TMA4240 Statistikk
Høst 2018

Norges teknisk-naturvitenskapelige universitet
Institutt for matematiske fag

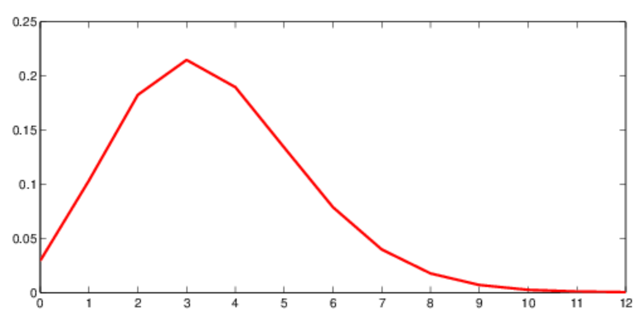
Anbefalt øving 8

Denne anbefalte øvingen er basert på stoffet som gjennomgås i den åttende uka med forelesninger. Det vil si kap. 8 og første del av kap. 9 i læreboka (Walpole og andre). Oppgavene handler om utvalgsfordelinger, sentralmål og sentralgrenseteoremet.

Oppgave 1

En lottorekke består av 7 tall krysset av blant tallene $1, 2, ..., 34$. Mer presist kan den oppfattes som et ikke-ordnet utvalg på 7 elementer blant tallene 1 til 34, der utvelgelsen skjer uten tilbakelegging. Hver lørdag trekker Norsk tipping ut ukens riktige lottorekke ved tilfeldig trekking slik at alle mulige lottorekker blir like sannsynlige. Ukens toppgevinst utbetales til innehaverne av innleverte rekker som er identiske med den riktige lottorekka (dvs. oppnå 7 riktige tall). Dersom ingen tippet den riktige rekka, blir premiepotten for 7 riktige overført til neste ukes lotto-omgang (som blir en "Gull-Lotto" omgang).

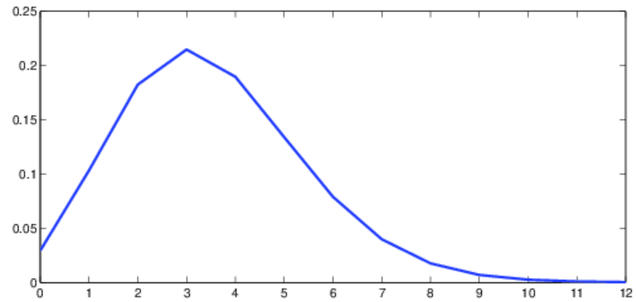
a) Hvor mange forskjellige lottorekker finnes det? Hva er sannsynligheten for at tallet 34 er med i den riktige lottorekka? Gjør rede for at enhver lottorekke vil oppnå 7 riktige ved lottotrekningen med sannsynlighet $p = 1.859 \cdot 10^{-7}$



Figur 1: Poissontetthetsfunksjonen, $f_X(x) = \frac{\mu^x}{x!}e^{-\mu}$ for heltallsverdier av x fra 0 til 12 med $\mu = 3.53$.

Anta at det en bestemt uke leveres inn tilsammen $n = 19\,000\,000$ rekker i lotto. La X være antall av disse rekkene som oppnår 7 riktige i ukens trekning. Det antas at X er binomisk fordelt med parametre n og p (med verdier som gitt ovenfor).

b) Hvorfor vil X med god tilnærmelse kunne regnes å være poissonfordelt? Hva blir parameteren i denne poissonfordelingen?



Figur 2: Binomisk tetthetsfunksjon, $f_X(x) = \binom{n}{x} p^x (1-p)^{n-x}$ for heltallsverdier av x fra 0 til 12 med $n = 19\,000\,000$ og $p = 1.859 \cdot 10^{-7}$.

I Figur 1 har vi plottet Poissontetthetsfunksjonen med denne parameteren, og i Figur 2 har vi plottet den binomiske tetthetsfunksjonen med parametere som oppgitt ovenfor. Bruk figurene til å kommentere Poisson-tilnærmingen.

c) Finn sannsynligheten for at ingen av de innleverte rekkene oppnår 7 riktige ved ukens trekning. Anta i det følgende at antall innleverte rekker holder seg konstant på 19 000 000 pr. uke. Hva blir da det forventede antall "Gull-Lotto" omganger pr. år? Hva blir sannsynligheten for at det i løpet av ett år ikke oppnås noen "Gull-Lotto"-omgang?

Norsk Tipping ønsker å øke hyppigheten av "Gull-Lotto"-omganger og vurderer derfor å øke antall valgbare tall fra 34 til m (>34), mens en rekke fremdeles skal bestå av 7 tall.

d) Hvor stor må m velges for at det med sannsynlighet minst 0.10 ikke finnes rekker med 7 riktige i en uke der det innleveres $n = 19\,000\,000$ rekker?

Oppgave 2

Medianen til et datasett, $\tilde{X}$, er den midterste verdien. Hvis vi har stokastiske (tilfeldige) variabler $X_1, X_2, ..., X_n$ og ordner dem etter størrelse slik at $X_{(1)} < X_{(2)} < ... < X_{(n)}$, så er medianen definert som

$$\tilde{X} = \begin{cases} X_{\left(\frac{n+1}{2}\right)} & \text{hvis } n \text{ er et oddetall,} \\ \frac{1}{2}\left(X_{\left(\frac{n}{2}\right)} + X_{\left(\frac{n}{2}+1\right)}\right) & \text{hvis } n \text{ er et partall.} \end{cases}$$

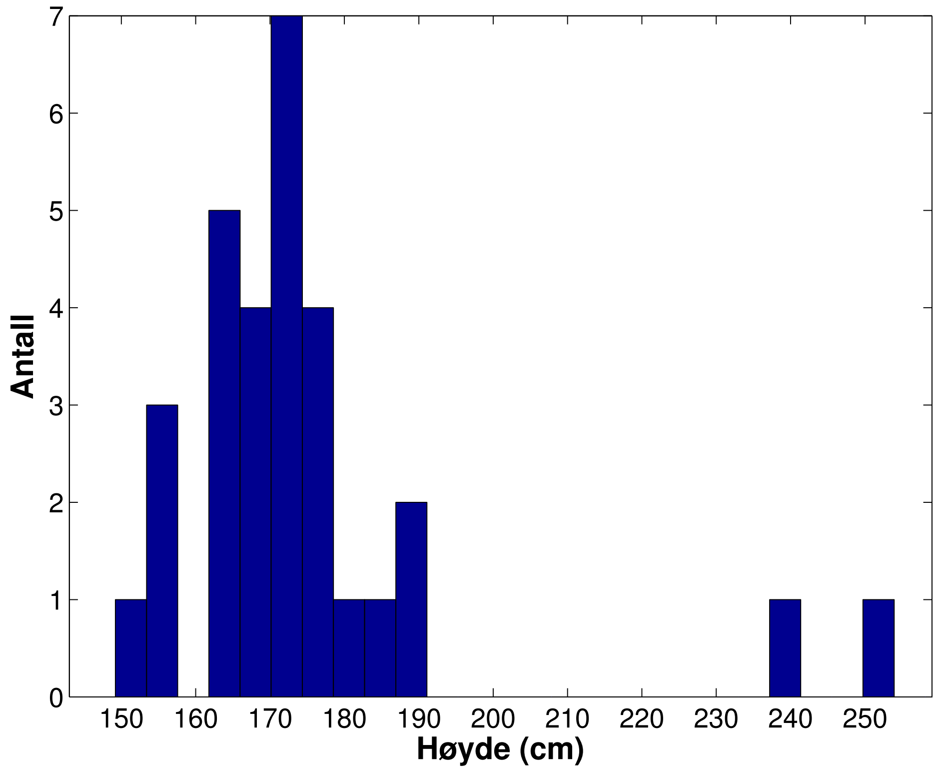
Når de stokastiske variablene våre er uavhengige og normalfordelte med forventningsverdi μ og varians σ^2, altså $X_i \sim N(\mu, \sigma^2)$, og vi har at antallet variabler, n, er stort, kan vi anta at variansen til medianen er

$$\mathrm{Var}(\tilde{X}) = \frac{1}{4n\big(f(\mu)\big)^2},$$

der $f(x)$ er sannsynlighetstettheten til normalfordelingen.

a) For dette tilfellet, vis at

$$\mathrm{Var}(\tilde{X}) = \frac{\pi}{2}\mathrm{Var}(\bar{X}),$$



Figur 3: Høydene til 30 rekrutter, kanskje fra 1814.

der gjennomsnittet $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$.

$\tilde{X}$ er en forventningsrett estimator for forventningsverdien μ. Hvorfor foretrekker vi vanligvis $\bar{X}$ framfor $\tilde{X}$ som estimator for μ?

Statistisk sentralbyrå har data for høydene til mannlige norske rekrutter til hæren hvert år tilbake til 1878. I denne oppgaven kan du anta at du vet med sikkerhet at høydene til rekruttene i et hvilket som helst år er normalfordelte.

På Terningmoen leir har løytnant Munthe funnet et skjema med høydene på 30 rekrutter som han mener må være fra 1814. Papiret er gulnet og blekket har falmet en del, men løytnanten får en av sine nåværende rekrutter til å skrive dataene inn i et regneark etter beste evne. Figur 3 viser et histogram av disse dataene.

b) For dette datasettet, vil medianen $\tilde{X}$ være større enn, mindre enn eller omtrent like stor som gjennomsnittet $\bar{X}$?

Ville du ha brukt medianen eller gjennomsnittet til å estimere forventningsverdien μ her? Begrunn svaret.

Oppgave 3

En bestemt målemetode for bestemmelse av pH-verdien i en løsning gir måleresultater som antas å være uavhengige og normalfordelte, med forventningsverdi μ lik virkelig pH og varians $\sigma^2 = 0.060^2$. La $X_1, \ldots, X_n$ være uavhengige målinger av pH i en bestemt løsning.

a) Anta (bare i dette punktet) at den virkelige pH-verdien i en løsning er 6.8.

Hva er sannsynligheten for at en bestemt måling gir et resultat som er under 6.74?

Hva er sannsynligheten for at en bestemt måling gir et resultat som er mellom 6.74 og 6.86?

Hva er sannsynligheten for at en bestemt måling, X, gir et resultat som avviker mer enn 0.06 fra μ, dvs bestem $P(|X - \mu| > 0.06)$?

Du skal estimere pH i en løsning, og bruker gjennomsnittet av 5 uavhengige målinger som estimat. La Y være gjennomsnittet av 5 uavhengige målinger.

b) Hva er sannsynligheten for at Y avviker mer enn 0.06 fra μ?

Fasit

1. **a**) 5379616, 7/34 **b**) 3.53 **c**) 0.029, 1.5, 0.22 **d**) 36

2. **b**) medianen er mindre enn gjennomsnittet

3. **a**) $0.159, 0.682, 0.318$ **b**) 0.026