A Stochastic Programming Model for Optimal Power Dispatch: Stability and Numerical Treatment

Nicole Gröwe and Werner Römisch
Humboldt-Universität zu Berlin
Fachbereich Mathematik
PSF 1297, 1086 Berlin

Abstract

The economic dispatch of electric power with uncertain demand is modeled as stochastic program with simple recourse. We analyze quantitative stability properties of the power dispatch model with respect to the L_1–distance of the marginal distribution functions of the demand vector. These stability results are used to derive asymptotic properties of the model if the (true) marginal distributions are replaced by smooth nonparametric estimates based on the kernel method. Finally, we discuss how smooth estimates can be used efficiently for the numerical treatment of simple recourse models by using nonlinear programming techniques. Numerical results are reported for Dantzig's Aircraft Allocation Problem.

1 Introduction

A model for the optimal dispatch of electric power to the units of an energy production system is considered that takes explicit account of the uncertainty of the electric power demand. This is done by introducing so-called expected recourse costs for under- and over-dispatching (similar to [5]) and leads to a large-scale convex stochastic program with simple recourse. For the uncertain demand, we suppose that a set of empirical data is given for the whole operating cycle. This motivates investigations in two directions:

K. Marti (ed.), *Stochastic Optimization*

(1) the stability analysis of stochastic programs with simple recourse to treat the situation of incomplete information (see Section 3), and

(2) the choice of appropriate estimators for the distribution functions of the random demand at each time interval.

The stability analysis is carried out by applying general results taken from [24], [25]. We also refer to [9] for a survey of stability results in stochastic programming, to [10], [16], [28], [31] for relevant statistical stability results and to [4] where the possibility of using density estimates in stochastic programming is outlined.

For our application, we motivate the use of nonparametric estimators for the distribution functions of the random demand. Nonparametric estimators based on the kernel method (see e.g. [7], [21]) are apparently favourable in our context, since they lead to stochastic programs having the property that the objective function is continuously differentiable and that function (gradient) values can be computed efficiently without numerical integration (cf. Section 5). Additionally, asymptotic properties of kernel-type estimators for distribution functions are comparable to those of the empirical distribution (see Section 4). In Section 4, we also outline how these asymptotic properties together with stability results of Section 3 lead to convergence rates for optimal solution sets of the power dispatch model if the sample size of observations for the demand tends to infinity. In Section 5, we discuss the numerical treatment of simple recourse models involving kernel-type estimators (for the unknown distribution) via standard nonlinear programming techniques. We report on the development of a program system and on numerical results for (a modified version of) Dantzig's Aircraft Allocation Problem.

2 A model for optimal power dispatch with uncertain demand

The problem of optimal power dispatch consists in allocating amounts of electric power to the generation units of an energy production system such that the total generation costs are minimal while the actual power demand is met and certain operational constraints (of the system) are satisfied. The system we shall consider comprises thermal power stations (*tps*), pumped hydro storage plants (*psp*) and an energy contract with connected systems.

The peculiarities of the system and the power dispatch model (cf. also [26]) are the following:

(a) *tps* and *psp* serve as base– and peak–load plants, respectively,

(b) the model is designed for a daily operating cycle and assumes that a unit commitment stage has been carried out before,

(c) the reserve levels and transmission losses are modeled by means of adjusted portions of the demand,

(d) the cost functions of the thermal plants are taken to be strictly convex and quadratic,

(e) the model takes explicit account of the uncertainty of the electric power demand by introducing an expected recourse action which is associated with the mismatch between scheduled generation and actual demand.

To give a more detailed description of the model, let K and M denote the number of *tps* and *psp*, respectively. Assume that the scheduling time–horizon consists of N intervals T_r $(r = 1, \ldots, N)$. Let $I_r \subset \{1, 2, \ldots, K\}$ denote the index set of available online *tps* within the time interval T_r $(r = 1, \ldots, N)$. The (unknown) outputs of the *tps* and *psp* at the interval T_r are y_{lr} $(l = 1, \ldots, K)$ and s_{jr} (generation mode of the *psp* $j \epsilon \{1, \ldots, M\}$), respectively. By w_{jr} we denote the input of the *psp* $j \epsilon \{1, \ldots, M\}$ during the pumping mode and by e_r the level of electric power which corresponds to the energy contract at time interval T_r. Denoting $x := (y, s, w, e)^T \epsilon \mathrm{R}^m$ with $m := N(K + 2M + 1)$ the model of optimal power dispatch reads

$$\min \{g(x) : x \epsilon C, Ax = z\} \tag{2.1}$$

Here, the total generation cost function g is convex quadratic on R^m and has the form

$$g(x) := \sum_{r=1}^{N} \left(\sum_{l \epsilon I_r} C_{lr}(y_{lr}) + d_r e_r \right)$$

where $C_{lr}(\cdot)$ are strictly convex quadratic cost functions for the *tps* l within T_r and d_r is the cost according to the contract at T_r ($l = 1, \ldots, K; r = 1, \ldots, N$).

The set $C \subset \mathrm{R}^m$ in (2.1) is a nonempty bounded convex polyhedron formed by the operational constraints of the system, e.g. bounds for the power output of the plants, balances between generation and pumping in the *psp*, balances for the *psp* over the whole time–horizon, fuel quotas of the *tps*.

The equation $Ax = z$ in (2.1) reads componentwise ($r = 1, \ldots, N$)

$$[Ax]_r := \sum_{l \epsilon I_r} y_{lr} + \sum_{j=1}^{M} (s_{jr} - w_{jr}) + e_r = z_r$$

and says that the total generated output meets the demand $z = (z_1, \ldots, z_N)^T$ at each time interval T_r. We consider the demand z as a random vector and denote by μ its probability distribution on R^N and by F_r the distribution function of z_r ($r = 1, \ldots, N$). Following [5] we introduce a penalty cost for the deviation of the scheduled output from the actual demand for under- and over-dispatching, respectively. To be more precise, we define

$$\tilde{Q}(t) := \sum_{r=1}^{N} \tilde{Q}_r(t_r) := \sum_{r=1}^{N} \begin{cases} q_r^+ t_r, & t_r \geq 0 \\ -q_r^- t_r, & t_r < 0 \end{cases} \qquad (t \epsilon \mathrm{R}^N) \tag{2.2}$$

where q_r^+ and q_r^- are the recourse costs for the under- and over-dispatching at time interval T_r ($r = 1, \ldots, N$), respectively. For a

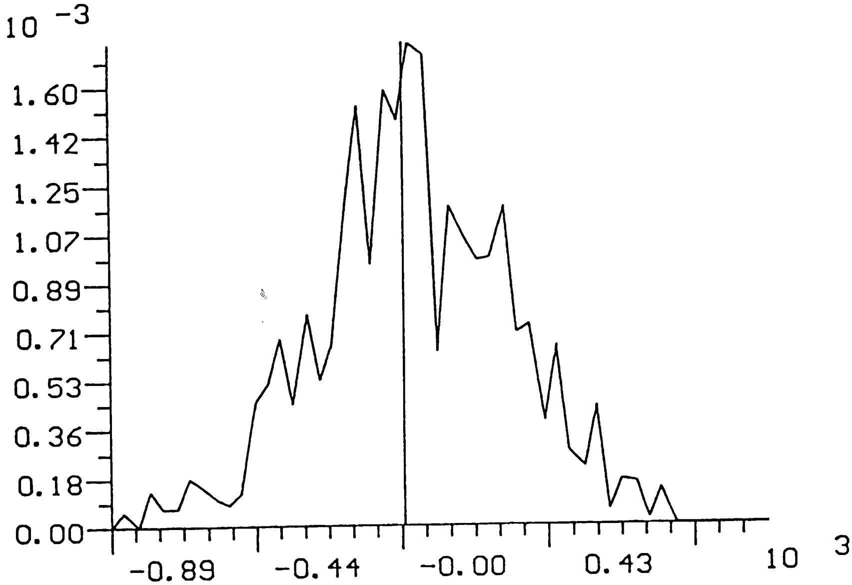


Fig. 2.1

discussion of the interpretation and choice of the recourse costs q_r^+ and q_r^- we refer to [5, pp. 181, 184].

Adding the expected recourse costs $\mathrm{E}\left[\tilde{Q}(z - Ax)\right]$ to the deterministic cost function g we arrive at the following stochastic power dispatch model :

$$\min \left\{ g(x) + \sum_{r=1}^{N} \int_{\mathbb{R}} \tilde{Q}_r(t - [Ax]_r)\, dF_r(t) \,:\, x \epsilon C \right\} \tag{2.3}$$

Similar power dispatch models are considered in [3] and [5]. More information on various aspects of power dispatch can be found in the volume [32] and in several papers of [11], Part IV. It is well-known that (2.3) is a particular stochastic program with simple recourse (see [13]). (2.3) is a large scale convex nonlinear program having linear constraints. If all distribution functions F_r $(r = 1, \ldots, N)$ have densities, the objective function of (2.3) is continuously differentiable (cf. [13], p.56 ff.). To give an idea how large (2.3) is for real-life applications, we mention that for the energy production system of East Germany typically $K := 26$, $M := 5$ which leads together with (say)

$N := 24$ to $m := 888$. For the uncertain demand, a set of empirical data (in practice, a medium - sized sample) is given (see also Section 4). It is suggested in [3] and [5] that the distribution of the random demand (at each time interval) can be chosen as (trimmed) normal. However, our tests with the available empirical data did not justify this hypothesis (especially for all time-intervals in the day-time). As an example, Figure 2.1 shows an estimate for the density of the centered demand during the hour 1 p.m. – 2 p.m. (of a day of normal category). The estimate is obtained according to formula (4.1) and by using the triangular kernel (with $b_n = 30$ and $n = 436$). Finally, we preferred the use of nonparametric estimators for the (unknown) distribution functions. This is described in more detail in Section 4 and 5.

3 Stability analysis

Consider the following (convex) stochastic program with simple recourse and random right–hand side

$$\min\{g(x) + Q_\mu(Ax) : x \epsilon C\} \tag{3.1}$$

where

$$\begin{aligned} Q_\mu(\chi) &:= \int_{R^s} \tilde{Q}(z-\chi)\,\mu(dz) \\ \tilde{Q}(t) &:= \min\{q^T y : (I,-I)y = t, y \in R^{2s}_+\} \end{aligned} \tag{3.2}$$

We assume that g is a real-valued convex function on R^m, C is a nonempty, closed, convex subset of R^m, A is a $s \times m$ matrix, $q \epsilon \mathrm{R}^{2s}$ and μ is a (Borel) probability measure on R^s.

Under the basic assumptions

$(A1)$ $\qquad q^+ + q^- \epsilon \mathrm{R}^s_+$, where $q = \begin{pmatrix} q^+ \\ q^- \end{pmatrix}$

$(A2)$ $\qquad \int_{\mathrm{R}^s} \|z\|\,\mu(dz) < +\infty$ $\quad$ ($\|\cdot\|$ denoting the Euclidean norm on R^s),

$\tilde{Q}$ has the representation (2.2) and Q_μ is a real-valued separable convex function on R^s having the shape

$$Q_\mu(\chi) = \sum_{r=1}^{s} \left[q_r^+(\bar{\mu}_r - \chi_r) - (q_r^+ + q_r^-) \int_{-\infty}^{\chi_r} (t - \chi_r)\, dF_r(t) \right] \tag{3.3}$$

where F_r $(r = 1, \ldots, s)$ are the one–dimensional marginal distribution functions of μ and $\bar{\mu}_r$ $(r = 1, \ldots, s)$ their mean values (cf. e.g.[13]).

This section deals with the stability of problem (3.1), when the underlying probability measure μ is subjected to (small) perturbations. Here, stability means that the optimal value $\varphi(\mu)$ and the optimal solution set $\psi(\mu)$ of problem (3.1) enjoy certain continuity properties with respect to variations of μ in a (properly selected) subset of probability measures on R^s equipped with a suitable distance ("probability metric").

To select a set of probability measures and a metric, we notice that, due to the separability structure of Q_μ (see (3.3)), problem (3.1) only depends on the marginal distributions μ_r $(r = 1, \ldots, s)$ of μ. Hence, we may assume that μ has independent one-dimensional marginal distributions.

Therefore we consider the following metric space $(\mathcal{M}(\mathrm{R}^s), d)$ where $\mathcal{M}(\mathrm{R}^s) := \{\nu : \nu$ is a probability measure on R^s having independent one-dimensional marginals ν_r $(r = 1, \ldots, s)$ and $\int_{\mathrm{R}} |t|\, \nu_r(dt) < \infty$ $(r = 1, \ldots, s)\}$ and

$$d(\nu_1, \nu_2) := \sum_{r=1}^{s} \int_{\mathrm{R}} |F_{1r}(t) - F_{2r}(t)|\, dt, \tag{3.4}$$

F_{1r} and F_{2r} $(r = 1, \ldots, s)$ denoting the one-dimensional marginal distribution functions of $\nu_1, \nu_2 \in \mathcal{M}(\mathrm{R}^s)$.

The first stability result asserts upper semicontinuity of the optimal set mapping $\psi(\cdot)$ and a local Lipschitz property of the optimal value function $\varphi(\cdot)$ of (3.1) at $\mu \in \mathcal{M}(\mathrm{R}^s)$.

Theorem 3.1

Fix $\mu \in \mathcal{M}(\mathrm{R}^s)$, suppose (A1) and let $\psi(\mu)$ be nonempty and bounded.

Then there exist constants $L > 0$, $\delta > 0$ such that

$$|\varphi(\mu) - \varphi(\nu)| \leq L\, d(\mu, \nu)$$

and $\psi(\nu)$ is nonempty whenever $\nu \in \mathcal{M}(\mathrm{R}^s)$, $d(\mu, \nu) < \delta$. The set-valued mapping $\psi(\cdot)$ from $(\mathcal{M}(\mathrm{R}^s), d)$ into R^m is upper semicontinuous at μ, i.e. for each open set $\mathcal{U}$ containing $\psi(\mu)$ there exists $\delta_0 > 0$ such that $\psi(\nu) \subset \mathcal{U}$ whenever $\nu \in \mathcal{M}(\mathrm{R}^s)$, $d(\mu, \nu) < \delta_0$.

Proof

We apply Theorem 2.4 and Remark 2.5 of [25] and obtain the assertion by using the Wasserstein metric W_1 (cf. Section 2 of [25]) instead of the metric d. It remains to notice that W_1 coincides with d on $\mathcal{M}(\mathrm{R}^s)$ if R^s is equipped with the norm $\|z\|_1 := \sum_{r=1}^{s} |z_i|$ $(z \in \mathrm{R}^s)$ (see Remark 2.11 in [25]).•

The following example shows that, under the assumptions of the above Theorem, ψ is in general not lower semicontinuous at μ even if μ has a density. Recall that lower semicontinuity of ψ at $\mu \in \mathcal{M}(\mathrm{R}^s)$ means that for each open set $\mathcal{U}$ satisfying $\mathcal{U} \cap \psi(\mu) \neq \emptyset$ there exists $\delta_0 > 0$ such that $\mathcal{U} \cap \psi(\nu) \neq \emptyset$ whenever $d(\mu, \nu) < \delta_0$.

Example 3.2

In (3.1), let $m = s = 1$, $g(x) \equiv 0$, $C := \mathrm{R}$, $q := (1, 1)^{\mathrm{T}}$, $A := 1$. Consider the family ν_ε, $\varepsilon \in [0, 1]$ of probability measures on R given by their densities

$$\Theta_\varepsilon(t) := \begin{cases} 1 - \varepsilon & , t \in [-1, -\frac{1}{2}] \\ \varepsilon & , t \in (-\frac{1}{2}, \frac{1}{2}) \\ 1 - \varepsilon & , t \in [\frac{1}{2}, 1] \\ 0 & , \text{otherwise.} \end{cases}$$

Then $\tilde{Q}(t) = |t| \quad (t \in \mathrm{R})$, (A1) is satisfied and $\nu_\varepsilon \in \mathcal{M}(\mathrm{R}^s)$ for all $\varepsilon \in [0, 1]$. We obtain from (3.3) that

$$\begin{aligned} \psi(\nu_\varepsilon) &= \{0\} \qquad \text{for all } \varepsilon \in (0,1], \\ \psi(\nu_0) &= [-\frac{1}{2},\frac{1}{2}] \qquad \text{and} \\ \varphi(\nu_\varepsilon) &= \frac{3}{4}-\frac{\varepsilon}{2} \qquad \text{for } \varepsilon \in [0,1]. \end{aligned}$$

Furthermore, we have $d(\nu_\varepsilon,\nu_0) \le \varepsilon$. Hence, we conclude that ψ is not lower semicontinuous at $\mu := \nu_0$.

Under a certain positivity condition for the one–dimensional marginal densities of μ, we now show that (at least) the sets of optimal tenders behave locally Hölder continuous at μ.

Theorem 3.3

Fix $\mu \in \mathcal{M}(\mathrm{R}^s)$, suppose $q_r^+ + q_r^- > 0$ $(r = 1,\dots,s)$ and let $\psi(\mu)$ be nonempty and bounded. Assume, in addition, that the one–dimensional marginal densities $\Theta_r(r = 1,\dots,s)$ of μ exist and that there exist real numbers $a_r, b_r, \varepsilon > 0$ $(r = 1,\dots,s)$ such that the conditions $A(\psi(\mu)) \subseteq \times_{r=1}^s(a_r,b_r)$ and $\Theta_r(t) \ge \varepsilon_r$ for all $t \in (a_r,b_r)$ and $r = 1,\dots,s$ hold.

Then the set $\{Ax : x \in \psi(\mu)\}$ is a singleton and there exist constants $L_1 > 0$ and $\delta_1 > 0$ such that for all $\nu \in \mathcal{M}(\mathrm{R}^s)$ with $d(\mu,\nu) < \delta_1$ we have

$$\sup_{x\in\psi(\nu)} \|Ax - \chi_*\| \le L_1 d(\mu,\nu)^{1/2}$$

where

$$\{\chi_*\} = \{Ax : x \in \psi(\mu)\}$$

Proof

We want to apply Theorem 4.3 in [24]. To this end we have to show that Q_μ is strongly convex on $V := \times_{r=1}^s(a_r,b_r)$. Let $\lambda \in [0,1]$ and χ,$\tilde{\chi}$ be chosen such that $\chi_r, \tilde{\chi}_r \in (a_r,b_r)$ for all $r = 1,\dots,s$. Then we obtain from (3.3)

$$Q_\mu(\lambda\chi + (1-\lambda)\tilde{\chi}) = \lambda Q_\mu(\chi) + (1-\lambda)Q_\mu(\chi) - G(\chi,\tilde{\chi};\lambda),$$

where

$$G(\chi,\tilde{\chi};\lambda) := \sum_{r=1}^{s}(q_r^+ + q_r^-)\{\lambda[h_r(\lambda\chi_r + (1-\lambda)\tilde{\chi}_r;\chi_r) - h_r(\chi_r;\chi_r)] \\ +(1-\lambda)[h_r(\lambda\chi_r + (1-\lambda)\tilde{\chi}_r;\tilde{\chi}_r) - h_r(\tilde{\chi}_r;\tilde{\chi}_r)]\}$$

and $h_r(u,v) := \int_{-\infty}^{u}(t-v)\Theta_r(t)\,dt \quad ,u,v \in \mathrm{R},\ r = 1,\dots,s$. Now, let $r\in\{1,\dots,s\}$ and assume without loss of generality that $\chi_r < \tilde{\chi}_r$. Then we have, setting $\chi_r(\lambda) := \lambda\chi_r + (1-\lambda)\tilde{\chi}_r$,

$$\begin{aligned} h_r(\chi_r(\lambda);\chi_r) - h_r(\chi_r;\chi_r) &= \int_{\chi_r}^{\chi_r(\lambda)}(t-\chi_r)\Theta_r(t)\,dt \\ &\geq \varepsilon_r \int_{\chi_r}^{\chi_r(\lambda)}(t-\chi_r)\,dt \\ &= \frac{\varepsilon_r}{2}(\chi_r(\lambda)-\chi_r)^2 \\ &= \frac{\varepsilon_r}{2}(1-\lambda)^2(\chi_r-\tilde{\chi}_r)^2. \end{aligned}$$

Analogously, we get the inequality

$$h_r(\chi_r(\lambda);\tilde{\chi}_r) - h_r(\tilde{\chi}_r;\tilde{\chi}_r) \geq \frac{\varepsilon_r}{2}\lambda^2(\chi_r-\tilde{\chi}_r)^2.$$

Altogether, we obtain

$$\begin{aligned} Q_\mu(\lambda\chi + (1-\lambda)\tilde{\chi}) &\leq \lambda Q_\mu(\chi) + (1-\lambda)Q_\mu(\tilde{\chi}) \\ &\quad -\frac{1}{2}\sum_{r=1}^{s}(q_r^+ + q_r^-)\varepsilon_r\lambda(1-\lambda)(\chi_r-\tilde{\chi}_r)^2 \\ &\leq \lambda Q_\mu(\chi) + (1-\lambda)Q_\mu(\tilde{\chi}) \\ &\quad -\frac{\kappa}{2}\lambda(1-\lambda)\|\chi-\tilde{\chi}\|^2 \end{aligned}$$

where $\kappa := \min_{r=1,\dots,s}(q_r^+ + q_r^-)\varepsilon_r > 0$ and Q_μ is strongly convex on V. Setting $\lambda = \frac{1}{2}$, this together with the convexity of g implies in particular

$$g(x) + Q_\mu(Ax) \geq \varphi(\mu) + \frac{\kappa}{4}\|Ax - Ax_*\|^2$$

for all $x \in C$ and $x_* \in \psi(\mu)$. This proves that the set $\{Ax : x \in \psi(\mu)\}$ is a singleton. The assertion now follows from Theorem 4.3 in [24] with the same argument concerning the metrics as in the proof of Theorem 3.1. •

Remark 3.4

Example 4.5 in [24] shows that the exponent 1/2 on the right-hand side in the assertion of Theorem 3.3 is optimal, and our Example 3.2 shows that the assertion of Theorem 3.3 is not true if $A(\psi(\mu))$ is not contained in the support of μ.

Theorem 3.5

Let, in addition to the assumptions of Theorem 3.3, g be convex quadratic and C be polyhedral.

Then there exist constants $L_2 > 0$ and $\delta_2 > 0$ such that

$$d_H(\psi(\mu), \psi(\nu)) \leq L_2 \, d(\mu, \nu)^{1/2}$$

whenever $\nu \in \mathcal{M}(\mathrm{R}^s)$, $d(\mu, \nu) < \delta_2$. (Here, d_H denotes the Hausdorff distance on subsets of R^m.)

Proof

The result follows from Theorem 2.7 in [25] by repeating the strong-convexity and metric arguments in the proof of Theorem 3.3. •

Remark 3.6

The discussion in Remark 2.9 in [25] shows that Theorem 3.5 does not remain true for a general convex constraint set C and for a general convex (deterministic) objective function g. Fortunately, the above results cover the situation of the power dispatch model in Section 2, if the marginal densities of μ fulfil the positivity condition imposed in Theorem 3.3.

Extensions of our stability results to more general recourse models may be found in [24], [25] and in the papers [14], [23], where qualitative stability results for general recourse problems are obtained with respect to the topology of weak convergence on the set of all probability measures (cf. [2]).

4 Smooth nonparametric distribution estimates and asymptotic analysis

In this section, we consider nonparametric estimates for univariate distribution functions and analyze their rates of convergence. In particular, we study smooth estimates which are obtained by integrating a density estimator of the kernel type. This is motivated by the stochastic power dispatch model (2.3), since there the distribution functions of the uncertain electric power demand at each time interval have to be estimated and since smooth estimates lead to a smooth nonlinear programming problem.

Let $(X_i)_{i \in \mathrm{N}}$ be a sequence of independent and identically distributed real–valued random variables with common distribution function F. By $\mathcal{F}_n$ we denote the empirical distribution function for sample size $n \in \mathrm{N}$, i.e.

$$\mathcal{F}_n(x) := n^{-1} \sum_{i=1}^{n} I_0(x - X_i) \qquad (x \in \mathrm{R})$$

where I_0 is the indicator function of the interval $[0, +\infty)$. A nonnegative function k having the property $\int_{\mathrm{R}} k(x)\, dx = 1$ is called *kernel*. Suppose (b_n) is a sequence of positive numbers (*"smoothing parameters"*) tending to zero. Then

$$\hat{f}_n(x) := (nb_n)^{-1} \sum_{i=1}^{n} k\left((x - X_i)b_n^{-1}\right) \quad (x \in \mathrm{R}) \tag{4.1}$$

is a *kernel estimate for the density* $f := F'$ and the corresponding kernel estimate of F is

$$\hat{\mathcal{F}}_n(x) := \int_{-\infty}^{x} \hat{f}_n(t)\, dt = \int_{-\infty}^{\infty} \mathcal{K}\left((x - t)b_n^{-1}\right)\, d\mathcal{F}_n(t) \tag{4.2}$$

where $\mathcal{K}(x) := \int_{-\infty}^{x} k(t)\, dt$. $\hat{\mathcal{F}}_n$ may be interpreted as a smoothed version of the empirical distribution function $\mathcal{F}_n$. For more information and background on kernel-type estimators it is referred to [7],[21] and [30].

In the following, a kernel k is called *class s kernel* for some $s \in \mathrm{N}$ if

$$\begin{aligned} \int_{\mathrm{R}} x^i k(x)\, dx &= 0 \quad , i = 1, \ldots, s-1, \\ \int_{\mathrm{R}} |x|^s k(x)\, dx &< \infty. \end{aligned}$$

If, in addition, the kernel k is symmetric (about 0), we need only consider even values of s. In that case, it is known that class 4 kernels of compact support do not exist (see [7, p.100], [30, p.66]). For a discussion of class s kernels which are possibly negative-valued see [7, Chapter 7.2].

Some kernels, which are, in fact, all symmetric class 2 kernels and will be considered in Section 5, and their cumulative distribution functions $\mathcal{K}$ are now listed.

Epanechnikov

$$k(t) = \begin{cases} \frac{3}{4\sqrt{5}}(1 - t^2/5) & (|t| \le \sqrt{5}) \\ 0 & otherwise \end{cases}$$

$$\mathcal{K}(t) = \begin{cases} 0 & (t \le -\sqrt{5}) \\ \frac{1}{2} + \frac{3t}{4\sqrt{5}} - \frac{t^3}{20\sqrt{5}} & (-\sqrt{5} < t < \sqrt{5}) \\ 1 & (t \ge \sqrt{5}) \end{cases}$$

Biweight

$$k(t) = \begin{cases} \frac{15}{16}(1 - t^2)^2 & (|t| \le 1) \\ 0 & otherwise \end{cases}$$

$$\mathcal{K}(t) = \begin{cases} 0 & (t \le -1) \\ \frac{1}{2} + \frac{15}{16}t - \frac{5}{8}t^3 + \frac{3}{16}t^5 & (-1 < t < 1) \\ 1 & (t \ge 1) \end{cases}$$

Triangular

$$k(t) = \begin{cases} 1 - |t| & (|t| \le 1) \\ 0 & otherwise \end{cases}$$

$$\mathcal{K}(t) = \begin{cases} 0 & (t \leq -1) \\ \frac{1}{2} + t + \frac{1}{2}t^2 & (-1 < t \leq 0) \\ \frac{1}{2} - t + \frac{1}{2}t^2 & (0 < t < 1) \\ 1 & (t \geq 1) \end{cases}$$

Let $C_b^s := C_b^s(\mathrm{R})$ denote the class of s–times continuously differentiable functions F such that $F^{(s)}$ is bounded on R. The following auxiliary result gives an estimate for the distance

$$\|\hat{\mathcal{F}}_n - F\|_\infty := \sup_{x \epsilon \mathrm{R}} |\hat{\mathcal{F}}_n(x) - F(x)|$$

where $\hat{\mathcal{F}}_n$ is the kernel type estimate (4.2) for a sufficiently smooth distribution function F. Its proof follows essentially the ideas developed in [34].

Lemma 4.1

Let $s \epsilon \mathrm{N}$ and assume that $F \epsilon C_b^s$ and k is a class s kernel. Then

$$\|\hat{\mathcal{F}}_n - F\|_\infty \leq C b_n^s + \|\mathcal{F}_n - F\|_\infty \qquad (n \epsilon \mathrm{N})$$

where

$$C := \frac{1}{s!}\|F^{(s)}\|_\infty \int_{\mathrm{R}} |x|^s k(x)\, dx.$$

Proof

From [34], Lemma 2.3 we have the estimate

$$\|\hat{\mathcal{F}}_n - F\|_\infty \leq \|\mathcal{F}_n - F\|_\infty + \sup_{x \epsilon \mathrm{R}} |\mathrm{E}\,\hat{\mathcal{F}}_n(x) - F(x)|$$

where E denotes the mean value with respect to the sample probability space. It remains to derive an estimate for the second term on the right-hand side. For each $x \epsilon \mathrm{R}$ we obtain by Taylor's expansion and using that k is a class s kernel

$$\begin{aligned} \mathrm{E}\,\hat{\mathcal{F}}_n(x) - F(x) &= \int_{\mathrm{R}} \mathcal{K}((x-t)b_n^{-1})\, dF(t) - F(x) \\ &= \int_{\mathrm{R}} [F(x - tb_n) - F(x)]\, k(t)\, dt \\ &= \int_{\mathrm{R}} (-tb_n)^s \frac{F^{(s)}(x - \theta t b_n)}{s!} k(t)\, dt \end{aligned}$$

with some $\theta \in (0,1)$ depending possibly on x, t and n. This finally yields the desired estimate.•

Another, but similar, technique for estimating

$$\|\hat{\mathcal{F}}_n - F\|_\infty$$

can be found in [29], Chapt. 23.2. The next result now asserts rates for the almost sure and mean convergence of $\left(\|\hat{\mathcal{F}}_n - F\|_\infty\right)_{n \in \mathrm{N}}$.

Proposition 4.2

Let $s \in \mathrm{N}$, assume that $F \in C_b^s$ and k is a class s kernel. Suppose that (b_n) is chosen such that

$$\limsup_{n\to\infty} b_n^s n^{\frac{1}{2}} < \infty.$$

(a) The following law of the iterated logarithm (LIL) holds

$$\limsup_{n\to\infty} (2n/\log\log n)^{\frac{1}{2}} \|\hat{\mathcal{F}}_n - F\|_\infty \le 1$$

almost surely.

(b)

$$\limsup_{n\to\infty} n^{\frac{1}{2}}\, \mathrm{E}(\|\hat{\mathcal{F}}_n - F\|_\infty < \infty.$$

Proof

Part (a) of the assertion follows from Lemma 4.1 and the Smirnov-Chung LIL for empirical distribution functions (see [12, Chapt.6.8], [29]; cf. also the proof of Theorem 3.2 in [34] which in fact is the particular case of our result for $s = 2$).

To establish (b), we again use Lemma 4.1. It remains to apply the following known result for empirical distribution functions (cf. [12, Chapt.3.3]):

$$\mathrm{E}\left[\|\mathcal{F}_n - F\|_\infty\right] \le 2\,\mathrm{E}\left[\sup_{t \in \mathrm{R}} \left[\mathcal{F}_n(t) - F(t)\right]\right] = O(n^{-\frac{1}{2}}) \quad \bullet$$

Less seems to be known for the convergence of $(\hat{\mathcal{F}}_n)$ to F in terms of the L_1-distance. Next we give a speed-of-convergence result in this direction for the case that the kernel k has compact support.

Proposition 4.3

In addition to the hypotheses of Proposition 4.2, let the kernel k have compact support and assume that the p-th absolute moment $M_p := \int_{\mathrm{R}} |t|^p \, dF(t)$ be finite for some $p > 1$. Then

$$\limsup_{n\to\infty} n^{\frac{1}{2}(1-\frac{1}{p})} \mathrm{E}\left(\int_{\mathrm{R}} |\hat{\mathcal{F}}_n(t) - F(t)| \, dt\right) < \infty.$$

Proof

Let $\varepsilon_n := n^{-\frac{1}{2}}$, $c_n := \varepsilon_n^{-\frac{1}{p}}$ for each $n \in \mathrm{N}$. Then we obtain

$$\begin{aligned} n^{\frac{1}{2}(1-\frac{1}{p})} \mathrm{E}\left(\int_{\mathrm{R}} |\hat{\mathcal{F}}_n(t) - F(t)| \, dt\right) &\leq c_n^{p-1} \mathrm{E}\Big(\int_{-\infty}^{-c_n} (\hat{\mathcal{F}}_n(t) + F(t)) \, dt \\ &+ 2c_n \|\hat{\mathcal{F}}_n - F\|_\infty \\ &+ \int_{c_n}^{\infty} ((1 - \hat{\mathcal{F}}_n(t)) + (1 - F(t))) \, dt\Big) \\ &= c_n^{p-1}\Big[\int_{c_n}^{\infty} (F(-t) + (1 - F(t))) \, dt \\ &+ \int_{c_n}^{\infty} (\mathrm{E}\,\hat{\mathcal{F}}_n(-t) + (1 - \mathrm{E}\,\hat{\mathcal{F}}_n(t)))dt\Big] \\ &+ 2\varepsilon_n^{-1} \mathrm{E}(\|\hat{\mathcal{F}}_n - F\|_\infty). \end{aligned}$$

In view of Proposition 4.2(b) we need to study only the first two terms in the last row. By Markov's inequality we have

$$F(-t) + (1 - F(t)) \leq 2M_p t^{-p} \quad \text{for all} \quad t > 0 \, .$$

Hence, for the first integral we obtain the estimate

$$c_n^{p-1} \int_{c_n}^{\infty} (F(-t) + (1 - F(t))) \, dt \leq c_n^{p-1} 2M_p \int_{c_n}^{\infty} t^{-p} \, dt = \frac{2M_p}{p-1}.$$

It remains to estimate the second integral. To this end, let I denote the compact support of k and let $n_o \in \mathrm{N}$ be such that $c_n \geq 1$ and $b_n I \subseteq [-\frac{1}{2}, \frac{1}{2}]$ for all $n \geq n_o$.

Let $n \geq n_o$ and $x \geq 1$. We obtain by Markov's inequality

$$\begin{aligned} \mathrm{E}\,\hat{\mathcal{F}}_n(-x) &= \int_{\mathrm{R}} F(-x - tb_n)k(t) \, dt = \int_I F(-(x + tb_n))k(t) \, dt \\ &\leq \int_I M_p(x + tb_n)^{-p} k(t) \, dt \leq M_p\Big(x - \frac{1}{2}\Big)^{-p} \leq M_p\Big(\frac{x}{2}\Big)^{-p}. \end{aligned}$$

Analogously, we have the estimate

$$1 - \mathrm{E}\,\hat{\mathcal{F}}_n(x) = \int_I (1 - F(x - tb_n))k(t)\,dt \le M_p\left(\frac{x}{2}\right)^{-p}.$$

The last two inequalities yield

$$c_n^{p-1}\int_{c_n}^{\infty}(\mathrm{E}\,\hat{\mathcal{F}}_n(-t)+(1-\hat{\mathcal{F}}_n(t)))\,dt \le 2^{p+1}M_p c_n^{p-1}\int_{c_n}^{\infty} t^{-p}\,dt = \frac{2^{p+1}M_p}{p-1}.$$

This completes the proof.•

Remark 4.4

It is clear from the proof of Proposition 4.3 that stronger moment conditions for F lead to more comfortable rates of convergence for

$$\mathrm{E}\left(\int_{\mathrm{R}} |\hat{\mathcal{F}}_n(t) - F(t)|\,dt\right) \quad \text{as } n \to \infty.$$

It is not known whether the rate $O\left(n^{-\frac{1}{2}}\right)$ as $n \to \infty$ can be attained as in Proposition 4.2(b).

If F is the uniform distribution on $[0,1]$ and $\mathcal{F}_n$ the corresponding empirical distribution function, then it is shown in [8, Chapter 6] that there exists a constant $C > 0$ such that

$$C^{-1}n^{-\frac{1}{2}} \le \mathrm{E}\left(\int_{\mathrm{R}} |\mathcal{F}_n(t) - F(t)|\,dt\right) \le Cn^{-\frac{1}{2}} \quad \text{for all } n.$$

Remark 4.5

The convergence results show that the speed of convergence of perturbed (or smoothed) empirical distribution function $\hat{\mathcal{F}}_n$ (obtained by the kernel method) to the (unknown) distribution function F is essentially the same as for $\mathcal{F}_n$ if the sequence (b_n) of smoothing parameters is chosen appropriately. If $F \in C_b^s$ and if a class s kernel is used, then $b_n := n^{-\alpha}$ with $\alpha \ge \frac{1}{2s}$ is an appropriate choice. For a thorough discussion of the choice of smoothing parameters (e.g. also for small sample sizes) we refer to [1] and [30] (see also Section

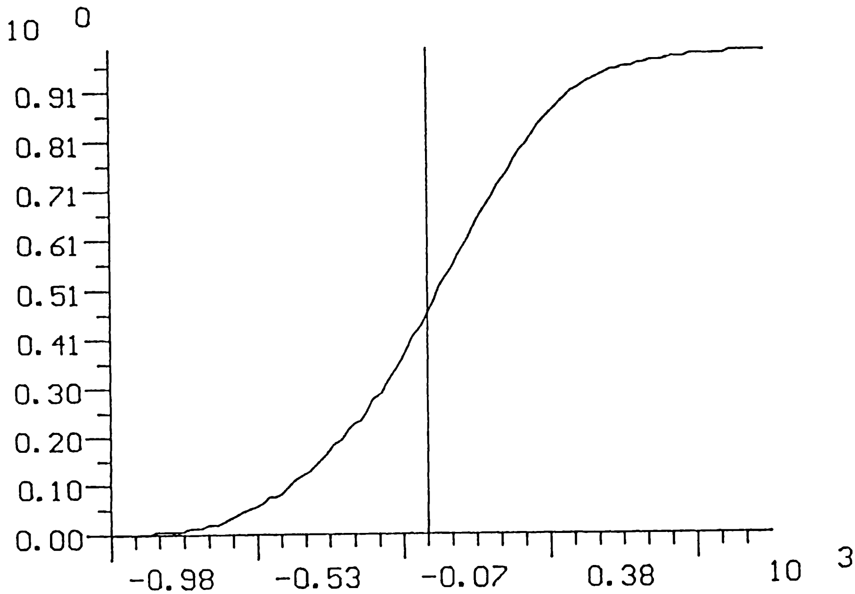


Fig. 4.1

5). Let us consider a stochastic program (3.1) with simple recourse and random right-hand side and suppose that for all components of the random vector a sample of n observations is given. We suppose that all marginal distribution functions are estimated by the kernel method((4.2)) which leads to an estimated distribution $\hat{\mu}_n$. Then the stability results of Section 3 together with Propositions 4.2 and 4.3 yield asymptotic properties for the almost sure and mean convergence of the sequences

$$(|\varphi(\mu) - \varphi(\hat{\mu}_n)|)_{n \in N},$$
$$(d_H(\psi(\mu), \psi(\hat{\mu}_n)))_{n \in N}.$$

For example, convergence rates for $\mathrm{E}(d_H(\psi(\mu), \psi(\hat{\mu}_n)))$ $(n \in N)$ can be obtained following the ideas of Corollary 2.12 in [25]. For the power dispatch model of Section 2 we get, in particular,

$$\mathrm{E}\left(d_H(\psi(\mu), \psi(\hat{\mu}_n))\right) = O\left(n^{-1/4}\right),$$

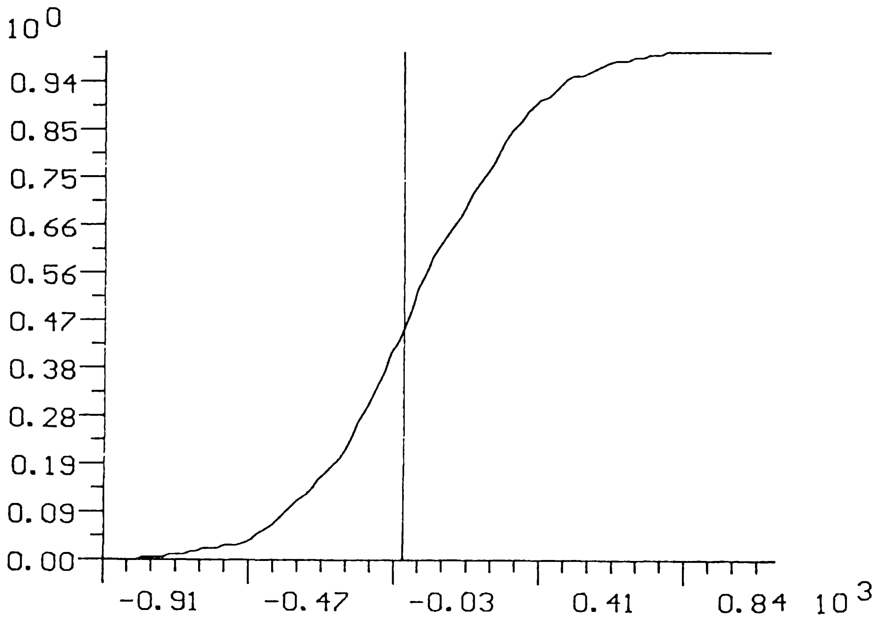


Fig. 4.2

since the support of the random demand is compact and the positivity condition for the marginal densities (see Theorem 3.3) in a neighbourhood of the optimal tender is (certainly) satisfied. This asymptotic argument also provides a certain justification for our numerical approach to the treatment of the economic dispatch in the energy production system in East Germany. The available empirical data is the difference of actual and predicted demand (so-called residuum) over several years. The data has been analyzed and classified (in particular, into data belonging to days with comparable demand curves) (for details see [20]). After performing the data analysis and adding the predicted demand, the sample (for the demand) may be viewed as independent. The following pictures show kernel estimates for the marginal distribution function of the demand during the hour 1 p.m.- 2 p.m. (of a day of normal category). In both cases the triangular kernel has been used with a sample size $n := 436$ and smoothing parameters $b_n = 30$ (Fig. 4.1) and $b_n = 50$ (Fig. 4.2).

5 Numerical treatment, implementation and test example

In this section, we deal with the numerical treatment of stochastic programs with simple recourse and random right–hand side

$$\min \{g(x) + Q_\mu(\chi) : x \epsilon C, Ax = \chi\} \tag{5.1}$$

where g is convex and continuously differentiable, $C \subset \mathrm{R}^m$ is a convex polyhedron, A an $s \times m$ – matrix, Q_μ is defined by (3.2) and has the special feature that only a sample of n observations (with common distribution μ) is available. Since Q_μ only depends on the marginal distribution functions F_r, $r = 1, \ldots, s$ (cf. (3.3)), we may assume that n real observations $X_{r1}, \ldots, X_{rn}$ with common distribution function F_r $(r = 1, \ldots, s)$ are given. Our approach is the following: For each $r = 1, \ldots, s$, F_r is estimated by a smooth nonparametric estimator $\hat{F}_r$ based on the kernel k (see Section 4) and F_r in (3.3) is replaced by $\hat{F}_r$. This leads to the following (convex) nonlinear program having continuously differentiable objective and linear constraints:

$$\min \left\{ g(x) + \hat{Q}_\mu(\chi) : x \epsilon C, Ax = \chi \right\} \tag{5.2}$$

$$\hat{Q}(\chi) := \sum_{r=1}^{s} [q_r^+(\hat{\mu}_r - \chi_r) - (q_r^+ + q_r^-)\frac{1}{n}\sum_{i=1}^{n}\{(X_{ri} - \chi_r)\,\mathcal{K}_1((\chi_r - X_{ri})b^{-1})$$
$$+ b\,\mathcal{K}_2((\chi_r - X_{ri})b^{-1})\}]$$

where $\hat{\mu}_r := \frac{1}{n}\sum_{i=1}^{n} X_{ri}$, $\mathcal{K}_1(u) := \int_{-\infty}^{u} k(t)\, dt$, $\mathcal{K}_2(u) := \int_{-\infty}^{u} tk(t)\, dt$ and b is the smoothing parameter.

We note that for many kernels k (especially those mentioned in Section 4) the functions $\mathcal{K}_1$ and $\mathcal{K}_2$ can be computed explicitly. Hence, no numerical integration has to be performed when $\hat{Q}$ and their partial derivatives

$$\frac{\partial \hat{Q}}{\partial \chi_r} := -q_r^+ + (q_r^+ + q_r^-)\frac{1}{n}\sum_{i=1}^{n}\mathcal{K}_1((\chi_r - X_{ri})b^{-1})$$

are evaluated. Hence, the nonlinear program (5.2) can be solved efficiently by using standard nonlinear programming systems, like e.g. the MINOS–system (see [17]).

The essential difference of our approach to those developed in several papers (e.g. [3], [5], [18], [19], [22], [33]) is that we do not consider F_r as discrete distribution functions but replace F_r by a smooth estimate $\hat{F}_r$ $(r = 1, \dots, s)$. This is certainly justified if (at least) medium – sized samples for F_r are given (as in our applications to power dispatch, see Section 2). We also refer to [11] as general reference for numerical methods in stochastic programming and to [18] and [27] for the description of program systems designed for simple recourse problems.

A program system STOCHOPT according to the above mentioned approach has been developed for IBM/PC AT computers. The programs are written in FORTRAN 77 and Turbo Pascal (user interface).

The system consists of the following main parts:

(*i*) User interface: it realizes an interactive facility for problem specification, the construction of samples (if not available a priorily), choice of smoothing parameters, graphical representation of the optimal solution.

(*ii*) Nonlinear programming part for solving (5.2).

In (*ii*) the alternative use of Epanechnikov, biweigth or triangular kernels is possible. Since these kernels have compact support, an appropriate ordering of the samples accelerates the evaluation of $\hat{Q}$ and its gradient considerably. The nonlinear programming part of STOCHOPT has been tested successfully on both modifications of problems taken from [15] and the model for optimal power dispatch. The latter model has been solved for a real-life situation (with $m = 888$). Numerical results will be reported elsewhere. Now, we report the solution of the classical Aircraft Allocation Problem due to Dantzig ([6]).

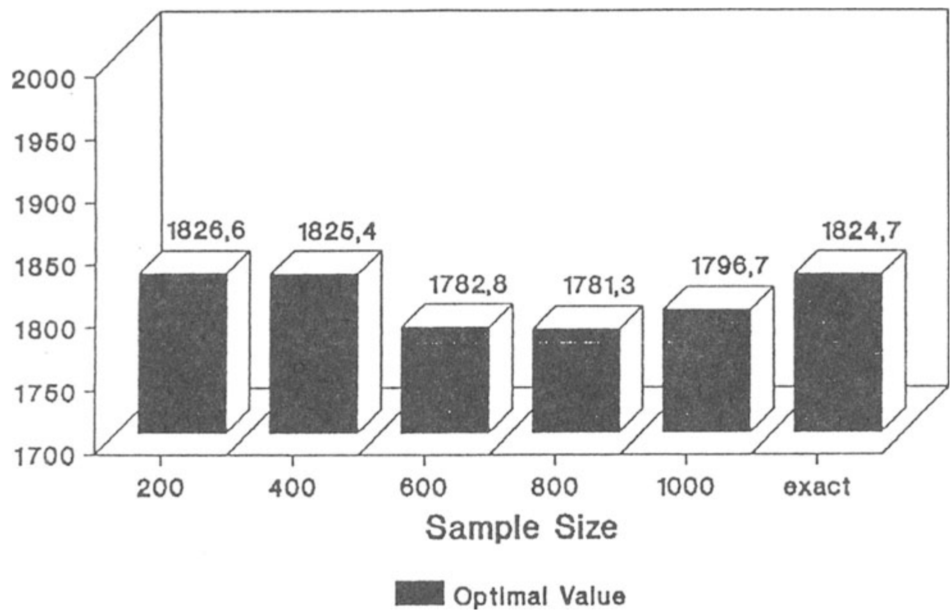


Fig. 5.1 Dependence of the optimal value on the sample size (with smoothing parameter b_{opt})

An airline wishes to allocate airplanes of various types among its routes to satisfy an uncertain passenger demand, while operating costs plus the lost revenue from passengers turned away are minimal.

The stochastic program with simple recourse is as follows

$$\min\left\{d^T x + Q(\chi) : Tx = e, Ax = \chi\right\} \tag{5.3}$$

where $x \in R^{17}_{+}$, $\chi \in R^5$, $A \in L\ (R^{17}, R^5)$, $T \in L\ (R^{17}, R^4)$ and

$d \in R^{17}$, $e \in R^4$:

$$A = \begin{pmatrix} 16 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 9 & 0 & 0 & 0 & 0 \\ 0 & 15 & 0 & 0 & 0 & 10 & 0 & 0 & 0 & 5 & 0 & 0 & 0 & 11 & 0 & 0 & 0 \\ 0 & 0 & 28 & 0 & 0 & 0 & 14 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 22 & 0 & 0 \\ 0 & 0 & 0 & 23 & 0 & 0 & 0 & 15 & 0 & 0 & 7 & 0 & 0 & 0 & 0 & 17 & 0 \\ 0 & 0 & 0 & 0 & 81 & 0 & 0 & 0 & 57 & 0 & 0 & 29 & 0 & 0 & 0 & 0 & 55 \end{pmatrix}$$

$$T = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}$$

$$d = (18 \quad 21 \quad 18 \quad 16 \quad 10 \quad 15 \quad 16 \quad 14 \quad 9 \quad 10 \quad 9 \quad 6 \quad 17 \quad 16 \quad 17 \quad 15 \quad 10)^T$$

$$e = (10 \quad 19 \quad 25 \quad 15)^T.$$

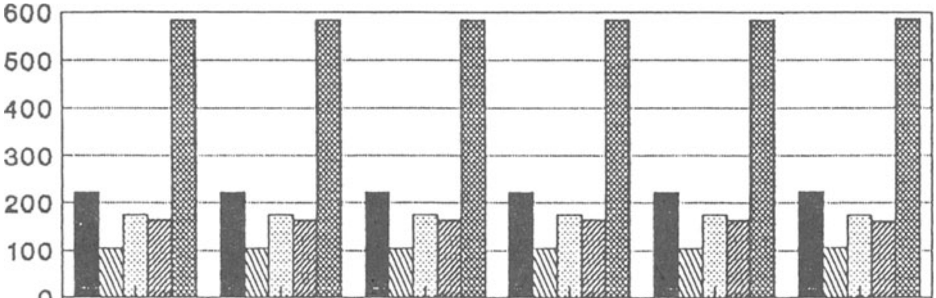
The revenue lost per passengers turned away on the r–th route ($r = 1, \ldots, 5$) is

$$\begin{aligned} q^+ &= (13 \quad 13 \quad 7 \quad 7 \quad 1)^T \\ q^- &= (0 \quad 0 \quad 0 \quad 0 \quad 0)^T \end{aligned}$$

Dantzig solved the program (5.3) with discretely distributed passenger demand z for each of the routes. We assume the passenger demand to have a continuous distribution:

$$\begin{aligned} z_1 &\sim U[200, 300] \\ z_2 &\sim U[50, 150] \\ z_3 &\sim U[140, 220] \\ z_4 &\sim U[10, 340] \\ z_5 &\sim U[580, 620] \end{aligned} \tag{5.4}$$

($U[a, b]$ denotes the uniform distribution on [a,b]). The optimal solution set is listed below:



	200	400	600	800	1000	exact
x1	223,4	223,5	223,4	223,4	223,4	223,9
x2	104,4	104,7	104,7	104,4	104,4	105,9
x3	175,1	174,9	175,1	175,1	175,1	173,8
x4	164,8	164,3	164,3	164,7	164,5	162,4
x5	584,3	584,1	584,7	584,5	585	585,8

Fig. 5.2 Optimal tenders for various sample sizes (with smoothing parameter b_{opt}) and the exact result

Aircrafttype→ ↓ Route	1	2	3	4	Tenders
1	$x_1 = 10$			$x_{13} = 7.1$	$\chi_1 = 224$
2	$x_2 = 0$	$x_6 = 8.2$	$x_{10} = 4.8$	$x_{14} = 0$	$\chi_2 = 106$
3	$x_3 = 0$	$x_7 = 0$		$x_{15} = 7.9$	$\chi_3 = 174$
4	$x_4 = 0$	$x_8 = 10.8$	$x_{11} = 0$	$x_{16} = 0$	$\chi_4 = 162$
5	$x_5 = 0$	$x_9 = 0$	$x_{12} = 20.2$	$x_{17} = 0$	$\chi_5 = 586$

The optimal value of (5.3) is 1824.7.

In order to test our numerical approach, a pseudo-random number generator has been used to simulate samples from the distributions (5.4). The problem (5.2) (as approximation for (5.3)) has been solved for different sample sizes n and smoothing parameters b, re-

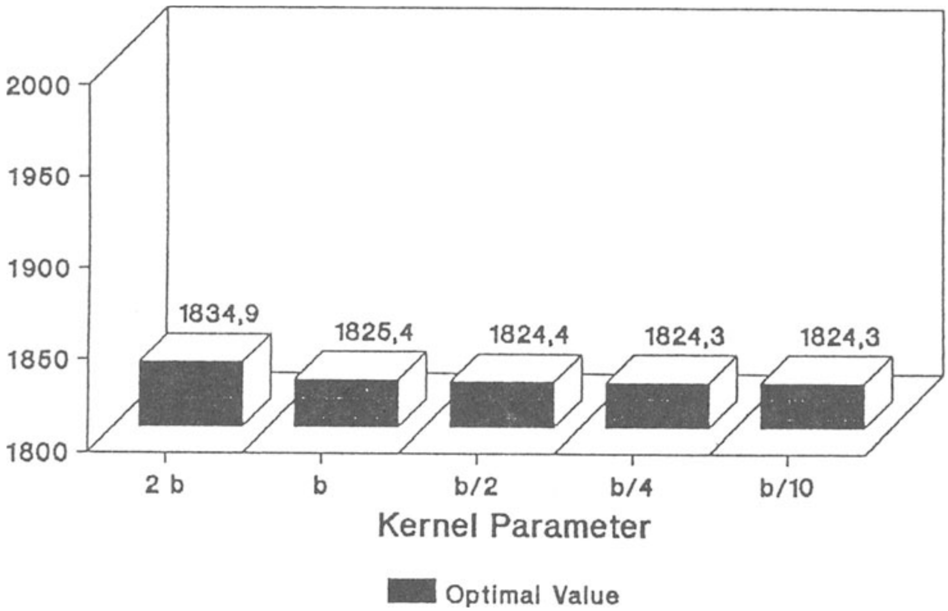


Fig. 5.3 Dependence of the optimal value on the choice of smoothing parameter ($b = b_{opt}$) for a fixed sample of 400 observations

spectively. According to the suggestion in [1], the special smoothing parameter $b_{opt} := 0.5\sigma n^{-\frac{1}{3}}$ (σ being the standard deviation of the unknown distribution) was used when varying the sample size n. We note that for this choice of b_n the asymptotic properties of Section 4 are valid (see Remark 4.5). The numerical results are summarized in the figures 5.1-5.3.

The results show that even for small sample sizes good approximations are obtained for optimal values and optimal tenders, respectively. Fig. 5.3 indicates that the optimal value behaves insensitive on the choice of smoothing parameters (of a relatively large bandwidth), but that a choice of smaller smoothing parameters (than that suggested in [1]) might be favourable.

Acknowledgement

We wish to thank Prof. Georg Pflug (University of Vienna) for beneficial discussions during the GAMM/IFIP–Workshop at Munich. We extend our gratitude to the referees for their constructive criticism.

References

[1] A. Azzalini, A note on the estimation of a distribution function and quantiles by a kernel method, Biometrika 68(1981), 326-328.

[2] P. Billingsley, Convergence of Probability Measures, Wiley, New York, 1968.

[3] J. Böttcher, Stochastische lineare Programme mit Kompensation, Mathematical Systems in Economics, vol. 115, Athenäum Verlag, Frankfurt am Main, 1989.

[4] C. Bouza and S. Allende, Density function estimation and the approximation of convergence rates in stochastic linear programming, Revista de Investigacion Operational 10(1989), 135-140.

[5] D.W. Bunn and S.N. Paschentis, Development of a stochastic model for the economic dispatch of electric power, European Journal of Operational Research 27(1986), 179-191.

[6] G. Dantzig, Linear Programming and Extensions, Princeton University Press, 1963.

[7] L. Devroye, A Course in Density Estimation, Birkhäuser, Boston, 1987.

[8] R.M. Dudley, The speed of the mean Glivenko-Cantelli-convergence, The Annals of Mathematical Statistics 40(1969), 40-50.

[9] J. Dupačová, Stability and sensitivity analysis for stochastic programming, Annals of Operations Research 27(1991), 115-142.

[10] J. Dupačová and R.J.-B. Wets, Asymptotic behaviour of statistical estimators and of optimal solutions of stochastic optimization problems, The Annals of Statistics 16(1988), 1517-1549.

[11] Yu. Ermoliev and R.J.-B. Wets (Eds.), Numerical Techniques for Stochastic Optimization, Springer-Verlag, Berlin, 1988.

[12] P. Gänssler and W. Stute, Wahrscheinlichkeitstheorie, Springer-Verlag, 1977.

[13] P. Kall, Stochastic Linear Programming, Springer-Verlag, Berlin, 1976.

[14] P. Kall, On approximations and stability in stochastic programming, Parametric Optimization and Related Topics (J. Guddat, H.Th. Jongen, B. Kummer, F. Nožička, Eds.), Akademie-Verlag, Berlin, 1987, 387-407.

[15] A.J. King, Stochastic Programming Problems: Examples from the literature, in [11], 543-567.

[16] A.J. King and R.J.-B. Wets, Epi-consistency of convex stochastic programs, Research Report, IBM Research Division, T.J. Watson Research Center, New York, 1989.

[17] B. Murtagh and M. Saunders, Large-scale linearly constrained optimization, Mathematical Programming 14(1978), 41-72.

[18] L. Nazareth, Design and implementation of a stochastic programming optimizer with recourse and tenders, in [11], 273-294.

[19] L. Nazareth and R.J.-B. Wets, Nonlinear Programming Techniques applied to stochastic programs with recourse, in [11], 95-121.

[20] J. Polzehl and K. May, Short term prediction of electric load in a power network, Manuscript, Fachbereich Mathematik der Humboldt-Universität zu Berlin (in preparation).

[21] B.L.S. Prakasa Rao, Nonparametric Functional Estimation, Academic Press, New York, 1983.

[22] A. Prékopa, Dual method for the solution of a one-stage stochastic programming problem with random rhs obeying a discrete probability distribution, Zeitschrift für Operations Research 34(1990), 441-461.

[23] S.M. Robinson and R.J.-B. Wets, Stability in two-stage stochastic programming, SIAM Journal on Control and Optimization 25(1987), 1409-1416.

[24] W. Römisch and R. Schultz, Stability of solutions for stochastic programs with complete recourse having $C^{1,1}$ data, Manuskript, Institut für Operations Research der Universität Zürich, 1989.

[25] W. Römisch and R. Schultz, Stability analysis of stochastic programs, Annals of Operations Research 29(1991) (to appear).

[26] W. Römisch and R. Schultz, Stochastic programs with complete recourse: Stability and an application to power dispatch, in: System Modelling and Optimization (H.-J. Sebastian, K. Tammer,Eds.), Proceedings 14th IFIP Conference (Leipzig, 1989), Lecture Notes in Control and Information Sciences vol.143, Springer-Verlag, Berlin, 1990.

[27] N. Roenko, V. Loskutov and S. Uryasyev, Stochastic nonlinear programming system, IIASA, Laxenburg, Working Paper WP-89-075(1989).

[28] A. Shapiro, Asymptotic properties of statistical estimators in stochastic programming, The Annals of Statistics 17(1989), 841-858.

[29] G.R. Shorack and J. Wellner, Empirical Processes with Applications to Statistics, Wiley, New York, 1986.

[30] B.W. Silvermann, Density Estimation for Statistics and Data Analysis, Chapman and Hall, London, 1986.

[31] S. Vogel, Stability results for stochastic programming problems, Optimization 19(1988), 269-288.

[32] Hj. Wacker (Ed.), Applied Optimization Techniques in Energy Problems, Teubner, Stuttgart, 1985.

[33] R.Wets, Solving stochastic programs with simple recourse, Stochastics 10(1983), 219-242.

[34] B.B. Winter, Convergence rate of perturbed empirical distribution functions, Journal of Applied Probability 16(1979), 163-173.