Grundläggande matematisk statistik

Diskreta och kontinuerliga slumpvariabler

Uwe Menzel, 2018

uwe.menzel@matstat.org

www.matstat.org

1

Diskreta och kontinuerliga slumpvariabler

Slumpvariabel (s.v.):

- variabel som antar slumpmässiga värden, beskriver ett slumpförsök
- stora bokstäver: X, Y, L osv.
- obestämt före slumpförsöket

Observation (för en slumpvariabel):

- resultat (realisation) av ett slumpförsök
- små bokstäver: k, x, l osv.

Diskret slumpvariabel:

- kan endast anta ett ändligt (eller uppräkneligt oändligt) antal värden
- ögontalet vid tärningskast; antal kast innan man får samma ögontal två gånger i rad; antal kunder i en kö; antalet användare kopplade till en server; antalet bilolyckor per år i Sverige osv.

Kontinuerlig slumpvariabel:

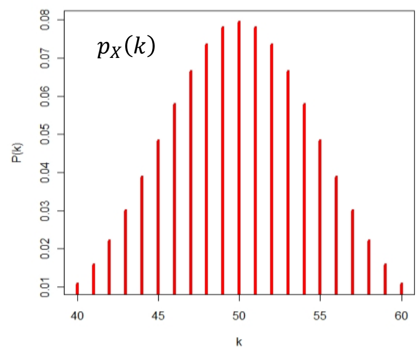
- inte diskret !
- antar reella värden (”oändligt tätt”: mellan två godtyckliga reella värden finns oändligt många reella tal)
- livslängd av en glödlampa; vikt; vindhastighet; elektrisk spänning; koncentration av en kemikalie; …

2

Diskreta slumpvariabler

Diskret slumpvariabel:

- kan endast anta ett ändligt (eller uppräkneligt oändligt) antal värden
- ögontalet vid tärningskast; antal kast innan man får samma ögontal två gånger i rad; antal kunder i en kö; antalet användare kopplade till en server; antalet bilolyckor per år i Sverige osv.



www.matstat.org

3

Sannolikhetsfunktion

- X : diskret slumpvariabel
- k : realisation (observation)

Beteckning för sannolikheten att slumpvariablen X antar värdet k :

$$\boxed{P(X = k) = p_X(k)}$$

Sannolikhetsfunktion: $P(X = k)$ för varje k från Ω_X

<u>Exempel</u>: tärning slumpvariabel X : ögontalet för ett kast

$$P(X = 4) = p_X(4) = \frac{1}{6}$$

www.matstat.org

4

Sannolikhetsfunktion

... kan beskrivas på tre sätt:

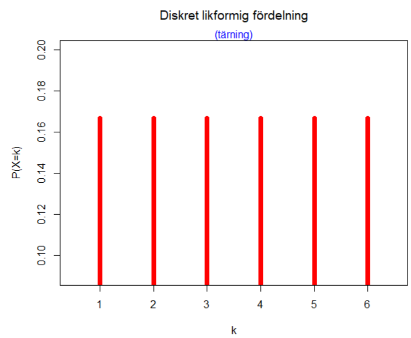
Exempel: tärning: $\Omega_X = \{1, 2, 3, 4, 5, 6\}$

1. **tabell:**

k	1	2	3	4	5	6
$p_X(k)$	$^1/_6$	$^1/_6$	$^1/_6$	$^1/_6$	$^1/_6$	$^1/_6$

2. **formel:** $p_X(k) = {}^1/_6 \qquad k = 1, 2, \ldots, 6$

3. **stolpdiagram**



www.matstat.org

5

Egenskaper hos sannolikhetsfunktionen

$0 \le p_X(k) \le 1$ för alla k (följer från axiom 1)

$\sum_{alla\ k} p_X(k) = 1$ axiom 2 $P(\Omega) = 1$

$P(A) = \sum_{k \in A} p_X(k)$ axiom 3 (utfall för olika k måste vara oförenliga)

summeras över alla utfall som ingår i händelse A

Exempel : tärning $A = \{2, 4, 6\}$ jämnt tal

$$P(A) = \sum_{k=2,4,6} p_X(k) = p_X(2) + p_X(4) + p_X(6) = \frac{3}{6}$$

www.matstat.org

6

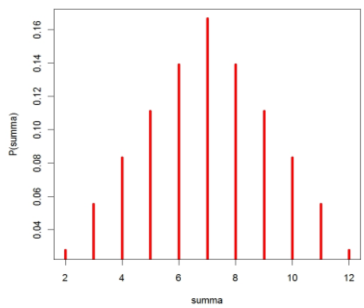
Egenskaper hos sannolikhetsfunktionen

Exempel : två tärningar   slumpvariabel $Z =$ ögonsumma $\Omega_Z = \{2, 3, \dots, 12\}$

k	2	3	4	5	6	7	8	9	10	11	12
$p_Z(k)$	$^1/_{36}$	$^2/_{36}$	$^3/_{36}$	$^4/_{36}$	$^5/_{36}$	$^6/_{36}$	$^5/_{36}$	$^4/_{36}$	$^3/_{36}$	$^2/_{36}$	$^1/_{36}$

kom ihåg hur tabellen skapades
(föreläsning **F1**)

	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



Har vi gjort detta på rätt sätt ? :

$0 \le p_Z(k) \le 1$ ✓

$$\sum_{k=2}^{12} p_Z(k) = 1$$ ✓

7

Egenskaper hos sannolikhetsfunktionen

Fortsättning, exempel : två tärningar

händelse **A** : summan av ögontalet är ett jämnt tal

$$P(A) = \sum_{k \in A} p_Z(k) = p_Z(2) + p_Z(4) + \ \dots + p_Z(12) = \underline{\underline{\frac{1}{2}}}$$

händelse **B** : summan av ögontalet ≤ 4

$$P(B) = P(Z \le 4) = p_Z(2) + p_Z(3) + \ p_Z(4) = \frac{1}{36} + \frac{2}{36} + \frac{3}{36} = \underline{\underline{\frac{1}{6}}}$$

	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

www.matstat.org

8

Egenskaper hos sannolikhetsfunktionen

Exempel 1, diskret sannolikhetsfunktion

k	1	2	3	4
$p_X(k)$	0.2	0.3	0.4	?

$\Omega = \{1, 2, 3, 4\}$

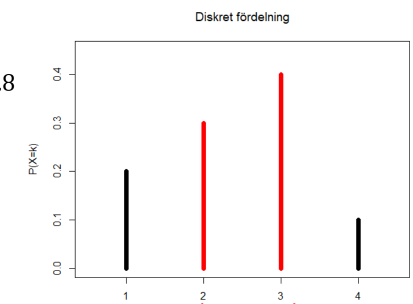
$\Rightarrow$ $p_X(4) = 0.1$ axiom 2

$P(X \leq 1) = p_X(1) = 0.2$

$P(X > 1) = p_X(2) + p_X(3) + p_X(4) = 0.8$

$P(1 < X \leq 3) = p_X(2) + p_X(3) = 0.7$

OBS!: Ettan är **inte** inkluderad! För diskreta slumpvariabel är det **mycket viktigt att skilja mellan "<" och "≤"**



$P(1 < X \leq 3)$

www.matstat.org

9

Egenskaper hos sannolikhetsfunktionen

Exempel 2, diskret sannolikhetsfunktion

k	1	2	3	4
$p_X(k)$	0.2·c	0.3·c	0.4·c	0.5·c

$\Omega = \{1, 2, 3, 4\}$

$c = ?$

$$\sum_{k=1}^{4} p_X(k) = 1.4 \cdot c = 1 \quad \Rightarrow \quad c = \frac{10}{14} = \frac{5}{7} \quad \text{(axiom 2)}$$

Sammanfattning:

- Sannolikhetsfunktionen anger sannolikheten för varje utfall av en diskret slumpvariabel
- kan representeras med formel, tabell, stolpdiagram

www.matstat.org

10

Fördelningsfunktionen för diskreta slumpvariabler

Definition fördelningsfunktion:

- X : slumpvariabel
- a : reellt tal

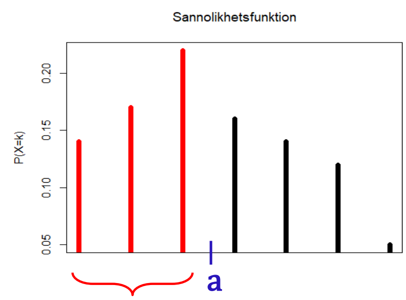
$$\boxed{F_X(a) = P(X \le a)}$$

tecknet "≤" är viktigt: "lika med eller mindre än a" !

Fördelningsfunktionen kan beräknas med hjälp av sannolikhetsfunktionen:

$$\boxed{F_X(a) = P(X \le a) = \sum_{k \le a} p_X(k)}$$

- Man summerar över alla $p_X(k)$ vars argument k är lika med eller mindre än det **reella talet a**
- (**OBS:** a måste alltså **inte** tillhöra utfallsrummet Ω !)
- fördelningsfunktionen är en kumulativ summa.



summera över dessa $p_X(k)$

www.matstat.org

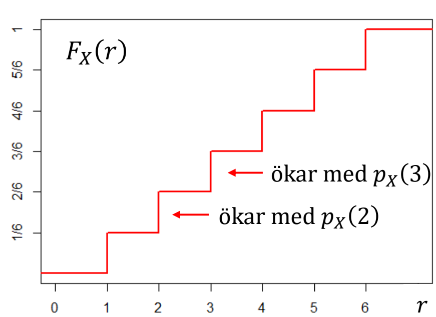
11

Fördelningsfunktionen för diskreta slumpvariabler

Exempel : tärning

utfall	k	1	2	3	4	5	6
Slh.funktion	$p_X(k)$	$^1/_6$	$^1/_6$	$^1/_6$	$^1/_6$	$^1/_6$	$^1/_6$
Fördelningsfkt.	$F_X(k)$	$^1/_6$	$^2/_6$	$^3/_6$	$^4/_6$	$^5/_6$	**1**

Fördelningsfunktionen (CDF) är alltså en kontinuerlig funktion, definerad för reella tal. (Tabellen ovan visar den bara för några utvalda tal.)



$F_X(0.3) = 0$

högerkontinuerlig:

$F_X(1) = P(X \le 1) = {^1/_6}$

$F_X(4.5) = {^2/_3}$

$F_X(1000) = 1$

F_X är 1 för alla tal som är större än det största utfallet

12

Fördelningsfunktion, egenskaper

$0 \le F_X(t) \le 1$ — $F_X(t)$ är ju själv en slh.: $F_X(a) = P(X \le a)$; måste vara mellan 0 och 1 (axiom 1)

$\lim_{t \to -\infty} F_X(t) = 0$ — ”$F_X(-\infty) = P(X \le -\infty)$” : är noll ”vänster om” det minsta värdet av utfallsrummet Ω

$\lim_{t \to +\infty} F_X(t) = 1$ — alla utfall måste väl vara mindre än ∞

$t_2 > t_1 \;\rightarrow\; F_X(t_2) \ge F_X(t_1)$ — F växer monotont (icke avtagande). Det adderas ju bara positiva tal.

Beräkning av sannolikhetsfunktionen med hjälp av fördelningsfunktionen
(t. ex. tärning):

$$F_X(3) = \underbrace{p_X(1) + p_X(2) + p_X(3)}$$

$$F_X(4) = \qquad F_X(3) \qquad + p_X(4)$$

$\Rightarrow$ $p_X(4) = F_X(4) - F_X(3)$

$$\boxed{p_X(k) = F_X(k) - F_X(k-1)}$$

$F_X(k)$ ökar med $p_X(k)$, se bild ovan
F_X är ofta tabellerad → möjlighet att beräkna $p_X(k)$

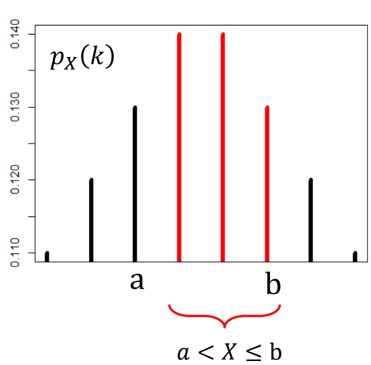
13

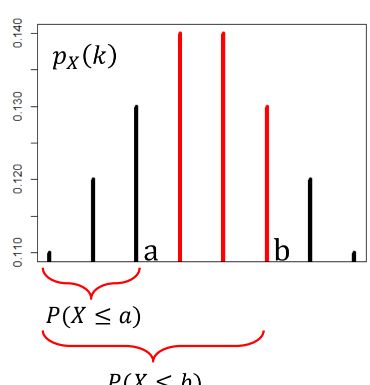
Fördeningsfunktionen och sannolikhet för händelser

X: diskret slumpvariabel ; F_X : dess fördelningsfunktion

mycket viktigt att använda tecknen korrekt!

$\boxed{P(a < X \le b) = F_X(b) - F_X(a)}$ alltså $P(a < X \le b) = P(X \le b) - P(X \le a)$



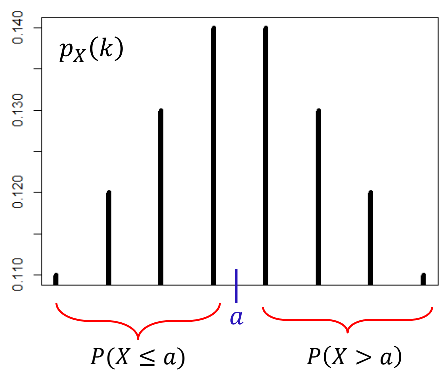


www.matstat.org

14

Fördeningsfunktionen och Komplementsatsen

$\boxed{P(X > a) = 1 - P(X \leq a)}$ $\quad P(X > a) = 1 - F_X(a)$



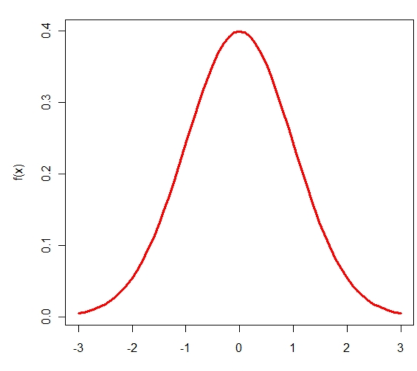
www.matstat.org

15

Kontinuerliga slumpvariabler

Kontinuerliga slumpvariabel:

- inte diskreta !
- utfall = reella värden (”som är oändligt täta”: mellan två godtyckliga reella tal finns oändligt många reella tal)
- livslängd av en glödlampa; vikt; vindhastighet; elektrisk spänning; koncentration av en kemikalie; …



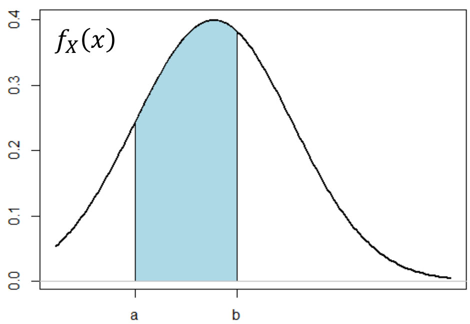
www.matstat.org

16

Täthetsfunktion för kontinuerliga slumpvariabler

En kontinuerlig slumpvariabel X kan beskrivas med hjälp av en täthetsfunktion $f_X(x)$

- X : kontinuerlig slumpvariabel
- x : argument



Sannolikheten att slumpvariablen X hamnar mellan de reella värden a och b är lika med arean under täthetsfunktionen mellan a och b:

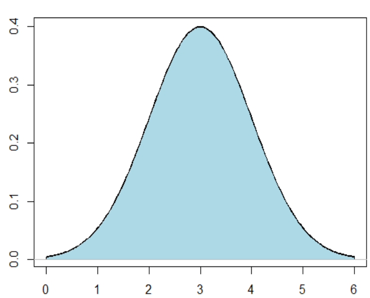
$$\boxed{P(a < X \le b) = \int_a^b f_X(x)\,dx}$$

= sannolikhet att X ligger i intervallet (a, b)

www.matstat.org

17

Täthetsfunktion



Arean under en täthetsfunktion måste vara 1 (axiom 2 : $P(\Omega) = 1$)

$$\boxed{\int_{-\infty}^{+\infty} f_X(x)\,dx = 1}$$

normering

detta är ju slh.:en att X hamnar mellan -∞ och +∞, något som med säkerhet måste inträffa.

Anmärkningar:

- Om utfallsrummet Ω är begränsat till ett visst interval (u, v) så kan integralgränserna i formeln ovan ersättas med dessa värden.
- En täthetsfunktion kan bara ange sannolikheten att slumpvariabeln X hamnar mellan två värden, ett enskilt värde har **sannolikheten noll** !

$\cancel{P(X = a)}$ inte definerad för kontinuerliga slumpvariabler!

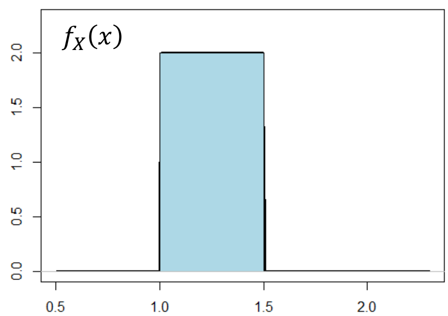
www.matstat.org

18

Täthetsfunktion

Kom ihåg: en sannolikhetsfunktion för en diskret slumpvariabel måste uppfylla kravet $0 \leq p_X(k) \leq 1$
I motsats till sannolikhetsfunktionen kan en täthetsfunktion anta värden som är större an 1, så länge arean under hela kurvan är 1.

Exempel: Likformig kontinuerlig fördelning på intervallet $(1, 1.5)$



$$\int_{-\infty}^{+\infty} f_X(x)\,dx = \int_{1}^{1.5} 2\,dx = [2x]_1^{1.5} = \underline{\underline{1}}$$

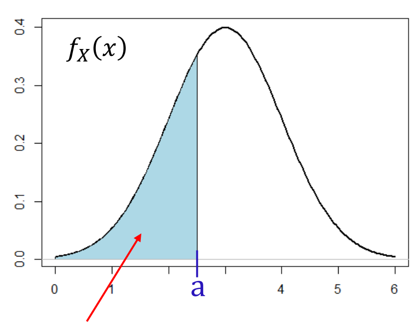
(mycket) enklare: $2 \cdot 0.5 = 1$

Fördelningsfunktionen för kontinuerliga slumpvariabler

X: kontinuerlig slumpvariabel ; F_X: dess fördelningsfunktion

$$\boxed{F_X(a) = P(X \leq a) = \int_{-\infty}^{a} f_X(x)dx}$$

En ny funktion defineras för varje värde på a: funktionsvärdet = arean under täthetsfunktionen ”vänster” om a.

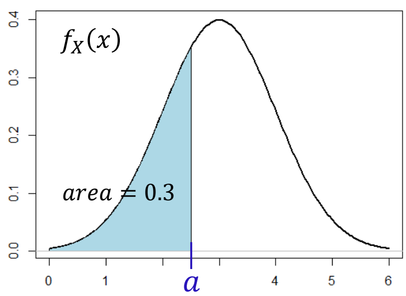


$F_X(a)$ är den blå markerade arean

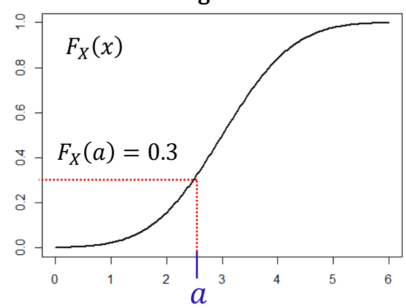
Sammanhanget mellan täthetsfunktionen och fördelningsfunktionen

$$F_X(a) = P(X \leq a) = \int_{-\infty}^{a} f_X(x)dx$$

täthetsfunktion



fördelningsfunktion



www.matstat.org

21

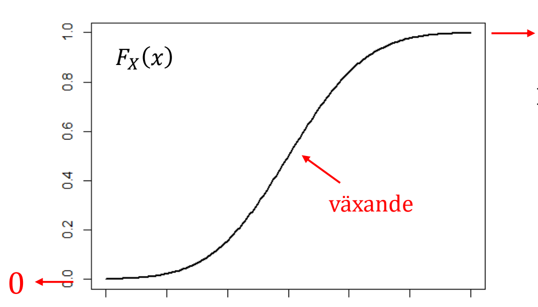
Egenskaper hos fördelningsfunktionen

X: kontinuerlig slumpvariabel; F_X : dess fördelningsfunktion

$\lim_{a \to -\infty} F_X(a) = 0$ "$F_X(-\infty) = P(X \leq -\infty)$" (OBS!: denna notation är inte korrekt)

$\lim_{a \to +\infty} F_X(a) = 1$ "$F_X(\infty) = P(X \leq \infty)$" (OBS!: denna notation är inte korrekt)

Fördelningsfunktionen är monotont växande (icke avtagande).



Monotoni:

om $a_1 < a_2$

måste $P(X \leq a_1) \leq P(X \leq a_2)$

alltså $F_X(a_1) \leq F_X(a_2)$

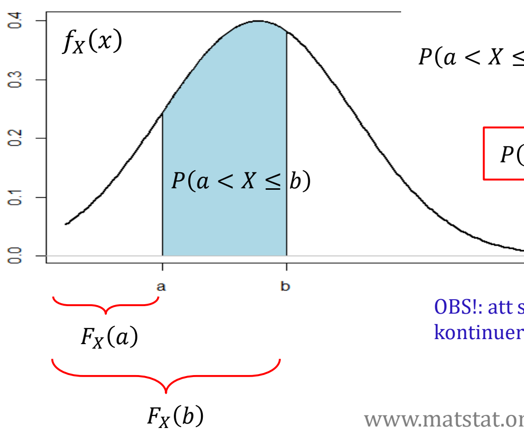
www.matstat.org

22

Beräkning av sannolikheter med hjälp av fördelningsfunktionen

X: kontinuerlig slumpvariabel; F_X : dess fördelningsfunktion

Kom ihåg: sannolikheten att slumpvariabeln X hamnar mellan de reella värden a och b är lika med arean under täthetsfunktionen mellan a och b:



$$P(a < X \leq b) = \int_a^b f_X(x)\,dx = F_X(b) - F_X(a)$$

$$\boxed{P(a < X \leq b) = F_X(b) - F_X(a)}$$

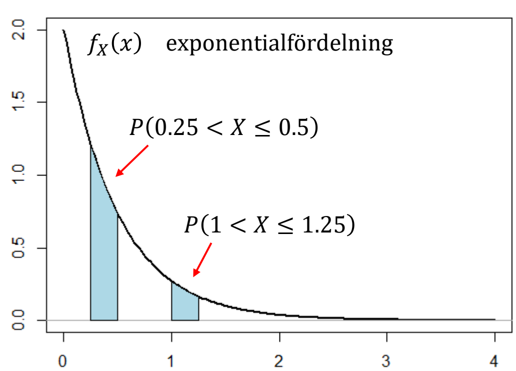
OBS!: att skilja mellan ”≤” och ”<” i formeln för kontinuerliga slumpvariabler är **inte** viktigt

www.matstat.org

23

Beräkning av sannolikheter med hjälp av fördelningsfunktionen

Exempel: $f_X(x) = \begin{cases} 2 \cdot e^{-2x} & x \geq 0 \\ 0 & x < 0 \end{cases}$ exponentialfördelning



www.matstat.org

24

Beräkning av sannolikheter med hjälp av fördelningsfunktionen

Exempel , fortsättning : $f_X(x) = \begin{cases} 2 \cdot e^{-2x} & x \geq 0 \\ 0 & x < 0 \end{cases}$

a) Kolla normeringen: $\int_{-\infty}^{+\infty} f_X(x)\, dx = 1$??

$$\int_{-\infty}^{+\infty} f_X(x)dx = \int_{0}^{+\infty} 2 \cdot e^{-2x}\, dx = \left[2 \cdot \frac{1}{-2} \cdot e^{-2x}\right]_0^{+\infty} = [e^{-2x}]_\infty^0 = \underline{\underline{1}}$$ ✓

www.matstat.org

25

Beräkning av sannolikheter med hjälp av fördelningsfunktionen

Exempel , fortsättning: $f_X(x) = \begin{cases} 2 \cdot e^{-2x} & x \geq 0 \\ 0 & x < 0 \end{cases}$

b) Beräkna fördelningsfunktionen:

$a > 0$ krävs!

$$F_X(a) = \int_{-\infty}^{a} f_X(x)dx = \int_{0}^{a} 2 \cdot e^{-2x}\, dx = [-e^{-2x}]_0^a = [e^{-2x}]_a^0 = \underline{\underline{1 - e^{-2a}}}$$

$$F_X(x) = \begin{cases} 1 - e^{-2x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

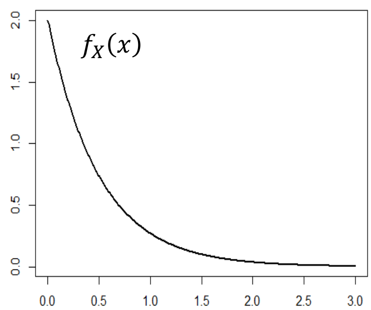
www.matstat.org

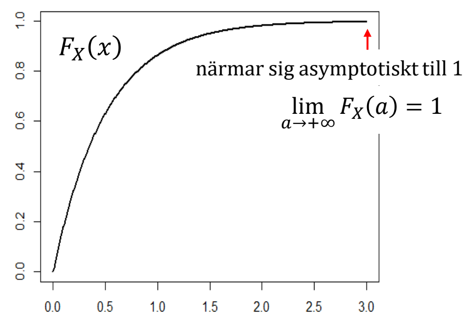
26

Beräkning av sannolikheter med hjälp av fördelningsfunktionen

Exempel , fortsättning:

$$f_X(x) = \begin{cases} 2 \cdot e^{-2x} & x \geq 0 \\ 0 & x < 0 \end{cases} \qquad F_X(x) = \begin{cases} 1 - e^{-2x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$



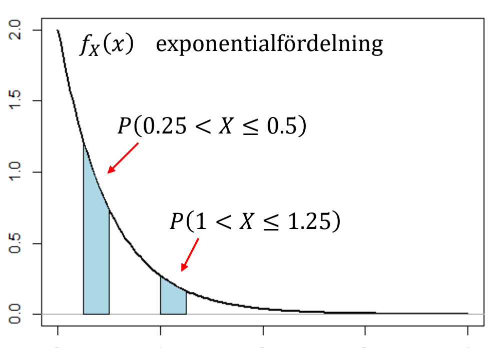


www.matstat.org

27

Beräkning av sannolikheter med hjälp av fördelningsfunktionen

Exempel , fortsättning: $F_X(x) = \begin{cases} 1 - e^{-2x} & x \geq 0 \\ 0 & x < 0 \end{cases}$



$$P(0.25 < X \leq 0.5) = F_X(0.5) - F_X(0.25) = (1 - e^{-1}) - (1 - e^{-0.5}) = e^{-0.5} - e^{-1} = 0.239$$

$$P(1 < X \leq 1.25) = F_X(1.25) - F_X(1) = (1 - e^{-2.5}) - (1 - e^{-2}) = e^{-2} - e^{-2.5} = 0.053$$

www.matstat.org

28

Sammanhanget mellan täthetsfunktionen och fördelningsfunktionen

$$F_X(a) = \int_{-\infty}^{a} f_X(x)dx \qquad \text{derivering } \frac{\partial}{\partial a}$$

$$\Rightarrow$$

$$\boxed{\frac{\partial}{\partial a} F_X(a) = f_X(a)}$$

derivering av parameterintegral → **Leibnizregel:**

$$\frac{\partial}{\partial a} \int_{u(a)}^{o(a)} h(x)\,dx = h[o(a)] \cdot \dot{o}(a) - h[u(a)] \cdot \dot{u}(a)$$

(om $h(x)$ inte explicit beror på a)

Exempel exponentialfördelning , fortsättning:

$$F_X(a) = 1 - e^{-2a} \quad \text{för } a > 0$$

$$f_X(a) = \frac{\partial}{\partial a} F_X(a) = \frac{\partial}{\partial a}(1 - e^{-2a}) = \underline{\underline{2 \cdot e^{2a}}} \quad ✓ \qquad (a > 0)$$

www.matstat.org

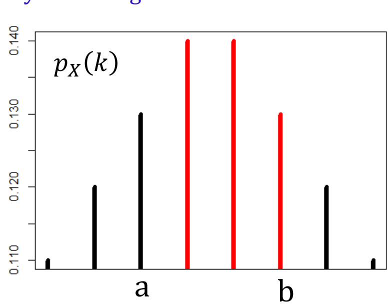
29

Beräkning av sannolikheter för diskreta och kontinuerliga slumpvariabler

Diskreta slumpvariabler

$$P(a < X \leq b) = F_X(b) - F_X(a)$$

mycket viktig!



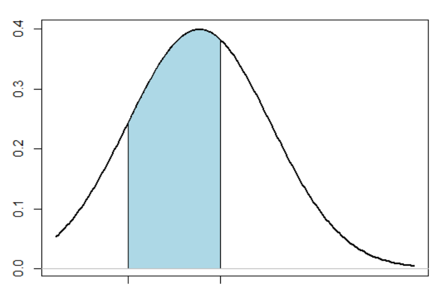
$$\underbrace{}_{a < X \leq b}$$

www.matstat.org

Kontinuerliga slumpvariabler

$$P(a < X \leq b) = P(a \leq X \leq b) = P(a < X < b)$$

spelar ingen roll



$$\underbrace{}_{F_X(a)}$$

$$\underbrace{}_{F_X(b)}$$

30

Beräkning av sannolikheter för kontinuerliga slumpvariabler

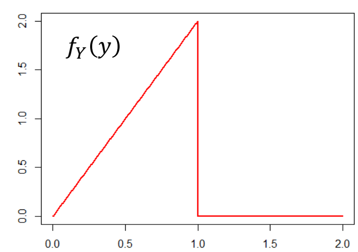
Exempel: täthetsfunktion $f_Y(y) = \left\{ \begin{array}{cc} c \cdot y & 0 < y \leq 1 \\ 0 & annars \end{array} \right\}$ c = konstant

a) Bestäm konstanten c

$$1 = \int_{-\infty}^{+\infty} f_Y(y)\, dy = \int_0^1 c \cdot y\, dy = c \cdot \left[\frac{1}{2} \cdot y^2\right]_0^1 = c \cdot \frac{1}{2} \Rightarrow \underline{\underline{c = 2}}$$

b) Rita upp täthetsfunktionen

$$f_Y(y) = \left\{ \begin{array}{cc} 2 \cdot y & 0 < y \leq 1 \\ 0 & annars \end{array} \right\}$$



www.matstat.org

31

Beräkning av sannolikheter för kontinuerliga slumpvariabler

c) Hitta fördelningsfunktionen $f_Y(y) = \left\{ \begin{array}{cc} 2 \cdot y & 0 < y \leq 1 \\ 0 & annars \end{array} \right\}$

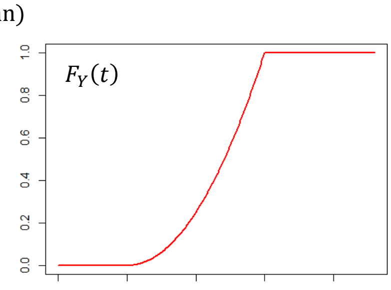
$$F_Y(t) = \int_{-\infty}^{t} f_Y(y) dy = \int_0^t 2 \cdot y\, dy = 2 \cdot \left[\frac{1}{2} \cdot y^2\right]_0^t = t^2 \qquad \text{för } 0 < t \leq 1$$

$F_Y(t) = 0$ för $t < 0$ (se bild ovan)

$F_Y(t) = 1$ för $t > 1$

$\Rightarrow$

$$\boxed{F_Y(t) = \left\{ \begin{array}{cc} 0 & t < 0 \\ t^2 & 0 < t \leq 1 \\ 1 & t > 1 \end{array} \right\}}$$



www.matstat.org

32

Jämförelse av diskreta och kontinuerliga slumpvariabler

Diskret slumpvariabel	Kontinuerlig slumpvariabel
sannolikhetsfunktion	täthetsfunktion
$P(X=k)=p_X(k)$	slh. är noll för ett enskilt reellt värde
$0 \le p_X(k) \le 1$	$f(x)$ kan bli större än 1
$\sum_{k\in\Omega} p_X(k)=1$	$\int_{-\infty}^{+\infty} f_X(x)\,dx=1$
$P(A)=\sum_{k\in A} p_X(k)$	$P(a<X\le b)=\int_a^b f_X(x)\,dx$

www.matstat.org

33

Jämförelse av diskreta och kontinuerliga s.v.

Diskret slumpvariabel	Kontinuerlig slumpvariabel
fördelningsfunktion	fördelningsfunktion
$F_X(a)=P(X\le a)$	$F_X(a)=P(X\le a)$
$F_X(a)=\sum_{k\le a} p_X(k)$	$F_X(a)=\int_{-\infty}^{a} f_X(x)\,dx$
$P(a<X\le b)=F_X(b)-F_X(a)$	$P(a<X\le b)=F_X(b)-F_X(a)$
$p_X(k)=F_X(k)-F_X(k-1)$	$f_X(x)=\frac{\mathrm{d}}{\mathrm{d}x}F_X(x)$
$P(X>a)=1-F_X(a)$	$P(X>a)=1-F_X(a)$
$F_X(x)$ monotont växande från 0 till 1	$F_X(x)$ monotont växande från 0 till 1

www.matstat.org

34