Tentamen – Bayesiansk inferens och maskininlärning (TIF385)

Tid och plats:	13 januari, 2023, fm, Johanneberg.
Hjälpmedel:	Physics Handbook, Beta Mathematics Handbook, typgodkänd kalkylator.
Examinator:	Christian Forssén.
Jourhavande lärare:	Christian Forssén (031–772 3261).

Tentamen består av sex uppgifter som kan ge maximalt 60 poäng totalt. För att bli godkänd med betyg 3 krävs 24 poäng, för betyg 4 krävs 36 poäng och för betyg 5 krävs 48 poäng.

Rättningsprinciper: Alla svar skall motiveras (om ej annat anges) och införda storheter skall förklaras. Lösningarna förväntas vara välstrukturerade och begripligt presenterade. Skriv och rita tydligt. Vid tentamensrättning gäller följande allmänna principer:

- För maximal (10) poäng krävs fullständigt korrekt och välmotiverad lösning.
- Mindre fel ger 1-3 poängs avdrag. Gäller även mindre brister i presentationen.
- Lösningar som inte går att följa (t.ex. avsaknad av relevant figur, ej definierade variabler, svårtydd, etc) renderar poängavdrag även om svaret verkar vara korrekt.
- Allvarliga principiella fel ger fullt poängavdrag.
- Allvarliga fel som leder till orimliga resultat kan ge lägre poängavdrag om orimligheten pekas ut.
- Även skisserade lösningar kan ge delpoäng.

Lycka till!

---

1. Valideringsdata och k-faldig korsvalidering.

   (a) Ge två exempel på vad valideringsdata kan användas till vid konstruktion av en maskininlärningsmodell. För full poäng skall du ange exempel där det konkreta användandet av valideringsdata framgår. (*4 poäng*)

   (b) Beskriv algoritmen för k-faldig korsvalidering och ge åtminstone två anledningar till varför den kan vara användbar. (*6 poäng*)

2. Antag att vi skall skapa en maskininlärningsmodell för att klassificera bilder av handritade siffror mellan 0–9. Varje bild definieras av 8×8 pixlar med ett värde på intensiteten för varje pixel. Vi har bestämt att använda ett fullständigt kopplat neuralt nätverk med fem gömda lager som har tjugo noder vardera samt ett utlager med tio noder. Varje nod karakteriseras av vikter som inkluderar en bias-term w_0. Vi använder sigmoidfunktionen, $1/(1+e^{-z})$, som aktivitetsfunktion i de gömda lagren.

   (a) Inför relevant notation och ge ett uttryck för utsignalerna från noderna i det första gömda lagret. Definiera storlekar på eventuella matriser och vektorer.

   (b) Hur många parametrar har vi totalt i vårt neurala nätverk?

   (c) Givet att vi vill skapa en mjuk klassificerare som ger sannolikheter för alla tio möjliga klasserna (siffrorna 0–9). Hur skulle vi kunna definiera utlagret och utsignalen från modellen?

   (3 poäng per korrekt besvarad deluppgift, 10 poäng för alla tre.)

3. Två studenter från Chalmers och tre från KTH befinner sig på en rymdexpedition till planeten Mars. Tre av dem går ombord på en landningskapsel medan två stannar ombord. Urvalet sker genom lottdragning där allas namn skrivs på lappar som läggs i en låda. Det första namnet som dras blir kapten på landningskapseln och de nästföljande två blir stödbesättning.

   (a) Beteckna den eftersökta sannolikheten i deluppgifterna (b–d) nedan med statistisk notation. (*2 poäng*)
   Tips: Inför gärna beteckningar för olika propositioner. Till exempel: $\mathrm{C}_!$ = Det första namnet som dras är en Chalmersstudent.

   (b) Vad är sannolikheten att kaptenen kommer från Chalmers? (*2 poäng*)

   (c) Vad är sannolikheten att kaptenen kommer från Chalmers givet att det andra namnet som dras är en student från Chalmers? Notera om denna sannolikhet är större, mindre eller lika stor som i (b)-uppgiften. (*3 poäng*)

   (d) Vad är sannolikheten att kaptenen kommer från Chalmers givet att åtminstone en i stödbesättningen kommer från Chalmers? (*3 poäng*)

4. Antag att vi har gjort bayesiansk inferens från data $\mathcal{D}$ med en modell $\tilde{y} = \theta_0 + \theta_1 x$ och finner att a posteriori-fördelningen för parametrarna är
$$p\left(\theta_0, \theta_1 \,|\, \mathcal{D}, I\right) \propto \exp\left[-\frac{1}{2}(\boldsymbol{\theta} - \bar{\boldsymbol{\theta}})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta} - \bar{\boldsymbol{\theta}})\right],$$
där $\boldsymbol{\theta} = \begin{pmatrix}\theta_0 \\ \theta_1\end{pmatrix}$ och $\bar{\boldsymbol{\theta}} = \begin{pmatrix}1 \\ 2\end{pmatrix}$ samt $\boldsymbol{\Sigma}^{-1} = 4\begin{pmatrix}2 & 1 \\ 1 & 5\end{pmatrix}$.

   (a) Beräkna varianserna $\mathrm{Var}\left(\theta_0\right)$, $\mathrm{Var}\left(\theta_1\right)$ samt kovariansen $\mathrm{Cov}\left(\theta_0, \theta_1\right)$.

   (b) Skissa a posteriori-fördelningen för parametrarna.

   (c) Givet att inferensen genomfördes med a priori fördelningen
$$p\left(\theta_0, \theta_1 \,|\, I\right) = \frac{1}{25(2\pi)} \exp\left[-(\theta_0^2 + \theta_1^2)/50\right],$$
   skulle du säga att slutresultatet domineras av trolighetsfunktionen eller av a priori fördelningen? (motivera ditt svar)

(3 poäng per korrekt besvarad deluppgift, 10 poäng för alla tre.)

5. Antag att en Markovkedja initialiseras med $p_{X_0}(x_0) = \mathcal{U}\left([0,1]\right)$ och uppdateras enligt
$$x_{n+1} = \frac{x_n + \xi}{2},$$
där $p(\xi) = \mathcal{U}\left([0,1]\right)$, dvs utfallet av en likformigt fördelad slumpvariabel.

   (a) Är detta en stationär kedja? (*2 poäng*)

   (b) Skissa $p_{X_1|X_0}\left(x_1|x_0\right)$ (*4 poäng*)

   (c) Beräkna $p_{X_1}(x_1)$. (*4 poäng*)

6. Betrakta den diskreta Poissonfördelningen (med $\lambda = 1$)
$$p(k) = \frac{e^{-1}}{k!}, \quad k = 0, 1, 2, \ldots$$
Konstruera och beskriv en Metropolis-Hastings-algoritm som vid konvergens kommer att generera stickprov ur denna fördelning. Utgå från en slumpvandring med steglängden 1 och beräkna de resulterande elementen $T(i,j)$ i övergångsmatrisen för $i, j \in \{0, 1, 2\}$. *(10 poäng)*

Tentamen – Bayesiansk inferens och maskininlärning (TIF385)

Tid och plats: 13 januari, 2023, fm, Johanneberg.
Hjälpmedel: Physics Handbook, Beta Mathematics Handbook, typgodkänd kalkylator.
Lösningsskiss: Christian Forssén.

Tentamen består av sex uppgifter som kan ge maximalt 60 poäng totalt. För att bli godkänd med betyg 3 krävs 24 poäng, för betyg 4 krävs 36 poäng och för betyg 5 krävs 48 poäng.

Rättningsprinciper: Alla svar skall motiveras (om ej annat anges) och införda storheter skall förklaras. Lösningarna förväntas vara välstrukturerade och begripligt presenterade. Skriv och rita tydligt. Vid tentamensrättning gäller följande allmänna principer:

- För maximal (10) poäng krävs fullständigt korrekt och välmotiverad lösning.
- Mindre fel ger 1-3 poängs avdrag. Gäller även mindre brister i presentationen.
- Lösningar som inte går att följa (t.ex. avsaknad av relevant figur, ej definierade variabler, svårtydd, etc) renderar poängavdrag även om svaret verkar vara korrekt.
- Allvarliga principiella fel ger fullt poängavdrag.
- Allvarliga fel som leder till orimliga resultat kan ge lägre poängavdrag om orimligheten pekas ut.
- Även skisserade lösningar kan ge delpoäng.

Detta är enbart en skiss av den fullständiga lösningen. Det kan innebära att vissa mellansteg i uträkningarna, som egentligen är nödvändiga för en komplett lösning, inte redovisas.

1. Valideringsdata och k-faldig korsvalidering.

   (a) Ge två exempel på vad valideringsdata kan användas till vid konstruktion av en maskininlärningsmodell. För full poäng skall du ange exempel där det konkreta användandet av valideringsdata framgår. (*4 poäng*)

   (b) Beskriv algoritmen för k-faldig korsvalidering och ge åtminstone två anledningar till varför den kan vara användbar. (*6 poäng*)

Lösning:

(a) Valideringsdata kan till exempel användas för att (i) bestämma hyperparametrar i maskininlärningsmodellen, eller (ii) för att uppskatta generaliseringsfelet. Notera dock att den inte kan användas för bägge dessa ändamål samtidigt. Då krävs en tredje datamängd.

Ett konkret exempel på (i) är följande: Betrakta ett neuralt nätverk med N noder i det enda gömda lagret. Utför träning av denna modell för olika värden på N och beräkna valideringsfelet för varje. Det värde på N som ger det lägsta uppnådda felet på valideringsdata är sedan att föredra för den slutliga modellen.

Ett konkret exempel på (ii) är följande: Antag istället att N är fixerat från början. Träna modellen på träningsdata och evaluera sedan felfunktionen på valideringsdata. Det genomsnittliga felet som uppnås ger då ett estimat på vilken precision som vi kan förvänta oss av modellen när den används för förutsägelser.

(b) Se till exempel avsnitt 17.4.2.1 i kompendiet. Några anledningar till att korsvalidering är användbar: (i) man har begränsat med data och vill anända denna så effektivt som möjligt för både träning och validering. (ii) valideringsmått är känsliga för vilket specifikt stickprov av valideringsdata som används. Det är därför bättre att betrakta ett flertal olika valideringsset. (iii) Träningen kan vara extra känslig för en viss delmängd av tillgänglig data. Med korsvalidering kommer all data att användas för träning.

---

2. Antag att vi skall skapa en maskininlärningsmodell för att klassificera bilder av handritade siffror mellan 0–9. Varje bild definieras av 8×8 pixlar med ett värde på intensiteten för varje pixel. Vi har bestämt att använda ett fullständigt kopplat neuralt nätverk med fem gömda lager som har tjugo noder vardera samt ett utlager med tio noder. Varje nod karakteriseras av vikter som inkluderar en bias-term w_0. Vi använder sigmoidfunktionen, $1/(1+e^{-z})$, som aktivitetsfunktion i de gömda lagren.

(a) Inför relevant notation och ge ett uttryck för utsignalerna från noderna i det första gömda lagret. Definiera storlekar på eventuella matriser och vektorer.

(b) Hur många parametrar har vi totalt i vårt neurala nätverk?

(c) Givet att vi vill skapa en mjuk klassificerare som ger sannolikheter för alla tio möjliga klasserna (siffrorna 0–9). Hur skulle vi kunna definiera utlagret och utsignalen från modellen?

(3 poäng per korrekt besvarad deluppgift, 10 poäng för alla tre.)

Lösning:

(a) Utsignalerna ges av vektorn $\boldsymbol{y} = f(\boldsymbol{z})$, vilket alltså innebär att utsignalen från nod i i det första gömda lagret är $y_i = f(z_i)$ där $f(z_i) = 1/(1 + e^{-z_i})$.

Både vektorn $\boldsymbol{y}$ och $\boldsymbol{z}$ har längden N där N är bredden på det första gömda lagret.

Aktiveringerna $\boldsymbol{z}$ ges av

$$\boldsymbol{z} = \boldsymbol{y}_{\text{in}} \boldsymbol{W} + \boldsymbol{w}_0,$$

där vektorn med insignaler, $\boldsymbol{y}_{\text{in}}$, har längden N_{in}, viktmatrisen för det första gömda lagret, $\boldsymbol{W}$, har dimensionen $N_{\text{in}} \times N$ samt vektorn med dess basvikter, $\boldsymbol{w}_0$, har längden N. För att vara extra tydlig bör man klargöra att formen ovan bygger på att samtliga vektorer är radvektorer, dvs $1 \times N$ eller $1 \times N_{\text{in}}$.

(b) Inlagret har bredden 64. Dessa signaler går till samtliga noder i det första gömda lagret som alltså måste ha $64 + 1$ vikter vardera. De 20 noderna i varje gömt lager ger varsin utsignal som går till varje nod i det efterföljande lagret som alltså behöver $20 + 1$ vikter vardera. Detta ger

$$20 \times (64 + 1) + (20 + 20 + 20 + 20 + 10) \times (20 + 1) = 3190$$

(c) Vi önskar en mjuk klassificerare vilket innebär att modellen skall ge en (diskret) sannolikhetsfördelning, dvs en vektor där elementen kan tolkas som sannolikheten att måldata tillhör den specifika klassen $p(t = i) = \hat{y}_i$.

För att åstadkomma detta kan vi till exempel använda sigmoid-funktionen även för noderna i utlagret. Dessa leder då till en utvektor $\tilde{\boldsymbol{y}}$ med tio element $\tilde{y}_i = f(z_i)$ som uppfyller $0 \le \tilde{y}_i \le 1$. Denna vektor måste dock normaliseras för att omvandlas till en diskret sannolikhetsfördelning $\hat{\boldsymbol{y}}$. Alltså ges utsignalen av

$$\hat{y}_i = \frac{f(z_i)}{\sum_{i=0}^{9} f(z_i)}.$$

3. Två studenter från Chalmers och tre från KTH befinner sig på en rymdexpedition till planeten Mars. Tre av dem går ombord på en landningskapsel medan två stannar ombord. Urvalet sker genom lottdragning där allas namn skrivs på lappar som läggs i en låda. Det första namnet som dras blir kapten på landningskapseln och de nästföljande två blir stödbesättning.

   (a) Beteckna den eftersökta sannolikheten i deluppgifterna (b–d) nedan med statistisk notation. (*2 poäng*)
   Tips: Inför gärna beteckningar för olika propositioner. Till exempel: C! = Det första namnet som dras är en Chalmersstudent.

   (b) Vad är sannolikheten att kaptenen kommer från Chalmers? (*2 poäng*)

   (c) Vad är sannolikheten att kaptenen kommer från Chalmers givet att det andra namnet som dras är en student från Chalmers? Notera om denna sannolikhet är större, mindre eller lika stor som i (b)-uppgiften. (*3 poäng*)

   (d) Vad är sannolikheten att kaptenen kommer från Chalmers givet att åtminstone en i stödbesättningen kommer från Chalmers? (*3 poäng*)

Lösning:

(a) Vi inför följande propositioner:

C_i = Det i:te namnet som dras är en Chalmersstudent.

K_i = Det i:te namnet som dras är en KTH-student.

C_B = Åtminstone en besättningsmedlem är en Chalmersstudent.

I = Given information kring besättningsens sammansättning och process för urvalet.

De eftersökta sannolikheterna i respektive deluppgift:
(b) $\mathbb{P}(C_1 \,|\, I)$, (c) $\mathbb{P}(C_1 \,|\, C_2, I)$, (d) $\mathbb{P}(C_1 \,|\, C_B, I)$

(b) $\mathbb{P}(C_1 \,|\, I) = \frac{2}{5}$

(c) Här gäller det att visa/argumentera att $\mathbb{P}(C_1 \,|\, C_2, I) = \mathbb{P}(C_2 \,|\, C_1, I)$ där den andra sannolikheten i likheten triviellt är 1/4. Likheten understryker att betingade sannolikheter inte beror på kausalitet. Dvs information om vad som händer senare påverkar sannolikheten för vad som hände först.

Vi skriver om sannolikheten med Bayes formel

$$\mathbb{P}(C_1 \mid C_2, I) = \frac{\mathbb{P}(C_2 \mid C_1, I)\,\mathbb{P}(C_2 \mid I)}{\mathbb{P}(C_1 \mid I)},$$

Utan information kring utfallet på någon av lottningarna så måste $\mathbb{P}(C_1 \mid I) = \mathbb{P}(C_2 \mid I)$ så att vi har visat att

$$\mathbb{P}(C_1 \mid C_2, I) = \mathbb{P}(C_2 \mid C_1, I) = \frac{1}{4}.$$

Givet informationen att en Chalmersstudent väljs som nummer två så blir sannolikheten att en Chalmersstudent valdes först *mindre* än vad den är om vi inte hade den informationen (uppgift b).

(d) Vi skriver först om den eftersökta sannolikheten med Bayes formel

$$\mathbb{P}(C_1 \mid C_B, I) = \frac{\mathbb{P}(C_B \mid C_1, I)\,\mathbb{P}(C_1 \mid I)}{\mathbb{P}(C_B \mid I)}. \tag{1}$$

Den andra termen i täljaren är triviell $\mathbb{P}(C_1 \mid I) = \frac{2}{5}$. För att beräkna den första termen utnyttjar vi komplementet och summaregeln $\mathbb{P}(C_B \mid C_1, I) = 1 - \mathbb{P}\left(\overline{C}_B \mid C_1, I\right)$. Eftersom $C_B = C_2 + C_3$ (där plustecknet skall tolkas som "eller") så är komplementet $\overline{C}_B = K_2, K_3$ (där kommatecknet skall tolkas som "och"). Dvs negationen till att någon av besättningsmedlemmarna är Chalmersstudent är att bägge är KTH-studenter. Vi får att

$$\begin{aligned}
\mathbb{P}(C_B \mid C_1, I) &= 1 - \mathbb{P}(K_2, K_3 \mid C_1, I) \\
&= 1 - \mathbb{P}(K_2 \mid C_1, I)\,\mathbb{P}(K_3 \mid K_2, C_1, I) \\
&= 1 - \left(\frac{3}{4}\right)\left(\frac{2}{3}\right) = \frac{1}{2},
\end{aligned}$$

där vi utnyttjar produktregeln på andra raden.

Analogt kan vi skriva nämnaren $\mathbb{P}(C_B \mid I) = 1 - \mathbb{P}\left(\overline{C}_B \mid I\right) = 1 - \mathbb{P}(K_2, K_3 \mid I)$. Eftersom ingen information ges kring resultatet på den första lotten så måste $\mathbb{P}(K_2, K_3 \mid I) = \mathbb{P}(K_1, K_2 \mid I)$. Samma sorts uträkning som ovan ger att

$$\begin{aligned}
\mathbb{P}(C_B \mid I) &= 1 - \mathbb{P}(K_1 \mid I)\,\mathbb{P}(K_2 \mid K_1 I) \\
&= 1 - \left(\frac{3}{5}\right)\left(\frac{2}{4}\right) = \frac{7}{10}.
\end{aligned}$$

Återstår bara att sätta in detta i Ekv. (1) och får

$$\mathbb{P}(C_1 \mid C_B, I) = \frac{2}{7}.$$

---

4. Antag att vi har gjort bayesiansk inferens från data $\mathcal{D}$ med en modell $\tilde{y} = \theta_0 + \theta_1 x$ och finner att a posteriori-fördelningen för parametrarna är
$$p(\theta_0, \theta_1 \,|\, \mathcal{D}, I) \propto \exp\left[-\frac{1}{2}(\boldsymbol{\theta} - \overline{\boldsymbol{\theta}})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{\theta} - \overline{\boldsymbol{\theta}})\right],$$
där $\boldsymbol{\theta} = \begin{pmatrix}\theta_0 \\ \theta_1\end{pmatrix}$ och $\overline{\boldsymbol{\theta}} = \begin{pmatrix}1 \\ 2\end{pmatrix}$ samt $\boldsymbol{\Sigma}^{-1} = 4\begin{pmatrix}2 & 1 \\ 1 & 5\end{pmatrix}$.

   (a) Beräkna varianserna $\mathrm{Var}(\theta_0)$, $\mathrm{Var}(\theta_1)$ samt kovariansen $\mathrm{Cov}(\theta_0, \theta_1)$.

   (b) Skissa a posteriori-fördelningen för parametrarna.

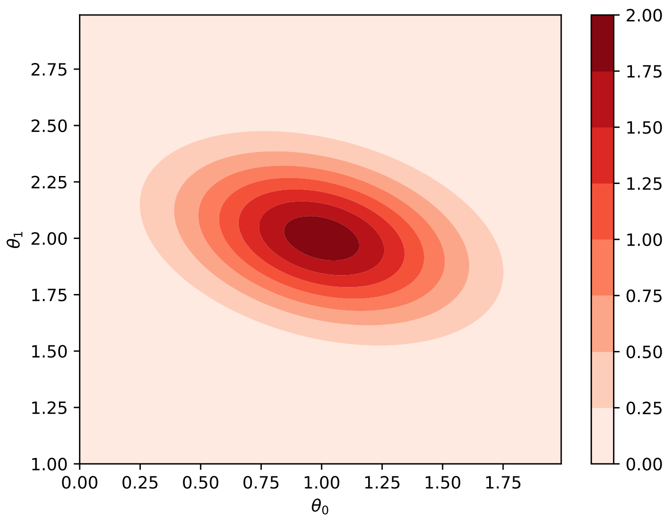
   (c) Givet att inferensen genomfördes med a priori fördelningen
$$p(\theta_0, \theta_1 \,|\, I) = \frac{1}{25(2\pi)} \exp\left[-(\theta_0^2 + \theta_1^2)/50\right],$$
skulle du säga att slutresultatet domineras av trolighetsfunktionen eller av a priori fördelningen? (motivera ditt svar)

(3 poäng per korrekt besvarad deluppgift, 10 poäng för alla tre.)

Lösning:

(a) Kovariansmatrisen är inversen av den givna matrisen,
$$\boldsymbol{\Sigma} = \frac{1}{4\cdot 9}\begin{pmatrix}5 & -1 \\ -1 & 2\end{pmatrix}.$$
Kommentar: Ett enkelt sätt att finna den är att börja med att ansätta de icke-diagonala matriselementen till -1. De diagonala följer då av att vi skall få nollorna i identitetsmatrisen medan normaliseringen följer av ettorna
$$\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma} = \left[4\begin{pmatrix}2 & 1 \\ 1 & 5\end{pmatrix}\right]\left[\frac{1}{4\cdot 9}\begin{pmatrix}5 & -1 \\ -1 & 2\end{pmatrix}\right] = \begin{pmatrix}1 & 0 \\ 0 & 1\end{pmatrix}$$
Från kovariansmatrisen får vi direkt $\mathrm{Var}(\theta_0) = 5/36$, $\mathrm{Var}(\theta_1) = 1/18$ samt kovariansen $\mathrm{Cov}(\theta_0, \theta_1) = -1/36$.

(b) Detta är en bivariat normalfördelning med moden i $(\theta_0, \theta_1) = (1, 2)$ och negativ korrelation, dvs en lutande ellips.



(c) A priorifördelningen är en produkt av två oberoende, univariata normalfördelningar med moden i $\theta_i = 0$ och variansen $\sigma_i = 5$ ($i \in \{0, 1\}$). Eftersom moden för a posteriorifördelningen har flyttat på sig en bra bit, och framförallt eftersom den har en mycket snävare utsträckning (mindre varianser samt parameterkorrelation) så är det uppenbart att trolighetsfunktionen har varit dominerande.

---

5. Antag att en Markovkedja initialiseras med $p_{X_0}(x_0) = \mathcal{U}([0,1])$ och uppdateras enligt

$$x_{n+1} = \frac{x_n + \xi}{2},$$

där $p(\xi) = \mathcal{U}([0,1])$, dvs utfallet av en likformigt fördelad slumpvariabel.

(a) Är detta en stationär kedja? (*2 poäng*)

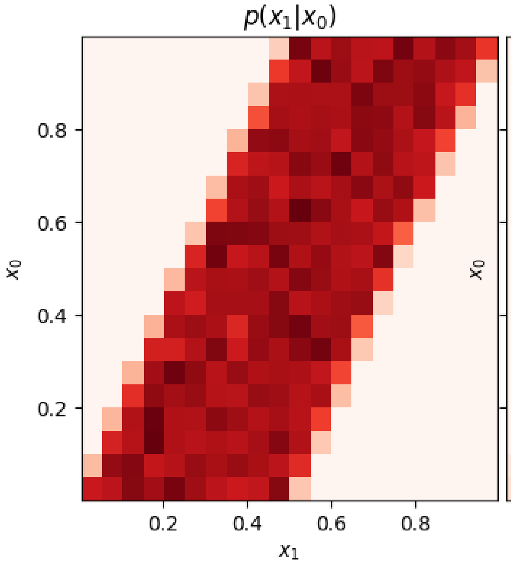
(b) Skissa $p_{X_1|X_0}(x_1|x_0)$ (*4 poäng*)

(c) Beräkna $p_{X_1}(x_1)$. (*4 poäng*)

Lösning:

(a) Detta är en stationär kedja eftersom det inte finns något beroende på tidsindex: $p_{X_{n+1}|X_n}(x|y) = p_{X_1|X_0}(x|y)$.

(b) Se figur nedan



(c) $p(x_1) = \int p(x_0, x_1)\, dx_0 = \int p(x_1 \,|\, x_0)\, p(x_0)\, dx_0$. Fördelningen $p(x_0) = \mathcal{U}([0,1])$, dvs likformig på intervallet $x_0 \in [0,1]$, medan $p(x_1 \,|\, x_0) = \mathcal{U}\left(\left[\frac{x_0}{2}, \frac{x_0+1}{2}\right]\right)$. Integralen löses för två olika fall $x_1 \leq 0.5$ och $x_1 > 0.5$ och man finner

$$p(x_1) = \begin{cases} 4x_1 & \text{för } x_1 \leq 0.5, \\ 4 - 4x_1 & \text{för } x_1 > 0.5. \end{cases}$$

---

6. Betrakta den diskreta Poissonfördelningen (med $\lambda = 1$)

$$p(k) = \frac{e^{-1}}{k!}, \quad k = 0, 1, 2, \ldots$$

Konstruera och beskriv en Metropolis-Hastings-algoritm som vid konvergens kommer att generera stickprov ur denna fördelning. Utgå från en slumpvandring med steglängden 1 och beräkna de resulterande elementen $T(i,j)$ i övergångsmatrisen för $i, j \in \{0, 1, 2\}$. *(10 poäng)*

Lösning:

- Vi utgår från den symmetriska stegförslagsfunktionen $S(i,j) = 0.5$ för $j = i \pm 1$ (och noll annars). Dvs en slumpmässig vandring med steglängden 1.

- De relevanta elementen i acceptansfunktionen

$$A(i,j) = \max\left(1, \frac{p(j)}{p(i)}\right)$$

  är: $A(i,i-1) = 1$, $A(i,i+1) = 1/(i+1)$.

  Vi definierar också elementet $A(0,-1) = 0$ eftersom stegfunktionen kan föreslå ett steg från $i = 0$ till $j = -1$ vilket dock inte är ett möjligt utfall.
- Icke-diagonala element i övergångsmatrisen $T(i,j) = S(i,j)A(i,j)$. De diagonala ges av $T(i,i) = 1 - \sum_{j\neq i} T(i,j)$.
- Låt oss betrakta den första raden som beskriver övergångar från tillstånd 0: $T(0,1) = 0.5 * 1/2 = 1/4$, $T(0,2) = 0 * A(0,2) = 0$, $T(0,0) = 1 - 1/4 - 0 - 0 - \ldots = 3/4$.
- Även om följande element inte efterfrågades så behövs det för att få det diagonala $T(2,2)$-elementet: $T(2,3) = 0.5 * 1/3 = 1/6$.

Ovanstående ger

$$T = \begin{pmatrix} 3/4 & 1/4 & 0 & \ldots \\ 1/2 & 1/4 & 1/4 & \\ 0 & 1/2 & 1/3 & \\ \vdots & & & \ddots \end{pmatrix}$$

Vi beskriver algoritmen med pseudokod:

(a) Initialisera kedjan med något tillstånd. För att vara konkreta väljer vi $x_0 = 1$. Detta är alltså iteration $n = 0$.

(b) Vid iteration $n \geq 1$ med $x_{n-1} = i$:

   i. Om $i \geq 1$: Singla en rättvis slant.
      - Om utfallet är krona, $x_n = i - 1$
      - Annars om utfallet är klave, dra ett slumptal $a \sim \mathcal{U}([0,1])$: $x_n = i + 1$ om $a \leq 1/(i+1)$, $x_n = 1$ annars.

   ii. Annars om $i = 0$: Singla en rättvis slant.
      - Om utfallet är krona, $x_n = 0$
      - Annars om utfallet är klave, singla slanten igen: $x_n = 1$ om krona, $x_n = 0$ om klave.