BP-203 Foundations for Mathematical Biology

Statistics Lecture IV

By Hao Li

Nov 13, 2001

Bayesian Inference

Example A revisited:

Toss a coin N times, observe m heads
N is small
Model: binomial distribution
e.g., N=2, m=0, would you infer $p = m / N = 0$?

There is a wide range of p that can produce the observed result
Our prior knowledge → $p = 0$ unlikely

Bayesian Inference:

infer a probability distribution of the parameters conditioned on the observed data
input our prior knowledge of the distribution of the parameters

Bayes' theorem

Let A and $B_1, B_2, ..., B_n$ be events where B_i are disjoint and exhaustive (cover all sample space), and $P(B_i) > 0$ for all i, then

$$P(B_j \mid A) = \frac{P(A|B_j)P(B_j)}{\sum_{j=1}^{n} P(A|B_j)P(B_j)}$$

prior knowledge of B_j

distribution of B_j conditioned on A

Example F: polygraph tests/lie detector tests (discrete sample space)

Events:

L: subject is lying

T: subject is telling truth

+: polygraph reading is positive (indicating that the subject is lying)

-: polygraph reading negative (indicating that the subject is telling truth)

Polygraph reliability → conditional probability (conditioned on L and T)

	L	T
+	0.88	0.14
-	0.12	0.86

One specific screen, prior P(T)= 0.99 P(L)= 0.01

What is P(T|+), the prob. that the reading is positive but the subj. is telling truth?

$$P(T \mid +) = \frac{P(+|T)P(T)}{P(+|T)P(T)+P(+|L)P(L)} = \frac{0.14\times0.99}{0.14\times0.99+0.88\times0.01} = 0.94$$

Bayesian inference for continuous parameters

prior distribution

$$P(\theta \mid Data) = \frac{P(Data|\theta)\pi(\theta)}{\int P(Data|\theta)\pi(\theta)d\theta}$$

posterior distribution

Example A revisited

$$P(m \mid p) = p^m (1-p)^{N-m}$$

posterior distribution

$$P(p \mid m) = \frac{P(m|p)\pi(p)}{\int P(m|p)\pi(p)dp}$$

conjugate prior: same functional form as the likelihood function

$$\pi(p) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1}(1-p)^{\beta-1}$$

Beta distribution

$$P(p \mid m) = const \times p^{m+\alpha-1}(1-p)^{N-m+\beta-1}$$

$$\mathrm{B}(a,b) \equiv \int_0^1 x^{a-1}(1-x)^{b-1}dx = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} \qquad \Gamma(x) = (x-1)\Gamma(x-1)$$

average over
posterior dis.

$$\overline{p} = \int pP(p \mid m)dp =$$

$$\frac{\int p^{m+\alpha}(1-p)^{N-m+\beta-1}dp}{\int p^{m+\alpha-1}(1-p)^{N-m+\beta-1}dp} = \frac{m+\alpha}{N+\alpha+\beta}$$

pseudo counts

Generalize to T alphabet

Multinomial $\quad P(\vec{m} \mid \vec{p}) = \prod_{i=1}^{T} p_i^{m_i} \qquad \sum_{i=1}^{T} m_i = N \qquad \sum_{i=1}^{T} p_i = 1$

Dirichlet distr $\quad \pi(\vec{p}) = \prod_{i=1}^{T} p_i^{\alpha_i - 1}$

$$\overline{p}_i = \frac{m_i+\alpha_i}{\sum_{i=1}^{T}(m_i+\alpha_i)}$$

Example C: Segmentation Revisited

a sequence of head (1) and tail (0) is generated by first using a coin with p_1 and then change to a coin with p_2 the change point unknown

Data = (0010100000000101111011111100010)

Infer the distribution of parameters as well as missing data

$$P(seq \mid p_1, p_2, x) = p_1^{h_1(x)}(1-p_1)^{t_1(x)} p_2^{h_2(x)}(1-p_2)^{t_2(x)}$$

x position right before the change

$h_1(x)$ number of heads up to x. $t_1(x)$ number of tails up to x

$h_2(x)$ number of heads after x. $t_2(x)$ number of tails after x

N total number of tosses

In the Bayesian approach, we treat the parameters and the missing data on the same footing

posterior dis. $$P(p_1, p_2, x \mid seq) = \frac{P(seq|p_1,p_2,x)\pi(p_1,p_2,x)}{\sum_x \iint dp_1 dp_2 P(seq|p_1,p_2,x)\pi(p_1,p_2,x)}$$

uniform on x

prior $$\pi(p_1, p_2, x) = p_1^{\alpha_1 - 1}(1-p_1)^{\beta_1 - 1} p_2^{\alpha_2 - 1}(1-p_2)^{\beta_2 - 1} / (N+1)$$

posterior dis. of x $$P(x \mid seq) = \iint dp_1 dp_2 P(p_1, p_2, x \mid seq)$$

$$P(x \mid seq) = \frac{\mathrm{B}(h_1(x)+\alpha_1, t_1(x)+\beta_1)\mathrm{B}(h_2(x)+\alpha_2, t_2(x)+\beta_2)}{\sum_x \mathrm{B}(h_1(x)+\alpha_1, t_1(x)+\beta_1)\mathrm{B}(h_2(x)+\alpha_2, t_2(x)+\beta_2)}$$

Problem sets

1. Suppose that $x_1, x_2, \ldots, x_N$ are independent and identically distributed (i.i.d) sample drawn from a normal distribution $N(\mu, \sigma^2)$

a) show that the maximum likelihood estimate (MLE) for the mean and variance is given by the following formula

$$\hat{\mu} = \bar{x} = \sum_{i=1}^{N} x_i / N$$

$$\hat{\sigma}^2 = \sum_{i=1}^{N} (x_i - \bar{x})^2 / N$$

b) calculate the mean and variance of $\hat{\mu}$ and $\hat{\sigma}^2$ under repeated sampling. Show that the MLE converge to the true values with $1/\sqrt{N}$ error

hint: recall from my second lecture that $N\hat{\sigma}^2 / \sigma^2$ has a chi-square distribution with (N-1) degrees of freedom

Problem sets

2. Maximum likelihood estimate for the parameters of multinomial distribution. Consider N independent trials, each can result in one of T possible outcomes, with probabilities $p_1, p_2, \ldots, p_T$
 the observed numbers for the T possible outcomes are $m_1, m_2, \ldots, m_T$
 calculate the MLEs for the probabilities.

 hint: write down the likelihood function $P(\vec{m} \mid \vec{p}) = const \prod_{i=1}^{T} p_i^{m_i}$
 and use Lagrangian multiplier to implement the constraint $\sum_{i=1}^{T} p_i = 1$

3. Entropic segmentation. Consider the example C in my previous lecture. use the some observed data. Suppose you know that the change point x is between 12 and 16. Find x that minimize the entropy (maximizing the likelihood)

4. Bayesian inference for a Poisson process. Suppose that four samples (1,1,0,3) are drawn from a Poisson distribution specified by a mean λ. Assuming a uniform prior distribution for λ.
 calculate the posterior distribution of λ, $\hat{\lambda}$ that maximize the posterior distribution, and $\bar{\lambda}$ which is the average over posterior distribution.