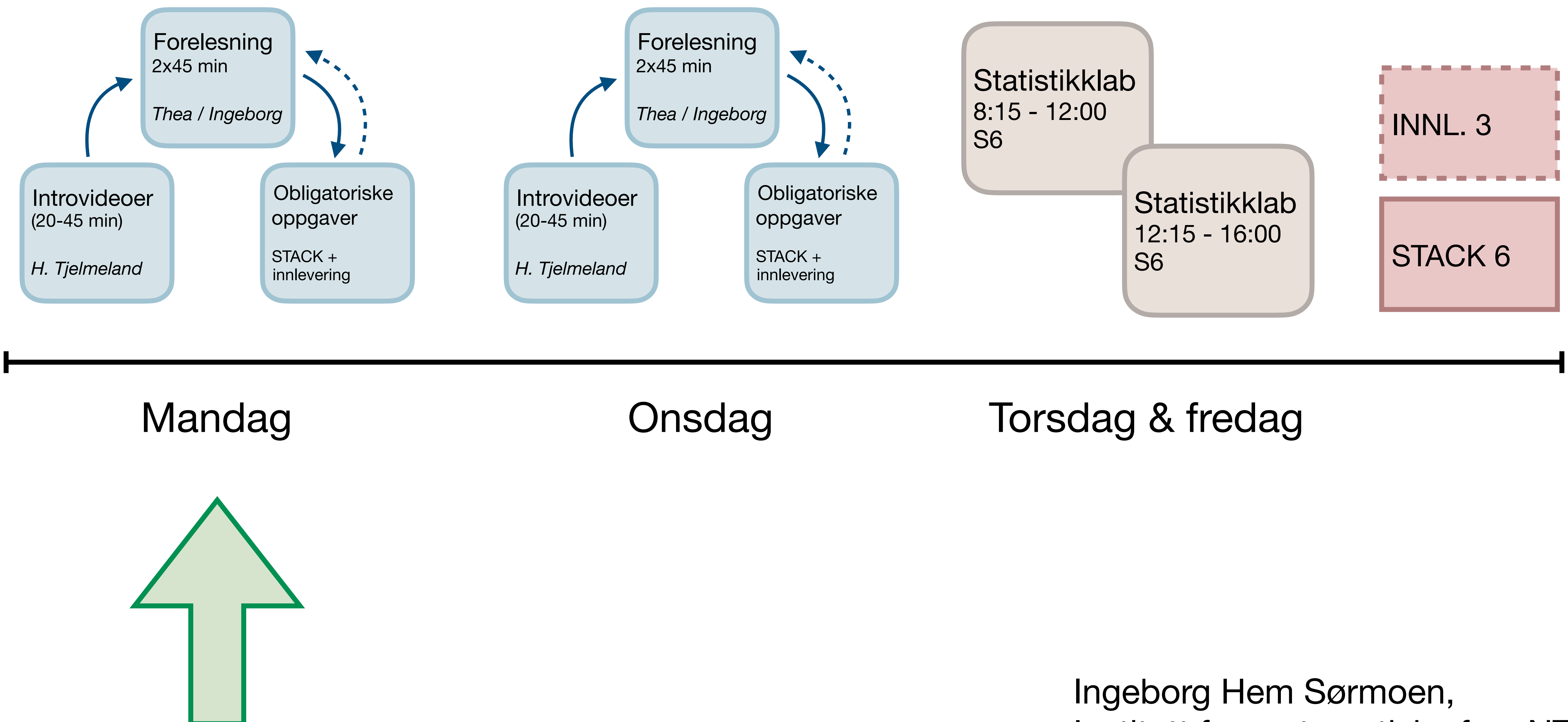
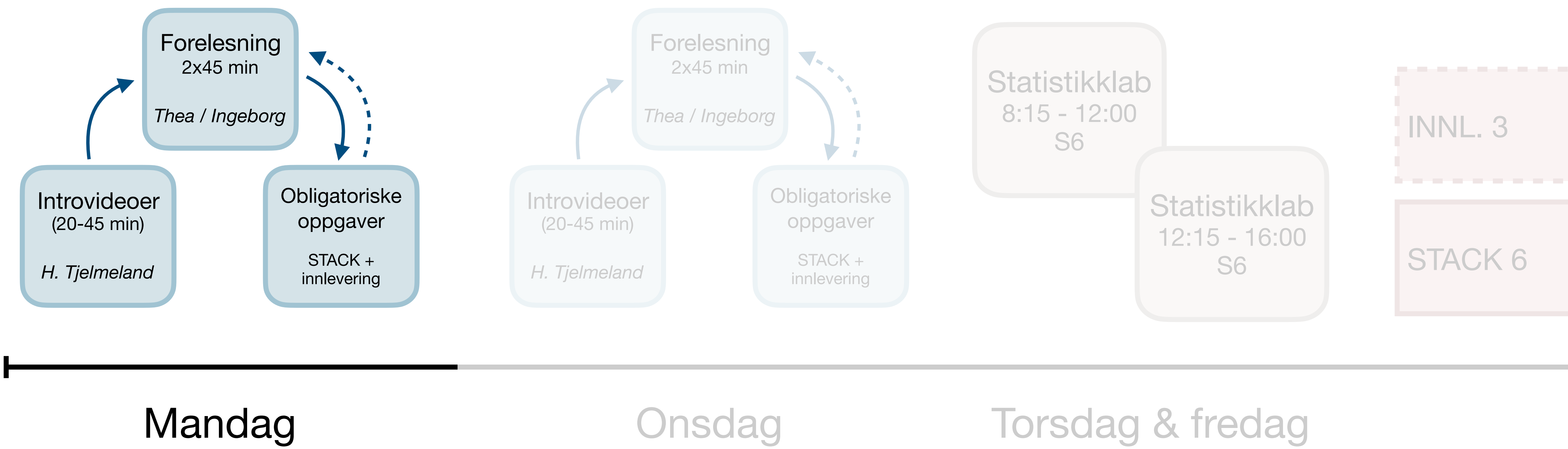
TMA4245 Statistikk

Uke 7 - mandag



Ingeborg Hem Sørmoen,
Institutt for matematiske fag, NTNU

Uke 7 (mandag 12. februar - fredag 16. februar)



Introvideoer

Eksponensialfordeling

TMA4240/TMA4245 Statistikk

Håkon Tjelmeland
Institutt for matematiske fag
Norges teknisk-naturvitenskapelige universitet

Gamma- og kjikvadrat-fordeling

TMA4240/TMA4245 Statistikk

Håkon Tjelmeland
Institutt for matematiske fag
...naturvitenskapelige universitet

Student *t*-fordeling

TMA4240/TMA4245 Statistikk

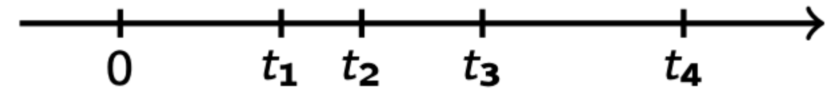
Håkon Tjelmeland
Institutt for matematiske fag
Norges teknisk-naturvitenskapelige universitet

Oppgaver

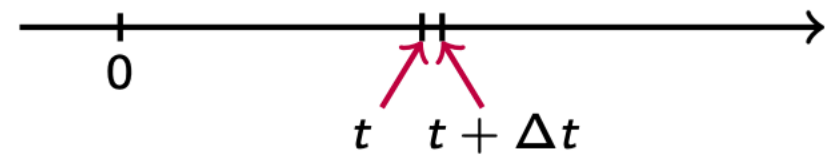
STACK 6: Oppg. 1-3
INNL. 3: Oppg. 3

Definisjon (Poissonprosess)

En prosess $\{N(t); t \geq 0\}$ *kalles in poissonprosess med intensitet* λ *hvis følgende krav er oppfylt*



1. $N(0) = 0$.
2. *For* $0 \leq t_1 < t_2 \leq t_3 < t_4$ *er* $N(t_2) - N(t_1)$ *og* $N(t_4) - N(t_3)$ *uavhengige.*
3. *For* $t \geq 0$ *og* $\Delta t > 0$ *har vi at*



$$P(N(t + \Delta t) - N(t) = 1) = \lambda \Delta t + o(\Delta t),$$
$$P(N(t + \Delta t) - N(t) \geq 2) = o(\Delta t),$$

der hver $o(\Delta t)$ *angir en funksjon som oppfyller*

$$\lim_{\Delta t \to 0} \frac{o(\Delta t)}{\Delta t} = 0.$$

Definisjon (Poissonfordeling)

La $N(t)$ *være en poissonprosess med intensitet* λ. *Fikser en verdi* $t > 0$ *og la* $X = N(t)$. *Vi sier da at* X *er poissonfordelt med parameter* λt.

Antall hendelser

Definisjon

La $N(t), t \geq 0$ *være en poissonprosess med intensitet* λ, *og la* X *være tidspunktet for den første hendelsen i denne prosessen, dvs*

$$X = \min\{t : N(t) = 1\}.$$

Vi sier da at X *er eksponensialfordelt med parameter* λ.

Tid mellom hendelser

Eksponentialfordeling

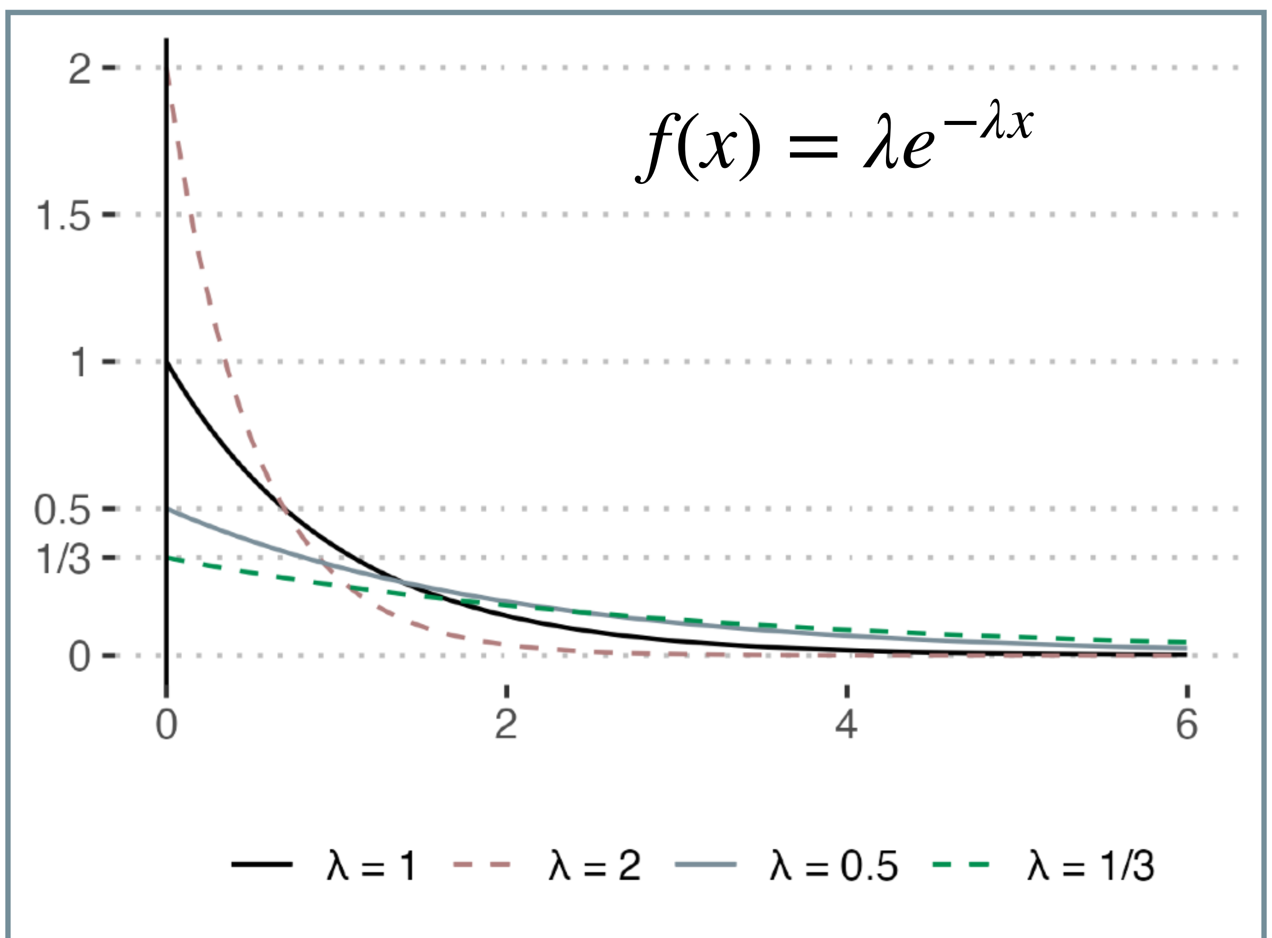
$$F(x) = \int_0^\infty \frac{1}{\beta} e^{-\frac{1}{\beta}x} dx = 1 - e^{-\frac{1}{\beta}x}$$

$$f(x) = \frac{1}{\beta} e^{-x/\beta}, \quad x \geq 0.$$

$$\mathrm{E}(X) = \beta, \quad \mathrm{Var}(X) = \beta^2, \quad M_X(t) = \frac{1}{1-\beta t} \text{ for } t < \frac{1}{\beta}.$$

Kommentar: Ofte vanlig å bruke parameteren $\lambda = 1/\beta$.

Har ikke egen tabell!



Hvis antall hendelser er Poissonfordelt med intensitet λ, så er tilhørende tid mellom hendelsene eksponensialfordelt med forventning $1/\lambda$.

Oppgave 1

Oppgave 2 Levetid

Anta at levetiden T, målt i timer (behøver ikke være et heltall), til en tilfeldig valgt elektronisk komponent er eksponentialfordelt med sannsynlighetstetthet

$$f(t;\beta) = \begin{cases} \frac{1}{\beta} e^{-t/\beta} & t \geq 0 \\ 0 & t < 0, \end{cases}$$

$$f(x) = \frac{1}{\beta} e^{-x/\beta}, \quad x \geq 0.$$

$$\mathrm{E}(X) = \beta, \quad \mathrm{Var}(X) = \beta^2,$$

der $\beta > 0$ er en parameter.

a) Anta (kun i dette punktet) at $\beta = 30$.

i) Vis at den kumulative fordelingsfunksjonen til T er gitt som

$$P(T \leq t) = \begin{cases} 1 - e^{-t/30} & t \geq 0 \\ 0 & t < 0. \end{cases}$$

$$F(t) = P(T \leq t) = \int_{-\infty}^{t} f(x)dx$$

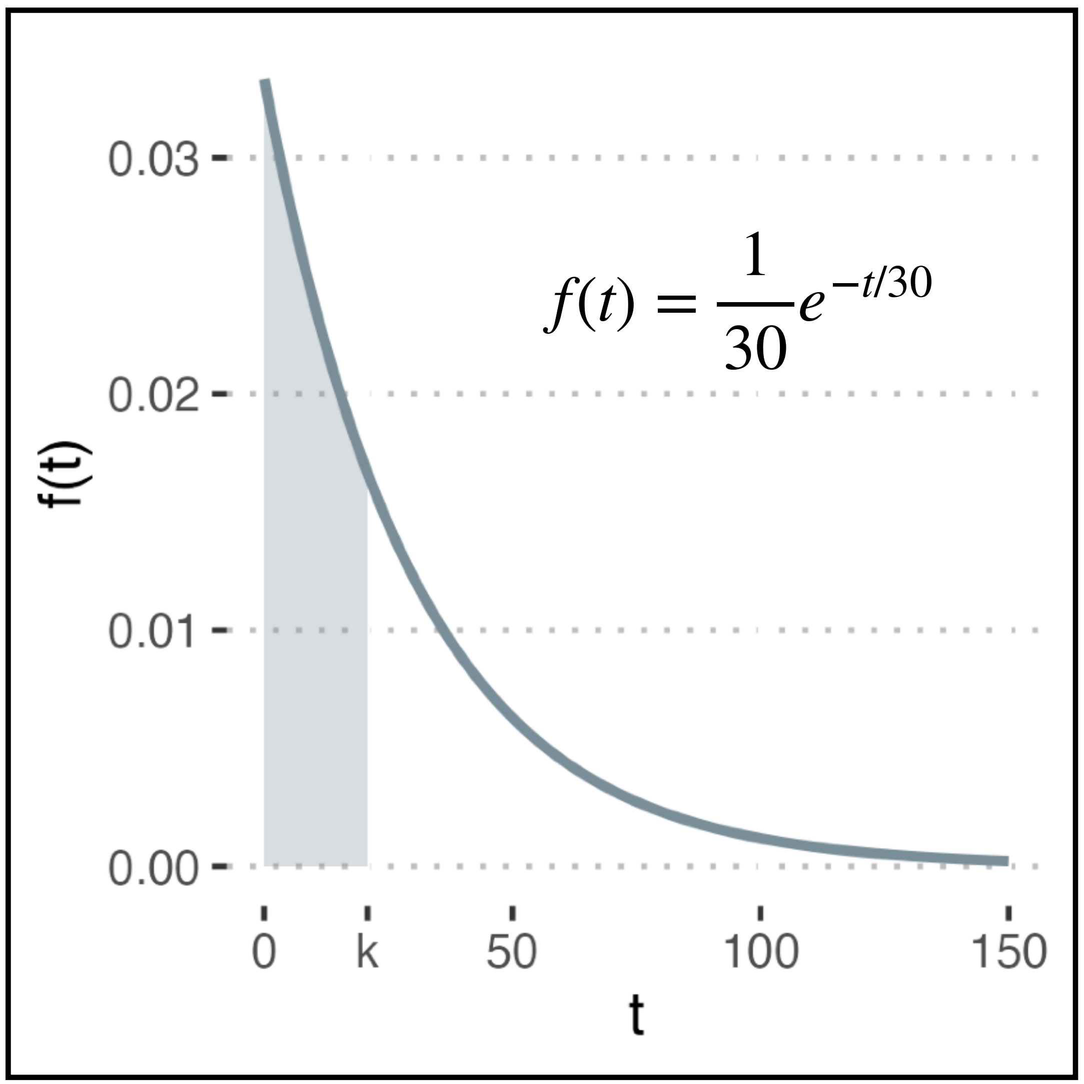
ii) Finn $P(T < 20)$ og $P(T < 20 \cup T > 40)$.

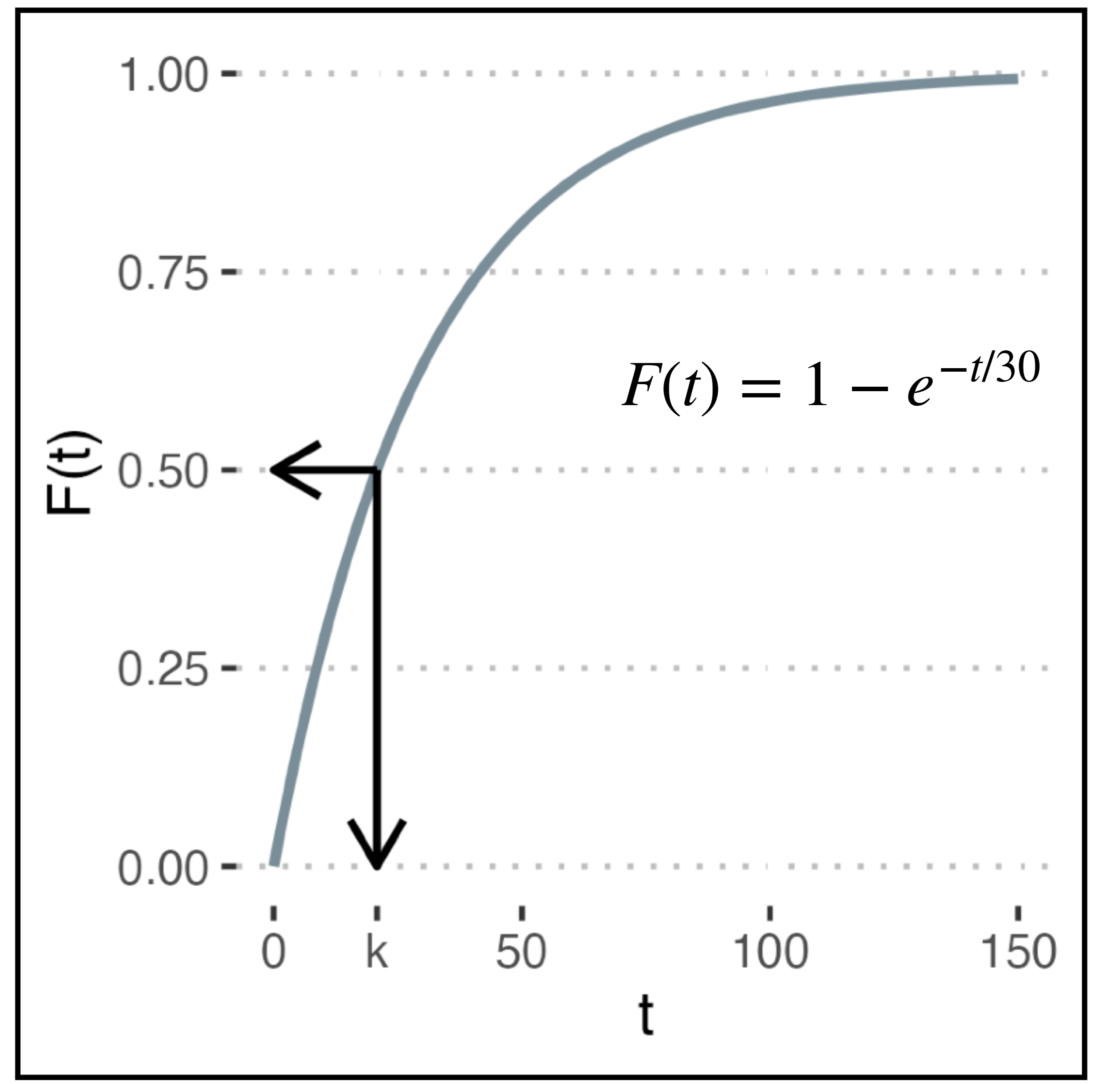
iii) Finn medianen til T, det vil si k slik at $P(T \leq k) = 0.5$.

Eksamen august 2019

Oppgave 1

Median i eksponensialfordelingen





Oppgave 1

Oppgave 2 Levetid

Anta at levetiden T, målt i timer (behøver ikke være et heltall), til en tilfeldig valgt elektronisk komponent er eksponentialfordelt med sannsynlighetstetthet

$$f(t;\beta) = \begin{cases} \frac{1}{\beta}e^{-t/\beta} & t \geq 0 \\ 0 & t < 0, \end{cases} \tag{1}$$

der $\beta > 0$ er en parameter.

b) Vis at eksponentialfordelingen er uten hukommelse, det vil si

$$P(T \geq t + s \mid T \geq t) = P(T \geq s)$$

for $s \geq 0$.

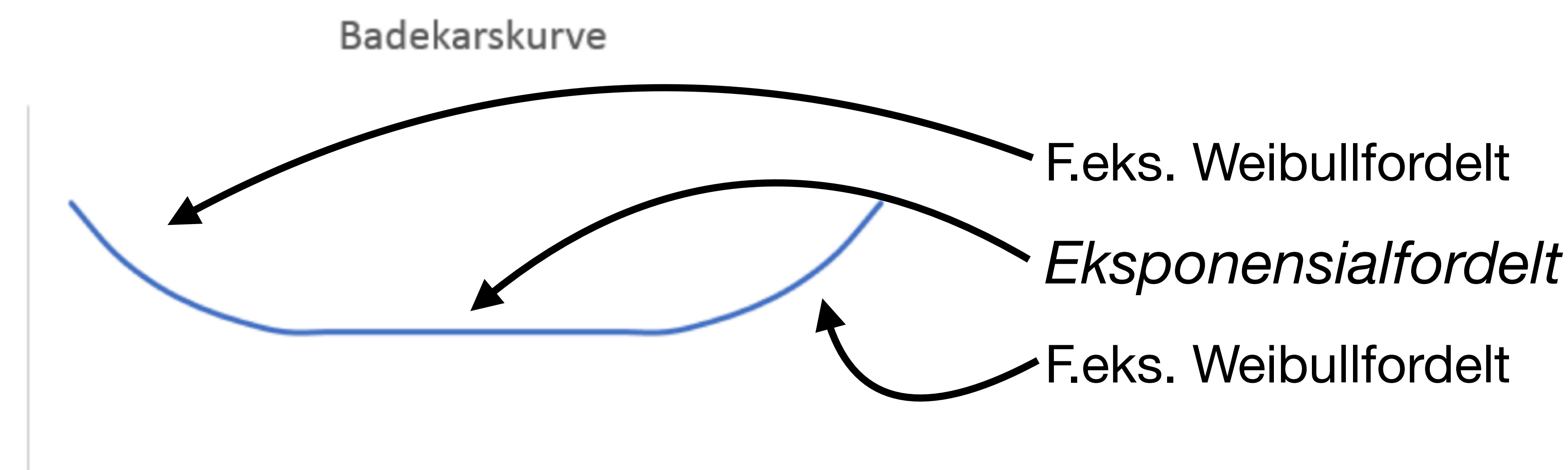
Eksamen august 2019

Betinget sannsynlighet:

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

Badekarskurven

Den velkjente badekarskurven er hyppig brukt som illustrasjon på et typisk sviktmønster. Og for folk i vedlikeholdsbransjen er det er lett å kjenne igjen mønsteret: Først avtar antall svikt etter å ha vært høyt rett etter installasjonsfasen, for så å bli etterfulgt av en lang, stabil fase. Til slutt går kurven oppover igjen og avspeiler et økende antall problemer typisk forbundet med økende alder på utstyret. Når de 3 ulike sviktmønstre settes sammen, fås en badekarskurve som i Figur 1.



Figur 1

Badekarskurven benyttes ofte som en fin illustrasjon på hvorfor tidsbaserte utskiftninger eller overhalinger ikke er velegnet i kampen for å forbedre driftssikkerheten. For når mesteparten av kurven er konstant eller fallende, er det heller ikke noe poeng med periodiske utskiftninger eller overhalinger i samme tidsrommet. Slik presenteres kurven ofte, men så enkelt er det ikke.

Kilde: pålitelighet.no/badekarskurven

$F(x)$ har et ikke et enkelt uttrykk

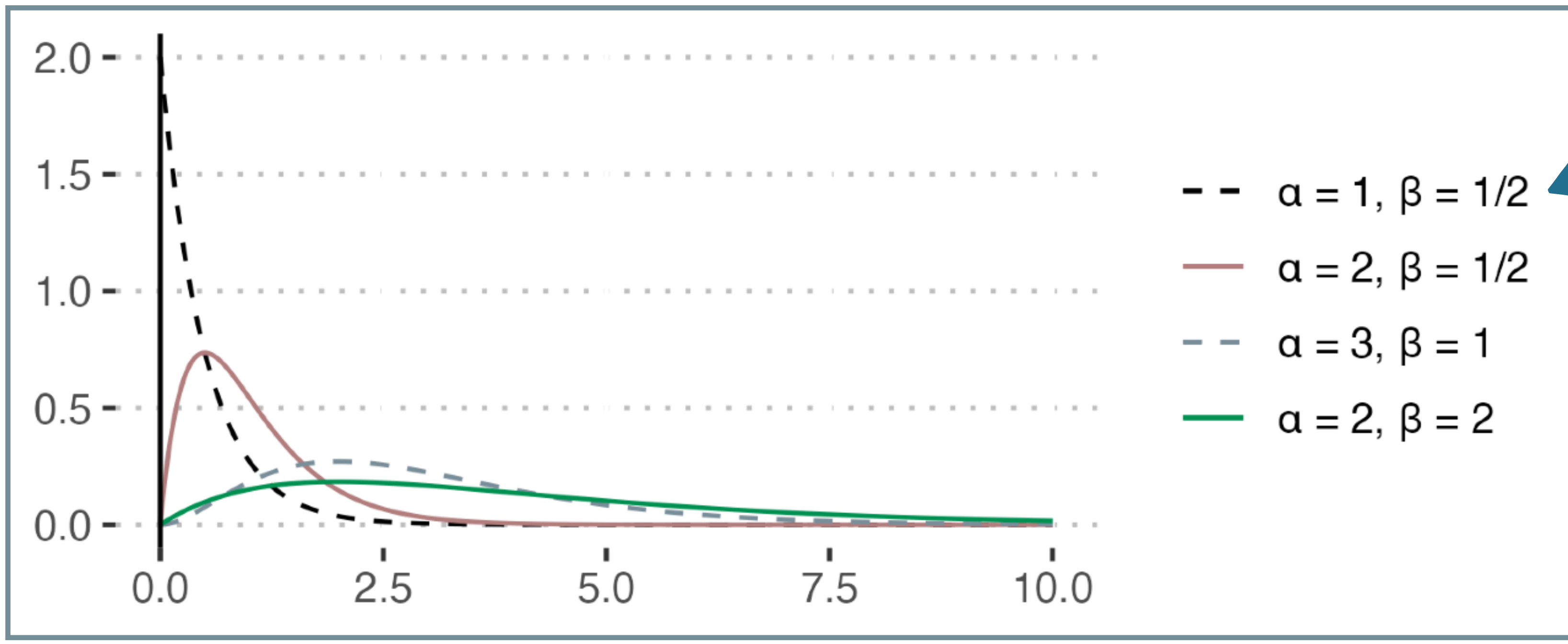
Gammafordeling

$$f(x) = \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta}, \quad x \geq 0.$$

$$\mathrm{E}(X) = \alpha\beta, \quad \mathrm{Var}(X) = \alpha\beta^2, \quad M_X(t) = \left(\frac{1}{1-\beta t}\right)^\alpha \text{ for } t < \frac{1}{\beta}.$$

Spesialtilfeller: $\alpha = 1$ gir eksponensialfordelingen.

Har ikke egen tabell!



eksponensialfordeling

Teorem

La $N(t); t \geq 0$ være en poissonprosess med intensitet λ, og la X være tidspunktet hvor hendelse nummer n i denne prosessen skjer. Da er X gammafordelt med parametre $\alpha = n$ og $\beta = \frac{1}{\lambda}$.

Oppgave 3a, innlevering 3

Betrakt en poissonprosess med intensitet λ. La Z_1 være tidspunkt for den første hendelsen i prosessen, la Z_2 være tidspunkt for den andre hendelsen i prosessen, og tilsvarende videre slik at Z_k er tidspunkt for hendelse nummer k i prosessen.

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{for } x \geq 0, \\ 0 & \text{ellers} \end{cases}$$

$$E[X] = \frac{1}{\lambda}$$

La $X_1 = Z_1$ og for $k = 2, 3, \dots$ la

$$X_k = Z_k - Z_{k-1}.$$

Dermed har vi altså at X_1 er tid til første hendelse i poissonprosessen, og for $k = 2, 3, \dots$ er X_k tid mellom to etterfølgende hendelser i poissonprosessen. Vi vet dermed at $X_1, X_2, \dots$ er uavhengige og ekseponensialfordelte, alle med samme parameterverdi λ.

Finn sannsynlighetene

$$P(X_1 \geq 2), \quad P(X_1 + X_2 \geq 4) \quad \text{og} \quad P(X_1 + X_2 \geq 4 | X_1 \geq 2).$$

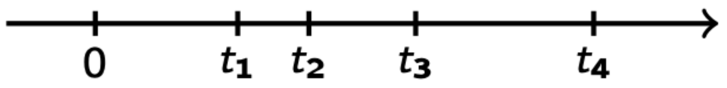
Finn først formler for sannsynlighetene som funksjon av λ, og sett deretter inn $\lambda = 0.5$ og regn ut en numerisk verdi.

Merk at sannsynlighetene med to stokastiske variabler kan beregnes ved å sette opp dobbeltintegraler over egnet område eller ved å omskrive sannsynligheten til å omhandle en annen stokastisk variabel med en annen sannsynlighetsfordeling.

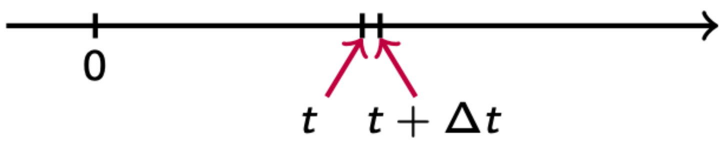
Oppgave 3a, versjon 1

Definisjon (Poissonprosess)

En prosess $\{N(t); t \geq 0\}$ *kalles in poissonprosess med intensitet* λ *hvis følgende krav er oppfylt*



1. $N(0) = 0$.
2. *For* $0 \leq t_1 < t_2 \leq t_3 < t_4$ *er* $N(t_2) - N(t_1)$ *og* $N(t_4) - N(t_3)$ *uavhengige.*
3. *For* $t \geq 0$ *og* $\Delta t > 0$ *har vi at*



$$P(N(t+\Delta t) - N(t) = 1) = \lambda \Delta t + o(\Delta t),$$
$$P(N(t+\Delta t) - N(t) \geq 2) = o(\Delta t),$$

der hver $o(\Delta t)$ *angir en funksjon som oppfyller*

$$\lim_{\Delta t \to 0} \frac{o(\Delta t)}{\Delta t} = 0.$$

Definisjon (Poissonfordeling)

La $N(t)$ *være en poissonprosess med intensitet* λ. *Fikser en verdi* $t > 0$ *og la* $X = N(t)$. *Vi sier da at* X *er poissonfordelt med parameter* λt.

$$f(x) = P(X = x) = \frac{(\lambda t)^x}{x!} e^{-\lambda t} \quad \textit{for } x = 0, 1, 2, \ldots$$

Oppgave 3a, versjon 2

Eksponentialfordeling

$$f(x) = \frac{1}{\beta} e^{-x/\beta} \,, \quad x \geq 0.$$

$$\mathrm{E}(X) = \beta \,, \quad \mathrm{Var}(X) = \beta^2 \,, \quad M_X(t) = \frac{1}{1 - \beta t} \text{ for } t < \frac{1}{\beta}.$$

Kommentar: Ofte vanlig å bruke parameteren $\lambda = 1/\beta$.

Oppgave 3a, versjon 3

Teorem
La $N(t); t \geq 0$ være en poissonprosess med intensitet λ, og la X være tidspunktet hvor hendelse nummer n i denne prosessen skjer. Da er X gammafordelt med parametre $\alpha = n$ og $\beta = \frac{1}{\lambda}$.

Parameters	• $k > 0$ shape • $\theta > 0$ scale
Support	$x \in (0, \infty)$
PDF	$f(x) = \frac{1}{\Gamma(k)\theta^k} x^{k-1} e^{-x/\theta}$
CDF	$F(x) = \frac{1}{\Gamma(k)} \gamma\left(k, \frac{x}{\theta}\right)$
Mean	$k\theta$

Kilde: https://en.wikipedia.org/wiki/Gamma_distribution

Parameters	$k \in \{1, 2, 3, \ldots\}$, shape $\lambda \in (0, \infty)$, rate alt.: $\beta = 1/\lambda$, scale
Support	$x \in [0, \infty)$
PDF	$\frac{\lambda^k x^{k-1} e^{-\lambda x}}{(k-1)!}$
CDF	$P(k, \lambda x) = \frac{\gamma(k, \lambda x)}{(k-1)!} = 1 - \sum_{n=0}^{k-1} \frac{1}{n!} e^{-\lambda x} (\lambda x)^n$

Kilde: https://en.wikipedia.org/wiki/Erlang_distribution

If k is a positive integer, then the distribution represents an Erlang distribution; i.e., the sum of k independent exponentially distributed random variables, each of which has a mean of θ.

Kilde: https://en.wikipedia.org/wiki/Gamma_distribution

Oppgave 3a, versjon 3

Teorem

La $N(t); t \geq 0$ være en poissonprosess med intensitet λ, og la X være tidspunktet hvor hendelse nummer n i denne prosessen skjer. Da er X gammafordelt med parametre $\alpha = n$ og $\beta = \frac{1}{\lambda}$.

Gammafordeling

$$f(x) = \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta}\,, \quad x \geq 0.$$

$$\mathrm{E}(X) = \alpha\beta\,, \quad \mathrm{Var}(X) = \alpha\beta^2\,, \quad M_X(t) = \left(\frac{1}{1-\beta t}\right)^\alpha \text{ for } t < \frac{1}{\beta}.$$

Spesialtilfeller: $\alpha = 1$ gir eksponensialfordelingen.

$$n \to \infty,\ p \to 0,\quad np = \lambda$$

Bernoulli forsøksrekke

- Bin(n, p) teller antall suksesser i n forsøk
- Geom(p) teller antall forsøk frem til og med første suksess
- NegBin(k, p) teller antall forsøk frem til og med suksess nr. k

Poissonprosess

- Poi(λt) teller antall hendelser i intervall t
- Exp(λ) måler tid til første hendelse
- Gamma$(k, 1/\lambda)$ måler tid til hendelse nr. k

χ^2-fordeling (kjikvadratfordeling)

Gamma$(\alpha = \nu/2,\ \beta = 2) = \chi^2_\nu$

$$f(x) = \frac{1}{2^{\nu/2}\Gamma(\nu/2)} x^{\nu/2-1} e^{-x/2}\,, \quad x \geq 0\,, \quad \nu = 1, 2, \ldots.$$

$$\mathrm{E}(X) = \nu\,, \quad \mathrm{Var}(X) = 2\nu\,, \quad M_X(t) = \left(\frac{1}{1-2t}\right)^{\nu/2} \text{ for } t < \frac{1}{2}.$$

Hører sammen med normalfordelingen:

La $Z_i \sim \mathrm{N}(0,1)$ og uavh. for $i = 1,\ldots,\nu$,

da er $\sum_{i=1}^{\nu} Z_i^2 \sim \chi^2_\nu$

Parameter $\nu \in \{1,2,\ldots\}$

Kritiske verdier i χ^2-fordelingen

$$P(\mathcal{X}^2 > \chi^2_{\alpha,\nu}) = \alpha$$

$\nu\backslash\alpha$	.995	.990	.975	.950	.050	.025	.
1	.000	.000	.001	.004	3.841	5.024	6.
2	.010	.020	.051	.103	5.991	7.378	9.
3	.072	.115	.216	.352	7.815	9.348	11.

Kritiske verdier i χ^2-fordelingen

$$P(\mathcal{X}^2 > \chi^2_{\alpha,\nu}) = \alpha$$

$\nu\backslash\alpha$	.995	.990	.975	.950	.050	.025	.010	.005
1	.000	.000	.001	.004	3.841	5.024	6.635	7.879
2	.010	.020	.051	.103	5.991	7.378	9.210	10.597
3	.072	.115	.216	.352	7.815	9.348	11.345	12.838
4	.207	.297	.484	.711	9.488	11.143	13.277	14.860
5	.412	.554	.831	1.145	11.070	12.833	15.086	16.750
6	.676	.872	1.237	1.635	12.592	14.449	16.812	18.548
7	.989	1.239	1.690	2.167	14.067	16.013	18.475	20.278
8	1.344	1.646	2.180	2.733	15.507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.940	18.307	20.483	23.209	25.188
11	2.603	3.053	3.816	4.575	19.675	21.920	24.725	26.757
12	3.074	3.571	4.404	5.226	21.026	23.337	26.217	28.300
13	3.565	4.107	5.009	5.892	22.362	24.736	27.688	29.819
14	4.075	4.660	5.629	6.571	23.685	26.119	29.141	31.319
15	4.601	5.229	6.262	7.261	24.996	27.488	30.578	32.801
16	5.142	5.812	6.908	7.962	26.296	28.845	32.000	34.267
17	5.697	6.408	7.564	8.672	27.587	30.191	33.409	35.718
18	6.265	7.015	8.231	9.390	28.869	31.526	34.805	37.156
19	6.844	7.633	8.907	10.117	30.144	32.852	36.191	38.582
20	7.434	8.260	9.591	10.851	31.410	34.170	37.566	39.997
21	8.034	8.897	10.283	11.591	32.671	35.479	38.932	41.401
22	8.643	9.542	10.982	12.338	33.924	36.781	40.289	42.796
23	9.260	10.196	11.689	13.091	35.172	38.076	41.638	44.181
24	9.886	10.856	12.401	13.848	36.415	39.364	42.980	45.559
25	10.520	11.524	13.120	14.611	37.652	40.646	44.314	46.928
26	11.160	12.198	13.844	15.379	38.885	41.923	45.642	48.290
27	11.808	12.879	14.573	16.151	40.113	43.195	46.963	49.645
28	12.461	13.565	15.308	16.928	41.337	44.461	48.278	50.993
29	13.121	14.256	16.047	17.708	42.557	45.722	49.588	52.336
30	13.787	14.953	16.791	18.493	43.773	46.979	50.892	53.672
40	20.707	22.164	24.433	26.509	55.758	59.342	63.691	66.766
50	27.991	29.707	32.357	34.764	67.505	71.420	76.154	79.490
60	35.534	37.485	40.482	43.188	79.082	83.298	88.379	91.952
70	43.275	45.442	48.758	51.739	90.531	95.023	100.425	104.215
80	51.172	53.540	57.153	60.391	101.879	106.629	112.329	116.321
90	59.196	61.754	65.647	69.126	113.145	118.136	124.116	128.299
100	67.328	70.065	74.222	77.929	124.342	129.561	135.807	140.169

Denne kommer tilbake om noen uker.

t-fordeling (Student t-fordeling)

$$f(x) = \frac{\Gamma[(\nu+1)/2]}{\Gamma(\nu/2)\sqrt{\pi\nu}}\left(1+\frac{x^2}{\nu}\right)^{-(\nu+1)/2}, \quad -\infty < x < \infty.$$

$$\mathrm{E}(X) = 0 \text{ hvis } \nu \geq 2\,, \quad \mathrm{Var}(X) = \frac{\nu}{\nu-2} \text{ hvis } \nu \geq 3\,, \quad M_X(t) \text{ eksisterer ikke.}$$

Spesialtilfeller: $\nu = 1$ gir Cauchyfordelingen.
$\nu = \infty$ gir Normalfordelingen. !

Hører sammen med normalfordelingen:

La $Z \sim \mathrm{N}(0,1)$ og $V \sim \chi^2_\nu$,

da er $\dfrac{Z}{\sqrt{V/\nu}} \sim t_\nu$

Parameter ν

Kritiske verdier i t-fordelingen

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

Kritiske verdier i t-fordelingen

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

$\nu\backslash\alpha$	.150	.100	.075	.050	.025	.010	.005	.001	.0005
1	1.963	3.078	4.165	6.314	12.706	31.821	63.657	318.309	636.619
2	1.386	1.886	2.282	2.920	4.303	6.965	9.925	22.327	31.599
3	1.250	1.638	1.924	2.353	3.182	4.541	5.841	10.215	12.924
4	1.190	1.533	1.778	2.132	2.776	3.747	4.604	7.173	8.610
5	1.156	1.476	1.699	2.015	2.571	3.365	4.032	5.893	6.869
6	1.134	1.440	1.650	1.943	2.447	3.143	3.707	5.208	5.959
7	1.119	1.415	1.617	1.895	2.365	2.998	3.499	4.785	5.408
8	1.108	1.397	1.592	1.860	2.306	2.896	3.355	4.501	5.041
9	1.100	1.383	1.574	1.833	2.262	2.821	3.250	4.297	4.781
10	1.093	1.372	1.559	1.812	2.228	2.764	3.169	4.144	4.587
11	1.088	1.363	1.548	1.796	2.201	2.718	3.106	4.025	4.437
12	1.083	1.356	1.538	1.782	2.179	2.681	3.055	3.930	4.318
13	1.079	1.350	1.530	1.771	2.160	2.650	3.012	3.852	4.221
14	1.076	1.345	1.523	1.761	2.145	2.624	2.977	3.787	4.140
15	1.074	1.341	1.517	1.753	2.131	2.602	2.947	3.733	4.073
16	1.071	1.337	1.512	1.746	2.120	2.583	2.921	3.686	4.015
17	1.069	1.333	1.508	1.740	2.110	2.567	2.898	3.646	3.965
18	1.067	1.330	1.504	1.734	2.101	2.552	2.878	3.610	3.922
19	1.066	1.328	1.500	1.729	2.093	2.539	2.861	3.579	3.883
20	1.064	1.325	1.497	1.725	2.086	2.528	2.845	3.552	3.850
21	1.063	1.323	1.494	1.721	2.080	2.518	2.831	3.527	3.819
22	1.061	1.321	1.492	1.717	2.074	2.508	2.819	3.505	3.792
23	1.060	1.319	1.489	1.714	2.069	2.500	2.807	3.485	3.768
24	1.059	1.318	1.487	1.711	2.064	2.492	2.797	3.467	3.745
25	1.058	1.316	1.485	1.708	2.060	2.485	2.787	3.450	3.725
26	1.058	1.315	1.483	1.706	2.056	2.479	2.779	3.435	3.707
27	1.057	1.314	1.482	1.703	2.052	2.473	2.771	3.421	3.690
28	1.056	1.313	1.480	1.701	2.048	2.467	2.763	3.408	3.674
29	1.055	1.311	1.479	1.699	2.045	2.462	2.756	3.396	3.659
30	1.055	1.310	1.477	1.697	2.042	2.457	2.750	3.385	3.646
35	1.052	1.306	1.472	1.690	2.030	2.438	2.724	3.340	3.591
40	1.050	1.303	1.468	1.684	2.021	2.423	2.704	3.307	3.551
50	1.047	1.299	1.462	1.676	2.009	2.403	2.678	3.261	3.496
60	1.045	1.296	1.458	1.671	2.000	2.390	2.660	3.232	3.460
80	1.043	1.292	1.453	1.664	1.990	2.374	2.639	3.195	3.416
100	1.042	1.290	1.451	1.660	1.984	2.364	2.626	3.174	3.390
120	1.041	1.289	1.449	1.658	1.980	2.358	2.617	3.160	3.373
∞	1.036	1.282	1.440	1.645	1.960	2.326	2.576	3.090	3.291

Denne kommer også tilbake om noen uker.