Applied Probability Trust (17 July 2003)

SEQUENTIAL SELECTION OF RANDOM VECTORS UNDER A SUM CONSTRAINT

MARIO STANKE,* *Georg-August-Universität Göttingen*

*

Abstract

We observe a sequence $X_1, X_2, \ldots, X_n$ of independent identically distributed coordinate-wise nonnegative d-dimensional random vectors sequentially. When a vector is observed it can either be selected or rejected but once met this decision is final. In each coordinate the sum of the selected vectors must not exceed a given constant. The problem is to find a selection policy that maximizes the expected number of selected vectors. For general absolutely continuous distribution of the X_i we determine the maximal expected number of selected vectors asymptotically and give a selection policy which asymptotically achieves optimality.

Above problem raises a question closely related to the following problem. Given an absolutely continuous measure μ on $Q = [0,1]^d$ and a $\tau \in Q$, find a set A of maximal measure $\mu(A)$ among all $A \subset Q$ whose center of gravity lies below τ in all coordinates. We will show that a simplicial section $\{\mathbf{x} \in Q \mid \langle x, \theta \rangle \leq 1\}$, where $\theta \in \mathbb{R}^d, \theta \geq 0$ satisfies a certain additional property, is a solution to this problem.

Keywords: online selection, sum constraint, threshold region

AMS 2000 Subject Classification: Primary 60G50
Secondary 62L15

1. Introduction

The one-dimensional case $d = 1$ has been treated by several authors before. The special one-dimensional case where the distribution function on the size of the X_i's is of the form $F(x) = Ax^\alpha \quad (\alpha > 0)$ has been solved by Coffman et al. [2]. Later, Rhee and Talagrand [11] generalized this result to arbitrary distributions. The generalization of the one-dimensional problem where the number n of observed random vectors is itself random has been treated by Gnedin [6].

In this paper we generalize the problem to X_i's of multi-dimensional size. This can be interpreted in the following way. We have d different types of resources and the j-th type of resource is limited by the constant $c_j \quad (j = 1, \ldots, d)$. The 'items' X_i require a certain amount $X_i^{(j)}$ of each resource j. For each resource type the total amount needed by the selected items must not exceed the given limit. By transforming the $X_i^{(j)}$ (via ${X'}_i^{(j)} = X_i^{(j)}/c_j$) we can assume without loss in generality that all the c_j's are 1.

A related multidimensional bin packing problem has been treated by Garey, Graham and Johnson [4]. As a special case of their setting they consider the problem of assigning *all* the sequentially observed vectors – which are not random in their problem definition – to multiple "bins" such that the sum of all vectors assigned to a bin is dominated by $(1, \ldots, 1)$. They consider the task of minimizing the number of used bins, whereas we only may use one "bin" and want to maximize the number of vectors "packed" into that one bin by possibly rejecting some of them.

* Postal address: Institut für Mikrobiologie und Genetik, Abteilung Bioinformatik, Goldschmidtstraße 1, 37077 Göttingen, Germany, Email: mstanke@gwdg.de

We now introduce some terms and notation. Let $Q = [\mathbf{0}, \mathbf{1}]^d$ be the d-dimensional interval with endpoints $\mathbf{0} = (0, \dots, 0)$ and $\mathbf{1} = (1, \dots, 1)$. Let $X_1, X_2, \dots X_n \geq \mathbf{0}$ be independent identically distributed d-dimensional random vectors with law μ on $\mathbb{R}^d_+$. (Inequalities between vectors should always be interpreted as a set of coordinate wise inequalities.) Let $\mathbf{X} = (X_1, X_2, \dots, X_n)$. We will speak of the X_i's as sizes of *items* and we speak of subintervals of Q as *space*. n is the number of items that are at disposal to be packed. All random variables are assumed to be defined on some common probability space with probability measure P.
We define a *selection policy* to be a function $\Psi = (\Psi_1, \Psi_2, \dots, \Psi_n) : \prod_{i=1}^{n} \mathbb{R}^d_+ \to \{0,1\}^n$ and an *online selection policy* is a selection policy Ψ where $\mathbf{x} = (x_1, x_2, \dots, x_n)$ is mapped to $\Psi(\mathbf{x}) = (\Psi_1(\mathbf{x}), \dots, \mathbf{\Psi_n}(\mathbf{x}))$ and $\Psi_i(\mathbf{x}) = \Psi_i(x_1, \dots, x_i)$ for $i = 1, 2, \dots, n$, that is, Ψ_i is a function of $x_1, x_2, \dots, x_i$ only.

We consider the restriction that the sum of the selected variables must stay within Q. We call those policies *admissible* that satisfy the sum constraint

$$\sum_{i=1}^{n} \Psi_i(\mathbf{x}) x_i \leq \mathbf{1} \qquad \text{for all } \mathbf{x} = (x_1, \dots, x_n) \quad x_i \geq 0, i = 1, \dots, n \tag{1}$$

A selection policy Ψ will be regarded as a function of the random sequence $\mathbf{X}$ and we will usually write Ψ instead of $\Psi(\mathbf{X})$ and Ψ_i for $\Psi_i(X)$. We say that item i of size X_i is selected by Ψ if $\Psi_i(X_1, \dots, X_i) = 1$ and it is rejected if $\Psi_i(X_1, \dots, X_i) = 0$.

We are interested in the expected number of selected variables

$$\mathcal{E}(\Psi) := E \sum_{i=1}^{n} \Psi_i \tag{2}$$

and want to maximize it.

Let $\mathcal{P}$ be the set of all admissible *online* selection policies and let $\mathcal{S}$ be the set of *all* admissible selection policies.

Define

$$\mathrm{Opt}_n := \sup_{\Psi \in \mathcal{P}} \mathcal{E}(\Psi) \quad \text{ and } \quad \mathrm{Proph}_n := \sup_{\Psi \in \mathcal{S}} \mathcal{E}(\Psi),$$

the maximal expected number of selected items within the respective class of policies. As Opt_n and Proph_n depend on the distribution of X_1 we will sometimes write $\mathrm{Opt}_n = \mathrm{Opt}_n(\mu)$ and $\mathrm{Proph}_n = \mathrm{Proph}_n(\mu)$.

Interpretation: For a policy in $\mathcal{P}$ the decision whether to select X_i or not depends only on X_i and the 'past': $X_1, X_2, \dots, X_{i-1}$. The items come one after the other and we have to decide *online*, i.e. without knowing the 'future' and without revoking a decision we have made before. In $\mathcal{S}$ the decision can depend on the whole sequence $X_1, X_2, \dots$. As has been done before we imagine a prophet who is given the task of selecting the items. The prophet knows the sizes of all the items in advance.
Clearly, the prophet can 'simply' select the largest subset I of $\{1, 2, \dots, n\}$ such that $\sum_{i \in I} X_i \leq \mathbf{1}$. The expected number of selected variables by the prophet then is

$$\mathrm{Proph}_n = E\,[\max\{\#I \mid \sum_{i \in I} X_i \leq \mathbf{1}\}].$$

We know $\mathrm{Opt}_n \leq \mathrm{Proph}_n$ as $\mathcal{P} \subset \mathcal{S}$.

In section 2 we will examine Opt_n for n becoming large and will show for absolute continuous μ that $\mathrm{Opt}_n(\mu) \sim \mathrm{Proph}_n(\mu)$ as $n \to \infty$ as in the one-dimensional case. We will give an asymptotically optimal selection policy and determine $\mathrm{Opt}_n(\mu)$. In section 3 we will give an example solution for a measure μ which is a direct generalization of Coffman et al.'s distribution: the multidimensional distribution function $F(x) = a x_1^{\alpha_1} \cdots x_d^{\alpha_d}$ for small $x_1 > 0, \ldots, x_d > 0$.
The proof in section 2 depends on a result about simplicial sections of Q and their barycentre which is given in section 4.

2. Sequential selection out of n random vectors under a sum constraint

In general, we cannot determine $\mathrm{Opt}_n(\mu, \mathbf{c})$ exactly. Instead, we we will focus on its asymptotic behavior when $n \to \infty$.
But first some definitions. We are given a probability measure μ on the d-dimensional unit cube $Q = [0,1]^d$. By a *simplicial section* we mean a set $\{\mathbf{x} \in Q \mid \langle \mathbf{x}, \boldsymbol{\theta} \rangle \leq 1\}$, where $\boldsymbol{\theta} \geq \mathbf{0}$.
Figure 3 on page 14 shows one. Assume the random vector Z has law μ. For a measurable set $A \subset Q$ let

$$\mathbf{g}(A) := E[Z\, \mathbb{1}_A(Z)] = \int_A \mathbf{x}\, d\mu(\mathbf{x}).$$

And let $\mathbf{c}(A)$ for $\mu(A) > 0$ denote the barycentre (center of gravity) of A with respect to μ, i.e.

$$\mathbf{c}(A) := E[Z \mid Z \in A] = \frac{\mathbf{g}(A)}{\mu(A)} = \frac{1}{\mu(A)} \int_A \mathbf{x}\, d\mu(\mathbf{x}).$$

We write g_i, c_i for the i-th coordinate of $\mathbf{g}$ and $\mathbf{c}$. To be able to prove the main result of this part, we need 3 lemmata.

Lemma 1. *Let $\mu(A) > 0$ for any neighborhood A of $\mathbf{0}$. Then*

$$\lim_{n\to\infty} \mathrm{Opt}_n(\mu) = \infty.$$

Proof. Let $M > 1$ be arbitrary, let $A := [\mathbf{0}, \frac{1}{M}\mathbf{1}]$ and consider the policy Ψ which selects all items with sizes in A unless the sum of the selected items would exceed $\mathbf{1}$. Then

$$\mathrm{Opt}_n(\mu) \geq \mathcal{E}(\Psi) \geq M\, P\Big(\sum_{j=1}^{n} \mathbb{1}_{\{X_j \in A\}} \geq M\Big) \to M,$$

since $\sum_{j=1}^{n} \mathbb{1}_{\{X_j \in A\}}$ is binomially distributed with parameters n and $\mu(A) > 0$. Therefore $\lim\limits_{n\to\infty} \mathrm{Opt}_n(\mu) \geq M$ and the claim follows as M was arbitrary.

If there is a neighborhood A of $\mathbf{0}$ in Q such that $\mu(A) = 0$, we can almost surely only select a bounded number of items. As this case doesn't seem very interesting in our asymptotical analysis we will exclude it from further consideration. From now on let $\mu(A) > 0$ for any neighborhood A of $\mathbf{0}$. For the asymptotic behavior of $\mathrm{Opt}_n(\mu)$ only the values of $\mu(A)$ for neighborhoods of $\mathbf{0}$ play a role.

Lemma 2. *Let $A \subset Q$ be such that $c_i(A) \leq s < 1$ for $(i = 1, 2, \ldots, d)$. Define $A' := \{\mathbf{x} \in A \,|\, \mathbf{x} \leq s^{2/3}\mathbf{1}\}$. Then*

$$\frac{\mu(A')}{\mu(A)} \geq 1 - d\, s^{1/3} \tag{3}$$

Remark: *We will need this lemma for small s and a simplicial section $A = \Delta$. See also figure 1 on page 7.*

Proof. For $i = 1, \ldots, d$ we have

$$s \geq c_i(A) = \frac{1}{\mu(A)} \int_A x_i \, d\mu(\mathbf{x}) \geq \frac{1}{\mu(A)} \int_{A \cap \{x_i \geq s^{2/3}\}} x_i \, d\mu(\mathbf{x}) \geq \frac{1}{\mu(A)} s^{2/3} \mu(A \cap \{x_i \geq s^{2/3}\}).$$

This implies

$$\frac{\mu(A \cap \{x_i \geq s^{2/3}\})}{\mu(A)} \leq s^{\frac{1}{3}}. \tag{4}$$

Now we get

$$1 - \frac{\mu(A')}{\mu(A)} = \frac{\mu(A \setminus A')}{\mu(A)} \leq \sum_{i=1}^{d} \frac{\mu(A \cap \{x_i \geq s^{2/3}\})}{\mu(A)} \leq d\, s^{1/3}$$

Lemma 3. (a Chernoff bound.) *Let $Z_1, Z_2, \ldots, Z_n$ be independent random variables with $E\,[Z_i] \leq \frac{1}{n}$ and $0 \leq Z_i \leq a$. Then we have for any natural number $m < n$*

$$P(\sum_{i=1}^{m} Z_i > 1) \leq \exp\,[-\frac{\delta^2}{2a}]$$

with $\delta := 1 - \frac{m}{n}$.

Proof. The proof follows the common idea of Chernoff bounds. By the Markov inequality and because of the independence we have for any $t > 0$

$$p := P(\sum_{i=1}^{m} Z_i > 1) \leq e^{-t} E\,[\exp(t \sum_{i=1}^{m} Z_i)] = e^{-t} \prod_{i=1}^{m} E\,[e^{tZ_i}].$$

As the function $x \mapsto e^{tx}$ is convex, we have $e^{tx} \leq 1 + x\frac{1}{a}(e^{ta} - 1)$ for $0 \leq x \leq a$. Plugging in Z_i for x in this general inequality and using $1 + x \leq e^x$ we get

$$p \leq \exp\,\Big[\frac{1}{a}((1-\delta)(e^{ta} - 1) - ta)\Big]$$

Now put $t = -\frac{1}{a}\ln(1-\delta) > 0$. Simplifying and using $\ln(1-x) \leq -x - x^2/2 \quad (0 \leq x < 1)$ yields $p \leq \exp(\frac{-\delta^2}{2a})$.

Next we will derive asymptotic results about Opt_n for $n \to \infty$. This is the main result of this part.
We know $\mathrm{Opt}_n \leq \mathrm{Proph}_n$ but it turns out that for large n the prophets policy is not much better than the optimal online selection policy:

Theorem 1. *Let μ be an absolutely continuous probability measure on $Q = [\mathbf{0}, \mathbf{1}]$. Let $\Delta = \Delta(n)$ be a simplicial section $\{\mathbf{x} \in Q \,|\, \langle \mathbf{x}, \boldsymbol{\theta} \rangle \leq 1\}$ such that $g_i(\Delta) \leq 1/n$ and $\theta_i = 0$ when $g_i(\Delta) < 1/n$. Then*

(1)

$$\mathrm{Opt}_n \sim \mathrm{Proph}_n \sim n\,\mu(\Delta).$$

(2) Let

$$s := \frac{1}{n\mu(\Delta)} \tag{5}$$

and $\Delta' := \{\mathbf{x} \in \Delta \,|\, \mathbf{x} \leq s^{2/3}\mathbf{1}\}$*. Let* Ψ *be the following policy (which depends on* n*).*
Accept X_j *if* $X_j \in \Delta'$ *and the sum of the variables selected so far plus* X_j *is still less than or equal to* $\mathbf{1}$*. But if this sum exceeds* $\mathbf{1}$ *in any coordinate then reject* X_j *and all subsequent variables* $X_{j+1}, X_{j+2}, \ldots$*.*
Then Ψ *is asymptotically optimal, i.e.* $\mathcal{E}(\Psi) \sim \mathrm{Opt}_n$*.*
Furthermore, for any $\varepsilon > 0$ *we have the error bound*

$$1 - \frac{\mathcal{E}(\Psi)}{n\,\mu(\Delta)} = O\left(s^{1/3-\varepsilon}\right) \qquad \text{as } n \to \infty. \tag{6}$$

Note: Theorem 3 of section 4 ensures that a Δ like above always exists.

Proof. For the upper bound on Opt_n we will give an upper bound on Proph_n which we can reduce to the one-dimensional case. For the lower bound we will show that Ψ asymptotically achieves the upper bound.

Upper Bound. Recall that, when choosing the variables X_i with $i \in I$, we had to comply with the constraint

$$\sum_{i\in I} X_i \leq \mathbf{1}. \tag{7}$$

Define $\delta := (\theta_1 + \theta_2 + \cdots + \theta_d)^{-1}$ and set $\boldsymbol{\alpha} := \delta\,\boldsymbol{\theta}$. Now consider the following one-dimensional relaxation of (7)

$$\sum_{i\in I} \langle X_i, \boldsymbol{\alpha}\rangle \leq 1. \tag{8}$$

(7) implies (8) and intuitively, (8) means that – instead of staying in the cube Q – the sum of the selected points must stay within a certain simplex given by a hyperplane that goes through $\mathbf{1}$.
Now, let

$$Y_i := \langle X_i, \boldsymbol{\alpha}\rangle\,.$$

When selecting the Y_i's under the relaxed constraint $\sum_{i\in I} Y_i \leq 1$ the prophet will do at least as good as under (7). Let F be the distribution function of the Y_i's. Then F is continuous because μ has a density.
We apply the one-dimensional result from [11] to the sequence $(Y_1, Y_2, \ldots)$ to get for $n > 1/\mathrm{E}\,[Y_1]$

$$\mathrm{Proph}_n \leq nF(\varepsilon), \tag{9}$$

for any ε such that

$$\int_0^{\varepsilon} x\, dF(x) = \frac{1}{n}. \tag{10}$$

We will show that $\varepsilon = \delta$ satisfies (10) and that $F(\delta) = \mu(\Delta)$, then (9) implies the upper bound $\mathrm{Proph}_n \lesssim n\mu(\Delta)$, meaning that $\limsup_n \mathrm{Proph}_n/(n\mu(\Delta) \leq 1$.
The latter is clear because

$$F(\delta) = P(Y_1 \leq \delta) = P(\langle X_1, \boldsymbol{\alpha}\rangle \leq \delta) = P(\langle X_1, \boldsymbol{\theta}\rangle \leq 1) = P(X_1 \in \Delta) = \mu(\Delta).$$

Now, we show that $\varepsilon = \delta$ satisfies (10).

$$\begin{aligned}
\int_0^\delta x\,dF(x) &= E\,[Y_1 \mathbb{1}_{\{Y_1 \le \delta\}}] \\
&= E\,[\langle X_1, \boldsymbol{\alpha}\rangle \ \mathbb{1}_{\{\langle X_1, \boldsymbol{\alpha}\rangle \le \delta\}}] \\
&= E\,[\sum_{j=1}^{d} \alpha_j X_1^{(j)} \mathbb{1}_{\{X_1 \in \Delta\}}] \\
&= \sum_{j=1}^{d} \alpha_j E\,[X_1^{(j)} \mathbb{1}_{\{X_1 \in \Delta\}}] \\
&= \sum_{j=1}^{d} \alpha_j g_j(\Delta) \\
&= \sum_{j=1}^{d} \alpha_j \frac{1}{n} && \text{because } g_j(\Delta) = \tfrac{1}{n} \text{ if } \alpha_j \neq 0 \\
&= \frac{1}{n}
\end{aligned}$$

We get the upper bound on Opt_n

$$\mathrm{Opt}_n \le \mathrm{Proph}_n \lesssim n\mu(\Delta). \tag{11}$$

Remark: *As we know already that* $\mathrm{Opt}_n \to \infty$ *by Lemma 1 we can now conclude that* $n\mu(\Delta) \to \infty$ *as well and therefore*

$$s \to 0 \qquad \text{as } n \to \infty \tag{12}$$

Lower Bound. First note that Ψ as defined in the theorem is an admissible online selection policy. And so the optimal expected number of selected items is at least $\mathcal{E}(\Psi)$:

$$\mathrm{Opt}_n \ge \mathcal{E}(\Psi).$$

In this part we will show the error bound (6). Since $s \to 0$ when $n \to \infty$ this will give us $\mathcal{E}(\Psi) \sim n\mu(\Delta)$. And together with the upper bound we get

$$n\mu(\Delta) \sim \mathcal{E}(\Psi) \le \mathrm{Opt}_n \le \mathrm{Proph}_n \lesssim n\mu(\Delta),$$

which proves the rest of the claim.

The stationary policy which uses Δ instead of Δ' as acceptance region seems to be more natural. Unfortunately, it is not always asymptotically optimal when $d > 1$. But the difference in measure between the two regions is asymptotically negligible:
$c_i(\Delta) = g_i(\Delta)/\mu(\Delta) \le 1/(n\mu(\Delta)) = s$, so Lemma 2 gives us

$$\frac{\mu(\Delta')}{\mu(\Delta)} \ge 1 - d\, s^{1/3}, \tag{13}$$

which converges to 1 as $n \to \infty$. We will need (13) later.

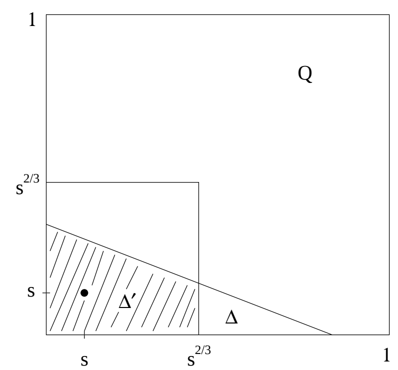

FIGURE 1: With the policy Ψ a point is selected when it is in the hatched region Δ'. (s is the largest coordinate of the barycentre of Δ.)

The upper bound on $\mathcal{E}(\Psi)$ is easy. Trivially, we have $\mathcal{E}(\Psi) \le E\,[\sum_{i=1}^n \mathbb{1}_{\{X_i \in \Delta'\}}] = n\,\mu(\Delta') \le n\,\mu(\Delta)$.

So $1 - \mathcal{E}(\Psi)/(n\,\mu(\Delta)) \ge 0$.

Now, we turn to the lower bound on $\mathcal{E}(\Psi)$. Introduce the stopping time

$$\rho := \inf\,\{k \mid \sum_{i=1}^{k} X_i \mathbb{1}_{\{X_i \in \Delta'\}} \not\le \mathbf{1}\}$$

and set $\rho = \infty$ if no such k exists. Then the number of selected variables by our strategy Ψ is

$$\sum_{i=1}^{(\rho-1)\wedge n} \mathbb{1}_{\{X_i \in \Delta'\}},$$

where $(\rho - 1) \wedge n$ denotes the minimum of n and $\rho - 1$. When using $\rho \wedge n$, which is also a stopping time, instead of $(\rho - 1) \wedge n$ as upper bound of the sum the error is at most 1 and will be asymptotically negligible. By Wald's equation

$$\begin{aligned} \mathcal{E}(\Psi) &\ge E\,[\sum_{i=1}^{\rho\wedge n} \mathbb{1}_{\{X_i \in \Delta'\}}] - 1 \\ &= \mu(\Delta') \cdot E\,[\rho \wedge n] - 1. \end{aligned} \tag{14}$$

For $\mu(\Delta')$ we already have a bound. Now, we want to bound $E[\rho \wedge n]$ from below. We will use that

$$E\,[\rho \wedge n] \ge m\,P(\rho > m) \qquad \text{for any } m < n \tag{15}$$

and will have to choose m suitably. Note that

$$\begin{aligned} P(\rho \le m) &= P(\sum_{i=1}^{m} X_i \mathbb{1}_{\{X_i \in \Delta'\}} \not\le \mathbf{1}) \\ &\le \sum_{j=1}^{d} P(\sum_{i=1}^{m} X_i^{(j)} \mathbb{1}_{\{X_i \in \Delta'\}} > 1) \end{aligned}$$

Now, apply Lemma 3 to $Z_i = X_i^{(j)} \mathbb{1}_{\{X_i \in \Delta'\}}$, $a = s^{2/3}$ and $m = \lfloor n(1 - s^{1/3-\varepsilon}) \rfloor$. This is possible because $\mathrm{E}\,[Z_i] = g_j(\Delta') \le g_j(\Delta) \le \frac{1}{n}$ and $Z_i \le a$ by definition of

Δ'.
We get with $\delta := 1 - \frac{m}{n} \geq s^{1/3-\varepsilon}$

$$\begin{aligned} P(\rho \leq m) &\leq d \exp[-\frac{\delta^2}{2a}] \\ &\leq d \exp[-\frac{s^{2/3-2\varepsilon}}{2s^{2/3}}] \\ &= d \exp[-\frac{1}{2}s^{-2\varepsilon}] \\ &= \mathrm{O}(s) \qquad \text{as } s \to 0. \end{aligned} \tag{16}$$

Now we are ready to prove the rest.

$$\begin{aligned} 1 - \frac{\mathcal{E}(\Psi)}{n\,\mu(\Delta)} &\leq 1 - \frac{\mu(\Delta')\mathrm{E}[\rho \wedge n] - 1}{n\,\mu(\Delta)} && \text{(by inequality (14))} \\ &\leq 1 - \frac{\mu(\Delta')}{\mu(\Delta)}\,\frac{m}{n}\,P(\rho > m) + s && \text{(by inequality (15))} \end{aligned}$$

All three factors on the right-hand side are less than 1 (and converge to 1) and we already have bounds for them. Use the inequality $1-abc \leq (1-a)+(1-b)+(1-c)$, which holds for all $0 \leq a,b,c \leq 1$ and apply (13), the definition of m and (16) to get

$$\begin{aligned} 1 - \frac{\mathcal{E}(\Psi)}{n\,\mu(\Delta)} &\leq ds^{\frac{1}{3}} + \mathrm{O}(s^{1/3-\varepsilon}) + \mathrm{O}(s) + s \\ &= \mathrm{O}(s^{1/3-\varepsilon}). \end{aligned}$$

3. Example

Consider a measure μ on Q with distribution function

$$F(\mathbf{x}) = a x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_d^{\alpha_d}$$

in a neighborhood U of $\mathbf{0}$, where $a, \alpha_1, \alpha_2, \ldots, \alpha_d$ are positive constants. This is the direct generalization of the distributions studied in the paper of Coffman and al. [2] to more than one dimension. This will also give the result for the Lebesgue-measure on Q as a special case.

Theorem 2.

$$\mathrm{Opt}_n \sim \gamma \cdot (an)^{1/(1+\alpha)} \tag{17}$$

where

$$\gamma := (1+\alpha)\left[\frac{\alpha_1 \cdots \alpha_d\,\Gamma(\alpha_1)\cdots\Gamma(\alpha_d)}{\alpha_1^{\alpha_1}\cdots\alpha_d^{\alpha_d}\,\Gamma(2+\alpha)}\right]^{1/(1+\alpha)}$$

and $\alpha := \alpha_1 + \alpha_2 + \cdots + \alpha_d$. Γ *here denotes the Gamma function.*

Proof. Let $\boldsymbol{\theta} > \mathbf{0}$ be defined by $\theta_i := \alpha_i/\alpha$. Let $\Delta_r := \{\mathbf{x} \in Q \mid \langle \mathbf{x}, \boldsymbol{\theta} \rangle \leq r\}$. Observe that there is a $r_0 > 0$ such that for $0 < r < r_0$ we have $\Delta_r \subset U$. In the following let $0 < r \leq r_0$.

Using Dirichlet's formula for the integration we get

$$\begin{aligned} g_i(\Delta_r) &= \int_{\Delta_r} x_i \, d\mu(\mathbf{x}) \\ &= a \int_{\Delta_r} \alpha_1 \alpha_2 \cdots \alpha_d \quad x_i \quad x_1^{\alpha_1 - 1} x_2^{\alpha_2 - 1} \cdots x_d^{\alpha_d - 1} \, d\mathbf{x} \\ &= a \left(\frac{r\alpha}{1+\alpha} \gamma \right)^{1+\alpha}. \end{aligned}$$

Now, by choosing $r = \frac{1+\alpha}{\gamma\alpha}(na)^{-1/(1+\alpha)}$, which is smaller than r_0 for n large enough, we get $g_i(\Delta_r) = \frac{1}{n}$ for $i = 1, \ldots, n$. So the hypothesis for Theorem 1 is satisfied with $\Delta = \Delta_r$. Very similar to above computation we can now compute

$$\mu(\Delta_r) = \frac{1}{n} \gamma (na)^{1/(1+\alpha)},$$

which gives the desired result.

We now get the one-dimensional example, $F(x) = ax^\alpha$ for some neighborhood of 0, as a special case. Plugging in the values we obtain for this case

$$\mathrm{Opt}_n \sim \left(\frac{1+\alpha}{\alpha} \right)^{\frac{\alpha}{1+\alpha}} (an)^{1/(1+\alpha)}.$$

Another example for μ is the d-dimensional Lebesgue-measure on Q, which gives

$$\mathrm{Opt}_n \sim \gamma \cdot n^{\frac{1}{1+d}} \qquad \text{with } \gamma = \frac{d+1}{((d+1)!)^{1/(d+1)}},$$

which itself specializes to $\sqrt{2n}$ in the one-dimensional case.

4. Sets of maximal volume under certain restrictions

In this part we will give the proof we left out in section 2. Theorem 3 below shows that for an absolutely continuous measure μ on Q, there always is a simplicial section Δ with the properties required in Theorem 1. As a side effect of this examination we will get results about optimization problems addressed by Mallows, Nair, Shepp and Vardi in [10].

The idea which leads to these questions is the following. In the selection problem the simplest strategy seems to choose a fixed acceptance region A and accept all the items with sizes in A as long as allowed by the sum constraint. We call this a *stationary* strategy. If there was no constraint the expected number of selected items would be $n\mu(A)$ and the expected space needed would be $n\mathrm{E}\left[X_1 \mathbb{1}_A(X_1)\right] = n \cdot \mathbf{g}(A)$. It seems natural to try to use an A such that $n\mu(A)$ - or equivalently $\mu(A)$ - is maximal under all A's such that the expected space needed is less than or equal to $\mathbf{1}$. It will turn out that such an A (which depends on n) indeed gives an asymptotically optimal admissible strategy when $n \to \infty$.
In this part we will deal with the problem of determining the shape of A.

Problem definition

Recall the definition of $\mathbf{g}(A)$ and $\mathbf{c}(A)$ on page 3. The problem of this section will be an optimization problem:

$$(P_1) \qquad \text{maximize } \mu(A) \quad \text{on } \{A \subset Q \,|\, \mathbf{g}(A) \leq \boldsymbol{\rho}, \quad A \text{ measurable}\}$$

for some $\boldsymbol{\rho} > \mathbf{0}$.
The solution of this problem will also give us a solution to the problem

$$(P_2) \qquad \text{maximize } \mu(A) \quad \text{on } \{A \subset Q \,|\, \mathbf{c}(A) \leq \boldsymbol{\tau}, \quad A \text{ measurable}\}$$

for some $\boldsymbol{\tau} > \mathbf{0}$.
As the solutions are trivial for $d = 1$ we will assume that $d > 1$. Problem (P_2) has also been treated in [10]. Unfortunately, their proof, attributed to Andrew Odlyzko, seems to have a gap when $d > 2$. We will use a different approach here. One that yields an additional property of the solution that we require in the first part of this paper.

It can be proved that these optimization problems actually have a solution (proof omitted).

Shape of the solution

Definition 1. A set $A \subset \mathbb{R}^d_+$ is a *lower layer* if $\mathbf{x} \in A, 0 \leq \mathbf{y} \leq \mathbf{x}$ implies $\mathbf{y} \in A$.

Lemma 4. *Let μ be absolutely continuous. If A is a solution to (P_1) or (P_2) then A is a lower layer up to sets of measure 0.*

Proof. Let A be a solution to (P_1). Assume A was not equal to a lower layer up to sets of measure 0. Then there would exists a $\mathbf{y} \in Q$ such that $\mu([\mathbf{y}, \mathbf{1}] \cap A) > 0$ and $\mu([\mathbf{0}, \mathbf{y}] \cap A^c) > 0$.

To see this assume that $\mathbf{y_1}, \mathbf{y_2}, ...$ was an enumeration of all the points in Q with just rational coordinates and for each i we had $\mu([\mathbf{y_i}, \mathbf{1}] \cap A) = 0$ or $\mu([\mathbf{0}, \mathbf{y_i}] \cap A^c) = 0$. If $\mu([\mathbf{y_i}, \mathbf{1}] \cap A) = 0$ set $N_i^+ = [\mathbf{y_i}, \mathbf{1}] \cap A$ otherwise set $N_i^+ = \emptyset$ and set $N_i^- = [\mathbf{0}, \mathbf{y_i}] \cap A^c$ if $\mu([\mathbf{0}, \mathbf{y_i}] \cap A^c) = 0$ otherwise let $N_i^- = \emptyset$. Then $A' := A \bigcup_i N_i^- \setminus \bigcup_i N_i^+$ differs from A just by a set of measure 0 and A' is a lower layer:

If A' was no lower layer then there were points $\mathbf{a} \leq \mathbf{b}, a \neq b$ with $a \notin A'$ and $b \in A'$. But then also $a \notin A$ and $b \in A$ and there existed an i such that $\mathbf{a} \leq \mathbf{y_i} \leq \mathbf{b}$ and $\mathbf{a} \neq \mathbf{y_i} \neq \mathbf{b}$. But then either $a \in N_i^+$ or $b \in N_i^-$ contradicting that $a \notin A'$ and $b \in A'$.

As this was assumed to be not so above $\mathbf{y}$ must indeed exist.

As μ is absolutely continuous, we can choose sets D and E such that $D \subset A^c, D < \mathbf{y}$ and $E \subset A, E > \mathbf{y}$ with $\mu(D) = \mu(E) > 0$. Then the set $A' = A \setminus E \cup D$ has the same measure as A but $g(A') < g(A) \leq \rho$ again because of the absolute continuity of μ. But this contradicts the optimality of A, as A' could be enlarged a little while still satisfying the constraint.
The proof for A being a solution to (P_2) is almost literally the same.

In their paper Mallows, Nair, Shepp and Vardi say that the solution to (P_2) is always a lower layer, even for an arbitrary probability measure μ. But the following example shows that the hypothesis demanded here is necessary.

To see this consider the counterexample shown in figure 2. Let A contain $\mathbf{a}$ and $\mathbf{c}$ but not $\mathbf{b}$. Let μ be the probability measure which puts the masses $\varepsilon, 1 - 2\varepsilon, \varepsilon$ onto the points $\mathbf{a}, \mathbf{b}$ and $\mathbf{c}$, respectively, for some very small ε. Then A is an optimal

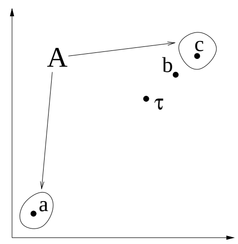

FIGURE 2: counterexample

solution for (P_2) because every set containing the point b has a center of gravity close to b and thus violates the constraint.
For arbitrary μ the solution to (P_1) is also not always a lower layer.

A lower layer A is in particular *starlike* with respect to the origin, i.e. $\mathbf{x} \in A \Rightarrow r\mathbf{x} \in A$ for $0 \leq r \leq 1$.
We will describe a starlike region A by a function in polar coordinates in order to be able to use calculus of variation.
Define a generalized polar coordinate transformation of the positive (and negative) orthant

$$\begin{array}{rcl} \alpha : M \times \mathbb{R} & \rightarrow & \mathbb{R}^d \\ (\boldsymbol{\varphi}, r) = (\varphi_1, \ldots, \varphi_{d-1}, r) & \mapsto & \mathbf{x} = \alpha(\boldsymbol{\varphi}, r), \end{array}$$

where $M := [0, \frac{\pi}{2}]^{d-1}$ and $x_1 = r\cos\varphi_1, \ldots, x_{d-1} = r\sin\varphi_1 \sin\varphi_2 \cdots \sin\varphi_{d-2}\cos\varphi_{d-1}$, $x_d = r\sin\varphi_1\sin\varphi_2\sin\varphi_{d-1}$. I.e. the φ_i are angles between 0 and 90° and r is the distance from the origin. Then $\mathbf{x}$ ranges over $\mathbb{R}^d_+$ when r and $\boldsymbol{\varphi}$ range over $\mathbb{R}_+$ and M, respectively. We know that the functional determinant is $\det D\alpha(\boldsymbol{\varphi}, r) = r^{d-1}\sin^{d-2}\varphi_1 \sin^{d-3}\varphi_2 \cdots \sin\varphi_{d-2}$ and does not vanish for $r > 0$ and $\boldsymbol{\varphi} \in \overset{\circ}{M}$.
Now, define the function $R(\boldsymbol{\varphi}) = \sup\{r \mid \alpha(\boldsymbol{\varphi}, r) \in A\}$. Then $A \subset \{\alpha(\boldsymbol{\varphi}, r) \mid \boldsymbol{\varphi} \in M, 0 \leq r \leq R(\boldsymbol{\varphi})\}$ and the two sets differ only by a set of measure 0. We have a one-to-one correspondence up to sets of measure 0 between the starlike regions in Q and positive functions in polar coordinates. We will call R the function *describing* A.

In terms of R the measure $\mu(A)$ and the coordinates of $\mathbf{g}(A)$ are functionals $J(R)$ and $G_j(R)$ $(1 \leq j \leq d)$, respectively.

$$\begin{array}{rcll} J(R) & := & \mu(A) & \\ & = & \displaystyle\int_{\mathbb{R}^d} \mathbb{1}_A(\mathbf{x})\, f(\mathbf{x})\, d\mathbf{x} & \\ & = & \displaystyle\int_{M\times\mathbb{R}} \mathbb{1}_A(\alpha(\mathbf{z}))\, f(\alpha(\mathbf{z}))\, |\det D\alpha(\mathbf{z})|\, d\mathbf{z} & \text{(transf. formula for Lebesgue integrals)} \\ & = & \displaystyle\int_M \int_0^{R(\boldsymbol{\varphi})} f(\alpha(\boldsymbol{\varphi}, r))\, |\det D\alpha(\boldsymbol{\varphi}, r)|\, dr\, d\boldsymbol{\varphi} & \text{(Fubinis theorem)} \\ & = & \displaystyle\int_M F(\boldsymbol{\varphi}, R(\boldsymbol{\varphi}))\, d\boldsymbol{\varphi}, & \end{array}$$

where

$$F(\boldsymbol{\varphi}, R) := \int_0^R f(\alpha(\boldsymbol{\varphi}, r))\, |\det D\alpha(\boldsymbol{\varphi}, r)|\, dr. \tag{18}$$

Similarly,

$$G_j(R) := g_j(A) = \int_{\mathbb{R}^d} x_j\, \mathbb{1}_A(\mathbf{x})\, f(\mathbf{x})\, d\mathbf{x} = \int_M F_j(\boldsymbol{\varphi}, R(\boldsymbol{\varphi}))d\boldsymbol{\varphi},$$

where

$$F_j(\boldsymbol{\varphi}, R) := \int_0^R \alpha_j(\boldsymbol{\varphi}, r)\, f(\alpha(\boldsymbol{\varphi}, r))\, |\det D\alpha(\boldsymbol{\varphi}, r)|\, dr. \tag{19}$$

Theorem 3. *Given a probability measure μ on Q with a density f with respect to Lebesgue measure and a $\boldsymbol{\rho} > \mathbf{0}$ there is a simplicial section $\hat{A} = \{\mathbf{x} \in Q \mid \langle \mathbf{x}, \boldsymbol{\theta}\rangle \leq 1\}$ which solves (P_1). That is, it maximizes $\mu(A)$ over all $A \subset Q$ which have $\mathbf{g}(A) \leq \boldsymbol{\rho}$. The optimal region is unique up to sets of measure 0.*
Furthermore, for all $i \in \{1, \dots, d\}$ such that the i-th constraint is inactive, i.e. $g_i(\hat{A}) < \rho_i$, we have $\theta_i = 0$.

Remark: In section 2 we only needed that there *is* a simplicial section satisfying the constraint and the latter statement starting “Furthermore ...”. For the proof it is not needed that it actually is optimal. Unfortunately, this weaker statement does not seem to be easier to prove.

Proof. **Step 1.** First assume that the density f is continuous on $\mathbb{R}^d$, $f(\mathbf{x}) > 0$ for $\mathbf{x}$ in $\mathring{Q}$ and $f(\mathbf{x}) = 0$ for $\mathbf{x} \notin Q$. In step 2 we will approximate the general density f of μ by such continuous densities.
Let $\hat{A} \subset Q$ be an *optimal* region and a lower layer given by the function $\hat{R}(\boldsymbol{\varphi})$. And let A be any (measurable) starlike region, given by the function R.
We want to apply the generalized Kuhn-Tucker theorem (e.g. see [9]). Let H be the vector space of all bounded measurable functions on M.
The optimal solution $\hat{R}$ minimizes $-J(R)$ (i.e maximizes $J(R)$) over all $R \in H$ satisfying the constraint

$$\mathbf{G}(R) - \boldsymbol{\rho} \leq \mathbf{0}.$$

The fact that H contains functions R which attain negative values or do not describe a subset of Q does not bring complications. If R is such that for some $\boldsymbol{\varphi} \in M$ $\alpha(\boldsymbol{\varphi}, R(\boldsymbol{\varphi})) \notin Q$ we can define $R^*(\boldsymbol{\varphi})$ to be 0 if $R(\boldsymbol{\varphi})$ is negative and maximal so that $\alpha(\boldsymbol{\varphi}, R^*(\boldsymbol{\varphi})) \in Q$ if $R(\boldsymbol{\varphi})$ was too large. Recall that we have $f(\mathbf{x}) = 0$ for $x \notin Q$, so $J(R) = J(R^*)$ and $\mathbf{G}(R) = \mathbf{G}(R^*)$.
We have to show that J and G are Gateaux differentiable functionals on H and that the variations are linear in their increments.
For any $R, h \in H$ the Gateaux-variation of J at R with increment h is (if it exists)

$$\delta J(R, h) = \frac{d}{d\varepsilon} J(R + \varepsilon h)\Big|_{\varepsilon=0} = \frac{d}{d\varepsilon} \int_M F(\boldsymbol{\varphi}, R + \varepsilon h)\, d\boldsymbol{\varphi}\Big|_{\varepsilon=0}. \tag{20}$$

The integrand $F(\boldsymbol{\varphi}, R + \varepsilon h)$ is differentiable with respect to ε: For every $\boldsymbol{\varphi}$ the integrand in the definition of F

$$f(\alpha(\boldsymbol{\varphi}, r))\, |\det D\alpha(\boldsymbol{\varphi}, r)|$$

is continuous in r, since α, f and $\det D\alpha(\boldsymbol{\varphi}, r)$ are continuous. And therefore we get

$$\begin{aligned}\frac{d}{d\varepsilon}F(\boldsymbol{\varphi}, R+\varepsilon h) &= \frac{\partial F}{\partial R}(\boldsymbol{\varphi}, R+\varepsilon h)\, h(\boldsymbol{\varphi}) \\ &= f(\alpha(\boldsymbol{\varphi}, R+\varepsilon h))\,|\det D\alpha(\boldsymbol{\varphi}, R+\varepsilon h)|\, h(\boldsymbol{\varphi}),\end{aligned}$$

which is bounded in $\boldsymbol{\varphi}$ and ε. Therefore we can differentiate (20) by differentiating the integrand and get

$$\delta J(R,h) = \int_M f(\alpha(\boldsymbol{\varphi}, R))\,|\det D\alpha(\boldsymbol{\varphi}, R)|\, h(\boldsymbol{\varphi})\, d\boldsymbol{\varphi}. \tag{21}$$

It was essential here that f is continuous.
In the same manner we get the Gateaux-variations of G_j $(j = 1, \ldots, d)$:

$$\delta G_j(R,h) = \int_M \alpha_j(\boldsymbol{\varphi}, R) f(\alpha(\boldsymbol{\varphi}, R))\,|\det D\alpha(\boldsymbol{\varphi}, R)|\, h(\boldsymbol{\varphi})\, d\boldsymbol{\varphi}. \tag{22}$$

Since limits in $\mathbb{R}^d$ are defined component wise, we get $\delta\mathbf{G}(R,h) = (\delta G_1(R,h), \ldots, \delta G_d(R,h))$ and we see that both $\delta J(R,h)$ and $\delta\mathbf{G}(R,h)$ are linear in the increment h.
The last thing we need to show to be able to apply the theorem of Kuhn-Tucker is that $\hat{R}$ is a regular point of the inequality $\mathbf{G}(R) - \boldsymbol{\rho} \le \mathbf{0}$.
Since $\mathbf{G}(\hat{R}) \le \boldsymbol{\rho}$ it is sufficient to give an $h \in H$ such that $\delta\mathbf{G}(\hat{R}, h) < \mathbf{0}$.
For this we can simply choose $h(\boldsymbol{\varphi}) = -1$ for all $\boldsymbol{\varphi} \in M$.
Now, by the Kuhn-Tucker theorem we have a $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_d) \ge \mathbf{0}$ such that the Lagrangian

$$L(R) := \langle \mathbf{G}(R) - \boldsymbol{\rho}, \boldsymbol{\theta}\rangle - J(R)$$

is stationary at $\hat{R}$ and

$$\left\langle \mathbf{G}(\hat{R}) - \boldsymbol{\rho}, \boldsymbol{\theta} \right\rangle = 0. \tag{23}$$

As $\hat{R}$ satisfies the constraint $\mathbf{G}(\hat{R}) - \boldsymbol{\rho} \le \mathbf{0}$ and as $\boldsymbol{\theta} \ge 0$, we can conclude by (23) that

$$\theta_i = 0 \quad \text{if} \quad g_i(\hat{A}) = G_i(\hat{R}) < \rho_i,$$

as desired.
It remains to prove that the simplicial section $\{\mathbf{x} \in Q \,|\, \langle x, \boldsymbol{\theta}\rangle \le 1\}$ is an optimal region.
By definition of a stationary point we have for every $h \in H$

$$\begin{aligned} 0 &= \delta L(\hat{R}, h) \\ &= \left\langle \delta\mathbf{G}(\hat{R}, h), \boldsymbol{\theta} \right\rangle - \delta\mathbf{J}(\hat{R}, h) \\ &= \int_M \underbrace{(\left\langle \alpha(\boldsymbol{\varphi}, \hat{R}), \boldsymbol{\theta} \right\rangle - 1)\; f(\alpha(\boldsymbol{\varphi}, \hat{R}))\,|\det D\alpha(\boldsymbol{\varphi}, \hat{R})|}_{=:l(\boldsymbol{\varphi})}\; h(\boldsymbol{\varphi})\, d\boldsymbol{\varphi}. \end{aligned} \tag{24}$$

Since h is arbitrary, we can use

$$h(\boldsymbol{\varphi}) := \begin{cases} 1 & , \text{ if } l(\boldsymbol{\varphi}) \ge 0 \\ -1 & , \text{ if } l(\boldsymbol{\varphi}) < 0 \end{cases}$$

Then $l\,h$ is nonnegative and vanishes exactly when l vanishes. With (24) we conclude that λ-almost surely $l(\boldsymbol{\varphi}) = 0$.

For $\alpha(\boldsymbol{\varphi}, \hat{R}) \in \overset{\circ}{Q}$, $f(\alpha(\boldsymbol{\varphi}, \hat{R})) > 0$ by hypothesis and $\det D\alpha(\boldsymbol{\varphi}, \hat{R}) \neq 0$. So, either $\alpha(\boldsymbol{\varphi}, \hat{R}) \in \partial Q$ or

$$\left\langle \alpha(\boldsymbol{\varphi}, \hat{R}), \boldsymbol{\theta} \right\rangle = 1$$

Let K be the hyperplane $\{\mathbf{x} \,|\, \langle \mathbf{x}, \boldsymbol{\theta} \rangle = 1\}$ (see Figure 3). Then $\partial \hat{A} = \{\alpha(\boldsymbol{\varphi}, \hat{R}(\boldsymbol{\varphi})) \,|\, \boldsymbol{\varphi} \in M\} \subset (K \cap Q) \cup \partial Q$. So as $\hat{A}$ is a lower layer we must have that $\hat{A} = \{\mathbf{x} \in Q \,|\, \langle \mathbf{x}, \boldsymbol{\theta} \rangle \leq 1\}$ or $\hat{A} = Q$, which is a trivial simplicial section.

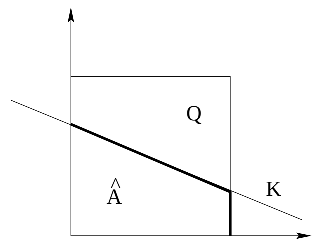

FIGURE 3: The simplicial section $\hat{A}$

Step 2. Now, drop the additional assumptions made in step 1 about the density f of μ. Let μ have the density f, which still is defined on $\mathbb{R}^d$ (that is $f(\mathbf{x}) = 0$ a.s. for $\mathbf{x} \notin Q$), but does not need to be positive on $\overset{\circ}{Q}$ or continuous.
Let $f_1, f_2, \ldots$ be a sequence of probability densities such that f_n satisfies for every n the hypothesis of part 1 and the sequence (f_n) converges to f in L_1, i.e.

- f_n is continuous
- $f_n(\mathbf{x}) > 0$ for $\mathbf{x} \in \overset{\circ}{Q}$, $f_n(\mathbf{x}) = 0$ for $\mathbf{x} \notin Q$
- $\int |f - f_n| \, d\lambda \to 0$

The existence of such a sequence of densities (f_n) can be shown using the fact that the set of continuous functions is dense with respect to the L_1-norm in the set of Lebesgue-integrable functions and the lemma of Urysohn.

Now define $\mu_n := f_n \lambda$ for $n = 1, 2, \ldots$. So μ_n is a probability measure on Q with density f_n and step 1 is applicable to μ_n. Also, for brevity, define $\mathbf{g}^n(A) := \int_A \mathbf{x} \, d\mu_n(\mathbf{x})$ and $\mathbf{c}^n(A) := \mathbf{g}^n(A)/\mu_n(A)$ in analogy to $\mathbf{g}(A)$ and $\mathbf{c}(A)$.
Since $f_n \to f$ in L_1 we now have

$$\mu_n(A) \to \mu(A) \qquad \text{and} \qquad \mathbf{g}^n(A) \to \mathbf{g}(A) \qquad (A \subset Q). \tag{25}$$

Let $\Delta_n = \{\mathbf{x} \,|\, \langle x, \boldsymbol{\theta}^{(n)} \rangle \leq 1\}$ be the simplicial section from part 1. I.e., it is an optimal region for the measure μ_n and $A = \Delta_n$ maximizes $\mu_n(A)$ over $\mathbf{g}^n(A) \leq \boldsymbol{\rho}$. The idea is that a subsequence of (Δ_n) converges to a simplicial section Δ in such a sense that we can conclude that this Δ is optimal for the probability measure μ. We will show that a subsequence of $\boldsymbol{\theta}^{(n)}$ converges in $\mathbb{R}^d$ to a vector $\boldsymbol{\theta}$.
First observe that, as $\boldsymbol{\rho} > \mathbf{0}$, we can show that the $\boldsymbol{\theta}^{(n)}$'s are bounded. Consider the i-th coordinate of $\boldsymbol{\theta}^{(n)}$. By part 1, either $\theta_i^{(n)} = 0$ or $\rho_i = g_i^n(\Delta_n) \leq c_i^n(\Delta_n) \leq \max\{x_i \,|\, \mathbf{x} \in \Delta_n\} \leq 1/\theta_i^{(n)}$. So $\theta_i^{(n)} \leq 1/\rho_i$. With $M := \max_i 1/\rho_i$ we have $\boldsymbol{\theta}^{(n)} \in [0, M]^d$.
Since this set is compact, there is a convergent subsequence of $(\boldsymbol{\theta}^{(n)})$. Let $\boldsymbol{\theta} \in \mathbb{R}_+^d$ be the limit. For simplicity of the notations assume that this subsequence was the sequence $(\boldsymbol{\theta}^{(n)})$ itself. Define

$$\Delta := \{\mathbf{x} \in Q \,|\, \langle \mathbf{x}, \boldsymbol{\theta} \rangle \leq 1\}. \tag{26}$$

We will now show that $\lambda(\Delta_n \diamond \Delta)$ converges to 0.
We can assume that $\boldsymbol{\theta} \neq \mathbf{0}$ (otherwise, $\Delta = Q$ and as $\Delta_n = Q$ for n large enough). Let $\|\boldsymbol{\theta}\|$, the euclidian norm of $\boldsymbol{\theta}$, be positive. By Cauchy-Schwarz's inequality for any $\varepsilon > 0$ there is an n_0 such that for $n > n_0$ $|\langle \mathbf{x}, \theta\rangle - \langle \mathbf{x}, \boldsymbol{\theta}^{(n)}\rangle| < \varepsilon \quad (x \in Q)$.
Now, suppose $\mathbf{x} \in \Delta \diamond \Delta_n$. Then $\langle \mathbf{x}, \boldsymbol{\theta}\rangle \leq 1$ and $\langle \mathbf{x}, \boldsymbol{\theta}^{(n)}\rangle > 1$ or the other way around. In any way we get $\Delta \diamond \Delta_n \subset \{\mathbf{x} \in Q \,|\, \langle \mathbf{x}, \theta\rangle \in (1-\varepsilon, 1+\varepsilon)\}$. The Lebesgue-measure of the right set converges to 0 when $\varepsilon \to 0$, so $\lim_{n\to\infty} \lambda(\Delta \diamond \Delta_n) = 0$. and since μ is continuous with respect to the Lebesgue measure we also get

$$\lim_{n\to\infty} \mu(\Delta \diamond \Delta_n) = 0. \tag{27}$$

As the mapping $\mathbf{g} : \mathcal{B}(Q) \to \mathbb{R}^d$ is continuous with respect to the pseudo norm $(M_1, M_2) \mapsto \mu(M_1 \diamond M_2)$ on $\mathcal{B}(Q)$ this implies that

$$\mathbf{g}(\Delta_n) \to \mathbf{g}(\Delta). \tag{28}$$

Finally we show that $A = \Delta$ maximizes $\mu(A)$ over $\mathbf{g}(A) \leq \boldsymbol{\rho}$ and $g_i(\Delta) < \rho_i \Rightarrow \theta_i = 0$.
Suppose for the sake of contradiction there was an $A \subset Q$ with $\mu(A) > \mu(\Delta)$ and $\mathbf{g}(A) \leq \boldsymbol{\rho}$. Then there also would be an A' with $\mu(A) > \mu(\Delta)$ and $\mathbf{g}(A) < \boldsymbol{\rho}$, because we simply could take away from A a small subset $B \subset A$ of positive measure.
We want to show that then some Δ_n could not have been optimal. First observe that

$$\begin{aligned} |\mu(\Delta) - \mu_n(\Delta_n)| &\leq |\mu(\Delta) - \mu(\Delta_n)| + |\mu(\Delta_n) - \mu_n(\Delta_n)| \\ &\leq \mu(\Delta \diamond \Delta_n) + \int |f - f_n| \, d\lambda \\ &\to 0 \end{aligned} \tag{29}$$

where we used (27) and made use of the choice of f_n.
Similarly,

$$\|\mathbf{g}(\Delta) - \mathbf{g}^n(\Delta_n)\| \to 0. \tag{30}$$

Now, define $\varepsilon := \mu(A') - \mu(\Delta)$ and choose n so large that

(1) $\mathbf{g}^n(A') < \boldsymbol{\rho}$ (possible, since $\mathbf{g}^n(A') \to \mathbf{g}(A') < \boldsymbol{\rho}$ and by (25))

(2) $\mu_n(A') > \mu(A') - \frac{\varepsilon}{2}$ (possible by (25))

(3) $\mu_n(\Delta_n) < \mu(\Delta) + \frac{\varepsilon}{2}$. (possible by (29))

Then

$$\mu_n(A') \overset{(2)}{>} \mu(A') - \frac{\varepsilon}{2} = \mu(\Delta) + \frac{\varepsilon}{2} \overset{(3)}{>} \mu_n(\Delta_n),$$

which contradicts together with (1) the assumption that Δ_n was an optimal region for the measure μ_n.
And suppose $g_i(\Delta) < \rho_i$ then $g_i^{(n)}(\Delta_n) < \rho_i$ for n large enough, because of (30). This implies $\theta_i^{(n)} = 0$ by step 1. So $\theta_i = \lim\limits_{n\to\infty} \theta_i^{(n)} = 0$, too.

It remains to show that the optimal region is unique up to sets of measure 0. Let Δ be as in (26) an optimal region and let B be any other optimal region. We will show that $\mu(\Delta \diamond B) = 0$.
Suppose $\mu(\Delta \diamond B) > 0$. Then $\mu(B \setminus \Delta) = \mu(\Delta \setminus B) =: m > 0$, as $\mu(\Delta) = \mu(B)$.
We have

$$\langle \mathbf{g}(B), \boldsymbol{\theta}\rangle - \langle \mathbf{g}(\Delta), \boldsymbol{\theta}\rangle = \int\limits_{B\setminus\Delta} \langle \mathbf{x}, \boldsymbol{\theta}\rangle \, d\mu(\mathbf{x}) - \int\limits_{\Delta\setminus B} \langle \mathbf{x}, \boldsymbol{\theta}\rangle \, d\mu(\mathbf{x}) \tag{31}$$

But for $\mathbf{x} \in B \setminus \Delta$ we have $\langle \mathbf{x}, \boldsymbol{\theta} \rangle > 1$ and for $\mathbf{x} \in \Delta \setminus B$ we have $\langle \mathbf{x}, \boldsymbol{\theta} \rangle \leq 1$. So the right integral is at most m. Plugging this into (31) yields

$$\langle \mathbf{g}(B), \boldsymbol{\theta} \rangle - \langle \mathbf{g}(\Delta), \boldsymbol{\theta} \rangle \geq \int_{B \setminus \Delta} \underbrace{\langle \mathbf{x}, \boldsymbol{\theta} \rangle - 1}_{>0} \, d\mu(\mathbf{x}) > 0$$

as $\mu(B \setminus \Delta) > 0$. We get

$$\langle \mathbf{g}(B), \boldsymbol{\theta} \rangle > \langle \mathbf{g}(\Delta), \boldsymbol{\theta} \rangle = \langle \boldsymbol{\rho}, \boldsymbol{\theta} \rangle \tag{32}$$

where the equation on the right holds because $\theta_i = 0$ if $g_i(\Delta) \neq \rho_i$. But (32) implies that $g_i(B) > \rho_i$ for some $i \in \{1, \ldots, d\}$, which is a contradiction. So $\mu(\Delta \diamond B) = 0$.

Corollary 1. *Given a probability measure μ on Q with a density f and given a $\boldsymbol{\tau} > \mathbf{0}$ there is a simplicial section $\hat{A} = \{\mathbf{x} \in Q \mid \langle \mathbf{x}, \boldsymbol{\theta} \rangle \leq 1\}$ which solves (P_2). That is, it maximizes $\mu(A)$ amongst all $A \subset Q$ which have $\mathbf{c}(A) \leq \boldsymbol{\tau}$.*
For all i such that the i-th constraint is inactive, i.e. $c_i(\hat{A}) < \tau_i$, we have $\theta_i = 0$.
Furthermore, the optimal region is unique up to sets of measure 0.

Proof. Let A be a solution to (P_2), i.e. $\mathbf{c}(A) \leq \boldsymbol{\tau}$ and whenever $\mathbf{c}(A') \leq \boldsymbol{\tau}$ we have $\mu(A') \leq \mu(A)$. We have $\mu(A) > 0$ because $\boldsymbol{\tau} > \mathbf{0}$. Define $\boldsymbol{\rho} > \mathbf{0}$ by

$$\boldsymbol{\rho} := \mu(A)\boldsymbol{\tau} \in Q.$$

We will show that A also maximizes $\mu(A)$ under the restriction $\mathbf{g}(A) \leq \boldsymbol{\rho}$.
First note that A satisfies this restriction since $\mathbf{g}(A) = \mu(A)\mathbf{c}(A) \leq \mu(A)\boldsymbol{\tau} = \boldsymbol{\rho}$.
Now, let A' be any set such that $\mathbf{g}(A') \leq \boldsymbol{\rho}$. Then we have

$$\begin{aligned} \mu(A')\mathbf{c}(A') &= \mathbf{g}(A') \\ &\leq \boldsymbol{\rho} \\ \mu(A')\mathbf{c}(A') &= \mu(A)\boldsymbol{\tau} \end{aligned} \tag{33}$$

Now, if $\mathbf{c}(A') \leq \boldsymbol{\tau}$, we have $\mu(A') \leq \mu(A)$ since $\mu(A)$ was maximal under that condition.
And if $\mathbf{c}(A') \not\leq \boldsymbol{\tau}$ there is at least one coordinate i such that $c_i(A') > \tau_i$. But then the i-th coordinate of inequality (33) tells us that $\mu(A') \leq \mu(A)$.
Since A' was arbitrary, A also maximizes $\mu(A)$ under the restriction $\mathbf{g}(A) \leq \boldsymbol{\rho}$.
Because the solution to that problem is unique and a lower-layer, by Theorem 3, we get that A is a simplicial section $\hat{A} = \{\mathbf{x} \in Q \mid \langle \mathbf{x}, \boldsymbol{\theta} \rangle \leq 1\}$ up to sets of measure 0. And since $\mu(A) = \mu(\hat{A})$ we have $\theta_i = 0$ for each i such that $g_i(\hat{A}) < \rho_i = \mu(\hat{A})\tau_i$.
Or equivalently,

$$\theta_i = 0 \qquad \text{if} \quad c_i(\hat{A}) < \tau_i.$$

Acknowledgement

I would like to thank Alexander V. Gnedin for his help and his support.

References

[1] Bruss, F. Thomas, Robertson, James B. (1991). *'Wald's Lemma' for Sums of Order Statistics of i.i.d Random Variables*, Adv. Appl. Prob. **23**, 612-623.

[2] Coffman, E.G., Flatto, L. and Weber, R.R. (1987). *Optimal selection of stochastic intervals under a sum constraint.*, Adv. Appl. Prob. **19**, 454-473.

[3] Coffman, E.G. jr., Lueker, G.S. (1991). *Probabilistic Analysis of Packing and Partitioning Algorithms*, John Wiley & Sons.

[4] Garey, M. R., Graham, R. L. and Johnson, D. S. (1976). *Resource Constrained Scheduling as Generalized Bin Packing*, Journal of Combinatorial Theory (A) **21**, 257-298.

[5] Gnedin, Alexander V., Baryshnikov, Yuliy M. (2000). *Sequential Selection of an Increasing Sequence from a Multidimensional Random Sample*, Ann. Appl. Probab. **10**, No.1, 258-267

[6] Gnedin, Alexander V. (1999). *Sequential Selection under a Sum Constraint when the Number of Observations is Random*, to appear in *Proc. 4th Hungarian Conf. on Limit Theorems in Probability and Statistics.*

[7] Gnedin, Alexander V. (2000). *Sequential Selection of an increasing subsequence from a random sample with geometrically distributed sample-size*, Lect. Notes of the Inst. of Math. Stat., Monogr. Ser. **35**, 101-109

[8] Gnedin, Alexander V. (1999). *Sequential selection of an increasing subsequence from a sample of random size* J. Appl. Probab. **36**, No.4, 1074-1085

[9] Luenberger, David G. *Optimization by Vector Space Methods* John Wiley & Sons

[10] Mallows C.L., Nair V.N., Shepp L.A. and Vardi Y. (1985). *Optimal Sequential Selection of Secretaries*, Mathematics of Operations Research, **10**, No. 4.

[11] Rhee, Wansoo, Talagrand, Michel (1991). *A Note on the Selection of Random Variables under a Sum Constraint*, J. Appl. Prob. **28**, 919-923