BAB III
VARIABEL ACAK DAN
DISTRIBUSI PROBABILITAS

3.1. Konsep variabel acak (random variable)

Sebuah variabel *y* adalah **variabel acak** jika nilai-nilai yang dimiliki oleh *y* adalah suatu kemungkinan atau peristiwa acak.

Suatu **variabel acak diskrit** adalah variabel yang dapat memiliki sejumlah nilai yang dapat dihitung. Dapat dihitung berarti bahwa nilai-nilai yang dimilikinya adalah bilangan bulat positif. Contoh: jumlah mobil mobil yang terjual setiap bulan, jumlah kecelakaan lalu lintas setahun, jumlah televisi yang diproduksi setiap tahun.

Suatu **variabel acak menerus (continue)** adalah variabel yang dapat memiliki nilai yang tak berhingga dengan titik-titik dalam suatu interval garis. Kata menerus berarti melakukan tanpa terputus-putus

Contoh: banyaknya minyak yang dipompa setiap jam dari sebuah sumur, banyaknya energi yang dihasilkan oleh PLN setiap hari, dll.

3.2 Distribusi Probabilitas Diskrit

Distribusi probabilitas variabel acak yang diskrit dapat dinyatakan dengan sebuah rumus, atau sebuah grafik yang memperlihatkan probabilitas yang berkaitan dengan setiap nilai dari variabel acak.

Kumpulan pasangan terurut (x, $f(x)$) adalah fungsi probabilitas/fungsi massa probabilitas atau distribusi probabilitas dari variavel acak diskrit X, bila untuk setiap hasil x, dipenuhi persyaratan berikut:

1. $f(x) \geq 0$

2. $\sum_{x} f(x) = 1$

3. $P(X = x) = f(x)$

Contoh 3.1: Dalam suatu pengiriman 8 komputer ke suatu toko terdapat 3 komputer yang rusak. Bila sekolah membeli 2 komputer tersebut secara acak, tentukan distribusi probabilitas untuk jumlah computer yang rusak.

Solusi:

$$f(0) = P(X = 0) = \frac{C_0^3 C_2^5}{C_2^8} = \frac{\binom{3}{0}\binom{5}{2}}{\binom{8}{2}} = \frac{10}{28}$$

$$f(1) = P(X = 1) = \frac{\binom{3}{1}\binom{5}{1}}{\binom{8}{2}} = \frac{15}{28}$$

$$f(2) = P(X = 2) = \frac{\binom{3}{2}\binom{5}{0}}{\binom{8}{2}} = \frac{3}{28}$$

Maka distribusi probabilitas X adalah:

X	0	1	2
f(x)	10/28	15/28	3/28

Contoh 3.2: Bila agen mobil berencana menjual 50% dari mobilnya yang dilengkapi dengan kantong udara, tentukan rumusan untuk menentukan distribusi probabilitas jumlah mobil yang dilengkapi dengan kantong udara diantara 4 mobil yang akan dijual.

Solusi: Karena probabilitas untuk menjual mobil yang dilengkapi dengan kantong udara adalah 0,5. Maka ruang sampel sebanyak 2^4 = 16 memiliki peluang yang sama. Untuk menjual *x* mobil dengan kantong udara dan 4- *x* mobil tanpa kantong udara dapat dilakukan dengan $C_x^4 = \binom{4}{x}$ cara. Sehingga distribusi probabilitas dapat dinyatakan dengan rumusan berikut:

$$f(x) = \frac{\binom{4}{x}}{16} \quad \text{untuk } x = 0,1,2,3,4.$$

Dalam banyak kasus, kita perlu mengetahui probabilitas dari suatu variabel acak yang lebih kecil atau sama dengan suatu nilai riil *x*; ditulis dengan $F(x) = P(X \le x)$. Nilai $F(x)$ tersebut dinamakan **distribusi kumulatif.**

Distribusi kumulatif $F(x)$ dari suatu variabel acak diskrit dengan distribusi probabilitas $f(x)$ adalah:

$$F(x) = P(X \le x) = \sum_{t \le x} f(t) \quad \text{untuk } -\infty < x < \infty.$$

Contoh 3.3: Tentukan distribusi kumulatif variabel acak *X* dari soal sebelumnya.

Solusi:

$$f(0) = \frac{\binom{4}{0}}{16} = \frac{1}{16}; \quad f(1) = \frac{1}{4}; \quad f(2) = \frac{3}{8}; \quad f(3) = \frac{1}{4}; \quad f(4) = \frac{1}{16}$$

$$F(0) = f(0) = \frac{1}{16}; \quad F(1) = f(0) + f(1) = \frac{5}{16}; \quad F(2) = f(0) + f(1) + f(2) = \frac{11}{16}$$

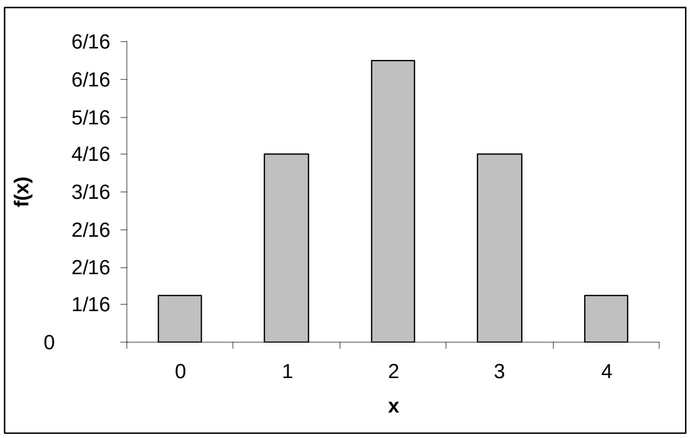
$$F(3) = f(0) + f(1) + f(2) + f(3) = \frac{15}{16};$$

$F(4) = f(0) + f(1) + f(2) + f(3) + f(4) = 1$

maka:

$$F(x) = \begin{cases} 0 & \text{untuk } x < 0 \\ 1/16 & \text{untuk } 0 \le x < 1 \\ 5/16 & \text{untuk } 1 \le x < 2 \\ 11/16 & \text{untuk } 2 \le x < 3 \\ 15/16 & \text{untuk } 3 \le x < 4 \\ 1 & \text{untuk } x \ge 4 \end{cases}$$

Untuk memudahkan Distribusi probabilitas dapat dibuat dalam bentul Bar Chart, dan histogram.



Histogram distribusi probabilitas

3.3 Distribusi probabilitas menerus

DEFINISI: Fungsi $f(x)$ adalah **fungsi kepadatan probabilitas (*probability density function*)** untuk variabel acak menerus, bila:

1. $f(x) \ge 0$ untuk
2. $\int_{-\infty}^{\infty} f(x)dx = 1$
3. $P(a < X < b) = \int_a^b f(x)dx$

Contoh 3.4: bila kesalahan perhitungan suhu dalam uji laboratorium adalah variabel acak dengan fungsi kepadatan probabilitas sbb:

$$f(x)=\begin{cases}\dfrac{x^2}{3}, & -1<x<2\\ 0, & \text{selainnya}\end{cases}$$

a. buktikan teorema point 2 diatas

b. Tentukan $P(0<X\leq 1)$

Solusi:

a. $\int_{-\infty}^{\infty} f(x)dx=\int_{-1}^{2}\frac{x^2}{3}dx=\left.\frac{x^3}{9}\right|_{-1}^{2}=1$

b. $P(0<X\leq 1)=\int_{0}^{1}\frac{x^2}{3}dx=\left.\frac{x^3}{9}\right|_{0}^{1}=1/9$

DEFINISI: Distribusi kumulatif (cumulative distribution) dari suatu variabel acak menerus X dengan fungsi kepadatan $f(x)$ adalah:

$$F(x)=P(X\leq x)=\int_{-\infty}^{x} f(t)dt \quad \text{untuk } -\infty<x<\infty$$

$$P(a<X<b)=F(b)-F(a) \quad \text{dan} \quad f(x)=\frac{dF(x)}{dx}$$

Contoh 3.5: Seperti soal sebelumnya, tentukan $F(x)$, dan gunakan untuk mengevaluasi $P(0<X\leq 1)$

Solusi:

untuk -1< x <2 ; $\quad F(x)=\int_{-\infty}^{x} f(t)dt=\int_{-1}^{x}\frac{t^3}{3}dt=\left.\frac{t^3}{9}\right|_{-1}^{x}=\frac{x^3+1}{9}$

Maka:

$$F(X)=\begin{cases}0; & x\le -1\\ \dfrac{x^3+1}{9}; & -1\le x<2\\ 0; & x\ge 2\end{cases}$$

Distribusi kumulatif $F(x)$:

$$P(0<x\le 1)=F(1)-F(0)=\frac{2}{9}-\frac{1}{9}=\frac{1}{9}$$

3.4 Distribusi probabilitas gabungan

DEFINISI: Bila X dan Y adalah variabel acak diskrit, distribusi probabilitas untuk suatu kejadian secara bersamaan dapat dinyatakan dengan suatu fungsi dengan nilai $f(x,y)$ untuk setiap pasang nilai (x,y). Pada umumnya fungsi diatas disebut dengan **distribusi probabilitas gabungan (joint probability distribution)** dari X dan Y.

Dalam kasus diskrit:

$$f(x,y)=P(X=x,Y=y);$$

nilai $f(x,y)$ adalah probabilitas X dan Y yang terjadi secara bersamaan.

Fungsi $f(x,y)$ adalah **distribusi probabilitas gabungan (joint probability distribution) atau fungsi masa probabilitas (probability mass function)** dari variabel acak diskrit X dan Y, bila:

1. $f(x,y)\ge 0$ untuk semua (x,y)
2. $\sum_x \sum_y f(x,y)=1$
3. $P(X=x,Y=y)=f(x,y)$

Untuk setiap daerah A dalam bidang xy, $P[(X,Y)\in A]=\sum_A \sum f(x,y)$

Contoh 3.6: dua buah pulpen dipilih secara acak dari suatu kotak yang berisi 3 pulpen biru, 2 pulpen merah, dan 3 pulpen hijau. Bila *X* adalah jumlah pulpen biru dan *Y* adalah jumlah pulpen merah yang dipilih. Tentukan:

a. Fungsi probabilitas gabungan *f*(*x*,*y*)
b. $P[(X,Y)\in A]$, A adalah daerah $\{(x,y)|x+y\le 1\}$

Solusi: a. Nilai pasangan yang mungkin adalah (0,0), (0,1), (1,0), (1,1), (0,2), (2,0). Jumlah cara untuk memilih dua pulpen dari 8 pulpen adalah $\binom{8}{2}=28$. Jumlah cara untuk memilih 1 pulpen merah dari 2 pulpen merah dan 1 pulpen hijau dari 3 pulpen hijau adalah: $\binom{2}{1}\binom{3}{1}=6$. Karena itu *f*(1,1) = 6/28 = 3/14. Sehingga formulasi matematisnya dapat dinyatakan sebagai berikut:

$$f(x,y)=\frac{\binom{3}{x}\binom{2}{y}\binom{3}{2-x-y}}{\binom{8}{2}}$$

Tabel: Distribusi probabilitas gabungan

f(*x*,*y*)		*X*			Total baris
		0	1	2	
y	0	3/28	9/28	3/28	15/28
	1	3/14	3/14		3/7
	2	1/28			1/28
Total kolom		5/14	15/28	3/28	1

untuk $x=0,1,2;\ y=0,1,2;\ 0\le x+y\le 2$

(b). $P[(X,Y)\in A] = P(X+Y\le 1) = f(0,0)+f(0,1)+f(1,0)$

$$= 3/28 + 3/14 + 9/28 = 9/14$$

DEFINISI: Fungsi *f*(*x*,*y*) adalah **fungsi kepadatan gabungan (joint density function)** dari suatu variabel acak menerus *X* dan *Y* bila;

1. $f(x,y)\ge 0 \quad \text{untuk semua } (x,y)$
2. $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)dxdy = 1$
3. $P[(X,Y)\in A] = \iint_A f(x,y)dxdy$

Contoh 3.7: Perusahaan permen mendistribusikan kotak-kotak coklat dengan campuran dari cream, kofi, kacang, yang dilapisi oleh coklat tipis dan tebal. Kotak dipilih secara acak, yaitu *X* dan *Y* masing-masing mewakili proporsi coklat cream yang dilapisi coklat tipis dan tebal. Bila diasumsikan fungsi kepadatan gabungan adalah:

$$f(x,y) = \begin{cases} \frac{2}{5}(2x+3y), & 0\le x\le 1, 0\le y\le 1 \\ 0, & \text{selainnya} \end{cases}$$

a. Buktikan point 2 definisi diatas.
b. Tentukan $P[(X,Y)\in A]$, A adalah daerah $\{(x,y) | 0 < x < \frac{1}{2}, \frac{1}{4} < y < \frac{1}{2}\}$

Solusi:

a. $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)dxdy = \int_0^1\int_0^1 \frac{2}{5}(2x+3y)dxdy = \int_0^1 \frac{2}{5}\left(x^2+3xy\right)\Big|_0^1 dy$

$$= \int_0^1\left(\frac{2}{5}+\frac{6y}{5}\right)dy = \frac{2y}{5}+\frac{3y^2}{5}\Big|_0^1 = \frac{2}{5}+\frac{3}{5} = 1$$

b. $P[(X,Y)\in A] = P\left(0 < x < \frac{1}{2}, \frac{1}{4} < y < \frac{1}{2}\right)$

$$= \int_{1/4}^{1/2}\int_0^{1/2} \frac{2}{5}(2x+3y)dxdy = \int_{1/4}^{1/2} \frac{2}{5}\left(x^2+3xy\right)\Big|_0^{1/2} dy$$

$$= \int_{1/4}^{1/2} \left(\frac{1}{10} + \frac{3y}{5} \right) dy = \frac{y}{10} + \frac{3y^2}{10} \Bigg|_{1/4}^{1/2} = 13/160$$

DEFINISI: **distribusi marginal (marginal distribution)** dari X saja dan Y saja adalah:

Untuk kasus diskrit: $g(x) = \sum_y f(x, y)$ dan $h(y) = \sum_x f(x, y)$

Untuk kasus menerus: $g(x) = \int_{-\infty}^{\infty} f(x, y) dy$ dan $h(y) = \int_{-\infty}^{\infty} f(x, y) dx$

Dalam kasus diskrit, istilah marginal digunakan untuk menghitung nilai dari $g(x)$ dan $h(y)$ untuk masing-masing kolom dan baris saja.

Contoh 3.8: Lihat tabel sebelumnya pada contoh 3.6, tentukan distribusi marginal dari *X* saja dan *Y* saja.

Solusi: Untuk variabel acak *X*,

$$P(X = 0) = g(0) = \sum_{y=0}^{2} f(0, y) = f(0,0) + f(0,1) + f(0,2) = \frac{3}{28} + \frac{3}{14} + \frac{1}{28} = 5/14$$

$$P(X = 1) = g(1) = \sum_{y=0}^{2} f(1, y) = f(1,0) + f(1,1) + f(1,2) = \frac{9}{28} + \frac{3}{14} + 0 = 15/28$$

$$P(X = 2) = g(2) = \sum_{y=0}^{2} f(2, y) = f(2,0) + f(2,1) + f(2,2) = \frac{3}{28} + 0 + 0 = 3/28$$

dengan cara yang sama $h(y)$ juga dapat dihitung; sehingga:

x	0	1	2
$g(x)$	5/14	15/28	3/28

y	0	1	2
$h(y)$	15/28	3/7	1/28

DEFINISI: Bila *X* dan *Y* adalah dua variabel acak, diskrit atau menerus, **distribusi bersyarat (conditional probability)** dari variabel acak *Y*, dengan syarat *X* = *x* adalah:

$$f(y|x) = \frac{f(x,y)}{g(x)}, \qquad g(x) > 0$$

Hal yang sama, distribusi bersyarat variabel acak *X*, dengan syarat *Y* = *y*, adalah:

$$f(x|y) = \frac{f(x,y)}{h(y)}, \qquad h(y) > 0$$

Bila ingin menentukan probabilitas dari variabel acak diskrit *X* yang terletak antara *a* dan *b*, bila diberikan *Y* = *y*, maka:

$$P(a < X < b | Y = y) = \sum_{x} f(x|y)$$

Bila *X* dan *Y* menerus:

$$P(a < X < b | Y = y) = \int_a^b f(x|y)dx$$

Contoh 3.9: Lihat contoh 3.6, tentukan probabilitas bersyarat dari *X*, bila diberikan *Y* =1, dan tentukan pula $P(X = 0|Y = 1)$

Solusi: mencari $f(x|y)$, bila *Y* = 1,

$$h(1) = \sum_{x=0}^{2} f(x,1) = \frac{3}{14} + \frac{3}{14} + 0 = \frac{3}{7}$$

$$f(x|1) = \frac{f(x,1)}{h(1)} = \frac{7}{3} f(x,1); \qquad x = 0,1,2$$

maka: $f(0|1) = \frac{7}{3} f(0,1) = \left(\frac{7}{3}\right)\left(\frac{3}{14}\right) = 1/2$

$$f(1|1) = \frac{7}{3} f(1,1) = \left(\frac{7}{3}\right)\left(\frac{3}{14}\right) = 1/2$$

$$f(2|1) = \frac{7}{3} f(2,1) = \left(\frac{7}{3}\right)(0) = 0$$

Sehingga distribusi bersyarat dari *X*, bila *Y* = 1 adalah:

x	0	1	2
$f(x\|1)$	½	½	0

Akhirnya, $P(X=0|Y=1)=f(0|1)=1/2$

Contoh 3.10: bila fungsi kepadatan gabungan dinyatakan sebagai berikut:

$$f(x,y)=\begin{cases}\dfrac{x(1+3y^2)}{4}, & 0<x<2, 0<y<1\\ 0, & \text{selainnya}\end{cases}$$

Tentukan $g(x), h(y), f(x|y)$ dan evaluasi $P\left(\frac{1}{4}<X<\frac{1}{2}\middle|Y=\frac{1}{3}\right)$

Solusi: $$g(x)=\int_{-\infty}^{\infty} f(x,y)\,dy=\int_0^1 \frac{x(1+3y^2)}{4}\,dy$$

$$=\frac{xy}{4}+\frac{xy^2}{4}\Bigg|_{y=0}^{y=1}=\frac{x}{2}, \qquad 0<x<2$$

$$h(y)=\int_{-\infty}^{\infty} f(x,y)\,dx=\int_0^2 \frac{x(1+3y^2)}{4}\,dx$$

$$=\frac{x^2}{8}+\frac{3x^2y^2}{8}\Bigg|_{x=0}^{x=2}=\frac{1+3y^2}{2}, \qquad 0<y<1$$

Maka: $$f(x|y)=\frac{f(x,1)}{h(1)}=\frac{x(1+3y^2)/4}{(1+3y^2)/2}=\frac{x}{2}; \qquad 0<x<2$$

dan $$P\left(\tfrac{1}{4}<X<\tfrac{1}{2}\middle|Y=\tfrac{1}{3}\right)=\int_{1/4}^{1/2}\frac{x}{2}\,dx=3/64$$

DEFINISI: bila X dan Y adalah dua variabel acak, diskrit atau menerus dengan distribusi probabilitas gabungan $f(x,y)$ dan distribusi marginal adalah $g(x)$ dan $h(y)$. Variabel acak X dan Y dikatakan bebas statistik (statistically independent) jika dan hanya jika;

$$f(x,y)=g(x)h(y) \text{ untuk semua } (x,y)$$

Contoh 3.11: Tunjukkan bahwa 3.6 tidak bebas statistik.

Solusi: dari tabel diketahui:

$$f(0,1) = 3/14$$

$$g(0) = \sum_{y=0}^{2} f(0,y) = \frac{3}{28} + \frac{3}{14} + \frac{1}{28} = \frac{5}{14}$$

$$h(1) = \sum_{x=0}^{2} f(x,1) = \frac{3}{14} + \frac{3}{14} + 0 = \frac{3}{7}$$

Bukti: $f(0,1) \neq g(0)h(1)$

DEFINISI: Bila X_1, X_2, X_3,..... X_n, adalah variabel acak diskrit atau menerus, dengan distibusi probabilitas gabungan $f(x_1, x_2, x_3,...x_n)$ dan distribusi marginal masing-masing $f_1(x_1), f_2(x_2), f_3(x_3),.... f_n(x_n)$. Variabel acak X_1, X_2, X_3,..... X_n disebut **saling bebas statistik (mutually statistically independent)** jika dan hanya jika:

$$f(x_1, x_2, x_3,...x_n) = f_1(x_1) f_2(x_2) f_3(x_3)..... f_n(x_n) \text{ untuk semua } (x_1, x_2, x_3,...x_n)$$

Contoh 3.12: bila umur papan almari adalah variabel acak dengan fungsi kepadatan probabilitas sbb:

$$f(x) = \begin{cases} e^{-x}, & \text{x} > 0 \\ 0 & \text{selainnya} \end{cases}$$

Bila X1, X2, X3, menunjukkan papan yang diambil secara acak, tentukan $P(X_1 < 2, 1 < X_2 < 3, X_3 > 2)$