TMA4245 Statistikk
Vår 2015

Norges teknisk-naturvitenskapelige universitet
Institutt for matematiske fag

Øving nummer 11, blokk II

Oppgave 1

Mange mennesker har i dag en lidenskapelig interesse for elitefotball og (såkalte) eksperter har ofte klare meninger om spillet. I denne oppgaven skal vi konsentrere oss om kamper mellom to spesielle lag, som vi benevner henholdsvis R og L. En ekspertkommentator på fjernsyn kom med følgende påstand om kamper mellom R og L: "som oftest vil det laget som får det første målet også vinne kampen". I denne oppgaven skal vi regne litt med utgangspunkt i denne påstanden.

For en fotballkamp mellom lagene R og L, la følgende hendelser være definert:

R: Lag R vinner kampen.
F: Lag R får mål før lag L.
I: Kampen ender målløs, dvs. 0-0.

a) I dette punktet skal du anta at $\mathrm{P}(R) = 0.4$, $\mathrm{P}(F) = 0.5$, $\mathrm{P}(R \cap F) = 0.3$ og $\mathrm{P}(I) = 0.05$.

Tegn hendelsene R, F og I i et venndiagram.

Bestem sannsynligheten for at lag R vinner gitt at lag R får mål før lag L, dvs. $\mathrm{P}(R \,|\, F)$.

Bestem sannsynligheten for at lag R vinner kampen gitt at kampen ikke ender målløs, dvs. $\mathrm{P}(R \,|\, I')$, hvor I' betegner komplementærhendelsen til I.

Vi skal videre kun analysere de kampene mellom R og L som ikke endte målløse. La p benevne sannsynligheten for at det laget som får det første målet også vinner kampen. Vi forutsetter at denne sannsynligheten ikke avhenger av om det er R eller L som har hjemmekamp. Vi skal estimere p ut fra resultatene i de siste n seriekampene mellom R og L (kun kamper med minst ett mål blir tatt med). La X benevne antall av de n kampene hvor laget som fikk det første målet også vant kampen. Vi antar at X er binomisk fordelt med parametre n og p og bruker estimatoren

$$\widehat{p} = \frac{X}{n}.$$

b) Hva er de generelle forutsetninger for en binomisk fordeling? Er det ut fra dette rimelig å anta at X er binomisk fordelt? (begrunn svaret)

Redegjør kort for det generelle resultatet i sentralgrenseteoremet.

Vis hvordan sentralgrenseteoremet gir at

$$\frac{\widehat{p} - p}{\sqrt{\frac{p(1-p)}{n}}}$$

er tilnærmet standard-normalfordelt, dersom n er stor.

Da ekspertkommentatoren som ble nevnt i begynnelsen av oppgaven ble bedt om å konkretisere sin påstand om at i kamper mellom R og L er det som oftest laget som får det første målet som vinner kampen, sa han at sannsynligheten p er minst lik 0.80. Vi ønsker nå å undersøke om vår observerte verdi for X gir grunnlag for si at ekspertens uttalelse er feil.

c) Formuler dette som et hypotesetestingsproblem. Velg signifikansnivå 5% og bestem en regel for når H_0 skal forkastes.

Hva blir konklusjonen på testen når $n = 24$ og $x = 17$? (Dette er resultater fra kamper mellom Rosenborg og Lillestrøm i perioden 1990-2001. Ingen av disse kampene endte forøvrig målløse.)

d) Anta at forkastningsregelen fra **c)** benyttes, men at p i virkeligheten er 0.7. Hvor mange kampobservasjoner må man da ha for at sannsynligheten for å oppdage at ekspertens uttalelse er feil skal være minst 0.9.

Oppgave 2

I forbindelse med valg og folkeavstemninger er det i dag vanlig at meningsmålingsinstitutter foretar såkalte "exit polls". Dette utføres ved at et tilfeldig utvalg av de som har avgitt stemme, i det de kommer ut fra stemmelokalet, blir spurt om hva de stemte. I denne oppgaven skal vi regne litt på denne situasjonen for et valg mellom kun to kandidater, som vi skal benevne henholdsvis G og B. Vi skal altså se bort fra at noen kan stemme blankt og at stemmer kan bli forkastet. Vi skal også se bort fra muligheten av at noen ikke vil svare hva de har stemt, eller at noen svarer usant.

La N betegne antall personer som avgir stemme og la p betegne andelen av disse som stemte på G. La videre n betegne antall personer som ble spurt av meningsmålingsinstituttet om sin stemmegivning og la X betegne antall av disse n som stemte på G.

a) Hvilke(n) betingelse(r) må være oppfylt for at X skal være tilnærmet binomisk fordelt ?

Gitt at denne betingelsen er oppfylt, bestem følgende sannsynligheter for $n = 20$ og $p = 0.50$:

$$\mathrm{P}(X = 9) \quad , \quad \mathrm{P}(X > 9) \quad \text{og} \quad \mathrm{P}(X > 9 | X \leq 12)$$

I resten av oppgaven skal vi forutsette at X er (tilnærmet) binomisk fordelt.

b) Utled sannsynlighetsmaksimeringsestimatoren (SME) for p.

Bestem estimatorens forventning og varians.

Som kjent kan en binomisk fordeling tilnærmes med en normalfordeling dersom n er stor nok. I resten av oppgaven skal vi forutsette at dette er tilfelle slik at vi har at

$$\frac{X - np}{\sqrt{np(1-p)}}$$

er (tilnærmet) standard-normalfordelt.

Anta at meningsmålingen utføres på oppdrag fra et fjernsynsselskap. Dersom et tilstrekkelig stort flertall av de spurte stemte på en av kandidatene, vil fjernesynsselskapet gå ut og erklære denne kandidaten som vinner av valget.

c) Formuler dette som en hypotesetest. Spesifiser nullhypotese og alternativ hypotese og bestem forkastningskriteriet når signifikansnivået er α.

Hva blir konklusjonen på hypotesetesten dersom 5 000 ble spurt, 2562 av disse hadde stemt på G, og en benytter signifikansnivå $\alpha = 0.10$?

Oppgave 3

Per er nylig ferdig med sine studier i Trondheim og på sin første arbeidsdag får han en interessant oppgave på laboratoriet. Når to væsker A og B blandes sammen, vil volumet av en ideell blanding av x deler A og $1-x$ deler B være lik volumet av x deler A pluss volumet av $(1-x)$ deler B. For en ikke-ideell blanding vil volumet avvike noe fra det ideelle tilfellet.

Per vil eksperimentelt bestemme avviket fra en ideell blanding når A og B blandes. La $x_1, x_2, \ldots, x_n$ være andeler av A som blandes med $1-x_1, 1-x_2, \ldots, 1-x_n$ andeler B, og la $Y_1, Y_2, \ldots, Y_n$ være avviket i volum som måles fra en ideell blanding. Per antar at avviket er på formen

$$Y_i = a x_i (1 - x_i) + \epsilon_i, \qquad i = 1, \ldots, n,$$

der $\epsilon_1, \ldots, \epsilon_n$ er uavhengige normalfordelte støyledd med forventningsverdi lik null og *kjent* varians $\sigma_0^2 = 0.025^2$ for alle i, slik at Y_i er normalfordelt med

$$\mathrm{E}(Y_i|x_i) = a x_i (1 - x_i) \quad \text{og} \quad \mathrm{Var}(Y_i|x_i) = \sigma_0^2, \qquad i = 1, \ldots, n$$

For en ideell blanding vil a være lik null.

Pers målinger gir følgende resultat:

i	1	2	3	4	5	6	7	8	9
x_i	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
y_i	$-.0561$	$-.0369$	$-.0651$	$-.0578$	$-.0941$	$-.0684$	$-.0546$	$-.0247$	$-.0072$

Det oppgis at $\sum_{i=1}^{9} x_i(1-x_i)y_i = -0.0945$ og $\sum_{i=1}^{9}(x_i(1-x_i))^2 = 0.3333$.

a) Vis at sannsynlighetsmaksimeringsestimatoren (SME) for a, $\hat{a}$, er

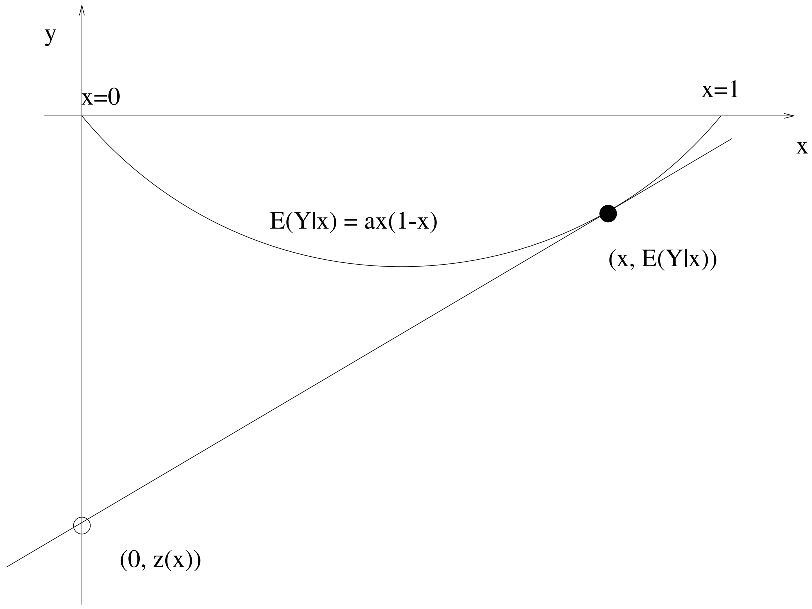
$$\hat{a} = \frac{\sum_{i=1}^{n} x_i(1-x_i)Y_i}{\sum_{i=1}^{n} \left(x_i(1-x_i)\right)^2}.$$

b) Finn forventningsverdi og varians til $\hat{a}$.
Hvilken fordeling har $\hat{a}$? (Begrunn svaret.)

Per vil undersøke om det er grunn til å påstå at han har å gjøre med en ikke-ideell blanding.

c) Formuler dette som et hypotesetestingsproblem og lag en test for dette formål med signifikansnivå α. Hva blir konklusjonen på hypotesetesten for dataene over og $\alpha = 5\%$?

En interessant størrelse for Pers sjef er det partielle molarvolumet for væske A. Denne størrelsen kan avledes fra $z(x)$; tangentlinja i punktet $(x, \mathrm{E}(Y|x))$ skjærer y-aksen i $(0, z(x))$, se figuren under.



d) Finn $z(x)$ uttrykt ved hjelp av a og angi deretter en rimelig estimator $\hat{z}(x)$ for $z(x)$. Finn et 95% konfidensintervall for $z(x)$. Beregn så et 95% konfidensintervall for $z(2/3)$ med data som gitt over.

Fasit

1. **a**) $\mathrm{P}(R|F) = 0.6$, $\mathrm{P}(R|I') = 0.421$ **c**) Forkaster ikke H_0 **d**) 156

2. **a**) 0.1602, 0.588, 0.525 **b**) $\frac{X}{n}$, θ, $\frac{\theta(1-\theta)}{n}$ **c**) Forkaster H_0

3. **c**) Forkast H_0 **d**) $z(x) = ax^2$, $[-0.164, -0.088]$