SF1920/SF1921 Sannolikhetsteori och statistik 6,0 hp
Föreläsning 5
Funktioner av stokastiska variabler
Flerdimensionella stokastiska variabler

Jörgen Säve-Söderbergh

Funktioner av en stokastisk variabel

Låt X vara en kontinuerlig stokastisk variabel.

Vi är intresserade av

$$Y = g(X).$$

Y är också en stokastisk variabel.

Om vi kan finna fördelningsfunktionen för Y

$$F_Y(y) = P(Y \leq y) = P[g(X) \leq y],$$

så kan vi finna täthetsfunktionen genom $F'_Y(y) = f_Y(y)$.

Exempel: Linjär transformation

Antag att X är en kontinuerlig stokastisk variabel, samt att

$$Y = aX + b,$$

där $a > 0$.

$$\begin{aligned} F_Y(y) &= P(Y \leq y) = P(aX + b \leq y) \\ &= P(aX \leq y - b) \\ &= P(X \leq \frac{y-b}{a}) \\ &= F_X\left(\frac{y-b}{a}\right). \end{aligned}$$

Tätheten blir då

$$f_Y(y) = f_X\left(\frac{y-b}{a}\right)\frac{1}{a}.$$

Antag att $g(\cdot)$ är en monoton funktion – växande

- Om $g(\cdot)$ är **växande**, så är

$$g(X) \leq y \quad \Leftrightarrow \quad X \leq g^{-1}(y).$$

$$\begin{aligned} F_Y(y) &= P(Y \leq y) = P[g(X) \leq y] \\ &= P\left[X \leq g^{-1}(y)\right] \\ &= F_X\left[g^{-1}(y)\right] \end{aligned}$$

$$\begin{aligned} f_Y(y) &= \frac{d}{dy} F_Y(y) = \frac{d}{dy} F_X\left[g^{-1}(y)\right] \\ &= f_X\left[g^{-1}(y)\right] \frac{d}{dy} g^{-1}(y) \end{aligned}$$

$g(\cdot)$ monoton – avtagande

- Om $g(\cdot)$ är **avtagande**

$$g(X) \leq y \quad \Leftrightarrow \quad X \geq g^{-1}(y),$$

så

$$\begin{aligned} F_Y(y) &= P[g(X) \leq y] \\ &= P\left[X \geq g^{-1}(y)\right] \\ &= 1 - P\left[X \leq g^{-1}(y)\right] \\ &= 1 - F_X\left[g^{-1}(y)\right] \end{aligned}$$

$$\begin{aligned} f_Y(y) &= \frac{d}{dy}\left(1 - F_X\left[g^{-1}(y)\right]\right) \\ &= -f_X\left[g^{-1}(y)\right]\frac{d}{dy}g^{-1}(y) \\ &= f_X\left[g^{-1}(y)\right]\left(-\frac{d}{dy}g^{-1}(y)\right) \end{aligned}$$

Om $g(\cdot)$ är monoton

- Om $g(\cdot)$ är **monoton**, d v s $g(\cdot)$ är växande eller avtagande, kan vi sammanfatta de bägge fallen ovan i en formel

$$f_Y(y) = f_X\left[g^{-1}(y)\right]\left|\frac{d}{dy}g^{-1}(y)\right|.$$

Tvådimensionell stokastisk variabel

Mät två egenskaper hos det utvalda elementet.

Example

På en av er kan vi mäta både längden och vikten.

Definition

En *tvådimensionell stokastisk variabel* $(X(\omega), Y(\omega))$ är en funktion definierad på ett utfallsrum Ω med värden i planet $\mathbb{R}^2$.

- (X, Y) är en förkortning för $(X(\omega), Y(\omega))$.
- Den stokastiska variabeln (X, Y) kommer att anta observerade värden (x, y) i $\mathbb{R}^2$ (som vi kan åskådliggöra).

Simultan fördelningsfunktion

Vi vill kunna räkna sannolikheter som $P((X, Y) \in A)$, där A är en delmängd av $\mathbb{R}^2$.

Definition

Funktionen

$$F_{X,Y}(x, y) = P(X \leq x, Y \leq y) = P[\{X(\omega) \leq x\} \cap \{Y(\omega) \leq y\}]$$

kallas *fördelningsfunktionen* för den tvådimensionella stokastiska variabeln (X, Y).

Diskreta tvådimensionella stokastiska variabler

Definition

Om både X och Y antar ett ändligt eller ett uppräkneligt oändligt antal olika värden kallar vi den tvådimensionella stokastiska variabeln (X, Y) *diskret*.

Det vanligaste fallet är att vi har icke-negativa tal.

Definition

Sannolikheten att $X = j$ och $Y = k$ betecknas

$$p_{X,Y}(j,k) = P(X = j, Y = k), \quad j = 0, 1, \ldots, \quad k = 0, 1, \ldots.$$

Storheterna $p_{X,Y}(j,k)$ kallas den *(simultana) sannolikhetsfunktionen* för (X, Y).

Storheterna $p_{X,Y}(j,k)$ har följande egenskaper:

(a) $0 \leq p_{X,Y}(j,k) \leq 1$.

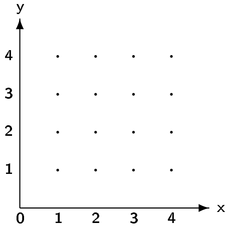
(b) $\sum_{j=0}^{\infty} \sum_{k=0}^{\infty} p_{X,Y}(j,k) = 1$.

$$P\left[(X,Y) \in A)\right] = \sum_{(j,k)\in A} p_{X,Y}(j,k), \quad \text{där } A \subset \mathbb{R}^2.$$

Example

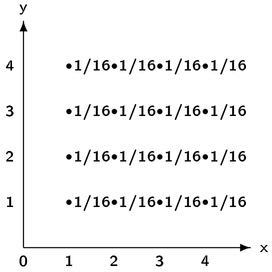
Kasta två fyrsidiga tärningar med olika färg, en röd och en svart.

Låt X beteckna den röda tärningen och Y den svarta.



Diskret fördelning över $(1,1),(1,2),\ldots,(4,4)$

Alla utfall har sannolikheten $1/16$.



Marginalfördelning

Låt X och Y vara simultant fördelade med sannolikhetsfunktionen $p_{X,Y}(j,k)$.

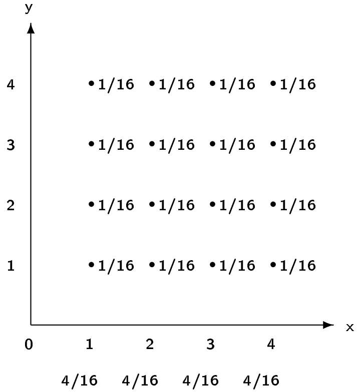
Sannolikhetsfunktionen för X kallas **marginella sannolikhetsfunktionen** för X och ges av

$$P(X=j)=p_X(j)=\sum_{k=0}^{\infty} p_{X,Y}(j,k)\,, \quad j=0,1,\ldots$$

Om vi summerar över j istället för k i $p_{X,Y}(j,k)$ får vi den marginella sannolikhetsfunktionen för Y.

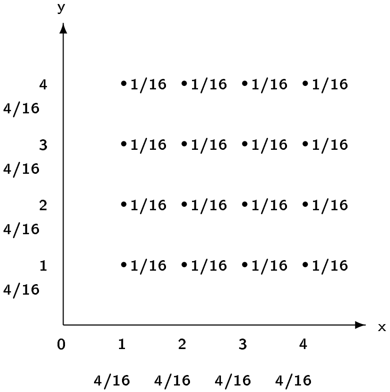
Diskret fördelning över $(1,1),(1,2),\ldots,(4,4)$

Vi beräknar den marginella sannolikhetsfunktionen för X nedan.



Diskret fördelning över $(1,1),(1,2),\ldots,(4,4)$

Den marginella sannolikhetsfunktionen för Y.



Kontinuerlig tvådimensionell stokastisk variabel

Definition

Om det finns en funktion $f_{X,Y}(x,y)$ sådan att

$$P\left[(X,Y)\in A\right] = \int\int_A f_{X,Y}(x,y)dxdy$$

för alla A säges den tvådimensionella stokastiska variabeln (X,Y) vara *kontinuerlig*.

Funktionen $f_{X,Y}(x,y)$ kallas *täthetsfunktionen* för (X,Y).

Täthetsfunktioner måste uppfylla

(a) $f_{X,Y}(x,y) \geq 0$;

(b) $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)dxdy = 1$

Kontinuerlig tvådimensionell stokastisk variabel

Om vi väljer A enligt sidan 83 har vi

$$F_{X,Y}(x,y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f_{X,Y}(u,v)dudv$$

vilket är *fördelningsfunktionen* för (X, Y).

Exempel Kontinuerlig tvådimensionell stokastisk variabel

$$f_{X,Y}(x, y) = 2e^{-x-y}, \quad 0 \leq x \leq y < \infty.$$

Eftersom exponentialfunktionen är icke-negativ, så har vi att $f_{X,Y}(x, y) \geq 0$.
Man kan visa att

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} 2e^{-x-y} dxdy = 1.$$

Alltså fungerar $f_{X,Y}(x, y)$ som täthetsfunktion för (X, Y).

Marginell täthetsfunktion

Om $f_{X,Y}(x,y)$ är täthetsfunktionen för en tvådimensionell stokastisk variabel (X,Y) kan vi definiera den *marginella täthetsfunktionen* för X genom

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy,$$

samt för Y

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx.$$

Marginell täthetsfunktion för tvådimensionell stokastisk variabel

$$f_{X,Y}(x,y) = 2e^{-x-y}, \quad 0 \le x \le y < \infty.$$

Välj ett godtyckligt x i $0 \le x < \infty$.

$$\begin{aligned}
f_X(x) &= \int_x^\infty 2e^{-x-y}\,dy \\
&= 2e^{-x} \int_x^\infty e^{-y}\,dy \\
&= 2e^{-x}\left[-e^{-y}\right]_{y=x}^{y=R} \\
&= 2e^{-x}\left[-e^{-R} + e^{-x}\right] \\
&\to 2e^{-x}e^{-x}, \text{ as } R \to \infty.
\end{aligned}$$

$$f_X(x) = \begin{cases} 2e^{-2x} & 0 \le x < \infty \\ 0 & \text{annars} \end{cases}$$

Marginell täthetsfunktion för tvådimensionell stokastisk variabel

Välj ett godtyckligt y i $0 \leq y < \infty$.

$$
\begin{aligned}
f_Y(y) &= \int_0^y 2e^{-x-y}\,dx \\
&= 2e^{-y}\int_0^y e^{-x}\,dx \\
&= 2e^{-y}\left[-e^{-x}\right]_{x=0}^{x=y} \\
&= 2e^{-y}\left[-e^{-y}+1\right].
\end{aligned}
$$

$$
f_Y(y) = \begin{cases} 2e^{-y}\left(1-e^{-y}\right) & 0 \leq y < \infty \\ 0 & \text{annars} \end{cases}
$$

De stokastiska variablerna X och Y säges vara **oberoende** om

$$P(X \in C, Y \in D) = P(X \in C)\, P(Y \in D).$$

Theorem

De stokastiska variablerna X och Y är oberoende om och endast

$$F_{X,Y}(x,y) = F_X(x)\, F_Y(y) \qquad \textit{för alla } x \textit{ och } y$$

$$p_{X,Y}(j,k) = p_X(j)\, p_Y(k) \qquad \textit{för alla } x \textit{ och } y$$

$$f_{X,Y}(x,y) = f_X(x)\, f_Y(y) \qquad \textit{för alla } x \textit{ och } y$$

Diskret fördelning över $(1,1),(1,2),\ldots,(4,4)$-continued

Är X or Y oberoende?
Oberoende kräver

$$p_{X,Y}(j,k) = p_X(j)\, p_Y(k)$$

för alla $j \in \{1,2,3,4\}$ and $k \in \{1,2,3,4\}$.

Om vi multiplicerar marginalerna för X och Y för godtyckliga värden på j och k

$$p_X(j)\, p_Y(k) = \frac{4}{16}\frac{4}{16} = \frac{1}{16} = p_{X,Y}(j,k)\,.$$

Det spelar ingen roll hur vi väljer j och k - alltid samma resultat.

Alltså är X och Y oberoende.

Kontinuerlig bivariat fördelning

(X, Y) har täthetsfunktionen

$$f_{X,Y}(x, y) = 2e^{-x-y}, \quad 0 \leq x \leq y < \infty.$$

Är X och Y oberoende?

$$f_{X,Y}(x, y) = f_X(x) f_Y(y) \qquad \text{för alla } x \text{ och } y$$

Om vi kan finna ett talpar (x, y) där detta inte gäller, har vi visat att X och Y är beroende. (Ett motexempel)

Kontinuerlig bivariat fördelning– motexempel

Vi ska visa att X och Y är beroende. Låt oss välja punkten $(1,1)$.

Marginalen för X i $(1,1)$

$$f_X(1) = 2e^{-2\cdot 1} = 2e^{-2}$$

Marginalen för Y i $(1,1)$

$$f_Y(1) = 2e^{-1}\left(1 - e^{-1}\right)$$

Den simultana tätheten i $(1,1)$

$$f_{X,Y}(1,1) = 2e^{-1-1} = 2e^{-2}$$

$$f_X(1)\, f_Y(1) = 2e^{-2}\, 2e^{-1}\left(1 - e^{-1}\right) = 4e^{-3}\left(1 - e^{-1}\right) \neq 2e^{-2} = f_{X,Y}(1,1)$$

Låt X och Y vara två oberoende stokastiska variabler.
Låt

$$Z = \max(X, Y).$$

$$Z \leq z \quad \Leftrightarrow \quad X \leq z \text{ och } Y \leq z$$

$$F_Z(z) = P(Z \leq z) = P(X \leq z \text{ och } Y \leq z) = F_X(z) F_Y(z)$$

X och Y behöver inte ha samma fördelning.

Minsta värdet av två

Återigen, låt X och Y vara två oberoende stokastiska variabler. Låt

$$Z = \min(X, Y).$$

Eftersom $Z > z$ om och endast om både $X > z$ och $Y > z$, har vi att

$$\begin{aligned} F_Z(z) &= P(Z \leq z) \\ &= 1 - P(Z > z) \\ &= 1 - P(X > z \text{ och } Y > z) \\ &= 1 - P(X > z)\, P(Y > z) \\ &= 1 - [1 - P(X \leq z)]\, [1 - P(Y \leq z)] \\ &= 1 - [1 - F_X(z)]\, [1 - F_Y(z)] \end{aligned}$$

Ej nödvändigt X och Y samma fördelning.

Summa av stokastiska variabler

Låt X och Y vara två diskreta stokastiska variabler som har en simultan fördelning $p_{X,Y}(i,j)$. (Inget antagande om oberoende!)
Låt

$$Z = X + Y.$$

Antag att X och Y endast antar icke-negativa värden.

$$p_Z(k) = P(Z = z) = P(X + Y = k)$$

Summa av stokastiska variabler

Händelsen $X + Y = k$ kan inträffa på många olika sätt

X	Y	Z
0	k	k
1	$k-1$	k
2	$k-2$	k
$\vdots$	$\vdots$	$\vdots$
k	0	k

De olika fallen är disjunkta händelser och händelsen $X + Y = k$ är unionen av dem.

Enligt Kolmogorovs tredje axiom ska vi summera över alla dessa fall för att beräkna $P(X + Y = k)$.

Summa av stokastiska variabler

$$P(X+Y=k) = \sum_{i+j=k} p_{X,Y}(i,j)$$

Alltså

$$p_Z(k) = \sum_{i+j=k} p_{X,Y}(i,j)$$

På samma sätt kan vi beräkna fördelningsfunktionen för Z;

$$F_Z(z) = \sum_{i+j\leq z} p_{X,Y}(i,j)$$

Summa av (**oberoende**) stokastiska variabler

Antag att X och Y är oberoende.
$p_{X,Y}(i,j) = p_X(i)\,p_Y(j)$

$$p_Z(k) = \sum_{i+j=k} p_X(i)\,p_Y(j)$$

Fördelningsfunktionen för Z;

$$F_Z(z) = \sum_{i+j\leq z} p_X(i)\,p_Y(j)$$