Maple 11 - Chi-i-anden test

Indledning

I dette dokument skal vi se hvordan Maple kan bruges til at løse opgaver indenfor χ^2 tests: χ^2-Goodness of fit test samt χ^2-uafhængighedstest.

Den nye *Gym*-pakke introduceret til Maple 16 indeholder en række gode værktøjer til netop dette emne.

Da emnet med Chi-i-anden tests er noget involveret, har jeg for begge typers vedkommende gennemgået hele processen, der fører frem til en accept eller en afvisning af hypotesen. Det er gjort for at man får en bedre forståelse af emnet, selv om noget lignende er gennemgået i flere lærebøger. Efter denne manuelle gennemgang (analyse af problemstilling) forklares det, hvordan man let løser opgaven med *Gym*-pakken.

Hvis du gerne vil vide lidt mere om teorien for Chi-i-anden tests, så kan du med fordel læse en af nedenstående dokumenter. De er udarbejdet af *Susanne Christensen*, Lektor i Statistik, Aalborg Universitet, og specielt henvendt til gymnasiet.

At træffe sine valg i en usikker verden - eller den statistiske modellerings rolle (Kan eventuelt bruges direkte i klassen)
Kursusmateriale til det nye statistikpensum (Kursusmateriale til et kursus for lærere - går mere i dybden)

Goodness of fit test

Lad os kigge på et eksempel. Før vi løser opgaven med *Gym*-pakken beskrives, analyseres situationen, så eleverne får en fornemmelse af, hvad der foregår. Det kan også være fornuftigt at løse opgaven "manuelt" skridt for skridt den første gang, inden man begiver sig ud i de automatiske løsninger.

Eksempel 1

En person deltager i et terningespil i en spilleklub. Personen er i tvivl om terningen er falsk, altså om antallet af øjne 1, 2, 3, 4, 5 og 6 forekommer med samme sandsynlighed. I al hemmelighed foretager han en række kast med terningen - i alt 120 gange slår han med terningen og hyppighederne af de enkelte antal øjne fremgår af tabellen nedenfor. Der ønskes en vurdering af om terningen kan være falsk.

Antal øjne	1	2	3	4	5	6
Hyppighed	10	27	16	18	24	25

Analyse af problemstilllingen

Eksemplet med terninger er velegnet, fordi det er tydeligt for enhver, at der er tale om tilfældigheder. Der er udtaget én *stikprøve*, som er resultatet efter at have slået 120 gange med en terning. Havde man gentaget forsøget, ville man efter al sandsynlighed have fået et andet stikprøve-resultat. Og det er også klart, at vi på ingen måde med sikkerhed kan afgøre om terningen er falsk eller ej. Det kunne jo i princippet hænde - ved en ekstrem tilfældighed - at en ægte terning ville vise for eksempel 2 øjne i alle 120 kast! For at udnytte oplysningen om stikprøven, vender vi derfor tingene på hovedet: Vi opstiller en såkaldt *nulhypotese*, kaldet H_0, som består i at antage, at terningen er ægte.

Alternativhypotesen, som er den modsatte hypotese af nulhypotesen og betegnes H_1, er da at terningen er falsk. Under antagelse af at nulhypotesen er sand, vil vi derefter udregne sandsynligheden for på denne baggrund at få den pågældende stikprøve eller noget der i en vis forstand er "mere skævt" end den observerede stikprøve. Vi *vedtager* da, at hvis denne sandsynlighed er mindst 5% eller 0.05, så vil vi *acceptere* nulhypotesen, mens vi vil *afvise* nulhypotesen, hvis sandsynligheden er mindre end 5%. Man siger, at vi sætter *signifikansniveauet* til 5%. I visse situationer vælger man et signifikansniveau på 1%. Det er de mest almindelige - med 5% som den mest anvendte.

H_0-hypotesen: Terningen er ægte
H_1-hypotesen: Terningen er falsk

Som et *mål* for, hvor langt en stikprøve er fra den forventede fordeling benyttes den såkaldte Q-værdi, indført af den britiske statistiker *Karl Pearson* (1857-1936) i år 1900:

$$Q = \sum \frac{(\textit{observeret antal} - \textit{forventet antal})^2}{\textit{forventet antal}}$$

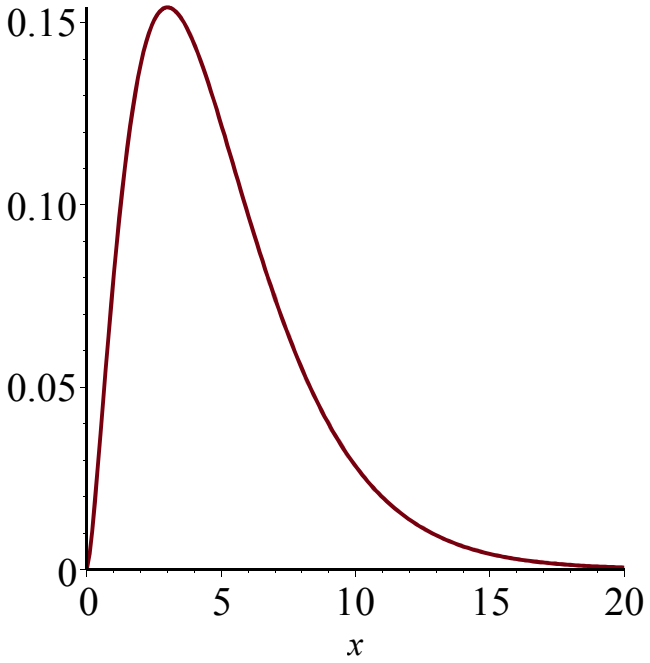
Der summeres over alle *mulige udfald*, dvs. for hvert muligt udfald (fx en et'er, en to'er, en tre'er, etc.) skriver man hvor mange, der blev observeret af hver slags og trækker det forventede antal fra, sætter i anden potens, og dividerer med det forventede antal. I øvrigt betegnes Q-værdien også *test statistikken*. Da vi slår 120 gange med en terning vil vi forventeligt få $120/6 = 20$ af hvert antal øjne. Dermed fås i dette tilfælde:

$$Q := \frac{(10-20)^2}{20} + \frac{(27-20)^2}{20} + \frac{(16-20)^2}{20} + \frac{(18-20)^2}{20} + \frac{(24-20)^2}{20} + \frac{(25-20)^2}{20}$$
$$= \frac{21}{2} \xrightarrow{\text{at 10 digits}} 10.50000000$$

Q giver kun 0, hvis for hvert udfald, det observerede antal er lig med det forventede antal. Der er altså fuld overensstemmelse i det tilfælde mellem de observerede antal og de forventede antal for hver eneste mulige udfald. Jo større Q er, jo dårligere stemmer antallet af observerede værdier overens med antallet af forventede værdier. Det er i hvert fald sådan, vi definerer det. Man kan vise - hvilket er uhyre svært - at Q tilnærmelsesvist er χ^2-fordelt (Chi-i-anden fordelt) med $n-1$ frihedsgrader, hvor n står for antallet af mulige udfald, her 6. Begrebet *antal frihedsgrader* kan opfattes som antallet af mulige udfald, hvis hyppigheder kan varieres frit i skemaet ovenfor. Da summen skal være 120, kan den sidste hyppighed ikke varieres frit, derfor er der $n-1$ frihedsgrader. Omtalte fordeling skal vi slet ikke kende til i detaljer - det er Uni stof! Vi kan dog godt få Maple til at tegne tæthedsfunktionen for Chi-i-anden fordelingen med $n-1 = 6-1 = 5$ frihedsgrader:

restart

```
with(Statistics) :
X := RandomVariable(ChiSquare(5)) :
f := x→PDF(X, x) :
plot(f(x), x = 0..20)
```



Vi udregnede en Q-værdi på 10.5 ovenfor. Sandsynligheden for at få en Q-værdi på 10.5 eller større svarer (tilnærmelsesvist) til arealet under grafen fra 10.5 til uendelig. Det fås som integralet af funktionen fra 10.5 til uendelig:

$$\int_{10.5}^{\infty} f(x)\,dx = 0.06224592809$$

Denne værdi kaldes p-værdien og er en tilnærmet værdi for sandsynligheden for at Q-værdien er 10.5 eller derover. Da p-værdien er over 0.05 *accepterer* vi nulhypotesen. Terningen ser altså ud til at være ægte. Da værdien imidlertid er meget tæt på de 0.05 kunne det måske være hensigtsmæssigt i en praktisk situation at gennemføre en stikprøve mere.

Det kunne også være interessant at bestemme den *kritiske værdi*, som er den Q-værdi, der giver anledning til en p-værdi på netop 0.05. Det kan klares ved en simpel ligningsløsning. Vi skal bestemme den "hale" under grafen for f, hvis areal er 0.05. Vi bruger *fsolve*, da der kun er én reel løsning:

$$fsolve\left(\int_{K}^{\infty} f(x)\,dx = 0.05, K\right)$$

$$11.07049769 \qquad \textbf{(2.1)}$$

Den kritiske værdi er altså 11.0705.

Løsning med Gym-pakken

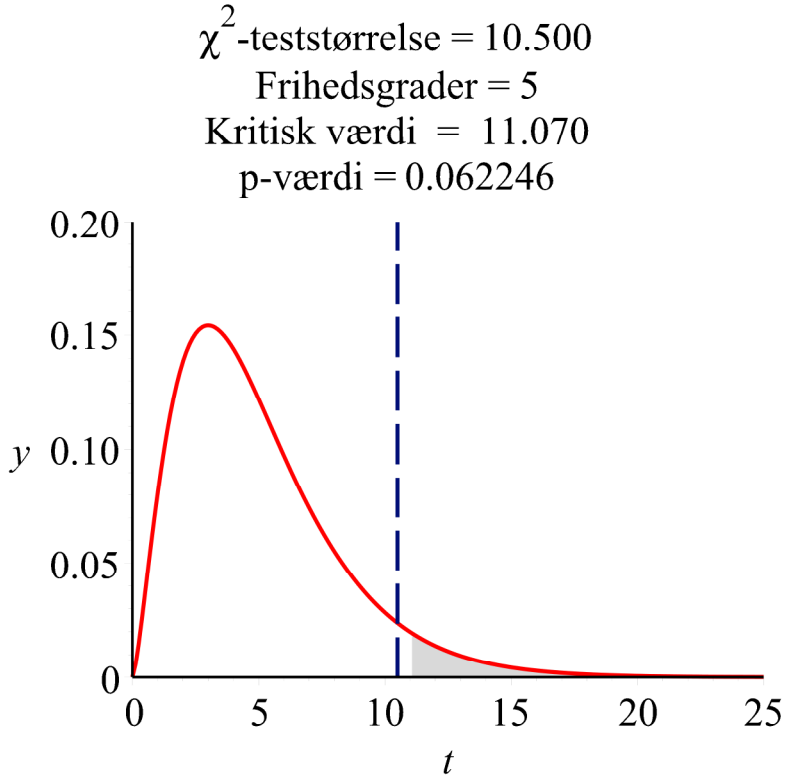
Med *Gym*-pakken er ovenstående proces yderst nem. Man laver en liste (husk firkantede parenteser)

over hyppighederne i den *observerede stikprøve* samt en liste over de forventede hyppigheder. De forventede hyppigheder er alle $120/6 = 20$. Husk at kalde *Gym*-pakken! Så fås alle udregninger straks, endda med en skraveret graf. Den *kritiske værdi* er 11.070 og svarer til den *Q*-værdi, som lige netop vil give anledning til, at nulhypotesen accepteres - svarende til en *p*-værdi på 0.05. På grafen svarer det til værdien udfor venstre kant af det skraverede område. Test statistikken er 10.5 og er markeret med en lodret blå linje på grafen. Da *p*-værdien i det konkrete tilfælde er $0.062246 > 0.05$ accepteres nulhypotesen.

H_0-hypotesen: Terningen er ægte

H_1-hypotesen: Terningen er falsk

```
restart
with(Gym) :
obs := [10, 27, 16, 18, 24, 25] :
forv := [20, 20, 20, 20, 20, 20] :
ChiKvadratGOFtest(obs, forv)
```



Bemærkninger 1

Der er nogle forhold, der er vigtige at gøre sig klart, når man foretager en Chi-i-anden test. Der er også nogle betingelser for at approksimationen (tilnærmelsen) med en Chi-i-anden fordeling overhovedet kan bruges.

- Der er tale om en *simpel* stikprøve, hvor stikprøven er udtrukket tilfældigt fra en population, hvor enhver samling af elementer fra populationen af den givne stik-prøvestørrelse har lige stor sandsynlighed for at blive udtrukket.
- De enkelte udfald skal være *uafhængige* af hinanden, dvs. hændelsen at et udfald forekommer har ingen indflydelse på sandsynligheden for at det pågældende udfald forekommer igen eller ej.
- Populationen er meget større end stikprøven.

• For at tilnærmelsen er god, skal hyppighederne af de enkelte observationer helst *alle* være mindst 5.

Bemærk, at *p*-værdien **ikke** angiver sandsynligheden for at terningen er ægte!!! *p*-værdien er derimod en tilnærmelsesvis værdi for sandsynligheden for at man får den pågældende stikprøve eller noget der afviger mere fra det forventede, **givet at nulhypotesen er korrekt**. Stikprøvens afvigelse fra det forventede er beskrevet ved *Q*-værdien.

Som nævnt i analysedelen ovenfor kan man sagtens komme ud for at drage en forkert konklusion på baggrund af Chi-i-anden testen. Man taler om de såkaldte *type I* og *type II* fejl. Ved type I fejl afvises en nulhypotese, som faktisk er korrekt. Ved type II fejl accepteres en nulhypotese, som faktisk er falsk. Vi kan skrive det op i et skema:

	H_0 er sand	H_0 er falsk
H_0 accepteres	Korrekt afgørelse	Type II fejl
H_0 afvises	Type I fejl	Korrekt afgørelse

Type I fejl forekommer især, hvis den stikprøve, man har opnået, er meget usædvanlig.

Bemærkninger 2

Hvis man for eksempel ønsker at foretage en hypotesetest med et andet signifikansniveau, fx 1% eller 0.01, så kan det gøres ved at tilføje en option til kommandoen:

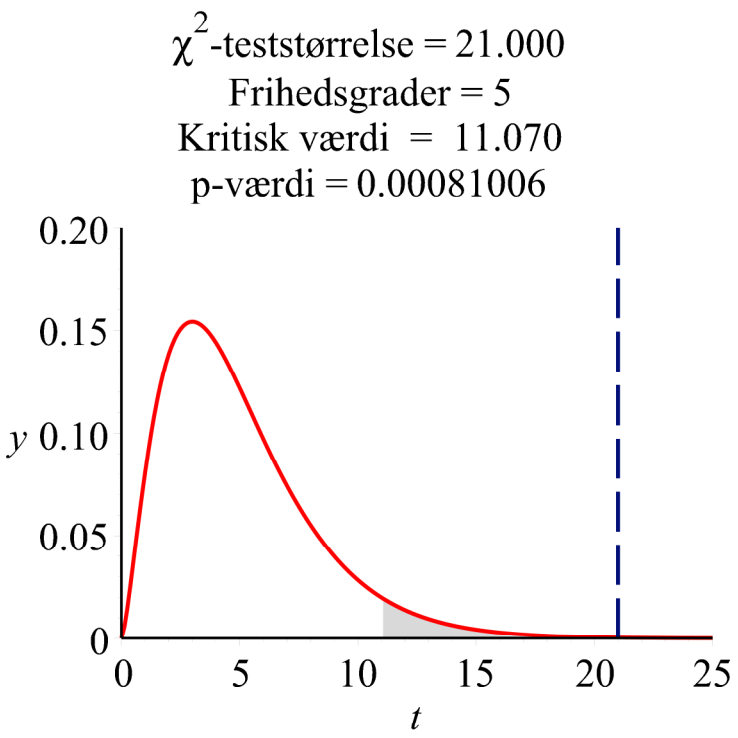
ChiKvadratGOFtest(*obs*, *forv*, *level* = 0.01)

Eksempel 2

Lad os prøve at se, hvad der ville ske, hvis alle hyppighederne i tabellen i eksempel 1 blev fordoblet, altså:

Antal øjne	1	2	3	4	5	6
Hyppighed	20	54	32	36	48	50

```
restart
with(Gym) :
obs := [20, 54, 32, 36, 48, 50] :
forv := [40, 40, 40, 40, 40, 40] :
ChiKvadratGOFtest(obs, forv)
```



Vi ser her, at Q-værdien er 21 og p-værdien er nede på 0.00081006. Det er altså ret usandsynligt, at vi ved en udvidelse af stikprøven til 240 slag med terningen får de samme indbyrdes forhold mellem de observerede værdier som i eksempel 1. Dette kan forklares med *de store tals lov*. Jo flere slag man foretager, jo større er tendensen til at tilfældighederne udjævner sig, og at antallet af observerede værdier falder tættere på antallet af forventede værdier, relativt set.

Uafhængighedstest

Der er en anden type Chi-i-anden test og det er den såkaldte *uafhængighedstest*. Vi kigger igen på problemstillingen via et eksempel og foretager - ligesom i forrige afsnit - en analyse af problemstillingen, så det uddybes hvad der egentlig foregår. Igen er det en god idé at udføre de enkelte trin i processen "manuelt", før man kaster sig over de mere automatiske løsninger.

Eksempel 1

I Politiets statistik for praktiske køreprøver til almindelig personbil fra året 2004 fremgår det, at der i Haderslev var 514, der dumpede, mens 915 bestod. I Slagelse dumpede 677, mens 1349 bestod. Vurder indenfor et signifikans-niveau på 5% eller 0.05, om man kan sige, at chancen for at dumpe er *uafhængig* af i hvilken af de to byer man bor.

Analyse af problemstilllingen

Mange ukyndige ville måske blot udregne beståelsesprocenterne de to steder:

Haderslev: $514 / (514 + 915) = 0.360 = 36\%$.
Slagelse: $677 / (677 + 1349) = 0.334 = 33.4\%$.

og derefter blot give et bud på svaret ved brug af sund fornuft. Det går som regel også godt, hvis de

to procenter er meget forskellige eller er meget ens. Men hvad når der er tale om en mellemting som her? Det er ikke muligt at formulere en regel om, at hvis forskellen på de to procenter er så og så stor, så kan man sige, at der er tale om tilfældigheder. I eksempel 2 i forrige afsnit så vi nemlig, at også *antallet* af observationer - ikke blot procenterne - spiller en stor rolle i afgørelsen. For at løse opgaven på en pålidelig og matematisk måde, må man gøre brug af den såkaldte *Chi-i-anden uafhængighedstest*. Her opstilles de to hypoteser:

H_0-hypotesen: Chancen for at bestå køreprøven er *uafhængig* af hvilken af de to byer man bor i.
H_1-hypotesen: Chancen for at bestå køreprøven er *ikke uafhængig* af hvilken af de to byer man bor i.

Man kan se på data for de to byer som en slags *stikprøve*. Det skal forstås på den måde, at hvis man lod 1429 *andre* personer fra Haderslev og 2026 *andre* personer fra Slagelse gå op til køreprøve, så ville man højst sandsynligt få et andet resultat. Der er altså noget *tilfældighed* indbygget. Spørgsmålet er, om man med rimelighed kan sige, at chancen for at dumpe i de to byer er den samme, eller om der måske gør sig noget andet gældende. Det kan for eksempel være, at det er særligt svært at køre rundt i den ene af byerne - dette ville nok være tilfældet hvis København var inde i billedet - så det er sværere at bestå. Dette kan også være, at de køresagkyndige i den ene by er mere hårde i bedømmelsen end de køresagkyndige i den anden by, etc. Mange forklaringer kan være mulige.

Ligesom i Chi-i-anden Goodness of fit testen i forrige afsnit griber vi opgaven modsat an: Vi antager at nulhypotesen er sand, dvs. at chancen for at bestå køreprøven er uafhængig af i hvilken af de to byer man tager prøven. Dernæst bestemmer vi - på baggrund af denne antagelse - sandsynligheden for at få den "stikprøve" vi allerede har opnået eller noget, der er længere fra det forventede end denne. Hvis denne sandsynlighed, kaldet *p*-værdien, er større end eller lig med signifikansniveauet på de 0.05, så vedtager vi at *acceptere* nulhypotesen. I modsat fald afvises nulhypotesen. Som et mål for, hvor langt en stikprøve ligger fra det forventede, bruger vi den samme størrelse som i forrige afsnit, nemlig *Q*-værdien givet ved:

$$Q = \sum \frac{(observeret\ antal - forventet\ antal)^2}{forventet\ antal}$$

og som blev indført af Karl Pearson i år 1900. De forventede antal er lidt sværere at bestemme end tilfældet var i eksempel 1 i forrige afsnit. Da vi ikke har anden viden end de fire data-værdier fra Haderslev og Slagelse, så er vi nødt til selv at konstruere et godt bud eller estimat på det forventede antal af hver type. Det mest overskuelige er at skrive data ind i et skema, en såkaldt *antalstabel*:

	Bestået	Ikke bestået	Sum
Haderslev	915	514	1429
Slagelse	1349	677	2026
Sum	2264	1191	3455

Observerede værdier

hvor vi har udregnet summerne i både vandret og lodret retning. Vi ser, at betragtet over *begge* byer, så er 2264 ud af 3455 bestået. Det giver en beståelsesprocent på $2264 / 3455 = 0.6553$, altså 65.53%. Det vil derfor være rimeligt at *forvente* at $1429 \cdot 0.6553 = 936.4$ består i Haderslev og $2026 \cdot 0.6553 = 1327.6$ består i Slagelse. Vi kan også foretage beregningen i et hug:

restart

$$\frac{2264}{3455}\cdot 1429 = \frac{3235256}{3455} \xrightarrow{\text{at 10 digits}} 936.3982634$$

$$\frac{2264}{3455}\cdot 2026 = \frac{4586864}{3455} \xrightarrow{\text{at 10 digits}} 1327.601737$$

På tilsvarende måde kan vi udregne de forventede værdier for antal *ikke beståede* i henholdvis Haderslev og Slagelse:

$$\frac{1191}{3455}\cdot 1429 = \frac{1701939}{3455} \xrightarrow{\text{at 10 digits}} 492.6017366$$

$$\frac{1191}{3455}\cdot 2026 = \frac{2412966}{3455} \xrightarrow{\text{at 10 digits}} 698.3982634$$

Disse forventede værdier, kan nu føjes ind i en tabel:

	Bestået	Ikke bestået	Sum
Haderslev	936.4	492.6	1429
Slagelse	1327.6	698.4	2026
Sum	2264	1191	3455

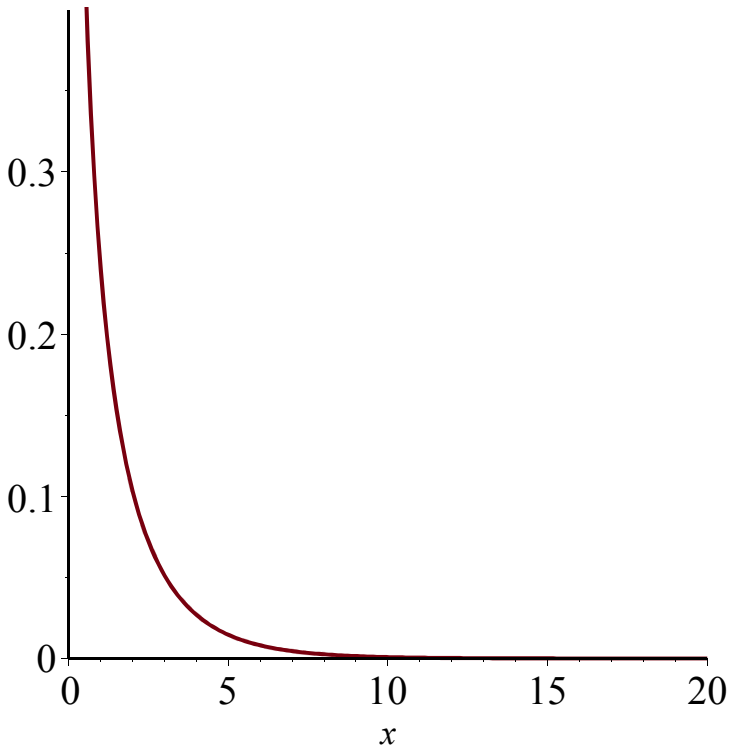
Forventede værdier

Vi ser at de vandrette og lodrette summer giver det samme som i tabellen med de observerede værdier. Det er ingenlunde et tilfælde - det sker altid op til afrunding. Du kan fundere over hvorfor? Vi har nu de fire observerede hyppigheder og de fire forventede hyppigheder. Dem indsætter vi i formlen for Q. Det giver:

$$Q := \frac{(915-936.4)^2}{936.4} + \frac{(514-492.6)^2}{492.6} + \frac{(1349-1327.6)^2}{1327.6} + \frac{(677-698.4)^2}{698.4} = 2.419424431$$

Teorien siger nu - og det er alt for svært at bevise i gymnasiet - at Q-værdien, også kaldet *test statistikken* eller *test størrelsen*, tilnærmelsesvist er Chi-i-anden fordelt med $(s-1)\cdot(r-1)$ frihedsgrader. Her betyder s antallet af *søjler* i den inderste del af tabellen, dvs. den med data uden summer. Tilsvarende betyder r antallet af *rækker* i samme tabel. Her har vi at gøre med to rækker og to søjler, så $(s-1)\cdot(r-1) = (2-1)\cdot(2-1) = 1\cdot 1 = 1$. Ligesom i afsnit 2 kan vi prøve at tegne tæthedsfunktionen for Chi-i-anden fordelingen med 1 frihedsgrad:

```
with(Statistics) :
X := RandomVariable(ChiSquare(1)) :
f := x→PDF(X, x) :
plot(f(x), x = 0..20)
```



Vi udregnede en Q-værdi på 2.419 ovenfor. Sandsynligheden for at få en Q-værdi på 2.419 eller større svarer (tilnærmelsesvist) til arealet under grafen fra 2.419 til uendelig. Det fås som integralet af funktionen fra 2.419 til uendelig:

$$\int_{2.419}^{\infty} f(x)\,\mathrm{d}x = 0.1198714301$$

Denne værdi kaldes p-værdien og er en tilnærmet værdi for sandsynligheden for at Q-værdien er 2.419 eller derover. Da p-værdien er over 0.05 *accepterer* vi nulhypotesen. Hypotesen om at der er lige stor chance for at bestå i de to byer kan dermed godt accepteres indenfor et signifikansniveau på 0.05.

Til slut kunne det være interessant at bestemme den kritiske værdi, som er grænsen for Q-værdien, der skiller en accepteret nulhypotese fra en afvist nulhypotese.

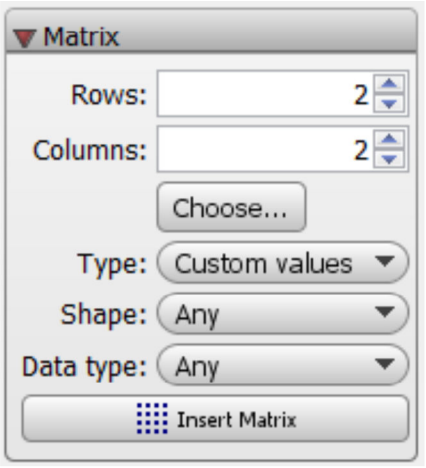
$$fsolve\left(\int_{K}^{\infty} f(x)\,\mathrm{d}x = 0.05,\ K\right)$$

$$3.841458821 \qquad \textbf{(3.1)}$$

Den kritiske værdi er altså 3.84, hvilket betyder, at hvis Q er større end 3.841, så afvises nulhypotesen og hvis den er 3.841 eller derunder, så accepteres den.

Løsning med Gym-pakken

Efter en *restart* og kald af *Gym*-pakken skal vi have skrevet de observerede hyppigheder ind i en matrix, som hentes i paletten *Matrix* ovre i venstre side. Den skal have dimensioner 2×2:



Problemstillingen er følgende:

H_0-hypotesen: Chancen for at bestå køreprøven er *uafhængig* af hvilken af de to byer man bor i.
H_1-hypotesen: Chancen for at bestå køreprøven er *ikke uafhængig* af hvilken af de to byer man bor i.

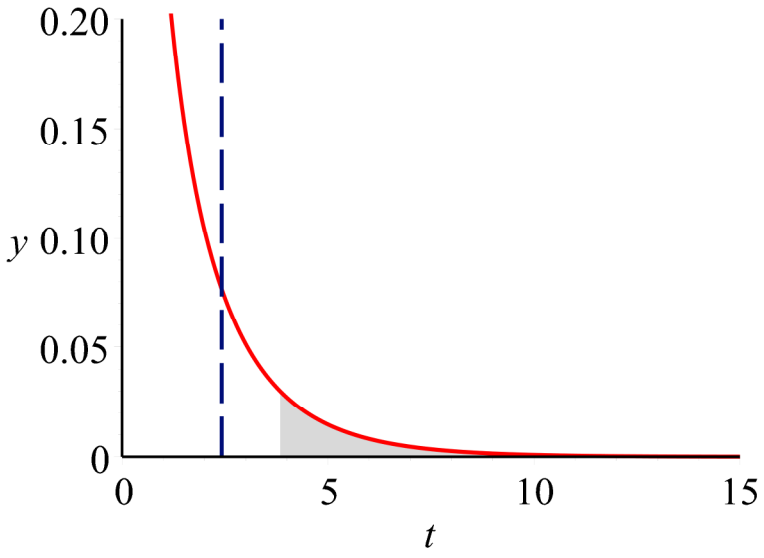
Vi skal afgøre om nulhypotesen kan accepteres ved et signifikansniveau på 5% eller 0.05.

restart
with(*Gym*) :

$$obs := \begin{bmatrix} 915 & 514 \\ 1349 & 677 \end{bmatrix} :$$

ChiKvadratUtest(*obs*)

χ^2-teststørrelse = 2.4190
Frihedsgrader = 1
Kritisk værdi = 3.8415
p-værdi = 0.11987



Da *p*-værdien er større end 0.05, accepteres nulhypotesen. Man kan altså godt acceptere hypotesen om at der er lige stor chance for at bestå køreprøven i de to byer.

Bemærkning

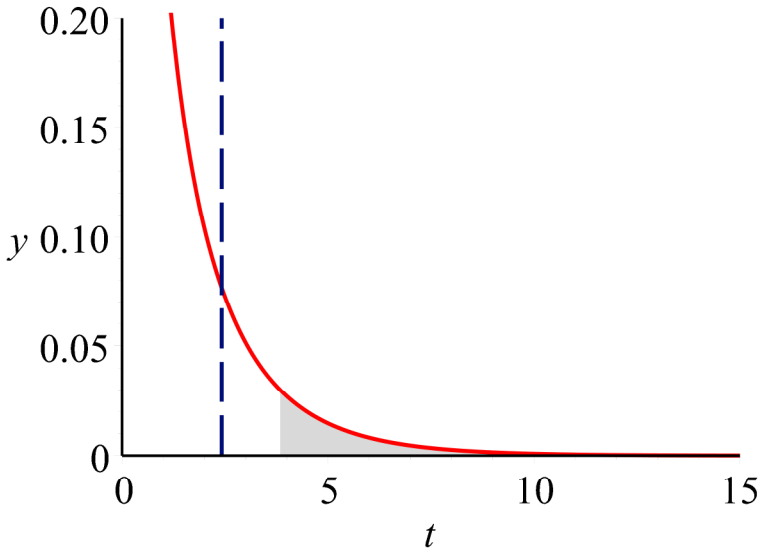
Hvis man ønsker, giver *Gym*-pakken også mulighed for at skrive lidt ekstra til i antalstabellen. Man skal blot huske at anbringe anførselstegn om forklaringerne. Man skal også være omhyggelig med at bruge *præcist* nedenstående ord: "Observeret", "i alt", endda med store/små bogstaver! :

restart
with(*Gym*) :

$$obs := \begin{bmatrix} \text{"Observeret"} & \text{"Bestået"} & \text{"Ikke bestået"} & \text{"i alt"} \\ \text{"Haderslev"} & 915 & 514 & 1429 \\ \text{"Slagelse"} & 1349 & 677 & 2026 \\ \text{"i alt"} & 2264 & 1191 & 3455 \end{bmatrix} :$$

ChiKvadratUtest(*obs*)

χ^2-teststørrelse = 2.4190
Frihedsgrader = 1
Kritisk værdi = 3.8415
p-værdi = 0.11987



Husk at hvis du får brug for en matrix med mere end 10 rækker eller 10 søjler (det sker nok sjældent), så skal du før du trykker på Insert Matrix skrive en kommando *visMatrix*(20), som også er i *Gym*-pakken. Med den udvides eksplicit visning af matricer til 20×20 matricer. Man kan også taste matricer ind på en anden måde. Kig eventuelt i hjælpemenuen til *Gym*-pakken under *antalstabel*. En sidste ting: Du kan også få tabellen med de forventede værdier frem:

forventet(*obs*)

"Forventet"	"Bestået"	"Ikke bestået"	"i alt"
"Haderslev"	936.40	492.60	1429
"Slagelse"	1327.6	698.40	2026
"i alt"	2264	1191	3455

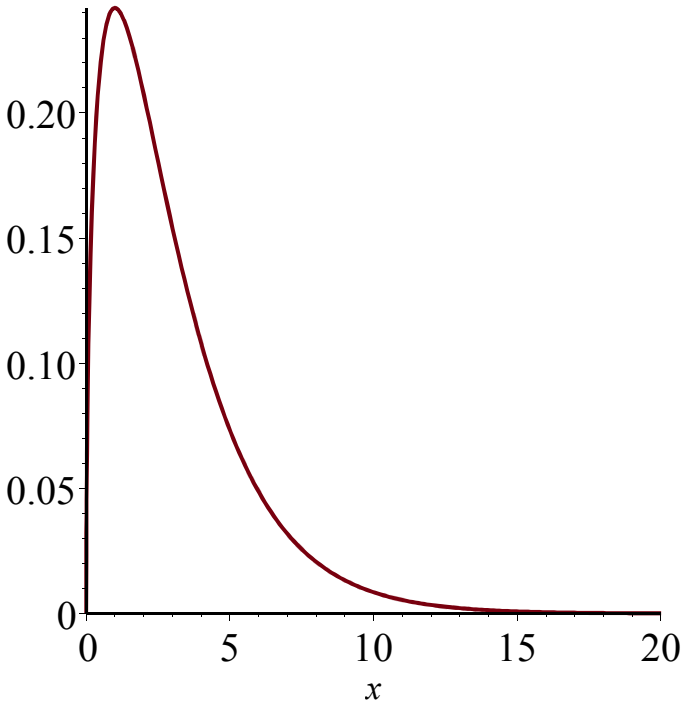
(3.2)

Hvis du skriver *obs* med forklaringer, kommer tabellen med de forventede værdier også med forklaringer!

Chi-i-anden fordelingen

I dette afsnit skal vi blot overfladisk betragte den såkaldte *Chi-i-anden fordeling med n frihedsgrader*, også skrevet med det lille græske bogstav χ for chi: χ^2*-fordelingen med n frihedsgrader*. Ligesom tilfældet er for *Normalfordelingen*, er der tale om en *kontinuert* sandsynlighedsfordeling. I Maple kan man definere en *stokastisk variabel X* ved hjælpe af kommandoen: *RandomVariable*. Som argument skal man fortælle hvilken fordeling denne stokastiske variabel skal have. Chi-i-anden fordelingen med n frihedsgrader hedder her $ChiSquare(n)$. Da kommandoerne ligger i pakken *Statistics* skal du huske at kalde denne pakke, før du går i gang. Endelig kan man referere til den stokastiske variabels *tæthedsfunktion* vis kommandoen *PDF*. Nedenfor har vi gjort alt dette og endda plottet fordelingen i tilfældet $n = 3$.

```
restart
with(Statistics) :
X := RandomVariable(ChiSquare(3)) :
f := x→PDF(X, x) :
plot(f(x), x = 0..20)
```



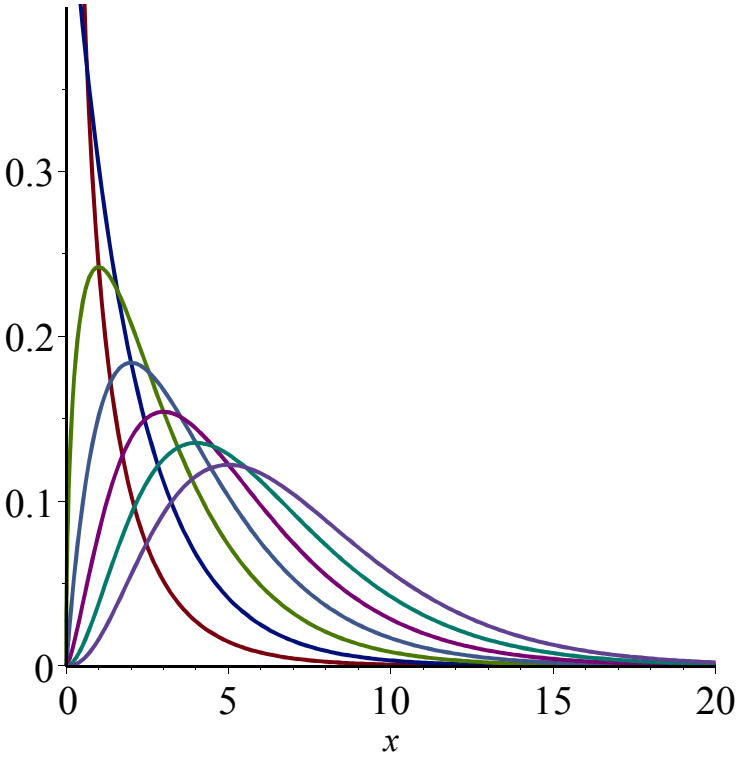
Chi-i-anden fordelingens tæthedsfunktion ser ret forskellig ud for de forskellige værdier af n. Vi kan

prøve at tegne tæthedsfunktionerne for tilfældene 1, 2, 3, 4, 5, 6 og 7 i samme koordinatsystem.

```
restart
with(Statistics) :

for k from 1 to 7 do
  X_k := RandomVariable(ChiSquare(k)) :
  f_k := x→PDF(X_k, x) :
end do:

plot([seq(f_k(x), k = 1 ..7)], x = 0 ..20)
```



Forklaring

Husk at man kan tegne graferne for flere funktioner i samme koordinatsystem, hvis man sætter dem ind i en liste - dvs. angivet med komma imellem og med firkantede parenteser omkring. Da der er systematik i det her, kan vi hurtigt definere tæthedsfunktionerne ved hjælp af en *for*-løkke. Bemærk at hele løkken står i ét matematikfelt, selv om det fylder flere linjer. Man kan skrive på flere linjer, hvis man skifter linje med **Shift+Enter**. Endelig kan man bruge kommandoen *seq* for "sequence" til at spare plads. $[seq(f_k(x), k = 1 ..7)]$ er det samme som $[f_1(x), f_2(x), f_3(x), f_4(x), f_5(x), f_6(x), f_7(x)]$.

Vil man ydermere have betegnelser på hvilke kurver, der hører til hvilke Chi-i-anden fordelinger, kan man enten tilføje en *legend* option, som vist nedenfor. Man kan dog også manuelt højreklikke på grafen og vælge *Legend > Show Legend.* Derved kommer der som default betegnelserne "Curve 1", "Curve 2", etc., men man kan hurtigt redigere i dem ved at klikke på dem og skrive det man vil. Man skal dog være opmærksom på, at de manuelle ændringer føres tilbage til default tilstanden ved genberegning. Det gør det derimod ikke, når man skriver direkte i kommandoen.

$$plot\left(\left[seq\left(f_k(x), k = 1..7\right)\right], x = 0..20, legend = \left["\lambda = 1", "\lambda = 2", "\lambda = 3", "\lambda = 4", "\lambda = 5", "\lambda = 6", "\lambda = 7"\right]\right)$$