Before reading the solutions, we encourage you to go through the exam first. To help guide your studying, we went through and tagged each of the exam questions with relevant topics. If you are struggling with a particular question, we encourage you to brush up on the tagged topic. Note that these are not comprehensive are rather intended to serve as a starting point for your studying!

As always, please come to Office Hours or ask on Piazza for clarifications! Good luck studying! ☺

Practice Final Solutions

1. Probabilities with card decks

2. a. Probability calculations
2. b. Probability calculations

3. a. Expectation
3. b. Sum of Poisson RVs
3. c. CDF calculations

4. a. Covariance
4. b. Variance Properties

5. Central limit theorem

6. a. Expectation and Variance
6. b. Expectation and Variance

7. a. Probability distribution
7. b. Expectation
7. c. Expectation

8. a. MLE and log-likelihood function
8. b. MLE

9. a. MLE
9. b. Naïve Bayes

Mehran Sahami
CS109

Handout #20
June 1, 2018

Practice Final Examination Solutions

1. Answer: 24/51 = 8/17. There are multiple ways to obtain this answer; here are two:

The first (common) method is to sum over all possibilities for the rank of the first card drawn multiplied by the probability that the second card has greater rank, given the rank of the first card. The first card drawn can be of any of the 13 ranks with equal probability (= 1/13). Let i be the rank of the first card. After the first card is chosen, 51 cards remain, of which $4(13 - i)$ have a rank greater than i.

$$\sum_{i=1}^{13} P(\text{Rank of first card} = i) P(\text{Rank of second card} > i \mid \text{Rank of first card} = i)$$

$$= \sum_{i=1}^{13} \frac{1}{13} \cdot \frac{4(13-i)}{51} = \frac{4}{(13)(51)} \sum_{i=1}^{13} (13-i) = \frac{4}{(13)(51)} \left(13^2 - \sum_{i=1}^{13} i \right)$$

$$= \frac{4}{(13)(51)} \left(13^2 - \frac{(13)(14)}{2} \right) = \frac{4}{51}(13-7) = \frac{24}{51} = \frac{8}{17}$$

The second method is to solve this problem using symmetry. After the first card is drawn, there are 51 cards remaining. Of those 51 cards there are 48 (= 51 – 3) that have a rank different rank than the first card drawn. For a randomly chosen rank for the first card, by symmetry, half of the remaining cards (24 = 48/2) will have a rank higher than the first card, giving us 24/51.

2.

a. In order to sell the share of HCI exactly 4 days after buying it, it means that the first two days after buying it must have included one day of increasing price (denote that as U) and one day of decreasing price (denoted that by D), then followed by two consecutive days of increasing or decreasing price. Thus the possible outcomes are:

$$\begin{aligned} \text{P(Sell on day 4)} &= \text{P(UDUU)} + \text{P(DUUU)} + \text{P(UDDD)} + \text{P(DUDD)} \\ &= p(1-p)p^2 + (1-p)p^3 + p(1-p)^3 + (1-p)p(1-p)^2 \\ &= 2p(1-p)(p^2 + (1-p)^2) \end{aligned}$$

b. There are two common ways to compute this. The first is to define a recurrence relation. Namely, the probability you eventually sell for a gain is the probability that you either have two U days in a row, or that you have a U day and a D day (so the stock is again at the starting price of \$10), multiplied by the probability that you eventually sell for a gain. Formally, this can be written as:

$$\begin{aligned} \text{P(sell for \$12)} &= \text{P(UU)} + \text{P(sell for \$12 and UD)} + \text{P(sell for \$12 and DU)} \\ &= p^2 + \text{P(sell for \$12)P(UD)} + \text{P(sell for \$12)P(DU)} \\ &= p^2 + \text{P(sell for \$12)}p(1-p) + \text{P(sell for \$12)}(1-p)p \\ &= p^2 + 2\text{P(sell for \$12)}p(1-p) \end{aligned}$$

Let X = P(sell for \$12), and solve for X yielding:

$$\begin{aligned} X &= 2Xp(1-p) + p^2 \\ 1 &= 2p(1-p) + p^2/X \\ p^2/X &= 1 - 2p(1-p) \\ X &= p^2/(1 - 2p(1-p)) \\ X &= p^2/(p^2 + (1-p)^2) \end{aligned}$$

So, $\text{P(sell for \$12)} = \dfrac{p^2}{p^2 + (1-p)^2}$

A second (simpler) way to compute this is using the odds that when we sell, we are selling for a gain. Here we essentially ignore (cancel out) pairs composed of a U and a D before the sale, and simply focus on whether the two days that determine the sale are Us or Ds. Formally, we have the following (which immediately gives us the answer):

$$\text{P(sell for \$12)} = \text{P(2 Us)}/(\text{P(2 Us)} + \text{P(2 Ds)}) = \frac{p^2}{p^2 + (1-p)^2}$$

3.

Let A = number of type W machines on "watch list"
Let B = number of type X machines on "watch list"

Note that A ~ Bin(10, 0.2) and B ~ Bin(10, 0.2).
Thus, we have: E[A] = 10(0.2) = 2 and E[B] = 10(0.2) = 2

a. Since A and W_i are independent and B and X_i are independent:

$$E[Y] = E[A]E[W_i] + E[B]E[X_j]$$
$$= (2)(4) + (2)(5) = 8 + 10 = 18$$

b. We can define random variable $C = W_a + W_b + W_c$. Noting that all W_i are independent, we have that: C ~ (Poi(4) + Poi(4) + Poi(4)) = Poi(12)

Here, Y = C, so we have:

$$P(Y \geq 20) = P(C \geq 20) = \sum_{i=20}^{\infty} \frac{e^{-12}12^i}{i!}$$

c. We can define random variable $D = X_a + X_b + X_c$. Noting that all X_i are independent, we have that: D ~ N(5, 3) + N(5, 3) + N(5, 3) = N(15, 9)

Since the X_j are Normally distributed to begin with, they are continuous variables and so is their sum. So, in computing a probability involving the sum, there is no need to approximate a discrete quantity using a continuity correction. Here, Y = D, so we have:

$$P(Y \geq 20) = P(D \geq 20) = 1 - P(D < 20) = 1 - P(Z < \frac{20-15}{\sqrt{9}}) = 1 - P(Z < 1.67)$$
$$= 1 - \phi(1.67) \approx 1 - 0.9525 = 0.0475$$

Here is what you would have gotten if you had used the continuity correction (which in this particular case, we gave full credit for when grading the problem):

$$P(Y \geq 20) = P(D \geq 19.5) = 1 - P(D < 19.5) = 1 - P(Z < \frac{19.5-15}{\sqrt{9}}) = 1 - P(Z < 1.5)$$
$$= 1 - \phi(1.5) \approx 1 - 0.9332 = 0.0668$$

4.

Let indicator variable $X_i = 1$ if the i-th integer generated is a 1, and 0 otherwise.
Let indicator variable $Y_j = 1$ if the j-th integer generated is a 5, and 0 otherwise.

Note that $X = \sum_{i=1}^{n} X_i$ and likewise $Y = \sum_{j=1}^{n} Y_j$

a. Note: $E[X_i] = P(X_i) = 1/5$ and likewise $E[Y_j] = P(Y_i) = 1/5$.

Also note: $E[X_i, Y_j] = 0$ whenever $i = j$, since a 1 and 5 cannot both be the i-th integer.

$\text{Cov}(X_i, Y_j) = E[X_i Y_j] - E[X_i]\,E[Y_j]$

$$= \begin{cases} -\frac{1}{25} & \text{when } i = j, \text{ since } E[X_i Y_j] = 0 \text{ when } i = j \\ 0 & \text{otherwise (when } i \neq j\text{), by independence} \end{cases}$$

$$\text{So, } \text{Cov}(X, Y) = \text{Cov}\left(\sum_{i=1}^{n} X_i, \sum_{j=1}^{n} Y_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} \text{Cov}(X_i, Y_j) = \sum_{i=1}^{n} \text{Cov}(X_i, Y_i) = (n)\frac{-1}{25} = -\frac{n}{25}$$

b.

By definition: $\rho(X, Y) = \dfrac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}$

We note that $X_i \sim \text{Ber}(p = 1/5)$ and likewise $Y_j \sim \text{Ber}(p = 1/5)$

Thus, $\text{Var}(X_i) = \text{Var}(Y_j) = p(1 - p) = (1/5)(4/5) = 4/25$

Since $X = \sum_{i=1}^{n} X_i$ and all the X_i are independent: $\text{Var}(X) = n\text{Var}(X_i) = 4n/25$

Also, $\text{Var}(Y) = \text{Var}(X) = 4n/25$

So, $\rho(X, Y) = \dfrac{-n/25}{\sqrt{(4n/25)^2}} = \dfrac{-n/25}{4n/25} = -\dfrac{1}{4}$

5. Let n = the number of machines we purchase. Let Y_i = the total number of weeks we use that the i-th machine purchased until it dies. Note that: $X = \sum_{i=1}^{n} Y_i$

We want to compute an expression for n, such that:

$$P(X > 2000) = P(\sum_{i=1}^{n} Y_i > 2000) \geq 0.95$$

Note that $E[X] = E\left[\sum_{i=1}^{n} Y_i\right] = n\, E[Y] = 100n$

Similarly, $Var(X) = Var\left[\sum_{i=1}^{n} Y_i\right]$. Since all Y_i are independent, we have:

$Var\left[\sum_{i=1}^{n} Y_i\right] = n\, Var(Y_i) = 25n$. Thus, $Var(X) = 25n$.

Now, we apply the Central Limit Theorem:

$$P(X > 2000) = P(\sum_{i=1}^{n} Y_i > 2000) = P(\frac{\sum_{i=1}^{n} Y_i - 100n}{\sqrt{25n}} > \frac{2000 - 100n}{\sqrt{25n}}) = P(Z > \frac{400 - 20n}{\sqrt{n}})$$

We want to have:

$$P(Z > \frac{400 - 20n}{\sqrt{n}}) \geq 0.95 \;\Rightarrow\; 1 - P(Z \leq \frac{400 - 20n}{\sqrt{n}}) \geq 0.95 \;\Rightarrow\; 1 - \Phi(\frac{400 - 20n}{\sqrt{n}}) \geq 0.95$$

Noting that $1 - \Phi(C) = \Phi(-C)$, we obtain:

$$\Phi(\frac{20n - 400}{\sqrt{n}}) \geq 0.95 \;\Rightarrow\; \frac{20n - 400}{\sqrt{n}} \geq 1.645, \text{ since } \Phi(1.645) \approx 0.95.$$

Here we want to determine the minimal value of n satisfying the inequality above. Clearly, n = 20 is too small, since $\frac{20(20) - 400}{\sqrt{20}} = 0$. We consider n = 21, giving us: $\frac{20(21) - 400}{\sqrt{21}} = \frac{420 - 400}{\sqrt{21}} = \frac{20}{\sqrt{21}}$. We know that $\sqrt{21} \leq 5$, so $\frac{20}{\sqrt{21}} \geq \frac{20}{5} = 4 \geq 1.645$, so 21 machines is sufficient to give us $P(X > 2000) \geq 0.95$.

6. Let X = value returned by `Near()`.

$E[X] = 1/4(2 + 4 + E[6 + X] + E[8 + X]) = 1/4(2 + 4 + 6 + E[X] + 8 + E[X])$
$= 1/4(20 + 2E[X]) = 5 + 1/2E[X]$

So, $E[X] = 10$

$E[X^2] = 1/4(2^2 + 4^2 + E[(6 + X)^2] + E[(8 + X)^2]$
$= 1/4(4 + 16 + 36 + 12E[X] + E[X^2] + 64 + 16E[X] + E[X^2])$
$= 1/4(120 + 28E[X] + 2E[X^2])$
$= 1/4(120 + 28(10) + 2E[X^2]) = 1/4(400 + 2E[X^2]) = 100 + 1/2E[X^2]$

So, $E[X^2] = 2(100) = 200$

a.

$E[Y] = 1/3(2 + E[2 + X] + E[4 + Y]) = 1/3(2 + 2 + E[X] + 4 + E[Y])$
$= 1/3(8 + E[X] + E[Y]) = 1/3(8 + 10 + E[Y]) = 18/3 + 1/3E[Y]$

So, $E[Y] = 9$

b.

$E[Y^2] = 1/3(2^2 + E[(2 + X)^2] + E[(4 + Y)^2]$
$= 1/3(4 + 4 + 4E[X] + E[X^2] + 16 + 8E[Y] + E[Y^2])$
$= 1/3(24 + 40 + E[X^2] + 8(9) + E[Y^2])$
$= 1/3(136 + 200 + E[Y^2])$
$= 1/3(336 + E[Y^2])$

So, $E[Y^2] = 336/2 = 168$

$Var(Y) = E[Y^2] – E[Y]^2 = 168 – (9)^2 = 168 – 81 = 87$

7.

a.

P(X = 1) = 0

P(X = 2) = 1/4 = 16/64

P(X = 3) = (3/4)(2/4) = 3/8 = 6/16 = 24/64

P(X = 4) = (3/4)(2/4)(3/4) = 9/32 = 18/64

P(X = 5) = (3/4)(2/4)(1/4) = 3/32 = 6/64

b. $E[X] = \sum_{i=1}^{5} i\,P(X=i)$ = 1(0) + 2(16/64) + 3(24/64) + 4(18/64) + 5(6/64)

= 238/64 = 103/32

c. In the general case: $P(X=i) = \left(\prod_{j=n-i+2}^{n \text{ or } n-1} \frac{j}{n}\right)\frac{i-1}{n}$, since we need to hash $i - 1$ strings without any collisions, and then get a collision on the last (i-th) string hashed.

Note that the product above could be written with either an n or $n - 1$ as the top index. Either form is equivalent, since the form with n as the top index just does an extra multiplication of the product by 1 (= n/n).

Using the definition of expectation, we have:

$$E[X] = \sum_{i=2}^{n+1} i\,P(X=i) = \sum_{i=2}^{n+1} i\left(\prod_{j=n-i+2}^{n \text{ or } n-1} \frac{j}{n}\right)\frac{i-1}{n}$$

$$= \sum_{i=2}^{n+1} \frac{i(i-1)n!}{n^i(n-i+1)!}$$

8.

a. The mass function for the Geometric distribution with given parameter p is $f(X_i \mid p) = p(1-p)^{X_i - 1}$, where $X_i \geq 0$.

The likelihood function to maximize is:

$$L(p) = \prod_{i=1}^{n} p(1-p)^{X_i - 1}$$

So, the log-likelihood function to maximize is:

$$LL(p) = \sum_{i=1}^{n} [\log p + (X_i - 1)\log(1-p)]$$

Taking the derivative of LL(p) w.r.t. p, and setting it to 0, yields:

$$\frac{\partial LL(p)}{\partial p} = \sum_{i=1}^{n} [\frac{1}{p} + (X_i - 1)\frac{-1}{1-p}] = 0$$

Solving for p gives us: $\frac{n}{p} = \frac{1}{1-p}\sum_{i=1}^{n}(X_i - 1) \quad \Rightarrow \quad \frac{1-p}{p} = \frac{1}{n}\sum_{i=1}^{n}(X_i - 1)$

$$\frac{1}{p} - 1 = \left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] - 1 \quad \Rightarrow \quad p_{\text{MLE}} = \frac{1}{\frac{1}{n}\sum_{i=1}^{n} X_i} = \frac{1}{\overline{X}}$$

b. We have: $p_{\text{MLE}} = \frac{1}{\frac{1}{n}\sum_{i=1}^{n} X_i} = \frac{1}{\frac{1}{5}(20)} = \frac{5}{20} = \frac{1}{4}$

9.

a. The likelihood function is the probability mass function of a Bernoulli with probability p_{ij}:

$$\text{Likelihood} = (p_{ij})^{S_{ij}} \cdot (1 - p_{ij})^{1 - S_{ij}}$$
$$\text{Log Likelihood} = S_{ij} \log(p_{ij}) + (1 - S_{ij}) \log(1 - p_{ij})$$

b. Using chain rule:

$$\frac{\partial \text{LL}}{\partial a_i} = \frac{\partial \text{LL}}{\partial p_{ij}} \cdot \frac{\partial p_{ij}}{\partial a_i}$$

Just like in a deep learning network:

$$\frac{\partial \boldsymbol{LL}}{\partial \boldsymbol{p}_{ij}} = \frac{S_{ij}}{p_{ij}} - \frac{(1 - S_{ij})}{(1 - p_{ij})}$$

Staring with the equation for p_{ij}

$$\frac{\partial p_{ij}}{\partial a_i} = p_{ij}(1 - p_{ij}) \cdot \frac{\partial}{\partial a_i}(a_i - d_j)$$
$$= p_{ij}(1 - p_{ij})$$

You can optionally reduce your equations further. If you substitute and cancel you will get that:

$$\frac{\partial p_{ij}}{\partial a_i} = S_{ij} - p_{ij}$$

c. Using chain rule:

$$\frac{\partial \text{LL}}{\partial d_j} = \frac{\partial \text{LL}}{\partial p_{ij}} \cdot \frac{\partial p_{ij}}{\partial d_j}$$

This part is the same:

$$\frac{\partial \boldsymbol{LL}}{\partial \boldsymbol{p}_{ij}} = \frac{S_{ij}}{p_{ij}} - \frac{(1 - S_{ij})}{(1 - p_{ij})}$$

Starting with the equation for p_{ij}:

$$\frac{\partial p_{ij}}{\partial d_j} = p_{ij}(1 - p_{ij}) \cdot \frac{\partial}{\partial d_j}(a_i - d_j)$$
$$= p_{ij}(1 - p_{ij})(-1)$$

You can optionally reduce your equations further. If you substitute and cancel you will get that:

$$\frac{\partial p_{ij}}{\partial d_j} = p_{ij} - S_{ij}$$

d. Use can estimate the value of all parameters using gradient ascent. Gradient ascent repeatedly takes a step along the gradient with a fixed step size. Just like when we implemented logistic regression, we can program our closed form mathematical solution for gradients to efficiently calculate the gradient for any values of our parameters.