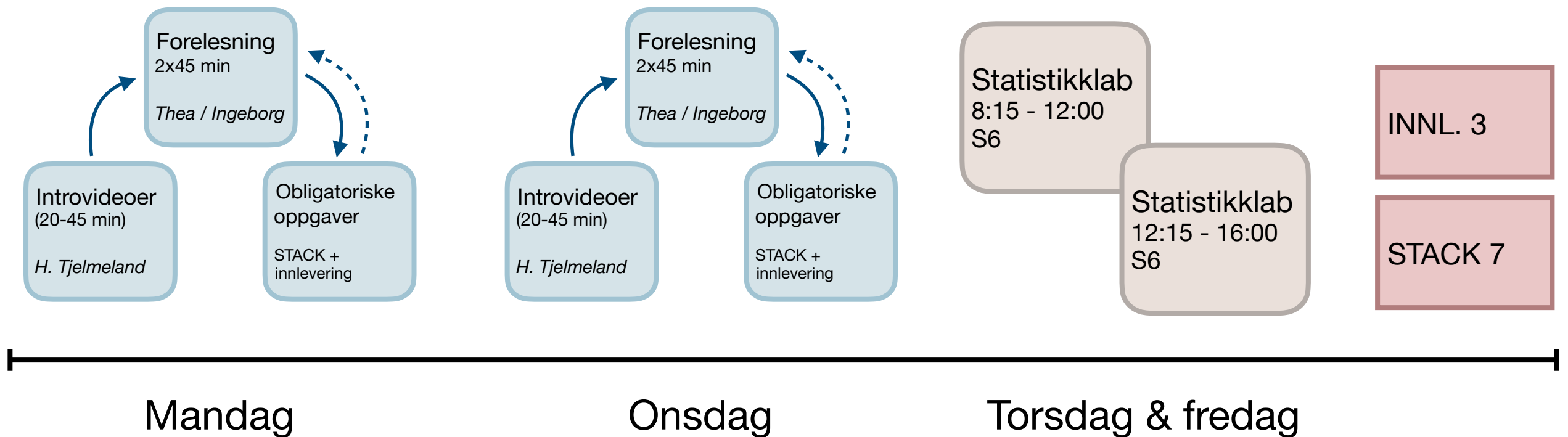


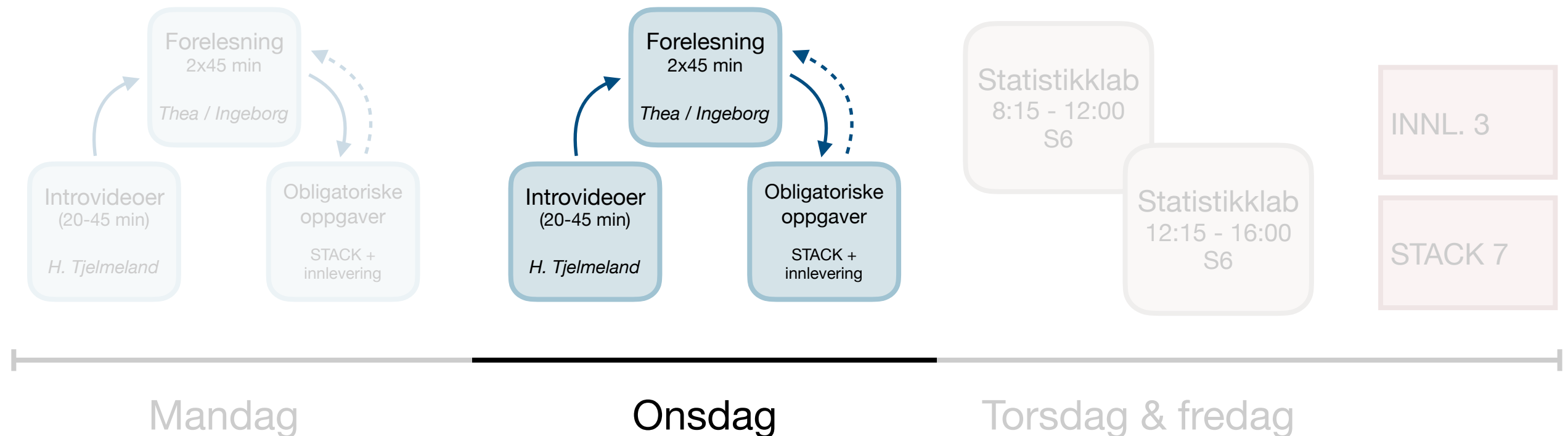
TMA4245 Statistikk

Uke 8 - onsdag

Siste forelesning
i Blokk 1!



Uke 8 (mandag 19. februar - fredag 23. februar)



Introvideoer

Sentralgrenseteoremet

TMA4240/TMA4245 Statistikk

Håkon Tjelmeland
Institutt for matematiske fag
Norges teknisk-naturvitenskapelige universitet

Normalfordeling som tilnærming til binomisk fordeling

TMA4240/TMA4245 Statistikk

Håkon Tjelmeland
Institutt for matematiske fag
Norges teknisk-naturvitenskapelige universitet

Oppgaver

STACK 7: Opppg. 4-6

INNL. 3: Opppg. 5

Teorem (Sentralgranseteoremet)

La $X_1, X_2, \dots, X_n$ være uavhengige og identisk fordelte variabler med forventningsverdi $E[X_i] = \mu$ og varians $\text{Var}[X_i] = \sigma^2$. La

$$Z = \frac{\bar{X} - \mu}{\sqrt{\frac{\sigma^2}{n}}},$$

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

der $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ er gjennomsnittet av X_i -ene. Da vil sannsynlighetsfordelingen til Z konvergere mot en standard normalfordeling når $n \rightarrow \infty$.

God approksimasjon hvis $n \geq 30$

Gjennomsnittet $\bar{X}$ blir normalfordelt, og $X_1, \dots, X_n$ har fremdeles sin originale fordeling!

Å simulere fra $\bar{X}$ betyr å simulere $X_1, \dots, X_n$ mange ganger, altså å repetere:

1. $X_i \sim f(x_i)$, $i = 1, \dots, n$
2. Regn ut $\bar{X}$

til vi har f.eks. 1000 simulerte gjennomsnitt.

Oppgave 1

En elektronikkbedrift produserer en elektrisk komponent. En fjerdedel av komponentene de lager har produksjonsfeil. Av disse har tre av fem defekter som kan repareres, mens to av fem må vrakes. Bedriften taper 100 kr per komponent som kan repareres, og 700 per komponent som må vrakes.

- a) Vi lar X betegne tapet på en enkelt komponent. Finn $E(X)$ og $\text{Var}(X)$.
- b) På en dag produserer bedriften 100 enheter. Hva er sannsynligheten for at totalt tap overstiger 10 000 kr?

$$E(X) = \sum_x xP(X = x)$$

$$\text{Var}(X) = \sum_x x^2P(X = x) - E(X)^2$$

Teorem (Sentralgranseteoremet)

La $X_1, X_2, \dots, X_n$ være uavhengige og identisk fordelte variabler med forventningsverdi $E[X_i] = \mu$ og varians $\text{Var}[X_i] = \sigma^2$. La

$$Z = \frac{\bar{X} - \mu}{\sqrt{\frac{\sigma^2}{n}}},$$

der $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ er gjennomsnittet av X_i -ene. Da vil sannsynlighetsfordelingen til Z konvergere mot en standard normalfordeling når $n \rightarrow \infty$.

God approksimasjon hvis $n \geq 30$

Standard normalfordeling

$$\Phi(z) = P(Z \leq z)$$

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441

Regneregler for forventning og varians

$X_i \sim N(\mu_i, \sigma_i^2)$ og uavh., da er

$$\sum_{i=1}^n a_i X_i + b \sim N\left(\sum_{i=1}^n a_i \mu_i + b, \sum_{i=1}^n a_i^2 \sigma_i^2\right)$$

$$E[aX] = aE[X]$$

$$E[b] = b$$

$$E[X + Y] = E[X] + E[Y]$$

$$\text{Var}[aX] = a^2 \text{Var}[X]$$

$$\text{Var}[b] = 0$$

$$\text{Var}[X + Y] = \text{Var}[X] + \text{Var}[Y] + 2\text{Cox}[X, Y]$$

Oppgave 2

Verdens største laksemerde inneholder en og en halv million laks. Vi antar at i denne har 12 prosent av laksene lus. Vi plukker ut fem hundre tilfeldige laks.

- a) Hvilken fordeling har antall laks med lus?
- b) Hva er forventet antall laks med lus?
- c) Hva er sannsynligheten for at maksimalt femti av dem har lus?

Binomisk fordeling

$$f(x) = b(x; n, p) = \binom{n}{x} p^x (1 - p)^{n-x}, \quad x = 0, 1, \dots, n.$$

$$E(X) = np, \quad \text{Var}(X) = np(1 - p), \quad M_X(t) = (pe^t + 1 - p)^n.$$

Normaltilnærming av binomisk fordeling

sentralgrense-

- ★ Praktisk konsekvens av teoremet: Når n er stor (nok) har vi

$$Z = \frac{\bar{X} - \mu}{\sqrt{\frac{\sigma^2}{n}}} \approx N(0, 1) \Leftrightarrow \sum_{i=1}^n X_i \approx N(n\mu, n\sigma^2)$$

- ★ La $Y \sim \text{binomisk}(n, p)$

- skriver $Y = \sum_{i=1}^n X_i$, der $X_i = \begin{cases} 1 & \text{hvis suksess i } i\text{-te forsøk,} \\ 0 & \text{ellers.} \end{cases}$
- $\mu = E[X_i] = 0 \cdot (1 - p) + 1 \cdot p = p$
- $\sigma^2 = \text{Var}[X_i] = E[X_i^2] - E[X_i]^2 = \dots = p(1 - p)$
- hvis n er stor (nok) gir sentralgrenseteoremet at

$$Y = \sum_{i=1}^n X_i \approx N(np, np(1 - p))$$

- tommelfingerregel: god approksimasjon hvis $np \geq 5$ og $n(1 - p) \geq 5$

Selve fordelingen blir tilnærmet, ikke “bare” gjennomsnittet/summen!

Oppgave 2

Verdens største laksemerde inneholder en og en halv million laks. Vi antar at i denne har 12 prosent av laksene lus. Vi plukker ut fem hundre tilfeldige laks.

- a) Hvilken fordeling har antall laks med lus?
- b) Hva er forventet antall laks med lus?
- c) Hva er sannsynligheten for at maksimalt femti av dem har lus?

$$Y = \sum_{i=1}^n X_i \approx N(np, np(1 - p))$$

– tommelfingerregel: god approksimasjon hvis $np \geq 5$ og $n(1 - p) \geq 5$

+ Kontinuitetskorreksjon: legg på 0.5 (se introvideo)

Oppgave 3

Frøya vil måle lengden (i cm) på en skjoteledning så nøyaktig som mulig. La μ betegne den ukjente lengden. Frøya gjør 5 uavhengige målinger med en meterstokk. Vi kan anta at hver måling X_i ($i = 1, \dots, 5$) er normalfordelt med forventningsverdi μ cm og standardavvik 1 cm. Senere får Frøya låne en lasermåler som hun bruker til å gjøre én måling. Vi kan anta at Y er normalfordelt med forventningsverdi μ cm og standardavvik 0.2 cm, og at Y er uavhengig av tidligere målinger.



Funksjoner av X_i og Y

For å anslå lengden μ vurderer Frøya fem ulike metoder:

$$\hat{\mu}_1 = Y, \hat{\mu}_2 = \frac{\sum_{i=1}^5 X_i}{5}, \hat{\mu}_3 = \frac{1}{2} \left(Y + \frac{\sum_{i=1}^5 X_i}{5} \right), \hat{\mu}_4 = \frac{Y + \sum_{i=1}^5 X_i}{6} \text{ og } \hat{\mu}_5 = 5Y - \sum_{i=1}^5 X_i$$

- a) Basert på forventningsverdi, hvilke(n) metode(r) er brukbar(e)?
- b) Basert på varians, hvilken metode er nyttigst?
- c) Hvilken metode er best?

Ukas forelesninger er ferdige!

Oppgaver

STACK 7: Hele
INN. 3: Hele

Torsdag

Statistikklab
8:15 - 12:00 S6

Fredag

Statistikklab
12:15 - 16:00 S6

INN.

Frist 18:00

STACK

Frist 18:00

