Obligatorisk innlevering 3 i STAT110

- Basert på kapittel 1-7 i boken.
- Alle svar skal begrunnes, og det skal være med så mye mellomregning at fremgangsmåten fremgår tydelig av besvarelsen.
- Noen oppgaver er basert på en tidligere eksamensoppgave. Hvis du trenger hjelp kan du kikke på disse, men merk at eksamensoppgavene **er endret**.

1. **Oljeleting** (Modifisert eksamen H2016, oppg. 2)

   Først en liten oppvarming. La E og F være to hendelser i et utfallsrom. Det oppgis at $P(E) = 0.4$, $P(F) = 0.5$ og $P(E \cup F) = 0.8$.

   (a) Er hendelsene E og F disjunkte? Er hendelsene E og F uavhengige?

   Så til oljeleting. Vi tenker oss to oljefelt. Hendelsen C = "oljefunn på felt 1", mens den komplementære hendelsen C' = "ingen olje på felt 1". Tilsvarende er hendelsen D = "oljefunn på felt 2", mens D' = "ingen olje på felt 2". Vi får oppgitt at $P(C \cap D) = 0.07$, $P(C' \cap D) = 0.11$, $P(C \cap D') = 0.14$ og $P(C' \cap D') = 0.68$.

   (b) Finn sannsynligheten for olje på felt 1. Anta at man har funnet olje på felt 2. Hva er nå sannsynligheten for olje på felt 1? Anta at man har påvist at felt 2 ikke inneholder olje. Hva er nå sannsynligheten for olje på felt 1? Er begivenhetene C og D uavhengige?

2. **Investeringer** (ikke ekte tall)

   En finansalytiker har kommet frem til at Orkla (ORK), en aksje på Oslo Børs, har forventet avkastning $E(X) = 4\%$ og varians til avkastningen lik $V(X) = 0.25$, mens Norsk Hydro (NHY), en annen aksje, har forventet avkastning $E(Y) = 6\%$ og varians $V(Y) = 0.49$. Korrelasjonen mellom disse to aksjenes avkastning er $\rho(X, Y) = 0.3$. Finansalytikeren vil investere en andel q $(0 < q < 1)$ i Orkla-aksjer og resten $(1 - q)$ i Hydroaksjer. Den kombinerte investeringen har derfor avkastning:

   $$U = qX + (1 - q)Y.$$

   (a) Sett $q = 0.3$ slik at $U = 0.3X + 0.7Y$. Regn ut forventning og varians til U. Hvordan er disse i forhold til forventning og og varians til hver av X og Y?

   Hint: Se på slides fra kapittel 5.3, og spesielt formelen for $V\left(\sum_{i=1}^{n} a_i X_i\right)$.

3. **QR-koder**

   Start med å lese teksten under Figur 1. Vi skal generelt betrakte X og Y som tilfeldige variabler. For den uthevede pikselen i figuren er gråfargen $X = 29$ og den sanne pikselverdien hvit, dvs. $Y = 0$. Vi antar at QR-koder er laget slik at det i gjennomsnitt er like mange hvite som sorte piksler, som betyr at $p_Y(0) = p_Y(1) = 1/2$.

   I denne situasjonen er X kontinuerlig fordelt $(0 \leq X \leq 100)$ og Y er diskret fordelt, men vi kan likevel snakke om simultanfordelingen til X og Y. Vi starter ut med å definere betinget tetthet til X, gitt verdien til Y:

   $$\begin{aligned} f_{X|Y}(x|0) &= \frac{3}{100}\left(1 - \frac{x}{100}\right)^2 \qquad \text{(piksel egentlig hvit)} \\ f_{X|Y}(x|1) &= \frac{3}{100}\left(\frac{x}{100}\right)^2 \qquad \text{(piksel egentlig sort)} \end{aligned}$$

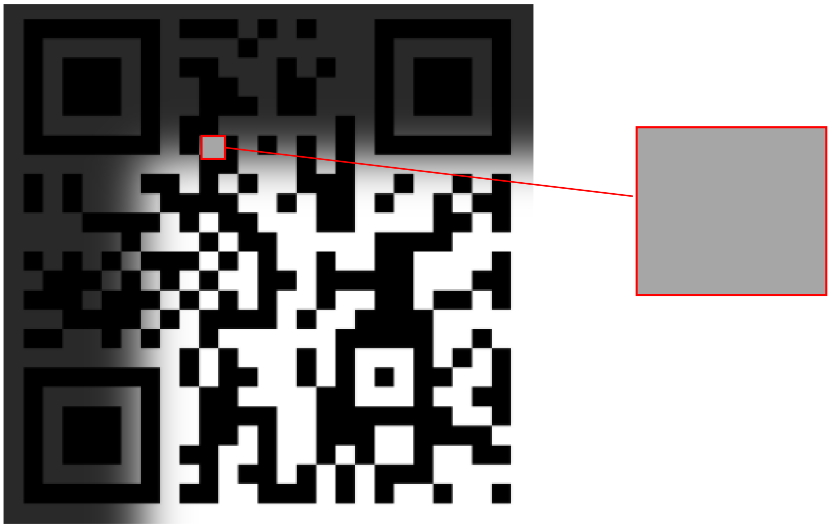

Figure 1: QR-kode fotografert i dårlig belysning, slik at kan være vanskelig å skille sorte og hvite piksler. Gråfargen (X) i hvert piksel er derfor kodet på en skala fra 0 (hvit) til 100 (sort). Det sanne pikselverdien (uten skygge) er kodet $Y = 0$ for hvit, og $Y = 1$ for sort.

(a) Lag et plott av marginaltettheten $f_X(x) = \sum_y f_{X|Y}(x|y)p_Y(y)$ for $0 \leq x \leq 100$.

Bayes formel tar i denne situasjonen formen:

$$P(Y = 0|X = x) = \frac{f_{X|Y}(x|0)p_Y(0)}{f_X(x)}$$

(b) Bruk Bayes formel til å finne sannsynligheten for hvit når $x = 29$ slik som på figuren.

4. **Sannsynlighetsmaksimering**

Vi har uavhengige tilfeldige variable X og Y med fordeling $X \sim \text{Poisson}(\theta)$ og $Y \sim \text{Poisson}(2\theta)$, og observasjoner $x = 2$ og $y = 5$ av disse.

(a) Vis at uttrykket for log-likelihood funksjonen er gitt ved

$$l(\theta) = [5\ln(2) - \ln(2!) - \ln(5!)] + 7\ln\theta - 3\theta.$$

Plott $l(\theta)$ for verdier av $\theta \in [0, 10]$ (en grov skisse er OK). For ca hvilken verdi av θ har $l(\theta)$ sitt maksimum?

(b) Beregn sannsynlighetsmaksimeringsestimatoren $\hat{\theta}$ for de observerte verdiene $x = 2$ og $y = 5$.

5. **Hvilken estimator er best?** (Modifisert eksamen V2017, oppg. 3)

   Et politisk parti finansierer sin virksomhet ved hjelp av et lotteri. Lotteriet er organisert slik at hver deltager har mulighet til å få gevinst i tre uavhengige spilleomganger. I følge reklamemateriellet fra partiet er det ulik sjanse for gevinst i de tre omgangene. Hvis vi betegner sannsynligheten for gevinst i første omgang med θ, er sannsynligheten for gevinst i andre omgang lik 2θ, og i tredje omgang 5θ. Her er θ en parameter som tilfredsstiller $\theta < 1/5$. Både i andre og tredje omgang har alle deltagerne de samme gevinstmulighetene, uansett om vedkommende allerede har vunnet gevinst i en tidligere omgang eller ikke.

   Vi innfører de tre tilfeldige variablene X_1, X_2 og X_3 som angir om du får gevinst i hver av de tre omgangene. Vi lar $X_1 = 1$ dersom første omgang gir gevinst, og $X_1 = 0$ ellers. Tilsvarende for X_2 i andre omgang ($X_2 = 1$ eller $X_2 = 0$), og X_3 i tredje omgang.

   (a) Vis at disse har forventning og varians:

$$\begin{aligned} E(X_1) &= \theta, & V(X_1) &= \theta(1-\theta) \\ E(X_2) &= 2\theta, & V(X_2) &= 2\theta(1-2\theta) \\ E(X_3) &= 5\theta, & V(X_3) &= 5\theta(1-5\theta) \end{aligned}$$

   Du vil bestemme et estimat for sannsynligheten θ ut fra de tre observasjonene X_1, X_2, X_3. Du foreslår at man skal bruke en av disse estimatorene:

$$\hat{\theta}_1 = \frac{1}{8}\left(X_1 + X_2 + X_3\right), \qquad \hat{\theta}_2 = \frac{1}{3}\left(X_1 + \frac{X_2}{2} + \frac{X_3}{4}\right).$$

   (b) Finn forventning og varians til $\hat{\theta}_1$ og $\hat{\theta}_2$ når $\theta = 0.12$. Avgjør hvilken av estimator som er best (når $\theta = 0.12$). Hint: Vi har

$$V(\hat{\theta}_2) = \frac{1}{3^2}\left(\theta(1-\theta) + \frac{1}{2^2}2\theta(1-2\theta) + \frac{1}{4^2}5\theta(1-5\theta)\right).$$

   som når vi evaluerer for $\theta = 0.12$ gir $V(\hat{\theta}_2) = 0.0185$.