Johdatus todennäköisyyslaskentaan
Kevät 2014

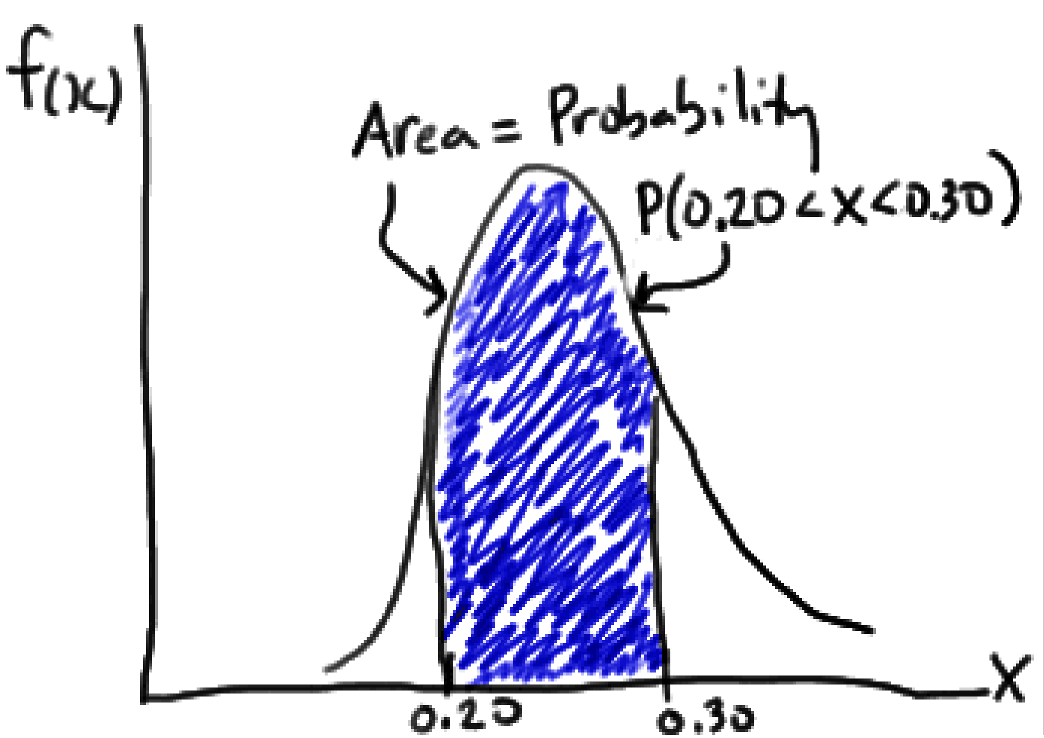
Luento 10 / 13

Jukka Kohonen
Matematiikan ja tilastotieteen laitos
Helsingin yliopisto

Harjoitus 5: Huomioita

- Tehtävä 2 (lentokone):
  - Yksittäinen järjestelmä rikkoutuu noin tn:llä 1 / 1000
  - Kaikki järjestelmät rikkoutuvat noin tn:llä 1 / 100 000
  - Selvästikään eivät riippumattomat. Riippumattomuuteen perustuva tn-arvio olisi pielessä kertoimella 10000

- Tehtävä 15 (Tuominen 2:45)
  - Muistuttaa harj. 4 nopanheittotehtävää
  - Pitää pystyä yhdistämään toisiinsa kolme asiaa: kertymäfunktio, tiheysfunktio, odotusarvo



JATKUVA SATUNNAISMUUTTUJA

Tiheysfunktio

- Jos X:n tiheysfunktio on f, niin f(x) **ilmaisee verrannollisuuskertoimen**: tn osua (lyhyelle) välille (x:n lähellä) on välin pituus kertaa f(x).

- Huom: f(x) **ei ole todennäköisyys**, että X=x
- Esim. ennuste lämpötilasta, joka tasajakautunut välillä (20, 30)
- Koska f(x) ei ole todennäköisyys, se voi olla suurempikin kuin 1. (Mitä tämä tarkoittaa?)

Kertymäfunktio

- Vaikka tiheysfunktio antaa **intuitiivisemman** käsityksen jakaumasta (mihin jakauma keskittyy), kertymäfunktio on keskeinen **laskennallinen** työkalu jakaumien käsittelyssä.
- F(x) vastaa kysymykseen
”mikä on tn, että $X \leq x$”.
- Esim. lämpötilalle F(25) = ½ tarkoittaa, että on ½ todennäköisyys, että lämpötila on enintään 25 astetta.

- Kertymäfunktion on pakko olla kasvava – miksi?
- Kertymäfunktio kasvaa arvojoukossa vasemmalta 0:sta oikealle 1:een – miksi?
- Kertymäfunktio on yleensä tf:n integraali, ja tf on kertymäfunktion derivaatta – miksi?

Epätasaisia jakaumia

- Eksponenttijakauma (Tuominen s. 59) muistuttaa geometrista jakaumaa, mutta onkin jatkuva

- Ts. ”yrityksiä” on jatkuvasti, ei vain kokonaislukujen kohdalla. Esim.
  - Alkeishiukkanen voi hajota milloin tahansa, mikä on hiukkasen elinikä?
  - Laite voi hajota milloin tahansa, miten kauan se toimii? (”Satunnainen” hajoaminen, ei kuluminen)
  - Tuulilasiin voi osua kivi milloin tahansa, miten kauan voidaan ajaa ennen kuin ensimmäinen kivi osuu?

- Huom: mediaani ei ole odotusarvon kohdalla (lasketaan kertymäfunktion avulla!)

  (mediaani = se piste, jonka molemmin puolin on puolet todennäköisyydestä, tarkemmin Tuominen s. 89)

MUUNNOKSEN $Y=g(X)$ JAKAUMA

Sovellusesimerkkejä

- Pitkulainen esine (pituus = L) nähdään kaukaa satunnaisessa kulmassa (θ).
  - Nähty projektio = $X = L \cdot \cos\theta$
  - Projektion jakauma: $X \sim ?$ (Harjoitustehtävät 4:9 ja 4:10)
  - Pitää siis osata ratkaista θ:n muunnoksen jakauma.

- **Käänteisongelma:** Sama pitkulainen esine (pituus <u>tuntematon</u>) nähdään monta kertaa satunnaisissa kulmissa.
  - Tunnetaan nähtyjen projektioiden jakauma
  - Mikä on pituuden jakauma?
  - Sovelletaan esim. Bayesin kaavaa.

- Epäsäännöllisen muotoinen asteroidi nähdään satunnaisissa kulmissa.
  - Tunnetaan nähtyjen (asteroidin heijastamien) kokonaiskirkkauksien jakauma ja/tai kehitys ajan kuluessa (kun asteroidi pyörii ja liikkuu)
  - Mikä on asteroidin muoto?

Erilaisia muunnostehtäviä

1. Tunnetaan **kertymäfunktio** F_X ja muunnosfunktio $Y = g(X)$.
   Mikä on Y:n kertymäfunktio?
   → Periaatteessa **helppo** tehtävä: Ratkaistaan $P(Y \leq y)$

2. Tunnetaan **tiheysfunktio** f_X ja muunosfunktio $Y = g(X)$.
   Mikä on Y:n tiheysfunktio?
   → **Ei aivan triviaalia**.
   On olemassa tiheysfunktion muuntokaava (ei käsitellä tällä kurssilla).

3. Osataan **simuloida** $X_1 \ldots X_n$ iid samasta jakaumasta kuin X,
   ja tunnetaan muunnosfunktio $Y = g(X)$.
   Miten simuloidaan Y:n jakaumasta?
   → **Todella helppoa**: lasketaan $Y_i = g(X_i)$

Kokeilemme muutamaa yksinkertaista muunnosta menetelmällä 3.

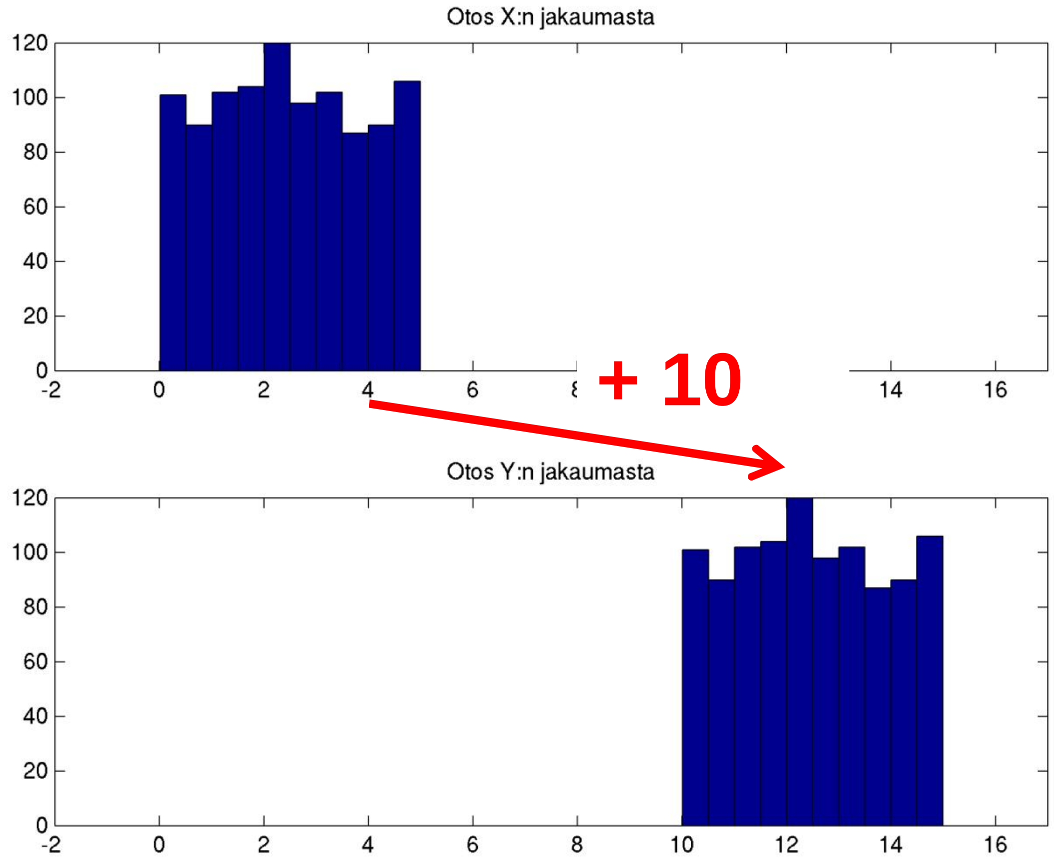
Muunnos: vakion lisääminen

X ~ Tas(0, 5)
Y = X **+ 10**

Y ~ ?

Ilmeisesti Y:in on tasajakautunut.

Millä parametreilla?



Muunnos: vakiolla kertominen

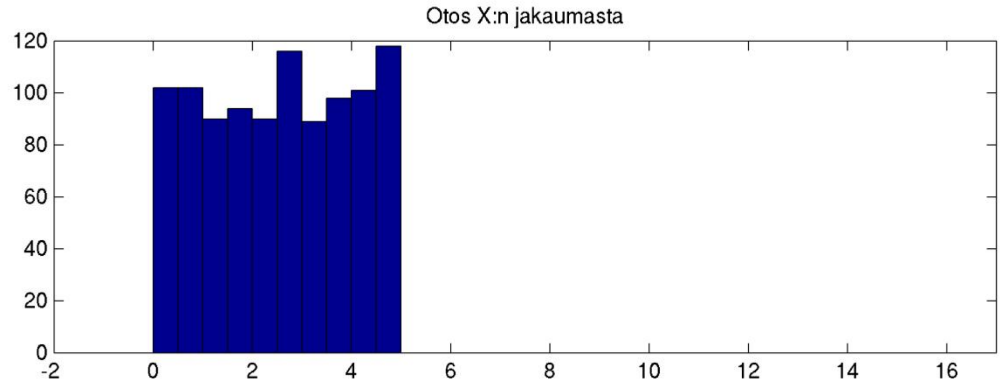
X ~ Tas(0, 5)
Y = **3 ·** X

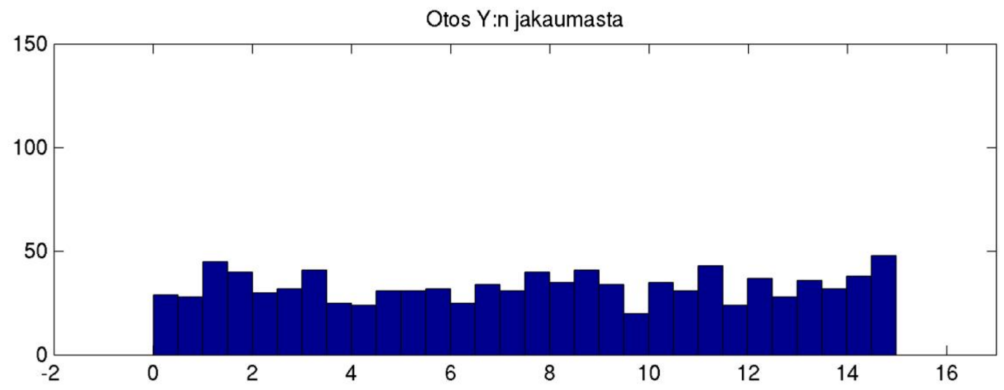
Y ~ ?

Ilmeisesti Y:in on tasajakautunut.

Millä parametreilla?

Huom. litistyminen, joka näkyy myös Y:n tiheysfunktiossa.





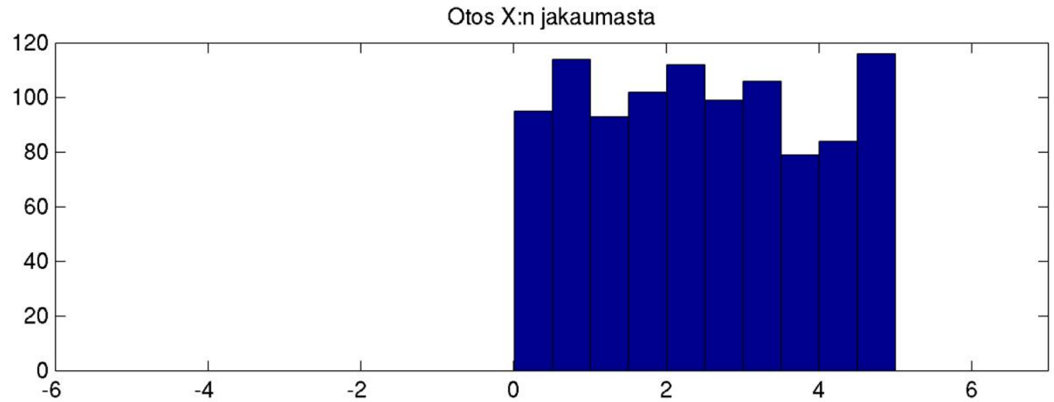
Muunnos: logaritmi

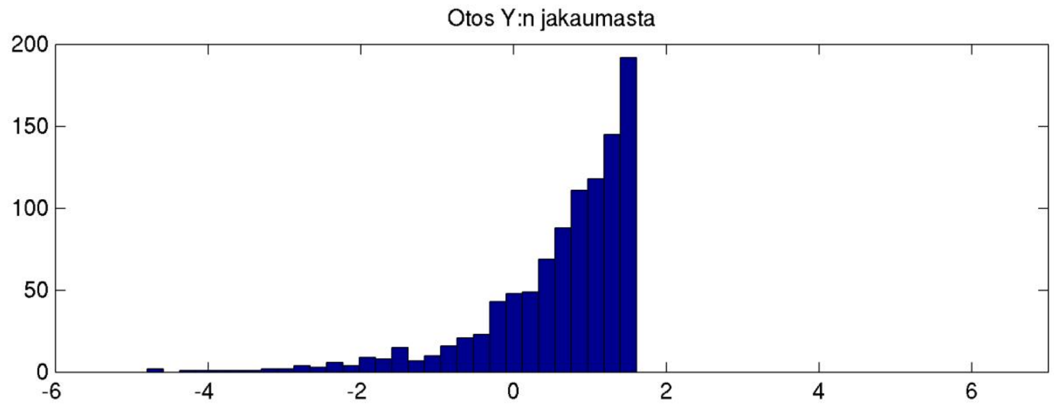
X ~ Tas(0, 5)
Y = **ln** X

Y ~ ?

Y:llä on ilmeisesti joku erikoisempi jakauma.

Osataan kyllä laskea (Tuominen s. 74)





Kertymäfunktion muuntaminen

Tuominen s. 74

$X \sim \mathrm{Tas}(0, 5)$
$Y = \ln X$
$Y \sim$?

Ratkaistaan Y:n kertymäfunktio.

$F_Y(a)$	$= P(Y$	$\le a)$	Kertymäfunktion määritelmä.
	$= P(\ln X$	$\le a)$	Koska $Y = \ln X$.
	$= P(X$	$\le \exp(a))$	Eksponenttifunktio aidosti kasvava.
	$= F_X(\exp(a))$		Kertymäfunktion määritelmä
	$= \exp(a) / 5$		Tasajakauman kertymäfunktio.

Derivoimalla saadaan Y:n tiheysfunktio.
$f_Y(a) \quad = (1/5) \exp(a)$

Huom. että Y:n jakauma sijoittuu välille $(-\infty, \ln 5)$, sillä sinne väli $(0, 5)$ kuvautuu logaritmimuunnoksessa!
Tulos tuntuu vastaavan otoshistogrammin muotoa.

SUMMAN $Y = X_1 + \ldots + X_n$ JAKAUMA

2.3 Satunnaismuuttujien summa

Herra K:n bussimatka koostuu osista

– bussin odotus	Y	$\sim$ Tas(0, 4)	$E(Y)=2$ min
– ajoaika	Z	$\sim$ Tas(15, 25)	$E(Z)=20$ min

Koko matka-aika **M** = Y+Z

– vähintään	0 + 15	= 15 min
– enintään	4 + 25	= 29 min
– odotusarvo E(M)	= E(Y) + E(Z)	= 22 min

Mikä on M:n jakauma välillä (15, 29)?

- Odotusarvo ei kerro kaikkea jakaumasta.
  Jakauma voi esim. olla hyvin keskittynyt odotusarvon lähelle, tai sitten ei

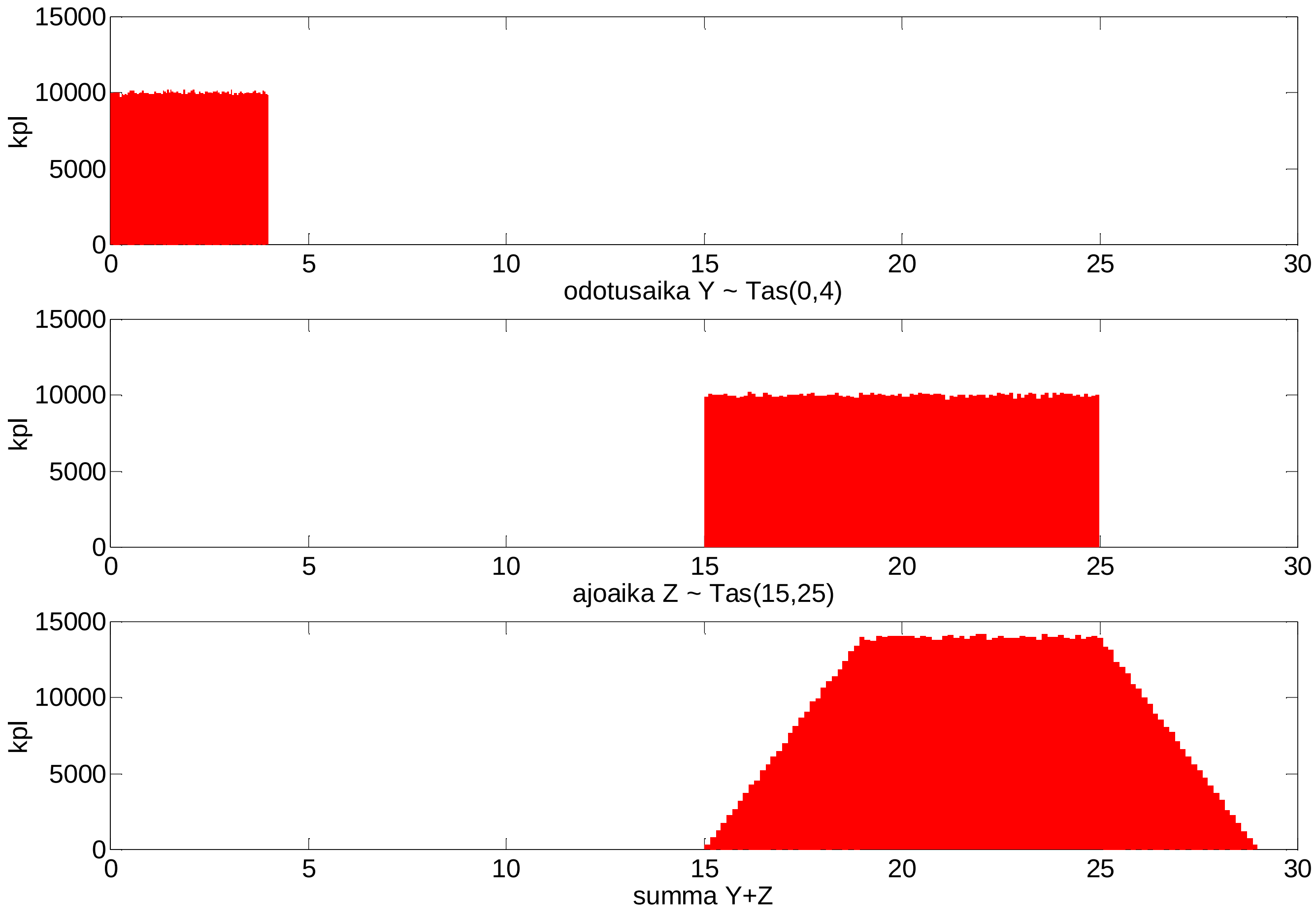
Empiirinen kokeilu

Bussin odotusaika Tas(0,4) ja ajoaika Tas(15,25). Miten summa on jakautunut?

Arvotaan miljoona kertaa Y~Tas(0,4) ja Z~Tas(15,25) ja lasketaan summat.

Empiirinen histogrammi antaa tällöin hyvän käsityksen tiheysfunktion muodosta.

Esim. Matlabissa Y = unifrnd(0,4,1,1e6); Z=unifrnd(15,25,1,1e6); M=Y+Z; hist(M)



Y ~ Tas(0,4)

Z ~ Tas(15,25)

M = Y+Z:
ei tasajakauma

(Miten poikkeaa?
Miksi?)

Jos halutaan laskea tarkasti

- Summan M=X+Y jakauma voidaan laskea ns. **konvoluutiolla** (Tuominen 71-72)
- Voidaan myös **järkeillä geometrisesti**, miten X ja Y voivat sijoittua: ajatellaan, että piste (X,Y) on tasajakautunut suorakulmiossa $(0, 4) \times (15, 25)$
- Tähän tarvitaan 2-ulotteisen jakauman käsite. Vrt. Tuominen s. 131 esimerkki 5.3.5: Tasainen jakauma alueessa *A*

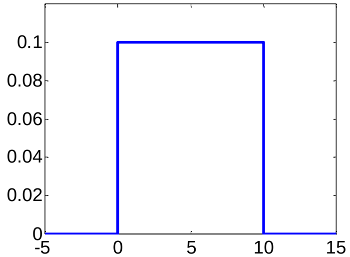
Jos halutaan approksimoida

- Summan voidaan **arvioida** olevan **suunnilleen normaalijakautunut**,
  ja lasketaan sen parametrit μ ja σ
- Tähän tarvitaan
  - normaalijakauman käsite
  - varianssin käsite
- jotka tulevat seuraavilla luennoilla.

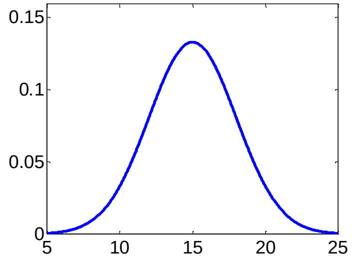
JATKUVIEN JAKAUMIEN KÄSITTELYÄ

Tiheys- ja kertymäfunktio; E(X)

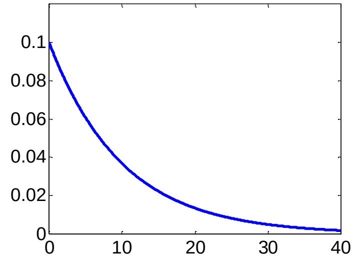
- Käydään läpi muutamia jatkuvia jakaumia



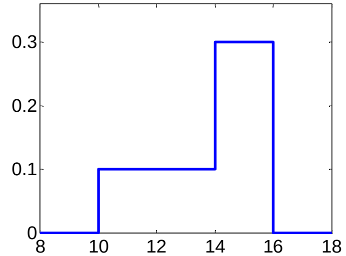
Tas(*a*, *b*)



N(μ, σ^2)



Exp(λ)

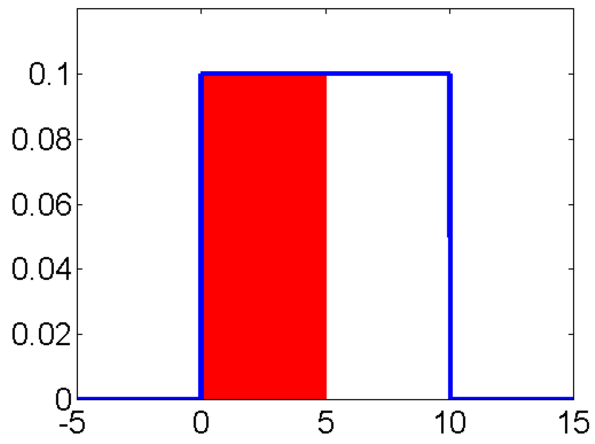


Mikä tahansa jakauma, jonka tiheys tunnetaan

- Miten lasketaan tietyn **välin todennäköisyys**
  - Tiheydestä: integroimalla Määritelmä 2.3.1 (s. 56)
  - Kertymästä: vähennyslaskulla Huomautus 2.3.5 (s. 57)

- Miten lasketaan tai todetaan jakauman **odotusarvo**
  - Tiheydestä: integroimalla Määritelmä 2.4.1 (s. 64)

Tasajakauma



- Bussin odotusaika **$X \sim$ Tas(0, 10)**
- Tiheys $f(x) = 1/10$ kun $0<x<10$
- Kertymä $F(x)= x/10$ kun $0<x<10$

Huom. F'(x)=f(x)

- Välien todennäköisyydet integroimalla tiheyttä (vakiota) ko. välin yli, **tai suoraan kertymäfunktiosta:**

$P(2 < X < 5) = F(5) - F(2) = 0.5 - 0.2 = 0.3$

$P(X > 7) = 1 - F(7) = 1.0 - 0.7 = 0.3$

$P(X < 5) = F(5) = 0.5$

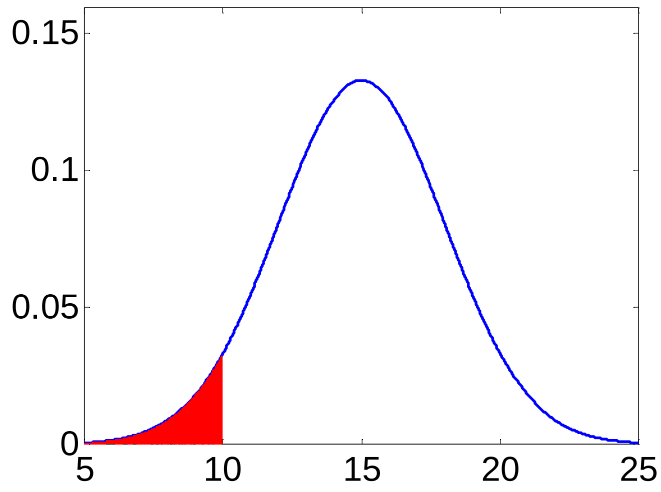
- Tasajakauman odotusarvolle on tunnettu lauseke (= päätepisteiden keskiarvo)

$E(X) = (0+10) / 2 = 5.0$

Mutta se voidaan myös laskea raa'asti integraalina:

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx = \int_{0}^{10} x \cdot 0.1 dx = 5.0$$

Normaalijakauma



- Bussin matka-aika **$X \sim$ N(15, 3^2)**
  Parametrit μ=15, σ=3

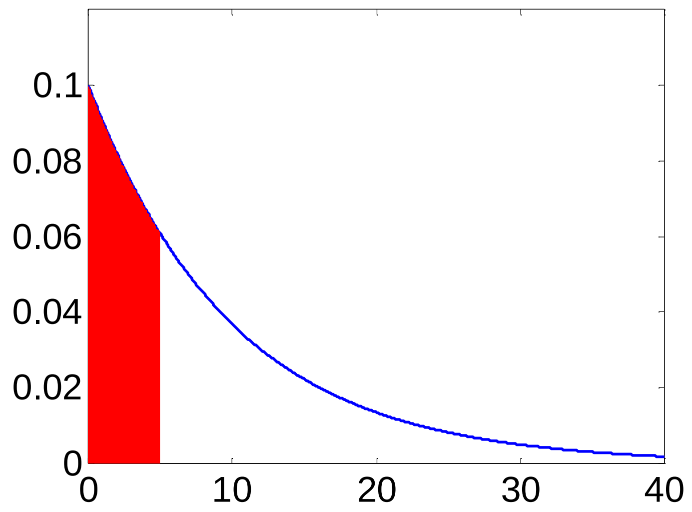
- Tiheys $f(x) = c \cdot \exp(- \ldots)$
- Kertymä $F(x) = \Phi((x - 15) / 3)$

- Tiheysfunktiota hankala integroida joten käytetään kertymäfunktiota.
  (Kertymäfunktio löytyy taulukoista, esim. Tuomisen kirjan takana, ja monista laskimista)
- Välien todennäköisyydet suoraan kertymäfunktiosta:

$P(10 < X < 15)$	$= F(15) - F(10)$	$= 0.500 - 0.048$	$= 0.452$
$P(X > 20)$	$= 1 \quad - F(20)$	$= 1.000 - 0.952$	$= 0.048$
$P(X < 10)$	$= F(10)$		$= 0.048$

- Normaalijakauman odotusarvolle on tunnettu lauseke (= eka parametri μ)
  $E(X) = 15$

Exp-jakauma



- Lampun kestoaika **$X \sim$ Exp(0.1)**
  Parametri λ=0.1

- Tiheys $f(x) = 0.1 \exp(-0.1x)$ kun $x>0$
- Kertymä $F(x) = 1 - \exp(-0.1x)$ kun $x>0$ Huom. $F'(x)=f(x)$

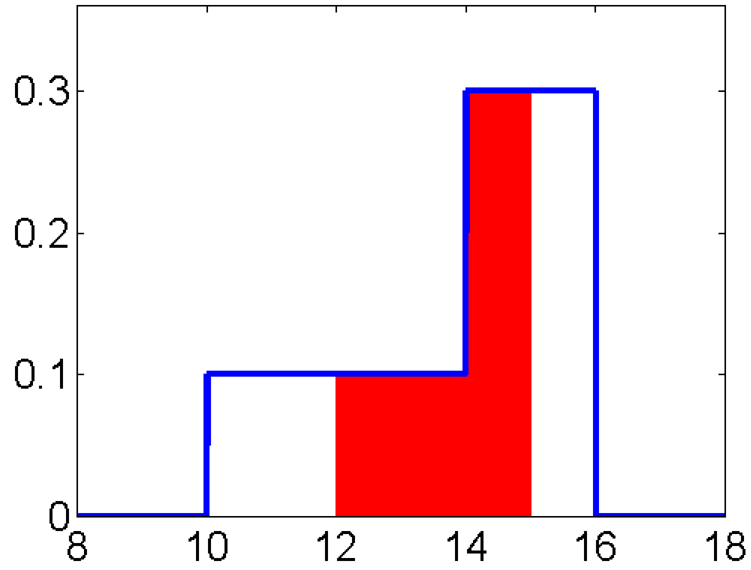
- Välien todennäköisyydet integroimalla tiheyttä ko. välin yli,
  tai suoraan kertymäfunktiosta:
  P($5 < X < 10$) $= F(10) - F(5)$ $= 0.632 - 0.394$ $= 0.239$
  P($X > 20$) $= 1 - F(20)$ $= 1.000 - 0.865$ $= 0.135$
  P($X < 5$) $= F(5)$ $= 0.394$

- Eksponenttijakauman odotusarvolle on tunnettu lauseke ($= 1 / \lambda$)
  E(X) $= 1 / 0.1 = 10$
  Sen voisi laskea myös integraalina (tarvitaan osittaisintegrointia)

Yleinen tiheysfunktio

- Meteorologi kertoo, että huomisaamun lämpötilalla *X* on tiheysfunktio

$$f(x) = \begin{cases} 0.1 & \text{kun } 10 \le x < 14 \\ 0.3 & \text{kun } 14 \le x < 16 \\ 0 & \text{muualla} \end{cases}$$



- Todennäköisyydet ja odotusarvo saadaan integroimalıa

$$P(12 < X < 15) = \int_{12}^{15} f(x)dx = \int_{12}^{14} 0.1\,dx + \int_{14}^{15} 0.3\,dx = 0.2 + 0.3 = 0.5$$

$$E(X) = \int_{10}^{16} x \cdot f(x)dx = \int_{10}^{14} x\,0.1\,dx + \int_{14}^{16} x\,0.3\,dx = 4.8 + 9.0 = 13.8$$