Tõenäosusteooria ja statistika II

Loeng 2

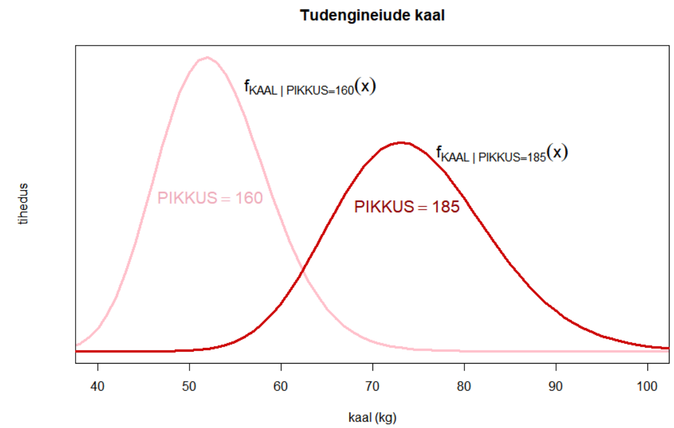
6. september 2021

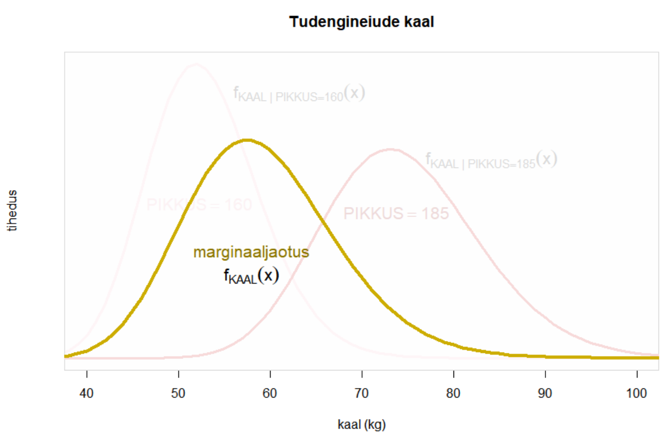
Ühisjaotus; tinglik jaotus; marginaaljaotus

Märt Möls

mart.mols@ut.ee

Kursuse koduleht:
www-1.ms.ut.ee/mart/TS2





Diskreetsete juhuslike suuruste ühisjaotus

Juhuslikku vektorit (*X,Y*) nimetatakse diskreetseks kui nii *X* kui ka *Y* on diskreetsed juhuslikud suurused.

Kui diskreetse juhusliku vektori (*X,Y*) puhul teame mistahes *X* ja *Y* väärtuste kombinatsiooni esinemistõenäosust siis ütleme, et me teame vektori (*X,Y*) jaotust ehk teame juhuslike suuruste *X* ja *Y* ühisjaotust.

Diskreetsete juhuslike suuruste ühisjaotus

Definitsioon 1 *Juhusliku vektorit* (X,Y) *nimetatakse* diskreetseks, *kui* X *ja* Y *on diskreetsed juhuslikud suurused. Diskreetse juhusliku suuruse korral nimetatakse kolmikuid* (x_i,y_j,p_{ij}), $i \in I$, $j \in J$, *kus* $p_{ij} = P(\{X = x_i, Y = y_j\})$ *ning* $\{x_i : i \in I\}$ *ja* $\{y_j : j \in J\}$ *on vastavalt juhuslike suuruste* X *ja* Y *väärtuste hulgad, juhusliku vektori* (X,Y) *jaotuseks ehk juhuslike suuruste* X *ja* Y *ühisjaotuseks.*

Märkus. Lihtsuse mõttes on eelnev definitsioon toodud kahest juhuslikust suurusest koosneva vektori kohta, kuid üldistus n-mõõtmelisele juhule on analoogiline.

Näide. Liiklusõnnetustesse (USA, 1997-2006) sattunud sõidukijuhid ja nende saatus

*X**Y*	1 (autos turvavööga)	2 (auto, turvavööta)	3 (mootorratas)
1 (=suri)	0,0053	0,0071	0,0017
0 (=ei surnud)	0,8021	0,1837	0,0001
Y-i marginaaljaotus	0,8074	0,1908	0,0018

Täistõenäosuse valem:
$P(Y=y_1) = P(Y=y_1 \cap X=x_1) + P(Y=y_1 \cap X=x_2)$

Milline on *X*-tunnuse marginaaljaotus?

Diskreetne juhuslik vektor

Juhusliku vektori (X,Y) käsitlemisel nimetatakse juhuslike suuruste X ja Y jaotusi *marginaaljaotusteks.*

Lemma 1 *Juhusliku vektori* (X,Y) *jaotuse* $\{(x_i,y_j,p_{ij}) : i \in I, j \in J\}$ *korral kehtivad võrdused*

$$p_{.j} := \sum_{i \in I} p_{ij} = P(\{Y = y_j\})$$
$$p_{i.} := \sum_{j \in J} p_{ij} = P(\{X = x_i\})$$
$$\sum_{i \in I,\, j \in J} p_{ij} = 1.$$

Näide. Liiklusõnnetustesse (USA, 1997-2006) sattunud sõidukijuhid ja nende saatus

*X**Y*	1 (autos turvavööga)	2 (auto, turvavööta)	3 (mootorratas)
1 (=suri)	0,0053	0,0071	0,0017
0 (=ei surnud)	0,8021	0,1837	0,0001
Y-i marginaaljaotus	0,8074	0,1908	0,0018

$$P(A|B) = \frac{P(A \bigcap B)}{P(B)}$$

P(*X*=1|*Y*=1) = 0,0053/0,8074 = 0,0065…

P(*X*=0|*Y*=1) = 0,8021/0,8074 = 0,9934…

Milline on juhusliku suuruse *X* tinglik jaotus juhul kui *Y*=1?

Näide. Liiklusõnnetustesse (USA, 1997-2006) sattunud sõidukijuhid ja nende saatus

$X \backslash Y$	1 (autos turvavööga)		2 (auto, turvavööta)	3 (mootorratas)
1 (=suri)	0,0053	0,0066	0,0071	0,0017
0 (=ei surnud)	0,8021	0,9934	0,1837	0,0001
Y-i marginaaljaotus	0,8074		0,1908	0,0018

juhusliku suuruse *X* tinglik jaotus juhul kui *Y*=1

$$P(A|B) = \frac{P(A \bigcap B)}{P(B)}$$

Üldjuhul:

Tinglik tõenäosus näha väärtust x_i juhul kui $Y=y_j$ on leitav valemiga $p_{x_i|y_j} = p_{ij}/p_{i.}$

Milline on juhusliku suuruse *X* tinglik jaotus juhul kui *Y*=1?

Näide. Liiklusõnnetustesse (USA, 1997-2006) sattunud sõidukijuhid ja nende saatus

$X \backslash Y$	1 (autos turvavööga)		2 (auto, turvavööta)		3 (mootorratas)
1 (=suri)	0,0053	0,0066	0,0071	0,0372	0,0017
0 (=ei surnud)	0,8021	0,9934	0,1837	0,9628	0,0001
Y-i marginaaljaotus	0,8074		0,1908		0,0018

Kui sattud liiklusõnnetusse ja sa ei kanna turvavööd, siis on sinu surmarisk enam kui 5 korda suurem võrreldes juhiga, kes kannab turvavööd

juhusliku suuruse *X* tinglik jaotus juhul kui *Y*=2

Milline on juhusliku suuruse *X* tinglik jaotus juhul kui *Y*=2? Leia!

Definitsioon

Diskreetseid juhuslikke suuruseid X ja Y nimetatakse sõltumatuteks, kui iga $i \in I$ ja iga $j \in J$ korral kehtib võrdus

$$p_{ij} = p_{i.}p_{.j}$$

Kuna statistika kasutamisel on tegemist tavaliselt rohkem kui kahe sõltumatu juhusliku suurusega, siis sõnastame sõltumatuse mõiste ka üldkujul.

Definitsioon

Diskreetseid juhuslikke suuruseid $X_1, X_2, \ldots, X_n$ nimetatakse sõltumatuteks, kui sündmused $\{X_1 = x_1\}, \ldots, \{X_n = x_n\}$ on täielikult sõltumatud iga X_1 võimaliku väärtuse x_1, X_2 võimaliku väärtuse x_2, ..., iga X_n võimaliku väärtuse x_n korral.

Näide. Liiklusõnnetustesse (USA, 1997-2006) sattunud sõidukijuhid ja nende saatus

$X \backslash Y$	1 (autos turvavööga)	2 (auto, turvavööta)	3 (mootorratas)	
1 (=suri)	0,0053	0,0071	0,0017	0,0141
0 (=ei surnud)	0,8021	0,1837	0,0001	0,9859
Y-i marginaaljaotus	0,8074	0,1908	0,0018	

juhusliku suuruse *X* marginaaljaotus

Kui tinglikud jaotused on samad, siis on juhuslikud suurused sõltumatud:

$$P(X = x_i | Y = y_j) = c \;\; \forall j, i$$

$$c = c \sum_j P(Y = y_j) = \sum_j cP(Y = y_j) = \sum_j P(X = x_i | Y = y_j) P(Y = y_j) = P(X = x_i)$$

$$P(X = x_i, Y = y_j) = P(X = x_i | Y = y_j) P(Y = y_j) = cP(Y = y_j) = P(X = x_i) P(Y = y_j)$$

Näide 2

Münti visatakse kaks korda.

Kas juhuslikud suurused

X: 1 – mõlema viske tulemused on samad
0 – mõlema viske tulemused on erinevad

Y: 1 – esimesel viskel tuleb KIRI
0 – esimesel viskel tuleb KULL

on sõltumatud?

X: 1 – mõlema viske tulemused on samad
0 – mõlema viske tulemused on erinevad
Y: 1 – esimesel viskel tuleb KIRI
0 – esimesel viskel tuleb KULL

Näide 2

katseseeria	p_{ij}	X	Y	$p_{i.} \cdot p_{.j}$
KULL KULL	1/4	1	0	1/4
KULL KIRI	1/4	0	0	1/4
KIRI KULL	1/4	0	1	1/4
KIRI KIRI	1/4	1	1	1/4

X ja *Y* on sõltumatud!

X-i marginaaljaotus

	0	1
$P(X=x_i)$	1/2	1/2

Y-i marginaaljaotus

	0	1
$P(Y=y_j)$	1/2	1/2

Näide 2a

Münti visatakse kaks korda.

Kas juhuslikud suurused

X: 1 – mõlema viske tulemused on samad
0 – mõlema viske tulemused on erinevad

Y: 1 – esimesel viskel tuleb KIRI
0 – esimesel viskel tuleb KULL

Z: 1 – teisel viskel tuleb KIRI
0 – teisel viskel tuleb KULL

on sõltumatud?

X: 1 – mõlema viske tulemused on samad
0 – mõlema viske tulemused on erinevad
Y: 1 – esimesel viskel tuleb KIRI
0 – esimesel viskel tuleb KULL
Z: 1 – teisel viskel tuleb KIRI
0 – teisel viskel tuleb KULL

Näide 2a

katseseeria	p_{ijk}	X	Y	Z	$p_{i..} \cdot p_{.j.} \cdot p_{..k}$
KULL KULL	1/4	1	0	0	1/8
KULL KIRI	1/4	0	0	1	1/8
KIRI KULL	1/4	0	1	0	1/8
KIRI KIRI	1/4	1	1	1	1/8

X-i marginaaljaotus

	0	1
$P(X=x_i)$	1/2	1/2

Y-i marginaaljaotus

	0	1
$P(Y=y_j)$	1/2	1/2

Z-i marginaaljaotus

	0	1
$P(Z=z_k)$	1/2	1/2

Kõik juhuslikud suurused pole diskreetsed…

Mida teha teistsuguste juhuslike suuruste korral?

Nii diskreetse kui ka pideva juhusliku vektori jaotust saab kirjeldada jaotusfunktsioonide abil. Tuletame meelde, et kahemõõtmelisel juhul oli jaotusfunktsioon defineeritud järgmiselt.

Definitsioon

Juhusliku vektori (X, Y) jaotusfunktsiooniks (ehk juhuslike suuruste X ja Y ühisjaotuse jaotusfunktsiooniks) nimetatakse funktsiooni

$$F_{X,Y}(x, y) = P(\{X \leq x,\ Y \leq y\}), \quad x, y \in \mathbb{R}.$$

Diskreetse juhusliku suuruse korral

$$F_{X,Y}(x,y) = \sum_{\{(i,j):\, x_i \leq x, y_j \leq y\}} p_{ij}$$

Nii diskreetse kui ka pideva juhusliku vektori jaotust saab kirjeldada jaotusfunktsioonide abil. Tuletame meelde, et kahemõõtmelisel juhul oli jaotusfunktsioon defineeritud järgmiselt.

Definitsioon

Juhusliku vektori (X, Y) jaotusfunktsiooniks (ehk juhuslike suuruste X ja Y ühisjaotuse jaotusfunktsiooniks) nimetatakse funktsiooni

$$F_{X,Y}(x, y) = P(\{X \leq x,\ Y \leq y\}), \quad x, y \in \mathbb{R}.$$

$$F(x,y) = 1/16(x^2y + xy^2), \;\; 0 \leq x,y \leq 2$$

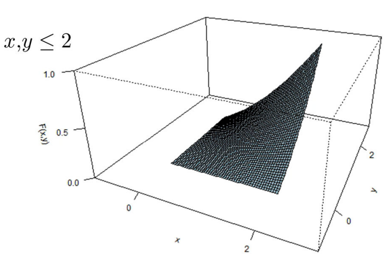
Nii diskreetse kui ka pideva juhusliku vektori jaotust saab kirjeldada jaotusfunktsioonide abil. Tuletame meelde, et kahemõõtmelisel juhul oli jaotusfunktsioon defineeritud järgmiselt.

Definitsioon

Juhusliku vektori (X, Y) jaotusfunktsiooniks (ehk juhuslike suuruste X ja Y ühisjaotuse jaotusfunktsiooniks) nimetatakse funktsiooni

$$F_{X,Y}(x, y) = P(\{X \leq x,\ Y \leq y\}), \quad x, y \in \mathbb{R}.$$

$$F(x,y) = 1/16(x^2y + xy^2), \;\; 0 \leq x,y \leq 2$$



Juhusliku vektori jaotusfunktsiooni omadused

1. $0 \le F_{X,Y}(x,y) \le 1 \ \forall (x,y) \in \mathbb{R}^2$,
2. $F_{X,Y}$ on kummagi muutuja järgi paremalt pidev igas punktis,
3. $\lim_{y\to\infty} F_{X,Y}(x,y) = F_X(x) \ \forall x \in \mathbb{R}$, $\lim_{x\to\infty} F_{X,Y}(x,y) = F_Y(y) \ \forall y \in \mathbb{R}$,
4. $\lim_{y\to-\infty} F_{X,Y}(x,y) = 0 \ \forall x \in \mathbb{R}$, $\lim_{x\to-\infty} F_{X,Y}(x,y) = 0 \ \forall y \in \mathbb{R}$.
5. $P(\{a < X \le b, c < Y \le d\}) = F_{X,Y}(b,d) - F_{X,Y}(a,d) - F_{X,Y}(b,c) + F_{X,Y}(a,c)$,
6. $F_{X,Y}(x,y) = F_{X|Y\le y}(x) F_Y(y)$.

$$\begin{aligned} F_{X,Y}(x,y) &= P(X \le x \cap Y \le y) \\ &= P(X \le x | Y \le y)\, P(Y \le y) \\ &= F_{X|Y\le y}(x) F_Y(y) \end{aligned}$$

Sõltumatud (diskreetsed) juhuslikud suurused

$$F_{X,Y}(x,y) = \sum_{\{(i,j):\, x_i \le x, y_j \le y\}} p_{ij}$$

Kui *X* ja *Y* on sõltumatud

$$F_{X,Y}(x,y) = \sum_{\{(i,j):\, x_i \le x, y_j \le y\}} p_{i.} p_{.j}$$

$$F_{X,Y}(x,y) = \sum_{\{i:\, x_i \le x\}} \sum_{\{j:\, y_j \le y\}} p_{i.} p_{.j}$$

$$F_Y(y) = \sum_{\{j:\, y_j \le y\}} p_{.j}$$

$$F_{X,Y}(x,y) = \sum_{\{i:\, x_i \le x\}} \left(p_{i.} \sum_{\{j:\, y_j \le y\}} p_{.j} \right)$$

$$F_{X,Y}(x,y) = \sum_{\{i:\, x_i \le x\}} p_{i.} F_Y(y)$$

$$F_X(x) = \sum_{\{i:\, x_i \le x\}} p_{i.}$$

$$F_{X,Y}(x,y) = F_Y(y) \sum_{\{i:\, x_i \le x\}} p_{i.} = F_Y(y) F_X(x)$$

Diskreetsed juhuslikud suurused

$$F_{X,Y}(x,y) = F_Y(y) F_X(x) \ \Rightarrow \ X \text{ ja } Y \text{ on sõltumatud?}$$

Olgu võimalikud väärtused järjestatud

$x_1 < x_2 < \ldots \qquad y_1 < y_2 < \ldots$

$$\begin{aligned} P(X = x_i, Y = y_j) &= P(\{x_{i-1} < X \le x_i, y_{j-1} < Y \le y_j\}) \\ &= F_{X,Y}(x_i,y_j) - F_{X,Y}(x_{i-1},y_j) - F_{X,Y}(x_i,y_{j-1}) + F_{X,Y}(x_{i-1},y_{j-1}) \\ &= F_X(x_i)F_Y(y_j) - F_X(x_{i-1})F_Y(y_j) - F_X(x_i)F_Y(y_{j-1}) + F_X(x_{i-1})F_Y(y_{j-1}) \\ &= F_Y(y_j)\,[F_X(x_i) - F_X(x_{i-1})] - F_Y(y_{j-1})\,[F_X(x_i) - F_X(x_{i-1})] \\ &= F_Y(y_j) P(X = x_i) - F_Y(y_{j-1}) P(X = x_i) \\ &= [F_Y(y_j) - F_Y(y_{j-1})]\, P(X = x_i) \\ &= P(Y = y_j)\, P(X = x_i) \end{aligned}$$

Üldine sõltumatuse definitsioon

Definitsioon

Juhuslikke suurusi X ja Y nimetatakse sõltumatuteks, kui iga $x, y \in \mathbb{R}$ korral on sündmused $\{X \le x\}$ ja $\{Y \le y\}$ sõltumatud.

Järeldus

Juhuslikud suurused on sõltumatud parajasti siis, kui nende ühisjaotuse jaotusfunktsioon avaldub kujul

$$F_{X,Y}(x,y) = F_X(x) F_Y(y) \ \forall x, y \in \mathbb{R},$$

kus F_X ja F_Y on vastavalt juhuslike suuruste X ja Y jaotusfunktsioonid.

Definitsioon

Juhuslikku vektorit (X, Y) nimetatakse pidevaks, kui tema jaotusfunktsioon avaldub kujul

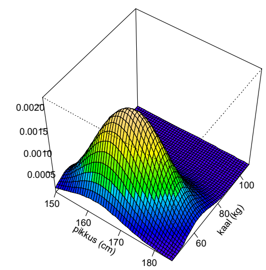
$$F_{X,Y}(x, y) = \int_{-\infty}^{x} \left(\int_{-\infty}^{y} f_{X,Y}(u, v)\, dv \right) du, \quad x, y \in \mathbb{R}$$

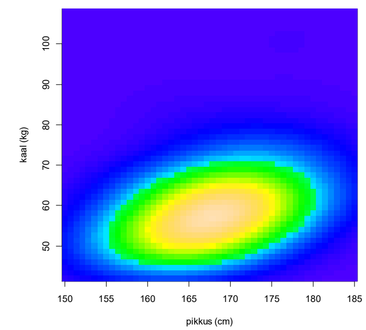
mingi funktsiooni $f_{X,Y} : \mathbb{R}^2 \to \mathbb{R}$ korral. Funktsiooni $f_{X,Y}$ nimetatakse sel juhul juhusliku vektori (X, Y) tihedusfunktsiooniks (ehk juhuslike suuruste X ja Y ühistiheduseks).

Kui $F_{X,Y}(x,y)$ on diferentseeruv punktis (x,y), siis

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

Kahemõõtmelise tihedusfunktsiooni näide – naistudengite pikkused ja kaalud





Lemma (Tihedusfunktsiooni omadused)

Olgu (X, Y) pidev juhuslik vektor jaotusfunktsiooniga $F_{X,Y}$ ja tihedusfunktsiooniga $f_{X,Y}$. Siis kehtivad järgmised omadused:

1. *Funktsioon $f_{X,Y}$ on mittenegatiivne, st $f_{X,Y}(x, y) \geq 0 \ \forall (x, y) \in \mathbb{R}^2$;*
2. *kehtivad võrdused*

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

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

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)\, dx\, dy = 1$$

$$f_{X,Y}(x,y) = f_{X|Y=y}(x) f_Y(y)$$

$$f_{X|Y=y}(x) = \frac{f_{X,Y}(x,y)}{f_Y(y)}, \text{ kui } f_Y(y) \neq 0$$

Kui $D \subset \mathbb{R}^2$ on esitatav loenduva arvu ristkülikute abil kasutades ühendeid, ühisosasid ja täiendeid (st Boreli σ-algebra suhtes mõõtuv hulk), siis

$$P(\{(X, Y) \in D\}) = \iint_D f_{X,Y}(x, y)\, dx\, dy.$$

Sõltumatus ja tihedusfunktsioon

Lemma (X ja Y sõltumatus tihedusfunktsioonide kaudu)

Pidevad juhuslikud suurused X ja Y on sõltumatud parajasti siis, kui nende ühisjaotuse tihedusfunktsioon avaldub kujul

$$f_{X,Y}(x,y) = f_X(x)f_Y(y)\ \forall x,y \in \mathbb{R}.$$

Kui X ja Y on sõltumatud, siis

$$F_{X,Y}(x,y) = F_Y(y)F_X(x)$$

$$f_{X,Y}(x,y) = \frac{\partial^2 F_{X,Y}(x,y)}{\partial x \partial y} = \frac{\partial}{\partial x}\left(\frac{\partial F_X(x)F_Y(y)}{\partial y}\right)$$

$$= \frac{\partial}{\partial x}\left(F_X(x)\frac{\partial F_Y(y)}{\partial y}\right) = f_X(x)f_Y(y)$$

Sõltumatus ja tihedusfunktsioon

Lemma (X ja Y sõltumatus tihedusfunktsioonide kaudu)

Pidevad juhuslikud suurused X ja Y on sõltumatud parajasti siis, kui nende ühisjaotuse tihedusfunktsioon avaldub kujul

$$f_{X,Y}(x,y) = f_X(x)f_Y(y)\ \forall x,y \in \mathbb{R}.$$

Kui $f_{X,Y}(x,y) = f_X(x)f_Y(y)$

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

$$= \int_{-\infty}^{x}\int_{-\infty}^{y} f_X(u)f_Y(v)\,dv\,du = \int_{-\infty}^{x} f_X(u)\int_{-\infty}^{y} f_Y(v)\,dv\,du$$

$$= \int_{-\infty}^{x} f_X(u)F_Y(y)\,du = F_Y(y)\int_{-\infty}^{x} f_X(u)\,du = F_Y(y)F_X(x)$$

Näide

Kahemõõtmelise pideva juhusliku suuruse tihedusfunktsioon on järgmine:

$$f_{X,Y}(x,y) = \begin{cases} (x+xy)/2, & 0 \le x \le 1, 0 \le y \le 2, \\ 0, & \text{muidu.} \end{cases}$$

Leia tõenäosus P(X≥Y)!

Näide

Kahemõõtmelise pideva juhusliku suuruse tihedusfunktsioon on järgmine:

$$f_{X,Y}(x,y) = \begin{cases} (x+xy)/2, & 0 \le x \le 1, 0 \le y \le 2, \\ 0, & \text{muidu.} \end{cases}$$

Leia tõenäosus P(X≥Y)!

Kas on tegemist tihedusfunktsiooniga?

Näide

Kahemõõtmelise pideva juhusliku suuruse tihedusfunktsioon on järgmine:

$$f_{X,Y}(x,y) = \begin{cases} (x+xy)/2, & 0 \le x \le 1, 0 \le y \le 2, \\ 0, & \text{muidu.} \end{cases}$$

Leia tõenäosus P(*X*≥*Y*)!

$$P(\{(X,Y) \in D\}) = \iint\limits_D f_{X,Y}(x,y)\,dx\,dy$$

$$P(X \ge Y) = \int\limits_0^1 \int\limits_0^x f_{X,Y}(x,y)\,dy\,dx$$

$$\begin{aligned} P(X \ge Y) &= \int\limits_0^1 \int\limits_0^x f_{X,Y}(x,y)\,dy\,dx \\ &= \int\limits_0^1 \int\limits_0^x (x+xy)/2\,dy\,dx \\ &= \int\limits_0^1 \left(xy + \frac{xy^2}{2}\right)/2 \Bigg|_{y=0}^{y=x} dx \\ &= \int\limits_0^1 (x^2 + x^3/2)/2\,dx \\ &= (x^3/3 + x^4/8)/2 \Bigg|_0^1 = (1/3+1/8)/2 = 11/48 \end{aligned}$$