Will Monroe
CS 109

Supplementary Handout #4
August 15, 2017

Final Review Session

Problems by Chris Piech

Topics

Event-based probability

- combinatorics (combinations, permutations for distinct/indistinct, divider method)
- Principle of Inclusion/Exlcusion, De Morgan's laws, Bayes' theorem, Law of Total Probability, chain rule $P(A) = P(A|B)P(B) + P(A|B^C)P(B^C)$
- conditional probability $P(A|B) = P(AB)/P(B)$
- independence $P(AB) = P(A)P(B)$, conditional independence $P(AB|C) = P(A|C)P(B|C)$

Random Variables

Discrete

- PMF $p_Y(k) = P(Y = k)$, CMF $F_Y(k) = P(Y \leq k)$
- expectation and variance $\text{Var}(X) = E[X^2] - E[X]^2$
- Bernoulli, Binomial, Poisson, Geometric, Negative Binomial (know what kind of processes each distribution models)

Continuous

- PDF $\int_a^b dk\, f_Y(k) = P(a \leq Y \leq b)$, CDF $F_Y(k) = P(Y \leq k)$
- Exponential, Uniform, Normal

Multivariate

- joint distribution $P_{X,Y}(x, y) = P(X = x, Y = y)$
- marginal distribution (we don't care about one of the variables anymore so sum over it)
- conditional distribution (fix one of the variables)
- independence $f_{X,Y}(x, y) = g(x)h(y)$
- covariance $\text{Cov}(X, Y) = E[XY] - E[X]E[Y]$
- correlation $\rho = \text{Cov}(X, Y)/\sqrt{\text{Var}(X)\text{Var}(Y)}$
- conditional expectation, law of total expectation

Sampling / Parameter Estimation

- unbiased, consistent estimators
- sample mean and sample variance, variance of the sample mean
- p-value and bootstrapping
- Markov and Chebychev bounds, weak and strong law of large numbers
- Central Limit Theorem
- Maximum Likelihood Estimator, Maximum A Posteriori Estimator
- Naive Bayes Classifier, Logistic Regression classifier

Practice Problems

Here are four exam questions which were either on previous exams or were candidates for them. We will go over as many answers as we can in the review session. We probably won't get through all of them, and if so, please use the extra problems as practice.

1. You want to know the probability that Steph Curry will make his next free throw. Let X be an indicator variable that is 1 if he makes it and model $X \sim \text{Ber}(p)$.

   a. In his first 10 attempts of the season, Curry has made 6 free throws. What is the maximum likelihood estimate for p, given this observation?

   b. Your friend Thomas Bayes informs you that Curry is actually quite a good free throw shooter and his average is most likely somewhere between 0.85 and 0.95. You want to account for this belief in your estimate of p, so you define a prior $p \sim \text{Beta}(90, 10)$. What is the maximum a posteriori estimate of p?

2. A colleague has collected samples of heights of corgis that live on two different islands. The colleague collects 30 samples from both islands and observes that the island A has a sample mean that is 3 cm greater than island B. The colleague wants to make a scientific claim that corgis on island A are significantly taller than corgis on island B. You are skeptical. It is possible that heights are identically distributed across both islands and that the observed difference in means was a result of chance and a small sample size (the null hypothesis).

   To calculate the probability of the null hypothesis, find the probability that two sets of 30 numbers (E, F) which are IID samples from *the same* Gaussian with mean 100 and variance 20 have a difference in means greater than or equal to the one your friend observed.

Figure 1: A sample of island-dwelling corgis. (Image by Meme Binge)

   a. How do you calculate the sample mean of a set of 30 numbers?

   b. What is the probability that the difference between E's sample mean and F's sample mean is greater than or equal to 3? Give an analytic solution.

3. Consider the following functions:

```
int Intersection() {
  int wait = 0;
  bool cars = randomBool(0.5); //equally likely to be true or false
  if(cars) {
    wait = randomInt(0, 2); //equally likely to be 0, 1 or 2
  }
  return wait;
}

int BikeSimulation() {
  int time = 0;
  for(int i = 0; i < 4; i++){
    time += Intersection();
  }
  return time;
}
```

Let W = the value returned by `Intersection()`.

a. What is $E[W]$? Give a numeric answer.

b. What is $\text{Var}(W)$? Give a numeric answer.

c. Let A be an indicator variable that is 1 if `cars` is true. What is $\text{Cov}(A, W)$?

d. Let T = the return value of `BikeSimulation()`. What is $E[T]$? You can express your answers in terms of $E[W]$ and $\text{Var}(W)$.

e. What is $\text{Var}(T)$? You can express your answers in terms of $E[W]$ and $\text{Var}(W)$.

4. You have a newly discovered document that dates back to the 1600s and you want to identify whether or not it was written by Shakespeare. Before doing any analysis, your prior belief is that the document is equally likely to be authored by Shakespeare or not by Shakespeare.

   To assist in your analysis, you have k documents written by Shakespeare $D_1, \ldots, D_k$. You also have m documents written by other authors $F_1, \ldots, F_m$. In your answers you may use a function `contains(X, W)` which returns 1 if the document X contains the word W.

   a. Write an expression for the probability that a document contains the word "eyeball" given that it was written by Shakespeare.

   b. The document contains n unique words $W_1, \ldots, W_n$. Write an expression for the probability that the document was written by Shakespeare given that it contains those n words. Use the Naive Bayes assumption.

Solutions

1. a. The MLE for p for a Bernoulli variable is

$$\hat{p} = \frac{1}{m}\sum_i X_i$$

So for this problem,

$$\hat{p} = \frac{6}{10} = 0.6$$

OR you could do it from first principles:

$$\hat{p} = \arg_p \max[P(X_1, ..., X_m|p)] = \arg_p \max[p^6(1-p)^4]$$

$$= \arg_p \max[6\log p + 4\log(1-p)]$$

$$6\frac{1}{p} - 4\frac{1}{1-p} = 0$$

$$\hat{p} = \frac{6}{6+4} = 0.6$$

b. Using MAP, we get

$$\hat{p} = \arg_p \max[P(p|X_1, ..., X_m)] = \arg_p \max[P(X_1, ..., X_m|p)f(p)]$$

$$P(X_1, ..., X_m|p) = p^6(1-p)^4$$
$$f(p) = Cp^{90-1}(1-p)^{10-1}$$

$$\hat{p} = \arg_p \max[p^6(1-p)^4 \times Cp^{90-1}(1-p)^{10-1}]$$
$$= \arg_p \max[Cp^{95}(1-p)^{13}]$$

$$\hat{p} = \frac{95}{95+13} \approx 0.88$$

2. a.

$$\bar{X} = \frac{1}{30}\sum_{i=1}^{30} X_i$$

b. The central limit theorem tells us that the mean of 30 IID samples with mean 100 and variance 20 follows a normal distribution. Specifically:

$$\bar{E} \sim \mathrm{N}(100, 20/30)$$

$$\bar{F} \sim \mathrm{N}(100, 20/30)$$

Since we're interested in the difference of sample means, we have:

$$\bar{E} - \bar{F} \sim \mathrm{N}(0, 40/30)$$

Now we calculate the probability of getting a more extreme difference (greater than 3).

$$P(|\bar{E} - \bar{F}| \geq 3) = 2P(\bar{E} - \bar{F} \geq 3) = 2P\left(Z \geq \frac{3}{\sqrt{40/30}}\right) = 2(1 - \Phi(2.60)) = 0.0094$$

Since the p-value is less than 0.05 we can reject the null hypothesis: these samples probably did not come from the same normal distribution with mean 100 and variance 20.

3. a.

$$E[W] = 0(0.5 + 0.5 \cdot (1/3)) + 1(0.5 \cdot (1/3)) + 2(0.5 \cdot (1/3)) = 1/2$$

b.

$$\mathrm{Var}(W) = E[W^2] - E[W]^2$$
$$E[W^2] = 0^2(0.5 + 0.5 \cdot (1/3)) + 1^2(0.5 \cdot (1/3)) + 2^2(0.5 \cdot (1/3)) = 5/6$$

$$\mathrm{Var}(W) = 5/6 - (1/2)^2 = 7/12$$

c.

$$\mathrm{Cov}(A, W) = E[AW] - E[A]E[W]$$
$$E[A] = 1/2$$
$$E[AW] = E[AW|A = 1]P(A = 1) + E[AW|A = 0]P(A = 0)$$
$$= [0(1/3) + 1(1/3) + 2(1/3)](1/2) + 0 = 1/2$$

$$\mathrm{Cov}(A, W) = 1/4$$

d. Let W_i be value returned by the ith call to `Intersection`.

$$E[T] = E\left[\sum_{i=1}^{4} W_i\right] = \sum_{i=1}^{4} E[W_i] = 4E[W]$$

e.

$$\mathrm{Var}[T] = \mathrm{Var}\left[\sum_{i=1}^{4} W_i\right] = \sum_{i=1}^{4} \mathrm{Var}(W_i) + \sum_{i \neq j} \mathrm{Cov}(W_i, W_j)$$

Since the Ws are independent, the second term is 0. then:

$$\mathrm{Var}(T) = 4\mathrm{Var}(W)$$

4. Let S be the event that a document was written by Shakespeare.

a. Let W be event that document contains the word "eyeball".

$$P(W|S) = \frac{1}{k} \sum_{i=1}^{k} \texttt{contains}(D_i, \text{"eyeball"})$$

b. By Bayes' theorem:

$$P(S|W_1, ...W_n) = \frac{P(W_1, ..., W_n|S)P(S)}{P(W_1, ..., W_n|S)P(S) + P(W_1, ..., W_n|S^C)P(S^C)}$$

Since the problem states "your prior belief is that the document is equally likely to be authored by Shakespeare or not by Shakespeare", $P(S) = P(S^C) = 0.5$.

Using the Naive Bayes assumption:

$$P(W_1, ..., W_n|S) = \Pi_{i=1}^{n} P(W_i|S)$$

$$P(W_1, ..., W_n|S^C) = \Pi_{i=1}^{n} P(W_i|S^C)$$

$P(W_i|S)$ can be calculated using the method from part a and $P(W_i|S^C)$ can be calculated using the method from part a (summing over the documents F instead of D).