RUTCOR
RESEARCH
REPORT

On the Concavity of Multivariate Probability Distribution Functions

András Prékopa [a]

RRR 16-99, August, 1999

RUTCOR
Rutgers Center for
Operations Research
Rutgers University
640 Bartholomew Road
Piscataway, New Jersey
08854-8003
Telephone: 732-445-3804
Telefax: 732-445-5472
Email: rrr@rutcor.rutgers.edu
http://rutcor.rutgers.edu/∼rrr

[a]Rutcor, Rutgers Center for Operation Research, 640 Bartholomew Road, Piscataway, NJ 08854-8003

Rutcor Research Report
RRR 16-99, August, 1999

On the Concavity of Multivariate Probability Distribution Functions

András Prékopa

Abstract. We prove that the multivariate standard normal probability distribution function is concave for large argument values. The method of proof allows for the derivation of similar statements for other types of multivariate probability distribution function too. The result have important application, e.g., in probabilistic constrained stochastic programming problems.

Keywords: probability distribution, normal distribution, Dirichelet distribution, concavity, stochastic programming

1 Introduction

Let $\Phi(z;R)$ be the n-variate normal probability distribution function with correlation matrix R. We asuume that R is nonsingular, i.e., the distribution is non-degenerate.

It is well-known (see, e.g., Prékopa [2]) that $\Phi(z;R)$ is logarithmically concave (logconcave) in the entire space R^n.

Logconcave probability distributions have many important applications. In probabilistic constrained stochastic programming we use them to prove the convexity of a large class of problems. If we have, e.g., a constraint of the form

$$P(Tx \geq \xi) \geq p, \tag{1.1}$$

where ξ is a normally distributed random vector, then the logconcavity property of its distribution function implies that the set of x vectors satisfying (1.1) is a convex set.

There are, however, cases (see Prékopa [3]) where logconcavity is not enough and where the stronger concavity property of the distribution function is needed.

The purpose of this paper is to prove that if the components of z are large, then $\Phi(z;R)$ is not only logconcave but also concave. The method of proof carries over to other probability distribution functions too.

2 The Main Theorem

In this section our main objective is to prove Theorem 2.2. To do this we need another theorem which is interesting in itself, too.

Let $I_1,\ldots,I_n$ be finite or infinite intervals on the real line, $D = I_1 \times \cdots \times I_n$, and $F(z)$ an n-variate probability distribution function.

Definition. We Say that $F(z)$ is concave in D in the positive direction, if for any $z_1 \leq z_2$, z_1, $z_2 \in D$ and $0 \leq \lambda \leq 1$, we have the inequality $F(\lambda z_1 + (1-\lambda z_2)) \geq \lambda F(z_1) + (1-\lambda)F(z_2)$.

Theorem 2.1 *If $F(z)$ is concave in D in the positive direction, then $F(z)$ is concave in D.*

Proof. Let $z_1, z_2 \in D$. If $z_1 \leq z_2$ or $z_2 \leq z_1$, then by assumption the concavity of F between the two points holds true. Otherwise, some components of z_1 are smaller than or equal to the corresponding components of z_2 while for the others the opposite inequalities hold. We may assume that z_1, z_2 can be partitioned as

$$z_1 = \begin{pmatrix} x_1 \\ y_1 \end{pmatrix}, z_2 = \begin{pmatrix} x_2 \\ y_2 \end{pmatrix},$$

where $x_1 \leq x_2$, $y_1 \geq y_2$. Let us partition any $z \in R^n$ as $z = \begin{pmatrix} x \\ y \end{pmatrix}$ and use $F(x,y)$ as an alternative notation for $F(z)$. If ζ is a random vector that has distribution function $F(z)$, then we partation ζ accordingly, $\zeta = \begin{pmatrix} \xi \\ \eta \end{pmatrix}$. Then we have $F(z) = F(x,y) = P(\xi \leq x, \eta \leq y)$.

We have the following relation:

$$\begin{aligned} & F(x_2,y_1) + F(x_1,y_2) - F(x_1,y_1) - F(x_2,y_2) \\ = \;& P(\xi \leq x_2, \eta \leq y_1) + P(\xi \leq x_1, \eta \leq y_2) \\ & -P(\xi \leq x_1, \eta \leq y_1) - P(\xi \leq x_2, \eta \leq y_2) \\ = \;& P(\xi \leq x_2, (\eta \leq y_1) \cap (\overline{\eta \leq y_2})) \\ & -P(\xi \leq x_1, (\eta \leq y_1) \cap (\overline{\eta \leq y_2})) \\ = \;& P((\xi \leq x_2) \cap (\overline{\xi \leq x_1}) \cap (\eta \leq y_1) \cap (\overline{\eta \leq y_2})) \geq 0, \end{aligned} \tag{2.1}$$

where the upper bars indicate that we take the complementary of the events.

Relations (2.1) imply that $F(x_2, y_1) + F(x_1, y_2) \geq F(x_1, y_1) + F(x_2, y_2)$. Using this, we conclude

$$\begin{aligned} & F(\lambda x_1 + (1-\lambda)x_2, \lambda y_1 + (1-\lambda)y_2) \\ \geq \;& \lambda^2 F(x_1, y_1) + \lambda(1-\lambda)F(x_1, y_2) \\ & +\lambda(1-\lambda)F(x_2, y_1) + (1-\lambda)^2 F(x_2, y_2) \\ \geq \;& \lambda^2 F(x_1, y_1) + (1-\lambda)^2 F(x_2, y_2) \\ & +\lambda(1-\lambda)(F(x_1, y_1) + F(x_2, y_2)) \\ = \;& \lambda F(x_1, y_1) + (1-\lambda)F(x_2, y_2). \end{aligned}$$

The next theorem states our main result. (For the special case of $n = 2$, the theorem has already been proved in Prékopa [3]).

Theorem 2.2 *The function $\Phi(z_1, \cdots, z_n; R)$ is concave in the set $\{z | z_i \geq \sqrt{n-1}, i = 1, \cdots, n\}$.*

Proof. If ζ has distribution function $\Phi(z; R)$, then it can be represented as $\zeta = A\gamma$ where A satisfies $R = AA^T$ and γ has independent, standard normally distributed components. Let $A_1, \ldots, A_n$ designate the rows of A. Since R is a correlation matrix, it follows that $A_1, \ldots, A_n$ are unit length vectors.

We have the equality

$$\Phi(z; R) = P(A\gamma \leq z) = \int \cdots \int_{Av \leq z} \phi(v_1) \cdots \phi(v_n) dv_1 \cdots dv_n. \qquad (2.2)$$

The hyperplanes $\{v | A_i v = \sqrt{n-1}\}$, $i = 1, \cdots, n$ are tangent to the sphere $\{v | v^T v = n-1\}$. Thus, the set $\{v | Av \leq z\}$ contains the ball $\{v | v^T v \leq n-1\}$, provided that $z_i \geq \sqrt{n-1}, i = 1, \cdots, n$.

Let us introduce polar coordinates in the integral (2.2):

$$\begin{aligned} v_1 &= w \sin \psi_1 \\ v_2 &= w \cos \psi_1 \sin \psi_2 \\ &\vdots \\ v_{n-1} &= w \cos \psi_1 \cdots \cos \psi_{n-2} \sin \psi_{n-1} \\ v_n &= w \cos \psi_1 \cdots \cos \psi_{n-2} \cos \psi_{n-1} \end{aligned}$$

$-\frac{\pi}{2} < \psi_i \leq \frac{\pi}{2}$, $i = 1, \ldots, n-2$, $-\pi < \psi_{n-1} \leq \pi$.

The determinant of the Jacobian of the transformation is

$$J = w^{n-1} \cos^{n-2} \psi_1 \cos^{n-3} \psi_2 \cdots \cos \psi_{n-2} = w^{n-1} K(\psi)$$

where $\psi = (\psi_1, \cdots, \psi_{n-2})$. With this transformation the integrand in (2.2) takes the form

$$K(\psi) w^{n-1} e^{-\frac{w^2}{2}}. \qquad (2.3)$$

For fixed ψ, the function of the variable w in (2.3) is increasing if $0 \leq w \leq \sqrt{n-1}$ and is decreasing if $w \geq \sqrt{n-1}$.

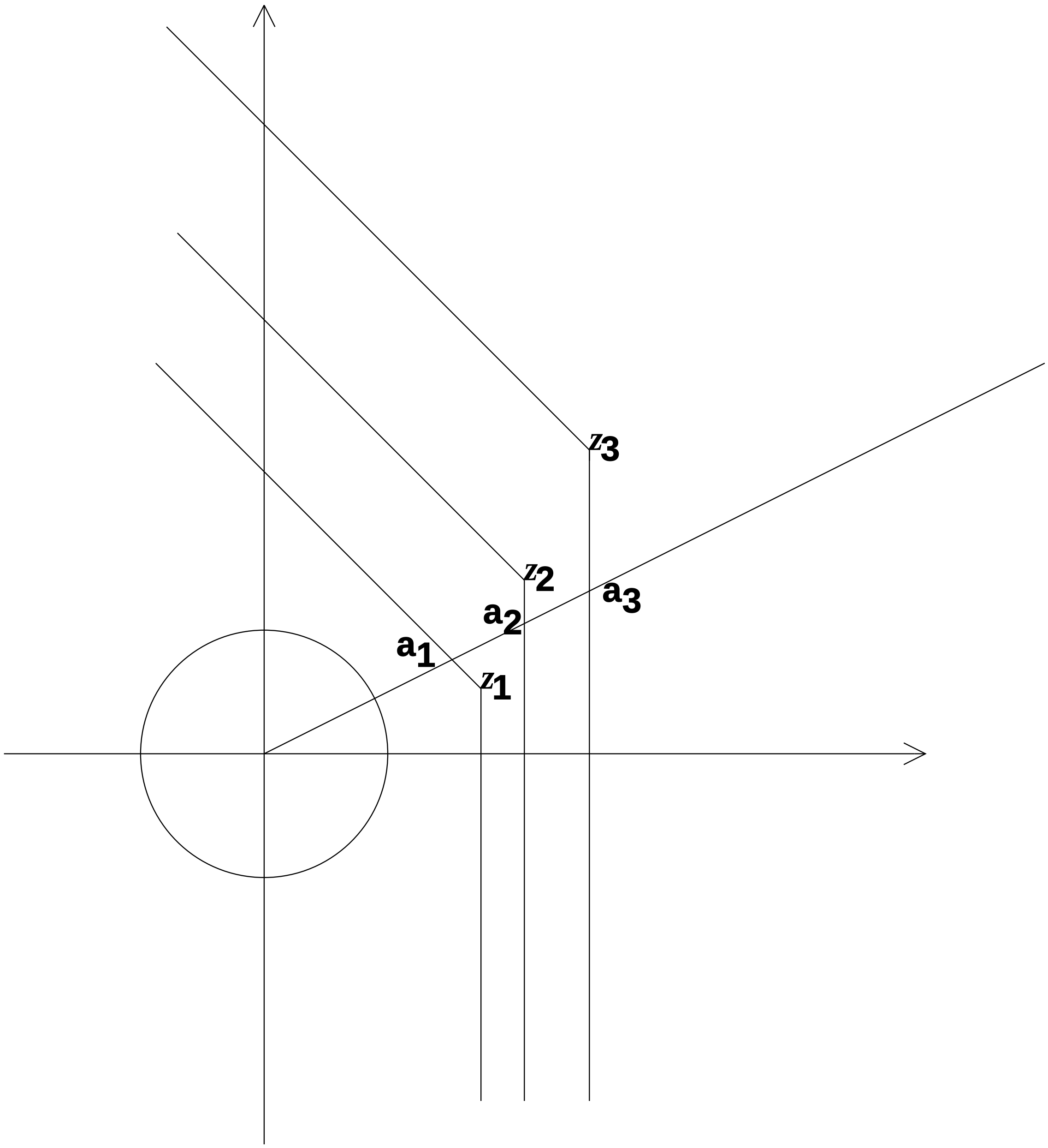
Let us introduce the notation

$$D(z) = \{v | Av \leq z\} \qquad (2.4)$$

Take two vectors z_1, z_2 and a λ such that $z_1 \leq z_2$, $0 < \lambda < 1$ and all components z_{1i} of z_1 satisfy $z_{1i} \geq \sqrt{n-1}$. Define z_3 as $z_3 = \lambda z_1 + (1-\lambda) z_2$. Take a unit length vector $c \in R^n$ and consider the ray

$$\{tc | t \geq 0\} \qquad (2.5)$$

Let $t(z)$ designate the length of the intersection of the boundary of the set (2.4) and the set (2.5). If there is one, then it is equal to the optimum value of the LP (for illustration see Figure 1):



The relation $z_2 = \lambda z_1 + (1-\lambda)z_2$ implies that $t(\lambda z_1 + (1-\lambda)z_2) \geq \lambda t(z_1) + (1-\lambda)t(z_2)$, where $t(z_i)$ is the distance between the origin and $a_i, i = 1, 2, 3$.

$$\begin{aligned} \text{Max}\, t \\ \text{subject to} \\ Atc \leq z \\ t \geq 0, \end{aligned} \tag{2.6}$$

where $z \geq 0$ is some fixed vector.

Since the optimum value of problem (2.6) is a concave function of the right hand side vector z, it follows that

$$t(z_3) \geq \lambda t(z_1) + (1-\lambda) t(z_2). \tag{2.7}$$

If we integrate the function (2.3) along the ray $\{tc|t \geq 0\}$ for fixed ψ, then by (2.7) we obtain

$$\begin{array}{l} K(\psi) \int_0^{t_3} w^{n-1} e^{-\frac{w^2}{2}} dw \\ \geq \lambda K(\psi) \int_0^{t_1} w^{n-1} e^{-\frac{w^2}{2}} dw \\ +(1-\lambda) K(\psi) \int_0^{t_2} w^{n-1} e^{-\frac{w^2}{2}} dw \end{array} \tag{2.8}$$

Integrating on both sides in (2.8) with respect to the angles in ψ, we get $\Phi(z_3; R) \geq \lambda \Phi(z; R) + (1-\lambda)\Phi(z_2; R)$. Thus $\Phi(z; R)$ is concave in $\{z|z_i \geq \sqrt{n-1}, i = 1, \ldots, n\}$.

In view of Theorem 2.1, the function $\Phi(z; R)$ is concave in the same set.

One can argue that the usefulness of the above result is restricted by the fact that the value of the distribution function becomes very large if the components of z are at least $\sqrt{n-1}$. This is certainly true for the univariate marginal distribution function $\Phi(z)$.

The n-variate standard normal probability distribution function, however, takes much smaller values if the components of z are around $\sqrt{n-1}$. For illustration we take the distribution function corresponding to independent components and define $a_n = \Phi^n(\sqrt{n-1})$. Then we have the figures: $a_1 = 0.50, a_2 = 0.71, a_3 = 0.78, a_4 = 0.84, a_5 = 0.89, a_6 = 0.93$. Thus, if the probability p in (1.1) is large enough but not irrealistically large, we have a chance that condition (1.1) implies that each component of Tx is at least $\sqrt{n-1}$ which, in turn, implies the concavity of $P(Tx \geq \xi)$ in the set $\{x|Tx \geq \sqrt{n-1}\}$. On the other hand, high dimensional probability distribution functions can be closely approximated by linear combinations of lower dimensional ones, at least in many cases (see [3] and the references there). Theorem 2.2 is very useful in some applications.

3 Miscellaneous Remarks

The lower bound (in Theorem 2.2) $\sqrt{n-1}$ for the components of z to ensure concavity of $\Phi(z; R)$ is not always the best one. If, e.g., $\rho = 0$, then the bivariate standard normal probability distribution function is concave for, $z_i \geq 0.51, i = 1, 2$. It is an open problem to determine that $\alpha = \alpha(R)$ for which $\Phi(z; R)$ is concave in the set $\{z|z_i \geq \alpha, i = 1, \ldots, n\}$ and then find the worst bound, i.e., the maximum of $\alpha(R)$ with respect to all correlation matrices R.

In the degenerate case where all entries of R are +1 or -1, $\Phi(z; R)$, $z \in R$, is concave in $\{z|z \geq 0\}$. The proof is very simple and is omitted.

There are other multivariate probability distribution functions too, for which the reasoning in Theorem 2.2 can be used with slight modification.

As an example, we look at the Dirichlet distribution with parameters $a_1 > 0, \ldots, a_n > 0, a_{n+1} > 0$, the probability density function of which is

$$\frac{\Gamma(a_1 + \cdots + a_{n+1})}{\Gamma(a_1) \cdots \Gamma(a_{n+1})} z_1^{a_1 - 1} \cdots z_n^{a_n - 1} (1 - z_1 - \cdots - z_n)^{a_{n+1} - 1} \tag{3.2}$$

for $z_i > 0, i = 1, \ldots, n, z_1 + \cdots + z_n \leq 1$.

Suppose that (3.2) is the probability distribution of $\xi = (\xi_1, \cdots, \xi_n)^T$ and let $\eta = (\eta_1, \cdots, \eta_n)^T$ designate the intersection of the ray $\{\lambda\xi|\lambda \geq 0\}$ and the simplex $\{z|z \geq 0, z_1 + \cdots z_n \leq 1\}$. Given $\eta = u$, the random vector ξ has the distribution given by the density function:

$$\frac{\Gamma(a_1 + \cdots + a_{n+1})}{\Gamma(a_1 + \cdots + a_n)\Gamma(a_{n+1})} \frac{1}{d} \left(\frac{x}{d}\right)^{a_1 + \cdots + a_n - 1} \left(1 - \frac{x}{d}\right)^{a_{n+1} - 1} \tag{3.3}$$

$0 < x < d$, on the line segment between the origin and u, where $d = \sqrt{u_1^2 + \cdots + u_n^2}$ (see, e.g., Wilks [4]).

The function (3.3) is decreasing in $(0, d)$, if (assume $a_1 + \cdots + a_n \neq 2$)

$$\alpha(d) = \frac{(a_1 + \cdots + a_n - 1)d}{a_1 + \cdots + a_{n+1} - 2} \leq 1. \tag{3.4}$$

If this value is positive, the function (3.3) is increasing for $z \leq \alpha(d)$ and is decreasing for $z \geq \alpha(d)$. The largest $\alpha(d)$ corresponds to $d = 1$. Let $\alpha = \alpha(1)$. If we apply the same reasoning what we have applied in Theorem 2.2, then we obtain:

Theorem 3.1 *The probability distribution function $F(z)$ of the Dirichlet distribution, given by the probability density function (3.2), is concave in $\{z|z \geq 0\}$, if $\alpha \leq 0$ and in $\{z|z_i \geq \alpha, i = 1, \ldots, n\}$, if $\alpha > 0$.*

References

[1] Prékopa, A. (1970). On Probabilistic Constrained Programming. *Proceedings of the Princeton Symposium on Mathematical Programming*, Princeton University Press, Princeton NJ, 113-138.

[2] Prékopa, A. (1995). *Stochastic Programming*, Kluwer Academic Publishers, Dordrecht

[3] Prékopa, A. (1999). The Use of Discrete Moment Problems in Probabilistic Constrained Stochastic Programming Models, *Annals of Operations Research* **85** 21-38.

[4] Wilks, S. S. (1962). *Mathematical Statistics*, Wiley, New York