Øving 11, ST1301

Leveringsfrist: Mandag 14. april, kl 12:15 på e-post til Yngvild. Teller 20% på karakteren. Øvingen besvares individuelt og leveres med studentnummer og uten navn.

Oppgave 1 Anta at et antall arter $i = 1, 2, \ldots, s$ utgjør andeler $p_1, p_2, \ldots, p_s$ av et dyresamfunn. Et mye brukt mål på et slikt samfunns artsmangfold (diversitets) er Simpsons indeks

$$H_S = 1 - \sum_{i=1}^{n} p_i^2. \tag{1}$$

Vis at denne indeksen er lik sannsynligheten for at to tilfeldig valgte individ trukket fra samfunnet ikke tilhører samme art. Dersom vi trekker et utvalg invidid av størrelse N fra samfunnet følger det at antallet av art $1, 2, \ldots, s$, vektoren $(X_1, X_2, \ldots, X_s)$, er multinomisk fordelt.

Anta at det i et samfunn er $s = 10$ arter tilstede og at artene utgjør andelene $0.5, 0.2, 0.1, 0.05, 0.05, 0.05, 0.02, 0.01, 0.01, 0.01$. Hva er da den sanne verdien av H_S? Beregn forventningsfeilen til estimatoren

$$\hat{H}_S = 1 - \sum_{i=1}^{s} \left(\frac{X_i}{N}\right)^2 \tag{2}$$

for utvalgstørrelser $N = 10$, $N = 100$ og $N = 1000$. Du vil få bruk for funksjonen `rmultinom` i R for å simulere fra multinomisk fordeling. Undersøk også forvetningsfeilen til estimatoren

$$\hat{H}'_S = 1 - \sum_{i=1}^{s} \frac{X_i(X_i - 1)}{N(N-1)} \tag{3}$$

for samme verdier av N. Hvilken estimator ser ut til å være å foretrekke? Kan vi estimere H_S uten å kjenne det totale antallet arter s i samfunnet?

Oppgave 2 Anta at vi observerer 100 trær i et plantefelt på 10 ganger 10 meter. Les inn koordinatene til trærne med kommandoen

```
x<-matrix(scan(file="http://www.math.ntnu.no/~jarlet/skog.dat"),ncol=2)
```

Objektet `x` blir en 100×2 matrise som inneholder koordinatene til hvert tre i skogen. Lag et plot av skogen med kommandoen `plot(x)`. En mulig modell er å tenke seg plantefeltet beskrevet som en Poisson-prossess i planet med rate λ. Hva forutsetter denne modellen? Er antakelsene biologisk realistiske?

Dersom plantefeltet er beskrevet ved en Poisson-prossess er antall invidid i kvadratet Poissonfordelt med parameter λA der A er arealet av kvadratet. Hva blir estimatet av λ når plantefeltet er på 10 ganger 10 meter?

Lag en funksjon som simulerer en ny realisasjon av plantefeltet under Poisson-prossess-antakelsen, for en gitt verdi av λ og som returnerer koordinatene i form av en $(N \times 2)$-matrise der N er antall trær. Hint: Simulerer først antall hendelser (hendelser) N i plantefeltet. Gitt dette antallet er da posisjonene uniformt fordelte på kvadratet på 10 ganger 10 meter. Lag et plot av en realisasjon og sammenlign visuelt med den observerte skogen. Ser du noen forskjell mellom det simulerte og observerte mønsteret? Vi ønsker å bruke Poisson-prossess-antakelsen som en null-hypotese og teste denne. Som testobservator T velger vi å bruke minste avstand mellom to nabotrær. Dette blir en funksjon av koordinatene til alle trærne som lett kan beregnes ved hjelp av funksjonene `dist` og `min` i R. Hvilken verdi T^* tar testobservatoren for den observerte plantefeltet?

Simulere mange realiasjoner av plantefeltet under H_0, beregn tilhørende verdier av testobservatoren T, og beregn til slutt testens p-verdi, $P(T > T^*)$. Kommenter resultatet.