

FORELESNING 10

ELVIND ERIKSEN

28 FEB 2012

MET 3431

STATISTIKK

PLAN:

① OPPSUMMERING: BINOMISK FØRDELING

② NORMALFØRDELING

[T] Kap. 6

Binomialfordelingen

Binomialfordelingen

- La x være antall suksesser i en binomisk forsøksrekke
- Da er x en diskret tilfeldig variabel
- Sannsynligheten for at vi totalt får a vellykkete forsøk er gitt ved binomialfordelingen

$$X = x$$

$$P(x = a) = \binom{n}{a} \cdot p^a \cdot (1 - p)^{n-a}$$

Binomisk forsøksrekke:

- * Et tilfeldig forsøk gjentas n ganger
- * Sannsynlighet for suksess er p hver gang
- * Forsøkene er uavhengige

$X =$ antall suksesser

$$P(x) = \frac{n!}{(n-a)!a!} \cdot p^a \cdot (1-p)^{n-a}$$

Antall utfall med akkurat a suksesser på n forsøk

Sannsynligheten for at de første a forsøkene er suksesser, og at de resterende $n-a$ er mislykkete.

$$\frac{10!}{8! \cdot 2!} = \frac{10 \cdot 9 \cdot 8 \cdot \dots \cdot 1}{8 \cdot 7 \cdot \dots \cdot 1 \cdot 2 \cdot 1}$$

Tre måter å regne ut binomialfordelingen

1. Formelen

$$P(X = a) = \binom{n}{a} \cdot p^a \cdot (1 - p)^{n-a}$$

Example

Sannsynligheten for å få '6' to ganger i ti terningkast er

$n = 10$
 $p = 1/6$
 $X = \text{antall 6'ere}$
 $p = 10 \text{ kast}$

$$P(X = 2) = \frac{10 \cdot 9}{2 \cdot 1} (1/6)^2 \cdot (5/6)^8 = 0.29$$

$$= \binom{10}{2} \cdot \left(\frac{1}{6}\right)^2 \cdot \left(1 - \frac{1}{6}\right)^8$$

$$= \frac{10 \cdot 9}{2 \cdot 1} \cdot \left(\frac{1}{6}\right)^2 \cdot \left(\frac{5}{6}\right)^8$$

Tre måter å regne ut binomialfordelingen

2. Tabell

Bruk tabell A-1 i boka. F.eks. med $n = 12$ og $p = 0.8$ så er sannsynligheten for 5 suksesser lik 0.003.

n	x	p	
		0.80	0.001
4	4	0.001	0.001
5	5	0.003	0.003
6	6	0.016	0.016
7	7	0.053	0.053

Example

Ti terningkast eksempelet fra forrige side.

- $n = 10$, $p = 1/6$ og $x = 2$
- Sjekk i tabell A-1
- Er sannsynligheten er 0.29?

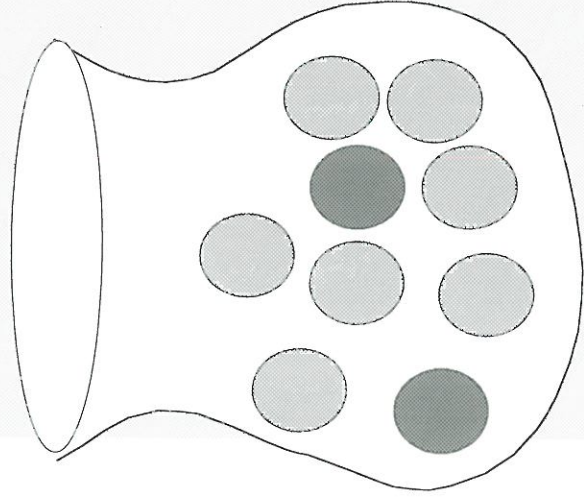
Tre måter å regne ut binomialfordelingen

3. Kalkulator/Excel/JMP

- JMP. Lag en søyle med formula editor Binomial distribution.
- Excel har funksjonene
 - $COMBIN(N; n)$ for $\binom{N}{n}$ $\swarrow$ $a=5, n=20, p=0.22$ punkt sannsynlighet $p(x=5)$
 - $BINOMDIST(5; 20; 0.22; 0)$ for binomiske punktsannsynligheter $\swarrow$ kumulativ $p(x \leq 5)$
 - $BINOMDIST(5; 20; 0.22; 1)$ for binomiske kumulative (maks 5) sannsynligheter
- Online kalkulatorer. Søk på 'binomial distribution pdf' på <http://www.wolframalpha.com>
- Texas BII plus kalkulator har også $\binom{N}{n}$

Binomialfordelingen

Example



- To røde og syv grønne baller
- Vi trekker en ball $n = 20$ ganger, med tilbakelegging.
- Hva er sannsynligheten for å få rød ball nøyaktig fem ganger?
- Svar: Dette er identiske og uavhengige forsøk med $p = \frac{2}{9}$:

$$P(5) = \binom{20}{5} \cdot \left(\frac{2}{9}\right)^5 \cdot \left(\frac{7}{9}\right)^{15} = 0.19$$

$$\frac{20 \cdot 19 \cdot 18 \cdot 17 \cdot 16}{5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}$$

eller

$$\boxed{nCr} : 20 \boxed{nCr} 5 =$$

Punktsannsynligheter og kumulative sannsynligheter

Kumulativ sannsynlighet

Punktsannsynlighet Sannsynligheten for nøyaktig a vellykkete forsøk.
 $P(X=a)$

Kumulativ sannsynlighet er derimot sannsynligheten for at høyst a forsøk er vellykkete: $P(X \leq a)$

$$P(\text{høyst } a) = P(0) + P(1) + P(2) + \dots + P(a)$$
$$P(X \leq a) = P(X=0) + P(X=1) + \dots + P(X=a)$$

Example

To røde og syv grønne baller. Sannsynligheten for rød ball høyst to ganger:

$$P(\text{høyst } 2) = P(0) + P(1) + P(2) =$$
$$\underbrace{\binom{20}{0} \cdot \left(\frac{2}{9}\right)^0 \cdot \left(\frac{7}{9}\right)^{20}}_{P(0)} + \underbrace{\binom{20}{1} \cdot \left(\frac{2}{9}\right)^1 \cdot \left(\frac{7}{9}\right)^{19}}_{P(1)} + \underbrace{\binom{20}{2} \cdot \left(\frac{2}{9}\right)^2 \cdot \left(\frac{7}{9}\right)^{18}}_{P(2)} = 0.146$$

Example

Ringe 12 tilfeldig valgte kunder for en markedsundersøkelse.

- vellykket $\leftrightarrow$ kunden svarer på spørsmålene
- Oppringningene uavhengige av hverandre
- Suksess-sannsynlighet $p = 0.2$
- Binomisk forsøksrekke med $n = 12$ og $p = 0.2$

Sannsynligheten for at minst to kunder svarer?

- Regn ut komplementet: at 0 eller 1 kunde svarer

- $P(0) + P(1) = 0.069 + 0.206 = 0.275$

- Sannsynligheten for at minst to kunder svarer er

$$1 - 0.275 = 0.725$$

$$1 - P(0) - P(1) = 0.725$$

Minst to svarer:
 $1 - P(0) - P(1)$

$$\begin{cases} P(0) + P(1) = P(X \leq 1) \\ P(2) + P(3) + \dots + P(12) \end{cases}$$

- 1 Section 6-2: Standard normalfordelingen
- 2 Section 6-3: Anvendelser av normalfordelingen
- 3 Section 6-4: Observator fordeling
- 4 Section 6-5: Sentralgrenseteoremet

Kontinuerlige tilfeldige variable

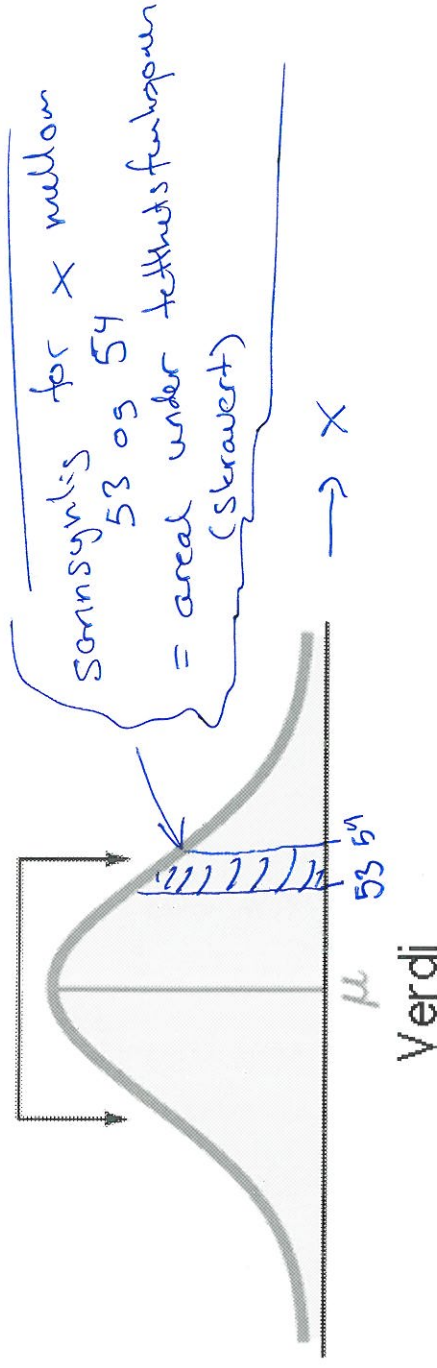
Kontinuerlige vs. diskrete variable

- I forrige kapittel handlet det om diskrete tilfeldige variable. F.eks.
 - Hvor mye du får når du triller en terning
 - Antall suksesser i en binomisk forsøksrekke - binomial fordelingen
- I kapittel 6 handler det om tilfeldige variable som er *kontinuerlige*. F.eks.
 - Vekten til en tilfeldig valgt pasient
 - Lønna til en tilfeldig valgt innbygger

Normalfordelingen

Normalfordelingen

Klokkeformet og symmetrisk



Sannsynlighetstettheten:

σ = std. avvik
 μ = gjennomsnitt

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]$$

Men hvorfor er normalfordelingen viktig?

Normalfordelingen

- Mange typer målinger fordeler seg som denne kurven
 - Høyden på mennesker
 - IQ er normalfordelt med $\mu = 100$ og $\sigma^2 = 10^2$
- Summen av tilfeldigheter er normalfordelt: *sentralgrenseteorem*

Tetthetskurven (eng: density curve)

Grafen til sannsynlighetstettheten

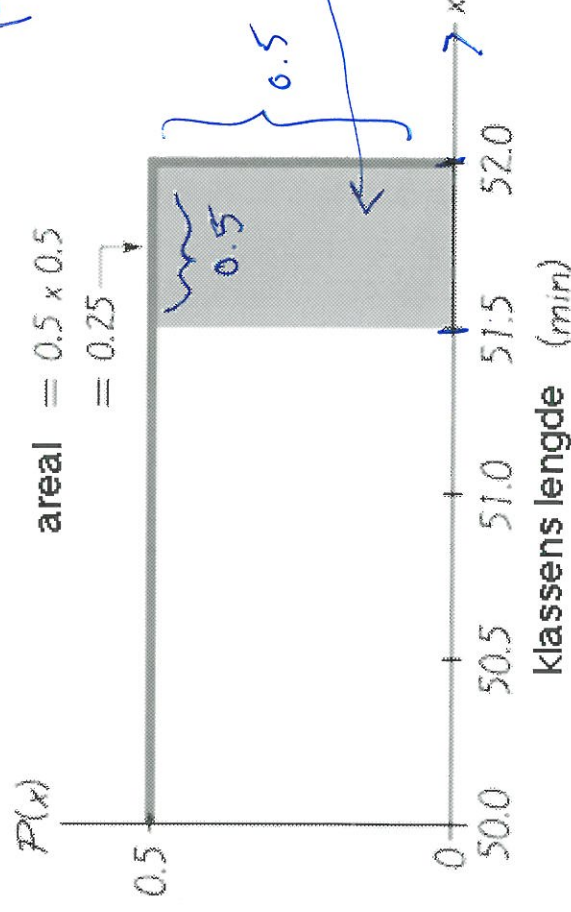
- Krav:

- 1 Totale areal under grafen er 1
 - 2 Kurven må ligge over x-aksen
- Areal $\leftrightarrow$ Sannsynlighet

X ligger mellom 51.5
og 52

$$P(51.5 \leq X \leq 52)$$

$$= 0.5 \cdot 0.5 = \underline{\underline{0.25}}$$

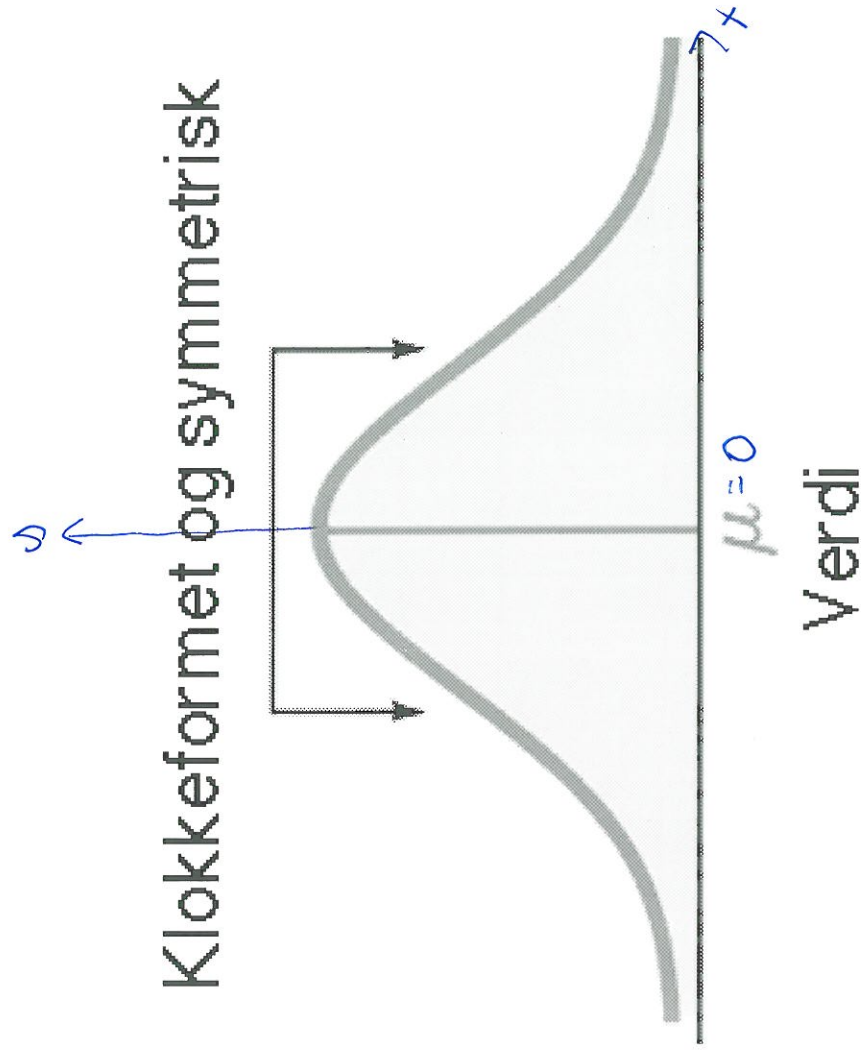


Figur: 6-3. Uniform tetthet.

Standard normalfordelingen

Kjennetegn ved standard normalfordeling

- Klokkeformet og symmetrisk
- Gjennomsnittet: $\mu = 0$
- Standardavviket: $\sigma = 1$



Finne arealet under grafen

Arealet under standard normalfordeling

- Viktig å kunne finne arealet til et område under grafen.
- Slike areal tolkes som sannsynligheter
- Bruk tabell A-2



TABLE A-2 Standard Normal (z) Distribution: Cumulative Area from the LEFT

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
-3.50 and lower	.0001									
-3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
-3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
-3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
-3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
-3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
-2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
-2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
-2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
-2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
-2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
-2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
-2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
-2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
-2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
-2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
-1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
-1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
-1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
-1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
-1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559

$$P(z \leq -3.33)$$

0.0004

z standard normalfordelt

$$P(z \leq -3.33)$$

$$= 0.0004$$

høyst -3.33
= kumulativ