TMA4245 Statistikk
Eksamen 9. desember 2013

Norges teknisk-naturvitenskapelige universitet
Institutt for matematiske fag

Oppgave 1

I kortspillet Blackjack får ein den høgaste gevinsten derssom dei to første korta ein får er eit ess saman med anten ein tiar, ein knekt, ei dame eller ein konge. Dette blir kalla «å få ein blackjack».

I kasino der ein spelar Blackjack for pengar, blandar ein gjerne 8 kortstokkar som ein trekkjer desse korta frå. (Dette gjer det vanskelegare å Â«telje kortÂ», det vil seie å rekne seg fram til kor mange ess og billetkort som er igjen i stokken.) Det er då så mange kort i stokken at vi kan sjå på dette som ein situasjon Â«med tilbakeleggingÂ», altså at ein trekkjer eitt og eitt kort frå ein vanleg kortstokk og legg forrige kort tilbake før ein trekkar eit nytt. Ein vanleg kortstokk har 52 kort, med 4 fargar som har kvar sin tiar, knekt, dame, konge og ess.

a) Kva er sannsynet for å få ein blackjack?

Kva er sannsynet for å få ein blackjack viss det første kortet du har fått er eit ess?

La oss seie at ein spelar som teljer kort har funne ut at sannsynet for å få blackjack no er 0.06, at sannsynet for å få eit ess som første kort er 0.1 og at sannsynet for at det første kortet ditt var eit ess dersom du har fått ein blackjack er 0.4. Bruk desse tala til å rekne ut kva sannsynet no er for å få ein blackjack dersom det første kortet du har fått er eit ess.

Lars er på ferie i Las Vegas, og går inn på det største kasinoet han finn. Han set seg ved Blackjack-bordet og sett seg føre å spele fram til han vinn og å doble innsatsen for kvart spel. Han satsar éin dollar i fyrste spel, to dollar i andre spel, og så vidare fram til han vinn.

Gå for enkelthets skuld ut frå at Lars alltid får igjen det dobbelte av det han satsa dersom han vinn, at sannsynet for at han vinn er 0.3 i kvart spel, og at Lars sluttar å spele etter å ha vunne éin gong.

b) La X vere talet på gonger Lars spelar før han gjev seg. Kva er sannsynsfordelinga til X?

Gå ut frå i resten av oppgåva at Lars går tom for pengar og dermed sluttar å spele viss han ikkje har vunne etter å ha spela fem gonger. La W vere talet på dollar Lars vinn i kasinoet (uavhengig av kor mykje han har satsa først). Kva er forventningsverdien til W?

La Y vere gevinsten Lars sit igjen med etter at innsatsen er trekt frå. Kva er forventningsverdien til Y?

Oppgave 2

Agent John Bang går regelmessig til skytetrening. Erfaring seier at sanssynet for eit treff er $p = 0.8$. I ein treningssesjon skyt han 20 skudd. Anta at skudda er uavhengige og at kvart enkeltskot er enten ein treff eller bom.

a) Kva er forventa antal treff?

Kva er sannsynet for at han treff flere enn forventa?

Kva er det betinga sannsynet for at han treff 20 skot når vi veit at han treff fleire enn forventa?

Sjefen bestemmer at John skal ha ein ny pistol. Dei håper at denne skal gje betre treffsannsyn. Dei ønskjer å undersøkje om dette holder, og John bruker den nye pistolen i ein vanlig treningssesjon med 20 skot.

b) Formuler problemet som ein hypotesetest.

Bruk den vanlige normalapproksimasjonen til å gjennomføre hypotesetesten på signifikansnivå $\alpha = 0.1$ når observert antal treff er 18.

c) Forklar korleis ein eksakt test vert ved bruk av den binomiske fordelinga.

Kva vert P-verdien til den eksakte testen når han treff på 18 skot?

Anta signifikansnivå $\alpha = 0.1$ for testen. Rekn ut teststyrken for den eksakte testen når sant treffsannsyn er $p = 0.9$.

Oppgave 3

Medianen til eit datasett, $\tilde{X}$, er den midterste verdien. Dersom vi har stokastiske (tilfeldige) variablar $X_1, X_2, ..., X_n$ og ordnar dei i stigande rekkjefølje 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 dei stokastiske variablane våre er uavhengige og normalfordelte med forventningsverdi μ og varians σ^2, altså $X_i \sim N(\mu, \sigma^2)$, og vi har at talet på variablar, n, er stort, kan vi gå ut frå at variansen til medianen er

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

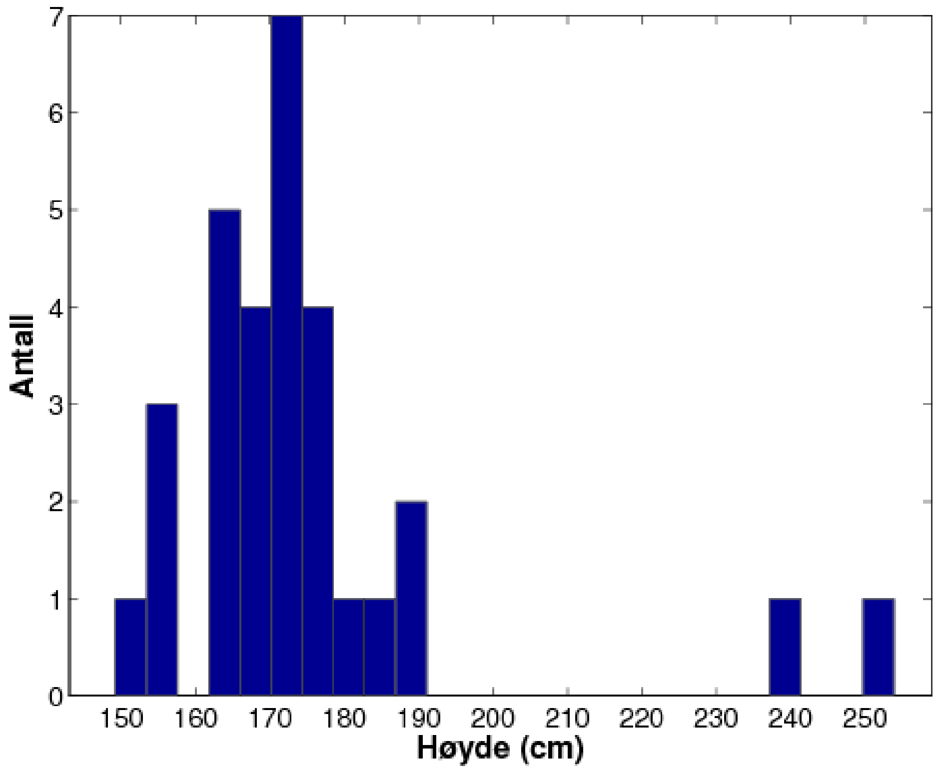
der $f(x)$ er sannsynstettleiken til normalfordelinga.

a) For dette tilfellet, vis at

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

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

$\tilde{X}$ er ein forventningsrett estimator for forventningsverdien μ. Kvifor føretrekkjer vi vanlegvis $\bar{X}$ framfor $\tilde{X}$ som estimator for μ?



Figur 1: Høgdene til 30 rekruttar, kanskje frå 1814.

Statistisk sentralbyrå har data for høgdene til mannlege norske rekruttar til hæren kvart år tilbake til 1878. I denne oppgåva kan du gå ut frå at høgdene til rekruttane i kva år som helst er normalfordelte.

På Terningmoen leir har løytnant Munthe funne eit skjema med høgdene på 30 rekruttar som han meiner må vere frå 1814. Papiret er gulna og blekket har falma ein del, men løytnanten får ein av dei noverande rekruttane sine til å skrive dataa inn i eit rekneark etter beste evne. Figur 1 viser eit histogram av desse data.

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

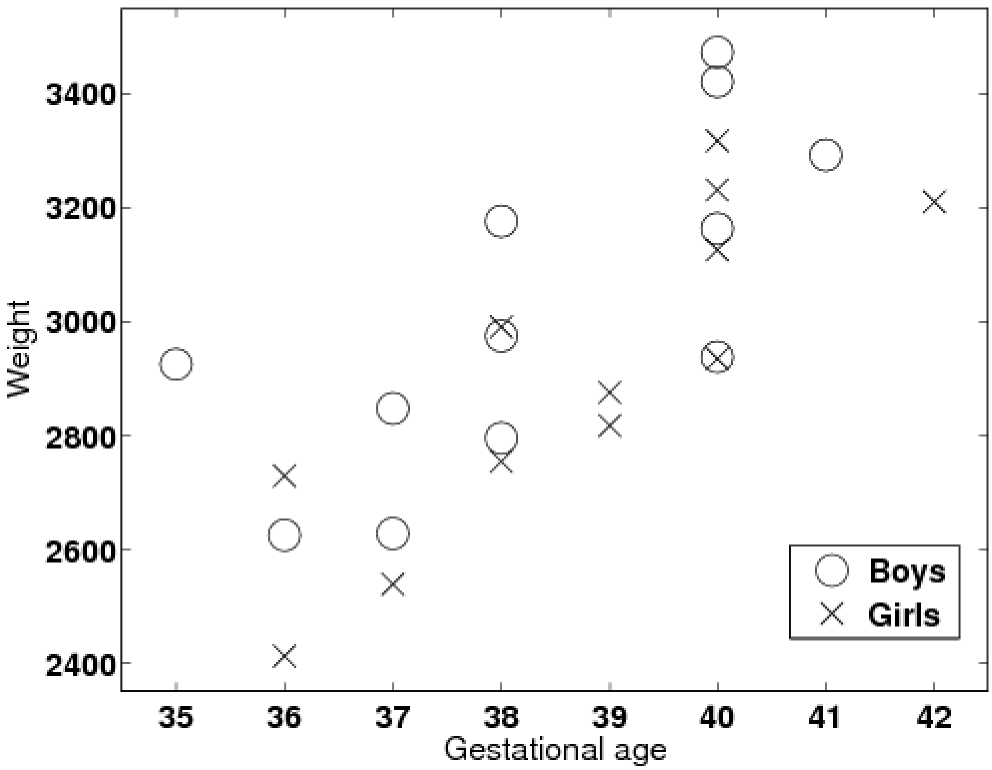
Ville du ha nytta medianen eller gjennomsnittet til å estimere forventningsverdien μ her? Grunngje svaret.

Oppgave 4

I medisin er det nyttig å studere vekta til nyfødde som ein funksjon av deira terminalder (*gestational age* eller tid sidan unnfanging). Data er her terminalder x_i (veker) og vekt y_i (gram) for $i = 1, \ldots, n$ babyer, og $n = 24$.

For dette datasettet har vi $\sum_{i=1}^{n} x_i y_i = 2\ 752\ 667$, $\sum_{i=1}^{n} x_i^2 = 35\ 727$, $\sum_{i=1}^{n} x_i = 925$ og $\sum_{i=1}^{n} y_i = 71\ 194$.

Anta ein lineær regresjonsmodell: $Y_i = \beta_0 + \beta_1 x_i + \epsilon_i$, $i = 1, \ldots, n$, der $\epsilon_1, \ldots, \epsilon_n$ blir antatt



Figur 2: Kryssplot av terminveke og fødselsvekt for 12 gutar og 12 jenter.

som uavhengige og normalfordelte med forventning 0 og varians σ^2.

a) Bruk oppsummeringa av talmateriale over til å rekne ut estimata for skjæringspunkt og stigningstalet for regresjonsmodellen: $\hat{\beta}_0$ og $\hat{\beta}_1$.

Vi reknar ut eit estimat for σ^2 ved $s^2 = \frac{1}{n-2}\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 = 194^2$.

Rekn ut eit 95 prosent konfidensinterval for stigningstalet.

b) Bruk data til å finne eit 90 prosent prediksjonsinterval for vekta til ein nyfødd i terminveke 40.

Kva er bredda på 90 prosent prediksjonsintervalet for terminveke 42 i forhold til det vi fann for veke 40?

Figur 2 viser eit kryssplott av terminveke og vekt. I dette plottet vert data delt inn i to grupper: gutar og jenter. Det er $n_b = 12$ gutar (nummerert 1 til n_b) og 12 jenter (nummerert $i = n_b + 1$ til n).

Vi foreslår følgjande modell for data

$$Y_i = \beta_b + \beta_1 x_i + \epsilon_i, \quad i = 1, \ldots, n_b.$$

$$Y_i = \beta_g + \beta_1 x_i + \epsilon_i, \quad i = n_b + 1, \ldots, n.$$

der vi fortsatt antek at $\epsilon_1, \ldots, \epsilon_n$ er uavhengige og normalfordelte med forventning 0 og varians σ^2.

c) Bruk plottet til å forklare kvifor denne modellen kan vere nyttig. Forklar vidare om nokre element av modellen kan vere uønska.

Rekn ut minste kvadratsums estimat (eller maximum likelihood estimat) for parametra i modellen, her benevnt ved $\hat{\beta}_b$, $\hat{\beta}_g$ og $\hat{\beta}_1$. I tillegg til summane tidligare i oppgåva har vi no at $\sum_{i=1}^{n_b} y_i = 36\ 258$, $\sum_{i=1}^{n_b} x_i = 460$, $\sum_{i=n_b+1}^{n} y_i = 34\ 936$ og $\sum_{i=n_b+1}^{n} x_i = 465$.

Fasit

1. **a**) 0.04734, 0.3077, 0.24 **b**) 6.567, -4.4

2. **a**) 16, 0.41, 0.028 **b**) ikkje forkast H_0 **c**) 0.21, 0.39

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

4. **a**) -1465, 115, [69,161] **b**) [2789,3481], 690, 732 **c**) -1587, -1747, 120.22