

BAB I

PENDAHULUAN

1.1 Latar Belakang

Statistika merupakan ilmu yang berfokus pada pengumpulan, penyajian, analisis dan interpretasi data dalam bentuk numerik, Tabel, grafik, maupun distribusi probabilitas. Secara umum, statistika terbagi menjadi dua cabang utama, yaitu statistika deskriptif dan statistika inferensial. Statistika deskriptif digunakan untuk menggambarkan atau menyajikan data secara ringkas maupun informatif. Sedangkan Statistika inferensi bertujuan untuk menarik kesimpulan atau generalisasi tentang suatu populasi berdasarkan data sampel yang tersedia (Usman dan Akbar, 2006).

Salah satu topik penting dalam statistika inferensi adalah estimasi parameter distribusi. Dalam konteks distribusi binomial, parameter yang perlu diestimasi adalah peluang keberhasilan dalam satu percobaan (p) dan jumlah percobaan (n). Estimasi parameter yang tepat sangat penting untuk membentuk model yang akurat dan menjadi dasar dalam proses pengambilan keputusan di berbagai bidang, seperti ekonomi, ilmu sosial, dan ilmu sains.

Dua metode estimasi parameter yang umum digunakan dalam statistika adalah Maximum Likelihood Estimation (MLE) dan Game Theory Estimation (GTE). MLE merupakan pendekatan klasik yang telah banyak diterapkan dalam berbagai disiplin ilmu. Metode ini bekerja dengan mencari nilai parameter yang memaksimalkan kemungkinan (likelihood) data yang diamati. Sementara itu, GTE merupakan pendekatan berbasis teori permainan yang mulai dikembangkan dalam konteks ekonometri dan situasi kompetitif antar pengambil keputusan. Meskipun memiliki dasar teoritis yang kuat, penggunaan GTE dalam statistika masih tergolong baru dan belum banyak dibandingkan secara langsung dengan MLE, terutama dalam konteks distribusi binomial.

Teori permainan sendiri merupakan cabang matematika yang digunakan untuk menganalisis strategi dalam situasi yang melibatkan dua atau lebih pihak dengan kepentingan saling bertentangan. Teori ini dikembangkan untuk menganalisa proses pengambilan keputusan dari situasi-situasi persaingan yang berbeda-beda dan melibatkan dua atau lebih kepentingan. Dalam konteks estimasi, GTE bertujuan untuk menghasilkan strategi estimasi parameter yang meminimalkan kerugian rata-rata (expected loss) dalam kondisi ketidakpastian.

Kajian literatur menunjukkan bahwa meskipun MLE dan GTE telah dibahas secara terpisah dalam berbagai studi, namun masih sedikit penelitian yang secara eksplisit membandingkan efektivitas keduanya dalam konteks distribusi binomial. Oleh karena itu, penelitian ini menjadi penting untuk mengisi celah tersebut dan memberikan pemahaman yang lebih komprehensif mengenai keunggulan dan kelemahan masing-masing metode.

Pemahaman yang lebih mendalam terhadap perbandingan antara GTE dan MLE tidak hanya

akan memperkaya literatur statistik, tetapi juga membantu para peneliti, analis data, dan praktisi di berbagai bidang dalam memilih metode estimasi yang tepat sesuai dengan konteks dan tujuan analisis mereka.

Oleh karena itu, penelitian ini memiliki latar belakang yang kuat dan relevan dalam memahami perbandingan antara GTE dan MLE dalam konteks estimasi parameter distribusi binomial dan potensial untuk memberikan kontribusi yang signifikan dalam bidang statistik dan aplikasinya dalam berbagai disiplin ilmu.

1.2 Batasan Masalah

Batasan masalah dalam penelitian ini adalah sebagai berikut.

- 1.2.1 Fungsi distribusi yang digunakan dalam penelitian ini adalah distribusi *uniform* dan distribusi dari strategi terbaik dalam permainan “Das Statistische Spiel”
- 1.2.2 Parameter n yang digunakan dalam distribusi binomial $b(n, p)$ dalam penelitian ini adalah $n = 10$, sedangkan parameter p dibangkitkan melalui distribusi *uniform* dan distribusi strategi terbaik.

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah yang telah diuraikan, tujuan penelitian inidiantaranya

- 1.3.1 Menentukan dan menghitung estimator parameter distribusi binomial dengan menggunakan *maximum likelihood estimation (MLE)* dan *game theory estimation (GTE)*.
- 1.3.2 Menganalisis perbandingan efektivitas antara *maximum likelihood estimation* dan *game theory estimation* dalam studi kasus permainan “Das Statistische Spiel”..

1.4 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat sebagai berikut :

- 1.4.1 Secara akademik, Penelitian ini memberikan kontribusi terhadap pengembangan literatur mengenai metode estimasi parameter distribusi, khususnya dalam membandingkan Game Theory Estimation (GTE) dan Maximum Likelihood Estimation (MLE)
- 1.4.2 Secara praktis, hasil penelitian ini dapat menjadi acuan bagi peneliti dan praktisi dalam memilih metode estimasi parameter distribusi binomial yang paling sesuai untuk kondisi data tertentu, sehingga dapat meningkatkan akurasi dalam pengambilan keputusan berbasis data.

1.5 Teori

1.5.1 Distribusi Binomial

Dalam n percobaan Bernoulli, dimana peluang suksesnya tetap sama. Misalkan X_i merupakan peubah acak Bernoulli yang bersesuaian dengan percobaan ke- i . Sebuah barisan dari n percobaan Bernoulli kemudian merupakan sebuah n -tuple dari 0 (nol) dan 1 (satu). Pilih peubah acak $Y = \sum_{i=1}^n X_i$ yang menyatakan banyaknya sukses dari n percobaan

tersebut, sehingga nilai dari Y mencakup $0, 1, 2, \dots, n$. Jika y sukses terjadi, maka $n - y$ gagal terjadi. Banyaknya cara memilih y posisi untuk y sukses dalam n percobaan adalah

$$\binom{n}{y} = \frac{n!}{y!(n-y)!} \quad (1.1)$$

karena setiap percobaan saling independen dan peluang dari sukses dan gagal tetapsama dari percobaan satu ke percobaan yang lain, yaitu p dan $1 - p$, sehingga peluang suatu y sukses dan $n - y$ gagal adalah $p^y(1 - p)^{n-y}$. *Probability Mass Function (pmf)* selanjutnya merupakan dari $\binom{n}{y}$ kejadian-kejadian yang saling eksklusif; yaitu (Hogg, 2019),

$$P(Y = y; p) = \begin{cases} \binom{n}{y} p^y (1 - p)^{n-y}, & y = 0, 1, 2, \dots, n \\ 0, & \text{lainnya} \end{cases} \quad (1.2)$$

1.5.2 Estimasi Maximum Likelihood

Misalkan dimiliki sampel acak $X_1, X_2, \dots, X_n$ pada peubah acak X yang memiliki *probability density function* (pdf) $f(x; \theta)$, $\theta \in \Omega$. Basis dari inferensi adalah fungsi *likelihood* yang diberikan oleh

$$L(\theta, x) = \prod_{i=1}^n f(x_i; \theta), \quad \theta \in \Omega \quad (1.3)$$

dengan $x = (x_1, x_2, \dots, x_n)'$. Adapun log dari fungsi *likelihood* ini diberikan oleh

$$\ell(\theta; x) = \log L(\theta; x) = \sum_{i=1}^n \log f(x_i; \theta), \quad \theta \in \Omega \quad (1.4)$$

Estimasi titik dari θ selanjutnya adalah $\hat{\theta} = \hat{\theta}(X_1, X_2, \dots, X_n)$, dimana $\hat{\theta}$ merupakan nilai θ yang memaksimumkan fungsi $\ell(\theta; x)$ yang juga sekaligus memaksimumkan fungsi $L(\theta; x)$ (Hogg, 2019).

1.5.3 Prinsip Minimax

Misalkan $Y = u(X_1, X_2, \dots, X_n)$ adalah statistik yang menjadi basis dari sebuah estimasi titik dari parameter $\theta \in \Omega$. Y merupakan peubah acak pada D_Y dengan distribusi $F_Y(y; \theta)$. Pilih $\delta(y)$, yaitu fungsi dari realisasi statistik Y , yang merupakan estimasi titik dari θ . Fungsi δ ini kemudian disebut sebagai *decision function*. Salah satu nilai dari fungsi ini, $\delta(y)$ dinamakan sebagai *decision*. Sehingga estimasi titik yang ditentukan secara numerik adalah sebuah keputusan.

Decision dapat saja benar atau salah. Oleh karena itu diperlukan sebuah ukuran untuk mengukur keseriusan dari beda antara θ dengan estimasi titik $\delta(y)$. Berdasarkan hal tersebut, setiap pasangan $[\theta, \delta(y)]$ dipasangkan dengan sebuah bilangan tak negatif

$\mathcal{L}[\theta, \delta(y)]$ yang merefleksikan ukuran keseriusan tersebut. Fungsi $\mathcal{L}$ kemudian disebut sebagai *loss function*. Mean dari *loss function* dinamakan *risk function* yang diberikan oleh

$$R_\delta(\theta) = E_\theta\{\mathcal{L}[\theta, \delta(y)]\} = \int_{D_Y} \mathcal{L}[\theta, \delta(y)] dF_Y(y; \theta) \quad (1.5)$$

yang menyatakan *risk function* saat θ merupakan parameternya. Prinsip *minimax* selanjutnya adalah memilih *decision function* terbaik $\delta_0(y)$ sedemikian sehingga untuk semua $\theta \in \Omega$,

$$\max_{\theta} R_{\delta_0}(\theta) \leq \max_{\theta} R_\delta(\theta)$$

Untuk setiap *decision function* $\delta(y)$ (Hogg, 2019).

Masalah estimasi titik selanjutnya dapat diformulasikan sebagai berikut. Sebuah peubah acak X memiliki distribusi sepanjang D_X menurut suatu fungsi distribusi F_X yang merupakan anggota dari *family distribution* $\mathcal{F}$. Kita ingin mengestimasi $g(F_X)$ yang merupakan fungsi dengan domain $\mathcal{F}$ dan range D_Y . Sebuah estimasi adalah statistik $\delta(X)$ yang memiliki nilai dalam D_Y . Selanjutnya, nyatakan $E\{\mathcal{L}[F_X, \delta(x)]\}$ sebagai *loss function* yang dihasilkan dari estimasi $\delta(x)$ saat F_X merupakan distribusinya, dan definisikan *risk function* dari estimasi δ sebagai

$$R_\delta(F_X) = E_{F_X}\{\mathcal{L}[g(F_X), \delta(X)]\} \quad (1.6)$$

Lalu pilih δ_0 yang meminimumkan $\sup_{F_X \in \mathcal{F}} \{R_\delta(F_X)\}$ (Hodges & Lehman, 1950).

Teorema 1 Misalkan $\{F_X(x; \theta)\}, \theta \in \Omega$ merupakan *parametric subfamily* dari $\mathcal{F}$ dan misalkan $F_\theta(\theta)$ merupakan distribusi sepanjang Ω . Jika f meminimumkan

$$\int_{\Omega} E_\theta\{\mathcal{L}[g(F_X(x; \theta)), \delta(X)]\} dF_\theta(\theta) \quad (1.7)$$

Dan

- (i) $E_\theta\{\mathcal{L}[g(F_X(x; \theta)), \delta_0(X)]\}$ adalah konstan (c) untuk semua $\theta \in \Omega$
- (ii) $E_{F_X}\{\mathcal{L}[g(F_X), \delta_0(X)]\} \leq c$ untuk semua F_X dalam $\mathcal{F}$

Maka δ_0 merupakan estimasi minimax untuk mengestimasi g .

1.5.4 Estimasi Game Theory

Misalkan suatu bilangan $p \in (0,1)$ dipilih secara acak berdasarkan suatu fungsi densitas $f(p)$ dengan $\int_0^1 f(p) dp = 1$. Nilai p ini kemudian merupakan parameter dari distribusi $bin(n, p)$ yang akan menghasilkan y sukses. Asumsikan bahwa hal tersebut dilakukan secara berulang kali (Q), namun n tetap untuk tiap percobaan. Estimasi game

theory selanjutnya adalah $\tilde{p}$ yang meminimumkan secara rata-rata $(p - \tilde{p})^2$ dari tiap sub barisan $Q_y (y = 0, 1, 2, \dots, n)$, dengan Q_y merupakan semua percobaan yang menunjukkan y sukses yang sama. Ekspektasi dari $(p - \tilde{p})^2$ untuk tiap sub barisan tersebut adalah

$$I(\tilde{p}, y) = \int_0^1 f(p) \binom{n}{y} p^y (1-p)^{n-y} (p - \tilde{p})^2 dp \quad (1.8)$$

Minimum dari fungsi tersebut adalah solusi $\tilde{p}$ dari persamaan $\frac{\partial I(\tilde{p}, y)}{\partial \tilde{p}} = 0$ yaitu suatu fungsi dari y dan n yang dinotasikan oleh $\tilde{p}(y, n)$ yang memiliki bentuk :

$$\tilde{p}(y, n) = \frac{\int_0^1 f(p) p^{y+1} (1-p)^{n-y} dp}{\int_0^1 f(p) p^y (1-p)^{n-y} dp}$$

Karena $I(\tilde{p}, y)$ merupakan fungsi kuadrat tak-negatif dari $\tilde{p}$, sehingga diperoleh $I''(\tilde{p}, y) > 0$, yaitu $\tilde{p}(y, n)$ merupakan minimum dari $I(\tilde{p}, y)$. (Steinhaus, 1957).

BAB II

METODOLOGI PENELITIAN

2.1 Sumber Data

Data dalam permainan ini berasal dari suatu studi kasus suatu permainan. Namadari permainan ini adalah “Das Statistische Spiel”. Permainan ini terdiri dari 2 (dua) pemain, yaitu S dan D . Pemain D memiliki koleksi koin yang cukup banyak dengantiap koin memiliki peluang p untuk memunculkan “kepala” dalam tiap toss. Asumsikan bahwa D memiliki cukup banyak koleksi koin dengan sembarangpeluang kemunculan “kepala” $p \in (0,1)$. Pemain D memilih koin sesuai yang ia inginkan dan membiarkan pemain S melemparkan koin tersebut sebanyak n kali; yang mana menghasilkan “kepala” sebanyak y kali. Lalu S akan memberikan sebuah estimasi $\hat{p}$ dari nilai p yang mana diketahui oleh D namun tidak diketahui oleh dirinya. Setelah hal ini dilakukan S membayar D sebanyak $$(\hat{p} - p)^2$$. Tujuan pemain S adalah mencari strategi yang dapat meminimumkan ekspektasi $(\hat{p} - p)^2$.

2.2 Variabel Penelitian

Peubah acak yang digunakan dalam penelitian ini adalah peubah acak $Y \sim \text{bin}(n, p)$ dengan $n = 10$ dan $0 < p < 1$ berasal dari distribusi yang memiliki densitas:

1. $f_1(p) = 1$
2. $f_2(p) = (p(1-p))^{\frac{\sqrt{n}}{2}-1} \cdot \frac{\Gamma(\frac{\sqrt{n}}{2})}{\left(\Gamma(\frac{\sqrt{n}}{2})\right)^2}$

2.3 langkah-Langkah Analisis

1. Bangkitkan observasi p dari densitas $f_1(p)$ dan $f_2(p)$
2. Bangkitkan $Y \sim \text{Bin}(10, p)$
3. Langkah 1-2 dilakukan sebanyak total simulasi yg diinginkan
3. Hitung estimasi maximum likelihood dan estimasi minimax dari setiap densitas
4. Hitung $(p - \hat{p})^2$ dan $(p - \tilde{p})^2$ dari setiap estimasi
5. Interpretasi Hasil

2.4 Diagram alir (Flowchart)

