STOCHASTIC MODELING OF ENVIRONMENTAL TIME SERIES

Richard W. Katz

LECTURE 1

(1) Environmental Motivation

(2) Probabilistic Background

(3) Statistical Background

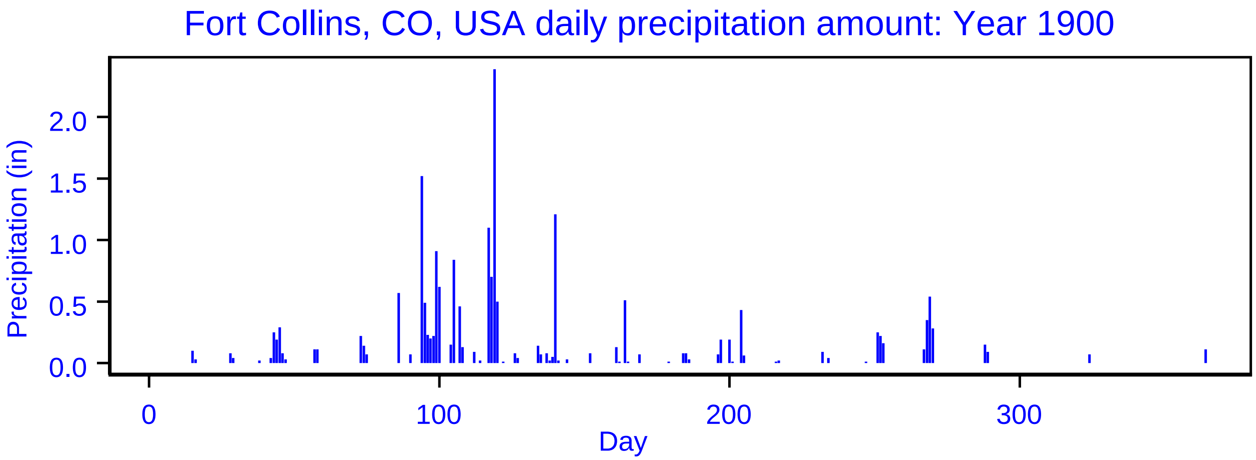
(1) Environmental Motivation

- **Example**

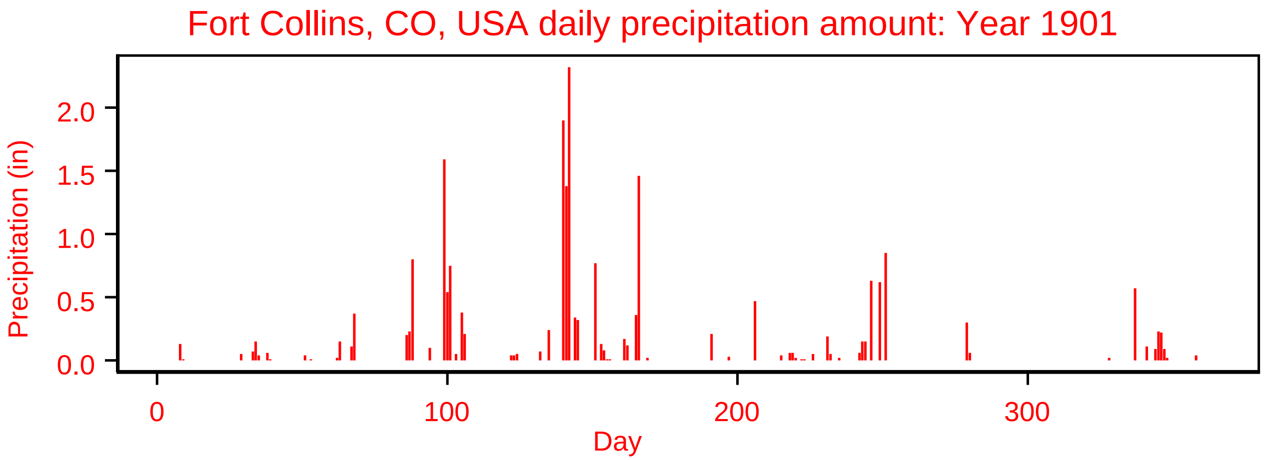
-- **Time series of daily precipitation amount at single location (Fort Collins, CO, USA: 1900 – 1999)**

- **Statistical Characteristics of Environmental Time Series**

- **Scientific Research Questions**



Fort Collins, CO, USA daily precipitation amount: Year 1900
Precipitation (in)
0.0
0.5
1.0
1.5
2.0
0
100
200
300
Day



Fort Collins, CO, USA daily precipitation amount: Year 1901
Precipitation (in)
0.0
0.5
1.0
1.5
2.0
0
100
200
300
Day

Statistical Characteristics

- **Intermittency**
- **Skewness**
- **Extremes (heavy tail?)**
- **Temporal dependence**
- **Seasonality**
- **Spatial dependence (not shown)**
- **Trend?**

Scientific Research Questions

Overdispersion

-- Stochastic models for daily precipitation tend to underestimate variance of monthly, seasonal, or annual total precipitation

Explanations

- **High Frequency Variations**

  -- Inadequate stochastic model for daily precipitation?

- **Low Frequency Variations**

  -- "Potential predictability" or climate change?

(2) Probabilistic Background

Preliminaries

-- To be used in statistical modeling of overdispersion phenomenon

- **Scaling/Aggregation**
- **Conditional Moments**
- **Effects of Intermittency**

Scaling/Aggregation

- **Stationary stochastic process** $\{X_t : t = 1, 2, \ldots\}$

-- Moments

$$\mu = \mathrm{E}(X_t), \qquad \sigma^2 = \mathrm{Var}(X_t), \qquad \rho_l = \mathrm{Corr}(X_t, X_{t+l}), \ \ l = 1, 2, \ldots$$

- **Sum** $S_T = X_1 + X_2 + \cdots + X_T$

-- Variance

$$(1/T)\ \mathrm{Var}(S_T) \longrightarrow \sigma^2\,[1 + 2(\rho_1 + \rho_2 + \ldots)] \equiv (\sigma^*)^2 \text{ as } T \longrightarrow \infty$$

-- Distribution $(S_T - T\mu) \,/\, (\sigma^* T^{1/2}) \longrightarrow N(0, 1)$ (in distribution)

- **Example: First-order autoregressive [AR(1)] process**

$$X_{t+1} - \mu = \varphi(X_t - \mu) + \varepsilon_t \,, \;\; t = 1, 2, \ldots$$

$$\mathrm{E}(\varepsilon_t) = 0, \quad \mathrm{E}(\varepsilon_t \, \varepsilon_{t+l}) = 0, \;\; l = 1, 2, \ldots$$

Here $\mu = \mathrm{E}(X_t)$, $\sigma^2 = \mathrm{Var}(X_t)$, $\varphi = \mathrm{Corr}(X_t, X_{t+1})$

-- Variance of sum S_T

$$(1/T)\ \mathrm{Var}(S_T) \longrightarrow (\sigma^*)^2 = \sigma^2\ [(1 + \varphi)/[(1 - \varphi)] \;\; \text{as } T \longrightarrow \infty$$

Central Limit Theorem holds with this normalizing constant σ^*

- **Example: Two-state, first-order Markov chain**

$$\{J_t: t = 1, 2, \ldots\}, \quad J_t = 0, 1$$

Transition probabilities:

$$P_{jk} = \Pr\{J_{t+1} = k \mid J_t = j\}; \; j, k = 0, 1$$

$$\pi = \Pr\{J_t = 1\} = P_{01} / (P_{10} + P_{01}), \quad d = \mathrm{Corr}(J_t, J_{t+1}) = P_{11} - P_{01}$$

Sum: $N(T) = J_1 + J_2 + \cdots + J_T$

Variance:

$$(1/T) \,\mathrm{Var}[N(T)] \longrightarrow (\sigma^*)^2 = \pi(1 - \pi) \,[(1 + d)/(1 - d)] \text{ as } T \longrightarrow \infty$$

Central Limit Theorem holds with this normalizing constant σ^*

Conditional Moments

- **Expected Value**

$$\mathrm{E}(X) = \mathrm{E}[\mathrm{E}(X|Y)]$$

- **Variance**

$$\mathrm{Var}(X) = \mathrm{E}[\mathrm{Var}(X|Y)] + \mathrm{Var}[\mathrm{E}(X|Y)]$$

- **Covariance**

$$\mathrm{Cov}(X, Y) = \mathrm{E}[\mathrm{Cov}(X, Y|Z)] + \mathrm{Cov}[\mathrm{E}(X|Z), \mathrm{E}(Y|Z)]$$

Effects of Intermittency

$\{N(t): t = 1, 2, \ldots\}$ number of events in time interval $[0, t]$ (corresponding to point process)

$\{X_t: t = 1, 2, \ldots\}$ independent and identically distributed [$\mu = \mathrm{E}(X_t)$, $\sigma^2 = \mathrm{Var}(X_t)$]

$\{X_t\}$ independent of $\{N(t)\}$

- **Random sum** $S_{N(T)} = X_1 + X_2 + \cdots + X_{N(T)}$

Mean: $\mathrm{E}[S_{N(T)}] = \mathrm{E}[N(T)]\,\mu$

Variance: $\mathrm{Var}[S_{N(T)}] = \mathrm{E}[N(T)]\,\sigma^2 + \mathrm{Var}[N(T)]\,\mu^2$

Central Limit Theorem: Still holds for $S_{N(T)}$

- **Extremes**

-- **Ordinary maximum** $M_T = \max\{X_1, X_2, \ldots, X_T\}$

Extreme value theorem: M_T**, suitably normalized, has asymptotically Generalized Extreme Value (GEV) distribution function**

$$G(x) = \exp\left(-\{1 + \xi\,[(x-\mu)/\sigma]\}^{-1/\xi}\right), \quad 1 + \xi\,[(x-\mu)/\sigma] > 0$$

-- **Maximum with random indices**

$$M_{N(T)} = \max\{X_1, X_2, \ldots, X_{N(T)}\}$$

Extreme value theorem still holds for $M_{N(T)}$ **if**

$$(1/T)\,N(T) \to \pi \ \text{ as } T \to \infty \ \text{ (in probability)}$$

But now limiting distribution is $[G(x)]^{\pi}$ **(i.e., GEV with adjusted parameters)**

(3) Statistical Background

Mixtures:

-- Chance mechanism for overdispersion

- **Mixture of Distributions**
- **Dependence Induced by Mixtures**

Mixture of Distributions

- **Finite mixture (two-component case):**

$$\Pr\{X \le x\} = \Pr\{X \le x \mid I = 0\}\, \Pr\{I = 0\} + \Pr\{X \le x \mid I = 1\}\, \Pr\{I = 1\}$$

I two-state random variable (generally only observe X, *not* I)

-- Notation $F(x) = \Pr\{X \le x\}$, $\quad F_i(x) = \Pr\{X \le x \mid I = i\}$, $\quad w = \Pr\{I = 1\}$,

$$\mu_i = \mathrm{E}(X \mid I = i), \qquad \sigma_i^2 = \mathrm{Var}(X \mid I = i), \quad i = 0, 1$$

-- Moments $\mathrm{E}(X) = (1 - w)\mu_0 + w\mu_1$

$$\mathrm{Var}(X) = (1 - w)\sigma_0^2 + w\sigma_1^2 + w(1 - w)(\mu_1 - \mu_0)^2$$

- **Example: Mixture of two normal distributions**

Conditional densities:

$$f_i(x) = [(2\pi)^{1/2}\sigma_i]^{-1} \exp(-\tfrac{1}{2}[(x-\mu_i)/\sigma_i]^2), \quad i = 0, 1$$

-- Unimodal or bimodal?

Only bimodal if difference between μ_0 and μ_1 large enough relative to σ_0 and σ_1

Sufficient condition for unimodal distribution:

$$|\mu_1 - \mu_0| \leq 2 \min(\sigma_0, \sigma_1)$$

- **Example: Mixture of two exponential distributions**

$$F_i(x) = 1 - (1/\sigma_i)\ \exp[-\ (x/\sigma_i)],\ \ x > 0,\ \ i = 0, 1$$

-- Longer tail than single exponential (*not* memoryless)

$$\Pr\{X \le u + x \mid X > u, I = i\} = \Pr\{X \le x \mid I = i\},\ \ i = 0, 1$$

But

$$\Pr\{X \le u + x \mid X > u\} \ne \Pr\{X \le x\}$$

-- Used to fit precipitation "intensity"

Conditional distribution of precipitation amount given occurrence:
$\Pr\{X \le x \mid X > 0\}$

- **Example: Infinite-dimensional mixture of exponentials**

-- Let X have conditional exponential distribution

$$\Pr\{X \leq x \mid \sigma\} = 1 - (1/\sigma)\exp[-(x/\sigma)],$$

where $Y = 1/\sigma$ has gamma distribution

$$\Pr\{Y \leq y\} = [\beta\Gamma(\alpha)]^{-1}(y/\beta)^{\alpha-1}\exp[-(y/\beta)],\ \ y > 0$$

Then unconditional distribution of X is Pareto

$$\Pr\{X \leq x\} = 1 - (1 + \beta x)^{-\alpha}$$

-- Chance mechanism by which heavy tail obtained from light tail

Dependence Induced by Mixtures

- **Example: Induced "spatial" dependence**

Let X & Y both have mixture distributions given same two-state conditioning variable I

Conditional means: $\mu_X(i) = \mathrm{E}(X \mid I = i), \quad \mu_Y(i) = \mathrm{E}(Y \mid I = i), \; i = 0, 1$

Conditional variances: $[\sigma_X(i)]^2 = \mathrm{Var}(X \mid I = i), \; [\sigma_Y(i)]^2 = \mathrm{Var}(Y \mid I = i)$

Assume X & Y conditionally independent given I [so $\mathrm{Cov}(X, Y|I) = 0$]

-- Unconditional covariance

$$\mathrm{Cov}(X, Y) = w(1 - w)[\mu_X(1) - \mu_X(0)][\mu_Y(1) - \mu_Y(0)]$$

- **Example: Hidden Markov model**

$\{X_t: t = 1, 2, \ldots\}$ has mixture distribution given J_t (two-state variable with conditional means μ_0 & μ_1, standard deviations σ_0 & σ_1)

$\{J_t: t = 1, 2, \ldots\}$ first-order Markov chain (with parameters π & d)

Assume $\{X_t\}$ conditionally independent given $\{J_t\}$

-- Unconditional autocorrelation function of $\{X_t\}$

$$\mathrm{Corr}(X_t, X_{t+l}) = a\, \mathrm{Corr}(J_t, J_{t+l}) = a\, d^{\,l}, \quad l = 1, 2, \ldots$$

Where $$a = [\pi(1-\pi)(\mu_1-\mu_0)^2] \,/\, \mathrm{Var}(X_t),$$

$$\mathrm{Var}(X_t) = (1-\pi)\sigma_0^{\,2} + \pi\sigma_1^{\,2} + \pi(1-\pi)(\mu_1-\mu_0)^2$$

- **Example: AR(1) process with shifting mean**

$\{X_t: t = 1, 2, \ldots\}$ **conditional AR(1) process given** $\{I_t: t = 1, 2, \ldots\}$

$$\mu_i = \mathrm{E}(X_t \mid I_t = i), \;\; i = 0, 1$$

Assume $$\sigma_i^2 = \mathrm{Var}(X_t \mid I_t = i) = \sigma^2, \;\; i = 0, 1,$$

$$\varphi_{ij} = \mathrm{Corr}(X_t, X_{t+1} \mid I_t = i, I_{t+1} = j) = \varphi, \;\; i, j = 0, 1$$

$\{I_t\}$ **independent & identically distributed** ($w = \Pr\{I = 1\}$)

-- Autocorrelation function of $\{X_t\}$

$$\mathrm{Corr}(X_t, X_{t+l}) = a\, \varphi^l, \;\; l = 1, 2, \ldots$$

where $a = \sigma^2 / \mathrm{Var}(X_t)$, $\mathrm{Var}(X_t) = \sigma^2 + w(1 - w)(\mu_1 - \mu_0)^2$

Derivation of autocorrelation function:

Because $\text{Cov}[\text{E}(X_t \mid I_t, I_{t+l}), \text{E}(X_{t+l} \mid I_t, I_{t+l})] = 0,$

$$\text{Cov}(X_t, X_{t+l}) = \text{E}[\text{Cov}(X_t, X_{t+l} \mid I_t, I_{t+l})]$$

So $\text{Corr}(X_t, X_{t+l}) = \text{E}[\text{Cov}(X_t, X_{t+l} \mid I_t, I_{t+l})] / \text{Var}(X_t)$

$$= \sigma^2\varphi^l / \text{Var}(X_t)$$

$$= \{\sigma^2 / [\sigma^2 + w(1 - w)(\mu_1 - \mu_0)^2]\}\varphi^l, \ \ l = 1, 2, \ldots$$

STOCHASTIC MODELING OF ENVIRONMENTAL TIME SERIES

Richard W. Katz

LECTURE 2

(1) Background on EM Algorithm

(2) Description & Properties of EM Algorithm

(3) Environmental Applications of EM Algorithm

(4) Maximum Likelihood Estimation for Mixtures

(1) Background on EM Algorithm

Expectation-Maximization (EM) Algorithm

- **Basic Idea of EM Algorithm**
- **Types of Estimation**
- **Incomplete Data**

Basic Idea of EM Algorithm

- **Observed mixture**

-- Maximum likelihood estimation of parameters straightforward if case for component distributions

- **Hidden mixture**

-- Parameter estimation much more complex

-- Exploit simpler structure of observed mixture (via conditional expectation)

Types of Estimation

More approach than "algorithm"

- **Maximum likelihood estimation (MLE)**

-- Primary use

- **Bayesian**

-- Bayesian orientation ("precursor" of MCMC)

-- Maximize posterior density estimation

-- Maximum penalized likelihood estimation

Incomplete Data

EM algorithm applies to data that are "incomplete"

- **Mixtures**
- **Missing data/ "imputation"**
- **Grouped data**
- **Censored or truncated data**
- **Random effects**
- **Hidden/Latent variables**

(2) Description & Properties of EM Algorithm

- **Description of EM Algorithm**

  -- Expectation Step (E-Step)

  -- Maximization Step (M-Step)

- **Properties of EM Algorithm**

  -- Monotonicity

  -- Convergence

Description of EM Algorithm

- **Conceptual example**

Mixture of two distributions with density function

$$f(x) = (1 - w)\, f_0(x) + w\, f_1(x)$$

For now, assume that only mixing proportion w is unknown

-- i.e., conditional densities $f_i(x)$, $i = 0, 1$, completely specified

- **Observed data**

Given observed random sample from density f : $x_1, x_2, \ldots, x_n$

Log likelihood for observed data, $\log L(w)$:

$$\log L(w) = \log[(1 - w) f_0(x_1) + w f_1(x_1)] + \cdots + \log[(1 - w) f_0(x_n) + w f_1(x_n)]$$

Differentiating log likelihood function with respect to w and equating result to zero does not yield explicit solution for mixing proportion w

- **Unobservable data**

Let $i_j = 1$ if x_j sampled from density f_1 ($i_j = 0$ otherwise)

If $i_1, i_2, \ldots, i_n$ were observable, then MLE of w would be:

$$(i_1 + i_2 + \cdots + i_n) / n$$

"Complete-data" log likelihood function, $\log L_C(w)$:

$$\log L_C(w) = (1 - i_1)\,[\log(1 - w) + \log f_0(x_1)] + i_1\,[\log(w) + \log f_1(x_1)]$$

$$+ \cdots + (1 - i_n)\,[\log(1 - w) + \log f_0(x_n)] + i_n\,[\log(w) + \log f_1(x_n)]$$

- **Eliminating unobservables**

Let $I_1, I_2, \ldots, I_n$ denote random variables corresponding to unobservable data $i_1, i_2, \ldots, i_n$

To eliminate unobservables, take conditional expectation of $I_1, I_2, \ldots, I_n$ given observed data $x_1, x_2, \ldots, x_n$

In case of mixture, complete data log likelihood function is linear in unobservable data $i_1, i_2, \ldots, i_n$

So only need to calculate:

$$\mathrm{E}(I_j \mid x_1, x_2, \ldots, x_n) = \Pr\{I_j = 1 \mid x_1, x_2, \ldots, x_n\}$$

$$= [w\, f_1(x_j)] \,/\, [(1 - w)\, f_0(x_j) + w\, f_1(x_j)]$$

[Bayes' Theorem: posterior probability jth member of sample from density f_1]

- **EM algorithm**

(k+1)th interation, $k = 0, 1, 2, \ldots$

-- Expectation Step (E-Step)

$$i_j^{(k)} = w^{(k)} f_1(x_j) \,/\, [(1 - w^{(k)}) f_0(x_j) + w^{(k)} f_1(x_j)]$$

-- Maximization Step (M-Step)

$$w^{(k+1)} = (i_1^{(k)} + i_2^{(k)} + \cdots + i_n^{(k)}) \,/\, n$$

Initial step of algorithm: Need to specify starting value for w, say $w^{(0)}$

Properties of EM Algorithm

- **Monotonicity**

Likelihood cannot decrease after an iteration of EM algorithm

-- Proof (concavity of logarithm & Jensen's inequality)

- **Convergence**

Under general conditions, convergence of EM algorithm to MLE

-- Issue of choosing starting values

- **Standard errors**

Not automatically produced (as with Newton-Raphson)

Approaches to obtain Hessian/observed information matrix

-- Direct numerical differential (if likelihood not too complicated)

-- Extraction of observed information matrix from EM algorithm (e.g., making use of "Missing Information principle")

-- Resampling

- **Rate of Convergence**

-- Methods for accelerating [e.g., generalized EM (GEM) algorithm]

(3) Environmental Applications of EM Algorithm

- **Example**

Mixture of two exponential distributions

-- Chico daily precipitation intensity

- **Example**

Mixture of two normal distributions

-- Yellowstone Geyser time between eruptions

Example: Mixture of two exponential distributions

- **EM algorithm for mixture of two exponentials (parameters w, σ_0, σ_1)**

E-step: Unchanged (same for any mixture)

M-step [Solution for $(k+1)$th step]: ($w^{(k+1)}$ as before)

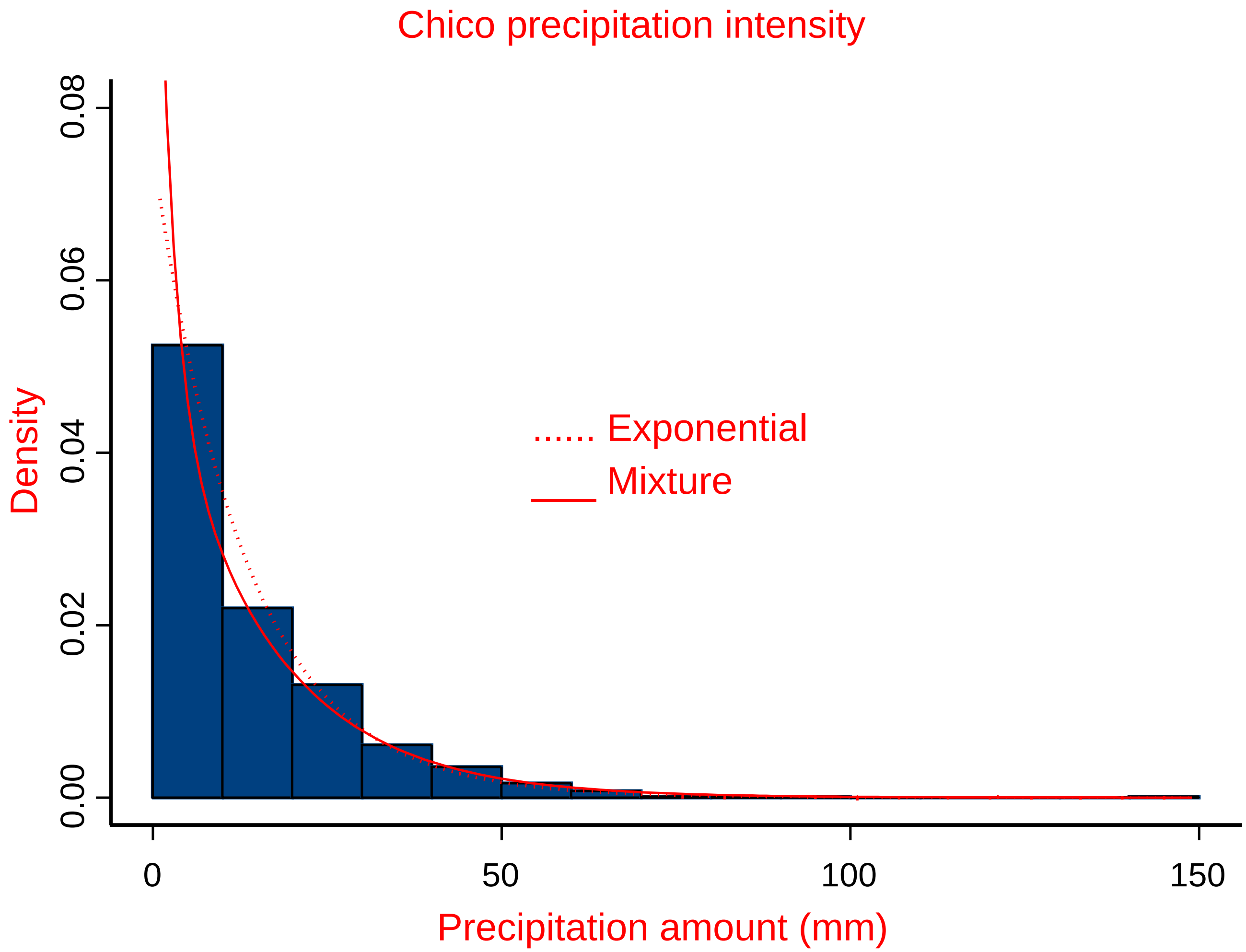
$$\sigma_1^{(k+1)} = [i_1^{(k)} x_1 + i_2^{(k)} x_2 + \cdots + i_n^{(k)} x_n] / [i_1^{(k)} + i_2^{(k)} + \cdots + i_n^{(k)}]$$

$$\sigma_0^{(k+1)} = [(1 - i_1^{(k)})\, x_1 + \cdots + (1 - i_n^{(k)})\, x_n] / [(1 - i_1^{(k)}) + \cdots + (1 - i_n^{(k)})]$$

- **Chico, CA, USA, daily precipitation intensity (January, 78 yrs.)**

n = 787, sample mean = 13.3596 mm

k	$w^{(k)}$	$\sigma_0^{(k)}$	$\sigma_1^{(k)}$	log L
0	0.50000	6.6798	20.0394	−2822.50817
1	0.51290	6.7995	19.5896	−2821.75749
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
86	0.81735	2.1704	15.8600	−2806.58896
87	0.81736	2.1703	15.8599	−2806.58896
SE's	(0.03296)	(0.3778)	(0.7757)	


Chico precipitation intensity
Density
0.00
0.02
0.04
0.06
0.08
0
50
100
150
Precipitation amount (mm)
...... Exponential
____ Mixture

Example: Mixture of Two Normal Distributions

- **EM algorithm for mixture of two normals (parameters $w, \mu_0, \mu_1, \sigma_0^2, \sigma_1^2$)**

E-step: Unchanged

M-step [Solution for $(k+1)$th step]: ($w^{(k+1)}$ as before)

$$\mu_1^{(k+1)} = [i_1^{(k)} x_1 + i_2^{(k)} x_2 + \cdots + i_n^{(k)} x_n] \,/\, [i_1^{(k)} + i_2^{(k)} + \cdots + i_n^{(k)}]$$

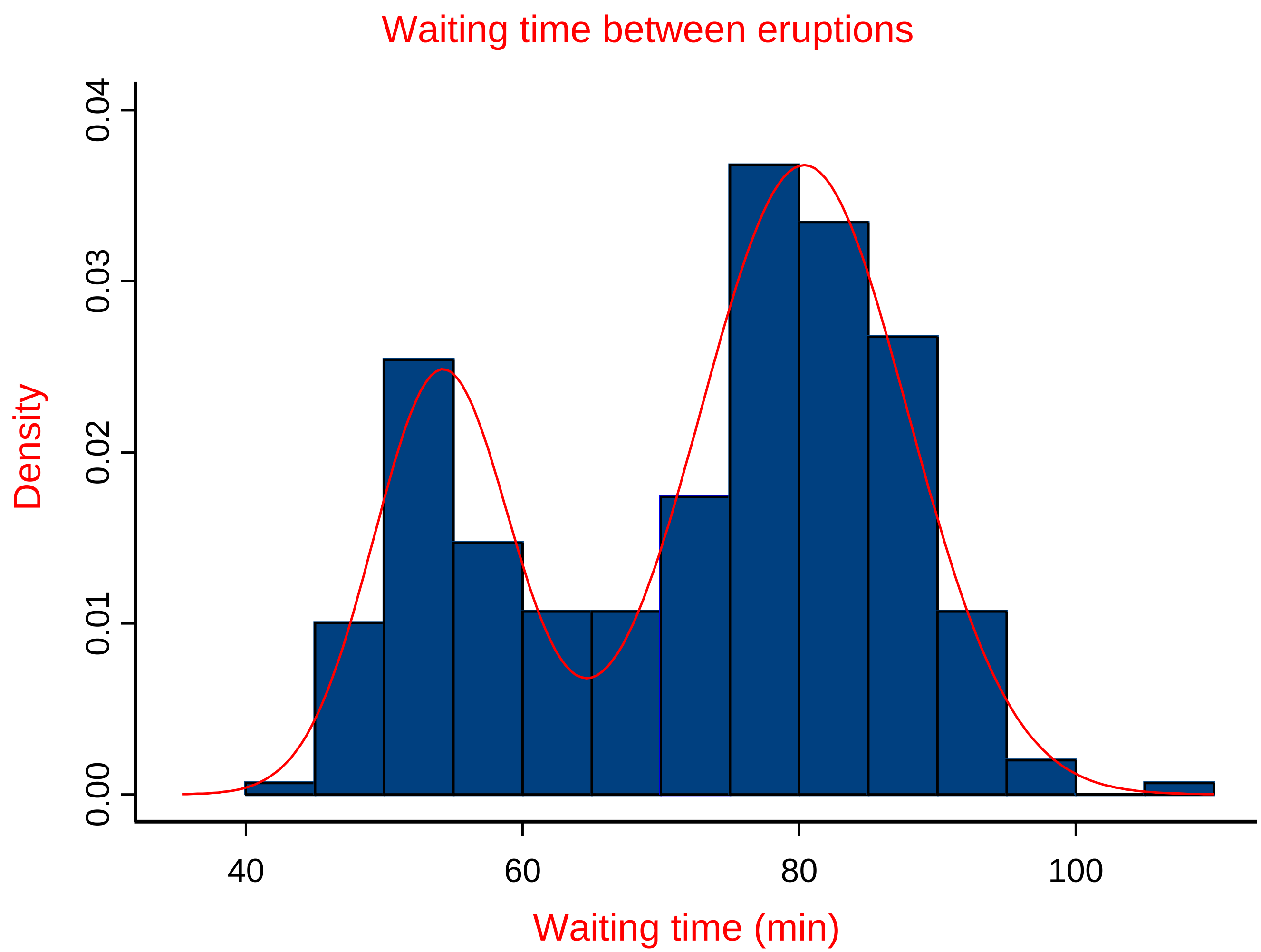
$$[\sigma_1^{(k+1)}]^2 = [i_1^{(k)} (x_1 - \mu_1^{(k+1)})^2 + \cdots + i_n^{(k)} (x_n - \mu_1^{(k+1)})^2] \,/\, [i_1^{(k)} + i_2^{(k)} + \cdots + i_n^{(k)}]$$

(similarly for $\mu_0^{(k+1)}$ & $[\sigma_0^{(k+1)}]^2$)

- **Yellowstone Geyser (Waiting Time Between Eruptions)**

n = 299, sample mean = 72.3144 min, standard dev. = 13.8671 min

k	$w^{(k)}$	$\mu_0^{(k)}$	$\mu_1^{(k)}$	$\sigma_0^{(k)}$	$\sigma_1^{(k)}$	log L
0	0.50000	65.3808	79.2479	12.0092	12.0092	−1208.30926
1	0.51087	64.8022	79.5069	13.0968	10.3160	−1200.61590
⋮	⋮	⋮	⋮	⋮	⋮	⋮
40	0.69238	54.2036	80.3611	4.9528	7.5069	−1157.54202
41	0.69238	54.2034	80.3609	4.9526	7.5071	−1157.54202
SE's	(0.03438)	(0.6831)	(0.6333)	(0.5183)	(0.5070)	


Waiting time between eruptions
Density
0.00
0.01
0.02
0.03
0.04
40
60
80
100
Waiting time (min)

(4) Maximum Likelihood Estimation for Mixtures

- **Point Estimation**

-- "Nonstandard" case

-- Likelihood can be infinite

-- Maximize largest local likelihood

- **Tests of Significance**

-- Likelihood ratio test (LRT)

No longer asymptotic chi-square distribution
Recalibrate (e.g., adjust degrees of freedom or resample)

- **Model Selection Criteria**

-- Akaike's Information Criterion (AIC) or Bayesian Information Criterion (BIC) still appear to be valid

Choose model for which AIC (or BIC) is minimum

$$\mathrm{AIC} = -2 \log L + 2p$$

$$\mathrm{BIC} = -2 \log L + p \log n$$

where p is number of parameters estimated,

L is maximized likelihood function,

n is sample size

-- Chico Precipitation Intensity ($n = 787$)

Model	p	Log Likelihood	AIC	BIC
Single exponential	1	−2827.089	5656.177	5660.846
Mixture of two exponentials	3	−2806.589	5619.178	5633.183

-- Yellowstone Geyser Waiting Time ($n = 299$)

Model	p	Log Likelihood	AIC	BIC
Single normal	2	−1210.488	2424.977	2432.378
Mixture of two normals	5	−1157.542	2325.084	2343.586

STOCHASTIC MODELING OF ENVIRONMENTAL TIME SERIES

Richard W. Katz

LECTURE 3

(1) High Frequency Approach to Time Series Modeling

(2) Low Frequency Approach to Overdispersed Time Series

(3) Low Frequency Approach: Observed Mixture/Covariate

(4) Low Frequency Approach: Hidden Mixture/Variable

(1) High Frequency Approach to Time Series Modeling

Basic Idea: Apparent overdispersion actually attributable to inadequate model for high frequency variations

- **High Frequency Approach to Modeling Precipitation**
- **Example: Chico Daily Precipitation**
- **Methodological Issues: Presence of Overdispersion**

High Frequency Approach to Modeling Precipitation

- **Chain-Dependent process**

$\{X_t$: $t = 1, 2, \ldots, T\}$ denotes amount of precipitation on tth day

-- Occurrence $\{J_t$: $t = 1, 2, \ldots, T\}$ modeled as two-state, first-order Markov chain ($J_t = 1$ if prec. occurs on tth day $J_t = 0$ otherwise)

(parameters P_{01}, P_{11} or π, d)

Number of wet days: $N(T) = J_1 + J_2 + \cdots + J_T$

-- Intensity (i.e., amount X_t given $J_t = 1$) assumed conditional i.i.d. given $\{J_t\}$

(intensity mean μ and variance σ^2)

-- Variance of total precipitation ($S_T = X_1 + X_2 + . . . + X_T$)

$$\mathrm{Var}(S_T) = \mathrm{E}[N(T)]\ \mathrm{Var}[X_t \mid J_t = 1] + \mathrm{Var}[N(T)]\ \{\mathrm{E}[X_t \mid J_t = 1]\}^2$$

$$\approx T\{\pi\sigma^2 + \pi(1 - \pi)\ [(1 + d)/(1 - d)]\ \mu^2\}$$

- **Extensions of Chain-Dependent process**

-- High-order Markov Chain (e.g., order 2, 3, . . .)

-- Conditionally dependent intensities [AR(1) process for transformed intensities]

-- Conditionally non-identically distributed intensities (distribution depends on J_{t-1})

Example: Chico Daily Precipitation

- **Chico, CA, USA**

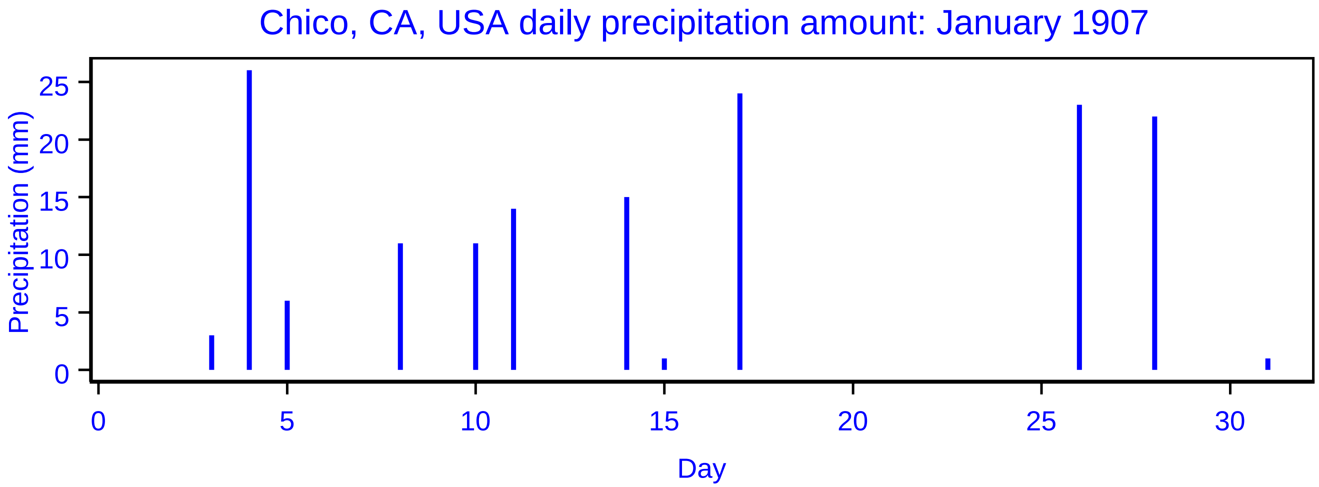
-- January time series of daily precipitation amount (mm)

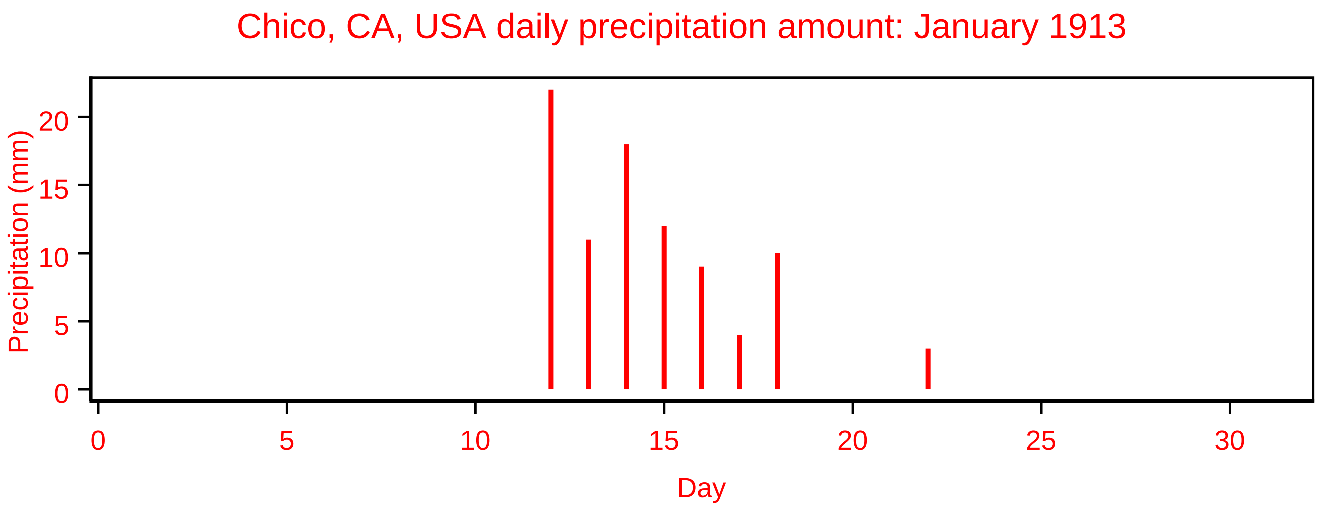
-- 78 total years during time period 1907-1998 (4 years eliminated because of missing observations)

- **Climate**

-- Winter is wet season (large annual cycle with summer dry)

-- Relatively persistent climate (precipitation governed by pressure patterns over Pacific Ocean adjacent to California coast)


Chico, CA, USA daily precipitation amount: January 1907
Precipitation (mm)
0
5
10
15
20
25
0
5
10
15
20
25
30
Day


Chico, CA, USA daily precipitation amount: January 1913
Precipitation (mm)
0
5
10
15
20
0
5
10
15
20
25
30
Day

- **Results for Chico**

Markov chain order	Intensity correlation	Identical intensity dist.	Number param.	$[\mathrm{Var}(S_T)]^{1/2}$
1	No	Yes	4	68.7 mm
2	No	Yes	6	71.1 mm
2	Yes	Yes	7	74.5 mm
2	Yes	No	9	76.4 mm
3	Yes	No	13	79.5 mm
4	Yes	No	21	80.6 mm
			***Observed*:**	**88.6 mm**

Methodological Issues: Presence of Overdispersion

- **Model selection criteria**

-- How do LRT, AIC, or BIC perform when overdispersion ignored?

-- Tend to overfit (because low-frequency overdispersion can induce apparent high-frequency effects)

-- Difficult to protect against

-- Effects on analysis of variance estimates (biased estimates of "potential predictability"?)

-- Compelled to explicitly model overdispersion

(2) Low Frequency Approach to Overdispersed Time Series

Basic idea

-- Two climate regimes between which parameters of daily precipitation time series randomly shift from year to year

-- Regimes may be treated as either *observed* or *hidden*

- **Description of Low Frequency Model**
- **Properties of Low Frequency Model**

Description of Low Frequency Model

- **Conditioning variable**

-- Two-state conditioning variable I with $w = \Pr\{I = 1\}$

-- Low frequency in that I only varies from year to year, *not* day to day

- **Conditional chain-dependent process**

-- Conditional on $I = i$

Markov chain for daily occurrence has transition probs. $P_{01}(i)$, $P_{11}(i)$ (or π_i, d_i)

Daily intensities i.i.d. with mean μ_i, variance σ_i^2

Properties of Low Frequency Model

- **Variance of total precipitation**

$$\mathrm{Var}(S_T) = (1 - w)\,\mathrm{Var}(S_T \mid I = 0) + w\,\mathrm{Var}(S_T \mid I = 1)$$

$$+\, w(1 - w)\,\{\mathrm{E}(S_T \mid I = 1) - \mathrm{E}(S_T \mid I = 0)\}^2$$

Here

$$\mathrm{E}(S_T \mid I = i) = T\pi_i\,\mu_i$$

$$\mathrm{Var}(S_T \mid I = i) \approx T\{\pi_i\,\sigma_i^2 + \pi_i(1 - \pi_i)\,[(1 + d_i)/(1 - d_i)]\,\mu_i^2\}$$

- **Interpretation**

-- Autocorrelation function for single chain-dependent process:

$$\mathrm{Corr}(X_t, X_{t+l}) = [\pi(1-\pi)\, d^{\,l}\mu^2] \,/\, \mathrm{Var}(X_t)$$

where

$$\mathrm{Var}(X_t) = \pi\sigma^2 + \pi(1-\pi)\mu^2$$

-- Autocovariance function for mixture of two chain-dependent processes:

$$\mathrm{Cov}(X_t\,, X_{t+l}) = (1-w)\,\pi_0(1-\pi_0)\, d_0^{\,l}\mu_0^{\,2} + w\,\pi_1(1-\pi_1)\, d_1^{\,l}\mu_1^{\,2}$$

$$+\, w(1-w)\,(\pi_1\,\mu_1 - \pi_0\,\mu_0)^2$$

Resembles second-order Markov chain, along with another term for shift in conditional mean of daily precipitation

(3) Low Frequency Approach: Observed Mixture/Covariate

Parameter estimation straightforward for observed covariate, but useful in formulating EM algorithm for hidden variable case

- **Parameter Estimation & Model Selection**

- **Application to Chico Precipitation**

  -- Also 13 locations across California

- **Overdispersion Results**

Parameter Estimation and Model Selection

- **Parameter estimation**

-- Occurrences

MLE's for transition probs. of Markov chain are based on transition counts:

e.g., $n_{01}/(n_{00} + n_{01})$ is MLE for P_{01} (n_{jk} no. times state j is followed by state k)

-- Intensities

Assume i.i.d. with power transform distribution:

$X_t^* = X_t^{\,s}$ normally distributed with mean μ^* and variance $(\sigma^*)^2$ (e.g., $s = ¼$)

MLE's just sample means and variances of transformed intensities

Log likelihood function for chain-dependent process:

$$\log L[P_{01}, P_{11}, \mu^*, (\sigma^*)^2] = \sum_j [n_{j0} \log(1 - P_{j1}) + n_{j1} \log P_{j1}]$$

$$- (n_{\cdot 1}/2) \log[2\pi(\sigma^*)^2] - \{1/[2(\sigma^*)^2]\} \sum_t (x_t^* - \mu^*)^2$$

where $n_{\cdot 1} = n_{01} + n_{11}$

$\sum_j$ **is over** $j = 0, 1$

$\sum_t$ **is over** $n_{\cdot 1}$ **terms for which** $x_t > 0$

- **Application to Chico January Precipitation**

-- Observed covariate

January mean sea level pressure at grid point off coast of California:

$I = 1$ if greater than sample mean, $I = 0$ otherwise

-- Model fitting for Chico

Conditioning	π	d	μ (mm)	σ (mm)	μ^*	σ^*
None	0.329	0.359	13.36	14.68	1.699	0.521
$I = 0$	0.413	0.334	15.16	15.50	1.777	0.515
$I = 1$	0.253	0.349	10.75	12.99	1.585	0.509

-- Model selection for Chico

Issue of how to choose "sample size" n for BIC:

Take $n = 78$ years (not $n = 78 \cdot 31 = 2418$ days)

More parsimonious to parameterize in terms of π, d (instead of P_{01}, P_{11})

AIC & BIC both select model in which π & μ^* varied with circulation index I

January precipitation at 13 sites across California (including Chico), still conditioning on same covariate: no. sites for which vary parameter with I

	π	d	μ^*	σ^*
AIC	12	4	9	1
BIC	9	1	4	0

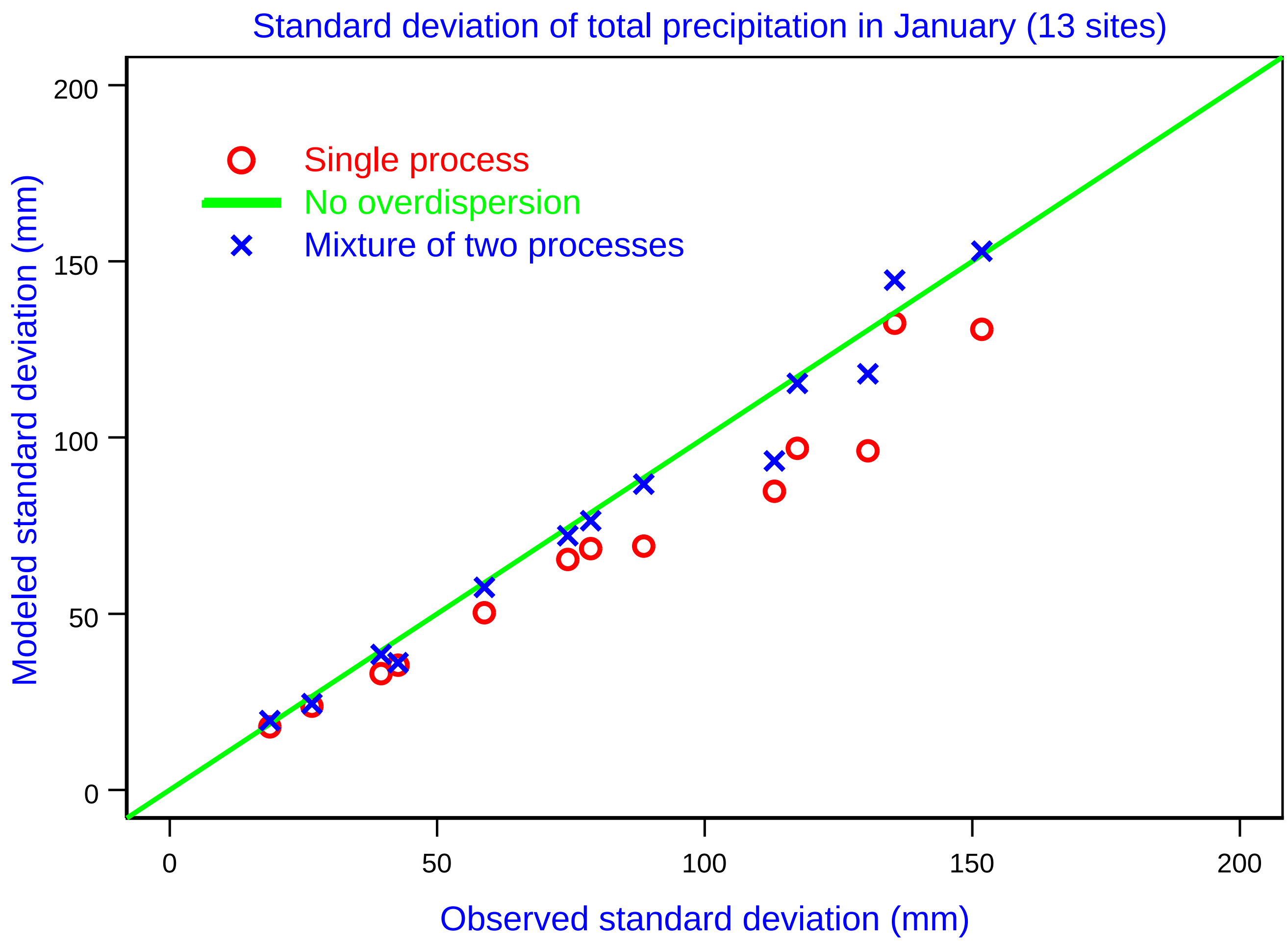
- **Overdispersion Results**

-- Chico Jan. precipitation (St. dev. of monthly total)

Single process	**69.2 mm**	**(Observed 88.6 mm)**
Mixture	**86.8 mm**	**(Observed 88.6 mm)**
Given $I = 0$	**80.8 mm**	**(Observed 82.4 mm)**
Given $I = 1$	**52.2 mm**	**(Observed 58.9 mm)**

-- All 13 sites in California

Apparent reduction in extent of overdispersion for majority of sites



Standard deviation of total precipitation in January (13 sites)
Single process
No overdispersion
Mixture of two processes
Modeled standard deviation (mm)
Observed standard deviation (mm)
0
50
100
150
200

(4) Low Frequency Approach: Hidden Mixture/Variable

Treat conditioning variable as hidden (ignore any observed covariates)

- **Formulation of EM algorithm**
- **Model Fitting & Selection**
- **Overdispersion Results**

Formulation of EM Algorithm

- **Complete-data likelihood function**

Observed data:

$\{x_t(m): t = 1, 2, . . ., T; m = 1, 2, . . ., M\}$ **prec. amount on tth day of mth yr.**

Unobservable data:

$\{i(m), m = 1, 2, . . ., M\}, i(m) = 0, 1$ **denoting index state for mth year**

Vector of parameters: $\Theta = (w, \Theta_0 , \Theta_1)$

where $\Theta_i = [P_{01}(i), P_{11}(i), \mu_i^*, (\sigma_i^*)^2], \; i = 0, 1$

Log-likelihood function (complete data):

$$\log L_C(\Theta) = \sum_m \{[1 - i(m)] \log[(1 - w)\, L_m(\Theta_0)] + i(m) \log[w\, L_m(\Theta_1)]\}$$

where $L_m(\Theta_i)$ denotes likelihood function for chain-dependent process given index state i evaluated for daily precipitation time series in mth year

MLEs for model parameters (complete data):

w: $[\sum_m i(m)] \,/\, M$

$P_{j1}(1),\; j = 0, 1$: $[\sum_m i(m)\, n_{j1}(m)] \,/\, [\sum_m i(m)\, n_{j\cdot}(m)]$

μ_1^*: $[\sum_m i(m)\, s_1(m)] \,/\, [\sum_m i(m)\, n_{\cdot 1}(m)]$

$(\sigma_1^*)^2$: $[\sum_m i(m)\, s_2(m)] \,/\, [\sum_m i(m)\, n_{\cdot 1}(m)] - (\mu_1^*)^2$

Here $s_1(m) = \sum_t x_t(m),\;\; s_2(m) = \sum_t [x_t(m)]^2$

EM algorithm:

(1) E-step

-- Estimate posterior probabilities by

$$\Pr\{I(m) = 1\} = w\, L_m(\Theta_1) \,/\, [(1 - w)\, L_m(\Theta_0) + w\, L_m(\Theta_1)],$$

$$m = 1, 2, \ldots, M$$

(2) M-step

-- Replace *i*(*m*) with estimated posterior probability in expressions of MLE's for complete data case

Model Fitting and Selection

- **Results for Chico**

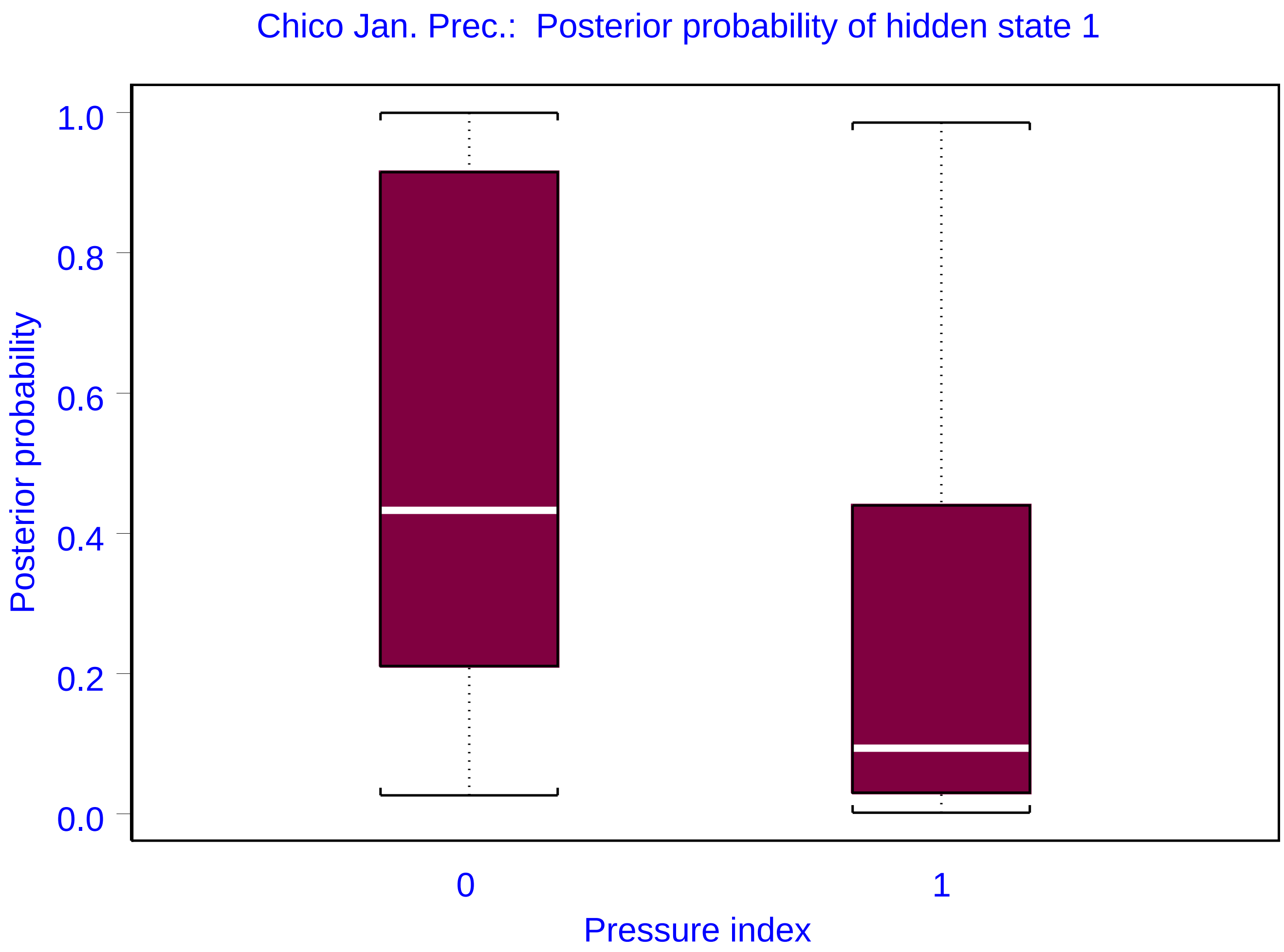
-- Fitted models

Model	w	$P_{01}(i)$, $i = 0, 1$	$P_{11}(i)$, $i = 0, 1$	μ_i^*, $i = 0, 1$	σ_i^*, $i = 0, 1$
Completely constrained	------	0.211 0.211	0.571 0.571	1.702 1.702	0.521 0.521
$P_{01}(0) = P_{01}(1)$ $\sigma_0^* = \sigma_1^*$	0.371	0.211 0.211	0.505 0.660	1.585 1.859	0.503 0.503
Completely unconstrained	0.378	0.214 0.205	0.505 0.662	1.587 1.861	0.504 0.502

-- Model selection

Again taking n = 78 years

Model	Number parameters	Log likelihood	AIC	BIC
Completely constrained	4	−1924.246	3856.49	3865.92
$P_{01}(0) = P_{01}(1)$ $\sigma_0^* = \sigma_1^*$	7	−1912.393	3838.79	3855.28
Completely unconstrained	9	−1912.386	3842.77	3863.98


Chico Jan. Prec.: Posterior probability of hidden state 1
Posterior probability
1.0
0.8
0.6
0.4
0.2
0.0
0
1
Pressure index

Overdispersion Results

- **Estimated variance of Chico January total precipitation:**

Completely constrained	**70.4 mm**
“Optimal model”	**89.4 mm**
Completely unconstrained	**88.9 mm**
Observed	**88.6 mm**

STOCHASTIC MODELING OF ENVIRONMENTAL TIME SERIES

Richard W. Katz

LECTURE 4

(1) Hidden Markov Models: Properties

(2) Hidden Markov Models: Motivational Example

(3) EM Algorithm for Hidden Markov Model

(1) Hidden Markov Models: Properties

- **Definition of Hidden Markov Model**

-- Finite-state observed variable

- **Properties of Hidden Markov Model**

-- Binary hidden Markov model

Definition of Hidden Markov Model

- **General case**

-- **Observed sequence of variables** $\{X_t: t = 1, 2, \ldots, T\}$

-- **Finite-state, first-order Markov chain** $\{J_t: t = 1, 2, \ldots, T\}$

Transition probabilities $P_{ij} = \Pr\{J_{t+1} = j \mid J_t = i\}, \; i, \; j = 0, 1, \ldots$

-- $\{X_t\}$ **conditionally i.i.d. given** $\{J_t\}$

Conditional distribution function $F_i(x) = \Pr\{X_t \leq x \mid J_t = i\}, \; i = 0, 1, \ldots$

[conditional mean $\mu_i = E(X_t \mid J_t = i)$**, variance** $\sigma_i^2 = \mathrm{Var}(X_t \mid J_t = i)$**]**

- **Finite-state observed variable**

-- This case receives most attention

$$\Pr\{X_t = x \mid J_t = i\} = p_i(x), \; x = 0, 1, \ldots; \; i = 0, 1, \ldots$$

-- For example, binary variable ($x = 0, 1$)

Set $p_i = p_i(1) = \Pr\{X_t = 1 \mid J_t = i\}, \; i = 0, 1$

Conditional mean $\mu_i = p_i, \; i = 0, 1, \ldots$

Conditional variance $\sigma_i^2 = p_i(1 - p_i), \; i = 0, 1, \ldots$

Properties of Hidden Markov Model

- **Properties of Markov chain (two-state, first-order)**

-- Variables: $\{J_t: t = 1, 2, \ldots, T\}$, $J_t = 0, 1$

-- Transition probs.: $P_{i1} = \Pr\{J_{t+1} = 1 \mid J_t = i\}$, $i = 0, 1$

-- Initial prob.: $\delta = \delta(1) = \Pr\{J_1 = 1\}$, $\delta(0) = 1 - \delta$

-- Stationary process: Set $\delta = \pi = P_{01} / (P_{10} + P_{01})$

-- Moments (stationary case): $\mathrm{E}(J_t) = \pi$, $\mathrm{Var}(J_t) = \pi(1 - \pi)$,

$\mathrm{Corr}(J_t, J_{t+l}) = d^l$, where $d = P_{11} - P_{01}$

-- l-step transition probs.:

Define $P_{i1}^{(l)} = \Pr\{J_{t+l} = 1 \mid J_t = i\}$, $i = 0, 1$, $l = 1, 2, \ldots$

Then $P_{01}^{(l)} = \pi\,(1 - d^{\,l})$, $P_{11}^{(l)} = \pi + (1 - \pi)\, d^{\,l}$,

-- Conditional independence

$$\Pr\{ J_{t+1} = j \mid J_t = i\} = \Pr\{ J_{t+1} = j \mid J_t = i, J_{t-1}, \ldots, J_1\}$$

But $\Pr\{ J_{t+1} = j \mid J_{t-1} = i\}$ still depends on i

-- Run length (geometric distribution)

$$\Pr(\text{run of state } i \text{ of length } l\} = P_{ii}^{\,l-1} P_{i,\,1-i}\,, \quad l = 1, 2, \ldots; \quad i = 0, 1$$

- **Hidden two-state Markov chain**

General observed variable (i.e., not necessarily discrete)

-- Moments

$E(X_t) = (1 - \pi)\mu_0 + \pi\mu_1,$

$\mathrm{Var}(X_t) = (1 - \pi)\sigma_0^2 + \pi\sigma_1^2 + \pi(1 - \pi)(\mu_1 - \mu_0)^2$

$\mathrm{Corr}(X_t, X_{t+l}) =$

$$\{[\pi(1 - \pi)(\mu_1 - \mu_0)^2] / [(1 - \pi)\sigma_0^2 + \pi\sigma_1^2 + \pi(1 - \pi)(\mu_1 - \mu_0)^2]\} d^l$$

-- Reduces to independent two-component mixture model if $d = 0$

- **Binary Hidden Two-State Markov Model**

Assume two-state observations, as well as two-state Markov chain

-- Moments

$$E(X_t) = \Pr\{X_t = 1\} = (1 - \pi)p_0 + \pi p_1,$$

$$\mathrm{Var}(X_t) = (1 - \pi)p_0(1 - p_0) + \pi p_1(1 - p_1) + \pi(1 - \pi)(p_1 - p_0)^2$$

$$\mathrm{Corr}(X_t, X_{t+l}) =$$

$$\left\{[\pi(1 - \pi)(p_1 - p_0)^2] \big/ [(1 - \pi)p_0(1 - p_0) + \pi p_1(1 - p_1) + \pi(1 - \pi)(p_1 - p_0)^2]\right\} d^{\,l}$$

-- Run length

Can be mixture of two geometric distributions

Capable of producing longer or shorter mean run length than comparable Markov chain model (i. e., $\Pr\{X_t = 1\}$ same for both models)

Examples: See MacDonald & Zucchini text (pp. 88-89)

-- Reduces to Markov chain (e. g., if $p_0 = 0$, $p_1 = 1$)

-- Hidden Markov model is not necessarily Markov chain

Examples: See Guttorp text (p. 105) & MacDonald & Zucchini text (pp. 82-83)

(2) Hidden Markov Models: Motivational Example

- **Stochastic "weather generator"**

  -- Markov chain not necessarily "hidden" (relax conditional independence condition)

  -- Definition of model

  $\{(X_t, J_t)$: $t = 1, 2, . . ., T\}$ where X_t denotes temperature (e.g., minimum or maximum) and J_t denotes precipitation occurrence on tth day at same location

  $\{J_t\}$ is assumed first-order, two-state Markov chain (parameters P_{01}, P_{11} or π, d)

Assume $X_t \mid J_t = i$ is distributed as $N(\mu_i, \sigma_i^2)$, $i = 0, 1$

Given $J_t = i$, set $Z_t = (X_t - \mu_i) / \sigma_i$ (random standardization)

Conditional on $\{J_t\}$, assume $\{Z_t: t = 1, 2, \ldots\}$ as AR(1) process with conditional first-order autocorrelation coefficient φ; that is,

$$Z_{t+1} = \varphi Z_t + \varepsilon_t, \text{ where } \varepsilon_t \text{ is } N(0, 1 - \varphi^2),$$

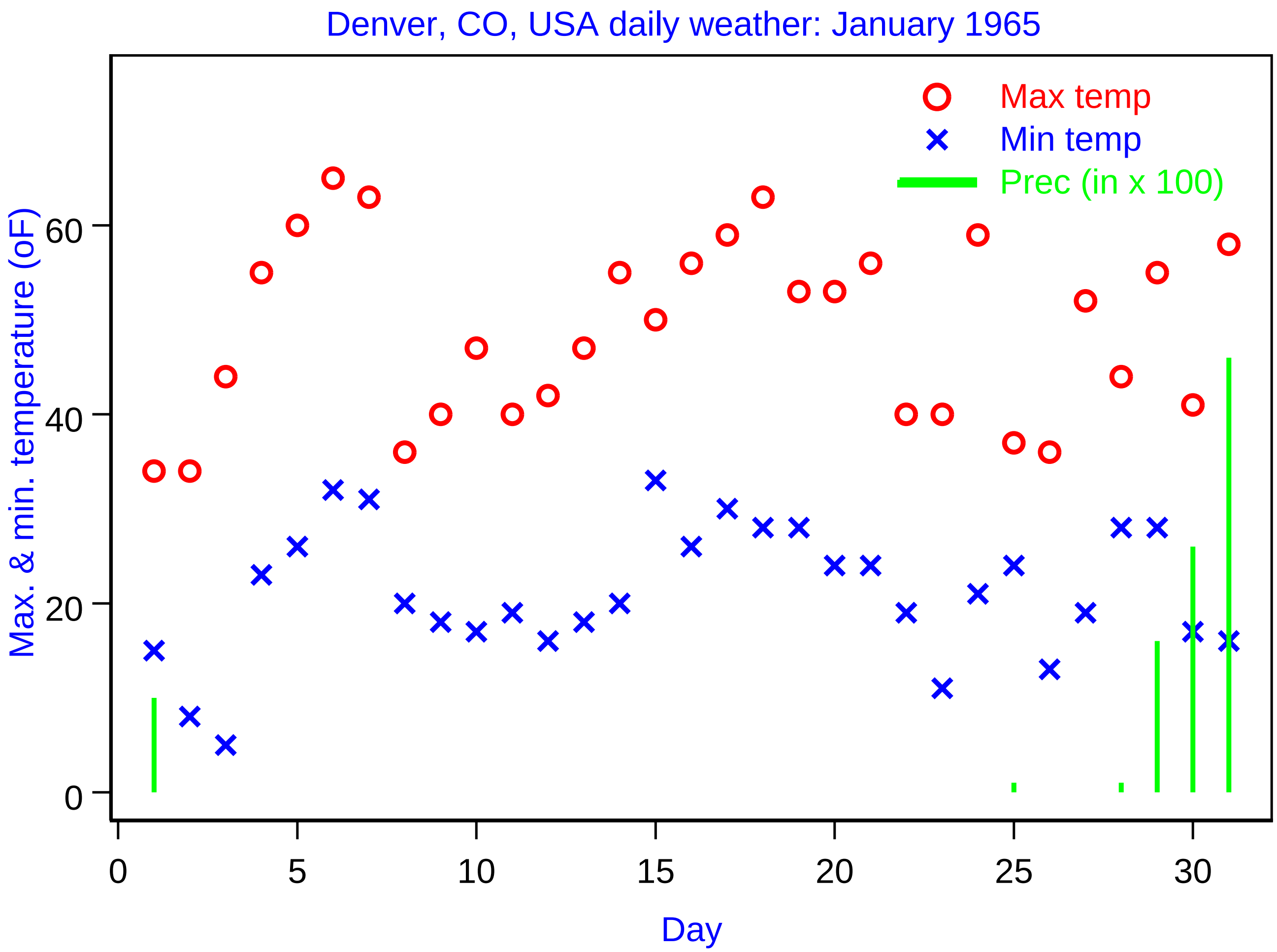
-- Properties of model

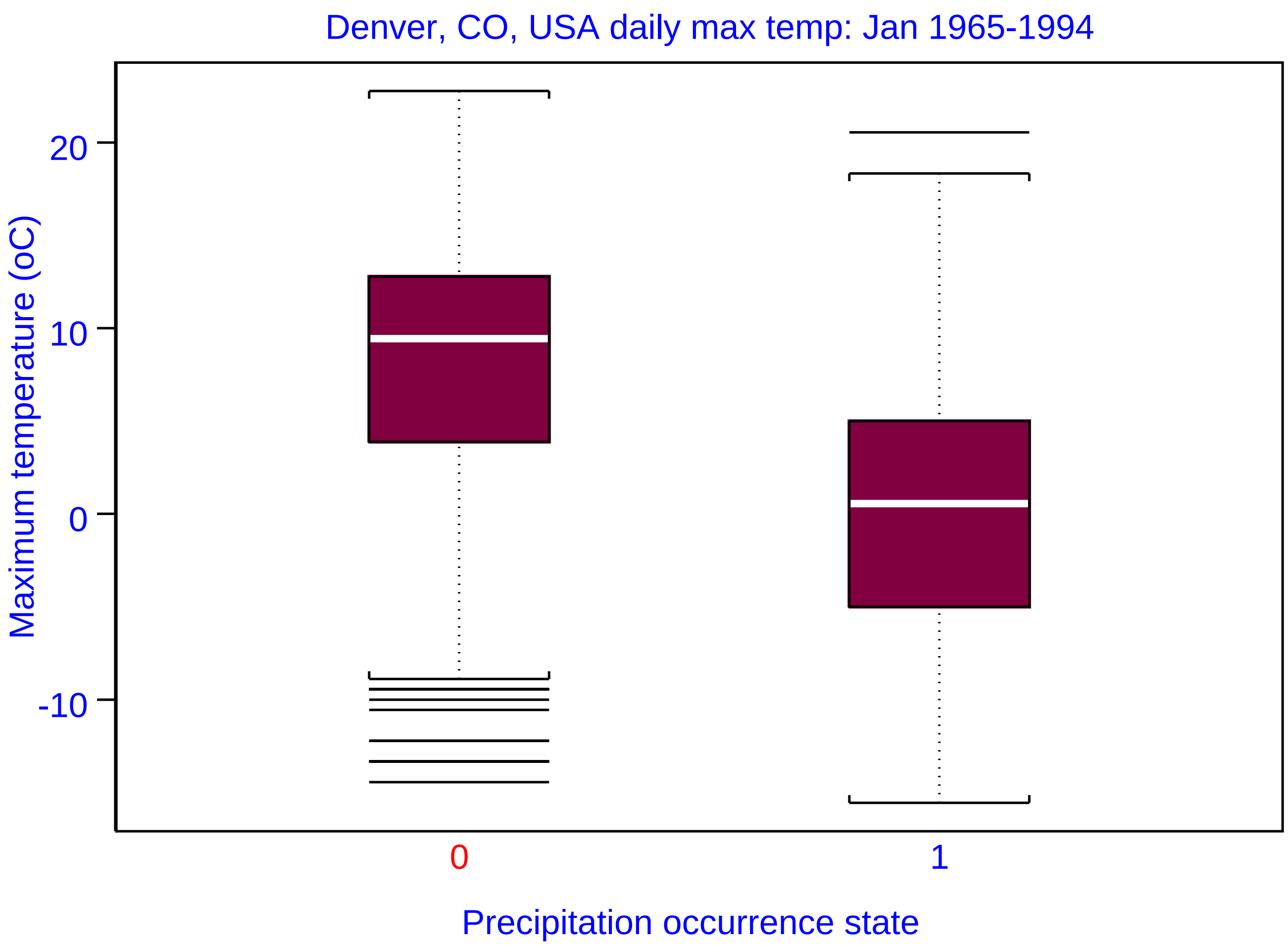
$$E(X_t) = (1 - \pi)\mu_0 + \pi\mu_1,$$

$$\text{Var}(X_t) = (1 - \pi)\sigma_0^2 + \pi\sigma_1^2 + \pi(1 - \pi)(\mu_1 - \mu_0)^2$$

$$\text{Cov}(X_t, X_{t+l}) =$$

$$\varphi^l [(1 - \pi)\sigma_0^2 + \pi\sigma_1^2] + \pi(1 - \pi)[d^l (\mu_1 - \mu_0)^2 - \varphi^l (1 - d^l)(\sigma_1 - \sigma_0)^2]$$


Denver, CO, USA daily weather: January 1965
Max temp
Min temp
Prec (in x 100)
Max. & min. temperature (oF)
0
20
40
60
0
5
10
15
20
25
30
Day


Denver, CO, USA daily max temp: Jan 1965-1994
Maximum temperature (oC)
20
10
0
-10
0
1
Precipitation occurrence state

- **Application**

-- Denver, Colorado, USA January daily maximum temperature (◦C) and precipitation occurrence, 1965-1994

Precipitation occurrence: $\pi = 0.178$, $d = 0.219$

Sample statistics	Unconditional	Given $J_t = 0$	Given $J_t = 1$
Mean (◦C)	6.5	8.0	0.0
Std. dev. (◦C)	7.4	6.7	7.0
Autocorrelation	0.62**	0.55*	0.55*

* Constrained to be same for $J_t = 0$ & $J_t = 1$

** Model calculation yields 0.49

(3) EM Algorithm for Hidden Markov Model

- **Software**

  -- Speech recognition orientation

- **Tutorial**

  -- Companion to software

- **Formulation of EM Algorithm**

Software for Fitting Hidden Markov Models

- **Download from WWW**

  -- Web address: ***www.cfar.umd.edu/~kanungo/software/software.html***

  -- Author: Tapas Kanungo, Center for Automation Research, Univ. of Maryland

  -- Software written in C

  -- Treats finite-state observed variable with finite-state, first-order hidden Markov chain

Tutorial on Hidden Markov Models

- **PDF File (*hmmtut.pdf*)**

  -- Posted on EPFL site for this course

  -- Notation

  Makes use of quite different notation

  (from engineering/speech processing literature)

- **Speech Recognition**

-- Utterance

Observed sequence is acoustic signal corresponding to "utterance"

-- Words

Assume utterance comes known vocabulary of "words"

-- State of hidden Markov chain represents configuration of vocal tract

-- Training Problem

For each word in vocabulary, fit hidden Markov chain to utterance

-- Classification Problem

Use training model to compute probability of each word (choose word with highest probability)

-- "Localized decoding"

Determine state of Markov chain at time (i.e., J_t) with highest posterior probability

-- "Global decoding"

Determine sequence of states of Markov chain (i.e., $J_1, J_2, \ldots, J_T$) with highest joint posterior probability

Formulation of EM Algorithm for Hidden Markov Model

Binary observed variable with two-state hidden Markov chain

- **Data**

  -- **Observed data** $\{x_t: t = 1, 2, \ldots, T\}$, $x_t = 0, 1$

  -- **Unobservable data** $\{j_t: t = 1, 2, \ldots, T\}$

  Here $j_t = 0$ **or** 1 **(state of hidden Markov chain** J_t**)**

  -- **Notation:** $n_{ij}(t) = 1$ **if** $j_t = i$ **and** $j_{t+1} = j$ **[**$n_{ij}(t) = 0$ **otherwise]**, $t = 1, 2, \ldots, T-1$

  $n_i(t) = 1$ **if** $j_t = i$ **[**$n_i(t) = 0$ **otherwise]**, $t = 1, 2, \ldots, T$

- **Complete Data Log-Likelihood Function**

$$\log L_C(\delta, P_{01}, P_{11}, p_0, p_1) = \log \delta(j_1) + \sum_{i,j,t} n_{ij}(t) \log P_{ij} + \sum_{i,t} n_i(t) \log p_i(x_t)$$

- **Incomplete Data Likelihood Function**

$$L_T = \Pr\{X_1 = x_1, \ldots, X_T = x_T; \delta, P_{01}, P_{11}, p_0, p_1\}$$

-- Computing L_T

$$L_T = a_T(0) + a_T(1)$$

where

$$a_t(i) = \Pr\{X_1 = x_1, \ldots, X_t = x_t, J_t = i\},$$

$$i = 0, 1; t = 1, 2, \ldots, T$$

-- "Forward recursion" for $\{a_t(i)\}$

(1) Initial step

$$a_1(i) = \delta(i)\, p_i(x_1), \quad i = 0, 1$$

(2) Iterative step

$$a_{t+1}(i) = p_i(x_{t+1})\, [a_t(0)\, P_{0i} + a_t(1)\, P_{1i}],$$

$$i = 0, 1; \quad t = 1, 2, \ldots, T - 1$$

[in practice, issue of avoiding underflow by rescaling, etc.]

- **E-Step of EM Algorithm**

-- Need to determine terms of form $\Pr\{J_t = i \mid x_1, x_2, \ldots, x_T\}$

Define

$$b_t(i) = \Pr\{X_{t+1} = x_{t+1}, \ldots, X_T = x_T \mid J_t = i\}$$

$$i = 0, 1; \quad t = 1, 2, \ldots, T$$

with convention that $b_T(i) = 1,\ i = 0, 1$

Then $\Pr\{J_t = i \mid x_1, x_2, \ldots, x_T\} = [a_t(i)\, b_t(i)] / L_T$

$i = 0, 1; \quad t = 1, 2, \ldots, T$

-- "Backward recursion" for $\{b_t(i)\}$

(1) Initial step

$$b_T(i) = 1,\ i = 0, 1$$

(2) Iterative step

$$b_t(i) = p_0(x_{t+1})\, b_{t+1}(0)\, P_{i0} + p_1(x_{t+1})\, b_{t+1}(1)\, P_{i1},$$

$$i = 0, 1;\ \ t = T-1, T-2, \ldots, 1$$

[in practice, issue of avoiding underflow by rescaling, etc.]

- **M-Step**

-- Parameter estimates can be obtained via output of forward & backward recursions

(1) Estimate of δ

$$j_1^{(k)} = [a_1(1)\ b_1(1)] \,/\, L_T$$

(estimate of $\Pr\{J_1 = 1 \mid x_1, x_2, ..., x_T\}$)

(2) Estimates of p_i ($i = 0, 1$)

$$\left[\sum_t n_i(t)^{(k)} x_t\right] / \sum_t n_i(t)^{(k)}$$

where $n_i(t)^{(k)} = [a_t(i)\, b_t(i)] / L_T,\ \ i = 0, 1;\ \ t = 1, 2, \ldots, T$

($n_i(t)^{(k)}$ **is estimate of** $\Pr\{J_t = i \mid x_1, x_2, \ldots, x_T\}$**)**

(3) Estimates of P_{i1} ($i = 0, 1$)

$$\sum_t n_{i1}(t)^{(k)} / \sum_t \left[n_{i0}(t)^{(k)} + n_{i1}(t)^{(k)}\right]$$

(summations over $t = 1, 2, \ldots, T - 1$**)**

where $n_{ij}(t)^{(k)} = \left[P_{ij}\, p_j(x_{t+1})\, a_t(i)\, b_{t+1}(j)\right] / L_T,\ \ i = 0, 1;\ \ j = 0, 1;\ \ t = 1, 2, \ldots, T - 1$

($n_{ij}(t)^{(k)}$ **is estimate of** $\Pr\{J_t = i, J_{t+1} = j \mid x_1, x_2, \ldots, x_T\}$**)**

- **Other Issues**

-- Standard errors

More complex than for independent mixtures

-- Model selection

Same issues arise as with independent mixtures

-- Assumptions about hidden Markov chain

Distribution of initial state: fixed, arbitrary or stationary?

-- Alternative of direct maximization

Requires efficient calculation of incomplete data likelihood function

-- Posterior probabilities

Localized decoding: Already computed as part of EM algorithm

Global decoding: Can use Viterbi Algorithm (a form of dynamic programming)

STOCHASTIC MODELING OF ENVIRONMENTAL TIME SERIES

Richard W. Katz

LECTURE 5

(1) Hidden Markov Models: Applications

(2) Hidden Markov Models: Viterbi Algorithm

(3) Non-Homogeneous Hidden Markov Model

(1) Hidden Markov Models: Applications

- **Duration of Old Faithful Eruptions**
- **Snoqualmie Falls Precipitation Occurrence**
- **Great Plains Precipitation Occurrence**
- **Other Applications**

Duration of Old Faithful Eruptions

- **Data**

-- Recall data set of eruptions of Old Faithful (fit mixture of two normal distributions to waiting time between eruptions)

-- Data set also contains duration of eruptions

299 eruptions (X_t = 0 or 1):

State 0 (“short”): 105 eruptions < 3 min.

State 1 (“long”): 192 eruptions > 68 min.

(2 eruptions of intermediate duration: with 3 ≤ duration ≤ 68; by convention, count as state 1)

- **Results**

-- Fitting binary hidden Markov model

Assume hidden Markov model has only two states

(i) Starting values

-- Used method of moments estimates (taking $X_t \approx J_t$)

$\Pr\{X_t = 1)$: 194/299 = 0.649

$\mathrm{Corr}(X_t, X_{t+1})$: − 0.538

$\delta = 0.65$, $P_{01} = 0.9$, $P_{11} = 0.35$, $p_0 = 0.01$, $p_1 = 0.99$

(ii) Forward probability recursion (for starting values)

$$a_t(i) = \Pr\{X_1 = x_1, \ldots, X_t = x_t, J_t = i\}, \quad i = 0, 1; \quad t = 1, 2, \ldots, T$$

T	x_t	$a_t(0)$	$a_t(1)$
1	1	0.003500	0.643500
2	0	0.414439	0.002284
3	1	0.000429	0.370056
.	.	.	.
.	.	.	.
.	.	.	.

(iii) Backward probability recursion (for starting values)

$$b_t(i) = \Pr\{X_{t+1} = x_{t+1}, \ldots, X_T = x_T \mid J_t = i\}, \; i = 0, 1; \; t = T, T-1, \ldots, 1$$

T	x_t	$b_t(0)$	$b_t(1)$
299	0	1.000000	1.000000
298	1	0.108000	0.647000
297	1	0.576585	0.224888
296	0	0.200095	0.081671
.	.	.	.
.	.	.	.
.	.	.	.

(iv) Parameter estimates (EM algorithm)

Iter.	δ	p_0	p_1	P_{01}	P_{11}	log like
1	0.6500	0.0100	0.9900	0.9000	0.3500	−150.136
2	0.9991	0.0240	0.9979	0.9979	0.4411	−133.414
⋮	⋮	⋮	⋮	⋮	⋮	⋮
30	1.0000	0.2246	1.0000	1.0000	0.1722	−126.7079
31	1.0000	0.2248	1.0000	1.0000	0.1719	−126.7078

- **Model Identification**

-- **Use results in MacDonald & Zucchini book**

(Different assumption about initial state of hidden Markov model; i.e., $\delta = \pi$)

Model	No. Par.	Log like	AIC	BIC
Independence	1	−193.80	389.60	393.31
First-order Markov chain	2	−134.24	272.48	279.88
Hidden Markov model	4	−127.31	262.62	277.42
Second-order Markov chain	4	−127.12	262.24	277.04

- **Model Properties**

-- Autocorrelation function

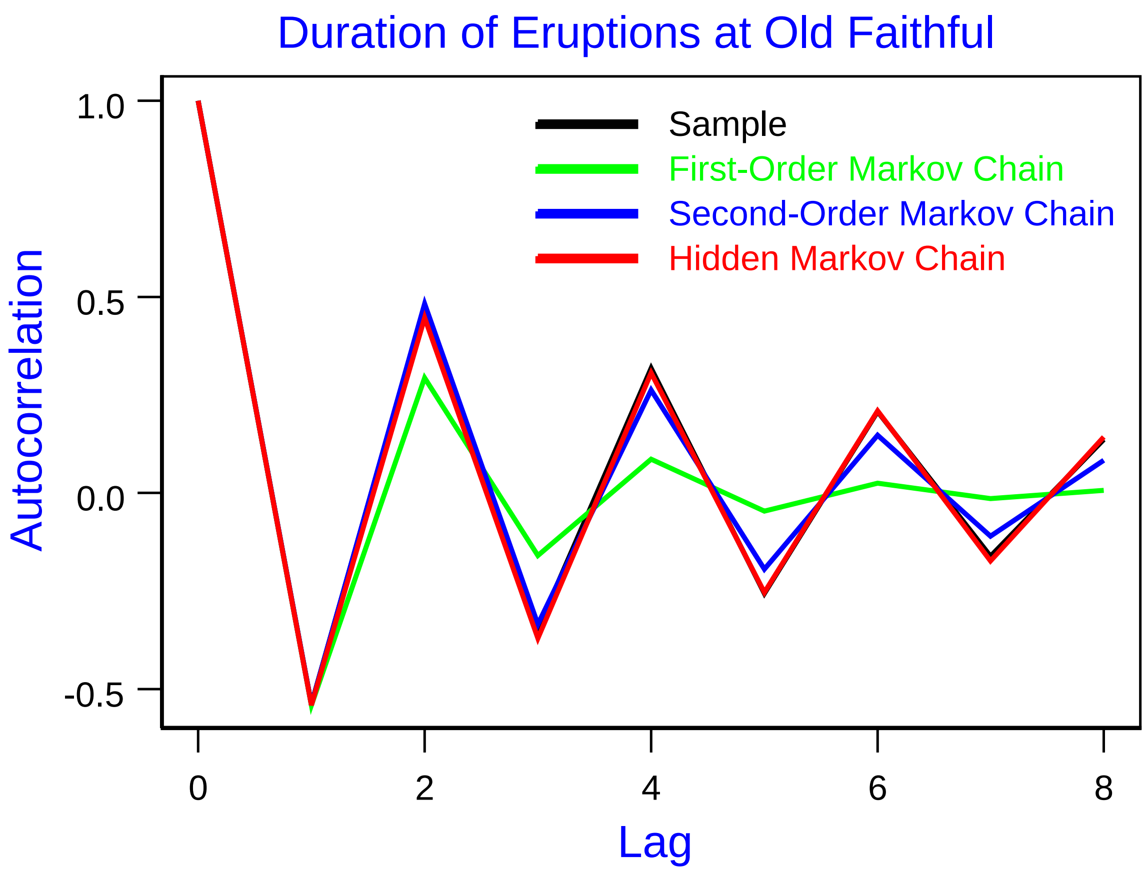
Have expressions for case of:

(1) First-order Markov chain (two-state)

(2) Second-order Markov chain (two-state)

Analogous to first-order Markov chain

(3) Hidden Markov model (binary, two-state)


Duration of Eruptions at Old Faithful
Sample
First-Order Markov Chain
Second-Order Markov Chain
Hidden Markov Chain
1.0
0.5
0.0
-0.5
Autocorrelation
0
2
4
6
8
Lag

Snoqualmie Falls Precipitation Occurrence

- **Data**

-- **Snoqualmie Falls, WA, USA**

-- **Time series of daily precipitation occurrence ($X_t = 0$ or 1)**

-- **January – March, 1948 – 1983 (90 values per yr, 36 yrs)**

-- **Assume each year represents independent time series**

- **Model Fitting**

-- **Direct numerical optimization (not EM algorithm) used by Guttorp (1995)**

-- **Parameter estimates (standard errors from Hughes, 1997)**

Parameter	Estimate	Standard error
P_{01}	0.3256	0.03008
P_{11}	0.8578	0.01549
p_0	0.0592	0.03578
p_1	0.9419	0.01614

- **Model Identification**

Model	No. Par.	Log like	AIC	BIC
First-order Markov chain	2	−898.32	1800.64	1812.81
Hidden Markov model	4	−895.10	1798.20	1822.53

- **Model Properties**

-- Distribution of length of dry spells

Geometric for Markov Chain: Underestimates observed frequency (for longer spells)

Hidden Markov model apparently fits better (figure in Guttorp book, p. 108)

Great Plains Precipitation Occurrence

- **Data**

-- Time series of daily precipitation occurrence

-- 3 sites in US (Omaha, NE; Des Moines, IO; Grand Island, NE)

-- Divided into 6 seasons of about 2 months length (most of length 60 days)

-- Period of record: 1949 – 1984 (i.e., 36 yrs)

- **Model**

-- $X_t(n) = 1$ if precipitation occurs on tth day at nth site, n = 1, 2, 3

-- Assume two-state hidden Markov model $\{J_t\}$

-- Given common hidden state, assume precipitation occurrence at given site conditionally independent of occurrence at other sites

-- Assume each year represents independent time series

-- Let $p_i^{[n]} = \Pr\{X_t(n) = 1 \mid J_t = i\}$, $i = 0, 1$; $n = 1, 2, 3$

- **Model Fitting**

-- Direct numerical optimization (not EM algorithm) used by Guttorp (1995)

-- Parameter estimates (Season 1: Jan-Feb.)

Markov chain: $P_{01} = 0.352$, $P_{11} = 0.665$

Hidden state J_t	$p_i^{[1]}$	$p_i^{[2]}$	$p_i^{[3]}$
$J_t = 0$	0.927	0.874	0.762
$J_t = 1$	0.039	0.211	0.139

- **Model Properties**

-- Common hidden states can induce unconditional spatial dependence (even if assume conditional independence)

-- Table in Guttorp book (p. 109)

Joint probability of precipitation occurrence at pairs of sites (and at all 3 sites) reproduced well by hidden Markov model

-- But not always case that sufficient spatial dependence can be induced

See Zucchini and Guttorp, 1991 paper: modeled additional sites for same region

Other Applications

- **Rainfall Rate** (Sansom, 1998)

-- Data essentially instantaneous measurements of precipitation "rate"

-- Fit hidden Markov models to precipitation rate time series (using EM algorithm)

-- Find strong evidence of existence of "breakpoints" between different precipitation rates

-- Able to provide physical interpretation to hidden states

(2) Hidden Markov Models: Viterbi Algorithm

- **Stochastic Dynamic Programming**
- **Viterbi Algorithm**
- **Example**

Stochastic Dynamic Programming

- **Background**

-- Dynamic optimization problem

-- Stochastic (with underlying Markovian structure)

-- Use "backwards induction"

Start with single occasion/ "static" problem

Solve optimization problem by recursion

Viterbi Algorithm

- **"Global Decoding"**

-- Want to determine sequence of states of hidden Markov chain ($i_1, i_2, \ldots, i_T$) maximizing conditional probability:

$$\Pr\{J_1 = i_1, J_2 = i_2, \ldots, J_T = i_T \mid X_1 = x_1, X_2 = x_2 \ldots, X_T = x_T\}$$

-- Initial step

Set

$$\xi_1(i) = \Pr\{J_1 = i, X_1 = x_1\},\ i = 0, 1$$

-- Recursion

Set

$$\xi_t(i) = \max_{i_1,\ldots,i_{t-1}} \Pr\{J_1 = i_1, \ldots, J_{t-1} = i_{t-1}, J_t = i, X_1 = x_1, X_2 = x_2, \ldots, X_t = x_t\}$$

$t = 2, 3, \ldots, T$

Then

$$\xi_{t+1}(j) = \max\{\xi_t(0)P_{0j}, \xi_t(1)P_{1j}\}\, p_j(x_{t+1})$$

$j = 0, 1;\ \ t = 1, 2, \ldots, T-1$

-- Required states (i_1^*, i_2^*, . . ., i_T^*)

(1) $t = T$

$$i_T^* = 1 \text{ if } \xi_T(0) < \xi_T(1),$$

$$i_T^* = 0 \text{ otherwise}$$

(2) $t = T - 1, T - 2, \ldots, 1$

$$i_t^* = 1 \text{ if } \xi_t(0)P(0, i_{t+1}^*) < \xi_t(1)P(1, i_{t+1}^*),$$

$$i_t^* = 0 \text{ otherwise}$$

[writing $P_{ij} = P(i, j)$]

Example

- **Data**

-- Duration of Old Faithful Eruptions

Use parameter estimates from termination of EM algorithm (i.e., iteration no. 31)

Recall posterior probabilities for "local decoding" produced automatically

- **Results**

Time	Observed State	Local decoding	Global decoding
1	1	1 (1.0000)	1
2	0	0 (1.0000)	0
3	1	1 (1.0000)	1
4	0	0 (0.8625)	0
.	.	.	.
.	.	.	.
162	1	0 (0.5895)	0
.	.	.	.
.	.	.	.
299	0	0 (1.0000)	0

(3) Non-Homogeneous Hidden Markov Models

- **Basic Non-Homogeneous Model**

-- **Hughes et al. (1999)**

- **Extensions**

-- **Conditional dependence (Hughes et al., 1999)**

-- **Inclusion of precipitation intensity (Bellone et al., 2000)**

Non-Homogeneous Model

- **Observed covariates**

-- In addition to hidden Markov model

-- Transition probabilities of hidden Markov model depend on covariates

-- Example

Network of sites at which precipitation occurrence modeled (as in Great Plains example)

But now observed meteorological covariates available as well (daily time scale)

- **Model**

-- Notation

Observed variable: $X_t(n) = 1$ **if precipitation occurs on** t**th day at** n**th site**

(Set $\mathbf{X}_t = [X_t(1), X_t(2), \ldots]$**)**

Hidden Markov chain: $\{J_t\}$

Covariates: $\mathbf{Z}_t$ **vector of atmospheric variables at time** t**,** $t = 1, 2, \ldots, T$

-- Concept

Hidden variable ("weather state") acts as link between two disparate scales (small-scale precipitation and larger-scale atmospheric circulation patterns)

-- Assumption 1

Precipitation occurrence process is conditionally independent given current hidden state:

$$\Pr\{X_t \mid J_1, \ldots, J_T, X_1, \ldots, X_{t-1}, Z_1, \ldots, Z_T\} = \Pr\{X_t \mid J_t\}$$

-- Assumption 2

Given history of hidden state up to time $t-1$ and entire past and future sequence of atmospheric data, hidden state at time t depends only on previous hidden state and current atmospheric data

$$\Pr\{J_t \mid J_1, \ldots, J_{t-1}, Z_1, \ldots, Z_T\} = \Pr\{J_t \mid J_{t-1}, Z_t\}$$

-- **Conditional independence**

Retain (for now) assumption that precipitation occurrence is conditionally spatially independent given hidden state

(i.e., as model for $\Pr\{X_t \mid J_t\}$)

-- **Atmospheric variables**

Assume multivariate normal distribution

-- **Ordinary "homogenous" hidden Markov model is special case**

- **Parameter Estimation**

-- EM algorithm

(not feasible to resort to direct maximization of likelihood function)

-- Model selection

Used BIC & physical interpretability of hidden states

- **Application**

-- Precipitation data

15-yr record (1978 – 1992) of daily winter precipitation occurrences at 30 stations in SW Australia

-- Atmospheric variables (gridded data)

Sea-level pressure, geopotential height, air temperature, dew point, wind speed components

-- Interpretation of hidden states

Each day classified into its most likely state using Viterbi algorithm

Extensions

- **Conditional dependence**

-- Relaxing assumption of conditional spatial independence

-- Replace with autologistic model for multivariate binary model

-- Complicates EM algorithm

Both E-step and M-step of EM algorithm become computationally intractable [used EM-MCML (Monte Carlo Maximum Likelihood) algorithm instead]

-- Hughes et al. (1999) applied model to same Australia data set (Improvement in induced spatial dependence in precipitation occurrence)

- **Inclusion of precipitation intensity (Bellone et al., 2000)**

-- Non-homogeneous hidden Markov model

-- Retain conditional spatial independence assumption for precipitation occurrence

-- Given precipitation occurrence, assume precipitation intensity can be modeled as spatially conditionally independent with gamma distribution

-- Used EM algorithm for parameter estimation

-- Used BIC for model identification

-- Used Viterbi algorithm in conjunction with interpretation of hidden states

-- Applied to precipitation occurrence at 24 sites across Washington state, USA (Atmospheric data: pressure heights, temperature, relative humidity)