

Oppgavesett nr. 1

“MAT110 Statistikk I”, 2020

Oppgave 1: (økonomi - uniformt utfallsrom , urne , kombinatorikk , $s = 1$)

En student som bor i studentboligene på Molde campus har fått akutt dårlig råd. Studenten har derfor problemer med å betale husleien og risikerer å bli kastet ut fra hybelen. I et desperat forsøk på skaffe penger så arrangerer studenten et lotteri. Dette lotteriet er slik at 7 av totalt 1000 lodd gir en pengepremie.

Lotteriet beskrevet i oppgaven representerer et uniformt utfallsrom. Da gjelder urnemodellen.

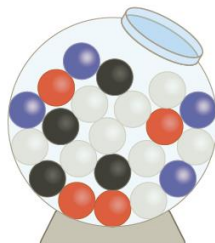
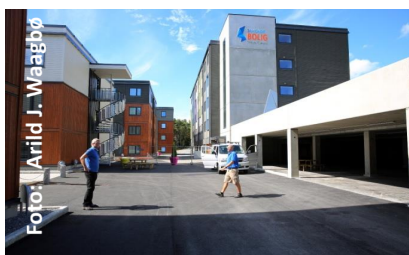
En person kjøper lodd hos studenten. Personen trekker lodd fra en urne.

Anta at personen kjøper kun *ett* lodd, dvs. $s = 1$. I dette tilfellet, når man kun trekker ett lodd, vet vi fra vedlegg B (vedlegg B finner dere bakerst i kompendiet, del 1) at de 4 situasjonene vi ser på i dette kurset i forbindelse med kombinatorikk, er sammenfallende:

$$\begin{array}{c} s = 1 \\ \Downarrow \\ \text{situasjon 1} = \text{situasjon 2} = \text{situasjon 3} = \text{situasjon 4} \end{array}$$

Dette betyr at formlene for antall kombinasjoner i de 4 situasjonene, er sammenfallende.

I denne oppgaven skal vi teste om dette virkelig stemmer for vårt konkrete tilfelle angående lodd.



Figur 1: Studentboliger og urne.

Eksperiment: trekke en kule (ett lodd) fra en urne.

Utfallsrom for sannsynlighetsmodellen:

$$\Omega = \left\{ u_{\text{tap}}, u_{\text{vinn}} \right\} \quad (1)$$

med utfallene

$$u_{\text{tap}} = \text{utfall som gir taperlodd} \quad (2)$$

$$u_{\text{vinn}} = \text{vinnerlodd} \quad (3)$$

og

$$p_{\text{tap}} = \text{sannsynligheten for at et tilfeldig valgt lodd er et taperlodd} \quad (4)$$

$$p_{\text{vinn}} = \text{vinnerlodd} \quad (5)$$

- a) Hva er sannsynligheten for at loddkjøperen trekker et vinnerlodd p_{vinn} ? ¹

Løs oppgaven ved å bruke formelen for antall kombinasjoner beskrevet av [situasjon 1](#) fra vedlegg *B* (side 392).

- b) Hva er sannsynligheten for at loddkjøperen trekker et vinnerlodd p_{vinn} ?

Løs oppgaven ved å bruke formelen for antall kombinasjoner beskrevet av [situasjon 2](#) fra vedlegg *B* (side 392).

- c) Hva er sannsynligheten for at loddkjøperen trekker et vinnerlodd p_{vinn} ?

Løs oppgaven ved å bruke formelen for antall kombinasjoner beskrevet av [situasjon 3](#) fra vedlegg *B* (side 392).

- d) Hva er konklusjonen?

Er situasjonene 1, 2 og 3 sammenfallende når vi kun har et trekk $s = 1$?



¹Tips: Fra utfallsrommet i lign.(1) ser vi at det er 2 typer kategorier lodd her: vinnerlodd og taperlodd.

Oppgave 2: (revisjon - uniformt utfallsrom , urne , kombinatoriske sannsynligheter)

XPRO er et firma som tilbyr flerfaglig prosjektadministrative tjenester og uavhengig kontroll, blant annet innen bygg og anlegg.

Du er ansatt i revisjonsfirmaet PwC og skal gjøre revisjon av dette nyoppstartede firmaet. Du ønsker å sjekke om bilagene i regnskapene til XPRO inneholder feil eller ikke. La oss dele regnskapsåret et i to, **første halvår** og **andre halvår**.

Du vurderer to alternative strategier for et gitt regnskapsår: ²

Strategi A: "få" stikkprøver i første halvår og "mange" stikkprøver i andre halvår:

- 2000 stikkprøver i første halvår gav 12 bilag med feil
- 8000 stikkprøver i andre halvår gav 24 bilag med feil

Strategi B: "mange" stikkprøver i første halvår og "få" stikkprøver i andre halvår:

- 4000 stikkprøver i første halvår gav 20 bilag med feil
- 1000 stikkprøver i andre halvår gav 2 bilag med feil

altså stragtegiene er "omvendt" sammenlignet med hverandre.



Figur 2: XPRO.

²Anta at tallene er representative i den forstand at de viser sannsynligheten for at selskapene finner feil i de to aktuelle årene.

Stikkprøvene representerer et uniformt utfallsrom. Da gjelder urnemodellen.
 Utfallsrom for sannsynlighetsmodell for strategi A i første halvår:

$$\Omega_{A1} = \left\{ u_{A1} , v_{A1} \right\} \quad (6)$$

altså to mulige utfall

$$u_{A1} = \text{utfall hvor bilag fra stikkprøvene i første halvår har feil} \quad (7)$$

$$v_{A1} = \text{rett} \quad (8)$$

og

$$p_{A1} = \text{sannsynligheten for at et bilag fra første halvår har feil v/bruk av strategi } A \quad (9)$$

$$q_{A1} = \text{rett} \quad (10)$$

Tilsvarende for strategi B .

Utfallsrom for sannsynlighetsmodell for strategi A for når man ser hele året under ett:

$$\Omega_A = \left\{ u_A , v_A \right\} \quad (11)$$

altså to mulige utfall

$$u_A = \text{utfall hvor bilag har feil når man ser hele året under ett} \quad (12)$$

$$v_A = \text{rett} \quad (13)$$

og

$$p_A = \text{sannsynlighet for at et bilag har feil ved bruk av strategi } A \quad (14)$$

$$q_A = \text{rett} \quad (15)$$

- a) Dersom du velger strategi A , hva er sannsynligheten for at et tilfeldig valgt bilag fra første halvår har feil, dvs. hva er p_{A1} ?³

Og for andre halvår, dvs. p_{A2} ?

- b) Dersom du velger strategi B , hva er sannsynligheten for at et tilfeldig valgt bilag fra første halvår har feil, dvs. hva er p_{B1} ?

Og for andre halvår, dvs. p_{B2} ?

Den strategien som gir **størst sannsynlighet** over en gitt periode anses som best i den aktuelle perioden.

- c) i) Hvilken er strategi er best for første halvår?
ii) Hvilken er strategi er best for andre halvår?

Istedet for å se på et halvår om gangen, som i oppgavene foran, la oss nå heller se på **hele året under ett**.

- d) Hvilken strategi er best når man ser hele året under ett?⁴

³Oppgaver dreier seg om ett trekk, $s = 1$. Da er situasjon 1, 2, 3 og 4 de samme. Det så vi i oppgave 1.

⁴Regn ut sannsynligheten p_A for strategi A når du ser hele året under ett. Gjør det samme for strategi B , dvs. regn ut p_B .

Dersom du har gjort oppgavene foran riktig så ser du at:

- strategi A ”vinner” både første halvår og andre halvår, ”taper” når man ser hele året under ett
- oppgave **2a** og **2b** viser at det er $\underbrace{\text{flest feil}}_{\text{første halvår}}$ i bilagene fra første halvår

e) Hvorfor må revisorselskapet bytte til strategi B når man ser hele året under ett? ⁵



PS:

Dette fenomenet er velkjent i statistikk og kalles **Yule-Simpsons paradoks**.

⁵Er det larest å ta flest stikkprøver i den perioden hvor det er flest feil? Er strategi B i så måte bedre tilpasset enn strategi A ?

Oppgave 3: (økonomi - betinget sannsynlighet , sannsynlighetsregning)

KPMG sine avdelingskontor i Møre & Romsdal har sammen utviklet et dataverktøy for å forutsi hvilke bedrifter som vil gå konkurs i løpet av året. Dataverktøyet har blitt testet på gamle data. Resultatet fra KPMG-modellen er:

- Av bedriftene som faktisk gikk konkurs ble 80% klassifisert som ikke levedyktig
- Av bedriftene som ikke gikk konkurs ble 95% klassifisert som levedyktige

Eksperiment:

KPMG velger en bedrift og predikerer hvorvidt bedriften er levedyktig eller ikke, og observerer om den faktisk går konkurs eller ikke.



Figur 3: KPMG i Molde.

Notasjon: (observasjoner)

$$l = \text{bedrift levedyktig} \quad (16)$$

$$\bar{l} = \text{bedrift ikke levedyktig} \quad (17)$$

$$\textcolor{red}{k} = \text{bedrift gikk konkurs} \quad (18)$$

$$\overline{\textcolor{green}{k}} = \text{bedrift gikk ikke konkurs} \quad (19)$$

Utfallsrom:

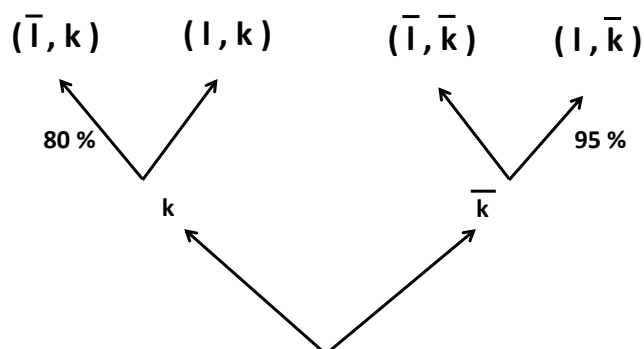
$$\Omega = \left\{ (l, k) , (\bar{l}, k) , (l, \bar{k}) , (\bar{l}, \bar{k}) \right\} \quad (20)$$

Definer begivenhetene: (observasjoner)

$$K = \left\{ (l, k) , (\bar{l}, k) \right\} \quad \text{bedrift gikk konkurs} \quad (21)$$

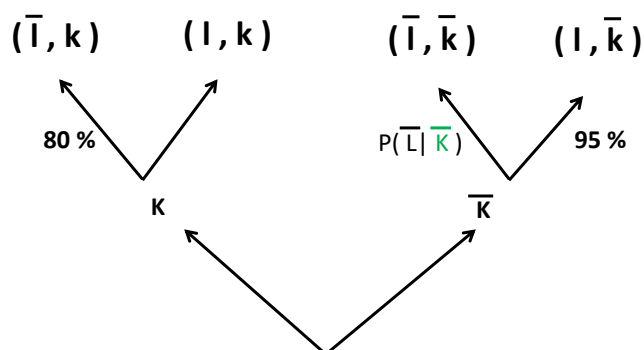
$$L = \left\{ (l, k) , (l, \bar{k}) \right\} \quad \text{bedrift klassifisert som levedyktig} \quad (22)$$

- a) Skriv opp informasjonen oppgitt i kulepunktene ovenfor på *matematisk form*.



Figur 4: Sannsynlighetstre.

- b) Hva er sannsynligheten for at bedriften blir klassifisert som $\overbrace{\text{ikke levedyktig}}^{\bar{L}}$ gitt at den faktisk ikke går konkurs, $P(\bar{L}|\bar{K})$? ⁶

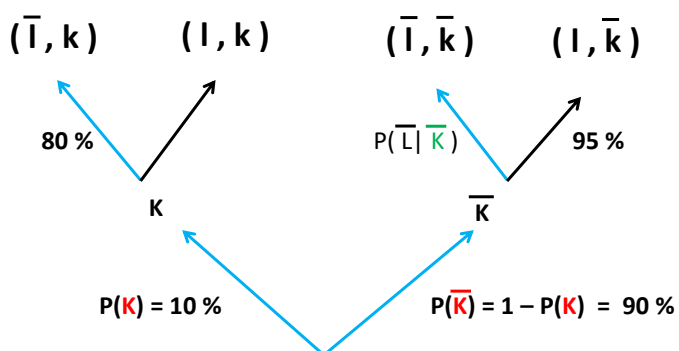


Figur 5: Sannsynlighetstre.

⁶Tips: Se figur 5 og bruk [komplementsetningen](#).

- c) Anta at 10% av bedriftene i Møre & Romsdal går konkurs i løpet av et gitt år, dvs. $P(K) = 0.10$.

Hvor stor sannsynlighet er det for at KPMG-modellen klassifiserer en bedrift som ikke levedyktig, dvs. $P(\bar{L})$? ⁷



Figur 6: Sannsynlighetstre.

⁷Tips: Se figur 6 og bruk formelen for [oppsplitting](#) av Ω .

d) Ifølge årboken til Statistisk sentralbyrå var det 1490 bedrifter i Møre & Romsdal i 2009.

i) Hvor mange bedrifter forventes å klassifiseres som ikke levedyktig ifølge modellen til KPMG? ⁸

ii) Hvor mange bedrifter forventes å gå konkurs i løpet av et gitt år?

e) En bedrift blir klassifisert som ikke levedyktig.
Hva er sannsynligheten, ifølge KPMG-modellen, for at denne bedriften faktisk går konkurs i løpet av året? ⁹

■

⁸Tips: En forventning kan regnes ut via: forventning = vekt · verdi, dvs.:

$$\text{antall ikke levedyktige bedrifter i 2009 ifølge KPMG} = P(\bar{L}) \cdot (\text{antall bedrifter i M\&R}) \quad (23)$$

⁹Dvs. finn $P(K|\bar{L})$. Tips: Bruk Bayes' lov.

Oppgave 4: (logistikk - utfallsom , begivenheter , disjunkt , urne , sannsynlighetsregning)

Glamox er en norsk produsent og leverandør av belysning.

Glamox har til enhver tid mange tilbud ute hos kunder som de venter tilbakemelding på. Hensikten med denne oppgaven er å finne relevante sannsynligheter for at tilbud aksepteres basert på to kriterier: **selger** som har laget tilbudet og **verdien** på tilbudet.

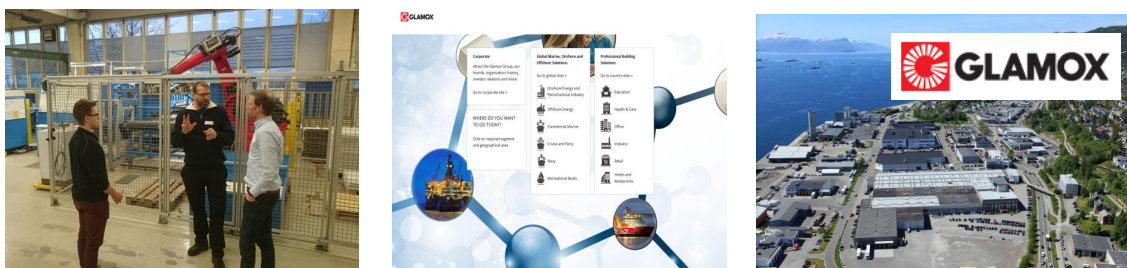
1) **Selgere:**

La oss kun se på 2 selgere hos Glamox, nemlig selgerne 1 og 2.

2) **Verdi:**

La oss for enkelhets skyld kun skille mellom store og små tilbud, dvs. tilbud som har høy (h) verdi eller lav ($\bar{h}$) verdi i kroner og øre. Det kan f.eks. være tilbud som er over eller under 10 000 NOK.

Et tilbud blir enten akseptert (a) eller ikke akseptert ($\bar{a}$). Utfallsrommet har da 8 utfall, se lign.(24) på neste side.



Figur 7: Glamox.

Glamox bruker 350 historiske data for å estimere relevante sannsynligheter.

Resultatene av de historiske dataene er at utfall $(1, h, a)$ inntraff 5 ganger, utfall $(1, \bar{h}, a)$ inntraff 40 ganger osv.

Sannsynlighetene $p_1, p_2, \dots, p_8$ for de 8 mulige utfallene i utfallsrommet er dermed:

$$\Omega = \left\{ \overbrace{(\textcolor{red}{1}, h, a)}^{p_1 = \frac{5}{350}}, \overbrace{(\textcolor{red}{1}, \bar{h}, a)}^{p_2 = \frac{40}{350}}, \overbrace{(\textcolor{red}{1}, h, \bar{a})}^{p_3 = \frac{25}{350}}, \overbrace{(\textcolor{red}{1}, \bar{h}, \bar{a})}^{p_4 = \frac{80}{350}}, \right. \\ \left. \overbrace{(\textcolor{blue}{2}, h, a)}^{p_5 = \frac{8}{350}}, \overbrace{(\textcolor{blue}{2}, \bar{h}, a)}^{p_6 = \frac{68}{350}}, \overbrace{(\textcolor{blue}{2}, h, \bar{a})}^{p_7 = \frac{22}{350}}, \overbrace{(\textcolor{blue}{2}, \bar{h}, \bar{a})}^{p_8 = \frac{102}{350}} \right\} \quad (24)$$

hvor utfallet $(\textcolor{red}{1}, h, a)$ betyr at selger $\textcolor{red}{1}$ lager et tilbud med høy (h) verdi som blir akseptert (a) . Tilsvarende for de øvrige utfallene.

Videre er $N_{\textcolor{red}{1}} = 150$ tilbud gitt av selger $\textcolor{red}{1}$ og $N_{\textcolor{blue}{2}} = 200$ gitt av selger $\textcolor{blue}{2}$.

Totalt er da: $N = N_1 + N_2 = 350$.

$\textcolor{blue}{a})$ Vis at sannsynlighetfordelingen $p_1, p_2, \dots, p_8$ i lign.(24) er en gyldig fordeling. ¹⁰

¹⁰Sannsynlighetsfordelingen er gyldig dersom $P(\Omega) = 1$, dvs. $\sum_{i=1}^8 p_i = 1$.

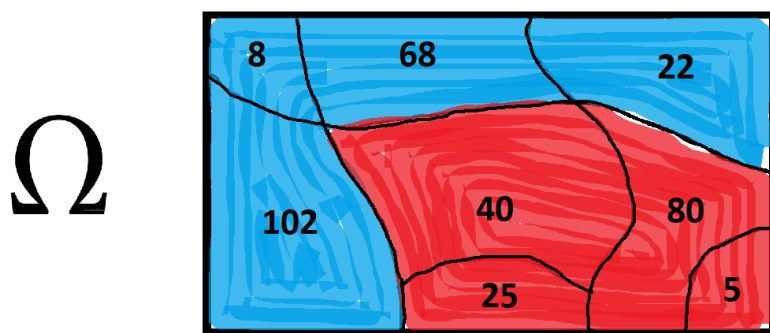
La oss definere følgende begivenhene:

$$S_1 = \left\{ (1, h, a), (1, \bar{h}, a), (1, h, \bar{a}), (1, \bar{h}, \bar{a}) \right\} \quad (25)$$

$$S_2 = \left\{ (2, h, a), (2, \bar{h}, a), (2, h, \bar{a}), (2, \bar{h}, \bar{a}) \right\} \quad (26)$$

- b) Figur 8 viser Venn-diagrammet for de 8 utfallene i utfallsrommet Ω . Tallene i figuren viser antall ganger de respektive utfallene inntreffer, se lign.(24). Den røde og den blå delen av utfallsrommet definerer begivenhetene S_1 og S_2 , henholdsvis.

Er begivenhetene S_1 og S_2 disjunkte? Gi en *kort* begrunnelse. ¹¹



Figur 8: Utfallsrom Ω .

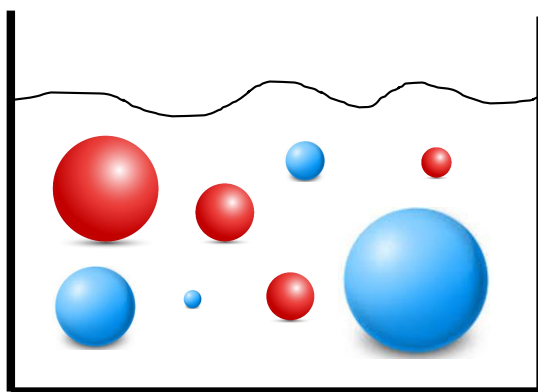
¹¹En setning er nok.

- c) La oss si at hvert av de 8 utfallene i utfallsrommet Ω er assosiert med ei kule. Altså totalt 8 kuler, se figur 9.

La oss videre anta at vi har to typer kuler, **røde** kuler og **blå** kuler. De røde kulene er assosiert med utfallene til selger **1**, og de blå kulene er assosiert med utfallene til selger **2**.

La oss anta at tilbudene laget av selger 1 er uavhengige av tilbudene laget av selger 2.

På tross av at tilbudene som de to selgerne har laget er uavhengige av hverandre, hvorfor kan vi ikke bruke urnemodellen til å f.eks. finne sannsynligheten $P(S_2)$ for at et tilbud er gitt av selger **2**?¹²



Figur 9: Urne.

¹²Svar kort på denne oppgaven. Én setning er nok. Ingen utregninger behøves. Hva må være oppfylt, i tillegg til uavhengighet, for at urnemodellen skal gjelde? Er det oppfylt i vårt tilfelle?

Man kan altså ikke bruke urnemodellen, men man kan likevel finne sannsynligheten $P(S_2)$ på følgende måte, via lign.(24):

$$\begin{aligned} \underline{P(S_2)} &= p_5 + p_6 + p_7 + p_8 \\ &= \frac{8}{350} + \frac{68}{350} + \frac{22}{350} + \frac{102}{350} = \frac{200}{350} = \underline{0.5714} \end{aligned} \tag{27}$$

Sannsynligheter for andre begivenheter finner man på tilsvarende måte.

La oss også definere følgende begivenheter:

$$A = \left\{ (\textcolor{red}{1}, h, a) , (\textcolor{red}{1}, \bar{h}, a) , (\textcolor{blue}{2}, h, a) , (\textcolor{blue}{2}, \bar{h}, a) \right\} \tag{28}$$

og

$$H = \left\{ (\textcolor{red}{1}, h, a) , (\textcolor{red}{1}, h, \bar{a}) , (\textcolor{blue}{2}, h, a) , (\textcolor{blue}{2}, h, \bar{a}) \right\} \tag{29}$$

d) Hva er sannsynligheten for at et tilbud blir akseptert, dvs. hva er $P(A)$?

La oss nå se på et tilbud med høy (h) verdi laget av selger 2.
 Den betingede sannsynligheten for at et slikt tilbud blir akseptert er gitt ved den generelle multiplikasjonssetningen:

$$P(A | \overbrace{S_2 \cap H}^{\text{vet}}) = \frac{P(S_2 \cap H \cap A)}{P(S_2 \cap H)} \quad (30)$$

e) Bruk lign.(30) og vis at $P(A | S_2 \cap H)$ er: ¹³

$$P(A | S_2 \cap H) = \frac{8}{30} \quad (= 0.2667) \quad (31)$$

f) Gi en tolkning av $P(A | S_2 \cap H)$,
 dvs. forklar *kort* hva det betyr på ”godt norsk”.

■

¹³Tips: Finn først $P(S_2 \cap H \cap A)$ og $P(S_2 \cap H)$ ved å bruke tilsvarende metode som i lign.(27).

Oppgave 5: (prognostisering - utfallsrom, begivenheter , sannsynlighetsregning)

Anta at Glamox gir 20 nye tilbud den første uken i mars 2018.

Glamox ønsker nå å bruke sannsynlighetsmodellen fra forrige oppgave til å regne ut relevante sannsynligheter for disse 20 tilbudene.

Anta at aksept av tilbudene er uavhengige av hverandre.

La oss definere begivenheten A_1 ved at kun ett av de 20 tilbudene blir akseptert. Det betyr altså at det er ett tilbud som blir akseptert (a_i) og at de resterende 19 tilbudene ikke blir akseptert ($\bar{a}_i$): ($i = 1, 2, 4, \dots, 20$)

$$A_1 = \left\{ \overbrace{(\underline{a_1}, \bar{a}_2, \bar{a}_3, \bar{a}_4, \dots, \bar{a}_{20})}^{p_1}, \overbrace{(\bar{a}_1, \underline{a_2}, \bar{a}_3, \bar{a}_4, \dots, \bar{a}_{20})}^{p_2}, \right. \\ \left. \overbrace{(\bar{a}_1, \bar{a}_2, \underline{a_3}, \bar{a}_4, \dots, \bar{a}_{20})}^{p_3}, \dots, \overbrace{(\bar{a}_1, \bar{a}_2, \bar{a}_3, \bar{a}_4, \dots, \underline{a_{20}})}^{p_{20}} \right\} \quad (32)$$

hvor vi bruker notasjonen p_1 for sannsynligheten at utfall ($\underline{a_1}, \bar{a}_2, \bar{a}_3, \bar{a}_4, \dots, \bar{a}_{20}$) inntreffer osv. Sannsynligheten for at kun ett av de 20 tilbudene blir akseptert er dermed:

$$P(A_1) = \sum_{i=1}^{20} p_i \quad (33)$$

hvor vi summerer opp til 20 siden det er 20 utfall i begivenheten A_1 i lign.(32).



Figur 10: Glamox.

- a) Hvordan vil du tolke sannsynligheten p_1 ?

Dersom du har gjort oppgave 4d rett så vet du at sannsynligheten for at et tilbud blir akseptert er:

$$P(A) = \frac{121}{350} \quad (34)$$

- b) Regn ut sannsynligheten p_1 ved å bruke lign.(34).¹⁴

- c) Bruk p_1 fra oppgave 5b til å regne ut tallverdien til $P(A_1)$ i lign.(33).¹⁵

¹⁴Husk at vi har antatt at aksept av tilbudene er uavhengige av hverandre. Da vet vi, ut fra den spesielle multiplikasjonssetningen, at sannsynligheten

$$p_1 = P(a_1, \bar{a}_2, \bar{a}_3, \bar{a}_4, \dots, \bar{a}_{20}) \quad (35)$$

er et produkt av totalt 20 sannsynligheter.

¹⁵Er sannsynlighetene $p_1, p_2, \dots, p_{20}$ like?

La oss definere begivenheten A_3 at 3 av de 20 tilbudene blir akseptert.
 Det betyr altså at 3 tilbud blir akseptert og at 17 tilbud ikke blir akseptert:

$$A_3 = \left\{ \overbrace{(a_1, a_2, a_3, \bar{a}_4, \dots, \bar{a}_{20})}^{\tilde{p}_1}, \overbrace{(\bar{a}_1, a_2, a_3, a_4, \dots, \bar{a}_{20})}^{\tilde{p}_2}, \right. \\ \left. \overbrace{(\bar{a}_1, \dots, a_{17}, a_{18}, a_{19}, \bar{a}_{20})}^{\tilde{p}_{M-1}}, \dots, \overbrace{(\bar{a}_1, \bar{a}_2, a_{18}, a_{19}, a_{20})}^{\tilde{p}_M} \right\} \quad (36)$$

hvor vi bruker notasjonen $\tilde{p}_1$ for sannsynligheten at utfall $(a_1, a_2, a_3, \bar{a}_4, \dots, \bar{a}_{20})$ inntreffer osv.

Sannsynligheten for at 3 av de 20 tilbudene blir akseptert er dermed:

$$P(A_3) = \sum_{i=1}^M \tilde{p}_i \quad (37)$$

hvor vi summerer opp til M siden det er M utfall i begivenheten A_3 i lign.(36).

- d)** Hvor mange måter kan vi oppnå at kun 3 aksepterte tilbud på? Altså finn M .¹⁶
- e)** Regn ut sannsynligheten $\tilde{p}_1$ ved å bruke lign.(34).¹⁷
- f)** Bruk svarene i oppgave **5d** og **5e** til å regne ut tallverdien til $P(A_3)$ i lign. (37).¹⁸

■

¹⁶Spiller rekkefølgen av tilbudene noe? Er det med eller uten tilbakelegging? Hva slags formel er det som da skal brukes for å finne antall kombinasjoner M ?

¹⁷Husk at vi har antatt at aksept av tilbudene er uavhengige av hverandre. Da vet vi, ut fra den spesielle multiplikasjonssetningen, at sannsynligheten

$$\tilde{p}_1 = P((a_1, a_2, a_3, \bar{a}_4, \dots, \bar{a}_{20})) \quad (38)$$

er et produkt av totalt 20 sannsynligheter.

¹⁸Er sannsynlighetene $\tilde{p}_1, \tilde{p}_2, \dots, \tilde{p}_M$ like?