CS258: Information Theory

Fan Cheng

Spring, 2018. chengfan@sjtu.edu.cn

Preface: Inference

- Bayesian Inference
- Occam's razor
- Monte Carlo Methods
- Ising model

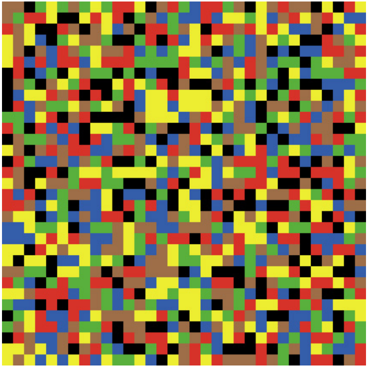
Discussion: Coin Prediction

Toss a coin seven times, we obtain that: TTFFTTT. What is the result of next toss.

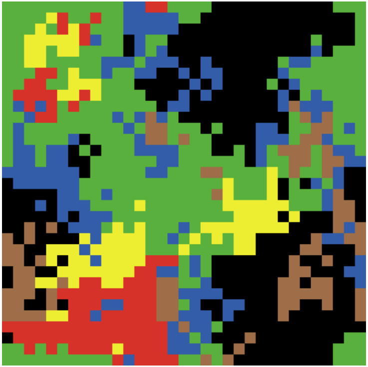
Discussion: Coin Prediction

Toss a coin seven times, we obtain that: TTFFTTT. What is the result of next toss.

Given a sequence, -1, 3, 7, 11, what is the next number.



(a) Initial



(b) Final

The rules of probability ensure that if two people make the same assumptions and receive the same data then they will draw identical conclusions.

Bayes inference

The rules of probability ensure that if two people make the same assumptions and receive the same data then they will draw identical conclusions.

you cannot do inference without making assumptions.

An urn contains K balls, of which B are black and $W = K - B$ are white. Fred draws a ball at random from the urn and replaces it, N times.

- (a) What is the probability distribution of the number of times a black ball is drawn, n_B?
- (b) What is the expectation of n_B? What is the variance of n_B? What is the standard deviation of n_B? Give numerical answers for the cases $N = 5$ and $N = 400$, when $B = 2$ and $K = 10$.

Forward probabilities

An urn contains K balls, of which B are black and $W = K - B$ are white. Fred draws a ball at random from the urn and replaces it, N times.

- (a) What is the probability distribution of the number of times a black ball is drawn, n_B?
- (b) What is the expectation of n_B? What is the variance of n_B? What is the standard deviation of n_B? Give numerical answers for the cases $N = 5$ and $N = 400$, when $B = 2$ and $K = 10$.

> Forward probability problems involve a generative model that describes a process that is assumed to give rise to some data; the task is to compute the probability distribution or expectation of some quantity that depends on the data.

An urn contains K balls, of which B are black and $W = K - B$ are white. We define the fraction $f_B := B/K$. Fred draws N times from the urn, obtaining n_B blacks, and computes the quantity

$$z = \frac{(n_B - f_B N)^2}{N f_B (1 - f_B)}$$

What is the expectation of z? In the case $N = 5$ and $f_B = 1/5$, what is the probability distribution of z? What is the probability that z ¡ 1?

Inverse probabilities

An urn contains K balls, of which B are black and $W = K - B$ are white. We define the fraction $f_B := B/K$. Fred draws N times from the urn, obtaining n_B blacks, and computes the quantity

$$z = \frac{(n_B - f_B N)^2}{N f_B (1 - f_B)}$$

What is the expectation of z? In the case $N = 5$ and $f_B = 1/5$, what is the probability distribution of z? What is the probability that z ¡ 1?

Like forward probability problems, inverse probability problems involve a generative model of a process, but instead of computing the probability distribution of some quantity produced by the process, we compute the conditional probability of one or more of the unobserved variables in the process, given the observed variables. This invariably requires the use of Bayes' theorem.

Answer

We need to compute:

$$P(u, n_B|N) = \frac{P(u)P(n_B|u, N)}{P(n_B|N)}$$

We call the marginal probability $P(u)$ the prior probability of u, and $P(n_B, u|N)$ is called the likelihood of u. It is important to note that the terms likelihood and probability are not synonyms. The quantity $P(n_B, u|N)$ is a function of both n_B and u. For fixed u, $P(n_B, u; N)$ defines a probability over n_B. For fixed n_B, $P(n_B, u; N)$ defines the likelihood of u.

Data and model

Never say 'the likelihood of the data'. Always say 'the likelihood of the parameters'. The likelihood function is not a probability distribution.

The conditional probability $P(u|n_B, N)$ is called the posterior probability of u given n_B. The normalizing constant $P(n_B, N)$ has no u-dependence so its value is not important if we simply wish to evaluate the relative probabilities of the alternative hypotheses u. However, in most data-modelling problems of any complexity, this quantity becomes important, and it is given various names: $P(n_B, N)$ is known as the evidence or the marginal likelihood.

If θ denotes the unknown parameters, D denotes the data, and H denotes the overall hypothesis space, the general equation:

$$P(\theta|D, H) = \frac{P(D|\theta, H)P(\theta|H)}{P(D|H)}$$

is written:

$$\text{posterior} = \text{likelihood} \times \text{prior} \, / \, \text{evidence}$$

The likelihood principles

Q1: Urn A contains three balls: one black, and two white; urn B contains three balls: two black, and one white. One of the urns is selected at random and one ball is drawn. The ball is black. What is the probability that the selected urn is urn A?
Q2: Urn A contains five balls: one black, two white, one green and one pink; urn B contains five hundred balls: two hundred black, one hundred white, 50 yellow, 40 cyan, 30 sienna, 25 green, 25 silver, 20 gold, and 10 purple. One of the urns is selected at random and one ball is drawn. The ball is black. What is the probability that the urn is urn A?

The likelihood principles

Q1: Urn A contains three balls: one black, and two white; urn B contains three balls: two black, and one white. One of the urns is selected at random and one ball is drawn. The ball is black. What is the probability that the selected urn is urn A?
Q2: Urn A contains five balls: one black, two white, one green and one pink; urn B contains five hundred balls: two hundred black, one hundred white, 50 yellow, 40 cyan, 30 sienna, 25 green, 25 silver, 20 gold, and 10 purple. One of the urns is selected at random and one ball is drawn. The ball is black. What is the probability that the urn is urn A?

Likelihood principle

The likelihood principle: given a generative model for data d given parameters θ, $P(d|\theta)$, and having observed a particular outcome d_1, all inferences and predictions should depend only on the function $P(d1|\theta)$.

Reading and Exercise

Reading: Ch. 2.2-2.3, Ch. 9 (David MacKay)