Distribusi Multivariate Gaussian

Ali Akbar Septiandri

Universitas Al Azhar Indonesia

June 25, 2019

Ulasan

Distribusi Gaussian/Normal

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

Apa itu "terdistribusi normal"?

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

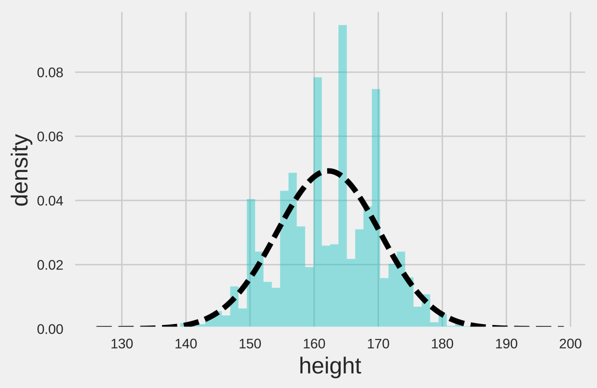
Apa itu "terdistribusi normal"?

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

Apa itu “terdistribusi normal”?

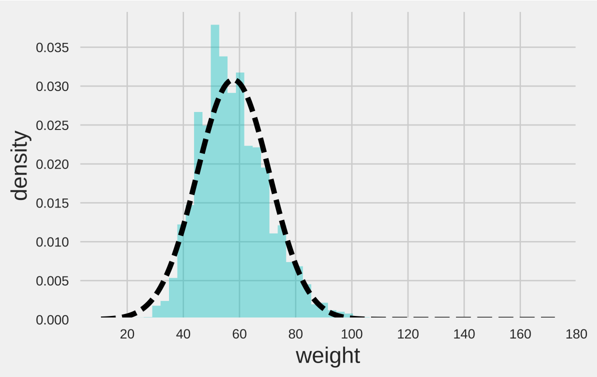
- Dapat ditemukan dalam berbagai fenomena di alam, e.g. tinggi badan, berat badan, ...
- Jumlah dari berbagai peubah acak yang independen
- Dengan jumlah sampel yang cukup, bisa menggambarkan populasi dengan baik

Contoh



Gambar: Hasil "pengukuran" tinggi badan

Contoh



Gambar: Hasil pengukuran berat badan

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

$Var[X] = \sigma^2$

CDF

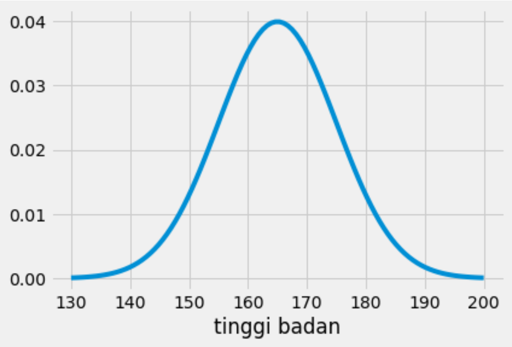
- Karena kita berurusan dengan distribusi kontinu, kita perlu cumulative density function
- CDF:

$$F(x) = \Phi\left(\frac{x - \mu}{\sigma}\right) = \int_{-\infty}^{x} f(x)dx$$

- e.g. Berapa peluangnya untuk mendapatkan orang dengan tinggi badan antara 150 dan 160?

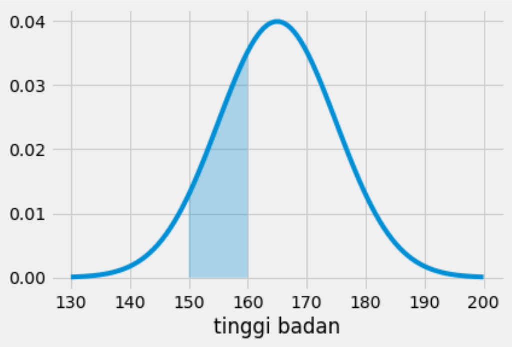
Solusi

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

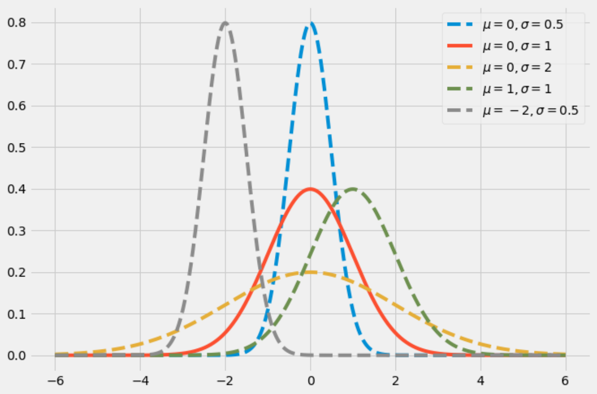


Solusi

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



Perhatikan bahwa...



Gambar: Distribusi Gaussian dengan berbagai nilai parameter

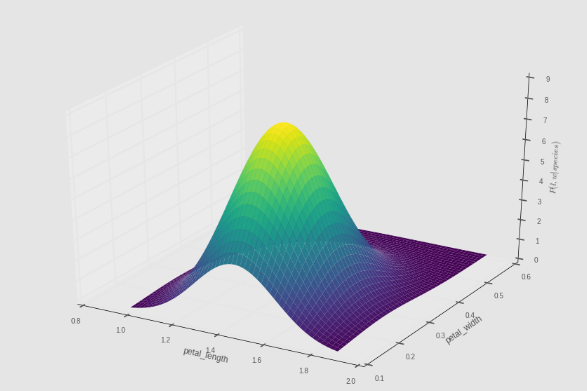
Multivariate Gaussian

Probabilitas Gabungan

	IQ = rendah	IQ = tinggi
Nilai = A	0.07	0.18
Nilai = B	0.28	0.09
Nilai = C	0.35	0.03

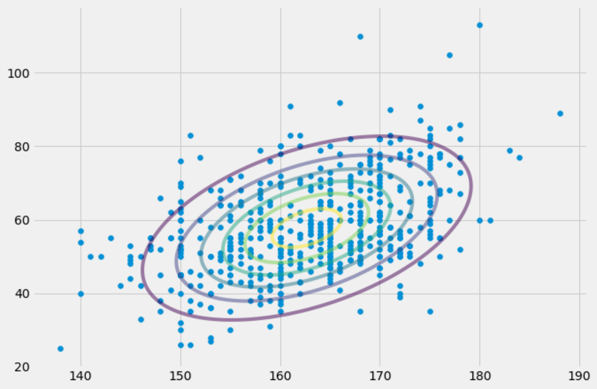
Tabel: Probabilitas gabungan dari dua peubah acak

Bivariate Gaussian



Gambar: Multivariate Gaussian dengan dua variabel dalam tiga dimensi

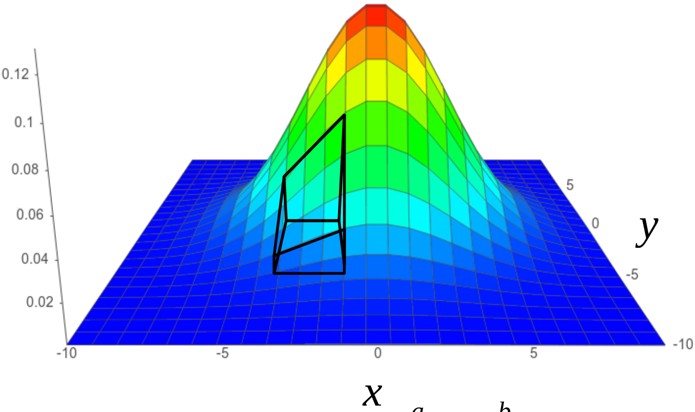
Bivariate Gaussian



Gambar: Tampak atas multivariate Gaussian dengan dua variabel

Joint probability density function

$$f_{X,Y}(x,y)$$



$$P(a_1 < X \le a_2, b_1 < Y \le b_2) = \int_{a_1}^{a_2} dx \int_{b_1}^{b_2} dy \boxed{f_{X,Y}(x,y)}$$

plot by Academo

Multivariate Gaussian

- Vektor $\mathbf{x}$ adalah multivariate Gaussian jika untuk *mean* μ dan *covariance matrix* Σ, nilainya terdistribusi menurut

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

- Univariate Gaussian adalah kasus khusus dari distribusi ini
- Σ adalah *covariance matrix*, i.e. setiap elemen $\sigma_{ij} = Cov(X_i, X_j)$ dengan

$$Cov(X_i, X_j) = \mathbb{E}[(X_i - \mu_i)(X_j - \mu_j)]$$

- Σ harus simetris

Kovariansi dan Korelasi

Kovariansi

- Berdasarkan linearitas ekspektasi

$$\begin{aligned} Cov(X_i, X_j) &= \mathbb{E}[(X_i - \mu_i)(X_j - \mu_j)] \\ &= \mathbb{E}[X_i X_j] - \mathbb{E}[X_i]\mathbb{E}[X_j] \end{aligned}$$

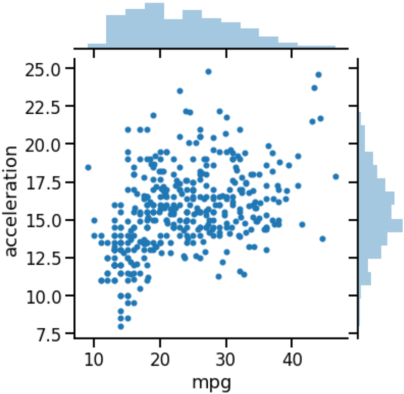
  Buktikan!
- Interpretasi: Seberapa besar perubahan bersama X_i dan X_j?

Perhitungan Kovariansi

mpg	acceleration	mpg $\times$ acceleration
23.0	18.5	425.5
16.0	18.0	288.0
14.0	15.5	217.0
31.0	19.0	589.0
21.0	17.0	357.0
18.0	16.0	288.0
18.0	16.5	297.0
32.0	21.0	672.0
19.0	17.7	336.3
14.0	14.5	203.0

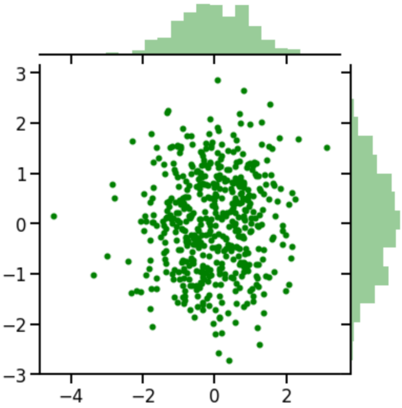
$$Cov(X_i, X_j) = \mathbb{E}[X_i X_j] - \mathbb{E}[X_i]\mathbb{E}[X_j]$$

Bivariate Gaussian - Kovariansi



Gambar: $Cov(mpg, acceleration) = 9.06$ — ada hubungan positif!

Bivariate Gaussian - Kovariansi



Gambar: $Cov(mpg, acceleration) = 0.06$ — hampir tidak berhubungan

Properti Kovariansi

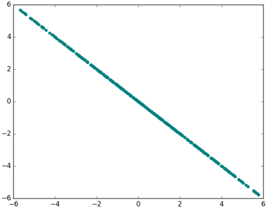
- $Cov(X, Y) = Cov(Y, X)$
- $Cov(X, X) = \mathbb{E}[XX] - \mathbb{E}[X]\mathbb{E}[X] = Var(X)$
- $Cov(aX + b, Y) = aCov(X, Y)$

Korelasi

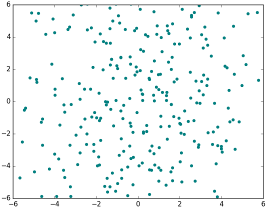
Korelasi dari dua variabel adalah ukuran dependensi linear antara keduanya yang dibuat dalam skala nilai -1 sampai 1.

$$\rho(X, Y) = \frac{Cov(X, Y)}{\sqrt{Var(X)Var(Y)}}$$

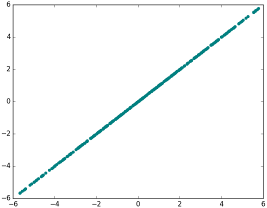
Important correlations



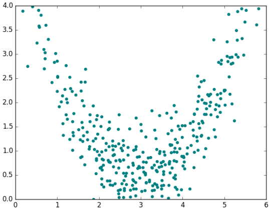
$$\rho(X, Y) = -1$$



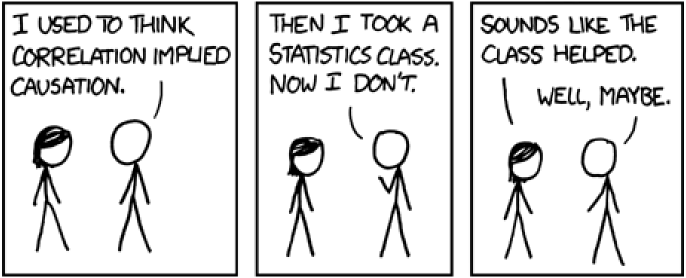
$$\rho(X, Y) = 0$$



$$\rho(X, Y) = 1$$



$$\rho(X, Y) = 0$$



Gambar: `https://xkcd.com/552/`

Pekan depan:

CLT, LLN, MLE, Bayesian stats

Terima kasih