LLN, CLT, & MLE

Universitas Al Azhar Indonesia

Ali Akbar Septiandri

Universitas Al Azhar Indonesia

July 2, 2019

Ulasan

Distribusi Gaussian/Normal

- Salah satu yang paling sering muncul untuk variabel kontinu
- Berhubungan dengan central limit theorem
- Dituliskan sebagai $X \sim \mathcal{N}(\mu, \sigma^2)$

Fakta

$$X \sim \mathcal{N}(\mu, \sigma^2)$$

PDF

$$f(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$$

Ekspektasi

$\mathbb{E}[X] = \mu$

Variansi

$Var[X] = \sigma^2$

Law of Large Numbers

Contoh

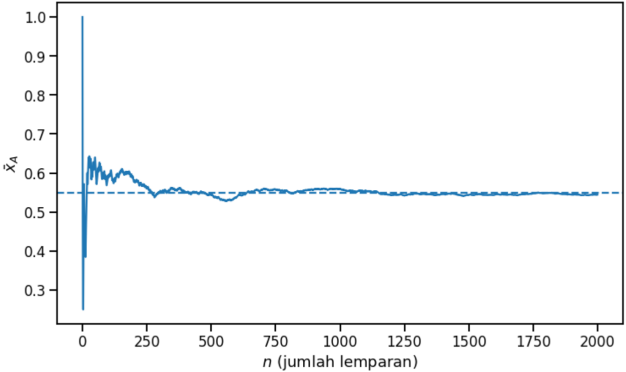
Anda diberikan koin yang bias dengan peluang munculnya angka adalah 0.55, i.e. $X \sim \mathit{Bernoulli}(0.55)$. Hitunglah nilai rata-rata kemunculan A setelah 1 lemparan, 2 lemparan, ..., hingga 2000 lemparan.

Ilustrasi

Hasil dari 4 lemparan adalah A G A A. Maka, nilai rata-rata yang Anda dapatkan adalah

- Lemparan 1: $\bar{x}_A = \frac{1}{1} = 1$,
- Lemparan 2: $\bar{x}_A = \frac{1+0}{2} = 0.5$,
- Lemparan 3: $\bar{x}_A = \frac{1+0+1}{3} = 0.67$, dan
- Lemparan 4: $\bar{x}_A = \frac{1+0+1+1}{4} = 0.75$.

Ilustrasi



Gambar: Nilai rata-rata dari sampel semakin mendekati nilai rata-rata sebenarnya seiring dengan jumlah sampelnya semakin besar

Law of Large Numbers

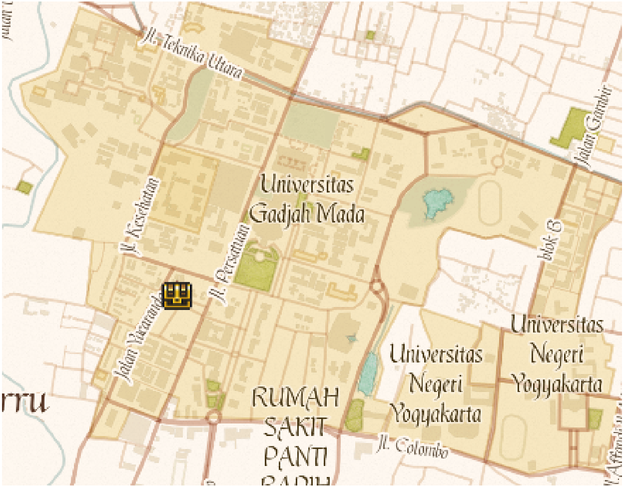
Nilai **rata-rata dari sampel** (*sample mean*) akan konvergen ke **nilai rata-rata sebenarnya** jika kita menggunakan *cukup banyak* sampel.

$$\lim_{n\to\infty} P(|\bar{X} - \mu| \geq \epsilon) = 0$$

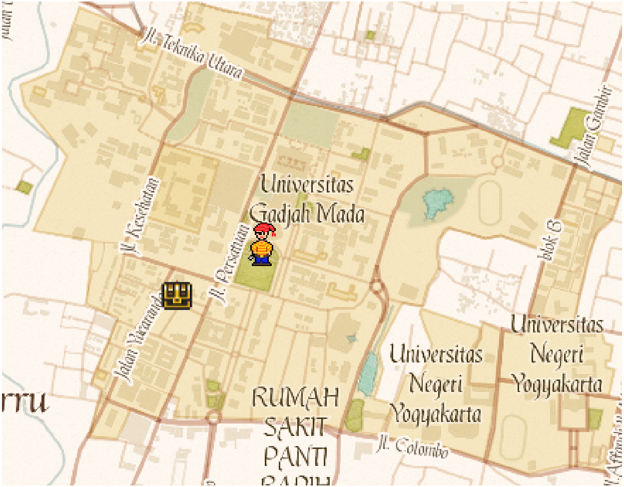
$$P(\lim_{n\to\infty}(\bar{X}) = \mu) = 1$$

Mari mencari harta karun!

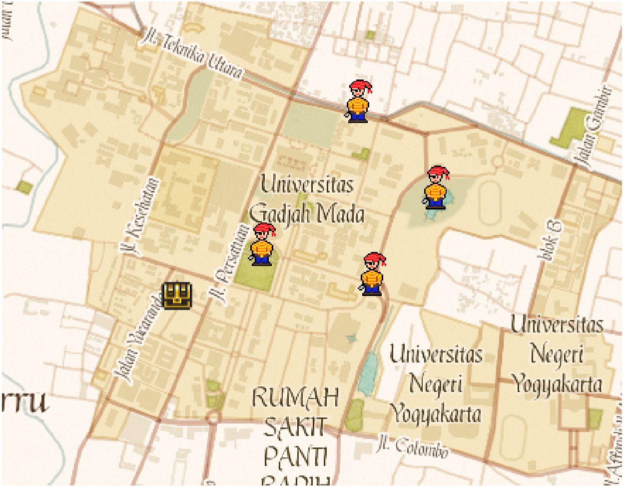
Mencari Harta Karun



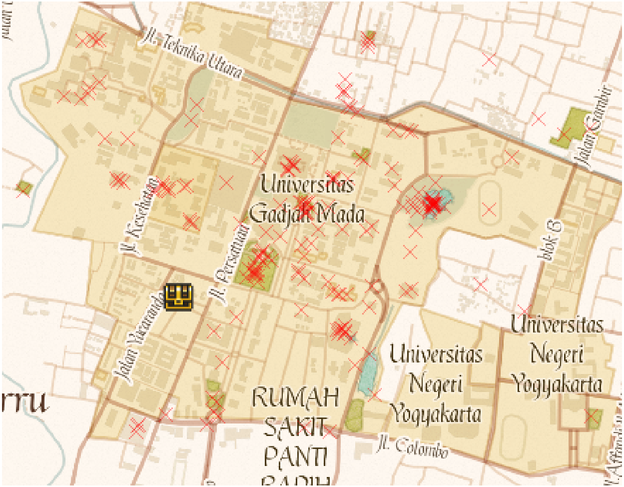
Mencari Harta Karun



Mencari Harta Karun



Mencari Harta Karun



Mencari Harta Karun

Gambar: Rata-rata jaraknya juga menjadi konvergen!

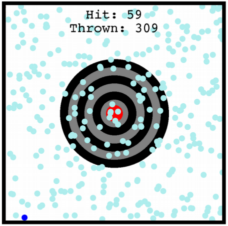
Law of Large Numbers

Nilai **rata-rata dari sampel** (*sample mean*) akan konvergen ke **nilai rata-rata sebenarnya** jika kita menggunakan *cukup banyak* sampel.

$$\lim_{n \to \infty} P(|\bar{X} - \mu| \geq \epsilon) = 0$$

$$P(\lim_{n \to \infty} (\bar{X}) = \mu) = 1$$

Frequentist probability and LLN

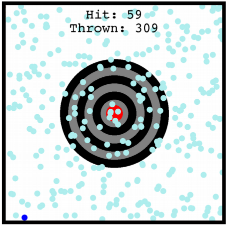


$$P(E) = \lim_{n \to \infty} \frac{\#(E)}{n}$$

indicator variables $X_i = \begin{cases} 1 & \text{if } E \text{ occurs on trial } i \\ 0 & \text{otherwise} \end{cases}$

$$\lim_{n \to \infty} \left(\frac{\overbrace{\boxed{X_1 + X_2 + \cdots + X_n}}^{=\#(E)}}{n} \right) = E[X_i]$$

Frequentist probability and LLN



$$P(E) = \lim_{n \to \infty} \frac{\#(E)}{n}$$

indicator variables $X_i = \begin{cases} 1 & \text{if } E \text{ occurs on trial } i \\ 0 & \text{otherwise} \end{cases}$

$$\lim_{n \to \infty} \left(\frac{\overbrace{\boxed{X_1 + X_2 + \cdots + X_n}}^{=\#(E)}}{n} \right) = P(E)$$

Gambler's fallacy

"I'm due for a win!"

Dampaknya apa?

Esensi LLN

1. Bagian esensial dari **simulasi, statistik, dan sains**
2. Saat melakukan **eksperimen** untuk mendapatkan data dan **menggeneralisasi hasilnya**, kita menggunakan LLN!

Central Limit Theorem

Ketika kita melempar koin 10 kali, kita dapat melihat **setiap lemparan** sebagai suatu peubah acak yang mengikuti distribusi Bernoulli dengan $\theta = 0.5$, i.e.

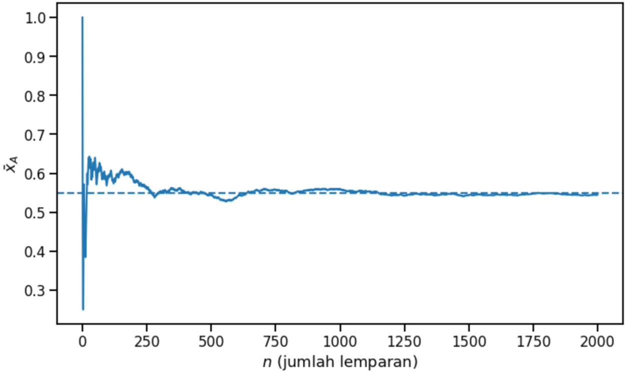
$$X_i \sim Bernoulli(0.5)$$

CLT

Central limit theorem menyatakan bahwa **penjumlahan** dan **rata-rata** dari peubah acak IID (*independent and identically distributed*) mengikuti **distribusi normal**. Dengan kata lain,

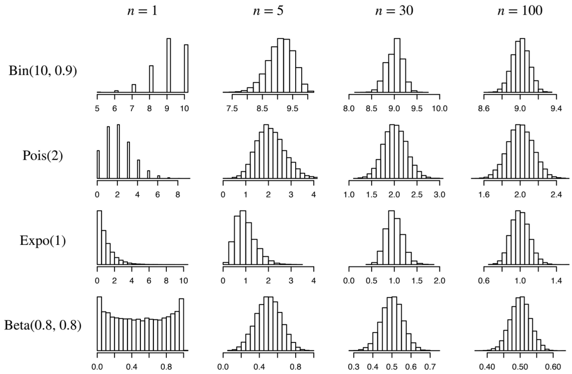
$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \sim \mathcal{N}(\mu, \frac{\sigma^2}{n})$$

Ingat!



Gambar: Distribusi dari rata-rata dari sampel mengikuti distribusi normal

Generalisasi



Gambar: Apapun distribusi awalnya, distribusi dari rata-rata sampel menjadi normal

Penjumlahan $\rightarrow$ CLT

Asumsikan ada variabel Y sedemikian sehingga

$$Y = \sum_{i=1}^{n} X_i$$

Ingat bahwa

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i \sim \mathcal{N}(\mu, \frac{\sigma^2}{n})$$

maka

$$Y = n \cdot \bar{X} \sim \mathcal{N}(n\mu, n\sigma^2)$$

saat $n \rightarrow \infty$.

Contoh

Example

Situs Anda mendapatkan $W \sim Poi(5/3)$ requests per detik. Server akan *crash* jika mendapatkan 120 atau lebih requests dalam satu menit.

Solusi 1

Nilai λ dapat kita sesuaikan agar unit waktunya sama, i.e. $X \sim Poi(\lambda = 60 \cdot 5/3 = 100)$. Jawabannya dapat dihitung dengan CDF dari Poisson:

$$P(Y \geq 120) = 1 - \sum_{i=0}^{120} \frac{e^{-100}(100)^i}{i!} \approx 0.0282$$

Contoh

Diketahui bahwa $Poi(\lambda) \approx \mathcal{N}(\lambda, \lambda)$ untuk nilai λ yang besar

Example

Situs Anda mendapatkan $W \sim Poi(5/3)$ requests per detik. Server akan *crash* jika mendapatkan 120 atau lebih requests dalam satu menit.

Solusi 2

Asumsikan Y adalah jumlah dari 60 RV yang IID dengan $\mu = \sigma^2 = 5/3$. Dengan demikian, hampiran untuk X adalah $X \approx Y \sim \mathcal{N}(60 \cdot (5/3), 60 \cdot (5/3))$.

$$
\begin{aligned}
P(X \geq 120) \approx P(Y > 119.5) &= P(\frac{Y - 100}{10} > \frac{119.5 - 100}{10}) \\
&= P(Z > 1.95) \\
&= 1 - \Phi(1.95) \approx 0.0256
\end{aligned}
$$

Maximum Likelihood Estimation

Likelihood

- **Asumsi** bahwa suatu peubah acak mengikuti distribusi tertentu

Likelihood

- **Asumsi** bahwa suatu peubah acak mengikuti distribusi tertentu
- Kita punya **data**, tapi tidak tahu distribusi dan parameter sebenarnya

Likelihood

- **Asumsi** bahwa suatu peubah acak mengikuti distribusi tertentu
- Kita punya **data**, tapi tidak tahu distribusi dan parameter sebenarnya
- Harus *cocokologi*?

Parameters

$$\theta$$

$$X \sim \begin{cases} Ber(p) & \theta = p \\ Poi(\lambda) & \theta = \lambda \\ Uni(a, b) & \theta = [a, b] \\ \mathcal{N}(\mu, sigma^2) & \theta = [\mu, \sigma^2] \end{cases}$$

Likelihood

$$L(\theta) = P(X_1, ..., X_n|\theta)$$

Dibaca: Berapa peluang mengobservasi datum pertama dan datum kedua dan ... dan datum ke-n jika diketahui bahwa datanya mengikuti distribusi tertentu dengan parameter θ.

Karena biasanya $X_1, X_2, ..., X_n$ diasumsikan IID, maka dapat ditulis juga sebagai

$$L(\theta) = \prod_{i=1}^{n} P(X_i|\theta)$$

“Jika diasumsikan bahwa data mengikuti distribusi tertentu, berapa nilai parameter yang dapat memaksimalkan nilai fungsi likelihood-nya.”

- Maximum Likelihood Estimation (MLE)

Tahapan Mencari MLE

1. Definisikan fungsi likelihood

$$L(\theta) = P(X_1, ..., X_n|\theta)$$

2. Kenakan fungsi log

$$L(\theta) = \log L(\theta)$$

3. Maksimalkan fungsi ini berdasarkan parameternya

$$\frac{d}{d\theta} LL(\theta) = 0$$

Maximum likelihood for Bernoulli

The maximum likelihood *p* for **Bernoulli** random variables is the **sample mean**.

$$\hat{p} = \frac{1}{m}\sum_{i=1}^{m} X_i$$

Derivation: MLE for Bernoulli

1. Compute the likelihood.

$$\theta = p$$

$$L(\theta) = P(X_1, \ldots, X_m | \theta)$$

$$= \prod_{i=1}^{m} P(X_i | \theta)$$ don't forget: IID means independent!

$$= \prod_{i=1}^{m} \begin{cases} p & \text{if } X_i = 1 \\ (1-p) & \text{if } X_i = 0 \end{cases}$$

Derivation: MLE for Bernoulli

1. Compute the likelihood.

$$\theta = p$$

$$\begin{aligned} L(\theta) &= P(X_1, \ldots, X_m | \theta) \\ &= \prod_{i=1}^{m} P(X_i | \theta) \\ &= \prod_{i=1}^{m} \begin{cases} p & \text{if } X_i = 1 \\ (1-p) & \text{if } X_i = 0 \end{cases} \\ &= \prod_{i=1}^{m} p^{X_i} (1-p)^{1-X_i} \end{aligned}$$

don't forget: IID means independent!

Derivation: MLE for Bernoulli

2. Take its log.

$$\theta = p$$

$$L(\theta) = \prod_{i=1}^{m} \theta^{X_i} (1-\theta)^{1-X_i}$$

$$LL(\theta) = \log \prod_{i=1}^{m} \theta^{X_i} (1-\theta)^{1-X_i}$$

$$= \sum_{i=1}^{m} \log\left[\theta^{X_i} (1-\theta)^{1-X_i}\right]$$

$$= \sum_{i=1}^{m} \left[X_i \log\theta + (1-X_i)\log(1-\theta)\right]$$

Derivation: MLE for Bernoulli

3. Maximize this as a function of the parameters.

$$\theta = p$$

$$LL(\theta) = \sum_{i=1}^{m} \left[X_i \log \theta + (1 - X_i) \log (1 - \theta) \right]$$

$$\hat{\theta} = \hat{p} = \arg\max_{\theta} LL(\theta)$$

$$\frac{d}{d\theta} LL(\theta) = \sum_{i=1}^{m} \left[\frac{X_i}{\theta} - \frac{1 - X_i}{1 - \theta} \right]$$

$$= \frac{1}{\theta} \sum_{i=1}^{m} X_i - \frac{1}{1-\theta} \sum_{i=1}^{m} (1 - X_i)$$

$$= \left(\frac{1}{\theta} + \frac{1}{1-\theta} \right) \left(\sum_{i=1}^{m} X_i \right) - \frac{m}{1-\theta} = 0$$

$$\left(\frac{1-\theta}{\theta} + 1 \right) \left(\sum_{i=1}^{m} X_i \right) = m$$

$$\frac{1}{\theta} \left(\sum_{i=1}^{m} X_i \right) = m$$

$$\theta = \frac{1}{m} \left(\sum_{i=1}^{m} X_i \right)$$

MLE untuk Normal

- Karena distribusi normal menggunakan dua parameter, gunakan turunan parsial!
- Hasil akhirnya sangat familiar:

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

$$\hat{\sigma^2} = \frac{1}{n} \sum_{i=1}^{n} (X_i - \mu)^2$$

Coba cek Wikipedia untuk MLE dari berbagai distribusi!

Terima kasih