Some MCMC tests

Elad Zippory

October 31st, 2016

Agenda

1. stationarity
2. Conditions for stationarity
3. Practical tests

basic definitions

- A Markov process is when $P(\theta_{t+1}|\theta_t, \theta_{t-1}, .., \theta_1) = P(\theta_{t+1}|\theta_t)$

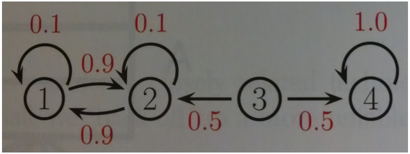
basic definitions

- A Markov process is when $P(\theta_{t+1}|\theta_t, \theta_{t-1}, .., \theta_1) = P(\theta_{t+1}|\theta_t)$
- Stationarity means that $\pi_t = \pi_{t-1}A$ where π is the vector of probabilities of states and A is the transition matrix.

basic definitions

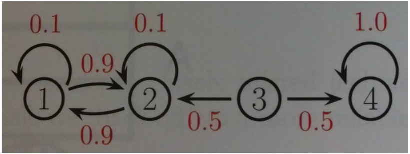
- A Markov process is when $P(\theta_{t+1}|\theta_t, \theta_{t-1}, .., \theta_1) = P(\theta_{t+1}|\theta_t)$
- Stationarity means that $\pi_t = \pi_{t-1}A$ where π is the vector of probabilities of states and A is the transition matrix.
- In general, most of the following conditions for stationarity cannot be a priori constructed and all the practical test are necessary but not sufficient.

irreducibility



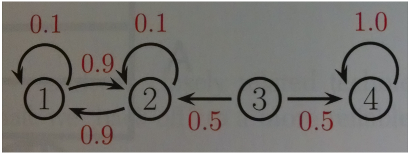
- The distribution is not stationary because the transition matrix is not 'irreducible': when you can get to any state from any state.

irreducibility



- The distribution is not stationary because the transition matrix is not ‘irreducible’: when you can get to any state from any state.
- Practically, we can check this with multiple initial conditions - multiple chains.

irreducibility



- The distribution is not stationary because the transition matrix is not 'irreducible': when you can get to any state from any state.
- Practically, we can check this with multiple initial conditions - multiple chains.
- keep in mid that you can have a stationary distribution but that is not unimodal - consider the oscillation between states 1 and 2, within each chain - stationarity does not mean unimodality of the distribution.

finite and infinite states

- When the states are finite the chain needs to be Irreducible and Aperiodic($\forall t, A^t_{i,i} > 0$). the two conditions allow to get from i to j in a finite number of iterations.

finite and infinite states

- When the states are finite the chain needs to be Irreducible and Aperiodic($\forall t, A_{i,i}^{t} > 0$). the two conditions allow to get from i to j in a finite number of iterations.
- But, what happens when the set is infinite?

finite and infinite states

- When the states are finite the chain needs to be Irreducible and Aperiodic($\forall t, A^{t}_{i,i} > 0$). the two conditions allow to get from i to j in a finite number of iterations.
- But, what happens when the set is infinite?
- A chain with infinite states that has a unique stationary distribution is called Ergodic.

finite and infinite states

- When the states are finite the chain needs to be Irreducible and Aperiodic($\forall t, A^t_{i,i} > 0$). the two conditions allow to get from i to j in a finite number of iterations.
- But, what happens when the set is infinite?
- A chain with infinite states that has a unique stationary distribution is called Ergodic.
- It needs an additional condition: recurrent and non-null, which means that the expected time to return to a state is finite.

finite and infinite states

- When the states are finite the chain needs to be Irreducible and Aperiodic($\forall t, A^t_{i,i} > 0$). the two conditions allow to get from i to j in a finite number of iterations.
- But, what happens when the set is infinite?
- A chain with infinite states that has a unique stationary distribution is called Ergodic.
- It needs an additional condition: recurrent and non-null, which means that the expected time to return to a state is finite.
- According to Jackman, ergodic also means that the finite time is proportional to the probability distribution.

burn-in

- A chain is ‘burnt’ or stationary when initial values do not matter anymore - $\pi_t = \pi_{t-1}A$

burn-in

- A chain is ‘burnt’ or stationary when initial values do not matter anymore - $\pi_t = \pi_{t-1}A$
- There are upper bounds of burn-in time, but to calculate it we need to estimate the transition matrix, which is usually intractable.

burn-in

- A chain is 'burnt' or stationary when initial values do not matter anymore - $\pi_t = \pi_{t-1}A$
- There are upper bounds of burn-in time, but to calculate it we need to estimate the transition matrix, which is usually intractable.
- In intuition, the minimal burn-in time is determined by the minimal probability, over all subsets of states, of transitioning from that set to its complament.

burn-in

- A chain is ‘burnt’ or stationary when initial values do not matter anymore - $\pi_t = \pi_{t-1}A$
- There are upper bounds of burn-in time, but to calculate it we need to estimate the transition matrix, which is usually intractable.
- In intuition, the minimal burn-in time is determined by the minimal probability, over all subsets of states, of transitioning from that set to its complament.
- AKA Conductance.

R hat - shrink factor

- One of the most famous tests, by Gelman and Rubin 1992

R hat - shrink factor

- One of the most famous tests, by Gelman and Rubin 1992
- $$W = \frac{1}{C} \sum_c \left[\frac{1}{S-1} \sum_s (\bar{y}_{sc} - \bar{y}_c)^2 \right]$$

R hat - shrink factor

- One of the most famous tests, by Gelman and Rubin 1992
- $$W = \frac{1}{C}\sum_{c}\left[\frac{1}{S-1}\sum_{s}(\bar{y}_{sc} - \bar{y}_{c})^2\right]$$
- W measures the within variation, should underestimate var(y) if the chains have not ranged over the entire distribution.

R hat - shrink factor

- One of the most famous tests, by Gelman and Rubin 1992
- $$W = \frac{1}{C}\sum_c \left[\frac{1}{S-1}\sum_s (\bar{y}_{sc} - \bar{y}_c)^2\right]$$
- W measures the within variation, should underestimate var(y) if the chains have not ranged over the entire distribution.
- Where the unbiased variance for a stationary distribution:

R hat - shrink factor

- One of the most famous tests, by Gelman and Rubin 1992
- $$W = \frac{1}{C}\sum_{c}\left[\frac{1}{S-1}\sum_{s}(\bar{y}_{sc}-\bar{y}_{c})^2\right]$$
- W measures the within variation, should underestimate var(y) if the chains have not ranged over the entire distribution.
- Where the unbiased variance for a stationary distribution:
- $$\hat{V} = \frac{S-1}{S}W + \frac{1}{S}B$$

R hat - shrink factor

- One of the most famous tests, by Gelman and Rubin 1992
- $$W = \frac{1}{C} \sum_c \left[\frac{1}{S-1} \sum_s (\bar{y}_{sc} - \bar{y}_c)^2 \right]$$
- W measures the within variation, should underestimate var(y) if the chains have not ranged over the entire distribution.
- Where the unbiased variance for a stationary distribution:
- $$\hat{V} = \frac{S-1}{S} W + \frac{1}{S} B$$
- $$B = \frac{C}{C-1} \sum_c (\bar{y}_c - \bar{y})^2$$

R hat - shrink factor

- One of the most famous tests, by Gelman and Rubin 1992
- $$W = \frac{1}{C}\sum_c \left[\frac{1}{S-1}\sum_s (\bar{y}_{sc} - \bar{y}_c)^2\right]$$
- W measures the within variation, should underestimate var(y) if the chains have not ranged over the entire distribution.
- Where the unbiased variance for a stationary distribution:
- $$\hat{V} = \frac{S-1}{S}W + \frac{1}{S}B$$
- $$B = \frac{C}{C-1}\sum_c (\bar{y}_c - \bar{y})^2$$
- $$\hat{R} = \sqrt{\frac{V}{W}} \approx 1$$

more tests

- **Effective Sample Size** calculates the information in the sample given the autocorrelation

more tests

- **Effective Sample Size** calculates the information in the sample given the autocorrelation
- **Autocorrelation** measures the dependency between samples. High dependency - high mixing and low conductance.

more tests

- **Effective Sample Size** calculates the information in the sample given the autocorrelation
- **Autocorrelation** measures the dependency between samples. High dependency - high mixing and low conductance.
- **Geweke's Convergence** posits that under stationarity $E[y_{.1}] - E[y_{.5}] = 0$. This gives a Z score per variable per chain.

more tests

- **Effective Sample Size** calculates the information in the sample given the autocorrelation
- **Autocorrelation** measures the dependency between samples. High dependency - high mixing and low conductance.
- **Geweke's Convergence** posits that under stationarity $E[y_{.1}] - E[y_{.5}] = 0$. This gives a Z score per variable per chain.
- **Heidelberger and Wlech** checks two things. Tries to reject the null of stationarity with a one sided Cramer-Von Mises. And if the sample size is large enough to measure the mean.