

SIMULASI DISTRIBUSI BINOMIAL UNTUK ANALISIS PROBABILITAS KEBERHASILAN BERULANG MENGGUNAKAN PYTHON

Arnah Ritonga¹, Naysla Aura Siffa², Rizal Muhaimin³, Khalista Amelia⁴, Maysi Puspita⁵, N. Aulia Asy Syifaa⁶, Adinda Saputri⁷

arnahritonga@unimed.ac.id¹, nayslaauraa@gmail.com², rizal.4243230008@mhs.unimed.ac.id³,
khalistaamelia2905@gmail.com⁴, maysipuspita8@gmail.com⁵, syifaaanun@gmail.com⁶,
adindasyahputri72@gmail.com⁷

Universitas Negeri Medan

ABSTRAK

Penelitian ini bertujuan untuk mengevaluasi kesesuaian distribusi Binomial empiris dengan distribusi teoretis melalui simulasi Monte Carlo menggunakan Python. Data dihasilkan secara sintesis dengan parameter jumlah percobaan ($n = 10$) dan probabilitas keberhasilan ($p = 0.3$). Sebanyak $N = 10.000$ simulasi dijalankan untuk menghitung frekuensi relatif jumlah keberhasilan. Hasil simulasi dibandingkan dengan probabilitas teoretis menggunakan uji Chi square dan selisih maksimum antara PMF empiris dan teoretis. Nilai statistik Chi square sebesar 7.57 (p value = 0.6704) dan selisih maksimum 0.0040 menunjukkan bahwa distribusi empiris tidak berbeda signifikan dari distribusi teoretis. Penelitian ini juga menunjukkan bagaimana penggunaan bahasa pemrograman Python secara praktis dapat membantu mahasiswa atau peneliti dalam memahami konsep probabilitas diskrit. Dengan memvisualisasikan hasil simulasi dan distribusi teoretis, pembelajaran konsep distribusi Binomial menjadi lebih intuitif dan menarik. Simulasi ini dapat menjadi alat bantu yang efektif dalam pembelajaran teori peluang di tingkat perguruan tinggi maupun riset eksperimental yang melibatkan distribusi diskrit. Selain itu, pendekatan Monte Carlo juga memberikan fleksibilitas dalam membangun skenario distribusi probabilitas lain untuk kebutuhan penelitian statistik dan data science.

Kata Kunci: Distribusi Binomial, Simulasi Monte Carlo, Python, Chi Square, Probabilitas.

ABSTRAK

This study aims to evaluate the suitability of the empirical Binomial distribution with the theoretical distribution through Monte Carlo simulation using Python. Data were generated synthetically with parameters of the number of trials ($n = 10$) and the probability of success ($p = 0.3$). A total of $N = 10,000$ simulations were run to calculate the relative frequency of the number of successes. The simulation results were compared with the theoretical probability using the Chi-square test and the maximum difference between the empirical and theoretical PMFs. The Chi-square statistical value of 7.57 (p value = 0.6704) and the maximum difference of 0.0040 indicate that the empirical distribution is not significantly different from the theoretical distribution. This study also shows how the use of the Python programming language can practically help students or researchers understand the concept of discrete probability. By visualizing the simulation results and theoretical distributions, learning the concept of the Binomial distribution becomes more intuitive and interesting. This simulation can be an effective tool in learning probability theory at the college level as well as experimental research involving discrete distributions. In addition, the Monte Carlo approach also provides flexibility in building other probability distribution scenarios for statistical and data science research needs.

Keywords: Binomial Distribution, Monte Carlo Simulation, Python, Chi Square, Probability.

PENDAHULUAN

Distribusi Binomial adalah salah satu distribusi diskrit yang banyak digunakan untuk memodelkan jumlah keberhasilan dalam sejumlah percobaan independen dengan probabilitas sukses konstan. Pemahaman dasar terhadap karakteristik empiris dan teoretis distribusi ini penting dalam berbagai aplikasi, seperti penilaian kualitas produk, riset klinis,

dan sistem telekomunikasi. Dengan kemudahan pemrograman Python dan paket-paket statistik seperti NumPy dan SciPy, simulasi Monte Carlo menjadi metode praktis untuk memvalidasi distribusi teoretis.

Dalam penelitian ini, penulis menerapkan konsep dari teori probabilitas dan statistika, yaitu distribusi binomial. Distribusi binomial merupakan salah satu jenis distribusi probabilitas diskret yang menggambarkan jumlah keberhasilan dalam serangkaian percobaan dengan dua kemungkinan hasil, yaitu berhasil atau gagal, yang bersifat saling bebas. Setiap percobaan memiliki peluang keberhasilan sebesar p . Percobaan dengan dua kemungkinan ini dikenal juga sebagai percobaan Bernoulli. Jika hanya terdapat satu percobaan ($n = 1$), maka distribusi binomial akan menjadi distribusi Bernoulli. Distribusi ini umum digunakan untuk memodelkan banyaknya keberhasilan atau kegagalan dari sejumlah sampel n dalam populasi N (Prasetyo et al., 2021).

Distribusi ini termasuk ke dalam jenis distribusi diskret yang sederhana namun memiliki pendekatan yang efektif dan sering diterapkan. Distribusi binomial umumnya dimanfaatkan untuk memodelkan jumlah keberhasilan dalam sampel berukuran n yang diambil dari populasi berukuran N . Penerapan distribusi ini membantu dalam menentukan probabilitas serta nilai harapan dari setiap peubah acak x . Nilai probabilitas suatu peristiwa disajikan secara kuantitatif dalam rentang antara 0 hingga 1, di mana nilai 0 menunjukkan peristiwa tersebut tidak mungkin terjadi, sedangkan nilai 1 menunjukkan peristiwa tersebut pasti terjadi (Diana, 2017). Probabilitas atau peluang dinotasikan dengan p dimana nilai untuk probabilitas berkisar antara 0 dan 1 atau $0 \leq p \leq 1$ (Yanti et al., 2024).

Distribusi binomial banyak digunakan untuk memodelkan jumlah keberhasilan dalam uji coba independen dengan probabilitas konstan. Distribusi ini memiliki aplikasi dalam menilai kualitas layanan (Nurhaliza et al., 2024;), memprediksi kehadiran mahasiswa (Hutahaean, 2018), dan membandingkan dengan distribusi lainnya (Manurung et al., 2013). Studi telah menggunakan distribusi binomial untuk memperkirakan probabilitas keberhasilan dalam uji coba kualitas layanan sistem informasi menggunakan dimensi E-ServQual (Nurhaliza et al., 2024). Simulasi dengan berbagai ukuran sampel dan probabilitas keberhasilan menunjukkan bahwa sampel yang lebih kecil dengan probabilitas keberhasilan yang lebih tinggi meningkatkan kemungkinan keberhasilan (Nurhaliza et al., 2024). Metode Monte Carlo, yang memanfaatkan pembangkitan bilangan acak, telah diterapkan untuk memprediksi kehadiran mahasiswa di lingkungan pendidikan tinggi (Hutahaean, 2018). Untuk n besar dan p kecil, distribusi binomial dapat didekati dengan distribusi Poisson, terutama ketika $n > 50$ dan $p \leq 0,1$ (Manurung et al., 2013).

TINJAUAN PUSTAKA

Distribusi Binomial

Distribusi Binomial adalah distribusi probabilitas diskrit yang menggambarkan jumlah keberhasilan dalam sejumlah percobaan Bernoulli yang independen, di mana setiap percobaan memiliki dua kemungkinan hasil: sukses (dengan probabilitas p) dan gagal (dengan probabilitas $1-p$). Fungsi probabilitas massanya (PMF) dinyatakan sebagai:

$$P(x = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

Di mana:

- n : jumlah percobaan
- k : jumlah keberhasilan ($0 \leq k \leq n$)
- p : probabilitas keberhasilan dalam satu percobaan

Distribusi Binomial memiliki nilai harapan (mean) dan variansi sebagai berikut:

- Nilai harapan: $E[x] = np$

- Variansi: $\text{Var}(x) = np(1-p)$

Distribusi ini banyak digunakan dalam statistika inferensial, khususnya untuk uji hipotesis dan estimasi parameter. Dalam konteks simulasi komputer, distribusi Binomial dapat disimulasikan menggunakan metode Monte Carlo, yaitu pendekatan berbasis pengulangan acak untuk menghitung estimasi probabilitas. Uji kesesuaian (goodness-of-fit) seperti uji Chi-square digunakan untuk mengevaluasi apakah data hasil simulasi konsisten dengan distribusi yang dihipotesiskan.

Metode Monte Carlo

Simulasi Monte Carlo merupakan sebuah metode stokastik yang sangat fleksibel dan banyak digunakan untuk menyelesaikan berbagai permasalahan kompleks. Metode ini bekerja dengan melakukan pengambilan sampel secara acak untuk merepresentasikan dan memodelkan sistem yang rumit, terutama ketika perhitungan analitis sulit dilakukan. Dengan menjalankan simulasi berulang kali menggunakan data acak, metode ini memungkinkan kita untuk memperkirakan kemungkinan hasil dari suatu proses atau fenomena yang melibatkan ketidakpastian (Hasyim, 2025).

Metode Monte Carlo merupakan pendekatan berbasis probabilitas yang menggunakan bilangan acak untuk mensimulasikan perilaku yang bersifat stokastik atau untuk memperkirakan solusi dari suatu permasalahan tertentu.

Beberapa keunggulan dari metode simulasi ini antara lain:

1. Memiliki tingkat fleksibilitas yang tinggi dalam menggambarkan sistem yang kompleks, sehingga dapat meningkatkan tingkat akurasi dalam merepresentasikan sistem tersebut.
2. Dapat digunakan sebagai alat pembanding antar alternatif solusi, sehingga membantu dalam menentukan pilihan yang paling sesuai dengan karakteristik sistem yang dikaji.
3. Hasil dari simulasi dapat dianalisis dan dipelajari dalam jangka waktu yang panjang, memberikan wawasan lebih mendalam terhadap perilaku sistem dari waktu ke waktu.
4. Simulasi memungkinkan penghematan waktu dalam proses analisis, serta memberikan kendali yang lebih baik terhadap variabel dan kondisi yang diuji.

(Mulia & Afif, 2024).

Rumus berikut digunakan sebagai dasar dalam penerapan metode Monte Carlo.

$$\text{PDF} = \frac{F}{J}$$

$$Z_{i+1} = (a \cdot Z_i + C) \bmod M$$

PDF merupakan distribusi frekuensi probabilitas, di mana F menyatakan frekuensi dan J adalah jumlah total data. Nilai ini digunakan untuk menghitung distribusi probabilitas kumulatif. Selanjutnya, rumus juga digunakan untuk menghasilkan bilangan acak dengan ketentuan bahwa a dan c lebih kecil dari M, serta Z_0 (nilai awal) harus lebih besar dari nol. Di sini, Z_i adalah nilai awal yang telah ditentukan, a merupakan konstanta perkalian, C adalah konstanta penambahan, Mod menunjukkan operasi modulus, dan M merupakan batas maksimum nilai bilangan acak yang dihasilkan. (Maulana, 2024)

METODOLOGI

Data Sintetis

Data dihasilkan secara sintetis menggunakan fungsi `numpy.random.binomial` dengan parameter:

- Jumlah trial per eksperimen: $n = 10$
- Probabilitas keberhasilan tiap trial: $p = 0.3$
- Jumlah eksperimen (simulasi): $N = 10.000$

Simulasi dan Perhitungan PMF

```
import numpy as np
from scipy.stats import binom

n, p, N = 10, 0.3, 10000
samples = np.random.binomial(n, p,
size=N)
# Hitung frekuensi relatif
counts = np.bincount(samples,
minlength=n+1)
empirical_pmf = counts / N
# PMF teoretis
k = np.arange(n+1)
theoretical_pmf = binom.pmf(k, n,
p)
```

Hasil empirical_pmf dan theoretical_pmf divisualisasikan menggunakan matplotlib.

Uji Goodness of Fit

Uji Chi square dilakukan untuk membandingkan frekuensi observasi (counts) dengan frekuensi ekspektasi (theoretical_pmf * N). Selain itu, dihitung selisih maksimum antara PMF empiris dan teoretis:

```
import numpy as np
from scipy.stats import binom

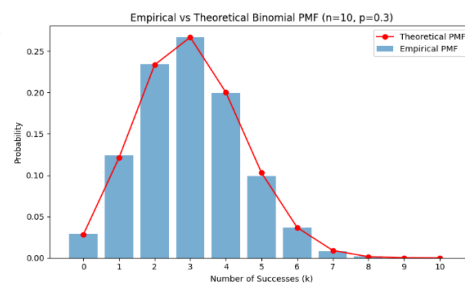
n, p, N = 10, 0.3, 10000
samples = np.random.binomial(n, p,
size=N)
# Hitung frekuensi relatif
counts = np.bincount(samples,
minlength=n+1)
empirical_pmf = counts / N
# PMF teoretis
k = np.arange(n+1)
theoretical_pmf = binom.pmf(k, n,
p)
```

HASIL DAN PEMBAHASAN

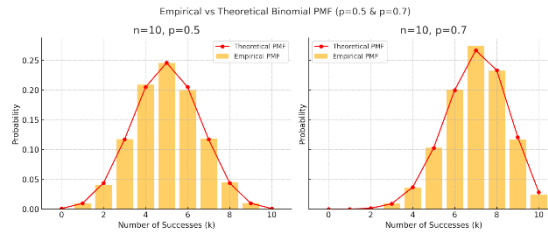
Hasil Penelitian

Chi-square Statistic: 7.57
p-value: 0.6704
Maximum absolute difference: 0.0040

Gambar 1: hasil output



Gambar 2: diagram



Gambar 3 : diagram perbandingan

Pembahasan

Hasil p value > 0.05 menunjukkan tidak ada bukti signifikan bahwa distribusi empiris berbeda dari distribusi Binomial teoretis. Selisih maksimum yang sangat kecil (< 0.01) juga menegaskan kesesuaian kedua distribusi. Ini menunjukkan bahwa metode simulasi Monte Carlo dengan jumlah simulasi yang cukup besar ($N = 10.000$) dapat merepresentasikan distribusi teoretis secara akurat.

Distribusi PMF empiris (batang biru) mengikuti dengan sangat baik bentuk distribusi teoretis (garis merah), terutama pada bagian tengah distribusi. Hasil ini mendukung pemahaman bahwa distribusi Binomial dengan $p = 0.3$ menghasilkan distribusi miring ke kiri karena peluang keberhasilan lebih kecil dari 0.5.

Dengan hasil ini, terbukti bahwa Python dapat digunakan sebagai alat bantu visualisasi dan validasi statistik. Teknik ini sangat bermanfaat dalam konteks pendidikan dan riset, karena memungkinkan eksplorasi konsep distribusi dengan pendekatan berbasis data dan simulasi langsung.

Perbandingan Distribusi dengan Nilai p yang Berbeda

Sebagai bagian dari eksplorasi lebih lanjut terhadap karakteristik distribusi Binomial, dilakukan simulasi tambahan dengan parameter $p=0.5$ dan $p=0.7$, masing-masing dengan jumlah percobaan tetap $n=10$ dan jumlah simulasi $N=10.000$. Tujuan dari simulasi ini adalah untuk mengamati bagaimana perubahan nilai probabilitas keberhasilan (p) memengaruhi bentuk distribusi.

Hasil visualisasi menunjukkan bahwa distribusi dengan $p=0.5$ bersifat simetris, dengan puncak distribusi berada di tengah (yakni di sekitar $k=5$). Hal ini sejalan dengan sifat teoretis distribusi Binomial bahwa ketika probabilitas keberhasilan dan kegagalan sama besar, distribusi akan membentuk simetri sempurna.

Sebaliknya, distribusi dengan $p=0.7$ terlihat condong ke kiri (left-skewed). Sebagian besar nilai keberhasilan terkonsentrasi di angka lebih tinggi (sekitar $k=7$), mencerminkan bahwa kemungkinan terjadinya keberhasilan lebih tinggi dibandingkan kegagalan. Distribusi ini memiliki ekor yang lebih panjang di sisi kiri, menunjukkan bahwa nilai keberhasilan rendah menjadi semakin jarang terjadi.

Perbandingan ini mempertegas pengaruh parameter p terhadap bentuk distribusi Binomial. Semakin besar nilai p , maka distribusi cenderung bergeser ke kanan dan menjadi lebih miring ke kiri. Sebaliknya, jika p mendekati 0, distribusi akan bergeser ke kiri dan miring ke kanan. Fakta ini penting dalam interpretasi hasil statistik, terutama dalam konteks pemodelan kejadian biner di dunia nyata.

Dari sisi simulasi, baik distribusi dengan $p=0.5$ maupun $p=0.7$ menunjukkan hasil empiris yang sangat dekat dengan distribusi teoretis masing-masing. Hal ini memperkuat bahwa pendekatan Monte Carlo efektif untuk merepresentasikan dan memverifikasi distribusi probabilitas diskrit dengan berbagai skenario parameter.

KESIMPULAN

Simulasi Monte Carlo menggunakan Python berhasil mereproduksi distribusi Binomial teoretis dengan baik. Uji Chi square dan analisis selisih maksimum

memperlihatkan bahwa data sintetis mengikuti distribusi Binomial yang ditentukan. Nilai p value sebesar 0.6704 menunjukkan tidak ada perbedaan signifikan antara distribusi empiris dan teoretis.

Dengan demikian, pendekatan ini dapat dijadikan sebagai alat bantu edukatif untuk mempelajari distribusi probabilitas diskrit secara lebih interaktif dan visual. Selain itu, pendekatan ini juga relevan digunakan dalam pengujian model statistika berbasis data sintetis dalam riset terapan.

Metode ini juga memperkuat pemahaman mahasiswa tentang distribusi Binomial dan teknik simulasi. Dalam jangka panjang, pendekatan berbasis Python dan simulasi Monte Carlo dapat dikembangkan untuk mengkaji distribusi lain seperti Poisson, Geometrik, dan distribusi kontinyu seperti Normal. Simulasi ini juga membuka peluang untuk mengembangkan eksperimen virtual yang dapat diakses luas melalui platform pembelajaran daring.

Secara keseluruhan, penggunaan Python dalam simulasi distribusi probabilitas menjadi jembatan antara teori dan praktik, memungkinkan mahasiswa dan peneliti untuk bereksperimen dan mengeksplorasi konsep statistik secara menyeluruh. Ini mendukung pengembangan literasi data dan kemampuan berpikir kuantitatif di era digital.

DAFTAR PUSTAKA

- Diana. (2017). DISTRIBUSI BINOMIAL SEBAGAI ESTIMASI PROBABILITAS KESUKSESAN PADA UJI COBA KUALITAS LAYANAN SISTEM INFORMASI. *Jurnal Ilmiah Matrik*, Vol.19(3), 227-228.
- Manurung, R., Ariswoyo, S., Sembiring, P. (2013). PERBANDINGAN DISTRIBUSI BINOMIAL DAN DISTRIBUSI POISSON DENGAN PARAMETER YANG BERBEDA. *Saintia Matematika*, vol. 1(3), 299-300.
- Nurhaliza, Rahmi, D., Kurniati, A., Yuniati, S. (2024). Model Distribusi Binomial dalam Mengukur Probabilitas Keberhasilan Uji Coba Kualitas Layanan Sistem Informasi. *Jurnal Teknologi dan Manajemen Industri Terapan (JTMIT)*, Vol. 3(4), 405-406.
- Prasetyo, A., A., Dianto, I., K., Andriani, L., Syarif, M., B., Ershaad, M., F., Syauqi, R., I., & Ayubi, S., A. (2021). Analisis Tingkat Keberhasilan Dan Kegagalan Dalam Pembuatan Tempe Dengan Menerapkan Distribusi Binomial. *Bulletin of Applied Industrial Engineering Theory*, vol. 2(2), 91-92.
- Yanti, R., Suryani, I., Putri, I. (2024). *Buku Ajar Statistik dan Probabilitas Dasar*. Jogjakarta: ANGGOTA IKAPI.
- Hasyim, W. (2025). Analisis Sensitivitas Kuat Leleh Tulangan Terhadap Kuat Lentur Balok Beton Bertulang. 15(01), 189–200.
- Hutahaean, H. D. (2018). Analisa Simulasi Monte Carlo Untuk Memprediksi Tingkat Kehadiran Mahasiswa dalam Perkuliahan. *Journal of Informatic Pelita Nusantara*, 3(1), 41–45.
- Maulana, S. A. (2024). PENERAPAN METODE MONTE CARLO DALAM MEMPREDIKSI PRODUKSI DAGING SAPI DI PROVINSI SULAWESI UTARA. 8(4), 4424–4429.
- Mulia, J. R., & Afif, A. (2024). Simulasi Monte Carlo untuk Memprediksi Jumlah Penerimaan Bantuan Sosial Pangan. *JR : Jurnal Responsive Teknik Informatika*, 7(02), 123–132. <https://doi.org/10.36352/jr.v7i02.758>