

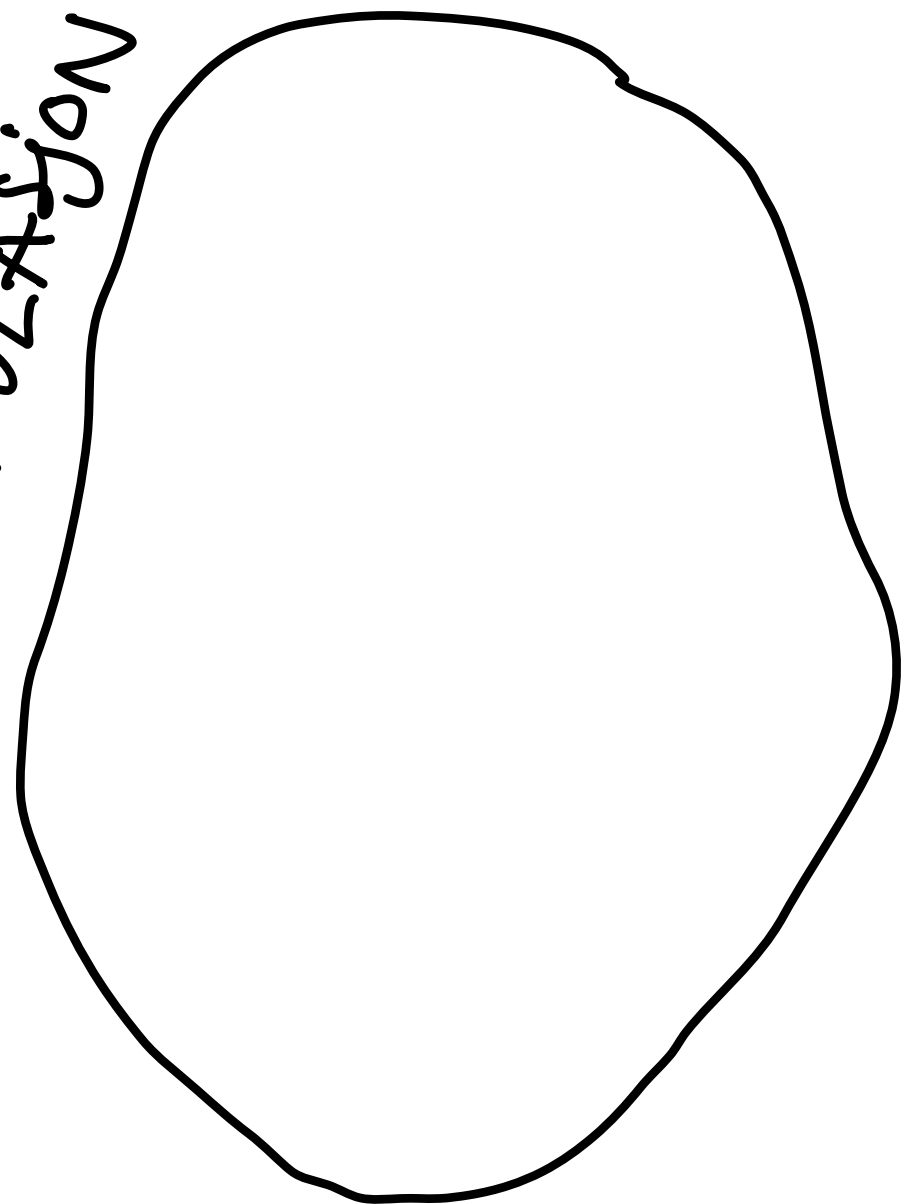


Thea Bjørnland
Institutt for matematiske fag
NTNU

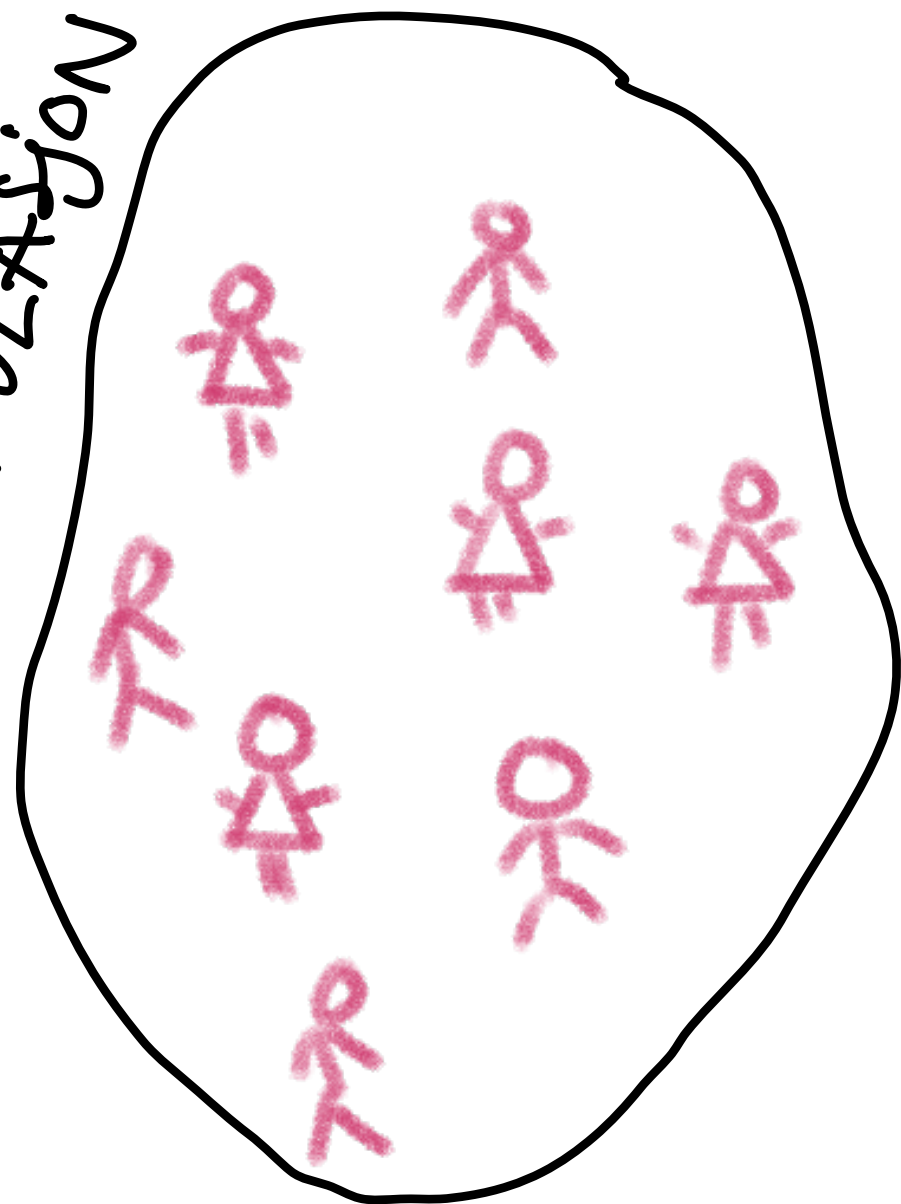
Husk!



POPULATION

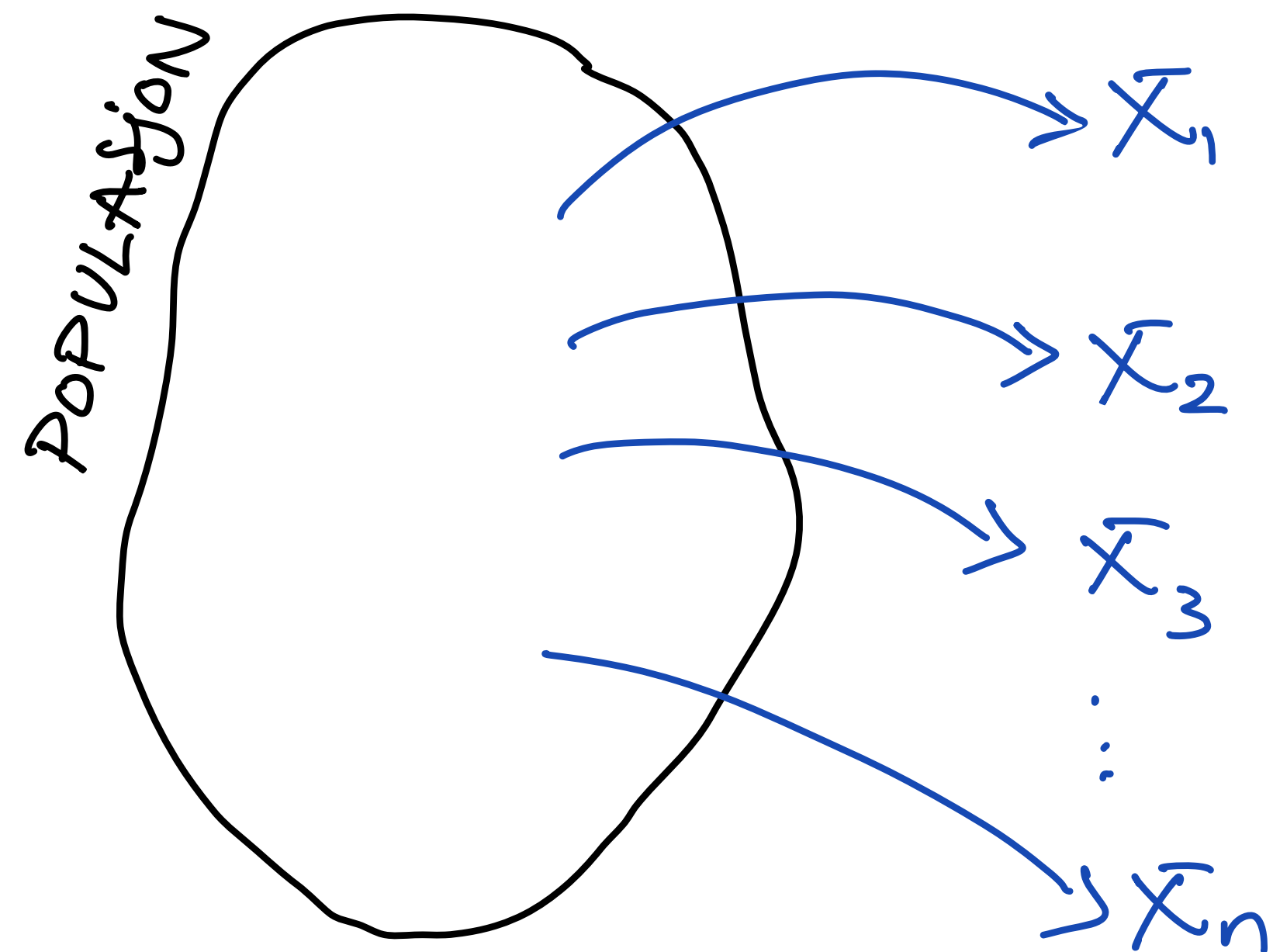


POPULASJON

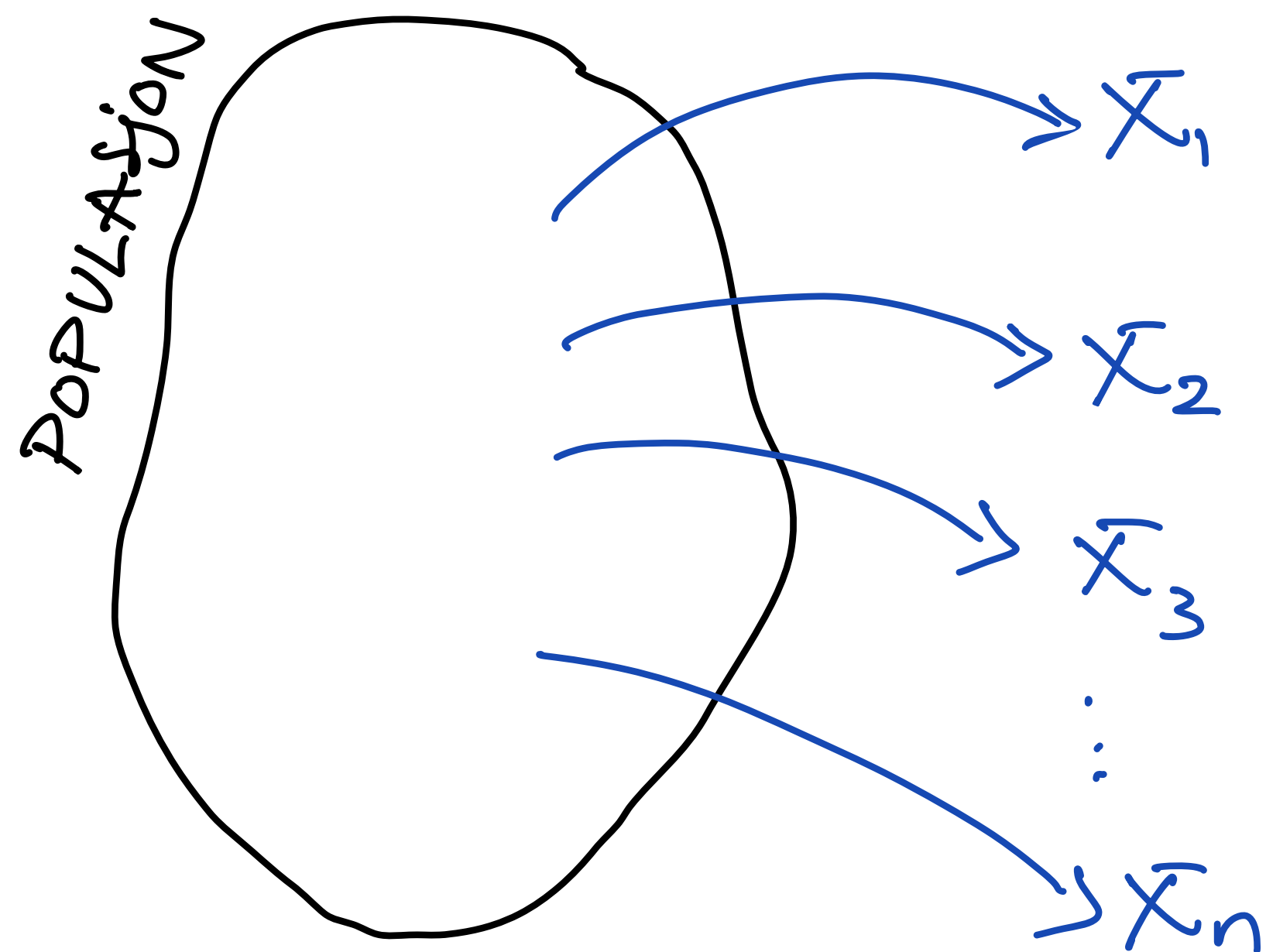


POPULASJON



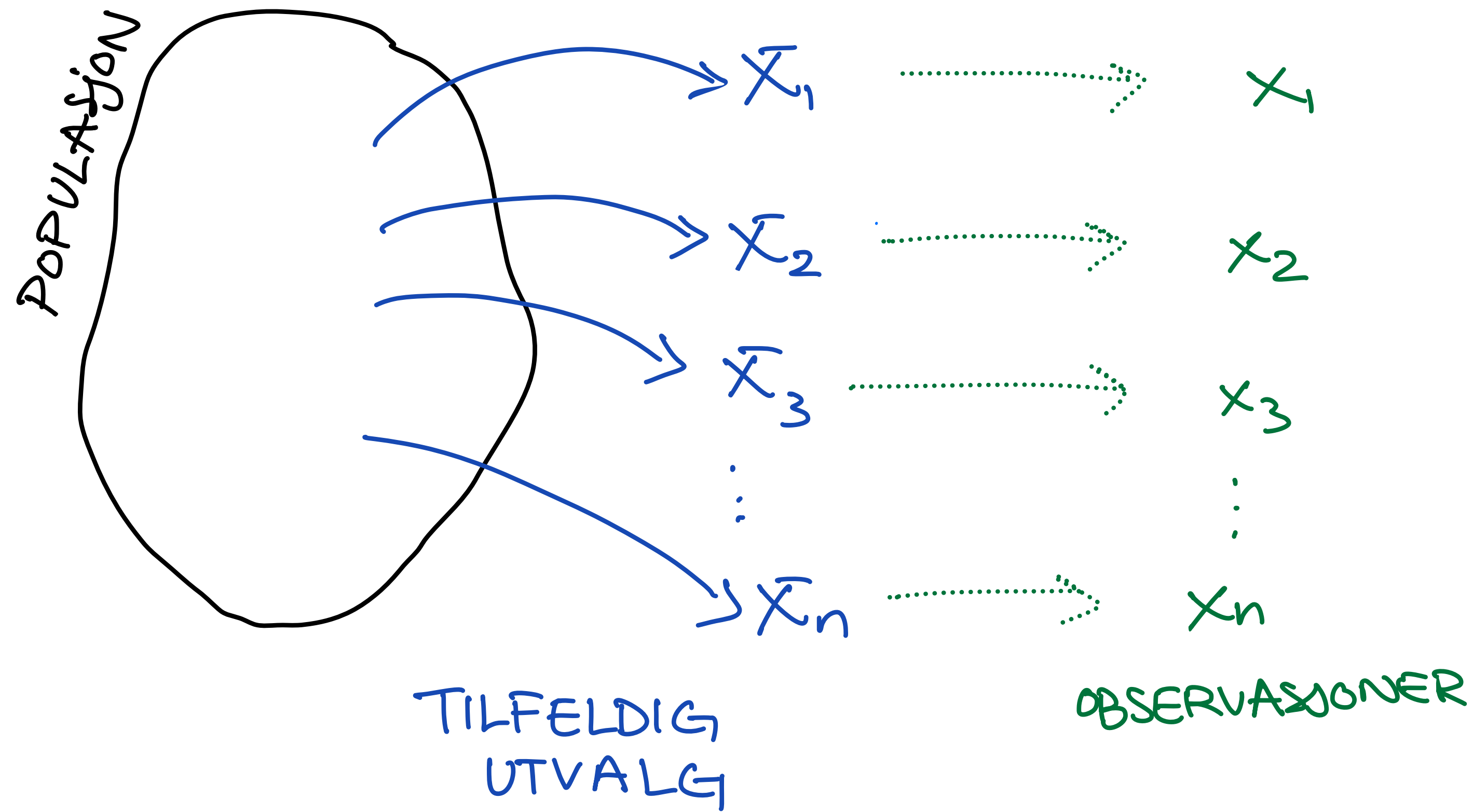


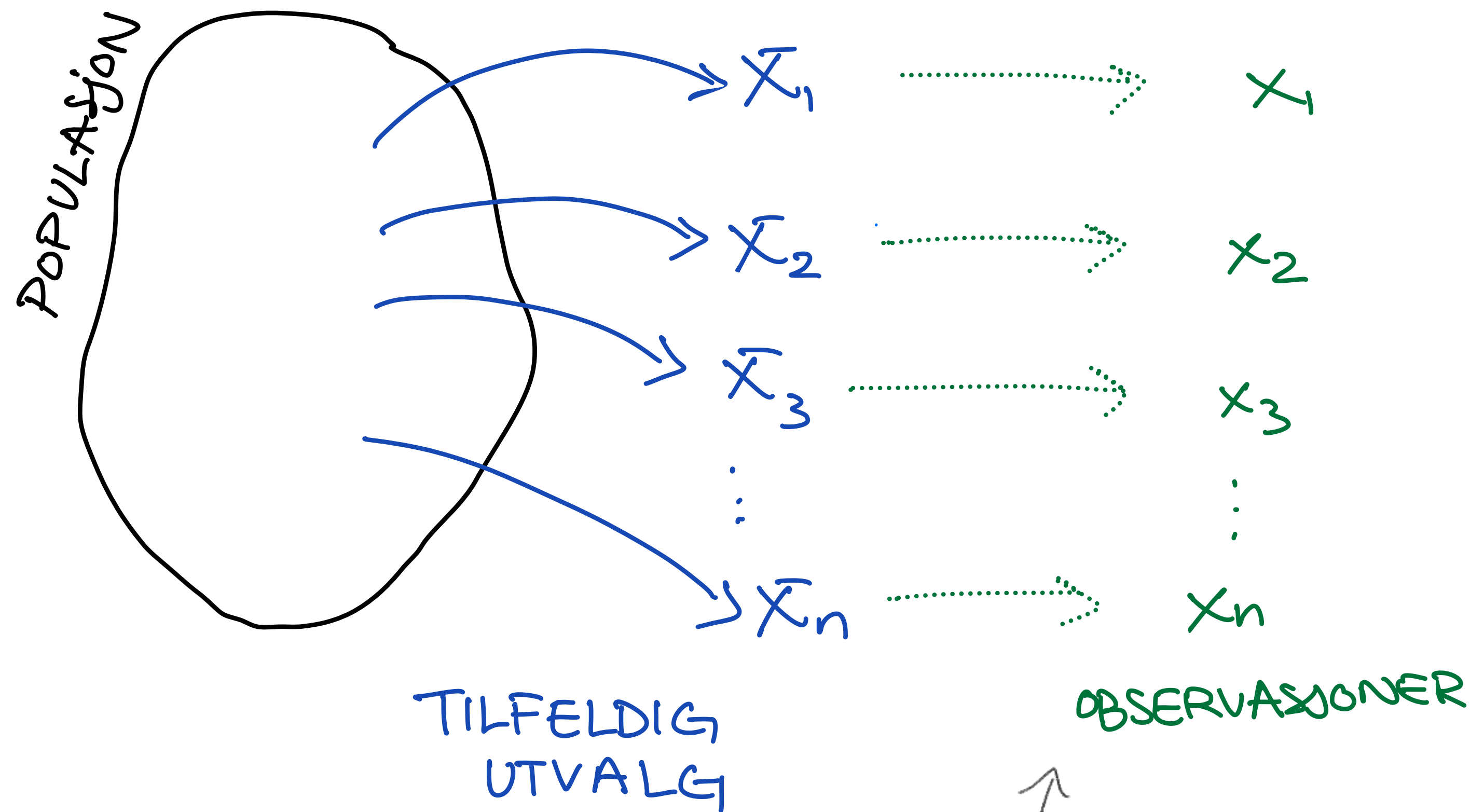
TILFELDIG
UTVALG



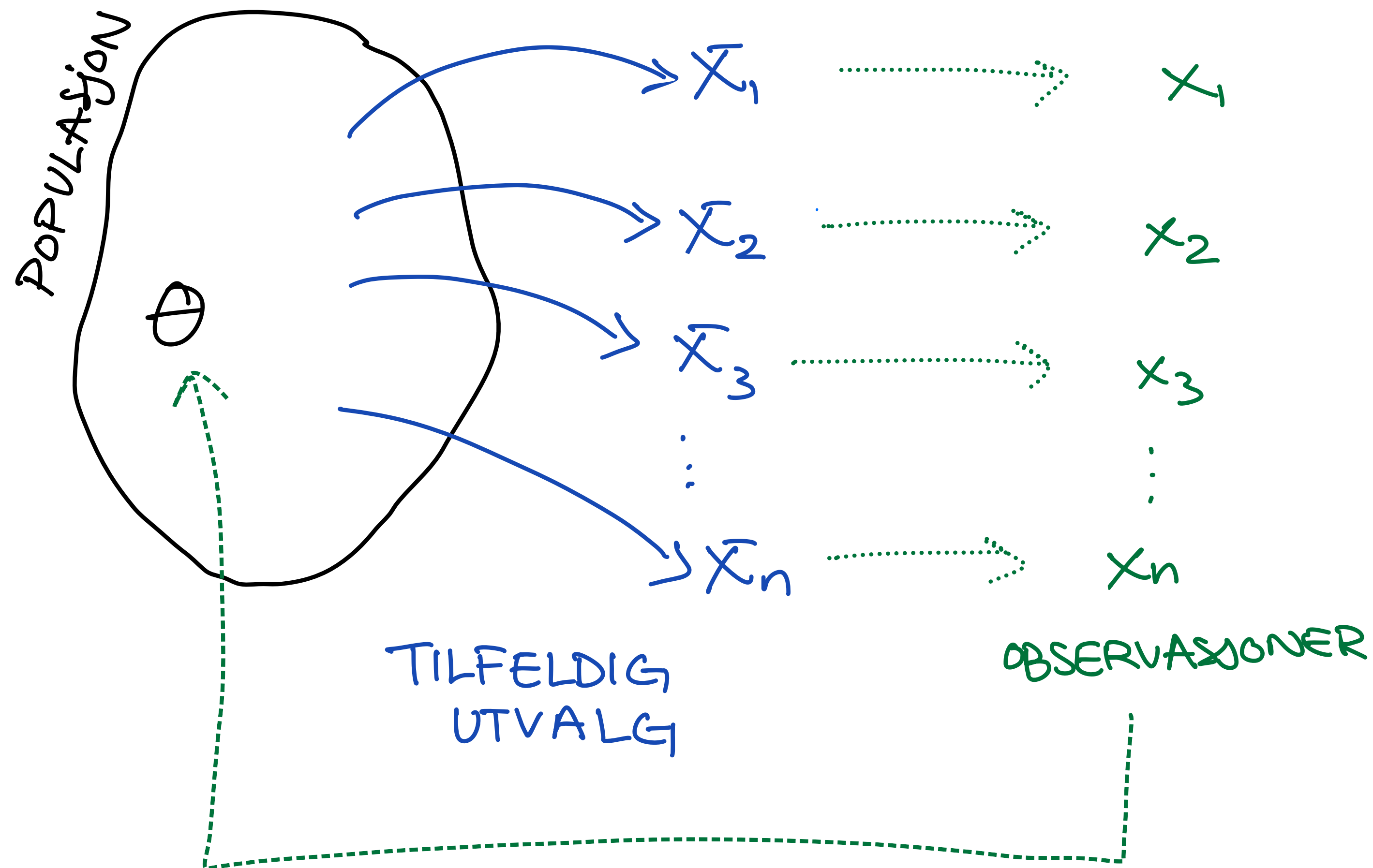
TILFELDIG
UTVALG

↖ Stokastiske variabler
som representerer n
uavhengige trekninger/målinger
fra en stor populasjon

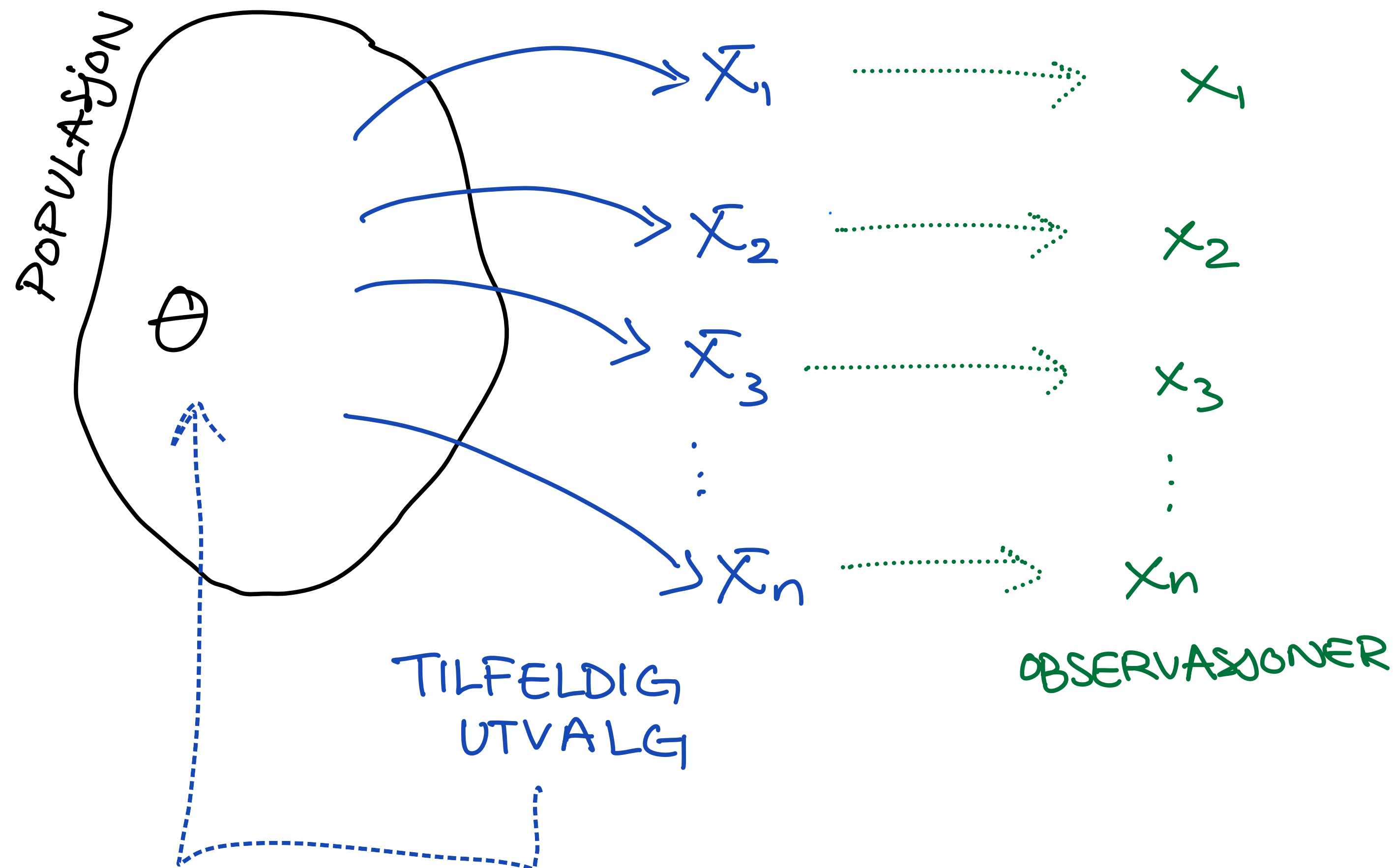




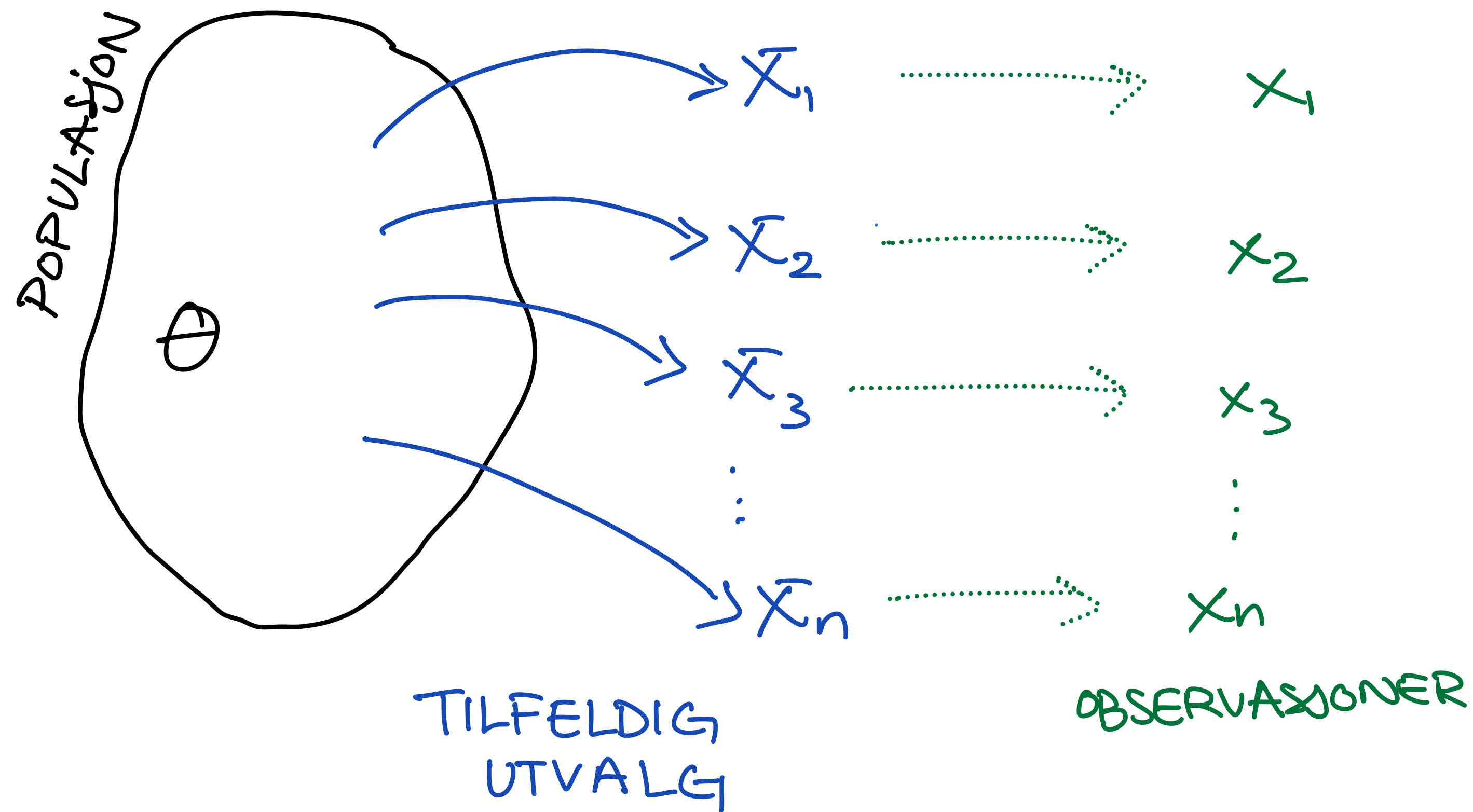
↑
de faktiske tallverdiene vi
får når vi gjør det
stokastiske forsøket
"trekke n fra populasjon"



MÅL 1: Estimere egenskap θ i populasjonen
basert på n observasjoner



MÅL 2: Kritisk vurdering av estimatet basert på egenskapene til det tilfeldige utvalget (og estimatorer basert på dette)



Hva er Θ ?

- Et populasjonsgjennomsnitt
- En andel
- En rate

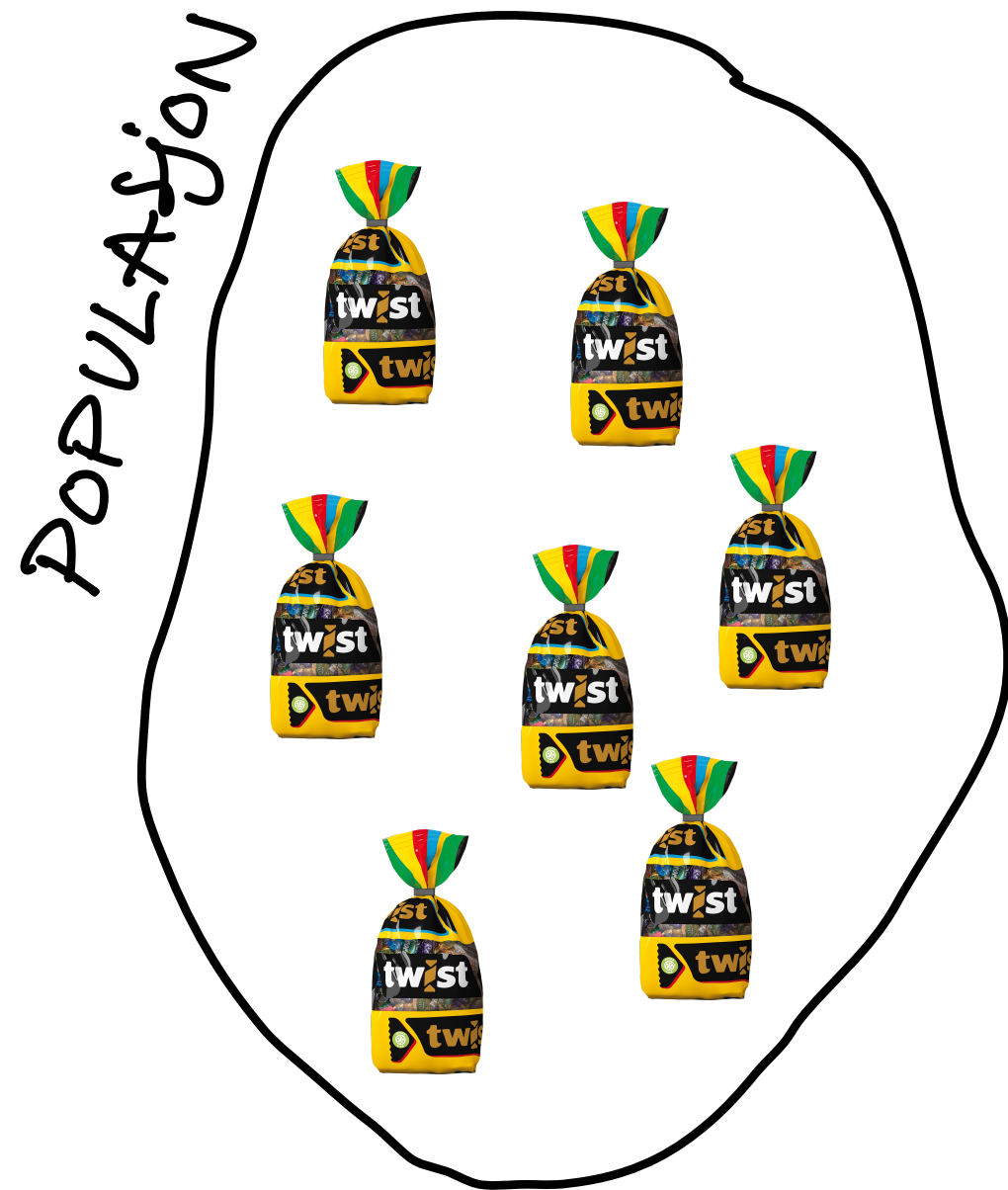
I dag skal vi...

1. Estimere et populasjons-gjennomsnitt
2. Estimere en populasjons-andel
3. Regne ut konfidensintervaller

- Hva er gjennomsnittsverdien av én bit?
- Og hva er variasjonen / standardavviket?
- Hvor stor andel er av typen banan?

Svarene våre skal være godt fundert i statistisk teori om estimering og konfidensintervaller





Populasjon:
alle twist fra fabrikk

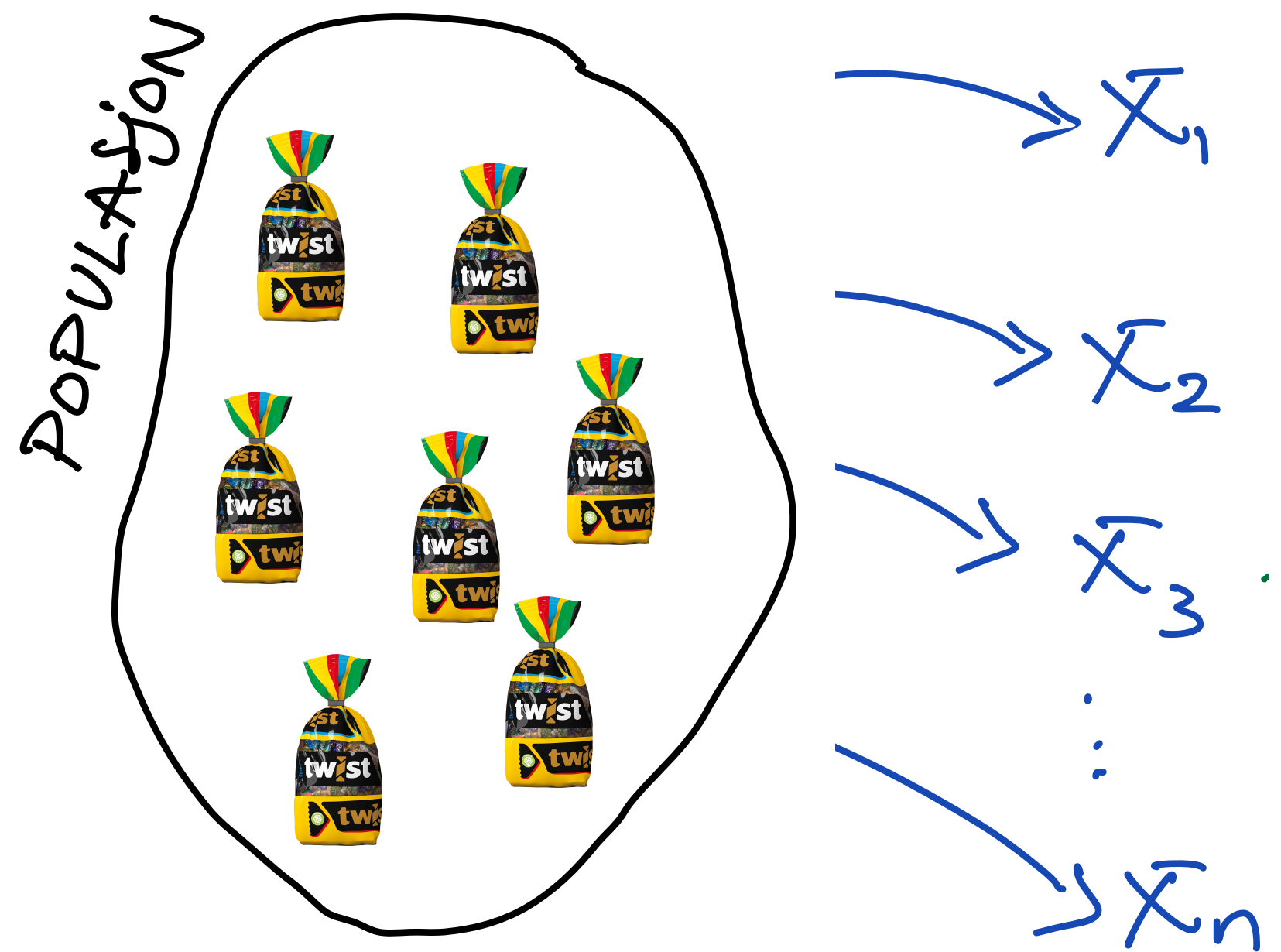
θ = populasjonsgjennomsnitt

La X representere vekten (i gram)
til en tilfeldig trukket twist-bit (med papir).

- a) Hvorfor er det rimelig å tenke seg at $E(X) = \theta$?
- b) Hva slags variasjon/støy representerer $SD(X) = \sigma$?

Mentimeter!

Over til en litt mer vitenskapelig
fundert metode for gjetting...
(estimering)



θ = populasjons-
gjennomsnitt

TILFELDIG
UTVALG

Trekke 20
twist
så tilfeldig som
mulig



$x_1 = 10\text{ g}$



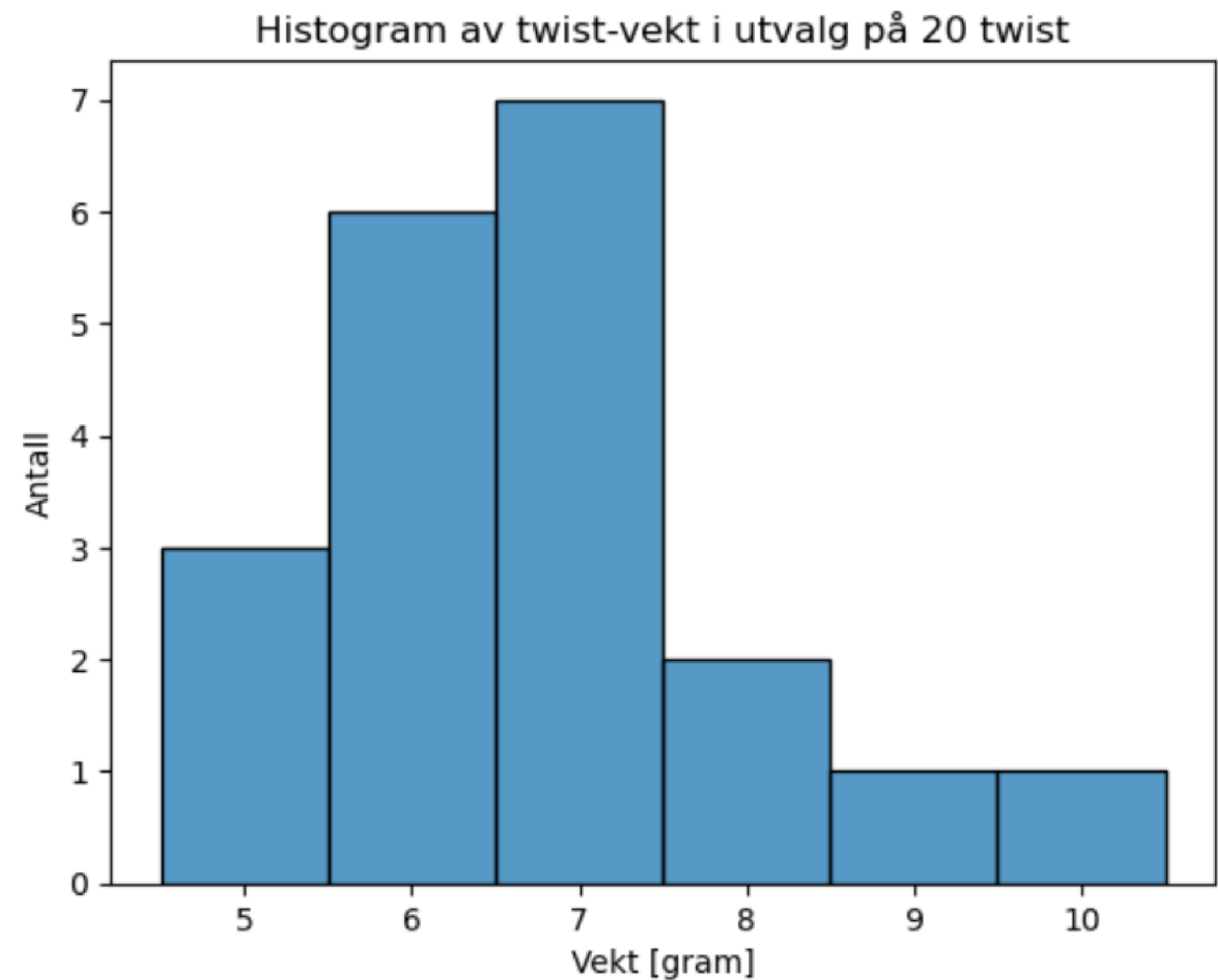
$x_2 = 7\text{ g}$



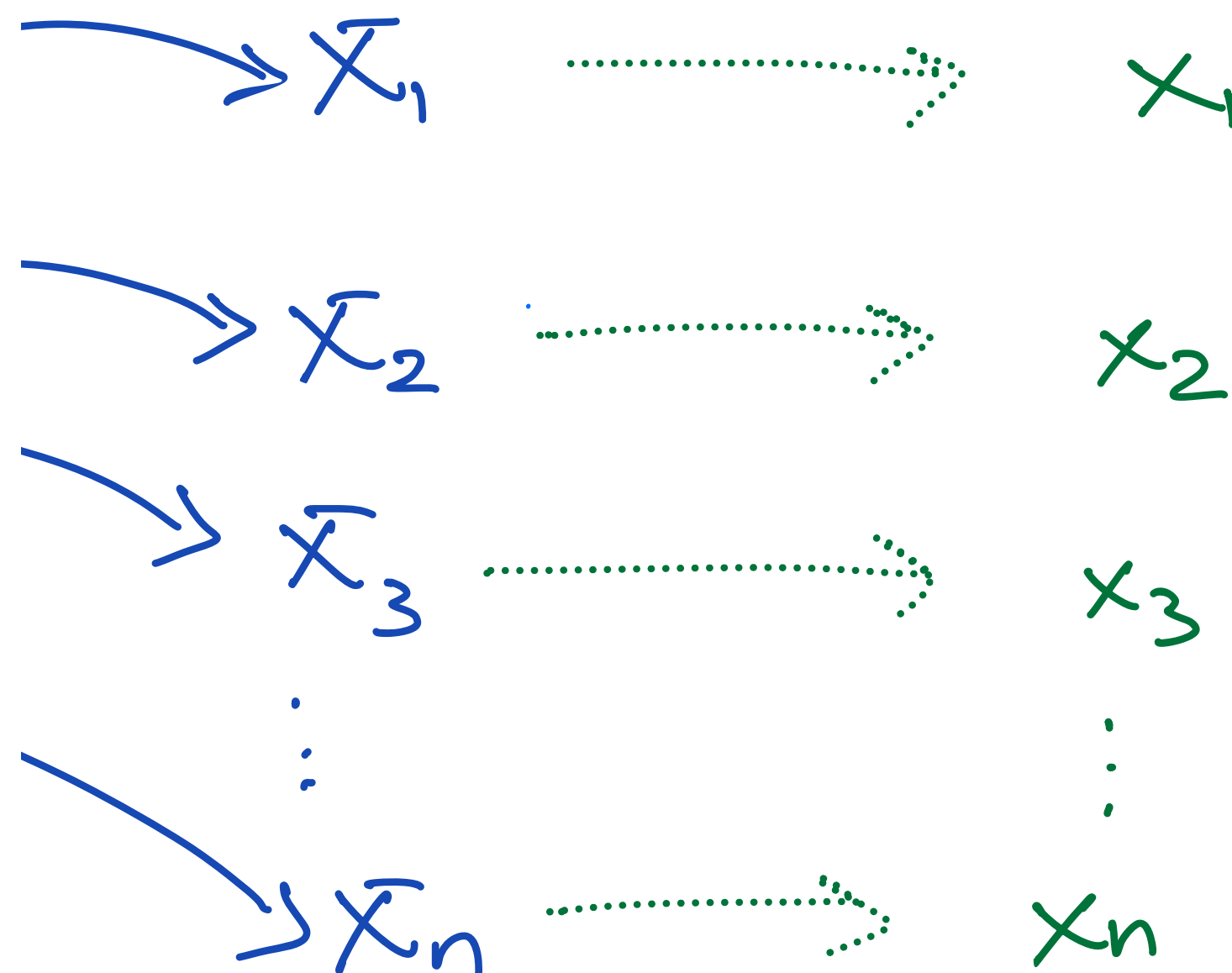
$x_3 = 6\text{ g}$

OSV...

Chocolate toffee $x_1 = 10g$	Japp $x_6 = 7g$	Lakris $x_{12} = 5g$	Daim $x_{17} = 5g$
Eclairs $x_2 = 7g$	Japp $x_7 = 6g$	Fransk nougat $x_{13} = 5g$	Golden toffee $x_{18} = 7g$
Daim $x_3 = 6g$	Cocos $x_8 = 6g$	F.noug $x_{14} = 6g$	Eclairs $x_{19} = 7g$
Banan $x_4 = 8g$	Eclairs $x_9 = 6g$	Caramel $x_{15} = 8g$	Nougat $x_{20} = 7g$
Lakris $x_5 = 6g$	Choc. toff $x_{10} = 9g$	Nougat visp $x_{16} = 7g$	
	Banan $x_{11} = 7g$		



Gjennomsnitt: 6.75 g
 Empirisk standardavvik: 1.29 g



θ = populasjons-
gjennomsnitt

TILFELDIG
UTVALG

OBSERVASJONER

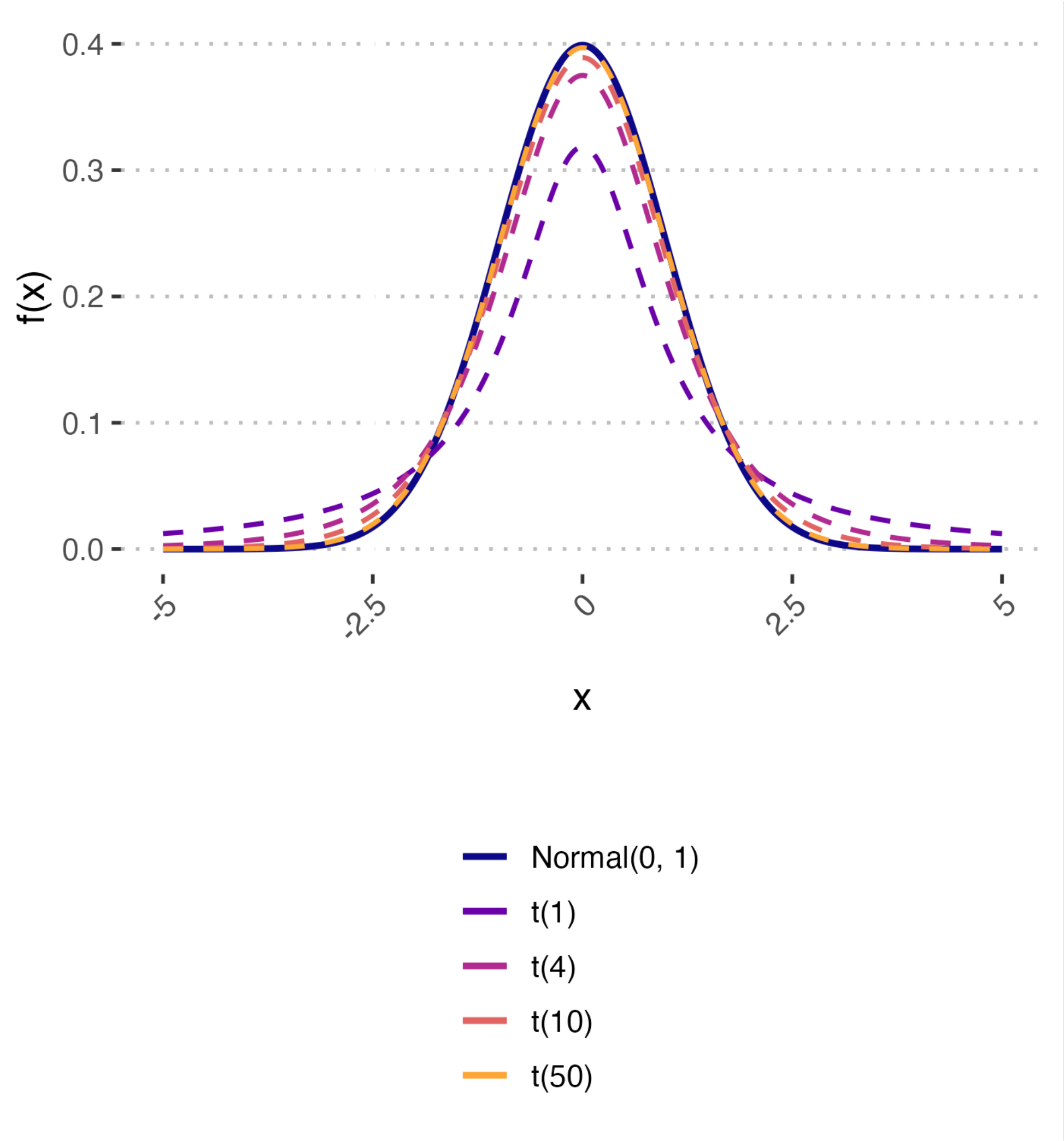
Hvor "god" er gjetningen $\hat{\theta} = 6.75$?

Kan vi supplere med et intervall som
sier noe om hvor på tallinja det er rimelig å
tenke seg at populasjonssnittet θ befinner seg?

Kritiske verdier i standard normalfordelingen

$$P(Z > z_\alpha) = \alpha$$

α	z_α
0.20	0.842
0.10	1.282
0.05	1.645
0.025	1.960
0.02	2.054
0.01	2.326
0.005	2.576
0.001	3.090
0.0005	3.291
0.0001	3.719



Kritiske verdier i t -fordelingen.

$$P(T > t_{\alpha,\nu}) = \alpha$$

ν	α							
	0.2	0.1	0.05	0.025	0.01	0.005	0.001	0.0005
1	1.376	3.078	6.314	12.706	31.821	63.657	318.309	636.619
2	1.061	1.886	2.920	4.303	6.965	9.925	22.327	31.599
3	0.978	1.638	2.353	3.182	4.541	5.841	10.215	12.924
4	0.941	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	0.920	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	0.906	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	0.896	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	0.889	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	0.883	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	0.879	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	0.876	1.363	1.796	2.201	2.718	3.106	4.025	4.437
19	0.861	1.328	1.729	2.093	2.539	2.861	3.579	3.883

- Hva er gjennomsnittsvekten av én bit?
- Og hva er variasjonen / standardavviket?
- **Hvor stor andel er av typen banan?**

Svarene våre skal være godt fundert i statistisk teori om estimering og konfidensintervaller



Chocolate toffee
 $x_1 = 10g$

Eclairs
 $x_2 = 7g$

Daim
 $x_3 = 6g$

Banan
 $x_4 = 8g$

Lakris
 $x_5 = 6g$

Fapp
 $x_6 = 7g$

Fapp
 $x_7 = 6g$

Cocos
 $x_8 = 6g$

Eclairs
 $x_9 = 6g$

Choc. toff
 $x_{10} = 9g$

Banan
 $x_{11} = 7g$

Lakris
 $x_{12} = 5g$

Fransh
nougat
 $x_{13} = 5g$

F. noug
 $x_{14} = 6g$

Caramel
 $x_{15} = 8g$

Nougat
wisp
 $x_{16} = 7g$

Daim
 $x_{17} = 5g$

Golden
toffee
 $x_{18} = 7g$

Eclairs
 $x_{19} = 7g$

Nougat
wisp
 $x_{20} = 7g$

Formelark

5 Punktestimering

5.1 Forventningsverdi og varians

For et tilfeldig utvalg $X_1, X_2, \dots, X_n$ der $E(X_i) = \mu$ og $\text{Var}(X_i) = \sigma^2$, for $i = 1, \dots, n$, så er

Estimator for forventningsverdien (μ):

$$\hat{\mu} = \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

Estimator for varians (σ^2):

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

Estimator for standardavviket (σ):

$$S = \sqrt{S^2}$$

5.2 Sannsynligheten p i binomisk fordeling

Dersom X teller antall suksesser i en binomisk forsøksrekke av n forsøk så er estimator for sannsynligheten for suksess (p):

$$\hat{p} = \frac{X}{n}$$

5.3 Raten λ i poissonfordelingen

Dersom X teller antall hendelser i en poissonprosess med rate λ over et intervall/område av lengde/størrelse t , så er estimator for raten (λ):

$$\hat{\lambda} = \frac{X}{t}$$

Formelark

6 Konfidensintervall

6.1 Forventningsverdi μ

For et tilfeldig utvalg $X_1, X_2, \dots, X_n$, $X_i \sim N(\mu, \sigma)$, $i = 1, \dots, n$, der standardavviket σ er *kjent*, så er

$$\left[\bar{X} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}, \quad \bar{X} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \right]$$

et $100(1 - \alpha)\%$ konfidensintervall for μ .

Dersom standardavviket σ er *ukjent*, så er

$$\left[\bar{X} - t_{\alpha/2, n-1} \cdot \frac{S}{\sqrt{n}}, \quad \bar{X} + t_{\alpha/2, n-1} \cdot \frac{S}{\sqrt{n}} \right]$$

et $100(1 - \alpha)\%$ konfidensintervall for μ .

6.2 Sannsynligheten p i binomisk fordeling

Under forutsetningen om at $n\hat{p}(1 - \hat{p}) \geq 5$, så er

$$\left[\hat{p} - z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}, \quad \hat{p} + z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}} \right]$$

et tilnærmet $100(1 - \alpha)\%$ konfidensintervall for p .

6.3 Raten λ i poissonfordeling

Under forutsetningen om at $\hat{\lambda}t > 10$, så er

$$\left[\hat{\lambda} - z_{\alpha/2} \cdot \sqrt{\frac{\hat{\lambda}}{t}}, \quad \hat{\lambda} + z_{\alpha/2} \cdot \sqrt{\frac{\hat{\lambda}}{t}} \right]$$

et tilnærmet $100(1 - \alpha)\%$ konfidensintervall for λ .

Tips

- Jobb gjennom notatboken om Estimering og konfidensintervaller under Uke 7 i jupyterhuben for å få bedre forståelse for oppførselen til konfidensintervaller og hva det betyr at grensene i et konfidensintervall er stokastiske.