UNIVERSITETET I OSLO

Det matematisk-naturvitenskapelige fakultet

Eksamen i:	STK1100 — Sannsynlighetsregning og statistisk modellering
Eksamensdag:	Mandag 7. juni 2021
Tid for eksamen:	09.00 – 13.00
Oppgavesettet er på 3 sider.	
Vedlegg:	Ingen
Tillatte hjelpemidler:	Alle hjelpemidler tillatt

Kontroller at oppgavesettet er komplett før du begynner å besvare spørsmålene.

Oppgave 1

På Blindern har det i våres vært utført testing for covid-19 på campus i stor skala. For hver test som tas, registreres det om den er positiv eller negativ. Sannsynligheten for at en tilfeldig student som tester seg på campus, tester positivt, er p. La X være antall tester som utføres til og med man har funnet en positiv test for første gang.

a) Forklar hvorfor X får punktsannsynlighet

$$P(X = k) = (1-p)^{k-1}p, \qquad k = 1, 2, ...$$

b) Vis at den momentgenererende funksjonen til X blir

$$M_X(t) = \frac{e^t p}{1-(1-p)e^t}$$

og spesifiser hvilke t dette uttrykket er gyldig for.

(*Hint:* Du kan få bruk for at $\sum_{m=0}^{\infty} r^m = 1/(1-r)$ hvis $\mid r \mid < 1$.)

c) Utled uttrykk for $\mathrm{E}(X)$ og $\mathrm{V}(X)$ og finn forventet antall tester til og med første positive test hvis $p = 0.001$.

La nå X_1 være antall tester til og med første positive, X_2 antall tester fra første positive til og med andre positive, og videre X_r antall tester fra $(r-1)$-te positive til og med r-te positive. Da er $Y = \sum_{i=1}^{r} X_i$ antall tester til og med r-te positive test.

d) Finn $\mathrm{E}(Y)$ og $\mathrm{V}(Y)$ og beregn forventet antall tester utført til og med den 10. positive, hvis p fremdeles er 0.001.

e) Finn et uttrykk for $P(Y = k)$, $k = r, r+1, ...$

(Fortsettes på side 2.)

f) Testen er ikke nøyaktig. For hurtigtesten benyttet på campus, viser studier at sannsynligheten for å teste positivt, gitt at man er smittet, bare er 0.75. Samtidig vet vi at testen kan gi falske positive. Sannsynligheten for at testen er positiv, selv om man ikke er smittet, er 0.0001. Vi antar fremdeles at sannsynligheten er $p = 0.001$ for at en student tester positivt. Vis at sannsynligheten for at en tilfeldig student som tester seg på campus faktisk er smittet, er 0.0012.

g) Finn sannsynligheten for at en tilfeldig student som tester seg på campus er smittet, gitt at testen er positiv.

Oppgave 2

En variabel X som er Weibull-fordelt med parametere $\alpha, \beta > 0$ har sannsynlighetstetthet

$$f_X(x) = \begin{cases} \frac{\alpha}{\beta^\alpha} x^{\alpha-1} e^{-\left(\frac{x}{\beta}\right)^\alpha}, & x > 0 \\ 0, & \text{ellers} \end{cases}.$$

a) Vis at den kumulative fordelingen til X er gitt ved

$$F_X(x) = \begin{cases} 1 - e^{-\left(\frac{x}{\beta}\right)^\alpha}, & x > 0 \\ 0, & \text{ellers} \end{cases},$$

og at medianen er $\tilde{\mu} = \beta(\ln(2))^{1/\alpha}$

(*Hint:* Du kan få bruk for variabeltransformasjonen $w = (x/\beta)^\alpha$.)

b) Vis at $Y = (X/\beta)^\alpha$ er eksponensialfordelt med parameter $\lambda = 1$, hvilket er det samme som at Y er gammafordelt med formparameter 1 og skalaparameter 1, altså $\text{gamma}(1, 1)$.

Oppgave 3

La $X_1, \dots, X_n$ være levetidene til n elektriske komponenter av samme type. Det er vanlig å anta at X_i-ene er uavhengige og Weibull(α, β)-fordelt, hvilket vi også vil anta her. Du kan dermed få bruk for noen av resultatene fra Oppgave 2 i denne oppgaven. Vi antar også i hele oppgaven at parameteren α er kjent, mens parameteren β er ukjent. En mulig estimator for β er da

$$\hat{\beta} = \left(\frac{1}{n} \sum_{i=1}^{n} X_i^\alpha \right)^{\frac{1}{\alpha}}.$$

a) Vis at en kan skrive $\hat{\beta} = \frac{\beta}{n^{1/\alpha}} \left(\sum_{i=1}^n Y_i\right)^{1/\alpha}$, der $Y_i = (X_i/\beta)^\alpha$. Bruk så det, sammen med resultatet fra Oppgave 2 b) til å vise at $n(\hat{\beta}/\beta)^\alpha \sim \text{gamma}(n, 1)$, altså gammafordelt med formparameter n og skalaparameter 1.

(*Hint:* Husk at fordelingen til en sum av uavhengige gamma(κ_i, ψ)-fordelte variabler med samme skalaparameter ψ er gamma$(\sum_i \kappa_i, \psi)$-fordelt).

(Fortsettes på side 3.)

b) Vis at et 95% konfidensintervall for β er gitt ved

$$\left(\hat{\beta}\left(\frac{n}{b}\right)^{1/\alpha}, \hat{\beta}\left(\frac{n}{a}\right)^{1/\alpha}\right),$$

der a og b er henholdsvis 2.5%- og 97.5%-persentilene i gamma$(n,1)$-fordelingen.

Målte levetider (i år) for 30 tilfeldig valgte komponenter av samme type, kalt $x_1, \ldots, x_{30}$, er gitt ved

```
5.46 1.96 2.45 2.09 0.37 2.57 0.84 5.14 1.54 0.68 0.70 5.85 1.37
0.76 0.81 3.35 0.11 1.91 3.27 1.86 0.87 4.10 2.75 5.07 0.96 1.51
3.83 2.61 2.54 1.68
```

Basert på tidligere studier har en funnet at $\alpha = 1.50$. Da er $\frac{1}{30}\sum_{i=1}^{30} x_i^\alpha = 4.08$.

c) Beregn $\hat{\beta}$, samt et 95% konfidensintervall for β, når det opplyses at 2.5%- og 97.5%-persentilene i gamma$(30,1)$-fordelingen er henholdsvis 20.24 og 41.65.

d) Vis at $\mathrm{E}(\hat{\beta}) = \beta\frac{\Gamma(n+1/\alpha)}{n^{1/\alpha}\Gamma(n)}$. Du kan få bruk for følgende resultat for en gamma(κ,ψ)-fordelt variabel Z, nemlig at $\mathrm{E}(Z^r) = \psi^r\Gamma(\kappa+r)/\Gamma(\kappa)$ for en hvilken som helst r som oppfyller $r > -\kappa$.
Argumentér for at $\hat{\beta}$ bare er forventningsrett for β dersom $\alpha = 1$.

e) Foreslå en estimator $\tilde{\beta}$ som er forventningsrett for β også når $\alpha \neq 1$.

Beregn $\tilde{\beta}$ basert på observasjonene over og sammenlign med $\hat{\beta}$. Det opplyses om at $\Gamma(n+1/\alpha)/\Gamma(n) = 9.62$ når $n = 30$ og $\alpha = 1.50$.

For å sammenligne egenskapene til estimatorene $\hat{\beta}$ og $\tilde{\beta}$ vil en estimere standardfeilene deres, $\sigma_{\hat{\beta}}$ og $\sigma_{\tilde{\beta}}$. Til dette kan en bruke parametrisk bootstrap, og med $B = 10000$ bootstrap-simuleringer fant en estimatene $s_{\hat{\beta}} = 0.310$ og $s_{\tilde{\beta}} = 0.308$.

f) Beskriv hvordan du kan gå fram for å estimere standardfeilene ved hjelp av parametrisk bootstrap. Hva forteller estimatene for standardfeilene deg om forskjellen på de to estimatorene $\hat{\beta}$ og $\tilde{\beta}$?

SLUTT