FernUniversität in Hagen

Quasi-Continuous Maximum Entropy Distribution Approximation with Kernel Density

Thomas Mazzoni und Elmar Reucher

Diskussionsbeitrag Nr. 447
Januar 2010

Diskussionsbeiträge der Fakultät für Wirtschaftswissenschaft
der FernUniversität in Hagen

Herausgegeben vom Dekan der Fakultät

Quasi-Continuous Maximum Entropy Distribution Approximation with Kernel Density

T. Mazzoni and E. Reucher

January 26, 2010

Abstract

This paper extends maximum entropy estimation of discrete probability distributions to the continuous case. This transition leads to a nonparametric estimation of a probability density function, preserving the maximum entropy principle. Furthermore, the derived density estimate provides a minimum mean integrated square error.

In a second step it is shown, how boundary conditions can be included, resulting in a probability density function obeying maximum entropy. The criterion for deviation from a reference distribution is the *Kullback-Leibler*-Entropy. It is further shown, how the characteristics of a particular distribution can be preserved by using integration kernels with mimetic properties.

Keywords: Maximum Entropy; *Kullback-Leibler*-Entropy; Kernel Density Estimation; Mean Integrated Square Error.

1. Introduction

Real decision situations are often highly complex, forcing the decision maker to rely on a decision support system. A research area in artificial intelligence (AI) is the representation and imitation of human knowledge in computer systems. Knowledge based systems, which started with purely deterministic rule processing in the early 1970s, are nowadays capable of handling uncertain, subjective and vague knowledge. One particular field of research uses rules, based on conditional probabilities, for communication between user and system. Such conditional probabilities have to be specified either by expert knowledge or by estimation from statistical data. At the beginning of the 1990s, knowledge representation in probabilistic networks received attention not only in academic research, but also industry wide (Lauritzen and Spiegelhalter, 1988; Jensen, 1996).

To conduct knowledge acquisition under uncertainty with maximum entropy (MaxEnt) principle is an axiomatically well-founded approach (Csiszàr, 1975; Kern-Isberner, 1998; Paris and Vencovská, 1997; Shore and Johnson, 1980). Incoming information about the conditional structure of all involved variables in a given domain is processed in a way that avoids generation of new and unintended dependencies (Paris and Vencovská, 1990; Schramm and Fronhöfer,

2001; Calabrese, 2004). This type of conservative knowledge processing provides a justification for MaxEnt as unique probabilistic inference process, satisfying a set of reasonable conditions (Paris and Vencovská, 1990, 1997; Shore and Johnson, 1980; Calabrese, 2004).

The remainder of this paper is organized as follows. Section 2 gives a short review of the milestones in AI. The focus in Section 3 lies on knowledge acquisition under maximum entropy for discrete models. For this case, the expert-system-shell SPIRIT[1] is a suitable tool, which allows to handle large knowledge bases with hundreds of variables in milliseconds. In Section 4 we show how to overcome the limitation on discrete variables by extending the MaxEnt inference principle to quasi continuous domains. We further show that the suggested approach is a generalization of the well established inference procedure in discrete domains. An illustrative example of this idea is given in Section 5. Section 6 summarizes the findings and draws conclusions. Further research activities on this field are motivated subsequently.

2. Milestones in Artificial Intelligence

The cradle of Artificial Intelligence (AI) was in Dartsmouth, USA, when 1956 some famous researcher organized a workshop to discuss, how to simulate human thinking patterns on a computer. In the following decade first expert-systems were developed (Harmon and King, 1985).

The "General Problem Sover" was created by Newell et al. (1959). It was based on their theoretical work on logical machines. The programm was the first to solve formalized symbolic problems, theorem proofs, geometric problems or to play chess, for instance. The "Advice Taker", developed by McCarthy (1958), was a hypothetical program and the first one allowing for the use of logic to represent information.

Some further programs followed in the 1960s, but at the beginning of the 1970s more user-friendly software-tools prevailed as mainstream in expert-systems. The most famous one was MYCIN, developed at the Stanford University in 1972. It was designed to identify bacteria causing severe infections, such as bacteremia and meningitis, and to recommend antibiotics. It was the first programm with the ability to handle conditionals of the type *if-then* with uncertainty factors. But due to ethical doubts, the system was never developed further. Other structural problems caused a decrease in enthusiasm in expert-systems over the following years.

At the end of the 1980s Lauritzen and Spiegelhalter (1988) revived the interest in the AI-community. Their work was seminal in that they discovered a method to handle large knowledge-bases efficiently in graphical structures. This knowledge launched a renaissance of expert-systems. The program HUGIN[2], developed by Jensen in the beginning of the 1990s, is a representative of *Bayes-*

[1]The shell SPIRIT is a java programm, which therefore is independent of the operating system. For more details about the algorithmic implementation and applications, illustrating the power of the shell, the reader is referred to Rödder et al. (2006).

[2]http://www.hugin.com

net-oriented programs. The system is used in different commercial applications. HUGIN is able to handle continuous variables with normal, beta, gamma, exponential, weibull, uniform, triangular and lognormal distribution. It is also able to handle discrete and (special) continuous variables simultaneously. The generation of knowledge and the inference process in *Bayes*-net-oriented systems is based on the *Bayes*-formula.

For two discrete variables X and Y with finite values x and y the probability P of the conditional $X = x|Y = y$ for any (x, y) is given by

$$\begin{aligned} P(X = x|Y = y) &= \frac{P(Y = y|X = x) \cdot P(X = x)}{P(Y = y)} \\ &= \frac{P(Y = y|X = x) \cdot P(Y = y)}{\sum_x P(Y = y|X = x) \cdot P(X = x)}. \end{aligned} \tag{1}$$

The inference principle in HUGIN is applicable for continuous X and Y, if both variables are Gaussian. This involves an additional restriction for handling continuous variables in HUGIN

- A continuous variable never can have discrete "parents". More precisely, for a continuous X the conditional $X|Y$ can only be defined, if Y is not a discrete variable.

We want to overcome these drawbacks and therefore focus on knowledge acquisition with MaxEnt, which is much more flexible in generating knowledge bases than Bayes-nets. For more details the reader is referred to Rödder et al. (2009). Theoretical results of MaxEnt knowledge acquisation in continuous domains are also obtained recently by Singer (2005)

3. Knowledge Acquisition by MaxEnt

In this section we give a short review of the ideas of the MaxEnt-concept, which is partially adopted in Rödder et al. (2009). To build a knowledge base one needs a finite set of finite valued variables $\mathbf{V} = \{V_1, ..., V_L\}$ with respective values v_l of V_l. The variables might be boolean, nominal or numerical. With help of literals of the form $V_l = v_l$, propositions $A, B, C, ...$ are formed by the junctors $\wedge$ (and), $\vee$ (or), $\neg$ (not) and by respective parentheses. Conjuncts of literals, such as $\mathbf{v} = v_1, ..., v_L$, are elementary propositions, $\mathbf{V}$ is the set of all $\mathbf{v}$. $|$ is the directed conditional operator; expressions like $B|A$ are conditionals. Such conditionals inside a knowledge domain are true up to a certain degree, which might be expressed by probabilities $p \in [0; 1]$; thus we write $B|A\,[p]$ for such conditional uncertain information. Regarding semantics, a model is a probability distribution P for which such conditional information is valid.

Given a set of rules $\mathcal{R} = \{B_i|A_i\,[p_i], i = 1, \ldots, I\}$, the knowledge acquisition process is conducted by solving the nonlinear optimization problem

$$P^* = \arg\min KL(Q||P^0), \quad s.t.\ Q \models \mathcal{R}, \tag{2a}$$

where

$$KL(Q||P^0) = \sum_{\mathbf{v}} Q(\mathbf{v}) \cdot ld \frac{Q(\mathbf{v})}{P^0(\mathbf{v})}. \tag{2b}$$

$\mathrm{KL}(Q||P^0)$ denotes *Kullback-Leibler*-divergence (cf. Csiszàr, 1975). Let ld denote the logarithm with base two, then the quantity KL has dimension bit. Here the arg-function determines the special probability distribution P^* among Q, minimizing $KL(Q||P^0)$ and satisfying the linear constrains $\mathcal{R}$. A set of conditionals $\mathcal{R} = \{B_i|A_i\,[p_i], i = 1, \ldots, I\}$ represents a convex polyhedron (Reucher and Kulmann, 2007). P^* is considered the epistemic state from which all valid conditionals can be evaluated. P^0 denotes the uniform distribution and is the solution of (2a) for $\mathcal{R} = \{\}$.
(2a) is equivalent to maximizing the entropy (Rödder, 2003):

$$P^* = \arg\max H(Q), \quad s.t.\ Q \models \mathcal{R}, \tag{3a}$$

where

$$H(Q) = -\sum_{\mathbf{v}} Q(\mathbf{v}) \cdot ld\, Q(\mathbf{v}). \tag{3b}$$

It is well known, that H measures the average uncertainty of any $\mathbf{v}$ being true.

The maximum entropy distribution P^* is uniquely determined as the one incorporating all information in $\mathcal{R}$. Furthermore, it is maximally unbiased with respect to missing information. So, P^* is characterized by the distribution expressing maximum uncertainty with respect to $\mathcal{R}$.

The principle of minimizing Kullback-Leibler-divergence, or maximum entropy, respectively, follows quite naturally from the axioms of Shore and Johnson (1980, p. 27). Given a continuous priori density and a set of constrains, there is only one posterior density satisfying all restrictions. It can be determined by a procedure satisfying the following axioms Shore and Johnson (1980, p. 29):

- Uniqueness
- Invariance
- System Independence
- Subset Independence

The unique posterior distribution can be obtained by minimizing the cross-entropy (2a). The principle of minimum cross-entropy is implemented in the expert-system-shell SPIRIT for discrete variable domains.

However, many knowledge domains consist of quasi-continuous variables. Such variables are height, age, etc. The following example demonstrates the principle of knowledge acquisition by maximum entropy with a quasi-continuous variable. It will be revisited in Section 5.2. Given a sample ($N = 10\,000$) of a random variable $X \sim N(5, 1.5)$. Clearly the interval $[0; 10]$ supports more than a 3σ-range to both sides of the expectation value. This should be sufficient for dividing the continuous domain of X into discrete classes.

In the first step we discretize the support into five intervals [0;2);[2;4);[4;6);[6;8);[8;10). Learning the sample in SPIRIT yields distribution P_5^* as a first approximation of $N(5, 1.5)$. Hence the difference $H(P_5^*) - H(N(5, 1.5)) = 2.7124 - 2.6321 = 0.0803$ bit are the 'costs' for the loss of knowledge due to the rough discretization. A finer dicretization of ten intervals,

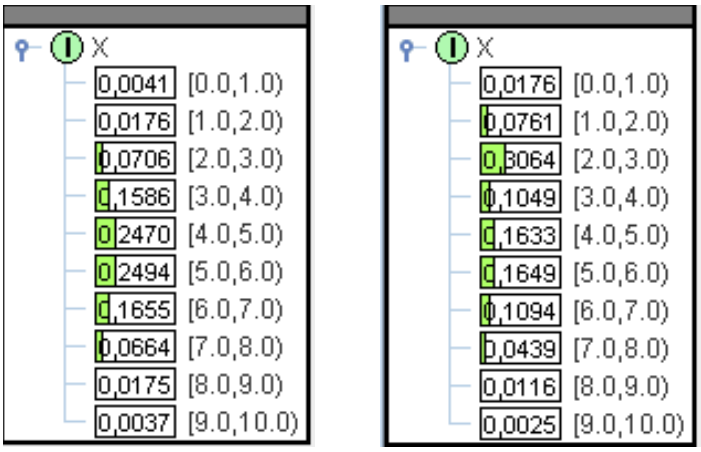

Figure 1: Sample-based and Rule-based Knowledge Acquisition is SPIRIT

each with length 1.0, yields the sample-based learned knowledge P^*_{10}. Figure 1 (left) shows a screenshot in SPIRIT.
The entropy of this distribution is $H(P^*_{10}) = 2.6413$ bit. In this case the acquired knowledge P^*_{10} differs just 0.0088 bit. Obviously, the approximation of $N(5, 1.5)$ becomes better, the finer the discretization is made.

Suppose now, we get the additional information $P(X \leq 3) = 0.4$. For the incorporation of this information in SPIRIT (2a) is solved with $P^0 = P^*_{10}$ and the rule $\mathcal{R} = \{X = [0;1) \vee X = [1;2) \vee X = [2;3) \ [.4]\}$. The posteriori distribution P^{**}_{10} is shown in figure 1 (right). Needless to say, the precision of the posteriori distribution depends on the discretization, too.

This example illustrates how quasi-continuous variables can be handled in the shell SPIRIT. But the discretization causes loss of knowledge. To overcome this problem, we will develop an alternative method for building knowledge bases with quasi-continuous variables under maximum entropy.

4. Explorative Density Approximation

Assume X is a discrete random variable defined on a probability space $(\Omega, \mathcal{F}, P)$, fulfilling all necessary requirements regarding measurability, as required. If the set of possible realizations $\{x_1, \ldots, x_N\}$ is large, it may be more convenient to extend the probability space in order to treat X as continuous random variable. Examples for such situations are body height, age, income and many more. In a continuous setting one has access to probability density functions, which, for large N, are far more efficient than sets of individual probabilities $\{\pi_1, \ldots, \pi_N\}$.

The transition between quasi continuous distributions and continuous probability densities can be formalized with Dirac's delta function

$$p(x) = \lim_{N \to \infty} \sum_{n=1}^{N} \pi_n \delta(x - x_n). \tag{4}$$

For finite N, (4) is an approximation, which exactly represents the moment

structure of the distribution. This can be shown easily for the first moment

$$E[X] = \int xp(x)dx \approx \sum_{n=1}^{N} \pi_n \int x\delta(x - x_n)dx = \sum_{n=1}^{N} \pi_n x_n \tag{5}$$

and for all higher moments, analogously. Unfortunately, the sum of delta functions, even though it generates a valid density with correct moment structure, is a poor approximation of the density function itself. In what follows, we derive a better representation with other integration kernels and show how to control the approximation error.

4.1. Nonparametric Density Approximation

In discrete setup, the maximum entropy distribution is a uniform distribution. So $\pi_1 = \ldots = \pi_N = \frac{1}{N}$ holds. If an additional observation is available, say x_m, the distribution changes to $\pi_n = \frac{1}{N+1}$ for all $n \neq m$ and $\pi_m = \frac{2}{N+1}$ (see Rödder, 2006). For quasi continuous random variables a density approximation is then realized in form of a histogram density. In continuous setup, the probability density is estimated similar to (4)

$$\hat{p}(x) = \frac{1}{N} \sum_{n=1}^{N} k(x - x_n, h). \tag{6}$$

Equation (6) should be discussed in some more detail. The hat above p indicates an estimator for the true but unknown continuous density. The function $k(x - x_n, h)$ is the kernel function, replacing Dirac's delta function in (4). This kernel function should fulfill the requirements of a probability density itself. The parameter h rules the bandwidth or window width of the kernel function. Usually this parameter is a function of N itself. Determination of an optimal bandwidth is a well known problem in kernel density estimation[3]. As measure for performance of the chosen bandwidth and kernel function, we introduce the mean integrated square error

$$\begin{aligned} \text{MISE} &= \int E\big[\hat{p}(x) - p(x)\big]^2 dx \\ &= \int \big(E\big[\hat{p}(x)\big] - p(x)\big)^2 dx + \int \text{Var}\big[\hat{p}(x)\big] dx. \end{aligned} \tag{7}$$

Equation (7) shows that the mean integrated square error can be broken up into a sum of integrated square bias and integrated variance (cf. Silverman, 1986, p. 36). It is easy to show that for zero bandwidth kernels, like the delta function in (4), the integrated bias vanishes but the integrated variance is large. For large bandwidths, $h \to \infty$, the variance of the density estimation vanishes but the estimator is seriously biased. Hence, choosing a proper bandwidth for the kernel function means finding an optimal tradeoff between bias and variance.

We will discuss this subject later in more detail. If we assume for the moment that we are able to calculate an optimal bandwidth, the immediate question

[3]For an excellent treatment on this subject see Silverman (1986).

of an appropriate kernel function arises. If an optimal bandwidth is chosen, then ceteris paribus the *Epanechnikov*-Kernel (Epanechnikov, 1969) causes the smallest MISE of all possible kernel functions (see Silverman, 1986, tab. 3.1). We will use a Gaussian kernel instead, because it is differentiable everywhere and its efficiency is about 95% of the *Epanechnikov*-Kernel

$$k(x - x_n, h) = \phi(x, x_n, h) = \frac{1}{\sqrt{2\pi h^2}} e^{-\frac{1}{2}\left(\frac{x-x_n}{h}\right)^2}. \tag{8}$$

Choosing a Gaussian kernel has one additional advantage. If we assume the true but unknown density to be Gaussian also, with standard deviation σ, then the optimal bandwidth can be calculated straight forward

$$h = \sqrt[5]{\frac{4}{3N}}\sigma, \tag{9}$$

(cf. Silverman, 1986, p. 45). Other methods for choosing an adequate bandwidth are subjective choice, cross validation and others. We will see later that characteristics of the approximated density function are also important for determination of kernel dispersion.

To demonstrate the advantage of smooth kernel functions an illustrative example is given in figure 2. Assume the drivers age in years (between 18 and 80) is observed in a spot check on traffic. The appropriate data was simulated, conditioned on a normal distribution with an expectation of 48 years and standard deviation of 8 years as the true distribution. The related probability density is indicated as black dashed line in figure 2. While the traditional maximum entropy density approximation is given as histogram density, the smooth kernel density estimation is indicated gray. Obviously, the area between the kernel density estimation and the true density is much smaller than that between the histogram density and the true one. The absolute value of this area is the square root of the MISE. Therefore, according to criterion (7), the kernel density approximation is far superior. The effect becomes even more distinct for larger samples (figure 2 right).

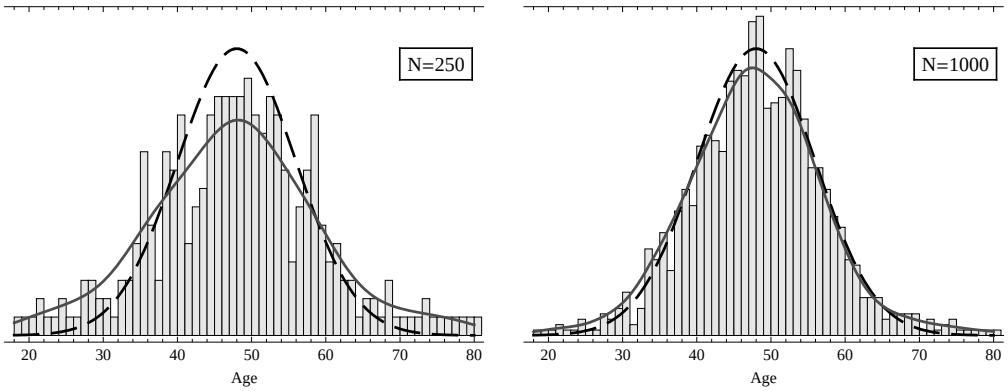

Figure 2: Maximum Entropy Density of Age Distribution – Histogram-Density and Kernel Density Estimation

Notice that the discrete maximum entropy distribution, prior to observation, is given by a uniform distribution. To achieve formal equivalence of both methods, in the kernel density framework every possible discrete realization has to be occupied with one artificial observation first. In doing so, one can reproduce the original histogram result by choosing a rectangular kernel function

$$k(x - x_n, h) = \begin{cases} \frac{1}{h} & \text{for } |x - x_n| < \frac{h}{2} \\ 0 & \text{else,} \end{cases} \tag{10}$$

with h indicating the width of the observation classes. The rectangular kernel is less efficient than the Gaussian kernel (Silverman, 1986, tab. 3.1 on page 43), which is why the kernel density method outperforms the traditional maximum entropy representation in this example.

4.2. Mimetic Hyperbolic Kernel Function

It may not always be appropriate to use a smooth kernel like the Gaussian one, because certain features of the approximated distribution may be blurred. In order to obtain more flexibility in approximating continuous densities with the discrete maximum entropy methodology, we introduce a new hyperbolic kernel function

$$\kappa(x - x_n, h, \nu) = \frac{\sinh\left[\frac{1}{2\nu}\right]}{h\left(\cosh\left[\frac{1}{2\nu}\right] + \cosh\left[\frac{1}{h\nu}(x - x_n)\right]\right)}. \tag{11}$$

Here, the bandwidth parameter h gives the width of one observation class, while the dispersion parameter ν determines, how much probability mass is located within the bandwidth. Thus, for $\nu \to 0$ we get a rectangular kernel function of width h, while for greater ν, it becomes more similar to a Gaussian kernel. If the dispersion parameter approaches infinity, the kernel function becomes completely non-informative, like a degenerated Gaussian density for $\sigma \to \infty$. Figure 3 shows a hyperbolic kernel function around $x_0 = 0$, with bandwidth $h = 1$ and various dispersions.

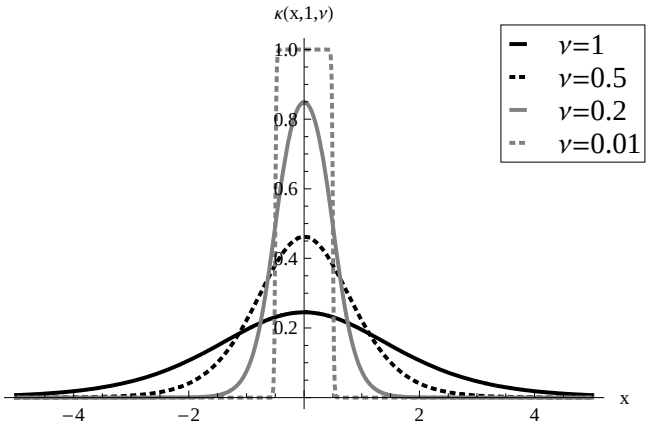

Figure 3: Hyperbolic Kernel Function with Different Dispersions

The problem of choosing an optimal bandwidth is a little more involved in this setup. While the raw bandwidth is determined by the width of observation classes, the primary task is to select an adequate dispersion parameter ν. On the one hand, the properties of the true underlying density have to be considered, in order to mimic the characteristics of the true density; on the other hand, one can calculate the efficiency of the hyperbolic kernel regarding the dispersion.

4.3. Efficiency of the Hyperbolic Kernel Function

The efficiency of an arbitrary kernel function is assessed in comparison to the *Epanechnikov*-Kernel, which is the optimal choice considering a minimal MISE. Silverman (1986, equation 3.26 on page 42) derives an explicit formula for the efficiency

$$\operatorname{Eff}\big[k(x)\big] = \frac{3}{5\sqrt{5}} \left(\int x^2 k(x) dx \right)^{-\frac{1}{2}} \left(\int k(x)^2 dx \right)^{-1}. \tag{12}$$

All arguments for bandwidth or dispersion, respectively, are suppressed in (12) for notational simplicity. Evaluating both integrals for the hyperbolic kernel shows

$$\int_{-\infty}^{\infty} x^2 \kappa(x, h, \nu) dx = \frac{h^2(4\pi^2 + \nu^2)}{12\nu^2} \tag{13a}$$

and

$$\begin{aligned} \int_{-\infty}^{\infty} \kappa(x, h, \nu)^2 dx &= \frac{2 + \nu + e^{\nu}(\nu - 2)}{h\nu(e^{\nu} - 1)} \\ &\approx \frac{\nu}{h(2 + \nu)}. \end{aligned} \tag{13b}$$

The approximation in (13b) results from a *Maclaurin*-Series expansion of the exponential, neglecting terms of $\mathcal{O}(\nu^3)$. Thus, the efficiency of the hyperbolic kernel is approximately

$$\operatorname{Eff}\big[\kappa(x, h, \nu)\big] \approx \frac{3\sqrt{12}}{5\sqrt{5}} \cdot \frac{2 + \nu}{\sqrt{4\pi^2 + \nu^2}}. \tag{14}$$

As required, the efficiency does not depend on the bandwidth h. By differentiating (14) and equating to zero, one finds $\nu = 2\pi^2$ which is the dispersion at which the hyperbolic kernel achieves its highes efficiency

$$\operatorname{Eff}\big[\kappa(x, h, 2\pi^2)\big] \approx 97.5\%. \tag{15}$$

It should be emphasized that this efficiency on the one hand results from a calculation, that does not account for specific properties of the underlying true density. On the other hand, a series expansion around $\nu = 0$ was conducted, which might cause significant error for large dispersions. Nevertheless, the hyperbolic kernel is more flexible than a rectangular or Gaussian kernel and seems to be more efficient under certain conditions.

5. Maximum Entropy and Boundary Conditions

The aim of this section is to extend the kernel density methodology to maximum (relative) entropy problems under boundary conditions. The *Kullback-Leibler*-Entropy for continuous variables x is definied as (Kullback and Leibler, 1951)

$$KL = \int p(x) \log \left[\frac{p(x)}{p_0(x)} \right] dx \tag{16}$$

In what follows, we do not minimize KL, but maximize $-KL$ instead and define $S = -KL$ to be the relative entropy.

$$\begin{aligned} S &= -\int p(x) \log \left[\frac{p(x)}{p_0(x)} \right] dx \\ &= \int p(x) \log\big[p_0(x)\big] dx - \int p(x) \log\big[p(x)\big] dx. \end{aligned} \tag{17}$$

The reference density is indicated as p_0 in equation (17), whereas p is approximated as kernel density

$$p(x) = \sum_{n=1}^{N} w_n k(x - x_n, h). \tag{18}$$

This is a slightly modified approach, because now every kernel has a weight w_n and the sum of all weights is one. In the following we will discuss, how these weights are determined optimally.

5.1. *Lagrange*-Method for Constrained Optimization

Let $p_0(x)$ be an arbitrary reference density, and $p(x)$ the demanded maximum relative entropy density. Further, assume that π is the probability mass located in the interval $[a, b]$. The maximum relative entropy density can now be found by *Lagrange*-Optimization. Taking approximation (18) and the relative entropy (17) one obtains the *Lagrange*-Function

$$\begin{aligned} L = &\sum_{m=1}^{N} w_m \int k(x - x_m) \log\big[p_0(x)\big] dx \\ &- \sum_{m=1}^{N} w_m \int k(x - x_m) \log \left[\sum_{n=1}^{N} w_n k(x - x_n) \right] dx \\ &+ \lambda \left(\sum_{m=1}^{N} w_m - 1 \right) + \lambda_1 \left(\sum_{m=1}^{N} w_m \int_a^b k(x - x_m) dx - \pi \right). \end{aligned} \tag{19}$$

Once again, all arguments of bandwidth and dispersion were suppressed for notational convenience. In order to find an optimal solution for an arbitrary

weight, (19) has to be differentiated with respect to w_m. One obtains

$$\begin{aligned}\frac{\partial L}{\partial w_m} &= \int k(x-x_m)\log\big[p_0(x)\big]dx \\ &\quad - \int k(x-x_m)\log\left[\sum_{n=1}^{N} w_n k(x-x_n)\right]dx \\ &\quad + \lambda_0 + \lambda_1 \int_a^b k(x-x_m)dx,\end{aligned} \tag{20}$$

with the multiplier substitution $\lambda_0 = \lambda - 1$. The next step is to approximate the integrals in (20), to get an analytic expression for the *Lagrange*-Derivative. This is accomplished by presuming the kernel functions strongly localized $k(x - x_m) \approx \delta(x - x_m)$. This assumption my be false for kernels, localized in direct neighborhood, but for more distant kernels the intersection is scarcely perceptible. The delta-approximation allows immediate calculation of the integrals and one obtains

$$\frac{\partial L}{\partial w_m} = \log\big[p(x_m)\big] - \log\left[\sum_{n=1}^{N} w_n k(x_m - x_n)\right] + \lambda_0 + \lambda_1 I_{[a,b)}(x_m), \tag{21a}$$

with the indicator function

$$I_{[a,b)}(x_m) = \begin{cases} 1 & \text{if } a \le x_m < b \\ 0 & \text{else.} \end{cases} \tag{21b}$$

Now we can equate (21a) to zero and after some algebraic manipulation the condition

$$\sum_{n=1}^{N} w_n k(x_m - x_n) = e^{\lambda_0 + \lambda_1 I_{[a,b)}(x_m)} p_0(x_m) \tag{22}$$

is obtained. Equation (22) is valid for all x_m with $m = 1, \ldots, N$. Therefore, the set of conditions can be written most conveniently in vector/matrix form. Furthermore, determining an optimal weight vector is obviously a linear problem to be solved by matrix inversion. The solution has the form

$$\mathbf{w} = \mathbf{\Psi}^{-1}\mathbf{R}\mathbf{p}_0 \approx \mathbf{R}\mathbf{\Psi}^{-1}\mathbf{p}_0. \tag{23}$$

In (23), $\mathbf{R}$ is a diagonal matrix, containing the boundary conditions or restrictions. Its elements are $R_{nn} = e^{\lambda_0 + \lambda_1 I_{[a,b)}(x_n)}$ and $R_{mn} = 0$ for $m \neq n$. $\mathbf{\Psi}$ is called the metrics, because it depends on the kernel function, or more precisely, on the distance measure induced by the kernel function. Its elements are $\Psi_{mn} = k(x_m - x_n)$. The vector $\mathbf{p}_0 = \big(p_0(x_1), \ldots, p_0(x_N)\big)'$ contains the function values of the reference density at $x_1, \ldots, x_N$. Generally two arbitrary matrices, $\mathbf{A}$ and $\mathbf{B}$, are not commutative regarding multiplication, $\mathbf{AB} \neq \mathbf{BA}$. Here, $\mathbf{R}$ is diagonal by definition and earlier we assumed the kernel function to be sharply localized, resulting in a nearly diagonal $\mathbf{\Psi}$. Thus, we can expect the commutator to be very small and the approximation in (23) to hold.

In order to determine $\mathbf{R}$, the *Lagrange*-Function (19) has to be differentiated with respect to its multipliers. By $\frac{\partial L}{\partial \lambda} = 0$,

$$\sum_{n=1}^{N} w_n = 1 \tag{24}$$

follows. Differentiating with respect to λ_1 and once again using the simplification $k(x - x_m) \approx \delta(x - x_m)$ yields

$$\sum_{n=1}^{N} w_n I_{[a,b)}(x_n) = \pi. \tag{25}$$

For the following calculations it is beneficial to collect the indicator terms into a vector $\boldsymbol{\chi} = \big(I_{[a,b)}(x_1), \ldots, I_{[a,b)}(x_N)\big)'$. We can then use the identity $\boldsymbol{\chi}'\mathbf{R} = e^{\lambda_0+\lambda_1}\boldsymbol{\chi}'$ to obtain

$$\boldsymbol{\chi}'\mathbf{w} = e^{\lambda_0+\lambda_1}\boldsymbol{\chi}'\boldsymbol{\Psi}^{-1}\mathbf{p}_0 = \pi. \tag{26}$$

Furthermore, define the unity vector $\mathbf{1} = (1, \ldots, 1)'$, then an expression similar to (26) can be derived for $(\mathbf{1} - \boldsymbol{\chi})'\mathbf{w} = 1 - \pi$. Summarizing these results, we obtain

$$e^{\lambda_0} = \frac{1-\pi}{(\mathbf{1}-\boldsymbol{\chi})'\boldsymbol{\Psi}^{-1}\mathbf{p}_0} \quad \text{and} \quad e^{\lambda_0+\lambda_1} = \frac{\pi}{\boldsymbol{\chi}'\boldsymbol{\Psi}^{-1}\mathbf{p}_0}. \tag{27}$$

Notice that the determination of kernel weights of an unrestricted maximum relative entropy density approximation is a special case of (27). In this case (23) simplifies to

$$\mathbf{w} = \frac{\boldsymbol{\Psi}^{-1}\mathbf{p}_0}{\mathbf{1}'\boldsymbol{\Psi}^{-1}\mathbf{p}_0}. \tag{28}$$

5.2. Uniform and Normal Distribution Example

To illustrate the capabilities of the suggested method, we approximate two standard distributions with and without constraints. In this example a uniform distribution in the interval $[0, 10]$ and a normal distribution with mean $\mu = 5$ and standard deviation $\sigma = 1.5$ is chosen. Both distributions provide completely different characteristics, and thus adequately illustrate certain aspects of the kernel density approximation.

To put a little stress on the method, a rather rough grid of ten equally spaced kernels was positioned inside the relevant interval, ranging from 0.5 to 9.5. Initially, the unconstrained probability densities were approximated. The hyperbolic kernel (11) was used as kernel function. Figure 4 shows the results of both approximations. The particular densities in figure 4 and 5 are coded as follows. The true, unconstrained density is indicated by a black dashed line. Kernel approximations with the hyperbolic kernel of bandwidth $h = 1$ are indicated gray. The solid gray function is generated by a kernel approximation with dispersion parameter $\nu = 0.001$, while the dotted gray line indicates a dispersion of $\nu = 0.5$.

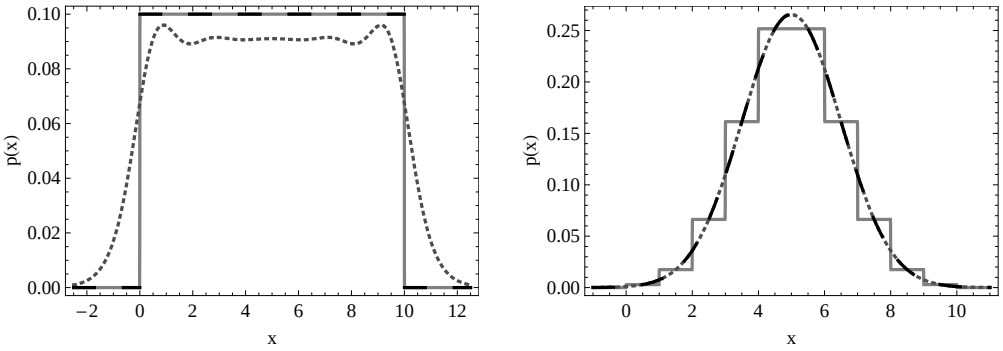

Figure 4: Unconstraint Uniform (Left) and Normal (Right) Probability Density Approximation

It is easily seen in figure 4 that the different characteristics of the distributions require different kernel shapes. While a sharp localized kernel is required to generate an appropriate representation of a uniform distribution or more generally, of a distribution with discontinuities in the density function, a more 'blurred' kernel is required for a smooth distribution like the normal one. Therefore, the dispersion parameter should always be selected with respect to the nature of the underlying distribution. Obviously the kernel approximation is a quite powerful method, even the smooth normal density coincides completely with the approximation if the dispersion is chosen properly.

Now we introduce a boundary condition. Suppose that due to external knowledge we have to assume that 40% of the probability mass is located beneath $x = 3$. This condition has the formal appearance $P(x \leq 3) = 0.4$. After calculating the restriction matrix $\mathbf{R}$ according to (27), the weight vector and kernel approximation of the maximum relative entropy density is available. Figure 5 shows the resulting approximations for both test distributions. Once again the importance of a proper choice of the dispersion parameter, with respect to the characteristics of the reference density, becomes obvious. The approximated maximum entropy density may serve as new reference density for incorporating additional boundary conditions due to external knowledge.

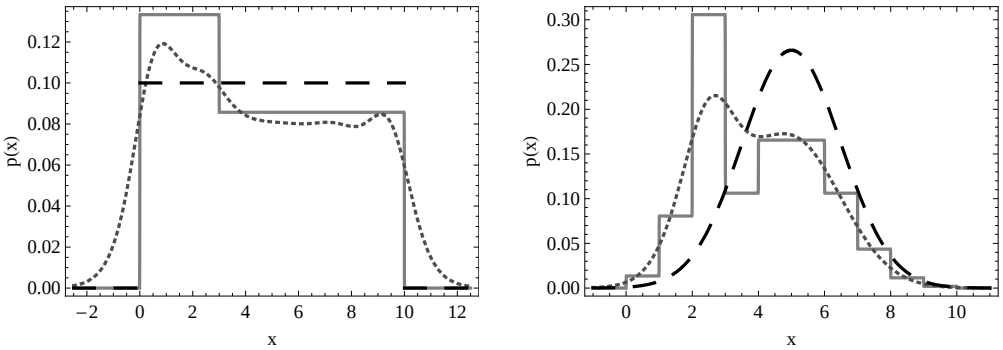

Figure 5: Constraint Uniform and Normal Probability Density Approximation – Boundary Condition $P(x \leq 3) = 0.4$

6. Maximum Entropy Density Approximation of Income under Constraints

In this section we present an example of quasi-continuous maximum entropy density estimation that illustrates the full potential of the method suggested in this paper. We want to approximate the maximum entropy distribution of income by use of various external information. We will show that in case of rare information the maximum entropy criterion fills the gaps and enables us to calculate the appropriate probability density.

Suppose we want to generate a maximum entropy density of the net-income of households in a certain period. Assume further that we are not provided with sample details, but only with some cumulated statistical quantities. In our example, the average income (in thousands of currency) may be given by $\mu = 20$. We are thereby provided with two clues; we know that the income has to be non-negative and we know its expectation. This is enough information to calculate a maximum entropy reference density in a first step.

Calculating a maximum entropy density is a *Lagrange*-Problem again. But contrary to the calculation we did before, we are dealing with a moment restriction here. The *Lagrange*-Function can be formulated as

$$\begin{aligned} L = &- \int_0^\infty p(x) \log\big[p(x)\big] dx + \lambda \left(\int_0^\infty p(x) dx - 1 \right) \\ &+ \lambda_1 \left(\int_0^\infty x p(x) dx - \mu \right). \end{aligned} \tag{29}$$

The first term on the right hand side of (29) is the entropy again. By calculating the functional derivative[4] and substituting $\lambda_0 = \lambda - 1$, one obtains

$$\frac{\eth L}{\eth p(x)} = -\log\big[p(x)\big] + \lambda_0 + \lambda_1 x, \tag{30}$$

which shows that after equating (30) to zero, the maximum entropy density has an exponential form. Differentiating the *Lagrange*-Function (29) with respect to λ and λ_1 provides solutions for the *Lagrange*-Multipliers and after some basic calculations one obtains the maximum entropy density

$$p(x) = \frac{1}{\mu} e^{-\frac{x}{\mu}}. \tag{31}$$

This density is indicated in figure 6 as dashed line. Furthermore, (31) serves as reference density, in case more information becomes available to be incorporated into a new maximum entropy probability density.

Now that we know the continuous maximum entropy density of the problem we can spread an adequate grid of kernels over the relevant interval. In this

[4] If a function value $\varphi(y)$ is assigned to the function $\varphi(x)$ at y, this assignment defines the functional $F[\varphi(x)] = \int \varphi(x)\delta(y-x)dx$. For the functional derivative one obtains $\frac{\eth F[\varphi(x)]}{\eth \varphi(x)} = \frac{\eth \varphi(y)}{\eth \varphi(x)} = \delta(y-x)$. If the functional is defined as $F[\varphi(x)] = \int f(\varphi(y))dy$, the derivative is obtained by applying the chain rule $\frac{\eth F[\varphi(x)]}{\eth \varphi(x)} = \int \frac{\eth f(\varphi(y))}{\eth \varphi(x)} dy = \int \frac{\partial f(\varphi(y))}{\partial \varphi(y)} \frac{\eth \varphi(y)}{\eth \varphi(x)} dy = \int \frac{\partial f(\varphi(y))}{\partial \varphi(y)} \delta(y-x) dy = \frac{\partial f(\varphi(x))}{\partial \varphi(x)}$.

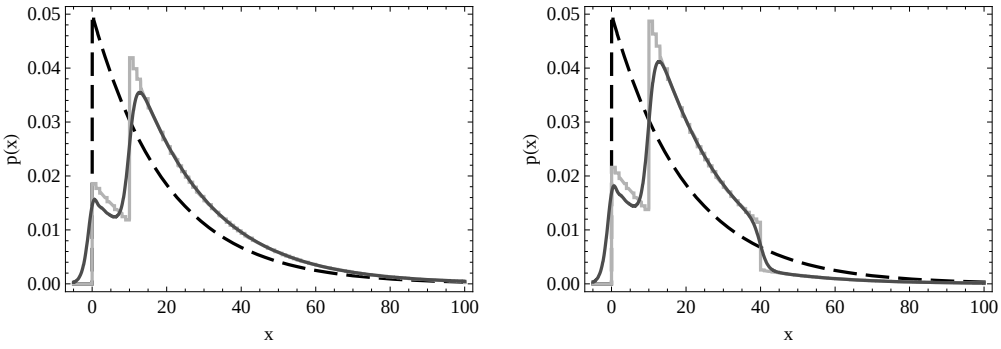

Figure 6: Maximum Entropy Density of Net Income under Constrains – Boundary Conditions $P(x \leq 10) = 0.15$ and $P(x \geq 40) = 0.05$

specific example we use one hundred kernels, equally distributed over the interval $[0, 100]$. Now suppose an additional information is available. Assume that only 15% of the investigated households have a net-income of 10 thousands of currency or less. This information can be incorporated easily as shown in the previous section. Figure 6 left shows the results. Again we have used the hyperbolic kernel (11) with dispersion $\nu = 1$ (dark gray) and $\nu = 0.001$ (light gray).

An inherent problem of smooth kernel functions becomes evident immediately. On the one hand, we are able to smoothly approximate the maximum entropy density, on the other hand some characteristics may be masked due to oversmoothing. This is the case for the discontinuity of the maximum entropy density at $x = 0$. A nearly rectangular kernel function can resolve this problem but again leaves us with a rough density approximation. This problem is known in kernel density estimation as well, where variable bandwidth selection is suggested as possible solution (cf. Silverman, 1986, sec. 2.6). An analogous modification for variable dispersion may result in better density approximations in this setup, but the present example is focused on incorporating boundary conditions.

Suppose an additional information is available. Let the proportion of hi-income households with a net-income of 40 thousands of currency or more be merely 5%. Now the previous kernel density approximation serves as maximum entropy reference and the new set of weights is calculated according to (23), as before. The results are given in figure 6 right. In fact, we can incorporate as many boundary conditions as become available to obtain a quasi-continuous approximation of the maximum entropy probability density with respect to the given information.

7. Conclusions

In this paper we provided a generic method for knowledge acquisition of continuous random variables under maximum entropy. We extended the common principle of MaxEnt to quasi continuous domains. The vehicle for this exten-

sion is the representation of probability densities with kernel density methods. A specific kernel with the ability to incorporate certain structural characteristics of the underlying problem was suggested. Its usefulness was demonstrated in a typical example of income estimation under constrains. Future work includes implementation of an algorithm in SPIRIT, allowing to process conditionals of the form $B|A$ in arbitrary combinations of continuous and discrete domains. Corresponding theoretical research is in progress.

References

Calabrese, P.G. (2004): Reflections on Logic & Probability in the Context of Conditionals. In: *Special Issue of LNCS*, Springer, Berlin, Heidelberg, New York, pp. 12–37.

Csiszàr, I. (1975): I-divergence Geometry of Probability Distributions and Minimization Problems. *The Annals of Probability*, 3(1):146–158.

Epanechnikov, V.A. (1969): Nonparametric Estimation of a Multidimensional Probability Density. *Theory of Probability and its Applications*, 14:153–158.

Harmon, P. and D. King (1985): *Expert Systems*. Wiley, New York.

Jensen, F.-V. (1996): *An Introduction to Bayesian Networks*. UCL Press Limited, London.

Kern-Isberner, G. (1998): Characterizing the principle of minimum cross-entropy within a conditional-logical framework. *Artificial Intelligence*, 98:169–208.

Kullback, S. and R.A. Leibler (1951): On information and sufficiency. *Annals of Mathematical Statistics*, 22:79–86.

Lauritzen, S.L. and D.J. Spiegelhalter (1988): Local computations with probabilities in graphical structures and their applications to expert systems. *Journal of the Royal Statistical Society B*, 50(2):415–448.

McCarthy, J. (1958): Programs with common sense. Symposium of Mechanization of Thought Processes, National Physical Laboratory, Teddington.

Newell, A.; J.C. Shaw; and H.A. Simon (1959): Report of a general problem-solving program. In: *Proceedings of the International Conference on Information Processing*, pp. 256–264.

Paris, J.B. and A. Vencovská (1990): A note on the inevitability of maximum entropy. *International Journal of Approximate Reasoning*, 4(3):193–224.

Paris, J.B. and A. Vencovská (1997): In Defence of the Maximum Entropy Inference Process. *International Journal of Approximate Reasoning*, 17(1):77–103.

Rödder, W. (2003): On the Measurability of Knowledge Acquisition and Query Processing. *International Journal of Approximate Reasoning*, 33(2):203–218.

Rödder, W. (2006): *Induktives Lernen bei unvollständigen Daten unter Wahrung des Entropieprinzips*. Tech. Rep. 388, FernUniversität in Hagen.

Rödder, W.; E. Reucher; and F. Kulmann (2006): Features of the Expert-System Shell SPIRIT. *Logic Journal of the IGPL*, 14(3):483–500.

Rödder, W.; E. Reucher; and F. Kulmann (2009): Where we Stand at Probabilistic Reasoning. Position paper presented at the 6th International Conference on Informatics in Control, Automation and Robotics – ICINCO in Milan. Vol. ICSO, pp. 394–397.

Reucher, E. and F. Kulmann (2007): Probabilistic Knowledge Processing and Remaining Uncertainty. In: *Proc. 20th International FLAIRS Conference - FLAIRS-20.* pp. 122–127.

Schramm, M. and B. Fronhöfer (2001): PIT: A System for Reasoning with Probabilities. Paper presented at the Workshop Uncertainty in AI at KI in Vienna.

Shore, J.E. and R.W. Johnson (1980): Axiomatic Derivation of the Principle of Maximum Entropy and the Principle of Minimum Cross Entropy. *IEEE Trans. Information Theory*, 26(1):26–37.

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

Singer, H. (2005): *Continuous-Discrete Unscented Kalman Filtering.* Tech. Rep. 384, FernUniversität in Hagen.

Die Diskussionspapiere ab Nr. 183 (1992) bis heute, können Sie im Internet unter http://www.fernuni-hagen.de/FBWIWI/ einsehen und zum Teil downloaden.
Die **Titel** der Diskussionspapiere von Nr 1 (1975) bis 182 (1991) können bei Bedarf in der Fakultät für Wirtschaftswissenschaft angefordert werden:
FernUniversität, z. Hd. Frau Huber oder Frau Mette, Postfach 940, 58084 Hagen .
Die Diskussionspapiere selber erhalten Sie nur in den Bibliotheken.

Nr	Jahr	Titel	Autor/en
322	2001	Spreading Currency Crises: The Role of Economic Interdependence	Berger, Wolfram Wagner, Helmut
323	2002	Planung des Fahrzeugumschlags in einem Seehafen-Automobilterminal mittels eines Multi-Agenten-Systems	Fischer, Torsten Gehring, Hermann
324	2002	A parallel tabu search algorithm for solving the container loading problem	Bortfeldt, Andreas Gehring, Hermann Mack, Daniel
325	2002	Die Wahrheit entscheidungstheoretischer Maximen zur Lösung von Individualkonflikten - Unsicherheitssituationen -	Mus, Gerold
326	2002	Zur Abbildungsgenauigkeit des Gini-Koeffizienten bei relativer wirtschaftlicher Konzentration	Steinrücke, Martin
327	2002	Entscheidungsunterstützung bilateraler Verhandlungen über Auftragsproduktionen - eine Analyse aus Anbietersicht	Steinrücke, Martin
328	2002	Die Relevanz von Marktzinssätzen für die Investitionsbeurteilung – zugleich eine Einordnung der Diskussion um die Marktzinsmethode	Terstege, Udo
329	2002	Evaluating representatives, parliament-like, and cabinet-like representative bodies with application to German parliament elections 2002	Tangian, Andranik S.
330	2002	Konzernabschluss und Ausschüttungsregelung im Konzern. Ein Beitrag zur Frage der Eignung des Konzernabschlusses als Ausschüttungsbemessungsinstrument	Hinz, Michael
331	2002	Theoretische Grundlagen der Gründungsfinanzierung	Bitz, Michael
332	2003	Historical background of the mathematical theory of democracy	Tangian, Andranik S.
333	2003	MCDM-applications of the mathematical theory of democracy: choosing travel destinations, preventing traffic jams, and predicting stock exchange trends	Tangian, Andranik S.
334	2003	Sprachregelungen für Kundenkontaktmitarbeiter – Möglichkeiten und Grenzen	Fließ, Sabine Möller, Sabine Momma, Sabine Beate

335	2003	A Non-cooperative Foundation of Core-Stability in Positive Externality NTU-Coalition Games	Finus, Michael Rundshagen, Bianca
336	2003	Combinatorial and Probabilistic Investigation of Arrow's dictator	Tangian, Andranik
337	2003	A Grouping Genetic Algorithm for the Pickup and Delivery Problem with Time Windows	Pankratz, Giselher
338	2003	Planen, Lernen, Optimieren: Beiträge zu Logistik und E-Learning. Festschrift zum 60 Geburtstag von Hermann Gehring	Bortfeldt, Andreas Fischer, Torsten Homberger, Jörg Pankratz, Giselher Strangmeier, Reinhard
339a	2003	Erinnerung und Abruf aus dem Gedächtnis Ein informationstheoretisches Modell kognitiver Prozesse	Rödder, Wilhelm Kuhlmann, Friedhelm
339b	2003	Zweck und Inhalt des Jahresabschlusses nach HGB, IAS/IFRS und US-GAAP	Hinz, Michael
340	2003	Voraussetzungen, Alternativen und Interpretationen einer zielkonformen Transformation von Periodenerfolgsrechnungen – ein Diskussionsbeitrag zum LÜCKE-Theorem	Terstege, Udo
341	2003	Equalizing regional unemployment indices in West and East Germany	Tangian, Andranik
342	2003	Coalition Formation in a Global Warming Game: How the Design of Protocols Affects the Success of Environmental Treaty-Making	Eyckmans, Johan Finus, Michael
343	2003	Stability of Climate Coalitions in a Cartel Formation Game	Finus, Michael van Ierland, Ekko Dellink, Rob
344	2003	The Effect of Membership Rules and Voting Schemes on the Success of International Climate Agreements	Finus, Michael J.-C., Altamirano-Cabrera van Ierland, Ekko
345	2003	Equalizing structural disproportions between East and West German labour market regions	Tangian, Andranik
346	2003	Auf dem Prüfstand: Die geldpolitische Strategie der EZB	Kißmer, Friedrich Wagner, Helmut
347	2003	Globalization and Financial Instability: Challenges for Exchange Rate and Monetary Policy	Wagner, Helmut

348	2003	Anreizsystem Frauenförderung – Informationssystem Gleichstellung am Fachbereich Wirtschaftswissenschaft der FernUniversität in Hagen	Fließ, Sabine Nonnenmacher, Dirk
349	2003	Legitimation und Controller	Pietsch, Gotthard Scherm, Ewald
350	2003	Controlling im Stadtmarketing – Ergebnisse einer Primärerhebung zum Hagener Schaufenster-Wettbewerb	Fließ, Sabine Nonnenmacher, Dirk
351	2003	Zweiseitige kombinatorische Auktionen in elektronischen Transportmärkten – Potenziale und Probleme	Pankratz, Giselher
352	2003	Methodisierung und E-Learning	Strangmeier, Reinhard Bankwitz, Johannes
353 a	2003	A parallel hybrid local search algorithm for the container loading problem	Mack, Daniel Bortfeldt, Andreas Gehring, Hermann
353 b	2004	Übernahmeangebote und sonstige öffentliche Angebote zum Erwerb von Aktien – Ausgestaltungsmöglichkeiten und deren Beschränkung durch das Wertpapiererwerbs- und Übernahmegesetz	Wirtz, Harald
354	2004	Open Source, Netzeffekte und Standardisierung	Maaß, Christian Scherm, Ewald
355	2004	Modesty Pays: Sometimes!	Finus, Michael
356	2004	Nachhaltigkeit und Biodiversität	Endres, Alfred Bertram, Regina
357	2004	Eine Heuristik für das dreidimensionale Strip-Packing-Problem	Bortfeldt, Andreas Mack, Daniel
358	2004	Netzwerkökonomik	Martiensen, Jörn
359	2004	Competitive versus cooperative Federalism: Is a fiscal equalization scheme necessary from an allocative point of view?	Arnold, Volker
360	2004	Gefangenendilemma bei Übernahmeangeboten? Eine entscheidungs- und spieltheoretische Analyse unter Einbeziehung der verlängerten Annahmefrist gem. § 16 Abs. 2 WpÜG	Wirtz, Harald
361	2004	Dynamic Planning of Pickup and Delivery Operations by means of Genetic Algorithms	Pankratz, Giselher

362	2004	Möglichkeiten der Integration eines Zeitmanagements in das Blueprinting von Dienstleistungsprozessen	Fließ, Sabine Lasshof, Britta Meckel, Monika
363	2004	Controlling im Stadtmarketing - Eine Analyse des Hagener Schaufensterwettbewerbs 2003	Fließ, Sabine Wittko, Ole
364	2004	Ein Tabu Search-Verfahren zur Lösung des Timetabling-Problems an deutschen Grundschulen	Desef, Thorsten Bortfeldt, Andreas Gehring, Hermann
365	2004	Die Bedeutung von Informationen, Garantien und Reputation bei integrativer Leistungserstellung	Prechtl, Anja Völker-Albert, Jan-Hendrik
366	2004	The Influence of Control Systems on Innovation: An empirical Investigation	Littkemann, Jörn Derfuß, Klaus
367	2004	Permit Trading and Stability of International Climate Agreements	Altamirano-Cabrera, Juan-Carlos Finus, Michael
368	2004	Zeitdiskrete vs. zeitstetige Modellierung von Preismechanismen zur Regulierung von Angebots- und Nachfragemengen	Mazzoni, Thomas
369	2004	Marktversagen auf dem Softwaremarkt? Zur Förderung der quelloffenen Softwareentwicklung	Christian Maaß Ewald Scherm
370	2004	Die Konzentration der Forschung als Weg in die Sackgasse? Neo-Institutionalistische Überlegungen zu 10 Jahren Anreizsystemforschung in der deutschsprachigen Betriebswirtschaftslehre	Süß, Stefan Muth, Insa
371	2004	Economic Analysis of Cross-Border Legal Uncertainty: the Example of the European Union	Wagner, Helmut
372	2004	Pension Reforms in the New EU Member States	Wagner, Helmut
373	2005	Die Bundestrainer-Scorecard Zur Anwendbarkeit des Balanced Scorecard Konzepts in nicht-ökonomischen Fragestellungen	Eisenberg, David Schulte, Klaus
374	2005	Monetary Policy and Asset Prices: More Bad News for ‚Benign Neglect“	Berger, Wolfram Kißmer, Friedrich Wagner, Helmut
375	2005	Zeitstetige Modellbildung am Beispiel einer volkswirtschaftlichen Produktionsstruktur	Mazzoni, Thomas

376	2005	Economic Valuation of the Environment	Endres, Alfred
377	2005	Netzwerkökonomik – Eine vernachlässigte theoretische Perspektive in der Strategie-/Marketingforschung?	Maaß, Christian Scherm, Ewald
378	2005	Diversity management`s diffusion and design: a study of German DAX-companies and Top-50-U.S.-companies in Germany	Süß, Stefan Kleiner, Markus
379	2005	Fiscal Issues in the New EU Member Countries – Prospects and Challenges	Wagner, Helmut
380	2005	Mobile Learning – Modetrend oder wesentlicher Bestandteil lebenslangen Lernens?	Kuszpa, Maciej Scherm, Ewald
381	2005	Zur Berücksichtigung von Unsicherheit in der Theorie der Zentralbankpolitik	Wagner, Helmut
382	2006	Effort, Trade, and Unemployment	Altenburg, Lutz Brenken, Anke
383	2006	Do Abatement Quotas Lead to More Successful Climate Coalitions?	Altamirano-Cabrera, Juan-Carlos Finus, Michael Dellink, Rob
384	2006	Continuous-Discrete Unscented Kalman Filtering	Singer, Hermann
385	2006	Informationsbewertung im Spannungsfeld zwischen der Informationstheorie und der Betriebswirtschaftslehre	Reucher, Elmar
386	2006	The Rate Structure Pattern: An Analysis Pattern for the Flexible Parameterization of Charges, Fees and Prices	Pleß, Volker Pankratz, Giselher Bortfeldt, Andreas
387a	2006	On the Relevance of Technical Inefficiencies	Fandel, Günter Lorth, Michael
387b	2006	Open Source und Wettbewerbsstrategie - Theoretische Fundierung und Gestaltung	Maaß, Christian
388	2006	Induktives Lernen bei unvollständigen Daten unter Wahrung des Entropieprinzips	Rödder, Wilhelm
389	2006	Banken als Einrichtungen zur Risikotransformation	Bitz, Michael
390	2006	Kapitalerhöhungen börsennotierter Gesellschaften ohne börslichen Bezugsrechtshandel	Terstege, Udo Stark, Gunnar
391	2006	Generalized Gauss-Hermite Filtering	Singer, Hermann

392	2006	Das Göteborg Protokoll zur Bekämpfung grenzüberschreitender Luftschadstoffe in Europa: Eine ökonomische und spieltheoretische Evaluierung	Ansel, Wolfgang Finus, Michael
393	2006	Why do monetary policymakers lean with the wind during asset price booms?	Berger, Wolfram Kißmer, Friedrich
394	2006	On Supply Functions of Multi-product Firms with Linear Technologies	Steinrücke, Martin
395	2006	Ein Überblick zur Theorie der Produktionsplanung	Steinrücke, Martin
396	2006	Parallel greedy algorithms for packing unequal circles into a strip or a rectangle	Timo Kubach, Bortfeldt, Andreas Gehring, Hermann
397	2006	C&P Software for a cutting problem of a German wood panel manufacturer – a case study	Papke, Tracy Bortfeldt, Andreas Gehring, Hermann
398	2006	Nonlinear Continuous Time Modeling Approaches in Panel Research	Singer, Hermann
399	2006	Auftragsterminierung und Materialflussplanung bei Werkstattfertigung	Steinrücke, Martin
400	2006	Import-Penetration und der Kollaps der Phillips-Kurve	Mazzoni, Thomas
401	2006	Bayesian Estimation of Volatility with Moment-Based Nonlinear Stochastic Filters	Grothe, Oliver Singer, Hermann
402	2006	Generalized Gauss-H ermite Filtering for Multivariate Diffusion Processes	Singer, Hermann
403	2007	A Note on Nash Equilibrium in Soccer	Sonnabend, Hendrik Schlepütz, Volker
404	2007	Der Einfluss von Schaufenstern auf die Erwartungen der Konsumenten - eine explorative Studie	Fließ, Sabine Kudermann, Sarah Trell, Esther
405	2007	Die psychologische Beziehung zwischen Unternehmen und freien Mitarbeitern: Eine empirische Untersuchung des Commitments und der arbeitsbezogenen Erwartungen von IT-Freelancern	Süß, Stefan
406	2007	An Alternative Derivation of the Black-Scholes Formula	Zucker, Max Singer, Hermann
407	2007	Computational Aspects of Continuous-Discrete Extended Kalman-Filtering	Mazzoni, Thomas
408	2007	Web 2.0 als Mythos, Symbol und Erwartung	Maaß, Christian Pietsch, Gotthard
409	2007	„Beyond Balanced Growth“: Some Further Results	Stijepic, Denis Wagner, Helmut

410	2007	Herausforderungen der Globalisierung für die Entwicklungsländer: Unsicherheit und geldpolitisches Risikomanagement	Wagner, Helmut
411	2007	Graphical Analysis in the New Neoclassical Synthesis	Giese, Guido Wagner, Helmut
412	2007	Monetary Policy and Asset Prices: The Impact of Globalization on Monetary Policy Trade-Offs	Berger, Wolfram Kißmer, Friedrich Knütter, Rolf
413	2007	Entropiebasiertes Data Mining im Produktdesign	Rudolph, Sandra Rödder, Wilhelm
414	2007	Game Theoretic Research on the Design of International Environmental Agreements: Insights, Critical Remarks and Future Challenges	Finus, Michael
415	2007	Qualitätsmanagement in Unternehmenskooperationen - Steuerungsmöglichkeiten und Datenintegrationsprobleme	Meschke, Martina
416	2007	Modernisierung im Bund: Akteursanalyse hat Vorrang	Pietsch, Gotthard Jamin, Leander
417	2007	Inducing Technical Change by Standard Oriented Evirnonmental Policy: The Role of Information	Endres, Alfred Bertram, Regina Rundshagen, Bianca
418	2007	Der Einfluss des Kontextes auf die Phasen einer SAP-Systemimplementierung	Littkemann, Jörn Eisenberg, David Kuboth, Meike
419	2007	Endogenous in Uncertainty and optimal Monetary Policy	Giese, Guido Wagner, Helmut
420	2008	Stockkeeping and controlling under game theoretic aspects	Fandel, Günter Trockel, Jan
421	2008	On Overdissipation of Rents in Contests with Endogenous Intrinsic Motivation	Schlepütz, Volker
422	2008	Maximum Entropy Inference for Mixed Continuous-Discrete Variables	Singer, Hermann
423	2008	Eine Heuristik für das mehrdimensionale Bin Packing Problem	Mack, Daniel Bortfeldt, Andreas
424	2008	Expected A Posteriori Estimation in Financial Applications	Mazzoni, Thomas
425	2008	A Genetic Algorithm for the Two-Dimensional Knapsack Problem with Rectangular Pieces	Bortfeldt, Andreas Winter, Tobias
426	2008	A Tree Search Algorithm for Solving the Container Loading Problem	Fanslau, Tobias Bortfeldt, Andreas
427	2008	Dynamic Effects of Offshoring	Stijepic, Denis Wagner, Helmut
428	2008	Der Einfluss von Kostenabweichungen auf das Nash-Gleichgewicht in einem nicht-kooperativen Disponenten-Controller-Spiel	Fandel, Günter Trockel, Jan
429	2008	Fast Analytic Option Valuation with GARCH	Mazzoni, Thomas
430	2008	Conditional Gauss-Hermite Filtering with Application to Volatility Estimation	Singer, Hermann
431	2008	Web 2.0 auf dem Prüfstand: Zur Bewertung von Internet-Unternehmen	Christian Maaß Gotthard Pietsch

432	2008	Zentralbank-Kommunikation und Finanzstabilität – Eine Bestandsaufnahme	Knütter, Rolf Mohr, Benjamin
433	2008	Globalization and Asset Prices: Which Trade-Offs Do Central Banks Face in Small Open Economies?	Knütter, Rolf Wagner, Helmut
434	2008	International Policy Coordination and Simple Monetary Policy Rules	Berger, Wolfram Wagner, Helmut
435	2009	Matchingprozesse auf beruflichen Teilarbeitsmärkten	Stops, Michael Mazzoni, Thomas
436	2009	Wayfindingprozesse in Parksituationen - eine empirische Analyse	Fließ, Sabine Tetzner, Stefan
437	2009	ENTROPY-DRIVEN PORTFOLIO SELECTION a downside and upside risk framework	Rödder, Wilhelm Gartner, Ivan Ricardo Rudolph, Sandra
438	2009	Consulting Incentives in Contests	Schlepütz, Volker
439	2009	A Genetic Algorithm for a Bi-Objective Winner-Determination Problem in a Transportation-Procurement Auction"	Buer, Tobias Pankratz, Giselher
440	2009	Parallel greedy algorithms for packing unequal spheres into a cuboidal strip or a cuboid	Kubach, Timo Bortfeldt, Andreas Tilli, Thomas Gehring, Hermann
441	2009	SEM modeling with singular moment matrices Part I: ML-Estimation of time series	Singer, Hermann
442	2009	SEM modeling with singular moment matrices Part II: ML-Estimation of sampled stochastic differential equations	Singer, Hermann
443	2009	Konsensuale Effizienzbewertung und -verbesserung – Untersuchungen mittels der Data Envelopment Analysis (DEA)	Rödder, Wilhelm Reucher, Elmar
444	2009	Legal Uncertainty – Is Hamonization of Law the Right Answer? A Short Overview	Wagner, Helmut
445	2009	Fast Continuous-Discrete DAF-Filters	Mazzoni, Thomas
446	2010	Quantitative Evaluierung von Multi-Level Marketingsystemen	Lorenz, Marina Mazzoni, Thomas
447	2010	Quasi-Continuous Maximum Entropy Distribution Approximation with Kernel Density	Mazzoni, Thomas Reucher, Elmar