Grundläggande matematisk statistik

Flerdimensionella slumpvariabler

Uwe Menzel, 2018

uwe.menzel@matstat.org

www.matstat.org

1

Flerdimensionella slumpvariabler

Ett slumpförsök kan ge upphov till **flera** slumpvariabler (s.v.):

- kast med två tärningar:
  - s.v. $X =$ ögontalet tärning 1
  - s.v. $Y =$ ögontalet tärning 2

  tvådimensionell s.v.
- kast med pil på en måltavla:
  - s.v. $X =$ höjden mätt från nedre kanten
  - s.v. $Y =$ horisontellt avstånd till vänster sida av måtavlan
- slumpmässigt vald person
  - s.v. $X =$ längd i cm
  - s.v. $Y =$ vikt i kg
  - s.v. $Z =$ blodtryck i mmHg

  tredimensionell s.v.
- slumpmässigt vald barnfamilj:
  - s.v. $X =$ antal flickor
  - s.v. $Y =$ antal pojkar

www.matstat.org

2

Diskreta tvådimensionella slumpvariabler

tvådimensionell slumpvariabel: funktion $\Omega \to \mathbb{R}^2$

$i = 0, 1, 2, \ldots$
$j = 0, 1, 2, \ldots$

$\boldsymbol{X}, \boldsymbol{Y}$ (stora bokstäver): slumpvariabler
$\boldsymbol{i}, \boldsymbol{j}$ (små bokstäver): reella värden (ofta heltal för diskreta slumpvariabler)

Slumpvariabeln (X, Y) kallas **diskret** om X och Y bara antar ett **ändligt eller uppräkneligt oändligt** antal olika värden.

Sannolikhetsfunktion: $\boxed{P(X = i, Y = j) = p_{X,Y}(i, j)}$ ”joint probability distribution”

Sannolikhetsfunktionen anger sannolikheten för varje kombination av i och j (alltså för hela utfallsrummet).

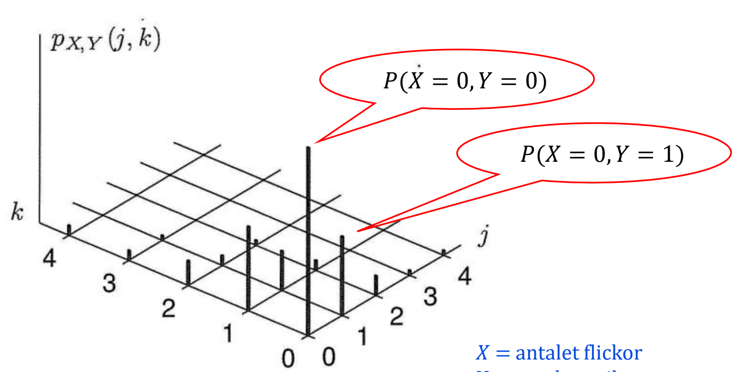
Normering: $\boxed{\sum_{i=0}^{\infty}\sum_{j=0}^{\infty} p_{X,Y}(i, j) = 1}$ [axiom 2: $P(\Omega) = 1$]

www.matstat.org

3

Diskreta tvådimensionella slumpvariabler

Sannolikhetsfunktionen kan åskådliggöras genom ett stolpdiagram (exempel barnkullar):



X = antalet flickor
Y = antalet pojkar

Bild: Blom et al, s. 85 (”barnkullar”)

www.matstat.org

4

Diskreta tvådimensionella slumpvariabler

Sannolikhetsfunktionen kan åskådliggöras genom en tabell (exempel barnkullar):

$P(X = 0, Y = 0)$

$P(X = 3, Y = 0)$

Y/X	$X = 0$	$X = 1$	$X = 2$	$X = 3$	$X = 4$
$Y = 0$	0.38	0.16	0.04	0.01	0.01
$Y = 1$	0.17	0.08	0.02	-	-
$Y = 2$	0.05	0.02	0.01	-	-
$Y = 3$	0.02	0.01	-	-	-
$Y = 4$	0.02	-	-	-	-

$P(X = 1, Y = 3)$

X = antalet flickor
Y = antalet pojkar

www.matstat.org

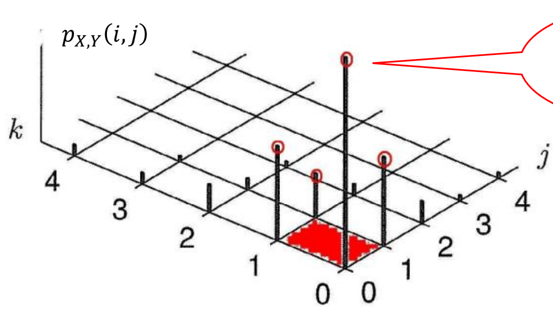
5

Diskreta tvådimensionella slumpvariabler

Fördelnings-funktion:

$$\boxed{F_{X,Y}(x,y) = P(X \leq x, Y \leq y) = \sum_{i \leq x} \sum_{j \leq y} p_{X,Y}(i,j)}$$

- analogt till endimensionell fördelningsfunktion
- x och y är reella tal, fördelningsfunktionen är alltså en kontinuerlig ”tvådimensionell trappfunktion”
- OBS!: $\leq$ tecknet viktigt för diskreta slumpvariabler



fyra slh:na måste summeras för att erhålla $F_{X,Y}(1,1)$

www.matstat.org

6

Diskreta tvådimensionella slumpvariabler

Marginalfördelningen (”margin” = rand)

Exempel: barnkullar (Blom et al.). Vi antar att vi känner sannolikhetsfunktionen $p_{X,Y}(i,j)$. (Antalet flickor X och antalet pojkar Y).
Vi vill nu bara veta sannolikheten för antalet flickor i en slumpmässigt vald familj, oavsett hur många pojkar det finns i familjen, dvs. vi behöver fördelningen för slumpvariablen X (flickor). Uppgiften är alltså att härleda den endimensionella fördelningen $p_X(i)$ från den tvådimensionella fördelningen $p_{X,Y}(i,j)$:

$$\boxed{p_X(i) = \sum_{j=0}^{\infty} p_{X,Y}(i,j)}$$

summeras över alla värden för s.v. Y, dvs. över j

X = antalet flickor
Y = antalet pojkar

$$\boxed{p_Y(j) = \sum_{i=0}^{\infty} p_{X,Y}(i,j)}$$

summeras över alla värden för s.v. X, dvs. över i

www.matstat.org

7

Diskreta tvådimensionella slumpvariabler

Marginalfördelningen (”margin” = rand)

X = antalet flickor
Y = antalet pojkar

Y/X	$X = 0$	$X = 1$	$X = 2$	$X = 3$	$X = 4$	Σ
$Y = 0$	0.38	0.16	0.04	0.01	0.01	0.60
$Y = 1$	0.17	0.08	0.02			0.27
$Y = 2$	0.05	0.02	0.01			0.08
$Y = 3$	0.02	0.01				0.03
$Y = 4$	0.02					0.02
Σ	0.64	0.27	0.07	0.01	0.01	1

$P(Y = 0)$
$P(Y = 2)$
$P(Y = 4)$
$P(X = 0)$
$P(X = 2)$
$P(X = 4)$

www.matstat.org

8

Diskreta tvådimensionella slumpvariabler

Multinomialfördelningen:

Exempel: Urna med kulor av r olika färg, drar n kulor med återläggning.
Hur stor är sannolikheten att dra

- k_1 kulor av färg 1, (slumpvariabel X_1: antalet dragna kulor med färg 1)
- k_2 kulor av färg 2, (slumpvariabel X_2: antalet dragna kulor med färg 2)
- ...
- k_r kulor av färg r ? (slumpvariabel X_r: antalet dragna kulor med färg r)

Sannolikhetsfunktion för multinomialfördelningen:

$$\boxed{p_{X_1, X_2, \cdots, X_r}(k_1, k_2, \cdots, k_r) = \frac{n!}{k_1! \cdot k_2! \cdot \ldots \cdot k_r!} \cdot {p_1}^{k_1} \cdot {p_2}^{k_2} \cdot \ldots \cdot {p_r}^{k_r}}$$

$$\sum_{i=1}^{r} p_i = 1$$ för att i varje dragning måste vi få **en** av de föreliggande färgarna

$$\sum_{i=1}^{r} k_i = n$$ vi drar totalt n gånger – antalen k_i måste summeras till n

www.matstat.org

9

Diskreta tvådimensionella slumpvariabler

Multinomialfördelningen, speciellt $r = 2$:

Om r (antalet färger) är bara 2 (t.ex. svart och vit): $r = 2$

$$p_{X_1, X_2}(k_1, k_2) = \frac{n!}{k_1! \cdot k_2!} \cdot {p_1}^{k_1} \cdot {p_2}^{k_2}$$ sannolikhetsfunktion

$k_2 = n - k_1$ antalet vita och svarta måste vara antalet dragna

$p_2 = 1 - p_1$ slh. att dra en vit eller svart måste adderas till 1

$$p_{X_1, X_2}(k_1, k_2) = \frac{n!}{k_1! \cdot (n-k_1)!} \cdot {p_1}^{k_1} \cdot (1-p_1)^{(n-k_1)} = \underbrace{\binom{n}{k_1} \cdot {p_1}^{k_1} \cdot (1-p_1)^{(n-k_1)}}_{\text{Binomialfördelning}}$$

Binomialfördelningen är ett speciellt fall av multinomialfördelningen för $r = 2$

www.matstat.org

10

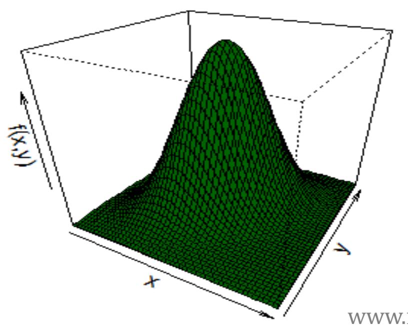
Kontinuerliga tvådimensionella slumpvariabler

Tvådimensionell slumpvariabel: funktion $\Omega \to \mathbb{R}^2$

$\boldsymbol{X}, \boldsymbol{Y}$ (stora bokstäver): slumpvariabler
$\boldsymbol{x}, \boldsymbol{y}$ (små bokstäver): reella värden

Slumpvariabeln (X, Y) kallas **kontinuerlig** om X och Y **inte** är diskreta.

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



Normering:

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

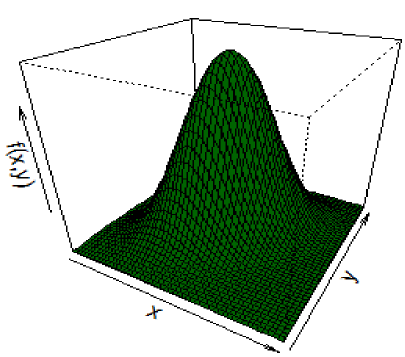
(axiom 2)

www.matstat.org

11

Kontinuerliga tvådimensionella slumpvariabler

Täthetsfunktion $f_{X,Y}(x, y)$



Beräkna sannolikheten att (X, Y) ligger inom ett område $\boldsymbol{A}$:

$$\boxed{P(X, Y \in A) = \int\int_A f_{X,Y}(x,y)\,dxdy}$$

täthetsfunktionen integreras över detta område (A)!

www.matstat.org

12

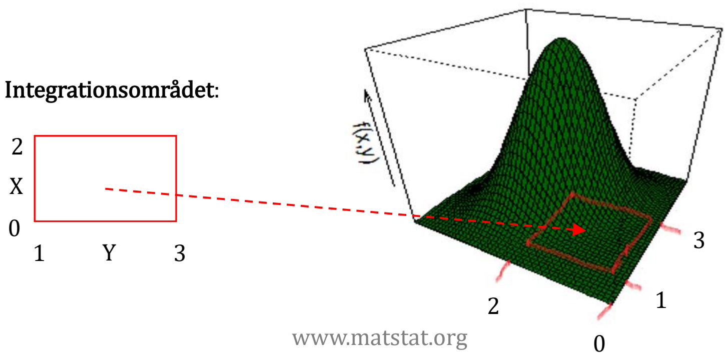
Kontinuerliga tvådimensionella slumpvariabler

Exempel: Låt $f_{X,Y}(x,y)$ vara given. Beräkna sannolikheten att X ligger mellan 0 och 2, och att Y ligger mellan 1 och 3:

$$P(X,Y \in A) = \int\int_A f_{X,Y}(x,y)\,dxdy \quad \Rightarrow$$

$$P(0 < X \leq 2, 1 < Y \leq 3) = \int_1^3 \int_0^2 f_{X,Y}(x,y)\,dxdy$$

Integrationsområdet:



Kontinuerliga tvådimensionella slumpvariabler

Fördelningsfunktion:

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

tvärtom gäller:

$$\boxed{f_{X,Y}(x,y) = \frac{\partial^2}{\partial x\,\partial y} F_{X,Y}(x,y)}$$

Kontinuerliga tvådimensionella slumpvariabler

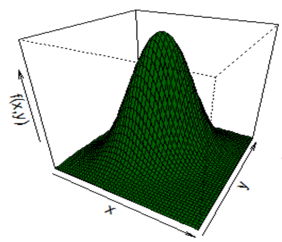
Marginalfördelningar:

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

Beräkning av marginalfördelningen för X från den simultana täthetsfunktionen

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

Beräkning av marginalfördelningen för Y från den simultana täthetsfunktionen



integration över
en dimension

15

Kontinuerliga tvådimensionella slumpvariabler

Marginalfördelning för fördelningsfunktionen:

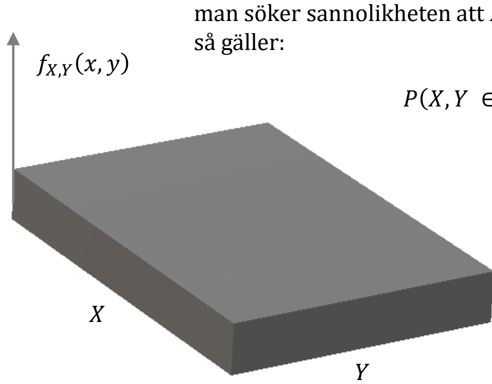
$$\boxed{F_X(x) = \lim_{y\to\infty} F_{X,Y}(x,y)} \qquad \boxed{F_Y(y) = \lim_{x\to\infty} F_{X,Y}(x,y)}$$

$$\begin{aligned}\lim_{y\to\infty} F_{X,Y}(x,y) &= \lim_{y\to\infty} \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(u,v)\,dv\,du \\ &= \int_{-\infty}^{x} \underbrace{\int_{-\infty}^{+\infty} f_{X,Y}(u,v)\,dv}_{f_X(u)}\,du = \int_{-\infty}^{x} f_X(u)\,du\end{aligned}$$

16

Likformigt kontinuerlig tvådimensionell fördelning

Simultan **täthetsfunktion**: $f_{X,Y}(x,y) = const$ för alla x, y



Likformig kontinuerlig 2-dimensionell fördelning:
Om X, Y är definerade på ett område med arean $\boldsymbol{B}$, och om man söker sannolikheten att X, Y ligger i ett delområde $\boldsymbol{A}$, så gäller:

$$P(X,Y \in A) = \frac{arean\ av\ A}{arean\ av\ B}$$

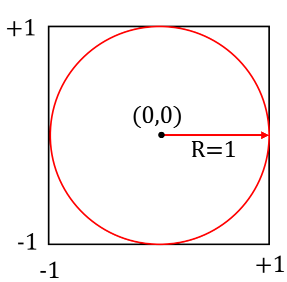
www.matstat.org

17

Likformigt kontinuerlig tvådimensionell fördelning

$$P(X,Y \in A) = \frac{arean\ av\ A}{arean\ av\ B}$$

Exempel: Blom et al., s. 88
En punkt väljs slumpmässigt i kvadraten $(X,Y) = (\pm 1, \pm 1)$.
Söker: sannolikheten $P(X^2 + Y^2 \leq 1)$



$X^2 + Y^2 \leq 1$

Kom ihåg: $X^2 + Y^2 = R^2$ är ekvationen för en cirkel med radien R och mittpunkt (0,0)

$$P(X^2 + Y^2 \leq 1) = \frac{arean\ av\ cirkeln}{arean\ av\ rektangeln} = \frac{\pi \cdot R^2}{2 \cdot 2} = \frac{\pi}{4}$$

www.matstat.org

18

Oberoende flerdimensionella slumpvariabler

Slumpvariablerna X och Y är oberoende, **om**:

diskreta tvådimensionella s.v.:

$p_{X,Y}(i,j) = p_X(i) \cdot p_Y(j)$ för **alla** i, j

kontinuerliga tvådimensionella s.v.:

$f_{X,Y}(x,y) = f_X(x) \cdot f_Y(y)$ för **alla** x, y

Kom ihåg:
$P(A \cap B) = P(A) \cdot P(B)$
om A och B oberoende

diskreta och kontinuerliga tvådimensionella slumpvariabler:

$F_{X,Y}(x,y) = F_X(x) \cdot F_Y(y)$ för **alla** x, y

www.matstat.org

19

Oberoende flerdimensionella slumpvariabler

diskreta tvådimensionella s.v.: $p_{X,Y}(i,j) = p_X(i) \cdot p_Y(j)$ för **alla** i, j

$p_{X,Y}(1,1)$ $p_Y(1)$

Y/X	$X=1$	$X=2$	$X=3$	marg.
$Y=1$	0.04	0.06	0.1	0.2
$Y=2$	0.02	0.03	0.05	0.1
$Y=3$	0.14	0.21	0.35	0.7
marg.	0.2	0.3	0.5	1

$p_X(1)$

$p_{X,Y}(i,j) = p_X(i) \cdot p_Y(j)$ för **alla** i, j

$$\left.\begin{array}{l} 0.04 = 0.2 \cdot 0.2 \\ 0.06 = 0.3 \cdot 0.2 \\ \ldots \\ \ldots \\ 0.35 = 0.5 \cdot 0.7 \end{array}\right\}$$

måste gälla för alla celler om oberoendet föreligger

www.matstat.org

20

Oberoende flerdimensionella slumpvariabler

Exempel: livstid hos två glödlampor, Blom s. 91

slumpvariabel X: livstid glödlampa 1
slumpvariabel Y: livstid glödlampa 2

antas vara oberoende, båda exponentiellt fördelade

$$f_X(x) = \left\{ \begin{array}{ll} a \cdot e^{-a \cdot x} & x > 0 \\ 0 & x \leq 0 \end{array} \right\} \qquad f_Y(y) = \left\{ \begin{array}{ll} b \cdot e^{-b \cdot y} & y > 0 \\ 0 & y \leq 0 \end{array} \right\}$$

täthets-funktioner

Söker: slh. att båda glödlamporna har en livstid mindre än t : $P(X \leq t, Y \leq t)$?

$$P(X \leq t, Y \leq t) = F_{X,Y}(t,t) = \int_0^t \int_0^t f_{X,Y}(x,y)\,dx\,dy$$

använd oberoendet

$$= \int_0^t \int_0^t f_X(x) \cdot f_Y(y)\,dx\,dy$$

använd integrationsregler

$$= \int_0^t f_X(x)\,dx \cdot \int_0^t f_Y(y)\,dy$$

fördelningsfunktion för $Exp(\mu)$

$$= F_X(t) \cdot F_Y(t) = \underline{\underline{\left(1 - e^{-a \cdot t}\right) \cdot \left(1 - e^{-b \cdot t}\right)}}$$

www.matstat.org

21

Maximumet av flera slumpvariabler

$X_1, X_2, \ldots, X_n$ **oberoende** slumpvariabler (med kända fördelningar)

Söker: fördelning för $Z = max(X_1, X_2, \ldots, X_n)$

- $max(X_i) \leq z \iff alla\ X_i \leq z$ (gäller i **båda** riktningar; händelserna har samma sannolikhet)

$$P[max(X_i) \leq z] = \mathrm{P}[(X_1 \leq z) \cap (X_2 \leq z) \cap \ldots \cap (X_n \leq z)]$$

$$F_Z(z) = P(Z \leq z) = P(X_1 \leq z) \cdot P(X_2 \leq z) \cdot \ldots \cdot P(X_n \leq z)$$

om alla X_i är parvist oberoende

$$= F_{X_1}(z) \cdot F_{X_2}(z) \cdot \ldots \cdot F_{X_n}(z)$$

definition för fördelningsfunktion F_X

$$\boxed{F_Z(z) = \prod_{i=1}^{n} F_{X_i}(z)}$$

fördelning för maximumet av flera oberoende slumpvariabler

$$F_Z(z) = F_X(z)^n$$

om alla X_i är oberoende och har samma fördelning

www.matstat.org

22

Maximumet av flera slumpvariabler

Exempel: livslängd hos ett elektroniskt instrument med två komponenter:
Ett elektroniskt instrument innehåller två komponenter vars livslängder X resp. Y är oberoende och exponentiellt fördelade:

$$F_X(x) = \begin{Bmatrix} 1 - e^{-a \cdot x} & x > 0 \\ 0 & x \leq 0 \end{Bmatrix} \qquad F_Y(y) = \begin{Bmatrix} 1 - e^{-b \cdot y} & y > 0 \\ 0 & y \leq 0 \end{Bmatrix}$$

Instrumentet fungerar så länge som minst en av båda komponenterna fungerar.
Livslängden hos instrumentet är alltså maximumet av livslängden för X och Y.

Söker: täthetsfunktion för $Z = max(X, Y)$

$$F_Z(z) = F_X(z) \cdot F_Y(z) = (1 - e^{-a \cdot z}) \cdot (1 - e^{-b \cdot z})$$

$$f_Z(z) = \frac{d}{dz} F_Z(z) = a \cdot e^{-a \cdot z} \cdot (1 - e^{-b \cdot z}) + b \cdot e^{-b \cdot z} \cdot (1 - e^{-a \cdot z})$$

för $z > 0$; 0 annars

24

Minimumet av flera slumpvariabler

$X_1, X_2, \dots, X_n$ **oberoende** slumpvariabler (med känd fördelning)

Söker: fördelning för $Z = min(X_1, X_2, \dots, X_n)$?

- $\min(X_i) > z \iff alla\ X_i > z$ (gäller i **båda** riktningar; händelserna har samma sannolikhet)

$$P[\min(X_i) > z] = P[(X_1 > z) \cap (X_2 > z) \cap \dots \cap (X_n > z)]$$

$$P(Z > z) = P(X_1 > z) \cdot P(X_2 > z) \cdot \dots \cdot P(X_n > z)$$ om alla X_i är parvist oberoende

$$1 - F_Z(z) = \left(1 - F_{X_1}(z)\right) \cdot \left(1 - F_{X_2}(z)\right) \cdot \dots \cdot \left(1 - F_{X_n}(z)\right)$$

kom ihåg:
$P(X_i > z) = 1 - F_{X_i}(z)$

$$\boxed{F_Z(z) = 1 - \prod_{i=1}^{n} [1 - F_{X_i}(z)]}$$

fördelning för minimumet av flera oberoende slumpvariabler

$$F_Z(z) = 1 - [1 - F_X(z)]^n$$

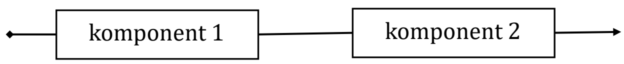
om alla X_i är oberoende och har samma fördelning

25

Minimumet av flera slumpvariabler

Exempel: livstid hos ett elektroniskt instrument, Blom et al. s. 93
Ett elektroniskt instrument innehåller två komponenter (X, Y) som är **seriekopplade**.
Livstiderna för båda komponenter är exponentiellt fördelade.
Söker: fördelningen för instrumentets livstid (Z)

$$F_X(x) = \begin{cases} 1 - e^{-a \cdot x} & x > 0 \\ 0 & x \le 0 \end{cases} \qquad F_Y(y) = \begin{cases} 1 - e^{-b \cdot y} & y > 0 \\ 0 & y \le 0 \end{cases}$$



Seriekopplung: intrumentets livstid = **minimumet** av komponenternas livstid!

$\boldsymbol{Z = min(X, Y)}$

$$F_Z(z) = 1 - [1 - F_X(z)] \cdot [1 - F_Y(z)] = 1 - [1 - (1 - e^{-a \cdot z})] \cdot [1 - (1 - e^{-b \cdot z})]$$

$$= 1 - e^{-a \cdot z} \cdot e^{-b \cdot z} = 1 - e^{-(a+b) \cdot z} \qquad z > 0$$

$$f_Z(z) = (a + b) \cdot e^{-(a+b) \cdot z} \qquad z > 0$$ täthetsfunktion, efter derivering

$$\boxed{Z \sim Exp(a + b)}$$ seriekoppling

www.matstat.org

26

Summa av två slumpvariabler

X, Y: **diskreta** s.v., $p_{X,Y}(i, j)$ antas vara känd (simultan sannolikhetsfunktion)

$Z = X + Y$ **summa**, slumpvariabel

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

summeras över alla i och j vars summa är k

$i + j = k \;\rightarrow\; j = k - i$

$$p_Z(k) = \sum_{i+j=k} p_{X,Y}(i, j) = \sum_{i=0}^{k} p_{X,Y}(i, k - i)$$

$$p_Z(k) = \sum_{i=0}^{k} p_X(i) \cdot p_Y(k - i)$$

... om X och Y är oberoende

Y\X	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

tärning, $k = 6$

www.matstat.org

27

Summa av två slumpvariabler

X, Y: **kontinuerliga** s.v., $f_{X,Y}(x,y)$ antas vara känd (simultan täthetsfunktion)

$Z = X + Y$ **summa**, slumpvariabel

$$F_Z(z) = P(Z \leq z) = P(X + Y \leq Z) = \iint\limits_{x+y \leq z} f_{X,Y}(x,y)\, dx\, dy$$

Täthetsfunktionen erhålls genom derivering.

integrationsområdet måste väljas noggrant

Om X och Y är oberoende erhåller man (härledning utelämnad, se Blom et al. s. 96):

$$f_Z(z) = \int_{-\infty}^{+\infty} f_X(x) \cdot f_Y(z - x)\, dx$$ faltningsformel

Anmärkning: differenser mellan 2 slumpvariabler, $Z = X - Y$, kan behandlas genom att formellt addera $-Y$.

www.matstat.org

28

Summa av slumpvariabler

Exempel: Låt bade X och Y vara **oberoende** och exponentiellt fördelade s.v. med samma parameter λ. Beräkna täthetsfunktionen för summan av X och Y !

$$f_X(x) = \begin{cases} \lambda \cdot e^{-\lambda \cdot x} & x > 0 \\ 0 & x \leq 0 \end{cases} \qquad f_Y(y) = \begin{cases} \lambda \cdot e^{-\lambda \cdot y} & y > 0 \\ 0 & y \leq 0 \end{cases}$$ $X, Y \sim Exp(\lambda)$

$$f_Z(z) = \int_{-\infty}^{+\infty} f_X(x) \cdot f_Y(z - x)\, dx = \int_{-\infty}^{+\infty} \lambda \cdot e^{-\lambda \cdot x} \cdot \lambda \cdot e^{-\lambda \cdot (z-x)}\, dx$$

OBS!: $z - x > 0$ erfordras!

$$= \int_0^z \lambda^2 \cdot e^{-\lambda \cdot z}\, dx = \lambda^2 \cdot e^{-\lambda \cdot z} \cdot \int_0^z 1\, dx = \lambda^2 \cdot e^{-\lambda \cdot z} \cdot z$$

$$\boxed{f_Z(z) = \lambda^2 \cdot e^{-\lambda \cdot z} \cdot z}$$

täthetsfunktion för summan av två **oberoende**, exponentiellt fördelade s.v. med samma λ $(z > 0)$.

Generalisering till **fler än två summander** $(z > 0)$:

$$\boxed{f_Z(z) = \frac{\lambda^n}{(n-1)!} \cdot z^{n-1} \cdot e^{-\lambda \cdot z}}$$

täthetsfunktion för summan av n **oberoende**,, exponentiellt fördelade s.v. (gammafördelning med $c = n$)

www.matstat.org

29