Journal of the Korean
Statistical Society
Vol. 22, No. 2, 1993

Robustness of Predictive Density and Optimal Treatment Allocation to Non-Normal Prior for The Mean

Ashok K. Bansal and Pankaj Sinha [1]

ABSTRACT

The predictive density function of a potential future observation and its first four moments are obtained in this paper to study the effects of a non-normal prior of the unknown mean of a normal population.

The derived predictive density function is modified to study changes in utility curves, used to choose the optimum treatment from a given set of treatments, at a given level of stimulus due to slight deviations from normality of the prior distribution. Numerical illustrations are provided to exhibit some effects.

KEYWORDS: Edgeworth series distribution, Barton Dennis region, U-tility, Non normality parameter, Predictive distribution.

1. INTRODUCTION

The concept of predictive distribution dates back, at least, to Laplace's rule of succession. The predictive approach which had long been neglected or entirely disregarded expresses statistical decisions in terms of potential observables. Geisser(1982) provides compelling arguments in favour of predictive approach to statistical inference problems. An excellent monograph of Aitchison and Dunsmore(1975) provides

[1] Department of Statistics, University of Delhi, Delhi, India

a detailed treatment of prediction analyses in Bayesian set up with a variety of illustrative applications.

Bayesian approach to prediction of the probability distribution of potential future sample data assumes availability of the prior distribution besides sample information. It is often felt that models generating the sample information will have some external validity, not possessed by priors on the parameters of the model. This implies that the probability density $f(x|\theta)$ will be known much more accurately than the prior distribution. Inference robustness with respect to prior (IRP) is, therefore, considered to be the bigger problem by non Bayesians. However, if $f(x|\theta)$ itself is uncertain, standard inference robustness (SIR) of the Bayesian inference/decision should be investigated along with IRP.

However, any Bayesian analysis based on a single prior is questionable due to subjectivity involved in its choice. Berger (1984, 1990) surveyed several approaches for examining robustness of Bayes decisions of possible misspecification of the prior distribution, while discussing robust Bayesian viewpoint. A reasonable approach is to consider a class of plausible priors which are in the neighbourhood of a specific assessed approximation to the " true " prior and study sensitivity of the decision as the prior varies over this class. The class of Edgeworth series distributions (ESD) has been employed by Bansal (1978 a, b, 1979, 1980) and Chakravarti and Bansal (1988) to investigate effects of non-normal prior on Bayes decisions and forecasts.

The investigators performing prediction analyses in Bayesian framework find mathematically and computationally convenient to use either conjugate class of priors or non-informative (or vague) priors. There is hardly any sensitivity study of decisive predictions to slight changes in prior from the commonly used conjugate prior. We obtain the predictive density function of a future potential sample mean, given an independent sample from a similar normal population, with an ESD prior for the unknown mean. Measures of skewness and kurtosis are computed to illustrate the effect of non-normality in the prior on the predictive density. Under the assumptions, given in Aitchison and Dunsmore (1975), treatment allocation problem is revisited to investigate effects of Edgeworth type of non-normality in the prior on the choice of optimum treatment at a given level of stimulus.

2. PREDICTIVE DENSITY FUNCTION

Let us consider a random sample $\underline{x}= (x_1, x_2, \ldots, x_n)$ from $N(\theta, r)$ population with unknown mean θ but known precision $r(>0)$. Suppose $\underline{Y}= (Y_1, Y_2, \ldots, Y_N)$ be an independent potential future sample from the same population. Further, assume that θ has a non-normal prior density function $\xi(\theta)$ belonging to the class of Edgeworth series distributions (ESD), given by

$$\xi(\theta) = (\tau/2\pi)^{1/2}\exp[-\tau(\theta-\mu)^2/2]H(\theta) \tag{2.1}$$

where,

$$H(\theta) = 1 + \frac{1}{6}\lambda_3 H_3(\sqrt{\tau}(\theta-\mu)) + \frac{1}{24}\lambda_4 H_4(\sqrt{\tau}(\theta-\mu)) + \frac{1}{72}\lambda_3^2 H_6(\sqrt{\tau}(\theta-\mu)) \tag{2.2}$$

$H_k(\cdot)$ is the Hermite polynomial of degree k; λ_3 and λ_4 are the measures of skewness and kurtosis, respectively. An ESD prior $\xi(\theta)$ is unimodal and a proper density function when (λ_3, λ_4) lies in the Barton-Dennis (1952) region. The advantage with Edgeworth series approach is that it represents a class of prior distributions which includes the normal conjugate prior $N(\mu, \tau)$ as one of the members for $\lambda_3 = \lambda_4 = 0$. With varying values of terms λ_3 and λ_4 (other than zero), it gives a variety of moderately non-normal uni-modal proper p.d.fs taking both skewness and kurtosis into consideration. We have found the class of ESD priors to be mathematically very convenient to work with, and study of sensitivity to non-normality in the prior is computationally simple.

The predictive density function of the potential future sample mean M, given the observed sample $\underline{x}$, is

$$\begin{aligned} p(M|\underline{x}) &= \int_{-\infty}^{\infty} \xi(\theta|\underline{x})p(M|\theta)d\theta, \quad M = \sum Y_i/N \\ &= (Nr\tau'/2\pi\tau^*)^{1/2}\exp[-\frac{1}{2}\frac{Nr\tau'}{\tau^*}(M-\mu')^2]H^*(z)/G \end{aligned} \tag{2.3}$$

where,

$$\begin{aligned} H^*(z) &= 1 + \frac{1}{6}\lambda_3\tau_1^{3/2}H_3(z) + \frac{1}{24}\lambda_4\tau_1^2 H_4(z) + \frac{1}{72}\lambda_3^2\tau_1^3 H_6(z), \\ G &= [1 + \frac{1}{6}\lambda_3 p^{3/2}H_3\{\sqrt{p\tau}(m-\mu)\} + \frac{1}{24}\lambda_4 p^2 H_4\{\sqrt{p\tau}(m-\mu)\} \\ &\quad + \frac{1}{72}\lambda_3^2 p^3 H_6\{\sqrt{p\tau}(m-\mu)\}] \\ z &= \{n(m-\mu) + N(M-\mu)\}\sqrt{\tau r/\tau^*(N+n)}, \\ \mu' &= (\tau\mu + nrm)/\tau', \quad p = nr\tau' \\ \tau_1 &= r(n+N)/\tau^*, \ \tau^* = \tau + r(N+n), \ \tau' = \tau + nr, \ m = \sum x_i/n \end{aligned}$$

and $\xi(\theta|\underline{x})$ is the posterior density of θ, given $\underline{x}$, w.r.t. ESD prior density of θ which was derived by Bansal (1978a).

It is well known that the effect of a prior distribution becomes insignificant on the posterior distribution as the likelihood function dominates the prior for large sample sizes. Bansal (1978a) , in particular, reported that $\xi(\theta|\underline{x})$ was insensitive to an ESD prior (2.1) for $n > 10$. Thus $p(M|\underline{x})$ should also be insensitive to moderate amount of nonnormality in the $\xi(\theta)$ for large values of n.

The moments of the predictive density function (2.3) with respect to n ESD prior are given by

$$m_1 = \int_{-\infty}^{\infty} M p(M|m) dM, \tag{2.4}$$

$$m_k = \int_{-\infty}^{\infty} (M - m_1)^k p(M|m) dM, \quad k = 2, 3, 4 \tag{2.5}$$

The measures of skewness and kurtosis are

$$\lambda_3^{*2} = m_3^2/m_2^3 \text{ and } \lambda_4^* = (m_4/m_2^2) - 3.$$

The expansions for m_1, m_2, m_3 and m_4 are given in the Appendix. It may be noted that for the normal prior $(\lambda_3 = \lambda_4 = 0)$ $m_3 = \lambda_3^* = 0$ and $\lambda_4^* = 0$.

Illustration 1. In order to illustrate effects of an ESD prior and observed sample data $\underline{x}$ on the predictive density function, we compute λ_3^* and λ_4^* for some ESD priors with common $\mu = 0$, $\tau = 3$. We take $r = 1$, $n = 1$, $N = 1$.

Table 1 suggests that non-normality in the prior passes on, to some extent, to the predictive density function. In particular, priors with $\lambda_4 = 2.0$, given $x = 2$, leads to highly leptokurtic asymmetric predictive density functions. However, for $x \leq 1$, the effect of non-normality seems to be less severe. The effect of kurtosis in the prior appears to be more serious on both the skewness and kurtosis of the predictive density function and it becomes more pronounced as x increases from 0 to 2.

3. PREDICTION FOR A REGRESSION MODEL

Suppose the future experiment records the response y (value of the dependent variable) made by some experimental unit to a known stimulus t (value of the explanatory variable). Assume that the common parameter be θ for each t and $p(y|t,\theta)$ be the density function of y. Let us denote the data set $(t_1, x_1), (t_2, x_2), \cdots, (t_n, x_n)$ by $\underline{z}$ which is obtained by n earlier independent and identical informative experiments. The predictive density function for y, given t and sample data $\underline{z}$, is

$$\begin{aligned} p(y|t,\underline{z}) &= \int_\Theta p(y|t,\theta) p(\theta|\underline{z}) d\theta \\ &= \int_\Theta p(y|t,\theta) p(\theta) \prod_{i=1}^{n} p(x_i|t_i,\theta) d\theta \Big/ \int_\Theta p(\theta) \prod_{i=1}^{n} p(x_i|t_i,\theta) d\theta \end{aligned} \tag{3.1}$$

For the normal linear regression case, y has a normal distribution with mean $\alpha + \beta t$ and precision $r(> 0)$. The least squares estimate of β is

$$\hat{\beta} = \sum(t_i - \bar{t})(x_i - \bar{x}) / \sum(t_i - \bar{t})^2; \quad \bar{t} = \sum t_i/n; \quad \bar{x} = \sum x_i/n$$

based on the data set $\underline{z}$. The regression line for X on t is

$$X = \bar{X} + \hat{\beta}(t - \bar{t}) \tag{3.2}$$

With $E(X) = \alpha + \beta t$ and $Var(X) = \frac{1}{r}[\frac{1}{n} + (t - \bar{t})^2 / \sum(t_i - \bar{t})^2]$
Thus, $\bar{X} + \hat{\beta}(t - \bar{t})$ is normally distributed with mean $\alpha + \beta t$ and precision kr where

$$k^{-1} = n^{-1} + (t - \bar{t})^2 / \sum(t_i - \bar{t})^2 \tag{3.3}$$

The predictive density function of the future response y for stimulus t when prior distribution for $\theta(= \alpha + \beta t)$ belongs to the ESD class (2.1) may be easily rewritten by replacing $1/n$ by $1/k$ and $N = 1$ in (2.3).

4. TREATMENT ALLOCATION

A treatment to an object is applied to change its present condition to a desirable future condition with respect to some aspect of the object.

Suppose the data obtained from n independent informative experiments is of the form $\underline{z} = \{(a_i, t_i, x_i) : i = 1, 2, \cdots, n\}$, where the triplet (a, t, x) denote the treatment, initial quality and final quality of a batch. Suppose a new batch of initial quality $t \in T$ awaits treatment $a \in A$. Let us assume that the possible density functions of the final quality $y \in Y$ are $p(y|a, t, \theta)$ and there is a linear regression of final quality y on initial quality t for each treatment a. The prognostic density function for treatment a with respect to the initial state t is the predictive density function of y

$$p(y|a, t, \underline{z}) = \int_{\Theta} p(y|a, t, \theta) p(\theta|\underline{z}) d\theta \tag{4.1}$$

If the advantage of improving quality from initial state t to final state y is $g(y) - g(t)$ and the cost of treatment a is C_a, let us consider the utility function

$$U(a, t, y) = g(y) - g(t) - C_a, (a \in A, \; t \in T, \; y \in Y), \tag{4.2}$$

The Bayes decisive prediction approach to treatment allocation problem provides the optimum treatment $a^*(t)$, corresponding to state t, which maximises the expected utility

$$U(a,t) = \int_Y U(a,t,y)p(y|a,t,\underline{z})dy \tag{4.3}$$

Illustration 2. We consider the example 12.1 of Aitchison and Dunsmore (1975) to illustrate sensitivity of optimum treatment allocation to possible misspecification of the prior distribution of the unknown mean. Table 2 shows treatments, initial and final qualities in 30 experimental runs of a quality improving process.

We suppose that the underlying stochastic model is $N(\alpha_a + \beta_a t, r_a)$ for treatment $a \in A$. Let us assume that the selling price is directly proportional to quality i.e., $g(y) = y$ with costs $C_1 = 4, C_2 = 5, C_3 = 3$. The relevant information on the parameters (α_a, β_a, r_a) comes from 10 results in each of the three treatments. The regression parameters (α_a, β_a) are unknown and we use their least squares estimates for the data set $\underline{z}$ in Table 2. The precisions r_a are assumed to be known and we take their maximum likelihood estimates as their true values. The values of the prior parameters μ and τ are chosen by ML-II prior method (cf. Berger, 1985).

The expected utility works out to be

$$\begin{aligned} U(a,t) &= \int_{-\infty}^{\infty} (y-t)p(y|a,t,\underline{z})dy - c_a \\ &= \hat{m}_1 - t - c_a, \quad a = 1,2,3 \end{aligned} \tag{4.4}$$

where $\hat{m}_1$ is the estimated value of the predictive mean with $n_1 = n_2 = n_3 = 10$, $N = 1, M = y$ and values of the parameters $(\mu_a, \alpha_a, \beta_a, r_a, \tau_a)$, $a = 1,2,3$, given in Table 3. In particular, for the normal prior $(\lambda_3 = \lambda_4 = 0)$ we have

$$U(a,t) = [\tau\mu + kr(\alpha + \beta t)]/(t + kr) - t - C_a, \tag{4.5}$$

where $k^{-1} = n^{-1} + (t-\bar{t})^2/\sum(t_i - \bar{t})^2$.

We plot the utility function (4.4) against t for each of the three treatments to illustrate effects of an Edgeworth type non-normality in the prior on utility curves. Table 4 gives optimum treatment for some members of ESD class of priors as $t = 26(4)42$ on the basis of their graphs. Some cases are illustrated in Figs. 1-4.

We note, for example, treatment 3 is best for $t < 34$ whereas treatment 1 should be assigned for $t > 34$ with a normal prior. Treatments 1 and 3 are equally good for values of t in the neighbourhood of 34 (see Fig. 1). In the case of priors with $(\lambda_3, \lambda_4) = (0.2, 2.0), (0.3, 2.0), (0.4, 2.0)$ treatment 2 is best for all values of t. A region of indifference occurs in the neighbourhood of $t = 34$ when $\lambda_4 = 0.8$ and $\lambda_3 \geq 0.2$.

5. CONCLUSIONS

The first illustration indicates that the predictive density function may be quite sensitive to kurtosis in the prior distribution when the observation x is wild (or extreme).

The choice of optimum treatment is affected by moderate amount of non-normality in the prior distribution. The graphs of utility curves in second illustration suggest counterbalancing effects of skewness and kurtosis in the prior on the expected utility of the treatment at each level of the treatment.

The decisions, therfore, based on predictive decisions may be quite sensitive to slight departures from normality assumption of the prior distribution.

ACKNOWLEDGEMENTS

We are grateful to the UGC for granting a Minor Research Project in Science to the first author and SRF (National Education Test) to the second author. We thank the referees for their useful suggestions.

APPENDIX

Let us write

$$I_k = \int_{-\infty}^{\infty} (M - \mu')^k p(M|m) dM, \quad k = 1, 2, 3, 4. \tag{A1}$$

Then

$$m_1 = I_1 + \mu' \tag{A2}$$
$$m_2 = I_2 - (m_1 - \mu')^2 \tag{A3}$$
$$m_3 = I_3 - 3(m_1 - \mu')m_2 - (m_1 - \mu')^3 \tag{A4}$$
$$m_4 = I_4 - 4(m_1 - \mu')m_3 - 6(m_1 - \mu')^2 m_2 - (m_1 - \mu')^4 \tag{A5}$$

The integrals I_k are evaluated by repeated use of the integral

$$\left(\frac{q}{2\pi}\right)^{\frac{1}{2}} \int_{-\infty}^{\infty} \exp[-\frac{1}{2}q(M - \mu')^2][M - \mu']^k z^\ell dM$$
$$= (AR)^\ell \left(\frac{q}{2\pi}\right)^{\frac{1}{2}} \int_{-\infty}^{\infty} \exp[-\frac{1}{2}q[M - \mu')^2] \sum_{j=0}^{\ell} \binom{\ell}{j} \left(\frac{B}{A}\right)^j (M - \mu')^{k+j} dM$$

$$= (AR)^\ell \sum_{j=0}^{\ell} \binom{\ell}{j} \left(\frac{B}{A}\right)^j \left(\frac{2}{q}\right)^{\frac{1}{2}(k+j)} \frac{1}{\sqrt{\pi}} \int_0^\infty t^{(j+k+1)\frac{1}{2}-1} e^{-t} dt$$

$$= (AR)^\ell \sum_{j=0}^{\ell} \binom{\ell}{j} \left(\frac{B}{A}\right)^j \left(\frac{2}{q}\right)^{\frac{k+j}{2}} \frac{1}{\sqrt{\pi}} \Gamma(\frac{j+k+1}{2}) \qquad \text{(A6)}$$

and after tedious algebra we find

$$I_1 = \frac{1}{6}(1-\frac{\tau}{\tau^*})^{3/2}\frac{BR}{q}[3\lambda_3\{(A^2+B_1^2)R^2-1\}+\lambda_4(1-\frac{\tau}{\tau^*})^{1/2} AR\{(A^2+3B_1^2)R^2-3\}+\frac{1}{2}\lambda_3^2(1-\frac{\tau}{\tau^*})^{3/2}AR\{R^4(A^4+10A^2B_1^2+15B_1^4) -10R^2(A^2+3B_1^2)+15\}]/G \qquad \text{(A7)}$$

$$I_2 = [1+\frac{1}{6}\lambda_3 AR\{R^2(A^2+9B_1^2)-3\}(1-\frac{\tau}{\tau^*})^{3/2}+\frac{1}{24}\lambda_4(1-\frac{\tau}{\tau^*})^2 \{R^4(A^4+18A^2B_1^2+15B_1^4-6R^2(A^2+3B_1^2)+3\}+\frac{1}{72}\lambda_3^2(1-\frac{\tau}{\tau^*})^3 \{R^6(A^6+45A^4B_1^2+225A^2B_1^4+105B_1^6)-15R^4(A^4+18A^2B_1^2+15B_1^4) +45R^2(A^2+3B_1^2)-15\}]/qG \qquad \text{(A8)}$$

$$I_3 = \frac{BR}{2q^2}(1-\frac{\tau}{\tau^*})^{3/2}[3\lambda_3(A^2R^2+5B_1^2R^2-1)+\lambda_4(1-\frac{\tau}{\tau^*})^{1/2}AR(A^2R^2 +5B_1^2R^2-3)+\frac{1}{6}\lambda_3^2(1-\frac{\tau}{\tau^*})^{3/2}\{R^4(3A^4+50B_1^2A^2+105B_1^4)-30R^2(A^2 +5B_1^2)+45\}]/G \qquad \text{(A9)}$$

$$I_4 = \frac{3}{q^2G}[1+\frac{1}{6}\lambda_3 AR\{R^2(A^2+15B_1^2)-3\}(1-\frac{\tau}{\tau^*})^{3/2}+\frac{1}{24}\lambda_4(1-\frac{\tau}{\tau^*})^2 \{R^4(A^4+30A^2B_1^2+35B_1^4)-6R^2(A^2+5B_1^2)+3\}+\frac{1}{72}\lambda_3^2(1-\frac{\tau}{\tau^*})^3 \{R^6(A^6+75A^4B_1^2+525A^2B_1^4+315B_1^6)-15R^4(A^4+30A^2B_1^2+35B_1^4) +45R^2(A^2+5B_1^2)-15\}] \qquad \text{(A10)}$$

Where,

$$A = \{N\mu' + n(m-\mu)\}R/\tau^*, \quad B = Nr/\tau^*, \quad B_1 = B/\sqrt{q},$$
$$R = \{\tau\tau^*/r(N+n)\}^{1/2}, \quad q = Nr\tau'/\tau^*.$$

REFERENCES

(1) Aitchison, J. and Dunsmore, I. R.(1975). *Statistical Prediction Analysis.* Cambridge University Press, Cambridge, London.

(2) Bansal, A. K.(1978a). Robustness of Bayes Estimator for the Mean of a Normal Population with Non-normal Prior. *Communications in Statistics* A7, 453-460.

(3) Bansal, A. K.(1978b). Robustness of Bayes Forecast to Non-normality. *Journal of the Korean Statistical Society*, Vol. 7, 11-16.

(4) Bansal, A. K.(1979). Effect of Non-normality on Bayes Decision Function for Testing Normal Mean. *Journal of the Korean Statistical Society*, Vol. 8, 15-21.

(5) Bansal, A. K.(1980). The Behaviour of Bayes Decisions for the Normal Mean under Non-standard Prior : Unknown Precision. In : S. Ikeda et. al.(eds.), *Statistical Climatology*, Elsvier Scientific Publishing Company, Amsterdam, 85-98.

(6) Barton, D. E. and Dennis, K. E.(1952). The Conditions under which Gram Charlier and Edgeworth Curves are Positive, Definite and Unimodal. *Biometrika* , Vol. 39, 425-427.

(7) Berger, J. O.(1984). Robust Bayesian Viewpoint. In : J. Kadane ed., *Robustness of Bayesian Analysis*, North Holland, Amsterdam, 63-124.

(8) Berger, J. O.(1985). *Statistical Decision Theory and Bayesian Analysis.* Springer-Verlag, New York.

(9) Berger, J. O.(1990). Robust Bayesian Analysis : Sensitivity to the Prior. *Journal of Statistical Planning and Inference*, North Holland, Vol. 25, 303-323.

(10) Chakravarti, S. and Bansal, A. K.(1988). Effect of Non-normal Prior for Regression Parameter on Bayes Decisions and Forecasts. *Journal of Quantitative Economics*, Vol. 4, 247-259.

(11) Geisser, S.(1982). Aspects of the Predictive and Estimative Approaches in the Determination of Probabilities. *Biometrics*, Vol. 38, 75-85.

Table 1 : Skewness and Kurtosis of the Predictive Density Function

λ_3	λ_4	x = 0	0.5	1.0	2.0
0.0	0.0	(0,0)	(0,0)	(0,0)	(0,0)
0.2	0.8	(0.07, 0.25)	(0.08, 0.31)	(0.10, 0.51)	(0.15, 1.48)
0.2	2.0	(0.08, 0.72)	(0.10, 0.91)	(0.14, 1.66)	(0.40, 6.45)
0.3	0.8	(0.11, 0.24)	(0.12, 0.31)	(0.13, 0.51)	(0.19, 1.49)
0.3	2.0	(0.12, 0.70)	(0.14, 0.90)	(0.18, 1.56)	(0.49, 6.39)
0.4	0.8	(0.15, 0.23)	(0.16, 0.30)	(0.17, 0.51)	(0.24, 1.52)
0.4	2.0	(0.16, 0.69)	(0.18, 0.89)	(0.23, 1.56)	(0.57, 6.37)

Table 2 : Treatment, initial and final qualities in 30 experimental runs of a quality improving process

Treatment	(Initial quality, Final quality)		
1 :	(30.9, 44.2),	(35.8, 48.6),	(28.2, 44.3),
	(40.5, 50.0),	(23.5, 43.0),	(47.4, 52.5),
	(51.2, 55.0),	(43.0, 51.8),	(37.7, 49.6),
	(33.8, 46.1)		
2 :	(33.3, 46.1),	(31.3, 46.7),	(23.9, 42.7),
	(42.2, 50.0),	(27.4, 45.0),	(50.3, 51.0),
	(35.8, 47.3),	(45.7, 51.0),	(39.8, 49.1),
	(37.6, 47.7)		
3 :	(39.8, 46.3),	(31.3, 38.6),	(41.0, 50.1),
	(51.2, 57.3),	(36.4, 43.1),	(45.7, 56.8),
	(26.0, 37.8),	(37.1, 47.3),	(43.0, 52.4),
	(35.2, 45.0)		

Table 3 : Estimates of the parameters (α_a, β_a, r_a), $a = 1, 2, 3$ and the prior parameters μ and τ

Treatment	Parameters				
a	$\hat{\mu}_a$	$\hat{\tau}_a$	$\hat{\alpha}_a$	$\hat{\beta}_a$	$\hat{r}_a$
1	48.50	0.25	31.51	0.457	0.1
2	47.67	0.25	35.89	0.321	0.33
3	47.47	0.25	12.52	0.904	0.02

Table 4 : Optimum treatment for a given value of stimulus t

		t				
λ_3	λ_4	26	30	34	38	42
0.0	0.0	3	3	1, 3	1	1
0.2	0.8	2	2	1, 2, 3	2	2
0.2	2.0	2	2	2	2	2
0.3	0.8	2	2	1, 2, 3	2	2
0.3	2.0	2	2	2	2	2
0.4	0.8	2	2	1, 2, 3	2	2
0.4	2.0	2	2	2	2	2

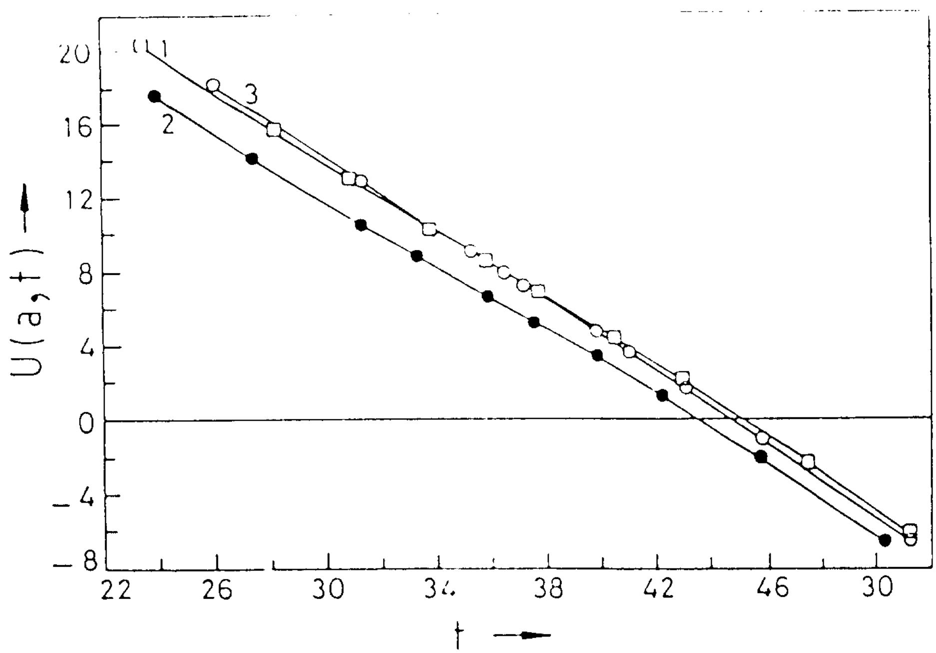

Figure 1. $\tau = 2.5$, $\lambda_3 = 0.0$, $\lambda_4 = 0.0$

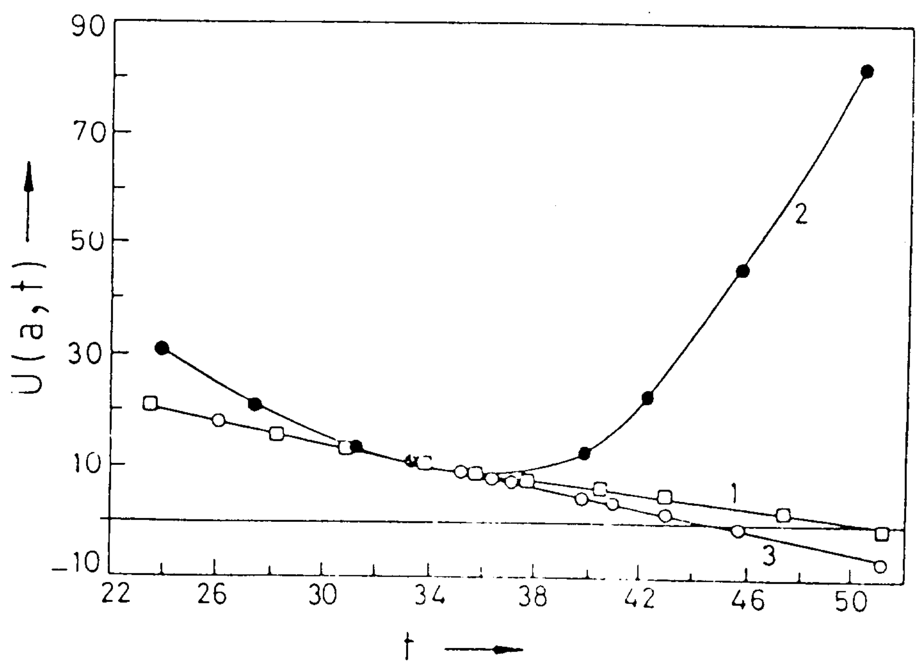

Figure 2. $\tau = 2.5$, $\lambda_3 = .2$, $\lambda_4 = .8$

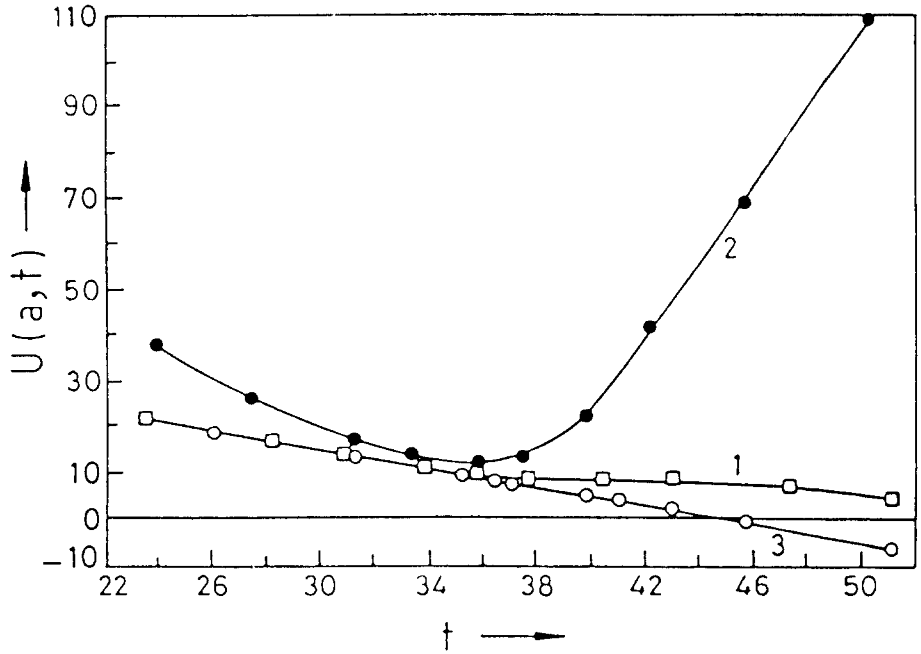

Figure 3. $\tau = 2.5$, $\lambda_3 = .2$, $\lambda_4 = 2.0$

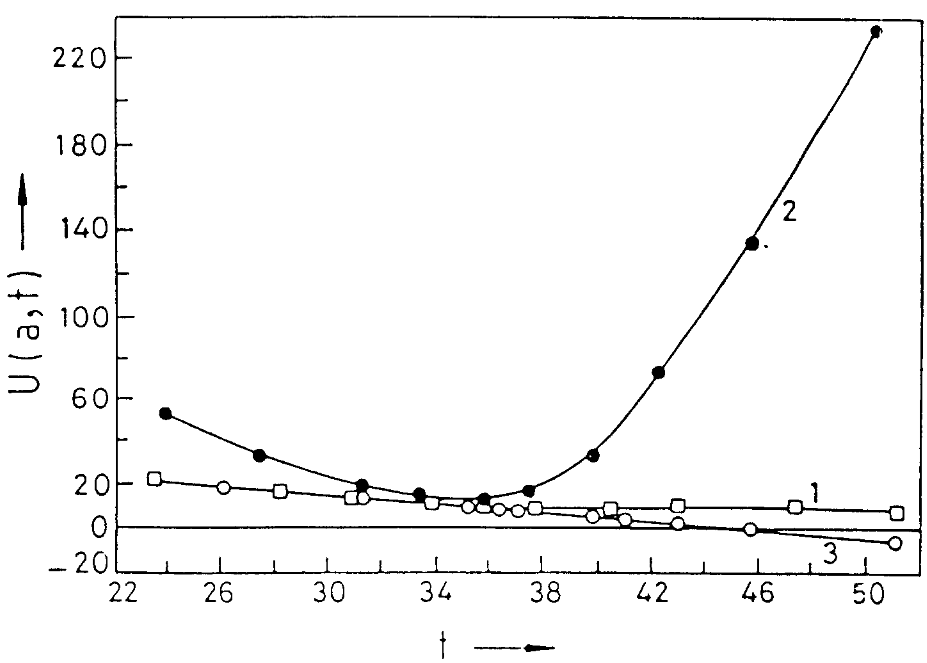

Figure 4. $\tau = 2.5$, $\lambda_3 = .4$, $\lambda_4 = 2.0$