

TEK5020/9020 Mønstergjenkjenning Høsten 2023

Forelesning 5 – Parametriske metoder (2)

Idar Dyrdal (idar.dyrdal@its.uio.no)

UiO : Institutt for teknologisystemer

9. september 2023

Innhold i kurset

- Introduksjon til mønstergjenkjenning
- Beslutningsteori (desisjonsteori)
- [Parametriske metoder \(forts.\)](#)
- Ikke-parametriske metoder
- Lineære og generaliserte diskriminantfunksjoner
- Evaluering av klassifikatorer
- Ikke-ledet læring
- Klyngeanalyse.

Bayesisk estimering

Metoden bruker a priori kunnskap i tillegg til informasjonen fra treningssettet \mathcal{X} . I tillegg betraktes parametervektoren som en stokastisk variabel med en gitt fordeling.

Anta nå

$$\begin{aligned} p(\mathbf{x}|\boldsymbol{\theta}) &= \text{antatt parametrisk fordeling,} \\ p(\boldsymbol{\theta}) &= \text{a priori parameterfordeling.} \end{aligned}$$

Tetthetsfunksjonen $p(\boldsymbol{\theta})$ representerer på sett og vis en initiell formodning om parametervektoren, og er noe man må velge ut fra kunnskap om problemet.

Maksimalpunktet i fordelingen vil normalt representere det man anser som den mest sannsynlige verdien, mens spredningen (variansen) til fordelingen sier noe om usikkerheten i denne antagelsen.

Bayesisk estimering (forts.)

Siden vi er interessert i å klassifisere ukjente objekter, ønsker vi å estimere tettheten i et vilkårlig punkt \mathbf{x} i egenskapsrommet, dvs. vi ønsker å finne tetthetsfunksjonen $p(\mathbf{x}|\mathcal{X})$ som er gitt ved

$$\begin{aligned} p(\mathbf{x}|\mathcal{X}) &= \int p(\mathbf{x}, \boldsymbol{\theta}|\mathcal{X}) d\boldsymbol{\theta} \\ &= \int p(\mathbf{x}|\boldsymbol{\theta}, \mathcal{X}) p(\boldsymbol{\theta}|\mathcal{X}) d\boldsymbol{\theta} \\ &= \int p(\mathbf{x}|\boldsymbol{\theta}) p(\boldsymbol{\theta}|\mathcal{X}) d\boldsymbol{\theta} \quad (\text{siden } \mathbf{x} \text{ er uavhengig av } \mathcal{X}), \end{aligned}$$

der Bayes regel gir at $p(\boldsymbol{\theta}|\mathcal{X})$ kan skrives som

$$p(\boldsymbol{\theta}|\mathcal{X}) = \frac{p(\mathcal{X}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{\int p(\mathcal{X}|\boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta}}.$$

Dette er *a posteriori parameterfordeling*, dvs. parameterfordelingen etter at treningssettet er introdusert.

Bayesisk estimering (forts.)

I uttrykket for a posteriori parameterfordeling inngår likelihoodfunksjonen

$$p(\mathcal{X}|\boldsymbol{\theta}) = \prod_{k=1}^n p(\mathbf{x}_k|\boldsymbol{\theta}),$$

som var utgangspunktet for maksimum likelihood metoden.

Hvis $p(\mathcal{X}|\boldsymbol{\theta})$ har en skarp topp i ett eller annet punkt $\hat{\boldsymbol{\theta}}$, dvs. maksimum likelihood løsningen, og $p(\boldsymbol{\theta})$ er tilnærmet uniform i nærheten av denne verdien, vil a posteriori parameterfordeling gitt ved

$$p(\boldsymbol{\theta}|\mathcal{X}) = \frac{p(\mathcal{X}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{\int p(\mathcal{X}|\boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta}},$$

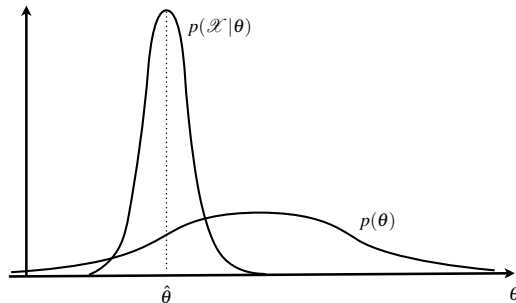
også ha en skarp topp i nærheten av $\hat{\boldsymbol{\theta}}$.

Bayesisk estimering (forts.)

Integralet i likningen for tettheten

$$p(\mathbf{x}|\mathcal{X}) = \int p(\mathbf{x}|\boldsymbol{\theta})p(\boldsymbol{\theta}|\mathcal{X})d\boldsymbol{\theta}$$

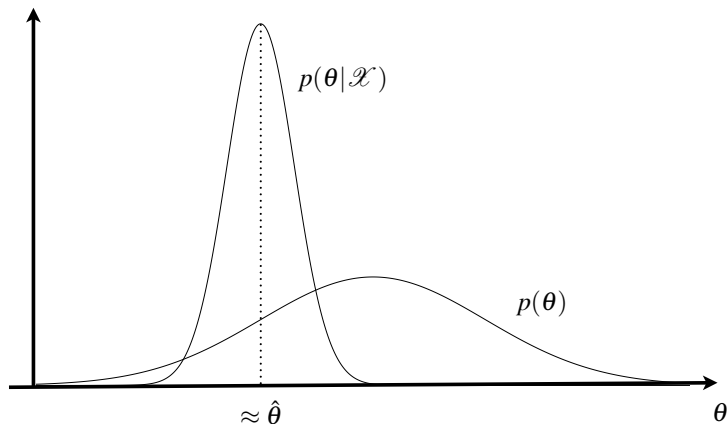
vil derfor lede til et resultat som tilsvarer hva man ville få med maksimum likelihoodmetoden, siden $\hat{\boldsymbol{\theta}}$ er tilnærmet den verdien som vektlegges sterkest i integralet.



Likelihoodfunksjonen og a priori parameterfordeling.

Selv om $p(\boldsymbol{\theta})$ er langt fra uniform vil a posteriori parameterfordeling uansett ha en skarp topp i nærheten av maksimum likelihoodløsningen når n er stor.

Bayesisk estimering (forts.)



A priori og a posteriori parameterfordelinger. A posteriorifordelingen vil typisk være mye skarpere enn a priorifordelingen.

Bayesisk estimering (forts.)

I det asymptotiske tilfellet ($n \rightarrow \infty$) vil $p(\boldsymbol{\theta}|\mathcal{X})$ i de fleste tilfeller konvergere mot en Dirac-deltafunksjon sentrert omkring $\hat{\boldsymbol{\theta}}$.

Tettheten i \mathbf{x} blir da

$$p(\mathbf{x}|\mathcal{X}) = \int p(\mathbf{x}|\boldsymbol{\theta})p(\boldsymbol{\theta}|\mathcal{X})d\boldsymbol{\theta} = \int p(\mathbf{x}|\boldsymbol{\theta})\delta(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})d\boldsymbol{\theta} = p(\mathbf{x}|\hat{\boldsymbol{\theta}}),$$

som er nettopp maksimum likelihoodløsningen. I det endelige tilfellet vil dette fremdeles være tilnærmet riktig, slik at

$$p(\mathbf{x}|\mathcal{X}) \approx p(\mathbf{x}|\hat{\boldsymbol{\theta}}),$$

dersom treningssettet er stort ($n \gg 1$).

Vi skal se på noen eksempler på bruk av Bayesisk estimering.

Bayesisk estimering av forventningsverdien i en univariat normalfordeling

La den ukjente tetthetsfunksjonen være en éndimensjonal normalfordeling $N(\mu, \sigma^2)$ der bare forventningsverdien μ er ukjent. I tillegg antar vi en a priori parameterfordeling, som også er univariat normal, og med kjent forventning og varians.

Utgangspunktet er altså følgende:

$p(x|\mu) = N(\mu, \sigma^2)$ der σ antas kjent og μ er ukjent,

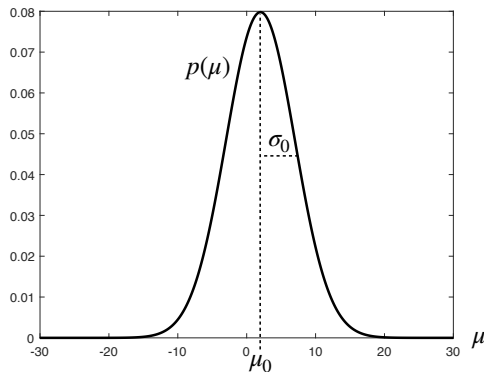
$p(\mu) = N(\mu_0, \sigma_0^2)$ der μ_0 og σ_0 er kjente,

$\mathcal{X} = \{x_1, x_2, \dots, x_n\}$ treningssampler fra aktuell klasse.

Her kan man se på μ_0 som en initiell antakelse om verdien til parameteren μ og σ_0 som usikkerheten i denne antakelsen.

Vi skal se hvordan Bayesisk estimering kan brukes til å oppnå et estimet av tettheten i et vilkårlig punkt x i egenskapsrommet.

Bayesisk estimering av forventningsverdien i en univariat normalfordeling (forts.)



A priori parameterfordeling $p(\mu)$ der forventningen μ_0 representerer den beste gjetningen om parameterverdien og standardavviket σ_0 representerer usikkerheten i denne antagelsen.

Bayesisk estimering av forventningsverdien i en univariat normalfordeling (forts.)

Likelihoodfunksjonen blir her

$$\begin{aligned} p(\mathcal{X}|\mu) &= \prod_{k=1}^n p(x_k|\mu) = \prod_{k=1}^n \frac{1}{\sqrt{2\pi}\sigma} \exp \left[-\frac{1}{2} \left(\frac{x_k - \mu}{\sigma} \right)^2 \right] \\ &= \frac{1}{(2\pi)^{n/2} \sigma^n} \exp \left[-\frac{1}{2} \sum_{k=1}^n \left(\frac{x_k - \mu}{\sigma} \right)^2 \right] \end{aligned}$$

A posteriori parameterfordeling blir da

$$\begin{aligned} p(\mu|\mathcal{X}) &= \frac{p(\mathcal{X}|\mu)p(\mu)}{\int p(\mathcal{X}|\mu')p(\mu')d\mu'} \\ &= \alpha \exp \left[-\frac{1}{2} \sum_{k=1}^n \left(\frac{x_k - \mu}{\sigma} \right)^2 \right] \times \exp \left[-\frac{1}{2} \left(\frac{\mu - \mu_0}{\sigma_0} \right)^2 \right] \end{aligned}$$

Bayesisk estimering av forventningsverdien i en univariat normalfordeling (forts.)

som videre gir

$$\begin{aligned} p(\mu|\mathcal{X}) &= \alpha \exp \left[-\frac{1}{2} \left\{ \sum_{k=1}^n \left(\frac{x_k - \mu}{\sigma} \right)^2 + \left(\frac{\mu - \mu_0}{\sigma_0} \right)^2 \right\} \right] \\ &= \alpha \exp \left[-\frac{1}{2} \left\{ \frac{1}{\sigma^2} \sum_{k=1}^n x_k^2 - 2 \frac{\mu}{\sigma^2} \sum_{k=1}^n x_k + n \frac{\mu^2}{\sigma^2} + \frac{\mu^2}{\sigma_0^2} - 2 \frac{\mu \mu_0}{\sigma_0^2} + \frac{\mu_0^2}{\sigma_0^2} \right\} \right] \\ &= \alpha' \exp \left[-\frac{1}{2} \left\{ \left(\frac{n}{\sigma^2} + \frac{1}{\sigma_0^2} \right) \mu^2 - 2 \mu \left(\frac{1}{\sigma^2} \sum_{k=1}^n x_k + \frac{\mu_0}{\sigma_0^2} \right) \right\} \right] \\ &= \alpha' \exp \left[-\frac{1}{2} \left(\frac{\mu^2 - 2 \mu \mu_n}{\sigma_n^2} \right) \right] \quad (\text{der hjelpestørrelsene } \mu_n \text{ og } \sigma_n \text{ er innført}) \\ &= \alpha'' \exp \left[-\frac{1}{2} \left(\frac{\mu - \mu_n}{\sigma_n} \right)^2 \right] = N(\mu_n, \sigma_n^2). \end{aligned}$$

Bayesisk estimering av forventningsverdien i en univariat normalfordeling (forts.)

Her er μ_n og σ_n gitt ved

$$\mu_n = \frac{\frac{1}{\sigma^2} \sum_{k=1}^n x_k + \frac{\mu_0}{\sigma_0^2}}{\frac{n}{\sigma^2} + \frac{1}{\sigma_0^2}} = \frac{\sigma_0^2 \sum_{k=1}^n x_k + \sigma^2 \mu_0}{n\sigma_0^2 + \sigma^2}$$

og

$$\sigma_n^2 = \frac{1}{\frac{n}{\sigma^2} + \frac{1}{\sigma_0^2}} = \frac{\sigma^2 \sigma_0^2}{n\sigma_0^2 + \sigma^2}.$$

Bayesisk estimering av forventningsverdien i en univariat normalfordeling (forts.)

Ved å sette inn for

$$m_n = \frac{1}{n} \sum_{k=1}^n x_k \quad \text{dvs. samplemiddel over treningssettet,}$$

får man da

$$\mu_n = \frac{n\sigma_0^2 m_n + \sigma^2 \mu_0}{n\sigma_0^2 + \sigma^2} = \frac{n\sigma_0^2}{n\sigma_0^2 + \sigma^2} m_n + \frac{\sigma^2}{n\sigma_0^2 + \sigma^2} \mu_0$$

for forventningsestimatet.

Legg merke til at μ_n er et veiet middel av m_n (målinger) og μ_0 (a priori antagelse), siden $\mu_n = \alpha_1 m_n + \alpha_2 \mu_0$ der $\alpha_1 + \alpha_2 = 1$.

Bayesisk estimering av forventningsverdien i en univariat normalfordeling (forts.)

Har nå funnet

$$p(\mu | \mathcal{X}) = \frac{1}{\sqrt{2\pi}\sigma_n} \exp \left[-\frac{1}{2} \left(\frac{\mu - \mu_n}{\sigma_n} \right)^2 \right] = N(\mu_n, \sigma_n^2) \quad (\text{a posteriori parameterfordeling}),$$

mens vi fra før har

$$p(\mu) = \frac{1}{\sqrt{2\pi}\sigma_0} \exp \left[-\frac{1}{2} \left(\frac{\mu - \mu_0}{\sigma_0} \right)^2 \right] = N(\mu_0, \sigma_0^2) \quad (\text{a priori parameterfordeling}).$$

Som vi ser av dette er konsekvensen av å inkludere treningssamplene at a priori parameterverdier erstattes av a posteriori verdier, dvs.:

$\mu_0 \rightarrow \mu_n$ der μ_n er beste gjetning om μ etter at treningssettet er introdusert, og
 $\sigma_0 \rightarrow \sigma_n$ der σ_n representerer usikkerheten i denne gjetningen.

Bayesisk estimering av forventningsverdien i en univariat normalfordeling (forts.)

Vi har altså funnet at a posteriori parameterfordeling er gitt ved

$$p(\mu|\mathcal{X}) = N(\mu_n, \sigma_n^2)$$

der

$$\mu_n = \frac{n\sigma_0^2}{n\sigma_0^2 + \sigma^2} m_n + \frac{\sigma^2}{n\sigma_0^2 + \sigma^2} \mu_0 \quad \text{der} \quad m_n = \frac{1}{n} \sum_{k=1}^n x_k,$$

og

$$\sigma_n^2 = \frac{\sigma^2 \sigma_0^2}{n\sigma_0^2 + \sigma^2}.$$

Her er μ_0 , σ_0 og σ kjente størrelser.

Bayesisk estimering av forventningsverdien i en univariat normalfordeling (forts.)

Av disse resultatene ser vi følgende:

- μ_n er veiet middel av m_n og μ_0
- $\lim_{n \rightarrow \infty} \mu_n = m_n$, dvs. størst tiltro til målingene når treningssettet er stort
- $\mu_n \approx \mu_0$ hvis $\sigma_0 \ll \sigma$ dvs. a priori kunnskap betyr mest hvis den er god (liten σ_0)
- $\mu_n \approx m_n$ hvis $\sigma_0 \gg \sigma$ dvs. nye målinger betyr mest hvis lite a priori kunnskap (stor σ_0)
- $\sigma_n \approx 0$ hvis $\sigma \ll 1$ dvs. liten usikkerhet i estimatet av skarp fordeling
- $\lim_{n \rightarrow \infty} \sigma_n = 0, \sigma_n \approx \frac{\sigma}{\sqrt{n}}$ når $n \gg 1$ dvs. Bayesisk læring.

Bayesisk estimering av forventningsverdien i en univariat normalfordeling (forts.)

Tettheten i punktet x kan nå beregnes

$$\begin{aligned} p(x|\mathcal{X}) &= \int p(x|\mu)p(\mu|\mathcal{X})d\mu \\ &= \int \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] \frac{1}{\sqrt{2\pi}\sigma_n} \exp\left[-\frac{1}{2}\left(\frac{\mu-\mu_n}{\sigma_n}\right)^2\right] d\mu \\ &\vdots \text{ (lang regning)} \\ &= \frac{1}{\sqrt{2\pi(\sigma^2 + \sigma_n^2)}} \exp\left[-\frac{(x - \mu_n)^2}{2(\sigma^2 + \sigma_n^2)}\right] \\ &= \underline{\underline{N(\mu_n, \sigma^2 + \sigma_n^2)}}. \end{aligned}$$

Bayesisk estimering av forventningsverdien i en univariat normalfordeling (forts.)

Den estimerte fordelingen for x blir en normalfordeling med forventning lik veiet middel av m_n og μ_0 og varians lik den kjente variansen for fordelingen med tillegg av usikkerheten i estimatet av forventningsverdien.

Siden

$$\sigma_n \rightarrow 0$$

og

$$\mu_n \rightarrow m_n$$

når $n \rightarrow \infty$ (det asymptotiske tilfellet) vil

$$p(x|\mathcal{X}) \rightarrow N(m_n, \sigma^2).$$

Dette betyr at Bayes løsning \rightarrow maksimum likelihoodløsningen når $n \rightarrow \infty$.

Rekursiv Bayesisk estimering

Skal her betrakte Bayesisk estimering som en gradvis prosess der treningssamplene introduseres ett og ett om gangen. Innfører her følgende notasjon for treningssettet

$$\mathcal{X}^n = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\} \quad \text{dvs. et treningssett med } n \text{ sampler, der } n > 1.$$

Kan da skrive likelihoodfunksjonen som

$$p(\mathcal{X}^n | \boldsymbol{\theta}) = p(\mathbf{x}_n | \boldsymbol{\theta}) \prod_{k=1}^{n-1} p(\mathbf{x}_k | \boldsymbol{\theta}) = p(\mathbf{x}_n | \boldsymbol{\theta}) p(\mathcal{X}^{n-1} | \boldsymbol{\theta})$$

Dette gir

$$p(\boldsymbol{\theta} | \mathcal{X}^n) = \frac{p(\mathcal{X}^n | \boldsymbol{\theta}) p(\boldsymbol{\theta})}{\int p(\mathcal{X}^n | \boldsymbol{\theta}) p(\boldsymbol{\theta}) d\boldsymbol{\theta}} = \frac{p(\mathbf{x}_n | \boldsymbol{\theta}) p(\mathcal{X}^{n-1} | \boldsymbol{\theta}) p(\boldsymbol{\theta})}{\int p(\mathbf{x}_n | \boldsymbol{\theta}) p(\mathcal{X}^{n-1} | \boldsymbol{\theta}) p(\boldsymbol{\theta}) d\boldsymbol{\theta}} \quad (I).$$

Rekursiv Bayesisk estimering (forts.)

Kan tilsvarende skrive:

$$p(\boldsymbol{\theta}|\mathcal{X}^{n-1}) = \frac{p(\mathcal{X}^{n-1}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{\int p(\mathcal{X}^{n-1}|\boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta}} \quad (II),$$

for et treningssett med $n-1$ sampler.

Innsetting av (II) i (I) gir da

$$\begin{aligned} p(\boldsymbol{\theta}|\mathcal{X}^n) &= \frac{p(\mathbf{x}_n|\boldsymbol{\theta})p(\boldsymbol{\theta}|\mathcal{X}^{n-1}) \int p(\mathcal{X}^{n-1}|\boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta}}{\int p(\mathbf{x}_n|\boldsymbol{\theta})p(\boldsymbol{\theta}|\mathcal{X}^{n-1})[\int p(\mathcal{X}^{n-1}|\boldsymbol{\theta}')p(\boldsymbol{\theta}')d\boldsymbol{\theta}']d\boldsymbol{\theta}} \\ &= \frac{p(\mathbf{x}_n|\boldsymbol{\theta})p(\boldsymbol{\theta}|\mathcal{X}^{n-1})}{\int p(\mathbf{x}_n|\boldsymbol{\theta})p(\boldsymbol{\theta}|\mathcal{X}^{n-1})d\boldsymbol{\theta}} \quad (III). \end{aligned}$$

Rekursiv Bayesisk estimering (forts.)

Definerer nå

$$p(\boldsymbol{\theta}|\mathcal{X}^0) = p(\boldsymbol{\theta}) \quad \text{der } \mathcal{X}^0 \text{ er et tomt treningssett}$$

slik at

$$p(\boldsymbol{\theta}), p(\boldsymbol{\theta}|\mathbf{x}_1), p(\boldsymbol{\theta}|\mathbf{x}_1, \mathbf{x}_2), \dots, p(\boldsymbol{\theta}|\mathcal{X}^n), \dots$$

er en følge av fordelinger som kan beregnes rekursivt ved hjelp av likning (III).

Typisk vil a posteriorifordelingen bli skarpere og skarpere når antall sampler øker.

Bayesisk læring betyr at $p(\boldsymbol{\theta}|\mathcal{X}^n) \rightarrow \delta(\boldsymbol{\theta} - \boldsymbol{\theta}_0)$ når $n \rightarrow \infty$.

Dette forutsetter at fordelingen er *identifiserbar*. Med dette menes at parametervektoren $\boldsymbol{\theta}$ er unikt bestemt av fordelingen $p(\mathbf{x}|\boldsymbol{\theta})$. De fleste kontinuerlige fordelinger er identifiserbare, mens mange diskrete fordelinger ikke er det. Mangel på identifiserbarhet er ikke noe problem i ledet læring, siden vi her midler over alle mulige parameterverdier for én og samme klasse. Det er imidlertid et fundamentalt problem innen ikke-ledet læring (mer om dette senere).

Estimering av forventning til univariat normalfordeling (Eksempel)

Skal se på et eksempel der ukjent tetthetsfunksjon og a priori parameterfordeling er hhv.

$p(x|\mu) = N(\mu, \sigma^2)$ der σ antas kjent og μ er ukjent, og

$p(\mu) = N(\mu_0, \sigma_0^2)$ der μ_0 og σ_0 er kjente.

A posteriori parameterfordeling blir da

$$p(\mu|\mathcal{X}) = N(\mu_n, \sigma_n^2).$$

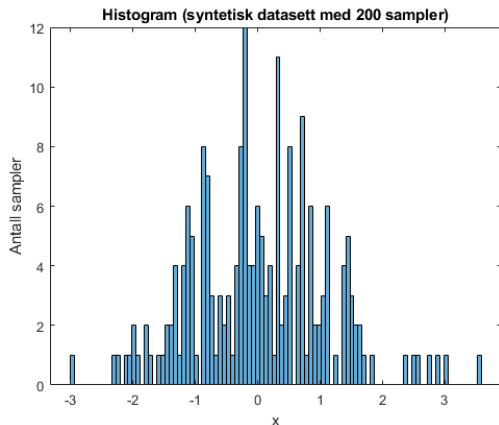
Her er estimatene av forventning og varians

$$\mu_n = \frac{n\sigma_0^2}{n\sigma_0^2 + \sigma^2} m_n + \frac{\sigma^2}{n\sigma_0^2 + \sigma^2} \mu_0 \quad \text{og} \quad \sigma_n^2 = \frac{\sigma^2 \sigma_0^2}{n\sigma_0^2 + \sigma^2},$$

der m_n er sampelmiddelet over treningssettet. Skal ut fra dette beregne det Bayesiske tetthetsestimatet

$$p(x|\mathcal{X}) = N(\mu_n, \sigma^2 + \sigma_n^2).$$

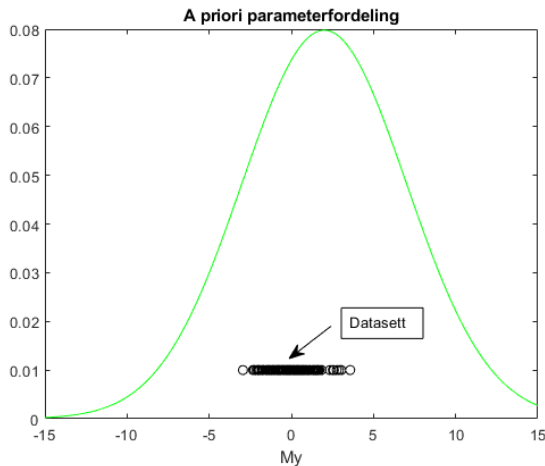
Estimering av forventning til univariat normalfordeling (forts.)



Histogram over datasettet (egenskap x)

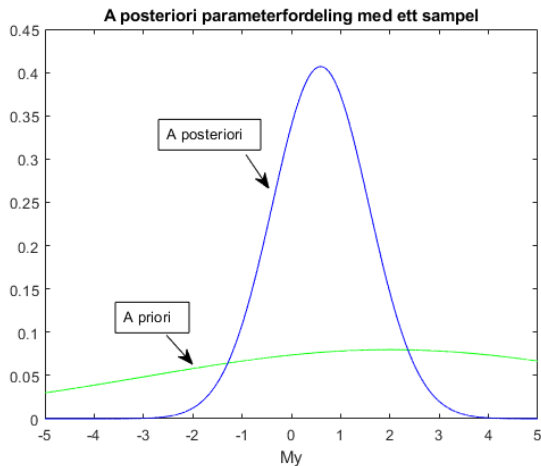
- Univariat datasett med 200 objekter, trukket fra normalfordeling med $\mu = 0$ og $\sigma = 1$ (samme som i forelesning 4).
- Antar normalfordeling med ukjent μ og kjent σ , dvs. $\sigma = 1$.
- Skal plote a posteriori parameterfordeling og finne μ_n og σ_n som funksjon av antall sampler n i treningssettet.

Estimering av forventning til univariat normalfordeling (forts.)



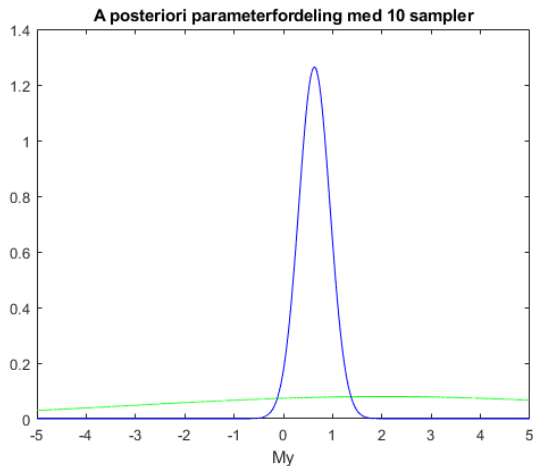
A priori parameterfordeling (for samme datasett som i forelesning 4). Her er det valgt en fordeling $N(\mu_0, \sigma_0^2)$ med $\mu_0 = 2,0$ og $\sigma_0 = 5,0$.

Estimering av forventning til univariat normalfordeling (forts.)



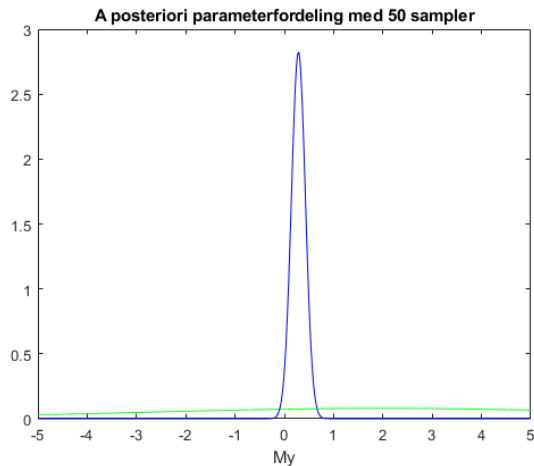
A posteriori parameterfordeling $p(\mu|\mathcal{X}^1)$ (blå kurve) og a priori fordeling (grønn kurve). Maksimum for $\mu = 0,594$.

Estimering av forventning til univariat normalfordeling (forts.)



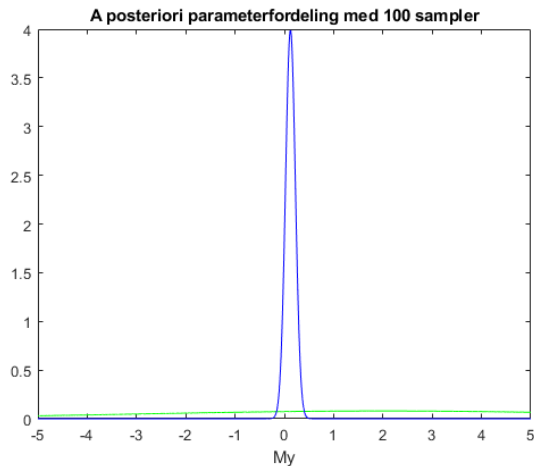
A posteriori parameterfordeling $p(\mu | \mathcal{X}^{10})$ (blå kurve) og a priori fordeling (grønn kurve). Maksimum for $\mu = 0,630$.

Estimering av forventning til univariat normalfordeling (forts.)



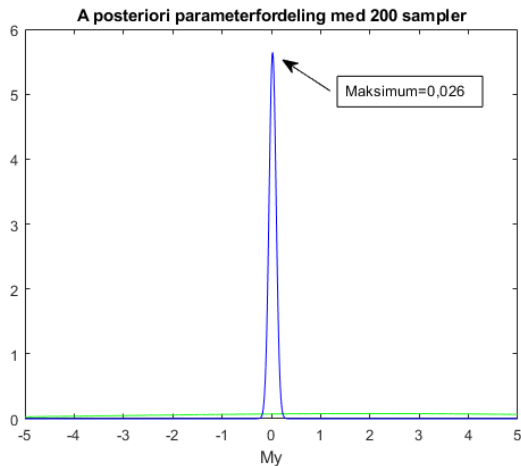
A posteriori parameterfordeling $p(\mu | \mathcal{X}^{50})$ (blå kurve) og a priori fordeling (grønn kurve). Maksimum for $\mu = 0,285$.

Estimering av forventning til univariat normalfordeling (forts.)



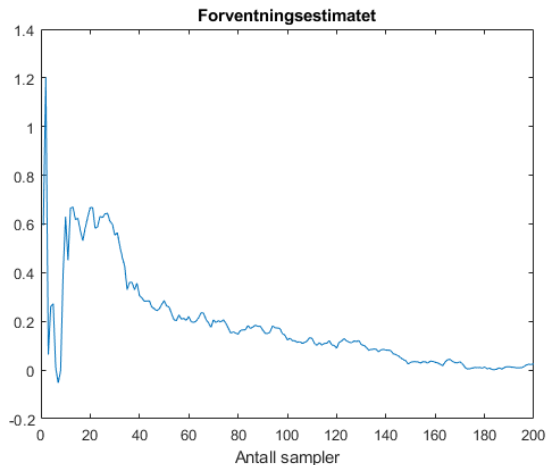
A posteriori parameterfordeling $p(\mu | \mathcal{X}^{100})$ (blå kurve) og a priori fordeling (grønn kurve). Maksimum for $\mu = 0,124$.

Estimering av forventning til univariat normalfordeling (forts.)



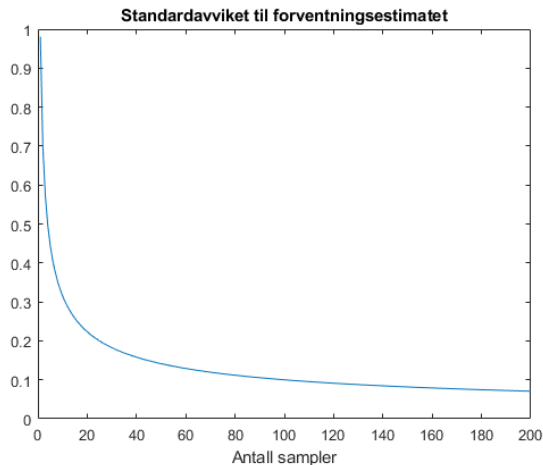
A posteriori parameterfordeling $p(\mu | \mathcal{X}^{200})$ (blå kurve) og a priori fordeling (grønn kurve). Maksimum for $\mu = 0,026$.

Estimering av forventning til univariat normalfordeling (forts.)



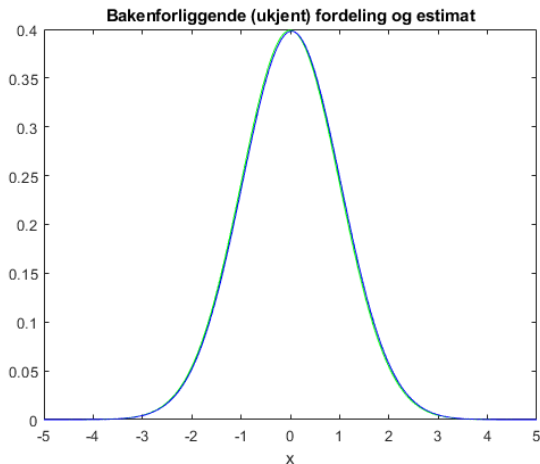
Forventningsestimatet μ_n som funksjon av antall sampler.

Estimering av forventning til univariat normalfordeling (forts.)



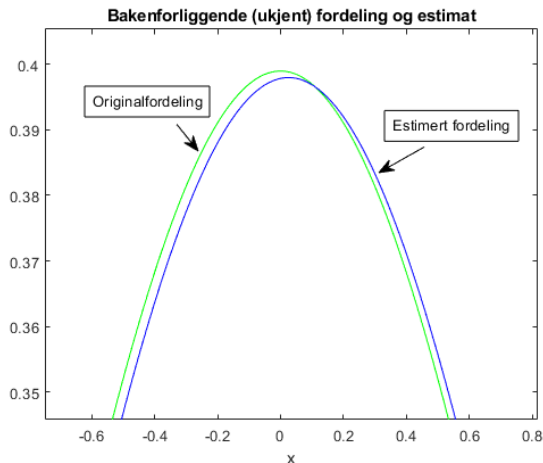
Standardavviket σ_n til forventningsestimatet som funksjon av antall sampler.

Estimering av forventning til univariat normalfordeling (forts.)



Fasit (ukjent fordeling) og tetthetsestimatet (kurvene overlapper).

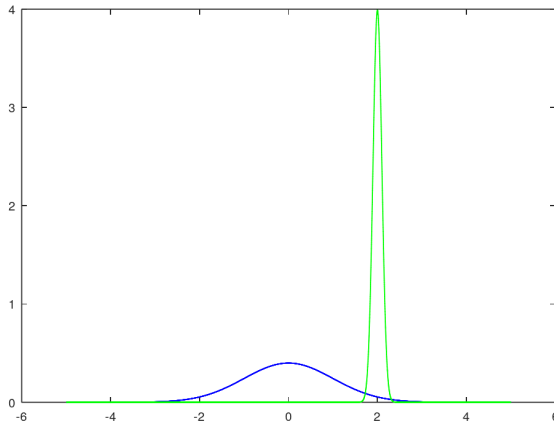
Estimering av forventning til univariat normalfordeling (forts.)



Fasit (ukjent fordeling) og tetthetsestimatet (forstørret utsnitt).

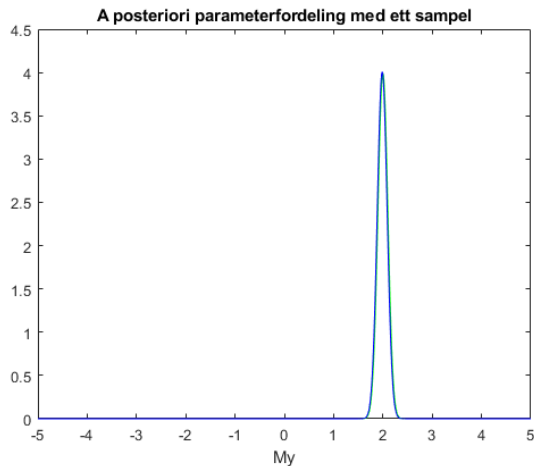
Estimering av forventning til univariat normalfordeling (Eksempel)

Skarpere a priori parameterfordeling



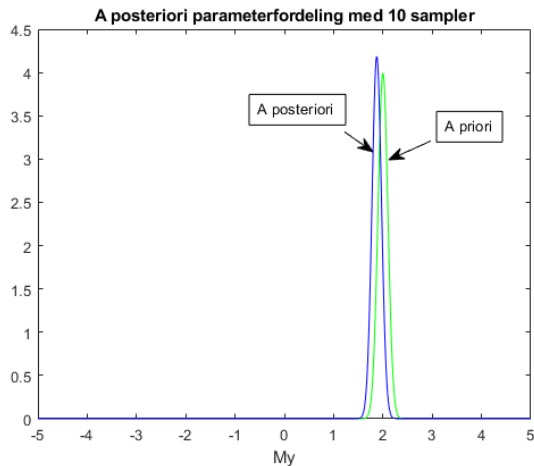
La nå a priori parameterfordeling være gitt ved $N(\mu_0, \sigma_0^2)$ med $\mu_0 = 2,0$ og $\sigma_0 = 0,1$ (grønn kurve). Til sammenlikning den ukjente fordelingen (blå kurve).

Estimering av forventning til univariat normalfordeling (forts.)



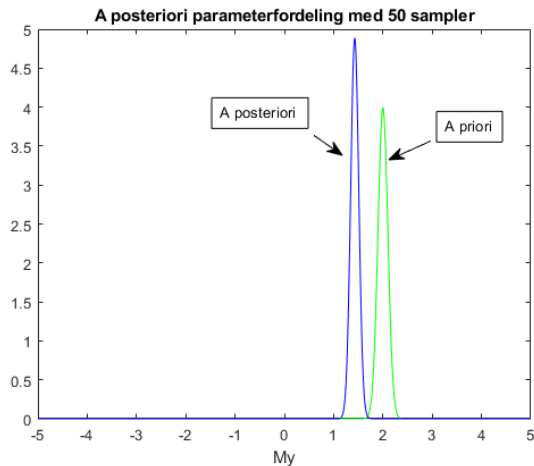
A posteriori parameterfordeling $p(\mu|\mathcal{X}^1)$ og a priori fordeling (kurvene overlapper).
Maksimum for $\mu = 1,986$.

Estimering av forventning til univariat normalfordeling (forts.)



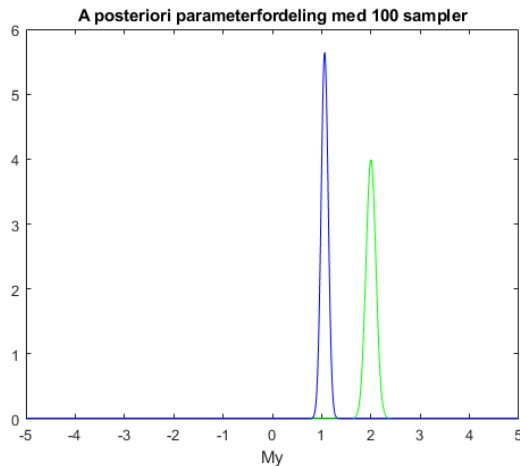
A posteriori parameterfordeling $p(\mu | \mathcal{X}^{10})$ (blå kurve) og a priori fordeling (grønn kurve). Maksimum for $\mu = 1,875$.

Estimering av forventning til univariat normalfordeling (forts.)



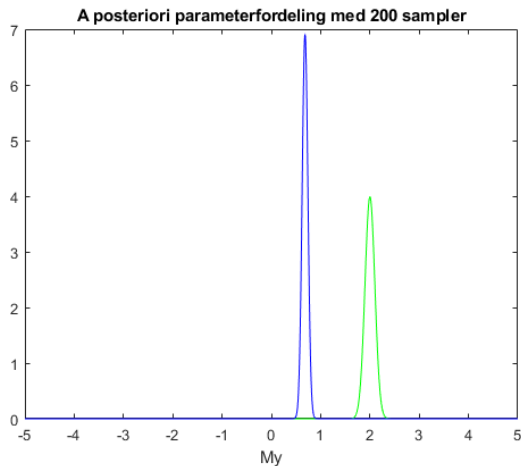
A posteriori parameterfordeling $p(\mu | \mathcal{X}^{50})$ (blå kurve) og a priori fordeling (grønn kurve). Maksimum for $\mu = 1,428$.

Estimering av forventning til univariat normalfordeling (forts.)



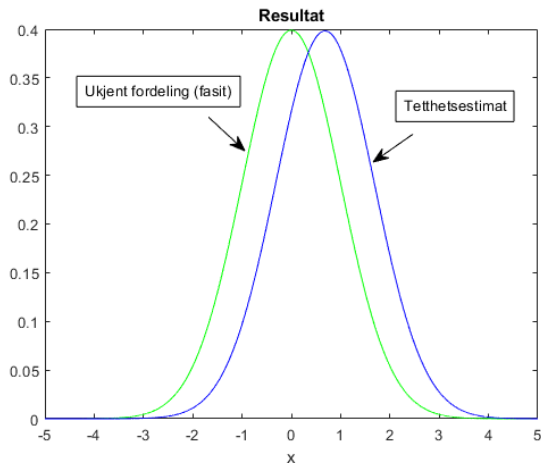
A posteriori parameterfordeling $p(\mu | \mathcal{X}^{100})$ (blå kurve) og a priori fordeling (grønn kurve). Maksimum for $\mu = 1,062$.

Estimering av forventning til univariat normalfordeling (forts.)



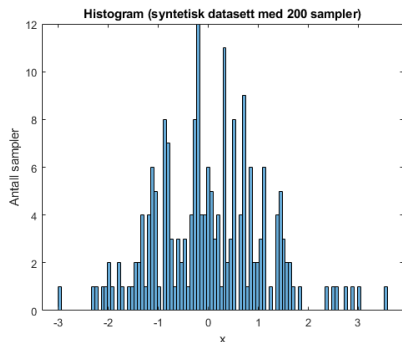
A posteriori parameterfordeling $p(\mu | \mathcal{X}^{200})$ (blå kurve) og a priori fordeling (grønn kurve). Maksimum for $\mu = 0,683$.

Estimering av forventning til univariat normalfordeling (forts.)



Ukjent fordeling (fasit) og tetthetsestimatet (normalfordeling med parametrene $\mu = 0,6835$ og $\sigma = 1,0017$).

Estimering av forventning og varians til univariat normalfordeling (Eksempel)



Univariat datasett med 200
sampler trukket fra normalfordeling
med $\mu = 0$ og $\sigma = 1$.

- Antar normalfordeling der *både* μ og σ er ukjente!
- Samme syntetiske datasett som tidligere.
- Antar a priori parameterfordeling:

$$p(\mu, \sigma) = \frac{1}{2\pi\sigma_1\sigma_2} \exp \left[-\frac{1}{2} \left(\frac{(\mu - \mu_0)^2}{\sigma_1^2} + \frac{(\sigma - \sigma_0)^2}{\sigma_2^2} \right) \right]$$

der

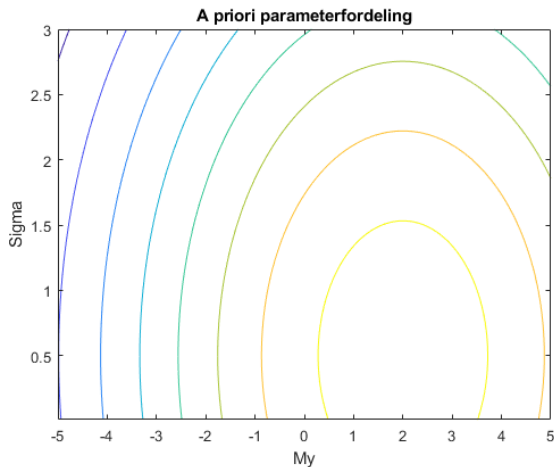
$\mu_0 = 2$ (beste gjetning om forventningsverdi)

$\sigma_0 = 0,5$ (usikkerhet i denne gjetningen)

$\sigma_1 = 5$ (usikkerhet i forventningsestimat)

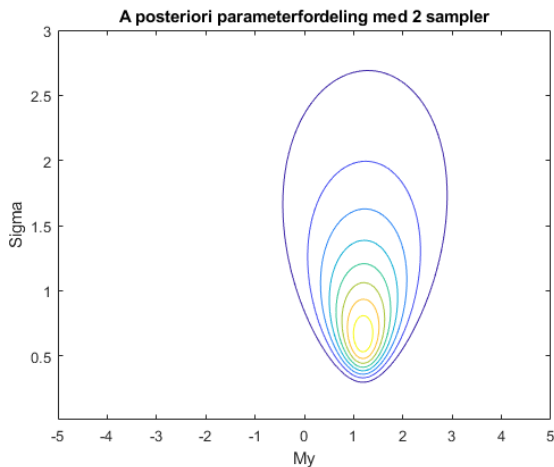
$\sigma_2 = 3$ (usikkerhet i estimat av standardavvik).

Estimering av forventning og varians til univariat normalfordeling (forts.)



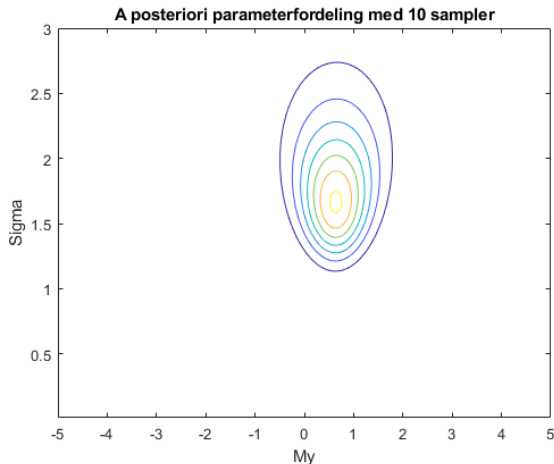
Konturplott av a priori parameterfordeling med maksimum for $\mu = 2$ og $\sigma = 0,5$.

Estimering av forventning og varians til univariat normalfordeling (forts.)



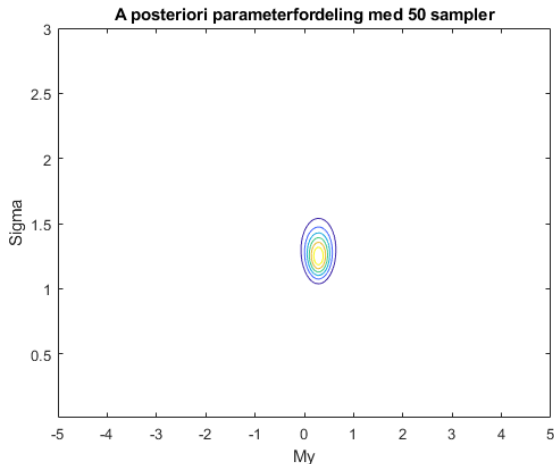
Konturplott av a posteriori parameterfordeling $p(\mu|\mathcal{X}^2)$. Maksimum for $\mu = 1,19$ og $\sigma = 0,65$.

Estimering av forventning og varians til univariat normalfordeling (forts.)



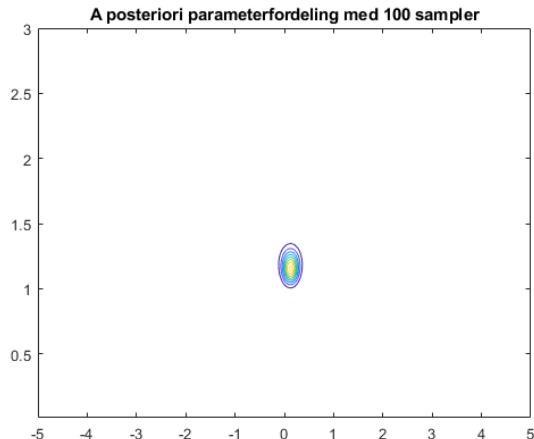
Konturplott av a posteriori parameterfordeling $p(\mu|\mathcal{X}^{10})$. Maksimum for $\mu = 0,64$ og $\sigma = 1,66$.

Estimering av forventning og varians til univariat normalfordeling (forts.)



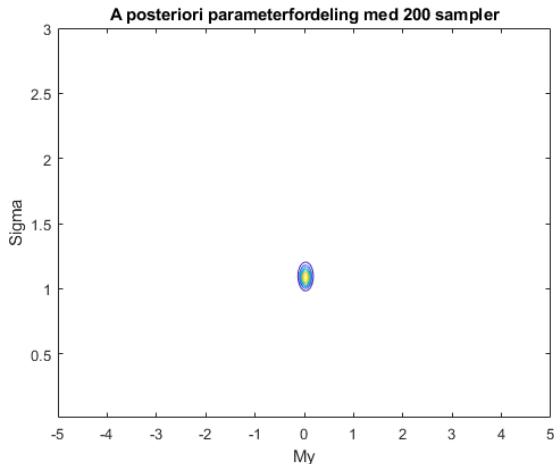
Konturplott av a posteriori parameterfordeling $p(\mu|\mathcal{X}^{50})$. Maksimum for $\mu = 0,29$ og $\sigma = 1,25$.

Estimering av forventning og varians til univariat normalfordeling (forts.)



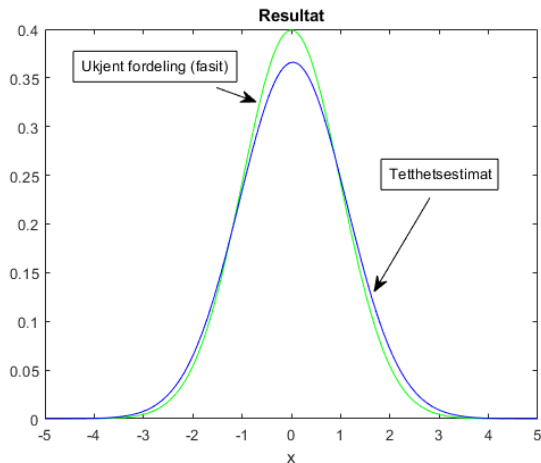
Konturplott av a posteriori parameterfordeling $p(\mu|\mathcal{X}^{100})$. Maksimum for $\mu = 0,12$ og $\sigma = 1,16$.

Estimering av forventning og varians til univariat normalfordeling (forts.)



Konturplott av a posteriori parameterfordeling $p(\mu|\mathcal{X}^{200})$. Maksimum for $\mu = 0,03$ og $\sigma = 1,09$.

Estimering av forventning og varians til univariat normalfordeling (forts.)



Fasit (ukjent fordeling) og tetthetsestimateret (normalfordeling med parametrene $\mu = 0,03$ og $\sigma = 1,09$).

Innhold i kurset

- Introduksjon til mønstergjenkjenning
- Beslutningsteori
- [Parametriske metoder \(fortsetter neste gang\)](#)
- Ikke-parametriske metoder
- Lineære og generaliserte diskriminantfunksjoner
- Evaluering av klassifikatorer
- Ikke-ledet læring
- Klyngeanalyse.