



A Shift of the Mean Level in a Sequence of Independent Normal Random Variables: A Bayesian Approach

Austin F. S. Lee; Sylva M. Heghinian

Technometrics, Vol. 19, No. 4. (Nov., 1977), pp. 503-506.

Stable URL:

<http://links.jstor.org/sici?sici=0040-1706%28197711%2919%3A4%3C503%3AASOTML%3E2.0.CO%3B2-5>

Technometrics is currently published by American Statistical Association.

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/about/terms.html>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/journals/astata.html>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

The JSTOR Archive is a trusted digital repository providing for long-term preservation and access to leading academic journals and scholarly literature from around the world. The Archive is supported by libraries, scholarly societies, publishers, and foundations. It is an initiative of JSTOR, a not-for-profit organization with a mission to help the scholarly community take advantage of advances in technology. For more information regarding JSTOR, please contact support@jstor.org.

A Shift Of The Mean Level In A Sequence Of Independent Normal Random Variables

—A Bayesian Approach—

Austin F. S. Lee and Sylva M. Heghinian

Department of Mathematics
Boston University
Boston, Massachusetts 02215

In this article, a study is made about a shift in the mean of a set of independent normal random variables with unknown common variance. The marginal and joint posterior distributions of the unknown time point and the amount of shift are derived. Small and large sample results are presented.

KEY WORDS:

Mean Shift
Time Point and Amount of Shift
Posterior Distribution

1. INTRODUCTION

Suppose that X_1, X_2, \dots, X_n are random variables which have the following structure:

$$X_i = \begin{cases} \mu + \epsilon_i, & i = 1, 2, \dots, \tau; \\ \mu + \delta + \epsilon_i, & i = \tau + 1, \dots, n; \end{cases} \quad (1.1)$$

where $\epsilon_i, i = 1, 2, \dots, n$, are independently and normally distributed random variables with mean zero and variance σ^2 ; τ, μ, δ and σ are unknown parameters, $1 \leq \tau \leq n - 1, -\infty < \mu < \infty, -\infty < \delta < \infty, \sigma > 0$.

In practice, situations often arise that observations are taken on the random variables X_1, X_2, \dots, X_n at consecutive time points. Possibly due to some exogenous factors, the first and second parts of the sequence of the random variables may operate at two different mean levels. It is of interest to make statistical inference about the time point and the amount of shift. In terms of the notations given in (1.1), the unknown parameters τ and δ represent the time point and the amount of shift, respectively.

Many studies concerning this shifting problem and related topics have appeared in the literature: Chernoff and Zacks [1], Kander and Zacks [7], Hinkley [6], Srivastava and Sen [9], Ferreira [4], etc. Most

recently, Feder [2] [3], Sen and Srivastava [8] have given bibliographies in this subject.

In this article a Bayesian approach is used to obtain the marginal and joint posterior distributions of τ and δ . Bayesian inference can thus be made marginally on τ or δ , or jointly. In section 2, the marginal and joint posterior distributions of τ and δ are derived. Examples using real data are provided in section 3.

2. POSTERIOR DISTRIBUTIONS OF τ AND δ

In matrix notation, (1.1) can be written as

$$\mathbf{X} = \mu \mathbf{1}_n + \boldsymbol{\epsilon} + \delta \mathbf{c}_\tau, \quad (2.1)$$

where

$$\mathbf{X} = (X_1, X_2, \dots, X_n)';$$

$$\boldsymbol{\epsilon} = (\epsilon_1, \epsilon_2, \dots, \epsilon_n)';$$

$\mathbf{1}_n$ is an n -dimensional vector with all components unity;

\mathbf{c}_τ is also an n -dimensional vector with the first τ components all zero and the last $n - \tau$ all unity.

The notation " ' " denotes the transpose of a vector. The object is to derive the joint and marginal posterior distributions of τ and δ . It is assumed that the prior distributions of $\tau, \delta, \mu, \sigma$ are

- 1) $p(\tau) = 1/(n - 1), \tau = 1, 2, \dots, n - 1$;
- 2) $p(\delta)$ is normal with mean 0 and variance σ_δ^2 ;
- 3) $p(\mu)$ is normal with mean 0 and variance σ_μ^2 ;
- 4) $p(\sigma) \propto \sigma^{-1}$,

respectively. Furthermore, $\tau, \delta, \mu, \sigma$ are independent.

2.1 Joint posterior distribution of τ and δ

In this section the joint posterior distribution of τ and δ is derived. The likelihood function $L(\tau, \delta, \sigma | \mathbf{x})$ of τ, δ and σ , given the observations $\mathbf{x} = (x_1, x_2, \dots, x_n)'$ on \mathbf{X} has the form of a normal probability function with mean $\delta\mathbf{c}_\tau$ and covariance matrix \mathbf{U} , where

$$\mathbf{U} = \sigma^2 [\mathbf{I}_n + (\sigma_\mu^2/\sigma^2)\mathbf{1}_n\mathbf{1}_n'] \tag{2.1.1}$$

\mathbf{I}_n denotes the identity matrix of order n . Following Bayes' formula, the joint posterior probability function of τ, δ and σ , given \mathbf{x} is

$$p(\tau, \delta, \sigma | \mathbf{x}) \propto \sigma^{-n} |\mathbf{U}|^{-1/2} \exp [-1/2\{(\mathbf{x} - \delta\mathbf{c}_\tau)' \mathbf{U}^{-1}(\mathbf{x} - \delta\mathbf{c}_\tau) + \delta^2/\sigma_\delta^2\}] \tag{2.1.2}$$

It may be shown (see Gardner [5]):

$$\mathbf{U}^{-1} = \sigma^{-2}[\mathbf{I}_n - (1 + r)^{-1} \mathbf{1}_n\mathbf{1}_n'/n] \tag{2.1.3}$$

where

$$r = (\sigma^2/n)/\sigma_\mu^2.$$

Furthermore,

$$|\mathbf{U}| = \sigma^{2n} |\mathbf{I}_n + (\sigma_\mu^2/\sigma^2)\mathbf{1}_n\mathbf{1}_n'| = n\sigma^{2(n-1)}\sigma_\mu^2(1 + r).$$

Thus,

$$|\mathbf{U}|^{-1/2} = n^{-1/2}\sigma^{-(n-1)}\sigma_\mu^{-1}(1 + r)^{-1/2}. \tag{2.1.4}$$

In case where the prior knowledge about the target value μ is vague relative to the information obtainable from the sample (even for small sample size), we may assume that $\sigma^2/n \ll \sigma_\mu^2$. For example, the precision due to human setting of certain machines may be less than the precision of the machines in producing the items according to that setting. Then (2.1.3) and (2.1.4) can be further simplified to

$$\mathbf{U}^{-1} \doteq \sigma^{-2}[\mathbf{I}_n - \mathbf{1}_n\mathbf{1}_n'/n] \tag{2.1.5}$$

and

$$|\mathbf{U}|^{-1/2} \doteq \sigma^{-(n-1)}/(n^{1/2}\sigma_\mu),$$

respectively. Hence, from (2.1.2) and (2.1.5) we have that, for $\sigma^2/n \ll \sigma_\mu^2$,

$$p(\tau, \delta, \sigma | \mathbf{x}) \propto \sigma^{-n} \exp [-1/(2\sigma^2)\{Q(\tau, \delta) + \sigma^2\delta^2/\sigma_\delta^2\}], \tag{2.1.6}$$

where

$$Q(\tau, \delta) = (\mathbf{x} - \delta\mathbf{c}_\tau)'[\mathbf{I}_n - \mathbf{1}_n\mathbf{1}_n'/n](\mathbf{x} - \delta\mathbf{c}_\tau). \tag{2.1.7}$$

With an extensive amount of algebra, it can be shown that

$$Q(\tau, \delta) = H(\tau) + \tau(n - \tau)(\delta - \hat{\delta}_\tau)^2/n, \tag{2.1.8}$$

where

$$\hat{\delta}_\tau = \bar{x}_{n-\tau} - \bar{x}_\tau; \tag{2.1.9}$$

and

$$H(\tau) = \sum_{i=1}^{\tau} (x_i - \bar{x}_\tau)^2 + \sum_{i=\tau+1}^n (x_i - \bar{x}_{n-\tau})^2, \tag{2.1.10}$$

or

$$H(\tau) = \sum_{i=1}^n (x_i - \bar{x}_n)^2 - \tau(n - \tau)\hat{\delta}_\tau^2/n. \tag{2.1.11}$$

If, in addition, $\sigma_\delta^2 \gg \sigma^2$, then

$$p(\tau, \delta, \sigma | \mathbf{x}) \propto \sigma^{-n} \exp [-Q(\tau, \delta)/(2\sigma^2)]. \tag{2.1.12}$$

This result is obtained by applying the well known identity

$$a(z - b)^2 + c(z - d)^2 = (a + c)(z - h)^2 + ac(b - d)^2/(a + c),$$

where

$$h = (ab + cd)/(a + c),$$

to the expression $\tau(n - \tau)(\delta - \hat{\delta}_\tau)^2/n + \sigma^2\delta^2/\sigma_\delta^2$ in (2.1.6) which, for $\sigma_\delta^2 \gg \sigma^2$, yields

$$\tau(n - \tau)(\delta - \hat{\delta}_\tau)^2/n + \sigma^2\hat{\delta}_\tau^2/\sigma_\delta^2.$$

Furthermore, the fact that $H(\tau) \geq 0$ leads to the inequality

$$\hat{\delta}_\tau^2 \leq n(n - 1)s_n^2/(\tau(n - \tau)), \tag{2.1.13}$$

where

$$s_n^2 = \sum_{i=1}^n (x_i - \bar{x}_n)^2/(n - 1). \tag{2.1.14}$$

Hence $\hat{\delta}_\tau^2$ is bounded from above for a given sample x_1, x_2, \dots, x_n . Thus, $(\sigma^2/\sigma_\delta^2)\hat{\delta}_\tau \rightarrow 0$ as $(\sigma^2/\sigma_\delta^2) \rightarrow 0$.

By integrating out σ in (2.1.12), the joint posterior probability function of τ and δ is

$$p(\tau, \delta | \mathbf{x}) \propto [Q(\tau, \delta)]^{-(n-1)/2} \quad 0 \leq \tau \leq n - 1, \quad -\infty < \delta < \infty, \tag{2.1.15}$$

where $Q(\tau, \delta)$ is defined in (2.1.8). A convenient way of computing (2.1.15) can be obtained by multiplying $p(\tau | \mathbf{x})$ and $p(\delta | \tau, \mathbf{x})$, which are given in (2.2.2) and (2.3.1), respectively.

2.2 The posterior distribution of τ

Integrating (2.1.15) with respect to δ results in the posterior probability function $p(\tau | \mathbf{x})$ of τ :

$$p(\tau | \mathbf{x}) \propto [n/(\tau(n - \tau))]^{1/2}[H(\tau)]^{-(n-2)/2}, \quad 0 \leq \tau \leq n - 1, \tag{2.2.1}$$

where $H(\tau)$ is defined in (2.1.10) or (2.1.11). An expression for $p(\tau | \mathbf{x})$ which is more convenient for computation and interpretation is given by

$$p(\tau | \mathbf{x}) \propto [n/(\tau(n - \tau))]^{1/2}[R(\tau)]^{-(n-2)/2}, \quad 0 \leq \tau \leq n - 1, \tag{2.2.2}$$

where

$$R(\tau) = H(\tau) / \sum_{i=1}^n (x_i - \bar{x}_n)^2, \\ = \left[\sum_{i=1}^{\tau} (x_i - \bar{x}_{\tau})^2 + \sum_{i=\tau+1}^n (x_i - \bar{x}_{n-\tau})^2 \right] \\ \div \sum_{i=1}^n (x_i - \bar{x}_n)^2. \tag{2.2.3}$$

Alternatively, one can write

$$R(\tau) = 1 - \tau(n - \tau)\hat{\delta}_{\tau}^2 / (n(n - 1)s_n^2), \tag{2.2.4}$$

where $\hat{\delta}_{\tau} = \bar{x}_{n-\tau} - \bar{x}_{\tau}$, as defined in (2.1.9). In general, the mode of $p(\tau|\mathbf{x})$ is the value of τ which gives rise to the minimal value of $R(\tau)$ or, equivalently, the maximum of $\tau(n - \tau)\hat{\delta}_{\tau}^2/n$, for all $\tau, \tau = 1, 2, \dots, n - 1$.

For large $n, p(\tau|\mathbf{x})$ of (2.2.2) can be approximated by

$$p(\tau|\mathbf{x}) \propto [n/(\tau(n - \tau))]^{1/2} \exp [-nR(\tau)/2], \\ 0 \leq \tau \leq n - 1. \tag{2.2.5}$$

This is obtained from (2.2.2) and repeated use of (2.2.4).

2.3 The posterior distribution of δ

From $p(\delta|\tau, \mathbf{x}) = p(\tau, \delta|\mathbf{x})/p(\tau|\mathbf{x})$, it can be shown that the conditional posterior distribution of δ , given τ is a t -distribution with mean $\hat{\delta}_{\tau}$, variance $S(\tau)/(n - 2)$ and $n - 2$ degrees of freedom. For large values of n , the conditional posterior distribution of δ , given τ approaches a normal distribution with mean $\hat{\delta}_{\tau}$ and variance $S(\tau)/(n - 2)$, where $S(\tau) = nH(\tau)/(\tau(n - \tau))$.

TABLE 1—The posterior probability of τ

m	x	p($\tau = m \mathbf{x}$)
1	1.38	0.0548
2	1.79	0.0733
3	0.49	0.0512
4	2.66	0.4531
5	-0.29	0.1217
6	0.06	0.0891
7	0.34	0.1118
8	-1.87	0.0449
9	0.54	--

The marginal posterior distribution of δ can be obtained from (2.2.2) as

$$p(\delta|\mathbf{x}) = \sum_{\tau=1}^{n-1} p(\delta|\tau, \mathbf{x})p(\tau|\mathbf{x}), \tag{2.3.1}$$

which is a weighted average of t -distributions.

3. APPLICATION TO ILLINOIS TRAFFIC DATA

For the annual data from 1962 to 1971 on traffic deaths in the State of Illinois, Srivastava and Sen [9] postulated the following model:

$$X_i = Y_{i+1} - Y_i = \mu_i + \epsilon_i,$$

where Y_i is the number (in hundreds) of traffic deaths for the i th year, and the ϵ_i 's are assumed to be independently and normally distributed with mean zero and a common unknown variance. To detect a possible change and its amount of change in μ_i , the posterior distributions of τ and δ , given $\mathbf{X} = \mathbf{x}$, are obtained using the results in Section 2. The results are listed in Tables 1 and 2. The posterior mode of τ is τ

TABLE 2—The Posterior Probability Density of δ

δ	p($\delta \mathbf{x}$)
-3.2	0.0698
-3.0	0.1026
-2.8	0.1489
-2.6	0.2106
-2.4	0.2852
-2.2	0.3637
-2.0	0.4300
-1.8	0.4671
-1.65	0.4702
-1.6	0.4663
-1.4	0.4316
-1.2	0.3758
-1.0	0.3129
-0.8	0.2526
-0.6	0.1999
-0.4	0.1562
-0.2	0.1212
0.0	0.0939
0.2	0.0729
0.4	0.0568
0.6	0.0446
0.8	0.0352
1.0	0.0279
1.2	0.0222
1.4	0.0179

$= 4$ with probability 0.453. That of δ is $\delta = -1.65$ with density 0.47. The joint posterior distribution of τ and δ is also obtained and the joint mode is also found to be $(\tau, \delta) = (4, -1.65)$ with joint density 0.2662.

REFERENCES

- [1] CHERNOFF, H. and ZACKS, S. (1964). Estimating the Current Mean of a Normal Distribution Which is Subject to Changes in Time, *Ann. Math. Statist.*, 35, 999-1018.
- [2] FEDER, P. I. (1975a). On Asymptotic Distribution Theory in Segmented Regression Problems—Identified Case, *Annals of Statistics*, 3, 49-83.
- [3] FEDER, P. I. (1975b). The Log Likelihood Ratio in Segmented Regression, *Annals of Statistics*, 3, 84-97.
- [4] FERREIRA, P. E. (1975). A Bayesian Analysis of a Switching Regression Model: Known Number of Regimes, *J. Amer. Statist. Assoc.*, 70, 370-374.
- [5] GARDNER, L. A. (1969). On Detecting Changes in the Mean of Normal Variates, *Ann. Math. Statist.*, 40, 116-126.
- [6] HINKLEY, D. V. (1970). Inference About the Change Point in a Sequence of Random Variables, *Biometrika*, 57, 1-17.
- [7] KANDER, Z. and ZACKS, S. (1966). Test Procedures for Possible Changes in Parameters of Statistical Distributions Occurring at Unknown Time Points, *Ann. Math. Statist.*, 37, 1196-1210.
- [8] SEN, A. and SRIVASTAVA, M. S. (1975). On Tests for Detecting Change in Mean, *Annals of Statistics*, 3, 98-108.
- [9] SRIVASTAVA, M. S. and SEN, A. (1975). Some One-Sided Tests for Change in Level, *Technometrics*, 17, 1, 61-64.