} \qquad (2.2.8a)$$

and

$$a = s - 1 \tag{2.2.8b}$$

that the following inequality holds for $0 \le \tau \le 1$ and $s \ge 2$

$$\tau\, F_s(\xi_1;\theta) + (1-\tau)\, F_s(\xi_2;\theta)$$

$$\ge \left[\int_X \tau \left| \frac{\partial}{\partial\theta} \log p_1(x;\theta) \right|^{\frac{s}{s-1}} p_1(x;\theta)\, dx + \right.$$

$$\left. + \int_X (1-\tau) \left| \frac{\partial}{\partial\theta} \log p_2(x;\theta) \right|^{\frac{s}{s-1}} p_2(x;\theta)\, dx \right]^{s-1} . \tag{2.2.9}$$

Secondly we note that

$$\left| \frac{u}{v} \right|^r v \tag{2.2.10}$$

is a convex two-variable function for $r \ge 1$. This follows from the convexity of $|w|^r$, $r \ge 1$ or

$$\alpha|w_1|^r + (1-\alpha)|w_2|^r \ge |\alpha w_1 + (1-\alpha)w_2|^r \tag{2.2.11}$$

if we substitute

$$w_1 = \frac{u_1}{v_1} , \tag{2.2.12a}$$

$$w_2 = \frac{u_2}{v_2} , \tag{2.2.12b}$$

$$\alpha = \frac{\tau v_1}{\tau v_1 + (1-\tau) v_2} , \tag{2.2.12c}$$

$$1 - \alpha = \frac{(1-\tau)v_2}{\tau v_1 + (1-\tau)v_2} \tag{2.2.12d}$$

into Eq. 2.2.11 to obtain

$$\tau \left| \frac{u_1}{v_1} \right|^r v_1 + (1-\tau) \left| \frac{u_2}{v_2} \right|^r v_2$$

$$\geq \left| \frac{\tau u_1 + (1-\tau)u_2}{\tau v_1 + (1-\tau)v_2} \right|^r \{\tau v_1 + (1-\tau)v_2\}. \tag{2.2.13}$$

In Eq. 2.2.13 we substitute

$$u_1 = \frac{\partial}{\partial\theta} p_1(x;\theta), \tag{2.2.14a}$$

$$u_2 = \frac{\partial}{\partial\theta} p_2(x;\theta), \tag{2.2.14b}$$

$$v_1 = p_1(x;\theta), \tag{2.2.14c}$$

$$v_2 = p_2(x;\theta), \tag{2.2.14d}$$

$$r = \frac{s}{s-1}, \tag{2.2.14e}$$

and integrate with respect to x. The left-hand side of this result is identical with the right-hand side of Eq. 2.2.9 to the power $1/(s-1)$, so that we finally obtain that

$$\tau F_s(\xi_1;\theta) + (1-\tau) F_s(\xi_2;\theta) \geq F_s(\tau p_1 + (1-\tau)p_2;\theta), \tag{2.2.15}$$

which proves the theorem. □

Equation 2.2.15 can also be interpreted as an upper bound on $F_s(\xi;\theta)$. This bound can easily be extended to the case in which the density function $p(x)$ is a mixture of n density functions $p_i(x)$, $i=1,2,\ldots,n$. We then have

$$p(x) = \sum_{i=1}^{n} a_i \, p_i(x), \tag{2.2.16}$$

in which the coefficients a_i satisfy

$$0 \le a_i \le 1 \tag{2.2.17a}$$

and

$$\sum_{i=1}^{n} a_i = 1. \tag{2.2.17b}$$

Then a mathematical induction applied to Eq. 2.2.15 results in the following corollary.

COROLLARY 2.2.1

If $p(x;\theta)$ satisfies Eq. 2.2.16, then

$$F_s(p;\theta) \le \sum_{i=1}^{n} a_i \, F_s(p_i;\theta). \tag{2.2.18}$$

Some applications of Corollary 2.2.1 will be considered in Section 4.2.

Another important property of $F_s(\xi;\theta)$ is invariance under a measurable transformation T. We will show that under this condition T has to be a sufficient statistic.

Let T be a measurable transformation, or statistic, of the observation ξ taking on values in the space Y. Then this transformation generates a class of Borel sets $\mathcal{B}$ whose inverse image

$$A = T^{-1}(B) = \{x\colon\, x \in X,\ T(x) \in B\} \tag{2.2.19}$$

is in $\mathcal{A}$ for all $B \in \mathcal{B}$.

Then $T = T(\xi)$ is a random variable over the measurable space $(Y, \mathcal{B})$. Furthermore, since $\xi \in T^{-1}(B)$ iff $T(\xi) \in B$, the probability distribution of $T(\xi)$ is for $\theta \in \Theta$ given by

$$Q_\theta(B) = Q_\theta(T \in B) = Q_\theta(\xi \in T^{-1}(B)) = P_\theta(T^{-1}(B)) = P_\theta(A) \quad (2.2.20)$$

and is said to be induced by T. It is sometimes denoted by $Q_\theta = T^{-1}P_\theta$. The density function of $T(\xi)$ will be denoted by $q(t;\theta)$ for $t \in Y$, $\theta \in \Theta$.

THEOREM 2.2.5

Let the family of density functions $\{P_\theta\} = \{p(x;\theta), \theta \in \Theta\}$ of the random variable ξ, defined over the probability space $(X, \mathcal{A}, \{P_\theta\})$, satisfy the conditions of Eqs. 2.1.1a and 2.1.1b and let the additional condition

(iii) for any $A \in \mathcal{A}$ and every $\theta \in \Theta$ the integral $\int p(x;\theta)\, dx$ can be differentiated under the integral sign,

$$\frac{\partial}{\partial\theta} \int_A p(x;\theta)\, dx = \int_A \frac{\partial}{\partial\theta} p(x;\theta)\, dx, \quad (2.2.21)$$

be satisfied.

Let $T = T(\xi)$ be a measurable transformation (or statistic) of ξ and let $\{Q_\theta\}$ be the family of density functions induced by the transformation T. We assume that $\{Q_\theta\}$ satisfies conditions similar to those of Eqs. 2.1.1a, 2.1.1b and 2.2.21.

Then $F_s(T(\xi);\theta)$ exists if $F_s(\xi;\theta)$ exists, and for $s \geq 1$ we have

$$F_s(T(\xi);\theta) \leq F_s(\xi;\theta). \quad (2.2.22)$$

Equality in Eq. 2.2.22 holds iff (for $s > 1$) or if (for $s = 1$) T is a sufficient statistic for the family $\{P_\theta\}$.

Proof

A proof of this theorem for the case s=2 can be found, e.g., in Stam [53] and Rao [46]. To prove the general case we proceed as follows (see also Stam [53]).

First we note that $Q_\theta = T^{-1}P_\theta$ and $A = T^{-1}B$, where $B \in \mathcal{B}$. Then for any $A \in \mathcal{A}$ we have, using Eq. 2.2.21, that

$$\begin{aligned}\int_A \frac{p'(x;\theta)}{p(x;\theta)}\, p(x;\theta)\, dx &= \frac{\partial}{\partial\theta}\int_A p(x;\theta)\, dx \\ &= \frac{\partial}{\partial\theta} P_\theta(A) \\ &= \frac{\partial}{\partial\theta} Q_\theta(B) \\ &= \frac{\partial}{\partial\theta}\int_B q(t;\theta)\, dt \\ &= \int_B \frac{\partial}{\partial\theta} q(t;\theta)\, dt \\ &= \int_B \frac{q'(t;\theta)}{q(t;\theta)}\, q(t;\theta)\, dt\ . \end{aligned} \qquad (2.2.23)$$

But, according to the definition of conditional expectation, we also have

$$\begin{aligned}\int_A \frac{p'(x;\theta)}{p(x;\theta)}\, p(x;\theta)\, dx &= \int_A \left[\int_B \left\{\frac{p'(x;\theta)}{p(x;\theta)}\ \Big|\, t\right\} q(t/x;\theta)\, dt\right] p(x;\theta)\, dx \\ &= \int_A\int_B \left\{\frac{p'(x;\theta)}{p(x;\theta)}\ \Big|\, t\right\} r(x,t;\theta)\, dx\, dt\end{aligned}$$

$$= \int_B \left[\int_A \left\{ \frac{p'(x;\theta)}{p(x;\theta)} \Big| t \right\} p(x/t;\theta)\, dx \right] q(t;\theta)\, dt \,. \tag{2.2.24}$$

Combining Eqs. 2.2.23 and 2.2.24 yields

$$\frac{q'(t;\theta)}{q(t;\theta)} = \int_A \left\{ \frac{p'(x;\theta)}{p(x;\theta)} \Big| t \right\} p(x/t;\theta)\, dx \quad \text{a.e. } Q_\theta$$

$$= E_\theta \left\{ \frac{p'(\xi;\theta)}{p(\xi;\theta)} \Big| t \right\} \quad \text{a.e. } Q_\theta . \tag{2.2.25}$$

We now have

$$E_\theta \left| \frac{p'(\xi;\theta)}{p(\xi;\theta)} \right|^{\frac{s}{s-1}} = E_\theta \left[E_\theta \left\{ \left| \frac{p'(\xi;\theta)}{p(\xi;\theta)} \right|^{\frac{s}{s-1}} \Big| T(\xi) \right\} \right] . \tag{2.2.26}$$

From Jensen's inequality (see Appendix A) we obtain

$$E_\theta \left| \frac{p'(\xi;\theta)}{p(\xi;\theta)} \right|^{\frac{s}{s-1}} \geq E_\theta \left[\left| E_\theta \left\{ \frac{p'(\xi;\theta)}{p(\xi;\theta)} \Big| T(\xi) \right\} \right|^{\frac{s}{s-1}} \right] , \tag{2.2.27}$$

and using Eq. 2.2.25 yields for Eq. 2.2.27:

$$E_\theta \left| \frac{p'(\xi;\theta)}{p(\xi;\theta)} \right|^{\frac{s}{s-1}} \geq E_\theta \left| \frac{q'(T(\xi);\theta)}{q(T(\xi);\theta)} \right|^{\frac{s}{s-1}} . \tag{2.2.28}$$

Finally, we find for $s \geq 1$ that

$$\left[E_\theta \left| \frac{q'(T(\xi);\theta)}{q(T(\xi);\theta)} \right|^{\frac{s}{s-1}} \right]^{s-1} \leq \left[E_\theta \left| \frac{p'(\xi;\theta)}{p(\xi;\theta)} \right|^{\frac{s}{s-1}} \right]^{s-1} , \tag{2.2.29}$$

which proves the first part of the theorem.

The equality condition is

$$\frac{p'(x;\theta)}{p(x;\theta)} = \frac{q'(T(x);\theta)}{q(T(x);\theta)} \qquad \text{a.e. } P_\theta \tag{2.2.30}$$

or

$$\frac{\partial}{\partial\theta} \log p(x;\theta) = \frac{\partial}{\partial\theta} \log q(T(x);\theta) \qquad \text{a.e. } P_\theta. \tag{2.2.31}$$

From Eq. 2.2.31 we obtain after integration with respect to θ that for $\theta \in \Theta$

$$\log p(x;\theta) = \log q(T(x);\theta) + \log h(x) \quad \text{a.e. } x \tag{2.2.32}$$

or

$$p(x;\theta) = q(T(x);\theta) \,.\, h(x) \qquad \text{a.e. } x. \tag{2.2.33}$$

This is precisely the condition under which T is a sufficient statistic for the family $\{P_\theta\}$, as follows from the well-known factorization theorem (see Kagan et al. [30]). For $s > 1$ it is an iff condition, whereas for $s = 1$ it is an if condition. □

The theorem shows the important result that for the Fisher information of order s a transformation can only lead to a loss of information. From the equality condition it follows that invariance of information is closely related to the concept of sufficiency. Since a statistic can be considered as a transformation, the same conclusions that hold for a transformation hold for a statistic.

We can also consider $F_s(\xi;\theta)$ as a function of its order parameter s. The properties which we will obtain follow from the relation of $F_s(\xi;\theta)$ to the mean of order $s/(s-1)$. Some inequalities for the mean of order α can be found in Appendix A.

THEOREM 2.2.6

For $1 \le s < m$ it holds that

$$F_s(\xi;\theta) > F_m(\xi;\theta)^{s/m}, \tag{2.2.34}$$

provided $F_s(\xi;\theta)$ exists.

Proof

The condition m > s is equivalent to the condition

$$\frac{s}{s-1} = 1 + \frac{1}{s-1} > 1 + \frac{1}{m-1} = \frac{m}{m-1} . \tag{2.2.35}$$

Then it follows from the order inequality for means of order α, as given in Appendix A, that Eq. 2.2.34 holds. We have strict inequality unless ξ does not depend on θ. □

THEOREM 2.2.7

For $1 \le r < s < m$ it holds that

$$F_s(\xi;\theta) < F_r(\xi;\theta)^{\frac{m-s}{m-r}} \cdot F_m(\xi;\theta)^{\frac{s-r}{m-r}}, \tag{2.2.36}$$

provided $F_r(\xi;\theta)$ exists.

Proof

As in the proof of Theorem 2.2.6 it follows that $r < s < m$ is equivalent to

$$\frac{r}{r-1} > \frac{s}{s-1} > \frac{m}{m-1} . \tag{2.2.37}$$

Setting

$$\alpha = \frac{m}{m-1} , \tag{2.2.38a}$$

$$\beta = \frac{s}{s-1} , \tag{2.2.38b}$$

and

$$\delta = \frac{r}{r-1} , \qquad (2.2.38c)$$

yields $$\delta - \alpha = \frac{m-r}{(r-1)(m-1)} \qquad (2.2.39a)$$

$$\delta - \beta = \frac{s-r}{(r-1)(s-1)} \qquad (2.2.39b)$$

and $$\beta - \alpha = \frac{m-s}{(s-1)(m-1)} \qquad (2.2.39c)$$

Using the mean of order α, as given in Appendix A, Eq. A.7, we find

$$M_{\alpha}^{\alpha} \left\{ \frac{p'(x;\theta)}{p(x;\theta)} \right\} = F_m(\xi;\theta)^{\frac{1}{m-1}} . \qquad (2.2.40)$$

Similarly we can relate M_β and M_δ to $F_s(\xi;\theta)$ and $F_r(\xi;\theta)$.

Substituting Eqs. 2.2.38a to 2.2.39c into inequality 6. of Appendix A yields

$$F_s(\xi;\theta)^{\frac{m-r}{(r-1)(s-1)(m-1)}} < F_r(\xi;\theta)^{\frac{m-s}{(r-1)(s-1)(m-1)}} . \, F_m(\xi;\theta)^{\frac{s-r}{(r-1)(s-1)(m-1)}} \qquad (2.2.41)$$

As a result of Eq. 2.2.41 we easily obtain Eq. 2.2.36, which proves the theorem. □

From this theorem we obtain two corollaries.

COROLLARY 2.2.2

For $s = \frac{1}{2}(r+m)$ it holds that

$$F_s(\xi;\theta) < \left[F_r(\xi;\theta) \, . \, F_m(\xi;\theta) \right]^{\frac{1}{2}} . \qquad (2.2.42)$$

Proof

Substitution of $s = \frac{1}{2}(r+m)$ into Eq. 2.2.36 proves Eq. 2.2.42. □

COROLLARY 2.2.3

For $s \geq 1$ it holds that $\log F_s(\xi;\theta)$ is a convex function of s.

Proof

From Eq. 2.2.36 we obtain by setting

$$\tau = \frac{m-s}{m-r} \tag{2.2.43}$$

and taking logarithms on both sides, that

$$\log F_s(\xi;\theta) < \tau \log F_r(\xi;\theta) + (1-\tau) \log F_m(\xi;\theta). \tag{2.2.44}$$

This proves the corollary. □

The next theorem establishes a relation between the accuracy of the estimators of some differentiable function $\tau(\theta)$ of the parameter θ and the Fisher information of order s. We assume that the density function $p(x;\theta)$ satisfies the conditions

(i) $p(x;\theta) > 0 \qquad x \in X,\ \theta \in \Theta$ (2.2.45a)

(ii) $\frac{\partial}{\partial\theta} p(x;\theta)$ exists for every $x \in X$, $\theta \in \Theta$ except possibly a finite number of points θ where the left and the right derivatives exist and are finite (2.2.45b)

(iii) $\int p(x;\theta)\, dx$ can be differentiated under the integral sign (2.2.45c)

(iv) $E_\theta \left| \frac{\partial}{\partial\theta} \log p(\xi;\theta) \right|^{\frac{s}{s-1}} > 0, \qquad \theta \in \Theta$ (2.2.45d)

(v) $\int t\, p(x;\theta)\, dx$ can be differentiated under the integral sign. (2.2.45e)

THEOREM 2.2.8

If $T = T(\xi)$ is an unbiased estimator of the continuously differentiable function $\tau(\theta)$ and if $p(x;\theta)$ satisfies the conditions (i) - (v), then for $s \geq 1$

$$E_\theta |T - \tau(\theta)|^s \geq \frac{\left|\frac{d}{d\theta}\tau(\theta)\right|^s}{F_s(\xi;\theta)} . \qquad (2.2.46a)$$

Equality in Eq. 2.2.46a holds iff (for $s > 1$) or if (for $s = 1$)

$$M(\theta)\{t - \tau(\theta)\} = K(\theta)\left|\frac{\partial}{\partial\theta}\log p(x;\theta)\right|^{\frac{1}{s-1}} \operatorname{sgn}\left[\frac{\partial}{\partial\theta}\log p(x;\theta)\right] \text{ a.e.} P_\theta,$$

(2.2.46b)

where $$E_\theta\left[\left|\frac{\partial}{\partial\theta}\log p(\xi;\theta)\right|^{\frac{1}{s-1}} \operatorname{sgn}\left[\frac{\partial}{\partial\theta}\log p(\xi;\theta)\right]\right] = 0 \qquad (2.2.46c)$$

and where the non-negative constants $M(\theta)$ and $K(\theta)$ are not both zero.

Proof

By definition we have

$$\int_X p(x;\theta)\, dx = 1. \qquad (2.2.47)$$

Differentiating both sides with respect to θ and using condition (iii) yields

$$\int_X \frac{\partial}{\partial\theta} p(x;\theta)\, dx = 0 \qquad (2.2.48)$$

or $$\int_X \frac{\partial}{\partial\theta} \log p(x;\theta)\ p(x;\theta)\ dx = 0 \tag{2.2.49}$$

From this it follows that

$$\int_X \tau(\theta) \frac{\partial}{\partial\theta} \log p(x;\theta)\ p(x;\theta)\ dx = 0, \tag{2.2.50}$$

since x does not depend on θ. The statistic T is unbiased, which means that

$$\int_X t\ p(x;\theta)\ dx = \tau(\theta). \tag{2.2.51}$$

If we differentiate Eq. 2.2.51 with respect to θ and use condition (v) we obtain

$$\int_X t \frac{\partial}{\partial\theta} \log p(x;\theta)\ p(x;\theta)\ dx = \frac{d}{d\theta} \tau(\theta) = \tau'(\theta), \tag{2.2.52}$$

where $\tau'(\theta) \neq 0$.

Combining Eqs. 2.2.50 and 2.2.52 results in

$$\int_X \{t - \tau(\theta)\} \frac{\partial}{\partial\theta} \log p(x;\theta)\ p(x;\theta)\ dx = \tau'(\theta). \tag{2.2.53}$$

Noting that

$$\left| \frac{t - \tau(\theta)}{\tau'(\theta)} \right| \left| \frac{\partial}{\partial\theta} \log p(x;\theta) \right| \geq \frac{t - \tau(\theta)}{\tau'(\theta)} \frac{\partial}{\partial\theta} \log p(x;\theta), \tag{2.2.54}$$

it follows from Hölder's inequality (see Appendix A) that

$$\left[E_\theta \left| \frac{T - \tau(\theta)}{\tau'(\theta)} \right|^s \right]^{\frac{1}{s}} \cdot \left[E_\theta \left| \frac{\partial}{\partial\theta} \log p(\xi;\theta) \right|^{\frac{s}{s-1}} \right]^{\frac{s-1}{s}} \geq 1 \tag{2.2.55}$$

$$\text{or} \qquad E_\theta \left| T - \tau(\theta) \right|^s \cdot \left[E_\theta \left| \frac{\partial}{\partial\theta} \log p(\xi;\theta) \right|^{\frac{s}{s-1}} \right]^{s-1} \geq \left| \tau'(\theta) \right|^s . \tag{2.2.56}$$

Using Definition 2.1.4 in Eq. 2.2.56 proves the first part of the theorem.

Equality in Eq. 2.2.46a is obtained iff the two conditions

$$\operatorname{sgn} [t - \tau(\theta)] = \operatorname{sgn}\left[\frac{\partial}{\partial\theta} \log p(x;\theta) \right] \tag{2.2.57a}$$

$$\text{and} \qquad M(\theta) \left| t - \tau(\theta) \right| = K(\theta) \left| \frac{\partial}{\partial\theta} \log p(x;\theta) \right|^{\frac{1}{s-1}} \quad \text{a.e. } P_\theta \tag{2.2.57b}$$

are satisfied. Here $M(\theta)$ and $K(\theta)$ are non-negative constants, not both zero.

By combining Eqs. 2.2.57a and 2.2.57b we obtain the necessary and sufficient condition

$$M(\theta)\{t - \tau(\theta)\} = K(\theta) \left| \frac{\partial}{\partial\theta} \log p(x;\theta) \right|^{\frac{1}{s-1}} \operatorname{sgn}\left[\frac{\partial}{\partial\theta} \log p(x;\theta) \right] \quad \text{a.e. } P_\theta . \tag{2.2.58}$$

However, since T is an unbiased estimator of $\tau(\theta)$, it follows that Eq. 2.2.46c must hold. □

Note that for $s = 2$ we have the familiar Cramer-Rao inequality. We thus have obtained a generalized version of the Cramer-Rao bound in terms of the s-th absolute central moment (s.a.c.m.) of the estimator T of $\tau(\theta)$.

As a special case we will consider the inference function $\tau(\theta) = \theta + b(\theta)$, where $b(\theta)$ is the bias of the estimator $T = \hat{\theta}$.

COROLLARY 2.2.4

Let $\hat{\theta}$ be a biased estimator of θ such that

$$E_\theta\{\hat{\theta}\} = \theta + b(\theta) \qquad \theta \in \Theta. \tag{2.2.59}$$

Then under conditions similar to those of Theorem 2.2.8, we have

$$E_\theta|\hat{\theta} - \theta|^s \geq \left[\frac{|1 + b'(\theta)|}{[F_s(\xi;\theta)]^{1/s}} - |b(\theta)| \right]^s \tag{2.2.60}$$

Proof

It follows from Theorem 2.2.8 by substitution of $T = \hat{\theta}$ and $\tau(\theta) = \theta + b(\theta)$ into Eq. 2.2.46a that

$$E_\theta|\hat{\theta} - \theta - b(\theta)|^s \geq \frac{|1 + b'(\theta)|^s}{F_s(\xi;\theta)} . \tag{2.2.61}$$

Noting that

$$\left[E_\theta|\hat{\theta} - \theta - b(\theta)|^s\right]^{1/s} \leq \left[E_\theta|\hat{\theta} - \theta|^s\right]^{1/s} + \left[E_\theta|b(\theta)|^s\right]^{1/s}$$

$$= \left[E_\theta|\hat{\theta} - \theta|^s\right]^{1/s} + |b(\theta)| \tag{2.2.62}$$

and combining Eqs. 2.2.61 and 2.2.62 completes the proof. □

If $\hat{\theta}$ is an unbiased estimator of θ, we find a simple relation.

COROLLARY 2.2.5

Let $\hat{\theta}$ be an unbiased estimator of θ, then under conditions similar to those of Theorem 2.2.8 we have

$$E_\theta|\hat{\theta} - \theta|^s \geq \frac{1}{F_s(\xi;\theta)} . \tag{2.2.63}$$

Proof

The proof follows from Corollary 2.2.4 by substitution of $b(\theta) = 0$ into Eq. 2.2.61. □

An estimator which achieves the bound of Eq. 2.2.46a will be called a minimum s.a.c.m. bound estimator, or MSB estimator. The equality condition 2.2.46b of Theorem 2.2.8 can be used to obtain some properties of MSB estimators.

THEOREM 2.2.9

If Eq. 2.2.46b is satisfied then T is an MSB estimator with the property that

$$E_\theta|T - \tau(\theta)|^s = \frac{|\tau'(\theta)|}{a(\theta)^{s-1}} . \tag{2.2.64}$$

Here $\quad a(\theta) = \dfrac{M(\theta)}{K(\theta)}$, (2.2.65)

where we assume that $K(\theta) > 0$. For the Fisher information of order s it holds that

$$F_s(\xi;\theta) = a(\theta)^{s-1} \; |\tau'(\theta)|^{s-1} . \tag{2.2.66}$$

Proof

From the equality condition 2.2.46b it follows, using Eq. 2.2.65, that

$$a(\theta)\{t - \tau(\theta)\} = \left| \frac{\partial}{\partial\theta} \log p(x;\theta) \right|^{\frac{1}{s-1}} \operatorname{sgn}\left[\frac{\partial}{\partial\theta} \log p(x;\theta)\right] \text{ a.e. } P_\theta . \tag{2.2.67}$$

Using Eq. 2.2.57a we then find

$$\left| \frac{\partial}{\partial\theta} \log p(x;\theta) \right|^{\frac{1}{s-1}} = a(\theta)\left|t - \tau(\theta)\right| , \tag{2.2.68}$$

for which it follows

$$E_\theta \left| \frac{\partial}{\partial\theta} \log p(\xi;\theta) \right|^{\frac{s}{s-1}} = a(\theta)^s \; E_\theta\left|T - \tau(\theta)\right|^s \tag{2.2.69}$$

and $$F_s(\xi;\theta) = a(\theta)^{s(s-1)} \left[E_\theta\left|T - \tau(\theta)\right|^s\right]^{s-1} . \tag{2.2.70}$$

Substitution of Eq. 2.2.70 into Eq. 2.2.46a yields for the equality case:

$$\left[E_\theta\left|T - \tau(\theta)\right|^s\right]^s = \frac{\left|\tau'(\theta)\right|^s}{a(\theta)^{s(s-1)}} , \tag{2.2.71}$$

from which we obtain Eq. 2.2.64. From Eq. 2.2.70 we find that

$$E_\theta\left|T - \tau(\theta)\right|^s = \frac{F_s(\xi;\theta)^{\frac{1}{s-1}}}{a(\theta)^s} . \tag{2.2.72}$$

Substitution of Eq. 2.2.72 into Eq. 2.2.46a yields

$$F_s(\xi;\theta)^{\frac{s}{s-1}} = a(\theta)^s \left|\tau'(\theta)\right|^s , \tag{2.2.73}$$

from which Eq. 2.2.66 follows. □

Based on the inequality of Theorem 2.2.8 we can give a modified definition for the efficiency of estimators.

DEFINITION 2.2.1

Let T be an unbiased estimator of the inference function $\tau(\theta)$. Then the efficiency of T is defined as

$$e_s = \frac{|\tau'(\theta)|^s}{F_s(\xi;\theta) \, . \, E_\theta |T - \tau(\theta)|^s} \, . \qquad (2.2.74)$$

It is clear that $0 \le e_s \le 1$, where $e_s = 1$ if T is an MSB estimator.

It is possible to obtain an expression for the density function $p(x;\theta)$ for which the equality in Eq. 2.2.46a is obtained.

THEOREM 2.2.10

Equality in Theorem 2.2.8 is obtained iff the density function $p(x;\theta)$ is of the form

$$p(x;\theta) = h(x) \exp\left[- \int_\Theta \frac{a(\theta)^{s-1}}{s\,\tau'(\theta)} \frac{\partial}{\partial\theta} |t - \tau(\theta)|^s \, d\theta \right] . \qquad (2.2.75)$$

Proof

From Eq. 2.2.46b we obtain, using Eq. 2.2.57a, and assuming that $K(\theta) > 0$, that

$$\frac{\partial}{\partial\theta} \log p(x;\theta) = \left[\frac{M(\theta)}{K(\theta)}\right]^{s-1} |t - \tau(\theta)|^{s-1} \operatorname{sgn}[t - \tau(\theta)].$$

(2.2.76)

If we set

$$a(\theta) = \frac{M(\theta)}{K(\theta)} \qquad (2.2.77)$$

and integrate Eq. 2.2.76 with respect to θ we obtain

$$\log p(x;\theta) = \int_{\Theta} a(\theta)^{s-1} \, |t - \tau(\theta)|^{s-1} \operatorname{sgn}[t - \tau(\theta)] \, d\theta + \log h(x) \tag{2.2.78}$$

or $$p(x;\theta) = h(x) \exp\left[\int_{\Theta} a(\theta)^{s-1} \, |t - \tau(\theta)|^{s-1} \operatorname{sgn}[t - \tau(\theta)] \, d\theta \right]. \tag{2.2.79}$$

Noting that

$$\frac{\partial}{\partial\theta} |t - \tau(\theta)|^{s} = - s \, \tau'(\theta) \, |t - \tau(\theta)|^{s-1} \operatorname{sgn}[t - \tau(\theta)] \tag{2.2.80}$$

leads to Eq. 2.2.75. □

Note that for s = 2 the density function of Eq. 2.2.75 reduces to the well-known exponential family (see Kagan et al. [30]).

Finally we will show that it is possible to improve the s.a.c.m. of a statistic T_1 if a sufficient statistic T for the family $\{P_\theta\}$ exists.

THEOREM 2.2.11

Let T_1 be an estimator of $\tau(\theta)$ and let T be a sufficient statistic. Then

$$E_{T_1} |T_1 - \tau(\theta)|^{s} \geq E_T \mid E_{T_1}\{T_1/T\} - \tau(\theta) \mid^{s}, \tag{2.2.81}$$

and the conditional expectation $E_{T_1}\{T_1/T\}$ is unbiased, if T_1 is unbiased.

Proof

The proof has essentially been given by Rao [46]. Since T is a sufficient statistic for θ it follows that $E_{T_1}\{T_1/T\}$ does not depend on θ. Then we see that

$$E_{T_1}\{T_1\} = E_T\left[E_{T_1}\{T_1/T\}\right]. \qquad (2.2.82)$$

This means that T_1 and $E_{T_1}\{T_1/T\}$ have the same expectation, from which the second assertion follows. Using Jensen's inequality we find

$$E_{T_1}\left[|T_1 - \tau(\theta)|^s \mid T\right] \geq |E_{T_1}\{T_1 - \tau(\theta)\}| \; T|^s$$

$$= |E_{T_1}\{T_1/T\} - \tau(\theta)|^s. \qquad (2.2.83)$$

From this inequality it follows that

$$E_T\left[E_{T_1}\left\{|T_1 - \tau(\theta)|^s \mid T\right\}\right] = E_{T_1}|T_1 - \tau(\theta)|^s$$

$$\geq E_T\left[|E_{T_1}\{T_1/T\} - \tau(\theta)|^s\right], \qquad (2.2.84)$$

which proves Eq. 2.2.81. Equality in Eq. 2.2.81 holds if

$$T_1 = E_{T_1}\{T_1/T\} \quad \text{a.e.} \qquad (2.2.85)$$

□

It follows from Theorem 2.2.11 that an unbiased estimator T_1 is an MSB estimator iff T_1 is a function of a sufficient statistic T. The reverse does not hold since there may exist a sufficient statistic T even if there is no MSB estimator. Therefore, the existence condition for a sufficient statistic is less restrictive than Condition 2.2.46b is on the existence of an MSB estimator.

Theorem 2.2.11 can be seen as a generalization of the Blackwell-Rao theorem which can be found e.g. in Rao [46].

2.3 MULTIPLE OBSERVATIONS

In this section we will consider the Fisher information of order s for multiple observations. We shall denote these observations by $\xi_1, \xi_2, \dots, \xi_n$ or by ξ and η in the bivariate case. The definition of the Fisher information of order s for this case is a direct extension of Definition 2.1.4. It is easy to see that the properties of $F_s(\xi;\theta)$ which were discussed in Section 2.2 also hold for $F_s(\xi_1, \dots, \xi_n; \theta)$. The main difference is that now the space $X \subset \mathbb{R}^n$ is n-dimensional.

An interesting property of an information measure is its additivity. It is well-known that the Fisher information measure is additive. For $F_s(\xi;\theta)$ this property does no longer hold for $s \geq 1$, $s \neq 2$. In the next theorems we will study its behaviour with respect to additivity, using the Definitions 2.1.5 and 2.1.6 for the joint and the conditional measures. For simplicity we will restrict ourself to bivariate observations.

THEOREM 2.3.1

Let the information measures $F_s(\xi;\theta)$ and $F_s(\eta/\xi;\theta)$, as given in the Definitions 2.1.4 and 2.1.6, exist. Then for $s > 1$ it holds that

$$F_s(\xi,\eta;\theta) \leq 2\left[F_s(\xi;\theta)^{\frac{1}{s-1}} + F_s(\eta/\xi;\theta)^{\frac{1}{s-1}}\right]^{s-1}. \qquad (2.3.1)$$

Proof

The proof is based on the C_α-inequality, see Appendix A, which for $\alpha > 1$ is given by

$$\int |g+h|^\alpha \, p \, du \leq 2^{\alpha-1}\left[\int |g|^\alpha \, p \, du + \int |h|^\alpha \, p \, du\right]. \qquad (2.3.2)$$

If we use

$$g = \frac{\partial}{\partial\theta} \log p(x;\theta) \qquad (2.3.3a)$$

and $\quad h = \frac{\partial}{\partial\theta} \log p(y/x;\theta)$, (2.3.3b)

we find that

$$g + h = \frac{\partial}{\partial\theta} \log p(x;\theta)\, q(y/x;\theta)$$

$$= \frac{\partial}{\partial\theta} \log r(x,y;\theta). \qquad (2.3.4)$$

Substitution of Eqs. 2.3.3a - 2.3.4 into Eq. 2.3.2 yields for $\alpha = s/(s-1)$, with $s > 1$, that

$$F_s(\xi,\eta;\theta)^{\frac{1}{s-1}} \le 2^{\frac{1}{s-1}} \left[F_s(\xi;\theta)^{\frac{1}{s-1}} + F_s(\eta/\xi;\theta)^{\frac{1}{s-1}} \right], \qquad (2.3.5)$$

from which we obtain Eq. 2.3.1. □

As a consequence of Theorem 2.3.1 it follows that the existence of $F_s(\xi;\theta)$ and $F_s(\eta/\xi;\theta)$ guarantees the existence of $F_s(\xi,\eta;\theta)$. In the next theorem we obtain another upper bound on $F_s(\xi,\eta;\theta)$.

THEOREM 2.3.2

For $s \ge 1$ we have, under conditions similar to those of Theorem 2.3.1, that

$$F_s(\xi,\eta;\theta) \le \left[F_s(\xi;\theta)^{1/s} + F_s(\eta/\xi;\theta)^{1/s} \right]^s. \qquad (2.3.6a)$$

Equality in Eq. 2.3.6a holds iff

$$M(\theta) \frac{\partial}{\partial\theta} \log p(x;\theta) = K(\theta) \frac{\partial}{\partial\theta} \log q(y/x;\theta) \quad \text{a.e.} \qquad (2.3.6b)$$

Here $M(\theta)$ and $K(\theta)$ are non-negative constants, not both zero.

Proof

We start with Definition 2.1.5 of $F_s(\xi,\eta;\theta)$. Noting that

$$\frac{\partial}{\partial\theta}\log r(x,y;\theta) = \frac{\partial}{\partial\theta}\log p(x;\theta) + \frac{\partial}{\partial\theta}\log q(y/x;\theta) \qquad (2.3.7)$$

and applying the Minkowski inequality (see Appendix A) it follows for $F_s(\xi,\eta;\theta)$ that

$$\begin{aligned}
F_s(\xi,\eta;\theta)^{\frac{1}{s}} &= \left[E_\theta\left|\frac{\partial}{\partial\theta}\log r(\xi,\eta;\theta)\right|^{\frac{s}{s-1}}\right]^{\frac{s-1}{s}} \\
&= \left[E_\theta\left|\frac{\partial}{\partial\theta}\log p(\xi;\theta) + \frac{\partial}{\partial\theta}\log q(\eta/\xi;\theta)\right|^{\frac{s}{s-1}}\right]^{\frac{s-1}{s}} \\
&\le \left[E_\theta\left|\frac{\partial}{\partial\theta}\log p(\xi;\theta)\right|^{\frac{s}{s-1}}\right]^{\frac{s-1}{s}} + \\
&\quad + \left[E_\theta\left|\frac{\partial}{\partial\theta}\log q(\eta/\xi;\theta)\right|^{\frac{s}{s-1}}\right]^{\frac{s-1}{s}} \\
&= F_s(\xi;\theta)^{1/s} + F_s(\eta/\xi;\theta)^{1/s}\,, \qquad (2.3.8)
\end{aligned}$$

from which we obtain Eq. 2.3.6a. Equality is achieved in the Minkowski inequality iff Eq. 2.3.6b is satisfied. □

The upper bound of Theorem 2.3.2 is tighter than that of Theorem 2.3.1.

In a similar way we can obtain a lower bound on the joint Fisher information of order s.

THEOREM 2.3.3

For $s \geq 1$ the following lower bound is valid:

$$F_s(\xi,\eta;\theta) \geq \max\left[0, \left\{F_s(\xi;\theta)^{1/s} - F_s(\eta/\xi;\theta)^{1/s}\right\}^s\right]. \qquad (2.3.9a)$$

Equality in Eq. 2.3.9a holds iff

$$M(\theta)\frac{\partial}{\partial\theta}\log r(x,y;\theta) = -K(\theta)\frac{\partial}{\partial\theta}\log q(y/x;\theta) \quad \text{a.e.} \qquad (2.3.9b)$$

Here $M(\theta)$ and $K(\theta)$ are non-negative constants, not both zero.

Proof

From Eq. 2.3.7 we find

$$\frac{\partial}{\partial\theta}\log p(x;\theta) = \frac{\partial}{\partial\theta}\log r(x,y;\theta) - \frac{\partial}{\partial\theta}\log q(y/x;\theta)\,. \qquad (2.3.10)$$

Using the Minkowski inequality, we find

$$F_s(\xi;\theta)^{\frac{1}{s}} = \left[E_\theta\left|\frac{\partial}{\partial\theta}\log p(\xi;\theta)\right|^{\frac{s}{s-1}}\right]^{\frac{s-1}{s}}$$

$$= \left[E_\theta\left|\frac{\partial}{\partial\theta}\log r(\xi,\eta;\theta) - \frac{\partial}{\partial\theta}\log q(\eta/\xi;\theta)\right|^{\frac{s}{s-1}}\right]^{\frac{s-1}{s}}$$

$$\leq \left[E_\theta\left|\frac{\partial}{\partial\theta}\log r(\xi,\eta;\theta)\right|^{\frac{s}{s-1}}\right]^{\frac{s-1}{s}} +$$

$$+ \left[E_\theta\left|\frac{\partial}{\partial\theta}\log q(\eta/\xi;\theta)\right|^{\frac{s}{s-1}}\right]^{\frac{s-1}{s}}$$

$$= F_s(\xi,\eta;\theta)^{1/s} + F_s(\eta/\xi;\theta)^{1/s}. \qquad (2.3.11)$$

Equality in Eq. 2.3.11 holds iff Eq. 2.3.9b is satisfied. From Eq. 2.3.11 we obtain

$$F_s(\xi,\eta;\theta) \geq \left[F_s(\xi;\theta)^{1/s} - F_s(\eta/\xi;\theta)^{1/s} \right]^s . \tag{2.3.12}$$

However, the right-hand side of Eq. 2.3.12 may be negative, and since $F_s(\xi,\eta;\theta)$ is non-negative, we may easily improve Eq. 2.3.12 to obtain Eq. 2.3.9a. □

With Theorems 2.3.1 - 2.3.3 we have obtained bounds on the Fisher information of order s for the bivariate case. It will be clear from these theorems that $F_s(\xi;\theta)$ is non-additive. For s = 2 additivity can be shown using a different technique (see Stam [53]). The extension of Theorem 2.3.2 to the multivariate case is straightforward. Therefore, we shall give it without proof for the case of independent random variables.

COROLLARY 2.3.1

Let $\xi_1, \xi_2, \ldots, \xi_n$ be independent random variables and let $F_s(\xi_i;\theta)$, i = 1, 2, ... , n exist. Then for s > 1,

$$F_s(\xi_1, \ldots, \xi_n;\theta) \leq \left[\sum_{i=1}^{n} F_s(\xi_i;\theta)^{1/s} \right]^s . \tag{2.3.13}$$

An interesting special case arises if all random variables are independent and identically distributed (i.i.d.).

COROLLARY 2.3.2

Let $\xi_1, \xi_2, \ldots, \xi_n$ be i.i.d. random variables. Then

$$F_s(\xi_i;\theta) = F_s(\xi;\theta) \qquad i = 1, 2, \ldots, n \tag{2.3.14}$$

and $$F_s(\xi_1, \ldots, \xi_n;\theta) \le n^s F_s(\xi;\theta). \tag{2.3.15}$$

Another important property of an information measure is its behaviour for dependent random variables. Because of Theorem 2.2.8 we may expect that on the average an increase of our knowledge, modelled by a condition in the density function, leads to a smaller s.a.c.m. of an estimator T and thus to an increased value of the Fisher information of order s. This is stated more precisely in the next theorem.

THEOREM 2.3.4

Let the random variables ξ and η with density functions $p(x;\theta)$ and $q(y;\theta)$ have the joint density function $r(x,y;\theta) = p(x;\theta).q(y/x;\theta) = q(y;\theta).p(x/y;\theta)$ and let the information measures $F_s(\xi;\theta)$ and $F_s(\xi/\eta;\theta)$ exist. If the following conditions hold

(i) $\int r(x,y;\theta)\, dx$ can be differentiated under the integral sign (2.3.16a)

(ii) $\int p(x/y;\theta) \frac{\partial}{\partial\theta} q(y;\theta)\, dy \le 0,$ (2.3.16b)

then for s > 1 we have

$$F_s(\xi;\theta) \le F_s(\xi/\eta;\theta). \tag{2.3.17}$$

Equality in Eq. 2.3.17 holds if ξ and η are independent.

Proof

We have

$$F_s(\xi/\eta;\theta)^{\frac{1}{s-1}} = E_\theta \left| \frac{\partial}{\partial\theta} \log p(\xi/\eta;\theta) \right|^{\frac{s}{s-1}} . \tag{2.3.18}$$

Using Jensen's inequality for Eq. 2.3.18 yields

$$F_s(\xi/\eta;\theta)^{\frac{1}{s-1}} \geq E_\theta \left| E_\theta \frac{\partial}{\partial\theta} \log p(\xi/\eta;\theta) \right|^{\frac{s}{s-1}}$$

$$= E_\theta \left| \int \frac{p'(\xi/y;\theta)}{p(\xi/y;\theta)} q(y/\xi;\theta)\, dy \right|^{\frac{s}{s-1}}$$

$$= E_\theta \left| \frac{1}{p(\xi;\theta)} \int p'(\xi/y;\theta)\, q(y;\theta)\, dy \right|^{\frac{s}{s-1}} . \quad (2.3.19)$$

Using the conditions (i) and (ii), we find

$$\int \frac{\partial}{\partial\theta} \{p(x/y;\theta)\}\, q(y;\theta)\, dy$$

$$\geq \int \frac{\partial}{\partial\theta} \{p(x/y;\theta)\}\, q(y;\theta)\, dy + \int p(x/y;\theta) \frac{\partial}{\partial\theta} q(y;\theta)\, dy$$

$$= \int \frac{\partial}{\partial\theta} \{p(x/y;\theta)\, q(y;\theta)\}\, dy$$

$$= \frac{\partial}{\partial\theta} \int p(x/y;\theta)\, q(y;\theta)\, dy$$

$$= \frac{\partial}{\partial\theta} p(x;\theta). \quad (2.3.20)$$

Substitution of Eq. 2.3.20 into Eq. 2.3.19 yields

$$F_s(\xi/\eta;\theta)^{\frac{1}{s-1}} \geq E_\theta \left| \frac{p'(\xi;\theta)}{p(\xi;\theta)} \right|^{\frac{s}{s-1}}$$

$$= F_s(\xi;\theta)^{\frac{1}{s-1}} \quad (2.3.21)$$

from which Eq. 2.3.17 follows. □

As the theorem shows the Fisher information of order s, under certain conditions, increases if the random variable ξ depends on η. Since $F_s(\xi;\theta)$ is inversely related to the accuracy of a parameter estimator T, which is based on the observation ξ, this means that the fact that ξ depends on η may lead to a better accuracy of this estimator.

2.4 DISCRETE OBSERVATIONS

The results for $F_s(\xi;\theta)$ which have been obtained in the previous sections of this chapter have been formulated in terms of continuous observations ξ with density function $p(x;\theta)$, $x \in X$, $\theta \in \Theta$.

However, in practical situations the observations may be discrete random variables with probability distribution

$$\{Pr_\theta\} = \{Pr(x_i;\theta),\ \theta \in \Theta,\ x_i \in X,\ i = 1, 2, \ldots, N\} \qquad (2.4.1)$$

We assume that the values x_i do not depend on θ. As in the case of continuous observations we can define a Fisher information of order s.

DEFINITION 2.4.1

Let the probability distribution $\{Pr_\theta\}$ satisfy

(i) $\quad Pr(x_i;\theta) > 0 \qquad i = 1, 2, \ldots, N \qquad$ (2.4.2a)

(ii) $\quad \frac{\partial}{\partial\theta} Pr(x_i;\theta)$ exists for $i = 1, 2, \ldots, N$, $\theta \in \Theta \qquad$ (2.4.2b)

Then the (discrete) Fisher information of order s is defined as

$$F_s(\xi;\theta) = \left[\sum_{i=1}^{N} \left|\frac{\partial}{\partial\theta} \log Pr(x_i;\theta)\right|^{\frac{s}{s-1}} Pr(x_i;\theta)\right]^{s-1} . \qquad (2.4.3)$$

Similarly we can obtain definitions for the joint and the conditional Fisher information of order s.

It can be expected that the properties of $F_s(\xi;\theta)$ are valid irrespective of the fact that the observations are discrete or continuous. This is indeed the case. However, in the proofs of some of the theorems of Section 2.2 we have to modify slightly some conditions if we consider discrete observations. Therefore we will state two major theorems explicitly, i.e., those which are concerned with the invariance of $F_s(\xi;\theta)$ and with the s.a.c.m. of estimators of θ.

THEOREM 2.4.1

Let the family of distributions $\{Pr_\theta\} = \{Pr(x_i;\theta),\ \theta \in \Theta,\ i = 1, 2, \ldots, N\}$ of the random variable ξ, defined over the probability space $(X, \mathcal{A}, \{Pr_\theta\})$, satisfy the conditions of Eqs. 2.4.2a and 2.4.2b and let the additional condition

(iii) for every $A \in \mathcal{A}$ and $\theta \in \Theta$ the sum $\Sigma\ \partial\ Pr(x_i;\theta)/\partial\theta$ converges uniformly in Θ (2.4.4)

be satisfied.

Let $T = T(\xi)$ be a measurable transformation of ξ. Then $F_s(T(\xi);\theta)$ exists if $F_s(\xi;\theta)$ exists, and for $s \geq 1$ we have

$$F_s(T(\xi);\theta) \leq F_s(\xi;\theta) \quad . \tag{2.4.5}$$

Equality in Eq. 2.4.5 holds iff (for $s > 1$) or if (for $s = 1$) T is sufficient for the family $\{Pr_\theta\}$.

Proof

The proof is analogous to that of Theorem 2.2.5 and will not be given here in detail. The condition (iii) allows differentiation under the summation sign, which is an essential step in the proof. □

In the next theorem we consider a bound on the s.a.c.m. of estimators of $\tau(\theta)$.

THEOREM 2.4.2

Let the distribution $Pr(x_i;\theta)$, $\theta \in \Theta$ of the random variable ξ satisfy

(i) $Pr(x_i;\theta) > 0 \qquad i = 1, \dots, N, \ \theta \in \Theta$ (2.4.6a)

(ii) $\frac{\partial}{\partial\theta} Pr(x_i;\theta)$ exists for every $x_i \in X$, $\theta \in \Theta$ except possibly a finite number of points θ where the left and right derivatives exist and are finite (2.4.6b)

(iii) $\sum_{i=1}^{N} \frac{\partial}{\partial\theta} Pr(x_i;\theta)$ converges uniformly in Θ (2.4.6c)

(iv) $\sum_{i=1}^{N} \left| \frac{\partial}{\partial\theta} \log Pr(x_i;\theta) \right|^{\frac{s}{s-1}} Pr(x_i;\theta) > 0, \ \theta \in \Theta$ (2.4.6d)

(v) $\sum_{i=1}^{N} t \frac{\partial}{\partial\theta} Pr(x_i;\theta)$ converges uniformly in Θ (2.4.6e)

Let T be an unbiased estimator of $\tau(\theta)$, where $\tau(\theta)$ is a continuously differentiable function of θ. Then for $s \geq 1$ it holds that

$$E_\theta |T - \tau(\theta)|^s \geq \frac{\left| \frac{d}{d\theta} \tau(\theta) \right|^s}{F_s(\xi;\theta)} . \tag{2.4.7}$$

Proof

The proof is similar to the proof of Theorem 2.2.8. The conditions (iii) and (v) permit us to perform differentiation under the summation sign. □

In a similar way we can obtain the other properties of $F_s(\xi;\theta)$ for discrete observations. For this reason we will assume in the sequel that the observations ξ are continuous, i.e., that they have a density function $p(x;\theta)$.

2.5 EQUIVALENT INFORMATION MEASURES

It is possible to obtain information measures which are closely related to the Fisher information of order s. In this section we shall introduce three such measures, which we shall denote by $F_s'(\xi;\theta)$, $F_s''(\xi;\theta)$ and $G(\xi;\theta)$, and discuss some aspects of these measures.

DEFINITION 2.5.1

Let the density function $p(x;\theta)$ of the random variable ξ satisfy

(i) $p(x;\theta) > 0 \qquad x \in X,\ \theta \in \Theta$ (2.5.1a)

(ii) $\frac{\partial}{\partial\theta} p(x;\theta)$ exists for all $x \in X$, $\theta \in \Theta$. (2.5.1b)

Then the information measure $F_s'(\xi;\theta)$ is defined for $s \geq 1$ as

$$F_s'(\xi;\theta) = \left[\int_X \left| \frac{\partial}{\partial\theta} p(x;\theta)^{\frac{s-1}{s}} \right|^{\frac{s}{s-1}} dx\right]^{s-1}, \qquad (2.5.2)$$

provided it exists.

Before we give some properties of this information measure we show in which way $F_s'(\xi;\theta)$ is related to $F_s(\xi;\theta)$.

THEOREM 2.5.1

Let $p(x;\theta)$ satisfy the conditions (i) and (ii) and let $F_s(\xi;\theta)$ and $F_s'(\xi;\theta)$ exist. Then for $s \geq 1$

$$F_s'(\xi;\theta) = \left(\frac{s-1}{s}\right)^s F_s(\xi;\theta)\,. \qquad (2.5.3)$$

Proof

We have

$$\frac{\partial}{\partial\theta} p(x;\theta)^{\frac{s-1}{s}} = \frac{s-1}{s} p(x;\theta)^{-\frac{1}{s}} \frac{\partial}{\partial\theta} p(x;\theta)$$

$$= \frac{s-1}{s} p(x;\theta)^{\frac{s-1}{s}} \frac{\partial}{\partial\theta} \log p(x;\theta) \qquad (2.5.4)$$

from which we obtain

$$\left| \frac{\partial}{\partial\theta} p(x;\theta)^{\frac{s-1}{s}} \right|^{\frac{s}{s-1}} = \left(\frac{s-1}{s}\right)^{\frac{s}{s-1}} p(x;\theta) \left| \frac{\partial}{\partial\theta} \log p(x;\theta) \right|^{\frac{s}{s-1}}. \qquad (2.5.5)$$

Integration of Eq. 2.5.5 with respect to x yields

$$\int_X \left| \frac{\partial}{\partial\theta} p(x;\theta)^{\frac{s-1}{s}} \right|^{\frac{s}{s-1}} dx = \left(\frac{s-1}{s}\right)^{\frac{s}{s-1}} F_s(\xi;\theta)^{\frac{1}{s-1}} \qquad (2.5.6)$$

from which Eq. 2.5.3 is obtained. □

For s = 2 the relation obtained in Eq. 2.5.3 has been given by Ibragimov et al. [26]. It is interesting to observe that $F_s'(\xi;\theta)$ is not an expectation like $F_s(\xi;\theta)$. As a result of Theorem 2.5.1 it is possible to obtain properties for $F_s'(\xi;\theta)$ which are similar to those of $F_s(\xi;\theta)$ which have been obtained in the previous chapters.

The information measure $F_s'(\xi;\theta)$ enables us to give a geometric interpretation of the Fisher information of order s, as will be discussed in Section 4.1. Note that the regularity conditions which are required for $p(x;\theta)$ are the same as those for $F_s'(\xi;\theta)$ and $F_s(\xi;\theta)$.

In the next definition a second information measure $F_s''(\xi;\theta)$, which is related to $F_s(\xi;\theta)$, is given in terms of higher order derivatives of $p(x;\theta)$. For this reason $F_s''(\xi;\theta)$ is slightly less general than $F_s(\xi;\theta)$.

DEFINITION 2.5.2

Let the density function $p(x;\theta)$ of the random variable ξ satisfy the conditions

(i) $p(x;\theta) > 0 \qquad x \in X,\ \theta \in \Theta$ (2.5.7a)

(ii) $\frac{\partial^i}{\partial\theta^i} \log p(x;\theta)$ exist for all $i = 1, 2, \ldots, s$ and

all $x \in X,\ \theta \in \Theta$ (2.5.7b)

Then for $s \in \mathbb{N}$ the information measure $F_s''(\xi;\theta)$ is defined as

$$F_s''(\xi;\theta) = \left[\int_X \left| \frac{\partial^s}{\partial\theta^s} \log p(x;\theta) \right|^{\frac{1}{s-1}} p(x;\theta)\, dx \right]^{s-1} \tag{2.5.8}$$

provided it exists.

The relation between $F_s''(\xi;\theta)$ and $F_s(\xi;\theta)$ is given in the next theorem.

THEOREM 2.5.2

Let $p(x;\theta)$ satisfy the conditions (i) and (ii) and an additional condition

(iii) $\int p(x;\theta)$ can be differentiated under the integral sign (2.5.9)

and let $F_s''(\xi;\theta)$ and $F_s(\xi;\theta)$ exist. Then for $s \in \mathbb{N}$

$$F_s''(\xi;\theta) = \Gamma(s)\, F_s(\xi;\theta). \tag{2.5.10}$$

Proof

This follows from a mathematical induction. Assume that for $s = 1, 2, 3 \ldots$

$$\frac{\partial^s}{\partial\theta^s} \log p(x;\theta) = (-1)^{s-1}\, \Gamma(s) \left[\frac{\partial}{\partial\theta} \log p(x;\theta)\right]^s \quad \text{a.e. } P_\theta \quad (2.5.11)$$

holds. Then

$$\frac{\partial}{\partial\theta}\left[\frac{\partial^{s-1}}{\partial\theta^{s-1}} \log p(x;\theta)\right] = \frac{\partial}{\partial\theta}\left[(-1)^{s-2}\, \Gamma(s-1)\left[\frac{\partial}{\partial\theta} \log p(x;\theta)\right]^{s-1}\right]$$

$$= (-1)^{s-2}\, (s-1)\, \Gamma(s-1) \left[\frac{\partial}{\partial\theta} \log p(x;\theta)\right]^{s-2} \frac{\partial^2}{\partial\theta^2} \log p(x;\theta)$$

$$\text{a.e. } P_\theta. \qquad (2.5.12)$$

Using condition (iii) we have

$$0 = \int_X \frac{\partial}{\partial\theta} p(x;\theta)\, dx$$

$$= \int_X \frac{\partial}{\partial\theta} \log p(x;\theta)\, p(x;\theta)\, dx. \qquad (2.5.13)$$

Differentiating Eq. 2.5.13 yields

$$0 = \frac{\partial}{\partial\theta}\int_X \frac{\partial}{\partial\theta} \log p(x;\theta)\, p(x;\theta)\, dx$$

$$= \int_X \left[\frac{\partial^2}{\partial\theta^2} \log p(x;\theta)\right] p(x;\theta)\, dx + \int_X \left[\frac{\partial}{\partial\theta} \log p(x;\theta)\right]^2 p(x;\theta)\, dx, \qquad (2.5.14)$$

from which we conclude that

$$\frac{\partial^2}{\partial\theta^2} \log p(x;\theta) = - \left[\frac{\partial}{\partial\theta} \log p(x;\theta)\right]^2 \quad \text{a.e. } P_\theta. \tag{2.5.15}$$

Substitution of Eq. 2.5.15 into Eq. 2.5.12 and noting that

$$(s-1)\ \Gamma(s-1) = \Gamma(s) \tag{2.5.16}$$

proves the assertion in Eq. 2.5.11. From Eq. 2.5.11 we obtain

$$\left| \frac{\partial^s}{\partial\theta^s} \log p(x;\theta) \right|^{\frac{1}{s-1}} = \Gamma(s)^{\frac{1}{s-1}} \left| \frac{\partial}{\partial\theta} \log p(x;\theta) \right|^{\frac{s}{s-1}} \quad \text{a.e. } P_\theta. \tag{2.5.17}$$

Taking expectations on both sides with respect to x and using the definition of $F_s''(\xi;\theta)$ leads to Eq. 2.5.10, which completes the proof.

□

The case s = 2 is well known. For s = 3, 4, ... we obtain an information measure which shows some resemblance to a measure which has been mentioned by Kobayashi [33] [34], who discussed distance measures and asymptotic relative efficiency. He found a quantity, given by

$$E_\theta \left[\frac{\partial^m}{\partial\theta^m} \log p(\xi;\theta) \right]^2, \tag{2.5.18}$$

and noted that this could be considered to be a generalization of the Fisher information measure. For m = 1 Eq. 2.5.18 is the usual Fisher information. With the information measure $F_s''(\xi;\theta)$ and Kobayashi's measure, it is easy to define the following measure.

DEFINITION 2.5.3

Let the density function $p(x;\theta)$ of the random variable ξ satisfy the conditions:

(i) $p(x;\theta) > 0 \qquad x \in X,\ \theta \in \Theta$ (2.5.19a)

(ii) $\dfrac{\partial^i}{\partial\theta^i} \log p(x;\theta)$ exists for all $i = 1, 2, \ldots, s$

and all $x \in X$, $\theta \in \Theta$. (2.5.19b)

Then the information measure $G(\xi;\theta)$ is defined for $s \in \mathbb{N}$, $m \in \mathbb{N}$ and $r \geq 0$ by

$$G(\xi;\theta) = \left[E_\theta \left| \frac{\partial^m}{\partial\theta^m} \log p(\xi;\theta) \right|^r \right]^{s-1} \tag{2.5.20}$$

provided it exists.

Note that for $r = 2$, $s = 2$ the measure $G(\xi;\theta)$ coincides with Kobayashi's measure. If we set $m = s$ and $r = (s-1)^{-1}$ in Eq. 2.5.20, we obtain $G(\xi;\theta) = F_s''(\xi;\theta) = \Gamma(s)\, F_s(\xi;\theta)$. The general case for $r \geq 0$ and $m \neq s$ will not be considered in detail. We only note that because of Eq. 2.5.10 the measure $G(\xi;\theta)$ can also be written as

$$G(\xi;\theta) = \Gamma(m)^{r(s-1)} \left[E_\theta \left| \frac{\partial}{\partial\theta} \log p(\xi;\theta) \right|^{mr} \right]^{s-1} \tag{2.5.21}$$

in terms of the first derivative.

2.6 RANDOM PARAMETERS

Sometimes the parameter θ is itself a random variable, as has been mentioned in Section 1.2. We shall denote this random variable by ϕ and assume that it has a density function $\pi(\theta)$, $\theta \in \Theta$. It turns out that there are two different ways to define an information measure of order s. We shall denote them by $\bar{F}_s(\xi;\phi)$ and $F_s(\xi,\phi)$ and discuss some of their properties. First we will introduce the information measure $\bar{F}_s(\xi;\phi)$.

DEFINITION 2.6.1

Let $q(x,\theta) = \pi(\theta).p(x;\theta)$ be the joint density function defined on $X \times \Theta$, and let the density function $p(x;\theta)$ satisfy the conditions

(i) $p(x;\theta) > 0 \qquad (x,\theta) \in X \times \Theta,$ (2.6.1a)

(ii) $\frac{\partial}{\partial\theta} p(x;\theta)$ exists for every $(x,\theta) \in X \times \Theta$. (2.6.1b)

Then the mean Fisher information of order s can be defined as

$$\bar{F}_s(\xi;\phi) = \left[\int_\Theta \int_X \left| \frac{\partial}{\partial\theta} \log p(x;\theta) \right|^{\frac{s}{s-1}} q(x,\theta)\, dx\, d\theta \right]^{s-1}$$

$$= \left[E_\phi \left[F_s(\xi;\theta)^{\frac{1}{s-1}} \right] \right]^{s-1} . \qquad (2.6.2)$$

The quantity $\bar{F}_s(\xi;\phi)$ can be considered to be the mean information contained in ξ about the random parameter ϕ. The case $s = 2$ has been discussed by Gart [22]. In general $\bar{F}_s(\xi;\phi)$ is not the expectation of $F_s(\xi;\theta)$. In fact we have the following relation between $F_s(\xi;\theta)$ and $\bar{F}_s(\xi;\phi)$.

THEOREM 2.6.1

Let $\bar{F}_s(\xi;\phi)$ and $F_s(\xi;\theta)$ exist. Then it holds that

$$\bar{F}_s(\xi;\phi) \geq E_\phi\{F_s(\xi;\theta)\} \qquad 1 \leq s < 2$$

$$= E_\phi\{F_s(\xi;\theta)\} \qquad s = 2$$

$$\leq E_\phi\{F_s(\xi;\theta)\} \qquad s > 2 \qquad (2.6.3)$$

Proof

The function $|u|^{s-1}$ is convex in u for s > 2 and concave for $1 \leq s \leq 2$. Setting

$$|u| = E_{\phi}\left[F_s(\xi;\theta)^{\frac{1}{s-1}}\right] \tag{2.6.4}$$

we find, when applying Jensen's inequality to Eq. 2.6.2, that for s > 2

$$\begin{aligned}\overline{F}_s(\xi;\phi) &= \left[E_{\phi}\left[F_s(\xi;\theta)^{\frac{1}{s-1}}\right]\right]^{s-1} \\ &\leq E_{\phi}\{F_s(\xi;\theta)\}.\end{aligned} \tag{2.6.5}$$

This proves the third part of Eq. 2.6.3. The proofs of the first two parts follow in a similar way. □

Based on Definition 2.6.1 we can derive a bound on the s.a.c.m. of an estimator T of $\tau(\phi)$, where $\tau(\theta)$ is a continuously differentiable function of θ. The precise form is given in the next theorem.

THEOREM 2.6.2

Let T be an unbiased estimator of the continuously differentiable function $\tau(\phi)$ and let $p(x;\theta)$ satisfy the conditions (i) and (ii) and the additional conditions:

(iii) $\frac{\partial}{\partial\theta} p(x;\theta)$ is absolutely integrable with respect to x and θ, (2.6.6a)

(iv) $$\int_{\Theta}\int_{X} \left| \frac{\partial}{\partial\theta} \log p(x;\theta) \right|^{\frac{s}{s-1}} q(x,\theta)\, dx\, d\theta > 0, \tag{2.6.6b}$$

(v) $t \frac{\partial}{\partial\theta} p(x;\theta)$ is absolutely integrable with respect to x and θ. (2.6.6c)

Then $$E_\phi\left|T - \tau(\phi)\right|^s \geq \frac{\left|E_\phi\{\tau'(\phi)\}\right|^s}{\overline{F}_s(\xi;\phi)} . \tag{2.6.7a}$$

Equality in Eq. 2.6.7a holds iff

$$M\{t - \tau(\theta)\} = K \left| \frac{\partial}{\partial\theta} \log p(x;\theta) \right|^{\frac{1}{s-1}} \operatorname{sgn}\left[\frac{\partial}{\partial\theta} \log p(x;\theta)\right] \text{ a.e. } Q. \tag{2.6.7b}$$

Here M and K are non-negative constants, not both zero, and

$$\int_\Theta \int_X \left| \frac{\partial}{\partial\theta} \log p(x;\theta) \right|^{\frac{1}{s-1}} \operatorname{sgn}\left[\frac{\partial}{\partial\theta} \log p(x;\theta)\right] q(x,\theta)\, dx\, d\theta = 0. \tag{2.6.8}$$

Proof

We have

$$\int_X p(x;\theta)\, dx = 1 \tag{2.6.9}$$

and $$\int_X t\, p(x;\theta)\, dx = \tau(\theta). \tag{2.6.10}$$

Multiplying Eq. 2.6.9 by $\tau(\theta)$ and subtracting this result from Eq. 2.6.10 yields

$$\int_X \{t - \tau(\theta)\}\, p(x;\theta)\, dx = 0 . \tag{2.6.11}$$

Differentiating Eq. 2.6.11 with respect to θ and using conditions (iii) and (v), we obtain

$$\int_X \{t - \tau(\theta)\} \frac{\partial}{\partial\theta} \log p(x;\theta)\ p(x;\theta)\ dx - \tau'(\theta) \int_X p(x;\theta)\ dx = 0. \tag{2.6.12}$$

Taking the expectation in Eq. 2.6.12 with respect to θ yields

$$\int_\Theta \int_X \{t - \tau(\theta)\} \frac{\partial}{\partial\theta} \log p(x;\theta)\ q(x,\theta)\ dx\ d\theta - \int_\Theta \tau'(\theta)\ \pi(\theta)\ d\theta = 0, \tag{2.6.13}$$

or $$\int_\Theta \int_X \{t - \tau(\theta)\} \frac{\partial}{\partial\theta} \log p(x;\theta)\ q(x,\theta)\ dx\ d\theta = E_\phi\{\tau'(\phi)\}. \tag{2.6.14}$$

Applying Hölder's inequality to Eq. 2.6.14 yields

$$\left[\int_\Theta \int_X |t - \tau(\theta)|^s\ q(x,\theta)\ dx\ d\theta\right]^{\frac{1}{s}} \cdot$$

$$\cdot \left[\int_\Theta \int_X \left|\frac{\partial}{\partial\theta} \log p(x;\theta)\right|^{\frac{s}{s-1}} q(x,\theta)\ dx\ d\theta\right]^{\frac{s-1}{s}} \geq E_\phi\{\tau'(\phi)\}, \tag{2.6.15}$$

from which Eq. 2.6.7a is obtained by using Definition 2.6.1. The equality condition is obtained in a manner similar to that in Theorem 2.2.8. Note that now the constants M and K do not depend on θ. □

COROLLARY 2.6.1

Let T be an unbiased estimator of ϕ and let $p(x;\theta)$ satisfy conditions (i) - (v) of Theorem 2.6.2. Then

$$E_\phi|T - \phi|^s \geq \frac{1}{\overline{F}_s(\xi;\phi)}. \tag{2.6.16}$$

Equality in Eq. 2.6.16 holds iff Eq. 2.6.7b holds for $\tau(\theta) = \theta$.

Proof

The proof follows from Eq. 2.6.7a by substitution of $\tau(\phi) = \phi$. □

In the next theorems we shall consider the equality condition 2.6.7b in Theorem 2.6.2 in more detail.

THEOREM 2.6.3

If T is an unbiased MSB estimator of ϕ then equality in Eq. 2.6.16 holds iff

$$P\{F_s(\xi;\theta) = c\} = 1 \tag{2.6.17}$$

where c is a real constant.

Proof

Since T is an MSB estimator we know that

$$E_\theta|T - \theta|^s = F_s^{-1}(\xi;\theta). \tag{2.6.18}$$

Taking expectations yields

$$E_\phi|T - \phi|^s = E_\phi\{F_s^{-1}(\xi;\theta)\}. \tag{2.6.19}$$

On the other hand, equality in Eq. 2.6.16 is obtained iff

$$E_\phi|T - \phi|^s = \bar{F}_s^{-1}(\xi;\phi)$$

$$= \left[E_\phi\left[F_s(\xi;\theta)^{\frac{1}{s-1}}\right]\right]^{1-s}. \tag{2.6.20}$$

If Eq. 2.6.19 and 2.6.20 are combined we only have to prove that, under the condition of Eq. 2.6.17,

$$E_\phi\{F_s^{-1}(\xi;\theta)\} = \left[E_\phi\left[F_s(\xi;\theta)^{\frac{1}{s-1}}\right]\right]^{1-s}, \tag{2.6.21}$$

or
$$\left[E_\phi\{F_s^{-1}(\xi;\theta)\}\right]^{\frac{1}{s}} \cdot \left[E_\phi\left[F_s(\xi;\theta)^{\frac{1}{s-1}}\right]\right]^{\frac{s-1}{s}} = 1. \tag{2.6.22}$$

Using Hölder's inequality

$$E|g(\theta)\ h(\theta)| \leq [E|g(\theta)|^\alpha]^{1/\alpha}\ [E|h(\theta)|^\beta]^{1/\beta} \tag{2.6.23}$$

with
$$g(\theta) = \{F_s(\xi;\theta)\}^{-\ 1/s}, \tag{2.6.24}$$

$$h(\theta) = \{F_s(\xi;\theta)\}^{1/s}, \tag{2.6.25}$$

$$\alpha = s, \qquad \beta = \frac{s}{s-1}, \tag{2.6.26}$$

we obtain from Eq. 2.6.22

$$E_\phi\{1\} = 1 \leq \left[E_\phi\{F_s^{-1}(\xi;\theta)\}\right]^{\frac{1}{s}} \cdot \left[E_\phi\left[F_s(\xi;\theta)^{\frac{1}{s-1}}\right]\right]^{\frac{s-1}{s}}. \tag{2.6.27}$$

Equality in Eq. 2.6.27 holds iff

$$M\{F_s(\xi;\theta)\}^{-\frac{1}{s}} = K\{F_s(\xi;\theta)\}^{\frac{1}{s}} \qquad \text{a.e. } \Theta, \tag{2.6.28}$$

or iff

$$P\{F_s(\xi;\theta) = c\} = 1. \tag{2.6.29}$$

where we have set $c = (\frac{K}{M})^{-\frac{2}{s}}$. (2.6.30)

But if Eq. 2.6.29 holds Eq. 2.6.22 holds too, which proves the theorem. □

THEOREM 2.6.4

Equality in Eq. 2.6.16 is achieved iff T is an unbiased MSB estimator of ϕ with an exponential power distribution $N_s(\theta, m_s)$, where the s-th absolute central moment m_s is equal to $\bar{F}_s^{-1}(\xi;\phi)$.

Proof

It follows from Theorem 2.6.2 and consequently from Corollary 2.6.1 that equality in Eq. 2.6.16 is attained iff

$$M\{t - \theta\} = K \left| \frac{\partial}{\partial\theta} \log p(x;\theta) \right|^{\frac{1}{s-1}} \operatorname{sgn}\left[\frac{\partial}{\partial\theta} \log p(x;\theta)\right] \quad \text{a.e. } Q. \tag{2.6.31}$$

Eq. 2.6.31 can also be written as

$$\frac{\partial}{\partial\theta} \log p(x;\theta) = \left(\frac{M}{K}\right)^{s-1} |t - \theta|^{s-1}. \tag{2.6.32}$$

Setting

$$\left(\frac{M}{K}\right) = a, \tag{2.6.33}$$

and noting that

$$\frac{\partial}{\partial\theta} |t - \theta|^s = - s\, |t - \theta|^{s-1}, \tag{2.6.34}$$

we find for Eq. 2.6.32 that

$$\frac{\partial}{\partial\theta} \log p(x;\theta) = - \frac{a^{s-1}}{s} \frac{\partial}{\partial\theta} |t - \theta|^s. \tag{2.6.35}$$

Integration with respect to θ yields

$$\log p(x;\theta) = - \frac{a^{s-1}}{s} |t - \theta|^s + \log h(x) \tag{2.6.36}$$

or $$p(x) = h(x) \exp\left[-\frac{a^{s-1}}{s} |t - \theta|^s \right]. \tag{2.6.37}$$

Integration with respect to x gives

$$\int_X h(x) \exp\left[-\frac{a^{s-1}}{s} |t - \theta|^s \right] dx = 1, \tag{2.6.38}$$

which after a change of variables leads to

$$\int_{\hat{\Theta}} g(t) \exp\left[-\frac{a^{s-1}}{s} |t - \theta|^s \right] dt = 1. \tag{2.6.39}$$

Here $\hat{\Theta}$ is the domain of T. We set

$$g(t) = \frac{s^{\frac{s-1}{s}}}{2\,\Gamma(1/s)\, m_s^{1/s}}, \tag{2.6.40}$$

and note that from Theorem 2.2.9 it follows for an unbiased MSB estimator that

$$m_s = a^{1-s} \tag{2.6.41}$$

Therefore Eq. 2.6.31 is equivalent to

$$\int_{\hat{\Theta}} \frac{s^{\frac{s-1}{s}}}{2\,\Gamma(1/s)\, m_s^{1/s}} \exp\left[-\frac{|t - \theta|^s}{s m_s} \right] dt = 1, \tag{2.6.42}$$

which proves the theorem.

Note that the proof of this theorem does not depend on $\pi(\theta)$. Therefore the equality in Eq. 2.6.16 holds regardless of the form of $\pi(\theta)$. □

As has been mentioned in the introduction of this section it is possible to define a Fisher information measure for a random parameter ϕ in a different way. The case s = 2 can be found, e.g., in Van Trees [56] and Young [60]. This information measure is defined as follows.

DEFINITION 2.6.2

Let the density function $q(x,\theta) = \pi(\theta).p(x;\theta)$ satisfy the conditions:

(i) $q(x,\theta) > 0 \qquad (x,\theta) \in X \times \Theta,$ (2.6.43a)

(ii) $\frac{\partial}{\partial\theta} p(x;\theta)$ exists for all $(x,\theta) \in X \times \Theta$. (2.6.43b)

Then the information measure $F_s(\xi,\phi)$ can be defined as

$$F_s(\xi,\phi) = \left[\int_\Theta \int_X \left| \frac{\partial}{\partial\theta} \log q(x,\theta) \right|^{\frac{s}{s-1}} q(x,\theta)\, dx\, d\theta \right]^{s-1} . \quad (2.6.44)$$

Note that this definition is essentially a joint information measure of ξ and ϕ, as follows from Definition 2.1.5. As a consequence of this point of view the relation between $\overline{F}_s(\xi;\phi)$ and $F_s(\xi,\phi)$ is obtained easily.

THEOREM 2.6.5

Let $p(x;\theta)$, $\pi(\theta)$ and $q(x,\theta)$ be continuously differentiable and positive density functions. Let $F_s(\theta)$ and $F_s(\xi;\theta)$ (defined as in Definition 2.1.5) exist. Then for $s \geq 1$ we have the following inequality

$$F_s(\xi,\phi) \leq \left[F_s(\theta)^{1/s} + \overline{F}_s(\xi;\phi)^{1/s} \right]^s . \quad (2.6.45)$$

Proof

First we note that $F_s(\xi,\phi)$ can be written as

$$F_s(\xi,\phi) = \left[\int_\Theta \int_X \left| \frac{\partial}{\partial\theta} \log \pi(\theta) + \frac{\partial}{\partial\theta} \log p(x;\theta) \right|^{\frac{s}{s-1}} q(x,\theta)\, dx\, d\theta \right]^{s-1} .$$

(2.6.46)

Then we find Eq. 2.6.45 as a consequence of Theorem 2.3.2. □

Thus we see that for $s \geq 1$ the quantity $F_s(\xi,\phi)$ is bounded by a quantity $F_s(\theta)$, which depends on the density function $\pi(\theta)$ only, and by the mean information $\overline{F}_s(\xi;\phi)$, which was introduced in the first part of this section.

With this definition we can also obtain a bound on the s.a.c.m. of estimators.

THEOREM 2.6.6

Let T be an unbiased estimator of the continuously differentiable function $\tau(\phi)$ and let $q(x,\theta) = \pi(\theta).p(x;\theta)$ satisfy the conditions (i) and (ii) and the additional conditions:

(iii) $\frac{\partial}{\partial\theta} q(x,\theta)$ is absolutely integrable with respect to x and θ, (2.6.47a)

(iv) $$\int_\Theta \int_X \left| \frac{\partial}{\partial\theta} \log q(x,\theta) \right|^{\frac{s}{s-1}} q(x,\theta)\, dx\, d\theta > 0, \qquad (2.6.47b)$$

(v) $t \frac{\partial}{\partial\theta} q(x,\theta)$ is absolutely integrable with respect to x and θ. (2.6.47c)

Then $$E_\phi |T - \tau(\phi)|^s \geq \frac{|E_\phi\{\tau'(\phi)\}|^s}{F_s(\xi,\phi)} . \qquad (2.6.48a)$$

Equality in Eq. 2.6.48a holds iff

$$M\{t - \tau(\theta)\} = K \left| \frac{\partial}{\partial\theta} \log p(\theta;x) \right|^{\frac{1}{s-1}} \operatorname{sgn}\left[\frac{\partial}{\partial\theta} \log p(\theta;x)\right] \quad \text{a.e. } Q.$$

(2.6.48b)

Here M and K are non-negative constants, not both zero, and

$$\int_\Theta \int_X \left| \frac{\partial}{\partial\theta} \log p(\theta;x) \right|^{\frac{1}{s-1}} \operatorname{sgn}\left[\frac{\partial}{\partial\theta} \log p(\theta;x)\right] q(x,\theta)\, dx\, d\theta = 0.$$

(2.6.49)

Proof

The proof is analogous to the proof of Theorem 2.6.2. If we multiply Eq. 2.6.11 by $\pi(\theta)$ and carry out the proof in a manner similar to that in Theorem 2.6.2 we easily obtain Eq. 2.6.48a. □

If T is a biased estimator with an unknown bias function, it is possible under certain conditions to obtain a bound on the s.a.c.m. of T which does not depend on the bias function.

THEOREM 2.6.7

Let T be a biased estimator such that

$$E_\theta\{T\} = \tau(\theta) + b(\theta) \tag{2.6.50}$$

for all $\theta \in \Theta$. And let the bias function satisfy

$$\lim_{|\theta|\to\infty} b(\theta)\, \pi(\theta) = 0. \tag{2.6.51}$$

Then, if $q(x,\theta)$ satisfies conditions (i) - (v), it holds that

$$E_\phi |T - \tau(\phi)|^s \geq \frac{|E_\phi\{\tau'(\phi)\}|^s}{F_s(\xi,\phi)} \tag{2.6.52}$$

Proof

We have

$$\int_X p(x;\theta)\, dx = 1 \tag{2.6.53}$$

and $$\int_X t\, p(x;\theta)\, dx = \tau(\theta) + b(\theta). \tag{2.6.54}$$

Multiplying Eq. 2.6.53 by $\tau(\theta)\ \pi(\theta)$ and Eq. 2.6.54 by $\pi(\theta)$ and subtracting these results we obtain

$$\int_X \{t - \tau(\theta)\}\ q(x,\theta)\ dx = b(\theta)\ \pi(\theta). \tag{2.6.55}$$

Differentiating with respect to θ and using conditions (iii) and (v) yields

$$-\ \tau'(\theta) + E_\theta\left[\{T - \tau(\theta)\}\ \frac{\partial}{\partial\theta} \log q(x,\theta)\right] = \frac{d}{d\theta}\, b(\theta)\ \pi(\theta). \tag{2.6.56}$$

Integrating with respect to θ and using Eq. 2.6.51 yields

$$\int_\Theta \int_X \{t - \tau(\theta)\}\ \frac{\partial}{\partial\theta} \log q(x,\theta)\ q(x,\theta)\ dx\ d\theta = E_\phi\{\tau'(\phi)\} \tag{2.6.57}$$

from which Eq. 2.6.52 is obtained in a manner similar to that used in Theorem 2.6.2. □

As is clear from the previous theorems, both quantities $\overline{F}_s(\xi;\phi)$ and $F_s(\xi,\phi)$ provide us with bounds on the s.a.c.m. of estimators. It is easy to see that in this respect $\overline{F}_s(\xi;\phi)$ is more useful than $F_s(\xi,\phi)$.

This is a consequence of Theorem 2.6.5. Noting that $F_s(\theta) \geq 0$ gives

$$\overline{F}_s(\xi;\phi) \leq F_s(\xi,\phi) \tag{2.6.58}$$

and $$\overline{F}_s^{-1}(\xi;\phi) \geq F_s^{-1}(\xi,\phi), \tag{2.6.59}$$

it follows that $\overline{F}_s(\xi;\phi)$ gives a tighter bound on the s.a.c.m. of estimators.

A useful property of $F_s(\xi,\phi)$ is that the bound on the s.a.c.m. does not depend on the bias function $b(\theta)$, provided a certain condition is satisfied, as is shown in Theorem 2.6.7. Thus, it is not necessary to have a detailed knowledge of $b(\theta)$.

CHAPTER 3

LOCATION AND SCALE PARAMETERS

In this chapter we shall apply some results, obtained in Chapter 2, to a particular class of parameters.

First we shall consider location parameters, which were defined in Section 1.2. We show that for these parameters there exists a class of density functions, called the exponential power distribution, for which the Fisher information is minimal. We also study the stability of this characterization and develop a related measure.

Secondly we shall consider scale parameters. Then we study location-scale parameters as a special case of a two-dimensional vector parameter.

Based on the notion of entropy power we develop a generalization, called the entropy moment of order s, and discuss some properties of this entropy moment of order s. We show that there exists an interesting analogy between the inverse of the Fisher information of order s and this entropy moment of order s.

3.1 LOCATION PARAMETERS

An important special case arises if θ is a location parameter. A definition has been given in Section 1.2. We will show that $F_s(\xi;\theta)$

possesses some interesting properties in this case. Some of these properties are extensions of similar properties for the case $s = 2$ obtained by Stam [53]. First, we will give an expression for $F_s(\xi;\theta)$, for the case that θ is a location parameter of the random variable ξ.

DEFINITION 3.1.1

Let the density function $q(y)$ of the random variable η satisfy the conditions

(i) $q(y) > 0 \qquad y \in Y$ (3.1.1a)

(ii) $\frac{d}{dy} q(y)$ exists for every $y \in Y$ except possibly a finite number of points where the left and right derivatives exist and are finite. (3.1.1b)

Then for $s \geq 1$ the quantity $I_s(\eta)$ is defined by

$$I_s(\eta) = \left[\int_Y \left| \frac{\frac{d}{dy} q(y)}{q(y)} \right|^{\frac{s}{s-1}} q(y)\, dy \right]^{s-1} . \tag{3.1.2}$$

The quantity $I_s(\eta)$ is related to $F_s(\xi;\theta)$ if θ is a location parameter.

THEOREM 3.1.1

Let $\theta \in \Theta$ be a location parameter for the density function of ξ. Then it holds for $\eta = \xi - \theta$ that

$$I_s(\eta) = F_s(\xi;\theta) . \tag{3.1.3}$$

Proof

By Definition 1.2.1 we have

$$p(x;\theta) = q(x-\theta), \tag{3.1.4}$$

from which it follows that

$$\frac{\partial}{\partial\theta} \log p(x;\theta) = \frac{\partial}{\partial\theta} \log q(x-\theta)$$

$$= - \frac{\frac{\partial}{\partial\theta} q(x-\theta)}{q(x-\theta)} . \qquad (3.1.5)$$

Substitution of this result into Eq. 2.1.9d yields

$$F_s(\xi;\theta) = \left[\int_X \left| - \frac{\frac{\partial}{\partial\theta} q(x-\theta)}{q(x-\theta)} \right|^{\frac{s}{s-1}} q(x-\theta)\, dx \right]^{s-1}$$

$$= \left[\int_Y \left| \frac{\frac{d}{dy} q(y)}{q(y)} \right|^{\frac{s}{s-1}} q(y)\, dy \right]^{s-1} \qquad (3.1.6)$$

which proves Eq. 3.1.3. □

From this theorem we find an interesting property, if we take into account Definition 3.1.1.

COROLLARY 3.1.1

If $\theta \in \Theta$ is a location parameter for the density function $p(x;\theta)$ of ξ, we have

$$I_s(\eta) = F_s(\xi;0) = F_s(\xi;\theta). \qquad (3.1.7)$$

Proof

Since $I_s(\eta) = F_s(\xi;\theta)$ for every $\theta \in \Theta$ the equality also holds for $\theta = 0$, provided $0 \in \Theta$. □

It thus follows that the Fisher information of order s does not depend on θ if θ is a location parameter. Therefore, the information in ξ concerning θ is independent of the value of θ.

An interesting property of I_s is that there exists a family of distributions, called the gaussian family, or exponential power distribution, for which I_s is minimal under a certain constraint.

THEOREM 3.1.2

Let the density function p(x) satisfy the conditions

(i) $p(x) > 0 \qquad x \in X,$ (3.1.8a)

(ii) $p'(x) = \dfrac{dp(x)}{dx}$ exists for every $x \in X$ except possibly at a finite number of points where the left and right derivatives exist and are finite, (3.1.8b)

(iii) $$\int_{-\infty}^{\infty} |x|^s \, p(x) \, dx < \infty, \qquad s > 1 \tag{3.1.8c}$$

(iv) $x\, p(x) \to 0$ if $|x| \to \infty$. (3.1.8d)

In the class of density functions which satisfy conditions (i) - (iv) and which have a given s-th absolute moment m_s, the exponential power distribution $N_s(0,\beta)$ with

$$p(x) = \frac{s^{\frac{s-1}{s}}}{2\,\Gamma(1/s)\,\beta^{1/s}} \cdot \exp\left\{-\frac{|x|^s}{s\beta}\right\} \qquad s > 1, \tag{3.1.9}$$

attains minimal information $I_s = 1/\beta$.

Proof

Integration by parts gives for $0 < K < \infty$

$$\int_{-K}^{K} x\, p'(x)\, dx = x\, p(x)\Big|_{-K}^{K} - \int_{-K}^{K} p(x)\, dx. \tag{3.1.10}$$

Because of condition (iv) we obtain for $K \to \infty$

$$\int_{-\infty}^{\infty} x\, p'(x)\, dx = -1 \tag{3.1.11}$$

or

$$\int_{-\infty}^{\infty} x \frac{d}{dx} \log p(x)\, p(x)\, dx = -1. \tag{3.1.12}$$

If we next apply Hölder's inequality to Eq. 3.1.12, we find

$$\left[\int_{-\infty}^{\infty} |x|^{s}\, p(x)\, dx \right]^{\frac{1}{s}} \cdot \left[\int_{-\infty}^{\infty} \left| \frac{d}{dx} \log p(x) \right|^{\frac{s}{s-1}} p(x)\, dx \right]^{\frac{s-1}{s}} \geq 1. \tag{3.1.13}$$

Using Eq. 3.1.2 we then have

$$\int_{-\infty}^{\infty} |x|^{s}\, p(x)\, dx \,.\, I_{s} \geq 1. \tag{3.1.14}$$

Equality in Eq. 3.1.14 holds iff

$$\operatorname{sgn}[x] = - \operatorname{sgn}[p'(x)] \tag{3.1.15a}$$

and

$$c\,|x| = \left| \frac{d}{dx} \log p(x) \right|^{\frac{1}{s-1}} \quad \text{a.e. } P, \tag{3.1.15b}$$

where c is a positive constant. Hence

$$\frac{d}{dx} \log p(x) = - c^{s-1}\, |x|^{s-1} \operatorname{sgn}[x]. \tag{3.1.16}$$

The general solution of this equation is given by

$$p(x) = M \exp\left\{ - \frac{|x|^{s}}{s\, c^{1-s}} \right\}. \tag{3.1.17}$$

Since p(x) is a density function, the constant M has to be chosen as

$$M = \frac{s^{\frac{s-1}{s}}}{2\,\Gamma(1/s)\,\beta^{1/s}} \tag{3.1.18}$$

where $c^{1-s} = \beta$, (3.1.19)

in order to have $\int p(x)\,dx = 1$.

Thus, we obtain the expression for p(x) in Eq. 3.1.9. From Eq. 3.1.14 it follows that for the exponential power distribution we have $I_s = 1/\beta$. This proves the theorem. □

The parameter β is exactly the s-th absolute moment m_s of the distribution. A discussion of the exponential power distribution will be given in Section 4.3. For s = 2 this distribution is equivalent with the gaussian distribution with $\beta = \sigma^2$.

Note that the exponential power distribution is a unique solution for $s \geq 2$ due to the convexity of I_s in p(x) for $s \geq 2$ (see Theorem 2.2.4). If the s-th absolute central moment m_s is given, we obtain a similar result.

COROLLARY 3.1.2

In the class of density functions which satisfy conditions (i) - (iv) and which have a given s.a.c.m. m_s, the exponential power distribution $N_s(\mu,\beta)$, with

$$p(x) = \frac{s^{\frac{s-1}{s}}}{2\,\Gamma(1/s)\,\beta^{1/s}} \exp\left\{ - \frac{|x-\mu|^s}{s\beta} \right\} \quad s > 1, \tag{3.1.20}$$

attains minimal information $I_s = 1/\beta$.

The parameter β is the s-th absolute central moment m_s of the distribution.

Theorem 3.1.2 can be seen as a characterization of the exponential power distribution, in the sense that it possesses minimal Fisher information of order s for a given s-th absolute (central) moment β. For s = 2 this result has been given in Stam [53] and Kagan et al. [30]. Thus, it follows that the exponential power distribution is a natural extension of the gaussian density if we consider the Fisher information of order s.

Next we will consider the stability of this characterization, using a technique developed by Shimizu [51]. To this end we introduce a quantity J_s.

DEFINITION 3.1.2

Let p(x) satisfy the conditions (i) - (iv) of Theorem 3.1.2. Then we define for $s \geq 1$

$$J_s = \left[\int_{-\infty}^{\infty} \left| \frac{g(x)}{p(x)} \right|^{\frac{s}{s-1}} p(x)\, dx \right]^{s-1} \tag{3.1.21}$$

where $g(x) = p'(x) + \operatorname{sgn}[x]\, |x|^{s-1}\, m_s^{-1}\, p(x).$ (3.1.22)

It is easy to see that

$$J_s \geq 0. \tag{3.1.23}$$

Equality in Eq. 3.1.23 holds iff $g(x) = 0$ a.e. P. This is equivalent to the condition that p(x) is the exponential power distribution $N_s(0,\beta)$ as can be seen from Eq. 3.1.16. Therefore, both J_s and I_s are minimized by the same density function $N_s(0,\beta)$.

It is possible to sharpen Eq. 3.1.23 by considering the relation of J_s with I_s.

THEOREM 3.1.3

Let the density function p(x) satisfy the conditions (i) - (iv) of Theorem 3.1.2. Then the following inequality holds

$$J_s^{1/s} \geq I_s^{1/s} - m_s^{-1/s} \geq 0. \tag{3.1.24}$$

Equality in Eq. 3.1.24 holds iff (for $s \geq 2$) or if (for $1 \leq s < 2$) p(x) is the exponential power density $N_s(0,\beta)$.

Proof

We have, using Definition 3.1.1,

$$I_s^{1/s} = \left[\int_X \left| \frac{p'(x)}{p(x)} \right|^{\frac{s}{s-1}} p(x)\, dx \right]^{\frac{s-1}{s}}$$

$$\leq \left[\int_X \left| \frac{p'(x)}{p(x)} + \operatorname{sgn}[x] \, |x|^{s-1} m_s^{-1} + \right. \right.$$

$$\left. \left. + |x|^{s-1} m_s^{-1} \right|^{\frac{s}{s-1}} p(x)\, dx \right]^{\frac{s-1}{s}}. \tag{3.1.25}$$

Applying Minkowski's inequality to Eq. 3.1.25 we obtain

$$I_s^{1/s} \leq \left[\int_X \left| \frac{p'(x)}{p(x)} + \operatorname{sgn}[x] \, |x|^{s-1} m_s^{-1} \right|^{\frac{s}{s-1}} p(x)\, dx \right]^{\frac{s-1}{s}} +$$

$$+ \left[\int_X \left| \, |x|^{s-1} m_s^{-1} \right|^{\frac{s}{s-1}} p(x)\, dx \right]^{\frac{s-1}{s}}$$

$$= J_s^{1/s} + \left[m_s^{\frac{s}{1-s}} \int_X |x|^s \, p(x) \, dx \right]^{\frac{s-1}{s}}$$

$$= J_s^{1/s} + m_s^{-1/s} \tag{3.1.26}$$

or $$J_s^{1/s} \geq I_s^{1/s} - m_s^{-1/s} \tag{3.1.27}$$

from which the first inequality follows. Using Eq. 3.1.14 proves the second inequality. □

We shall need an upper bound on J_s in terms of I_s. This bound is given in the next theorem.

THEOREM 3.1.4

Let p(x) satisfy the conditions (i) - (iv) of Theorem 3.1.2. Then J_s exists if I_s exists and

$$J_s \leq 2^s \, I_s. \tag{3.1.28}$$

Proof

Using Definition 3.1.2 we find

$$J_s = \left[\int_X \left| \frac{p'(x)}{p(x)} + \mathrm{sgn}\,[x] \, |x|^{s-1} \, m_s^{-1} \right|^{\frac{s}{s-1}} p(x) \, dx \right]^{s-1}$$

$$\leq \left[\int_X \left\{ \left| \frac{p'(x)}{p(x)} \right| + |x|^{s-1} \, m_s^{-1} \right\}^{\frac{s}{s-1}} p(x) \, dx \right]^{s-1}$$

$$\leq 2^s \left[\int_X \left\{ \max \left[\left| \frac{p'(x)}{p(x)} \right| , \frac{|x|^{s-1}}{m_s} \right] \right\}^{\frac{s}{s-1}} p(x) \, dx \right]^{s-1} . \tag{3.1.29}$$

We consider two cases for Eq. 3.1.29:

a. $$\left| \frac{p'(x)}{p(x)} \right| > \frac{|x|^{s-1}}{m_s} . \qquad (3.1.30)$$

Then it follows from Eq. 3.1.29 that

$$J_s \le 2^s \, I_s. \qquad (3.1.31)$$

b. $$\left| \frac{p'(x)}{p(x)} \right| < \frac{|x|^{s-1}}{m_s} . \qquad (3.1.32)$$

Then we find from Eq. 3.1.29 that

$$J_s \le 2^s \left[\int_X m_s^{\frac{s}{1-s}} |x|^s \, p(x) \, dx \right]^{s-1} = 2^s \, m_s^{-1}. \qquad (3.1.33)$$

Combining Eqs. 3.1.31 and 3.1.33 and using Theorem 3.1.2 yields Eq. 3.1.28, which proves the theorem. □

For simplicity we shall set $m_s = \beta = 1$ in the next two lemmas and in Theorem 3.1.5. This is no restriction of the results obtained.

Before we consider the stability of the characterization of $N_s(0,1)$ we shall prove two lemmas.

LEMMA 3.1.1

For any positive number x we have

$$\int_{-x}^{x} |g(u)| \, du \le J_s^{1/s} \qquad (3.1.34)$$

with g(u) as given in Definition 3.1.2.

Proof

We have

$$\int_{-x}^{x} |g(u)| \, du = \int_{-x}^{x} \frac{|g(u)|}{p(u)} p(u) \, du. \tag{3.1.35}$$

Using Hölder's inequality we then find

$$\int_{-x}^{x} |g(u)| \, du \le \left[\int_{-x}^{x} \left\{ \frac{|g(u)|}{p(u)} \right\}^{\frac{s}{s-1}} p(u) \, du \right]^{\frac{s-1}{s}} \cdot \left[\int_{-x}^{x} p(u) \, du \right]^{\frac{1}{s}}$$

$$\le \left[\int_{-x}^{x} \left\{ \frac{|g(u)|}{p(u)} \right\}^{\frac{s}{s-1}} p(u) \, du \right]^{\frac{s-1}{s}}$$

$$\le J_s^{1/s} . \tag{3.1.36}$$

□

LEMMA 3.1.2

For $x > 0$ it holds that

$$\int_{0}^{\infty} e^{-\frac{x^s}{s}} dx \int_{-x}^{x} |g(u)| \, e^{\frac{|u|^s}{s}} du \le K(s) \, J_s^{1/s} \tag{3.1.37a}$$

where $K(s) = (3s - 3)^{1/s} \cdot \frac{s}{s-1}$. (3.1.37b)

Proof

The proof consists of two parts:

a. For all $0 < x \leq \alpha$ we have

$$\int_0^\alpha e^{-\frac{x^s}{s}} dx \int_{-x}^{x} |g(u)|\, e^{\frac{|u|^s}{s}} du \leq \int_0^\alpha e^{-\frac{x^s}{s}} dx \,.\, e^{\frac{x^s}{s}} \int_{-x}^{x} |g(u)|\, du$$

$$= \int_0^\alpha \int_{-x}^{x} |g(u)|\, du\, dx$$

$$\leq \alpha\, J_s^{1/s} \tag{3.1.38}$$

where we have used Lemma 3.1.1.

b. For $0 < \alpha < x$ we introduce the function

$$A(x) = x^{1-s} \int_{-x}^{x} |g(u)|\, e^{\frac{|u|^s}{s}} du \tag{3.1.39}$$

for which it holds that

$$0 \leq A(x) \leq x^{1-s}\, e^{\frac{x^s}{s}} \int_{-x}^{x} |g(u)|\, du$$

$$\leq \alpha^{1-s}\, e^{\frac{x^s}{s}}\, J_s^{1/s}, \tag{3.1.40}$$

where once again we have used Lemma 3.1.1. Furthermore, we have

$$\frac{d}{dx} A(x) = (1-s)\, x^{-s} \int_{-x}^{x} |g(u)|\, e^{\frac{|u|^s}{s}} du +$$

$$+ x^{1-s} e^{\frac{x^s}{s}} \{|g(x)| + |g(-x)|\} \tag{3.1.41}$$

from which it follows that

$$|A'(x)| \le |1-s| \; x^{-s} \; e^{\frac{x^s}{s}} \; J_s^{1/s} + \alpha^{1-s} \; e^{\frac{x^s}{s}} \{|g(x)| + |g(-x)|\}. \tag{3.1.42}$$

Using Eqs. 3.1.40 and 3.1.42 we can now obtain a bound for the interval (α,∞) in the integral of Eq. 3.1.37a. We denote this integral by $L_{(\alpha,\infty)}$ and consider

$$L_{(\alpha,\infty)} = \int_{\alpha}^{\infty} e^{-\frac{x^s}{s}} dx \int_{-x}^{x} |g(u)| \; e^{\frac{|u|^s}{s}} du$$

$$= \int_{\alpha}^{\infty} e^{-\frac{x^s}{s}} A(x) \; x^{s-1} \; dx. \tag{3.1.43}$$

Integration by parts yields

$$L_{(\alpha,\infty)} = e^{-\frac{\alpha^s}{s}} A(\alpha) + \int_{\alpha}^{\infty} e^{-\frac{x^s}{s}} A'(x) \; dx$$

$$\le e^{-\frac{\alpha^s}{s}} A(\alpha) + \int_{\alpha}^{\infty} e^{-\frac{x^s}{s}} |A'(x)| \; dx. \tag{3.1.44}$$

Using Eqs. 3.1.40 and 3.1.42 gives

$$L_{(\alpha,\infty)} \le \alpha^{1-s} J_s^{1/s} + |1-s|\, J_s^{1/s} \int_\alpha^\infty x^{-s}\, dx +$$

$$+ \alpha^{1-s} \int_\alpha^\infty \{|g(x)| + |g(-x)|\}\, dx$$

$$= 2\,\alpha^{1-s} J_s^{1/s} + \alpha^{1-s} \int_\alpha^\infty |g(x)|\, dx + \alpha^{1-s} \int_{-\infty}^{-\alpha} |g(x)|\, dx$$

$$\le 2\,\alpha^{1-s} J_s^{1/s} + \alpha^{1-s} \int_{-\infty}^{\infty} |g(x)|\, dx$$

$$\le 3\,\alpha^{1-s} J_s^{1/s}. \tag{3.1.45}$$

From Eq. 3.1.38 and 3.1.45 we finally obtain

$$\int_0^\infty e^{-\frac{x^s}{s}}\, dx \int_{-x}^{x} |g(u)|\, e^{\frac{|u|^s}{s}}\, du$$

$$\le C(\alpha,s)\, J_s^{1/s} \tag{3.1.46a}$$

where $C(\alpha,s) = \alpha + 3\,\alpha^{1-s}.$ (3.1.46b)

However, since $\alpha > 0$ is arbitrary, we can choose α such that $C(\alpha,s)$ is minimized. We then find from

$$\frac{d}{d\alpha} C(\alpha,s) \Big|_{\alpha=\alpha_o} = 0 \tag{3.1.47}$$

that $\alpha_0 = (3s - 3)^{1/s}$ (3.1.48) and $$C(\alpha_0, s) = K(s) = \frac{s}{s-1}(3s - 3)^{1/s} \tag{3.1.49}$$

which proves the lemma. □

Using the previous two lemmas we are now able to prove the next theorem on the stability of the characterization of the exponential distribution.

THEOREM 3.1.5

Let p(x) satisfy the conditions (i) - (iv) of Theorem 3.1.2 and let q(x) and Q(x) be the density function and distribution function of the exponential power distribution $N_s(0,1)$. Then

$$|p(x) - q(x)| \leq k(s)\, J_s^{1/s} \tag{3.1.50}$$

and $$|P(x) - Q(x)| \leq 2\, K(s)\, J_s^{1/s} \tag{3.1.51}$$

where $$k(s) = 1 + K(s) \frac{s^{\frac{s-1}{s}}}{2\,\Gamma(1/s)} \tag{3.1.52}$$

and K(s) is as defined in Eq. 3.1.37b.

<u>Proof</u>

Solving

$$g(x) = p'(x) + \operatorname{sgn}[x] \,.\, |x|^{s-1}\, p(x) \tag{3.1.53}$$

for p(x) we obtain

$$p(x) = C\, q(x) + e^{-\frac{|x|^s}{s}} \int_0^x g(u)\, e^{\frac{u^s}{s}}\, du \,, \tag{3.1.54}$$

where

$$C = \frac{2\,\Gamma(1/s)}{s^{s-1}}\, p(0) \tag{3.1.55}$$

or
$$p(x) - C\, q(x) = e^{-\frac{|x|^s}{s}} \int_0^x g(u)\, e^{\frac{u^s}{s}}\, du. \tag{3.1.56}$$

First, we shall prove that

$$\int_{-\infty}^{\infty} |p(x) - C\, q(x)|\, dx \le K(s)\, J_s^{1/s}. \tag{3.1.57}$$

To this end we note that

$$\int_{-\infty}^{\infty} |p(x) - C\, q(x)|\, dx \le \int_{-\infty}^{\infty} e^{-\frac{|x|^s}{s}}\, dx \left| \int_0^x |g(u)|\, e^{\frac{|u|^s}{s}}\, du \right|$$

$$= \int_{-\infty}^{0} e^{-\frac{|x|^s}{s}}\, dx \left| \int_0^x |g(u)|\, e^{\frac{|u|^s}{s}}\, du \right| +$$

$$+ \int_0^{\infty} e^{-\frac{|x|^s}{s}}\, dx \left| \int_0^x |g(u)|\, e^{\frac{|u|^s}{s}}\, du \right|$$

$$= \int_0^{\infty} e^{-\frac{|x|^s}{s}}\, dx \left| \int_{-x}^{0} |g(u)|\, e^{\frac{|u|^s}{s}}\, du \right| +$$

$$+ \int_0^{\infty} e^{-\frac{x^s}{s}}\, dx \int_0^x |g(u)|\, e^{\frac{u^s}{s}}\, du$$

$$= \int_0^\infty e^{-\frac{x^s}{s}} dx \int_{-x}^{x} |g(u)| \, e^{\frac{u^s}{s}} du. \qquad (3.1.58)$$

Using Lemma 3.1.2 then leads to Eq. 3.1.57. From this result we find

$$\begin{aligned} |1-C| &= |P(\infty) - C\, Q(\infty)| \\ &\le \sup_x |P(x) - C\, Q(x)| \\ &\le \int_{-\infty}^{\infty} |p(x) - C\, q(x)| \, dx \\ &\le K(s)\, J_s^{1/s}. \qquad (3.1.59) \end{aligned}$$

We now obtain

$$\begin{aligned} |p(x) - q(x)| &\le |p(x) - C\, q(x)| + |1 - C| \, q(x) \\ &\le e^{-\frac{|x|^s}{s}} \left| \int_0^x |g(u)| \, e^{\frac{u^s}{s}} du \right| + K(s)\, J_s^{1/s}\, q(x) \\ &\le J_s^{1/s} \left\{ 1 + K(s) \frac{s^{s-1}}{2\, \Gamma(1/s)} \right\} \\ &= k(s)\, J_s^{1/s} \qquad (3.1.60) \end{aligned}$$

thus proving Eq. 3.1.50. Finally it follows that

$$|P(x) - Q(x)| \le \int_{-\infty}^{\infty} |p(x) - q(x)| \, dx$$

$$\leq \int_{-\infty}^{\infty} |p(x) - C\, q(x)|\, dx + |1 - C| \int_{-\infty}^{\infty} q(x)\, dx$$

$$\leq 2\, K(s)\, J_s^{1/s}. \tag{3.1.61}$$

□

The upper bound obtained in Theorem 3.1.5 is tight if $p(x) = q(x)$ since in this case $J_s = 0$. As a consequence of the theorem we obtain the next corollary.

COROLLARY 3.1.3

Let $p(x)$ satisfy the conditions (i) - (iv) of Theorem 3.1.2 and let $q(x)$ and $Q(x)$ be the density function and distribution function of the exponential power distribution $N_s(0,1)$. Then

$$\sup_x |p(x) - q(x)| \leq k(s)\, J_s^{1/s} \tag{3.1.62}$$

$$\sup_x |P(x) - Q(x)| \leq 2\, K(s)\, J_s^{1/s}. \tag{3.1.63}$$

Proof

This follows from Theorem 3.1.5 since Eqs. 3.1.50 and 3.1.51 hold for all x. □

If we do not have one observation ξ but n observations $\xi_1, \ldots, \xi_n$, we can express I_s as follows.

THEOREM 3.1.6

Let $p_i(x_i;\theta_i)$, $i = 1, 2, \ldots, n$ be continuously differentiable density functions, with $p_i(x_i;\theta_i) > 0$ for $x \in X$, of the independent random variables $\xi_1, \ldots, \xi_n$. Let the density functions depend on location parameters $\theta_i = \alpha_i \theta$ where the α_i are real constants. Then for the random vector $\tilde{\xi} = (\xi_1, \ldots, \xi_n)$, we have the following expression for $I_s(\tilde{\xi})$

$$I_s(\tilde{\xi}) = \left[\int_{X^n} \left| \sum_{i=1}^{n} \alpha_i \frac{d}{dx_i} \log p_i(x_i) \right|^{\frac{s}{s-1}} r(\tilde{x})\, d\tilde{x} \right]^{s-1} . \quad (3.1.64)$$

Proof

From the independence of $\xi_1, \ldots, \xi_n$ it follows that

$$r(x_1, \ldots, x_n; \theta) = \prod_{i=1}^{n} p_i(x_i - \alpha_i \theta) \quad (3.1.65)$$

and $$\frac{\partial}{\partial\theta} \log r(x_1, \ldots, x_n; \theta) = \sum_{i=1}^{n} \frac{\partial}{\partial\theta} \log p_i(x_i - \alpha_i \theta)$$

$$= \sum_{i=1}^{n} \frac{\partial \log p_i(x_i - \alpha_i\theta)}{\partial(x_i - \alpha_i\theta)} \cdot \frac{\partial(x_i - \alpha_i\theta)}{\partial\theta}$$

$$= - \sum_{i=1}^{n} \alpha_i \frac{d}{dx_i} \log p_i(x_i) . \quad (3.1.66)$$

Substituting Eq. 3.1.66 into Eq. 2.1.9c and using Theorem 3.1.1, Corollary 3.1.1 and Definitions 2.1.4 and 3.1.1 completes the proof. □

In the next theorem we consider sums of random variables which result in a convolution of density functions.

THEOREM 3.1.7

Let the random variable ζ with density function $q(z;\theta)$ be the sum of n independent random variables ξ_i with density functions $p_i(x_i - \theta_i)$ and let the density functions be continuously differentiable and positive. If

$$I_s(\xi_i) = \left[\int_{X_i} \left| \frac{d}{dx_i} \log p_i(x_i) \right|^{\frac{s}{s-1}} p(x_i)\, dx_i \right]^{s-1} \tag{3.1.67}$$

exists for all $i = 1, 2, \ldots, n$, then, for any set of real numbers $\alpha_1, \ldots, \alpha_n$, we have

$$\left| \sum_{i=1}^{n} \alpha_i \right| I_s(\zeta)^{1/s} \leq \sum_{i=1}^{n} |\alpha_i|\, I_s(\xi_i)^{1/s}. \tag{3.1.68a}$$

Equality in Eq. 3.1.68a holds iff for all $i = 1, 2, \ldots, n$

$$\sum_{i=1}^{n} \alpha_i \frac{d}{dz} \log q(z) = n \,.\, \frac{d}{dx_j} \log p_j(x_j) \,.\, \sum_{j=1}^{n} \alpha_j C_{ij} \tag{3.1.68b}$$

in which the C_{ij} are positive constants and j is arbitrary.

Proof

Consider the random variables ξ_i with density functions $p_i(x_i - \alpha_i\theta)$ where $\alpha_i\theta = \theta_i$. Then

$$F_s(\xi_i;\theta) = \left[\int_{X_i} \left| \frac{\partial}{\partial\theta} \log p(x_i - \alpha_i\theta) \right|^{\frac{s}{s-1}} p(x_i - \alpha_i\theta)\, dx_i \right]^{s-1}$$

$$= \left[\int_{X_i} \left| \frac{\partial \log p(x_i - \alpha_i\theta)}{\partial(x_i - \alpha_i\theta)} \,.\, \frac{\partial(x_i - \alpha_i\theta)}{\partial\theta} \right|^{\frac{s}{s-1}} p(x_i - \alpha_i\theta)\, dx_i \right]^{s-1}$$

$$= |\alpha_i|^s\, I_s(\xi_i). \tag{3.1.69}$$

From Corollary 2.3.1 it follows that, for the joint density function $r(x_1, \ldots, x_n;\theta)$, it holds that

$$F_s(\zeta;\theta) \le \left[\sum_{i=1}^{n} F_s(\xi_i;\theta)^{1/s} \right]^s$$

$$= \left[\sum_{i=1}^{n} |\alpha_i| \, I_s(\xi_i)^{1/s} \right]^s . \qquad (3.1.70)$$

The density function $q(z;\theta)$ is continuously differentiable and positive. Using Theorem 2.2.5 yields

$$F_s(\zeta;\theta) \le F_s(\xi;\theta). \qquad (3.1.71)$$

Since $q(z;\theta) = q(z - \sum_{i=1}^{n} \alpha_i\theta)$ it follows that

$$F_s(\zeta;\theta) = \left| \sum_{i=1}^{n} \alpha_i \right|^s I_s(\zeta). \qquad (3.1.72)$$

Combination of Eqs. 3.1.70, 3.1.71 and 3.1.72 yields

$$\left| \sum_{i=1}^{n} \alpha_i \right|^s I_s(\zeta) \le \left[\sum_{i=1}^{n} |\alpha_i| \, I_s(\xi_i)^{1/s} \right]^s \qquad (3.1.73)$$

from which Eq. 3.1.68a follows. The equality condition follows from the equality conditions in Eqs. 3.1.70 and 3.1.71. Therefore, we have equality in Eq. 3.1.68a iff

$$a_i(\theta) \frac{\partial}{\partial\theta} \log p_i(x_i - \alpha_i\theta) = a_j(\theta) \frac{\partial}{\partial\theta} \log p_j(x_j - \alpha_j\theta)$$

$$\forall\ i,j \qquad \text{a.e. } x \qquad (3.1.74a)$$

and $$\frac{\partial}{\partial\theta} \log q(z;\theta) = \sum_{i=1}^{n} \frac{\partial}{\partial\theta} \log p_i(x_i - \alpha_i\theta) \qquad \text{a.e. } x \qquad (3.1.74b)$$

are satisfied. Then, after some algebraic manipulations, we find for Eqs. 3.1.74a and 3.1.74b:

$$a_i \alpha_i \frac{d}{dx_i} \log p_i(x_i) = a_j \alpha_j \frac{d}{dx_j} \log p_j(x_j) \qquad \forall\, i,j \qquad \text{a.e. } x \tag{3.1.75a}$$

and
$$\sum_{i=1}^{n} \alpha_i \frac{d}{dz} \log q(z) = \sum_{i=1}^{n} \alpha_i \frac{d}{dx_i} \log p_i(x_i) \qquad \text{a.e. } x. \tag{3.1.75b}$$

Combining Eqs. 3.1.75a and 3.1.75b leads to Eq. 3.1.68b if we set $a_j/a_i = C_{ij}$ for $i,j = 1, 2, \ldots, n$. □

As a result of this theorem we can obtain an inequality between $I_s(\zeta)$ and $I_s(\xi_i)$ by a proper choice of the constants α_i.

COROLLARY 3.1.4

Under the conditions given in Theorem 3.1.7, it holds that

$$I_s(\zeta)^{-1/s} \geq \frac{1}{n} \sum_{i=1}^{n} I_s(\xi_i)^{-1/s}. \tag{3.1.76}$$

Equality in Eq. 3.1.76 holds iff Eq. 3.1.68b is satisfied.

Proof

If we set

$$\alpha_i = a\, I_s(\xi_i)^{-1/s}, \tag{3.1.77}$$

where a is a positive constant, we obtain, using Eq. 3.1.68a, that

$$\left| \sum_{i=1}^{n} a\, I_s(\xi_i)^{-1/s} \right| \cdot I_s(\zeta)^{1/s} \leq \sum_{i=1}^{n} a = n \cdot a \tag{3.1.78}$$

from which Eq. 3.1.76 follows easily. □

3.2 SCALE PARAMETERS

As a second class of parameters, we shall consider scale parameters. In Section 1.2 a definition has been given. First, we shall give an expression for $F_s(\xi;\theta)$ if θ is a scale parameter for the distribution of the random variable ξ.

DEFINITION 3.2.1

Let the density function $q(y)$ of the random variable η satisfy the conditions

(i) $q(y) > 0 \qquad y \in Y$ (3.2.1a)

(ii) $\frac{d}{dy} q(y)$ exists for all $y \in Y$. (3.2.1b)

Then for $s \geq 1$ the quantity $I_s^*(\eta)$ is defined by

$$I_s^*(\eta) = \left[\int_Y \left| 1 + y \frac{q'(y)}{q(y)} \right|^{\frac{s}{s-1}} q(y)\, dy \right]^{s-1} . \tag{3.2.2}$$

The quantity $I_s^*(\eta)$ is related to $F_s(\xi;\theta)$ if θ is a scale parameter. In fact we can prove the following theorem.

THEOREM 3.2.1

Let $\theta \in \Theta$ be a scale parameter for the density function of ξ. Then it holds for $\eta = \xi/\theta$ that

$$I_s^*(\eta) = \theta^s F_s(\xi;\theta). \tag{3.2.3}$$

Proof

By Definition 1.2.2 we have

$$p(x;\theta) = \frac{1}{\theta} q(\frac{x}{\theta}) \tag{3.2.4}$$

which gives

$$\frac{\partial}{\partial\theta}\log p(x;\theta) = \frac{\partial}{\partial\theta}\log\frac{1}{\theta}\, q(\frac{x}{\theta})$$

$$= -\frac{1}{\theta} - \frac{x}{\theta^2}\frac{1}{q(\frac{x}{\theta})}\frac{\partial\, q(\frac{x}{\theta})}{\partial(\frac{x}{\theta})} \quad . \qquad (3.2.5)$$

Substitution of Eq. 3.2.5 into Eq. 2.1.9c yields

$$F_s(\xi;\theta) = \left[\int_X \left| -\frac{1}{\theta} - \frac{x}{\theta^2}\frac{1}{q(\frac{x}{\theta})}\frac{\partial\, q(\frac{x}{\theta})}{\partial(\frac{x}{\theta})}\right|^{\frac{s}{s-1}} \frac{1}{\theta}\, q(\frac{x}{\theta})\, dx\right]^{s-1} . \qquad (3.2.6)$$

If we set $y = \frac{x}{\theta}$ and use Definition 3.2.1 we obtain Eq. 3.2.3 which completes the proof. □

It follows from this theorem that the information in ξ concerning the parameter θ can be expressed in a particular form.

COROLLARY 3.2.1

If $\theta \in \Theta$ is a scale parameter for the density function $p(x;\theta)$ of ξ we have

$$I_s^*(\eta) = F_s(\xi;1) = \theta^s\, F_s(\xi;\theta) \ . \qquad (3.2.7)$$

Proof

The proof follows by substitution of $\theta = 1$ into Eq. 3.2.3. □

Thus, we see that in this case the information in ξ concerning θ depends on θ only through a factor θ^{-s}. Note that for decreasing values of θ the information contained in ξ about θ increases.

If the observation is a multivariate random variable $\tilde{\xi}$ with marginals

$\xi_1, \xi_2, \ldots, \xi_n$, we have the following expression for I_s^*.

THEOREM 3.2.2

Let $p_1(x_1), \ldots, p_n(x_n)$ be positive, continuously differentiable density functions of the independent random variables $\xi_1, \xi_2, \ldots, \xi_n$ and let $r(\tilde{x};\theta) = r(x_1, \ldots, x_n;\theta)$ be the density function of $\tilde{\xi}$. Then if the density functions $p_i(x_i)$ depend on a scale parameter such that $p_i(x_i) = \frac{1}{\theta} \cdot q_i(x_i/\theta)$, we have

$$I_s^*(\tilde{\xi}) = n^s \left[\int_{X^n} \left| 1 + \frac{1}{n} \sum_{i=1}^{n} x_i \frac{d}{dx_i} \log p_i(x_i) \right|^{\frac{s}{s-1}} r(\tilde{x})\, d\tilde{x} \right]^{s-1} . \tag{3.2.8}$$

<u>Proof</u>

We have

$$r(x_1, \ldots, x_n;\theta) = \prod_{i=1}^{n} \frac{1}{\theta} q_i\left(\frac{x_i}{\theta}\right) \tag{3.2.9}$$

and $$\frac{\partial}{\partial\theta} \log r(x_1, \ldots, x_n;\theta) = \sum_{i=1}^{n} \frac{\partial}{\partial\theta} \log \frac{1}{\theta} q_i\left(\frac{x_i}{\theta}\right)$$

$$= -\frac{n}{\theta} - \sum_{i=1}^{n} \frac{x_i}{\theta^2} \frac{d \log q_i(x_i/\theta)}{d(x_i/\theta)} . \tag{3.2.10}$$

From this we find

$$F_s(\tilde{\xi};\theta) = \left[\int_{Y^n} \left| \frac{n}{\theta} + \frac{1}{\theta} \sum_{i=1}^{n} y_i \frac{d}{dy_i} \log q_i(y_i) \right|^{\frac{s}{s-1}} q(\tilde{y})\, d\tilde{y} \right]^{s-1}$$

$$= \frac{n^s}{\theta^s} \left[\int_{Y^n} \left| 1 + \frac{1}{n} \sum_{i=1}^{n} y_i \frac{d}{dy_i} \log q_i(y_i) \right|^{\frac{s}{s-1}} q(\tilde{y})\, d\tilde{y} \right]^{s-1} .$$

(3.2.11)

From Corollary 3.2.1 it follows that

$$F_s(\tilde{\xi};\theta) = \frac{1}{\theta^s} I_s^*(\tilde{\eta}). \qquad (3.2.12)$$

Combining Eqs. 3.2.11 and 3.2.12 yields Eq. 3.2.8. □

If we consider a random variable ζ, which is the sum of n independent random variables, we obtain the following theorem.

THEOREM 3.2.3

Let the random variable ζ with density function $q(z;\theta)$ be the sum of n independent random variables $\xi_1, \ldots, \xi_n$ with density functions $p_i(x_i;\theta) = \frac{1}{\theta} q_i(x_i/\theta)$, $i = 1, \ldots, n$ and let the density functions be positive and continuously differentiable. Then if

$$I_s^*(\xi_i) = \left[\int_{X_i} \left| 1 + x_i \frac{p_i'(x_i)}{p_i(x_i)} \right|^{\frac{s}{s-1}} p_i(x_i)\, dx_i \right]^{s-1} \qquad (3.2.13)$$

exists for all $i = 1, \ldots, n$, we have

$$\left[\int_{-\infty}^{\infty} \left| 1 + z \frac{q'(z)}{q(z)} \right|^{\frac{s}{s-1}} q(z)\, dz \right]^{\frac{s-1}{s}}$$

$$\leq \sum_{i=1}^{n} \left[\int_{-\infty}^{\infty} \left| 1 + x_i \frac{p_i'(x_i)}{p_i(x_i)} \right|^{\frac{s}{s-1}} p_i(x_i)\, dx_i \right]^{\frac{s-1}{s}} \qquad (3.2.14a)$$

where $q(z)$ is the convolution of $p_1(x_1), \dots, p_n(x_n)$. Equality in Eq. 3.2.14a holds iff for some $j = 1, 2, \dots, n$

$$1 + z\,\frac{q'(z)}{q(z)} = \sum_{i=1}^{n} k_{ij} \left[1 + x_j \frac{p_j'(x_j)}{p_j(x_j)} \right] \qquad \text{a.e. } Q, \tag{3.2.14b}$$

where k_{ij} are positive constants and j is arbitrary.

Proof

The random variable ζ has a density function $1/\theta\, q(z/\theta)$ which is also differentiable and positive. From this we have

$$I_s^*(\zeta) = \left[\int_{-\infty}^{\infty} \left| 1 + z\,\frac{q'(z)}{q(z)} \right|^{\frac{s}{s-1}} q(z)\, dz \right]^{s-1} . \tag{3.2.15}$$

From Corollary 2.3.1 it follows that for the joint density function $r(x_1, \dots, x_n)$ of $\tilde{\xi}$ it holds that

$$I_s^*(\tilde{\xi}) \le \left[\sum_{i=1}^{n} I_s^*(\xi_i)^{1/s} \right]^{s} . \tag{3.2.16}$$

Using Theorem 2.2.5 yields

$$I_s^*(\zeta) \le I_s^*(\tilde{\xi}) . \tag{3.2.17}$$

Combining Eqs. 3.2.15, 3.2.16 and 3.2.17 proves Eq. 3.2.14a. Equality in Eq. 3.2.14a holds iff

$$1 + z\,\frac{q'(z)}{q(z)} = \sum_{i=1}^{n} \left[1 + x_i \frac{p_i'(x_i)}{p_i(x_i)} \right] \qquad \text{a.e. } Q \tag{3.2.18a}$$

and $$k_i \left[1 + x_i \frac{p_i'(x_i)}{p_i(x_i)} \right] = k_j \left[1 + x_j \frac{p_j'(x_j)}{p_j(x_j)} \right] \qquad \text{a.e. } Q \tag{3.2.18b}$$

both hold. Setting $k_{ij} = k_j/k_i$ we obtain, after combining Eqs. 3.2.18a and 3.2.18b, the condition as given in Eq. 3.2.14b.

3.3 $F_s(\xi;\theta)$ FOR SOME DISTRIBUTIONS

In this section we shall consider some distributions and compute their Fisher information of order s. We shall use the expressions for $F_s(\xi;\theta)$ obtained in Sections 3.1 and 3.2 for location and scale parameters. Because of the relations which have been obtained in Theorems 3.1.1 and 3.2.1 it is sufficient to consider the information measures $I_s(\xi)$ and $I_s^*(\xi)$. The two most important distributions which we consider are the exponential power distribution and a generalized Weibull distribution which each include several well-known distributions as special cases.

First, we shall consider the exponential power distribution for the case that θ is a location parameter.

EXAMPLE 3.3.1

Consider the exponential power distribution $N_\nu(\mu,\beta)$ with density function

$$p(x;\mu,\beta,\nu) = \frac{\nu^{\frac{\nu-1}{\nu}}}{2\ \Gamma(1/\nu)\ \beta^{1/\nu}} \exp\left\{-\frac{|x-\mu|^\nu}{\nu\beta}\right\}, \tag{3.3.1}$$

where $\nu > 0$ is a shape parameter, called the order of this distribution, and β is the s-th absolute central moment. Then it holds for the location parameter $\theta=\mu$ that for $s > 1$

$$I_s(\xi) = \beta^{-\frac{s}{\nu}} \left[\nu^{\nu-1}\right]^{\frac{s}{\nu}} \left[\frac{\Gamma\left(\frac{s\nu-1}{s\nu-\nu}\right)}{\Gamma(1/\nu)}\right]^{s-1}. \tag{3.3.2}$$

Proof

From Corollary 3.1.1 it follows that we can set $\mu=0$ without loss of generality. Then we have

$$\frac{d}{dx} p(x) = - p(x) \cdot \frac{1}{\beta} |x|^{\nu-1} \operatorname{sgn}[x], \tag{3.3.3}$$

from which it follows that

$$\left| \frac{p'(x)}{p(x)} \right| = \frac{1}{\beta} |x|^{\nu-1} \tag{3.3.4}$$

and

$$I_s(\xi) = \left[\int_{-\infty}^{\infty} \left\{ \frac{|x|^{\nu-1}}{\beta} \right\}^{\frac{s}{s-1}} \frac{\nu^{\frac{\nu-1}{\nu}}}{2\beta^{1/\nu}\,\Gamma(1/\nu)} \exp\left\{ - \frac{|x|^{\nu}}{\nu\beta} \right\} dx \right]^{s-1} . \tag{3.3.5}$$

Using Gradshteyn et al. [24] it follows that for $s > 1$

$$I_s(\xi) = \frac{2^{s-1}}{\beta^s} \left[\frac{1}{\beta}(\nu\beta)^{\frac{1}{\nu}\left(1 + \frac{s(\nu-1)}{s-1}\right)} \Gamma\left(\frac{\frac{s(\nu-1)}{s-1} - 1}{\nu} \right) \frac{\nu^{\frac{\nu-1}{\nu}}}{2\beta^{1/\nu}\,\Gamma(1/\nu)} \right]^{s-1} , \tag{3.3.6}$$

which after some algebraic manipulations leads to Eq. 3.3.2. □

This example leads to some special cases which we will mention explicitly in the next three examples.

EXAMPLE 3.3.1a

If the density function $p(x;\theta)$ is an exponential power distribution of order s, we have

$$I_s(\xi) = \frac{1}{\beta} . \tag{3.3.7}$$

Proof

This follows at once from Eq. 3.3.2 by substitution of $\nu = s$ and also from the proof of Theorem 3.1.2. □

Note that $1/\beta$ is also the minimum value of $I_s(\xi)$ that can be attained by $p(x)$ for a given s-th absolute central moment. If we set $s = 2$ we have the familiar result that for a gaussian distribution $N(\mu,\sigma^2)$ the Fisher information is equal to σ^{-2}.

EXAMPLE 3.3.1b

If the density function is a gaussian distribution with

$$p(x;\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}, \tag{3.3.8}$$

then
$$I_s(\xi) = \frac{\sqrt{2}}{\sigma}\left[\frac{\sqrt{2}\ \Gamma\left(\frac{s-\frac{1}{2}}{s-1}\right)}{\sigma\sqrt{\pi}}\right]^{s-1}. \tag{3.3.9}$$

Proof

Substitution of the values $\beta = \sigma^2$ and $\nu = 2$ into Eq. 3.3.2 yields 3.3.9. □

EXAMPLE 3.3.1c

If the density function is a Laplace distribution with

$$p(x;\mu,\lambda) = \frac{1}{2\lambda} \exp\left\{-\frac{|x-\mu|}{\lambda}\right\}, \tag{3.3.10}$$

then
$$I_s(\xi) = \frac{1}{\lambda^s}. \tag{3.3.11}$$

Proof

In this case we have $\nu = 1$ and $\beta = \lambda$. Substitution of these values into Eq. 3.3.2 leads to Eq. 3.3.11. Note that the density function is not differentiable for $x = \mu$. However, the left and right derivatives exist and are finite, so that the result obtained for $I_s(\xi)$ is meaningful. □

As follows from the previous results, the exponential power distribution leads to simple expressions for $I_s(\xi)$. The next two examples lead to more complex results. First we consider a generalized Cauchy distribution, which was introduced by Miller et al. [42] as a model for non-gaussian signals. We use a simpler notation.

EXAMPLE 3.3.2

Let us consider a generalized Cauchy distribution with density function

$$p(x;\rho,\mu,\nu,c) = \frac{c}{2\rho\ B(\nu,1/c)} \left[1 + \left(\frac{|x-\mu|}{\rho} \right)^c \right]^{-\left(\nu + \frac{1}{c}\right)} \tag{3.3.12a}$$

where $$\rho = \lambda\ \nu^{1/c} \left[\frac{\Gamma(1/c)}{\Gamma(3/c)} \right]^{\frac{1}{2}} \tag{3.3.12b}$$

and where λ, c and ν are positive parameters. Then for $s > 1$ it holds that for the location parameter $\theta = \mu$:

$$I_s(\xi) = \left[\frac{\nu c + 1}{\rho} \right]^s \frac{B^{s-1}\left(\frac{cs-1}{cs-c}, \nu + \frac{s}{cs-c} \right)}{B^{s-1}(\nu,1/c)} . \tag{3.3.13}$$

<u>Proof</u>

We may set $\mu = 0$. Then we have

$$p'(x) = p(x) . -(\nu c + 1)\rho^{-c}\ |x|^{c-1}\ \mathrm{sgn}\ [x]\ . \left[1 + \left(\frac{|x|}{\rho} \right)^c \right]^{-1} \tag{3.3.14}$$

from which we obtain

$$\left| \frac{p'(x)}{p(x)} \right| = \frac{(\nu c + 1)\ |x|^{c-1}}{\rho^c \left[1 + \left(\frac{|x|}{\rho} \right)^c \right]} . \qquad (3.3.15)$$

Thus
$$I_s(\xi) = \left[\int_{-\infty}^{\infty} \left[\frac{(\nu c + 1)\ |x|^{c-1}}{\rho^c \left[1 + \left(\frac{|x|}{\rho} \right)^c \right]} \right]^{\frac{s}{s-1}} \cdot \frac{c}{2\rho\ B(\nu,1/c)} \right.$$

$$\left. \cdot \left[1 + \left(\frac{|x|}{\rho} \right)^c \right]^{-\left(\nu + \frac{1}{c}\right)} dx \right]^{s-1} \qquad (3.3.16)$$

from which we find with $x/\rho = y$ that

$$I_s(\xi) = \frac{c^{s-1} (\nu c + 1)^s}{\rho^s\ B^{s-1}(\nu,1/c)} \left[\int_0^{\infty} \frac{y^{\frac{s(c-1)}{s-1}}}{(1 + y^c)^{\nu + \frac{1}{c} + \frac{s}{s-1}}}\, dy \right]^{s-1} . \qquad (3.3.17)$$

Using Gradshteyn et al. [24] finally leads to Eq. 3.3.13. □

As a first special case we consider the following distribution.

EXAMPLE 3.3.2a

Let the density function be given by

$$p(x;\rho,\mu,c) = \frac{c}{2\rho\ \Gamma(1/c)\ \Gamma(\frac{c-1}{c})} \left[1 + \left(\frac{|x-\mu|}{\rho} \right)^c \right]^{-1} , \qquad (3.3.18a)$$

where
$$\rho = \lambda \left(\frac{c-1}{c} \right)^{1/c} \left[\frac{\Gamma(1/c)}{\Gamma(3/c)} \right]^{\frac{1}{2}} . \qquad (3.3.18b)$$

Then for $s > 1$ we have

$$I_s(\xi) = \left(\frac{c}{\rho}\right)^s \cdot \frac{B^{s-1}\left(\frac{s-1/c}{s-1}, \frac{s-1+1/c}{s-1}\right)}{B^{s-1}\left(\frac{c-1}{c}, \frac{1}{c}\right)} . \qquad (3.3.19)$$

Proof

Substitution of $\nu = (c-1)/c$ into Eq. 3.3.13 leads to Eq. 3.3.19. □

As a second special case we consider the usual Cauchy distribution.

EXAMPLE 3.3.2b

If the density function is a Cauchy distribution with

$$p(x;\lambda,\mu) = \frac{1}{\pi\lambda} \frac{1}{1 + \left(\frac{x-\mu}{\lambda}\right)^2} , \qquad (3.3.20)$$

then $$I_s(\xi) = \frac{2^s}{\pi^{s-1}\lambda^s} B^{s-1}\left(\frac{s-\frac{1}{2}}{s-1}, \frac{s-\frac{1}{2}}{s-1}\right). \qquad (3.3.21)$$

Proof

If we set $c = 2$ and $\nu = \frac{1}{2}$ in Eq. 3.3.13, we obtain Eq. 3.3.21. □

Note that since $B(3/2,3/2) = \pi/8$ the case $s = 2$ leads to the well-known result (see Cohen [15]),

$$I_2(\xi) = \frac{1}{2\lambda^2} \qquad (3.3.22)$$

where I_2 is the usual Fisher information.

The next distribution is a generalized Beta distribution, which has been proposed by Miller et al. [42].

EXAMPLE 3.3.3

Consider the generalized beta distribution with density function

$$p(x;\rho,\mu,\nu,c) = \frac{c}{2\rho \; B\left(\nu + 1 + \frac{1}{c}, \frac{1}{c}\right)} \left[1 - \left(\frac{|x-\mu|}{\rho}\right)^c\right]^{\nu + \frac{1}{c}} \qquad |x-\mu| < \rho$$

$$= 0 \qquad |x-\mu| \geq \rho$$

(3.3.23a)

$$\text{where} \quad \rho = \lambda \; \nu^{1/c} \left[\frac{\Gamma(1/c)}{\Gamma(3/c)}\right]^{\frac{1}{2}} \qquad (3.3.23b)$$

and where λ, c and ν are positive parameters. Then it holds for $s > 1 + c/(1 + \nu c)$ that for the location parameter $\theta = \mu$

$$I_s(\xi) = \left[\frac{\nu c+1}{\rho}\right]^s \frac{B^{s-1}\left(\frac{cs-1}{cs-c}, 1 + \nu + \frac{1}{c} - \frac{s}{s-1}\right)}{B^{s-1}\left(\nu + 1 + \frac{1}{c}, \frac{1}{c}\right)} . \qquad (3.3.24)$$

Proof

We have

$$p'(x) = p(x). -(\nu c+1) |x|^{c-1} \operatorname{sgn}[x] \left[1 - \left(\frac{|x|}{\rho}\right)^c\right]^{-1} \cdot \frac{1}{\rho^c} \qquad (3.3.25)$$

so that

$$\left| \frac{p'(x)}{p(x)} \right| = \frac{(\nu c + 1) \; |x|^{c-1}}{\rho^c \left| 1 - \left(\frac{|x|}{\rho}\right)^c \right|} . \qquad (3.3.26)$$

Therefore, we find

$$I_s(\xi) = \left[\int_{-\rho}^{\rho} \left[\frac{(\nu c + 1)\ |x|^{c-1}}{\rho^c \left| 1 - \left(\frac{|x|}{\rho}\right)^c \right|} \right]^{\frac{s}{s-1}} \cdot \frac{c}{2\rho\ B\left(\nu + 1 + \frac{1}{c}, \frac{1}{c}\right)} \cdot \right.$$

$$\left. \cdot \left[1 - \left(\frac{|x|}{\rho}\right)^c \right]^{\nu + \frac{1}{c}} dx \right]^{s-1} \tag{3.3.27}$$

from which we obtain for $c > 1$

$$I_s(\xi) = \frac{(\nu c + 1)^s\ c^{s-1}}{\rho^s\ B^{s-1}\left(\nu + 1 + \frac{1}{c}, \frac{1}{c}\right)} \left[\int_0^1 y^{\frac{s(c-1)}{s-1}} (1-y^c)^{\nu + \frac{1}{c} - \frac{s}{s-1}} dy \right]^{s-1} . \tag{3.3.28}$$

Using Gradsteyn et al. [24] yields Eq. 3.3.24 provided

$$\frac{s}{s-1} < 1 + \nu + \frac{1}{c} \tag{3.3.29}$$

or $$s > \frac{1 + c + \nu c}{1 + \nu c} = 1 + \frac{c}{1 + \nu c} . \tag{3.3.30}$$

□

As a special case we consider the following distribution.

EXAMPLE 3.3.3a

Let us consider the density function

$$p(x;\rho,\mu,c) = \frac{c^2}{2\rho(c+1)} \left[1 - \left(\frac{|x-\mu|}{\rho}\right)^c \right] \qquad |x-\mu| < \rho$$

$$= 0 \qquad |x-\mu| \geq \rho \tag{3.3.31a}$$

$$\text{where} \quad \rho = \lambda\left(\frac{c-1}{c}\right)^{1/c} \left[\frac{\Gamma(1/c)}{\Gamma(3/c)}\right]^{\frac{1}{2}} . \tag{3.3.31b}$$

Then for $s > 2$ we have

$$I_s(\xi) = \frac{c^{3s-2}}{\rho^s (c+1)^{s-1}} B^{s-1}\left(\frac{sc-1}{sc-c}, \frac{s-2}{s-1}\right) . \tag{3.3.32}$$

Proof

If we substitute $\nu = (c-1)/c$ into Eq. 3.3.24 we obtain Eq. 3.3.32.

□

As a second special case we consider a certain class of density functions from the beta distribution.

EXAMPLE 3.3.3b

Let the density function be a double-sided beta distribution, with

$$\begin{aligned} p(x;\rho,\mu,b) &= \frac{b}{2\rho}\left[1 - \frac{|x-\mu|}{\rho}\right]^{b-1} & |x-\mu| < \rho \\ &= 0 & |x-\mu| \geq \rho \end{aligned} \tag{3.3.33a}$$

$$\text{where} \quad \rho = \frac{\lambda}{\sqrt{2}} (b-1) . \tag{3.3.33b}$$

Then for $s > b/(b-1)$ it holds that

$$I_s(\xi) = \left(\frac{b-1}{\rho}\right)^s \left[\frac{b(s-1)}{b(s-1) - s}\right]^{s-1} . \tag{3.3.34}$$

Proof

Setting $c = 1$ and $\nu = b-2$ and substituting these values into Eq. 3.3.24 yields Eq. 3.3.34. □

The previous results have been obtained for $I_s(\xi)$, so for the case that θ is a location parameter. In the following part of this section we shall assume that the parameter θ is a scale parameter. First we shall consider a distribution which we shall call a generalized Weibull distribution.

EXAMPLE 3.3.4

Consider the generalized Weibull distribution with density function

$$p(x;\rho,\mu,c) = \frac{c}{\rho^{\mu}\,\Gamma(\frac{\mu}{c})}\, x^{\mu-1} \exp\left[-\left(\frac{x}{\rho}\right)^{c}\right]. \qquad x \geq 0 \qquad (3.3.35)$$

where ρ, μ and c are positive parameters. Then for $s > 1$ it holds in the case that $\theta = \rho$ is a scale parameter, that

$$I_s^*(\xi) = c^s\, \Gamma^{1-s}\left(\frac{\mu}{c}\right)\left[\left(\frac{\mu}{c}\right)^{\frac{s}{s-1}+\frac{\mu}{c}} B\left(\frac{s}{s-1}, \frac{\mu}{c}\right) .\right.$$

$$\cdot\ {}_1F_1\left(\frac{\mu}{c}, \frac{s}{s+1}+\frac{\mu}{c}+1; -\frac{\mu}{c}\right) + \left(\frac{\mu}{c}\right)^{\frac{s-\frac{1}{2}}{s-1}+\frac{\mu}{2c}-1} .\,\Gamma\left(\frac{2s-1}{s-1}\right)e^{-\frac{\mu}{2c}} .$$

$$\left.\cdot\ W_{\frac{\mu}{2c}-\frac{s-\frac{1}{2}}{s-1},\ \frac{1}{2}-\frac{s-\frac{1}{2}}{s-1}-\frac{\mu}{2c}}\left(\frac{\mu}{c}\right)\right]^{s-1}, \qquad (3.3.36)$$

Here ${}_1F_1(.)$ is the degenerate hypergeometric function and $W(.)$ is Whittaker's function (see Appendix B).

Proof

Because of Corollary 3.2.1 we can set $\rho = 1$. We then have

$$p(x) = \frac{c}{\Gamma(\frac{\mu}{c})}\, x^{\mu-1} \exp\{-x^{c}\} \qquad (3.3.37)$$

and

$$p'(x) = p(x)\left[\frac{\mu-1}{x} - c\,x^{c-1}\right] \tag{3.3.38}$$

from which we can obtain

$$\left| 1 + x\,\frac{p'(x)}{p(x)} \right| = |\,\mu - c\,x^{c}\,|. \tag{3.3.39}$$

Then it follows that

$$I_s^*(\xi) = \left[\int_0^\infty |\mu - c\,x^c|^{\frac{s}{s-1}} \frac{c}{\Gamma(\frac{\mu}{c})}\, x^{\mu-1}\, e^{-x^c}\, dx\right]^{s-1}$$

$$= \frac{c^{2s-1}}{\Gamma^{s-1}(\frac{\mu}{c})} \left[\int_0^\infty \left|\frac{\mu}{c} - x^c\right|^{\frac{s}{s-1}} x^{\mu-1}\, e^{-x^c}\, dx\right]^{s-1}. \tag{3.3.40}$$

If we denote the integral between braces by L we find

$$L = \int_0^{\left(\frac{\mu}{c}\right)^{1/c}} \left(\frac{\mu}{c} - x^c\right)^{\frac{s}{s-1}} x^{\mu-1}\, e^{-x^c}\, dx +$$

$$+ \int_{\left(\frac{\mu}{c}\right)^{1/c}}^{\infty} \left(x^c - \frac{\mu}{c}\right)^{\frac{s}{s-1}} x^{\mu-1}\, e^{-x^c}\, dx$$

$$= c^{-1} \int_0^{\frac{\mu}{c}} \left(\frac{\mu}{c} - z\right)^{\frac{s}{s-1}} z^{\frac{\mu}{c} - 1}\, e^{-z}\, dz +$$

$$+ c^{-1} \int_{\frac{\mu}{c}}^{\infty} \left(z - \frac{\mu}{c}\right)^{\frac{s}{s-1}} z^{\frac{\mu}{c} - 1}\, e^{-z}\, dz. \tag{3.3.41}$$

Using Gradshteyn et al. [24], we obtain

$$L = \left(\frac{\mu}{c}\right)^{\frac{s}{s-1} + \frac{\mu}{c} - 1} c^{-1} B\left(\frac{s}{s-1}, \frac{\mu}{c}\right) {}_1F_1\left(\frac{\mu}{c}, \frac{s}{s-1} + \frac{\mu}{c} + 1; -\frac{\mu}{c}\right) +$$

$$+ c^{-1} \left(\frac{\mu}{c}\right)^{\frac{s-\frac{1}{2}}{s-1} + \frac{\mu}{2c} - 1} \Gamma\left(\frac{2s-1}{s-1}\right) e^{-\frac{\mu}{2c}} W_{\frac{\mu}{2c} - \frac{s-\frac{1}{2}}{s-1}, \frac{1}{2} - \frac{s-\frac{1}{2}}{s-1} - \frac{\mu}{2c}} \left(\frac{\mu}{c}\right).$$

(3.3.42)

Substitution of Eq. 3.3.42 into Eq. 3.3.40 leads to Eq. 3.3.36. □

We shall consider some special cases of this distribution in the next four examples.

EXAMPLE 3.3.4a

If the density function is the usual Weibull distribution with

$$p(x;\rho,c) = \frac{c}{\rho^c} x^{c-1} \exp\left[-\left(\frac{x}{\rho}\right)^c\right] \qquad x \geq 0 \qquad (3.3.43)$$

it holds that

$$I_s^*(\xi) = c^s e^{1-s} \left[\Gamma\left(\frac{2s-1}{s-1}\right) + (-1)^{\frac{s}{s-1}} \gamma\left(\frac{2s-1}{s-1}, -1\right)\right]^{s-1} \qquad (3.3.44)$$

where $\gamma(.)$ is the incomplete gamma function.

Proof

Substitution of $\mu = c$ into Eq. 3.3.36 leads after some algebraic manipulations to Eq. 3.3.44. □

A special case of the Weibull distribution is the Rayleigh distribution.

EXAMPLE 3.3.4b

If the density function is a Rayleigh distribution with

$$p(x;\rho) = \frac{2}{\rho^2}\, x \exp\left[-\left(\frac{x}{\rho}\right)^2\right] \qquad x \geq 0 \tag{3.3.45}$$

then $$I_s^*(\xi) = 2^s\, e^{1-s} \left[\Gamma\left(\frac{2s-1}{s-1}\right) + (-1)^{\frac{s}{s-1}}\, \gamma\left(\frac{2s-1}{s-1}, -1\right)\right]^{s-1} . \tag{3.3.46}$$

Proof

This result follows from the substitution of c = 2 into Eq. 3.3.44.

□

Another special case which can be obtained from the generalized Weibull distribution is the one-sided exponential power distribution for $x \geq 0$, which we obtain by setting $\mu = 1$ in Eq. 3.3.35.

EXAMPLE 3.3.4c

For the one-sided exponential power distribution of order c, where

$$p(x;\rho,c) = \frac{c}{\rho\, \Gamma(1/c)} \exp\left[-\left(\frac{x}{\rho}\right)^c\right] \qquad x \geq 0, \tag{3.3.47}$$

it holds that

$$I_s^*(\xi) = c^s\, \Gamma^{1-s}(1/c) \left[c^{-\frac{s}{s-1} - \frac{1}{c}}\, B\left(\frac{s}{s-1}, \frac{1}{c}\right) \cdot {}_1F_1\left(\frac{1}{c}, \frac{s}{s-1} + \frac{1}{c} + 1; -\frac{1}{c}\right) + c^{-\frac{s-\frac{1}{2}}{s-1} - \frac{1}{2c} + 1}\, \Gamma\left(\frac{2s-1}{s-1}\right) \cdot e^{-\frac{1}{2c}}\, W_{\frac{1}{2c} - \frac{s-\frac{1}{2}}{s-1},\, \frac{1}{2} - \frac{s-\frac{1}{2}}{s-1} - \frac{1}{2c}}\left(\frac{1}{c}\right) \right]^{s-1} . \tag{3.3.48}$$

Proof

Substitution of $\mu = 1$ into Eq. 3.3.36 yields Eq. 3.3.48. □

This includes the exponential distribution for $c = 1$. For this distribution we find that

$$I_s^*(\xi) = e^{1-s}\left[\Gamma\left(\frac{2s-1}{s-1}\right) + (-1)^{\frac{s}{s-1}}\,\gamma\left(\frac{s}{s-1}, -1\right)\right]^{s-1} . \qquad (3.3.49)$$

Finally we shall consider the gamma distribution which can also be obtained from the generalized Weibull distribution as a special case.

EXAMPLE 3.3.4d

If the density function is a gamma distribution with

$$p(x;\rho,\mu) = \frac{1}{\rho^{\mu}\,\Gamma(\mu)}\, x^{\mu-1} e^{-\frac{x}{\rho}} \qquad x \geq 0, \qquad (3.3.50)$$

then $$I_s^*(\xi) = \Gamma^{1-s}(\mu)\left[\mu^{\frac{s}{s-1}+\mu} B\left(\frac{s}{s-1}, \mu\right)\, {}_1F_1\left(\mu, \frac{s}{s-1} + \mu + 1; -\mu\right) + \right.$$

$$\left. + \mu^{\frac{s-\frac{1}{2}}{s-1} + \frac{\mu}{2} - 1}\,\Gamma\left(\frac{2s-1}{s-1}\right) e^{-\frac{\mu}{2}}\, W_{\frac{\mu}{2} - \frac{s-\frac{1}{2}}{s-1},\, \frac{1}{2} - \frac{s-\frac{1}{2}}{s-1} - \frac{\mu}{2}}(\mu)\right]^{s-1} .$$

(3.3.51)

Proof

Substitution of $c = 1$ into Eq. 3.3.36 leads to Eq. 3.3.51. □

As has been shown in this section the computation of the quantity $I_s(\xi)$, which is equal to $F_s(\xi;\theta)$ if θ is a location parameter, leads to simple results. The quantity $I_s^*(\xi)$, which is related to $F_s(\xi;\theta)$ if θ is a scale parameter, is more complicated than $I_s(\xi)$, as is shown in the Examples 3.3.4 - 3.3.4d. Since the families are quite general, they also include densities which are not differentiable everywhere. This is not of crucial importance, since in those cases both the left and the right derivatives exist and are finite. As a result the expressions obtained remain valid.

In Chapter 4 we will use the results of this section and consider some applications. We summarize the results in the Tables 1 and 2 for further reference.

3.4 ENTROPY MOMENT OF ORDER S

It is of interest to compare the Fisher information of order s with a quantity which can be obtained from the Shannon entropy, called the entropy moment of order s. For s = 2 the Fisher information has been related to the entropy power by Stam [53]. Based on this relation Stam was able to give a rigorous proof of the so-called convolution inequality. For $s \geq 1$ such an inequality cannot be obtained. However, there still exist some interesting analogies, as will be shown in this section.

For a random variable ξ with density function $p(x;\theta)$, where $\theta \in \Theta$ is a non-random parameter, the Shannon entropy is defined as

$$H(\xi;\theta) = - \int_{-\infty}^{\infty} p(x;\theta) \log p(x;\theta)\, dx \qquad (3.4.1)$$

provided the integral exists.

A problem which has been considered by several authors (see Boekee [9]) is the maximization of $H(\xi)$ given one or more constraints on the density function $p(x)$ of the form

name	$p(x;\mu)$	$I_s(\xi)$
exponential power $N_\nu(\mu,\beta)$	$\dfrac{\nu^{\frac{\nu-1}{\nu}}}{2\,\Gamma(1/\nu)\,\beta^{1/\nu}}\ \exp\left[-\dfrac{\lvert x-\mu\rvert^{\nu}}{\nu\beta}\right]$	$\beta^{-\frac{s}{\nu}}\left[\nu^{\nu-1}\right]^{\frac{s}{\nu}}\left[\dfrac{\Gamma\left(\frac{s\nu-1}{s\nu-\nu}\right)}{\Gamma(1/\nu)}\right]^{s-1}$
exponential power $N_s(\mu,\beta)$	$\dfrac{s^{\frac{s-1}{s}}}{2\,\Gamma(1/s)\,\beta^{1/s}}\ \exp\left[-\dfrac{\lvert x-\mu\rvert^{s}}{s\beta}\right]$	$\dfrac{1}{\beta}$
gaussian	$\dfrac{1}{\sigma\sqrt{2\pi}}\ \exp\left[-\dfrac{(x-\mu)^2}{2\sigma^2}\right]$	$\dfrac{\sqrt{2}}{\sigma}\left[\dfrac{\sqrt{2}\ \Gamma\left(\frac{s-\frac{1}{2}}{s-1}\right)}{\sigma\sqrt{\pi}}\right]^{s-1}$
Laplace	$\dfrac{1}{2\lambda}\exp\left[-\dfrac{\lvert x-\mu\rvert}{\lambda}\right]$	$\dfrac{1}{\lambda^s}$
generalized Cauchy	$\dfrac{c}{2\rho\ B(\nu,1/c)}\left[1+\left(\dfrac{\lvert x-\mu\rvert}{\rho}\right)^{c}\right]^{-(\nu+\frac{1}{c})}$	$\left[\dfrac{\nu c+1}{\rho}\right]^{s}\dfrac{B^{s-1}\left(\frac{cs-1}{cs-c},\ \nu+\frac{s}{cs-c}\right)}{B^{s-1}(\nu,1/c)}$
Cauchy	$\dfrac{1}{\pi\lambda}\ \dfrac{1}{1+\left(\frac{x-\mu}{\lambda}\right)^2}$	$\dfrac{2^s}{\pi^{s-1}\,\lambda^s}\ B^{s-1}\left(\frac{s-\frac{1}{2}}{s-1},\ \frac{s-\frac{1}{2}}{s-1}\right)$
generalized beta	$\dfrac{c}{2\rho\ B(\nu+1+\frac{1}{c},\ \frac{1}{c})}\left[1-\left(\dfrac{\lvert x-\mu\rvert}{\rho}\right)^{c}\right]^{\nu+\frac{1}{c}}$ $\lvert x-\mu\rvert < \rho$	$\left[\dfrac{\nu c+1}{\rho}\right]^{s}\dfrac{B^{s-1}\left(\frac{cs-1}{cs-c},\ 1+\nu+\frac{1}{c}-\frac{s}{s-1}\right)}{B^{s-1}(\nu+1+\frac{1}{c},\ \frac{1}{c})}$
double-sided beta	$\dfrac{b}{2\rho}\left[1-\dfrac{\lvert x-\mu\rvert}{\rho}\right]^{b-1}$ $\lvert x-\mu\rvert < \rho$	$\left(\dfrac{b-1}{\rho}\right)^{s}\left[\dfrac{b(s-1)}{b(s-1)-s}\right]^{s-1}$

TABLE 1. The quantity $I_s(\xi)$ for some distributions.

name	$p(x;\rho)$	$I_s^*(\xi)$
generalized Weibull	$\frac{c}{\rho^{\mu}\,\Gamma(\frac{\mu}{c})}\,x^{\mu-1}\exp\left[-\left(\frac{x}{\rho}\right)^{c}\right]$	$c^{s}\,\Gamma^{1-s}\left(\frac{\mu}{c}\right)\left[\left(\frac{\mu}{c}\right)^{\frac{s}{s-1}+\frac{\mu}{c}} B\left(\frac{s}{s-1},\frac{\mu}{c}\right)\,{}_1F_1\left(\frac{\mu}{c},\frac{s}{s+1}+\frac{\mu}{c}+1;-\frac{\mu}{c}\right)+\right.$ $\left.+\left(\frac{\mu}{c}\right)^{\frac{s-\frac{1}{2}}{s-1}+\frac{\mu}{2c}-1}\Gamma\left(\frac{2s-1}{s-1}\right)e^{-\frac{\mu}{2c}}\,W_{\frac{\mu}{2c}-\frac{s-\frac{1}{2}}{s-1},\,\frac{1}{2}-\frac{s-\frac{1}{2}}{s-1}-\frac{\mu}{2c}}\left(\frac{\mu}{c}\right)\right]^{s-1}$
Weibull	$\frac{c}{\rho^{c}}\,x^{c-1}\exp\left[-\left(\frac{x}{\rho}\right)^{c}\right]$	$c^{s}\,e^{1-s}\left[\Gamma\left(\frac{2s-1}{s-1}\right)+(-1)^{\frac{s}{s-1}}\gamma\left(\frac{2s-1}{s-1},-1\right)\right]^{s-1}$
Rayleigh	$\frac{2}{\rho^{2}}\,x\exp\left[-\left(\frac{x}{\rho}\right)^{2}\right]$	$2^{s}\,e^{1-s}\left[\Gamma\left(\frac{2s-1}{s-1}\right)+(-1)^{\frac{s}{s-1}}\gamma\left(\frac{2s-1}{s-1},-1\right)\right]^{s-1}$
one-sided exponential power	$\frac{c}{\rho\,\Gamma(1/c)}\exp\left[-\left(\frac{x}{\rho}\right)^{c}\right]$	$c^{s}\,\Gamma^{1-s}\left(\frac{1}{c}\right)\left[c^{-\frac{s}{s-1}-\frac{1}{c}} B\left(\frac{s}{s-1},\frac{1}{c}\right)\,{}_1F_1\left(\frac{1}{c},\frac{s}{s-1}+\frac{1}{c}+1;-\frac{1}{c}\right)+\right.$ $\left.+\,c^{-\frac{s-\frac{1}{2}}{s-1}-\frac{1}{2c}+1}\Gamma\left(\frac{2s-1}{s-1}\right)e^{-\frac{1}{2c}}\,W_{\frac{1}{2c}-\frac{s-\frac{1}{2}}{s-1},\,\frac{1}{2}-\frac{s-\frac{1}{2}}{s-1}-\frac{1}{2c}}\left(\frac{1}{c}\right)\right]^{s-1}$
exponential	$\frac{1}{\rho}\exp\left[-\frac{x}{\rho}\right]$	$e^{1-s}\left[\Gamma\left(\frac{2s-1}{s-1}\right)+(-1)^{\frac{s}{s-1}}\gamma\left(\frac{s}{s-1},-1\right)\right]^{s-1}$
gamma	$\frac{1}{\rho^{\mu}\,\Gamma(\mu)}\,x^{\mu-1}\exp\left[-\frac{x}{\rho}\right]$	$\Gamma^{1-s}(\mu)\left[\mu^{\frac{s}{s-1}+\mu} B\left(\frac{s}{s-1},\mu\right)\,{}_1F_1\left(\mu,\frac{s}{s-1}+\mu+1;-\mu\right)+\right.$ $\left.+\,\mu^{\frac{s-\frac{1}{2}}{s-1}+\frac{\mu}{2}-1}\Gamma\left(\frac{2s-1}{s-1}\right)e^{-\frac{\mu}{2}}\,W_{\frac{\mu}{2}-\frac{s-\frac{1}{2}}{s-1},\,\frac{1}{2}-\frac{s-\frac{1}{2}}{s-1}-\frac{\mu}{2}}(\mu)\right]^{s-1}$

TABLE 2. The quantity $I_s^*(\xi)$ for some distributions.

$$\int_{-\infty}^{\infty} h_i(x)\ p(x)\ dx = C_i \qquad i = 1, 2, \ldots, k. \tag{3.4.2}$$

Since the Fisher information of order s is related to the s.a.c.m. of distributions, we shall consider the density function p(x) which maximizes $H(\xi)$ under the constraint

$$\int_{-\infty}^{\infty} |x-\mu|^s\ p(x)\ dx = m_s \tag{3.4.3}$$

where μ is the expectation of ξ.

THEOREM 3.4.1

Let Λ be the class of density functions which for some $s > 0$ satisfy the conditions

$$p(x) > 0 \qquad -\infty < x < \infty \tag{3.4.4a}$$

$$\int_{-\infty}^{\infty} |x-\mu|^s\ p(x)\ dx = m_s < \infty\ . \tag{3.4.4b}$$

Then the Shannon entropy $H(\xi)$ is maximal iff p(x) is the exponential power distribution of order s

$$p(x) = \frac{s^{\frac{s-1}{s}}}{2\ \Gamma(1/s)\ \beta^{1/s}}\ \exp\left\{-\ \frac{|x-\mu|^s}{s\beta}\right\} \qquad s > 0, \tag{3.4.5}$$

where $\beta = m_s$. In that case

$$H(\xi) = \frac{1}{s}\log \frac{2^s\ e\ \Gamma^s(1/s)\ \beta}{s^{s-1}}\ . \tag{3.4.6}$$

Proof

The (directed) divergence between two density functions $p(x)$ and $q(x)$ is defined as

$$D(p,q) = \int_X q(x) \log \frac{q(x)}{p(x)} dx. \tag{3.4.7}$$

It can be shown, see Kullback [35], that

$$D(p,q) \geq 0 \tag{3.4.8}$$

with equality iff $p(x) = q(x)$ a.e. x. From Eqs. 3.4.7 and 3.4.8 we obtain the inequality

$$-\int_X q(x) \log q(x)\, dx \leq -\int_X q(x) \log p(x)\, dx. \tag{3.4.9}$$

For the density function $p(x)$ we choose the exponential power distribution, as given in Eq. 3.4.5. The density function $q(x)$ is any member of Λ. Then the inequality becomes, after some algebraic manipulations

$$-\int_{-\infty}^{\infty} q(x) \log q(x)\, dx \leq \frac{1}{s} \log \frac{2^s\, e\, \Gamma^s(1/s)\, \beta}{s^{s-1}}. \tag{3.4.10}$$

The right-hand side of Eq. 3.4.10 is finite for $s > 0$. It thus follows that the Shannon entropy is bounded from above for any $q(x) \in \Lambda$. The upper bound is attained iff $q(x) = p(x)$ a.e. x, so iff $q(x)$ is also an exponential power distribution of order s. In that case Eq. 3.4.6 follows at once from Eq. 3.4.10. □

It follows from the theorem that the exponential power distribution is characterized by the property that it has a maximal Shannon entropy given the s-th absolute central moment. We are now able to introduce the entropy moment of order s.

DEFINITION 3.4.1

The entropy moment of order s of a random variable ξ with density function p(x) having the Shannon entropy $H(\xi)$ is defined for $s > 0$ by

$$M_s(\xi) = \frac{s^{s-1}}{e\, 2^s\, \Gamma^s(1/s)} \exp\{s\, H(\xi)\}. \tag{3.4.11}$$

Note that the case s = 2 is the familiar entropy power, as introduced by Shannon [50], defined as

$$M_2(\xi) = \frac{1}{2\pi e} \exp\{2H(\xi)\}. \tag{3.4.12}$$

The entropy power is related to the variance of ξ. For the entropy power of order s we have the following property.

THEOREM 3.4.2

Let ξ be a random variable with given s.a.c.m. m_s. Then for $s > 0$

$$M_s(\xi) \le m_s. \tag{3.4.13}$$

Equality in Eq. 3.4.13 holds iff ξ has an exponential power distribution $N_s(\mu,\beta)$, in which case $\beta = m_s$.

Proof

From Theorem 3.4.1 we have $H(\xi) < H_{max}$ unless p(x) satisfies Eq. 3.4.5, in which case we have equality. Using Definition 3.4.1 proves the theorem. □

The entropy moment of order s can be seen as the s.a.c.m. of an exponential power distribution which has the same entropy as the distribution p(x) of ξ. Since the exponential power distribution has maximum entropy for a given s.a.c.m., the entropy moment of order s is less than or equal to its actual s-th absolute central moment as is stated in Theorem 3.4.2. Some examples of $M_s(\xi)$ are given in Table 3.

name	$p(x)$	$M_s(\xi)$
exponential power $N_\nu(\mu,\beta)$	$\dfrac{\nu^{\frac{\nu-1}{\nu}}}{2\ \Gamma(1/\nu)\ \beta^{1/\nu}} \exp\left[-\dfrac{\lvert x-\mu\rvert^\nu}{\nu\beta}\right]$	$e^{\frac{\nu}{s}-1}\ \nu^{\nu-1}\ s^{\frac{\nu}{s}-\nu}\ \dfrac{\Gamma^\nu(1/s)}{\Gamma^\nu(1/\nu)}\ \beta^{\frac{\nu}{s}}$
exponential power $N_s(\mu,\beta)$	$\dfrac{s^{\frac{s-1}{s}}}{2\ \Gamma(1/s)\ \beta^{1/s}} \exp\left[-\dfrac{\lvert x-\mu\rvert^s}{s\beta}\right]$	β
gaussian	$\dfrac{1}{\sigma\sqrt{2\pi}} \exp\left[-\dfrac{(x-\mu)^2}{2\sigma^2}\right]$	$2\ \pi^{-1}\ e^{\frac{2}{s}-1}\ s^{\frac{2}{s}-2}\ \Gamma^2(1/s)\ \sigma^{4/s}$
Cauchy	$\dfrac{1}{\pi\lambda}\ \dfrac{1}{1+\left(\frac{x-\mu}{\lambda}\right)^2}$	$\dfrac{s^{s-1}\ e^{s-1}\ \pi^s\ \lambda^s}{2^s\ \Gamma^s(1/s)}$
beta	$\dfrac{1}{B(a,b)}\ x^{a-1}\ (1-x)^{b-1}$	$\dfrac{s^{s-1}}{e\ 2^s\ \Gamma^s(1/s)}\ B^s(a,b)\ e^{s\alpha}$ where $\alpha = (a+b-2)\ \psi(a+b) - (a-1)\ \psi(a) + - (b-1)\ \psi(b)$ *)
Weibull	$\dfrac{c}{\rho^c}\ x^{c-1} \exp\left[-\left(\dfrac{x}{\rho}\right)^c\right]$	$\dfrac{s^{s-1}\ \rho^s}{e\ 2^s\ \Gamma^s(1/s)\ c^s\ \gamma^{s/a}}$ **)
Rayleigh	$\dfrac{x}{\rho^2} \exp\left[-\dfrac{x^2}{2\rho^2}\right]$	$\dfrac{s^{s-1}\ \rho^s\ \gamma^{\frac{s}{2}}}{e^{1-s}\ 2^{\frac{3s}{2}}\ \Gamma^s(1/s)}$ **)
gamma	$\dfrac{1}{\rho^\mu\ \Gamma(\mu)}\ x^{\mu-1} \exp\left[-\dfrac{x}{\rho}\right]$	$\dfrac{\rho^s\ s^{s-1}}{e 2^s}\ \dfrac{\Gamma^s(\mu)}{\Gamma^s(1/s)} \exp\{s\mu - s(\mu-1)\ \psi(\mu)\}$ *)
Maxwell	$\dfrac{\sqrt{2}}{\rho^3\sqrt{\pi}}\ x^2 \exp\left[-\dfrac{x^2}{2\rho^2}\right]$	$\dfrac{s^{s-1}\ \rho^s\ \pi^{s/2}}{e^{1+\frac{s}{2}}\ 2^{\frac{s}{2}}\ \Gamma^s(1/s)}$
lognormal	$\dfrac{1}{x\sigma\sqrt{2\pi}} \exp\left[-\dfrac{(\log x-\mu)^2}{2\sigma^2}\right]$	$\dfrac{s^{s-1}\ e^{\frac{s}{2}-1}\ \pi^{\frac{s}{2}}}{2^{s/2}\ \Gamma^s(1/s)}\ \sigma^s\ e^{s\mu}$

*) $\psi(.)$ is Euler's psi-function, see Appendix B.

**) $\gamma = e^C$, where C is Euler's constant, see Appendix B.

TABLE 3. $M_s(\xi)$ for some distributions.

In the next two theorems we shall study $M_s(\xi)$ for location and scale parameters. To emphasize its dependence on the parameter θ we shall use the notation $M_s(\xi;\theta)$.

THEOREM 3.4.3

Let θ be a location parameter for the distribution of the random variable ξ. Then

$$M_s(\xi;\theta) = M_s(\xi;0). \tag{3.4.14}$$

Proof

We have $p(x;\theta) = q(x-\theta)$ from which it follows that

$$\begin{aligned} H(\xi;\theta) &= -\int_{-\infty}^{\infty} q(x-\theta) \log q(x-\theta)\, dx \\ &= -\int_{-\infty}^{\infty} q(y) \log q(y)\, dy \\ &= H(\xi;0). \end{aligned} \tag{3.4.15}$$

Substitution of Eq. 3.4.15 into Eq. 3.4.11 completes the proof. □

THEOREM 3.4.4

Let θ be a scale parameter for the distribution of the random variable ξ. Then

$$M_s(\xi;\theta) = \theta^s M_s(\xi;1). \tag{3.4.16}$$

Proof

By the definition of a scale parameter we have $p(x;\theta) = 1/\theta\ q(x/\theta)$. From this we obtain

$$H(\xi;\theta) = -\int_{-\infty}^{\infty} \frac{1}{\theta}\, q(\frac{x}{\theta})\ \log\left[\frac{1}{\theta}\, q(\frac{x}{\theta})\right] dx$$

$$= -\int_{-\infty}^{\infty} q(y)\ \log\left[\frac{1}{\theta}\, q(y)\right] dy$$

$$= \log\theta - \int_{-\infty}^{\infty} q(y)\ \log q(y)\ dy$$

$$= \log\theta + H(\xi;1). \tag{3.4.17}$$

Substitution of Eq. 3.4.17 into Eq. 3.4.11 yields Eq. 3.4.16. □

The quantities $M_s(\xi;\theta)$ and $F_s^{-1}(\xi;\theta)$ show much resemblance:

1. If θ is a location parameter, then $M_s(\xi;\theta)$ and $F_s^{-1}(\xi;\theta)$ do not depend on θ.
2. If θ is a scale parameter, then $M_s(\xi;\theta)$ and $F_s^{-1}(\xi;\theta)$ only depend on θ through a factor θ^s.
3. For the exponential power distribution $N_s(\mu,\beta)$, we have

$$M_s(\xi;0) = F_s^{-1}(\xi;0) = \beta. \tag{3.4.18}$$

4. If ξ has an s.a.c.m. m_s, we have

$$M_s(\xi;0) \le m_s \tag{3.4.19}$$

$$F_s^{-1}(\xi;0) \le m_s\ . \tag{3.4.20}$$

Equality in Eqs. 3.4.19 and 3.4.20 is obtained iff ξ has an exponential power distribution $N_s(\mu,\beta)$. For $1 \le s < 2$ equality in Eq. 3.4.20 only holds under an if condition, as follows from Theorem 2.2.4.

Thus we see that the entropy moment of order s and the inverse of the Fisher information of order s behave in a similar way for location and scale parameters. They are both less than the s.a.c.m. of a random variable. The fact that they are maximal as well as equal if the random variable ξ has an exponential power distribution $N_s(\mu,\beta)$, is of particular interest. It shows that this distribution can be associated with two concepts. The first follows from $M_s(\xi;\theta)$ and can be formulated as a maximum uncertainty if only the s.a.c.m. of a distribution is known. The second follows from $F_s^{-1}(\xi;\theta)$ and can be expressed as a worst case situation in terms of the minimum accuracy of estimators which can be obtained.

CHAPTER 4

FURTHER RESULTS AND APPLICATIONS

The purpose of this chapter is to give some further results for the Fisher information of order s and to give some applications to the estimation of signal parameters.

First we will consider the relation of the Fisher information of order s to some probabilistic distance measures. To this end we use the f-divergence between two density functions with parameters θ and $\theta + \Delta\theta$. We establish a relation which shows that these probabilistic distance measures are related to the Fisher information of order s in a simple way.

Secondly we consider the Fisher information of order s when the likelihood function is a mixture distribution. Then the computation of the Fisher information of order s will in general be complicated. We give upper and lower bounds for the Fisher information of order s. We discuss their application and illustrate them by two examples.

Finally we consider the estimation of signal parameters. First we consider the estimation of the average power of a certain random signal, as an example of scale parameter estimation. Secondly we consider the estimation of a signal parameter of an otherwise completely specified signal in additive, non-gaussian noise. This leads to location parameter estimation. We also discuss some aspects of the density functions which are used in this section.

4.1 SOME GEOMETRIC ASPECTS

In this section we shall study some geometric aspects of $F_s(\xi;\theta)$ and a related information measure $F_s'(\xi;\theta)$ which was obtained in Section 2.5. This allows us to use some familiar concepts as a tool for understanding these two information measures. Since the geometric considerations are essentially based on probabilistic distance measures, or measures of divergence, we shall consider a certain class of these measures and show that it is related to the measures $F_s(\xi;\theta)$ and to $F_s'(\xi;\theta)$. This approach is an extension of work reported by Kagan [29] and Aggarwal [1] for the case s = 2 and of the χ^{α}-divergence which was introduced by Vajda [55].

A probabilistic distance measure is a measure for the distance between two distributions or density functions. It is non-negative, and zero iff the two distributions coincide. A probabilistic distance measure is generally not a metric, since some widely used measures do not satisfy the triangle inequality. An example of such a distance measure is the Bhattacharyya distance, which for two density functions p(x) and q(x) is defined as

$$\rho_B = - \log \int_X \sqrt{p(x)\ q(x)}\ dx. \tag{4.1.1}$$

Another example is the directed divergence, which was used in the proof of Theorem 3.4.1. It is given, under certain conditions, as

$$D(p,q) = \int_X q(x) \log \frac{q(x)}{p(x)}\ dx. \tag{4.1.2}$$

We also mention the Matusita distance, which is given by

$$D_M(p,q) = \left[\int_X \left| \sqrt{p(x)} - \sqrt{q(x)} \right|^2 dx \right]^{1/2}. \tag{4.1.3}$$

We shall not go into detail about these probabilistic distance measures, but mention only that some of them have been applied to the signal selection problem in communication theory and to the feature selection problem in pattern recognition. A very general probabilistic distance measure is Csiszár's f-divergence, see Csiszár [17], which includes many other distance measures. It is defined as follows.

DEFINITION 4.1.1

Let the function $f(u)$ be a continuous, convex function defined on $(0,\infty)$ and let it satisfy

(i) $$f(0) = \lim_{u \downarrow 0} f(u) \tag{4.1.4a}$$

(ii) $$0 \,.\, f(\tfrac{0}{0}) = 0 \tag{4.1.4b}$$

(iii) $$0 \,.\, f(\tfrac{a}{0}) = \lim_{\varepsilon \downarrow 0} \varepsilon \,.\, f(\tfrac{a}{\varepsilon}) = a \lim_{u \to \infty} \frac{f(u)}{a} \qquad 0 < a < \infty \,.\, . \tag{4.1.4c}$$

Then the f-divergence of two density functions $p(x)$ and $q(x)$ is defined by

$$D_f(p,q) = \int_X f\left(\frac{p(x)}{q(x)}\right) q(x)\, dx. \tag{4.1.5}$$

It is easy to show, Csiszár [17], that

$$D_f(p,q) \geq f(1), \tag{4.1.6}$$

Equality in Eq. 4.1.6 holds iff $p(x) = q(x)$ a.e. x, provided that $f(u)$ is strictly convex at $u_0 = 1$. It follows that the quantity $D_f(p,q) - f(1)$ can be interpreted as a measure of the affinity, or distance, of $p(x)$ and $q(x)$, since this quantity is non-negative, and zero iff $p(x) = q(x)$ under the conditions mentioned above.

In this thesis we are concerned with the information about a parameter θ. To apply the concept of f-divergence we consider the density functions $p(x;\theta)$ and $p(x;\theta+\Delta\theta)$. This permits us to measure the rate of change of $p(x;\theta)$ as a result of a change $\Delta\theta$ in θ and also to give a geometrical interpretation of $F_s(\xi;\theta)$. To this end we give the next definition as a modification of Definition 4.1.1.

DEFINITION 4.1.2

Let the function $f(u)$ be a continuous convex function defined on $(0,\infty)$ which satisfies the conditions (i) - (iii). Let the density functions $p(x;\theta)$ and $p(x;\theta+\Delta\theta)$ be positive for every $x \in X$, $\theta \in \Theta$. Then for θ, $\theta + \Delta\theta \in \Theta$ the f-divergence with respect to θ is defined as

$$D_f(\theta+\Delta\theta,\theta) = \int_X f\left(\frac{p(x;\theta+\Delta\theta)}{p(x;\theta)}\right) p(x;\theta)\, dx. \tag{4.1.7}$$

The f-divergence with respect to θ is the basis for a definition of an information measure which can be related to the Fisher information of order s. We define it as a function of $D_f(\theta+\Delta\theta,\theta)$.

DEFINITION 4.1.3

The D_f-information of order s with respect to θ, $\theta \in \Theta$ is defined as

$$D_{f,s}(\theta) = \lim_{\Delta\theta\to 0} \frac{1}{|\Delta\theta|^s} \left[D_f(\theta+\Delta\theta,\theta) - f(1)\right]^{s-1}, \tag{4.1.8}$$

provided the limit exists.

As a consequence of Eqs. 4.1.6 and 4.1.7 it follows that $D_{f,s}(\theta)$ is non-negative. The information measure $D_{f,s}(\theta)$ also has the property that any transformation can only result in a loss of information, as stated in the next theorem.

THEOREM 4.1.1

Let the family of density functions $\{P_\theta\} = \{p(x;\theta),\ \theta \in \Theta\}$ of the random variable ξ be defined on the probability space $(\mathcal{X}, \mathcal{A}, \{P_\theta\})$. Let $T = T(\xi)$ be a measurable transformation of ξ and let $\{Q_\theta\}$ be the family of density functions induced by the transformation T. We denote the information measures associated with ξ and $T(\xi)$ by $D_{f,s}(\xi;\theta)$ and $D_{f,s}(T(\xi);\theta)$ respectively. Then $D_{f,s}(T(\xi);\theta)$ exists if $D_{f,s}(\xi;\theta)$ exists and for $s \geq 1$ we have

$$D_{f,s}(T(\xi);\theta) \leq D_{f,s}(\xi;\theta) . \tag{4.1.9}$$

Equality in Eq. 4.1.9 holds iff (for $s > 1$) or if (for $s = 1$) T is a sufficient statistic for the family $\{P_\theta\}$, provided the function $f(.)$ is a strictly convex function.

Proof

The proof is a direct consequence of the corresponding property of the f-divergence, see Csiszár [17], and will be omitted here. □

The information measure $D_{f,s}(\theta)$ is a generalization of Aggarwal's measure [1], which was developed for the case $s = 2$.

We shall not discuss other properties of $D_{f,s}(\theta)$ which are comparable with those of $F_s(\xi;\theta)$ but consider a useful relation which, under certain conditions, exists between $D_{f,s}(\theta)$ and $F_s(\xi;\theta)$. This relation is an important motivation for the definition of $D_{f,s}(\theta)$. In the remainder of this section we shall use the notation $F_s(\theta)$ for $F_s(\xi;\theta)$.

THEOREM 4.1.2

Let $p(x;\theta)$ be a positive and $(r+1)$-times continuously differentiable density function for all $x \in X$, $\theta \in \Theta$, where $r = \frac{s}{s-1} = 2, 3, 4, \ldots$. Let S be the set $\{2, 3/2, 4/3, \ldots\}$. Let $f(u)$ be a convex function defined on $(0,\infty)$, which is $(r+1)$-times differentiable around the point $u_0 = 1$ and satisfies the conditions

(i) $f^{(N)}(1) = 0, \quad N = 1, 2, \ldots, (r-1)$, where $f^{(N)}(1)$ is the N-th derivative in $u_0 = 1$, (4.1.10a)

(ii) $\operatorname{sgn}[f^{(r)}(1)] = \operatorname{sgn}[(u-1)^r]$. (4.1.10b)

Then for $s \in S$ it holds that

$$D_{f,s}(\theta) = \left[\frac{\left| f^{(\frac{s}{s-1})}(1) \right|}{\Gamma(\frac{2s-1}{s-1})} \right]^{s-1} . F_s(\theta). \tag{4.1.11}$$

Proof

The function $f(u)$ can be expanded in a Taylor series around the point $u_0 = 1$, to obtain

$$f(u) = f(1) + (u-1)\, f^{(1)}(1) + \frac{(u-1)^2}{\Gamma(3)} f^{(2)}(1) + \ldots$$

$$+ \frac{(u-1)^r}{\Gamma(r+1)} f^{(r)}(1) + R_{r+1}(u), \tag{4.1.12}$$

where the Lagrange remainder $R_{r+1}(u)$ is given by

$$R_{r+1}(u) = \frac{(u-1)^{r+1}}{\Gamma(r+2)} f^{(r+1)}(1+k_1(u-1)) \qquad 0 < k_1 < 1. \tag{4.1.13}$$

Using condition (i) and (ii) in Eq. 4.1.12 yields

$$f(u) = f(1) + \frac{|u-1|^r}{\Gamma(r+1)} . \left| f^{(r)}(1) \right| + R_{r+1}(u). \tag{4.1.14}$$

In a similar way the density function $p(x;\theta+\Delta\theta)$ can be expanded in a Taylor series around $p(x;\theta)$. This gives

$$p(x;\theta+\Delta\theta) = p(x;\theta) + \Delta\theta \frac{\partial}{\partial\theta} p(x;\theta) + R_2(\Delta\theta). \tag{4.1.15}$$

The Lagrange remainder $R_2(\Delta\theta)$ is given by

$$R_2(\Delta\theta) = \frac{(\Delta\theta)^2}{2} \frac{\partial^2}{\partial\theta^2} p(x;\theta+k_2\Delta\theta) \qquad 0 < k_2 < 1, \tag{4.1.16}$$

and is of the order $\mathcal{O}(\Delta\theta)^2$. From Eq. 4.1.15 it follows that

$$\frac{p(x;\theta+\Delta\theta)}{p(x;\theta)} = 1 + \frac{\Delta\theta}{p(x;\theta)} \frac{\partial}{\partial\theta} p(x;\theta) + \frac{\mathcal{O}(\Delta\theta)^2}{p(x;\theta)} . \tag{4.1.17}$$

If we set

$$u = \frac{p(x;\theta+\Delta\theta)}{p(x;\theta)} \tag{4.1.18}$$

and use Eq. 4.1.14, we arrive at the following expression for the f-divergence $D_f(\theta+\Delta\theta,\theta)$ with respect to θ

$$\begin{aligned} D_f(\theta+\Delta\theta,\theta) &= \int_X f\left(\frac{p(x;\theta+\Delta\theta)}{p(x;\theta)}\right) p(x;\theta)\, dx \\ &= \int_X \left[f(1) + \frac{|f^{(r)}(1)|}{\Gamma(r+1)} \cdot \left|\frac{p(x;\theta+\Delta\theta)}{p(x;\theta)} - 1\right|^r + \right. \\ &\qquad \left. + R_{r+1}(\Delta\theta) \right] p(x;\theta)\, dx. \end{aligned} \tag{4.1.19}$$

The Lagrange remainder $R_{r+1}(\Delta\theta)$ is given by

$$\begin{aligned} R_{r+1}(\Delta\theta) &= \left[\frac{p(x;\theta+\Delta\theta) - p(x;\theta)}{p(x;\theta)}\right]^{r+1} \cdot \frac{1}{\Gamma(r+2)} \cdot \\ &\quad \cdot f^{(r+1)}\left(1 + k_1 \frac{p(x;\theta+\Delta\theta) - p(x;\theta)}{p(x;\theta)}\right) \qquad 0 < k_1 < 1, \end{aligned} \tag{4.1.20}$$

and is of the order $\mathcal{O}(\Delta\theta)^{r+1}$. Substitution of Eq. 4.1.17 into Eq. 4.1.19 yields

$$D_f(\theta+\Delta\theta,\theta) = f(1) + \frac{|f^{(r)}(1)|}{\Gamma(r+1)} \cdot \int_X \left[\left| \frac{\Delta\theta}{p(x;\theta)} \frac{\partial}{\partial\theta} p(x;\theta) + \frac{o(\Delta\theta)^2}{p(x;\theta)} \right|^r \right] p(x;\theta)\, dx + o(\Delta\theta)^{r+1}. \qquad (4.1.21)$$

Dividing both sides of Eq. 4.1.21 by $|\Delta\theta|^r$ leads to

$$\frac{1}{|\Delta\theta|^r} \left[D_f(\theta+\Delta\theta,\theta) - f(1) \right] = \frac{|f^{(r)}(1)|}{\Gamma(r+1)} \cdot \int_X \left[\left| \frac{1}{p(x;\theta)} \frac{\partial}{\partial\theta} p(x;\theta) + \frac{o(\Delta\theta)}{p(x;\theta)} \right|^r \right] p(x;\theta)\, dx + o(|\Delta\theta|)$$

$$= \frac{|f^{(r)}(1)|}{\Gamma(r+1)} \cdot \int_X \left| \frac{1}{p(x;\theta)} \frac{\partial}{\partial\theta} p(x;\theta) \right|^r p(x;\theta)\, dx + o(|\Delta\theta|). \qquad (4.1.22)$$

From Eq. 4.1.22 we obtain by taking the limit for $\Delta\theta \to 0$ and using Definition 4.1.3 that

$$D_{f,s}(\theta) = \lim_{\Delta\theta\to 0} \left[\frac{|f^{(r)}(1)|}{\Gamma(r+1)} \cdot \int_X \left| \frac{1}{p(x;\theta)} \frac{\partial}{\partial\theta} p(x;\theta) \right|^r \cdot p(x;\theta)\, dx + o(|\Delta\theta|) \right]^{s-1}$$

$$= \left[\frac{|f^{(r)}(1)|}{\Gamma(r+1)} \right]^{s-1} \cdot \left[\int_X \left| \frac{1}{p(x;\theta)} \frac{\partial}{\partial\theta} p(x;\theta) \right|^r p(x;\theta)\, dx \right]^{s-1}. \qquad (4.1.23)$$

Noting that $r = s/(s-1)$ and using the definition of $F_s(\theta)$ finally leads to

$$D_{f,s}(\theta) = \left[\frac{\left| f^{(\frac{s}{s-1})}(1) \right|}{\Gamma(\frac{2s-1}{s-1})} \right]^{s-1} . F_s(\theta), \tag{4.1.24}$$

which completes the proof. □

Theorem 4.1.2 shows a close relation between $D_{f,s}(\theta)$ and $F_s(\theta)$ for $s \in S$, provided some conditions are satisfied. In particular it shows that a whole class of information measures, based on different choices of the function $f(u)$, are proportional to $F_s(\theta)$. Thus the information measure $D_{f,s}(\theta)$ can be seen as a generalized Fisher information measure.

It is of interest to consider $D_{f,s}(\theta)$ for some choices of the convex function $f(u)$. This leads to some special cases of the f-divergence. We shall also show how they are related to the Fisher information of order s and to some other distance measures and information measures which have been proposed.

EXAMPLE 4.1.1

As a first choice of the function $f(u)$ we will consider the function

$$f_1(u) = |1-u|^{\frac{s}{s-1}} \qquad s \geq 1. \tag{4.1.25}$$

The corresponding f-divergence with respect to θ is given by

$$D_{f_1}(\theta+\Delta\theta,\theta) = \int_X \left| 1 - \frac{p(x;\theta+\Delta\theta}{p(x;\theta)} \right|^{\frac{s}{s-1}} p(x;\theta)\, dx \qquad s \geq 1. \tag{4.1.26}$$

For this particular choice of $f(u)$, we find after some algebraic manipulations for the s/(s-1)-th derivative of $f(u)$, where $s \in S$, that

$$\left| f_1^{\left(\frac{s}{s-1}\right)}(1) \right| = \Gamma\left(\frac{2s-1}{s-1}\right). \qquad (4.1.27)$$

Substitution of Eq. 4.1.27 into Eq. 4.1.11 gives as a final result

$$D_{f_1,s}(\theta) = F_s(\theta) \qquad s \in S. \qquad (4.1.28)$$

It follows that, for this choice of $f(u)$, the D_f-information of order s is equal to the Fisher information of order s.

A related result was obtained by Vajda [55]. He introduced the χ^α-divergence, which in our notation is equal to $D_{f_1}(\theta+\Delta\theta,\theta)$ for $\alpha = s/(s-1)$. Based on this χ^α-divergence he defined a Fisher information of order α, $\alpha \geq 1$, as

$$F_\alpha(\theta) = \lim_{\Delta\theta\to 0} \frac{1}{|\Delta\theta|^\alpha} D_{f_1}(\theta+\Delta\theta,\theta), \qquad (4.1.29)$$

Vajda also studied properties of his information measure $F_\alpha(\theta)$ which shows a relation to some of the results obtained for $F_s(\theta)$ in Chapter 2 of this thesis.

Bij setting $s = 2$ in Eq. 4.1.28 we see that $D_{f_1,2}(\theta)$ is equal to the usual Fisher information $F_2(\theta)$. This case has been studied by Aggarwal [1], Kagan [29] and Chapman and Robbins [12] who considered several aspects of $D_{f_1,2}(\theta)$ and $F_2(\theta)$.

EXAMPLE 4.1.2

A second choice of the function $f(u)$ which will be considered is

$$f_2(u) = \left|1-u^{\frac{s-1}{s}}\right|^{\frac{s}{s-1}} \qquad s \geq 1. \qquad (4.1.30)$$

This leads to a special case of the f-divergence with respect to θ, which is given by

$$D_{f_2}(\theta+\Delta\theta,\theta) = \int_X \left| 1 - \left(\frac{p(x;\theta+\Delta\theta)}{p(x;\theta)}\right)^{\frac{s-1}{s}} \right|^{\frac{s}{s-1}} p(x;\theta)\, dx$$

$$= \int_X \left| p(x;\theta)^{\frac{s-1}{s}} - p(x;\theta+\Delta\theta)^{\frac{s-1}{s}} \right|^{\frac{s}{s-1}} dx. \qquad (4.1.31)$$

For this choice of $f_2(u)$ it follows that

$$\left| f_2^{\left(\frac{s}{s-1}\right)}(1) \right| = \left(\frac{s}{s-1}\right)^{\frac{s}{1-s}} \Gamma\left(\frac{2s-1}{s-1}\right). \qquad (4.1.32)$$

Substitution of Eq. 4.1.32 into Eq. 4.1.11 gives the relation

$$D_{f_2,s}(\theta) = \left(\frac{s-1}{s}\right)^s F_s(\theta). \qquad (4.1.33)$$

If we set s = 2 we have

$$f_2(u) = |1 - \sqrt{u}|^2, \qquad (4.1.34)$$

which leads to a special form of the f-divergence

$$D_{f_2}(p,q) = \int_X \left| 1 - \sqrt{\frac{p(x)}{q(x)}} \right|^2 q(x)\, dx$$

$$= \int_X |\sqrt{p(x)} - \sqrt{q(x)}|^2\, dx. \qquad (4.1.35)$$

The quantity $D_{f_2}(p,q)$ of Eq. 4.1.35 is closely related to the Matusita distance $D_M(p,q)$ of Eq. 4.1.3. From Theorem 4.1.2 it follows that the Matusita distance can be related to the Fisher information. We have

$$D_M(\theta) = \frac{1}{2}\sqrt{F_2(\theta)}. \tag{4.1.36}$$

Thus the f-divergence which is based on the function $f_2(u)$ of Eq. 4.1.30 can be seen as a generalization of the Matusita distance, and Eq. 4.1.33 relates this distance measure to the Fisher information of order s. From Eqs. 4.1.8 and 4.1.31 it follows that for $s \in S$, and for $p(x;\theta)$ a differentiable function of θ, that

$$D_{f_2,s}(\theta) = \lim_{\Delta\theta\to 0}\left[\int_X \left|\frac{p(x;\theta)^{\frac{s-1}{s}} - p(x;\theta+\Delta\theta)^{\frac{s-1}{s}}}{\Delta\theta}\right|^{\frac{s}{s-1}} dx\right]^{s-1}$$

$$= \left[\int_X \left|\frac{\partial}{\partial\theta}\, p(x;\theta)^{\frac{s-1}{s}}\right|^{\frac{s}{s-1}} dx\right]^{s-1}$$

$$= F_s{}'(\theta). \tag{4.1.37}$$

Here $F_s{}'(\theta)$ is the information measure which was introduced in Section 2.5, Definition 2.5.1.

Thus we see that under certain conditions the information measure $D_{f_2,s}(\theta)$ equals $F_s{}'(\theta)$, in which case they have the same relation to the quantity $F_s(\theta)$.

EXAMPLE 4.1.3

In the final example we consider a function $f(u)$ of the form

$$f_3(u) = \left|\frac{1-u}{1+u}\right|^{\frac{s}{s-1}}, \qquad s \geq 1 \tag{4.1.38}$$

Then the f-divergence with respect to θ is given by

$$D_{f_3}(\theta+\Delta\theta;\theta) = \int_X \left|\frac{p(x;\theta) - p(x;\theta+\Delta\theta)}{p(x;\theta) + p(x;\theta+\Delta\theta)}\right|^{\frac{s}{s-1}} p(x;\theta)\, dx. \tag{4.1.39}$$

After some algebraic manipulations we find for this case that

$$\left|f_3^{(\frac{s}{s-1})}(1)\right| = 2^{-\frac{s}{s-1}}\, \Gamma(\frac{2s-1}{s-1}). \tag{4.1.40}$$

From this result we finally obtain

$$D_{f_3,s}(\theta) = 2^{-s} F_s(\theta). \tag{4.1.41}$$

The case s = 2 has been given by Aggarwal [1].

A comparison of the results obtained for $D_{f_1,s}(\theta)$ and $D_{f_3,s}(\theta)$ reveals that $D_{f_3,s}(\theta) \leq D_{f_1,s}(\theta)$. This can also be deducted from Eqs. 4.1.25 and 4.1.38 since for $u \geq 0$

$$\left|\frac{1-u}{1+u}\right| \leq |1-u|. \tag{4.1.42}$$

In a similar way we can see that $D_{f_2,s}(\theta) \leq D_{f_1,s}(\theta)$.

As will be clear from these examples we have obtained an information measure $D_{f,s}(\theta)$ which is proportional to the Fisher information of order s. It is based on the concept of divergence, or distance, between the density functions $p(x;\theta)$ and $p(x;\theta+\Delta\theta)$. The use of $D_{f,s}(\theta)$ can be of importance if θ is a set of discrete parameter points $\theta_1, \ldots, \theta_k$. In this case we could omit the limit operation from Definition 4.1.3 of

$D_{f,s}(\theta)$.

It has been shown that the information measure $D_{f,s}(\theta)$ includes some well-known probabilistic distance measures, as well as some new ones. An interesting conclusion is that with respect to the parameter θ these measures all result in the Fisher information of order s, up to a constant which depends only on the order parameter s.

Another way of looking at $F_s(\theta)$ is to consider an infinite dimensional space $\mathbb{R}^\infty$ with the metric

$$d(g(x),\ h(x)) = \left[\int_X \left| g(x)^{1/r} - h(x)^{1/r} \right|^r dx \right]^{1/r} \qquad r=1,2,3,\ldots , \tag{4.1.43}$$

for any two functions g(x) and h(x) defined over X.

Since for any density function p(x) it holds that $\int p(x)\, dx = 1$, each density function defines a point on the unit hypersphere. If we consider the density function $p(x;\theta)$ and let θ vary then this results in a curve υ on this hypersphere as is illustrated in Fig. 4.1.1.

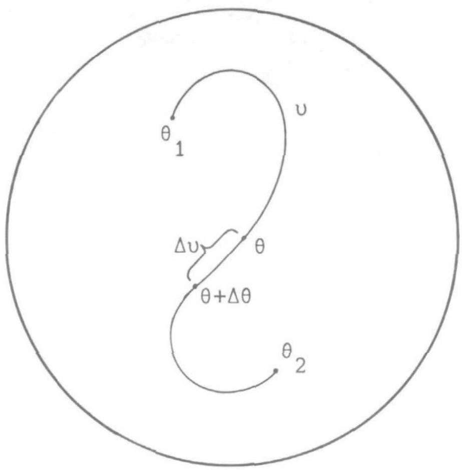

Fig. 4.1.1 Curve υ on hypersphere

Let a change of the parameter θ into $\theta + \Delta\theta$ result in a change $\Delta\upsilon$ along this curve. Then it follows, using the metric of Eq. 4.1.43, that

$$|\Delta\upsilon|^r = \int_X \left| p(x;\theta)^{1/r} - p(x;\theta+\Delta\theta)^{1/r} \right|^r dx. \tag{4.1.44}$$

This can also be written as

$$|\Delta\upsilon|^r = \int_X \left| 1 - \left(\frac{p(x;\theta+\Delta\theta)}{p(x;\theta)}\right)^{1/r} \right|^r p(x;\theta)\, dx. \tag{4.1.45}$$

The expression on the right-hand side is just the f-divergence with respect to θ $D_{f_2}(\theta+\Delta\theta,\theta)$, for $r = s/(s-1)$, which has been considered in Example 4.1.2 of this section. As has been shown in Theorem 4.1.2, this quantity can be associated with the Fisher information of order s. To this end we write for Eq. 4.1.45

$$\left|\frac{\Delta\upsilon}{\Delta\theta}\right|^s = \frac{1}{|\Delta\theta|^s} \left[\int_X \left| 1 - \left(\frac{p(x;\theta+\Delta\theta)}{p(x;\theta)}\right)^{\frac{s-1}{s}} \right|^{\frac{s}{s-1}} p(x;\theta)\, dx \right]^{s-1} . \tag{4.1.46}$$

Taking the limit for $\Delta\theta \to 0$ on both sides of Eq. 4.1.46 results in

$$\left|\frac{d\upsilon}{d\theta}\right|^s = D_{f_2,s}(\theta) = \left(\frac{s-1}{s}\right)^s F_s(\theta) = F_s'(\theta). \tag{4.1.47}$$

Thus it follows that the information measures $F_s(\theta)$ and $F_s'(\theta)$ are related to the change of the point on the unit hypersphere associated with the density function $p(x;\theta)$. Therefore, one may consider $F_s(\theta)$ to be a measure of the sensitivity of $p(x;\theta)$ for changes in θ through a simple geometric interpretation.

4.2 MIXTURE DISTRIBUTIONS

Quite often the density functions which are studied do not have a form which is easy to handle. Instead of performing rather complicated calculations one may want to simplify the problem by approximating such a density function by a finite mixture of density functions of a more useful form, like gaussian density functions. Such a density function is then called a *finite mixture* or *composite density function.*

In this section we will mainly use the exponential power distribution as an elementary density function. In a practical situation the mixing parameters have to be estimated, as well as the number of elementary density functions that have to be used. A good review of methods for estimating these parameters has been given by Macdonald [38]. For simplicity we shall assume that these parameters are known.

The finite mixture model can be used as an appropriate model for certain classes of noisy observations, e.g., impulse noise, and is of importance in such areas as communication theory and pattern recognition. Some applications are given by Friedrich [21] and Young et al. [60].

Another application has been given by Alspach [2] and Sorenson et al. [52]. They use a finite mixture method to simplify the analysis of certain nonlinear and non-gaussian filtering problems when either the a priori density function of the state of the system or the observation noise is non-gaussian. In this way they obtain a simplified form of the a posteriori density function of the state of the system.

Another approach has been suggested by Tukey [54]. He considers a density function which is contaminated by a second one and in this way obtains a class of non-gaussian density functions.

In this section we shall apply some of the results of Chapters 2 and 3 to obtain upper and lower bounds on $F_s(\xi;\theta)$ if the density function $p(x;\theta)$ is such a mixture of density functions. The lower bounds are given by the moments of the mixing density functions. The upper bounds are given by the information about θ contained in each of the mixing density functions. Throughout this section we shall assume that θ is a location parameter. This allows us to use the information measure $I_s(p)$ instead of $F_s(\xi;\theta)$. We shall use the notation $I_s(p)$ to emphasize its dependence on the density function $p(x)$.

Let $q_i(x;\theta_i)$, $i = 1, 2, \dots, n$ be positive and continuously differentiable density functions for every $x \in X$, $\theta \in \Theta$. We consider the density function

$$p(x;\theta) = \sum_{i=1}^{n} a_i\, q_i(x;\theta_i) \tag{4.2.1}$$

where the coëfficiënts a_i satisfy the conditions

$$0 \le a_i \le 1, \quad \sum_{i=1}^{n} a_i = 1 \tag{4.2.2}$$

and where θ_i, $i = 1, 2, \dots, n$ are location parameters. Then $p(x;\theta)$ is also a positive and continuously differentiable density function.

THEOREM 4.2.1

If the density function $p(x;\theta)$ is of the form of Eq. 4.2.1 and θ is a location parameter, we have

$$I_s(p) \le \sum_{i=1}^{n} a_i\, I_s(q_i) \tag{4.2.3}$$

where $I_s(p)$ and $I_s(q_i)$ are as given in Definition 3.1.1.

Proof

From Corollary 3.1.1 it follows for the density functions $p(x;\theta)$ and $q_i(x;\theta)$ that

$$F_s(p_\theta;\theta) = I_s(p) \tag{4.2.4}$$

$$F_s(q_{\theta_i};\theta) = I_s(q_i) \qquad i = 1, \dots, n. \tag{4.2.5}$$

If we next apply Corollary 2.2.1, we find the desired result. □

Since the upper bound results from the convexity of $I_s(p)$ in p, we shall refer to it as the convexity bound.

If $\theta_i = \theta$ for all $i = 1, \ldots, n$, in which case we have a set of contaminating density functions, we can also obtain a simple lower bound on $I_s(p)$.

THEOREM 4.2.2

If $p(x;\theta)$ satisfies Eq. 4.2.1 and if $\theta_i = \theta$, $i = 1, \ldots, n$, then $I_s(p)$ is lower bounded as follows

$$\left[\sum_{i=1}^{n} a_i\, m_{s_i}\right]^{-1} \le I_s(p). \tag{4.2.6}$$

Proof

From Corollary 3.1.2 it follows that

$$\frac{1}{m_s} \le I_s(p) \tag{4.2.7}$$

with equality if p(x) is the exponential power distribution $N_s(\mu, m_s)$. Furthermore, we have

$$m_s = \int_X |x-\theta|^s\, p(x)\, dx$$

$$= \int_X |x-\theta|^s \sum_{i=1}^{n} a_i\, q_i(x)\, dx. \tag{4.2.8}$$

By interchanging integration and summation we obtain

$$m_s = \sum_{i=1}^{n} a_i \int_X |x-\theta|^s\, q_i(x)\, dx$$

$$= \sum_{i=1}^{n} a_i \, m_{s_i}. \tag{4.2.9}$$

□

We shall refer to this lower bound as the moment bound. The use of the bounds from Theorems 4.2.1 and 4.2.2 will be illustrated by an example, which shows that one can obtain a simple bound on $I_s(p)$ in cases where the computation of $I_s(p)$ itself is quite complicated. For simplicity we shall only consider a mixture of two density functions $q_1(x)$ and $q_2(x)$ and shall use the notation $a_1 = a$ and $a_2 = 1 - a$. We then have the mixture

$$p(x) = a \, q_1(x) + (1-a) \, q_2(x) \tag{4.2.10}$$

for which we find the following corollary.

COROLLARY 4.2.1

If p(x) is given by Eq. 4.2.10, then $I_s(p)$ is bounded as follows

$$[a \, m_{s_1} + (1-a) \, m_{s_2}]^{-1} \le I_s(p) \le a \, I_s(q_1) + (1-a) \, I_s(q_2). \tag{4.2.11}$$

Note that the upper bound is valid for $\theta_i = \theta$ as well as for $\theta_i \neq \theta_j$, i,j = 1, 2, ... , n; whereas the lower bound is valid for $\theta_i = \theta$, i = 1, 2, ... , n only.

EXAMPLE 4.2.1

We shall consider a situation in which $q_1(x)$ and $q_2(x)$ are exponential power distributions. Let

$$q_1(x) = N_\nu(0,\gamma) \tag{4.2.12}$$

and

$$q_2(x) = N_s(0,\beta). \tag{4.2.13}$$

From Example 3.3.1 it follows that

$$I_s(q_1) = \gamma^{-\frac{s}{\nu}} \nu^{\frac{s(\nu-1)}{\nu}} \left[\frac{\Gamma(\frac{s\nu-1}{s\nu-\nu})}{\Gamma(1/\nu)} \right]^{s-1}, \tag{4.2.14}$$

and from Example 3.3.1a we find that

$$I_s(q_2) = \frac{1}{\beta}. \tag{4.2.15}$$

The s-th absolute central moment of $q_1(x)$ is given by

$$m_s(q_1) = \int_{-\infty}^{\infty} |x|^s \frac{\nu^{\frac{\nu-1}{\nu}}}{2\gamma^{1/\nu}\,\Gamma(1/\nu)} \exp\left[- \frac{|x|^\nu}{\nu\gamma} \right] dx. \tag{4.2.16}$$

Using Gradshteyn et al. [24] we find after some algebraic manipulations that

$$m_s(q_1) = (\nu\gamma)^{\frac{s}{\nu}} \frac{\Gamma(\frac{s+1}{\nu})}{\Gamma(1/\nu)} . \tag{4.2.17}$$

We find by substitution of $\nu = s$ and $\beta = \gamma$ into Eq. 4.2.17 that

$$m_s(q_2) = \beta. \tag{4.2.18}$$

Substitution of these results into the bound of Corollary 4.2.1 finally yields

$$\left[(1-a)\ \beta + a\ (\nu\gamma)^{\frac{s}{\nu}} \frac{\Gamma(\frac{s+1}{\nu})}{\Gamma(1/\nu)} \right]^{-1} \leq I_s(p)$$

$$\leq \frac{1-a}{\beta} + a\ \gamma^{-\frac{s}{\nu}} \nu^{\frac{s(\nu-1)}{\nu}} \left[\frac{\Gamma(\frac{s\nu-1}{s\nu-\nu})}{\Gamma(1/\nu)} \right]^{s-1} . \tag{4.2.19}$$

Using expression 4.2.19 we shall consider two special cases in more detail.

First we will consider the case in which

$$q_1(x) = N_s(0,\gamma) \tag{4.2.20}$$

and $$q_2(x) = N_s(0,1). \tag{4.2.21}$$

For this choice of $q_1(x)$ and $q_2(x)$ we find, respectively, that

$$I_s(q_1) = \frac{1}{\gamma}, \tag{4.2.22}$$

$$I_s(q_2) = 1, \tag{4.2.23}$$

and $$m_s = 1 + a\ (\gamma-1). \tag{4.2.24}$$

This leads to the bound

$$[1 + a\ (\gamma-1)]^{-1} \leq I_s(p) \leq 1 - a\ .\ \frac{\gamma-1}{\gamma}. \tag{4.2.25}$$

It is clear from Eq. 4.2.25 that in this case the upper and lower bounds do not depend on the order parameter s.

In Fig. 4.2.1 the bounds are given for $\gamma = 9$. In this figure also a numerical estimate of $I_s(p)$ is given for some values of s.

As this figure shows the upper bound, based on the convexity of $I_s(p)$ in p, is tighter for small values of s; whereas the lower bound, based on the moments of $q_i(x)$, becomes looser. For a = 0 and a = 1 the bounds coïncide with $I_s(p)$ since in that case the density function p(x) is an exponential power distribution which satisfies Theorem 3.1.2.

As a second special case we consider the density functions

$$q_1(x) = N(0,\sigma^2) \tag{4.2.26}$$

and $$q_2(x) = N_s(0,\beta). \tag{4.2.27}$$

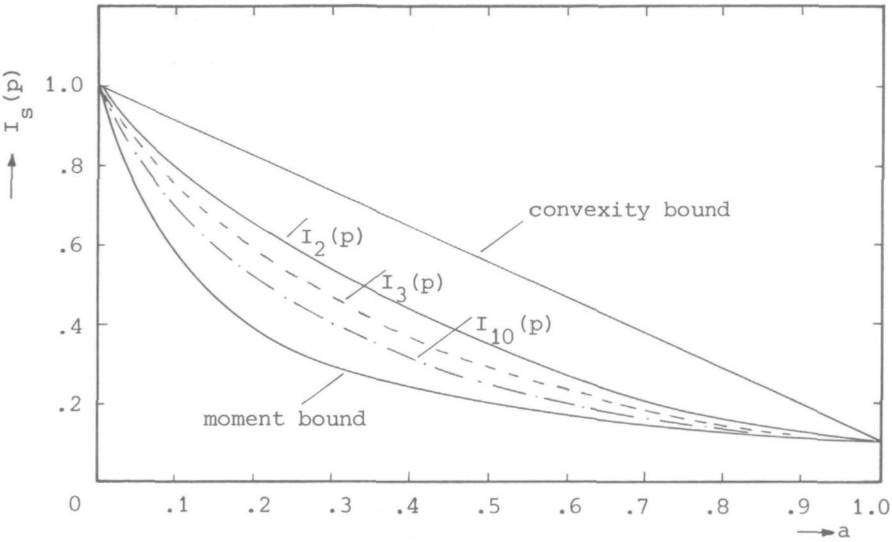

Fig. 4.2.1 $I_s(p)$ for s=2,3,10 and γ=9

Then we find that

$$I_s(q_1) = \frac{\sqrt{2}}{\sigma} \left[\frac{\sqrt{2}\ \Gamma\left(\frac{s-\frac{1}{2}}{s-1}\right)}{\sigma\sqrt{\pi}} \right]^{s-1}, \tag{4.2.28}$$

$$I_s(q_2) = \frac{1}{\beta}, \tag{4.2.29}$$

and $$m_s = (1-a)\ \beta + a\ \sqrt{\frac{2^s}{\pi}}\ \Gamma\left(\frac{s+1}{2}\right) \sigma^s, \tag{4.2.30}$$

from which we obtain the bounds

$$\left[(1-a)\ \beta + a\ \sqrt{\frac{2^s}{\pi}}\ \Gamma\left(\frac{s+1}{2}\right) \sigma^s \right]^{-1} \le I_s(p)$$

$$\le \frac{1-a}{\beta} + \frac{a\sqrt{2}}{\sigma} \left[\frac{\sqrt{2}\ \Gamma\left(\frac{s-\frac{1}{2}}{s-1}\right)}{\sigma\sqrt{\pi}} \right]^{s-1}. \tag{4.2.31}$$

In Fig. 4.2.2 - 4.2.4 the bounds are given, together with a numerical estimate of $I_s(p)$ for $\sigma = 1$ and $\beta = 1$ and for different values of s. It follows that in this situation the upper bound is much tighter than the lower bound. For $a = 0$ the bounds coïncide with $I_s(p)$, like in the previous special case. For increasing values of a, the lower bound becomes inferior to the upper bound. This behaviour can be explained by noting that from Eq. 3.1.14 of Theorem 3.1.2 it follows that $I_s(p)$ will only be equal to m_s^{-1} if $p(x)$ is an exponential power distribution of order s. Since for $a \to 1$ the gaussian density function $q_1(x)$ will become more dominant in $p(x)$ this will give rise to an increasing discrepancy between $I_s(p)$ and m_s^{-1}. This results in a looser lower bound on $I_s(p)$ for increasing values of the mixing coefficient a.

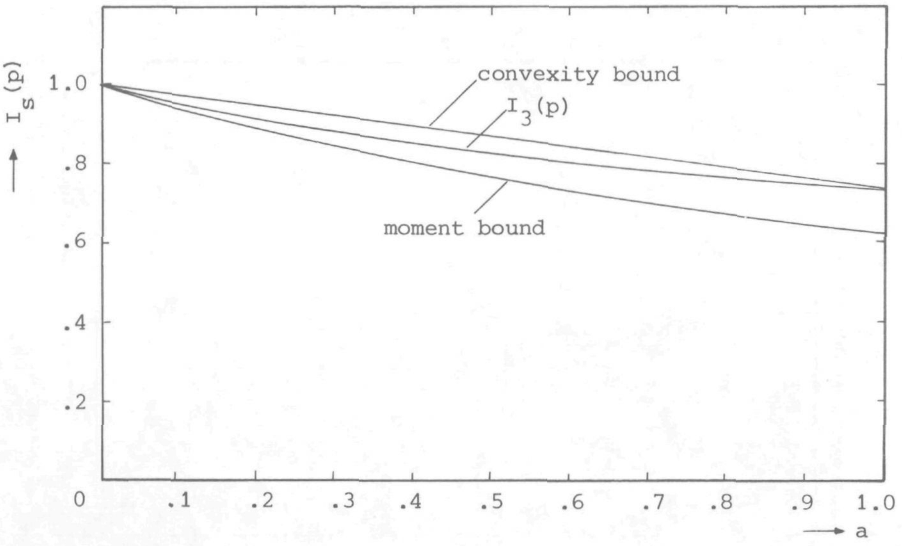

Fig. 4.2.2 $I_s(p)$ for $s=3$, $\sigma=1$

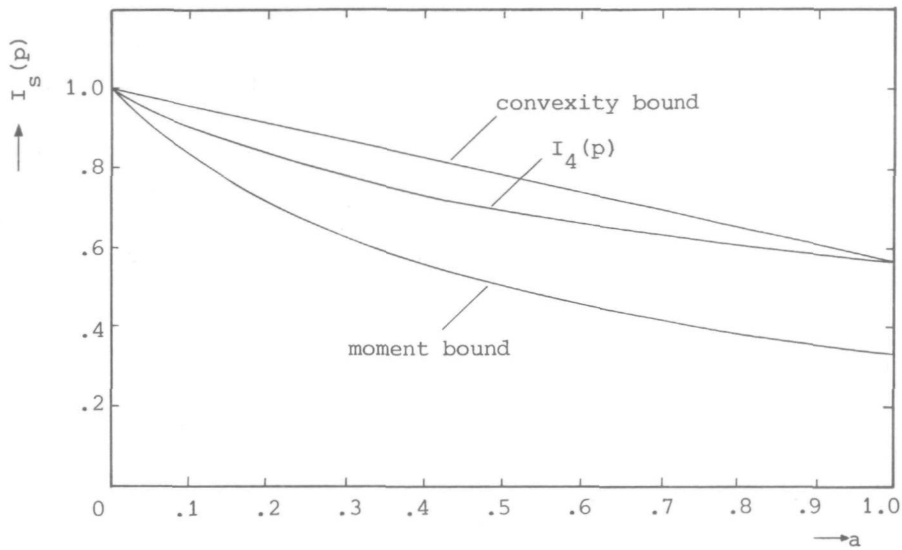

Fig. 4.2.3 $I_s(p)$ for $s=4$, $\sigma=1$

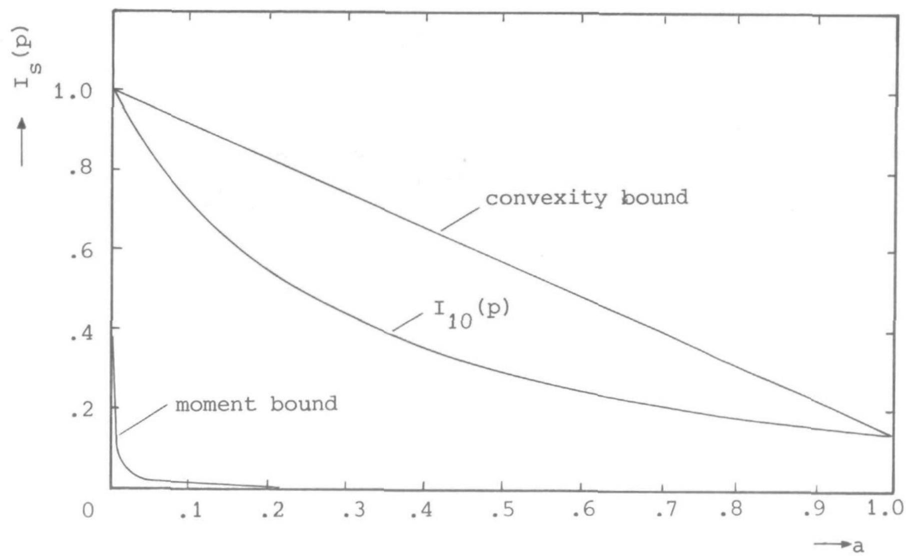

Fig. 4.2.4 $I_s(p)$ for $s=10$, $\sigma=1$

As a conclusion of this section we note that the convexity bound is very useful since it is much tighter than the moment bound, as has been illustrated with an example. If we want to obtain a lower bound on the s.a.c.m. of an estimator of θ, we can use the convexity bound on $I_s(p)$ instead of a numerical estimate of $I_s(p)$. Due to the tightness of the convexity bound we then obtain a lower bound on the s.a.c.m. of an estimator of θ which is still useful.

As the results indicate the convexity bound is favourable in particular if a is close to zero or to one. This situation arises, for example, if the density function $p(x;\theta)$ consists of the sum of a density function and a contaminating density function.

Another advantage of the convexity bound over the moment bound is its independence of the actual values of θ_i.

4.3 ESTIMATION OF SIGNAL PARAMETERS

In this section we will consider the application of the Fisher information of order s to the estimation of signal parameters. We will restrict ourself to non-random parameters. We will also mention some applications in which the density functions which were considered in Section 3.3 are useful.

Basically we can distinguish between situations which lead to location parameter estimation and those which lead to scale parameter estimation. The first situation arises if the observations which we use to estimate an unknown signal parameter, like the amplitude, are disturbed by additive noise. The second situation arises for example if we want to estimate the average power of a random signal, since this power is related to a scale parameter.

As an illustration of the second case we shall consider the estimation of the average power of the signal envelope in the short wave band. This signal envelope has an m-distribution, or Nakagami distribution, and was introduced by Nakagami [44]. Ivankin et al. [27]

have studied this problem, using a mean-square error criterion for the estimator performance. The m-distribution is given by

$$p(x;\rho,m) = \frac{2}{\Gamma(m)}\left(\frac{m}{\rho}\right)^m x^{2m-1} \exp\left[-\frac{m}{\rho}x^2\right] \tag{4.3.1}$$

where $x \geq 0$ and $m \geq 0.5$. The parameter ρ is the average power of the signal envelope. If we set $2m = \mu$, $c = 2$ and $\rho = \sqrt{\rho/m}$ in Eq. 3.3.35, we see that the m-distribution is a special case of the generalized Weibull distribution which was introduced in Example 3.3.4 (see also Table 2).

Let $\hat{\rho}$ be an estimator of ρ. Then the following bound on the s.a.c.m. of $\hat{\rho}$ can be given:

$$E_\rho|\hat{\rho} - \rho|^s \geq \frac{1}{F_s(\xi;\rho)} . \tag{4.3.2}$$

Here $F_s(\xi;\rho)$ is given by

$$F_s(\xi;\rho) = \rho^{-s} I_s^*(\xi)$$

$$= \rho^{-s} 2^s \Gamma^{1-s}(m) \left[m^{\frac{s}{s-1}+m} B(\frac{s}{s-1}, m) \; . \right.$$

$$. \; {}_1F_1(m, \frac{s}{s+1} + m + 1; - m) + m^{\frac{s-\frac{1}{2}}{s-1} + \frac{m}{2} - 1} \Gamma(\frac{2s-1}{s-1})$$

$$\left. . \; e^{-\frac{m}{2}} W_{\frac{m}{2} - \frac{s-\frac{1}{2}}{s-1}, \frac{1}{2} - \frac{s-\frac{1}{2}}{s-1} - \frac{m}{2}}(m) \right]^{s-1} . \tag{4.3.3}$$

Equation 4.3.3 can easily be obtained from Eq. 3.3.36 by substitution of $2m = \mu$, $c = 2$ and $\rho = \sqrt{\rho/m}$ into Eq. 3.3.36.

In Fig. 4.3.1 the m-distribution is given for $\rho = 1$ and some values of m. The case $m = 1$ leads to the Rayleigh distribution. Nakagami [44] has shown that the parameter m is equal to the inverse of the normalized variance $\rho^2/\mathrm{var}(x^2)$.

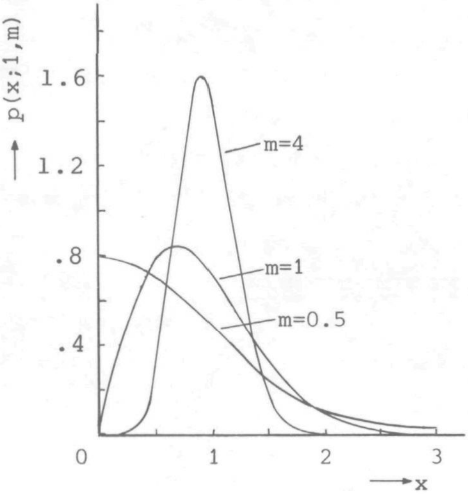

Fig. 4.3.1 The m-distribution

For this reason he suggests that m be used as a measure of fading range.

Perhaps the most frequently met situation is the estimation of a parameter in additive noise. Usually only gaussian noise is considered, but we will assume that the noise has an exponential power distribution. Before we study the estimation problem in more detail, we will mention some aspects of this distribution, as well as some applications, since it is not widely known.

The exponential power distribution was first mentioned by Diananda [18], and was suggested to him by Bartlett. Diananda used it as an example for certain properties of maximum likelihood estimates. In the field of Bayesian statistics, it was studied extensively by Box et al. [10] and Box et al. [11]. They emphasized its use as a 'nearly' gaussian distribution which can be applied to test the stability of models which use a gaussian assumption. Miller et al. [42] applied the exponential power distribution, or the gaussian family, as they called it, to study certain nonlinear detection problems. We also mention that Wyner et al. [59] have applied it as a model in information theory.

Recent applications of the exponential power distribution can be found in Chhikara et al. [14] and Whorwood et al. [58]. Chhikara et al. applied it as a model for independent noise samples. They studied certain Bayes analysis problems in remote sensing. They mainly used the values s = 1 and s = ∞ which in their analysis then lead to piecewise linear discriminant functions and, therefore, are easier to handle than the gaussian case.

Whorwood et al. found that the exponential power distribution is a good approximation to the density function of the first derivative of certain speech signals which are of importance in communication problems.

As has been stated earlier, the density function of the exponential power distribution is given by

$$p(x;\mu,\beta,s) = \frac{s^{\frac{s-1}{s}}}{2\,\Gamma(1/s)\,\beta^{1/s}} \exp\left[-\frac{|x-\mu|^{s}}{s\beta}\right] \qquad (4.3.4)$$

where β is the s.a.c.m. of the density function and s is the order parameter. In Fig. 4.3.2 we give the density function for $\beta = 1$ and some values of s.

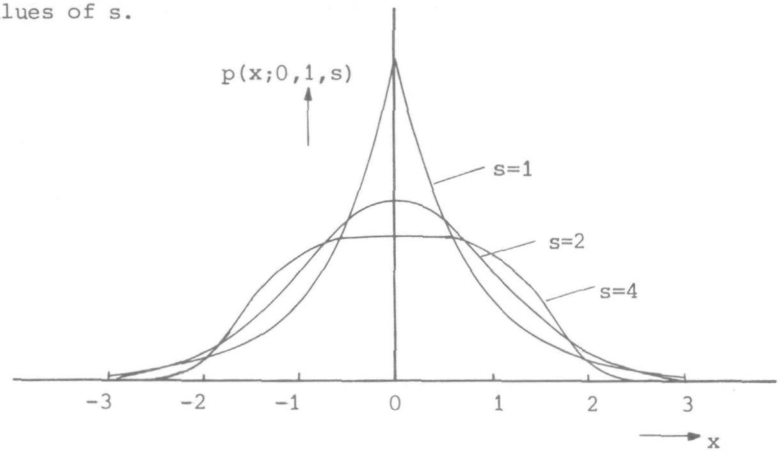

Fig. 4.3.2 Exponential power distribution of order s

This figure shows that for $s < 2$ the density function becomes peaked (leptokurtic) and that for $s > 2$ it becomes flatter (platykurtic) than the gaussian distribution ($s = 2$). Note that for $s = 1$ we have the double-sided exponential distribution.

As has been shown in Chapter 3, the exponential power distribution also has interesting properties in terms of information. It has a minimal Fisher information and a maximal Shannon information.

We now return to the estimation of an unknown parameter in additive noise. For a mean-square error criterion, the Cramér-Rao bound has been applied to signal processing, radar estimation and certain communication problems. In this context the rate-distortion bound, obtained from rate-distortion theory, has also been used. Some more details can be found, for example, in Van Trees [56] and Nahi [43]. Suppose that we have a set of observations $\xi_1, \dots, \xi_n$. If we consider discrete-time signals, the random variables $\xi_1, \dots, \xi_n$ can be the samples of this signal. The parameter θ modulates a time signal g_i, $i = 1, \dots, n$, which yields a signal $g_i(\theta)$. Furthermore, we assume that we have independent noise samples $\eta_1, \dots, \eta_n$ which have known density functions.

The most general formulation of the estimation problem is now to estimate the parameter θ, given the observations

$$\xi_i = g_i(\theta, \eta_i), \quad i = 1, 2, \dots, n. \tag{4.3.5}$$

If we restrict ourself to additive noise, we obtain a much simpler model, which is generally used in applications. We then have

$$\xi_i = g_i(\theta) + \eta_i, \quad i = 1, 2, \dots, n. \tag{4.3.6}$$

In the case of a linear estimation problem $g_i(\theta)$ is a linear function of θ. A discussion of this linear problem can be found, for example, in Van Trees [56] and Nahi [43]. If $g_i(\theta)$ is a nonlinear function of θ we can linearize this function. This leads to realistic lower bounds only for large signal-to-noise ratios. For smaller signal-to-noise ratios other bounds become more useful. Several authors, e.g., Glave [23], McAuley et al. [41] and Seidman [49], have studied this problem using Barankin bounds. Bellini et al. [5] and Ziv et al. [61] have studied a bound which is based on decision theoretic considerations and is known as the Ziv-Zakai bound.

If we adopt the mean s-th absolute error as a performance criterion, we obtain the following bound for unbiased estimators T of θ:

$$E_\theta |T - \theta|^s \geq \frac{1}{F_s(\tilde{\xi};\theta)} . \tag{4.3.7}$$

Here $\tilde{\xi}$ is an n-dimensional observation vector. Since we have additive noise, we can consider $\tilde{g}(\theta)$ to be a location parameter for the distribution of $\tilde{\xi}$, where $\tilde{g}(\theta)$ is a vector with elements $g_i(\theta)$. We then have

$$F_s(\tilde{\xi};\theta) = \left[\int_{\tilde{X}} \left| \frac{\partial}{\partial\theta} \log p(\tilde{x} - \tilde{g}(\theta)) \right|^{\frac{s}{s-1}} p(\tilde{x} - \tilde{g}(\theta))\, d\tilde{x} \right]^{s-1} . \tag{4.3.8}$$

Due to our assumption of independent noise samples, we have

$$p(\tilde{x} - \tilde{g}(\theta)) = \prod_{i=1}^{n} p(x_i - g_i(\theta)). \tag{4.3.9}$$

Differentiation with respect to θ yields

$$\frac{\partial}{\partial\theta} \log p(\tilde{x} - \tilde{g}(\theta)) = \sum_{i=1}^{n} \frac{\partial}{\partial\theta} \log p(x_i - g_i(\theta))$$

$$= \sum_{i=1}^{n} \frac{\partial \log p(x_i - g_i(\theta))}{(x_i - g_i(\theta))} \cdot \frac{\partial(x_i - g_i(\theta))}{\partial\theta}$$

$$= - \sum_{i=1}^{n} \frac{\partial}{\partial\theta} g_i(\theta) \cdot \frac{d}{dx_i} \log p(x_i). \qquad (4.3.10)$$

Substituting Eq. 4.3.10 into Eq. 4.3.8 and using Eq. 4.3.7, gives the following bound

$$E_\theta |T - \theta|^s \geq \left[\int_{\tilde{X}} \left| \sum_{i=1}^{n} \frac{\partial}{\partial\theta} g_i(\theta) \cdot \frac{d}{dx_i} \log p(x_i) \right|^{\frac{s}{s-1}} p(\tilde{x})\, d\tilde{x} \right]^{1-s}$$

(4.3.11)

If the function $g_i(\theta)$ is a linear function of θ, say $g_i(\theta) = \alpha_i \theta$, we find

$$E_\theta |T - \theta|^s \geq \left[\int_{\tilde{X}} \left| \sum_{i=1}^{n} \alpha_i \frac{d}{dx_i} \log p(x_i) \right|^{\frac{s}{s-1}} p(\tilde{x})\, d\tilde{x} \right]^{1-s} .$$

(4.3.12)

Finally, if we have $g_i(\theta) = \theta$, we obtain the simplest case. Then all random variables ξ_i are independent and identically distributed and we find, using Corollary 2.3.2, that in that case

$$E_\theta |T - \theta|^s \geq \frac{1}{n^s F_s(\xi_i;\theta)} . \qquad (4.3.13)$$

If the noise samples each have an exponential power distribution $N_s(0,\beta)$, we find, using Eq. 4.3.13, that

$$E_\theta |T - \theta|^s \geq \frac{\beta}{n^s} . \qquad (4.3.14)$$

If the noise samples each have a gaussian distribution we obtain instead of Eq. 4.3.14 the following bound

$$E_\theta|T-\theta|^s \geq \frac{\sqrt{2}}{n^s \sigma}\left[\frac{\sqrt{2}\,\Gamma(\frac{s-\frac{1}{2}}{s-1})}{\sigma\sqrt{\pi}}\right]^{s-1}. \qquad (4.3.15)$$

Using results obtained earlier, we thus have obtained bounds on $E_\theta|T-\theta|^s$ for various choices of the signal samples $g_i(\theta)$ and some distributions of the noise samples.

It is possible to obtain a bound on $E_\theta|T-\theta|^s$ which is less tight but takes on a simpler form than the bound of Eq. 4.3.11. To this end we proceed as follows.

For the observation ξ_i we have

$$F_s(\xi_i;\theta) = \left[\int_{X_i} \left|\frac{\partial}{\partial\theta}\log p(x_i - g_i(\theta))\right|^{\frac{s}{s-1}} p(x_i - g_i(\theta))\, dx_i\right]^{s-1}$$

$$= \left[\int_{X_i} \left|\frac{\partial \log p(x_i - g_i(\theta))}{\partial(x_i - g_i(\theta))} \cdot \frac{\partial(x_i - g_i(\theta))}{\partial\theta}\right|^{\frac{s}{s-1}} \cdot\right.$$

$$\left. \cdot\; p(x_i - g_i(\theta))\, dx_i\right]^{s-1}$$

$$= \left|\frac{\partial}{\partial\theta} g_i(\theta)\right|^s I_s(\eta_i). \qquad (4.3.16)$$

By applying Corollary 2.3.1, we find, using Eq. 4.3.16, that

$$F_s(\tilde{\xi};\theta) \leq \left[\sum_{i=1}^{n} \left|\frac{\partial}{\partial\theta} g_i(\theta)\right| I_s(\eta_i)^{1/s}\right]^s. \qquad (4.3.17)$$

Then the bound on $E_\theta|T-\theta|^s$ becomes

$$E_\theta |T - \theta|^s \geq \frac{1}{\left[\sum_{i=1}^{n} \left| \frac{\partial}{\partial \theta} g_i(\theta) \right| I_s(\eta_i)^{1/s} \right]^s} . \tag{4.3.18}$$

If we assume that the noise samples η_i are independent and identically distributed, we find from Eq. 4.3.18 the simple bound

$$E_\theta |T - \theta|^s \geq \frac{1}{\left[\sum_{i=1}^{n} \left| \frac{\partial}{\partial \theta} g_i(\theta) \right| \right]^s I_s(\eta_i)} . \tag{4.3.19}$$

In particular we have for noise samples with an exponential power distribution $N_s(0,\beta)$ that

$$E_\theta |T - \theta|^s \geq \frac{\beta}{\left[\sum_{i=1}^{n} \left| \frac{\partial}{\partial \theta} g_i(\theta) \right| \right]^s} . \tag{4.3.20}$$

It will be clear that the evaluation of Eq. 4.3.18 is much simpler than that of Eq. 4.3.11. The bound can be simplified further by applying Hölder's inequality to the right-hand side of Eq. 4.3.17. If we give Hölder's inequality in its discrete form

$$\sum_{i=1}^{n} a_i \, b_i \leq \left[\sum_{i=1}^{n} a_i^{\,s} \right]^{1/s} \cdot \left[\sum_{i=1}^{n} b_i^{\frac{s}{s-1}} \right]^{\frac{s-1}{s}} \tag{4.3.21}$$

and substitute

$$a_i = \left| \frac{\partial}{\partial \theta} g_i(\theta) \right| \tag{4.3.22a}$$

and $$b_i = I_s(\eta_i)^{1/s} \tag{4.3.22b}$$

into Eq. 4.3.21, we obtain the inequality

$$\left[\sum_{i=1}^{n}\left|\frac{\partial}{\partial\theta}g_i(\theta)\right| I_s(\eta_i)^{1/s}\right]^s \le \sum_{i=1}^{n}\left|\frac{\partial}{\partial\theta}g_i(\theta)\right|^s .$$

$$.\left[\sum_{i=1}^{n} I_s(\eta_i)^{\frac{1}{s-1}}\right]^{s-1} . \qquad (4.3.23)$$

Using this result we finally obtain the following bound, which is weaker than the bound of Eq. 4.3.18:

$$E_\theta\left|T-\theta\right|^s \ge \frac{1}{\sum_{i=1}^{n}\left|\frac{\partial}{\partial\theta}g_i(\theta)\right|^s . \left[\sum_{i=1}^{n} I_s(\eta_i)^{\frac{1}{s-1}}\right]^{s-1}} . \qquad (4.3.24)$$

If, once again, we assume that the noise samples are independent and identically distributed, we find that

$$E_\theta\left|T-\theta\right|^s \ge \frac{1}{\frac{1}{n}\sum_{i=1}^{n}\left|\frac{\partial}{\partial\theta}g_i(\theta)\right|^s .n^s\ I_s(\eta_i)} . \qquad (4.3.25)$$

For noise samples with an exponential power distribution of order s, we find

$$E_\theta\left|T-\theta\right|^s \ge \frac{\beta}{n^s} . \frac{1}{\frac{1}{n}\sum_{i=1}^{n}\left|\frac{\partial}{\partial\theta}g_i(\theta)\right|^s} . \qquad (4.3.26)$$

From the expressions in Eqs. 4.3.25 and 4.3.26 we can see that the bound depends on the signal samples $g_i(\theta)$. This influence is represented by the arithmetic mean of the s-th power of $|\partial g_i(\theta)/\partial\theta|$.

As an illustration we give an example of the bounds of Eqs. 4.3.25 and 4.3.26 for a specific choice of the signal $g_i(\theta)$.

EXAMPLE 4.3.1

Let us consider the estimation of the unknown signal amplitude of a cosine in additive noise. Then we have the following expression for the observations ξ_i:

$$\xi_i = \theta \cos \omega t_i + \eta_i \qquad i = 1, 2, \dots , n. \tag{4.3.27}$$

We assume that the noise samples η_i are independent and identically distributed. Then, from Eq. 4.3.25 we obtain the following bound on unbiased estimators T of θ:

$$E_\theta |T - \theta|^s \geq \frac{1}{\frac{1}{n} \sum_{i=1}^{n} |\cos \omega t_i|^s . n^s \; I_s(\eta_i)} . \tag{4.3.28}$$

If the noise samples have a gaussian distribution we can substitute the result for $I_s(\eta_i)$, as given in Table 1, into Eq. 4.3.28. In that case we obtain

$$E_\theta |T - \theta|^s \geq \frac{\sigma}{n^s \sqrt{2}} \left[\frac{\sigma\sqrt{\pi}}{\sqrt{2}\; \Gamma(\frac{s-\frac{1}{2}}{s-1})} \right]^{s-1} . \frac{1}{\frac{1}{n} \sum_{i=1}^{n} |\cos \omega t_i|^s} . \tag{4.3.29}$$

If the noise samples have an exponential power distribution of order s, we find

$$E_\theta |T - \theta|^s \geq \frac{\beta}{n^s} \; \frac{1}{\frac{1}{n} \sum_{i=1}^{n} |\cos \omega t_i|^s} . \tag{4.3.30}$$

Since the exponential power distribution of order s has a minimal Fisher information of order s the bound as given in Eq. 4.3.30 can be seen as a worst case situation for the estimation of θ.

Note that for s = 2 the bound of Eq. 4.3.28 depends on the average signal power. For $s \geq 1$ the bound depends on the average of the s-th absolute power of the signal samples.

As we have seen we can obtain bounds on the performance of estimators in terms of the signal samples $g_i(\theta)$ for the case that we estimate a non-random parameter θ in additive noise. The simplest results are obtained if the noise samples have an exponential power distribution of order s, as could be expected from results obtained earlier.

In a similar manner we can obtain bounds on the performance of estimators of a random parameter ϕ in additive noise. Then we can apply the results obtained in Section 2.6. This problem is usually referred to as the *filtering problem.*

APPENDIX A

INEQUALITIES

The purpose of this appendix is to bring together some inequalities which are used in this thesis. We will give them without proof. For the proofs as well as for further details we refer to Beckenbach et al. [4], Hardy et al. [25] and Loève [37]. We will present the inequalities in the form in which they are used in this thesis.

Let p(x) be a non-negative density function with support X and distribution function P, and let g(x) and h(x) be integrable functions. Then we can obtain the following inequalities.

1. JENSEN INEQUALITY

Let $f(u)$ be a convex function of u, in the sense that for a, b > 0, a + b = 1, the function $f(u)$ satisfies

$$f(ax_1 + bx_2) \le a\, f(x_1) + b\, f(x_2) \tag{A.1}$$

for all $x_1, x_2 \in X$. Then for the function g(x) it holds that

$$f\left[\int g(x)\ p(x)\ dx\right] \le \int f\{g(x)\}\ p(x)\ dx, \tag{A.2a}$$

whenever the right-hand side exists.

Equality in Eq. A.2a holds iff

$$g(x) = c \qquad \text{a.e. } P, \tag{A.2b}$$

where c is a constant.

2. HÖLDER INEQUALITY

Let α and β be two positive constants such that $1/\alpha + 1/\beta = 1$. Then, whenever the integrals on the right-hand side exist, it holds for $\alpha > 1$ that

$$\int |g(x)\ h(x)|\ p(x)\ dx \le \left[\int |g(x)|^{\alpha}\ p(x)\ dx\right]^{1/\alpha} \cdot \left[\int |h(x)|^{\beta}\ p(x)\ dx\right]^{1/\beta}. \tag{A.3a}$$

Equality in Eq. A.3a holds if

(i) for two non-negative constants M and K, not both zero, we have

$$M \cdot g(x)^{\alpha} = K \cdot h(x)^{\beta} \qquad \text{a.e. } P, \tag{A.3b}$$

or if:

(ii) $$g(x) \cdot h(x) = 0 \qquad \text{a.e. } P. \tag{A.3c}$$

An inequality which is used in Section 3.1, is found from the Hölder inequality by applying Jensen's inequality to the left-hand side of Eq. A.3a, for $f(u) = |u|$. We then obtain the inequality

$$\left|\int g(x)\ h(x)\ p(x)\ dx\right| \le$$

$$\leq \left[\int |g(x)|^{\alpha}\, p(x)\, dx\right]^{1/\alpha} \cdot \left[\int |h(x)|^{\beta}\, p(x)\, dx\right]^{1/\beta} \qquad (A.4)$$

3. MINKOWSKI INEQUALITY

For $\alpha \geq 1$ it holds that

$$\left[\int |g(x) + h(x)|^{\alpha}\, p(x)\, dx\right]^{1/\alpha} \leq \left[\int |g(x)|^{\alpha}\, p(x)\, dx\right]^{1/\alpha} + \left[\int |h(x)|^{\alpha}\, p(x)\, dx\right]^{1/\alpha}, \qquad (A.5a)$$

provided the integrals on the right-hand side exist. The inequality A.5a is strict unless for two non-negative constants M and K, not both zero, we have

$$M \cdot g(x) = K \cdot h(x) \qquad \text{a.e. } P. \qquad (A.5b)$$

4. C_{α}-INEQUALITY

For $\alpha \geq 1$ it holds that

$$\int |g(x) + h(x)|^{\alpha}\, p(x)\, dx \leq 2^{\alpha-1}\left[\int |g(x)|^{\alpha}\, p(x)\, dx + \int |h(x)|^{\alpha}\, p(x)\, dx\right], \qquad (A.6)$$

provided the integrals on the right-hand side exist.

The integrals which appear at the right-hand side of the inequalities 2 - 4 are called means of order α (or β):

$$M_\alpha = M_\alpha\{g(x)\} = \left[\int |g(x)|^\alpha \; p(x) \; dx \right]^{1/\alpha} \tag{A.7}$$

By $M_{-\infty}$ and M_∞ we understand the essential infimum and essential supremum, and for M_0 we find the geometric mean

$$M_0 = \exp\left[\int \{\log|g(x)|\} \; p(x) \; dx \right]. \tag{A.8}$$

For such means of order α we have the following inequalities.

5. ORDER INEQUALITY FOR M_α

For $0 < \alpha < \beta \leq \infty$ it holds, provided $M_\beta\{g(x)\}$ exists, that

$$M_\alpha\{g(x)\} \leq M_\beta\{g(x)\}, \tag{A.9a}$$

with strict inequality unless

$$g(x) = c \qquad \text{a.e. } P, \tag{A.9b}$$

where c is a constant.

6. LOG CONVEXITY INEQUALITY FOR M_α

For $0 < \alpha < \beta < \delta$, it holds that

$$M_\beta^{\beta(\delta-\alpha)} < M_\alpha^{\alpha(\delta-\beta)} \cdot M_\delta^{\delta(\beta-\alpha)} \;, \tag{A.10}$$

unless $g(x) = 0$ a.e. P for a subset $X_0 \subset X$ and $g(x) = c$ a.e. P in $X - X_0$.

From setting $\beta = \frac{1}{2}(\delta+\alpha)$ it follows that

$$\log M_\alpha^\alpha\{g(x)\} = \alpha \log M_\alpha\{g(x)\}$$

is a convex function of α.

APPENDIX B

SPECIAL FUNCTIONS

In this appendix we present some special functions and some of their properties, which are used in this thesis. For further details the reader is referred to Gradshteyn et al. [24]. Most of the functions which are mentioned here can be defined in several ways. We shall give their definitions in terms of integrals, since this is the form through which these functions appear in this thesis.

1. THE GAMMA FUNCTION

The gamma function can be defined for $x > 0$ as

$$\Gamma(x) = \int_0^\infty e^{-t}\, t^{x-1}\, dt. \tag{B.1}$$

If x is an integer we have

$$\Gamma(x) = (x-1)! \tag{B.2}$$

A recursive relation which is used frequently in this thesis is the following

$$\Gamma(x+1) = x\,\Gamma(x). \tag{B.3}$$

Some values which are of interest are

$$\Gamma(\tfrac{1}{2}) = \sqrt{\pi}$$

$$\Gamma(1) = \Gamma(2) = 1.$$

2. THE INCOMPLETE GAMMA FUNCTION

A definition of the incomplete gamma function is

$$\gamma(\alpha,x) = \int_0^x e^{-t}\, t^{\alpha-1}\, dt \qquad \alpha > 0. \tag{B.4}$$

It is easy to see that

$$\gamma(\alpha,\infty) = \Gamma(\alpha). \tag{B.5}$$

3. THE BETA FUNCTION

This function can be defined as

$$B(x,y) = \int_0^1 t^{x-1}\, (1-t)^{y-1}\, dt \qquad x,\ y > 0. \tag{B.6}$$

It is related to the gamma function by

$$B(x,y) = \frac{\Gamma(x)\ \Gamma(y)}{\Gamma(x+y)}. \tag{B.7}$$

4. THE PSI FUNCTION

This function is obtained from the gamma function and is defined for $x > 0$ as

$$\Psi(x) = \frac{d}{dx} \log \Gamma(x) = \int_0^\infty \left[\frac{e^{-t}}{t} - \frac{e^{-xt}}{1 - e^{-t}} \right] dt. \tag{B.8}$$

A value of special interest is Euler's constant, which is given by

$$C = -\Psi(1) = 0{,}577215664\ldots$$

Sometimes it is convenient to use the constant

$$\gamma = e^{C} = 1{,}78107242\ldots$$

The previous special functions are related to the gamma function. The next two special functions are related to the hypergeometric function.

5. THE DEGENERATE HYPERGEOMETRIC FUNCTION

This function is a special case of the hypergeometric function and is for $0 < \alpha < \gamma$ given by

$$_1F_1(\alpha,\gamma;x) = \frac{x^{1-\gamma}}{B(\alpha,\gamma-\alpha)} \int_0^x e^t \, t^{\alpha-1} \, (x-t)^{\gamma-\alpha-1} \, dt \, . \tag{B.9}$$

6. THE WHITTAKER FUNCTION $W_{\lambda,\alpha}(x)$

This Whittaker function is defined for $\alpha - \lambda > -\frac{1}{2}$ by

$$W_{\lambda,\alpha}(x) = \frac{x^{\lambda} \, e^{-\frac{x}{2}}}{\Gamma(\alpha-\lambda+\frac{1}{2})} \int_0^\infty t^{\alpha-\lambda-\frac{1}{2}} \, e^{-t} \, (1 - \frac{t}{x})^{\alpha+\lambda-\frac{1}{2}} \, dt \, . \tag{B.10}$$

BIBLIOGRAPHY

[1] N.L. AGGARWAL. Sur l'information de Fisher. In "Théories de l'information", J. Kampé de Fériet, ed., Springer Verlag, Berlin, 1974, pp. 111-117.

[2] D.L. ALSPACH. Gaussian sum approximations in nonlinear filtering and control. Information Sciences, Vol.7, 1974, pp. 271-290.

[3] E.W. BARANKIN. Locally best unbiased estimates. Ann. Math. Stat., Vol.20, 1949, pp. 477-501.

[4] E.F. BECKENBACH, R. BELLMAN. Inequalities. Springer Verlag, Berlin, 1971.

[5] S. BELLINI, G. TARTARA. Bounds on the error in signal parameter estimation. IEEE Trans. Commun., COM-22, 1974, pp. 340-342.

[6] A. BHATTACHARYYA. On some analogues of the amount of information and their use in statistical estimation, part 1 and 2. Sankhya, Vol.8, 1946, pp. 1-14, 201-218.

[7] D.E. BOEKEE. A Fisher information measure of order s. Techn. Report, Dept. of Electrical Engineering, Delft University of Technology, Delft, The Netherlands, July 1975.

[8] D.E. BOEKEE. An extension of the Fisher information measure. Proc. Colloquium on Information Theory, pp. 113-123, Aug. 25-29, 1975, Kesthely, Hungary (to appear).

[9] D.E. BOEKEE. Maximum information in continuous systems with constraints. Proc. VIIIth Intern. Congress on Cybernetics, Namur, Belgium, 1976 (to appear).

[10] G.E.P. BOX, G.C. TIAO. A further look at robustness via Bayes's theorem. Biometrika, Vol.49, 1962, pp. 419-432.

[11] G.E.P. BOX, G.C. TIAO. Bayesian inference in statistical analysis. Addison-Wesley Publ. Company, Reading, 1973.

[12] D.G. CHAPMAN, H. ROBBINS. Minimum variance estimation without regularity assumptions. Ann. Math. Stat., Vol.22, 1951, pp. 581-586.

[13] D. CHAZAN, M. ZAKAI, J. ZIV. Improved lower bounds on signal parameter estimation. IEEE Trans. Inform. Theory, Vol. IT-21, 1975, pp. 90-93.

[14] R.S. CHHIKARA, P.L. ODELL. Discriminant analysis using certain normed exponential densities with emphasis on remote sensing application. Pattern Recognition, Vol.5, 1973, pp. 259-272.

[15] M.L. COHEN. The Fisher information and convexity. IEEE Trans. Inform. Theory, Vol. IT-14, 1968, pp. 501-502.

[16] H. CRAMÉR. Mathematical methods of statistics. Princeton University Press, Princeton, 1974.

[17] I. CSISZÁR. Information-type measures of difference of probability distributions and indirect observations. Studia Sci. Math. Hungar., Vol.2, 1967, pp. 299-318.

[18] H. DIANANDA. Note on some properties of maximum likelihood estimates. Proc. Cambridge Phylos. Soc., Vol.45, 1949, pp. 536-544.

[19] TH. S. FERGUSON. Mathematical statistics. Academic Press Inc., New York, 1967.

[20] D.A.S. FRASER. On information in statistics. Ann. Math. Stat., Vol.36, 1965, pp. 890-896.

[21] R. FRIEDRICH. Beitrag zur Signalerkennung beim Vorliegen nicht-gausscher stochastischer Störungen. Nachrichtentechnik, Band 25, 1975, pp. 374-378.

[22] J.J. GART. An extension of the Cramér-Rao inequality. Ann. Math. Stat., Vol.30, 1959, pp. 367-380.

[23] F.E. GLAVE. A new look at the Barankin lower bound. IEEE Trans. Inform. Theory, Vol. IT-18, 1972, pp. 349-355.

[24] I.S. GRADSHTEYN, I.M. RYZHIK. Table of integrals, series and products. Academic Press Inc., New York, 1965.

[25] G.H. HARDY, J.E. LITTLEWOOD, G. POLYA. Inequalities. Cambridge University Press, Cambridge, 1973.

[26] I.A. IBRAGIMOV, R.Z. HASMINSKI. On the information in a sample about a parameter. Proc. 2nd Intern. Symp. on Inform. Theory, Tsahkadsor, USSR, 1971, pp. 295-309.

[27] P.A. IVANKIN, YE.V. LEBEDINSKY, V.B. CHERNENKO. An estimate of the average power of a random process, the values of which have a Nakagami probability distribution. Radio Eng. & Electron. Physics, May 1975, pp. 139-141.

[28] E.T. JAYNES. Prior probabilities. IEEE Trans. Syst. Sci. Cyb., SSC-4, 1968, pp. 227-241.

[29] A.M. KAGAN, On the theory of Fisher's amount of information. Soviet Math. Dokl., Vol.4, 1963, pp. 991-993.

[30] A.M. KAGAN, YU.V. LINNIK, C.R. RAO. Characterization problems in mathematical statistics. John Wiley & Sons Inc., New York, 1973.

[31] TH. KAILATH. The divergence and Bhattacharyya distance measures in signal selection. IEEE Trans. Comm. Techn., COM-15, 1967, pp. 52-60.

[32] M.G. KENDALL, A. STUART. The advanced theory of statistics. Charles Griffin & Company Ltd., London, 1968.

[33] H. KOBAYASHI, J.B. THOMAS. Distance measures and related criteria. Proc. 5th Ann. Allerton Conf. Circuits and Systems, University of Illinois, Urbana, 1967, pp. 491-500.

[34] H. KOBAYASHI. Distance measures and asymptotic relative efficiency. IEEE Trans. Inform. Theory, Vol. IT-16, 1970, pp. 288-291.

[35] S. KULLBACK. Information theory and statistics. Dover Publications, Inc., New York, 1968.

[36] D.V. LINDLEY. On a measure of the information provided by an experiment. Ann. Math. Stat., Vol.27, 1956, pp. 986-1005.

[37] M. LOÈVE. Probability theory. Van Nostrand Reinhold Company, New York, 1963.

[38] P.D.M. MACDONALD. Estimation of finite distribution mixtures. In "Applied Statistics", R.P. Gupta, ed., North-Holland Publ. Company, Amsterdam, 1975.

[39] C.L. MALLOWS. The information in an experiment. J. Royal Stat. Soc., Vol.21B, 1959, pp. 67-72.

[40] A.M. MATHAI. Dispersion and information. Metron, Vol.26, 1967, pp. 314-325.

[41] R.J. MCAULEY, E.M. HOFSTETTER. Barankin bounds on parameter estimation. IEEE Trans. Inform. Theory, Vol. IT-17, 1971, pp. 669-676.

[42] J.H. MILLER, J.B. THOMAS. Detectors for discrete-time signals in non-gaussian noise. IEEE Trans. Inform. Theory, Vol. IT-18, 1972, pp. 241-250.

[43] N.E. NAHI. Estimation theory and applications. Robert E. Krieger Publ. Company, Inc., Huntington, N.Y., 1976.

[44] M. NAKAGAMI. The m-distribution - A general formula of the intensity distribution of rapid fading. In "Statistical methods of radio wave propagation", W. Hoffman, ed., Pergamon Press, 1960, pp. 3-36.

[45] P. PAPANTONI-KAZAKOS. A generalized evaluation criterion in parameter estimation. IEEE Trans. Inform. Theory, Vol. IT-22, 1976, pp. 599-603.

[46] C.R. RAO. Linear statistical inference and its applications. John Wiley & Sons, Inc., New York, 1973.

[47] A. RÉNYI. Statistics and information theory. Studia Sci. Math. Hung., Vol.2, 1967, pp. 249-256.

[48] L.P. SEIDMAN. An upper bound on average estimation error in nonlinear systems. IEEE Trans. Inform. Theory, Vol. IT-14, 1968, pp. 243-250.

[49] L.P. SEIDMAN. Performance limitations and error calculations for parameter estimation. Proc. IEEE, Vol.58, 1970, pp. 644-652.

[50] C.E. SHANNON, W. WEAVER. The mathematical theory of communication. University of Illinois Press, Urbana, 1949.

[51] R. SHIMIZU. On Fisher's amount of information for location family. In "Statistical distributions in scientific work", Vol.3, G.P. Patil et al., ed., Reidel Publ. Company, Dordrecht, 1975, pp. 305-312.

[52] H.W. SORENSON, D.L. ALSPACH. Recursive Bayesian estimation using gaussian sums. Automatica, Vol.7, 1971, pp. 465-479.

[53] A.J. STAM. Some mathematical properties of quantities of information. Ph.D. Thesis, Delft University of Technology, Delft, The Netherlands, 1959.

[54] J.W. TUKEY. A study of sampling from contaminated distributions. In "Contributions to probability and statistics", I. Olkin, ed., Stanford University Press, Stanford, Calif., 1960, pp. 448-485.

[55] I. VAJDA. χ^{α}-divergence and generalized Fisher's information. Trans. 6th Prague Conf. on Information Theory, Statistical Decision Functions, Random Processes, Prague, 1973, pp. 873-886.

[56] H.L. VAN TREES. Detection, estimation and modulation theory, part I. John Wiley & Sons, Inc., New York, 1968.

[57] M.T. WASAN. Parametric estimation. McGraw-Hill Book Company, Inc., New York, 1970.

[58] R.W. WHORWOOD, K.G. WHITEHOUSE. Statistics of the first derivative of speech. Acustica, Vol.31, 1974, pp. 286-288.

[59] A.D. WYNER, J. ZIV. On communication of analog data from a bounded source space. Bell Syst. Techn. J., Vol.48, 1969, pp. 3139-3172.

[60] T.Y. YOUNG, TH.W. CALVERT. Classification, estimation and pattern recognition. Elsevier Publ. Company, New York, 1974.

[61] J. ZIV, M. ZAKAI. Some lower bounds on signal parameter estimation. IEEE Trans. Inform. Theory, Vol. IT-15, 1969, pp. 386-391.

SAMENVATTING

De Fisher informatiemaat is een bekende informatiemaat. Hij kan worden opgevat als een maat voor de informatie over een onbekende parameter die bevat is in een aantal waarnemingen. De informatiemaat voldoet aan bepaalde fundamentele voorwaarden die nodig zijn om hem te maken tot een bruikbare maat in toepassingen. Hij hangt samen met het begrip nauwkeurigheid in de schattingstheorie omdat hij een ondergrens verschaft voor de variantie van schatters van een parameter. Dit is de bekende Cramér-Rao-ongelijkheid.

In dit proefschrift zal een generalisatie van de Fisher informatiemaat worden beschouwd, namelijk de Fisher informatie van de orde s. Hij is gebaseerd op een bepaalde convexe functie van de eerste afgeleide van de logaritme van de aannemelijkheidsfunctie. Deze maat omvat de gebruikelijke Fisher informatiemaat als een bijzonder geval.

In hoofdstuk 1 worden enkele basisbegrippen geïntroduceerd die in het proefschrift zullen worden gebruikt. Ook wordt een korte inleiding gegeven op de problemen die zullen worden bestudeerd.

In hoofdstuk 2 wordt de Fisher informatie van de orde s geïntroduceerd en worden enkele basiseigenschappen van deze informatiemaat afgeleid. Een belangrijke eigenschap is zijn invariantie onder een meetbare uitputtende transformatie. Vervolgens wordt een ongelijkheid afgeleid voor het s-de absoluut centrale moment van zuivere parameterschatters, die gebaseerd is op de Fisher informatie van de orde s. We beschouwen enkele consequenties van deze ongelijkheid, en geven ook een gegeneraliseerde versie van de Blackwell-Rao-stelling.

Vervolgens wordt de Fisher informatie van de orde s bestudeerd voor meerdimensionale waarnemingen en worden boven- en ondergrenzen afgeleid voor dit geval. Tevens worden discrete waarnemingen kort besproken. Twee equivalente informatiematen worden geïntroduceerd en hun relatie tot de Fisher informatiemaat van de orde s wordt toegelicht. Als een laatste bijzonder geval worden stochastische parameters beschouwd. Het is mogelijk om voor dit geval twee informatiematen te definiëren door verschillende wijzen van middelen. Van deze twee maten worden enkele basiseigenschappen afgeleid.

In hoofdstuk 3 worden in het vorige hoofdstuk verkregen resultaten toegespitst op twee belangrijke klassen parameters, namelijk locatie- en schaalparameters. Allereerst worden uitdrukkingen afgeleid voor de Fisher informatie van de orde s voor deze twee gevallen. Voor de klasse van locatieparameters wordt aangetoond dat er een familie verdelingen bestaat, s^e orde exponentiële verdeling of gaussische familie geheten, die een minimale informatie bezit. Voor schattingsproblemen betekent dit dat waarnemingen die deze verdeling hebben zo min mogelijk informatie bevatten, en daarom tot de slechtste schattingskwaliteit (in termen van het s^e absoluut centrale moment van schatters) leiden. Ook wordt aangetoond dat deze karakterisering van de s^e orde exponentiële verdeling stabiel is.

Van de locatie- en schaalparameters worden tevens enkele meerdimensionale aspecten besproken. Vervolgens worden uitdrukkingen voor de Fisher informatie van de orde s bepaald voor locatie- en schaalparameters, door enkele voorbeelden te beschouwen van verdelingen die o.a. bij het schatten van signaalparameters worden gehanteerd.

Gebaseerd op Shannon's informatiemaat wordt een informatiemoment van de orde s gedefiniëerd. Hiervan worden enige resultaten gegeven voor locatie- en schaalparameters. Er wordt aangetoond dat er een analogie bestaat tussen dit informatiemoment van de orde s en de inverse van de Fisher informatie van de orde s.

In hoofdstuk 4 wordt het verband tussen de Fisher informatie van de orde s en enkele probabilistische afstandsmaten beschouwd. Aangetoond wordt dat er enkele maten zijn die op een eenvoudige wijze gerelateerd kunnen worden aan de Fisher informatie van de orde s. Dit betekent dat zij tevens gebruikt kunnen worden om grenzen voor het s^e absoluut centrale moment van parameterschatters te verkrijgen. De verkregen relaties stellen ons ook in staat om een geometrische interpretatie van de Fisher informatie van de orde s te geven.

Voor mengverdelingen worden grenzen voor de Fisher informatie van de orde s gegeven. Aangetoond wordt dat de bovengrens, die berust op de convexiteit van de Fisher informatie van de orde s, enkele aantrekkelijke eigenschappen bezit.

Tenslotte wordt de Fisher informatie van de orde s toegepast bij het schatten van signaalparameters. Hierbij wordt het schatten van een onbekende parameter van een verder volledig bekend signaal beschouwd, indien er sprake is van storing door niet-gaussische ruis. Tevens worden enkele modellen voor niet-gaussische signalen, die werden geïntroduceerd in hoofdstuk 3, besproken.

STELLINGEN

behorende bij het proefschrift van

D.E.BOEKEE

Delft, 28 september 1977

1 Door de verschillende aspecten van het begrip informatie bij uitstek te toetsen aan de door Shannon ontwikkelde maat voor informatie loopt men het gevaar eigenschappen van deze maat ook als vanzelfsprekende eigenschappen van het begrip informatie te zien.

2 Gezien het feit dat kwaliteitscriteria die op de gemiddeld kwadratische fout zijn gebaseerd niet altijd overeenkomen met criteria die door de mens worden gehanteerd verdient het onderzoek naar andere criteria groter aandacht dan doorgaans het geval is.

3 Het gebruik van benamingen als maximum informatie en maximum entropie is misleidend indien men zich beperkt tot signalen of systemen die worden beschreven met gaussische verdelingen en een gemiddeld kwadratisch foutcriterium en moet derhalve worden ontraden.

4 Een informatiemaat die is gebaseerd op een subjectief kansbegrip is niet daardoor een maat voor subjectieve informatie.

5 De gebruikelijke formulering dat een uitputtende steekproefgrootheid alle informatie benut die relevant is voor het onderhavige probleem is slechts dan zinvol indien tevens wordt aangegeven op welke wijze deze informatie wordt gemeten.

6 Het begrip informatiemetriek kan ook worden toegepast op andere informatiematen dan die van Shannon en leidt tot een andere formulering van het informatieverwerkingstheorema.

7 Bij het schatten van stochastische parameters kunnen twee verschillende Fisher informatiematen worden ontwikkeld, die leiden tot verschillende toepassingen (dit proefschrift).

8 De s^e orde exponentiële verdeling $N_s(\mu,\beta)$ wordt bij een gegeven s^e absoluut centrale moment gekarakteriseerd door de eigenschap dat deze verdeling een minimale Fisher informatie $I_s(p)$ bezit (dit proefschrift).

9 Het verdient aanbeveling om studenten bij het technisch-wetenschappelijk onderwijs tijdens hun studie een cursus spreek- en schrijfvaardigheid in de Engelse taal te laten volgen.

10 Het publiceren van de resultaten van onderzoeken naar het stemgedrag van kiezers, voorafgaande aan parlementsverkiezingen, zou men kunnen opvatten als een communicatiemodel met informatie-terugmelding. Gezien de resultaten die men voor een dergelijk model kan afleiden lijkt de conclusie gewettigd dat publikatie van deze resultaten het stemgedrag van de kiezers beïnvloedt.

11 Door technisch-wetenschappelijke onderzoekers mede verantwoordelijk te stellen voor bepaalde negatieve ontwikkelingen in de techniek en haar toepassingen overschat men de invloed die deze onderzoekers op deze ontwikkelingen hebben.

12 De te verwachten toename van de werkloosheid onder de beroepsbevolking in Nederland zal de emancipatie van de vrouw voor zover deze betrekking heeft op gelijke rechten op werk en de hieraan verbonden sociale voorzieningen, nadelig beïnvloeden.