PURDUE

ECE 640141

Final Exam, December 16, Fall 2025

NAME ____

PUID ____

Exam instructions:
- A fact sheet is included **at the end of this exam** for your use.
- You have 120 minutes to work the exam.
- This is a closed-book and closed-note exam. You may not use or have access to your book, notes, any supplementary reference, a calculator, or any communication device including a cell-phone or computer.
- You may not communicate with any person other than the official proctor during the exam.

To ensure Gradescope can read your exam:
- Write your full name and PUID above and on the top of every page.
- Answer all questions in the area designated for each problem.
- Write only on the front of the exam pages.
- DO NOT run over to the next question.

Name/PUID: Key

Problem 1.(25pt) Maximum Likelihood Estimate

Let Y_n for $n = 0, \ldots, N-1$ be i.i.d. random variables with distribution $N(\mu, \sigma^2)$ and define $\theta = (\mu, \sigma^2)$.

1a) Derive the expression for the negative log likelihood, $l(\theta) = -\log_\theta p(y)$.

1b) What are the natural sufficient statistics, $T(y)$, for this family of distributions?

1c) Find an expression for

$$\hat{\mu} = \arg\min_{\mu} l((\mu, \sigma^2)) \ ,$$

and show that the arg min does not depend on σ^2.

1d) Find an expression for $l^*(\sigma) = \min_\mu l((\mu, \sigma^2))$.

1e) Use the result of part c) above to solve for the maximum likelihood estimate of θ.

$$(\hat{\mu}, \hat{\sigma}) = \arg\min_{(\mu,\sigma)} l((\mu, \sigma)) \ .$$

Solution:

Q1a:

$$p(y) = \prod_{n=0}^{N-1} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ -\frac{1}{2\sigma^2}(y_n - \mu)^2 \right\}$$

$$l(\theta) = \sum_{n=0}^{N-1} \left\{ \frac{1}{2\sigma^2}(y_n - \mu)^2 + \frac{1}{2}\log\left(2\pi\sigma^2\right) \right\}$$

Q1b: The natural sufficient statistics are

$$\begin{aligned} b &= \sum_{n=0}^{N-1} y_n \\ S &= \sum_{n=0}^{N-1} y_n^2 \end{aligned}$$

Q1c: The minimum of $l(\mu, \sigma^2)$ is the solution to

$$\begin{aligned} 0 = \frac{dl(\mu, \sigma^2)}{d\mu} &= \sum_{n=0}^{N-1} \frac{1}{\sigma^2}(\mu - y_n) \\ 0 = \frac{dl(\mu, \sigma^2)}{d\mu} &= \sum_{n=0}^{N-1} (\mu - y_n) \ . \end{aligned}$$

which implies that

$$\begin{aligned} \sum_{n=0}^{N-1} \hat{\mu} &= \sum_{n=0}^{N-1} y_n \\ N\hat{\mu} &= \sum_{n=0}^{N-1} y_n \\ \hat{\mu} &= \frac{1}{N} \sum_{n=0}^{N-1} y_n \ , \end{aligned}$$

which is not a function of σ^2.

Q1d:

$$
\begin{aligned}
l(\hat{\mu}, \sigma^2) &= \sum_{n=0}^{N-1} \left\{ \frac{1}{2\sigma^2}(y_n - \hat{\mu})^2 + \frac{1}{2}\log\left(2\pi\sigma^2\right) \right\} \\
&= \frac{1}{2\sigma^2} \sum_{n=0}^{N-1} (y_n - \hat{\mu})^2 + N\frac{1}{2}\log\left(2\pi\sigma^2\right) \\
&= \frac{1}{2\sigma^2} \left\{ \sum_{n=0}^{N-1} y_n^2 - \sum_{n=0}^{N-1} \hat{\mu}^2 \right\} + N\frac{1}{2}\log\left(2\pi\sigma^2\right) \\
&= \frac{1}{2\sigma^2} \left\{ S - \sum_{n=0}^{N-1} \left(\frac{b}{N}\right)^2 \right\} + \frac{N}{2}\log\left(2\pi\sigma^2\right) \\
&= \frac{N}{2\sigma^2} \left\{ \frac{S}{N} - \left(\frac{b}{N}\right)^2 \right\} + \frac{N}{2}\log\left(2\pi\sigma^2\right)
\end{aligned}
$$

Q1e: The minimum of $l(\hat{\mu}, \gamma)$ for $\sigma^2 = \gamma$ is the solution to

$$
\begin{aligned}
0 &= \frac{dl(\hat{\mu}, \gamma)}{d\gamma} \\
&= \frac{-N}{2\gamma^2} \left\{ \frac{S}{N} - \left(\frac{b}{N}\right)^2 \right\} + \frac{N}{2}\frac{1}{\gamma}
\end{aligned}
$$

which implies that

$$
\begin{aligned}
\frac{1}{2\gamma^2} \left\{ \frac{S}{N} - \left(\frac{b}{N}\right)^2 \right\} &= \frac{1}{2}\frac{1}{\gamma} \\
\left\{ \frac{S}{N} - \left(\frac{b}{N}\right)^2 \right\} &= \gamma
\end{aligned}
$$

So we have that

$$
\begin{aligned}
\hat{\sigma}^2 &= \frac{S}{N} - \left(\frac{b}{N}\right)^2 \\
\hat{\mu} &= \frac{b}{N} \ .
\end{aligned}
$$

Name/PUID: ______________________________

Problem 2.(15pt) Maximum Likelihood Estimate

Let Z_n for $n = 0, \ldots, N-1$ be i.i.d. random variables with distribution $P\{Z_n = m\} = \pi_m$ for $m = \{0, \ldots, M-1\}$ and define $\theta = (\pi_0, \ldots, \pi_{M-1})$.

2a) Derive the expression for the negative log likelihood, $l(\theta) = -\log_\theta p(z)$.

2b) What are the natural sufficient statistics, $T(z)$, for this family of distributions?

2c) Find an expression for $\hat{\pi} = \arg\min_\pi l(\pi)$.

Solution:

Q2a:

$$l(\theta) = -\log \prod_{n=0}^{M-1} \pi_m^{N_m}$$
$$= \sum_{n=0}^{M-1} -N_m \log \pi_m \ ,$$

where

$$N_m = \sum_{n=0}^{N-1} \delta(Z_n - m) \ .$$

Q2b: The natural sufficient statistics are given by

$$N_m = \sum_{n=0}^{N-1} \delta(Z_n - m) \ ,$$

for $m = 0, \ldots, M - 1$.

Q2c:

$$\hat{\pi}_m = \frac{N_m}{N}$$

Name/PUID: ______________________________

Problem 3.(30pt) Proximal Maps and PnP

Let

$$Y = AX + \sigma W \ ,$$

where $A \in \mathbb{R}^{M\times N}$, $X \in \mathbb{R}^N$, $W \sim N(0, I)$, and $X \sim p(x)$.

Furthermore, let

$$Z = X + \rho\tilde{W} \ ,$$

where $\tilde{W} \sim N(0, I)$. Then the minimum measure squared error (MMSE) denoiser is given by

$$H_\rho(Z) = E[X|Z] \ .$$

3a) Show that for the special case when $X \sim p(x)$ is Gaussian, then the MMSE estimate of X given Z is the same as the MAP estimate of X given Z.

3b) Derive an expression for the function $H_\rho(\cdot)$ as the MAP estimate.

3c) Derive an expression for the function $f(x) = -\log p(y|x)$.

3d) Write out an explicit expression for the proximal map

$$F(v) = \arg\min_x \left\{ f(x) + \frac{1}{2\rho^2}\|x - v\|^2 \right\} \ .$$

3e) Let $\hat{H}$ be a learned approximation to H_ρ. Describe how $\hat{H}$ can be designed using training data.

3f) Specify the Plug-and-Play (PnP) algorithm in terms of forward model proximal map, F, and the denoiser $\hat{H}$.

Solution:

Q3a: Since X and Z are jointly Gaussian, the conditional distribution of X given Z must also be Gaussian. Since a Gaussian distribution is unimodal with a symmetric distribution, then we have that the mode must equal the mean. In other words,

$$\arg\max_{z} p(x|z) = \int xp(x|z)dx \ .$$

Q3b:

$$H_\rho(z) = \arg\min_x \left\{ \frac{1}{2\rho^2}\|z - x\|^2 - \log p(x) \right\}$$

Q3c:

$$f(x) = \frac{1}{2\sigma^2}\|y - Ax\|^2 + N\log(2\pi\sigma^2)$$

Q3d:

$$F(v) = \arg\min_x \left\{ \frac{1}{2\sigma^2}\|y - Ax\|^2 + \frac{1}{2\rho^2}\|x - v\|^2 \right\}$$

Q3e: Generate training pairs, (X_k, Z_k) with $Z_k = X_k + \rho W$ with $W \sim N(0, I)$. Then form a loss function

$$L(\theta) = \sum_k \|X_k - H_\theta(Z_k)\|^2 \ .$$

Then use an optimization algorithm to compute

$$\hat{\theta} = \arg\min_\theta L(\theta) \ .$$

Then the trained estimator is given by $H_{\hat{\theta}}$.

Q3f:

$v \leftarrow 0$

$u \leftarrow 0$

repeat {

$x \leftarrow F(v - u)$

$v \leftarrow \hat{H}(x + u)$

$u \leftarrow u + (x - v)$

}

Name/PUID: ______________________________

Problem 4.(25pt) EM Algorithm for Poisson Observations

Let X_n for $n = 1, \cdots, N$ be a series of i.i.d. multinomial random variables with distribution $P\{X_n = m\} = \pi_m$, and let

$$P\{Y_n = k | X_n = m\} = \frac{e^{-\lambda_m} \lambda_m^k}{k!} ,$$

be conditionally independent random variables given $X_n = m$, and let $\theta = \{\pi_0, \lambda_0, \cdots, \pi_{M-1}, \lambda_{M-1}\}$ parameterize the joint distribution.

Problem 4a) Calculate $\hat{\pi}_m$, the maximum likelihood estimate of π_m given $\{X_n, Y_n\}_{n=1}^N$.

Problem 4b) Calculate $\hat{\lambda}_m$, the maximum likelihood estimate of λ_m given $\{X_n, Y_n\}_{n=1}^N$.

Problem 4c) Use Bayes' rule to calculate an expression for $f(m|y_n) = P\{X_n = m | Y_n = y_n\}$.

Problem 4d) Specify the E-step of the EM algorithm for the estimation of θ for this specific problem.

Problem 4e) Specify the M-step of the EM algorithm for the estimation of θ for this specific problem.

Solution:

Q4a: First compute

$$N_m \leftarrow \sum_{n=1}^{N} \delta(X_n = m)$$

Then compute

$$\hat{\pi}_m \leftarrow \frac{N_m}{N} \ .$$

Q4b: First compute

$$b_m \leftarrow \sum_{n=1}^{N} Y_n \delta(X_n = m)$$
$$N_m \leftarrow \sum_{n=1}^{N} \delta(X_n = m) \ .$$

Then compute

$$\hat{\lambda}_m \leftarrow \frac{b_m}{N_m} \ .$$

Q4c:

$$\begin{aligned} f(m|y_n) &= P\{X_n = m | Y_n = y_n\} \\ &= \frac{1}{Z} \frac{e^{-\lambda_m} \lambda_m^{y_n}}{y_n!} \pi_m \ , \end{aligned}$$

where

$$Z = \sum_{m=0}^{M-1} \frac{e^{-\lambda_m} \lambda_m^{y_n}}{y_n!} \pi_m \ .$$

Q4d: The E-step is given by computing the conditional expectation of the sufficient statistics.

$$\bar{N}_m \leftarrow \sum_{n=1}^{N} f(m|y_n)$$
$$\bar{b}_m \leftarrow \sum_{n=1}^{N} Y_n f(m|y_n) \ .$$

Q4e: The M-step is given by updating the parameters.

$$\hat{\pi}_m \leftarrow \frac{\bar{N}_m}{N}$$
$$\hat{\lambda}_m \leftarrow \frac{\bar{b}_m}{\bar{N}_m} \ .$$

Name/PUID: ______________________________

Problem 5.(25pt) Markov Chains

Let $X_n \in \{0, \ldots, M-1\}$ be an irreducible Markov chain with transition probabilities $P \in \Re^{M \times M}$. Furthermore, assume that $\forall i, 1 > P_{i,i} > 0$, and $P = P^t$.

Problem 5a) Is X_n a homogeneous Markov chain? Prove your answer.

Problem 5b) Is X_n an aperiodic Markov chain? Prove your answer.

Problem 5c) Is X_n an ergodic Markov chain? Prove your answer.

Problem 5d) Is X_n a reversible Markov chain? Prove your answer.

Problem 5e) What is the stationary distribution of the Markov chain? Prove your answer.

Solution:

Q5a: Yes. This is true because the parameters of the MC are not a function of time.

Q5b: Yes. Since $P_{i,i} > 0$, each state must have the same period of 1.

Q5c: Since MC is irreducible, aperiodic, and has a finite number of states, it must be ergodic.

Q5d: The MC is reversible if and only if it solves the detailed balance equations.
Let $\pi_m = 1/M$ for all states $m \in \{0, \ldots, M-1\}$. Then we have that

$$\begin{aligned} \pi_i P_{i,j} &= (1/M) P_{i,j} \\ &= (1/M) P_{j,i} \\ &= \pi_j P_{j,i} \ . \end{aligned}$$

So the detailed balance equations hold, which implies that the MC is reversible.

Q5e: Since the distribution $\pi_i = (1/M)$ solves the DBE, it must also solve the FBE. So the stationary distribution of the MC is given by

$$pi_i = \frac{1}{M} \ .$$

So this implies that when the transition probabilities are symmetric, then the stationary distribution of the MC is uniform over the states.

ECE641 Fact Sheet

Probability Background

Total Probability
$P(A) = \sum_n P(A|B_n)P(B_n)$

Total Probability for Conditional Probabilities
$P(A|C) = \sum_n P(A|B_n, C)P(B_n|C)$

Bayes' Rule
$P(B|A) = \frac{P(A|B)P(B)}{P(A)}$

Conditional Joint Probability
$P(A, B|C) = P(A|B, C)P(B|C)$

Maximum Likelihood (ML) Estimator (Frequentist)

$$\hat{\theta} = \arg\max_{\theta\in\Omega} p_\theta(Y) = \arg\max_{\theta\in\Omega} \log p_\theta(Y)$$

$$0 = \nabla_\theta p_\theta(Y)|_{\theta=\hat{\theta}}$$

$$\hat{\theta} = T(Y)$$

$$\bar{\theta} = \mathbb{E}_\theta[\hat{\theta}]$$

$$\text{bias}_\theta = \bar{\theta} - \theta \qquad \text{var}_\theta = \mathbb{E}_\theta[(\hat{\theta} - \bar{\theta})^2]$$

$$\text{MSE} = \mathbb{E}_\theta[(\hat{\theta} - \theta)^2] = \text{var}_\theta + (\text{bias}_\theta)^2$$

For $Y = AX + W$, where X and W are independent zero mean Gaussian distributed with R_X and R_W, respectively. Then the ML estimate is find by maximizing $log(p_{y/x}(y/x))$:

$$\hat{X}_{ML} = (A^t R_W^{-1} A)^{-1} A^t R_W^{-1} y$$

Maximum A Posteriori (MAP) Estimator

$$\begin{aligned}\hat{X}_{MAP} &= \arg\max_{x\in\Omega} p_{x|y}(x|Y) \\ &= \arg\max_{x\in\Omega} \log p_{x|y}(x|Y) \\ &= \arg\min_{x\in\Omega} \{-\log p_{y|x}(y|x) - \log p_x(x)\}\end{aligned}$$

For $Y = AX + W$, where X and W are independent zero mean Gaussian distributed with R_X and R_W, respectively. Then the MAP or equivalently MMSE estimate is:

$$\hat{X}_{MAP} = (A^t R_W^{-1} A + R_X^{-1})^{-1} A^t R_W^{-1} y$$

Power Spectral Density (zero-mean WSS Gaussian process)

1D DTFT:

$$S_X(e^{j\omega}) = \sum_{n=-\infty}^{\infty} R(n) e^{-j\omega n}$$

2D DSFT:

$$S_X(e^{j\omega_1}, e^{j\omega_2}) = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} R(m, n) e^{-j\omega_1 m - j\omega_2 n}$$

Causal Gaussian Models

$\sigma_n^2 \triangleq \mathbb{E}[\mathcal{E}_n^2]$, $\hat{X} = HX$, $\mathcal{E} = (I - H)X = AX$, $\mathbb{E}[\mathcal{E}\mathcal{E}^t] = \Lambda$, $\Lambda = \text{diag}\{\sigma_1^2, \sigma_2^2, ..., \sigma_N^2\}$

$p_x(x) = |det(A)| p_{\mathcal{E}}(Ax)$, $|det(A)| = 1$,
$R_X = (A^t \Lambda^{-1} A)^{-1}$

1-D Gaussian AR models:

- Toeplitz $H_{i,j} = h_{i-j}$
- Circulant $H_{i,j} = h_{(i-j) mod N}$
- P^{th} order IIR filter $X_n = \mathcal{E}_n + \sum_{i=1}^{P} X_{n-i} h_i$, $R_{\mathcal{E}}(i - j) = \mathbb{E}[\mathcal{E}_i \mathcal{E}_j] = \sigma_c^2 \delta_{i-j}$
- $R_X(n) * (\delta_n - h_n) * (\delta_n - h_{-n}) = R_{\mathcal{E}}(n) = \sigma_c^2 \delta_n$, $S_X = \frac{\sigma_c^2}{|1 - H(\omega)|^2}$

2-D Gaussian AR:

- $\mathcal{E}_s = X_s - \sum_{r\in W_p} h_r X_{s-r}$,
- Toeplitz block Toeplitz $H_{mN+k, nN+l} = h_{m-n, k-l}$

Non-causal Gaussian Models

- $\sigma_n^2 \triangleq \mathbb{E}[\mathcal{E}_n^2 | X_i, i \neq n]$, $B_{i,j} = \frac{1}{\sigma_i^2}(\delta_{i-j} - g_{i,j})$, $\sigma_n^2 = (B_{n,n})^{-1}$, $g_{n,i} = \delta_{n-i} - \sigma_n^2 B_{n,i}$ (homogeneous: $g_{i,j} = g_{i-j}$, $\sigma_i^2 = \sigma_{NC}^2$)
- $G_{i,j} = g_{i,j}$, $\Gamma = \text{diag}\{\sigma_1^2, \sigma_2^2, ..., \sigma_N^2\}$, $B = \Gamma^{-1}(I - G)$, $\Gamma = \text{diag}(B)^{-1}$, $G = I - \Gamma B$, $\mathbb{E}[\mathcal{E}_n X_{n+k}] = \sigma_{NC}^2 \delta_k$
- $R_X(n) * (\delta_n - g_n) * (\delta_n - g_{-n}) = R_{\mathcal{E}}(n) = \sigma_{NC}^2(\delta_n - g_n)$, $S_X = \frac{\sigma_{NC}^2}{1 - G(\omega)}$, $R_X(n) * (\delta_n - g_n) = \sigma_{NC}^2 \delta_n$
- **Relationship b/w AR and GMRF:** $\sigma_{NC}^2 = \frac{\sigma_c^2}{1 + \sum_{n=1}^{P} h_n^2}$, $g_n = \delta_n - \frac{(\delta_n - h_n) * (\delta_n - h_{-n})}{1 + \sum_{n=1}^{P} h_n^2} (= \frac{\rho}{1+\rho^2}(\delta_{n-1} + \delta_{n+1}),\ P = 1)$

Surrogate Function

Our objective is to find a surrogate function $\rho(\Delta; \Delta')$, to the potential function $\rho(\Delta)$.

Maximum Curvature Method

Assume the surrogate function of the form

$$\rho(\Delta;\Delta') = \alpha_1\Delta + \frac{\alpha_2}{2}(\Delta-\Delta')^2$$

where $\alpha_1 = \rho'(\Delta')$ and $\alpha_2 = \max_{\Delta\in\mathbb{R}} \rho''(\Delta)$.

Symmetric Bound Method

Assume that potential function is bounded by symmetric and quadratic function of Δ, then the surrogate function is

$$\rho(\Delta;\Delta') = \frac{\alpha_2}{2}\Delta^2$$

which results in the following symmetric bound surrogate function:

$$\rho(\Delta;\Delta') = \begin{cases} \frac{\rho'(\Delta')}{2\Delta'}\Delta^2 & \text{if} \quad \Delta' \neq 0 \\ \frac{\rho''(0)}{2}\Delta^2 & \text{if} \quad \Delta' = 0 \end{cases}$$

Review of Convexity in Optimization

Definition A.6. ***Closed, Bounded, and Compact Sets***

Let $\mathcal{A} \subset \mathbb{R}^N$, then we say that $\mathcal{A}$ is:

- **Closed** if every convergent sequence in $\mathcal{A}$ has its limit in $\mathcal{A}$.
- **Bounded** if $\exists M$ such that $\forall x \in \mathcal{A}$, $\|x\| < M$.
- **Compact** if $\mathcal{A}$ is both closed and bounded.

Definition A.11. ***Closed Functions***

We say that function $f : \mathbb{R}^N \to \mathbb{R}\cup\{\infty\}$ is **closed** if for all $\alpha \in \mathbb{R}$, the sublevel set $\mathcal{A}_\alpha = \{x \in \mathbb{R}^N : f(x) \leq \alpha\}$ is closed set.

Theorem A.6. ***Continuity of Proper, Closed, Convex Functions***

Let $f : \mathbb{R}^N \to \mathbb{R}\cup\{\infty\}$ be a proper convex function. Then f is closed if and only if it is lower semicontinuous.

Optimization Methods:

Gradient Descent: $x^{(k+1)} = x^{(k)} - \beta\nabla f(x^{(k)})$
Gradient Descent with Line Search:
$d^{(k)} = -\nabla f(x^{(k)})$

α solves the equation : $0 = \frac{\partial f(x^{(k)}+\alpha d^{(k)})}{\partial\alpha} = [\nabla f(x^{(k)}+\alpha d^{(k)})]^t d^{(k)}$.

Update: $x^{(k+1)} \leftarrow x^k + \alpha\frac{\|d^{(k)}\|^2}{\|d^{(k)}\|_Q^2}d^{(k)}$ where $Q = A^t\Lambda A + B$

Coordinate Descent :
$\alpha = \frac{(y-Ax)^t\Lambda A_{*,s} - x^t B_{*,s}}{\|A_{*,s}\|_\Lambda^2 + B_{s,s}}$ (for $Y|X \sim N(AX, \Lambda^{-1})$)
$x_s \leftarrow x_s + \frac{(y-Ax)^t A_{*,s} - \lambda(x_s - \Sigma_{r\in\partial s} g_{s-r}x_r)}{\|A_{*,s}\|^2+\lambda}$, $\lambda = \frac{\sigma^2}{\sigma_x^2}$

Pairwise quadratic form identity

$x^tBx = \sum\limits_{s\in S} a_s x_s^2 + \frac{1}{2}\sum\limits_{s\in S}\sum\limits_{r\in S} b_{s,r}|x_s - x_r|^2$, $a_s = \sum\limits_{r\in S} B_{s,r}$, $b_s = -B_{s,r}$

Miscellaneous

For any invertible matrix A, 1. $\frac{\partial|A|}{\partial A} = |A|A^{-1}$ 2. $\frac{\partial tr(BA)}{\partial A} = B$ 3. $tr(AB) = tr(BA)$

Plug and Play

(non-expansive map)

(CE equations)

$x^* = F(x^* - u^*)$

$x^* = H(x^* + u^*)$

(Douglas-Rachford algorithm)

set $\rho \in (0,1)$

initialize w_1

repeat{

$w_1' \leftarrow Tw_1$

$w_1 \leftarrow (1-\rho)w_1' + \rho w_1$

}

return w_1

Note that here $w_1 = x - u$, $w_2 = x + u$, and $x = \frac{w_1+w_2}{2}$, so then $(2F - I)w_1 = w_2$. And, $T = (2H - I)(2F - I)$.

(Convergence of Douglas-Rachford algorithm)

When F and H are proximal maps of proper closed convex functions f and h then Douglas-Rachford algorithm converges to both the CE solution and the MAP estimate.

EM algorithm

General EM Algorithm:
E-step : $Q(\theta;\theta^{(k)}) = \mathbb{E}[\log(p(y,X|\theta))|Y=y,\theta^{(k)}]$
M-step : $\theta^{(k+1)} = \arg\max_{\theta\in\Omega} Q(\theta;\theta^{(k)})$
(ML estimate for Gaussian mixture)
$\log p(y,x|\theta) = \Sigma_{n=1}^{N} \log p(y_n,x_n|\theta) = \Sigma_{n=1}^{N}\Sigma_{m=0}^{M-1}\delta(x_n - m)\{\log p(y_n|\mu_m,\sigma_m) + \log \pi_m\}$

(Exponential Family)
A family of density functions $p_\theta(y)$ for y and θ is said to be a exponential family if there exists functions $\eta(\theta)$, $s(y)$, and $d(\theta)$ and natural statistic $T(y)$ such that $p_\theta(y) = exp\{\langle\eta(\theta), T(y)\rangle + d(\theta) + s(y)\}$

(sufficient statistic)
$T(Y)$ is a sufficient statistic for the family of distributions $p_\theta(y)$ if the density functions can be written in the form $p_\theta(y) = h(y)g(T(y),\theta)$ where g and h are any two functions.

Markov Chains

Parameter Estimation for Markov Chains: $N_j = \delta(X_0 - j)$, $K_{i,j} = \Sigma_{n=1}^{N}\delta(X_n - j)\delta(X_{n-1} - i)$
$\log(p(x)) = \sum_{j\in\Omega}\{N_j \log(\tau_j) + \sum_{i\in\Omega} K_{i,j}\log(P_{i,j})\}$
ML Estimate $\hat{\tau_j} = N_j$ and $\hat{P_{i,j}} = \frac{K_{i,j}}{\sum_{j\in\Omega} K_{i,j}}$
Marginal density at any time n: $\pi^{(n)} = \pi^{(0)}P^n$ and $\pi^{(\infty)} = \pi^{(0)}P^\infty$
Log likelihood of HMM (MAP Estimate):
$\hat{x} = \arg\max_{x\in\Omega^N}\{\log\tau_{x_0} + \Sigma_{n=1}^{N}\{\log f(y_n|x_n) + \log P_{x_{n-1},x_n}\}\}$
State Sequence Estimation and Dynamic Programming:
$L(j,n) = \max_{x_{>n}}\{\log p(y_{>n}, x_{>n}|x_n = j)\}$ and $L(j,N) = 0$
$L(i,n-1) = \max_{j\in\Omega}\{\log f(y_n|j) + \log P_{i,j} + L(j,n)\}$
$\hat{x_0} = \arg\max_{j\in\Omega}\{\log\tau_j + L(j,0)\}$
$\hat{x_n} = \arg\max_{j\in\Omega}\{\log P_{\hat{x_{n-1}},j} + \log f(y_n|j) + L(j,n)\}$
State Probability and the Forward-Backward Algorithm:
$\alpha_n(j) = p(x_n = j, y_n, y_{<n})$ $\beta_n(j) = p(y_{>n}|x_n = j)$
$p(x_{n-1} = i, x_n = j|y) = \frac{\alpha_{n-1}(i)P_{i,j}f(y_n|j)\beta_n(j)}{p(y)}$
$\alpha_n(j) = \sum_{i\in\Omega}\alpha_{n-1}(i)P_{i,j}f(y_n|j)$
$\beta_n(i) = \sum_{j\in\Omega}P_{i,j}f(y_{n+1}|j)\beta_{n+1}(j)$

(Irreducible Markov Chain). A discrete-time, discrete-space homogeneous Markov chain is said to be irreducible if for all states $i,j\in\Omega$, i and j communicate.

(Communicating States). States $i,j\in\Omega$ of a discrete-time, discrete-space homogeneous Markov chain are said to communicate if there exists integers $m > 0$ and $n > 0$ such that $[P^m]_{i,j} > 0$ and $[P^n]_{j,i} > 0$.

(period of state) State $i\in\Omega$ of a discrete-time, discrete-space homogeneous Markov chain has period $d(i) = gcd\{n\in\mathbb{N}_+|[P^n]_{i,i} > 0\}$.
State i is aperiodic if $d(i) = 1$ and periodic if $d(i) > 1$.
Ergodic MC : $\pi_j = lim_{n\to\infty}[P^n]_{i,j} > 0$

(detailed balance equations)
$\pi_i P_{i,j} = \pi_j P_{j,i}$
$\sum_{i\in\Omega}\pi_i = 1$

(full balance equations)
$\pi^\infty = \pi^\infty P$ or $\pi_j = \sum_{i\in\Omega}\pi_i P_{i,j}$
$\sum_{i\in\Omega}\pi_i = 1$

Stochastic Sampling

(inverse transform sampling)
$X \leftarrow F^{-1}(U)$ where $U \leftarrow Rand([0,1])$ and $F^{-1}(u) = \inf\{x|F(x) \geq u\}$ generates a sample from random variable X with CDF $F(x) = P\{X \leq x\}$

(Metropolis algorithm)

initialize X^0

for k from 0 to $K-1\{$

$U \leftarrow Rand([0,1])$

$W \leftarrow Q^{-1}(U|X^{(k)})$

$\alpha \leftarrow \min\{1, e^{-[u(W)-u(X^{(k)})]}\}$

$U \leftarrow Rand([0,1])$

if $U < \alpha$ then $X^{(k+1)} \leftarrow W$

else $X^{(k+1)} \leftarrow X^{(k)}$

$\}$

Note: where $Q^{-1}(\cdot|x^{(k)})$ is the inverse CDF corresponding to proposal density $q(w|x^{(k)})$