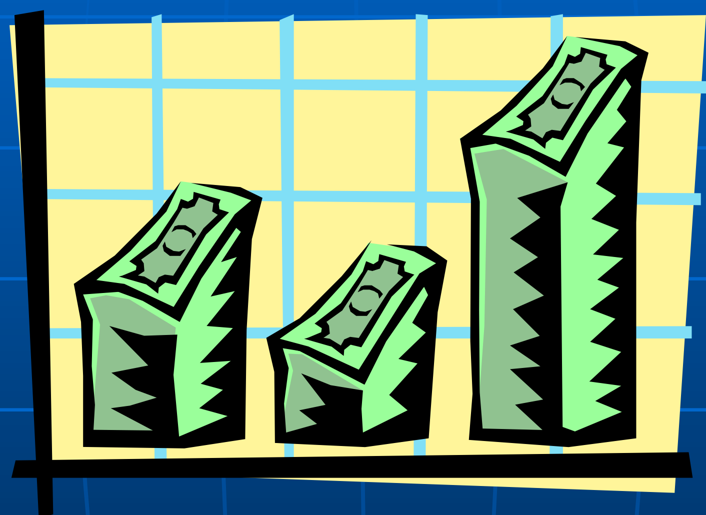
Speech Signal Processing

Dasar Statistik untuk Pemodelan dan Simulasi



oleh:
Tri Budi Santoso

Signal Processing Group
Electronic Engineering Polytechnic Institute of Surabaya-ITS

1. Probabilitas

Probabilitas=Peluang, bisa diartikan juga sebagai besarnya kemungkinan munculnya suatu kejadian.

Soal 1:

Saya memiliki sebuah koin yang memiliki satu muka berupa gambar (G) dan yang satunya angka (A). Saat saya lempar ke atas, maka saat koin jatuh kemungkinan untuk mendapatkan muka G adalah p(x) = ½.

Dalam hal ini x={G,A}

Kemudian saya lemparkan lagi, maka probabilitas mendapatkan A berapa?

Soal 2:

Untuk undian TOGEL minggu ini saya membeli angka 00, 15, 15, 20. Seperti kita ketahui bahwa dalam undian TOGEL kemungkinan angka keluar adalah dari 00 sampai 99 danmasing-masing memiliki peluang yang sama (*equ-probable)* untuk keluar. Maka berapa probabilitas saya mendapat undian?

2. Probability density function

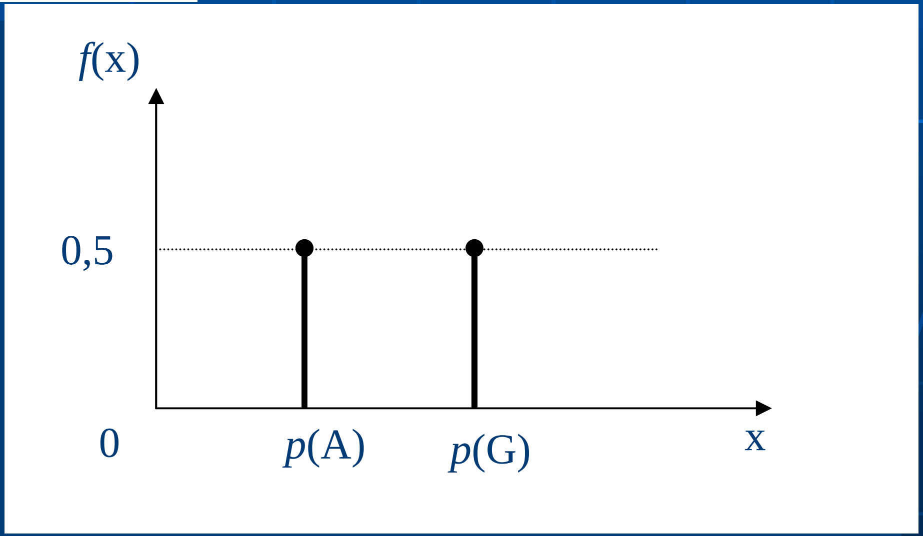
Probability density function (pdf) → simbolnya f(x):
Menyatakan fungsi kerapatan probabilitas, yaitu keadaan yang menjamin bahwa suatu probabilitas selalu bernilai positif.

Untuk kasus sistem kontinyu: $\int_A f(x)dx = 1$

Untuk kasus sistem diskrit: $\sum_A f(x) = 1$

Gambaran Grafik
untuk kasus Contoh 1 adalah sbb:

Bagaimana cara mendapatkan
gambar disamping?



Dengan memahami bahwa kasus pelemparan koin merupakan contoh sistem diskrit, maka kita dapat mengacu pada persamaan kedua pada fungsi kerapatan probabilitas.

Dari masing-masing muka (Angka dan Gambar) memiliki probabilitas untuk keluar sebesar p(x) = f(x) =1/2.
Masing-masing disajikan dalam bentuk grafik, maka kita dapatkan seperti pada halaman sebelumnya.

Dengan mengacu pada persamaan untuk menghitung fungsi kerapatan probabilitas, akan didapatkan:

$$\sum_{A} f(x) = f(A) + f(G) = 1/2 + 1/2 = 1$$

3. Fungsi Distribusi Kumulatif

Untuk kasus kontinyu: $F(x) = P(X \leq x) = \int_{-\infty}^{x} f(x)$

Untuk kasus diskrit: $F(x) = P(X \leq x) = \sum_{-\infty}^{x} f(x)$

Contoh 1:
Pada kasus pelemparan 2 koin secara bersamaan maka kemungkinan yang muncul adalah: 0G2A, 1G 1A, 2G0A. Dimana masing-masing memiliki probabilitas sebesar: p(0G) = 1/4, p(1G) = 2/4, p(2G) =1/4. Coba anda berikan grafik untuk menyatakan fungsi kerapatan probabilitas dan fungsi distribusi kumulatif kasus ini.

Penyelesaian:
Penggambaran fungsi kerapatan probabilitas untuk munculnya muka gambar dinyatakan sebagai f(x) yang dalam hal ini x = G. Untuk kasus ini dapat dituliskan p(0G) = f(x) dimana x =0 memiliki nilai ¼. Demikian pula untuk f(x)=p(1G) = ½, dan f(x)=p(2G)= ¼.

Fungsi distribusi kumulatifnya….

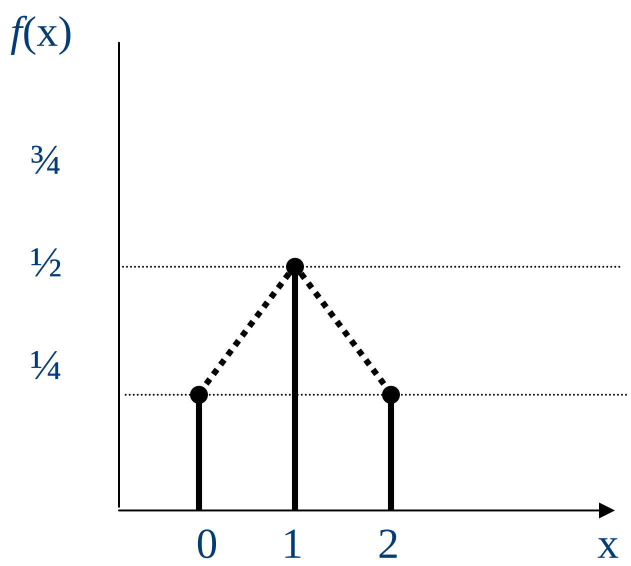
Fungsi distribusi kumulatif dicari satu persatu.

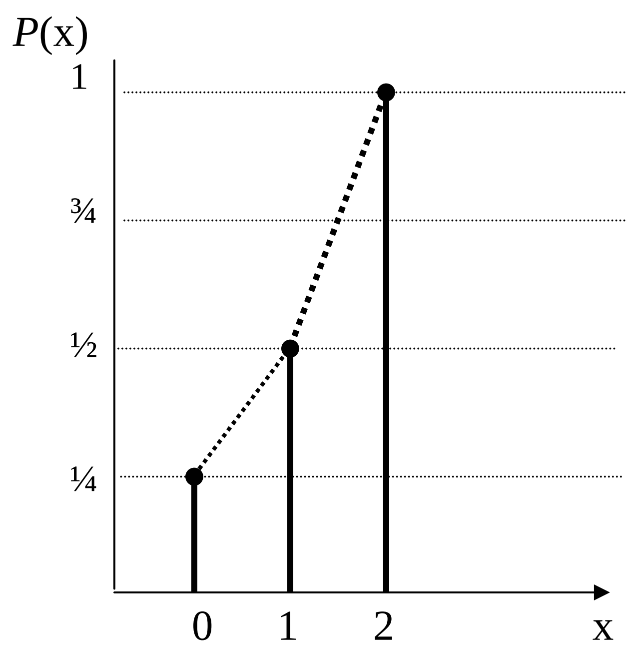
Untuk F(0), berarti dihitung berapa jumlahan semua probabilitas untuk $x \leq 0$.
Dalam hal ini adalah f(x) = f(0G) = p(0G) = ¼

Untuk F(1), berarti dihitung berapa jumlahan semua probabilitas untuk $x \leq 1$.
Dalam hal ini adalah f(0) + f(1) = f(0G) + f(1G) = ¼ + ½ = ¾

Untuk F(2), berarti dihitung berapa jumlahan semua probabilitas untuk $x \leq 2$.
Dalam hal ini adalah f(0) + f(1) = f(0G) + f(1G) + f(2G) = ¼ + ½ + ¼ = 1

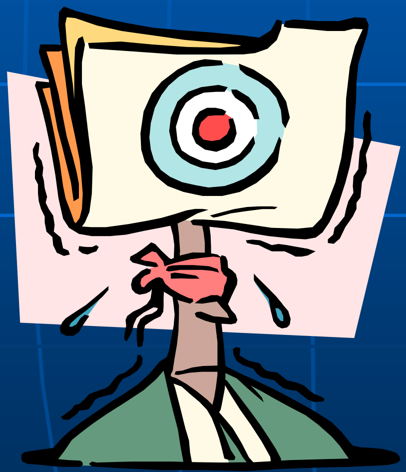
Grafik untuk f(x) dan F(x) adalah seperti berikut…





Soal 3:

Pada kasus pelemparan 3 koin secara bersamaan maka kemungkinan yang muncul adalah: 3G 0A, 2G 1A, 1G 2A, dan 0G3A. Dimana masing-masing memiliki probabilitas sebesar: p(3G) = 1/8, p(2G) = 3/8, p(1G) =3/8, p(0G) = 1/8. Coba anda berikan grafik untuk menyatakan fungsi kerapatan probabilitas dan fungsi distribusi kumulatif kasus ini.



4. Ekspektasi Matematika

Yang seringkali dibicarakan adalah:

Ekspektasi Pertama:

- Menyatakan pemusatan data, yang menyatakan satu nilai yang mewakili keseluruhan data
- Ada tiga nilai yang dapat mewakilinya, yaitu mean, median dan modus

$$E(x) = \sum_{A} x f(x)$$

Ekspektasi Kedua:

- Menyatakan penyebaran data, yang menyatakan rentang data yang kita miliki
- Ada dua nilai yang dapat mewakilinya, yaitu varians dan standar deviasi

$$E(x^2) = \sum_{A} x^2 f(x)$$

Ekspektasi Pertama

Mean (rata-rata)→ E(x)

Untuk sejumlah N nilai data diberikan sebagai $E(x) = \frac{1}{N}\sum_{i=1}^{i=N} x_i$

Medan (nilai tengah)→ P(X≤x) = ½

Untuk sejumlah N nilai data diberikan sebagai nilai tengah yang muncul dari seluruh data yang ada.

Modus

Untuk sejumlah N nilai data diberikan sebagai nilai mana yang paling sering muncul.

Contoh 2:

Dari 15 keluarga yang tinggal di RT1/RW2 Gebang Lor, jumlah anak yang mereka miliki berbeda-beda seperti berikut ini: 5,6,7,4,5,4,1,2,10,3,5,1,5,6,11.
Dari data tersebut, coba anda cari nilai mean, median dan modusnya.

Penyelesaian:

Dari data tersebut dapat kita buat sebuah tabel seperti berikut

Nilai	Muncul
1	2
2	1
3	1
4	2
5	4
6	2
7	1
10	1
11	1

- **Mean:**

x_rat = (1/15)(5 + 6 + 7 +….+ 6 + 11)

= 75/15 = 5

- **Median:**

(x_max - x_min)/2 =(11-1)/2 = 5

- **Modus:**

Dalam hal ini yang sering muncul adalah angka 5 sebanyak 4 kali.

Soal 4:
Data statistik dari 10 desa yang memiliki jumlah lulusan S1 adalah sebagai berikut:
Desa-1: 10 orang
Desa-2: 5 orang
Desa-1: 4 orang
Desa-2: 19 orang
Desa-1: 2 orang
Desa-2: 10 orang
Desa-1: 2 orang
Desa-2: 5 orang
Desa-1: 5 orang
Desa-2: 2 orang

Dari data tersebut, coba anda cari nilai mean, median dan modus tentang jumlah lulusan S1 yang dimiliki oleh desa-desa tersebut.

Ekspektasi Kedua

Varians:
memiliki persamaan dasar sebagai berikut:

$$E\left((x-\mu)^2\right) = \sum_A (x-\mu)^2 . f(x)$$

Untuk pemakaian pada sekumpulan data sebanyak N, maka formulasinya menjadi seperti berikut

$$\sigma^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - x\_rat)^2$$

Sedangkan untuk nilai deviasi dinyatakan sebagai $\sigma = \sqrt{\sigma^2}$

Contoh 3:
Dari contoh 2 yang telah diberikan, coba anda cari nilai varians dan deviasinya

Penyelesaian contoh 6:

Dengan menggunakan persamaan dasar diatas, maka didapatkan nilai variansnya sbb.

$$\sigma^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - x\_rat)^2 \quad = \left(\frac{1}{15-1}\right)\left\{\begin{array}{l}(5-5)^2 + (7-5)^2 + (4-5)^2 + \ldots\ldots\ldots \\ \ldots\ldots\ldots\ldots\ldots + (5-5)^2 + (6-5)^2 + (11-5)^2\end{array}\right\}$$

$$= \left(\frac{1}{14}\right)(114) = 8{,}14$$

Sedangkan nilai deviasi diberikan sebagai:

$$\sigma = \sqrt{\sigma^2} = \sqrt{8{,}14} = 2{,}853$$

Soal 5:
Dari nilai probabilitas yang anda dapatkan pada masing keadaan pada contoh 3 coba anda cari nilai mean, varians dan deviasi

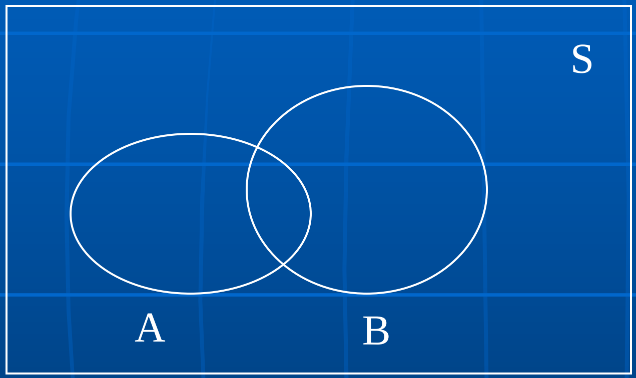
Soal 6:
Dalam suatu undian sepuluh angka berikut ini: 01, 02, 08, 11, 12, 24, 23, 31, 33, 45, 50, 98, 00 memiliki peluang yang tidak sama: 0,1% ; 0,02% ; 0,1% ; 0,02% ; 0,1% ; 0,02% ; 0,1% ; 0,02% ; 0,07% ; 0,02% ; 0,01% ; 0,01% ; 0,07% .
Dari data tersebut, cari nilai mean, modus, varians dan deviasi terhadap nilai probabilitas (peluang) yang ada.

Soal 7:
Dari beberapa soal yang telah anda selesaikan, pilih salah satu dan buat program untuk mendapatkan nilai mean, varians dan deviasi. Dalam hal ini anda bisa menggunakan Matlab atau bahasa C untuk menyelesaikannya.
Selamat belajar…..

5. Probabilitas Bersyarat

Probabilitas keluarnya A, jika diketahui informasi B. Bisa juga didefinisikan propabilitas A dengan syarat B

Perumusannya: $$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

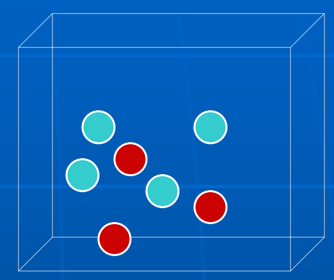


Contoh 4:
Probabilitas saya akan kenyang apabila saya makan satu piring.
Dalam hal ini kita bisa formulasikan bahwa saya makan adalah B, sedangkan saya amakan satu piring dan kenyang adalah A. Apabila dalam hal ini probabilitas B adalah 0.9 sedangkan probabilitas A adalah 0.5. Maka probabilitas saya kenyang apabila saya makan dapat dicari dengan cara:
P(A|B) = 0.5/0.9 = 0.555.

Apabila dari kasus diatas saya tidak makan, maka berapa P(A|B)?

Contoh 5:

Pada suatu kotak terdapat 4 kelereng kuning dan 3 kelereng merah. Akan dilakukan pengambilan secara acak beberapa kali, dimana setelah suatu pengambilan dilakukan kelerengnya tidak dikembalikan.

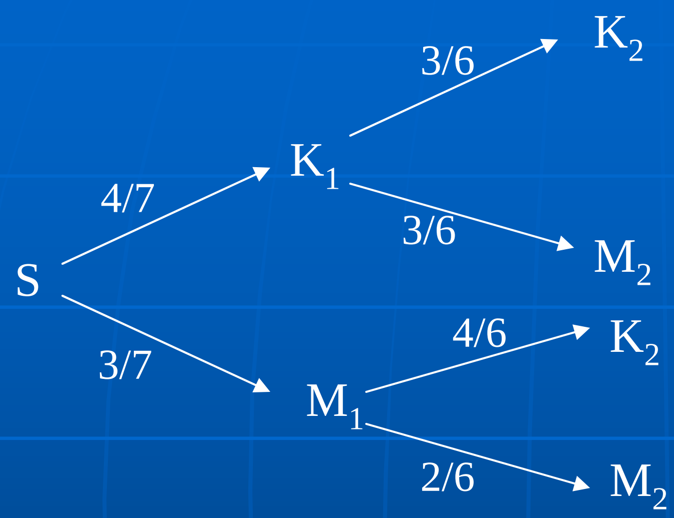


Pada pengambilan pertama:
p(kuning) = 4/7
p(merah) = 3/7

Bila pengambilan pertama didapat ***kelereng kuning***, maka untuk pengambilan kedua:
p(kuning)=3/6
p(merah)=3/6

Bila pengambilan pertama didapat ***kelereng merah***, maka untuk pengambilan kedua:
p(kuning)=4/6
p(merah)=2/6

Kondisi ini bisa digambarkan sbb….



Dari gambaran tersebut, probabilitas untuk mendapatkan K2 adalah

$$
\begin{aligned}
p(K_2) &= p(K_2 \cap K_1) + p(K_2 \cap M_1) \\
&= p(K_2 \mid K_1)\, p(K_1) + p(K_2 \mid M_1)\, p(M_1) \\
&= \left(\frac{3}{6}\right)\left(\frac{4}{7}\right) + \left(\frac{4}{6}\right)\left(\frac{3}{7}\right) = \frac{24}{42}
\end{aligned}
$$

Berapa probabilitas untuk mendapatkan M2 ?

6. Kovarians dan Korelasi

Kovarians:

$$\begin{aligned}\operatorname{cov}(x,y) &= E\left\lfloor (x-m_x)(y-m_y)\right\rfloor \\ &= \frac{1}{N}\sum_{i=1}^{i=N}(x_i-m_x)(y_i-m_y)\end{aligned}$$

Korelasi:

$$corr(x,y)=\frac{\operatorname{cov}(x,y)}{dev(x)dev(y)}$$

dimana dev(x) = s_x = akar var(x)

Contoh 6:
Diketahui dua variabel x dan y yang memiliki nilai-nilai acak sebagai berikut:
x (i) = 0, 4, 8, 3, 7, 3, 7, 9, 4, 2 dan y(i) = 1, 0, 4, 7, 5, 4, 6, 9, 7, 0

Coba anda cari nilai kovarian (x,y) dan korelasi (x,y) dari kedua data tersebut.

Penyelesaian:…….

Penyelesaian Contoh 6:

Langkah pertama adalah mencari nilai mean dari x dan y. Dengan penghitungan biasa, anda dapatkan nilai mean untuk x adalah $m_x = 4{,}7$ dan untuk y didapatkan mean $m_y = 4{,}3$.

Langkah kedua adalah menghitung nilai deviasi. Dengan menggunakan penghitungan varians dan dilanjutkan mengambil nilai akarnya, didapat:

deviasi x = $s_x = 2{,}9078$

deviasi y = $s_y = 3{,}1287$

Selanjutnya adalah mencari nilai kovarian (x,y). Dengan cara memasukkan nilai-nilai tersebut ke persamaan kovarian didapatkan hasil

$$\begin{aligned} \mathrm{cov}(x, y) &= \frac{1}{10}\sum_{i=1}^{10}(x_i - 4{,}7)(y_i - 4{,}3) \\ &= 4{,}89 \end{aligned}$$

Korelasi (x,y) didapatkan dengan memasukkan nilai kovarian (x,y) dan membagi deviasi x dan deviasi y, sbb.

$$\begin{aligned} corr(x, y) &= \frac{\mathrm{cov}(x, y)}{dev(x)dev(y)} \\ &= \frac{4{,}89}{(2{,}9078)(3{,}1287)} = 0{,}5375 \end{aligned}$$

Soal 8:

Cari nilai kovarian dan korelasi dua sinyal $s_1(i)$ dan $s_2(i)$ berikut ini:

$s_1(i)$ = 0.1951 0.3827 0.5556 0.7071 0.8315 0.9239 0.9808 1.0000 0.9808 0.9239 0.8315
0.7071 0.5556 0.3827 0.1951 0.0000 -0.1951 -0.3827 -0.5556 -0.7071 -0.8315 -0.9239
-0.9808 -1.0000 -0.9808 -0.9239 -0.8315 -0.7071 -0.5556 -0.3827 -0.1951 -0.0000

$s_2(i)$ = -0.1951 -0.3827 -0.5556 -0.7071 -0.8315 -0.9239 -0.9808 -1.0000 -0.9808 -0.9239 -0.8315
-0.7071 -0.5556 -0.3827 -0.1951 -0.0000 0.1951 0.3827 0.5556 0.7071 0.8315 0.9239
0.9808 1.0000 0.9808 0.9239 0.8315 0.7071 0.5556 0.3827 0.1951 0.0000

Dengan cara manual?
Dengan menggunakan program?