Introduction to Spatio-temporal Statistics

Adway Mitra

AI60002: Machine Learning for Earth System Sciences

18 January 2021

Groups of Random Variables

- We are recording hourly temperature at a particular location, every day
- Denote by $x_{d,t}$, the temperature reading on day d, hour t
- We wish to consider x as a random variable
- Option 1: $X_{d,t} \sim f$
- Option 2: $X_{d,t} \sim f_{d,t}$
- Option 3: $X_{d,t} \sim f_d$
- Option 4: $X_{d,t} \sim f_t$

Groups of Random Variables

- We are recording hourly temperature at a particular location, every day
- Denote by $X_{d,t}$, the temperature reading on day d, hour t
- We wish to consider X as a random variable
- Option 1: $X_{d,t} \sim f$ all observations are IID
- Option 2: $X_{d,t} \sim f_{d,t}$ all observations are separate RVs
- Option 3: $X_{d,t} \sim f_d$ Separate distribution for each day
- Option 4: $X_{d,t} \sim f_t$ Separate distribution for each hour

Groups of Random Variables

- We are recording hourly temperature at a particular location, every day
- Denote by $X_{d,t}$, the temperature reading on day d, hour t
- We wish to consider X as a random variable
- Option 1: $X_{d,t} \sim f$ fails to capture variations
- Option 2: $X_{d,t} \sim f_{d,t}$ infeasible, not beneficial
- Option 3: $X_{d,t} \sim f_d$ fails to capture hourly variations
- Option 4: $X_{d,t} \sim f_t$ fails to capture seasonal variations

Temporal Auto-correlation

- Suppose we are focusing on one season only.
- We go for Option 4: $X_{d,t} \sim f_t$
- $\{x_{1,t}, x_{2,t}, \ldots\}$ are realizations of X_t
- Missing out: relationship between hours!
- $Corr(X_{t_i}, X_{t_j})$: correlation coefficient between the Random variable for two different hours
- Example of **temporal autocorrelation**!
- Autocorrelation may be high or low, based on t_i and t_j

Temporal Auto-correlation

- Consider a set of temporal variables $\{X_{t1}, X_{t2}, \dots\}$
- **Mean stationarity:** $E(X_{ti}) = m$, i.e. constant
- **Covariance stationarity:** $Cov(X_{ti}, X_{tj}) = C_t(ti - tj)$,
- C_t is called **Temporal Covariance Function**
- Covariance stationarity implies *temporal autocorrelation* between X_{ti}, X_{tj} only a function of $(ti - tj)$
- **Weak stationarity**: Mean stationarity + Covariance stationarity + finite $E(|X_{ti}|^2)$

Temporal Auto-regression

- Can we express one temporal variable as a function of others?
- If $Corr(X_{ti}, X_{tj}) \neq 0$, can we have $X_{tj} = f(X_{ti})$?
- Simplest assumption: linear relation
- $X_{tj} = aX_{ti} + b$, where b is a random variable (eg. white noise)
- *Order-1 autoregressive process*: $X_{t,i+1} = aX_{t,i} + b$
- *Order-K autoregressive process*: $X_{t,i+1} = \sum_{k=0}^{K-1} a_k X_{t,i-k} + b$

Spatial Autocorrelation

- Consider the rainfall measured every day at locations S_1, S_2, ...
- x_{st} = rainfall at location s, day t
- $X_{st} \sim f_s$
- *Spatial autocorrelation*: $Corr(X_{si}, X_{sj})$
- *Mean Stationarity*: $E(X_s) = c$ (constant)
- *Covariance stationarity*: $Cov(X_{si}, X_{sj}) = C_S(||s_i - s_j||)$, $||.||$ denotes distance
- C_S is called **Spatial Covariance Function**
- Implies: spatial autocorrelation between any two points is a function of their distance!

Spatial Covariance Function

- **First Law of Geography**: *everything is related to everything else, but near things are more related than distant things*
- $Corr(X_{si}, X_{sj})$ should have high magnitude if $||si - sj||$ is low
- $Corr(X_{si}, X_{sj})$ should have low magnitude if $||si - sj||$ is high
- Possible covariance function: $Cov(X_{si}, X_{sj}) = k.exp(-\gamma||si - sj||^2)$
- k can be positive or negative, γ: scaling constant
- Temporal covariance function may be defined analogously

Variogram

- Defined as the variance of the difference between the variable at two different locations
- Measure of spatial smoothness of X
- $\gamma(si, sj) = \frac{1}{2} E((X_{si} - X_{sj})^2)$
- In case of *weakly stationery process*, this reduces to $\gamma(si, sj) = Var(X_{si}) + Var(X_{sj}) - 2Cov(X_{si}, X_{sj})$
- Further, $Var(X_{si}) = Var(X_{sj}) = C_S(0)$, and $Cov(X_{si}, X_{sj}) = C_S(||si - sj||)$
- So, for weakly stationery process, $\gamma(si, sj) = C_S(0) - C_S(||si - sj||)$

Spatial Autoregression

- Can we express one spatial variable as a function of others?
- If $Corr(X_{si}, X_{sj}) \neq 0$, can we have $X_{sj} = g(X_{si})$?
- Simplest assumption: linear relation
- $X_{sj} = aX_{si} + b$, where b is a random variable (eg. white noise)
- *Order-K autoregressive process*: $X_{sj} = \sum_{k=0}^{K-1} a_k X_{i_k} + b$

Autoregression Parameter Estimation

- How to estimate the coefficients like a?
- For each day t, we have $b_t = X_{sj,t} - aX_{si,t}$
- $b_t \sim \mathcal{N}(0, 1)$
- The likelihood $p(b_1, b_2, \dots) \propto \prod_t exp(\frac{1}{2} b_t^2)$
- Log-likelihood function $\mathcal{L}(a) \propto -\sum_t (X_{sj,t} - aX_{si,t})^2$
- Take derivative of log-likelihood, equate to 0, solve for a