CE 608: Reliability based Structural Design

Jan-May 2016

Dr. Budhaditya Hazra
Room: N-307
Department of Civil Engineering

Ref : Pattern recognition and machine learning
by
Christopher. M. Bishop

Maximum Likelihood

- Data are sampled from a probability distribution p(x ,y)
- The form of the probability distribution p is known but its parameters are unknown
- There is a training set $D = \{(x_1, y_1); \ldots\ldots; (x_m, y_m)\}$ of examples sampled *i.i.d.* according to p(x, y)

Task

Estimate the unknown parameters of p from training data $\mathcal{D}$.

IID sampling

- Independent: each example is sampled independently from others
- Identically distributed: all examples are sampled from the same distribution
- The joint probability over D decomposes into a product as examples are i.i.d (thus independent of each other given the distribution)

Liklihood estimator

- Training data $D = \{(x_1, y_1); \ldots\ldots; (x_m, y_m)\}$ of *i.i.d.* examples for the target class **y** is available
- Assume the parameter vector (θ) has a fixed but unknown value
- Estimate of θ: Maximize its **likelihood** with respect to the training data

$$\theta^* = \mathrm{argmax}_\theta p(\mathcal{D}|\theta) = \mathrm{argmax}_\theta \prod_{j=1}^{n} p(\mathbf{x}_j|\theta)$$

Gaussian with known sigma

- the log-likelihood is:

$$\sum_{j=1}^{n} \ln p(\mathbf{x}_j|\theta) = \sum_{j=1}^{n} -\frac{1}{2}(\mathbf{x}_j - \mu)^t \Sigma^{-1} (\mathbf{x}_j - \mu) - \frac{1}{2} \ln (2\pi)^d |\Sigma|$$

- The gradient wrt to the mean is:

$$\nabla_\mu \sum_{j=1}^{n} \ln p(\mathbf{x}_j|\theta) = \sum_{j=1}^{n} \Sigma^{-1} (\mathbf{x}_j - \mu)$$

- Setting the gradient to zero gives:

$$\sum_{j=1}^{n} \Sigma^{-1} (\mathbf{x}_j - \mu^*) = \mathbf{0} \qquad \Rightarrow \qquad \mu^* = \frac{1}{n} \sum_{j=1}^{n} \mathbf{x}_j$$

Gaussian with unknown mean & sigma

- the log-likelihood is:

$$\mathcal{L} = \sum_{j=1}^{n} -\frac{1}{2\sigma^2}(x_j - \mu)^2 - \frac{1}{2}\ln 2\pi\sigma^2$$

- The gradient is:

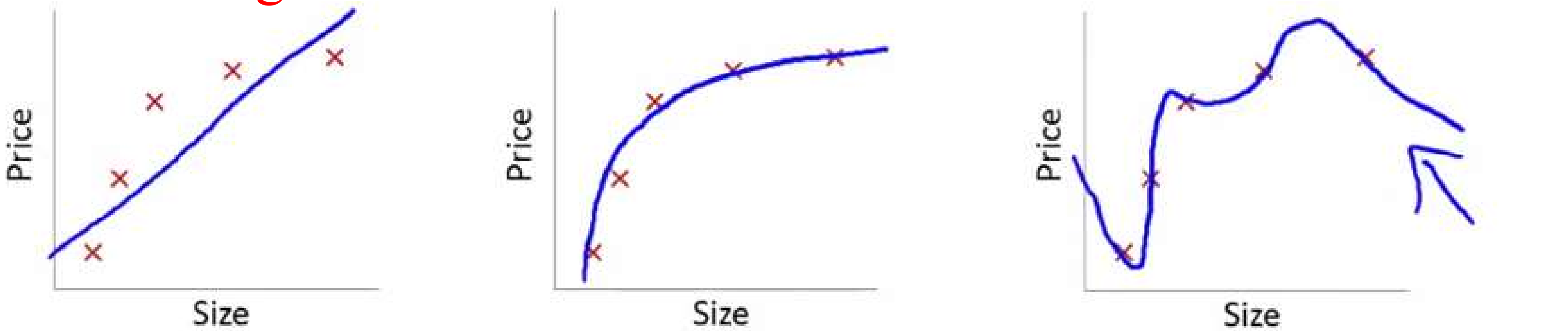
$$\nabla_{\mu,\sigma^2}\mathcal{L} = \begin{bmatrix} \sum_{j=1}^{n} \frac{1}{\sigma^2}(x_j - \mu) \\ \sum_{j=1}^{n} -\frac{1}{2\sigma^2} + \frac{(x_j-\mu)^2}{2\sigma^4} \end{bmatrix} = 0$$

$$\mu^* = \frac{1}{n}\sum_{j=1}^{n} x_j \qquad \sigma^2 = \frac{1}{n}\sum_{j=1}^{n}(x_j - \mu^*)^2$$

Question: Work out the case where sigma is known and varies at each point

Issues with Maximum Likelihood

Overfitting !



$\rightarrow \theta_0 + \theta_1 x$ $\rightarrow \theta_0 + \theta_1 x + \theta_2 x^2$ $\rightarrow \theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3 + \theta_4 x^4$

Accurate in training but poor in prediction

Overfitting with Maximum Likelihood

Given: $\mathcal{D} = \{x_1, \ldots, x_N\}$, m heads (1), $N - m$ tails (0)

$$p(\mathcal{D}|\mu) = \prod_{n=1}^{N} p(x_n|\mu) = \prod_{n=1}^{N} \mu^{x_n}(1-\mu)^{1-x_n}$$

$$\ln p(\mathcal{D}|\mu) = \sum_{n=1}^{N} \ln p(x_n|\mu) = \sum_{n=1}^{N} \{x_n \ln \mu + (1-x_n)\ln(1-\mu)\}$$

$$\mu_{\text{ML}} = \frac{1}{N}\sum_{n=1}^{N} x_n = \frac{m}{N}$$

Suppose we fit a MLE with **3** observations of a Bernoulli trial

$$\mathcal{D} = \{1, 1, 1\} \rightarrow \mu_{\text{ML}} = \frac{3}{3} = 1$$

Prediction: *all* future tosses will land heads up

Bayesian update

$$p(\mathbf{x}|\mathcal{D}) = \int_{\theta} p(\mathbf{x}, \theta|\mathcal{D}) d\theta = \int p(\mathbf{x}|\theta) p(\theta|\mathcal{D}) d\theta$$

- $p(\mathbf{x}|\theta)$ can be easily computed (we have both form and parameters of distribution, e.g. Gaussian)
- need to estimate the parameter posterior density given the training set:

$$p(\theta|\mathcal{D}) = \frac{p(\mathcal{D}|\theta)p(\theta)}{p(\mathcal{D})}$$

Bayesian update

$$p(\theta|\mathcal{D}) = \frac{p(\mathcal{D}|\theta)p(\theta)}{p(\mathcal{D})}$$

- $p(\mathcal{D})$ is a constant independent of θ (i.e. it will no influence final Bayesian decision)
- if final *probability* (not only decision) is needed we can compute:

$$p(\mathcal{D}) = \int_{\theta} p(\mathcal{D}|\theta)p(\theta)d\theta$$

Example: Bernoulli trial

$$p(\mu|a_0, b_0, \mathcal{D}) \propto p(\mathcal{D}|\mu) \, p(\mu|a_0, b_0)$$

Likelihood: $p(\mathcal{D}|\mu)$

Prior: $p(\mu|a_0, b_0)$

$$= \left(\prod_{n=1}^{N} \mu^{x_n} (1-\mu)^{1-x_n} \right) \mathrm{Beta}(\mu|a_0, b_0)$$

$$\propto \mu^{m+a_0-1} (1-\mu)^{(N-m)+b_0-1}$$

$$\propto \mathrm{Beta}(\mu|a_N, b_N)$$

$$a_N = a_0 + m \qquad b_N = b_0 + (N-m)$$

Bernoulli trial: Posterior

As N increases

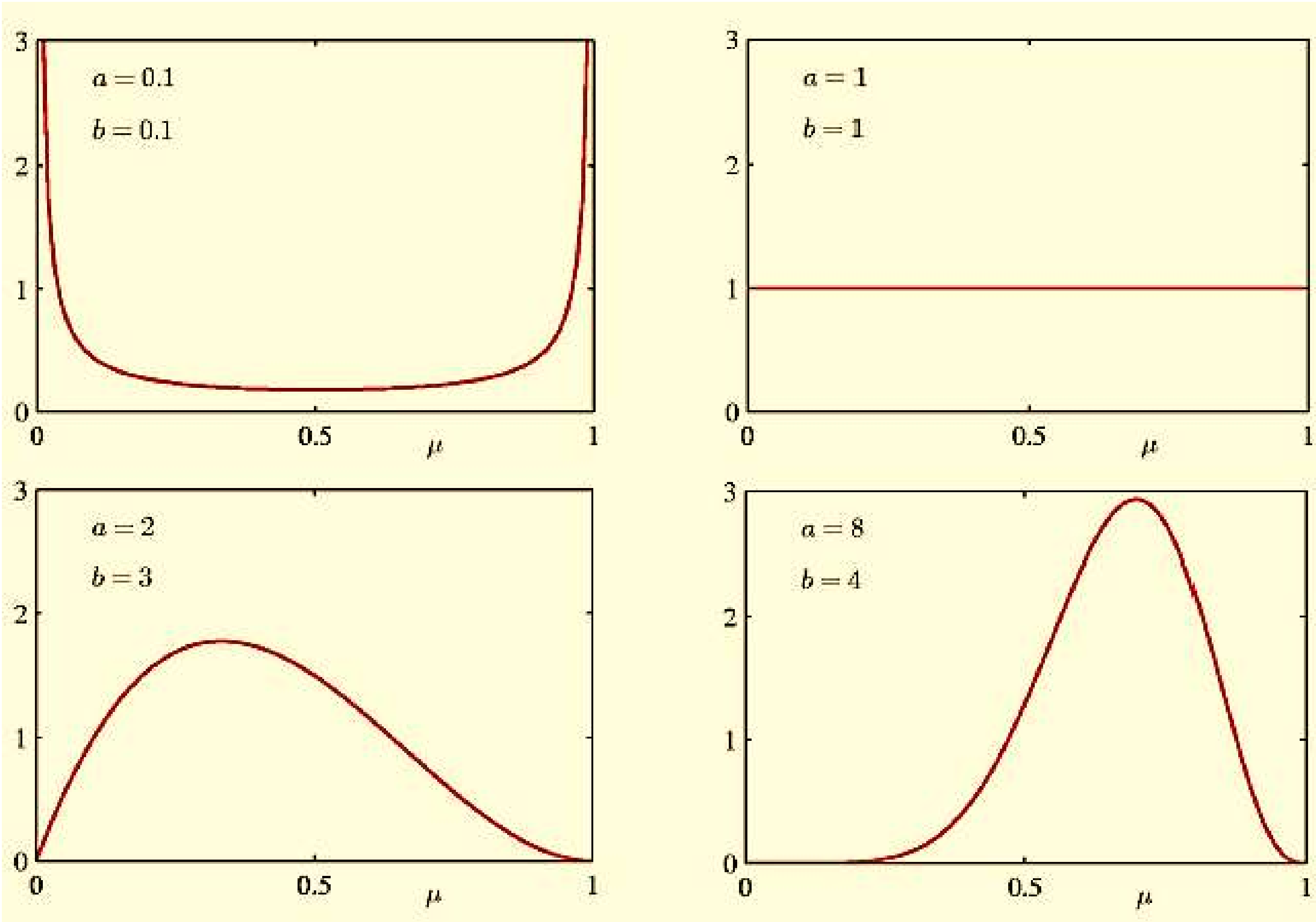
$$
\begin{aligned}
a_N &\rightarrow m \\
b_N &\rightarrow N - m \\
\mathbb{E}[\mu] &= \frac{a_N}{a_N + b_N} \rightarrow \frac{m}{N} = \mu_{\mathrm{ML}} \\
\mathrm{var}[\mu] &= \frac{a_N b_N}{(a_N + b_N)^2 (a_N + b_N + 1)} \rightarrow 0
\end{aligned}
$$

Prediction

What is the probability that the next coin toss will land heads up?

$$
\begin{aligned}
p(x=1|a_0,b_0,\mathcal{D}) &= \int_0^1 p(x=1|\mu)p(\mu|a_0,b_0,\mathcal{D})\,\mathrm{d}\mu \\
&= \int_0^1 \mu p(\mu|a_0,b_0,\mathcal{D})\,\mathrm{d}\mu \\
&= \mathbb{E}[\mu|a_0,b_0,\mathcal{D}] = \frac{a_N}{b_N}
\end{aligned}
$$

Prior: Beta distribution



Example: Gaussian with known σ

- The likelihood function for μ is given by

$$p(\mathbf{x}|\mu) = \prod_{n=1}^{N} p(x_n|\mu) = \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left\{ -\frac{1}{2\sigma^2} \sum_{n=1}^{N} (x_n - \mu)^2 \right\}$$

- This has a Gaussian shape as a function of μ (but it is *not* a distribution over μ)

- Combine with a Gaussian prior over μ

$$p(\mu|\mathbf{x}) \propto p(\mathbf{x}|\mu)p(\mu)$$

Example: Gaussian with known σ

This yields

$$p(\mu|\mathbf{x}) = \mathcal{N}\left(\mu|\mu_N, \sigma_N^2\right)$$

where

$$\mu_N = \frac{\sigma^2}{N\sigma_0^2 + \sigma^2}\mu_0 + \frac{N\sigma_0^2}{N\sigma_0^2 + \sigma^2}\mu_{\mathrm{ML}}, \qquad \mu_{\mathrm{ML}} = \frac{1}{N}\sum_{n=1}^{N} x_n$$

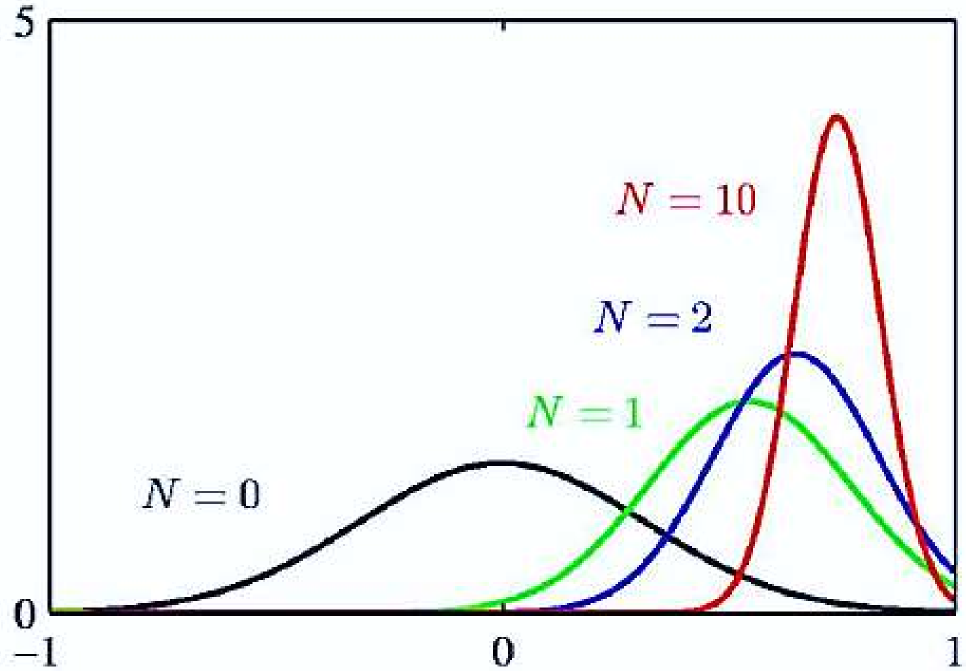
$$\frac{1}{\sigma_N^2} = \frac{1}{\sigma_0^2} + \frac{N}{\sigma^2}.$$

Question: Prove these results ?

	$N = 0$	$N \rightarrow \infty$
μ_N	μ_0	μ_{ML}
σ_N^2	σ_0^2	0

Example: Gaussian with known σ

Ex: $p(\mu|\mathbf{x}) = \mathcal{N}(\mu|\mu_N, \sigma_N^2)$ for $N = 0,\ 1,\ 2$ and 10



Question: Check this plot using a simple MATLAB code. What happens when N is 100 ?

Example: Gaussian with known σ

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

Example: Gaussian with known μ

- The likelihood function for $\lambda = \frac{1}{\sigma^2}$ is given by

$$p(\mathbf{x}|\lambda) = \prod_{n=1}^{N} \mathcal{N}(x_n|\mu, \lambda^{-1}) \propto \lambda^{N/2} \exp\left\{-\frac{\lambda}{2}\sum_{n=1}^{N}(x_n - \mu)^2\right\}$$

- This has a Gamma shape as a function of λ

Example: Gaussian with known μ

Prior → $\mathrm{Gam}(\lambda|a_0, b_0)$.

Posterior

$$p(\lambda|\mathbf{x}) \propto \lambda^{a_0-1}\lambda^{N/2}\exp\left\{-b_0\lambda - \frac{\lambda}{2}\sum_{n=1}^{N}(x_n-\mu)^2\right\}$$

Recognize the posterior becomes :

$$\mathrm{Gam}(\lambda|a_N, b_N)$$

$$a_N = a_0 + \frac{N}{2}$$

$$b_N = b_0 + \frac{1}{2}\sum_{n=1}^{N}(x_n-\mu)^2 = b_0 + \frac{N}{2}\sigma^2_{\mathrm{ML}}$$

Question: Prove these results ?

Gaussian with μ and σ unknown

$$p(\mathbf{x}|\mu,\lambda) = \prod_{n=1}^{N} \left(\frac{\lambda}{2\pi}\right)^{1/2} \exp\left\{-\frac{\lambda}{2}(x_n-\mu)^2\right\}$$

$$\propto \left[\lambda^{1/2}\exp\left(-\frac{\lambda\mu^2}{2}\right)\right]^N \exp\left\{\lambda\mu\sum_{n=1}^{N}x_n - \frac{\lambda}{2}\sum_{n=1}^{N}x_n^2\right\}$$

Gauss-Gamma likelihood

$$p(\mu,\lambda) = \mathcal{N}(\mu|\mu_0,(\beta\lambda)^{-1})\mathrm{Gam}(\lambda|a,b)$$

$$\propto \underbrace{\exp\left\{-\frac{\beta\lambda}{2}(\mu-\mu_0)^2\right\}}\ \underbrace{\lambda^{a-1}\exp\{-b\lambda\}}$$

- Quadratic in μ.
- Linear in λ.

- Gamma distribution over λ.
- Independent of μ.

Question: Prove this result ?

Conjugate Priors

- Given a likelihood function $p(x|\theta)$
- Given a prior distribution $p(\theta)$
- $p(\theta)$ is a *conjugate prior* for $p(x|\theta)$ if the posterior distribution $p(\theta|x)$ is in the same family as the prior $p(\theta)$

PRIOR	LIKELIHOOD	POSTERIOR
NORMAL	NORMAL	NORMAL
BETA	BINOMIAL	BETA
GAMMA	POISSON	GAMMA
GAMMA	EXPONENTIAL	GAMMA

Gaussian with known σ: Details

Likelihood

Let $D = (x_1, \ldots, x_n)$ be the data.

$$p(D|\mu, \sigma^2) = \prod_{i=1}^{n} p(x_i|\mu, \sigma^2) = (2\pi\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2\right\}$$

Define sample mean and variance

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2$$

$$\sum_i (x_i - \mu)^2 = \sum_i [(x_i - \overline{x}) - (\mu - \overline{x})]^2 = ns^2 + n(\overline{x} - \mu)^2$$

$$\sum_i (x_i - \overline{x})(\mu - \overline{x}) = (\mu - \overline{x})\left((\sum_i x_i) - n\overline{x}\right) = (\mu - \overline{x})(n\overline{x} - n\overline{x}) = 0$$

$$\begin{aligned} p(D|\mu, \sigma^2) &= \frac{1}{(2\pi)^{n/2}} \frac{1}{\sigma^n} \exp\left(-\frac{1}{2\sigma^2}\left[ns^2 + n(\overline{x} - \mu)^2\right]\right) \\ &\propto \left(\frac{1}{\sigma^2}\right)^{n/2} \exp\left(-\frac{n}{2\sigma^2}(\overline{x} - \mu)^2\right) \exp\left(-\frac{ns^2}{2\sigma^2}\right) \end{aligned}$$

If σ^2 is const, we can further write as:

$$p(D|\mu) \propto \exp\left(-\frac{n}{2\sigma^2}(\overline{x} - \mu)^2\right) \propto \mathcal{N}(\overline{x}|\mu, \frac{\sigma^2}{n})$$

Prior

Since the likelihood has the form

$$p(D|\mu) \propto \exp\left(-\frac{n}{2\sigma^2}(\overline{x}-\mu)^2\right) \propto \mathcal{N}(\overline{x}|\mu, \frac{\sigma^2}{n})$$

the **natural conjugate prior** has the form

$$p(\mu) \propto \exp\left(-\frac{1}{2\sigma_0^2}(\mu-\mu_0)^2\right) \propto \mathcal{N}(\mu|\mu_0, \sigma_0^2)$$

Do not confuse σ_o^2, which is the variance of the prior, with σ^2

Posterior

$$
\begin{aligned}
p(\mu|D) &\propto p(D|\mu,\sigma)p(\mu|\mu_0,\sigma_0^2) \\
&\propto \exp\left[-\frac{1}{2\sigma^2}\sum_i (x_i-\mu)^2\right] \times \exp\left[-\frac{1}{2\sigma_0^2}(\mu-\mu_0)^2\right] \\
&= \exp\left[\frac{-1}{2\sigma^2}\sum_i (x_i^2+\mu^2-2x_i\mu) + \frac{-1}{2\sigma_0^2}(\mu^2+\mu_0^2-2\mu_0\mu)\right]
\end{aligned}
$$

Since the product of two Gaussians is a Gaussian, we will rewrite this in the form

$$
\begin{aligned}
p(\mu|D) &\propto \exp\left[-\frac{\mu^2}{2}\left(\frac{1}{\sigma_0^2}+\frac{n}{\sigma^2}\right) + \mu\left(\frac{\mu_0}{\sigma_0^2}+\frac{\sum_i x_i}{\sigma^2}\right) - \left(\frac{\mu_0^2}{2\sigma_0^2}+\frac{\sum_i x_i^2}{2\sigma^2}\right)\right] \\
&\stackrel{\text{def}}{=} \exp\left[-\frac{1}{2\sigma_n^2}(\mu^2-2\mu\mu_n+\mu_n^2)\right] = \exp\left[-\frac{1}{2\sigma_n^2}(\mu-\mu_n)^2\right]
\end{aligned}
$$

Matching coefficients of μ^2, we find σ_n^2 is given by

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

$$
\mu_n = \frac{\sigma^2}{n\sigma_0^2+\sigma^2}\mu_0 + \frac{n\sigma_0^2}{n\sigma_0^2+\sigma^2}\overline{x} = \sigma_n^2\left(\frac{\mu_0}{\sigma_0^2}+\frac{n\overline{x}}{\sigma^2}\right)
$$

```
n = 20; %Sample size
sigma = 20; % Gaussian distribution with known sigma
load data_gaussian_mu50_sig20.mat  %save it in a file so that the random sequence doesnt change everytime you run it
 % %==========================================================================
mu = 30;    %hyperparameters
tau = 20;

 theta = linspace(-40, 100, 500); dth=(100-(-40))/500;

y1 = normpdf (theta, mean(x),sigma/sqrt(n));  % Liklihood: Derive the formula yourself, sigma known

y2 = normpdf(theta,mu,tau);                % Prior

 postMean = tau^2*mean(x)/(tau^2+sigma^2/n) + sigma^2*mu/n/(tau^2+sigma^2/n); % Using formula

postSD = sqrt(tau^2*sigma^2/n/(tau^2+sigma^2/n)); % Using formula for known sigma

 y3 = normpdf(theta, postMean, postSD); % Posterior

  y_post_Nr=y1.*y2;   % Likelihood x prior

sum=sum(y_post_Nr)*dth; % Denominstor
y_post=y_post_Nr/sum; %Posterior using Bayes rule

 plot(theta,y1,'m:',theta,y2,'k--',theta,y3,'b','linewidth',2); hold on
 plot(theta,y_post,'o-r','linewidth',1); xlim([-30 80])
```

A discrete RV example

A plant decides to replace all old temperature sensors in a nuclear reactor with new sensors that are highly reliable. Because of lack of information, the supervising engineer assumes that the failure rate for the equipment is uniformly distributed between 0 and 6 failures per operating year

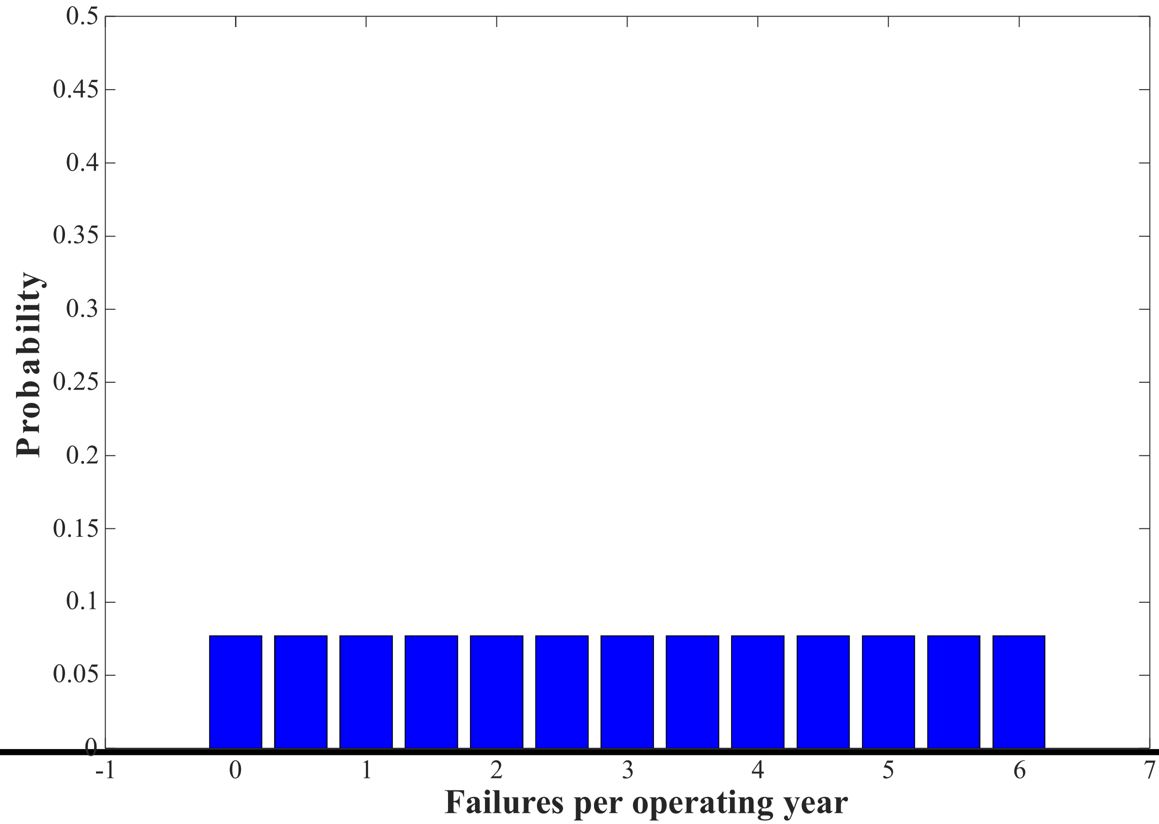
- After 3 years operation, the plant has experienced 5 sensor blips due to the failure of the new temperature sensing system.

- Compute the posterior distribution of the failure rate given the observed failures

Prior: Constant (0 to 6 failures per year discretized coarsely into equal intervals of 0.5 failures per year)

0 to 6 failures per year discretized into equal intervals so p = 1/13 = 0.077

Prior = 0.077 [1 1 1 1 1 1 1 1 1 1 1 1 1]



Likelihood: Poisson process

$$L(D/\lambda) = \frac{(\lambda t)^x}{x!}(e)^{-\lambda t}$$

λ= failure rate = discrete values between 0 to 6 with equal probability

Lambda = [0 0.5 1 1.5 2 2.5 3 3.5 4 4.5 5 5.5 6];

x = 5 and t = 3

Likelihood = (((3*lambda).^5)/factorial(5)). *exp(-3*lambda)

```
%Implementation of the nuclear reactor problem
  prior=0.077*[1 1 1 1 1 1 1 1 1 1 1 1 1];

% 0 to 6 failures per year discretized into equal intervals so p = 1/13 = 0.077

lambda=[0 0.5 1 1.5 2 2.5 3 3.5 4 4.5 5 5.5 6];

% 0 to 6 failures /year discretized into equal  intervals

L1=(((3*lambda).^5)/factorial(5)) ;
L2= exp(-3*lambda);

L=L1.*L2;                          %Liklihood

L_p= L.*prior;                   % Posterior

posterior= L_p/sum(L_p);    % Normalized
```

Car-experiment Data

In some simple problems such as the previous normal mean inference example, it is easy to figure out the posterior distribution in a closed form.

In general problems that involve nonconjugate priors, the posterior distributions are difficult or impossible to compute analytically.

We will consider logistic regression as an example.

This example involves an experiment to help model the proportion of cars of various weights that fail a mileage test.

The data include observations of weight, number of cars tested, and number failed.

Car-experiment Data

```
% A set of car weights
weight = [2100 2300 2500 2700 2900 3100 3300 3500 3700 3900 4100 4300]';
weight = (weight-2800)/1000; % recenter and rescale

% The number of cars tested at each weight
total = [48 42 31 34 31 21 23 23 21 16 17 21]';

% The number of cars that have poor mpg performances at each weight
poor = [1 2 0 3 8 8 14 17 19 15 17 21]';
```

Logistic Regression Model

Logistic regression, a special case of a generalized linear model, is appropriate for these data since the response variable is binomial. The logistic regression model can be written as:

$$P = \frac{e^{Xb}}{1 + e^{Xb}}$$

where X is the design matrix and b is the vector containing the model parameters

Car-experiment Data

```
% A set of car weights
weight = [2100 2300 2500 2700 2900 3100 3300 3500 3700 3900 4100 4300]';
weight = (weight-2800)/1000;            % re-center and rescale

% The number of cars tested at each weight
total = [48 42 31 34 31 21 23 23 21 16 17 21]';
 % The number of cars that have poor mpg performances at each weight
poor = [1 2 0 3 8 8 14 17 19 15 17 21]';

 logitp = @(b,x) exp(b(1)+b(2).*x)./(1+exp(b(1)+b(2).*x));

prior1 = @(b1) normpdf(b1,0,20);          % prior for intercept
prior2 = @(b2) normpdf(b2,0,20);          % prior for slope
post = @(b) prod(binopdf(poor,total,logitp(b,weight))) ...        % likelihood
                * prior1(b(1)) * prior2(b(2));                     % priors

                b1 = linspace(-2.5, -1, 50); b2 = linspace(3, 5.5, 50);

                simpost = zeros(50,50);
                    for i = 1:length(b1)
                        for j = 1:length(b2)
                            simpost(i,j) = post([b1(i), b2(j)]);
                            mesh(b2,b1,simpost);   drawnow
                            xlabel('Slope'); ylabel('Intercept');zlabel('Posterior density')
                            view(-110,30)
                        end;
                    end;
```

Y = normpdf (X,MU,SIGMA)

Parameter update vs data update

$$p(\mathbf{x}|\mathcal{D}) = \int_\theta p(\mathbf{x}, \theta|\mathcal{D})d\theta = \int p(\mathbf{x}|\theta)p(\theta|\mathcal{D})d\theta$$

- $p(\mathbf{x}|\theta)$ can be easily computed (we have both form and parameters of distribution, e.g. Gaussian)
- need to estimate the parameter posterior density given the training set:

$$p(\theta|\mathcal{D}) = \frac{p(\mathcal{D}|\theta)p(\theta)}{p(\mathcal{D})}$$

Parameter update vs data update

Gaussian with known σ

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

Case study: 2 parameter update

$f(\mathbf{x})$	$\boldsymbol{\theta}$	$f(\boldsymbol{\theta})$	$f(\boldsymbol{\theta}\vert\mathbf{D})$	$f(\mathbf{x}\vert\mathbf{D})$
$N(\mu, \sigma^2)$ (Normal with known variance)	μ	$N(\mu_\mu, \sigma_\mu^2)$	$N(\mu_{\mu\vert\mathbf{D}}, \sigma_{\mu\vert\mathbf{D}}^2)$, where $\mu_{\mu\vert\mathbf{D}} = \dfrac{\mu_\mu \sigma^2/n + \bar{\mathbf{d}}\sigma_\mu^2}{\sigma^2/n + \sigma_\mu^2}$ $\sigma_{\mu\vert\mathbf{D}} = \sqrt{\dfrac{\sigma_\mu^2 \sigma^2/n}{\sigma^2/n + \sigma_\mu^2}}$	$N(\mu_{\mu\vert\mathbf{D}}, \sigma_{\mu\vert\mathbf{D}}^2 + \sigma^2)$
$MV(\boldsymbol{\mu}, \mathbf{C})$ (Multivariate normal with known covariance matrix)	$\boldsymbol{\mu}$	$MV(\boldsymbol{\mu}_\mu, \mathbf{C}_\mu)$	$MV(\boldsymbol{\mu}_{\mu\vert\mathbf{D}}, \mathbf{C}_{\mu\vert\mathbf{D}})$, where $\boldsymbol{\mu}_{\mu\vert\mathbf{D}} = \left(\mathbf{C}_\mu^{-1} + n\mathbf{C}^{-1}\right)^{-1} \times \left(\mathbf{C}_\mu^{-1}\mu_\mu + n\mathbf{C}^{-1}\bar{\mathbf{d}}\right)$ $\mathbf{C}_{\mu\vert\mathbf{D}} = \left(\mathbf{C}_\mu^{-1} + n\mathbf{C}^{-1}\right)^{-1}$	$MV(\boldsymbol{\mu}_{\mu\vert\mathbf{D}}, \mathbf{C}_{\mu\vert\mathbf{D}} + \mathbf{C})$

*$\mathbf{x}$: basic random variables; $\boldsymbol{\theta}$: the parameters to be updated; $N(\mu, \sigma^2)$: the normal PDF with a mean of μ and a variance of σ^2; $MV(\boldsymbol{\mu}, \mathbf{C})$: the multivariate normal PDF with a mean of $\boldsymbol{\mu}$ and covariance matrix of $\mathbf{C}$; $\bar{\mathbf{d}}$: the sample mean.

Case study: 2 parameter update

Let $\mathbf{x} = \{c, \varphi\}$ with c and φ denoting the cohesion (kPa) and the friction angle (°) of the soil, respectively. Suppose that at a site $\mathbf{x}$ follow the bivariate normal distribution with a mean of $\boldsymbol{\mu}$ and a covariance matrix of $\mathbf{C} = \{16\ 0;\ 0\ 9\}$. The prior knowledge about $\boldsymbol{\mu}$ is that it is normally distributed with a mean of $\boldsymbol{\mu}_{\boldsymbol{\mu}} = \{12, 28\}^T$ and a covariance matrix of $\mathbf{C}_{\boldsymbol{\mu}} = [9\ 0;\ 0\ 8]$.

Known:

Data covariance matrix: C = [16 0;
0 9];

Hyperparameters: μ_μ and C_μ

Unknown:

μ

Case study: 2 parameter update

$$p(x|\mathcal{D}) = \int p(x|\mu)p(\mu|\mathcal{D})d\mu$$

$f(\mathbf{x})$	$\boldsymbol{\theta}$	$f(\boldsymbol{\theta})$	$f(\boldsymbol{\theta}\|\mathbf{D})$	$f(\mathbf{x}\|\mathbf{D})$
MV($\boldsymbol{\mu}$, **C**) (Multivariate normal with known covariance matrix)	$\boldsymbol{\mu}$	MV($\boldsymbol{\mu}_\mu$, $\mathbf{C}_\mu$)	MV($\boldsymbol{\mu}_{\mu\|D}$, $\mathbf{C}_{\mu\|D}$), where $\boldsymbol{\mu}_{\mu\|D} = \left(\mathbf{C}_\mu^{-1} + n\mathbf{C}^{-1}\right)^{-1} \times \left(\mathbf{C}_\mu^{-1}\mu_\mu + n\mathbf{C}^{-1}\bar{\mathbf{d}}\right)$ $\mathbf{C}_{\mu\|D} = \left(\mathbf{C}_\mu^{-1} + n\mathbf{C}^{-1}\right)^{-1}$	MV($\boldsymbol{\mu}_{\mu\|D}$, $\mathbf{C}_{\mu\|D}$+**C**)

$$\boldsymbol{\mu}_x = \boldsymbol{\mu}_\mu = \begin{Bmatrix} 12 \\ 28 \end{Bmatrix} \text{ and } \mathbf{C}_x = \mathbf{C}_\mu + \mathbf{C} = \begin{bmatrix} 25 & 0 \\ 0 & 17 \end{bmatrix}$$

where $\boldsymbol{\mu}_x$ = mean of $\boldsymbol{\mu}$, and $\mathbf{C}_x$ = covariance matrix of **x**.

Case study: 2 parameter update

Suppose n values of $\{c, \varphi\}$ are observed, as shown in Table

Test No.	c (kPa)	φ (°)
1	10.3	34.2
2	12.2	31.1
3	8.5	35.7
4	14.2	30.8

What are we trying to seek here ?

Ans: Parameter update (i.e. mean update).

So what is the type of information we are looking at ?

Case study: 2 parameter update

The sample mean of the measured data in Table is:

$$\bar{\mathbf{d}} = \{11.3 \quad 32.95\}^T$$

posterior mean ($\boldsymbol{\mu}_{\mu|D}$) and posterior covariance matrix ($\mathbf{C}_{\mu|D}$) of $\boldsymbol{\mu}$

$$\boldsymbol{\mu}_{\mu|D} = \left(\mathbf{C}_{\mu}^{-1} + n\mathbf{C}^{-1}\right)^{-1}\left(\mathbf{C}_{\mu}^{-1}\boldsymbol{\mu}_{\mu} + n\mathbf{C}^{-1}\bar{\mathbf{d}}\right)$$

$$= \left(\begin{bmatrix} 9 & 0 \\ 0 & 8 \end{bmatrix}^{-1} + 4\times\begin{bmatrix} 16 & 0 \\ 0 & 9 \end{bmatrix}^{-1}\right)^{-1} \times \left(\begin{bmatrix} 9 & 0 \\ 0 & 8 \end{bmatrix}^{-1} \times \begin{bmatrix} 12 \\ 28 \end{bmatrix} + 4\times\begin{bmatrix} 16 & 0 \\ 0 & 9 \end{bmatrix}^{-1} \times \begin{Bmatrix} 11.3 \\ 32.95 \end{Bmatrix}\right) = \begin{Bmatrix} 11.515 \\ 31.863 \end{Bmatrix}$$

$$\mathbf{C}_{\mu|D} = \left(\mathbf{C}_{\mu}^{-1} + n\mathbf{C}^{-1}\right)^{-1} = \left(\begin{bmatrix} 9 & 0 \\ 0 & 8 \end{bmatrix}^{-1} + 4\times\begin{bmatrix} 16 & 0 \\ 0 & 9 \end{bmatrix}^{-1}\right)^{-1} = \begin{bmatrix} 2.769 & 0 \\ 0 & 1.756 \end{bmatrix}$$

Case study: 2 parameter update

posterior mean ($\boldsymbol{\mu}_{\mu|D}$) and posterior covariance matrix ($\mathbf{C}_{\mu|D}$) of $\boldsymbol{\mu}$

$$\boldsymbol{\mu}_{\mu|D} = \left(\mathbf{C}_{\mu}^{-1} + n\mathbf{C}^{-1}\right)^{-1}\left(\mathbf{C}_{\mu}^{-1}\boldsymbol{\mu}_{\mu} + n\mathbf{C}^{-1}\overline{\mathbf{d}}\right)$$

$$= \left(\begin{bmatrix} 9 & 0 \\ 0 & 8 \end{bmatrix}^{-1} + 4\times\begin{bmatrix} 16 & 0 \\ 0 & 9 \end{bmatrix}^{-1}\right)^{-1} \times \left(\begin{bmatrix} 9 & 0 \\ 0 & 8 \end{bmatrix}^{-1} \times \begin{bmatrix} 12 \\ 28 \end{bmatrix} + 4\times\begin{bmatrix} 16 & 0 \\ 0 & 9 \end{bmatrix}^{-1} \times \begin{Bmatrix} 11.3 \\ 32.95 \end{Bmatrix}\right) = \begin{Bmatrix} 11.515 \\ 31.863 \end{Bmatrix}$$

$$\mathbf{C}_{\mu|D} = \left(\mathbf{C}_{\mu}^{-1} + n\mathbf{C}^{-1}\right)^{-1} = \left(\begin{bmatrix} 9 & 0 \\ 0 & 8 \end{bmatrix}^{-1} + 4\times\begin{bmatrix} 16 & 0 \\ 0 & 9 \end{bmatrix}^{-1}\right)^{-1} = \begin{bmatrix} 2.769 & 0 \\ 0 & 1.756 \end{bmatrix}$$

posterior statistics of $\mathbf{x}$

$$\boldsymbol{\mu}_{x|D} = \boldsymbol{\mu}_{\mu|D} = \begin{Bmatrix} 11.515 \\ 31.863 \end{Bmatrix}, \quad \mathbf{C}_{x|D} = \mathbf{C}_{\mu|D} + \mathbf{C} = \begin{bmatrix} 16 & 0 \\ 0 & 9 \end{bmatrix} + \begin{bmatrix} 2.769 & 0 \\ 0 & 1.756 \end{bmatrix} = \begin{bmatrix} 18.769 & 0 \\ 0 & 10.756 \end{bmatrix}$$

Case study: 2 parameter update

Numerical implementation

Posterior statistics

$$\mu_{i|\mathbf{D}} = \int \theta_i f(\theta_i \mid \mathbf{D})\,\mathrm{d}\theta_i$$

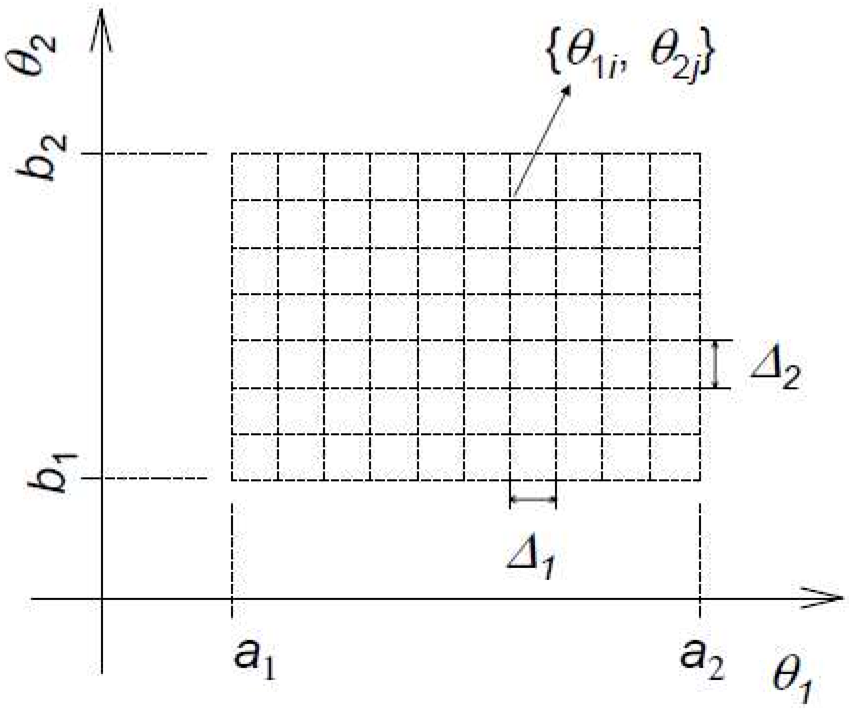
$$\sigma_{i|\mathbf{D}}^2 = \int (\theta_i - \mu_{i|\mathbf{D}})^2 f(\theta_i \mid \mathbf{D})\,\mathrm{d}\theta_i$$

$$\mathrm{COV}(\theta_i, \theta_j \mid \mathbf{D}) = \int\int (\theta_i - \mu_{i|\mathbf{D}})(\theta_j - \mu_{j|\mathbf{D}}) f(\theta_i, \theta_j \mid \mathbf{D})\,\mathrm{d}\theta_i \mathrm{d}\theta_j$$

where $f(\theta_i|\mathbf{D})$ = posterior PDF of the ith element of $\boldsymbol{\theta}$, $f(\theta_i, \theta_j|\mathbf{D})$ =posterior joint PDF of θ_i and θ_j, $\mu_{i|\mathbf{D}}$ = posterior mean of θ_i, $\sigma_{i|\mathbf{D}}$ = posterior standard deviation of θ_i, and $\mathrm{COV}(\theta_i, \theta_j|\mathbf{D})$ = posterior covariance of θ_i and θ_j, respectively.

Case study: 2 parameter update

Numerical implementation using grid integration



Case study: 2 parameter update

Let $q(\boldsymbol{\theta}_{ij})$ denote the value of the unnormalized posterior PDF at the point $\boldsymbol{\theta}_{ij} = \{\theta_{1i}, \theta_{2j}\}$, where θ_{1i} denotes the ith point of θ_1, and θ_{2j} denote the jth point of θ_2. The values of the $f(\theta_1, \theta_2|\mathbf{D})$, $f(\theta_1|\mathbf{D})$ and $f(\theta_2|\mathbf{D})$ can be evaluated on discrete points using the following equations

$$f\left(\theta_{1i}, \theta_{2j} \mid \mathbf{D}\right) = \frac{q\left(\boldsymbol{\theta}_{ij}\right)}{\Delta_1 \Delta_2 \sum_{j=1}^{n_2} \sum_{i=1}^{n_1} q\left(\boldsymbol{\theta}_{ij}\right)}$$

$$f\left(\theta_{1i} \mid \mathbf{D}\right) = \frac{\sum_{j=1}^{n_2} q\left(\boldsymbol{\theta}_{ij}\right)}{\Delta_1 \sum_{j=1}^{n_2} \sum_{i=1}^{n_1} q\left(\boldsymbol{\theta}_{ij}\right)}$$

$$f\left(\theta_{2j} \mid \mathbf{D}\right) = \frac{\sum_{i=1}^{n_1} q\left(\boldsymbol{\theta}_{ij}\right)}{\Delta_2 \sum_{j=1}^{n_2} \sum_{i=1}^{n_1} q\left(\boldsymbol{\theta}_{ij}\right)}$$

$$\mu_{\theta_1|\mathbf{D}} = \Delta_1 \sum_{i=1}^{n_1} \theta_{1i} f\left(\theta_{1i} \mid \mathbf{D}\right)$$

$$\sigma^2_{\theta_1|\mathbf{D}} = \Delta_1 \sum_{i=1}^{n_1} \left(\theta_{1i} - \mu_{1|\mathbf{D}}\right)^2 f\left(\theta_{1i} \mid \mathbf{D}\right)$$

$$\operatorname{cov}\left(\theta_1, \theta_2 | \mathbf{D}\right) = \Delta_1 \Delta_2 \sum_{j=1}^{n_2} \sum_{i=1}^{n_1} \left(\theta_{1i} - \mu_{1|\mathbf{D}}\right)\left(\theta_{2j} - \mu_{2|\mathbf{D}}\right) f\left(\theta_{1i}, \theta_{2i} \mid \mathbf{D}\right)$$