Dette er mine egne opgavebesvarelser til SS18-19. Jeg tager forbehold for fejl og dårlige formuleringer. Jeg vil opdatere dette dokument hver uge.

Bemærk at jeg har udeladt hvilken variabel der bliver integreret med hensyn til, dette må man abstrahere for.

Indhold

1 Uge 1

1.1 Øvelser tirsdag den 21. november kl 15.15-17

1.1.1 BH. 1.2

A How many 7-digit phone numbers are possible, assuming that the first digit can't be a 0 or a 1?
Svar: $8 \cdot 10^6$. Vi har ikke 10^7 muligheder da det første tal skal være $2-8$. Vi bruger her multiplikationsreglen, theorem 1.4.1.

B Re-solve (a), except now assume also that the phone number is not allowed to start with 911 (since this is reserved for emergency use, and it would not be desirable for the system to wait to see whether more digits were going to be dialed after someone has dialed 911).
Svar: Vi har 10^7 forskellige telefonnumre. Der er dog 10^4 forskellige telefonnumre der begynder med 911. Igen har vi brugt multiplikationsreglen, theorem 1.4.1. Vi justerer for at have talt for mange telefonnumre med og får $8 \cdot 10^6 - 10^4$ telefonnumre der ikke begynder med 911.

1.1.2 BH. 1.3

Fred is planning to go out to dinner each night of a certain week, Monday through Friday, with each dinner being at one of his ten favorite restaurants.

A How many possibilities are there for Fred's schedule of dinners for that Monday through Friday, if Fred is not willing to eat at the same restaurant more than once?
Svar: $10 \cdot 9 \cdot 8 \cdot 7 \cdot 6 = 30240$. Dette følger af multiplaktionsreglen, theorem 1.4.1 eller theorem 1.4.6, udtrækninger uden tilbagelægning.

B How many possibilities are there for Fred's schedule of dinners for that Monday through Friday, if Fred is willing to eat at the same restaurant more than once, but is not willing to eat at the same place twice in a row (or more)?
Svar: Fred kan mandag vælge i mellem 10 mulige restauranter. Tirsdag kan han ikke vælge den han spiste på mandag, han kan derfor kun vælge i mellem 9 mulige restauranter. Dette forsætter onsdag, torsdag og fredag. Vi får dermed $10 \cdot 9 \cdot 9 \cdot 9 \cdot 9 = 65610$ muligheder. Dette følger af multiplaktionsreglen, theorem 1.4.1.

1.1.3 BH.1.21

Three people get into an empty elevator at the first floor of a building that has 10 floors. Each presses the button for their desired floor (unless one of the others has already pressed that button). Assume that they are equally likely to want to go to floors 2 through 10 (independently of each other). What is the probability that the buttons for 3 consecutive floors are pressed? (Vink: Hvor mange valg kan de tre personer foretage? Hvor mange tripler af etager er mulige for at få tre etager i træk, og hvor mange muligheder er der for at hver af disse tripler vælges?)
Svar: Lad os følge det givne hint. Hver person kan vælge i mellem 9 mulige etager, dette giver $9^3 = 729$ mulige valg per multiplaktionsreglen, theorem 1.4.1. Der er 7 forskellige tripler af etager. Disse er givet ved

- 2,3,4
- 3,4,5
- 4,5,6
- 5,6,7
- 6,7,8
- 7,8,9

- 8,9,10

Der er $3! = 6$ forskellige måder at permuterer 3 tal på. Dette følger af eksempel 1.4.7 (permutationer og fakultet) eller multiplaktionsreglen, theorem 1.4.1. Det er 7 tripler. Vi får dermed $7 \cdot 6 = 42$ muligheder for at vælge en triple. Per def. 1.3.1 (den naive definition af sandsynlighed) får vi sandsynligheden for at 3 på hinanden etager bliver valgt til $42/729 \approx 0.0576$

1.1.4 BH. 1.27(a)

(probability that the total after rolling 4 fair dice is 21) or (probability that the total after rolling 4 fair dice is 22)

Svar: For at få 21 skal vi slå en permutation, dvs. det er ligegyldigt hvilken rækkefølge vi slår øjnene, af enten (6,6,6,3) (4 muligheder), (6,5,5,5) (4 muligheder) eller (6,6,5,4) (4!/2=12 muligheder. Vi har her 4! forskellige permutationer, men vi kan ikke kende forskel på 6'erne, derfor dividerer vi med 2).

For at få 22 skal vi enten slå (6,6,6,4) (4 muligheder) eller (6,6,5,5) ($4!/2^2$=6 muligheder. Igen kan vi ikke se forskel på alle tallene, 6'erne og 5'erne. Vi dividerer derfor med 2^2).

Videre per multiplaktionsreglen, theorem 1.4.1, er i alt $6^4 = 1296$ forskellige terninge slag med 4 terninger. Per def. 1.3.1 (den naive definition af sandsynlighed) får vi

$$\frac{4+4+12}{1296} \quad \text{og} \quad \frac{4+6}{1296},$$

hvor det let ses at 4/1296 må være det største tal, således

$$\frac{20}{1296} > \frac{10}{1296}.$$

Så her er vi altså mindst ligeså klog som Leibniz.

1.1.5 BH. 1.29

Elk dwell in a certain forest. There are N elk, of which a simple random sample of size n are captured and tagged. The captured elk are returned to the population, and then a new sample is drawn, this time with size m. What is the probability that exactly k of the m elk in the new sample were previously tagged? (Vink: På hvor mange måder kan de m elge udvælges? På hvor mange måder kan de k elge med tags udvælges? På hvor mange måder kan de m-k elge uden tags udvælges? Hvordan skal tallene kombineres?)

Svar: Igen følger vi hintet. Der er $\binom{N}{m}$ måder at udvælge m elge fra en population på N (Binomial koefficientens definition kan findes ved def. 1.4.12). Der er n elge med tags, vi kan dermed udvælge k elge i denne population på $\binom{n}{k}$ måder. Videre er der N-n elge uden tags. Vi kan altså udvælge m-k elge i denne population på $\binom{N-n}{m-k}$ måder. I alt får vi

$$\binom{n}{k}\binom{N-n}{m-k}$$

forskellige måder at vælge præcis k elge i den nye udtrækning med m elge i alt. Sammen med det første resultat får vi sandsynligheden til at være per def. 1.3.1 (den naive definition af sandsynlighed)

$$\frac{\binom{n}{k}\binom{N-n}{m-k}}{\binom{N}{m}}.$$

Dette er også kendt som den hypergeometriske fordeling. Hvis du er en ivrig kortspiller er du måske stødt på den, da du kan bruge fordelingen til at regne dine outs ud for at vinde en given hånd/spil. Noter at dette kun giver mening sålænge $k \leq n$ og $m - k \leq N - n$. Det giver også god fysisk mening idet at hvis k er større end n er det ikke k elge som har et mærkat på sig. I disse ekstreme tilfælde er sandsynligheden 0.

1.1.6 BH. 1.30

Four cards are face down on a table. You are told that two are red and two are black, and you need to guess which two are red and which two are black. You do this by pointing to the two cards you're guessing are red (and then implicitly you're guessing that the other two are black). Assume that all configurations are equally likely, and that you do not have psychic powers. Find the probability that exactly j of your guesses are correct, for j = 0,1,2,3,4.

Vink: Skriv alle (hvor mange?) muligheder op for hvordan kortene kan ligge.

Svar: Vi følger hint: Der er $\binom{4}{2} = 6$ muligheder at lægge kortene (Idet du positionen af f.eks. de røde kort, vælger du implicit også positionen af de sorte og vice versa). Da der kun er 6 kan de let opskrives, dette kunne man evt. også bare starte med. Disse ville så være

- RRSS
- RSRS
- RSSR
- SSRR
- SRSR
- SRRS.

Af symmetri årsager må P(0)=P(4), P(1)=P(3). Det er umuligt at få 1 eller 3 rigtige, da man nødvendigvis kender det sidste korts kulør såfremt man kender de 3 første. Videre er der kun en måde at gætte alle kortene rigtigt på. Dette giver at der også kun er en måde at gætte alle kortenes kulør forkert. Dette giver at P(0)=P(4)=1/6. Der er nu 4/6 sandsynlighedsmasse tilbage, hvilket betyder må betyde at P(2)=4/6. Så samlet P(0)=P(4)=1/6, P(1)=P(3)=0 og P(2)=4/6.

Alternativt: Da alle konfigurationer er lige sandsynlige kan vi uden tab af generalitet antage en af konfigurationer. En hurtig optælling viser at sandsynlighederne er P(0)=P(4)=1/6, P(1)=P(3)=0 og P(2)=4/6.

1.2 Øvelser fredag den 24. november kl 8.15-10

1.2.1 SS.1

Vi udfører følgende eksperiment: Vi kaster først den grønne mønt. Hvis vi får plat, kaster vi derefter den røde mønt; ellers kaster vi den blå mønt. Vi registrerer til sidst om vi fik krone eller plat i andet kast, som altså enten var med den røde eller den blå mønt.

1 Givet en mønt er rød/blå, så har vi en sandsynlighed for at få krone på 0.3 og 0.4 per antagelse. Videre, da den grønne mønt er symmetrisk, så sandsynligheden er den samme for at få blå eller rød.

2 Hvad er vores udfaldsrum S? Det er alle kombinationer af udfald, dvs. (hvor jeg har farvekodet) (plat, plat), (plat, krone), (krone, plat), (krone, krone).

Vi vælger at lave en partition af S alt efter om det den blå eller røde mønt vi slår med til sidst. Vi har lige argumenteret for at $P(blaa) + P(roed) = 1$ og per eksperiment konstruktion er $P(blaa \cap roed) = \varnothing$. Vi kan altså bruge det som en partition i LOTP (I selve sætningen ville dette være A_i'erne og krone ville være B). Vi opskriver summen

$$P(krone) = P(krone|blaa)P(blaa) + P(krone|roed)P(roed) = 0.4 \cdot 0.5 + 0.3 \cdot 0.5 = 0.2 + 0.15 = 0.35.$$

Det er relativt intuitivt. Da den grønne mønt er symmetrisk ændrer det ikke noget om vi kan derfor se det som (0.3+0.4)/2=0.35.

3 Vi bruger Bayes

$$P(roed|krone) = \frac{P(krone|rød)P(roed)}{P(krone)} = \frac{0.3 \cdot 0.5}{0.35} \approx 0.429.$$

Dette tal siger at givet du ved det er krone, altså du kender eksperimentets udfald, hvad er så chancen for at det var den røde mønt du kastede. Om det er super intuitivt at det lige præcis er dette tal er det måske ikke, men det giver meget god mening det er lidt mindre end 0.5 da der er større sandsynlighed for at få krone med den blå mønt end den røde.

1.2.2 SS.2

1 mest naturlige udfaldsrum ville dog nok være at sige dit udfaldsrum var $\{1, 2, \ldots, 6\}^2$, dvs. $(1, 1), (1, 2), \ldots, (2, 1), \ldots, (6, 6)$. Hvilke værdier kan X indtage? Der er 18 unikke elementer. Lidt stenet at skrive op, kan simuleres nemt i R, f.eks. ved

```
> terninge_produkt = rep(1:6, times = 1, each = 6)*1:6
> S = unique(terninge_produkt)
> S
 [1]  1  2  3  4  5  6  8 10 12  9 15 18 16 20 24 25 30 36
```

Dette er en diskret stokastisk variabel.

2 Y kan have værdierne [0,3/2]. Den kan indtage alle værdier i dette interval. Den er kontinuert.

1.2.3 BH. 1.54(a)-(c)

A (number of ways to choose 5 people out of 10) (number of ways to choose 6 people out of 10)

Svar: Per Binomial koefficientens definition, 1.4.12, får vi at der er $\binom{10}{5}$ måder at vælge 5 personer ud af 10. Tilsvarende får vi at der er $\binom{10}{6}$ måder at vælge 6 personer ud af 10. Alt i alt får vi

$$\binom{10}{5} = \frac{10!}{(10-5)!5!} = \frac{10!}{5!5!} = 252 > 210 = \frac{10!}{4!6!} = \frac{10!}{(10-6)!6!} = \binom{10}{6}.$$

Dette munder ud at $6 > 5$. Videre er $\binom{k}{n}$ maximeret ved $n = k/2$ for k lige. Dette kan illustreres ved Pascal trekant.

B (number of ways to break 10 people into 2 teams of 5) (number of ways to break 10 people into a team of 6 and a team of 4)

Svar: Fra tidligere ved vi at der er $\binom{10}{5}$ måder at lave et hold på 5. Vi har dog her talt alle holdene to gange, da vi har et hold 1 eller 2. Vi får derfor $\frac{1}{2}\binom{10}{5}$. Det andet udsagn udregnes som $\binom{10}{4}$ eller $\binom{10}{6}$. Bemærk at disse giver det samme (example 1.5.1 eller symmetrien i Pascal trekant). Her tæller vi ikke dobbelt, da $4 \neq 6$. Dette giver

$$\frac{1}{2}\binom{10}{5} = \frac{10!}{2(10-5)!5!} = \frac{10!}{2 \cdot 5!5!} = 126 < 210 = \frac{10!}{4!6!} = \frac{10!}{(10-6)!6!} = \binom{10}{6} = \binom{10}{4}.$$

C (probability that all 3 people in a group of 3 were born on January 1) (probability that in a group of 3 people, 1 was born on each of January 1, 2, and 3)

Svar: Der er en måde hvorpå alle 3 personer kan have fødselsdag den samme dag, navnligt hvis alle 3 har fødselsdag den samme dag. Der er derimod $3! = 6$ forskellige måder 3 mennesker kan have fødselsdag på henholdsvis 1, 2 og 3 januar per multiplaktionsreglen, theorem 1.4.1. Def. 1.3.1 (den naive definition af sandsynlighed) giver os nu, hvor vi udnytter at der er 365^3 forskellige måder at give 3 personer fødselsdage på (Vi tager ikke højde ikke for skudår. Vi tager heller ikke højde for alle septemberbørnene eller andre uregelmæssigheder i fødselsdage)

$$\frac{1}{365^3} < \frac{6}{365^3}.$$

Igen er vi klogere end Leibniz.

1.2.4 BH. 1.42

Let A and B be events. The difference B A is defined to be the set of all elements of B that are not in A. Show that if $A \subseteq B$, then

$$P(B - A) = P(B) - P(A).$$

Svar: (Husk tegning). Vi kigger på

$$P(A \cup B - A) = P(A) + P(B - A) - P(A \cap B - A).$$

Vi ser at $A \cup B - A = B$ da $A \subseteq B$ og $A \cap B - A = \varnothing$. Vi bruger at $P(\varnothing) = 0$ per def. Dette giver os

$$\begin{aligned} P(B) &= P(A) + P(B - A) \\ P(B - A) &= P(B) - P(A) \end{aligned}$$

som ønsket. Ofte bliver notationen $B \setminus A$ brugt i stedet for $B - A$. Denne notation er måske en smule mindre forvirrende idet man ikke kan blive i tvivl om det er et *mængdeminus* eller et *almindeligt minus*.

Alternatvt: (Husk tegning). Andet aksiom giver at, givet de to mængder er disjunkte, så er

$$P(C \cup D) = P(C) + P(D).$$

Vi ser først at mængderne $A \cap B$ og $B - A$ er disjunkte. Kan indses vha. af tegning eller at $B - A$ er B uden A, specielt er det uden den del af B som A deler med B. Vi får dermed at

$$P((A \cap B) \cup (B - A)) = P(A \cap B) + P(B - A).$$

Vi kan nu udnytte vores antagelse om at $A \subseteq B$. Således er $A \cap B = A$ og dermed har vi også at $A \cap B \cup B - A = A \cup (B - A) = B$. Dette giver os

$$\begin{aligned} P(B) &= P(A) + P(B - A) \\ P(B - A) &= P(B) - P(A) \end{aligned}$$

som ønsket.

1.2.5 BH. 1.43

Find den symmetriske differens

Svar: (Husk tegning). Vi skal altså finde sandsynligheden af $P(A \triangle B) = P(A \cup B - A \cap B)$. Vi indser at vi er det setup som opgave 1.42 siger noget om. Vi har derfor

$$P(A \triangle B) = P(A \cup B - A \cap B) = P(A \cup B) - P(A \cap B).$$

Videre bruger vi theorem 1.6.2, umiddelbare konsekvenser af definitionen af sandsynlighed, således $P(A \cup B) = P(A) + P(B) - P(A \cap B)$. Samlet får vi

$$P(A \triangle B) = P(A) + P(B) - P(A \cap B) - P(A \cap B) = P(A) + P(B) - 2P(A \cap B)$$

som ønsket.

1.2.6 BH. 1.36

Tyrion, Cersei, and ten other people are sitting at a round table, with their seating arrangement having been randomly assigned. What is the probability that Tyrion and Cersei are sitting next to each other? Find this in two ways:

A using a sample space of size 12!, where an outcome is fully detailed about the seating;

Svar: Vi tæller de gunstige udfald. Der er 12 pladser. Vi kan derfor sætte Tyrion på 12 forskellige pladser. Da Cersei skal sidde på en plads ved siden af Tyrion, er der 2 forskellige pladser Cerseo kan sidde på. Vi er ligeglade med resten, vi kan derfor sætte dem på 10! forskellige pladser per. example 1.4.7. Dette giver os per def. 1.3.1 (den naive definition af sandsynlighed) at vi har en sandsynlighed på

$$\frac{2 \cdot 12 \cdot 10!}{12!} = \frac{2}{11}.$$

B using a much smaller sample space, which focuses on Tyrion and Cersei.

Svar: Vi kan også tænke på problemet på en anden måde. Efter vi har sat Tyrion på en vilkårlig plads, har vi 11 forskellige pladser tilbage. Der er to pladser der gør at Cersei sidder ved siden af Tyrion, den til venstre og den til højre. Vi får derfor per def. 1.3.1 (den naive definition af sandsynlighed) at der er en sandsynlighed på

$$\frac{2}{11}.$$

Vi får dermed den samme sandsynlighed, hvilket er meget heldigt.

Alternativt: Vi kunne også kigget på alle konfigurationer Tyrion og Cersei kunne have siddet på, dvs. $\binom{12}{2}$. Videre er 12 konfigurationer hvor Tyrion og Cersei sidder vi siden af hinanden, $(1, 2), \dots, (11, 12), (12, 1)$. Dette ville give

$$\frac{12}{\binom{12}{2}} = \frac{2}{11},$$

så der er altså mange veje til Rom.

1.2.7 BH. 1.6

There are 20 people at a chess club on a certain day. They each find opponents and start playing. How many possibilities are there for how they are matched up, assuming that in each game it does matter who has the white pieces?

Vink: (1) Indse først at der er $20!/(2^{10} * 10!)$ muligheder hvis man man ikke skelner mellem hvem der spiller hvid hhv. sort. Tænk fx på følgende måde: På hvor mange måder kan første par udvælges, udtrykt ved en binomialkoefficient? Andet par, tredje par, osv? Skriv binomialkoefficienterne op, og se at ting forkorter ud. Husk derefter at rækkefølgen af parrene er ligegyldig. (2) Hvordan skal antallet korrigeres når man skelner mellem hvem der spiller hvid hhv. sort?

Svar: Vi bruger hintet. Det første par kan vælges på $\binom{20}{2}$ måder, den næste par kan vælges på $\binom{18}{2}$ måder. Helt generelt får vi $\binom{20-2i}{2}$, hvor $i = 0 \dots 9$. Dette kan skrives ud, leddene bliver $\frac{20!}{18!2!} \cdot \frac{18!}{16!2!} \cdot \dots \frac{2!}{2!}$ (hvor vi bruger multiplikationsreglen). Det ses at leddende går ud med hinanden og vi får $\frac{20!}{2!^{10}} = \frac{20!}{2^{10}}$. Med denne metode har vi dog valgt en specifik rækkefølge, vi har dermed overtalt, tænk på det som om vi har valgt at et specifikt par skulle sige på et specifikt bord. Der er 10 borde og dermed 10! forskellige måder at permutere disse borde. Alt i alt får vi

$$\frac{20!}{2!^{10} \cdot 10!}$$

som ønsket.

Dette var dog ikke helt det vi var ude i. Vi er også interesseret i hvilket sæt brikker personerne har. Per multiplikationsreglen er der 2^{10} måder at vælge hvem der har det sorte og det hvide sæt. Igen per multiplikationsreglen får vi at der er

$$\frac{20!}{2!^{10} \cdot 10!} \cdot 2^{10} = \frac{20!}{10!}$$

måder i alt.

Alternativt: Example 1.5.4 giver os også svaret. Så skal vi bare korrigere med 2^{10} da det betyder noget hvem der har hvid og sort.

2 Uge 2

2.1 Øvelser tirsdag den 28. november kl 15.15-17

2.1.1 SS.3-5

Find .R filen på min drev `https://drive.google.com/open?id=1pYa6Tm9qehn8M8TcDpclNA96jznVDipW`

SS.5.3 Svar: X er ligefordelt, dvs. alle sandsynligheder er lige sandsynlige med sandsynligheden 1/5 i dette tilfælde. Dette giver

$$P(X=2)=\frac{1}{5}.$$

Videre har vi pgså 1

$$P(X\leq 3)=P(X=1)+P(X=2)+P(X=3)=\frac{1}{5}+\frac{1}{5}+\frac{1}{5}=\frac{3}{5}.$$

2.1.2 BH 2.1

A spam filter is designed by looking at commonly occurring phrases in spam. Suppose that 80% of email is spam. In 10% of the spam emails, the phrase "free money" is used, whereas this phrase is only used in 1% of non-spam emails. A new email has just arrived, which does mention "free money". What is the probability that it is spam?

Svar: Vi har lige fået en mail hvor der står 'free money', hvad er sandsynligheden for det er spam? Oversat mere matematisk svarer dette til at finde P(spam—free), hvor free betyder der står 'free money' i mailen. Per Bayes kan vi skrive dette som

$$P(spam|free)=\frac{P(free|spam)P(spam)}{P(free)}.$$

P(free—spam) har vi allerede fået opgivet til at være 0.1. P(spam) har vi også fået opgivet. Vi kan finde P(free) vha. LOTP. Vi kan dele alle mails op i spam og ikke spam. Yderligere antages det at en mail kan ikke være og være spam på samme tid. LOTP giver da

$$P(free)=P(free|spam)P(spam)+P(free|spam^c)P(spam^c)=0.1\cdot 0.8+0.01\cdot(1-0.8)=0.082.$$

Vi ender med at få

$$P(spam|free)=\frac{0.1\cdot 0.8}{0.082}\approx 0.976.$$

2.1.3 BH 2.11

Consider an election with two candidates, Candidate A and Candidate B. Every voter is invited to participate in an exit poll, where they are asked whom they voted for; some accept and some refuse. For a randomly selected voter, let A be the event that they voted for A, and W be the event that they are willing to participate in the exit poll. Suppose that P(W—A) = 0.7 but P(W—Ac) = 0.3. In the exit poll, 60% of the respondents say they voted for A (assume that they are all honest), suggesting a comfortable victory for A. Find P(A), the true proportion of people who voted for A.

Svar: Først opsummere vi hvad vi ved. Vi har fået givet at $P(W\mid A)=0,7$, $P(W\mid A^c)=0,3$ og $P(A\mid W)=0,6$, da vi havde at 60% sagde de stemte på A i den givne rundspørgen.

Vi følger det givne hint. Hvis vi deler udfaldsrummet op i dem som har stemt på A og dem som ikke har, giver LOTP (Noter her at $A^c\cap A=\varnothing$ og $A^c\cup A=S$, hvor S betegner hele udfaldsrummet)

$$P(W)=P(W|A)P(A)+P(W|A^c)P(A^c).$$

Per Bayes har vi også

$$P(A|W) = \frac{P(W|A)P(A)}{P(W)}.$$

Vi har nu to ligninger med to ubekendte, P(W) og P(A) (husk på at $P(A^c) = 1 - P(A)$). Vi vil gerne kende P(A). Vi isolerer P(W) og sætter de to udtryk lig med hinanden og løser for P(A).

$$\begin{aligned}
P(W|A)P(A) + P(W|A^c)P(A^c) &= \frac{P(W|A)P(A)}{P(A|W)} \\
P(W|A)P(A) + P(W|A^c)(1 - P(A)) &= \frac{P(W|A)P(A)}{P(A|W)} \\
P(W|A)P(A) - P(W|A^c)P(A) + P(W|A^c) &= \frac{P(W|A)P(A)}{P(A|W)} \\
P(A)(P(W|A) - P(W|A^c)) + P(W|A^c) &= \frac{P(W|A)P(A)}{P(A|W)} \\
(P(W|A) - P(W|A^c)) + \frac{P(W|A^c)}{P(A)} &= \frac{P(W|A)}{P(A|W)} \\
\frac{P(W|A^c)}{P(A)} &= \frac{P(W|A)}{P(A|W)} - (P(W|A) - P(W|A^c)) \\
P(A) &= \frac{P(W|A^c)}{\frac{P(W|A)}{P(A|W)} - (P(W|A) - P(W|A^c))} = \frac{0.3}{\frac{0.7}{0.3 \cdot 0.6} - (0.7 - 0.3)} = 0.39.
\end{aligned}$$

Dette er en rimelig grim udregning, kunne også være lavet i maple/wolfram alpha. Resultatet viser at A faktisk har fået færrest stemmer til trods for at A havde havde vundet den givne exitpoll.

2.1.4 BH 2.30

A family has 3 children, creatively named A,B, and C.

A Discuss intuitively (but clearly) whether the event “A is older than B” is independent of the event “A is older than C”.

Svar: Du har 3! = 6 måder børnenes rækkefølge kan være. Vi antager alle måder er lige sandsynligheder. Disse kan opskrives

- ABC
- ACB
- BAC
- BCA
- CAB
- CBA

Vi ser at der er to måder hvorpå A både er ældre end B og C. Der er hver især 3 måder A bare er ældre end B og ligeså for C. Dette giver per def. af naiv sandsynlighed

$$P(B < A \cap C < A) = \frac{2}{6} = \frac{1}{3} \neq \frac{1}{4} = \frac{3}{6} \cdot \frac{3}{6} = P(B < A)P(C < A).$$

Vi har altså ikke $P(A \cap B) = P(A)P(B)$, og dermed er ikke afhængige per def 2.5.1.

B Find the probability that A is older than B, given that A is older than C.
Svar: Vi oversætter det til betingede sandsynligheder. Vi får

$$P(B < A|C < A).$$

Vi har allerede fået givet at $C < A$. I stedet for at kigge på hele udfaldsrummet, kan vi kigge på delrummet hvor $C < A$. Her har vi 3 kombinationer hvor hvor $B < A$ i 2 af dem. Per def. af naiv sandsynlighed får vi

$$P(B < A) = \frac{2}{3},$$

hvor P er et sandsynlighedsmål på delrummet hvor $C < A$.

2.1.5 BH 2.2

A woman is pregnant with twin boys. Twins may be either identical or fraternal (non- identical). In general, 1/3 of twins born are identical. Obviously, identical twins must be of the same sex; fraternal twins may or may not be. Assume that identical twins are equally likely to be both boys or both girls, while for fraternal twins all possibilities are equally likely. Given the above information, what is the probability that the woman's twins are identical?

Svar: Lad os betegne D for dreng og P for P. Oversat til en betinget sandsynlighed vil vi gerne finde ud af P(enæggede—DD). Bayes giver os

$$P(enaeggede|DD) = \frac{P(DD|enaeggede)P(enaegedde)}{P(DD)}.$$

Vi har fået opgivet P(enæggede)=1/3. Videre er det også givet at alle kombinationer er lige sandsynligheder, således P(DD—enæggede)=1/2, da der er to muligheder da kønnet skal være det samme. Vi bruger LOTP til at udregne DD, hvor vi deler udfaldsrummet op i enægget og tveægget tvillinger

$$\begin{aligned} P(DD) &= P(DD|enaeggede)P(enaeggede) + P(DD|tveaeggede)P(tveaeggede) \\ &= 1/2 \cdot 1/3 + 1/4 \cdot 2/3 = 1/6 + 2/12 = 1/3. \end{aligned}$$

Samet får vi

$$P(enaeggede|DD) = \frac{1/2 \cdot 1/3}{1/3} = \frac{1}{2}.$$

2.2 Øvelser fredag den 1. december kl 10.15-12

2.2.1 SS.6

1 X og Y er ikke uafhængige, f.eks. er $P(X = 2, Y = 1) = 0,15 \neq 0,1 = 0.2 \cdot 0,5 = P(X = 2)P(Y = 1)$. Måske sjovt, hvis man lavede en ny variabel Z som indtog værdien 1 for $X = 1$ og værdien 2 for $X = \{2, 4\}$, så ville Z og Y være uafhængige.

2 Per def. 4.1.1 BH får vi

$$E(X) = 1 \cdot 0,2 + 2 \cdot 0,2 + 4 \cdot \cdot 0,6 = 0,2 + 0,4 + 2,4 = 3.$$

2.2.2 SS.7

Find .R filen på min drev `https://drive.google.com/open?id=1ILHgeS8JJgbc6PsQMNqdslJZDz4ALoti`

1 Vi kan betragte alle 100 mennesker som popolationen, dvs. N=100. Vi har 38 piger, dette svarer til de elge som havde en mærkat på sig i elgeopgaven. Der er så 62 drenge og vi trækker 25 personer, dette bliver $X \sim HGeom(38, 68, 25)$

5 Der er en binomialfordeling med antalsparametre 38, da der er 38 piger, og en sandsynlighedsparameter 0,25. Y er binomialfordelt da hvert valg er en Bernouilli variable, hvor success er 0,25 (antaget at det er en nice fest Kasper holder).

2.2.3 SS.8

Find .R filen på min drev `https://drive.google.com/open?id=1ILHgeS8JJgbc6PsQMNqdslJZDz4ALoti`

1 S kan indtage værdierne 3-18. Nej, den er ikke ligefordelt, f.eks. er der kun en måde at få 3 på, navnligt (1,1,1). Derimod er der klart flere måder at få 8 på, f.eks. (6,1,1) og (5,2,1). Her ville man også bruge def. af naiv sandsynlighed, da man bruger at de forskellige kombinationer af triplen er lige sandsynlige.

2 Der er kun en måde at få 18 på, (6,6,6). Under antagelse at terningerne er uafhængige får vi

$$P(S = 18) = P(X_1 = 6, X_2 = 6, X_3 = 6) = 1/6^3 = 1/216.$$

Vi har lidt flere måder at få 5 på. Vi kan have (3,1,1) og (2,2,1). da vi er ligeglade med hvilken specifik terning der har slået hvad får vi at der 6 måder i alt (3 forskellige måder af hver). Vi får derfor

$$P(S = 5) = 6/6^3 = 1/36.$$

3 Givet summen er 5, er der 6 forskellige muligheder. Disse kan vi opskrive, (1,1,3),(1,2,2),(1,3,1),(2,1,2),(2,2,1),(3,1,1). Der er 3 af dem hvor $X_1 = 1$, derfor

$$P(X_1 = 1|S = 5) = 3/6.$$

2.2.4 BH 3.7

Svar: PMF, probability mass function, sandsynlighedsfunktionen for en **diskret** stokastisk variabel. Der er p_1 sandsynlighed for at Bob går videre til den næste bane. Dette betyder at sandsynligheden for at Bob ikke når videre er $1 - p_1$. Dette kan skrives som $P(X = 1) = p_1$. Videre er sandsynligheden for at Bob når videre fra første bane, men ikke videre fra anden bane $P(X = 2) = p_1(1 - p_2)$. Dette kan udvides og vi får

$$P(X = j) = p_1 p_2 \cdots p_{j-1}(1 - p_j)$$

for $2 \leq x \leq 6$. For x=7 bliver det, da der ikke er flere baner

$$P(X = 7) = p_1 p_2 \cdots p_6.$$

Samlet får vi

$$P(X = j) = \begin{cases} 1 - p_1, & \text{for } j = 1 \\ p_1 p_2 \cdots p_{j-1}(1 - p_j), & \text{for } 2 \leq j \leq 6 \\ p_1 p_2 \cdots p_6, & \text{for } j = 7. \end{cases}$$

2.2.5 BH 3.17

Svar: Der er en Bernouilli fordeling med 110 antalparametre (antal personer) og en successandsynlighed på 0,9. Dette giver os per 3.3.5

$$P(X \leq 100) = \sum_{k=0}^{100} \binom{110}{k} 0,9^k 0,1^{110-k} = 0,671$$

Vi regnede det specifikke kvantil ud i R. Find .R filen på mit drev `https://drive.google.com/open?id=1ILHgeS8JJgbc6PsQMNqdslJZDz4ALoti`

3 Uge 3

3.1 Øvelser tirsdag den 28. november kl 15.15-17

3.1.1 SS.10

1 Vi bruger theorem 5.1.5, den skal opfylde to kriterier. Det er klart at $x/6$ er positiv for $x \in (2,4)$. Vi skal også vise at den integrerer til 1. Vi udregner integralet

$$\int_{-\infty}^{\infty} f = \int_2^4 \frac{x}{6} = \frac{1}{6}\left[\frac{x^2}{2}\right]_2^4 = \frac{1}{6}(8-2) = 1,$$

som vi ønskede. Bemærk at integralet er defineret over hele $\mathbb{R}$, men den har kun support i $x \in (2,4)$.

2 Vi vil gerne udregne $P(3 < x < 3,5)$. Vi udregner integralet i dette interval.

$$\int_3^{7/2} \frac{x}{6} = \frac{1}{6}\left[\frac{x^2}{2}\right]_3^{7/2} = \frac{1}{6}\left(\frac{7^2}{2^3} - \frac{3^2}{2}\right) = \frac{13}{3 \cdot 2^4}.$$

3 Vi beregner EX, def. 5.1.9

$$E(X) = \int_{-\infty}^{\infty} x \cdot f = \int_2^4 \frac{x^2}{6} = \frac{1}{6}\left[\frac{x^3}{3}\right]_2^4 = \frac{1}{6}\left(\frac{4^3}{3} - \frac{2^3}{3}\right) = \frac{28}{9}.$$

3.1.2 BH. 4.3

a De er ligefordelt, vi får dermed per 4.1.1

$$\begin{aligned} E(X) &= 1 \cdot P(X=1) + 2 \cdot P(X=2) + 3 \cdot P(X=3) + 4 \cdot P(X=4) + 5 \cdot P(X=5) + 6 \cdot P(X=6) \\ &= \frac{1}{6}(1+2+3+4+5+6) = \frac{21}{6} = 3,5. \end{aligned}$$

Eksemplet ses også på s.138.

b Givet at de er uafhængige, får vi per linearitet af middelværdi, hvor vi kalder terning 1,2,3,4 for X_1, X_2, X_3 og X_4, at

$$E(X_1 + X_2 + X_3 + X_4) = 4E(X_1) = 4 \cdot 3,5 = 14$$

hvor vi brugte at alle de 4 terninger alle er ligefordelte.

3.1.3 SS.11

1 Vi indskriver det i en tabel, vi kan bare gange de 2 saandsynligheder sammen da de er antaget uafhængige.

	Y=0	Y=1
X=0	(1-p)/2	(1-p)/2
X=1	p/2	p/2

2 Vi vil gerne finde P(Z=0) og P(Z=1). Vi kigger i skemaet og får

$$P(Z=0) = (1-p)/2 + p/2 = 1/2 \quad \text{og} \quad P(Z=1) = (1-p)/2 + p/2 = 1/2.$$

Z er altså også en Bernouilli variabel med successandsynlighed 1/2.

3 Vi kan kigge i tabellen og vi får

$$\begin{aligned}
P(X=0,Z=0)=P(X=0,Y=0)=(1-p)/2=(1-p)\cdot 1/2=P(X=0)P(Z=0)\\
P(X=1,Z=0)=P(X=0,Y=1)=p/2=p\cdot 1/2=P(X=1)P(Z=0)\\
P(X=0,Z=1)=P(X=0,Y=1)=(1-p)/2=(1-p)\cdot 1/2=P(X=0)P(Z=0)\\
P(X=1,Z=1)=P(X=1,Y=1)=p/2=p\cdot 1/2=P(X=0)P(Z=0),
\end{aligned}$$

vi får dermed at Z og X er uafhængige.

4 Da Y og Z begge har successandsynlighed på 1/2 får vi at produktet er 2 sandsynligheder er 1/4. Vi skal dermed tjekke om P(Y=y,Y=z)=1/4. Vi kan igen referer til tabellen og vi får følgende ligninger

$$P(Y=0,Z=0)=P(Y=1,Z=1)=(1-p)/2 \quad \text{og} \quad P(Y=0,Z=1)=P(Y=1,Z=0)=p/2.$$

Disse ligninger er lig med 1/4 hvis og kun hvis p=1/2, hvilket var det vi ville vise.

3.1.4 SS.12

1 I leksiografisk orden får vi $(1,1),(1,2),\ldots,(2,1),(2,2),\ldots,(5,5)$. Der er 25 forskellige muligheder. Da terninger er ligefordelte og uaghængie får vi at hvert par en sandsynlighed på 1/25. Dette er tilsvarende nogle af de eksempler vi har set med 6 sidet terninger.

2 Z kan indtage værdier $Z=\{0,1,2,3,4\}$.

3 Den mest naive måde at løse opgaven på er simpelt at tælle de forskellige kombinationer. Så får vi at P(Z=0)=5/25, P(Z=1)=8/25, P(Z=2)=6/25, P(Z=3)=4/25 og P(Z=4)=2/25 (lav selv evt. en tabel). Dette bliver

$$P(Z=z)\begin{cases} 5/25, & \text{hvis } z=0\\ 8/25, & \text{hvis } z=1\\ 6/25, & \text{hvis } z=2\\ 4/25, & \text{hvis } z=3\\ 2/25, & \text{hvis } z=4, \end{cases}$$

hvor vi også ser at summen er lig med 1 og de er positive, det er dermed en lovlig sandsynlighedsfunktion.

4 Tegn selv. (Husk der er fordelingsfunktionen, ikke tætheden.)

5 Vi får

$$\begin{aligned}
E(Z)&=0\cdot P(Z=0)+1\cdot P(Z=1)+2\cdot P(Z=2)+3\cdot P(Z=3)+4\cdot P(Z=4)\\
&=0+8/25+12/25+12/25+8/25=40/25=8/5.
\end{aligned}$$

3.1.5 SS.13

1-2 Se mit google drev, link: `https://drive.google.com/open?id=1SQonew-aJ_yUjkDoAmReR67HK-p-LxV8`

3 Kast en terning en gang. Dette er bare gennemsnittet af en enkelt terning som vi passende kan kalde X. Dette bliver:

$$\begin{aligned}
E(X)&=1\cdot P(W=1)+2\cdot P(W=2)+3\cdot P(W=3)+4\cdot P(W=4)+5\cdot P(W=5)\\
&=\frac{1}{5}(1+2+3+4+5)=\frac{15}{5}=3.
\end{aligned}$$

4-6 Se mit google drev, link: `https://drive.google.com/open?id=1SQonew-aJ_yUjkDoAmReR67HK-p-LxV8`

3.1.6 Ekstra: BH 4.31

Vi laver bare første del. Vi følger hintet. Givet at fødselsdagene er uafhængige og ligefordelte vi har sandsynligheden 1/365 for at et givent par fødselsdag samme dag. (Den første person har fødselsdag en dag, der er så 1/365 chance for at den anden person har den samme dag). Videre har vi $\binom{50}{2}$ forskellige par. Vi bruger vores uafhængighedsantagelse og lineraitet af middelværdi giver

$$\binom{50}{2}\frac{1}{365} \approx 3,36,$$

så vi har lidt over 3 par i gennemsnit.

Anden del, ikke gennemgået til øvelser. Skitse Lav en indikator for hver dag, 365 hver dag. Find sandsynligheden for hver af disse dage. Det er lettere at regne på komplementet, således vi skal regne $1 - P(A_1)$, hvor A er ingen eller 1 fødselsdag på den givne dag. 1 kunne f.eks. så symbolisere 1 januar. Vi har 365 af disse, således sandsynligheden bliver (hvor vi antager at fødselsdagene er ligefordelte på alle dage) 365(1-P(A)). Svaret vil blive: $365(1 - (\frac{50 \cdot 364^{49}}{395^{50}} + (\frac{364}{365})^{50}))$.

3.2 Øvelser fredag den 7. december kl 10.15-12

3.2.1 SS.14

1 Se google drev: `https://drive.google.com/open?id=1ILHgeS8JJgbc6PsQMNqdslJZDz4ALoti`. Man kunne også være lidt hurtig og lave et relativt godt estimat med 68-95-99,7 reglen (theorem 5.4.5). Dette giver, hvor vi benytter at $X \sim \mathcal{N}(10, 9)$

$$0,95 \approx P(|X - 10| < 2 \cdot 3) = P(4 < X < 16),$$

således bliver a=4 og og b=16. Skitse: Tegn selv, husk på symmetrien og hvad det vi lige har regnet betyder.

2-3 Se google drev: `https://drive.google.com/open?id=1ILHgeS8JJgbc6PsQMNqdslJZDz4ALoti`

4 Dette får vi til

$$Y = 2X + 1 \approx \mathcal{N}(2 \cdot 10, 2^2 \cdot 9) + 1 = \mathcal{N}(21, 36).$$

Bemærk den lille smule abuse of notation i det vi plusser med 1. Husk vi skal gange 2 ind i middelværdien og 2 kvadreret i variansen. Vi plusser med 1, det ændrer ikke på variansen.

Vi får altså middelværdien til 21, variansen til 36 og spredning på 6.

3.2.2 SS.15

1 Vi bestemmer først EX.

$$EX = 1 \cdot 0,1 + 3 \cdot 0,4 + 5 \cdot 0,5 = 3,8.$$

Vi bestemmer så VX. Husk på $VX = EX^2 - (EX)^2$. Vi bruger LOTUS for at beregne EX^2.

$$EX^2 = 1^2 \cdot 0,1 + 3^2 \cdot 0,4 + 5^2 \cdot 0,5 = 16,2.$$

Vi får da VX til

$$VX = EX^2 - (EX)^2 = 16,2 - 3,8^2 = 1,76.$$

Spredningen bliver da $SD = \sqrt{VX} = \sqrt{1,76} \approx 1,33$.

2 LOTUS giver

$$E(X-2)^4 = (1-2)^4 \cdot 0,1 + (3-2)^4 \cdot 0,4 + (5-2)^4 \cdot 0,5 = 41.$$

3.2.3 BH 5.8

a CDF er givet ved, hvor vi benytter at en beta fordeling har support på $x \in (0,1)$

$$F(x) = \int_{-\infty}^{x} f = \int_{0}^{x} 12x^2(1-x) = \int_{0}^{x} 12x^2 - \int_{0}^{x} 12x^3 = 4x^3 - 3x^4 = 12\left(\frac{x^3}{3} - \frac{x^4}{4}\right).$$

b Find P(0¡X¡1/2). Vi kan bruge den CDF vi lige har udregnet

$$P(0 < X < 1/2) = F(1/2) - F(0) = 5/16 = 0,3125.$$

c Vi finder EX

$$EX = \int_{-\infty}^{\infty} xf = \int_{0}^{1} x(12x^2(1-x)) = \int_{0}^{1} 12x^3 - \int_{0}^{1} 12x^4 = 3 - 12/5 = 3/5 = 0,6.$$

Vi finder VX, først EX^2 per LOTUS

$$EX^2 = \int_{-\infty}^{\infty} x^2 f = \int_{0}^{1} x^2(12x^2(1-x)) = \int_{0}^{1} 12x^4 - \int_{0}^{1} 12x^5 = 12/5 - 2 = 2/5 = 0,4.$$

Vi får dermed VX til

$$VX = EX^2 - (EX)^2 = 2/5 - (3/5)^2 = 1/25 = 0,04.$$

Bonus: Def. af en beta fordeling:

$$\frac{x^{\alpha-1}(1-x)^{\beta-1}}{\mathrm{B}(\alpha,\beta)}$$
$$\text{hvor } \mathrm{B}(\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

3.2.4 BH 4.3 videre

Ekstra 1 (En terning) Vi finder EX^2 for at finde varians per LOTUS får vi

$$\begin{aligned} EX^2 &= 1^2 \cdot P(X=1) + 2^2 \cdot P(X=2) + 3^2 \cdot P(X=3) + 4^2 \cdot P(X=4) + 5^2 \cdot P(X=5) + 6^2 \cdot P(X=6) \\ &= \frac{1}{6}(1+4+9+16+25+36) = \frac{91}{6} \approx 15,167. \end{aligned}$$

Vi får nu variansen til, hvor vi benytter os af 4.3 fra i tirsdags.

$$VX = EX^2 - (EX)^2 - 15,167 - (3,5)^2 = 35/12 \approx 2,92.$$

Spredningen bliver da $SD = \sqrt{VX} = \sqrt{2,92} \approx 1,71$.

Ekstra 2 (4 terninger) Vi vil nu gerne finde variansen. Dvs, hvor vi bruger S om summen af de 4 terninger ligesom i tirsdags

$$V(S) = V(X_1 + X_2 + X_3 + X_4) = V(X_1) + V(X_2) + V(X_3) + V(X_4) = 4V(X_1) = 4 \cdot 2,92 = 11,68,$$

hvor vi brugte at alle terningerne var uafhængige således vi kunne linearitet af variansen.

3.2.5 SS.16

1 Vi husker på eksponentialfunktionen, $f(x) = \lambda e^{-\lambda x}$. (Overlevelsesfunktion/hldbarhedsfunktion: Hvor er sandsynligheden for at en patient/ting holder til et vist tidspunkt). Vi udregner P(X¿x). Dette giver

$$P(X > x) = \int_x^\infty f = \lim_{b\to\infty} \int_x^b \lambda e^{-\lambda x} = \lim_{b\to\infty} \lambda \left[\frac{e^{-\lambda x}}{-\lambda}\right]_x^b = \lim_{b\to\infty} e^{-\lambda b} + e^{-\lambda x} = e^{-\lambda x}.$$

Dette viser altså den memoryless property som eksponentialfordelingen besidder. Lige meget hvor lang tid der går, er der stadig samme sandsynlighed for en given ting holder.

2 Z har samme støtte som X og Y, dvs $(0, \infty)$. Vi vil gerne finde P(Z¿z). Da Z er defineret som minimum af have specielt have at

$$P(Z > z) = P(\min(X, Y) > z) = P(X > z, Y > z) = P(X > z)P(Y > z) = P(X > z)^2 = (e^{-\lambda z})^2 = e^{-2\lambda z},$$

hvor vi brugte at X og Y er uafhængige, vi får derfor at Z også er eksponentialfordelt også.

3 Vi finder fordelingen, CDF. Hvis vi kalder fordelingen af Z for g(z) får vi

$$F(Z) = P(Z \leq z) = 1 - P(Z > z) = 1 - e^{-2\lambda z}$$

Dette kunne man også have slået op i starten af starten af afsnit 5.5.

4 W har støtte på $(0, \infty)$. Vi finder fordelingens funktionen af W

$$F(w) = P(W \leq w) = P(\max(X, Y) \leq w) = P(X \leq w, Y \leq w) = P(Z \leq w)P(Y \leq w) = (P(X \leq w))^2 = \left(1 - e^{-\lambda w}\right)^2,$$

hvor vi brugte at X og Y er uafhængige og da vi kendte CDF'en for en alm. eksponentialfordelingen. Dette viser sig faktisk at være lig med $P(X + Y/2 < w)$, hvilket måske (måske ikke) er sjovt.

Vi får også af vide, at vi skal bestemme tætheden. Vi differentierer CDF'en således vi får PDF, dvs. sandsynlighedstætheden eller bare tætheden.

$$F'(w) = \left(\left(1 - e^{-\lambda w}\right)^2\right)' = 2(1 - e^{-\lambda w}) \cdot \lambda e^{-\lambda w} = 2\lambda e^{-2\lambda w}(e^{\lambda w} - 1),$$

men pænere kan det heller ikke gøres.

3.2.6 Ekstra: BH 4.7.

a Vi kalder variablen X. Vi finder PMF vha. LOTP. Vi betegner $1f$ for 1 barns familier og 1 for førstefødte barn, tilsvarende for 2 og 3. Vi ser at $1f \cup 2f \cup 3f = S$, da alle familier er en af de to typer, yderligere $1f \cap 2f \cap 3f = \varnothing$, da en familie ikke kan være to typer på samme tid. LOTP giver da

$$P(1) = P(1|1f)P(1f) + P(1|2f)P(2f) + P(1|3f)P(3f) = 1 \cdot 3/10 + 1/2 \cdot 5/10 + 1/3 \cdot 2/10 = 37/60$$
$$P(2) = P(2|1f)P(1f) + P(2|2f)P(2f) + P(2|3f)P(3f) = 0 \cdot 3/10 + 1/2 \cdot 5/10 + 1/3 \cdot 2/10 = 19/60$$
$$P(3) = P(3|1f)P(1f) + P(3|2f)P(2f) + P(3|3f)P(3f) = 0 \cdot 3/10 + 0 \cdot 5/10 + 1/3 \cdot 2/10 = 4/60.$$

Vi ser også her at det en lovlig PMF da summen er lig med 1 og alle punktsandsynligheder er positive, rimelig heldigt (lav evt. en gaffelfunktion). Vi regner EX ud

$$EX = 1 \cdot P(1) + 2 \cdot P(2) + 3 \cdot P(3) = 37/60 + 2 \cdot 19/60 + 3 \cdot 4/60 = 29/20 = 1,45.$$

Vi finder også VX, først EX^2.

$$EX^2 = 1^2 \cdot P(1) + 2^2 \cdot P(2) + 3^2 \cdot P(3) = 37/60 + 2^2 \cdot 19/60 + 3^3 \cdot 4/60 = 221/60 \approx 3,68.$$

Vi får da VX til

$$VX = EX^2 - (EX)^2 = 221/60 - (29/20)^2 = 1897/1200 \approx 1,58.$$

b Vi vælger nu et tilfældigt barn (intuition: Det giver mening at P(X = 1) fra (a) er mindre end P(Y = 1), siden metoden fra (a) giver alle familier lige sandsynlighed, hvor metoden i (b) giver børnene lige sandsynlighed, hvilket gør at der er større chance de kommer fra en stor familie (da der er flere børn i dem)). Vi kalder variablen Y.

Der er $30 + 2 \cdot 50 + \cdot 3 \cdot 30 = 190$ børn. Der 100 børn som er født først, 70 børn som er født nr.2 og 20 som er født nr.3. Da vi har antaget det tilfældigt giver naiv sandsynlighed at

$$
\begin{aligned}
P(1) &= 100/190 = 10/19 \\
P(2) &= 70/190 = 7/19 \\
P(3) &= 20/190 = 2/19.
\end{aligned}
$$

(lav evt. en gaffelfunktion) Vi kan udregne EY ud fra dette.

$$EY = 1 \cdot 10/19 + 2 \cdot 7/19 + 3 \cdot 2/19 = 30/19 \approx 1,58.$$

Videre, som altid, udregner vi EY^2.

$$EY^2 = 1^2 \cdot 10/19 + 2^2 \cdot 7/19 + 3^2 \cdot 2/19 = 56/19 \approx 2,95.$$

Vi finder VY

$$VY = 56/19 - (30/19)^2 = 164/361 \approx 0,45.$$

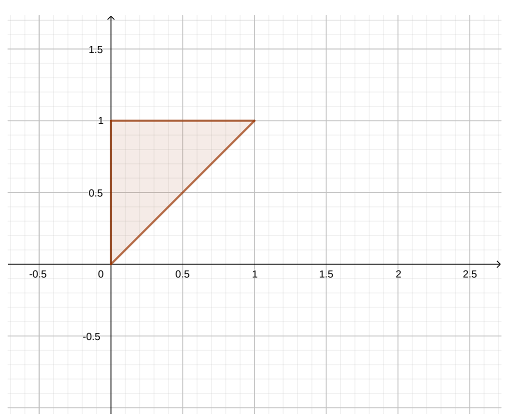
4 Uge 4

4.1 Øvelser tirsdag den 11. december kl 15.15-17

4.1.1 Revideret BH 7.17 (Laves til øvelser)

Lad $f_{X,Y}(x,y) = cxy$, for $0 < x < y < 1$. Bemærk den gemte antagelse at $c > 0$

i Tegn det område hvor f(x,y)¿0. Dette bliver: (bemærk at linjerne rent faktisk ikke er med da vi har skarpe uligheder, ikke fordi det rent praktisk betyder noget)



ii Vi skal vise at f er en tæthed hvis $c = 8$. Vi opskriver integralet

$$\int_0^1 \int_x^1 8xy\ dydx = 8x\left[1/2y^2\right]_x^1 = \int_0^1 8x \cdot (1/2 - 1/2x^2\ dx = \int_0^1 4x - 4x^3)\ dx = \left[2x^2 - x^4\right]_0^1 = 2 - 1 + 0 - 0 = 1,$$

det er dermed en tæthed.

iii Vi husker på at vi kan finde den marginale tæthed for X ved at differentierere med hensyn til Y og vice versa. Vi har allerede fået X's tæthed gratis i (ii). Vi får derfor

$$f_X(x) = 4x - 4x^3,$$

for $0 < x < 1$. Y's bliver (bemærk vi nu har skiftet parametrisering):

$$f_Y(y) = \int_0^y 8xy \; dx = y \left[4y^2\right]_0^y = 4y^3.$$

4.1.2 SS.17

1-3 Se google drev: `https://drive.google.com/open?id=1BoDlWKlXGArDpOPUTWWOhSD3YcbCZhGJ`

4.1.3 BH 5.5

a Areal af en cirkel er givet ved $A = \pi R^2$. Vi avender LOTUS

$$E(A) = \int_0^1 \pi r^2 = \pi \left[1/3r^3\right]_0^1 = \pi/3. \tag{1}$$

Middelværdien er altså $\pi/3$. Vi udregner også variansen. Hertil skal vi bruge andet momentet. Vi kan igen bruge LOTUS. Her bruger vi LOTUS på hele A således alt bliver kvadreret. Vi får

$$E(A^2) = \pi^2 \int_0^1 r^4 = \pi^2 \left[1/5r^5\right]_0^1 = \pi^2/5.$$

Samlet får vi variansen til:

$$V(A) = E(A^2) - (E(A))^2 = \pi^2/5 - (\pi/3)^2 = 4\pi^2/45.$$

b Vi finder først CDF, fordelingsfunktionen. Vi bemærker at A har support i $(0, \pi)$.

$$P(A \leq a) = P(\pi R^2 \leq a) = P(R \leq \sqrt{a/\pi}) = \sqrt{a/\pi},$$

hvor vi brugte at sandsynligheden er propertionelt med længden. Dette er meget let at regne for $X \sim Unif(0,1)$. Se eksempel 5.2.4. Vi får altså at fordelingsfunkitionen er

$$F(a) = \begin{cases} 0, & \text{for } a < 0 \\ \sqrt{a/\pi}, & \text{for } 0 \leq a \leq \pi \\ 1 & \text{for } a > 1. \end{cases}$$

Vi kan nu finde PDF'en, sandsynlighedstætheden, ved at differentiere CDF'en.

$$F'(a) = f(a) = (\sqrt{a/\pi})' = \frac{1}{2\sqrt{a\pi}},$$

for a i supporten, 0 ellers.

4.1.4 SS.18

Se google drev: `https://drive.google.com/open?id=1BoDlWKlXGArDpOPUTWWOhSD3YcbCZhGJ`

4.1.5 BH 5.21

1 $Z \sim \mathcal{N}(0,1)$. Vi kan gøre ligesom i tirsdags. Betragt $Y = 2Z + 1$. Denne vil $Y \sim \mathcal{N}(1,4)$. Vi kan udregne middelværdi og varians for at være sikker. Middelværdi:

$$EY = E(2Z + 1) = 2EZ + E1 = 0 + 1 = 1,$$

hvor vi brugte at Z har middelværdi 0. Videre er varians givet ved

$$Var(Y) = Var(2Z + 1) = Var(2Z) = 4Var(Z) = 4,$$

hvor vi brugte at Z har varians 1.

4.2 Øvelser fredag den 14. december kl 10.15-12

4.2.1 SS.19 (Laves til øvelser)

1 Da X og Y er uafhængige, kan vi gange de to tætheder sammen og få den simultane. Da Y er 1/2 på $(0, 2)$ og 0 alle andre steder, får vi at (X,Y) kun har positiv tæthed på $(0, 2)$. Videre er $f(x) = e^{-x}$ på hele $\mathbb{R}_+$. Dette giver

$$f_{X,Y}(x, y) = \begin{cases} 1/2e^{-x}, & \text{hvis } y \in (0, 2), x \in (0, \infty), \\ 0 \text{ ellers.} \end{cases}$$

2 Vi opsrkiver integralet og bruger 2D LOTUS

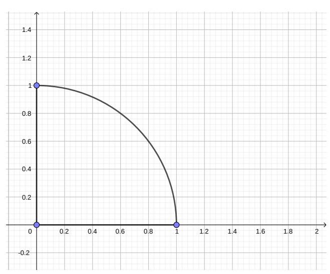
$$EZ = E(\log(X + Y)) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \log(x + y)f_{X,Y}(x, y)\ dxdy = \int_0^{\infty}\int_0^{2} \log(x + y)e^{-x}\ dydx.$$

Vi behøves ikke at regne det, så vi stopper her.

3 Vi har fået givet

$$f(u, v) = \begin{cases} c \cdot e^{-u}e^{-v^2/2} & \text{hvis } 0 < u < 1, 0 < v < 1, 0 < u^2 + v^2 < 1 \\ 0 & \text{ellers} \end{cases}$$

Vi ser at vi kan faktorisere f(u,v) til to funktioner som kun afhænger u henholdsvis v, men det gælder ikke for alle u og v, de er dermed ikke uafhængige.



4 Normalfordelinger er super seje i det de fungerer som man naivt tror fordeligner fungerer. Vi får dermed at

$$Z = X_1 + X_3 \sim \mathcal{N}(2 + 2, 9 + 9) = \mathcal{N}(4, 18).$$

Vi adderer simpelthen bare middelværdierne og variansen sammen (example 6.6.3).

4.2.2 SS.20

Vi har fået givet $\mathrm{E}(X) = 1, \quad \mathrm{Var}(X) = 2, \quad \mathrm{E}(Y) = 3, \quad \mathrm{Var}(Y) = 4.$

1 Vi vil gerne udregne midelværdi og varians af $Z = 2X - 3Y$.

$$EZ = E(2X - 3Y) = 2EX - 3Y = 2 - 3 \cdot 3 = -7,$$

hvor vi brugte uaghængighed af integralet. Man kunne evt godt bare udregne andet momentet af så køre videre derfra, men da X og Y er uafhængige er variansen lineær. Vi får da

$$Var(Z) = Var(2X - 3Y) = 4Var(X) + 9Var(Y) = 4 \cdot 2 + 9 \cdot 4 = 44.$$

2 Hvis X=Y får vi

$$Var(X+Y) = Var(2X) = 4Var(X).$$

Videre får vi at

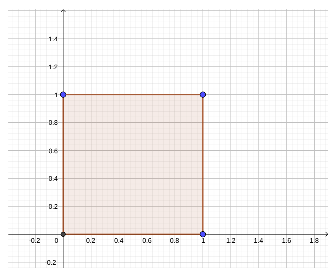
$$Var(X) + Var(X) = 2Var(X).$$

Bemærk at X og X faktisk godt kan være uafhængig med sig selv hvis X har en degeneret fordeling, dvs. X=c er 1 for $c \in \mathbb{R}$.

4.2.3 BH 7.16

dog ikke spørgsmål (d). Start med at tegne området hvor f(x,y)¿0, og lav derefter spørgsmålene i rækkefølgen (a), (c), (b). Bemærk at vi først snakker om uafhængighed fredag, men kig gerne på resultatet lige under Definition 7.1.18 (side 290).

a Vi tegner først hvor $f(x, y) > 0$, dette bliver: Vi ser at $f_{X,Y}$ er positiv og vi håber den integrerer til 1



(Spoiler: det gør den)

$$\int_0^1 \int_0^1 x + y = 1/2 + 1/2 = 1,$$

det er altså en lovlig PDF.

c Vi ser at der er symmetri i f, vi nøjes derfor bare at regne den ene marginale tæthed.

$$f_X(x) \int_0^1 x + y \; dy = x + 1/2$$

og analogt $f_Y(y) = y + 1/2$.

b Vi har ikke at $f_X \cdot f_Y = f_{X,Y}$. De er derfor ikke uafhængige.

4.2.4 SS.21

Vi har fået givet funktionen:

$$f(x) = \begin{cases} \frac{1}{2}x^{-3/2} & \text{for } x > 1 \\ 0 & \text{ellers} \end{cases}$$

1 Vi viser det er en sandsynlighedstæthed.

$$\int_1^\infty 1/2x^{-3/2} = 1/2\left[-2/\sqrt{x}\right]_1^\infty = \lim_{b\to\infty} -1/\sqrt{b} + 1 = 1$$

2 Vi vil gerne udregne

$$P(X < x) = \int_1^x 1/2x^{-3/2} = 1/2\left[-2/\sqrt{x}\right]_1^x = 1 - x^{-1/2}.$$

på X's support. Den er 0 ellers.

3 Vi har ligningen $1 - x'^{-1/2} = 0,5$. Vi får da $x = 4$.

4 Pareto fordelingen er altid god til at komme med modeksempler da det opfører sig rimelig iregulært.

$$\int_{-\infty}^{\infty} xf(x)\ dx = \int_1^{\infty} 1/2x^{-1/2} = \left[\sqrt{x}\right]_1^{\infty} = \lim_{b\to\infty} \sqrt{b} - \sqrt{1} = \infty.$$

Den har dermed ingen middelværdi. Bemærk at dette også betyder at den ej heller har andre momenter, specielt har den heller ikke varians.

5 For $y \in (0,1)$ får v

$$\begin{aligned} y &= 1 - 1/\sqrt{x} \\ y\sqrt{x} &= \sqrt{x} - 1 \\ \sqrt{x}(y-1) &= -1 \\ F^{-1}(y) &= 1/(1-y)^2. \end{aligned}$$

Hvorfor kan vi gøre dette? 5.3.1, univsersality of the norm

6-7 Se google drev: `https://drive.google.com/open?id=1zq7K-08TNcxSAhtGxVJ5pxCKxUCWD5fo`.

Ekstra Pareto fordelingen er givet ved $\frac{\alpha x_m^\alpha}{x^{\alpha+1}}$, hvor $x_m > 0$ og $a > 0$. I vores tilfælde er $a = 1/2$ og $x_m = 1$.

4.2.5 Tæthed på side 9 i et notat

Vi har fået givet:

$$\tilde{f}(x,y) = \begin{cases} 3y & \text{for } 0 < x < y < 1 \\ 0 & \text{ellers} \end{cases}$$

1 Vi udregner først den marginale tæthed for X. Dette bliver

$$f_X(x) = \int_x^1 3y\ dy = 3/2 - 3/2x^2$$

og

$$f_Y(y) = \int_0^y 3y\ dx = 3y^2.$$

5 Uge 5

5.1 Øvelser tirsdag den 18. december kl 15.15-17

5.1.1 SS.23

1 Vi bruger LOTUS

$$\int_{-\infty}^{\infty} 1/(y+1)f(y)dy = \int_0^1 1/(y+1)dy = [log(1+y)]_0^1 = \log(2)$$

2 Da X og Y er uafhængige er alle rimelige funktioner af X og Y uafghængige. Dette giver da

$$E(X+1/Y+1) = E(X+1)\cdot E(1/Y+1)$$

Vi har allerede regnet det andet led i (1). Det første led bliver 3/2, da X er standard uniform. Vi får

$$3/2\log(2).$$

3-5 Da X og Y er uafhængige får vi den simultane tæthed til at være

$$F_{X,Y}(x,y) = \int_0^1 \int_0^1 1dydx.$$

Vi bruger LOTUS til at finde følgende (Overvej hvad der sker når henholdsvis X og Y er størt)

$$\begin{aligned} E(|X-Y|) &= \int_0^1 \int_0^1 |x-y|dydx = \int_0^1 \left(\int_0^x x-ydy + \int_x^1 y-xdy\right)dx = \int_0^1 (x^2/2)+x^2/2-x+1/2dx \\ &= \int_0^1 x^2-x+1/2dx = 1/3. \end{aligned}$$

6 Se google drev: `https://drive.google.com/file/d/1acrygNQayz53JSXbmP1echdW2CzguGO9/view?usp=sharing`

5.1.2 BH. 8.2 (Laves til øvelser)

1 Vi vil gerne finde PDF'en for X^7 hvor $X \sim exp(\lambda)$. Vi bruger Change of variable/transformationssætningen. $Y = X^7$, $y = x^7$ og $x = y^{1/7}$ (8.1.1 s.341)

$$f_X(y) = f_X(g^{-1}(y))|(g^{-1}(y))'| = \lambda e^{-\lambda y^{1/7}} \cdot 1/7y^{-6/7} = \lambda/7 \cdot y^{-6/7}e^{-\lambda y^{1/7}}$$

5.1.3 SS.22 (Laves til øvelser)

1 Vi bruger transformationssætningen. Bemærk at $X = \sin(U)$, $x = \sin(u)$ og $\sin^{-1}(x) = u$. Bemærk at det er en lovlig transformation da sinus er voksende på intervallet.

$$f_X(x) = f_X(\sin^{-1}(x))|(\sin^{-1}(x))'| = 2/\pi \cdot 1/\sqrt{1-x^2} = 2/(\pi\sqrt{1-x^2}).$$

Bemærk at U har support i $(0, \pi/2)$ således X har support i $(0,1)$. $\sqrt{1-x^2}$ er dermed altid positiv.

2 Vi finder EX på to møder. Først vha. tætheden vi lige har udregnet:

$$EX = \int_0^1 2x/(\pi\sqrt{1-x^2})$$

Vi gætter på en stamfunktion $-\sqrt{1-x^2}$ og se den differentieret giver det vi gerne vil have. Evt substituition med $1-x^2$. Dette giver

$$EX = 2/\pi[-\sqrt{1-x^2}]_0^1 = 2/\pi.$$

Vi bruger nu LOTUS

$$E(g(U)) = EX = \int_0^{\pi/2} 2/\pi\sin(u) = 2/\pi[-\cos(u)]_0^{\pi/2} = 2/\pi.$$

5.1.4 BH.5.13 (a) og (b) (Laves til øvelser)

a X er længden af den korte del, mens Y er længden af den lange ende. Vi definerer nu $R = X/Y$. Det ses af $Y = 1 - X$. Bemærk R har støtte på $r \in (0, 1)$. Vi får da

$$P(R \le r) = P(X/(1 - X) \le r) = P(X \le 1/(1 + r)).$$

Vi kender dog ikk X, denne kan vi dog udregne. Vi ser at $X = min(U, 1 - U)$. Den korteste af de to pinde som er blevet skåret af.

$$\begin{aligned} P(X \le x) &= 1 - P(X > x) = 1 - P(U > x, U - 1 > x) \\ &= 1 - P(x < U < 1 - x) = 1 - ((1 - x) - x) = 2x. \end{aligned}$$

hvor vi brugte at U er standard uniform. Vi ser at X er konstant på $0 \le x \le 1/2$ og 0 hvis $x < 0$ og 1 hvis $1 < x$. Den er altså uniform fordelt på intervallet $(0, 1/2)$. Vi husker fra sidste uge at så er fordelingsfunktionen propertionel, vi får da

$$F(r) = P(X \le 1/(1 + r)) = 2r/(1 + r).$$

Vi har nu fundet CDF'en. Vi finder nu PDF'en ved differentiering.

$$F'(r) = \frac{2(1 + r) - 2r}{(1 + r)^2} = 2/(1 + r)^2.$$

b Vi finder ER. Vi benytter substitionen $1 + r = s$

$$ER = \int_0^1 2r/(1 + r)^2 = 2\int_1^2 (s - 1)/s^2 = 2\int_1^2 1/s - 1/s^2 = 2[\log(s) + s^{-1}]_1^2 = 2\log(2) - 1$$

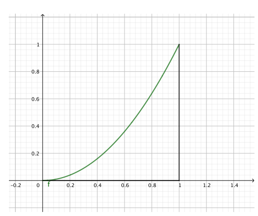
5.2 Øvelser fredag den 21. december kl 10.15-12

5.2.1 Januar 18, opg.1

Vi har fået givet:

$$f(x, y) = \begin{cases} 6y^{1/2} & \text{for } 0 < x < 1, 0 < y < x^2 \\ 0 & \text{ellers} \end{cases}$$

1.1 Vi tegner først området A.



Vi integerer for at vise det er en sandsynlighedstæthed

$$\int_0^1 \int_0^{x^2} 6y^{1/2} dy dx = 6\int_0^1 [2/3y^{3/2}]_0^{x^2} dx = 4\int_0^1 x^3 dx = 1$$

1.2 Vi finder de marginale tætheder. Vi har allerede i 1.1 fundet den marginale for X og vi ser at det er rigtigt. Vi omparameterisere det en smule. og får $0 < y < 1$, $\sqrt{y} < x < 1$.

$$\int_{\sqrt{y}}^1 6y^{1/2} dx = 6y^{1/2}(1 - \sqrt{y}).$$

X og Y er ikke uafhængige da produktet ikke er lig med den simultane tæthed, vi har ej heller at supporten er en produktmængde.

1.3 Vi definerer nu $Z = -\log(X^3)$. Vha. change of variable/transformationssætningen findes tætheden. Bemærk ift. til sætningen $Z = -\log(X^3)$, $(z = -\log(x^3)$ og $e^{-z/3} = x$. Bemærk at X har support i $(0, 1)$, betydende at Z har support in $(0, \infty)$.

$$f_Z(z) = f_X(e^{-z/3})|-1/3e^{-z/3}| = 4e^{-z} \cdot 1/3e^{-z/3} = 4/3e^{-4/3z}.$$

for $z \in (0, \infty)$. 0 ellers. Vi ser yderligere at dette er en eksponentialfordeling med $\lambda = 4/3$.

1.4 Vi bestemmer middelværdi og varians

$$EX = \int_0^1 4x^4 = 4/5$$

Vi bestemmer også varians. Først andet moment

$$EX^2 = \int_0^1 4x^5 = 4/6 = 2/3.$$

Varians bliver da

$$VX = EX^2 - (EX)^2 = 2/3 - (4/5)^2 = 2/3 - 16/25 = 2/75.$$

1.5 Vi ser at fordelingenfunktionen af X er

$$G(x) = \int_{-\infty}^x f_x x = \int_0^x 4x^3 = x^4.$$

Den inverse bliver da $y^{1/4} = x$, for $y \in (0, 1)$. Se resten i google drev: `https://drive.google.com/file/d/1lBLGKmj3LZotF62rV-mPLyO8RHNAYTfl/view?usp=sharing`

5.2.2 Januar 16, opg.2

2.1 Vi bestemmer EX

$$EX = -3 \cdot 0 + 0 \cdot 0 + 2 \cdot 0.1 + 4 \cdot 0.2 + 7 \cdot 0.5 + 10 \cdot 0.15 + 12 \cdot 0.05 = 6.6,$$

så svaret er A.

2.2 Vi bestemmer spredningen, så vi vil gerne bestemme variansen og dermed andet momentet.

$$EX^2 = -3^2 \cdot 0 + 0^2 \cdot 0 + 2^2 \cdot 0.1 + 4^2 \cdot 0.2 + 7^2 \cdot 0.5 + 10^2 \cdot 0.15 + 12^2 \cdot 0.05 = 50.3$$

Så spredningen bliver da

$$\sqrt{VX} = \sqrt{EX^2 - (EX)^2} = \sqrt{50.3 - (6.6)^2} \approx 2,6,$$

svaret er dermed B.

2.3 Vi vil gerne bestemme $P(X = 2, Y = 2)$. De er uafhængige, vi får derfor $P(X = 2, Y = 2) = P(X = 2)P(Y = 2) = 0.1 \cdot 0.05 = 0.005$. Svaret er da A.

2.4 Vi vil gerne finde $P(X = Y)$. Da de er uafhænige får vi

$$\begin{aligned} P(X = Y) &= P(X = -3, Y = -3) + P(X = 0, Y = 0) + \ldots + P(X = 12)P(Y = 12) \\ &= 0 \cdot 0 + 0 \cdot 0 + 0.1 \cdot 0.05 + 0.2 \cdot 0.1 + 0.5 \cdot 0.15 + 0.15 \cdot 0.5 + 0.05 \cdot 0.2 = 0.185, \end{aligned}$$

svaret er dermed C.

2.5 Vi indfører nu Z. Vi ser at $dansk \cup engelsk = 1$, $dansk \cap engelsk = \varnothing$

$$P(Z = 10) = P(dansk) \cdot P(10|dansk) + P(engelsk) \cdot P(10|englsk) = 0.6 \cdot 0.15 + 0.4 \cdot 0.5 = 0,29.$$

2.6 Vi bruger Bayes

$$P(dansk|10) = \frac{P(10|dansk) \cdot P(dansk)}{P(10)} = \frac{0.15 \cdot 0.6}{0,29} = 0,31.$$

5.2.3 SS.24 (Laves til øvelser)

1-2 Se google drev: `https://drive.google.com/file/d/1lBLGKmj3LZotF62rV-mPLyO8RHNAYTfl/view?usp=sharing`

3 Per def. 3.2 (da vi har antaget en kendt varians): Modellen består af udfaldsrummet $\mathbb{R}^{73}$, da vi har 73 observationer, samt familien

$$\mathcal{P} = \{N_\mu^{73} : \mu \in \mathbb{R}\}$$

af tætheder på $\mathbb{R}^{73}$ hvor (per 3.1)

$$N_\mu^{73} = \frac{1}{(2 \cdot 7^2 \cdot \pi)^{73/2}} \exp\left(-\frac{1}{2 \cdot 7^2} \sum_{i=1}^{73} (y_i - \mu)^2\right), \quad y = (y_1, y_2, \ldots y_n) \in \mathbb{R}^n.$$

Alternativt: $Y_1, Y_2, \ldots, Y_{73}$ er uafhængige og identisk normalfordelte stokastiske variable, $Y_i \sim N(\mu, 7^2)$, hvor $\mu \in \mathbb{R}$.

4 Se google drev: `https://drive.google.com/file/d/1lBLGKmj3LZotF62rV-mPLyO8RHNAYTfl/view?usp=sharing`

5 Se google drev: `https://drive.google.com/file/d/1lBLGKmj3LZotF62rV-mPLyO8RHNAYTfl/view?usp=sharing`

Det er vigtigt hvordan man fortolker dette. Vi bruger en frekventistisk tankegang i dette kursus, eller en frekvensfortolkning. I frekvensfortolkningen definerer man sandsynligheden for en hændelse som den frekvens, hvormed hændelsen optræder relativt til en referenceklasse. Referenceklassen kan fx være 'slag med denne terning' og hændelsen kan være '6'ere'. For at finde ud af, hvad sandsynligheden for at sla en 6'er med terningen er, skal jeg blot sla med terningen et antal gange og dividere antallet af 6'ere med det samlede antal slag.

Det medfører at fortolkningen af konfindesintervallet er denne s.38 HS: **Hvis eksperimentet gentages mange gange og intervallet regnes for hvert eksperiment, så vil ca 095% af disse intervallet indeholde den sande middelværdi** μ. Frekvensfortolkningen gør altså vi betragter μ, den sande middelværdi som fikset og dataet som tilfældigt.

5 ekstra - Subjektiv fortolkning (slet ikke relevant ift. SS) Dette er dog ikke den eneste fortolkning af sandsynlighed. Den store anden fortolkningsgren er den subjektive fortolkning, også kaldet Bayes fortolkning. Den grundlæggende ide i den subjektive fortolkning af sandsynlighed er, at sandsynligheder udtrykker en subjektiv grad af overbevisning om, at noget vil ske. Hvis jeg fx hævder, at sandsynligheden for at sla en 6'er med en given terning er 1/6, er det blot et udtryk for min subjektive vurdering af, hvorvidt det vil være muligt eller umuligt at sla en 6'er i næste slag.

Ift. konfidensintervaller får vi en lidt anden fortolkning. Denne bliver: Der er 95% sandsynlighed for at den sande middelværdi μ er inde i vores interval.

5.2.4 SS.25 (Laves til øvelser)

1 Brug 8.1.1, transformationssætningen til at vise det samme som i obligatorisk opgave. Bemærk at $Y = e^X$, $y = e^x$, $x = \log(y)$

$$\begin{aligned} f_Y(y) = f_x((\log(y))) \cdot |\log(y)'| &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(\log(y)-\mu)^2}{2\sigma^2}\right) \cdot \left|\frac{1}{y}\right| \\ &= \frac{1}{y\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(\log(y)-\mu)^2}{2\sigma^2}\right), \end{aligned}$$

for $y > 0$. Dette giver altså det samme, hvor heldigt.

2-3 Se google drev: `https://drive.google.com/file/d/1lBLGKmj3LZotF62rV-mPLyO8RHNAYTfl/view?usp=sharing`

4 Hver variabel kan have 2 værdier 1 og 0, alt efter om den givne person har regnet sudukoen på under eller over 3 min. Det er en sum af en masse bernouilli viriable, dvs. det er en binomialfordeling. Denner er beskrevet på en tidligere ugeseddel, men også i BH eller s.12 HS. Per sætning 2 fåes, meget naturligt, at MLE er x/n, hvor n er antallet af observationer og x er er antal succeser, vi får da estimatet for p til per sætning 1.2

$$\hat{p} = 81/104 \approx 0.779,$$

den estimerede fordeling $104\hat{p}$ er $(104, 81/104)$. og den har spredning

$$s(\hat{p}) = \sqrt{\hat{p}(1-\hat{p})/104} \approx 0,04.$$

6 Uge 6

6.1 Øvelser fredag den 4. januar kl 10.15-12

6.1.1 DS 1.3 (hører til kursusuge 4) (laves til øvelser)

1 Hvis man kun er interesseret i hvorvidt en aktie er faldet. Du kan fortolke p til at være sandsynligheden for at en given er faldet i værdi. Vi bliver også nødt til at antage uafhængighed.

2 Vi bruger def. 1.1. til at opstille en statistisk model. Vi bruger den alternative formulering. Dette bliver: X er en stokastisk variabel med udfaldsrum $\{0, 1, \ldots, 10\}$ og vi antager at $X \sim bin(10, p)$, hvor $p \in (0, 1)$.

Vi udregner nu MLE for p.

$$\hat{p} = \frac{8}{10}.$$

Denne estimator har fordelingen sætning 1.3

$$n\hat{p} \sim bin(10, p),$$

det vil altså sige det samme som vores model. Spredningen bliver videre nederst sætning 1.3.

$$s(\hat{p}) = \sqrt{\frac{\hat{p}(1-\hat{p})}{10}} \approx 0,13.$$

3 Hvis $p = 1/2$ er der lige stor chance for om aktie stiger eller falder. Det er derfor interessant at vide hvorvidt p er mindre end 0,5, da man så ikke vil forvente et tab.

4 Se google drev: `https://drive.google.com/open?id=1n4dtMrbw9ObRbJjk_FvboPQB7l22WciV`

5 Ja, det ser ud til der generelt har været en negativ udvikling. Da vores tidligere udregning har vist der kun er ca. 1% sandsynlighed for at få et sådan resultat eller værre. Såfremt p=1/2 er den sande værdi

6.1.2 SS.29 (hører til kursusuge 4) (Laves til øvelser)

1 Vi vil gerne bruge modellen for en normalfordeling med kendt varians. Vi antager kropstemraturen er uafhængig og identitisk fordelt fra person til person. Vi får da per def. 3.2 at $Y_1, Y_2, \ldots Y_{130}$ er uafhængige identisk fordelte stokastiske variable over kropstempratur hvor $Y_i \sim \mathcal{N}(\mu, (0,4)^2)$, hvor $\mu \in \mathbb{R}$ og er ukendt.

2-4 Se google drev: `https://drive.google.com/open?id=1n4dtMrbw90bRbJjk_FvboPQB7l22WciV`

6.1.3 DS 4.1, spørgsmål 1-3 (hører til dagens forelæsninger) (Laves til øvelser)

1 Vi bruger definition 4.1. Bemærk nu er variansen også ukendt. Vi benytter den alternative formulering. Lad $Y_1, Y_2, \ldots Y_{15}$ være uafhænige identisk normalfordelte variable hvor $Y_i \sim \mathcal{N}(\mu, \sigma^2)$ hvor $\mu \in \mathbb{R}$ og $\sigma^2 > 0$ er ukendte parametre.

2 Bemærkning 4.5 giver at $\hat{mu} = 197$, $\tilde{\sigma} = s^2 = (23.11)^2 = 534.07$. Videre får vi fra sætning 4.3 at de marginale fordelinger er

$$\hat{Y} \sim \mathcal{N}(\mu, \sigma^2/15)$$

for mu og

$$\hat{\sigma}^2 \sim \sigma^2/n \cdot \chi^2_{n-1}.$$

spredningen bliver s.60 øverst

$$SE(\hat{\mu}) = 23.11/\sqrt{15} \approx 5.97.$$

3 Vi beregner 95% og 90% konfidensintervaller. først 95

$$197 \pm 2.145 \cdot 5.97,$$

hvilket giver os et interval på $(184.194; 209.806)$ Ligeså med 90,

$$197 \pm 1.761 \cdot 5.97,$$

hvilket giver os et interval på $(186.487; 207.513)$.

1-3 i R Se google drev: `https://drive.google.com/open?id=1n4dtMrbw90bRbJjk_FvboPQB7l22WciV`

6.1.4 DS 1.4 (hører til kursusuge 4) (laves hjemme)

1 Vi er i afsnittet om binomialfordelinger, så det er da også denne vi gerne vil bruge. Vi finder første sensitiviteten. Sætning 1.2 giver os en MLE på

$$\hat{p}_1 = \frac{57}{62} \approx 0,92.$$

MLE for specificiteten er da

$$\hat{p}_2 = \frac{127}{131} \approx 0,97.$$

2 Vi beregner de estimerede spredninger for estimatorene.

$$s(\hat{p}_1) = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{62}} = 0,035,$$

og tilsvarende:

$$s(\hat{p}_2) = \sqrt{\frac{\hat{p}_2(1-\hat{p}_2)}{131}} = 0,015.$$

3 Vi har den samme MLE, $\hat{p}_1$. Vi får ligningen

$$\begin{aligned}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n}} &= 0,02\\ \hat{p}_1(1-\hat{p}_1) &= n(0,02)^2\\ n &= \frac{\hat{p}_1(1-\hat{p}_1)}{(0,02)^2} \approx 185,4,\end{aligned}$$

vi skal altså bruge mindst 186 kødprøver for at få en spredning mindre end eller lig med $0,02$.

6.1.5 SS.28 (hører til kursusuge 4) (laves hjemme)

1-2 Se google drev: `https://drive.google.com/open?id=1n4dtMrbw9ObRbJjk_FvboPQB7l22WciV`

7 Uge 7

7.1 Øvelser tirsdag den 8. januar kl 15.15-17

7.1.1 SS.30 (Laves til øvelser)

1-6 Se google drev: `https://drive.google.com/open?id=1JwMU7dpw4cOpe66Wz_nUccGu5A7s4Wt9`

7.1.2 Januar 2016, opgave 3

1-5 Se google drev: `https://drive.google.com/open?id=1JwMU7dpw4cOpe66Wz_nUccGu5A7s4Wt9`

6 Vi skal have et interval på 0,5. Vi antager at vi kan bruge $t_{436;0,975}$ fraktilen, selvom den naturligvis vil ændre sig en smule når n også ændrede sig. Vi skal dermed have at

$$\begin{aligned}0,5 &\approx 2 \cdot t_{436;0.975} \cdot \frac{s}{\sqrt{n}}\\ n &\approx \left(\frac{2 \cdot 1.965 \cdot 3.07}{0.5}\right)^2 \approx 582.\end{aligned}$$

Svaret er derfor A, cirka 580.

7 Det letteste er at lave vores model om til $\mathcal{N}(0,1)$ fordeling. Vi trækker derfor $\hat{\mu}$ fra og dividerer med s. Vi får da

$$\begin{aligned}P(|Y| < 4) = P(-4 < Y < 4) &= P\left(\frac{-4-\hat{\mu}}{s} < Z < \frac{4-\hat{\mu}}{s}\right)\\ &= P(1.276227 < Z < 1.333101).\end{aligned}$$

I R kan vi bruge pnorm direkte da Z er standard normalfordelt. Vi får da

$$pnorm(1.333101) - pnorm(-1.276227) \approx 0,808,$$

så svaret er A.

7.1.3 SS.31

1-7 Se google drev: `https://drive.google.com/open?id=1JwMU7dpw4cOpe66Wz_nUccGu5A7s4Wt9`

7.1.4 April 2017, opgave 2

1 2/15+2/15+1/15=5/15=1/3.

2 3/25+1/15=4/15

3 Vi skal kun kigge i sidste kollonne, vi får da 1/15/(1/15+3/15)=1/4.

4 Lad A være af et barn har en BMI på over 25. Lad B være hændelsen at et tilfældigt barn kun bor hos den ene af sine forældre.

$$P(A)P(B) = 1/3 \cdot 4/15 = 4/45 \neq 1/15 = P(A \cap B)$$

De er derfor ikke uafhængige.

5 Lad nu yderligere C være hændelsen et barn bor hos begge sin forældre. Vi har da

$$P(A)P(C) = 1/3 \cdot 6/15 = 2/15 = P(A \cap C),$$

de er dermed begge uafhængige.

6 Dette følger opgaven som vi havde en elge. Vi har en population af børn. Vi udtager så en stikprøve af den og kigger på egenskab. Det følger derfor en hypergeometrisk fordeling, $Hgeom(6, 12, 5)$. Vi får $P(X = 2)$ til

$$P(X = 2) = \frac{\binom{6}{2}\binom{12}{3}}{\binom{18}{5}} = 0.3852$$

7.1.5 Ekstra, April '15 1.4-1.5 (skitse)

1.4 følger af 2D LOTUS. 1.5: Tegn først området. A er en enhedsterning. $(xy) \leq w$, så $y \leq w/x$, så vi tager ligesom en bid af firkanten, alt efter hvor stor w er. (Find evt. tegning i eksamensbesvarelsen). Per def. af fordelingensfunktionen vil vi gerne finde

$$P(W \leq w) = P(XY \leq w).$$

Hvad betyder dette? Vi har at $XY \leq w \iff (X, Y) \in B$, hvor B er defineret som $B = \{(x, y) \in \mathbb{R}^2 | xy \leq w\} \cap A$. Vi får da

$$P(XY \leq w) = P((X, Y) \in B) = \int_B 6/5(x + y^2) d(x, y) = \int_0^w \int_0^1 6/5(x + y^2) dydx + \int_w^1 \int_0^{w/x} 6/5(x + y^2) dydx$$

Dette er super kedeligt at integrerer, men det bliver $9/5 - 3/5 \cdot w^2 - w^3/5$ for $w \in (0, 1)$. Vi får samlet:

$$F_W = P(W \leq w) = \begin{cases} 0 & \text{for } w \leq 0 \\ \frac{9}{5}w - \frac{3}{5}w^2 - \frac{1}{5}w^3 & \text{for } w \in (0,1) \\ 1 & \text{for } w \geq 1 \end{cases}.$$

7.2 Øvelser fredag den 11. januar kl 10.15-12

7.2.1 DS 6.2 (laves til øvelser)

7.2.2 SS.32

1,2, 4-7 Se google drev: `https://drive.google.com/open?id=1hbcwjXoF8Lb1hWaK1HMlv5FUKU1hrz0B`

3 Def. 5.1: Vores statistiske model: $X_1, X_2, \dots X_{73}$ og $Y_1, Y_2, \dots Y_{111}$ er uafhængige normalfordelte stokastiske variable, hvor $X_i \sim \mathcal{N}(\mu_1, \sigma^2)$ og $Y_j \sim \mathcal{N}(\mu_2, \sigma^2)$ og hvor $\mu_1, \mu_2 \in \mathbb{R}$ og $\sigma^2 > 0$ er ukendte parametre.

7.2.3 August 2015, opgave 3

1-7 Se google drev: `https://drive.google.com/open?id=1hbcwjXoF8Lb1hWaK1HMlv5FUKU1hrz0B`

7.2.4 April 2016, opgave 3

1-6 Se google drev: `https://drive.google.com/open?id=1hbcwjXoF8Lb1hWaK1HMlv5FUKU1hrz0B`

8 Uge 8

8.1 Øvelser tirsdag den 15. januar kl 15.15-17

8.1.1 August 2015, opgave 4 (Laves til øvelser)

1-8 Se google drev: `https://drive.google.com/open?id=1BLzn6mNsmQvbBYIKQOen4ibYE4gtBIjg`

8.1.2 SS.33 (Laves til øvelser)

1-4 Se google drev: `https://drive.google.com/open?id=1BLzn6mNsmQvbBYIKQOen4ibYE4gtBIjg`

8.1.3 April 2016, opgave 4

1-9 Se google drev: `https://drive.google.com/open?id=1BLzn6mNsmQvbBYIKQOen4ibYE4gtBIjg`

10 : Vi har at

$$\hat{\beta} = \frac{SPD_{xy}}{SSD_x}.$$

Vi kender $\hat{\beta}$ og vi kan beregne SSD. SSD udregens i R. Vi får

$$\hat{\beta} \cdot SSD_x \approx -11 \cdot 208 \approx -2287,$$

svaret er B.

8.1.4 SS.34

1 Vi må forvente at ægtepar bor sammen og spiser sammen. Det er derfor nok ikke realistisk at tro kalorieindtag er uafhængigt mellem mand og dame i et ægtepar, dog er der nok uafhængighed mellem ægtepar. De er dermed parrede.

2 Same, uafhængihed mellem potter, men nok ikke i potter.

3 Ikke parret.

8.1.5 April 2016, opgave 1

1 Det bliver en halvcirkel. Og vi har at $x > 0$, således $p(x, y) > 0$. Vi skal dermed kun intergrere til 1 for at vise det er en sandsynlighedstæthed.

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(x,y)dydx = \int_0^1 \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} 3/2x = \int_0^1 3x\sqrt{1-x^2}dx.$$

Vi kan bruge substituionen $u = 1 - x^2$. og får $-dx/(2x) = dx$

$$\int_1^0 -3/2\sqrt{u}du = \int_0^1 3/2\sqrt{u}du = 1$$

2 Y har support på (-1,1). Den marginale tæthed for Y er derfor 0 for udenfor (-1,1). På supporten har vi den marginale tæthed er givet ved:

$$h(y) = \int_{\mathbb{R}} p(x,y)dx = \int_0^{\sqrt{1-y^2}} \frac{3}{2}xdx = \left[\frac{3}{4}x^2\right]_0^{\sqrt{1-y^2}} = \frac{3}{4}\left(1-y^2\right),$$

som vi gerne ville vise. Support for X er (0,1). Den marginale tæthed for X er dermed 0 for X udenfor (0,1)i I (0,1) har vi

$$g(x) = \int_{\mathbb{R}} p(x,y)dy = \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} \frac{3}{2}xdy = \frac{3}{2}x2\sqrt{1-x^2} = 3x\sqrt{1-x^2}$$

De er ikke uafh. Det er ikke en produktmængde f.eks.

3 Vi udregner middelværdi og varians.

$$E(Y) = \int_{\mathbb{R}} yh(y)dy = \int_{-1}^1 \frac{3}{4}y\left(1-y^2\right)dy = \left[\frac{3}{8}y^2 - \frac{3}{16}y^4\right]_{-1}^1 = 0.$$

Andet momentet:

$$\begin{aligned} E\left(Y^2\right) &= \int_{\mathbb{R}} y^2h(y)dy = \int_{-1}^1 \frac{3}{4}y^2\left(1-y^2\right)dy = \left[\frac{3}{12}y^3 - \frac{3}{20}y^5\right]_{-1}^1. \\ &= 2\frac{3}{12} - 2\frac{3}{20} = \frac{1}{5} \end{aligned}$$

Variansen er dermed

$$V(Y) = E\left(Y^2\right) - (E(Y))^2 = \frac{1}{5}.$$

4 Vi bruger transformationssætningen BH 8.1.1. Vi Ser at log er strent voksende funktion, -log er dermed strengt aftagende. Videre Ser vi at billedmængden og dermed supporten for Z er $(0, \infty)$. Vi har at $z = -log(y + 1/2)$, så $t^{-1}(z) = y = 2\exp -z - 1$ og videre $(t^{-1}(z))^{-1} = -2\exp -z$. Tætheden for $z > 0$ er dermed

$$f_Z(z) = 3/4(1 - (2\exp -z - 1)^2) \cdot 2\exp -z = 3/4(4e^{-z} - 4e^{-2z})2e^{-z} = 6(e^{-2z} - e^{-3z}).$$

5 Minder meget om afleveringsopgaven i sidste uge. Mængden bliver en pizzaskive der starten fra toppen af halvciklen, Vi finder $P(Y > X)$

$$P(Y > X) = \int_B p(x,y)d(x,y) = \int_0^{\sqrt{2}} \int_x^{\sqrt{1-x^2}} 3/4xdydx = \ldots = 1/2(1 - \sqrt{1/2}).$$