Qatar Univ. Sci. J. (1991), 11: 19 - 26

PREDICTION INTERVALS FOR FUTURE SAMPLE MEAN FROM INVERSE GAUSSIAN DISTRIBUTION

By

MUHAMMAD S. ABU-SALIH and RAFIQ K. AL-BATTAT
Department of Statistics, Yarmouk University, Irbid, Jordan

ABSTRACT

A random sample $X_1, X_2, \ldots, X_n$ from Inverse Gaussian distribution $I(\mu,\lambda)$ is observed. On the basis of this observed sample a 100β prediction interval of the mean Y of a future sample $Y_1, \ldots, Y_m$ from $I(\mu,\lambda)$ has been constructed when either or both μ & λ are unknown.

INTRODUCTION

A prediction interval is an interval which contains the results of a future sample from a population depending on the results of a past sample from the same population with a specified probability.

Prediction intervals play an important role in quality control and reliability. In statistics, they are being used in goodness of fit tests, hypothesis testing, classifying observations, sample surveys, ... etc. (Englehardt & Bain 1979). This topic has been discussed by many authors, of whom we mention a few who derived prediction intervals for derived prediction intervals for the future sample mean.

Lawless (1972) gave prediction limits for $\overline{Y}$ in the case of the exponential distribution.

Kaminsky and Nelson (1974) gave prediction interval for the mean of a future sample using susbsets of an observed sample when the samples are from the exponential distribution.

Hahn (1975) gave a prediction interval to contain the difference between the sample means of two future samples from normal distributions.

Meeker and Hahn (1980) gave prediction interval for the ratio of the means of two future samples from normal distributions.

Azzam and Awad (1981) gave prediction intervals for the difference and ratio of the means of two future samples under the assumption that the samples are from a normal and an exponential distribution.

Vee-Ming (1984) constructed prediction intervals for the mean life time of a future sample based on incomplete data taken from a 2-parameters exponential distribution.

In this work we derive prediction intervals for the future sample mean from the inverse Gaussian distribution. Chhiekara and Guttman (1982) used an observed sample from the inverse Gaussian distribution to construct a prediction interval to contain the next future observation.

The probability density function (pdf) f(.) of an inverse Gaussian distribution denoted by $I(\mu,\lambda)$ is given by

$$f(x,\mu,\lambda) = \begin{cases} (\lambda/2\pi x^3)^{1/2} \exp(-\lambda(x-\mu)^2/2\mu^2 x) & x>0,\ \mu,\lambda>0 \\ 0 & \text{o.w.} \end{cases} \tag{1.1}$$

where λ is shape parameter (Tweedi, 1957). The pdf is skewed unimodal and is a member of the exponential family. Inverse Gaussian distribution is used as a life time model (Chhikara & Folks, 1977).

It is known that if $X \sim I(\mu,\lambda)$, then $E(X)=\mu$, and $Var(X)=\mu^3/\lambda$.

PREDICTION INTERVAL FOR THE FUTURE SAMPLE MEAN

Let $X_1, \ldots, X_n$ be a random sample from an inverse Gaussian distribution $I(\mu,\lambda)$ with density function $f(x,\mu,\lambda)$, as given in (1.1).

Let $Y_1, \ldots, Y_m$ be a future random sample from the same population with the same parameters μ and λ. Assume that the two samples are independent of one another and that both μ and λ are unknown parameters.

Shuster (1968) showed that

$$\frac{\lambda(X-\mu)^2}{\mu^2 X} \sim \chi_1^2 \tag{2.1}$$

So $$\frac{\lambda}{\mu^2}\left[\sum_{i=1}^{n}\frac{(X_i-\mu)^2}{X_i} + \sum_{j=1}^{m}\frac{(Y_j-\mu)^2}{Y_j}\right] \sim \chi^2_{n+m}$$

Now, $$\frac{\lambda}{\mu^2} \sum_{i=1}^{n} \frac{(X_i - \mu)^2}{X_i} = \lambda \sum_{i=1}^{n} \left(\frac{1}{X_i} - \frac{1}{\overline{X}}\right) + \frac{n\lambda(\overline{X}-\mu)^2}{\mu^2 \overline{X}} \qquad (2.2)$$

Also, $$\frac{\lambda}{\mu^2} \sum_{j=1}^{m} \frac{(Y_j - \mu)^2}{Y_j} = \lambda \sum_{j=1}^{m} \left(\frac{1}{Y_i} - \frac{1}{\overline{Y}}\right) + \frac{m\lambda(\overline{Y}-\mu)^2}{\mu^2 \overline{Y}} \qquad (2.3)$$

where

$$\overline{X} = \sum_{i=1}^{n} X_i/n \ , \ \overline{Y} = \sum_{j=1}^{m} Y_i/m$$

It may be seen that $\overline{X}$ and $\sum_{i=1}^{n} (1/X_i - 1/\overline{X})$ are the maximum likelihood estimators of μ and λ respectively and the two statistics are independent, (Tweedi, 1957),

Furthermore, $X \sim I(\mu, n\lambda)$ and $\lambda \sum_{i=1}^{n} (1/X_i - 1/\overline{X}) \sim \chi^2_{n-1}$ (Wasan, and Roy, 1969).

Clearly $\frac{n\lambda(\overline{X} - \mu)^2}{\mu^2 \overline{X}} \sim \chi^2_1$ in view of (2.1).

Thus from (2.2) and (2.3)

$$\frac{\lambda}{X_i} \left[\sum_{i=1}^{n} \frac{(X_i - \mu)^2}{X_i} + \sum_{j=1}^{m} \frac{(Y_j - \mu)^2}{Y_j} \right] = Q_1 + Q_2 + Q_3 \qquad (2.4)$$

where

$$Q_1 = \lambda \sum_{i=1}^{n} (1/X_i - 1/\overline{X}) + \sum_{j=1}^{m} (1/Y_i - 1/\overline{Y}) \qquad (2.5)$$

$$Q_2 = \frac{n\lambda(X - \mu)^2}{\mu^2 \overline{X}} \text{ and } Q_3 = \frac{m\lambda(Y - \mu)^2}{\mu^2 \overline{Y}}$$

are independently distributed as χ^2_{n+m-2}, χ^2_1 and χ^2_1 respectively (Hogg & Craig, 1978).

Now, Q_2 and Q_3 can be combined differently so that the right side of (2.4) equals $Q_1 + A + B$.

where

$$A = \frac{nm\lambda(\overline{X} - \overline{Y})^2}{\overline{X}\overline{Y}(n\overline{X} + m\overline{Y})} \tag{2.6}$$

and

$$B = \frac{\lambda[(n\overline{X} + m\overline{Y}) - \mu(n + m)]^2}{\mu^2\ (n\overline{X} + m\overline{Y})} \tag{2.7}$$

In view of (2.1), it can be seen that B has χ^2_1.

Therefore, A, B given in (2.6) and (2.7) are independently distributed each as χ^2_1 (Hogg & Craig, 1978, p. 279). Notice that Q_1 and A do not involve the parameter μ and can be used to obtain 100β percent prediction intervals for $\overline{Y}$ when μ is unknown and λ is known. In this case we use the statistic

$$A = nm\lambda(\overline{X} - \overline{Y}) / \overline{X}\overline{Y}(n\overline{X} + m\overline{Y}) \tag{2.8}$$

for the prediction of $\overline{Y}$.

Since $A \sim \chi^2_1$ the 100β% prediction interval for $\overline{Y}$ will be obtained by solving $P[A \leqslant \chi^2_{1},\beta] = 1 - \beta$ for $\overline{Y}$. The prediction interval for $\overline{Y}$ when λ is known will be given by:

$$\left[\left[\frac{1}{\overline{\overline{X}}} + \frac{1}{2m\lambda}\chi^2_{1},\beta\right] \mp \left[\frac{(n+m)}{nm\lambda\overline{\overline{X}}}\chi^2_{1},\beta + \frac{1}{4m^2\lambda^2}(\chi^2_{1,\beta})^2\right]\right]^{-1} \tag{2.9}$$

When μ and λ are unknown, consider the statistic

$R = \dfrac{(n + m - 2)\ A}{Q_1}$, so that

substituting for A and Q_1, we get:

$$R = (n + m - 2)\,(\overline{X} - \overline{Y})^2 / \overline{X}\overline{Y}(n\overline{X} + m\overline{Y})\,V \tag{2.10}$$

where $\quad V = Q_1 / nm$ (2.11)

R does not depend on any parameter μ or λ and has F distribution with 1 and (n+m−2) degrees of freedom. Hence, R can be used to construct prediction intervals for $\overline{Y}$ when both μ and λ are unknown.

The 100β prediction interval for $\overline{Y}$ will be obtained by inverting the inequality

$R \leqslant F_{1,n+m-,\beta}$ where $F_{1,n+m-,\beta}$ is the 100β% point of the $F_{1,n+m-2}$ distribution.

Using (2.10) and simplifying we get the 100β% prediction interval for $\overline{Y}$ to be:

$$\left[\frac{1}{\overline{X}} + \frac{(F_{1,n+m-2,\beta})\, nV}{2(n+m-2)} \mp \left\{\frac{(nVF_{1,n+m-2,\beta})^2}{4(n+m-2)^2} + \frac{V(n+m)\, F_{1,n+m-2,\beta}}{(n+m-2)\overline{X}}\right\}\right]^{-1} \quad (2.12)$$

When μ is known, and λ is unknown, consider the following statistic

$$T = \frac{mn(\overline{Y} - \mu)^2}{\overline{Y}\, Q}$$

where

$$Q = \sum_{i=1}^{n} (X_i - \mu)^2 / X_i$$

It is easy to see that T is distributed as $F_{1,n}$. Again the 100β% prediction interval for $\overline{Y}$ (for μ known) will be obtained by solving $T \leq F_{1,n,\beta}$ and is given by

$$\left[\mu + \frac{Q}{2mn} F_{1,n,\beta}\right] \mp \frac{Q}{2mn}\left[\frac{4mn\mu}{Q} F_{1,n,\beta} + F^2_{1,n,\beta}\right]^{1/2} \quad (2.13)$$

Note that A, R and T do not always provide two sided intervals because there is a possibility that the difference in the two terms of (2.9), (2.12) and (2.13) may be negative. Since $y > 0$ it can be seen that only the one-sided intervals with ∞ as the upper limit are admissible and these are obtained by restricting the solution of an inequality to the positive real line.

Example:

Consider the data given by Chhikara and Folks (1977) with $n = 46$, $\hat{\mu} = \overline{x} = 3.61$ and $\lambda = 1.66$, $V = \hat{\lambda}^{-1} = 0.66$. Based on these data a 100β percent prediction limits using (2.9) and (2.12) were computed for various values of β. These limits are given in Table 1 and Table 2 respectively.

Table 1

Prediction intervals for varying λ for $m = n = 46$, $x = 3.61$ (formula (2.9))

	$\lambda = 1.66$		$\lambda = 2.0$	
β	Lower	Upper	Lower	Upper
0.900	2.29	6.515	2.38	6.127
0.950	2.12	7.500	2.22	6.927
0.975	1.84	10.332	1.94	9.095
0.990	1.74	11.904	1.85	10.221

β	λ = 2.5 Lower	λ = 2.5 Upper	λ = 3.0 Lower	λ = 3.0 Upper
0.900	2.48	5.747	2.55	5.491
0.950	2.32	6.383	2.40	6.025
0.975	2.05	8.016	2.15	7.349
0.990	1.96	8.814	2.06	7.973

β	λ = 3.5 Lower	λ = 3.5 Upper	λ = 4.0 Lower	λ = 4.0 Upper
0.900	2.62	5.304	2.67	5.160
0.950	2.47	5.769	2.53	5.574
0.975	2.22	6.891	229	6.555
0.990	2.14	7.407	2.20	6.997

Comment:

As we can see from the results given above that the length of the interval becomes smaller as λ increases, which is a desirable property.

Table 2

Prediction intervals for varying m for n = 46, x = 3.61 (formula 2.12)

β	m = 4 Lower	m = 4 Upper	m = 16 Lower	m = 16 Upper
0.900	0.46	57.02	0.54	∞
0.950	0.45	138.48	0.41	∞
0.975	0.28	1416.28	0.33	∞
0.990	0.21	∞	0.26	∞

β	m = 36 Lower	m = 36 Upper	m = 76 Lower	m = 76 Upper
0.900	0.63	∞	0.74	∞
0.950	0.49	∞	0.60	∞
0.975	0.40	∞	0.51	∞
0.990	0.32	∞	0.42	∞

Comment:

It can be seen from the above table that as m increases the upper limit goes to ∞ and the lower limit increases.

If we compare the result of Table 1 and 2, we infer that we get shorter prediction intervals when λ is known, rather than when λ is unknown.

REFERENCES

Azzam, M.M. and **A.M. Awad, 1981.** Prediction intervals for a future sample mean and for the difference or ratio of two future sample means. *Dirasat,* 8: 21-27.

Chhikara, R.S. and **I. Guttman, 1982.** Prediction limits for the inverse Gaussian distribution. *Technometrics,* 24: 319-324.

Engelhardt, M. and **L.J. Bain, 1979.** Prediction limits and two-sample problems with complete or censored Weibull data. *Technometrics,* 21: 233-237.

Hahn, G.J. 1975. A simultaneous prediction limit on the mean of future samples from an exponential distribution. *Technometrics,* 17: 341-344.

Hogg, R.V. and **A.T. Craig, 1978.** *Introduction to Mathematical Statistics,* Fourth Ed. MacMillan Publishing Co. Inc. New York.

Kaminsky, K.S. and **P.I. Nelson, 1974.** Prediction intervals for the exponential distribution using subsets of the data. *Technometrics,* 16: 57-59.

Lawless, J.F. 1972. On prediction intervals for samples from the exponential distribution and prediction limits for system survival. *Sankhya,* Ser B, 34: 1-12.

Meeker, J.V. and **G.J. Hahn, 1980.** Prediction interval for the ratio of normal distribution samples mean. *Technometrics,* 22: 357-366.

Shuster, J.J. 1968. On the inverse Gaussian distribution function. *J. Amer. Statist. Ass.,* 63: 1514-1516.

Tweedie, M.C.K. 1957. Statistical Properties of Inverse Gaussian Distribution *I. Ann. Math. Statist.* 28: 362-377.

Wasan, M.T. and **L.K. Roy,** (1969). Tables of Inverse Gaussian Percentage Points. *Technometrics.* 11: 591-604.

فترات التنبؤ لوسط عينة مستقبلية
من توزيع جـــاوس العكسي

محمد أبو صـــالح و رفيـــق البطـــاط

يتناول هذا البحث مسألة إيجاد فترة تنبؤية للوسط $\overline{Y}$ لعينة مستقبلية $Y_1,\ldots,Y_n$ مأخوذة من توزيع جاوس العكسي $I(\mu,\lambda)$ في حالة عدم معرفة أي من أو كلا المعلمتين μ و λ وذلك استناداً على عينة سابقة $Y_1,\ldots,Y_n$ من نفس التوزيع .