Transformasi :Jurnal Pendidikan Matematika dan Matematika
Volume 6, No. 1, Bulan Juni Tahun 2022, pp. 75-85
ISSN 2549-1164 (online)
DOI: 10.36526/tr.v%vi%i.1934
Available online at https://ejournal.unibabwi.ac.id/index.php/transformasi

APLIKASI METODE MAXIMUM LIKELIHOOD ESTIMATION PADA DATA BINOMIAL INTERVAL-TERSENSOR

Bernadhita H. S. Utami[1], Dwi Herinanto[2], Miswan Gumanti[3]
[1,2,3] Program Studi Sistem Informasi, STMIK Pringsewu
email korespondensi**:** bernadhita.herindri.s@mail.ugm.ac.id

Diterima : 03-05-2022, ***Revisi***: 19-06-2022, ***Diterbitkan***: 23-06-2022

ABSTRAK

Penelitian ini bertujuan untuk menentukan estimasi data interval-tersensor dengan distribusi binomial. Penelitian ini mengestimasi parameter pada distribusi binomial interval-tersensor menggunakan metode *maximum likelihood estimation* dan menunjukkan sifat-sifat estimator pada distribusi binomial interval-tersensor. Hasil penelitian menunjukkan bahwa estimator statistik yang cukup dan bersifat simetris.
Kata kunci: interval-tersensor, distribusi Binomial, *maximum likelihood estimation*.

ABSTRACT

This study is to estimating of interval-censored data with the binomial distribution. This research uses quantitative methods, the steps are estimating parameters on the interval-censored binomial distribution using the Maximum Likelihood Estimation method. The second step shows the properties of the estimator on the interval-censored binomial distribution. The results showed that the estimator is a sufficient statistic and symmetric.
Key words: *interval-censored, Binomial distribution, maximum likelihood estimation*

Pendahuluan

Distribusi probabilitas binomial adalah distribusi probabilitas diskrit yang paling sering digunakan untuk merepresentasikan kejadian dalam kehidupan sehari-hari (Zhou et al., 2018). Ciri-ciri distribusi binomial yaitu memiliki dua hasil yang mungkin terjadi dalam sebuah percobaan, variabel acak berupa banyaknya kejadian sukses yang dihitung dari sejumlah percobaan, probabilitas dari suatu kejadian sukses tetap sama meskipun percobaannya diulang beberapa kali, dan setiap percobaan saling independen, yang artinya hasil dari suatu percobaan tidak mempengaruhi hasil dari percobaan yang lain (Hanneman et al., 2013).

Dalam statistika inferensial, estimasi titik merupakan suatu metode untuk

KONTAK: Bernadhita H.S. Utami, bernadhita.herindri.s@mail.ugm.ac.id, STMIK Pringsewu

menentukan nilai tunggal yang berasal dari sampel dan digunakan untuk memperkirakan parameter populasi (Utami et al., 2021). Berkaitan dengan estimasi parameter, pada tahun 1800 Karl Pearson memperkenalkan Metode Momen yang merupakan metode relatif sederhana dan menghasilkan estimator yang konsisten. Meskipun demikian, metode momen seringkali menghasilkan estimator yang bersifat bias sehingga untuk memecahkan permasalahan ini Ronald Fisher pada tahun 1913 memperkenalkan metode *maximum likelihood estimation* dengan prinsip memaksimumkan fungsi kemungkinan (*likelihood*) dengan syarat sampel acak berdistribusi probabilitas tertentu (Busemeyer & Wang, 2018).

Jika X suatu variabel acak berdistribusi binomial (n, p) dengan banyak percobaan *n* diketahui, maka parameter probabilitas kejadian sukses *p* akan diestimasi berdasarkan informasi mengenai X (Gupta, 2016). Untuk memperoleh estimator *p* dengan banyak kejadian sukses diketahui secara tepat dengan menggunakan metode *maximum likelihood estimation* diperoleh estimator $\hat{p} = \frac{X}{n}$, dan Xmenyatakan kejadian sukses teramati (Peace et al., 2010). Akan tetapi jika diketahui hanya banyak percobaan sedangkan banyak kejadian sukses hanya diketahui terletak dalam suatu interval misalkan $[x_1, x_2]$ dengan x_1 dan x_2 merupakan bilangan bulat antara 0 dan *n* maka diperoleh estimasi yang sedikit berbeda.

Salah satu permasalahan dalam estimasi parameter adalah adanya pengamatan yang tidak lengkap, yang secara umum dapat dikelompokkan menjadi data tersensor (*censored*) dan data terpotong (*truncated*) (James, 2013). Ketidaklengkapan data yang dapat disebabkan karena beberapa faktor seperti keterbatasan informasi, keterbatasan sumber daya, maupun terjadi hal yang tidak terduga (Fay & Shaw, 2014). Perbedaan keadaan dapat menghasilkan tipe tersensor yang berbeda pula (Ma, 2010). Salah satu yang dapat diketahui dari interval-tersensor adalah sebuah jarak (*range*) yang berada pada saat terjadinya peristiwa.

Kejadian yang menghasilkan data tersensor erat kaitannya dengan analisis survival yaitu analisis data yang memanfaatkan informasi kronologis dari suatu peristiwa (*event*). Respon yang diperhatikan adalah waktu terjadinya suatu *event* sedangkan waktu yang dibutuhkan objek untuk bertahan selama periode pengamatan disebut *survival time* atau *failure time*. Ketidaklengkapan informasi memunculkan permasalahan dalam inferensi yang meliputi estimasi parameter variabel acak berdistribusi binomial. Dalam penelitian ini dibahas mengenai para-

meter distribusi binomial interval-tersensor dengan menggunakan metode *maximum likelihood estimation*, sifat estimator, dan studi kasus data interval-tersensor pada analisis survival.

Beberapa penelitian sebelumnya, memperkenalkan metode *maximum likelihood estimation* untuk membuat model kelangsungan hidup semiparametrik dengan data tersensor pada interval (Zhou & Chellappa, 2006). Baltagi & Liu(2012) menentukan sifat sampel besar dari estimator yang dihasilkan dan mengevaluasi kinerja sampel menggunakan metode *maximum likelihood estimation* pada analisis data konversi diabetes yang dikumpulkan dari studi intervensi individu yang berisiko tinggi terkena diabetes. Selanjutnya, Anderson-Bergman (2017) membangun fungsi *likelihood* dan memperoleh estimasi *maximum likelihood* nonparametrik dari parameter regresi menggunakan algoritma maksimalisasi ekspektasi. Dengan demikian, melalui penelitian ini, penulis menentukan estimasi data interval-tersensor dengan distribusi Binomial. Penulis menggunakan metode *maximum likelihood estimation* dan menentukan karakteristik apakah estimator tersebut merupakan estimator tak bias dan estimator varians minimum.

Metode Penelitian

Penelitian ini merupakan studi literatur yang bertujuan untuk mengembangkan model matematis dan teori yang berkaitan dengan data interval-tersensor. Langkah-langkah yang dilakukan dalam penelitian ini adalah melakukan estimasi parameter dan mendeskripsikan sifat-sifat estimator metode *maximum likelihood estimation* pada distribusi binomial interval-tersensor.

Hasil dan Pembahasan

Estimasi Parameter Distribusi Binomial Interval-Tersensor

Suatu variabel acak X berdistribusi Binomial dengan parameter n dan p memiliki fungsi densitas probabilitas sebagai berikut (Walpole et al., 2012).

$$b(x) = \binom{n}{x} p^x (1-p)^{n-x} x = 0,1,2,\cdots,n \qquad (1)$$

Jika banyak pengamatan n diketahui dan variabel acak X yang menyatakan banyak kejadian sukses hanya diketahui berada dalam interval $[x_1, x_2]$ dengan x_1 dan x_2 merupakan bilangan bulat antara 0 dan n atau dinyatakan dengan

$0 < x_1 < x_2 < n$ maka dapat ditentukan estimator dengan menggunakan metode *maximum likelihood estimation*.

Oleh karena X variabel acak dari observasi tunggal maka fungsi *likelihood* yang diperoleh bukan merupakan perkalian fungsi densitas probabilitas sampel acak $X_1, X_2, \cdots, X_n$ (Casella et al., 2006) yang dinyatakan sebagai:

$$L(\theta; X) = \prod_{i=1}^{n} f(X_i) \tag{2}$$

melainkan sebagai jumlahan dari X_1 sampai X_2 dari fungsi densitas probabilitas yang dinyatakan sebagai:

$$L(b) = \sum_{i=x_1}^{X_2} \binom{n}{i} p^i (1-p)^{n-i} \tag{3}$$

Sekarang, tinjau interval $0 < x_1 \leq x_2 < n$. Dengan menggunakan definisi statistik tataan (*order statistic*) (Casella et al., 2006), jika variabel acak berdistribusi independen identik $X_1, X_2, \cdots, X_n$ disusun dan ditulis sebagai $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$ maka $X_{(i)}$ disebut sebagai tataan ke-*I* untuk $i = 1,2, \cdots, n$ maka fungsi distribusi kumulatif dan fungsi densitas probabilitas dari statistik tataan ke-r dinyatakan sebagai berikut.

1. Fungsi distribusi kumulatif

$$F_{(r)}(X) = Pr\left(X_{(r)} \leq x\right) = \sum_{i=r}^{n} \binom{n}{i} F^i(x)(1-F(x))^{n-i} \tag{4}$$

sehingga fungsi distribusi kumulatif statistik tataan ke-1 dinotasikan dengan $F_{(1)}(x)$ dan dinyatakan dalam persamaan (5).

$$F_{(1)}(x) = 1 - (1-F(x))^n \tag{5}$$

Fungsi distribusi kumulatif statistik tataan ke-*n* dinotasikan dengan $F_{(n)}(x)$ dan dinyatakan dalam persamaan (6).

$$F_{(n)}(x) = \left(F(x)\right)^n \tag{6}$$

2. Fungsi densitas probabilitas

Fungsi densitas probabilitas dari statistik tataan ke-*r* didefinisikan sebagai berikut.

$$f_r(x) = \frac{n!}{(r-1)!\,(n-r)!} \left(F(x)\right)^{r-1} \cdot (1-F(x))^{n-r} \cdot f(x) \tag{7}$$

Transformasi (Jurnal Pendidikan Matematika dan Matematika)
Volume 6, No. 1, Bulan Juni Tahun 2022, pp. 75-85
ISSN 2549-1664 (online)

Dengan $f(x)$ fungsi densitas probabilitas variabel acak maka fungsi densitas probabilitas statistik tataan ke-1 dinyatakan sebagai:

$$f_1(x) = n \cdot (1 - F(x))^{n-1} \cdot f(x) \quad (8)$$

Fungsi densitas probabilitas statistik tataan ke-*n* yaitu $X_{(n)}$ dinyatakan sebagai:

$$f_n(x) = n \cdot \big(F(x)\big)^{n-1} \cdot f(x) \quad (9)$$

Selanjutnya persamaan (3) dapat dinyatakan sebagai berikut:

$$L(b) = \sum_{i=x_1}^{n} \binom{n}{i} p^i (1-p)^{n-i} - \sum_{i=x_2+1}^{n} \binom{n}{i} p^i (1-p)^{n-i} \quad (10)$$

Persamaan terakhir merupakan selisih dari bentuk fungsi distribusi kumulatif statistik tataan sehingga masing-masing memiliki fungsi densitas probabilitas sebagai berikut.

$$f_{x_1}(x) = \frac{n!}{(x_1 - 1)!\,(n - x_1)!} p^{x_1 - 1} (1-p)^{n - x_1} \quad (11)$$

dan

$$f_{x_2}(x) = \frac{n!}{x_2!\,(n - x_2 - 1)!} p^{x_2} (1-p)^{n - x_2 - 1} \quad (12)$$

Berdasarkan fungsi Beta

$$B(\alpha, \beta) = \int_0^1 t^{\alpha - 1} (1-t)^{\beta - 1} dt \quad , \alpha > 0, \beta > 0 \quad (13)$$

dan

$$B(k, n - k + 1) = \int_0^1 t^{k-1} (1-t)^{n-k} dt = \frac{(k-1)!\,(n-k)!}{n!} \quad (14)$$

Sehingga dengan menggunakan hubungan antara distribusi Binomial dengan distribusi Beta, fungsi *likelihood* pada persamaan (10) dapat dinyatakan sebagai:

$$L(b) = I_p(x_1, n - x_1 + 1) - I_p(x_2 + 1, n - x_2)$$

$$L(b) = \int_0^p \frac{n!}{(x_1-1)!(n-x_1)!} y^{x_1 - 1} (1-y)^{n - x_1}\, dy - \int_0^p \frac{n!}{x_2!(n-x_2-1)!} y^{x_2} (1-y)^{n - x_2 - 1} dy \quad (15)$$

Pendiferensialan $L(b)$ terhadap p adalah:

$$\frac{dL(b)}{dp} = \frac{d}{dp} \int_0^p \frac{n!}{(x_1-1)!(n-x_1)!} y^{x_1 - 1} (1-y)^{n - x_1}\, dy$$

$$- \frac{d}{dp} \int_0^p \frac{n!}{x_2!(n-x_2-1)!} p^{x_2} (1-p)^{n - x_2 - 1} dy$$

$$\frac{dL(b)}{dp} = \frac{n!}{(x_1-1)!(n-x_1)!}p^{x_1-1}(1-p)^{n-x_1} - \frac{n!}{x_2!(n-x_2-1)!}p^{x_2}(1-p)^{n-x_2-1} \quad (16)$$

Dengan membuat $\frac{dL(b)}{dp}$ sama dengan nol diperoleh estimator $\hat{p}$.

$$\frac{\hat{p}}{1-\hat{p}} = \left(\frac{x_2!\,(n-x_2-1)!}{(x_1-1)!\,(n-x_1)!}\right)^{\frac{1}{x_2-x_1+1}} \quad (17)$$

Persamaan (17) tersebut ekivalen dengan persamaan berikut.

$$\frac{\hat{p}}{1-\hat{p}} = \left(\frac{x_2!}{(n-x_1)!}\right)^{\frac{1}{x_2-x_1+1}}$$

$$\Leftrightarrow \frac{\hat{p}}{1-\hat{p}} = \left(\frac{x_2\cdot(x_2-1)\cdots(x_1+1)\cdot x_1\cdot(x_1-1)!(n-x_2-1)!}{(n-x_1)(n-x_1-1)\cdots(n-x_2)(n-x_2-1)!(x_1-1)!}\right)^{\frac{1}{x_2-x_1+1}}$$

$$\Leftrightarrow \frac{\hat{p}}{1-\hat{p}} = \left(\left(\frac{x_1}{n-x_1}\right)\left(\frac{x_1+1}{n-x_1-1}\right)\cdots\left(\frac{x_2}{n-x_2}\right)\right)^{\frac{1}{x_2-x_1+1}} \quad (18)$$

Persamaan (18) ekivalen dengan

$$\Leftrightarrow \log\left(\frac{\hat{p}}{1-\hat{p}}\right) = \frac{1}{x_2-x_1+1}\log\left(\left(\frac{x_1}{n-x_1}\right)\left(\frac{x_1+1}{n-x_1-1}\right)\cdots\left(\frac{x_2}{n-x_2}\right)\right)$$

$$\Leftrightarrow \log\left(\frac{\hat{p}}{1-\hat{p}}\right) = \frac{1}{x_2-x_1+1}\left(\sum_{i=x_1}^{x_2}\log\frac{\frac{i}{n}}{1-\frac{i}{n}}\right) \quad (19)$$

Adapun nilai estimator $\hat{p}$ diperoleh dengan menyelesaikan persamaan (19) sebagai berikut.

$$\hat{p} = \frac{\exp\left(\frac{1}{x_2-x_1+1}\left(\sum_{i=x_1}^{x_2}\log\frac{\frac{i}{n}}{1-\frac{i}{n}}\right)\right)}{1+\exp\left(\frac{1}{x_2-x_1+1}\left(\sum_{i=x_1}^{x_2}\log\frac{\frac{i}{n}}{1-\frac{i}{n}}\right)\right)} \quad (20)$$

Untuk $0 < x_1 \leq x_2 < n$ dengan x_1, x_2, n bilangan bulat maka $(x_1 - x_2 - 1)\left(\frac{1}{\hat{p}(1-\hat{p})}\right) < 0$.

Statistik Cukup Distribusi Binomial Interval-Tersensor

Fungsi densitas probabilitas distribusi Binomial dengan parameter (n, θ) adalah:

$$f(x|\theta) = \binom{n}{x}\theta^x(1-\theta)^{n-x} = \binom{n}{x}(1-\theta)^n\left(\frac{\theta}{(1-\theta)}\right)^x$$

$$f(x|\theta) = \binom{n}{x}(1-\theta)^n\exp\left(x\log\left(\frac{\theta}{(1-\theta)}\right)\right) \quad (21)$$

Transformasi (Jurnal Pendidikan Matematika dan Matematika)
Volume 6, No. 1, Bulan Juni Tahun 2022, pp. 75-85
ISSN 2549-1664 (online)

sehingga $f(x|\theta)$ dapat diuraikan menjadi bentuk:

$$h(x) = \begin{cases} \binom{n}{x} & ; x = 0,1,2,\cdots,n \\ 0 & ; selainnya. \end{cases} \quad (22)$$

$$c(\theta) = \begin{cases} (1-\theta)^n & ; \ 0 < \theta < 1 \\ 0 & ; selainnya. \end{cases} \quad (23)$$

$$w_1(\theta) = \begin{cases} \log\left(\frac{\theta}{(1-\theta)}\right) & ; \ 0 < \theta < 1 \\ 0 & ; selainnya. \end{cases} \quad (24)$$

$$t_1(\theta) = x \quad (25)$$

Oleh karena distribusi Binomial dapat dinyatakan sebagai perkalian $h(x)$, $c(\theta)$, $w_1(\theta)$, dan $t_1(\theta)$ maka tergolong dalam kelas keluarga eksponensial. Berdasarkan teorema yang menyatakan bahwa suatu kelas keluarga eksponensial yang mempunyai fungsi statistik $T(\underline{X})$ maka fungsi tersebut merupakan statistik cukup bagi parameter θ (Subanar, 2013). Dengan demikian dapat diperoleh

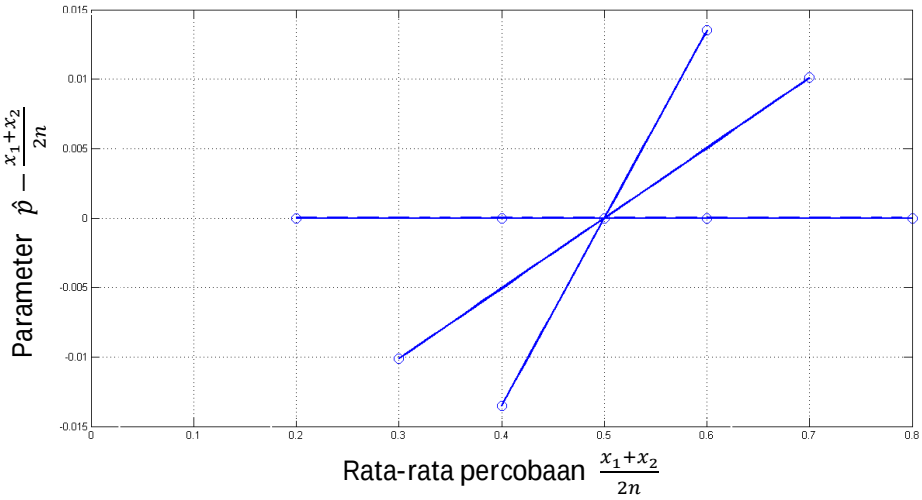
$$\hat{p}(\theta) = \frac{\exp\left(\frac{1}{x_2-x_1+1}\left(\sum_{i=x_1}^{x_2} \log\frac{\frac{i}{n}}{1-\frac{i}{n}}\right)\right)}{1+\exp\left(\frac{1}{x_2-x_1+1}\left(\sum_{i=x_1}^{x_2} \log\frac{\frac{i}{n}}{1-\frac{i}{n}}\right)\right)} \quad (26)$$

merupakan statistik cukup.

Sifat Simetri Estimator Distribusi Binomial Interval-Tersensor

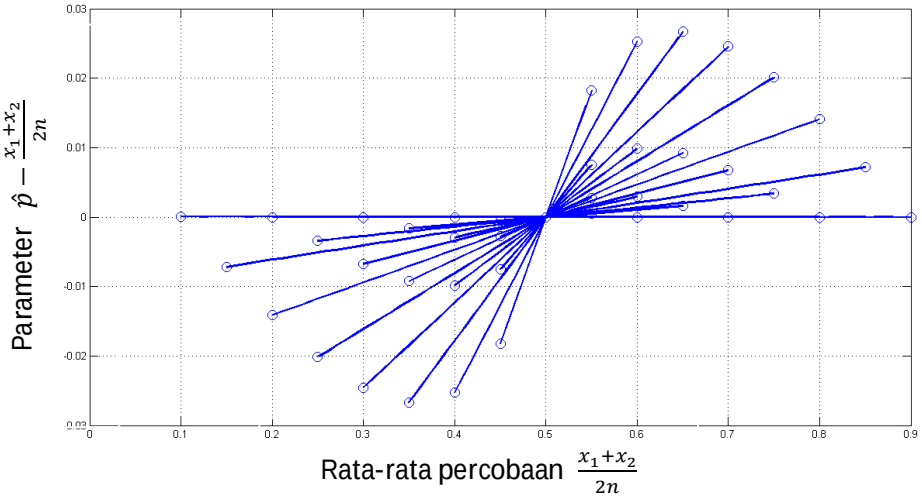
Berdasarkan hasil estimasi untuk$p$ pada persamaan (20) diperoleh sifat pertama bahwa estimator $\hat{p}$ selalu terletak di antara $\frac{x_1}{n}$ dan $\frac{x_2}{n}$ untuk $x_1 < x_2$. Sifat kedua yaitu fungsi $\log\left(\frac{p}{1-p}\right)$ bersifat simetri terhadap $p = 0{,}5$. Lebih lanjut, nilai $\hat{p}$ akan bernilai sama dengan $\frac{x_1+x_2}{2n}$ jika dan hanya jika $x_1 + x_2 = n$ atau $x_1 = x_2$. Untuk banyak percobaan$n$ maka dapat ditentukan semua kemungkinan interval $[x_1, x_2]$ dalam n dengan $0 < x_1 \leq x_2 < n$. Dengan demikian dapat ditunjukkan sifat simetri dari estimator $\hat{p}$untuk data binomial interval-tersensor berdasarkan selisih $\hat{p} - \frac{x_1+x_2}{2n}$. Berikut ini plot selisih $\hat{p} - \frac{x_1+x_2}{2n}$ terhadap ra-

ta-rata $\frac{x_1+x_2}{2n}$ untuk banyak percobaan n = 5, n = 10, dan n = 20 yang ditunjukkan pada Gambar 1 hingga Gambar 3.



Gambar 1. Plot $\hat{p}$ untuk n = 5

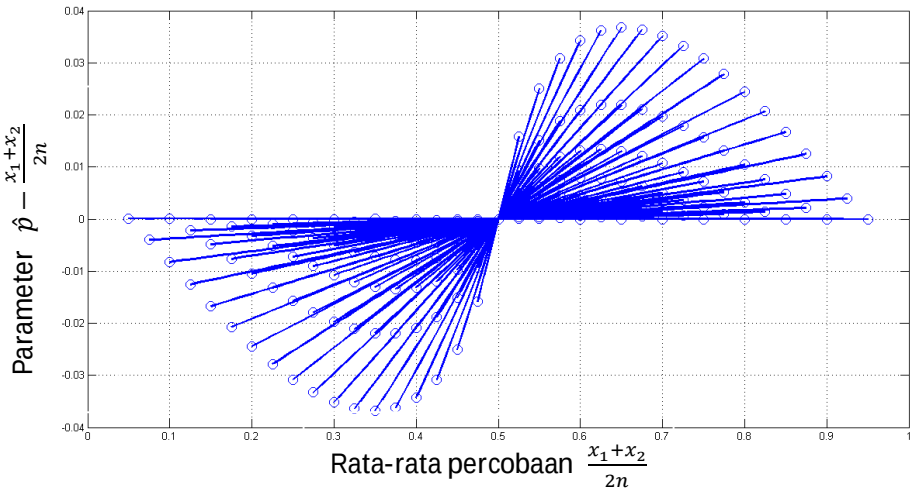
Untuk n = 5 terdapat $\binom{5}{2}$ = 10 kemungkinan interval $[x_1, x_2]$ yang dapat dipilih. Pada Gambar 1 ditunjukkan selisih maksimum ada pada saat rata-rata bernilai 0,6 yaitu 0,01351179. Nilai ini diperoleh saat interval [2,4]. Tampak bahwa plot bersifat simetris terhadap rata-rata yang bernilai 0,5.



Gambar 2. Plot $\hat{p}$ untuk n = 10

Transformasi (Jurnal Pendidikan Matematika dan Matematika)
Volume 6, No. 1, Bulan Juni Tahun 2022, pp. 75-85
ISSN 2549-1664 (online)

Pada Gambar 2 ditampilkan plot untuk n = 10 dengan $\binom{10}{2}$ = 45 kemungkinan interval $[x_1, x_2]$ yang dapat dibuat. Dari plot ditunjukkan bahwa selisih maksimum ada pada saat rata-rata bernilai 0,65 yaitu 0,026661. Nilai ini diperoleh saat interval [4,9]. Tampak bahwa plot bersifat simetris terhadap rata-rata yang bernilai 0,5. Hal ini menunjukkan bahwa semakin banyakpercobaan yang dilakukan maka nilai parameter akan berulang kembali di titik 0,5.



Gambar 3. Plot $\hat{p}$ untuk n = 20

Dari Gambar 1 hingga Gambar 3 diperoleh kesimpulan bahwa untuk banyak percobaan n yang semakin besar, plot menunjukkan pola cekung ke atas untuk nilai probabilitas 0 sampai dengan 0,5 dan menunjukkan pola cekung ke bawah untuk nilai probabilitas 0,5 sampai dengan 1.

Kesimpulan

Berdasarkan pembahasan, diperoleh kesimpulan bahwa distribusi Binomial interval-tersensor berasal dari keluarga distribusi Binomial yang merupakan kelas keluarga eksponensial sehingga estimatornya merupakan statistik cukup dan bersifat simetris.

Ucapan Terimakasih

Ucapan terimakasih ditujukan kepada Ketua STMIK Pringsewu, KabupatenPringsewu, Lampung yang telah mendukung keterlaksanaan penelitian ini.

Daftar Pustaka

Anderson-Bergman, C. (2017). Regression Models for Interval Censored Data in R. *Journal of Statistical Software*, *81*(12), 1–10. https://doi.org/10.18637/jss.v081.i12

Baltagi, B. H., & Liu, L. (2012). The Hausman-Taylor panel data model with serial correlation. *Statistics and Probability Letters*, *82*(7), 1401–1406. https://doi.org/10.1016/j.spl.2012.03.016

Busemeyer, J. R., & Wang, Z. (2018). Hilbert space multidimensional theory. *Psychological Review*, *125*(4), 572–591. https://doi.org/10.1037/rev0000106

Casella, G., Fienberg, S., & Olkin, I. (2006). Springer Texts in Statistics. In *Design* (Vol. 102). https://doi.org/10.1016/j.peva.2007.06.006

Fay, M. P., & Shaw, P. A. (2014). Censored Data : The interval R package. *Journal of Statistical Software*, *36*(2), 1–34. https://www.jstatsoft.org/v36/i02/.

Gupta, B. (2016). Introduction to Basic Statistics. *Interview Questions in Business Analytics*, 23–35. https://doi.org/10.1007/978-1-4842-0599-0_3

Hanneman, R. A., Kposowa, A. J., & Riddle, M. D. (2013). *Basic Statistics for Social Research* (1st ed.). Wiley and Sons, Inc.: San Fransisco.

James, G. (2013). *An Introduction to Statistical Learning with Applications in R* (1st ed.). Springer: New York.

Ma, S. (2010). Mixed case interval censored data with a cured subgroup. *Statistica Sinica*, *20*(3), 1165–1181.

Peace, K. E., Sun, J., & Chen, D.-G. (2010). Interval-Censored Time-to-Event Data : Methods and Applications. In *Interval-Censored Time-to-Event Data : Methods and Applications* (1st ed., p. 39). Springer: New York.

Utami, B. H. S., Irawan, A., Gumanti, M., & Primajati, G. (2021). Hausman and Taylor Estimator Analysis on The Linear Data Panel Model. *Varian*, *5*(1), 81–88.

Walpole, R. E., Myers, R. H., Myers, S. L., & Ye, K. (2012). *Probability and Statistics for Engineers and Scientistist* (9th ed.). Prentice Hall Pearson: New York.

Zhou, J., Zhang, J., & Lu, W. (2018). Computationally Efficient Estimation for the Generalized Odds Rate Mixture Cure Model With Interval-Censored Data. In *Journal of Computational and Graphical Statistics* (Vol. 27, Issue 1). https://doi.org/10.1080/10618600.2017.1349665

Zhou, S. K., & Chellappa, R. (2006). From Sample Similarity to Ensemble Similarity: Probabilistic Distance Measures in Reproducing Kernel Hilbert Space. *IEEE*

Transactions on Pattern Analysis and Machine Intelligence, *28*(6), 917–929. https://doi.org/10.1109/TPAMI.2006.120