PRØVEEKSAMEN STK 1100, MAI 2026

ved Nils Lid Hjort og Emil Aas Stoltenberg

Lagt ut på verdensveven man 11. mai; gjennomgått man 18. og ons. 20. mai

Forslag til løsning (med noen randbemerkninger underveis)

Oppgave 1: din statistiske verktøykasse

La oss sjekke om ting er på stell.

(a) Vi trenger (i) at f har integral 1 og (ii) at den er ikkenegativ over $[0,1]$. (i) holder, ser vi, så lenge nevneren ikke er null, altså $a \neq -2$. Hvis $a < -1$ går det galt; da skifter telleren fortegn underveis. Men alle andre verdier er ok; altså $a \geq -1$. Tettheten er pen og lineær, og tilfellet $a = 0$ svarer til den uniforme fordelingen. Forventningen er

$$\mathrm{E}\, X = \frac{1}{1+\frac{1}{2}a}\int_0^1 x(1+ax)\,\mathrm{d}x = \frac{\frac{1}{2}+\frac{1}{3}a}{1+\frac{1}{2}a}.$$

Med 100 observasjoner X_i kan vi bruke momentmetoden: det empiriske gjennomsnitt $\bar{X} = (1/100)\sum_{i=1}^{100} X_i$ settes lik det teoretiske gjennomsnitt, altså forventningen. Løsningen $\widehat{a}$ av ligningen

$$\bar{X} = \frac{\frac{1}{2}+\frac{1}{3}a}{1+\frac{1}{2}a}$$

er momentestimatoren. Litt regning gir

$$\bar{X} + \tfrac{1}{2}a\,\bar{X} = \tfrac{1}{2} + \tfrac{1}{3}a, \quad \bar{X} - \tfrac{1}{2} = \tfrac{1}{3}a - \tfrac{1}{2}a\bar{X}, \quad \text{og dermed } \widehat{a} = \frac{\bar{X}-\frac{1}{2}}{\frac{1}{3}-\frac{1}{2}\bar{X}} = \frac{6\bar{X}-3}{2-3\bar{X}}$$

(sannsynligheten er null for å få $\bar{X} = 2/3$). Hvis $\bar{X}$ er i nærheten av $\frac{1}{2}$, er estimatet i nærheten av null, som passer med den uniforme fordelingen. Vi kan også se på maximum likelihood-metoden. ML-estimatoren er den som maksimerer funksjonen

$$\ell(a) = \sum_{i=1}^{n}\{\log(1+aX_i) - \log(1+\tfrac{1}{2}a)\}.$$

Dette er lett nok numerisk, «plott funksjonen og se hvor den er størst», eller «ta den deriverte og se hvor den er null», men det er ingen eksplisitt formel. Under modellens betingelser vil ML være mer presis.

(b) Det empiriske gjennomsnittet er $\bar{X} = (1/24)\cdot 179.234 = 7.468$. Vi har videre, med $n = 24$,

$$Z = \sum_{i=1}^{n}(X_i - \bar{X})^2 = \sum_{i=1}^{n} X_i^2 - n\bar{X}^2 = 1362.573 - 24\cdot 7.468^2 = 24.0385,$$

og dermed det empiriske standardavviket

$$\widehat{\sigma} = \sqrt{\frac{Z}{n-1}} = \Big\{\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2\Big\}^{1/2} = 1.0223.$$

Nevneren er $n-1$ for å sørge for at $\mathrm{E}\,\widehat{\sigma}^2 = \sigma^2$, hvilket henger i hop med at $\mathrm{E}\,Z = (n-1)\sigma^2$; se under. For å lage konfidensintervall er det to relevante brøker å studere:

$$Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}, \qquad T = \frac{\bar{X}-\mu}{\widehat{\sigma}/\sqrt{n}}.$$

Den første er eksakt $\mathrm{N}(0,1)$-fordelt. Altså er $\Pr(-z_0 \le Z \le z_0) = 0.90$ for $z_0 = 1.645$, 0.95-kvantilen for $\mathrm{N}(0,1)$. Løser man disse to ulikheter, for å få μ i midten, får man

$$\Pr(\bar{X} - z_0\sigma/\sqrt{n} \le \mu \le \bar{X} + z_0\sigma/\sqrt{n}) = 0.90.$$

Dette er jo fint (og helt korrekt) – men i prinsippet ikke brukbart, siden σ er ukjent. Hvis man plugger inn estimatet 1.0223 finne man $[L,U] = [7.125, 7.811]$. Dette er jo det man får ved å bruke T, ikke Z. Problemet er at T, som er den vi (faktisk) bruker, bare er en (god) tilnærming til Z. Den virkelige fordelingen til T er en t-fordeling med $n = 1 = 23$ frihetsgrader, jfr. bokens Ch. 8.3, men det er akkurat utenfor pensum (og kommer i STK1110 m.m.). Så «det forventede riktige svar», for STK1100, er å komme frem til disse tallene, men huske på å si *tilnærmet* 90 prosent (fordi fordelingen til T er i nærheten av den for Z, med høy sannsynlighet). Gjør man det nøye, via T, trenger man 0.95-kvantilen 1.714, noe større enn 0.95-kvantilen 1.645 for $\mathrm{N}(0,1)$, og intervallet blir litt større, men ikke så meget: $[L', U'] = [7.110, 7.826]$.

Fotnote: den som oppdaget at dette faktisk er et problem, «plugg inn $\widehat{\sigma}$ og håp det beste», het W.S. Gosset, og var tallmann i Londons ølindustri. Han kalte seg «Student» (fint navn), skrev artikkelen *The Probable Error of a Mean* i *Biometrika* i 1908, og er evig verdensberømt; t-fordelingen er oppkalt etter ham og hans arbeid, og ordene «studentise», «studentify» er i bruk (vi må dele på en god nevner og så regne på konsekvensene).

(c) Når $X_i \sim \mathrm{N}(\mu, \sigma^2)$, er $\mathrm{E}\,(X_i-\mu)^2 = \sigma^2$. Så $\mathrm{E}\,U = 24$. ML-estimatoren for μ er $\widehat{\mu} = \bar{X}$. Som over har vi $Z = \sum_{i=1}^n (X_i - \bar{X})^2 = \sum_{i=1}^n X_i^2 - n\bar{X}^2$, med forventning

$$\sum_{i=1}^n (\mu^2 + \sigma^2) - n(\mu^2 + \sigma^2/n) = n\sigma^2 - \sigma^2 = (n-1)\sigma^2.$$

Altså er $\mathrm{E}\,V = 23$. Det er også slik at $U \sim \chi^2_{24}$, mens $V \sim \chi^2_{23}$ («den har mistet en frihetsgrad»); mer om chi-squared og om t-fordelinger i de neste kurs.

(d) Vi ser at den deriverte av $\exp(-y^4)$ er $-4\exp(-y^4)y^3$. Dermed finner vi den kumulative $H(y) = 1 - \exp(-y^4)$. Denne er lik $\frac{1}{2}$ når $y^4 = \log 2$, så medianen er $(\log 2)^{1/4} = 0.912$. Tettheten for $Z = Y^4$ finner vi fra transformasjonssetningen:

$$g(z) = h(z^{1/4})(1/4)z^{1/4-1} = 4\exp(-z)(z^{1/4})^3(1/4)z^{-3/4} = \exp(-z)$$

for $z > 0$. Den er altså standard expo-fordelt.

(e) La X ha Poisson-fordelingen med parameter λ. Da er dens mgf lik

$$M(t) = \mathrm{E}\,\exp(tX) = \sum_{x=0}^{\infty} \exp(-\lambda)(\lambda e^t)^x/x! = \exp(-\lambda + \lambda e^t).$$

Hvis nå X og Y er uavhengige Poisson, med parametere λ_1 og λ_2, har $X+Y$ mgf lik

$$\begin{aligned} M_{X+Y}(t) &= \mathrm{E}\,\exp(t(X+Y)) = \mathrm{E}\,\exp(tX)\exp(tY) \\ &= M_X(t)M_Y(t) = \exp(-\lambda_1+\lambda_1 e^t - \lambda_2 + \lambda_2 e^t). \end{aligned}$$

Men dette er sannelig det samme som $\exp(-\lambda+\lambda e^t)$, med $\lambda=\lambda_1+\lambda_2$. Altså er $X+Y$ også Poisson, med summen av parameterne (siden mgf-er karakteriserer fordelinger).

Fotnote: merk at dette passer riktig fint med $\mathrm{E}\,(X+Y)$ og $\mathrm{Var}\,(X+Y)$.

Oppgave 2: en mynt og en annen mynt

Det er to mynter i en hatt, og de ser prikk like ut. Den ene er «vanlig», med sjanse en halv for Kron; den andre har derimot sjanse to tredjedeler for Kron.

(a) Det er to muligheter, den vanlige V og den uvanlige U, hver med sjanse $\frac12$, så vi finner

$$p = \tfrac12\,\Pr(K\,|\,V) + \tfrac12\,\Pr(K\,|\,U) = \tfrac12\tfrac12 + \tfrac12(2/3) = 1/4 + 1/3 = 7/12,$$

en liten tanke høyere enn den vanlige $\frac12$.

(b) Du får Kron. Da finner man

$$q = \Pr(V\,|\,K) = \frac{\Pr(V,K)}{\Pr(K)} = \frac{\frac12\frac12}{\frac12\frac12+\frac12\frac23} = \frac{1}{1+4/3} = \frac37.$$

Dette er Bayes' regel, men man kan jo finne svaret uten å tenke på eller vite om det.

(c) Spenningen stiger!, og det ble Kron igjen.

$$r = \Pr(V\,|\,K_1,K_2) = \frac{\Pr(V,K_1,K_2)}{\Pr(K_1,K_2)} = \frac{\frac12\frac12\frac12}{\frac12\frac12\frac12+\frac12\frac23\frac23} = \frac{1}{1+16/9} = 9/25.$$

Dette tallet, 0.360 er noe lavere enn isted, 0.428; jo flere kron, jo rimeligere er det at du har funnet den uvanlige U, ikke den vanlige V.

(d) Begivenhetene K_1 og K_2 har begge sannsynlighet $p = 7/12$. Vi må sjekke om $\Pr(K_1)\Pr(K_2) = p^2$ er det samme som

$$\Pr(K_1\cap K_2) = \tfrac12(\tfrac12)^2 + \tfrac12(2/3)^2 = \frac18 + \frac{4}{2\cdot 81} = \frac18 + \frac29 = \frac{25}{72}.$$

Nei, $25/72 = 0.347$ er ikke det samme som $49/144 = 0.340$, men forskjellen er mindre enn det den skrivende hadde rukket å forestille seg.

Oppgave 3: en mynt, en terning, og et sannsynlighetsrom

Jeg ser på de ludene heste og på utfallsrommet

$$\Omega = \{(\omega_1,\omega_2)\colon \omega_1\in\{k,m\}, \omega_2\in\{1,2,3,4,5,6\}\}.$$

Navnet «potensmengden» er ikke så viktig her, men vi ser altså på alle mulige delmengder A av Ω.

(a) Dette handler om en mynt, med utfall k, m, og en terning, med utfall $1, 2, 3, 4, 5, 6$; det er i alt 12 elementer i utfallsrommet. Funksjonen

$$\Pr(A) = \frac{|A|}{12}, \quad \text{for } A \in 2^{\Omega},$$

teller opp antall elementer i A, og deler på det totale antall. Vi har $\Pr(\emptyset) = 0$, $\Pr(\Omega) = 1$, og $\Pr(A \cup B) = \Pr(A) + \Pr(B)$, for disjunkte A og B. Altså har vi et sannsynlighetsrom, $(\Omega, \Pr)$.

(b) Vi ser på

$$A = \{(\omega_1, \omega_2) \colon \omega_1 \in \{k, m\}, \omega_2 \in \{1, 3, 5\}\},$$
$$B = \{(\omega_1, \omega_2) \colon \omega_1 \in \{k, m\}, \omega_2 \in \{5, 6\}\}.$$

som kan oversettes til «terningen gir 1 eller 3 eller 5» og «terningen git 5 eller 6»; hva som skjer med myntkastet affiserer ikke A eller B. Vi teller i vei, og finner $\Pr(A) = 6/12 = 3/6 = 1/2$ og $\Pr(B) = 4/12 = 2/6 = 1/3$. Dessuten $\Pr(A \cap B) = 2/12 = 1/6$. Er det nå slik at $\frac{1}{2} \cdot \frac{1}{3}$ er lik $\frac{1}{6}$? Ja, så sannelig; de er uavhengige.

(c) Funksjonen

$$X(\omega_1, \omega_2) = \begin{cases} 1 & \text{hvis } \omega_1 = k, \\ 0 & \text{hvis } \omega_1 = m \end{cases}$$

er virkelig en funksjon fra sannsynlighetsrommet Ω inn i de reelle tall – ja, faktisk, bare inn på $\{0, 1\}$. Altså en tilfeldig variabel. Dens fordeling er

$$\Pr(X = 1) = 6/12 = 1/2, \quad \Pr(X = 0) = 6/12 = 1/2,$$

formelt sett ved telling av antall muligheter i Ω.

(d) Vi ser på $Y = X + I_A + I_B$, de to siste indikatorvariable, i likhet med X. De mulige verdier for Y er 0, 1, 2, 3. La $g(x, a, b) = \Pr(X = x, I_A = a, I_b = b)$. Da finner vi, nærmest per opptelling og bokholderi,

$$\Pr(Y = 0) = g(0, 0, 0),$$
$$\Pr(Y = 1) = g(1, 0, 0) + g(0, 1, 0) + g(0, 0, 1),$$
$$\Pr(Y = 2) = g(1, 1, 0) + g(1, 0, 1) + g(0, 1, 1),$$
$$\Pr(Y = 3) = g(1, 1, 1),$$

som per videre bokholderi leder oss til

$$\Pr(Y = 0) = (1/2)(2/6) = 2/12,$$
$$\Pr(Y = 1) = (1/2)(2/6) + (1/2)(2/6) + (1/2)(1/6) = 5/12,$$
$$\Pr(Y = 2) = (1/2)(2/6) + (1/2)(1/6) + (1/2)(1/6) = 4/12,$$
$$\Pr(Y = 3) = (1/2)(1/6) = 1/12.$$

Dens kumulative fordeling har verdiene 2/12 i 0, 7/12 i 1, 11/12 i 2, og 12/12 i 3, og er stykkevis lineær.

Fotnote: situasjonen i oppgaven er egentlig «temmelig enkel», og spørsmålene kan besvares uten den noe tunge formalismen: det handler om en mynt og om en terning, der hva som skjer med den ene er uavhengig av den andre. Stat- og prob-folk i farta vil neppe sette opp disse spørsmålene på denne måten. Men oppgavens poeng er «trening og klargjøring» i denne type formalisme, fordi man trenger dette i de langt mer komplekse situasjoner.

Oppgave 4: en øvre grense å estimere

Vi har kommet til funksjonen gitt ved

$$f(x) = \frac{3x^2}{\theta^3} \quad \text{for } x \in [0, \theta],$$

og $f(x) = 0$ ellers, der θ er en ukjent positiv parameter vi ønsker å estimere.

(a) Integralet $\int_0^\theta 3x^2 \, \mathrm{d}x$ er θ^3, så f er virkelig en tetthet. Dens kumulative blir

$$F(x) = \begin{cases} 0 & \text{for } x < 0, \\ (x/\theta)^3 & \text{for } 0 \le x \le \theta, \\ 1 & \text{for } x > \theta. \end{cases}$$

(b) Vi finner

$$\mathrm{E}\, X = \int_0^\theta 3x^3 \, \mathrm{d}x/\theta^3 = (3/4)\theta,$$

$$\mathrm{E}\, X^2 = \int_0^\theta 3x^4 \, \mathrm{d}x/\theta^3 = (3/5)\theta^2,$$

dermed også

$$\mathrm{Var}\, X = \{(3/5) - (3/4)^2\}\theta^2 = (3/80)\theta^2.$$

(c) Med n uavengige observasjoner X_i, med snitt $\bar{X}_n = (1/n)\sum_{i=1}^n X_i$, vil $\mathrm{E}\, \bar{X}_n = (3/4)\theta$, så den forventningsrette estimatoren er $\widehat{\theta} = a\bar{X}_n$, med $a = 4/3$.

(d) Tiden er kommen for $M_n = \max(X_1, \ldots, X_n)$. Dens kumulative blir

$$\Pr(M_n \le z) = \Pr(X_1 \le z, \ldots, X_n \le z) = F(z)^n = (z/\theta)^{3n},$$

for $z \in [0, \theta]$. Vi deriverer og finner tettheten

$$g_n(z) = 3nz^{3n-1}/\theta^{3n} \quad \text{for } z \in [0, \theta].$$

Forventningen blir

$$\mathrm{E}\, M_n = \int_0^\theta 3nz^{3n} \, \mathrm{d}z/\theta^{3n} = \frac{3n}{3n+1}\theta,$$

så den forventningsrette versjonen blir

$$\widetilde{\theta}_n = b_n M_n = \frac{3n+1}{3n} M_n.$$

(d) Siden begge estimatorene er forventningsrette, skal vi altså sjekke de to varianser. For den første har vi

$$\mathrm{Var}\, \widehat{\theta}_n = a^2 \, \mathrm{Var}\, \bar{X}_n = (4/3)^2(3/80)/n = (1/15)/n.$$

Videre finner vi

$$\mathrm{E}\, M_n^2 = \int_0^\theta 3nz^{3n+1}\,\mathrm{d}z/\theta^{3n} = \frac{3n}{3n+2}\theta^2,$$

som medfører at

$$\mathrm{Var}\,\widetilde{\theta}_n = b_n^2 \mathrm{Var}\, E_n^2 - (\mathrm{E}\, b_n M_n)^2 = c_n\theta^2,$$

med faktoren

$$c_n = \Big(\frac{3n+1}{3n}\Big)^2 \frac{3n}{3n+2} - 1 = \frac{(3n+1)^2 - 3n(3n+2)}{3n(3n+2)} = \frac{1}{3n(3n+2)}.$$

Det er denne som vinner – variansbrøken blir

$$\frac{1}{3n(3n+2)}\Big/\frac{1/15}{n} = \frac{15n}{3n(3n+2)} = \frac{5}{3n+2},$$

som også er drastisk mindre enn 1 når n øker.

Fotnote: prøv gjerne å vise at

$$\Pr(n(\theta - M_n) \le w) \to 1 - \exp(-3w/\theta),$$

som betyr at variabelen $W_n = n(\theta - M_n)$ konvergerer i fordeling mot expo-fordelingen med parameter $3/\theta$.

Oppgave 5: lineær trend pluss støy

For gitte målepunkter $x_1, \ldots, x_n$, alle positive, anta at $Y_i \sim \mathrm{N}(\beta x_i, 1)$.

(a) Siden vi alle kan normaltetthetssformelen, blir likelihood-funksjonen

$$L = \prod_{i=1}^n \Big[\frac{1}{(2\pi)^{1/2}} \exp\{-\tfrac{1}{2}(y_i - \beta x_i)^2\}\Big] = \frac{\exp\{-\frac{1}{2}Q(\beta)\}}{(2\pi)^{n/2}},$$

der $Q(\beta) = \sum_{i=1}^n (Y_i - \beta x_i)^2$ er kvadratsummen. Logaritmen blir

$$\ell(\beta) = -\tfrac{1}{2}Q(\beta) - \tfrac{1}{2}n\log(2\pi),$$

og vi forstår at ℓ er maksimal der Q er minimal. Altså er ML-estimatoren og OLS, Ordinary Least Squares, minste kvadratsum-etimatoren, identiske, her. Vi har

$$Q'(\beta) = -2\sum_{i=1}^n (Y_i - \beta x_i)x_i,$$

som er null når $\sum_{i=1}^n x_i Y_i = \beta \sum_{i=1}^n x_i^2 = \beta M$. Altså

$$\widehat{\beta} = \frac{\sum_{i=1}^n x_i Y_i}{\sum_{i=1}^n x_i^2} = \sum_{i=1}^n x_i Y_i/M.$$

(b) Fra dette finner vi

$$\mathrm{E}\,\widehat{\beta} = \sum_{i=1}^n x_i\,\mathrm{E}\, Y_i/M = \sum_{i=1}^n x_i\beta x_i/M = \beta,$$

$$\mathrm{Var}\,\widehat{\beta} = \sum_{i=1}^n \mathrm{Var}\,(x_i Y_i)/M^2 = \sum_{i=1}^n x_i^2/M^2 = 1/M.$$

(d) Dessuten er jo $\widehat{\beta}$ normalfordelt, qua lineærkombinasjon av uavhengige normale: $\widehat{\beta} \sim \mathrm{N}(\beta, 1/M)$. Dette betyr

$$\Pr(-1.96 \le \frac{\widehat{\beta} - \beta}{\sqrt{1/M}} \le 1.96) = 0.95.$$

Men begivenheten, med denne sannsynligheten, er jo det samme som

$$-1.96/M^{1/2} \le \beta - \widehat{\beta} \le 1.96/M^{1/2},$$

etter mild omordning, som betyr at $\widehat{\beta} \pm 1.96/M^{1/2}$ er et 95% konfidensintervall.

Fotnote: all regningen er altså nokså pen, i denne idylliske enkle regresjonsmodellen – det er bare én β, og variansen er kjent, lik 1. Poenget er at dette kan løftes temmelig fagert til den generelle lineær-normale modellen, med mer kompliserte, men tross alt nokså analoge, formler. Og igjen er ML = OLS, for denne type pene modeller (men ikke ellers, som for logistisk regresjon).

Oppgave 6: firkanter i prosess

Verden står ikke helt stille, og mange objekter er i vedvarende endring. Så våre firkanter er $F_n = [0, X_n] \times [0, Y_n]$, med

$$X_n = U_1 \cdots U_n, \quad Y_n = V_1 \cdots V_n,$$

der U_i-er og V_i-er er uavhengige uniformt over $[0, 2]$. Det er også praktisk å skrive

$$U_i = 2\,U_i', \quad V_i = 2\,V_i',$$

der U_i' og V_i' er standardiserte uniforme, over $[0, 1]$.

(a) Her er $\mathrm{E}\,U_i = \mathrm{E}\,V_i = 1$. Så variansene er $4\,\mathrm{Var}\,U_i' = 4/12 = 1/3$. Dermed har vi

$$\mathrm{E}\,X_n^2 = \mathrm{E}\,U_1^2 \cdots U_n^2 = \prod_{i=1}^{n} (1^2 + 1/3) = (4/3)^n,$$

med en temmelig stor varians, $\mathrm{Var}\,X_n = (4/3)^n - 1$, og tilsvarende for Y_n. For arealet $A_n = X_n Y_n$ har vi dermed

$$\mathrm{E}\,A_n = 1, \quad \mathrm{Var}\,A_n = \mathrm{E}\,(X_n Y_n)^2 - 1 = (4/3)^{2n} - 1.$$

Så riktignok kan vi «forvente» areal 1, men variansen er diger.

(b) At F_n er større enn F_{n-1} (altså i begge retninger, ikke bare i areal) betyr at $U_n > 1, V_n > 1$, og sjansen for det er $\frac{1}{2}\frac{1}{2} = 1/4$.

(c) Vi har

$$A_n = X_nY_n = X_{n-1}U_n \cdot Y_{n-1}V_n = A_{n-1}U_nV_n,$$

så at arealet er større betyr at $U_nV_n = 4\,U_n'V_n' > 1$. Vi skal altså finne sannsynligheten

$$p = \Pr(U'V' > \tfrac14) \quad \text{der } U' \text{ og } V' \text{ er uavhengige og uniforme på } [0,1].$$

Vi tar oss tid til så mange som tre gode spor her (med samme svar). (i) Paret (U', V') har konstant tetthet 1 på enhetskvadratet, så p er arealet av den veldefinerte mengden $\{u', v')\colon u'v' > \frac14\}$ – tegn opp $v' = \frac14/u'$, og overbevis deg om at

$$p = \int_{1/4}^1 (1 - \tfrac14/u')\,\mathrm{d}u' = 3/4 - \tfrac14\Big|_{1/4}^1 \log u' = 3/4 + 1/4\log(1/4) = 3/4 - (1/2)\log 2,$$

som er 0.4034. (ii) Vi kan like godt interessere oss for fordelingen til $W = U'V'$, og så lese ut $\Pr(W \ge \frac14)$ fra dette. Utfallsrommet for denne W er $[0,1]$. Kjør

$$\begin{aligned}\Pr(W \le w) &= \int_0^1 \Pr(U'V' \le w \,|\, U' = u)\,1\,\mathrm{d}u \\ &= \int_0^1 \Pr(V' \le w/u)\,\mathrm{d}u \;= \int_0^w 1\,\mathrm{d}u + \int_w^1 w/u\,\mathrm{d}u = w - w\log w.\end{aligned}$$

Sannsynligheten vi etterspør er altså $p = \Pr(W \ge \frac14)$, og vi finner samme svar som over. (iii) Vi kan også lett nok simulere oss frem til en god numerisk approksimasjon av denne p. Lage 1 million U' og 1 million V', og tell opp hvor ofte produktet er over $\frac14$; se under.

(d) Over svarte vi på «er arealet blitt større»; det er lettere å svare på «er firkanten blitt større (i begge retninger)». Det handler om $X_n > X_{n-1}$ og $Y_n > Y_{n-1}$, altså $U_n > 1$ og $V_n > 1$, og sannsynligheten for det er jo $\frac12\frac12 = \frac14$, som vi også så på over. Sjansen for at dette skal skje fire ganger på rad blir $(\frac14)^4 = 1/256$. Det blir samme svar som for fire ganger på rad med krymping i begge retninger.

(e) Vi bryr oss nå om

$$q_n = \Pr(F_n \text{ er større enn } [0,1] \times [0,1]),$$

og skal beregne denne for $n = 3$: vi skal ha

$$X_3 = U_1U_2U_3 = 8\,U_1'U_2'U_3' > 1, \quad Y_3 = V_1V_2V_3 = 8\,V_1'V_2'V_3' > 1.$$

Enkel simulering av en million slike produkter $U_1'U_2'U_3'$ av uavhengige standard uniforme viser at $X_3 > 1$ med hyppighet $p_3 = 0.3443$, så $q_3 = p_3^2 = 0.118$, modulo simuleringsusikkerhet. Vi kan være mer opplyste: skriv $U' = \exp(-T)$, der $T = -\log U'$ får fordelingen $\Pr(T \le t) = \Pr(U \ge e^{-t}) = 1 - e^{-t}$, altså en standard expo. Altså er

$$X_3 = U_1U_2U_3 = 8\exp(-T_1 - T_2 - T_3) = 8\exp(-G_3),$$

der $G_3 = T_1 + T_2 + T_3 \sim \text{gamma}(3,1)$. Den eksakte sannsynligheten blir

$$p_3 = \Pr(\log 8 - G_3 > 0) = \Pr(G_3 < \log 8) = \texttt{pgamma}(\texttt{log}(\texttt{8}), \texttt{3}, \texttt{1}) = 0.34482.$$

Mer generelt får vi altså utmerket kontroll på dette via gamma-fordelinger: $q_n = p_n^2$, der

$$\begin{aligned} p_n &= \Pr(X_n > 1) = \Pr(2^n U_1' \cdots U_n' > 1) \\ &= \Pr(2^n \exp(-T_1 - \cdots - T_n) > 1) = \Pr(n \log 2 > G_n) = \texttt{pgamma}(\texttt{n} * \texttt{log}(\texttt{2}), \texttt{n}, \texttt{1}). \end{aligned}$$

(f) Arealet er $A_n = X_n Y_n$, og

$$\Pr(A_n \geq \varepsilon) \leq \Pr((X_n \geq \sqrt{\varepsilon}) \cup (Y_n \geq \sqrt{\varepsilon})) \leq \Pr(X_n \geq \sqrt{\varepsilon}) + \Pr(Y_n \geq \sqrt{\varepsilon}),$$

så det er tilstrekkelig å vise at X_n, og dermed Y_n med samme fordeling, går mot null i sannsynlighet. Nøkkelen er nå å oversette lange produkter til lange summer, for der har vi LLN og CLT:

$$X_n = 2^n U_1' \cdots U_n', \qquad Z_n = \log X_n = n \log 2 + \sum_{i=1}^n \log U_i' = n(\log 2 - \bar{T}_n),$$

der $\bar{T}_n = (1/n) \sum_{i=1}^n T_i$, gjennomsnittet av mange $T_i = -\log U_i'$. Disse har forventning $\int_0^1 (-\log u)\, \mathrm{d}u = 1$, jfr. det over, så Z_n er n ganger noe som er svært nær $\log 2 - 1 = -0.307$, altså negativt. Vi forstår at Z_n er på full fart mot $-\infty$, som betyr X_n mot null.

(g) Vi lar nå $U_i = bU_i'$ og $V_i = bV_i'$, igjen med U_i' og V_i' fra den standardiserte uniforme over $[0,1]$. Da er

$$X_n = b^n U_1' \cdots U_n' = \exp\Big(n \log b + \sum_{i=1}^n \log U_i'\Big) = \exp\{n(\log b - \bar{T}_n)\},$$

med $\bar{T}_n$ som over, på vei mot 1, ved Store Talls Lov. Så dersom $b > e$ konvergerer $\log b - \bar{T}_n$ mot $\log b - 1$, et tall større enn null, og dermed eksploderer firkantene:

$$\Pr(X_n \geq d, Y_n \geq d) \to 1, \quad \text{for enhver diger skranke } d.$$

Fotnote: finere regning må til for grensetilfellet $b = e$, og det viser seg at også da vil $A_n \to_{\text{pr}} 0$, men langsommere enn når $b < e$.