วิธีการประมาณค่าฐานนิยมของข้อมูลที่มีการแจกแจงอินเวอร์สเกาส์เซียน
The Mode Estimation Method of Inverse Gaussian Data

วราฤทธิ์ พานิชกิจโกศลกุล[1]*

[1]ผู้ช่วยศาสตราจารย์ ภาควิชาคณิตศาสตร์และสถิติ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์ จ.ปทุมธานี 12121

บทคัดย่อ

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อเสนอวิธีการประมาณค่าฐานนิยมของข้อมูลที่มีการแจกแจงอินเวอร์สเกาส์เซียน และเปรียบเทียบวิธีการประมาณค่าฐานนิยม 3 วิธี คือ 1) วิธีอย่างง่าย 2) วิธีปรับค่าฐานนิยมด้วยเทอมค่าคงที่ (n+1)/n และ 3) วิธีปรับค่าฐานนิยมด้วยเทอมค่าคงที่ (n+2)/n โดยการเปรียบเทียบค่าสัมบูรณ์ของความเอนเอียง (|Bias|) และค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (Mean Square Error: MSE) ของค่าประมาณ กำหนดขนาดตัวอย่าง เท่ากับ 5, 10, 20, 30 และ 50 กำหนดค่าพารามิเตอร์ μ เท่ากับ 1 และพารามิเตอร์ λ เท่ากับ 1, 3, 5, 10, 15 และ 20 การวิจัยครั้งนี้ใช้วิธีการจำลองแบบมอนติคาร์โล และทำการทดลองซ้ำๆ กัน 50,000 ครั้งในแต่ละสถานการณ์ ผลการวิจัยสรุปได้ดังนี้ วิธีปรับค่าฐานนิยมด้วยเทอมค่าคงที่ (n+2)/n ให้ค่า |Bias| และค่า MSE ต่ำที่สุด เกือบทุกกรณีที่ศึกษา รองลงมาคือ วิธีปรับค่าฐานนิยมด้วยเทอมค่าคงที่ (n+1)/n และวิธีอย่างง่าย ตามลำดับ

Abstract

The objectives of this research are to propose the mode estimation method of Inverse Gaussian data and to compare three mode estimation methods. Those methods are simple method, adjusted mode method with term (n+1)/n, and adjusted mode method with term (n+2)/n. The performance of each method can be measured by the absolute bias (|Bias|) and the mean square errors (MSE). This study was performed by using different sample sizes (n) of 5, 10, 20, 30, and 50 whereas the parameter μ is fixed to be 1 and parameters λ are fixed to be 1, 3, 5, 10, 15, and 20. This research used the Monte Carlo Simulation technique. The experiment was repeated 50,000 times for each condition. The results showed that the adjusted mode method with term (n+2)/n gave the lowest value of |Bias| and MSE in most of the criterions of sample sizes and parameter values.

คำสำคัญ : การประมาณค่าฐานนิยม การแจกแจงอินเวอร์สเกาส์เซียน

Key words : The mode estimation, Inverse Gaussian distribution

* *ผู้นิพนธ์ประสานงาน ไปรษณีย์อิเล็กทรอนิกส์ : wararit@mathstat.sci.tu.ac.th โทร. 0-2564-4440-59 ต่อ 2101-3*

1. บทนำ

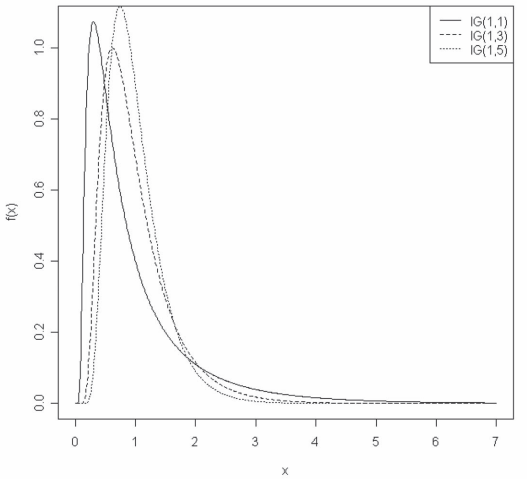
การแจกแจงอินเวอร์สเกาส์เซียน (Inverse Gaussian distribution) เป็นการแจกแจงหนึ่งที่นำมาประยุกต์ใช้กับปัญหาต่างๆ เช่น การวิเคราะห์ราคาหุ้น ปัญหาด้านชีววิทยา อุทกภูมิศาสตร์ อุตุนิยมวิทยา และความเชื่อถือได้ของผลิตภัณฑ์ เป็นต้น (Cohen and Whitten, 1988) โดยมีนักวิจัยหลายท่านที่สนใจศึกษาเกี่ยวกับการแจกแจงดังกล่าว อาทิ Halphen, Tweedie, Wald และ Wasan, Chikara, Folks, Seshadri, Cohen และ Jorgenson เป็นต้น (Bala-krisnan and Chen, 1997)

ฟังก์ชันหนาแน่นของความน่าจะเป็น (Probability density function) ของการแจกแจงอินเวอร์สเกาส์เซียน เขียนได้ดังนี้ (Krishnamoorthy, 2006)

$$f(x/\mu,\lambda) = \sqrt{\frac{\lambda}{2\pi x^3}}\exp\left(\frac{-\lambda(x-\mu)^2}{2\mu^2 x}\right)$$

โดยที่ $x>0$, $\mu,\lambda>0$

กราฟของฟังก์ชันความหนาแน่นของความน่าจะเป็นแสดงดังรูปที่ 1



รูปที่ 1 ฟังก์ชันความหนาแน่นของการแจกแจงอินเวอร์สเกาส์เซียน เมื่อ (μ = 1 และ λ = 1, 3 และ 5

สมบัติที่สำคัญของการแจกแจงอินเวอร์สเกาส์-เซียน มีดังนี้

ค่าเฉลี่ย μ

ความแปรปรวน μ^3/λ

ฐานนิยม $\mu\left[\sqrt{1+\frac{9\mu^2}{4\lambda^2}}-\frac{3\mu}{2\lambda}\right]$

วิธีการประมาณค่าพารามิเตอร์ μ และ λ สามารถประมาณค่าได้หลายวิธี เช่น วิธีความควรจะเป็นสูงสุด (Maximum Likelihood method: ML) และวิธีการประมาณแบบไม่เอนเอียงซึ่งมีความแปรปรวนต่ำสุด (Minimum Variance Unbiased Estimation method: MVUE) เป็นต้น (Evans, Hastings, and Peacock, 2000) เมื่อได้ค่าประมาณพารามิเตอร์ $\hat{\mu}$ และ $\hat{\lambda}$ แล้ว ขั้นตอนต่อไปจะนำค่าประมาณพารามิเตอร์ไปคำนวณค่าสถิติต่างๆ เช่น ค่าเฉลี่ย ความแปรปรวน หรือฐานนิยม เป็นต้น วิธีดังกล่าวเรียกว่า วิธีอย่างง่าย (Simple method) ผู้วิจัยจึงมีแนวคิดที่จะนำเทอมค่าคงที่มาปรับให้ค่าประมาณของฐานนิยมมีความคลาดเคลื่อนน้อยลง โดยทดลองกับเทอมค่าคงที่ในรูปแบบต่างๆ

ดังนั้นในการวิจัยครั้งนี้จึงเสนอวิธีการประมาณค่าฐานนิยมของข้อมูลที่มีการแจกแจงอินเวอร์สเกาส์เซียน และเปรียบเทียบวิธีการประมาณค่าฐานนิยม 3 วิธี คือ 1) วิธีอย่างง่าย 2) วิธีปรับค่าฐานนิยมด้วยเทอมค่าคงที่ (n+1)/n และ 3) วิธีปรับค่าฐานนิยมด้วยเทอมค่าคงที่ (n+2)/n โดยการเปรียบเทียบค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (Mean Square Error: MSE) และค่าสัมบูรณ์ของความเอนเอียง (Absolute Bias : |Bias|) ของค่าประมาณ

สำหรับขอบเขตของการวิจัยครั้งนี้มีดังต่อไปนี้

1) กำหนดขนาดตัวอย่างที่ใช้ (n) เป็นเท่ากับ 5, 10, 20, 30 และ 50

2) กำหนดค่าพารามิเตอร์ μ เท่ากับ 1

3) กำหนดค่าพารามิเตอร์ λ เท่ากับ 1, 3, 5, 10, 15 และ 20

4) ประมาณค่าพารามิเตอร์ μ และ λ ด้วยวิธีความควรจะเป็นสูงสุด (Maximum Likelihood method)

การวิจัยครั้งนี้ใช้วิธีการจำลองแบบการทดลองด้วยเทคนิคมอนติคาร์โลด้วยเครื่องไมโครคอมพิวเตอร์ โดยใช้โปรแกรม R เวอร์ชัน 2.5.0 และกำหนดให้เครื่องคอมพิวเตอร์ทำงานซ้ำๆ กัน 50,000 ครั้งในแต่ละสถานการณ์

2. วิธีการศึกษา

ในการวิจัยครั้งนี้มีวิธีการศึกษาดังนี้

2.1 การจำลองข้อมูลที่ใช้ในการวิจัย

จำลองข้อมูลที่ใช้โดยการสร้างตัวแปรสุ่ม X ซึ่งมีการแจกแจงอินเวอร์สเกาส์เซียนด้วยพารามิเตอร์ μ และ λ ตามที่กำหนดไว้ในขอบเขตของการวิจัย

2.2 ประมาณค่าพารามิเตอร์ μ และ λ ด้วยวิธีความควรจะเป็นสูงสุด

เมื่อสร้างตัวแปรสุ่ม X ที่มีการแจกแจงอินเวอร์สเกาส์เซียนแล้ว ประมาณค่าพารามิเตอร์ μ และ λ ด้วยวิธีความควรจะเป็นสูงสุด (Evans, Hastings, and Peacock, 2000)

ตัวประมาณความควรจะเป็นสูงสุดของ μ คือ

$$\hat{\mu} = \bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

ตัวประมาณความควรจะเป็นสูงสุดของ λ คือ

$$\hat{\lambda} = \frac{n}{\left(\sum_{i=1}^{n} X_i^{-1} - (\bar{X})^{-1}\right)}$$

2.3 ประมาณค่าฐานนิยมของข้อมูลด้วยวิธีการที่ศึกษา

เมื่อประมาณค่าพารามิเตอร์ μ และ λ นำค่าประมาณ $\hat{\mu}$ และ $\hat{\lambda}$ ที่ได้ ประมาณค่าฐานนิยม ด้วยวิธีการที่ศึกษา

1) วิธีอย่างง่าย (วิธี S) ตัวประมาณค่าฐานนิยม คือ

$$\hat{M}_S = \hat{\mu}\left[\sqrt{1+\frac{9\hat{\mu}^2}{4\hat{\lambda}^2}} - \frac{3\hat{\mu}}{2\hat{\lambda}}\right]$$

2) วิธีปรับค่าฐานนิยมด้วยเทอมค่าคงที่ (n+1)/n (วิธี C1) ตัวประมาณค่าฐานนิยม ($\hat{M}_{C1}$) คือ

$$\hat{M}_{C1} = \left(\frac{n+1}{n}\right)\hat{\mu}\left[\sqrt{1+\frac{9\hat{\mu}^2}{4\hat{\lambda}^2}} - \frac{3\hat{\mu}}{2\hat{\lambda}}\right]$$

3) วิธีปรับค่าฐานนิยมด้วยเทอมค่าคงที่ (n+2)/n (วิธี C2) ตัวประมาณค่าฐานนิยม ($\hat{M}_{C2}$) คือ

$$\hat{M}_{C2} = \left(\frac{n+2}{n}\right)\hat{\mu}\left[\sqrt{1+\frac{9\hat{\mu}^2}{4\hat{\lambda}^2}} - \frac{3\hat{\mu}}{2\hat{\lambda}}\right]$$

2.4 คำนวณค่าสัมบูรณ์ของความเอนเอียงและค่าความคลาดเคลื่อนกำลังสองเฉลี่ยของค่าประมาณของแต่ละวิธี แล้วทำการเปรียบเทียบ

ค่าสัมบูรณ์ของความเอนเอียง ($|\text{Bias}|$) ของค่าประมาณ คำนวณจาก

$$|\text{Bias}| = \left| \frac{\sum_{i=1}^{50,000} \hat{M}_{ki}}{50,000} - M \right|$$

ค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) ของค่าประมาณ คำนวณจาก

$$\text{MSE} = \frac{\sum_{i=1}^{50,000} (M - \hat{M}_{ki})^2}{50,000}$$

เมื่อ M แทน ค่าฐานนิยมของประชากร

$\hat{M}_{ki}$ แทน ค่าประมาณของฐานนิยม ซึ่งคำนวณจากวิธีที่ k ในการทำซ้ำครั้งที่ i

k แทน วิธีการประมาณค่าฐานนิยม

3. ผลการศึกษา

ผลการเปรียบเทียบค่าสัมบูรณ์ของความเอนเอียง และค่าความคลาดเคลื่อนกำลังสองเฉลี่ยของค่าประมาณ นำเสนอดังตารางที่ 1 สรุปรายละเอียดดังนี้

3.1 การเปรียบเทียบค่าสัมบูรณ์ของความเอนเอียง

ในทุกระดับของขนาดตัวอย่าง (n = 5, 10, 20, 30 และ 50) และทุกระดับของค่าพารามิเตอร์ (λ = 1, 3, 5, 10, 15 และ 20) วิธีปรับค่าฐานนิยมด้วยเทอมค่าคงที่ (n+2)/n ให้ |Bias| ต่ำที่สุด รองลงมาคือ วิธีปรับค่าฐานนิยมด้วยเทอมค่าคงที่ (n+1)/n และวิธีอย่างง่าย ตามลำดับ

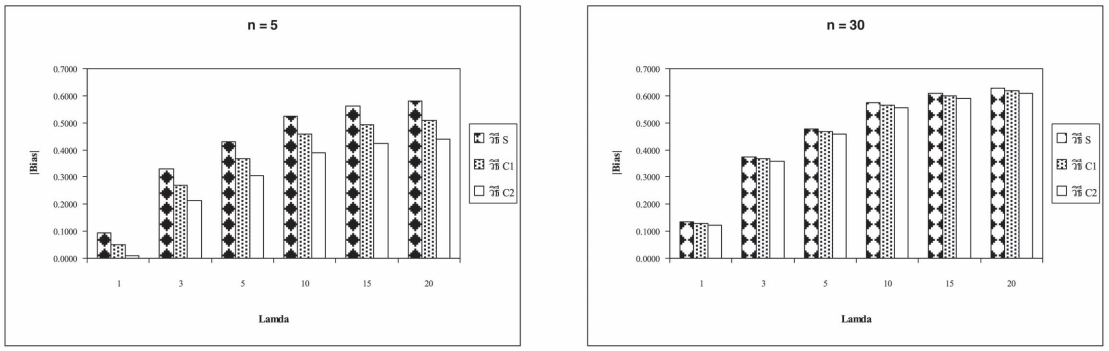
3.2 การเปรียบเทียบค่าความคลาดเคลื่อนกำลังสองเฉลี่ย

ในเกือบทุกระดับของขนาดตัวอย่าง และเกือบทุกระดับของค่าพารามิเตอร์ วิธีปรับค่าฐานนิยมด้วยเทอมค่าคงที่ (n+2)/n ให้ค่า MSE ต่ำที่สุด ยกเว้นในกรณีที่ขนาดตัวอย่างเท่ากับ 5 และพารามิเตอร์ λ เท่ากับ 1 วิธีปรับค่าฐานนิยมด้วยเทอมค่าคงที่ (n+1)/n ให้ค่า MSE ต่ำที่สุด

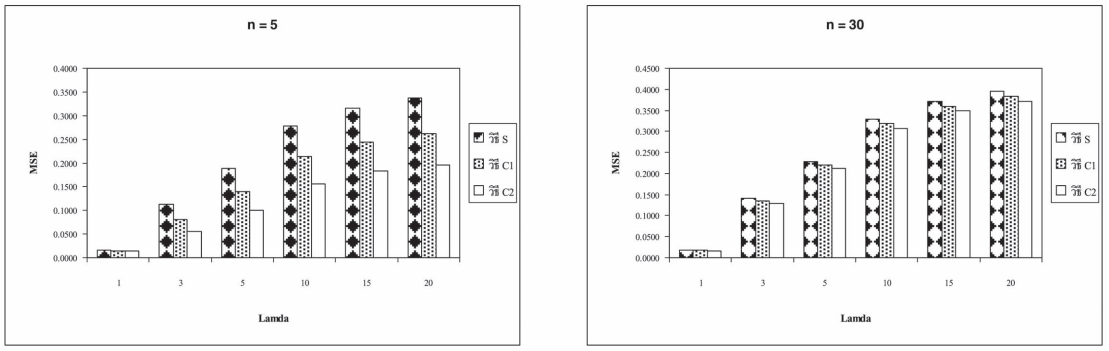
ตารางที่ 1 ค่าสัมบูรณ์ของความเอนเอียง (|Bias|) และค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) ของค่าประมาณฐานนิยมของข้อมูลที่มีการแจกแจงอินเวอร์สเกาส์เซียน เมื่อค่า μ เท่ากับ 1

n	λ	|Bias|			MSE		
		วิธี S	วิธี C1	วิธี C2	วิธี S	วิธี C1	วิธี C2
5	1	0.0933	0.0515	0.0096*	0.0164	0.0137*	0.0151
	3	0.3290	0.2712	0.2134*	0.1137	0.0814	0.0562*
	5	0.4294	0.3665	0.3036*	0.1883	0.1399	0.0998*
	10	0.5248	0.4575	0.3903*	0.2777	0.2126	0.1567*
	15	0.5611	0.4923	0.4235*	0.3163	0.2446	0.1824*
	20	0.5793	0.5096	0.4399*	0.3368	0.2614	0.1958*
10	1	0.1197	0.1013	0.0830*	0.0170	0.0136	0.0108*
	3	0.3579	0.3319	0.3059*	0.1302	0.1127	0.0966*
	5	0.4593	0.4309	0.4024*	0.2126	0.1876	0.1642*
	10	0.5559	0.5253	0.4948*	0.3099	0.2771	0.2462*
	15	0.5918	0.5605	0.5291*	0.3508	0.3149	0.2809*
	20	0.6104	0.5787	0.5470*	0.3731	0.3355	0.2999*
20	1	0.1309	0.1223	0.1137*	0.0183	0.0162	0.0143*
	3	0.3704	0.3580	0.3456*	0.1381	0.1292	0.1206*
	5	0.4729	0.4593	0.4457*	0.2243	0.2117	0.1995*
	10	0.5693	0.5547	0.5401*	0.3245	0.3081	0.2922*
	15	0.6054	0.5905	0.5755*	0.3668	0.3490	0.3315*
	20	0.6244	0.6092	0.5941*	0.3901	0.3714	0.3532*
30	1	0.1341	0.1285	0.1229*	0.0187	0.0173	0.0159*
	3	0.3743	0.3662	0.3581*	0.1407	0.1347	0.1289*
	5	0.4769	0.4680	0.4591*	0.2279	0.2195	0.2113*
	10	0.5736	0.5640	0.5545*	0.3293	0.3184	0.3077*
	15	0.6098	0.5999	0.5901*	0.3720	0.3601	0.3484*
	20	0.6288	0.6189	0.6089*	0.3956	0.3832	0.3709*
50	1	0.1368	0.1335	0.1302*	0.0191	0.0182	0.0174*
	3	0.3774	0.3726	0.3678*	0.1428	0.1392	0.1357*
	5	0.4803	0.4750	0.4697*	0.2309	0.2259	0.2209*
	10	0.5770	0.5713	0.5656*	0.3331	0.3265	0.3201*
	15	0.6133	0.6075	0.6016*	0.3763	0.3691	0.3621*
	20	0.6321	0.6262	0.6203*	0.3996	0.3922	0.3848*

* หมายถึง วิธีการประมาณค่าที่ให้ค่า |Bias| ต่ำที่สุด หรือค่า MSE ต่ำที่สุด



รูปที่ 2 ค่าสัมบูรณ์ของความเอนเอียงของค่าประมาณฐานนิยม เมื่อขนาดตัวอย่างเท่ากับ 5 และ 30



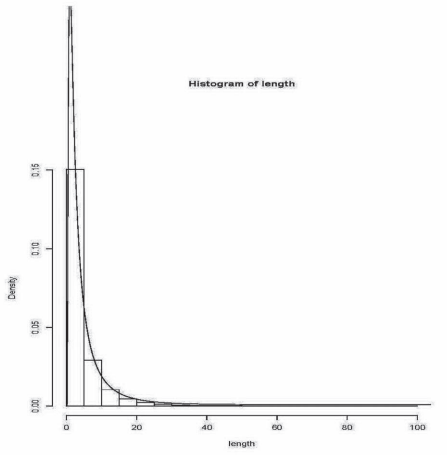
รูปที่ 3 ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยของค่าประมาณฐานนิยม เมื่อขนาดตัวอย่างเท่ากับ 5 และ 30

4. การประยุกต์ใช้ผลการศึกษา

การประมาณค่าฐานนิยมของข้อมูลที่มีการแจกแจงอินเวอร์สเกาส์เซียน สามารถนำไปประยุกต์ใช้ได้ดังนี้

ในการศึกษาเกี่ยวกับระยะเวลาที่ผู้ใช้เข้าเว็บไซต์ msbn.com แต่ละครั้ง โดยเก็บข้อมูลในวันที่ 28 กันยายน 2542 ซึ่งมีผู้ใช้บริการ 989,818 ราย (Sanchez and He, 2003) ระยะเวลาที่ผู้ใช้เข้าเว็บไซต์แสดงดังรูปที่ 4

จากข้อมูลประมาณค่าพารามิเตอร์ของการแจกแจงอินเวอร์สเกาส์เซียน ดังนี้



รูปที่ 4 ระยะเวลาที่ผู้ใช้เข้าเว็บไซต์ msbn.com

$$\hat{\mu} = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{4,698,794}{989,818} = 4.747129$$

$$\hat{\lambda} = \frac{n}{\left(\sum_{i=1}^{n} X_i^{-1} - (\overline{X})^{-1}\right)} = \frac{989,818}{321,169.6} = 3.081917$$

ขั้นต่อไปทำการประมาณค่าฐานนิยม เพื่อศึกษาว่าโดยส่วนใหญ่แล้วผู้ใช้เข้าเว็บไซต์เป็นระยะเวลาเท่าใด ดังนี้

$$\hat{M} = \left(\frac{n+2}{n}\right)\hat{\mu}\left[\sqrt{1+\frac{9\hat{\mu}^2}{4\hat{\lambda}^2}} - \frac{3\hat{\mu}}{2\hat{\lambda}}\right]$$

$$= \left(\frac{989,818+2}{989,818}\right)4.747129\left[\sqrt{1+\frac{9(4.747129)^2}{4(3.081917)^2}} - \frac{3(4.747129)}{2(3.081917)}\right]$$

= 0.983237 นาทีต่อครั้ง

5. สรุปผล

ผลการศึกษาครั้งนี้ สรุปได้ดังนี้

วิธีปรับค่าฐานนิยมด้วยเทอมค่าคงที่ (n+2)/n ให้ค่าสัมบูรณ์ของความเอนเอียง (|Bias|) และค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) ต่ำที่สุดเกือบทุกกรณีที่ศึกษา รองลงมาคือ วิธีปรับค่าฐานนิยมด้วยเทอมค่าคงที่ (n+1)/n และวิธีอย่างง่าย ตามลำดับ

ดังนั้นในการประมาณค่าฐานนิยมของข้อมูลที่มีการแจกแจงอินเวอร์สเกาส์เซียน ควรใช้วิธีปรับค่าฐานนิยมด้วยเทอมค่าคงที่ (n+2)/n ซึ่งมีสูตรดังนี้

$$\hat{M} = \left(\frac{n+2}{n}\right)\hat{\mu}\left[\sqrt{1+\frac{9\hat{\mu}^2}{4\hat{\lambda}^2}} - \frac{3\hat{\mu}}{2\hat{\lambda}}\right]$$

โดยที่ $\hat{\mu} = \overline{X}$

$$\hat{\lambda} = \frac{n}{\left(\sum_{i=1}^{n} X_i^{-1} - (\overline{X})^{-1}\right)}$$

6. เอกสารอ้างอิง

Bala-Krisnan, N. and Chen, W.S. 1997. **Tables for Order Statistics from Inverse Gaussian Distributions with Applications.** Florida: CRC Press.

Cohen, A.C. and Whitten, B.J. 1988. **Parameter Estimation in Reliability and Life Span Models.** New York: Marcel Dekker.

Evans, M., Hastings, N., and Peacock, B. 2000. **Statistical Distributions.** New York: John Wiley & Sons.

Krishnamoorthy, K. 2006. **Handbook of Statistical Distributions with Applications.** Florida: Chapman & Hall.

Sanchez, J. and He, Y. 2003. **Internet Data Analysis for the Undergraduate Statistics Curriculum.** Retrieved September 25, 2007. from http://preprints.stat.ucla.edu/370/ams2004.pdf