hc5hq Huilin Chang
DS6014

Homework 1:
Probability Review and Priors

Huilin Chang
hc5hq@virginia.edu

1 (15)

You are a data Scientist and are choosing between three approaches, A, B, and C to a problem. With approach A you will spend a total of four days coding and running an algorithm and it will not produce useful results. With approach B you will spend a total of three days coding and running an algorithm and it will not produce useful results. With approach C you will spend over day coding and running an algorithm and it will give the results you are looking for. You are equality likely to choose among unselected options. What is the expected time in days for you to obtain the results you are looking for? What is the variance on this time?

Response:

Considering all the possible approach chains

If starting with approach A:

- A-B-C: 4+3+1 = 8 days
- A-C: 4+1 = 5 days

If starting with approach B:

- B-A-C: 3+4+1= 8 days
- B-C : 3+1 = 4 days

If staring with approach C

- C: need one day = 1 days

Let E to represent the days to solve the question = {8, 5, 8, 4, 1} and the mean value is 5.2

The variance of time based on the variance formula:

$S^2 = \frac{\sum(x_i - \bar{x})^2}{n-1} = \frac{34.8}{4} =$ 8.7, the variance = 2.95

2 (15)

Suppose if it is sunny or not in Charlottesville depends on the weather of the last three days. Show how this can be modeled as Markov chain.

Response:

We can think about "Markov chain" in this way

- The next term in a sequence could depend on all the previous terms
  - If it only depends on the previous term it is called "first-order" Markov
  - If it depends on the two previous terms it is "second-order" Markov

We can form a Markov chain as follows, take the weather states R(rain), (nice), S(Sunny) to form transition probabilities.

$$P = \begin{bmatrix} P11 & P12 & P13 \\ P21 & P22 & P23 \\ P31 & P32 & P33 \end{bmatrix}$$

$$P_{ij} \geq 0, i = 1, \ldots\ldots n, j = 1, \ldots\ldots n \ \ and \ \sum_{j=1}^{n} P_{ij} = 1$$

For example:

$$P = \begin{array}{c} \\ R \\ N \\ S \end{array} \begin{array}{c} \begin{array}{ccc} R & N & S \end{array} \\ \begin{pmatrix} 0.5 & 0.4 & 0.1 \\ 0.3 & 0.4 & 0.3 \\ 0.2 & 0.3 & 0.5 \end{pmatrix} \end{array}$$

The entries in the first row of the matrix P represent the probabilities for the various kinds of weather following a rainy day. Similarly, the entries in the second and third rows represent the probabilities for the various kinds of weather following nice and snowy day, respectively.

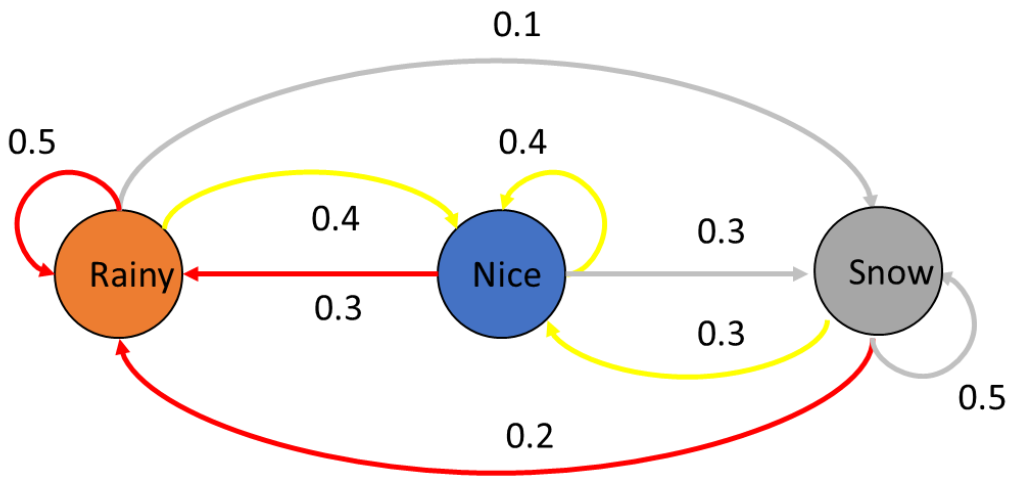
Considering given the chain is in state i today, it will be in state j two days from now given this probability by p_{ij}

hc5hq Huilin Chang
DS6014

Markov property: The state of the system at time t+1 only depends on thee state of the system at time t

$P[X_{t+1} = x_{t+1}|X_t = x_t, X_{t-1} = x_{t-1,\ldots\ldots\ldots\ldots}, X_1 = x_1, X_0 = x_0] =$

$P[X_{t+1} = x_{t+1}|X_t = x_t]$



Weather forecasting

- Two days: $\begin{pmatrix} 0.5 & 0.4 & 0.1 \\ 0.3 & 0.4 & 0.3 \\ 0.2 & 0.3 & 0.5 \end{pmatrix}\begin{pmatrix} 0.5 & 0.4 & 0.1 \\ 0.3 & 0.4 & 0.3 \\ 0.2 & 0.3 & 0.5 \end{pmatrix} = \begin{pmatrix} 0.39 & 0.39 & 0.22 \\ 0.33 & 0.37 & 0.30 \\ 0.29 & 0.35 & 0.36 \end{pmatrix}$
- Four days:

$$\begin{pmatrix} 0.5 & 0.4 & 0.1 \\ 0.3 & 0.4 & 0.3 \\ 0.2 & 0.3 & 0.5 \end{pmatrix}^2 \begin{pmatrix} 0.5 & 0.4 & 0.1 \\ 0.3 & 0.4 & 0.3 \\ 0.2 & 0.3 & 0.5 \end{pmatrix}^2 = \begin{pmatrix} 0.3446 & 0.3734 & 0.2820 \\ 0.3378 & 0.3706 & 0.2916 \\ 0.3330 & 0.3686 & 0.2984 \end{pmatrix}$$

The graph 3 (15)

Assume a Gaussian distribution for observations, $X_{i,\ i=1,\ldots\ldots\ldots N}$ with unknown mean, M and known variance 5. Suppose the prior for M is Gaussian with variance 10. How large a random sample must be taken (i.e., what is the minimum value for N) to specify an interval having unit length of 1 such that the probability that M lies in this interval is 0.95?

hc5hq Huilin Chang
DS6014

Response:

Gaussian with Unknown mean and known variance

From lecture

- Likelihood with N trials, $x = (x_1, \ldots\ldots\ldots, x_N)$ with unknown mean M and know variance σ^2
  - $f(x|m, \sigma^2) \propto \frac{1}{\sigma} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{N}(x_i - m)^2\right)$
- Prior for M is $N(\mu_0, \sigma_0^2)$
- Posterior is $N(\mu_N \; and \; \sigma_{post}^2) where$
  - $\mu_N = \frac{\mu_0 \, \sigma^2 + N\bar{x} \; \sigma_0^2}{\sigma^2 + N\sigma_0^2}$
  - $\sigma_{post}^2 = \frac{\sigma_0^2 \sigma^2}{\sigma^2 + N\sigma_0^2}$

The variance of the Gaussian distribution is 5 (σ) and the variance for the prior is 10 (σ_0^2)

Since the posterior is a function of N, we need to find an N that makes this variance (here is 5) that the probability M is centered at the posterior mean $\mu_N = 0.95$

$$\sigma_{post}^2 = \frac{\sigma_0^2 \sigma^2}{\sigma^2 + N\sigma_0^2} \rightarrow \sigma_{post}^2 = \frac{10*5}{5+N10} = \frac{50}{5+10N}$$

$$CDF = \frac{1}{2}\left[1 + erf\left(\frac{x - \mu}{\sigma\sqrt{2}}\right)\right]$$

$$0.975 = \frac{1}{2}\left[1 + erf\left(\frac{0.5}{\sqrt{\frac{50}{5+10N}}\sqrt{2}}\right)\right]$$

$$1.3859 * \left(\sqrt{\frac{50}{5+10N}}\sqrt{2}\right) = 0.5$$

$$1.9207 * \left(\frac{100}{5+10N}\right) = 0.25 \text{ , N = } 77$$

4 (15)

You have started an online business selling books that are of interest to your customers. A publisher has just given you a large book with photos from famous 20th century photographers. You think this book will appeal to people who have bought art books, history books and coffee table books. In an initial offering of the new book you collect data on purchases of the new book and combine these data with data from the past purchases (see ArtHistBooks.csv).

Use Baysesian analysis to give the posterior probabilities for purchases of art books, history books and coffee table books, as well as, the separate probabilities for purchases of new books given each possible combination of prior purchases of art books, history books and coffee table books. Do this by first using beta priors with values of the hyperparameters that represent lack of prior information. Then compute these probabilities again with beta priors that show strong weighting for low likelihood of a book purchase. Compare your results.

Response: see the notebook

5 (15)

The data set CHDdata.csv contains cases of coronary heart disease(CHD) and variables associated with the patient's condition: systolic blood pressure, yearly tobacco use (in kg), how density lipoprotein (ldl), adiposity, family history (0 or 1), type A personality score (typea), obesity (body mass index), alcohol use, and the diagnosis of CHD (0 or 1). Perform a Bayesian analysis of these data that finds the posterior marginal probability distributions for the means for the data of patients with and without CHD. You should first standard scale (substract the man and divide by the standard deviation) all the numeric variables (remove family history and do not scale CHD). Then separate the data into two sets, one for patients with CHD and one for patients without CHD.

Your priors for both groups should assume means of 0 for all variables and a correlation of 0 between all pairs of variables. You should assume all variances for the variables are 1. Use a prior alpha equal to one plus the number of predictor

variables. Compute and compare the Bayesian estimates for the posterior means for each group.

For 5 extra credit points, compute the probability of observing a point at least as extreme as the posterior mean of patients without coronary heart disease under the posterior distribution for the patients with coronary heart disease. Then compute the probability of observing a point at least as extreme as the posterior mean of patients with coronary heart disease under the posterior distribution for the patients without coronary heart disease

Response: see the notebook

6 (10)

For each of the following types of distributions, state the support type (single or multivariable and discrete or continuous), the formula for the PMP or PDF, the parameters, the support, the mean, and some typical uses of the distribution. You may use whatever source(s) you want, including for example Wikipedia.

(a) Bernoulli Distribution
(b) Binomial Distribution
(c) Poisson Distribution
(d) Uniform Distribution
(e) Beta Distribution
(f) Gamma Distribution
(g) Gaussian Distribution
(h) t Distribution
(i) Cauchy Distribution
(j) Multinomial Distribution
(k) Dirichlet Distribution
(l) Multivariate Gaussian Distribution
(m) Multivariate t Distribution
(n) Wishart Distribution

hc5hq Huilin Chang
DS6014

Distribution, PMP, PDF, support type

a	Bernoulli Distribution	Discreate probability distribution Support k ∈ {0, 1} PMF $\begin{cases} q = 1 - p \ if \ k = 0 \\ p \ if \ k = 1 \end{cases}$ Mean = p Bernoulli is the discrete probability distribution of a random variable which takes the value 1 with probability p and the value 0 with probability q=1-p. A Bernoulli distribution can be thought of as a model for the set of possible outcomes of any single experiment that asks a yes/true/one with probability p, and failure/no/false/zero with probability q such as coin toss problem.
b	Binomial Distribution	Binomial distribution is the discrete probability distribution. Support k ∈ $\{0, 1, \dots \dots \dots . n\}$ – number of successes PMF $\binom{n}{k} p^k q^{n-k}$ Mean np Thee binomial distribution is frequently used to model the number of successes in a sample of size n drawn with replacement from a population of size N.
c	Poisson Distribution	The Poisson distribution is a discrete probability distribution Support k ∈ N_0, Natural numbers starting from 0 PMF = $\frac{\lambda^k e^{-\lambda}}{k!}$ Mean = λ For instance: expresses the probability of a given number of events occurring in a fixed interval of time or space if these events occur with a known constant mean rate and independently of the time since the last event.
d	Uniform Distribution	Uniform distribution is the continuous uniform distribution or rectangular distribution Support $x \in [a, b]$

		PDF $\begin{cases} \frac{1}{b-a} \; for \; x \in [a,b] \\ 0 \; otherwise \end{cases}$ Mean ½(a+b) For instance, density function, uniform probability density function
e	Beta Distribution	The beta distribution is a family of continuous probability distributions defined on the interval[0, 1] parameterized by two positive shape parameters, denoted by α and β, that appear as exponents of the random variable and control t he shape of the distribution. Support x $\in [0,1] \; or \; x \in (0,1)$ PDF$\frac{X^{\alpha-1}(1-X)^{\beta-1}}{B(\alpha,\beta)}$ Where B(α, β) =$\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$ and Γ is the Gamma function Mean: $E[X] = \frac{\alpha}{\alpha+\beta}$ $E[\ln X] = \varphi(\alpha) - \varphi(\alpha+\beta)$ $E[XlnX] = \frac{\alpha}{\alpha+\beta}[\varphi(\alpha+1) - \varphi(\alpha+\beta+1)]$
f	Gamma Distribution	Gamma distribution is a two-parameter family of continuous probability distributions. Support $x \in (0,\infty)$ PDF: $f(x) = \frac{1}{\Gamma(k)\theta^k}x^{k-1}e^{-\frac{x}{\theta}}$ (k>0, $\theta > 0$) $f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\beta x}$ (α>0, $\theta > 0$) Mean :kθ (k>0, $\theta > 0$) $\frac{\alpha}{\beta}$ (α>0, $\theta > 0$) For instance: in life testing, the waiting time until death is a random variable that is frequently modeled with a gamma distribution.
g	Gaussian Distribution	Gaussian distribution is a type of continuous probability distribution Support $x \in R$ PDF = $\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$ Mean σ
h	t Distribution	t-distribution is an member of family of continuous probability distribution support $x \in (-\infty, \infty)$

		PDF $\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\Gamma\left(\frac{\nu}{2}\right)}(1+\frac{x^2}{\nu})^{-\frac{\nu+1}{2}}$ Mean 0 for $\nu > 1$
i	Cauchy Distribution	The Cauchy distribution is a continuous probability distribution. Support $x \in (-\infty, \infty)$ PDF = $\frac{1}{\pi\gamma\left[1+\left(\frac{x-x_0}{\gamma}\right)^2\right]}$ Mean undefined
j	Multinomial Distribution	In a probability theory, the multinominal distribution is a generalization of the binomial distribution. For example, it models the probability of counts for each side of a k-sided die rolled n times. For n independent trials each of which leads to a success for exactly one of k categories, with each category having a given fixed success probability, the multinomial distribution gives the probability of any particular combination of numbers of successes for the various categories. Support $x_i \in \{0, \ldots . n\}, i \in \{1, \ldots \ldots k\}$ $\sum x_i = n$ PMF $\frac{n!}{x_1!\ldots\ldots x_{k!}} p_1^{x_1} \ldots\ldots\ldots . . p_k^{x_k}$ Mean E(X_i) = np_i
k	Dirichlet Distribution	The Dirichlet distribution is a family of continuous multivariance probability distributions parametrized by a vector α of positive reals. Support $x_1, \ldots \ldots \ldots x_K$ where $x_i \in (0,1) and \sum_{i=1}^{K} x_i = 1$ PDF: $\frac{1}{B(\alpha)} \prod_{i=1}^{K} x_i^{\alpha_i - 1}$ Where B(α) = $\frac{\prod_{i=1}^{K} \Gamma(\alpha_i)}{\Gamma(\sum_{i=1}^{K} \alpha_i)}$ Where $\alpha = (\alpha_1, \ldots \ldots \ldots \alpha_k)$ Mean $E[X_i] = \frac{\alpha_i}{\Gamma(\sum_{k=1}^{K} \alpha_k)}$ $E[lnX_i] = \varphi(\alpha_i) - \varphi(\sum_{k} \alpha_k)$

l	Multivariate Gaussian Distribution	In probability theory, the multivariate normal distribution is a generalization of the one-dimensional normal distribution to higher dimensions. Support $x \in \mu + span(\Sigma) \subseteq R^k$ PDF $(2\pi)^{-\frac{k}{2}}\det(\Sigma)^{-\frac{1}{2}}e^{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)}$ exists only when Σ is positive-define Mean μ
m	Multivariate t Distribution	In statistics, the multivariate t-distribution is a multivariate probability distribution. Support $x \in R^p$ PDF $\frac{\Gamma[(\nu+p)/2]}{\Gamma(\nu/2)\nu^{p/2}\pi^{p/2}\lvert\Sigma\rvert^{1/2}}\left[1+\frac{1}{\nu}(x-\mu)^T\Sigma^{-1}(x-\mu)\right]^{-(\nu+p)/2}$ Mean μ if $\nu > 2$, else undefined.
n	Wishart Distribution	In statistics, the Wishart distribution is a generalization to multiple dimensions of the gamma distribution. Support X(p × p) PDF $f_x(x) = \frac{\lvert x\rvert^{(n-p-1)/2}e^{-tr(V^{-1}x)/2}}{2^{\frac{np}{2}}\lvert V\rvert^{\frac{n}{2}}\Gamma_p(\frac{n}{2})}$ Mean $E[X]$ = nV

7 (10)

Using the Python Notebook http://www.kaggle.com/billbasener/pt2-probabilities-likelihoods-and-bayes-theorem, complete the challenge questions from Seciton6: Modify the code from Section 5 to add the ability to use the posterior_from_conjugate_prior function to output the posterior probability parameters given parameters and for a Gaussian Likelihood with known variance σ^2, and use your modified function to create the Prior, Likelihood, Posterior plots as in Section 5 of the notebook.

See the notebook in details

Regarding normal prior and normal likelihood

A normal prior is conjugate to a normal likelihood with known σ

- Data: $x_1, x_2, \ldots\ldots\ldots x_n$
- Normal likelihood $x_1, x_2 \ldots\ldots, x_n \sim N(\theta, \sigma^2)$
- Normal prior $\theta \sim N(\mu_{prior}, \sigma^2_{prior})$
- Normal Posterior $\theta \sim N(\mu_{post}, \sigma^2_{post})$