IEOR165 Discussion
Week 10

Sheng Liu

University of California, Berkeley

Apr 1, 2016

Outline

Basics

A Markov process is a process in which the probability of being in a future state conditioned on the present state and past states is equal to the probability of being in a future state conditioned only on the present state:

$$\begin{aligned} &P(X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, \dots, X_1 = i_1, X_0 = i_0) \\ =&P(X_{n+1} = j | X_n = i) = P_{ij} \end{aligned}$$

If the process is in state i, there is a fixed probability that it will move to state j with probability P_{ij} and it is independent of the past history.

If the weather tomorrow only depends on the weather today, then it is reasonable to model the weather transitions as a Markov process.

Basics

Key elements: state space, time and initial conditions

- State space: discrete or continuous
  - Weather: {S,R,C}
  - Location on a 2D plane: $\mathbb{R} \times \mathbb{R}$
- Time:
  - Discrete: $X_n, \; n = 0, 1, 2, \ldots$
  - Continuous: $X(t), \; t \geq 0$
- Initial conditions
  - $P(X_0 = i) = \alpha_i$
  - $\sum_i \alpha_i = 1$ (the process must be in some state initially)

The process must make a transition to some state

$$\sum_j P_{ij} = 1 \quad \forall \; i$$

Transition matrix and transition graph

Transition Matrix & Graph

Example: Assume a Markov chain can be used to model a student's progress at UC Berkeley. In each academic year, if an undergraduate student is at UC Berkeley, then at the end of the same year he/she is two times more likely to successfully pass the academic year than failing the year (under failing we mean he/she will stay at the same grade next year). Also, he is three times less likely to leave UC Berkeley than failing the academic year (leaving means leaving before graduation).
$\rightarrow$ Show the transition matrix and transition graph.

Estimate Transition Probability

MLE estimator:

$$\hat{P}_{ij} = \frac{N_{ij}}{\sum_j N_{ij}}$$

In the last example, assume we have 5000 freshmen this year and suppose 4600 of them successfully pass the first year and 380 of them fail, what transition probabilities can you estimate from this data?

Basics

- Hypothesis is a statement about a **population** parameter
- Null hypothesis and alternative hypothesis, one of them is true
  - Collect a random sample
  - Develop a test statistic and calculate it based on the sample
  - Use this test statistic to make the decision whether to accept ot reject the null hypothesis
- p-value: the probability of obtaining a result equal to or more extreme than what was actually observed, assuming that the model is true.
  - Large p-value means the observation is more likely to happen under the null hypothesis (accept)
  - Small p-value indicates the observation can not be well explained by the null hypothesis (reject)
- Level of significance α: the probability of (falsely) rejecting the null hypothesis when the null hypothesis is true. Popular choices: 0.05 and 0.01, what's the effect of choosing a smaller α?

Basics

Specify a significance level α and compare the calculated p-value with α

- p-value$\leq \alpha$: reject the null hypothesis
- p-value$> \alpha$: accept the null hypothesis

In the coin-flipping example

- The null hypothesis is that the coin is fair:

$$H_0 : \ p(H) = 0.5 \quad H_1 : \ p(H) \neq 0.5$$

- The test statistic is the number of heads in the experiment: 40 (the distribution of this test statistic is known: binomial)
- p-value is the probability of getting more extreme test statistics than the observation under the null hypothesis: $P(N_H \leq 40 \text{ or } N_H \geq 60 | \ p(H) = 0.5) = 0.0569$
- With significance level $\alpha = 0.05$, accept the null hypothesis

One-Tailed vs. Two-Tailed

Berkeley UNIVERSITY OF CALIFORNIA

Two-tailed test: deviation is in both directions

$$H_0: \ \mu = 1 \quad H_1: \ \mu \neq 1$$

One-tailed test: deviation is in one direction

$$H_0: \ \mu = 1 \ (\mu \leq 1) \quad H_1: \ \mu > 1$$

or

$$H_0: \ \mu = 1 \ (\mu \geq 1) \quad H_1: \ \mu < 1$$

Example 1: The manufacturer of a new fiberglass tire claims that its average life will be at least 40000 miles.

Example 2: A public health official claims that the mean home water use is 350 gallons a day.

One-Tailed vs. Two-Tailed

p-value in One-Tailed and Two-Tailed tests

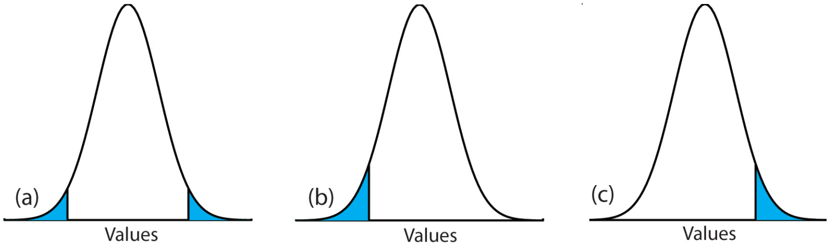

Figure: The distribution of test statistics.

Match (a), (b), (c) to the aforementioned three hypothesis tests

Z-statistics vs. t-statistics

Assume the data follows the normal distribution

- If the population variance is known, we use the Z-statistic:

$$\frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0,1)$$

- If the population variance is not known, we use the t-statistic:

$$\frac{\overline{X} - \mu_0}{s/\sqrt{n}} \sim t_{n-1}$$

where s^2 is the sample variance

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

1. Both standard normal distribution and t-distribution are symmetric!
2. Even when the underlying distribution is not normal, we may also use Z-statistics or t-statistics because they can nicely approximate the p-value when sample size is very large.

Z-statistics vs. t-statistics

Berkeley UNIVERSITY OF CALIFORNIA

Example 8.3h: A public heath official claims that the mean home water use is 350 gallon a day. To verify this statement, a study of 20 randomly selected homes was instigated with the result that the average daily waters uses of these 20 homes were as follows:

340	344	362	375
356	386	354	364
332	402	340	355
362	322	372	324
318	360	338	370

Calculation Details

Consider

$$H_0: \ \mu = 10 \quad H_1: \ \mu \neq 10$$

You obtain a Z-statistic as 1.87, then the p-value is

$$\begin{aligned} P(|Z| > 1.87) &= P(Z > 1.87) + P(Z < -1.87) \\ &= 2P(Z > 1.87) = 2(1 - P(Z < 1.87)) \\ &= 2(1 - 0.9693) = 0.0614 \end{aligned}$$

With $\alpha = 0.05$, we fail to reject the null hypothesis.

Z-table

Berkeley
UNIVERSITY OF CALIFORNIA

Normal probability table

	Second decimal place of Z									
Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767

Calculation Details

Consider

$$H_0:\ \mu \geq 10 \quad H_1:\ \mu < 10$$

You obtain a Z-statistics -1.92, then the p-value is

$$\begin{aligned} P(Z < -1.92) &= P(Z > 1.92) \\ &= 1 - P(Z < 1.92) = 0.0274 \end{aligned}$$

With $\alpha = 0.05$, we reject the null hypothesis.

References