

Beberapa Distribusi Kontinu

Ali Akbar Septiandri

Universitas Al-Azhar Indonesia

aliakbars@live.com

April 1, 2019

Selayang Pandang

- ① Ulasan
- ② Distribusi Uniform
- ③ Distribusi Beta
- ④ Distribusi Gaussian
 - Univariate Gaussian
 - Maximum Likelihood Estimation
 - Multivariate Gaussian

Ulasan

Ekspektasi Kontinu dan Variansi

Ekspektasi

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x \cdot p(x) dx$$

Ekspektasi Kontinu dan Variansi

Ekspektasi

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x \cdot p(x) dx$$

Variansi

$$\text{Var}[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$$

Tunjukkan!

Sudah enroll ke e-learning?

Distribusi Uniform

Contoh



The image shows the Spotify interface for Taylor Swift's album 'reputation'. The album cover features a black and white photo of Taylor Swift with a newspaper clipping overlay. The title 'reputation' is in a large, bold, white font. Below the title, it says 'Par Taylor Swift' and '2017 • 15 titres, 55 min'. There are buttons for 'LIRE' (green) and 'SAUVEGARDER' (white outline). Below the album information is a tracklist table.

#	TITLE	Duration
1	...Ready For It?	3:28
2	End Game - Ed Sheeran, Future	4:05
3	I Did Something Bad	3:58
4	Don't Blame Me	3:56
5	Delicate	3:52
6	Look What You Made Me Do	3:32
7	So It Goes...	3:48
8	Gorgeous	3:30
9	Getaway Car	3:54

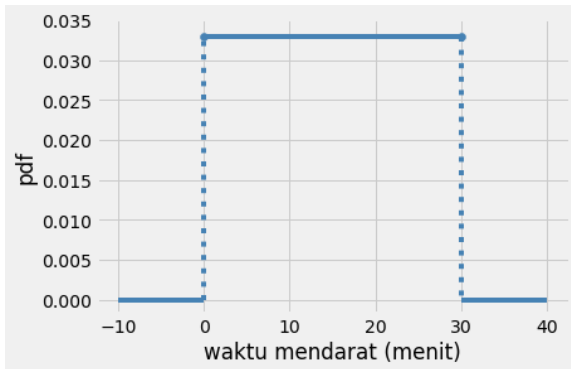
Gambar: Durasi lagu-lagu Taylor Swift dalam album Reputation cenderung terdistribusi seragam

Peubah Acak Seragam

- Semua bernilai sama peluangnya dalam interval tertentu
- Dituliskan sebagai $X \sim Uni(\alpha, \beta)$
- Bisa berupa diskrit maupun kontinu
- PDF:

$$p(x) = \begin{cases} \frac{1}{\beta - \alpha}, & \text{if } x \in [\alpha, \beta] \\ 0, & \text{otherwise} \end{cases}$$

Contoh

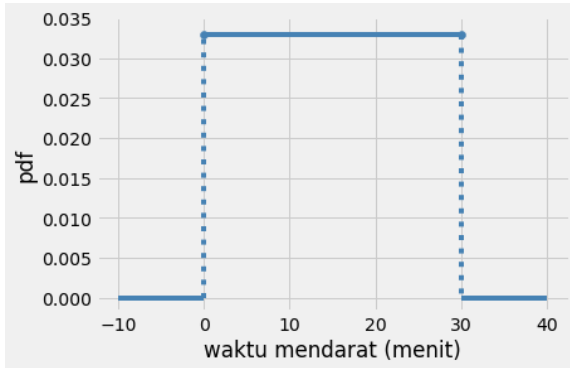


Gambar: Berapa probabilitas pesawat mendarat ada di antara 25-30 menit?

Mengapa PDF di satu titik bisa > 1 ?

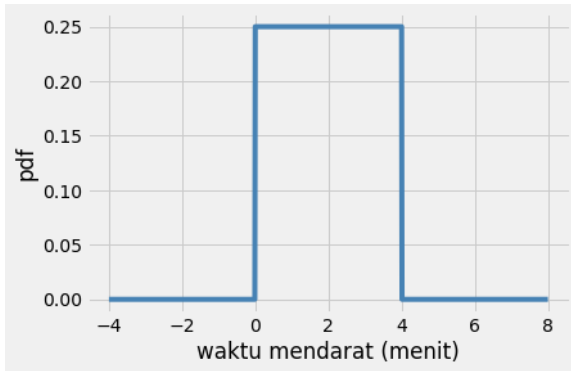
Perhatikan!

Ingat bahwa $\int_{-\infty}^{\infty} p(x)dx = 1$ sehingga luas area di bawah (persegi panjang) harus bernilai 1.

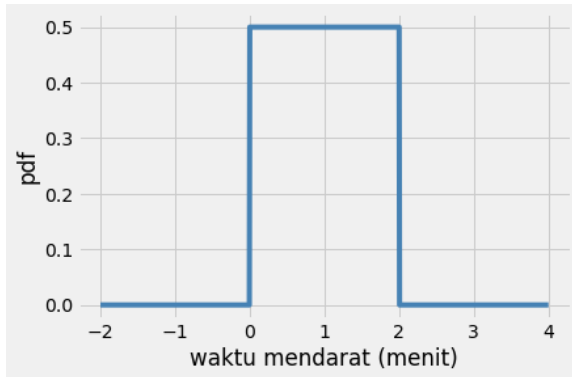


Apa yang terjadi kalau intervalnya kita perkecil?

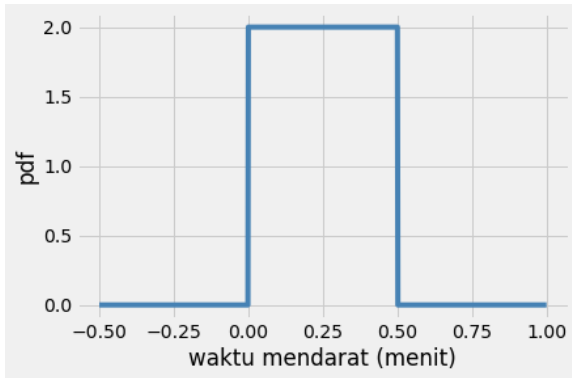
Perhatikan!



Perhatikan!



Perhatikan!



Ekspektasi dan Variansi

Ekspektasi

$$\mathbb{E}[X] = \frac{\alpha + \beta}{2}$$

Variansi

$$\text{Var}[X] = \frac{(\beta - \alpha)^2}{12}$$

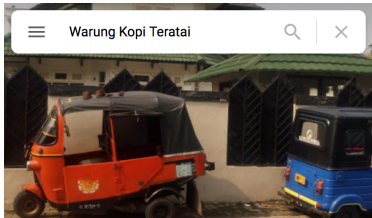
Distribusi Beta



Dua Coffee Shop @cipete

4.5 ★★★★★ (1,020) · \$\$

Coffee shop



Warung Kopi Teratai

4.6 ★★★★★ (18) · \$

Cafe

Gambar: Mana yang lebih Anda percaya?

Kita ingin memasukkan unsur
ketidakpastian

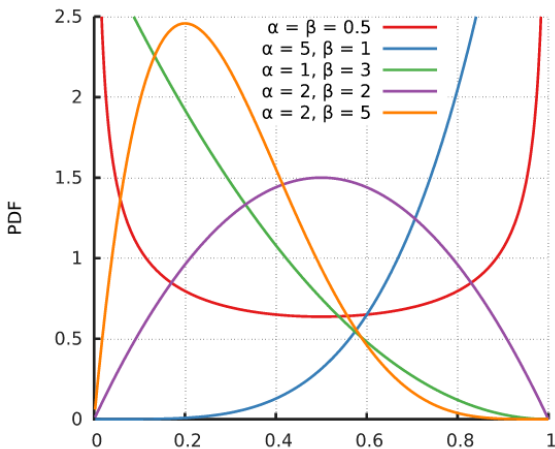
Distribusi Beta

- Merepresentasikan keluaran dari persentase atau proporsi
- Berguna untuk dipakai sebagai *prior probability*
- Jika X adalah RV yang bernilai $x \in [0, 1]$,
- dan

$$p(X = x|a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{(a-1)}(1-x)^{(b-1)}$$

- maka $X \sim \text{Beta}(a, b)$.

Grafik PDF



Gambar: PDF dari distribusi Binomial

Ekspektasi dan Variansi

Ekspektasi

$$\mathbb{E}[X] = \frac{a}{a+b}$$

Variansi

$$\text{Var}[X] = \frac{ab}{(a+b)^2(a+b+1)}$$

Example

Kafe Romeo diberikan nilai 5, 4, 2, 1, 4, 5, 5, 5, 4, 4, 5, 5, 5, 5, 5.
Kafe Juliet diberikan nilai 5, 5, 4. Mana yang lebih baik?

Solution

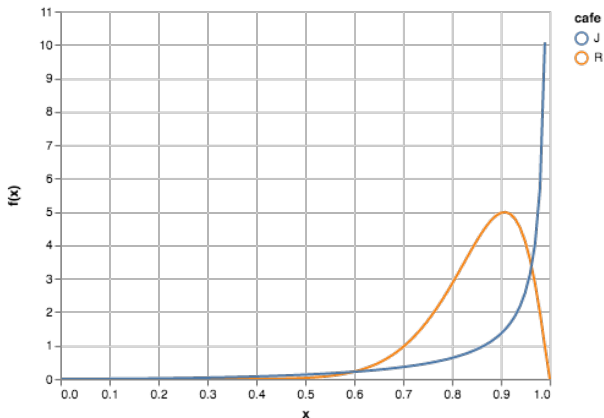
Jika $X \sim \text{Beta}(a, b)$, maka $X \in [0, 1]$. Jadi, ubah skala nilainya terlebih dahulu, e.g. $2 \rightarrow 0.4$.

$$a = 1 + S$$

$$b = 1 + N - S$$

dengan N adalah jumlah orang yang memberi nilai, dan S adalah jumlah nilai yang telah diubah skalanya.

Solusi



$$X_R \sim \text{Beta}(a_R, b_R) = \text{Beta}(12.8, 2.2)$$

$$X_J \sim \text{Beta}(a_J, b_J) = \text{Beta}(2.8, 0.2)$$

Solusi (lanjutan)

Dengan asumsi **central limit theorem**, kita dapat menghampiri distribusi Beta dengan Gaussian. Lalu, anggaplah kita mau melihat **kemungkinan terburuk**. Dengan demikian, nilai di kuantil .5 dapat dihitung dengan (95% *least plausible value*):

$$score = \mu - 1.65\sigma$$

atau

$$score = \frac{a}{a+b} - 1.65 \sqrt{\frac{ab}{(a+b)^2(a+b+1)}}$$

sehingga

$$score_R = 3.54$$

$$score_J = 3.64$$

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, ...

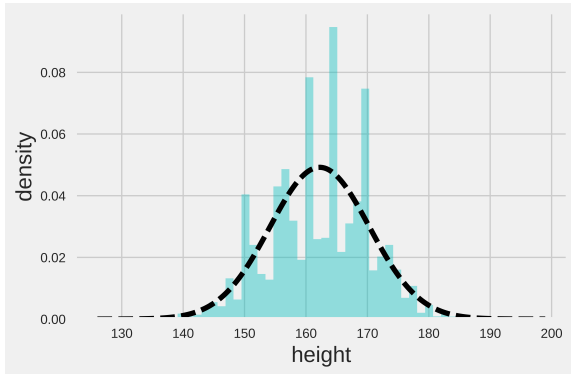
Apa itu “terdistribusi normal”?

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

Apa itu “terdistribusi normal”?

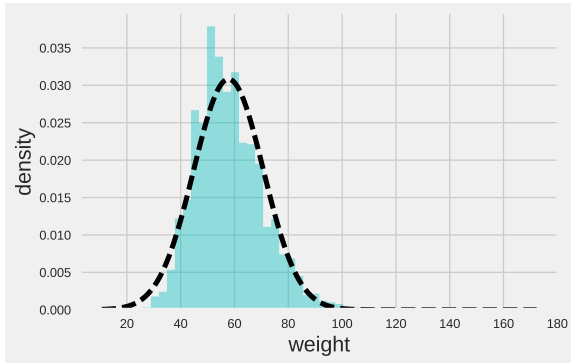
- Dapat ditemukan dalam berbagai fenomena di alam, e.g. tinggi badan, berat badan, ...
- Jumlah dari berbagai peubah acak yang independen
- Dengan jumlah sampel yang cukup, bisa menggambarkan populasi dengan baik

Contoh



Gambar: Hasil “pengukuran” tinggi badan

Contoh



Gambar: Hasil pengukuran berat badan

Fakta

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

PDF

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

Ekspektasi

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

Variansi

$$\text{Var}[X] = \sigma^2$$

CDF

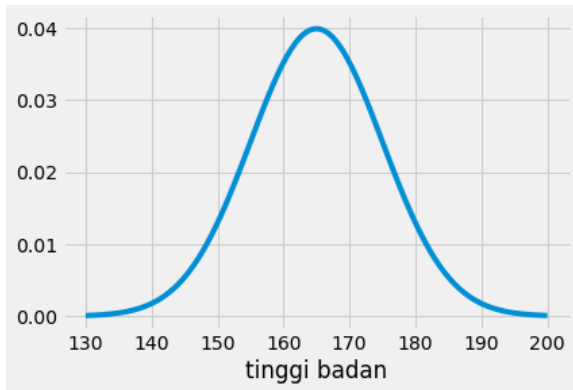
- Karena kita berurusan dengan distribusi kontinu, kita perlu **cumulative density function**
- CDF:

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

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

Solusi

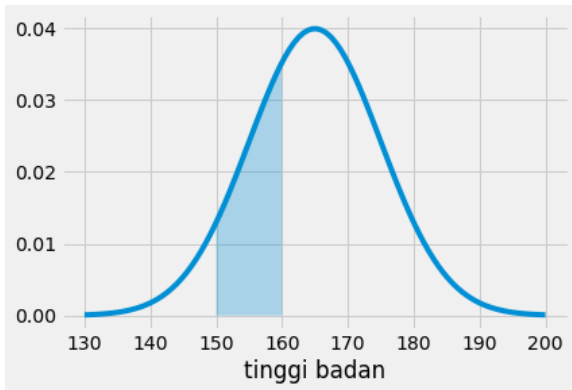
$$p(150 < X < 160) = F(160) - F(150)$$



Gambar: Distribusi tinggi badan

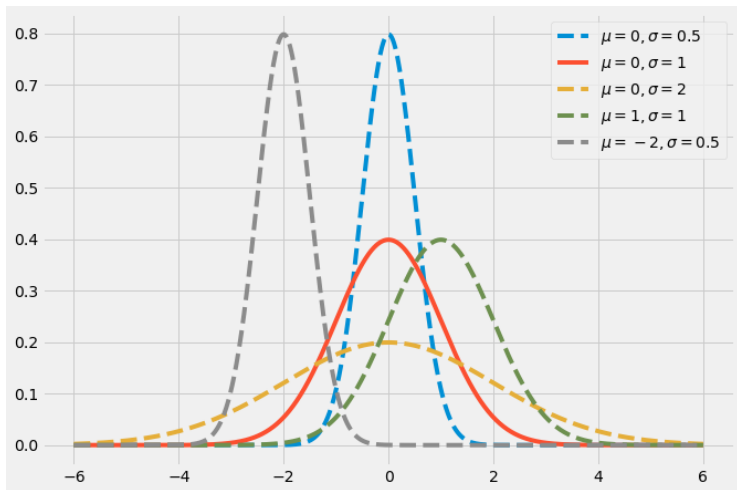
Solusi

$$p(150 < X < 160) = F(160) - F(150)$$



Gambar: Distribusi tinggi badan

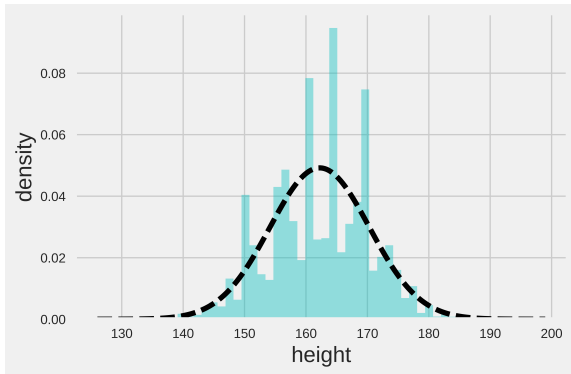
Perhatikan bahwa...



Gambar: Distribusi Gaussian dengan berbagai nilai parameter

Jika semua nilai hasil observasi sama, bagaimana grafiknya?

Contoh



Gambar: Bagaimana cara mendapatkan distribusi Gaussian¹?

¹digambarkan 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

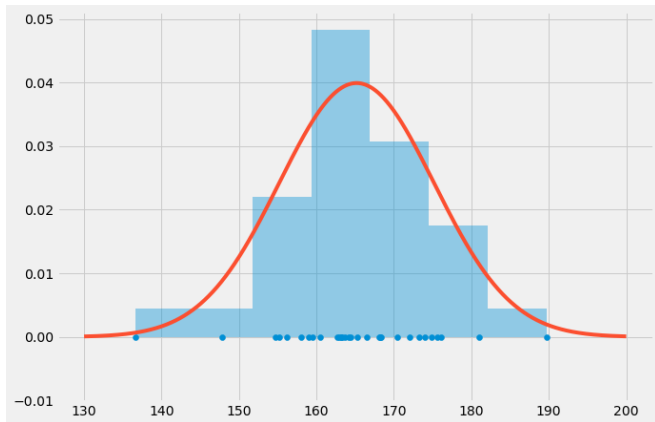
$$\log p(X|\mu, \gamma) = -\frac{1}{2} \sum_{i=1}^N \gamma (x_i - \mu)^2 - \frac{N}{2} \log(2\pi) + \frac{N}{2} \log \gamma$$

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

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

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

Multivariate Gaussian

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

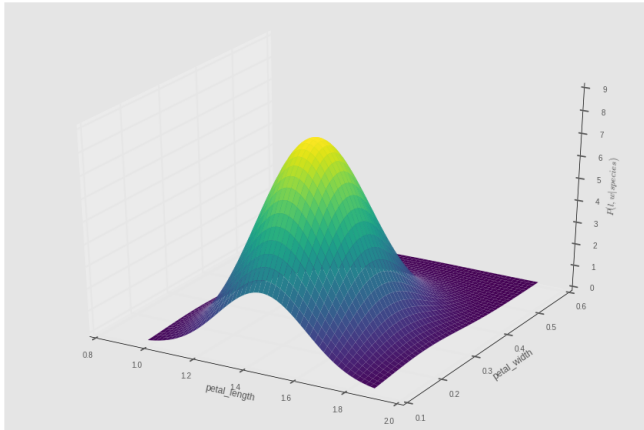
$$p(\mathbf{x}|\mu, \Sigma) = \frac{1}{|(2\pi)\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x} - \mu)^T \Sigma^{-1}(\mathbf{x} - \mu)\right)$$

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

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

- Σ harus simetris

Bivariate Gaussian

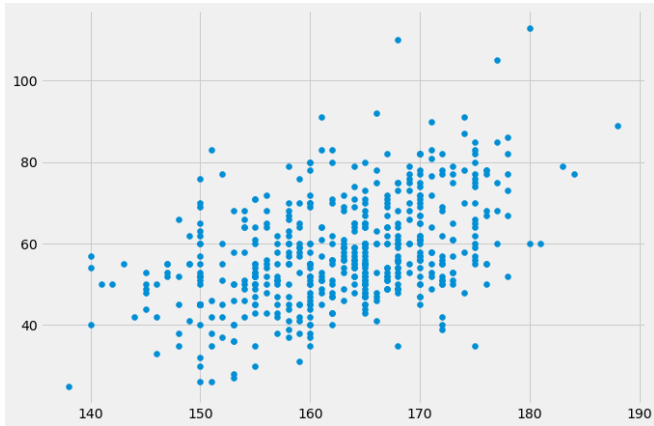


Gambar: Multivariate Gaussian dengan dua variabel yang dibuat dalam tiga dimensi

Maximum Likelihood

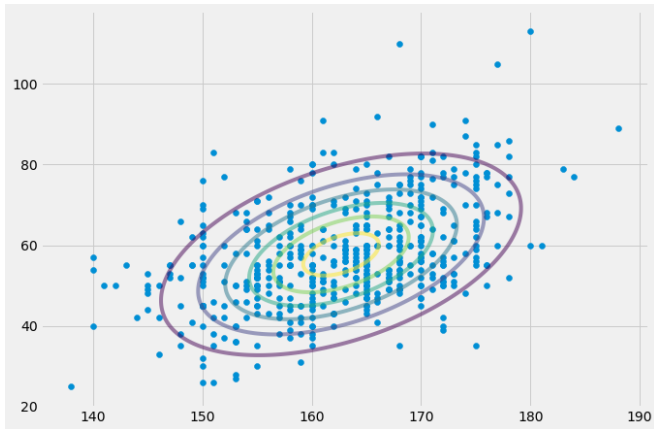
- Sama dengan kasus univariate
- $\mu = \frac{1}{N} \sum_{i=1}^N \mathbf{x}_i$
- $\Sigma = \frac{1}{N} \sum_{i=1}^N (\mathbf{x}_i - \mu)(\mathbf{x}_i - \mu)^T$

Grafik MLE



Gambar: Data tinggi vs berat badan

Grafik MLE



Gambar: MLE dari data

Poin Penting

- Jumlah dari Gaussian RVs adalah Gaussian
- Model yang terdiri dari kombinasi linear Gaussian akan memiliki *joint distribution* berupa Gaussian
- Jika $p(x, y)$ adalah multivariate Gaussian, maka $p(x)$ maupun $p(y)$ serta $p(x|y)$ dan $p(y|x)$ adalah Gaussian

Ikhtisar

- Distribusi uniform
- Distribusi Beta
- Distribusi normal/Gaussian

Pertemuan Berikutnya

- Bayes classifier
- Naïve Bayes
- Conditional independence

Office hours minggu depan ditiadakan

Referensi



Will Monroe (Jul. 2017)

The Normal Distribution

<http://web.stanford.edu/class/archive/cs/cs109/cs109.1178/lectureHandouts/110-normal-distribution.pdf>



Chris Williams (Sep. 2015)

Probability - Machine Learning and Pattern Recognition

<https://www.inf.ed.ac.uk/teaching/courses/mlpr/2015/>



Chris Williams (Sep. 2015)

The Gaussian Distribution - Machine Learning and Pattern Recognition

<https://www.inf.ed.ac.uk/teaching/courses/mlpr/2015/>

Terima kasih