A NEW FORM OF JENSEN'S INEQUALITY AND ITS APPLICATION TO STATISTICAL EXPERIMENTS

D. J. F. NONNENMACHER and R. ZAGST[1]

Abstract

Jensen's inequality for the expectation of a convex function of a random variable is proved for a wide class of convex functions defined on a space of probability measures. The result is applied to statistical experiments using the concept of Blackwell-sufficiency. In particular, we show a monotonicity result for the expected information of Poisson-experiments. As an application to economics we consider the introduction of new production technologies.

1. Introduction

The importance of Jensen's inequality for the expectation of a convex real-valued function in probability theory and mathematical statistics is well recognized and requires no discussion. One form of this inequality is the following:

Let S be a convex subset of the n-dimensional Euclidean space $\mathbb{R}^n$ and let X be an integrable random variable with range in S. Then for any convex function $f : S \to \mathbb{R}$ the expected value $Ef(X)$ of $f(X)$ exists, $EX \in S$ and

$$f(EX) \leq Ef(X).$$

In this paper we prove Jensen's inequality for a class of convex functions $f : \mathbb{P}(\Theta) \to [-\infty, \infty)$, where $\mathbb{P}(\Theta)$ denotes the space of all probability measures defined on the Borel σ-algebra of a separable and complete metric space Θ. Since in general we face an infinite-dimensional situation we require f in addition to be lower semicontinuous. (Recall that in the n-dimensional case the convexity of f implies the continuity of f in the interior of S.)

The formulation of our theorem is free of the notion of Bochner integration and related concepts, and the relevant quantity EX is explicitly stated in given terms. To prove the theorem we embed $\mathbb{P}(\Theta)$ in a suitable Banach space, we identify EX to

[1]Dept of Mathematics & Economics, Universität Ulm, Helmholtzstr. 18, D-89069 Ulm, Germany.

be the Bochner expected value of the random variable X in this Banach space, and we then apply the strong law of large numbers for Banach-valued random variables. For a detailed discussion of Jensen's inequality in infinite-dimensional linear spaces see [16].

In the applications we consider Θ to be an index space and we use the stated inequality to prove some sensitivity results on the expected information (relative to a function f) of a special class of statistical experiments, the so-called Poisson-experiments. Here we make use of the concept of Blackwell-sufficiency, which is well-examined in statistical sufficiency theory, to show how the expected posterior distributions can be calculated. The results are then applied to an interesting economical problem concerning the introduction of new production technologies.

2. Preliminaries

For any topological space Y we denote by $\mathcal{B}(Y)$ the σ-algebra generated by its open sets, that is, the Borel σ-algebra.

A separable and complete metric space (M, d) will be called a Polish space, and the set of all probability measures defined on $\mathcal{B}(M)$ is denoted by $\mathbb{P}(M)$. We will always consider $\mathbb{P}(M)$ as a topological space with the weak topology, and if $\mu \in \mathbb{P}(M)$ is the weak limit of a sequence of probability measures μ_n we write $\mu_n \xrightarrow{w} \mu$. Of course this means that $\int_M f \, d\mu_n$ converges to $\int_M f \, d\mu$ for each $f \in C(M)$, where $C(M)$ denotes the space of all real-valued bounded and continuous functions on M.

On the set $\mathbb{B}V(M)$ ($\supseteq \mathbb{P}(M)$) of all Borel measures of bounded variation on M the dual bounded Lipschitz norm $\| \cdot \|_{BL_*}$ can be defined (see for example, [1], page 10) and the associated metric is denoted by β. In particular, the subspace $(\mathbb{P}(M), \beta)$ is a Polish space, and β metrizes the weak topology (see for example, [1], pages 11, 18). We embed the space $\mathbb{B}V(M)$ in the usual way into a Banach space $(\mathcal{X}, \| \cdot \|)$ such that $\|\mu\| = \|\mu\|_{BL_*}$ for all $\mu \in \mathbb{B}V(M)$.

If (Ω, Σ, P) denotes a probability space, a map X from Ω into a topological space Y is called a random variable if $X^{-1}(A) \in \Sigma$ for each $A \in \mathcal{B}(Y)$. As usual P_X denotes the distribution of X with respect to P (see for example, [2]).

For any random variable $X : \Omega \to Y$, Y now being a Banach space, which is Bochner integrable (see for example, [10]) we denote by EX the Bochner integral of X, and we call EX the Bochner expected value of X.

Obviously every random variable $X : \Omega \to \mathbb{P}(M)$ is also a random variable considered as a mapping from Ω into $\mathcal{X}$ (note that $\mathcal{B}(\mathbb{P}(M)) = \mathcal{B}(\mathcal{X}) \cap \mathbb{P}(M)$) and since $\|\mu\|_{BL_*} \leq 1$ for each $\mu \in \mathbb{P}(M)$ we see by [10], Theorems 3.5.3 and 3.7.4, that X is Bochner integrable. Consequently X is Pettis integrable and thus $l(EX) = \int_\Omega l \circ X \, dP$ holds for all $l \in \mathcal{X}^*$, where $\mathcal{X}^*$ denotes the dual space of $\mathcal{X}$. Evidently

EX is uniquely determined by this property.

3. Jensen's inequality

THEOREM 3.1. *Suppose* (Θ, d) *to be a Polish space, let* (Ω, Σ, P) *be a probability space and assume* $X : \Omega \to \mathbb{P}(\Theta)$ *to be a random variable.*
Then a probability measure $\mu_X \in \mathbb{P}(\Theta)$ *can be defined by*

$$\mu_X(A) = \int_\Omega X(\omega)(A)\, dP(\omega) \qquad (A \in \mathcal{B}(\Theta)),$$

and for any convex and lower semi-continuous map $f : \mathbb{P}(\Theta) \to [-\infty, \infty)$ *we have Jensen's inequality*

$$f(\mu_X) \leq \int_\Omega f \circ X\, dP,$$

provided that $\int_\Omega |f \circ X|\, dP$ *remains finite.*

REMARK. The proof of the theorem will show that μ_X can be identified to be the Bochner expected value of $X : \Omega \to \mathcal{X}$ ($\mathcal{X}$ denoting a Banach space according to the preliminaries).

Within the proof of Theorem 3.1 we will need the following well-known result (see for example, [3], Theorem 4, page 237).

LEMMA 3.2. *Let* (M, d) *be a Polish space. Then each probability measure* $\mu \in \mathbb{P}(M)$ *is the weak limit of a sequence of discrete probability measures in* $\mathbb{P}(M)$.

(The latter are those measures which can be represented in a form $\sum \alpha_i \varepsilon_{x_i}$ by means of finitely many points $x_i \in M$ and non-negative real numbers α_i, where ε_{x_i} denotes the probability measure which is equal to one for each $A \in \mathcal{B}(M)$ containing x_i and zero else.)

PROOF OF THEOREM 3.1. First note that for any $A \in \mathcal{B}(\Theta)$ the map $F_A : \mathbb{P}(\Theta) \to \mathbb{R}$ defined by $F_A(\mu) = \mu(A)$ is (Borel-) measurable.
For, if $G \subseteq \Theta$ denotes an open set, and if $\mu_n \overset{w}{\to} \mu$ we have $\liminf F_G(\mu_n) \geq F_G(\mu)$, see [3], Theorem 2.1, page 11. Thus F_G is lower semi-continuous, hence measurable. Now set $\lambda = \{A \in \mathcal{B}(\Theta) : F_A \text{ is measurable}\}$ and let π consist of all open subsets of Θ. Applying Dynkin's π-λ-theorem (see [4], Theorem 3.2, page 37) then yields $\mathcal{B}(\Theta) = \lambda$.

Consequently, we see that for any $A \in \mathscr{B}(\Theta)$ the map $\omega \mapsto X(\omega)(A)$, $\omega \in \Omega$, is a random variable, and we therefore may define a probability measure $\mu_X \in \mathbb{P}(\Theta)$ according to

$$\mu_X(A) = \int_\Omega X(\omega)(A) \, dP(\omega) \qquad (A \in \mathscr{B}(\Theta)).$$

Since the distribution P_X of X is a probability measure on $\mathscr{B}(\mathbb{P}(\Theta))$ and since $(\mathbb{P}(\Theta), \beta)$ is a Polish space we can find by Lemma 3.2 a sequence of discrete probability measures $P_n \in \mathbb{P}(\mathbb{P}(\Theta))$ converging weakly to P_X.

Writing $P_n = \sum_j \alpha_j^n \varepsilon_{s_j^n}$ with $s_j^n \in \mathbb{P}(\Theta)$, $\alpha_j^n \geq 0$ we prove that $\mu_n = \sum_j \alpha_j^n s_j^n \xrightarrow{w} \mu_X$ in $\mathbb{P}(\Theta)$ for which it suffices to show that $\int_\Theta f \, d\mu_n$ converges to $\int_\Theta f \, d\mu_X$ for each $f \in C(\Theta)$.

Thus for $f \in C(\Theta)$ we define the function $g : \mathbb{P}(\Theta) \to \mathbb{R}$ by setting

$$g(\nu) = \int_\Theta f \, d\nu, \qquad \nu \in \mathbb{P}(\Theta).$$

Obviously $g \in C(\mathbb{P}(\Theta))$ and using standard arguments we conclude that

$$\int_\Theta f \, d\mu_X = \int_\Omega g \circ X \, dP = \int_{\mathbb{P}(\Theta)} g \, dP_X = \lim_{n \to \infty} \int_{\mathbb{P}(\Theta)} g \, dP_n$$

$$= \lim_{n \to \infty} \sum_j \alpha_j^n g(s_j^n) = \lim_{n \to \infty} \int_\Theta f \, d\mu_n$$

as desired.

As explained in the preliminaries the Bochner expected value EX of X exists as an element in an associated Banach space $\mathscr{X}$. Taking any $l \in \mathscr{X}^*$ and observing that

$$l(EX) = \int_\Omega l \circ X \, dP = \int_{\mathbb{P}(\Theta)} l \, dP_X = \lim_{n \to \infty} \int_{\mathbb{P}(\Theta)} l \, dP_n = \lim_{n \to \infty} l(\mu_n) = l(\mu_X)$$

we conclude that $EX = \mu_X$.

Now, having identified μ_X to be the Bochner expected value of X, the desired inequality is a simple consequence of the strong law of large numbers for Banach-valued random variables (see for example, [8], Theorem 1.1), see also [16], Theorem 4.1.

First assume the given function f to be finite and choose a sequence of $\mathscr{X}$-valued independent random variables X_i (on a suitable probability space) such that each X_i assumes only values in $\mathbb{P}(\Theta)$ and the distribution of X_i coincides with P_X. (The existence of such a sequence is guaranteed according to [2], Corollary 5.4.5.)

By the convexity of f and $\mathbb{P}(\Theta)$ we have $f(\sum_{i=1}^n \frac{1}{n} X_i) \leq \sum_{i=1}^n \frac{1}{n} f \circ X_i$, $n \in \mathbb{N}$, and since $\sum_{i=1}^n \frac{1}{n} X_i \to \mu_X$ and $\sum_{i=1}^n \frac{1}{n} f \circ X_i \to \int_\Omega f \circ X \, dP$ (a.s.) the lower semi-continuity of f implies:

$$f(\mu_X) \leq \liminf f \left(\sum_{i=1}^n \frac{1}{n} X_i \right) \leq \int_\Omega f \circ X \, dP.$$

Finally let f be as in Theorem 3.1.

We then introduce the finite, convex and lower semi-continuous functions $f_n = \max(-n, f)$, $n \in \mathbb{N}$, which converge monotone to f, and applying the monotone convergence theorem we get

$$f(\mu_X) \leq f_n(\mu_X) \leq \int_\Omega f_n \circ X \, dP \to \int_\Omega f \circ X \, dP.$$

4. Poisson-experiments

By $\mathbb{R}_{++}$ we denote the set of all positive real numbers, $\mathbb{R}_+ = \mathbb{R}_{++} \cup \{0\}$, and we let Λ be an arbitrary non-empty subset of $\mathbb{R}_{++}$ while we assume $\Theta \subseteq \mathbb{R}_{++}$ to be non-empty and closed in $\mathbb{R}$ (which is necessary and sufficient for Θ to be Polish, concerning the usual Euclidean metric in $\mathbb{R}$). With each $\lambda \in \Lambda$ we associate a family $q_\lambda = \{q(\cdot|\theta, \lambda) : \mathbb{N}_0 \to \mathbb{R}, \theta \in \Theta\}$ and, following the definition of statistical experiments (see [6], [9], [19]), we call q_λ a Poisson-experiment if for each $\theta \in \Theta$ the function $q(\cdot|\theta, \lambda)$ is the density of a Poisson distribution with parameter $\lambda\theta$. In what follows we will assume q_λ to be a Poisson-experiment for all $\lambda \in \Lambda$.

Suppose that the information on the parameter $\theta \in \Theta$ can be expressed by a so-called prior distribution $\rho \in \mathbb{P}(\Theta)$, let $\lambda \in \Lambda$ and assume $\omega \in \mathbb{N}_0$ to be the result of the Poisson-experiment q_λ.

Then the so-called posterior distribution after performing the experiment q_λ is given by the operator $\Phi_\lambda : \mathbb{P}(\Theta) \times \mathbb{N}_0 \to \mathbb{P}(\Theta)$, known as Bayes-operator, and defined by

$$\Phi_\lambda(\rho, \omega)(B) = \int_B \frac{q(\omega|\theta, \lambda)}{q(\omega|\rho, \lambda)} \, d\rho(\theta) \qquad (B \in \mathscr{B}(\Theta)),$$

where $q(\omega|\rho, \lambda) = \int_\Theta q(\omega|\theta, \lambda) \, d\rho(\theta) \, (\in \mathbb{R}_{++})$.

To quantify the information we get by performing a Poisson-experiment $q_\lambda, \lambda \in \Lambda$, we use a slight generalization of the definition given in [7]:

Let $f : \mathbb{P}(\Theta) \to [-\infty, \infty)$ be a convex lower semi-continuous map. The expected information $I_f(\rho, \lambda)$ of a Poisson-experiment $q_\lambda, \lambda \in \Lambda$, relative to f and given a prior distribution $\rho \in \mathbb{P}(\Theta)$ is defined by (existence provided)

$$I_f(\rho, \lambda) = \sum_{\omega=0}^{\infty} q(\omega|\rho, \lambda) \, f(\Phi_\lambda(\rho, \omega)) - f(\rho).$$

$I_f(\rho, \lambda)$ is the increment we get by a transition from the prior distribution ρ to the posterior distribution $\Phi_\lambda(\rho, \omega)$ measured in values of the map f.

The examination of the expected information is not only important for statistical information and entropy theory (see for example, [6], [7], [19]) but also for optimization theory concerning the Bayesian information models where f can be identified

to be the so-called value function of a finite horizon decision problem (see Section 5 and [17]).

THEOREM 4.1. *Let ρ be an element of $\mathbb{P}(\Theta)$ and $f : \mathbb{P}(\Theta) \to [-\infty, \infty)$ be a convex and lower semi-continuous map such that $\sum_{\omega=0}^{\infty} q(\omega|\rho, \lambda)\,|f(\Phi_\lambda(\rho, \omega))|$ remains finite for all $\lambda \in \Lambda$. Then we have*

$$0 \leq I_f(\rho, \lambda) \leq I_f(\rho, \lambda')$$

for each pair $(\lambda, \lambda') \in \Lambda \times \Lambda$ with $\lambda \leq \lambda'$.

This means that the expected information of a Poisson-experiment is always non-negative and increasing in $\lambda \in \Lambda$.

To prove Theorem 4.1 we need an auxiliary lemma. Within the proof of this lemma we establish a relation between two experiments q_λ and $q_{\lambda'}$ which was theoretically introduced for statistical experiments in [5] and is known as Blackwell-sufficiency (see also [6], [9], [13], [19]).

LEMMA 4.2. *For each $\rho \in \mathbb{P}(\Theta)$ and for each pair $(\lambda, \lambda') \in \Lambda \times \Lambda$ with $\lambda < \lambda'$ there exists a map $h' : \mathbb{N}_0 \times \mathbb{N}_0 \to \mathbb{R}_+$ such that $\sum_{\omega'=0}^{\infty} h'(\omega, \omega') = 1$ for all $\omega \in \mathbb{N}_0$ and such that the following two statements are fulfilled:*

a) $\displaystyle \sum_{\omega=0}^{\infty} h'(\omega, \omega')q(\omega|\rho, \lambda) = q(\omega'|\rho, \lambda')$ *for all $\omega' \in \mathbb{N}_0$.*

b) $\displaystyle \sum_{\omega'=0}^{\infty} \Phi_{\lambda'}(\rho, \omega')(A)\, h'(\omega, \omega') = \Phi_\lambda(\rho, \omega)(A)$ *for each $A \in \mathscr{B}(\Theta)$ and for all $\omega \in \mathbb{N}_0$.*

PROOF. Suppose $\rho \in \mathbb{P}(\Theta)$ and $(\lambda, \lambda') \in \Lambda \times \Lambda$ with $\lambda < \lambda'$ to be given.
Define a function $h : \mathbb{N}_0 \times \mathbb{N}_0 \to \mathbb{R}_+$ by

$$h(\omega, \omega') = \binom{\omega'}{\omega} \left(\frac{\lambda}{\lambda'}\right)^{\omega} \left(1 - \frac{\lambda}{\lambda'}\right)^{\omega'-\omega}$$

for $\omega \leq \omega'$ and by zero otherwise. Observe that for $\omega' \in \mathbb{N}_0$ we have $\sum_{\omega=0}^{\infty} h(\omega, \omega') = 1$ and $\sum_{\omega'=0}^{\infty} h(\omega, \omega')\, q(\omega'|\theta, \lambda') = q(\omega|\theta, \lambda)$ for all $\omega \in \mathbb{N}_0, \theta \in \Theta$. Now define $h' : \mathbb{N}_0 \times \mathbb{N}_0 \to \mathbb{R}_+$ by $h'(\omega, \omega') = h(\omega, \omega')q(\omega'|\rho, \lambda')/q(\omega|\rho, \lambda), \omega, \omega' \in \mathbb{N}_0$, and prove the statements straightforward.

PROOF OF THEOREM 4.1. Let $\rho \in \mathbb{P}(\Theta)$ and f as stated in the theorem. We may assume $f(\rho)$ to be finite since otherwise $I_f(\rho, \lambda) = \infty$ for all choices of λ. We take

$(\lambda, \lambda') \in \Lambda \times \Lambda$ with $\lambda < \lambda'$, and, using Lemma 4.2 we have

$$\sum_{\omega=0}^{\infty} q(\omega|\rho, \lambda) \sum_{\omega'=0}^{\infty} h'(\omega, \omega') |f(\Phi_{\lambda'}(\rho, \omega'))| = \sum_{\omega'=0}^{\infty} q(\omega'|\rho, \lambda') |f(\Phi_{\lambda'}(\rho, \omega'))| < \infty$$

by assumption.

Thus for any $\omega \in \mathbb{N}_0$ we have $\sum_{\omega'=0}^{\infty} h'(\omega, \omega') |f(\Phi_{\lambda'}(\rho, \omega'))| < \infty$ and considering the probability space $(\mathbb{N}_0, \mathscr{P}(\mathbb{N}_0), P_\omega)$ (where $\mathscr{P}(\mathbb{N}_0)$ denotes the power set of $\mathbb{N}_0$ and the probability measure P_ω is defined by $P_\omega(\{\omega'\}) = h'(\omega, \omega')$, $\omega' \in \mathbb{N}_0$) and the random variable $\Phi_{\lambda'}(\rho, \cdot) : \mathbb{N}_0 \to \mathbb{P}(\Theta)$ we get, using again the previous lemma and Jensen's inequality, that

$$f(\Phi_\lambda(\rho, \omega)) \leq \sum_{\omega'=0}^{\infty} h'(\omega, \omega') f(\Phi_{\lambda'}(\rho, \omega')).$$

Multiplying the inequality with $q(\omega|\rho, \lambda)$, summing over ω and subtracting $f(\rho)$ yields

$$I_f(\rho, \lambda) \leq I_f(\rho, \lambda')$$

as desired.

To prove that $I_f(\rho, \lambda)$ is non-negative take $\lambda \in \Lambda$. This time we consider the probability space $(\mathbb{N}_0, \mathscr{P}(\mathbb{N}_0), P)$ where the probability measure P is defined by $P(\{\omega\}) = q(\omega|\rho, \lambda)$, $\omega \in \mathbb{N}_0$, and the random variable $\Phi_\lambda(\rho, \cdot)$. Using the easily checked martingale property of Φ_λ, stating that $\sum_{\omega=0}^{\infty} q(\omega|\rho, \lambda) \Phi_\lambda(\rho, \omega)(A) = \rho(A)$ $for\ all$ $A \in \mathscr{B}(\Theta)$, again Jensen's inequality yields

$$f(\rho) \leq \sum_{\omega=0}^{\infty} q(\omega|\rho, \lambda) f(\Phi_\lambda(\rho, \omega)),$$

which completes the proof.

Looking at the proofs above we see that the statements in Lemma 4.2 are essential. It should be noted that the theorem can be proved analogously for all statistical experiments which can be compared by using the concept of Blackwell-sufficiency (see for example, [7], [14], [15], [21]).

5. New production technologies

Consider a company having the choice between two production techniques 1 (representing an old production process) and 2 (representing a new process). The profit function of each technique is assumed to be a linear function of the capital invested

in it. The company has to decide upon the quantity (proportion) of the two processes to be used in production. The problem is complicated by the returns of process 2 which are not known by certainty. In contrary, the returns of the old technique are well-known and are assumed to be $r_1 \in \mathbb{R}_+$. Suppose λ_1 and λ_2 to be the quantities of process 1 and 2 in production, and suppose that the return ω per unit of technology 2 can be described as the realization of a random variable with distribution $Q(\theta, \lambda_2; \cdot)$ on a probability space (Ω, Σ) where $\theta \in \Theta$ is an unknown parameter and Θ is a compact subset of $\mathbb{R}$. Suppose that $q(\cdot|\theta, \lambda_2) > 0$ is a ν-density of $Q(\theta, \lambda_2; \cdot)$ with respect to a σ-finite measure ν on $\mathscr{B}(\Omega)$ which is continuous in θ (ν is assumed to be independent of θ and λ_2). Furthermore, we assume the existence of a measurable map $g : \Omega \to \overline{\mathbb{R}}$ with $\int_\Omega g(\omega) \, d\nu(\omega) < \infty$ such that

$$q(\omega|\theta, \lambda) \le g(\omega) \quad \text{for all } \omega \in \Omega, \ \theta \in \Theta, \ \lambda \in \Lambda.$$

Most of these assumptions are needed in order to give a rigorous proof of the results to follow, in particular the continuity of $V_n(\cdot)$.

There is an investment budget of M dollars in each period. The (relative) prices (costs) per quantity used are 1 dollar for the old technique and $p \in \mathbb{N}$ dollars for the new one. Therefore, $\lambda_1 = M - p\lambda_2$ and $\lambda_2 \in \Lambda := \{\lambda \in \mathbb{N}_0 : \lambda \le M/p\}$. The aim of the company is to maximize the expected discounted total profit over $N \in \mathbb{N}$ periods where $\beta \in \mathbb{R}_{++}$ is an appropriate discount factor (see for example, [11], page 43). Let $\rho \in \mathbb{P}(\Theta)$ be the company's information on the unknown parameter. Then the expected periodic return is given by $r(\rho, \lambda_2) = \int_\Theta r(\theta, \lambda_2) \, d\rho(\theta)$ with

$$r(\theta, \lambda_2) = r_1 \cdot (M - p\lambda_2) + \lambda_2 \int_\Omega \omega \, dQ(\theta, \lambda_2; \omega).$$

We assume that $\theta \mapsto r(\theta, \lambda_2) \in C(\Theta)$ for all $\lambda_2 \in \Lambda$.

The following results are well-known from stochastic dynamic programming (see for example, [11], page 40):

1. The *maximum expected discounted total reward* V_N over N periods can be computed by the so-called *value iteration*

$$V_n(\rho) = \max_{\lambda_2 \in \Lambda} \left\{ r(\rho, \lambda_2) + \beta \int_\Omega V_{n-1}(\Phi_{\lambda_2}(\rho, \omega)) \, dQ(\rho, \lambda_2; \omega) \right\},$$

$1 \le n \le N$, $\rho \in \mathbb{P}(\Theta)$ with $V_0 \equiv 0$, and $\Phi_{\lambda_2} : \mathbb{P}(\Theta) \times \Omega \to \mathbb{P}(\Theta)$ (see Section 4) is defined by

$$\Phi_{\lambda_2}(\rho, \omega)(B) = \int_B \frac{q(\omega|\theta, \lambda_2)}{q(\omega|\rho, \lambda_2)} \, d\rho(\theta) \quad (B \in \mathscr{B}(\Theta)),$$

where $q(\omega|\rho, \lambda_2) = \int_\Theta q(\omega|\theta, \lambda_2) \, d\rho(\theta) (\in \mathbb{R}_{++})$ is the ν-density of $Q(\rho, \lambda_2; \cdot) = \int_\Theta Q(\theta, \lambda_2; \cdot) \, d\rho(\theta)$.

2. Let $f_n(\rho)$ be the greatest maximum point of the map

$$\lambda_2 \mapsto r(\rho, \lambda_2) + \beta \int V_{n-1}(\Phi_{\lambda_2}(\rho, \omega)) \, dQ(\rho, \lambda_2; \omega), \qquad \lambda_2 \in \Lambda$$

for $\rho \in \mathbb{P}(\Theta)$, $1 \leq n \leq N$. Then the so-called policy $\pi = (f_N, f_{N-1}, \ldots, f_1)$ is optimal, that is, the expected discounted total reward is maximized by investing $f_n(\rho)$ at the beginning of period $N - n + 1$ in the new technology if the prior distribution on the unknown parameter is $\rho \in \mathbb{P}(\Theta)$.

The convexity and continuity of V_n in ρ ($1 \leq n \leq N$) can easily be established using [12], Theorem 19.9. For example, taking [18], Lemma 4.4 into account, we obtain:

THEOREM 5.1. *Suppose Λ to be totally ordered by some ordering $\leq_\Lambda$, let $\rho \in \mathbb{P}(\Theta)$ and assume that $I_{V_{n-1}}(\rho, \lambda_2) \leq I_{V_{n-1}}(\rho, \lambda_2')$ holds for each $1 \leq n \leq N$ and each pair $(\lambda_2, \lambda_2') \in \Lambda \times \Lambda$ with $\lambda_2 \leq_\Lambda \lambda_2'$. Then we have*

$$f_1(\rho) \leq_\Lambda f_n(\rho), \quad 1 \leq n \leq N,$$

that is, $f_1(\rho)$ is a lower bound for the optimal investment in the new technology for the information ρ on the unknown parameter and for each period $1 \leq n \leq N$. (Here the expected information is analogously defined as in Section 4.)

Recall that according to Section 4 (using Theorem 3.1) the assumption in Theorem 5.1 is always satisfied in the case of Blackwell-sufficiency.

EXAMPLE 1. Let $\Theta \subseteq \mathbb{R}_{++}$, $\Omega = \mathbb{N}_0$ and $Q(\theta, \lambda_2; \cdot)$ be a Poisson distribution with parameter $\lambda_2 \theta$, $\lambda_2 \in \Lambda \cap \mathbb{N}$. Then it was proved in Section 4 that if $\leq_\Lambda$ denotes the natural ordering, the future expected rewards

$$\int_\Omega V_{n-1}(\Phi_{\lambda_2}(\rho, \omega)) \, dQ(\rho, \lambda_2; \omega), \qquad 1 \leq n \leq N$$

are increasing as the invested amount of money in the new technology is increased and therefore the so-called myopic investment $f_1(\rho)$ is a lower bound for the optimal investment.

EXAMPLE 2. (See [20].) Let $\Omega = \mathbb{R}$ and let $Q(\theta, \lambda_2; \cdot)$ be a normal distribution with unknown mean θ and a variance of $1/\lambda_2$, $\lambda_2 \in \Lambda \cap \mathbb{N}$. Thus the uncertainty on the return of the new process is decreasing as the quantity of this process used in production is increased.

Again Theorem 5.1 applies (see [18], [21]) which means that larger investments λ_2 (again with respect to the natural ordering) lead to more certainty about the new

process and therefore to higher future profits. This is because the company not only obtains returns from the process itself but also gains information on the unknown technique which will be useful in the following periods. Therefore, as in Example 1, the so-called myopic investment $f_1(\rho)$ is a lower bound for the optimal investment.

Acknowledgement

References

[1] A. Araujo and E. Giné, *The central limit theorem for real and Banach valued random variables* (John Wiley & Sons, New York, 1980).

[2] H. Bauer, *Probability theory and elements of measure theory* (Academic Press, London, 1981).

[3] P. Billingsley, *Convergence and probability measures* (John Wiley & Sons, New York, 1968).

[4] P. Billingsley, *Probability and measure* (John Wiley & Sons, New York, 1986).

[5] D. Blackwell, "Equivalent comparisons of experiments", *Ann. Math. Stat.* **24** (1953) 265–272.

[6] M. H. DeGroot, *Optimal statistical decisions* (McGraw-Hill, New York, 1970).

[7] M. H. DeGroot, "Changes in utility as information", *Theory and Decision* **17** (1984) 287–303.

[8] D. P. Giesy, "Strong law of large numbers for independent sequences of Banach space-valued random variables", Lecture Notes in Mathematics, No. 526, (Springer, New York, 1976).

[9] H. Heyer, *Theory of statistical experiments* (Springer, New York, 1982).

[10] E. Hille and R. S. Phillips, "Functional analysis and semi-groups", AMS Colloquium Publications Vol. 31, (AMS, Providence, Rhode Island, 1957).

[11] K. Hinderer, *Foundations of non-stationary dynamic programming with discrete time parameter* (Springer, Berlin, 1970).

[12] K. Hinderer, *Grundbegriffe der wahrscheinlichkeitstheorie* (Springer, Berlin, 1985).

[13] L. Le Cam, *Asymptotic methods in statistical decision theory* (Springer, New York, 1986).

[14] J. Marschak, "Economics of information systems", *J. Amer. Statist. Assoc.* **66** (1971) 192–219.

[15] J. Marschak and K. Miyasawa, "Economic comparability of information systems", *Internat. Econom. Rev.* (1968) 137–174.

[16] M. D. Perlman, "Jensen's inequality for a convex vector-valued function on an infinite-dimensional space", *J. Multivariate Anal.* **4** (1974) 52–65.

[17] U. Rieder and H. Wagner, "Structured policies in the sequential design of experiments", *Ann. Oper. Res.* **32** (1991) 165–188.

[18] U. Rieder and R. Zagst, "Monotonicity and bounds for convex stochastic control models", *ZOR - Methods and Models of Operations Research* **39** (1994) .

[19] H. Strasser, *Mathematical theory of statistics* (De Gruyter, Berlin, 1985).

[20] I. Tonks, "Bayesian learning and the optimal investment decision of the firm", *The Economic Journal* **93** (1983) 87–98.

[21] R. Zagst, "Blackwell-Informativität in stochastischen Kontrollmodellen", Ph. D. Thesis, University of Ulm, 1991.