

## HW 13

### Sundry

Before you start your homework, write down your team. Who else did you work with on this homework? List names and email addresses. (In case of homework party, you can also just describe the group.) How did you work on this homework? Working in groups of 3-5 will earn credit for your "Sundry" grade.

Please copy the following statement and sign next to it:

*I certify that all solutions are entirely in my words and that I have not looked at another student's solutions. I have credited all external sources in this write up.*

### 1 Normal Distribution

Recall the following facts about the normal distribution: if  $X \sim \mathcal{N}(\mu, \sigma^2)$ , then the random variable  $Z = (X - \mu)/\sigma$  is standard normal, i.e.  $Z \sim \mathcal{N}(0, 1)$ . There is no closed-form expression for the CDF of the standard normal distribution, so we define  $\Phi(z) = \mathbb{P}[Z \leq z]$ . You may express your answers in terms of  $\Phi(z)$ .

The average jump of a certain frog is 3 inches. However, because of the wind, the frog does not always go exactly 3 inches. A zoologist tells you that the distance the frog travels is normally distributed with mean 3 and variance 1/4.

- (a) What is the probability that the frog jumps more than 4 inches?
- (b) What is the probability that the distance the frog jumps is between 2 and 4 inches?

### 2 Binomial CLT

In this question we will explicitly see why the central limit theorem holds for the binomial distribution as the number of coin tosses grows.

Let  $X$  be the random variable showing the total number of heads in  $n$  independent coin tosses.

- (a) Compute the mean and variance of  $X$ . Show that  $\mu = \mathbb{E}[X] = n/2$  and  $\sigma^2 = \text{var} X = n/4$ .
- (b) Prove that  $\mathbb{P}[X = k] = \binom{n}{k}/2^n$ .
- (c) Show by using Stirling's formula that

$$\mathbb{P}[X = k] \simeq \frac{1}{\sqrt{2\pi}} \left(\frac{n}{2k}\right)^k \left(\frac{n}{2(n-k)}\right)^{n-k} \sqrt{\frac{n}{k(n-k)}}.$$

In general we expect  $2k$  and  $2(n-k)$  to be close to  $n$  for the probability to be non-negligible. When this happens we expect  $\sqrt{\frac{n}{k(n-k)}}$  to be close to  $\sqrt{\frac{n}{(n/2) \times (n/2)}} = \frac{2}{\sqrt{n}}$ . So replace that part of the formula by  $2/\sqrt{n}$ .

- (d) In order to normalize  $X$ , we need to subtract the mean, and divide by the standard deviation. Let  $Y = (X - \mu)/\sigma$  be the normalized version of  $X$ . Note that  $Y$  is a discrete random variable. Determine the set of values that  $Y$  can take. What is the distance  $d$  between two consecutive values?
- (e) Let  $X = k$  correspond to the event  $Y = t$ . Then  $X \in [k - 0.5, k + 0.5]$  corresponds to  $Y \in [t - d/2, t + d/2]$ . For conceptual simplicity, it is reasonable to assume that the mass at point  $t$  is distributed uniformly on the interval  $[t - d/2, t + d/2]$ . We can capture this with the idea of a "probability density" and say that the probability density on this interval is just  $\mathbb{P}[Y = t]/d = \mathbb{P}[X = k]/d$ .

Compute  $k$  as a function of  $t$ . Then substitute that for  $k$  in the approximation you have from part c to find an approximation for  $\mathbb{P}[Y = t]/d$ . Show that the end result is equivalent to:

$$\frac{1}{\sqrt{2\pi}} \left[ \left(1 + \frac{t}{\sqrt{n}}\right)^{1+t/\sqrt{n}} \left(1 - \frac{t}{\sqrt{n}}\right)^{1-t/\sqrt{n}} \right]^{-n/2}$$

- (f) As you can see, we have expressions of the form  $(1+x)^{1+x}$  in our approximation. To simplify them, write  $(1+x)^{1+x}$  as  $\exp((1+x)\ln(1+x))$  and then replace  $(1+x)\ln(1+x)$  by its Taylor series.

The Taylor series up to the  $x^2$  term is  $(1+x)\ln(1+x) \simeq x + x^2/2 + \dots$  (feel free to verify this by hand). Use this to simplify the approximation from the last part. In the end you should get the familiar formula that appears inside the CLT:

$$\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right).$$

(The CLT is essentially taking a sum with lots of tiny slices and approximating it by an integral of this function. Because the slices are tiny, dropping all the higher-order terms in the Taylor expansion is justified.)

### 3 Why Is It Gaussian?

Let  $X$  be a normally distributed random variable with mean  $\mu$  and variance  $\sigma^2$ . Let  $Y = aX + b$ , where  $a$  and  $b$  are non-zero real numbers. Show explicitly that  $Y$  is normally distributed with mean  $a\mu + b$  and variance  $a^2\sigma^2$ . (Your proof should be more explicit than what's in the class notes. One approach is to start with the cumulative distribution function of  $Y$  and use it to derive the probability density function of  $Y$ .)

### 4 Deriving Chebyshev's Inequality

Recall Markov's Inequality, which applies for non-negative  $X$  and  $\alpha > 0$ :

$$\mathbb{P}[X \geq \alpha] \leq \frac{\mathbb{E}[X]}{\alpha}$$

Use an appropriate substitution for  $X$  and  $\alpha$  to derive Chebyshev's Inequality, where  $\mu$  denotes the expected value of  $Y$ .

$$\mathbb{P}[|Y - \mu| \geq k] \leq \frac{\text{var}(Y)}{k^2}$$

### 5 Markov's Inequality and Chebyshev's Inequality

A random variable  $X$  has variance  $\text{var}(X) = 9$  and expectation  $\mathbb{E}[X] = 2$ . Furthermore, the value of  $X$  is never greater than 10. Given this information, provide either a proof or a counterexample for the following statements.

- (a)  $\mathbb{E}[X^2] = 13$ .
- (b)  $\mathbb{P}[X = 2] > 0$ .
- (c)  $\mathbb{P}[X \geq 2] = \mathbb{P}[X \leq 2]$ .
- (d)  $\mathbb{P}[X \leq 1] \leq 8/9$ .
- (e)  $\mathbb{P}[X \geq 6] \leq 9/16$ .
- (f)  $\mathbb{P}[X \geq 6] \leq 9/32$ .

### 6 Practical Confidence Intervals

- (a) It's New Year's Eve, and you're re-evaluating your finances for the next year. Based on previous spending patterns, you know that you spend \$1500 per month on average, with a standard deviation of \$500, and each month's expenditure is independently and identically distributed. As a poor college student, you also don't have any income. How much should you have in your bank account if you don't want to go broke this year, with probability at least 95%?

- (b) As a UC Berkeley CS student, you're always thinking about ways to become the next billionaire in Silicon Valley. After hours of brainstorming, you've finally cut your list of ideas down to 10, all of which you want to implement at the same time. A venture capitalist has agreed to back all 10 ideas, as long as your net return from implementing the ideas is positive with at least 95% probability.

Suppose that implementing an idea requires 50 thousand dollars, and your start-up then succeeds with probability  $p$ , generating 150 thousand dollars in revenue (for a net gain of 100 thousand dollars), or fails with probability  $1 - p$  (for a net loss of 50 thousand dollars). The success of each idea is independent of every other. What is the condition on  $p$  that you need to satisfy to secure the venture capitalist's funding?

- (c) One of your start-ups uses error-correcting codes, which can recover the original message as long as at least 1000 packets are received (not erased). Each packet gets erased independently with probability 0.8. How many packets should you send such that you can recover the message with probability at least 99%?

## 7 Quadratic Regression

In this question, we will find the best quadratic estimator of  $Y$  given  $X$ . First, some notation: let  $\mu_i$  be the  $i$ th moment of  $X$ , i.e.  $\mu_i = \mathbb{E}[X^i]$ . Also, define  $\beta_1 = \mathbb{E}[XY]$  and  $\beta_2 = \mathbb{E}[X^2Y]$ . For simplicity, we will assume that  $\mathbb{E}[X] = \mathbb{E}[Y] = 0$  and  $\mathbb{E}[X^2] = \mathbb{E}[Y^2] = 1$ . (Note that this poses no loss of generality, because we can always transform the random variables by subtracting their means and dividing by their standard deviations.) We claim that the best quadratic estimator of  $Y$  given  $X$  is

$$\hat{Y} = \frac{1}{\mu_3^2 - \mu_4 + 1} (aX^2 + bX + c)$$

where

$$\begin{aligned} a &= \mu_3\beta_1 - \beta_2, \\ b &= (1 - \mu_4)\beta_1 + \mu_3\beta_2, \\ c &= -\mu_3\beta_1 + \beta_2. \end{aligned}$$

Your task is to prove the Projection Property for  $\hat{Y}$ .

- (a) Prove that  $\mathbb{E}[Y - \hat{Y}] = 0$ .  
 (b) Prove that  $\mathbb{E}[(Y - \hat{Y})X] = 0$ .  
 (c) Prove that  $\mathbb{E}[(Y - \hat{Y})X^2] = 0$ .

Any quadratic function of  $X$  is a linear combination of 1,  $X$ , and  $X^2$ . Hence, these equations together imply that  $Y - \hat{Y}$  is orthogonal to any quadratic function of  $X$ , and so  $\hat{Y}$  is the best quadratic estimator of  $Y$ .

## 8 LLSE and Graphs

Consider a graph with  $n$  vertices numbered 1 through  $n$ , where  $n$  is a positive integer  $\geq 2$ . For each pair of distinct vertices, we add an undirected edge between them independently with probability  $p$ . Let  $D_1$  be the random variable representing the degree of vertex 1, and let  $D_2$  be the random variable representing the degree of vertex 2.

- (a) Compute  $\mathbb{E}[D_1]$  and  $\mathbb{E}[D_2]$ .
- (b) Compute  $\text{var}(D_1)$ .
- (c) Compute  $\text{cov}(D_1, D_2)$ .
- (d) Using the information from the first three parts, what is  $L(D_2 | D_1)$ ?