Sensorrettleiing eksamen ECON2130

Vår 2024

Oppgåve 1

a)

$$P(5\,spar) = \frac{\binom{13}{5}}{\binom{52}{5}} \approx 0.0005$$

I R:

I oppgåve a) til c) brukar vi framgangsmåten: 1. trekk 5 kort mage gonger, 2. finn kor stor del av gongane vi får 'suksess'. Det viktigaste er at ein klarer å implementere ein algoritme som simulerer sannsynet, men for full uttelling bør det vere noko grad av presisjon. Altså kan ikkje talet trekk vere veldig lågt. Vi definerer ein kortstokk:

$kortstokk <- c(rep(1,13), rep(2,13), rep(3,13), rep(4,13))$

Vi trekker fem kort 10^5 gonger og finn andelen gonger alle fem korta er lik spar:

$b <- replicate(1e5, sum(sample(kortstokk,5)) == 20)$

$mean(b)$

b)

$$P(5\,like) = 4 * \frac{\binom{13}{5}}{\binom{52}{5}} \approx 0.0020$$

I R:

Vi trekker fem kort 10^5 gonger og finn andelen gonger alle fem korta er like:

$c <- replicate(1e5, length(unique(sample(kortstokk,5))) == 1)$

$mean(c)$

c)

$$P(3\,spar\,og\,2\,ruter) = \frac{\binom{13}{3}\binom{13}{2}}{\binom{52}{5}} \approx 0.0086$$

I R:

Vi definerer ein funksjon som trekk 5 kort og gir verdien 1 om vi trekk 3 spar og 2 ruter:

$d\_funksjon <- function()\{$

$trekk <- sample(kortstokk,5)$

$suksess <- length(trekk[trekk == 4]) == 3\ \&\ length(trekk[trekk == 3]) == 2$

$return(suksess)$

$\}$

Vi trekker fem kort 10^5 gonger og finn andelen gonger vi har 3 spar og 2 ruter:

$d <- replicate(1e5, d\_funksjon())$

$mean(d)$

Om du gjer oppgåve 1 med simulering vil alle ekvivalente metodar som gjev omtrent same svar gi full uttelling.

Oppgåve 2

a) X er binomisk fordelt. X er summen av n uavhengige Bernoulli-forsøk med konstant sannsyn p.

b) Sidan X er binomisk fordelt er $EX = np$. Vi brukar rekneреglar for forventing og set inn for EX:

$$E[\pi] = E[bX - kn^2] = bE[X] - kn^2 = bpn - kn^2$$

Sidan $E[\pi]$ er ein konkav funksjon av n, finn vi maksimum ved å derivere og sette lim 0:

$$bp - 2kn^* = 0 \Longrightarrow n^* = p * \frac{b}{2k}$$

c)

$$E[\hat{n^*}] = E[\frac{X}{n} * \frac{b}{2k}] = \frac{b}{2kn} * E[X] = \frac{b}{2kn} * np = p * \frac{b}{2k} = n^*$$

$\hat{n^*}$ er forventningsrett.

Sidan X er binomisk fordelt er $Var(X) = np(1-p)$. Vi brukar reknereglar for varians og set inn for $Var(X)$:

$$Var[\hat{n^*}] = Var[\frac{X}{n} * \frac{b}{2k}] = (\frac{b}{2kn})^2 * Var[X] = (\frac{b}{2kn})^2 * np(1-p) = (\frac{b}{2k})^2 * \frac{p(1-p)}{n}$$

Estimatet er lik:

$$\hat{n^*} = \frac{60}{500}\frac{1500}{2 * 0.1} = 900$$

d) Sidan $\hat{n^*}$ er gjennomsnittet av n stokastiske variablar ganga med ein kosntant veit vi gjennom sentralgrenseteoremet at $\hat{n^*}$ blir tilnærma normalfordelt når n blir stor.

For våre observasjonar er

$$E(\hat{n^*}) = 900$$

Sidan vi ikkje veit den faktiske verdien av p må vi bytte ut p i variansen med estimatet $\hat{p} = X/n = 0.12$. Variansen blir då:

$$Var(\hat{n^*}) = (\frac{1500}{2 * 0.1})^2 * \frac{0.12(1-0.12)}{500} = 11880$$

og

$$SE(\hat{n^*}) = \sqrt{Var(\hat{n^*})} = 109$$

Vi brukar formelen for konfidensintervall:

$$KI = [\hat{n^*} - z_{\alpha/2}SE(\hat{n^*}), \hat{n^*} + z_{\alpha/2}SE(\hat{n^*})]$$

Vi finn z_α ved å slå opp i statistikktabell eller ved å bruke $'qnorm(0.995)$ og finn ein z-verdi lik 2.58. Ved å sette inn verdiane i formelen for konfidensintervall får vi:

$$KI = [900 - 2.57 * 109, 900 + 2.57 * 109]$$

$$KI = [620, 1180]$$

e) Vi vil teste om n^* er mindre eller lik 1000, eller om n^* er større enn 1000. Naturlege hypoteser er derfor:

$$H_0 : n^* \geq 1000,\ H_1 : n^* < 1000$$

Om du i staden set opp ein tosidig hypotesetest vil du få noko, men ikkje full uttelling.

Vi brukar formelen for standardfeil frå oppgåve c) for å finne standardfeilen under nullhypotesa:

$$SE(\hat{n_0^*}) = \sqrt{(\frac{1500}{2*0.1})^2 * \frac{0.13(1-0.13)}{500}} = 113$$

Vi konstruerer teststatistikken:

$$z = \frac{900 - n_0^*}{SE(\hat{n_0^*})} = \frac{900 - 1000}{113} = -0.88$$

Her skal du helst velje signifikansnivå på førehand. Det høgste vanlege konfidensnivået er 10%, og då er kritisk z-verdi lik -1.28, og vi forkastar ikkje nullhypotesa. Vi vil då heller ikkje forkaste nullhypotesa for lågare signifikansnivå.

f) For å finne p-verdien brukar vi formelen for p-verdiar ved ei einsidig hypotese:

$$pverdi = \Phi(z) = \Phi(-0.88) = 0.17$$

P-verdien er dermed lik 0.17. Vi finn verdien av $\Phi(z)$ ved å enten slå opp i statistikktabell eller ved å bruke 'pnorm(z)' i R.

Oppgåve 3

a) Vi brukar føresetnadane for at ein variabel skal vere Poisson-fordelt på vårt eksempel:

- Kor mange fisk du får innan eit gitt tidsintervall er uavhengig av kor mange fisk du får i andre intervall
- Det forventa talet fisk du får på kvar fisketur på fire timar er lik λ
- Du kan ikkje få to fisk nøyaktig samtidig

For full uttelling må alle dei tre føresetnadane vere nemnd.

b) $\hat{\lambda}$ er gjennomsnittet av $n = 50$ stokastiske variabalar og vi kan bruke sentralgrenseteoremet som gir at $\hat{\lambda}$ er normalfordelt med $E(\hat{\lambda})$ og $Var(\hat{\lambda})$. Sidan $\hat{\lambda}$ er gjennomsnittet av $n = 50$ uavhengige variablar med forventning $\lambda = 4$ og varians $\lambda = 4$ er $E(\hat{\lambda}) = \lambda = 4$ og $Var(\hat{\lambda}) = \lambda/n = 4/50 = 0.08$.

Vi kan derfor bruke tilnærminga $\hat{\lambda} \sim \mathcal{N}(4, 0.08)$

c)

$$P(\hat{\lambda} \leq 3.5) = P(\frac{\hat{\lambda} - E(\hat{\lambda})}{\sqrt{Var(\hat{\lambda})}} \leq \frac{3.5 - 4}{\sqrt{0.08}}) = P(Z \leq -1.77) = 1 - P(Z \leq 1.77)$$

Vi slår opp i statistikktabell eller brukar 'pnorm(1.76)' i R og finn at:

$$P(\hat{\lambda} \leq 3.5) = 1 - 0.96 = 0.04$$

Ein kan også gjere heile utrekninga i R ved å bruke 'pnorm(3.5,4,sqrt(0.08))'.

For å finne det eksakte sannsynet utnyttar vi at $X \sim Poisson(4 * 50)$. Vi finn derfor sannsynet:

$$P(\hat{\lambda} \leq 3.5) = P(X/n \leq 3.5) = P(X \leq 3.5 * n) = P(X \leq 175)$$

Dette sannsynet finn vi i R ved å bruke 'ppois(175,200)' og vi får:

$$P(\hat{\lambda} \leq 3.5) = 0.04$$

Vi ser at normalfordelinga er ei god tilnærming då resultata er identiske med to desimalar.

d) Ein type I-feil er å forkaste ei riktig nullhypotese. Signifikansnivået α er sannsynet for å gjere ein type I-feil ved ein hypotesetest.

e) Ein type II-feil er å ikkje forkaste ei gal nullhypotese. Viss ein reduserer signifikansnivået α aukar ein samtidig krava for å forkaste nullhypotesa, og ein aukar dermed sannsynet for type II-feil.

f) Det er fleire måtar å simulere teststyrken i R. Nedanfor er to av metodane som vart gjennomgått på forelesing vist. Begge metodane gir svar på omtrent 0.18. Dette vil sei at sannsynet for å gjere type II-feil ved denne hypotesetesten er omtrennt 0.18. Alle metodar som gjer det same som det er spurt om i oppgåva gir full uttelling. Viss du i staden for å trekke verdiar av X trekk verdiar av normalfordelt $\hat{\lambda}$, vil du får noko, men ikkje full uttelling.

Med for-loop:

$forkast < -vector()$

$for(i\, in\, 1 : 1000)\{$

$x < -rpois(50, 4)$

$test < -t.test(x, mu = 3.5)$

$forkast[i] < -test\$p.value <= 0.01$

$\}$

$mean(forkast)$

Med ein kombinasjon av ’function’ og ’replicate’:

$forkast < -function()\{$

$x < -rpois(50, 4)$

$test < -t.test(x, mu = 3.5)$

$return(test\$p.value <= 0.01)$

$\}$

$sim < -replicate(1000, forkast())$

$mean(sim)$