TMA4240 Statistikk
2014

Norges teknisk-naturvitenskapelige universitet
Institutt for matematiske fag

Øving nummer 4, blokk I

Oppgave 1

La X være en diskret fordelt stokastisk variable med punktsannsynlighet gitt i følgende tabell

x	-2	-1	0	1	2
$f(x)$	0.1	0.1	0.5	0.2	0.1

Regn ut forventningsverdien til X.

Bestem sannsynlighetene

$$\mathrm{P}(X \geq 0) \quad \text{og} \quad \mathrm{P}(X \geq 0 | X \leq 1).$$

Oppgave 2

Simultanfordelingen, $f(x, y)$, til de to diskrete stokastiske variablene X og Y er gitt i følgende tabell:

	y=0	y=1	y=2
x=-1	$\frac{1}{6}$	$\frac{1}{12}$	$\frac{1}{12}$
x=0	$\frac{1}{12}$	$\frac{1}{6}$	$\frac{1}{12}$
x=1	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{6}$

Finn marginalfordelingen til X og til Y, og beregn forventning og varians til X og til Y.

Beregn kovariansen mellom X og Y, $\mathrm{Cov}(X, Y)$. Er X og Y uavhengige? Svaret skal begrunnes.

Oppgave 3

Det har kommet en effektiv behandling for en dødelig sykdom der en tidligere ikke hadde noen behandling. Det planlegges en masseundersøkelse av alle voksne nordmenn, dvs. 3 mill. individer. En vet at sannsynligheten for at en voksen nordmann har denne sykdommen er 0.003. En ønsker altså å identifisere de individer som har sykdommen med en laboratoriumsprøve.

En vet at for en person med denne sykdommen så vil prøven slå ut med sannsynlighet 0.99. På den andre side så vil prøven slå ut feilaktig på personer uten sykdommen med sannsynlighet 0.005.

Hendelsen at en person har sykdommen benevnes D og hendelsen at prøven slår ut benevnes A.

a) Hvor mange voksne nordmenn ville en forvente har sykdommen?

Hvor mange av disse vil en forvente ikke blir avslørt av prøven?

Benytt Bayes regel og utled svarene i den notasjon som er gitt over før tallsvar regnes ut:

b) Hva er sannsynligheten for at en person som prøven har slått ut for, ikke har sykdommen?

Hva er sannsynligheten for at en person som prøven ikke har slått ut for, har sykdommen?

Gi en kort kommentar til svarene.

For å forbedre undersøkelsen, utføres en tilleggsprøve på de personene hvor den første prøven slo ut. For disse personene som den første prøven har slått ut for, vet en at tilleggsprøven har effektivitet 10, dvs. at tilleggsprøven slår ut med ti ganger større sannsynlighet for de som har sykdommen, enn for de som ikke har sykdommen.

Hendelsen at tilleggsprøven slår ut benevnes B.

Benytt Bayes regel og utled svaret i den notasjon som er gitt over før tallsvar regnes ut:

c) Gitt en person som begge prøvene har slått ut for, hva er sannsynligheten for at personen ikke har sykdommen?

Oppgave 4

Det skal tas blodprøve av samtlige soldater i en militæravdeling for å finne ut om noen er smittet av en bestemt sykdom. Sannsynligheten for at blodprøven fra en tilfeldig valgt soldat er positiv (dvs. at han er smittet) antas å være lik p, og testresultatene for ulike individer antas å være uavhengige.

Man går frem på følgende måte: Blodprøver tas av $k(> 1)$ soldater om gangen, blodprøvene blandes og blandingen analyseres. Hvis den gir positiv reaksjon, noe som betyr at minst en av soldatene er smittet, tas nye blodprøver av de k, og prøvene analyseres enkeltvis. Hvis blandingen gir negativ reaksjon, blir det ikke tatt flere prøver.

a) Begrunn hvorfor antall smittede i en k-gruppe er binomisk fordelt og vis at sannsynligheten for at en blanding av k blodprøver skal gi positiv reaksjon er

$$1 - (1 - p)^k.$$

b) Hvordan defineres betinget sannsynlighet for en hendelse A gitt en annen hendelse B?

Soldat 19 Haugen får beskjed om at hans k-gruppe har vist positiv reaksjon. Hva er da sannsynligheten for at han er positiv?

c) Anta at en avdeling består av m grupper, hver med k soldater, slik at det totale antallet soldater er mk. La X være antall blodprøver som må analyseres før hele avdelingen er ferdig kontrollert. Vis at

$$E(X) = mk\left(1 + \frac{1}{k} - (1-p)^k\right).$$

Hvis $k = 4$, for hvilke verdier av p er den benyttede fremgangsmåten å foretrekke i stedet for å analysere blodprøvene enkeltvis med en gang?

Oppgave 5

Vis at

$$\begin{aligned}\mathrm{Var}\left(\sum_{i=1}^n a_i X_i\right) &= \sum_{i=1}^n \sum_{j=1}^n a_i a_j \mathrm{Cov}(X_i, X_j) \\ &= \sum_{i=1}^n a_i^2 \mathrm{Var}(X_i) + 2\sum_{i=1}^n \sum_{j<i} a_i a_j \mathrm{Cov}(X_i, X_j)\end{aligned}$$

Det er nok at du viser det i det tilfellet at $E(X_i) = 0$ for $i = 1, 2, \ldots, n$. Da er $\mathrm{Var}\left(\sum_{i=1}^n a_i X_i\right) = E\left[\left(\sum_{i=1}^n a_i X_i\right)^2\right]$

Fasit

1. $0.1, 0.8, 0.78$

2. $\frac{1}{6}$

3. **a**) 9000, 90 **b**) 0.6266, $3.02 \cdot 10^{-5}$ **c**) 0.1437

4. **b**) $p/(1-(1-p)^k)$ **c**) $p < 1 - (1/2)^{(1/2)} \approx 0.29$