

NPS55-83-037

# NAVAL POSTGRADUATE SCHOOL

## Monterey, California



AN AUTOREGRESSIVE PROCESS FOR BETA

RANDOM VARIABLES

by

Ed. McKenzie

December 1983

Approved for public release; distribution unlimited

Prepared for:

FedDocs  
D 208.14/2  
NPS-55-83-037

Naval Postgraduate School  
Monterey, California 93943

NAVAL POSTGRADUATE SCHOOL  
Monterey, California

Commodore R. H. Shumaker  
Superintendent

David A. Schrady  
Provost

This work was supported by the Naval Postgraduate School Foundation  
Research Program under contract with the National Research Council.

Reproduction of all or part of this report is authorized.

## REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS  
BEFORE COMPLETING FORM

1. REPORT NUMBER NPS55-83-037		2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) AN AUTOREGRESSIVE PROCESS FOR BETA RANDOM VARIABLES		5. TYPE OF REPORT & PERIOD COVERED Technical Report	
7. AUTHOR(s) Ed. McKenzie		6. PERFORMING ORG. REPORT NUMBER	
9. PERFORMING ORGANIZATION NAME AND ADDRESS Naval Postgraduate School Monterey, CA 93943		8. CONTRACT OR GRANT NUMBER(s)	
11. CONTROLLING OFFICE NAME AND ADDRESS		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 61152N; RR000-01-100 N0001483WR30104	
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		12. REPORT DATE December 1983	
		13. NUMBER OF PAGES 25	
		15. SECURITY CLASS. (of this report) Unclassified	
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE	
16. DISTRIBUTION STATEMENT (of this Report) Approved for public release; distribution unlimited.			
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)			
18. SUPPLEMENTARY NOTES			
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) Beta Autoregression simulation bivariate Beta distributions bivariate Uniform distribution Dirichelet distribution			
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) Two stationary first-order autoregressive processes with Beta marginal distributions are presented. They are both linear, additive processes but the coefficients are Beta random variables. Their autocorrelation functions are investigated: One is positive and the other alternates in sign. The usefulness of the models in simulation is discussed. The Bivariate Beta distributions of two consecutive observations are considered in some detail. Several examples are given, including a Bivariate Uniform process which is also			

EXAMINED IN DETAIL. The relationship of these Bivariate Beta distributions to the Dirichelet distribution is discussed.

An Autoregressive Process for Beta Random Variables

Ed. McKenzie  
Department of Mathematics  
University of Strathclyde  
Glasgow  
Scotland

and

Department of Operations Research  
Naval Postgraduate School  
Monterey, California  
U.S.A.

December 1983



## ABSTRACT

Two stationary first-order autoregressive processes with Beta marginal distributions are presented. They are both linear, additive processes but the coefficients are Beta random variables. Their autocorrelation functions are investigated: one is positive and the other alternates in sign. The usefulness of the models in simulation is discussed. The Bivariate Beta distributions of two consecutive observations are considered in some detail. Several examples are given, including a Bivariate Uniform process which is also examined in detail. The relationship of these Bivariate Beta distributions to the Dirichelet distribution is discussed.





## 1. INTRODUCTION

The Beta distribution is the most versatile and useful distribution available on a bounded interval. Despite this there are very few practical models for describing correlation between pairs of Beta random variables or serial correlation in sequences of them. This is unfortunate since there is a natural interest in modelling sequences of dependent Beta variates. These arise in a variety of ways but are often associated with the stochastic behaviour of a proportion or probability over time. One example is the study of market share, e.g. the proportion of the market held by a particular product. (See, for example, Wichern and Jones (1977)).

Some efforts have been made to model such behaviour. Azzalini (1982) developed a simple Markov model for use in a quality control context where the proportions defective in adjacent batches sampled may be expected to be dependent. He used the Product Autoregressive process discussed by McKenzie (1983). This model has limited usefulness for the Beta distribution, however, since it requires that of the two parameters of that distribution one must be integral and the other rational. Another approach has been suggested by Souza and Harrison (1981). They attribute to the proportion a Beta distribution which is revised with each observation by a procedure based on Bayesian and Information Theoretic concepts. This approach appears to offer some promise for forecasting but is fairly complex in general. A more traditional time-series approach to such a data set would be to treat the proportions  $X_t$  not as Beta's but as Logistic-normal variates. Thus, we would transform  $X_t$  to log-odds, i.e.  $Y_t = \ln[X_t/(1-X_t)]$ , and attempt to model  $\{Y_t\}$  as an autoregressive moving-average process. This approach is suggested by the work of Aitchison and Shen (1980) and Aitchison (1982), but does not appear to have been investigated as a time-series procedure yet. It discards the Beta marginal distribution which is our main interest here. Further, it appears to be extremely difficult to

reverse the procedure and generate a sequence of proportions with a specific correlation structure.

The purpose of this work is to present a discrete time Markov process with a Beta marginal distribution. It is constructed in the spirit of the recent work on modelling of non-Gaussian time series illustrated, for example, by Lawrance and Lewis (1980, 1982) and Jacobs and Lewis (1983). Since simulation is also a major motivation for such work, we seek models which are simple and flexible and whose parameters are few and physically meaningful. The aspect of simulation is important here for, as Schmeiser and Lal (1980) noted in a recent survey (1980), there are few practical ways of generating dependent Beta random variables. This is because the usual multivariate Beta distributions are closely related to the Dirichlet distribution and so constrain their vector variates in a way which is undesirable in general.

We present here two simple discrete-time stochastic processes whose marginal distributions are Beta random variables. The processes are linear, additive autoregressions with random coefficients. The coefficients themselves are also Beta variates. There is a single free parameter which corresponds in a simple way to the correlation in each model and a wide range of correlation is possible. Because of their simplicity, the models provide a powerful way of generating sequences of dependent Beta variates using only independent Betas. As noted above, there is a scarcity of practical multivariate Beta distributions. Thus, the bivariate distributions associated with the processes are discussed in some detail. They exhibit a number of interesting features. We also examine in detail the particular case of the bivariate Uniform distribution. It plays an important role in the simulation of dependent pairs of random variables.

Before describing the models we may note that there is a well-known continuous-time Markov process with Beta marginals. It arises in genetics and is one of the forms of the Wright-Fisher gene frequency models described in detail in the books by Karlin and Taylor (1975, 1981). It is a diffusion process and, in different forms, has found applications in sociology, psychology and marketing. It is also derived by Massey et al. (1970) as a stochastic response model. They obtain it as the limiting form of the "contagious binomial" distribution developed by Coleman (1964) to model voting behaviour. Of course, such processes are in continuous time and it is by no means clear how they can be restructured in discrete time. Nor is it clear whether the processes presented here represent some discrete time formulations of the diffusion processes.

## 2. THE MODELS

2.1. A random variable (r.v.)  $X$  is said to have a Beta distribution with parameters  $(\alpha, \beta)$  if it has probability density function (p.d.f.)

$$f(x) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha, \beta)}, \quad 0 < x < 1; \quad \alpha, \beta > 0.$$

For convenience in what follows, we shall write such a random variable as  $Be(\alpha, \beta)$ . We note for later use that for such  $X$ ,  $E(X) = \alpha/(\alpha+\beta)$ ,  $Var(X) = \alpha\beta/[(\alpha+\beta)^2(\alpha+\beta+1)]$  and the third moment about the mean is given by  $m_3 = 2\alpha\beta(\beta-\alpha)/[(\alpha+\beta)^3(\alpha+\beta+1)(\alpha+\beta+2)]$ .

The models presented here use the following results:

$$1 - Be(\alpha, \beta) = Be(\beta, \alpha) \tag{1}$$

$$Be(\alpha, \beta) \cdot Be(\alpha+\beta, \gamma) = Be(\alpha, \beta+\gamma) \tag{2}$$

The first of these two results is well known and easily demonstrated. The second result states that the product of two independent Beta r.v.s. with parameters as specified is itself a Beta r.v. The result may be verified by considering the Mellin Transform (Widder, 1946), i.e.  $E(X^S)$ , of the product on the left-hand side of (2). It is

$$E(X^S) = \frac{B(\alpha+s, \beta)}{B(\alpha, \beta)} \cdot \frac{B(\alpha+\beta+s, \gamma)}{B(\alpha+\beta, \gamma)} = \frac{\Gamma(\alpha+s)\Gamma(\alpha+\beta+\gamma)}{\Gamma(\alpha)\Gamma(\alpha+\beta+\gamma+s)} = \frac{B(\alpha+s, \beta+\gamma)}{B(\alpha, \beta+\gamma)}$$

which is the transform of the right-hand side of (2). We shall refer to the application of (2) to change one Beta r.v. into another as the Beta-Beta transformation.

Two distinct models are presented here. One is for positively correlated pairs of Beta r.v.s. and is denoted by PBAR and the other is for negatively correlated pairs of Beta's and is denoted by NBAR. Both models are linear and additive and have random coefficients.

2.2. The PBAR model is given by

$$X_t = 1 - U_t(1 - W_t X_{t-1}) \quad (3)$$

where  $\{U_t\}$  and  $\{W_t\}$  are independent sequences of independent identically distributed (i.i.d.) r.v.s., independent of previous  $X$ 's, and  $U_t$  is  $Be(\beta, \alpha-p)$  and  $W_t$  is  $Be(p, \alpha-p)$ , ( $0 < p < \alpha$ ). Now if  $X_{t-1}$  is  $Be(\alpha, \beta)$  then  $W_t X_{t-1}$  is  $Be(p, \alpha+\beta-p)$  from (2) and  $1 - W_t X_{t-1}$  is  $Be(\alpha+\beta-p, p)$  from (1). Further use of (2) shows that  $U_t(1 - W_t X_{t-1})$  is  $Be(\beta, \alpha)$  and so  $X_t$  given by (3) is  $Be(\alpha, \beta)$  by (1). Thus, equation (3) defines a stationary process  $\{X_t\}$  with a  $Be(\alpha, \beta)$  marginal distribution.

Further, there is a single free parameter in this scheme, viz.  $p$ . As we shall see, the value of  $p$  determines the correlation structure of the

process  $\{X_t\}$ . From the structure of (3), the process is a first-order linear autoregression with random coefficients. Direct calculation yields the autocorrelation function of the process as  $\rho_X(k) = \rho^k$ ,  $k = 0, 1, \dots$ , where

$$\rho = E(U)E(W) = p\beta/\alpha(\alpha+\beta-p). \quad (4)$$

Now  $\rho$  as defined by (4) is a monotonic increasing function of  $p$  for fixed  $(\alpha, \beta)$ . Further, since  $\beta > 0$ ,  $0 < p < \alpha$ , we may deduce that  $0 < \rho < 1$ . Thus, the entire range of positive correlation is possible for any values of  $\alpha$  and  $\beta$ , i.e. any Beta marginal distribution.

There are two limiting cases for the parameter  $p$  we may consider. When  $p$  is zero  $W_t = 0$  with probability one and  $U_t$  is  $Be(\beta, \alpha)$ . Thus,  $X_t$  is independent of  $X_{t-1}$  and  $\rho = 0$ . When  $p = \alpha$  both  $U_t$  and  $W_t$  are unity with probability one. Thus,  $X_t = X_{t-1}$  and  $\rho = 1$ . The process is not ergodic in this case.

It is important to notice that with this model (3)  $\rho = 0$  implies independence.

2.3. The model for the NBAR process is given by

$$X_t = V_t(1 - W_t X_{t-1}). \quad (5)$$

As before,  $\{V_t\}$  and  $\{W_t\}$  are independent sequences of i.i.d., r.v.'s independent of  $X_{t-1}$  and  $V_t$  is  $Be(\alpha, \beta-p)$  and  $W_t$  is  $Be(p, \alpha-p)$ . Again, it is easily verified using (1) and (2) that equation (5) will generate a stationary process whose marginal distribution is  $Beta(\alpha, \beta)$ . The NBAR process is also a linear autoregression with random coefficients. It too has autocorrelation function of the form  $\rho^k$ ,  $k = 0, 1, \dots$ , but now

$$\rho = -p/(\alpha+\beta-p). \quad (6)$$

Notice that the specification of the distributions of  $V_t$  and  $W_t$  requires that  $0 < \rho < \alpha$  and  $0 < \rho < \beta$ . From (6),  $\rho$  is a monotonic decreasing function of  $p$  for fixed  $(\alpha, \beta)$ , and so we may deduce that for the NBAR process(5)  $-\max(\alpha/\beta, \beta/\alpha) < \rho < 0$ .

The upper extreme of zero is again attainable when  $p = 0$  and  $X_t = V_t$ , which is independent of  $X_{t-1}$ . As with the PBAR process, it is important to note that  $\rho = 0$  implies independence. The lower limit  $-\max(\alpha/\beta, \beta/\alpha)$  is also attainable. If  $\beta < \alpha$  it is attained when  $p = \beta$  and so  $V_t = 1$ . If  $\beta > \alpha$  it occurs where  $p = \alpha$  which corresponds to  $W_t = 1$ . When  $\alpha = \beta$  the lower limit is  $-1$  which corresponds to the usual antithetic relationship,  $X_t = 1 - X_{t-1}$ , given by  $V_t = W_t = 1$ .

2.4. The models PBAR and NBAR given by (3) and (5) yield random coefficient autoregressions of order one. Further, the first-order correlation  $\rho$  satisfies

$$-\max\left(\frac{\alpha}{\beta}, \frac{\beta}{\alpha}\right) \leq \rho \leq 1. \quad (7)$$

This is not the greatest possible range for general  $(\alpha, \beta)$ . For example, if  $\alpha = 2, \beta = 1$  we find that  $-0.5 < \rho < 1$  for these models. On the other hand, it may be deduced from Moran (1967) that the correlation between two  $Be(2,1)$  r.v.s. is bounded below by  $(9\pi - 32)/4$ , i.e. approximately  $-0.9314$ . How far the lower bound given by (7) is from the minimum correlation possible is not known for general  $(\alpha, \beta)$ .

We may note, however, that in the symmetric case, i.e.  $\alpha = \beta$ , the range is  $(-1, 1)$  as would be hoped. As before, the two extremes may be attained: the upper limit of 1 from  $X_t = X_{t-1}$  and the lower limit of  $-1$  from the usual antithetic relationship  $X_t = 1 - X_{t-1}$ . This latter is obtained from NBAR with  $p = \alpha = \beta$  so that  $V_t = Z_t = 1$ . In particular, it is now

possible to generate r.v.s. uniform on  $(0,1)$  with any first-order autoregressive correlation, i.e. any  $\rho \in [-1,1]$ . We return to this point later.

### 3. BIVARIATE DISTRIBUTIONS OF $(X_t, X_{t-1})$

3.1. For convenience, we rewrite the PBAR model (3) as  $Y = 1 - U(1-WX)$ .

Initially, consider the joint p.d.f. of  $(Y,X)$  conditional on  $U$ . It can be written in the form

$$f_{Y,X|U}(y,x|u) = \frac{1}{xu} \cdot f_X(x) \cdot f_W\left(\frac{y+u-1}{xu}\right), \quad 1-y < u < \frac{1-y}{1-x}.$$

The joint p.d.f. of  $Y, X$  and  $U$  may now be derived and from it the p.d.f. of  $(Y,X)$  is obtained in the form

$$f_{Y,X}^+(y,x) = \frac{(1-x)^{\beta-1}}{C_+} \int_0^{m(y,x)} s^{p-1} [x(1-y) - (1-x)s]^{\alpha-p-1} (1-y+s)^{\beta-\alpha} (y-s)^{\alpha-p-1} ds \quad (8)$$

where  $m(y,x) = \min(y, x(1-y)/(1-x))$  and  $C_+ = B(\alpha, \beta)B(p, \alpha-p)B(\beta, \alpha-p)$ .

A change of variable  $t = (1-x)s/(1-y)$  in the integral in (8) yields the result  $f_{Y,X}^+(y,x) = f_{Y,X}^+(x,y)$ . Thus, the bivariate p.d.f., which is defined on the unit square, is symmetric about  $y = x$ . Note also that  $m(y,x) = y$  if  $y < x$ . Thus, the two forms of the integral exist on either side of  $y = x$ , and we can define

$$f_{Y,X}^+(y,x) = \begin{cases} g(y,x) & y < x \\ g(x,y) & y > x \end{cases} \quad (9)$$

where  $g(y,x)$  is given by the right hand side of (8) with  $m(y,x)$  replaced by  $y$ .

This symmetry is also important from the viewpoint of modelling or identification of the PBAR process. The importance arises from the idea of time-reversibility of a stationary process. The concept is discussed in detail by Weiss (1975). A discrete-time stationary process  $\{X_t\}$  is time-reversible if the joint distributions of  $\{X_1, X_2, \dots, X_t\}$  and  $\{X_t, X_{t-1}, \dots, X_1\}$  are identical for every  $t$ . In the case of a first-order autoregression, as here, the joint p.d.f. can be expressed as

$$f(x_1, x_2, \dots, x_t) = \prod_{i=1}^{t-1} f(x_i, x_{i+1}) / \prod_{i=1}^{t-1} f(x_i)$$

Such a process is time reversible then whenever the bivariate distribution is symmetric. Thus, the PBAR process is time-reversible.

Indeed, the time reversed process is given by

$$X_t = 1 - U'_t(1 - W'_t X_{t+1})$$

where  $\{U'_t\}$  and  $\{W'_t\}$  have exactly the same properties as  $\{U_t\}$  and  $\{W_t\}$  respectively defined by (3).

3.2. For the NBAR process, writing  $Y = V(1-WX)$  and proceeding as for the PBAR process yields

$$f_{Y,X}^-(y,x) = \frac{(1-x)^{\beta-1}}{C_-} \int_0^{m(1-y,x)} s^{p-1} [xy - (1-x)s]^{\alpha-p-1} (1-y-s)^{\beta-p-1} ds \quad (10)$$

where  $C_- = B(\alpha, \beta)B(p, \alpha-p)B(\alpha, \beta-p)$ .

Again, the change of variable  $t = (1-x)s/(1-y)$  shows that the bivariate p.d.f. is symmetric about  $y = x$  and the NBAR process is time-reversible. The structure of the density, however, is a little more complex than for the PBAR process. The upper limit of the integral in (10) is  $m(1-y, x) = 1 - y$



if  $x + y > 1$ . Thus, the form of  $f^-$  depends upon which side of the line  $x + y = 1$  we are on.

Define

$$f_{Y,X}^-(y,x) = \begin{cases} g_1(y,x) & x + y < 1, \\ g_2(y,x) & x + y > 1. \end{cases} \quad (11)$$

From symmetry about  $y = x$ , we know that  $g_1(y,x) = g_1(x,y)$  and  $g_2(y,x) = g_2(x,y)$ . Of more immediate interest is the relationship between  $g_1$  and  $g_2$ . A further change of variable in the integral in (10) yields the following result. Using an obvious notation to denote dependence upon the parameters of the marginal distribution:

$$g_1(1-y, 1-x; \alpha, \beta) = g_2(y, x; \beta, \alpha) \quad (12)$$

Using (12) to evaluate  $f^-$  in both the triangular regions induced by the line  $x + y = 1$  yields

$$f_{Y,X}^-(1-y, 1-x; \alpha, \beta) = f_{Y,X}^-(y, x; \beta, \alpha) \quad (13)$$

This result (13) is an obvious two-dimensional analogue of the relationship specified by equation (1), viz.  $f_X(1-x; \alpha, \beta) = f_X(x; \beta, \alpha)$ .

Further, by the symmetry about  $y = x$ , equation (12) yields

$$g_1(1-x, 1-y; \alpha, \beta) = g_2(y, x; \beta, \alpha) \quad (14)$$

which specifies the nature of the relationship between  $g_1$  and  $g_2$  across the line  $x + y = 1$ . In particular, note that when  $\alpha = \beta$  the bivariate p.d.f. is symmetric about both  $y = x$  and  $x + y = 1$ .

## 4. MOMENTS

Although it is difficult to obtain the p.d.f.s in an explicit form, much information can be derived directly from the structural relationships, (3) and (5). This is also true of the conditional distributions. In both cases, the conditional expectation is linear. For the PBAR process, using (3) yields

$$E(X_{t+1}|X_t = x) = [(\alpha + \beta - \rho - \alpha\beta) + \rho\beta x] / \alpha(\alpha + \beta - \rho), \quad (15)$$

and for the NBAR process, (5) yields

$$E(X_{t+1}|X_t = x) = (\alpha - \rho x) / (\alpha + \beta - \rho). \quad (15a)$$

Further, the time-reversibility of the processes ensures that the inverse regressions are identical to (15) and (15a), i.e.  $E(X_t|X_{t+1} = x) = E(X_{t+1}|X_t = x)$ . In both cases the conditional variances are quadratic and of the form  $\text{var}(X_{t+1}|X_t = x) = \sigma^2(1 - mx)^2 + (\sigma^2 + m^2)\sigma_w^2 x^2$  where  $\sigma_w^2 = \text{var}(W)$  and  $m$  and  $\sigma^2$  are the mean and variance of  $U$  for PBAR and of  $V$  for NBAR.

Higher order moments are important in the identification of non-standard time-series models and we note here two of particular interest. The first is  $C_{21}(k) = \text{Cov}(\tilde{X}_t^2, \tilde{X}_{t-k})$  where  $\tilde{X}_t = (X_t - \mu)$ . For time reversible processes  $C_{21}(k) = C_{21}(-k)$ . For both the PBAR and NBAR processes,  $C_{21}(k) = m_3 \rho^k$ ,  $k = 0, 1, 2, \dots$ , where  $m_3$  is the third moment of  $X_t$  about its mean,  $\mu$  say, and these are given in Section 2.1.

Another moment of particular value in residual analysis is based on the residuals from minimum mean square error prediction, i.e.

$R_t = (X_t - \mu) - \rho(X_{t-1} - \mu)$ . Such residuals are uncorrelated for PBAR and NBAR processes. Of more interest is the behaviour of  $\text{Cov}(R_t^2, R_{t-k})$  which is useful in distinguishing between constant and random coefficient models. For both the PBAR and NBAR processes

$$\begin{aligned} \text{Cov}(R_t^2, R_{t-k}) &= 0 & m_3(1-\rho)(1-\rho^2)\rho^k, & \quad k = 1, 2, \dots \\ & & m_3(1+2\rho)(1-\rho)^2, & \quad k = 0. \end{aligned}$$

Further details and discussion of the usefulness of these moments may be found in Lewis and Lawrance (1983b).

## 5. EXAMPLES

To illustrate the nature of the bivariate p.d.f. the functions  $g$  from (9) and  $g_1$  and  $g_2$  from (11) are evaluated explicitly for a few specific cases below.

### PBAR

P1:  $\alpha = \beta = 2; \rho = 1$  i.e.  $\rho = 1/3$ ,

$$g(y, x) = 12y(1-x)$$

P2:  $\alpha = 3, \beta = 2; \rho = 1$  i.e.  $\rho = 1/6$ ,

$$g(y, x) = 72(1-x)[(1-x)y(2-y) - 2(1-y)\ln(1-y) - 2y(2-x-y)]$$

P3:  $\alpha = 4, \beta = 2; \rho = 3$  i.e.  $\rho = 1/2$ ,

$$g(y, x) = 120(1-x)[1 + 2(1-y)\ln(1-y) - (1-y)^2]$$

### NBAR

N1:  $\alpha = \beta = 2; \rho = 1$  i.e.  $\rho = -1/3$ ,

$$g_1(y, x) = 12xy$$

$$g_2(y, x) = 12(1-x)(1-y)$$

N2:  $\alpha = 3, \beta = 2; \rho = 1$  i.e.  $\rho = -1/4$

$$g_1(y, x) = 36x^2y^2$$

$$g_2(y, x) = 36(1-x)(1-y)(x+y + xy-1)$$

N3:  $\alpha = 5, \beta = 13/3; \rho = 4$  i.e.  $\rho = -3/4$

$$g_2(y, x) = A(1-x)^{10/3}(1-y)^{10/3}$$

$$\begin{aligned} g_1(y, x) &= g_2(y, x) - B(1-x-y)^{1/3} \left\{ 3(1-y)^3(1-x)^3 - \frac{9}{4}(1-y)^2(1-x)^2(1-x-y) \right. \\ &\quad \left. + \frac{9}{7}(1-y)(1-x)(1-x-y)^2 - \frac{3}{10}(1-x-y)^3 \right\} \end{aligned}$$

where  $A = 3532100/2187 = 1615.04$ , and  $B = 140A/283 = 798.96$ .

Plots of the contours and the surface of  $P1$  are displayed in Figures 1(a) and (b). If they are rotated through  $90^\circ$  about the point  $(0.5, 0.5)$  the corresponding plots are obtained for  $N1$ . Contour plots are displayed for  $P2$  and  $N2$  in Figures 2 and 3.

### Figures 1,2,3

The most remarkable feature of these distributions is the appearance of a ridge in the surface of  $P1$  (and so  $N1$ ) but not in  $P2$  or  $N2$ . The ridge is due to the fact that both densities have two forms as described by equations (9) and (11). Thus the ridge, if it occurs, corresponds to the line  $y = x$  for PBAR p.d.f.s and  $x + y = 1$  for NBAR's. From the definition, the p.d.f.s will be continuous but their derivatives need not be so. Any points of discontinuity occur on these two lines. To determine the general conditions for occurrence of a ridge we can examine the behaviour of the derivatives of the p.d.f.s on both sides of the two lines. This procedure yields the following results. For the PBAR density (8) there is no ridge provided  $\rho < (\alpha-1)$ . Using (4), we may deduce that no ridge occurs provided  $\rho$  satisfies

$$0 \leq \rho < \beta(\alpha-1)/\alpha(\beta+1) . \quad (16)$$

For the NBAR density (10) there is no ridge provided  $\rho < (\alpha+\beta-2)/2$  i.e. provided

$$0 \leq \rho > [2 - (\alpha+\beta)]/(2+\alpha+\beta) . \quad (17)$$

These conditions (16) and (17) are violated by  $P3$  and  $N3$  respectively and so both densities have a ridge. These are illustrated in the surface plots of  $P3$  and  $N3$  displayed in Figures 4 and 5, respectively.

### Figures 4,5

## 6. THE UNIFORM PROCESS

A bivariate distribution of particular interest is that with Uniform marginals. Apart from the natural interest in modelling data from such a distribution it has an important application in simulation. By using the inverse distribution function transformation of the Uniforms bivariate distributions with any other marginals can be obtained. This is an important approach to the generation of pairs of dependent random variables.

Some recent development of Bivariate Uniform distributions appears in the papers by Barnett (1980) and Lewis and Lawrance (1983a). The former exhibits several different Bivariate Uniform distributions but they are generally complex, have limited correlation, and require distribution function transformations to obtain the Uniforms. The latter work gives several procedures for generating a pair of dependent Uniforms from a random coefficient regression on independent Uniforms. The procedures are simple and the entire range of correlation can be attained. However, to achieve this breadth the correlations are usually complex functions of two parameters.

The Bivariate Uniform with p.d.f.s given by (8) and (10) is particularly useful in this kind of application. The entire range of correlation  $[-1, 1]$  is available to the Uniforms and so the entire possible range will be available to the transformed variates. Further, the correlation is a simple function of a single parameter, i.e.  $\rho = \pm p/(2-p)$ , ( $0 \leq p \leq 1$ ) and so any desired correlation is easily achieved. Finally, the generation of the desired pair is straightforward involving only the additional generation of two independent Beta r.v.s.

Since the Bivariate Uniform corresponds to (8) and (10) with  $\alpha = \beta = 1$  considerable simplification is possible. In this case,  $f_{Y,X}^-(y,x) = f_{Y,X}^+(1-y,x)$  and so only  $f^+$  is considered here. Making a suitable change of variable in the integral yields a more useful expression for the density given by (9), viz.

$$g(y,x) = \frac{2(1-p)}{B(p,1-p)x^p(1-y)^p} \int_0^{\pi/2} (\tan\theta)^{2p-1} \left[1 - \frac{y(1-x)}{x(1-y)} \sin^2\theta\right]^{-p} d\theta, (y < x) .$$

The density has a line singularity on  $y = x$  as can be seen from the contour and surface plots shown in Figure 6 for the case  $p = 1/2$  .

Figure 6

## 7. RELATIONSHIP WITH THE DIRICHELET DISTRIBUTION

The best-known "Bivariate Beta" distribution is the Dirichelet distribution. It is described in detail by Johnson and Kotz (1972). It is defined not on the Unit square as are the densities given by (8) and (10) but on the Unit simplex, i.e.  $\{(x,y): 0 < x, y < 1, x + y < 1\}$  . As such it plays a natural role as the joint distribution of two proportions from a single population. It seems to be generally regarded as a Bivariate Beta because both the marginal and conditional distributions are Beta. The joint p.d.f. in the case of identical  $Be(\alpha,\beta)$  marginals is

$$f(y,x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\alpha)\Gamma(\beta-\alpha)} \cdot x^{\alpha-1}y^{\alpha-1}(1-y-x)^{\beta-\alpha-1} ,$$

and the correlation between  $X$  and  $Y$  is  $\rho = -\alpha/\beta$  ( $\alpha < \beta$ ) . Note that  $\rho$  here depends explicitly upon the parameters of marginal distribution and so is fixed for any particular marginal distribution. Further,  $-\alpha/\beta$  is the minimum correlation attainable from the NBAR process. As noted, it is attained when  $p = \alpha$  and  $Y = V(1-X)$  where  $V$  is  $Be(\alpha,\beta-\alpha)$  and independent of  $X$  . Thus,  $Y/(1-X)$  is independent of  $X$  , and symmetry ensures that  $X/(1-Y)$  is independent of  $Y$  . This characterizes  $(X,Y)$  as having a Dirichelet distribution by a result of Darroch and Ratcliff (1971). Thus, we may view the Dirichelet distribution as a limiting form of the NBAR process distribution, as  $p \rightarrow \min(\alpha,\beta)$  .

## 8. EXTENSIONS

It is possible to extend the time-series models (3) and (5) to higher orders of dependence. However, two simple and more immediate extensions lie closer to the area of simulation and are noted here. The fact that the PBAR and NBAR models yield simple but powerful methods of generating sequences of correlated Beta r.v.s. has been emphasized. Such sequences are stationary so that each Beta r.v. has the same distribution. An obvious extension is to the generation of pairs of dependent Beta variates with different distribution. This may be achieved in a variety of ways using (1) and (2) and (3) or (5) if we wish.

A second simple generalization is to bounded intervals other than (0,1). Since an alternative sample space is achieved by a linear transformation the procedure is straightforward and all correlations are unaffected. Thus, suppose we wish to develop a PBAR process for  $Be(\alpha, \beta)$  r.v.s. defined on (a,b) rather than (0,1). By considering the  $X_t$ 's transformed to  $Y_t$ 's on (0,1) and  $\{Y_t\}$  satisfying (3) we find that the PBAR for  $\{X_t\}$  is given by

$$X_t = b - U_t[b-a-W_t(X_t-a)] .$$

## ACKNOWLEDGEMENT

I would like to thank Professor P. A. W. Lewis and Dr. A. J. Lawrance for letting me see some of their papers before publication and for the help and encouragement afforded me in many discussions. I also gratefully acknowledge the support of a National Research Council Associateship at the Naval Postgraduate School in Monterey, California, where this work was completed.

The graphs were produced by an experimental APL package GRAFSTAT which the Naval Postgraduate School is using under a test agreement with IBM Watson Research Center, Yorktown Heights, N.Y. I am grateful to Dr. P. D. Welch and Dr. P. Heidelberger for making GRAFSTAT available to me.

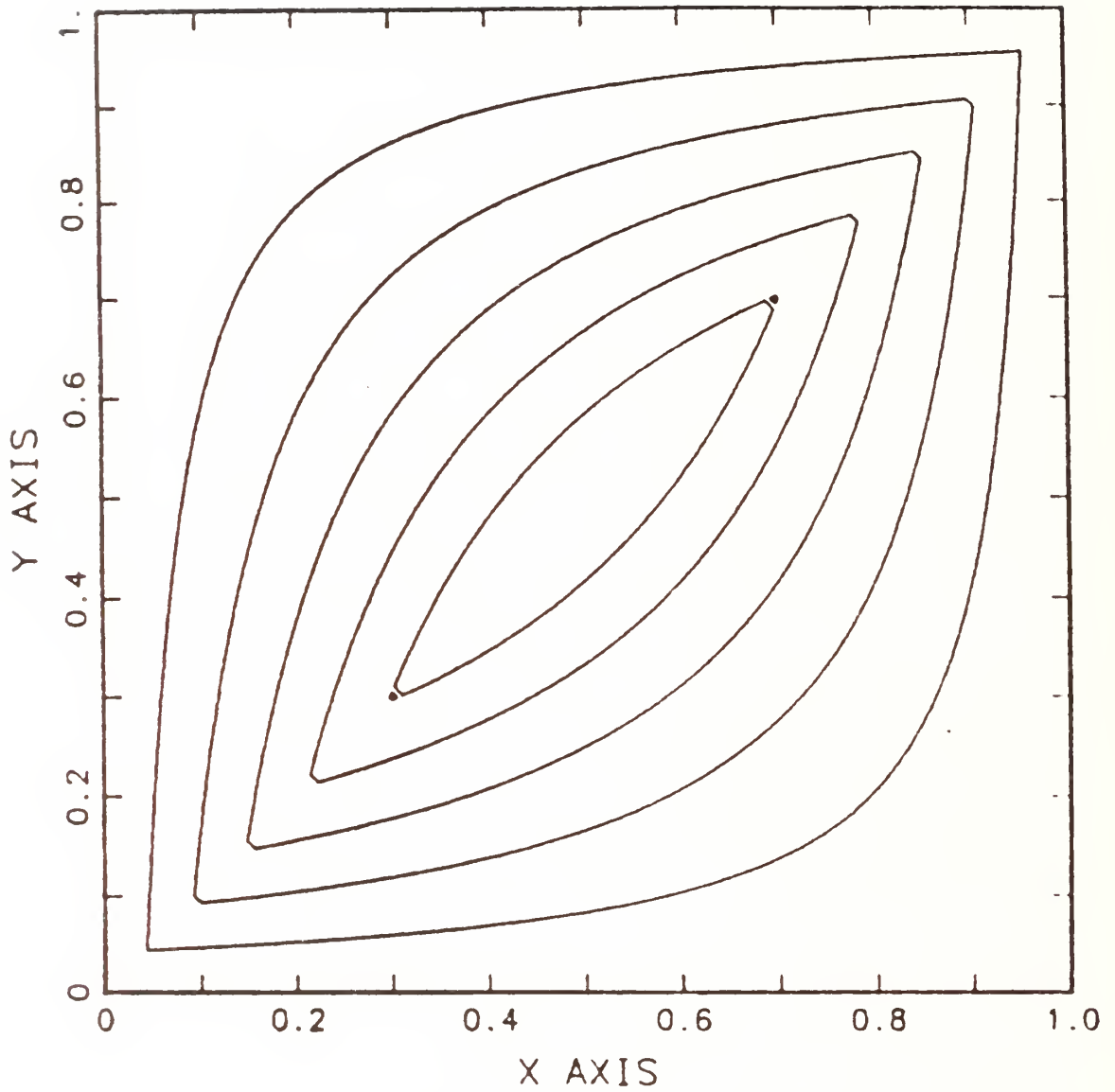


Figure 1(a). Bivariate Beta P1,  $\rho=0.33$ , contours .5(.5)2.5



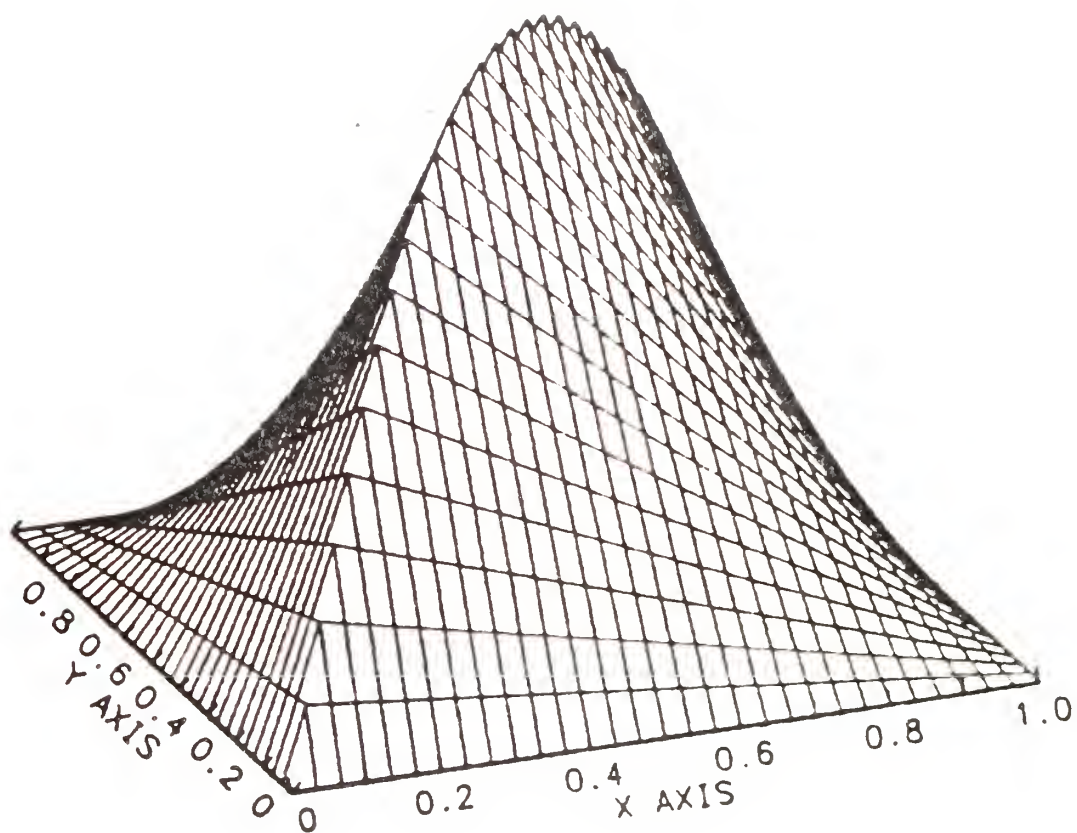


Figure 1(b). Bivariate Beta P1,  $\rho=0.33$ , Surface Plot

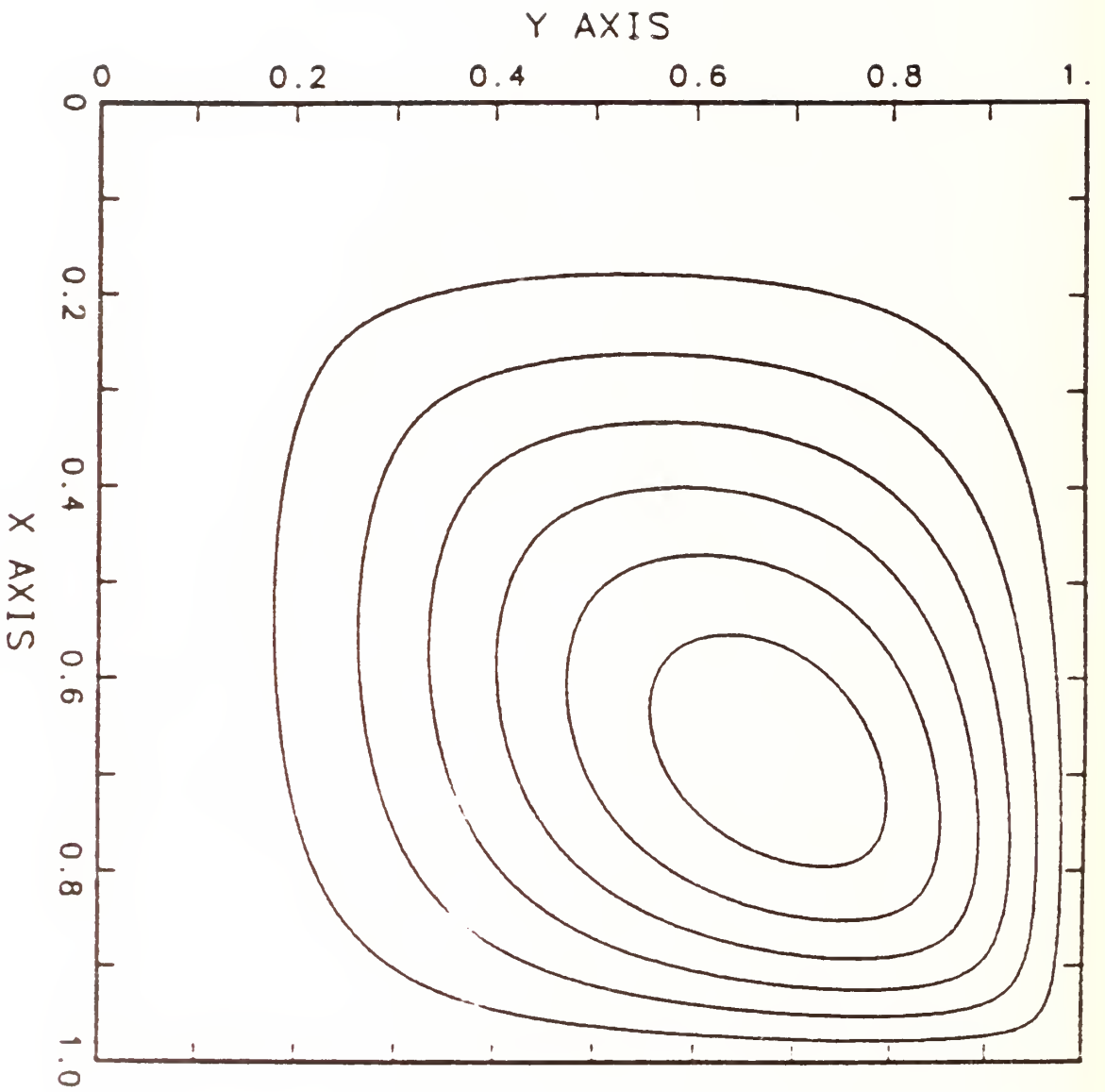


Figure 2. Bivariate Beta P2,  $\rho=0.1667$ , contours .5(.5)3

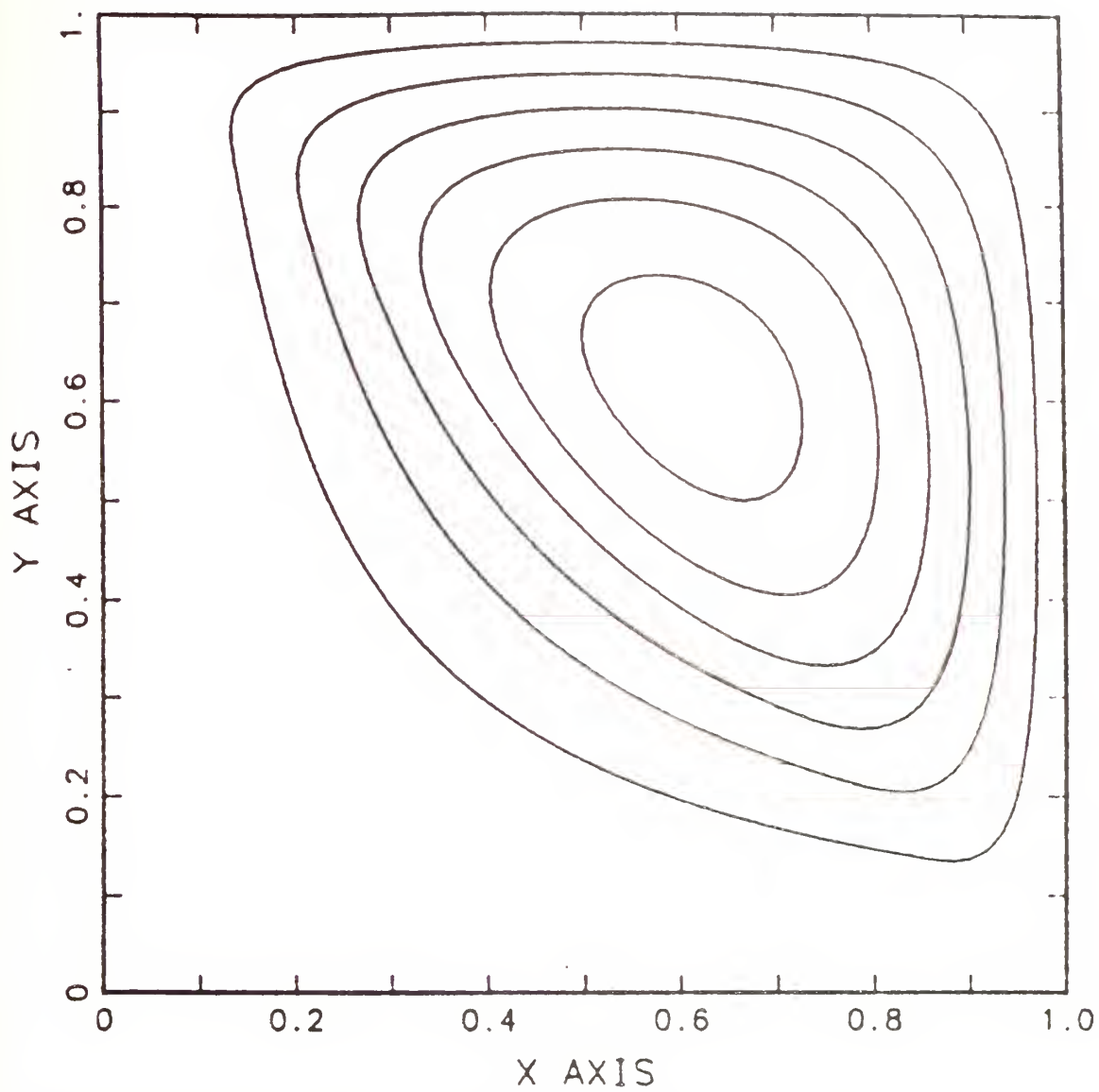


Figure 3. Bivariate Beta  $N_2$ ,  $\rho = -0.25$ , contours .5(.5)3

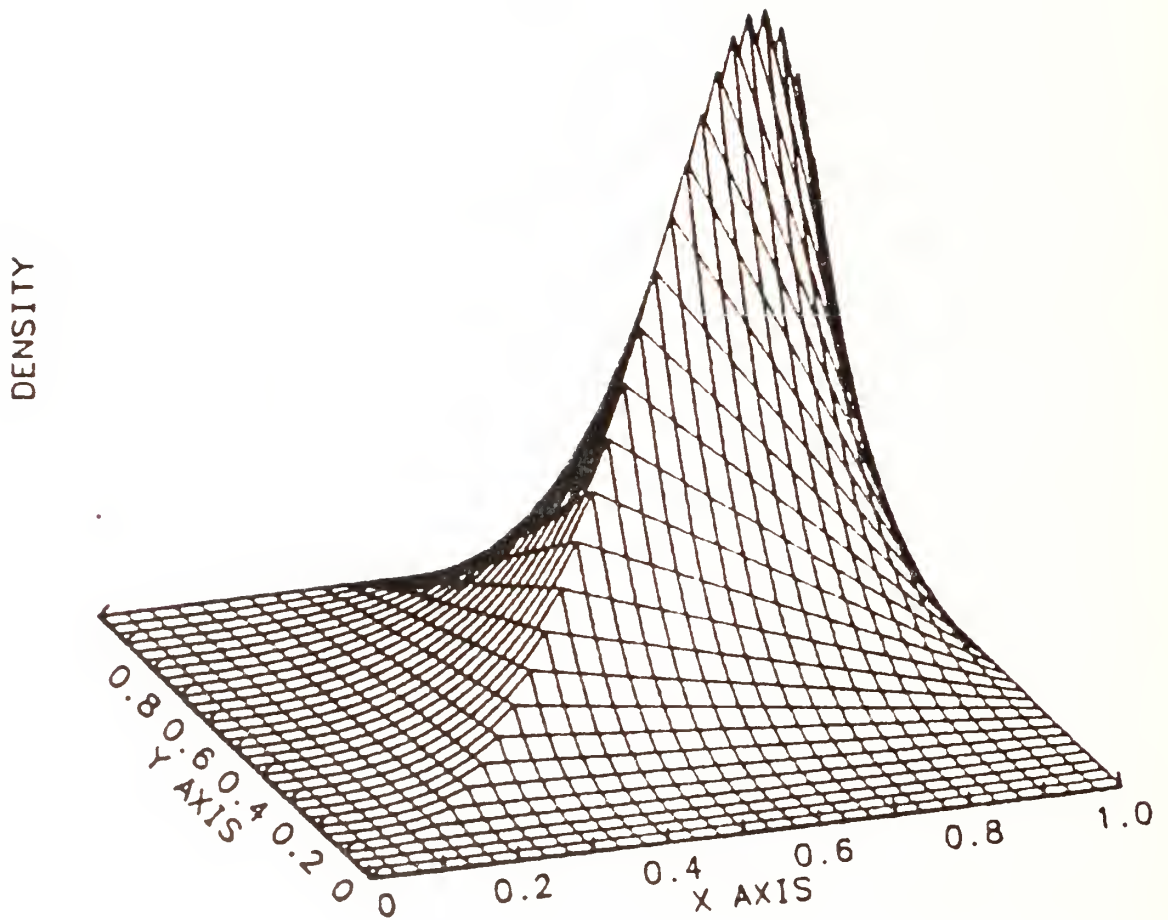


Figure 4. Bivariate Beta P3,  $\rho=0.5$

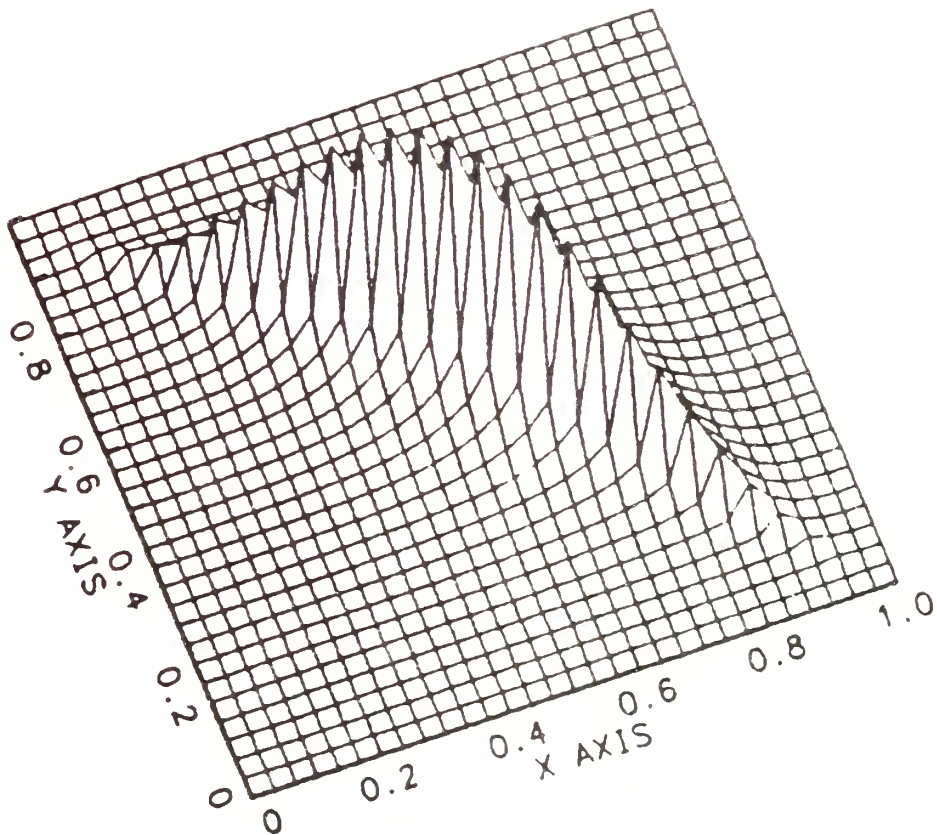


Figure 5. Bivariate Beta N3,  $\rho = -0.75$

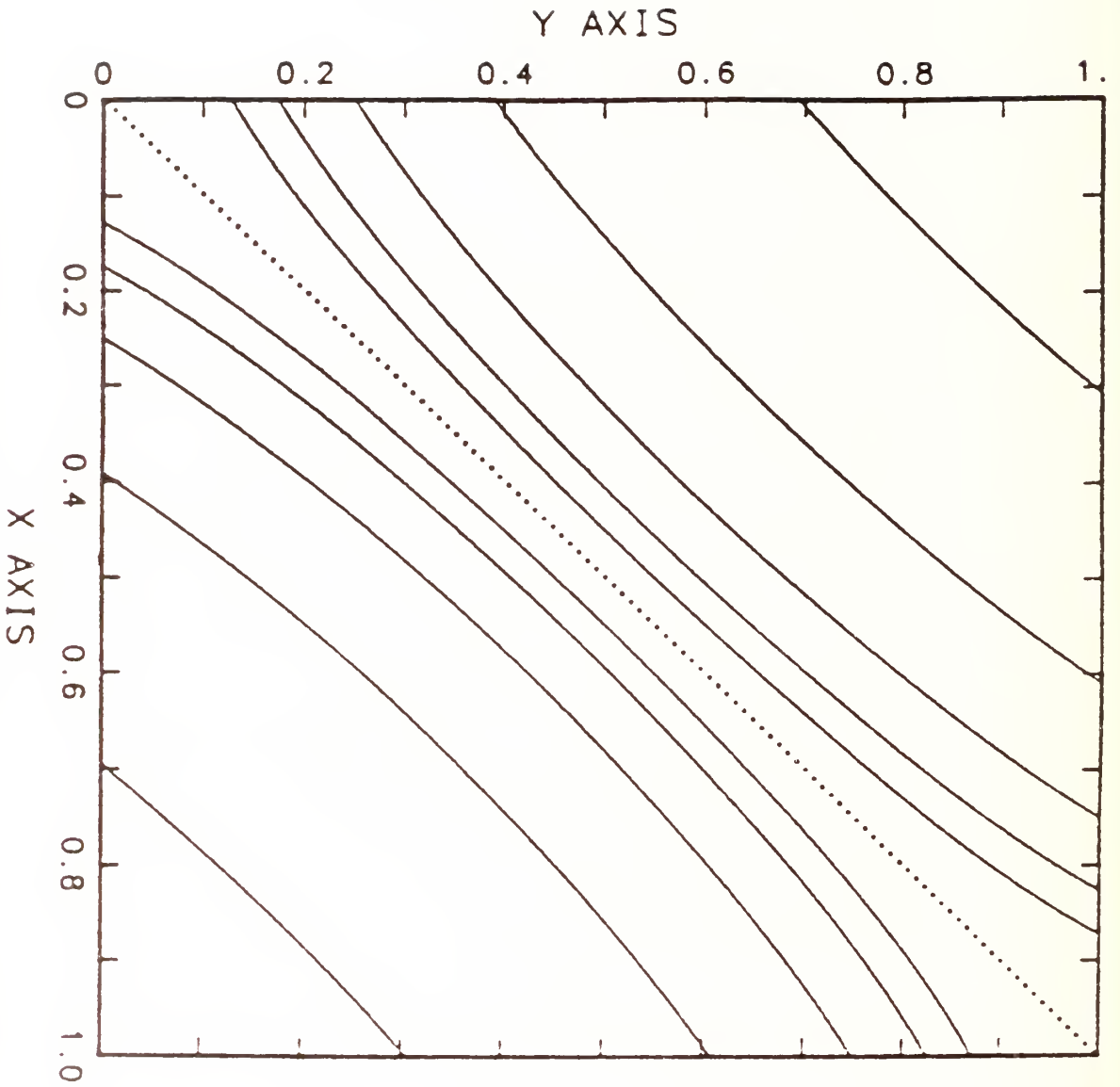


Figure 6. Bivariate Uniform,  $\rho=0.333$ , contours .5(.5)3

DENSITY

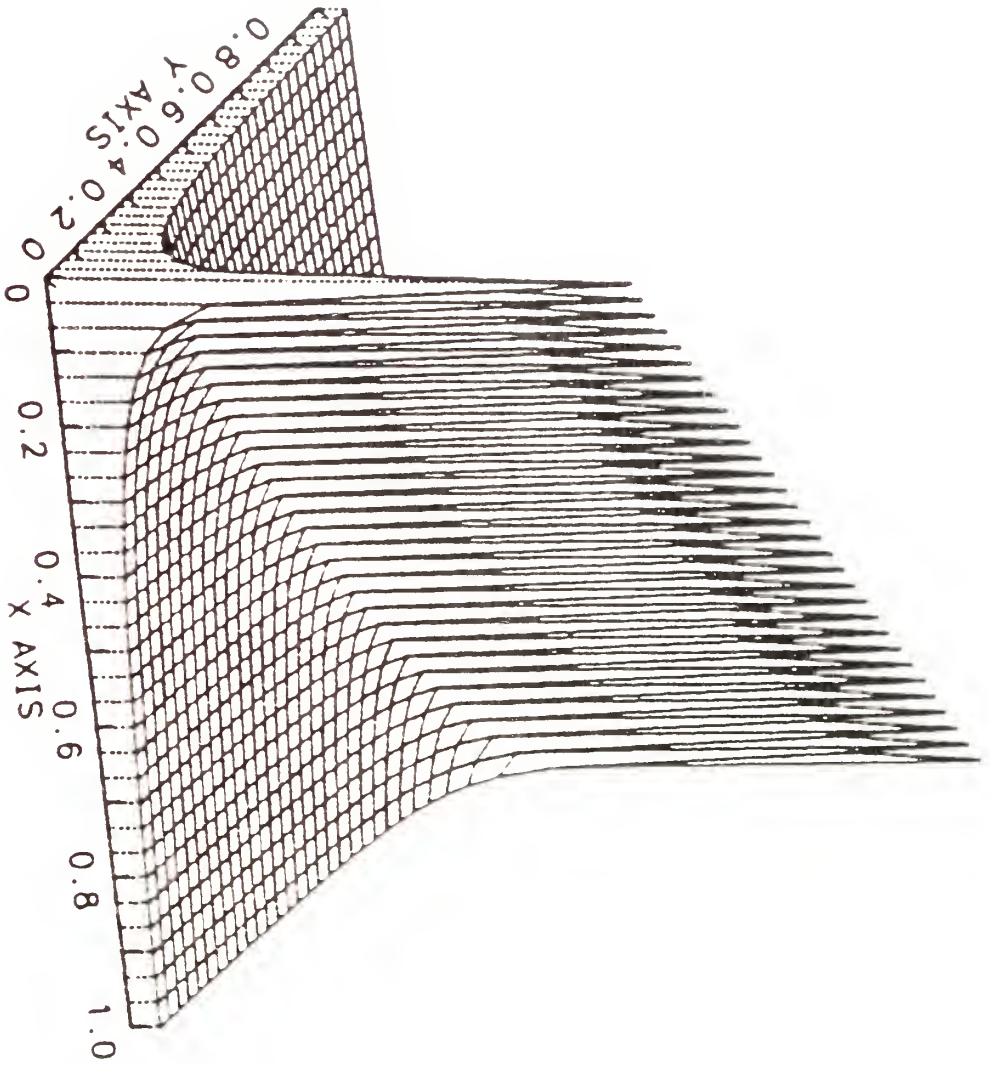


Figure 6. Bivariate Uniform,  $\rho=0.333$ , contours .5(.5)3

## REFERENCES

- AITCHISON, J. and SHEN, S. M. (1980). Logistic Normal distributions: some properties and uses. *Biometrika*, 67, 2, 261-272.
- AITCHISON, J. (1982). *The Statistical Analysis of Compositional Data*, J. R. Statist. Soc. B., 44, 139-177.
- AZZALINI, A. (1981). A Markov process with Beta marginal distribution, unpublished note, Instituto di Statistica, Universita di Padova, Italy.
- BARNETT, V. (1980). Some Bivariate Uniform Distributions, *Commun. Statist. Theor. Method.*, A9(4), 453-461.
- COLEMAN, J. S. (1964). *Introduction to mathematical sociology*, Free Press of Glencoe, Collier-Macmillan Ltd., London.
- DARROCH, J. N. and RATCLIFF, D. (1971). Characterization of the Dirichelet Distribution, *J. Amer. Statist. Assoc.*, 66, 641-643.
- JACOBS, P. A. and LEWIS, P. A. W. (1983). Stationary discrete autoregressive moving average time-series generated by mixtures, *J. Time Series Anal.*, 4, 1, 19-36.
- JOHNSON, N. L. and KOTZ, S. (1972). *Distributions in Statistics: Continuous Multivariate Distributions*, John Wiley and Sons, New York.
- KARLIN, S. and TAYLOR, H. M. (1975). *A First Course in Stochastic Processes* (2nd. Ed.), Academic Press, New York.
- \_\_\_\_\_ (1981). *A Second Course in Stochastic Processes*, Academic Press, New York.
- LAWRANCE, A. J. and LEWIS, P. A. W. (1980). The exponential autoregressive moving average EARMA(p,q) process, *J. R. Statist. Soc. B*, 42, 2, 150-161.
- \_\_\_\_\_ (1982). A mixed exponential time-series model, *Mgmt. Science*, 28, 9, 1045-1053.
- LEWIS, P. A. W. and LAWRANCE, A. J. (1983a). Simple dependent pairs of exponential and uniform random variables, *Operations Research*, to appear.
- \_\_\_\_\_ (1983b). A Residual Analysis for time-series with autoregressive correlation structure, to appear.
- MASSEY, W. F., MONTGOMERY, D. B. and MORRISON, D. G. (1970). *Stochastic Models of Buyer Behaviour*, M.I.T. Press, Cambridge, Mass.
- McKENZIE, Ed. (1982). Product Autoregression: a time-series characterization of the Gamma distribution, *J. Appl. Prob.*, 19, 463-468.



- MORAN, P. A. P. (1967). Testing for correlation between non-negative variates, *Biometrika*, 54, 385-394.
- SCHMEISER, B. and LAL, R. (1980). Multivariate Modelling in Simulation: a survey, in 33rd, Annual Technical Conference Transactions, American Society for Quality Control (Atlanta), 252-261.
- SOUZA, R. C. and HARRISON, P. J. (1981). A Bayesian-Entropy approach to forecasting: the steady-state model. Technical Report No. 33, Dept. of Statistics, University of Warwick, Coventry, U.K.
- WEISS, G. (1975). Time-reversibility of linear stochastic processes, *J. Appl. Prob.*, 12, 831-836.
- WICHERN, D. W. and JONES, R. H. (1977). Assessing the impact of market disturbances using intervention analysis, *Management Sci.*, 24, 3, 329-337.

## DISTRIBUTION LIST

	NO. OF COPIES
Defense Technical Information Center Cameron Station Alexandria, VA 22314	2
Library Code 0142 Naval Postgraduate School Monterey, CA 93943	2
Research Administration Code 012A Naval Postgraduate School Monterey, CA 93943	1
Library Code 55 Naval Postgraduate School Monterey, CA 93943	1
Professor Ed McKenzie Code 55 Naval Postgraduate School Monterey, CA 93943	60
Professor P. A. W. Lewis Code 55Lw Naval Postgraduate School Monterey, CA 93943	10



DUDLEY KNOX LIBRARY



3 2768 00332765 1

□