
STK 1100 våren 2026

Sannsynlighetsregning og statistisk modellering: OBLIG II

Dette er OBLIG II, STK 1000, Sannsynlighetsregning og statistisk modellering, vår-semesteret 2026. Den har fem oppgaver og er på fem sider (i tillegg til denne innledende side 0). **Print ut** til ekte papir, for best lesbarhet.

Er gjort tilgjengelig: fra fredag 10. april, på kursets nettside; informasjon blir også gitt på forelesningene.

Innleveringsfrist: torsdag 23. april, 14:00, i *Canvas*-systemet.

Instruksjoner: du velger selv om du skriver besvarelsen for hånd og skanner besvarelsen, eller om du skriver løsningen direkte inn per datamaskin, for eksempel ved bruk av TeX eller LaTeX. Besvarelsen skal leveres som én pdf-fil. Skannede ark må være godt lesbare. Besvarelsen skal inneholde navn, emne og oblignummer. I oppgaver der du blir bedt om eller velger å programmere, må du legge programkoden inn i pdf-en sammen med resten av besvarelsen.

Skriver I, Karle: sett i gang, på bokmål, nynorsk, riksmål, latin, eller engelsk.

Samarbeid: og alle slags hjelpemidler er tillatt, men den innleverte besvarelsen skal være skrevet av deg og reflektere din forståelse av stoffet. Oppgi hvilke hjelpemidler du har brukt. Er vi i tvil om du virkelig har forstått det du har levert inn, kan vi be deg om en muntlig redegjørelse.

Om du vil: skriv gjerne en halvside om hvordan prosjektet har forløpt, hva som har vært strevsomt, hva som kanskje har vært lett, om du har «likt» oppgavene, om eventuelle frustrasjoner, osv. Du kan også gjerne skrive om grad av samarbeide med andre her. Vi leser alt dette med stor interesse.

Dessuten: (a) Av grunner fakultetet har grunnet seg frem til, er det lov å levere «blankt», men også slike må registreres i Canvas-systemet, for at studenten skal få lov til å ta eksamen. (b) Vi anbefaler alle å levere en fysisk papirversjon til en gruppelærer, altså i tillegg til pdf-leveringen i Canvas, for da får man *konkrete tilbakemeldinger*. Om det er praktisk for vanskelig å komme seg til kampus, kan man sende en mail til sin gruppelærer, med pdf-en som vedheng.

De to obliger i dette kurset må registreres og godkjennes, i dette samme semester, før man kan møte til den avsluttende eksamen.

Complete guidelines about delivery of mandatory assignments, along with a ‘log on to Canvas’, can be found here:

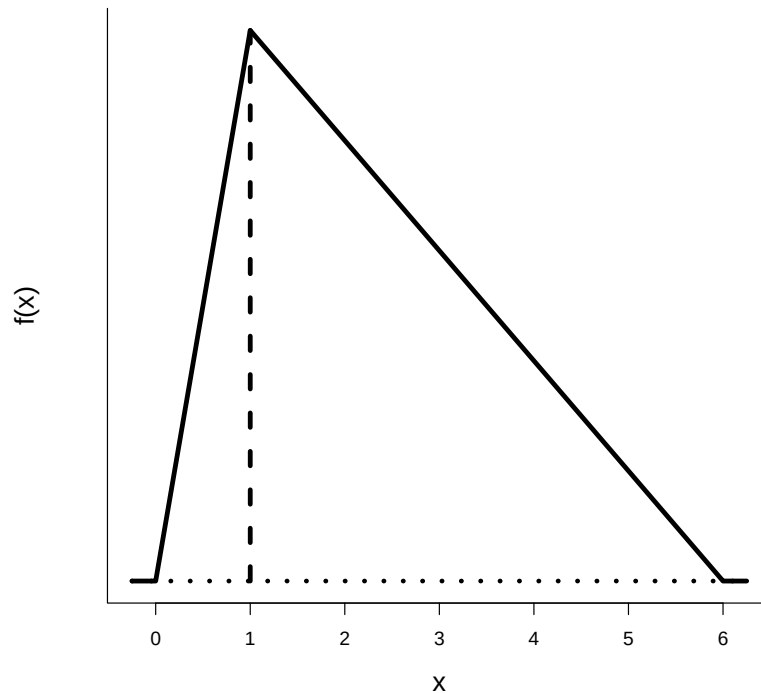
www.uio.no/english/studies/admin/compulsory-activities/mn-math-mandatory.html

Lær i vei!

Nils Lid Hjort, Emil Aas Stoltenberg

1. En trekantfordeling

I'VE DECIDED TO GET INTO SHAPE, and the shape I've selected is a triangle. Vi skal se på en fordeling med tetthet $f(x)$ gitt som i figuren: lineært oppover fra 0 til høyde h over $[0, 1]$, så lineært nedover fra denne høyde h til 0 over $[1, 6]$.



Trekanttettheten $f(x)$, positiv over $[0, 6]$, max-punkt i 1, null utenfor.

- (a) Skriv dette som $f(x) = ax$ over $[0, 1]$ og så $b(6 - x)$ over $[1, 6]$, der de to bitene skal henge sammen på toppen. Finn a og b og topphøyden h . Skalaene på x- og y-aksen er ikke den samme her, for denne figuren, så man kan ikke simpelthen måle seg frem til hva høyden er.
- (b) Hvis X har denne fordelingen, beregn $\Pr(X \leq 1)$ og $\Pr(X \geq 5)$.
- (c) Finn forventning og median for denne fordelingen.
- (d) Anta at $n = 1000$ observasjoner $X_1, \dots, X_n$ genereres fra denne fordelingen. Omtrent hvor mange havner i $[0, 1]$, og omtrent hvor mange i $[5, 6]$? Prøv også å lage et intervall, si $[A, B]$, som er slik at det er ca. 90 prosent sannsynlig at antallet N av datapunktene i $[0, 1]$ er innenfor denne $[A, B]$.

2. Uniform på sirkelen

WHEN IN DOUBT OR DANGER, run in circles, scream and shout. Her skal vi studere en sannsynlighetsfordeling $f(x, y)$ som skal være enkel nok å forstå, men der noen av de relevante utregninger kanskje ikke blir lette, likevel. Tettheten er $f(x, y) = 1/\pi$ på enhetssirkelen, altså de (x, y) der $x^2 + y^2 \leq 1$, og null ellers.

- (a) Bruk et par setninger til å forklare i ord hva denne fordelingen er, og om hva slags utfall (X, Y) man kan regne med å få.
- (b) Hvis (X, Y) trekkes fra denne fordelingen, finn da sannsynlighetene for at variabelen $R = \sqrt{X^2 + Y^2}$ havner i $[0.1, 0.2]$, og i $[0.8, 0.9]$. Finn dessuten forventningen til R .
- (c) Finn sannsynligheten for $\{X \geq \frac{1}{2}, Y \geq \frac{1}{2}\}$, altså for at begge hendelser inntreffer.
- (d) Finn så tettheten $f_1(x)$ i marginalfordelingen for X , og tilsvarende tettheten $f_2(y)$ for Y . Er X og Y uavhengige? Finn tettheten for X , gitt at $Y = \frac{1}{2}$.
- (e) Beregn sannsynligheten for at $X \leq \frac{1}{2}$.

3. Batterier

ETTER EKSAMEN KAN DET VÆRE TID FOR å lade batteriene. Du skal reise til og oppholde deg på en nokså øde øy i 12 måneder, uten butikker og uten mange hjelpemidler, og trenger batterier for at ditt radiosamband skal fungere. Levetiden X for ett batteri er eksponensialfordelt med parameter $\lambda = 1$, altså med tettheten $f(x) = e^{-x}$ for $x \geq 0$, der tidsenheten er 1 måned. Straks et batteri er utladet, setter du med et lite sukk inn det neste.

- (a) Vis at den momentgenererende funksjon for denne fordelingen blir

$$M(t) = E e^{tX} = \frac{1}{1-t} \quad \text{for } t < 1.$$

Vis fra dette at forventning og varians for X er lik 1.

- (b) Finn også den såkalte halveringstiden, medianen i fordelingen, og forklar forskjellen mellom forventning og median i denne situasjonen.
- (c) Dersom du tar med deg m batterier, er den samlede levetid $S_m = X_1 + \dots + X_m$, der $X_1, \dots, X_m$ er uavhengige fra denne eksponentialfordelingen med parameter 1. Finn forventning og standardavvik for S_m .
- (d) En god tommelfingerregel sier at for en variabel Z , med forventning ξ og standardavvik σ , gjelder det at ca. 95 prosent av alle utfall av Z vil havne i intervallet $[\xi - 2\sigma, \xi + 2\sigma]$ (dersom fordelingen for Z ikke er særlig skjev eller $\sigma/\sqrt{m}$ er særlig høy). Anvend denne regelen her, til å sjekke i hvilket intervall S_m altså antagelig vil havne, for $m = 12$ (en for hver måned i ditt øy-år) og for $m = 24$. Hvor mange batterier bør du ha med deg, for å være ganske sikker på at du fortsatt har radiosambandet i orden etter ett år?
- (e) Regnestykket over er ikke presist, siden vi der bare bruker denne tommelfingerregelen. Men læreboken definerer i Ch. 4 en *gamma-fordeling*, med positive parametre (α, β) , ved tettheten

$$g(x, \alpha, \beta) = \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} \exp(-x/\beta) \quad \text{for } x > 0,$$

og viser at dens genererende funksjon er $M(t) = 1/(1 - \beta t)^\alpha$ for $t < 1/\beta$. Bruk dette til å vise at S er gamma-fordelt med parametre $(m, 1)$.

- (f) Ved hjelp av dette, og passende kode fra `python` eller `R` for den kumulative gammafordelingen, undersøk sannsynligheten $p_m = \Pr(S_m \geq 12)$ for at du fortsatt har radiosamband fra din øde øy etter 12 måneder. Hvor mange batterier må du ha med deg, for at sannsynligheten er 99 prosent for at alt er i orden etter disse 12 måneder?

4. Poisson-estimering

ONE MAN'S FISH is another man's Poisson. Anta at X, Y, Z er uavhengige observasjoner, relatert til en ukjent Poisson-rate θ , men med ulike eksponeringstider; de har parametre $a\theta, b\theta, c\theta$, der a, b, c er kjente positive tall (f.eks. som i antall timer som ligger bak de tre observasjonstallene).

- (a) Skriv ned den fulle sannsynlighetsfordelingen, ved $p(x, y, z) = \Pr(X = x, Y = y, Z = z)$.
- (b) Det er kjent fra pensum & forelesninger at hvis V er Poisson-fordelt med parameter λ , så er både forventning og varians for V lik denne λ . Anvend dette til å vise at $X/a, Y/b, Z/c$ er forventningsrette estimatorer for θ .
- (c) Argumenter for at estimator A, gitt ved

$$\hat{\theta}_A = \frac{X/a + Y/b + Z/c}{3},$$

er fornuftig. Finn dens forventning og varians.

- (d) Hva er fordelingen til $X + Y + Z$?
- (e) Se også på estimatoren

$$\hat{\theta}_B = \frac{X + Y + Z}{a + b + c}.$$

Finn dens forventning og varians. Hvilken av de to estimatorer vil du foretrekke, og hvorfor?

5. Hvilken karakter får du på eksamen?

I DETTE KURSET FÅR DU EN KARAKTER, der oppskriften lærere og sensorer forholder seg til sier følgende: først beregnes $z = (1/3)x + (2/3)y$, der x og y er dine poengsummer fra henholdsvis midtveis i mars og avsluttende i juni, begge transformert til 0-100-skalaen, hvorpå z blir noe brutalt forenklet til bokstavkarakterer (altså fra en temmelig kontinuerlig skala til en med seks bokser):

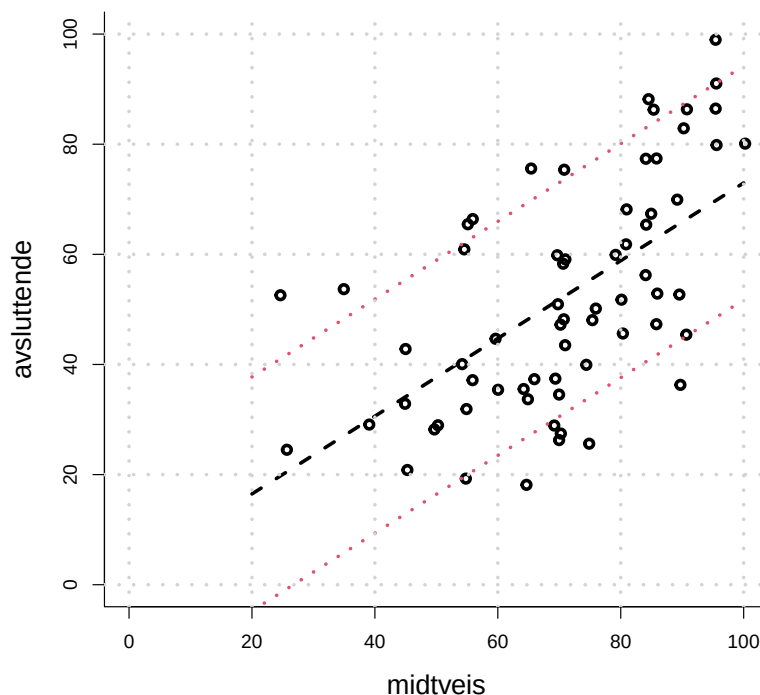
92-100 er A, 77-91 er B, 58-76 er C, 46-57 er D, 40-45 er E, 0-39 er F.

Nils og Emil skal nok lage seg et datasett for dere, STK 1100-studenter vårsemesteret 2026, i juni, med n omtrent 170. Men for å gjennomføre denne oppgaven ser vi på et analogt datasett fra 2013, da Nils ga kurset STK 1000. Antall studenter da var omtrent 250, men Nils har i sine filer bare (x, y) for $n = 65$ studenter, se plottet.

Modellen vi bruker i denne oppgaven (andre varianter kan diskuteres senere) er den klassiske lineær-normale regresjon, som sier at

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \quad \text{for } i = 1, \dots, n,$$

for de $n = 65$ studenter hin gang, der ε_i -ene er uavhengige og $N(0, \sigma)$ -fordelte, altså med standardavvik σ .



Her er $x = \text{midtveiskarakter}$, skalert opp til 0-100 skala, og $y = \text{avsluttende eksamenskarakter}$, for $n = 65$ studenter i STK 1000, som Nils ga i 2013. Korrelasjon er 0.60, regresjonslinjen er estimert til $\hat{y} = 2.37 + 0.71x$, og $\hat{\sigma} = 16.57$. Det er sannsynlighet 0.80 for å havne innenfor de prikkede linjer.

- Skriv noen setninger om hva denne modellen innebærer, og diskuter kort om den ser realistisk ut.
- I kursets og lærebokens Kapittel 12 lærer vi diverse om denne evig berømte og evig brukbare modell, inkludert den nødvendige kunst det er å estimere de tre parameterne (to for linjen, en for variansen), med presise metoder som tar dissens usikkerhet med i analyser, osv. Men her skal vi tillatte oss å se på den estimerte regresjonslinje

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x = 2.37 + 0.71x$$

som gitt og korrekt, også med den estimerte $\hat{\sigma} = 16.57$ som riktig verdi. – Anta nå at du fikk 17 av 20 poeng på midtveiseksamen, altså $x = 85$. Finn fordelingen for y , og etterpå for den avgjørende total-score z . Hva er sannsynligheten for at du får karakteren A i dette kurset, under modellens forutsetninger?

- (c) Anta at du har vært så flink at du fikk 20 av 20 på midtveis (gratulerer). Hva er sannsynligheten for at du «bare» får karakteren B (igjen, under modellens antagelser)?
- (d) Vi ser her at den enkle lineær-normale modellen kan anvendes til prediksjon av y ut fra x . Den er neppe perfekt: gjør deg noen tanker om hvordan en slik modell kanskje kan modifiseres, eventuelt med mulig tilleggsinformasjon, for å lage bedre prediksjoner for hvordan du gjør det på eksamen, altså ut fra hvordan du har gjort det på midtveiseksamen.
- (–) En *footnote* anbringes her, om temaet som tas opp i denne oppgaven: Lærere og sensorer er også instruert til å utvise godt skjønn, ved fastsettelse av sluttkarakterer, så den matematiske algoritmen over, altså transformasjon av $(1/3)x + (2/3)y$ til A–F, blir ikke fulgt slavisk. «Svært gode svar på vanskelige punkter», der ikke bare det såkalt riktige svar er landet på, men der den forklarende prosa også er konsis og klar, er sensorer alltid svake for, for eksempel. Det er full anledning til å øve seg på dette for nærværende Oblig II.

Litt litteratur

- Ladislaus von Bortkiewicz (Владислав Иосифович Борткевич) (1898). *Das Gesetz der kleinen Zahlen* [de små talls lov]. B.G. Teubner, Leipzig. Poisson-fordelingen defineres på første side (mens selve navnet Poisson-fordelingen kom senere). På sidene 23–25 finner vi: «Beispiel: Die durch Schlag eines Pferdes im preussischen Heere Getöteten», der Poisson-fordelingen passer presist til fordelingen av antall soldater døde av hestespark. Det passer jo med «høy n , lav p » i en binomisk fordeling.
- Peter David (1998, men handler jo om det 24de århundre): *Triangle*. Star Trek: The Next Generation. Mass Market Paperback, Los Angeles.
- Jay L. Devore, Kenneth N. Berk, Matthew A. Carlton (2021): *Modern Mathematical Statistics with Applications* [3rd ed.], Springer.
- Nils Lid Hjort og Emil Aas Stoltenberg (2026): *Statistical Inference: 600 Exercises and 100 Stories*. Cambridge University Press, Cambridge, UK. [Ferdig om ikke så svært lang tid.]
- Nils Lid Hjort (2026). Det ultimate femtusenmeterskeiserennet. Kronikk, *Dag og Tid*, 16. mars.
- Daniel Kehlmann (2005): *Die Vermessung der Welt*. Reinbek bei Hamburg. Roman (!), om Carl Friedrich Gauß (1777–1855; følg med, når han finner på minste kvadratsummetoden og dermed regresjonsanalyse), og Alexander von Humboldt (1769–1859; han som ifølge Andrea Wulfs bok var den som fant opp *naturen*). I kinofilm-versjonen kommer Gauß på minste kvadratsum midt under bryllupsnatten.
- Judith Schalansky (2009): *Atlas der abgelegenen Inseln: fünfzig Inseln, auf denen ich nie war und niemals sein werde*. Mare Verlag, Berlin.
- Jan Erik Vold (1994): *En sirkel is*. Gyldendal.