UPPSALA UNIVERSITET
Matematiska Institutionen
Annika Heckel

PROV I MATEMATIK
Sannolikhet och statistik DV, 1MS321
2023-05-30

Hjälpmedel: Räknedosa, formel- och tabellsamling och engelsk-svensk ordlista för kursen 1MS321. Goda resonemang och motiveringar vägs in vid poängsättning. För betygen 3, 4, resp. 5 krävs normalt minst 18, 25, resp. 32 poäng.

Fråga 1: Den diskreta slumpvariabeln X kan endast anta värden i mängden $\{1,2,3,4\}$. Vi vet följande om sannolikhetsfunktionen $p(x)$ tillhörande X,

$$p(1) = 0{,}1 \qquad p(3) = 0{,}3 \qquad p(4) = 0{,}2.$$

a) Beräkna $E[X]$ och $V[X]$. (2p)

b) Beräkna $P(X \leqslant 3 \mid X \geqslant 3)$. (1p)

c) Är händelserna $\{X \geqslant 3\}$ och $\{X \leqslant 3\}$ oberoende? Motivera ditt svar! (1p)

d) Beräkna $C\big(X, \frac{1}{X}\big)$, det vill säga kovariansen mellan slumpvariablerna X och $\frac{1}{X}$. (2p)

Fråga 2: Antag att X är en slumpvariabel med täthetsfunktion

$$f(x) = \frac{1}{4}x^3 \text{ när } 0 \leqslant x \leqslant 2,$$

och $f(x) = 0$ annars.

a) Beräkna fördelningsfunktionen samt väntevärdet av X. (2p)

b) Simulera tre pseudoslumptal från sannolikhetsfördelningen av X med hjälp av följande tre pseudoslumptal från Re(0,1):

$$u_1 = 0{,}6755 \qquad u_2 = 0{,}1438 \qquad u_3 = 0{,}8224$$

(3p)

Fråga 3: Anna driver ett glasstånd och säljer glass för 10 kronor per skopa. 30% av alla kunder köper 1 skopa glass, 50% av alla kunder köper 2 skopor glass, samt 20% av alla kunder köper 3 skopor glass. På en dag har 400 kunder köpt glass och deras beställningar antas vara oberoende av varandra. Approximera sannolikheten att Anna har sålt glass för minst 7500 kronor under den dagen.

När du svarar på denna fråga, definiera lämpliga slumpvariabler och motivera noggrant eventuella approximationer. (6p)

Fråga 4: Låt $n \geqslant 2$ vara ett heltal. Vi har ett slumpmässigt stickprov $x_1, x_2, \ldots, x_n$ där motsvarande slumpvariabler $X_1, X_2, \ldots, X_n$ är oberoende med okänt väntevärde μ och varians 1. Oberoende av detta har vi ett stickprov $y_1, y_2, \ldots, y_n$ där motsvarande slumpvariabler $Y_1, Y_2, \ldots, Y_n$ är oberoende med väntevärde μ och varians 100.

Man vill skatta μ, och följande skattningar är föreslagna:

$$\hat{\mu}_1 = 2x_1 - x_2 \qquad \hat{\mu}_2 = \frac{\bar{x} + \bar{y}}{2}$$

a) Visa att både skattningarna är väntevärdesriktiga. (2p)

b) För vilka värden av n är $\hat{\mu}_1$ en effektivare skattning än $\hat{\mu}_2$? (3p)

Fråga 5: Ett mjukvaruföretag mäter svarsfördröjningen (eng. latency) på information, med olika storlek, som skickas till en API. De skickar 100 informationspaket till APIn, med respektive storlek $x_1, \dots, x_{100}$ mätt i antalet tecken. De mäter svarsfördröjningarna $y_1, \dots, y_{100}$ i millisekunder, där y_i är svarsfördröjningen tillhörande informationspaketet x_i.

För att modellera förhållandet mellan informationsstorlek och svarsfördröjning, tillämpas en enkel linjär regressionsmodel genom att definiera

$$Y_i = m + kx_i + \epsilon_i \quad \text{för} \quad i = 1, \dots, 100,$$

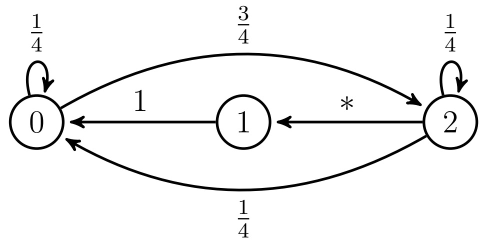
där $\epsilon_i \sim N(0, \sigma^2)$ är oberoende. Företaget använder färdiggjord mjukvara för att skatta parametrarna i modellen från tidigare nämnd data vilket ger $\hat{m} = 240$ och $\hat{k} = 10{,}1$, med förklaringsgrad $R^2 = 88{,}2\%$.

a) Företaget skickar information som är 55 tecken lång till APIn. Använd modellen för att prediktera svarsfördröjningen för det informationspaket. (1p)

b) I figurerna 1-4 ges fyra spridningsdiagram, där en av dem motsvarar företagets data. Vilken? Motivera ditt svar! (2p)

c) Företaget har alternativet att byta till en ny server för deras API. Den här servern ligger längre bort och har därför en större svarsfördröjning som en inneboende egenskap av avståndet, mer specifikt är svarsfördröjningen dubbelt så stor för en signal med 0 tecken. Däremot antas servern vara dubbelt så snabb, dvs. den del av svarsfördröjningen som kommer från informationsstorleken blir hälften så stor. Skatta parametrarna k_{new} och m_{new} för en linjär regressionsmodel som modellerar förhållandet mellan informationsstorlek och svarsfördröjning tillhörande den nya servern. Formulera ett beslutsprotokoll: Vilka informationspaket (dvs. beroende på deras längd) bör gå till den gamla servern samt till den nya för att göra svarsfördröjningen så liten som möjligt? (3p)

Fråga 6: Låt $X_0, X_1, X_2, \dots$ vara en Markovkedja med tillståndsrum $E = \{0, 1, 2\}$ och med övergångsdiagram



Markovkedjan startar vid tidpunkt $t = 0$ antingen vid tillståndet 0 eller 2, båda med samma sannolikhet.

a) Vad ska det stå istället för $*$ i övergångsdiagrammet ovan? (1p)

b) Vad är övergångsmatrisen tillhörande denna Markovkedja? (1p)

c) Beräkna $P(X_1 = 2)$. (1p)

d) Beräkna $P(X_0 = 0 \mid X_1 = 2)$. (1p)

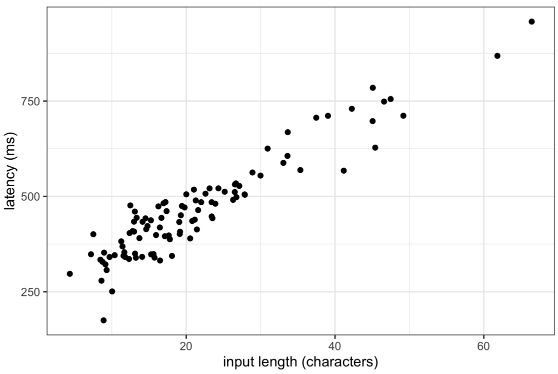
e) Bestäm Markovkedjans stationära fördelning. (Du kan anta att en unik stationär fördelning existerar.) (2p)

Fråga 7: Antalet e-post du får per timma kan modelleras som en Poissonfördelning med intensitet $\lambda = 2$ per timma. Vissa av dessa är spam och tas bort av ett spamfilter. Sannolikheten att en godtycklig e-post är spam är 20%, oberoende av alla andra e-post.

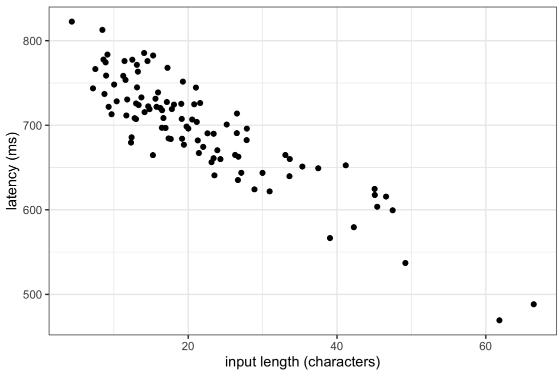
a) Om du struntar att ta spam i beaktning, vad är sannolikheten att du får precis 12 spamfria e-post under de följande 8 timmarna? (2p)

b) Din e-postleverantör råkade ut för flera systemavbrott under vilka inga e-post blev levererade. Det första systemavbrottet varade i 3 timmar, det andra varade i 1 timma, och det tredje varade i 2 timmar. Vad är fördelningen av totala antalet e-post (både spam och ej spam) du förlorat under alla tre systemavbrotten? Motivera ordentligt ditt svar. (2p)

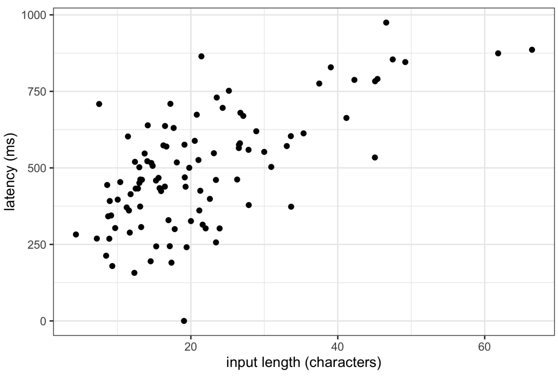
c) Du börjar övervaka din spam-mapp på morgonen. Vad är den förväntade väntetiden, i timmar, att du får ditt första spam den dagen? (2p)



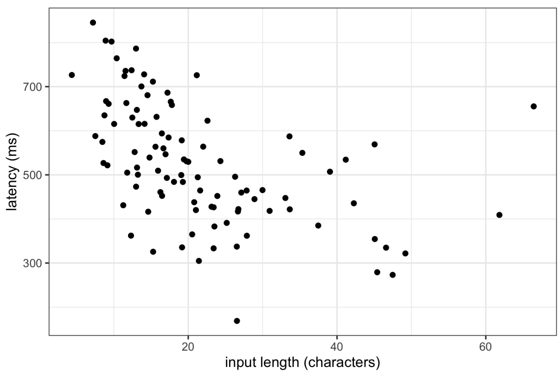
Figur 1



Figur 2



Figur 3



Figur 4