Distribusi Peluang Diskrit

Distribusi Uniform Diskrit, Binomial, dan Geometri

3.1 Distribusi Uniform Diskrit, Binomial, dan Geometri

	Uniform Diskrit	Binomial	Geometri
Karakteristik distribusi	Distribusi ini menghasilkan suatu luaran acak dari sekian banyak titik sampel yang masing-masing memiliki tingkat kemunculan atau peluang yang sama.	Distribusi ini menghitung banyaknya percobaan acak identik dan saling bebas yang sukses dari antara sekian banyak percobaan yang telah ditentukan sebelumnya.	Distribusi ini menyatakan banyak rentetan dari suatu percobaan acak identik dan saling bebas yang diperlukan hingga diperoleh suatu hasil yang diharapkan.
Fenomena Ketidakpastian	Luaran dari sebuah dadu yang dilempar.	Banyaknya angka 3 yang muncul dari 9 dadu yang dilempar.	Banyaknya pelemparan dadu yang dibutuhkan hingga muncul mata dadu 3 untuk pertama kalinya.
Notasi	$X \sim U(a, b)$	$X \sim \text{Bin}(N, p)$	$X \sim \text{Geom}(p)$
Parameter	Batas bawah a; Batas atas b.	Banyaknya percobaan N; Peluang sukses p.	Peluang sukses p.

CALVIN INSTITUTE OF TECHNOLOGY

Support	$\{a, a+1, \ldots, b\}$	$\{0, 1, 2, \ldots, N\}$	$\{1, 2, \ldots\}$
Intepretasi Kejadian $\{X = k\}$	Luaran acak yang muncul adalah bernilai k	Dari N kali percobaan Bernoulli, terdapat k di antaranya yang sukses.	Percobaan Bernoulli yang dilakukan sebanyak k kali dengan $k-1$ di antaranya gagal dan percobaan ke-k yang pertama kali sukses.
PMF: $f_X(x)$	$\dfrac{1}{b-a+1}$	$\binom{N}{x} p^x (1-p)^{N-x}$	$p(1-p)^{x-1}$
CDF: $F_X(x)$	$\dfrac{x}{b-a+1}$	$\sum_{k \le x} \binom{N}{x} p^k (1-p)^{N-k}$	$1-(1-p)^x$
Ciri khas bentuk PMF $f_X(x)$	Rata di semua titik.	Berbentuk ’lonceng’.	Monoton turun.

CALVIN INSTITUTE OF TECHNOLOGY

Uniform diskrit: $f_X(x) = \dfrac{1}{b-a+1}$, $S_X := \{\underbrace{a, a+1, a+2, \dots, b}_{\text{barisan aritmatika}}\}$

$U_1 = a$ & $U_n = b$

beda=1

$U_n = U_1 + (n-1)\text{beda}$

$b = a + (n-1) \cdot 1$

$n = b-a+1$

- Apakah $f_X(x)$ fs. peluang? $\sum_{x \in S_X} f_X(x) \overset{??}{=} 1$.

total peluang=1

Bukti: $\sum_{x \in S_X} \underbrace{\frac{1}{b-a+1}}_{\text{konstanta}} = \frac{1}{b-a+1} \sum_{x \in S_X} 1$

$= \frac{1}{b-a+1} (\underbrace{1+1+\dots+1}_{b-a+1 \text{ kali}}) = \frac{1}{\cancel{b-a+1}} [\cancel{(b-a+1)} \cdot 1] = 1$ //

$\therefore f_X(x) = \frac{1}{b-a+1}$ adalah fungsi massa peluang. (memenuhi)

	Uniform Diskrit	Binomial	Geometri
Ekspektasi	$\dfrac{a+b}{2}$ ✓ derimana?	Np	$\dfrac{1}{p}$
Variansi	$\dfrac{(b-a+1)^2-1}{12}$	$Np(1-p)$	$\dfrac{1-p}{p^2}$

Bukti. Diketahui $f_X(x) = \dfrac{1}{b-a+1}$, $S_X := \{a, a+1, \ldots, b\}$

$$\mathbb{E}[X] = \sum_{x \in S_X} x \cdot \underbrace{\frac{1}{b-a+1}}_{f_X(x)} = a \cdot \frac{1}{b-a+1} + (a+1) \cdot \frac{1}{b-a+1} + \cdots + b \cdot \frac{1}{b-a+1}$$

$$= \frac{a+(a+1)+(a+2)+\cdots+b}{b-a+1} = \frac{\left[(\overset{u_1}{a}+\overset{u_n}{b}) \cdot \frac{\cancel{b-a+1}^{\,n}}{2}\right]}{\cancel{b-a+1}} = \frac{a+b}{2} \quad // \quad \blacksquare$$

deret Aritmatika

1. Kita bisa menerima lampu merah pada sebuah lalu lintas dengan peluang 60% secara saling bebas. Dari sepuluh lampu lalu lintas yang harus dilewati, jika ternyata dapat setidaknya empat di antara yang kena lampu merah, maka kita terlambat. Berapa peluang kita terlambat?

Misalkan X variabel acak yg menyatakan byk lampu merah yg dilalui

$X \sim \text{binom}\left(10, \frac{6}{10}\right)$ → ditanya: $\mathbb{P}(\underbrace{X \geq 4}_{\text{terlambat}}) = ?$

N : byk percobaan

p : peluang "sukses" : kena lampu merah

$\mathbb{P}(X \geq 4) = 1 - \mathbb{P}(X < 4) = 1 - \mathbb{P}(X \leq 3)$ fungsi dist. kumulatif $F_X(3)$

$= 1 - [\mathbb{P}(X=0) + \mathbb{P}(X=1) + \mathbb{P}(X=2) + \mathbb{P}(X=3)]$

$= 1 - \left[\binom{10}{0}\left(\frac{6}{10}\right)^0\left(\frac{4}{10}\right)^{10} + \binom{10}{1}\left(\frac{6}{10}\right)^1\left(\frac{4}{10}\right)^9 + \dots + \binom{10}{3}\left(\frac{6}{10}\right)^3\left(\frac{4}{10}\right)^7\right]$

"sukses" "gagal"

$\approx 0{,}9452 = 94{,}52\%$ (kalkulator saintifik)

modal:

$f_X(x) = \mathbb{P}(X = x)$

$F_X(x) = \mathbb{P}(X \leq x)$

$f_X(x) = \mathbb{P}(X = x) = \binom{N}{x} p^x (1-p)^{N-x}$

https://colab.research.google.com/drive/172YwpKAHLioGUx_mtzk0L7ao-YjRf7qf?usp=sharing.

1. Kita bisa menerima lampu merah pada sebuah lalu lintas dengan peluang 60% secara saling bebas. Dari sepuluh lampu lalu lintas yang harus dilewati, jika ternyata dapat setidaknya empat di antara yang kena lampu merah, maka kita terlambat. Berapa peluang kita terlambat?

```
# X ~ Bin(N, p)
# N adalah jumlah percobaan, p adalah probabilitas keberhasilan dalam setiap percobaan.
N, p = 10, 0.6
x = 3

# Menghitung probabilitas kumulatif P(X <= x), yaitu probabilitas X <= x.
print("P( X <= ", x, ") = ", binom.cdf(x, N, p))
```

```
P( X <=  3 ) =  0.05476188160000002
```

$\therefore \mathbb{P}(X \geq 4) = 1 - \mathbb{P}(X \leq 3) \overset{\text{Python}}{=} 1 - \underbrace{0{,}0547618\ldots}_{\text{CDF}} \approx 0{,}9452$ //

2. Sebuah *playlist* berisi seratus lagu dan lagu-lagu tersebut akan diputar secara acak. Salah satu lagu di antaranya adalah sebuah lagu favorit. Berapa kali pemutaran acak paling sedikit dilakukan untuk menjamin lagu tersebut akan pernah terputar dengan peluang setidaknya 90%?

"pertama kali muncul

Misalkan X menyatakan banyaknya pemutaran lagu favorit (muncul pertama kali)

$X \sim \text{Geom}(p) \rightarrow \text{Geom}\left(\frac{1}{100}\right) \rightarrow$ ditanya $\mathbb{P}(X \leq \boxed{y}) \geq 90\%$

p: peluang kejadian

y: minimum

kejadian : dari y kali pemutaran ada lagu favorit.

90%: α

$\rightarrow$ Invers CDF $= \text{ppf}(90\%) = \text{ppf}(0,9)$

$= 230$

$\boxed{} = \mathbb{P}(X \leq \text{tahu})$: CDF

$\text{tahu} = \mathbb{P}(X \leq \boxed{})$: Invers CDF / ppf (percent point function)

$\boxed{y} = \text{ppf}(\boxed{90\%})$

$\text{cdf}(\boxed{y}) = \boxed{90\%}$

α

no.2

```
[ ] # Menentukan nilai k terkecil sehingga P(X <= k) >= alpha.
    # Fungsi ppf (percent point function) digunakan untuk menghitung nilai tersebut.
    p = 0.01
    alpha = 0.9
    print("ppf_X(", alpha, ") = ", geom.ppf(alpha, p))
```

```
ppf_X( 0.9 ) =  230.0
```

3. Diketahui pada sebuah *dataset* ada kira-kira 2% *instance data* yang memiliki kecacatan. Suatu sistem mengakses *instance dataset* tersebut satu demi satu. Pada akses *instance* ke berapakah dapat kita harapkan *dataset* tersebut akan menemukan kecacatan untuk pertama kalinya?

~ ekspektasi!!

Misalkan Z menyatakan banyaknya pemeriksaan dataset hingga ditemukan dataset yang cacat.

$Z \sim \text{Geom}\left(\frac{2}{100}\right) \to p$

kira²

$\to$ ditanya: $\mathbb{E}[Z] = \frac{1}{p} = \frac{1}{2/100} = 50$ //

∴ Pada akses instance ke-50, dapat ditemukan kecacatan utk pertama kalinya.

Sebuah kuis terdiri dari 15 soal pilihan ganda. Masing-masing soal memiliki lima opsi jawaban. Andina menjawab setiap soal ini secara acak dan saling bebas untuk setiap nomornya.

(a) Berapa peluang sebuah soal terjawab dengan benar?

(b) Berapa ekspektasi nilai kuis yang dijawab Andina?

Jawab.

(a) Misalkan X menyatakan pilihan jawaban. $S_X \in \{A, B, C, D, E\}$

$X \sim U\{A, B, C, D, E\}$

$\mathbb{P}[\text{Jawab Benar}] = \frac{1}{5}$ //

(b) Misalkan Y menyatakan banyak jawaban yg benar.

$Y \sim \text{Bin}(\underbrace{15}_{N}, \underbrace{\tfrac{1}{5}}_{\text{Peluang "sukses"} \to \text{Jawab benar}})$

$\mathbb{E}[Y] = Np = 15 \cdot \frac{1}{5} = 3$ soal //

Sebuah kuis terdiri dari 15 soal pilihan ganda. Masing-masing soal memiliki lima opsi jawaban. Andina menjawab setiap soal ini secara acak dan saling bebas untuk setiap nomornya.

(c) Untuk lulus, mahasiswa harus benar setidaknya 10 soal. Berapa peluang Andina bisa lulus?

(d) Jika seandainya diperbolehkan mengambil kuis berkali-kali hingga mahasiswa lulus, berapa kali percobaan yang diperlukan agar Andina lulus dengan peluang lebih besar dari 90%?

Jawab: (c) Variabel acak dan distribusi sama spt (a) → ditanya: $\mathbb{P}(Y \geq 10)$?

$$\mathbb{P}(Y \geq 10) = 1 - \mathbb{P}(Y < 10) = 1 - \mathbb{P}(Y \leq 9) = 1 - \underbrace{F_Y(9)}_{0{,}9998867} = 0{,}0001\ldots$$

Python

(d) Misalkan Z adalah banyak kuis yg diperlukan hingga lulus

$$Z \sim \text{Geom}(\underbrace{0{,}0001\ldots}_{p:\ \text{peluang lulus di (c)}}) \longrightarrow \text{ppf} = 20336$$

Python

```python
[8]  # X ~ Bin(N, p)
     # N adalah jumlah percobaan, p adalah probabilitas keberhasilan dalam setiap percobaan.
     N, p = 15, 0.2
     x = 9

     # Menghitung probabilitas kumulatif P(X <= x), yaitu probabilitas X <= x.
     print("P( X <= ", x, ") = ", binom.cdf(x, N, p))
```

```
P( X <=  9 ) =  0.999886774337536
```

```python
[9]  print(" 1 - P( X <= ", x, ") = ", 1-binom.cdf(x, N, p))
```

```
 1 - P( X <=  9 ) =  0.000113225662463967150
```

```python
[10] # Menentukan nilai k terkecil sehingga P(X <= k) >= alpha.
     # Fungsi ppf (percent point function) digunakan untuk menghitung nilai tersebut.
     p = 1-binom.cdf(x, N, p)
     alpha = 0.9
     print("ppf_X(", alpha, ") = ", geom.ppf(alpha, p))
```

```
ppf_X( 0.9 ) =  20336.0
```

CALVIN INSTITUTE OF TECHNOLOGY

$$\underbrace{1-(1-p)^y}_{\mathbb{P}(Y \geq y)} \geq 90\%$$

$$1-(1-p)^y \geq 0{,}9$$

$$0{,}1 \geq (1-p)^y$$

$$0{,}1 \geq (\boxed{0{,}99988677433753 6})^y$$

aproksimasi!

$$\ln 0{,}1 \leq y \cdot \ln(\boxed{})$$

$$20335.09615\ldots = \frac{\ln 0{,}1}{\ln \boxed{}} \leq y \rightsquigarrow y_{min} = 20336$$

approks!

Distribusi Peluang Diskrit

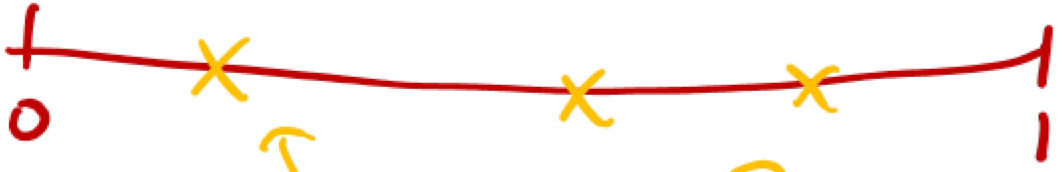
Distribusi Poisson

3.2 Distribusi Poisson

kejadian Poisson

Diketahui bahwa rata-rata maskapai penerbangan Udara Asia mengalami tiga kali kecelakaan per tahunnya. Berapa peluang tahun ini Udara Asia tidak mengalami kecelakaan? Ini merupakan contoh penerapan distribusi Poisson dengan variabel acak diskrit dalam rentang waktu yang kontinu.

mengamati sesuatu didalam interval waktu kontinu.



3 kejadian dalam 1 tahun

$\lambda = 3$

Misalkan X banyak kecelakaan dalam satu tahun

$$\rightarrow \mathbb{P}(\underbrace{X=0}_{\text{tdk ada kecelakaan}}) = f_X(0) = e^{-3} \cdot \frac{3^0}{0!} = \frac{1}{e^3} \approx \underline{\underline{0,0497}}$$

$$f_X(x) = e^{-\lambda} \cdot \frac{\lambda^x}{x!}$$

Definisi. Misalkan X berdistribusi Poisson $X \sim$ Poi(λ) dengan $S_X = \{0, 1, 2, \ldots\}$. Variabel acak diskrit X ini menyatakan banyak kejadian dalam suatu interval. Fungsi massa peluang, fungsi distribusi, ekspektasi, dan variansinya dapat didefinisikan sebagai

(a) $f_X(x) = e^{-\lambda} \cdot \dfrac{\lambda^x}{x!}$,

PMF

(b) $F_X(x) = \underbrace{\sum_{k=0}^{x} e^{-\lambda} \cdot \dfrac{\lambda^k}{k!}}_{\text{Jumlahin tiap support}}$,

CDF

(c) $\mathbb{E}[X] = \lambda$ dan Var$[X] = \lambda$.

Ekspektasi dan variansi

Parameter dalam distribusi Poisson adalah λ yang dapat diintepretasikan sebagai

- λ adalah rata-rata banyak kejadian dalam interval,
- $\dfrac{1}{\lambda}$ adalah laju terjadinya kejadian.

↳ seberapa cepat kejadian tjd dalam suatu interval.

~ waktu antar kejadian!

Untuk menjelaskan atau membuktikan hal-hal di atas, kita perlu mengingat deret (Maclaurin)

$$e^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \ldots = \sum_{k=0}^{\infty} \frac{x^k}{k!}. \rightarrow e^{\square} = \sum_{k=0}^{\infty} \frac{\square^k}{k!}$$

Dapatkah Anda menunjukkan sifat fungsi massa peluang bahwa $\sum_{x=0}^{\infty} f_X(x) = 1$?

$$f_X(x) = e^{-\lambda} \cdot \frac{\lambda^x}{x!}$$

Bukti:

$$\sum_{x=0}^{\infty} f_X(x) = \sum_{x=0}^{\infty} \underbrace{e^{-\lambda}}_{\text{konstanta}} \frac{\lambda^x}{x!} = e^{-\lambda} \cdot \underbrace{\sum_{x=0}^{\infty} \frac{\boxed{\lambda}^x}{x!}}_{} = e^{-\lambda} \cdot e^{\boxed{\lambda}} = e^0 = 1 \quad //$$

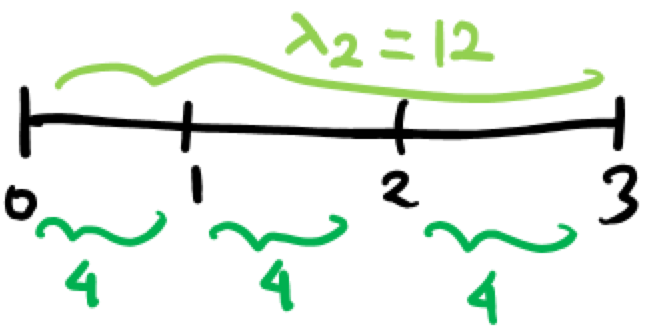
Sebuah ATM BTA di lokasi tertentu diketahui digunakan layanannya sebesar kira-kira empat pelanggan per jam, mengikuti distribusi Poisson. Berapa peluang tidak ada pelanggan yang datang selama satu jam? Berapa peluang tidak lebih dari dua pelanggan selama tiga jam?

$= \lambda_1$

Jawab: (1) Misalkan X_1 menyatakan banyak pelanggan dalam satu jam.

$$X_1 \sim \text{Poi}(\underbrace{4}_{\lambda_1}) \rightarrow \text{dit: } \mathbb{P}(X_1 = 0) = f_{X_1}(0) = e^{-4} \cdot \frac{4^0}{0!} = \frac{1}{e^4} \approx 0{,}0183 \; //$$

(2)



Misalkan X_2 menyatakan banyak pelanggan dalam tiga jam

$$X_2 \sim \text{poi}(\underbrace{12}_{\lambda_2})$$

$$\rightarrow \text{dit: } \mathbb{P}(X_2 \leq 2) = \mathbb{P}(X_2 = 0) + \mathbb{P}(X_2 = 1) + \mathbb{P}(X_2 = 2)$$

$$= e^{-12} \cdot \frac{12^0}{0!} + e^{-12} \cdot \frac{12^1}{1!} + e^{-12} \cdot \frac{12^2}{2!} = \frac{1}{e^{12}} [1 + 12 + 72]$$

$$\approx 0{,}0005 \; //$$

https://colab.research.google.com/drive/172YwpKAHLioGUx_mtzk0L7ao-YjRf7qf?usp=sharing.

Sebuah ATM BTA di lokasi tertentu diketahui digunakan layanannya sebesar kira-kira empat pelanggan per jam, mengikuti distribusi Poisson. Berapa peluang tidak ada pelanggan yang datang selama satu jam? Berapa peluang tidak lebih dari dua pelanggan selama tiga jam?

```
[1] from scipy.stats import poisson

    # X ~ Poi(mu)
    # mu adalah rata-rata jumlah kejadian dalam interval tertentu.
    mu = 12
    x = 2

    # Menghitung probabilitas kumulatif P(X <= x), yaitu probabilitas X <= x.
    print("P( X <= ", x, ") = ", poisson.cdf(x, mu))
```

```
P( X <=  2 ) =  0.0005222580500328981
```

Teorema. Secara teoritis, distribusi Binomial ($\mathrm{Bin}(N,p)$) memiliki hubungan dengan distribusi Poisson ($\mathrm{Poi}(\lambda)$). Saat $N \to \infty$, maka $Np \to \lambda$ konstan dan dapat dituliskan sebagai

$$\begin{aligned}\lim_{N\to\infty} \mathbb{P}(\mathrm{Bin}(N,p) = k) &= \lim_{N\to\infty} \binom{N}{k} p^k (1-p)^{N-k} \\ &= e^{-\lambda} \cdot \frac{\lambda^k}{k!} \\ \lim_{N\to\infty} \mathbb{P}(\mathrm{Bin}(N,p) = k) &= \mathbb{P}(\mathrm{Poi}(\lambda) = k).\end{aligned}$$

$$\lim_{N\to\infty} \binom{N}{k} p^k (1-p)^{N-k} = e^{-\lambda} \cdot \frac{\lambda^k}{k!}$$???

Bagi yang berminat, silahkan pelajari video: https://www.youtube.com/watch?v=sVBOSwT5K8I

CALVIN INSTITUTE OF TECHNOLOGY