

TMS136: Dataanalys och statistik – Tentamen 2010-01-11

Examinator och jour: Erik Jakobsson, tel. 031-772 53 79

Hjälpmedel: Chalmersgodkänd räknare och formelsamling (formelsamling delas ut med tentan).

Betygsgränser: För betyg **3** krävs 14 poäng, för **4** krävs 20 poäng och för betyg **5** krävs 27 poäng.

Fullständiga och välmotiverade lösningar skall ges till varje uppgift.

1. Låt A , B och C vara tre (icke-tomma) händelser där A och B är disjunkta.
 - (a) Rita Venndiagram (och förklara i ord eller formler) så att $P(A|C) = 1$ och $P(B|C) = 0$. (1p)
 - (b) För exakt vilka händelser C har vi att $P(C|A) = P(C|B) = 1$? (Rita och förklara!) (1p)
 - (c) Kan A och B vara beroende? Motivera. (1p)
2.
 - (a) Beskriv en situation där det skulle kunna vara lämpligt att använda en binomialfördelad variabel med parametrarna p och $n = 10$ som modell. Dvs $X \sim \text{Bin}(10, p)$. (2p)
 - (b) För det p du valt i (a), beräkna $P(X \leq 2)$. (1p)
3. En låda innehåller $n - 1$ vita bollar och 1 svart boll (dvs totalt n bollar). Vi drar k bollar, utan återläggning, vad är sannolikheten att vi dragit den svarta? (3p)
4. Tiden mellan två samtal till en telefonväxel är exponentialfördelad med väntevärde 2 minuter. Vad är sannolikheten att det kommer exakt 3 samtal under de 2 första minuterna? (3p)
5. Låt X vara en *diskret* stokastisk variabel med massfunktion $f_\theta(x)$, där θ är en okänd parameter, och låt $X_1, \dots, X_n$ vara ett stickprov från X .
 - (a) Hur skulle du förklara den intuitiva ideén bakom likelihoodfunktionen $L(\theta) = \prod_{1 \leq i \leq n} f(x_i|\theta)$ för någon som inte hört talas om den förut? (2p)
 - (b) Låt nu X vara geometriskt fördelad med parameter p . Bestäm maximum likelihoodskattaren för p givet stickprovet $X_1, \dots, X_n$. Maximum likelihoodskatta sedan p med dessa data: 6, 7, 14, 5, 4, 2, 11, 2, 5, 7. (3p)
 - (c) Betrakta ditt skattade p och fundera över tolkningen av en geometriskt fördelad variabel. Hur tror du att jag genererat data ovan? (1p)

6. Låt $X_1, X_2, \dots, X_n$ vara ett stickprov från en normalfördelad variabel X med parametrarna $\mu = E(X)$ och $\sigma^2 = Var(X)$.
- Definiera begreppet skattare. (1p)
 - $\bar{X} = \frac{1}{n} \sum_1^n X_i$ är ju den vanligaste skattaren för μ , men det finns förstås många. Hitta på en egen och motivera varför den är en skattare. (1p)
 - Beräkna $E(\bar{X})$ och $Var(\bar{X})$ och gör likadant för din egen skattare. Är din skattare väntevärdesriktig? Vilken av skattarna är bäst? Och vad betyder egentligen 'bäst' här? (3p)
 - Vad har $\bar{X}$ för fördelning? Varför? Om det är lätt, ge även fördelningen för din egen skattare. (1p)
 - En av Blossas nya glöggsorter som står på mitt skrivbord påstås ha en alkoholhalt per flaska som är normalfördelad med väntevärde 21% och en standardavvikelse på $\sqrt{2}$ procentenheter. Om jag skulle hälla 5 flaskor i en karaff och blanda, vad är sannolikheten att blandningen skulle få en alkoholhalt på mer än 22%? (2p)
 - Vi tror varken på Blossas påstående om 21% i genomsnitt eller på deras uppskattning av variansen. Så vi köper 5 flaskor glögg och mäter alkoholhalten hos var och en noggrant och får resultaten 17.64, 21.68, 21.02, 21.50, 18.11 (%).
Undersök, och diskutera vad vi kan dra för slutsatser ur dessa data. (3p)
7. En slumpgenerator jag har i min dator påstår sig ge observationer från en likformig fördelning på intervallet $[0, 1]$. Vi tror inte riktigt på detta och bestämmer oss för att undersöka påståendet. Vi genererar därför 200 slumpstal och indelar dem i de 10 delintervallen $[0, 0.1)$, $[0.1, 0.2)$, $[0.2, 0.3)$, $[0.3, 0.4)$, ... $[0.8, 0.9)$, $[0.9, 1]$.
- Undersök om vi bör skriva och klaga på tillverkaren av slumpgeneratorn då vi observerar följande frekvenser i vardera delintervall: 19, 29, 20, 21, 19, 18, 18, 16, 21, 19. (4p)
 - Hmm... men verkar det inte som om generatorn genererar tal i intervallet $[0.1, 0.2)$ oftare än i de andra intervallen? Om vi nu gör ett nytt test, med samma data som ovan, för att se om vi har skäl att tro att propotionen tal som genereras i $[0.1, 0.2)$ är större än 10% så kommer testet bekräfta detta (med ett p-värde på 0.017). Hur är det möjligt? Motsäger detta resultatet i (a)? (2p)

Lycka till!