Distribusi Gaussian

Ali Akbar Septiandri

Universitas Al-Azhar Indonesia

aliakbars@live.com

April 16, 2020

Selayang Pandang

1. Peubah Acak Kontinu
2. Distribusi Gaussian
3. Maximum Likelihood Estimation
4. Aproksimasi

Peubah Acak Kontinu

Peubah Acak Kontinu

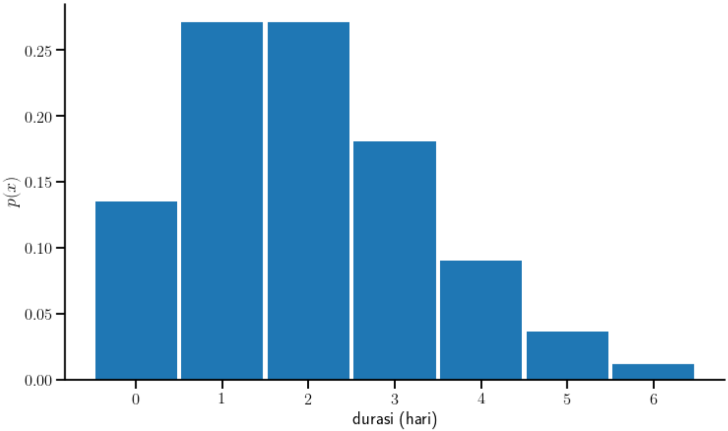
Peubah acak kontinu memiliki nilai berupa bilangan riil sehingga kita harus mengubah penjumlahannya menjadi integral.

$$p(a < X \leq b) = F(b) - F(a)$$
$$F(a) = \int_{-\infty}^{a} f(x)dx$$

Probability Density Function (PDF)

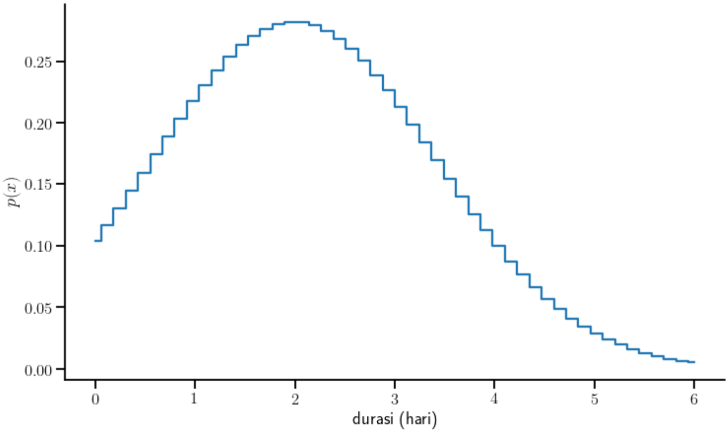
$$p(a < X \leq b) = \int_a^b f(x)dx$$

Ilustrasi PDF



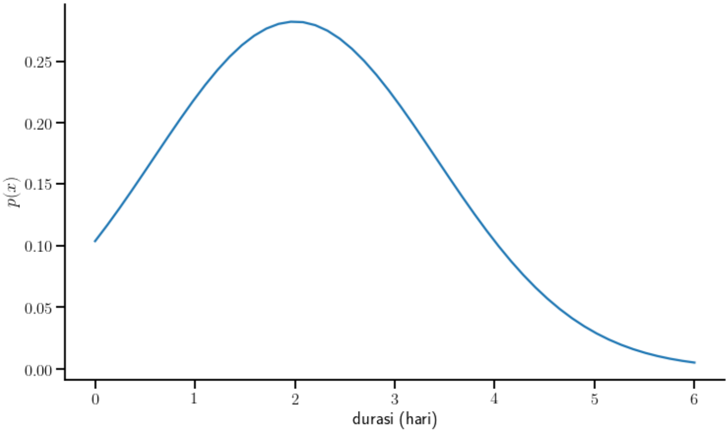
Gambar: Peluang durasi pengiriman barang

Ilustrasi PDF



Gambar: Peluang durasi pengiriman barang

Ilustrasi PDF



Gambar: Peluang durasi pengiriman barang

Probability Density Function (PDF)

Unit dari probabilitas dibagi dengan unit dari X.

$$p(a < X \leq b) = \int_a^b f(x)dx$$

Properti dari PDF

Integral dari PDF yang akan memenuhi aksioma probabilitas:

$$0 \leq \int_a^b f(x)dx \leq 1$$

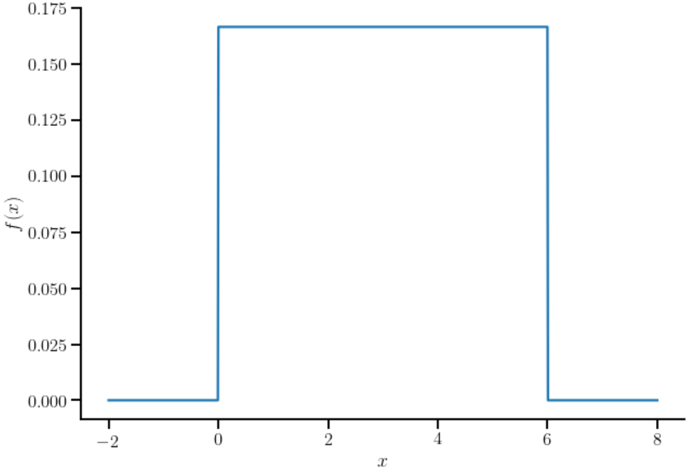
$$\int_{-\infty}^{\infty} f(x)dx = 1$$

Peubah Acak Seragam

$$X \sim Uni(\alpha, \beta)$$

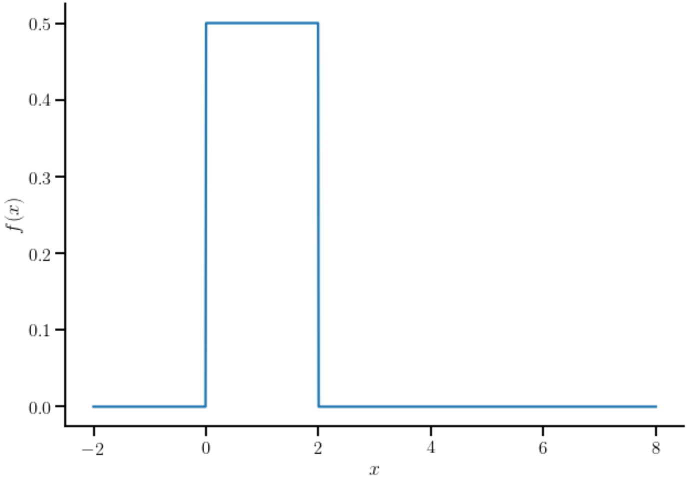
$$f(x) = \begin{cases} \frac{1}{\beta - \alpha} & \text{jika } x \in [\alpha, \beta] \\ 0 & \text{sebaliknya} \end{cases}$$

Ilustrasi Nilai PDF



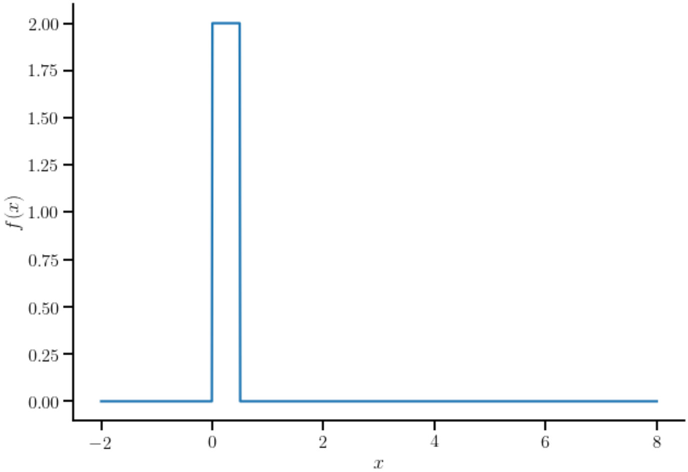
Gambar: Peluang durasi pengiriman barang

Ilustrasi Nilai PDF



Gambar: Peluang durasi pengiriman barang

Ilustrasi Nilai PDF



Gambar: Peluang durasi pengiriman barang

Jadi, $f(x)$ bukanlah probabilitas!

Distribusi Gaussian

Distribusi Gaussian/Normal

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

Apa itu "terdistribusi normal"?

- Dapat ditemukan dalam berbagai fenomena di alam, e.g. tinggi badan, berat badan, ...

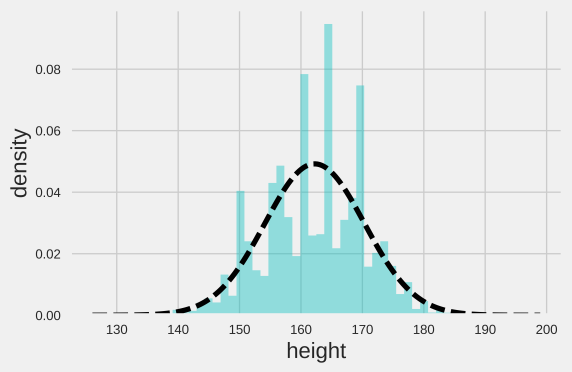
Apa itu "terdistribusi normal"?

- Dapat ditemukan dalam berbagai fenomena di alam, e.g. tinggi badan, berat badan, ...
- Jumlah dari berbagai peubah acak yang independen

Apa itu "terdistribusi normal"?

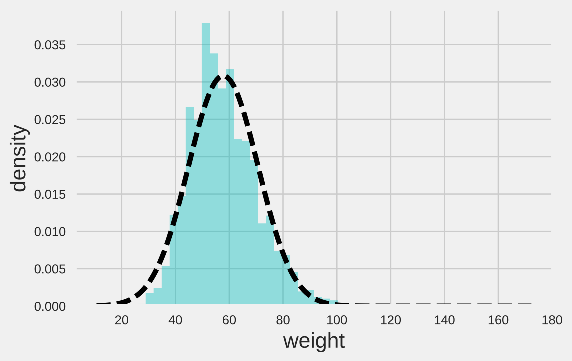
- Dapat ditemukan dalam berbagai fenomena di alam, e.g. tinggi badan, berat badan, ...
- Jumlah dari berbagai peubah acak yang independen
- Dengan jumlah sampel yang cukup, bisa menggambarkan populasi dengan baik

Contoh



Gambar: Hasil “pengukuran” tinggi badan

Contoh



Gambar: Hasil pengukuran berat badan

Mass for Shut-ins (podcast)
@gin_and_tacos

Follow

The thing I will miss most about teaching is not getting to use this pic of a weight machine to teach normal distribution

10:15 PM - 6 Jul 2019

1,042 Retweets **6,454** Likes

47 1.0K 6.5K

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

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

$Var[X] = \sigma^2$

Distribusi Normal Standar

$$Z \sim \mathcal{N}(0,1)$$

Dapat berupa transformasi dari peubah acak $X \sim \mathcal{N}(\mu, \sigma^2)$.
Petunjuk: Ingat tentang linearitas dari ekspektasi!

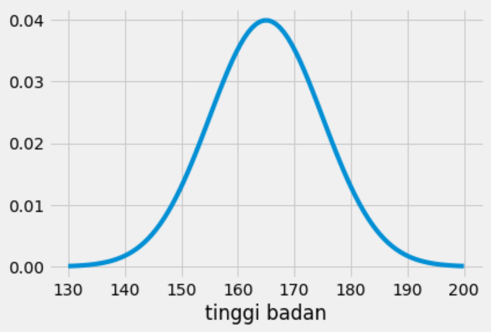
$$X = \sigma Z + \mu$$
$$Z = \frac{X - \mu}{\sigma}$$

CDF

- Karena kita berurusan dengan distribusi kontinu, kita perlu cumulative distribution function
- CDF:
$$F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right) = \int_{-\infty}^{x} f(x)dx$$
- e.g. Berapa peluangnya untuk mendapatkan orang dengan tinggi badan antara 150 dan 160?

Solusi

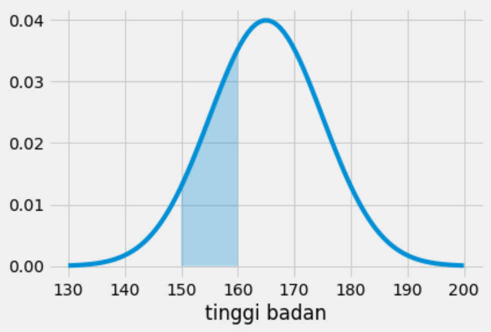
$$p(150 \leq X \leq 160) = F(160) - F(150)$$



Gambar: Distribusi tinggi badan

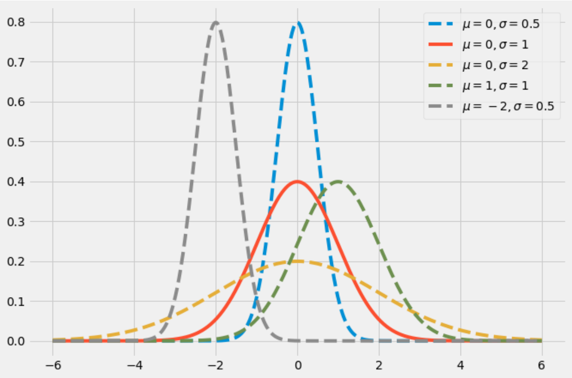
Solusi

$$p(150 \leq X \leq 160) = F(160) - F(150)$$



Gambar: Distribusi tinggi badan

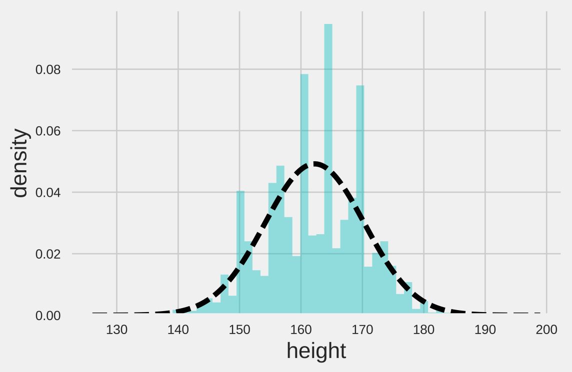
Perhatikan bahwa...



Gambar: Distribusi Gaussian dengan berbagai nilai parameter

Jika semua nilai hasil observasi sama, bagaimana grafiknya?

Maximum Likelihood Estimation



Gambar: Bagaimana cara mendapatkan distribusi Gaussian[1]?

[1] digambarkan dengan garis putus-putus

Maximum Likelihood Estimation

- Coba berbagai model $\mathcal{M}$ yang dapat memaksimalkan nilai *likelihood*, i.e. *maximum likelihood estimation*
- Dalam kasus distribusi Gaussian

$$L(\mathcal{M}) = p(X|\mu, \sigma^2) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} exp \left\{ -\frac{(x_i - \mu)^2}{2\sigma^2} \right\}$$

- Atur $\gamma = 1/\sigma^2$, lalu cari titik optimumnya.

Maximum Likelihood Estimation

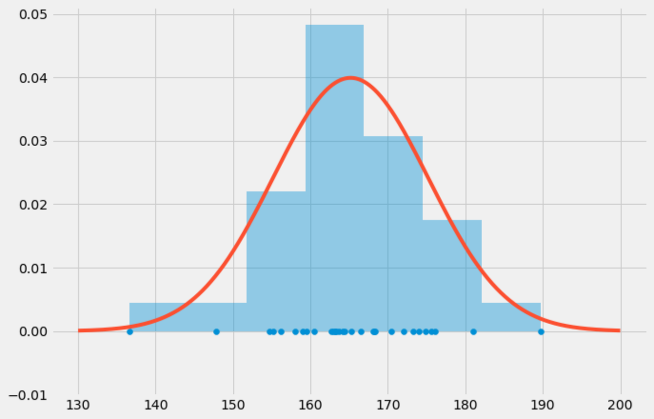
$$\log p(X|\mu,\gamma) = -\frac{1}{2}\sum_{i=1}^{N}\gamma(x_i-\mu)^2 - \frac{N}{2}\log(2\pi) + \frac{N}{2}\log\gamma$$

$$\frac{\partial p(X|\mu,\gamma)}{\partial \mu} = \gamma\sum_{i=1}^{N}(x_i-\mu)$$

$$\frac{\partial p(X|\mu,\gamma)}{\partial \gamma} = -\frac{1}{2}\sum_{i=1}^{N}(x_i-\mu)^2 + \frac{N}{2\gamma}$$

sehingga pada titik maksimum: $\hat{\mu} = \frac{1}{N}\sum_{i=1}^{N} x_i$ dan $\hat{\sigma}^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i-\mu)^2$. Terlihat familiar?

Grafik MLE



Gambar: Gaussian MLE dari 30 objek dalam data

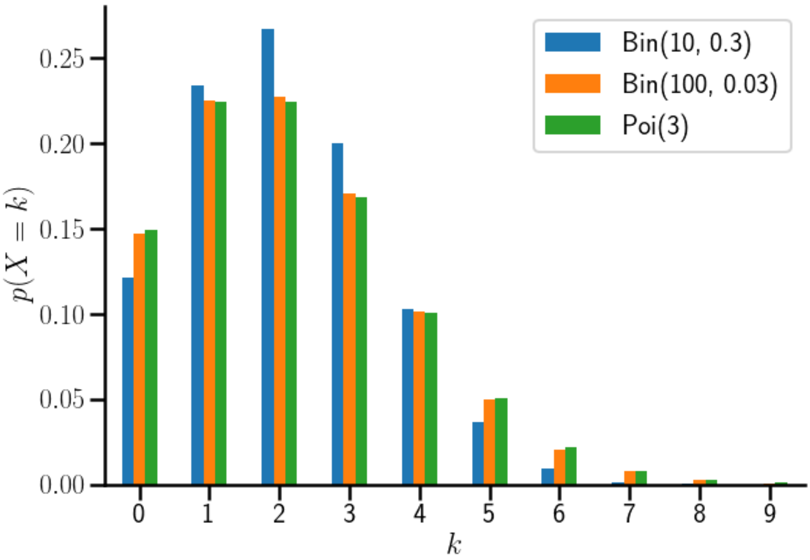
Aproksimasi

Aproksimasi Binomial dengan Poisson

$$Bin(n, \theta) \approx Poi(\lambda = n\theta)$$

jika n besar dan θ kecil

Perbandingan Distribusi

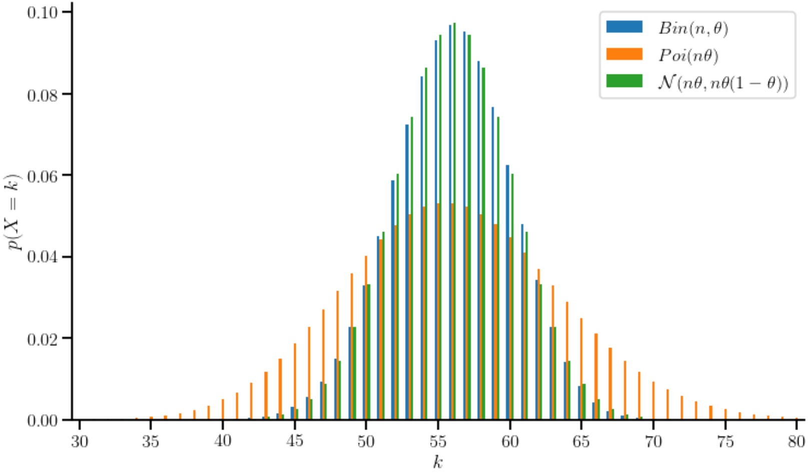


Aproksimasi Binomial dengan Gaussian

$$Bin(n, \theta) \approx \mathcal{N}(\mu, \sigma^2)$$

jika n besar dan θ sedang

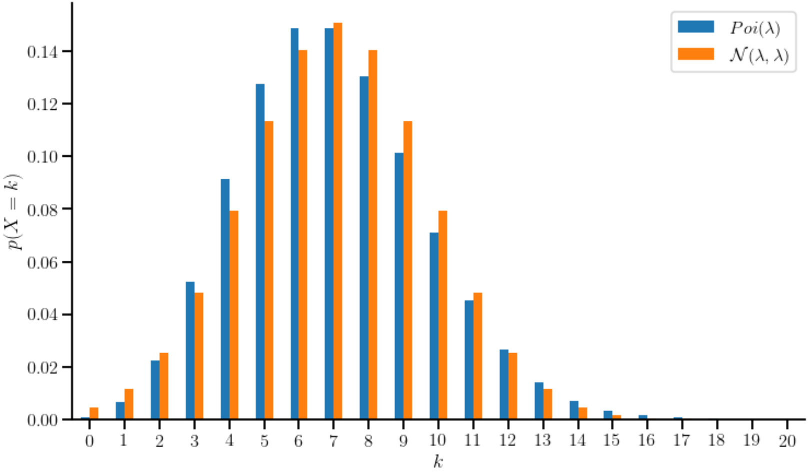
Perbandingan Distribusi



Aproksimasi Poisson dengan Gaussian

$$Poi(\lambda) \approx \mathcal{N}(\lambda, \lambda)$$

Perbandingan Distribusi



Ikhtisar

1. Dalam peubah acak kontinu, aksioma probabilitas dipenuhi saat $f(x)$ diintegralkan
2. Nilai probability density function (PDF) bisa lebih dari 1
3. Z-score merupakan hasil perubahan ke distribusi Gaussian standar
4. MLE distribusi Gaussian seperti rata-rata dan variansi dari sampelnya
5. Distribusi Gaussian dapat digunakan untuk menghampiri distribusi binomial dan Poisson

Terima kasih