Solutions

CSE 312
Foundations, II
Final Exam

Anna Karlin

June 11, 2014

DIRECTIONS:

- Closed book, closed notes except for one 8.5×11 sheet.
- Time limit 110 minutes.
- Calculators allowed.
- Grading will emphasize problem set-up over calculation.
- If possible, answer all problems on these sheets.
- Put your NAME on each sheet.
- Do not turn the page until I tell you to.
- Good luck!

1	/ 30
2	/ 70
3	/ 45
4	/ 30
5	/ 25
Total	/ 200

The following short table of values of the cumulative distribution function for the standard normal distribution may be useful.

x	-10	-8	-6	-4	-2	-1.96	-1.64	-1.00	0
$\Phi(x) = P(X < x)$	7.62e-24	6.22e-16	9.87e-10	3.17e-5	0.0228	0.025	0.05	0.159	.5

1. (30 points)

Here are some choices for solutions to the short answer questions below. They are in no particular order. Enter the solution number to each question in the box provided. You may need to use an answer more than once.

(1) $\frac{n!}{(n-m)!}$	(2) $\binom{n+m}{m}$	(3) $(n-m)!$	(4) m^n
(5) $\binom{n-1+m}{n}$	(6) $\binom{n-1+m}{m}$	(7) 2^{mn}	(8) n^m

(a) $\boxed{2}$ How many solutions over the nonnegative integers are there to the inequality

$$x_1 + x_2 + \cdots + x_n \leq m?$$

(b) $\boxed{1}$ How many length m words can be formed from an n-letter alphabet, if no letter is used more than once?

(c) $\boxed{1}$ How may injective (one-to-one) functions are there from set A to set B when $|A| = m$ and $|B| = n \geq m$.

(d) $\boxed{8}$ How many ways are there to place a total of m distinguishable balls into n distinguishable urns, with some urns possibly empty or with several balls?

(e) $\boxed{6}$ How many ways are there to place a total of m indistinguishable balls into n distinguishable urns, with some urns possibly empty or with several balls?

(f) $\boxed{1}$ How many ways are there to place a total of m distinguishable balls into n distinguishable urns, with at most one ball in each urn? (Assume $m \leq n$.)

2. (70 points)

(a) The Chinese Appetizer Problem: n people are each eating a different appetizer. All the appetizers are arranged in front of the people on a circular, rotating Chinese banquet tray. Someone spins the tray so that each person receives a random appetizer.

- What is the expected number of people who get their appetizer back?

$$E(X) = n \cdot \frac{1}{n} + 0 \cdot \left(1 - \frac{1}{n}\right) = 1$$

$$X = \begin{cases} n & \frac{1}{n} \\ 0 & 1 - \frac{1}{n} \end{cases}$$

- What is the variance of the number of people who get their appetizer back?

$$E\left[(X - E(X))^2\right] = \frac{1}{n}(n-1)^2 + \left(1 - \frac{1}{n}\right)(-1)^2 = n-1$$

- What is the probability that exactly 1 person gets his/her appetizer back?

0

- Are the events {person 1 gets his appetizer back} and {person 2 gets her appetizer back} independent?

no

- What is the conditional probability that person 1 gets his appetizer back given that person 2 gets her appetizer back?

1

- Use the Markov Inequality to give an upper bound on the probability that all n people get their appetizers back.

$$\Pr(X \geq n) \leq \frac{1}{n}$$

- What is the true probability that all n people get their appetizers back?

$$\frac{1}{n}$$

- Suppose that after spinning the tray, each person who does not get their own appetizer back eats the appetizer in front them with probability 1/2 (independently). After that, the tray is spun again, with each of the n possible rotations equally likely. (The second spin is independent of the first.) What is the expected number of people who get their appetizer back now? (Note that now a person may not get their appetizer back either because their plate didn't land in front of them after the second spin, or because it did, but someone ate their appetizer after the first spin.) You do not need to simplify your answer.

$$E(X) = E(X \mid 1^{st} \text{ spin good})\frac{1}{n} + E(X \mid 1^{st} \text{ spin bad})\left(1-\frac{1}{n}\right)$$

$$= \frac{n}{n} + \left(1-\frac{1}{n}\right) n \frac{1}{2} \cdot \frac{1}{n}$$

$$= 1 + \frac{1}{2} - \frac{1}{2n} = \frac{3}{2} - \frac{1}{2n}$$

(b) The Hat-Check Problem: n people check their hats when they arrive at a banquet. Each of them gets a random hat back. (All $n!$ permutations equally likely.)

- What is the expected number of people who get their hats back?

1

- Are the events {person 1 gets his hat back} and {person 2 gets her hat back} independent?

No, $Pr(\text{person } i \text{ gets hat back}) = \frac{1}{n}$ $Pr(1 \,\&\, 2 \text{ get hat back}) = \frac{1}{n(n-1)}$

- What is the conditional probability that person 1 gets his hat back given that person 2 gets her hat back?

$\frac{1}{n-1}$

- Use the Markov Inequality to give an upper bound on the probability that all n people get their hats back.

$Pr(X \geq n) \leq \frac{1}{n}$

- What is the true probability that all n people get their hats back?

$\frac{1}{n!}$

- What is $E(X_i X_j)$, where X_i is an indicator variable for the i-th person getting his/her own hat back?

$$E(X_i X_j) = \frac{1}{n(n-1)}$$

- Let $S_n = \sum_i X_i$, i.e., the number of people who get their hats back. Show that $E(S_n^2) = 2$. Hint: $X_i^2 = X_i$.

$$\left(\sum X_i\right)^2 = \sum X_i^2 + 2\sum_{i<j} X_i X_j$$

$$E(S_n^2) = 1 + 2\binom{n}{2}\frac{1}{n(n-1)} = 2$$

- What is the variance of S_n?

$$2 - 1 = 1$$

- Use Chebychev's inequality to show that there is at most a 1% chance that more than 10 people get their own hat back.

$$\Pr(X > 10) \le \Pr(|X - 1| \geq 10) \le \frac{1}{100}$$

3. (45 points)

Sam throws darts at a circular target of radius r and is equally likely to hit any point in the target. Let X be the distance of Sam's hit from the center.

(a) What is the cumulative distribution function of X?

$$F_X(x) = Pr(X \le x) = \frac{\pi x^2}{\pi r^2} = \frac{x^2}{r^2} \quad 0 \le x \le r$$

$$F_X(x) = \begin{cases} 1 & x > r \\ \frac{x^2}{r^2} & 0 \le x \le r \\ 0 & x < 0 \end{cases}$$

(b) What is the probability density function of X?

$$f_X(x) = \frac{d}{dx} F_X(x) = \begin{cases} 0 & x > r \\ \frac{2x}{r^2} & 0 \le x \le r \\ 0 & x < 0 \end{cases}$$

(c) What is the expectation of X?

$$E(X) = \int_0^r \frac{2x^2}{r^2} dx = \frac{2x^3}{3r^2}\Big|_0^r = \frac{2}{3} r$$

(d) What is the variance of X?

$$E(X^2) = \int_0^r \frac{2x^3}{r^2} dx = \frac{2x^4}{4r^2}\Big|_0^r = \frac{r^2}{2}$$

$$Var(X) = \frac{r^2}{2} - \left(\frac{2}{3} r\right)^2 = \frac{r^2}{18}$$

(e) Suppose that each time Sam tosses the dart, he gets \$1 if his dart hits the target at a distance at most 2 inches from the center. Assume that the radius r is 5 inches. Let W be his winnings if he tosses the dart independently 100 times. Use the Chernoff bound below to get an upper bound on the probability that he wins at least \$24. (You do not need to simplify your answer, but you do need to be clear on what μ and δ you are using in the Chernoff bound.)

$$E(W) = 100 \cdot \frac{2^2}{5^2} = 16$$

$$Pr(W \geq 24) = Pr\left(W \geq \left(1+\frac{1}{2}\right)16\right) \leq e^{-\frac{16}{4 \cdot 3}}$$

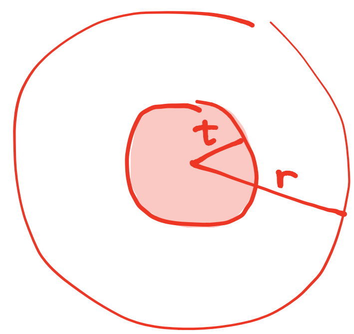
A Chernoff Bound: Suppose that $X_1, \ldots, X_n$ are independent Bernoulli random variables each with probability p of success. Let $X = \sum_{i=1}^{n} X_i$ and let $\mu = E(X)$. Then for any $0 < \delta \leq 1$,

$$Pr(X \geq (1+\delta)\mu) \leq e^{-\frac{\mu\delta^2}{3}}.$$

(f) The game changes now: The target has an inner circle of radius t and an outer circle of radius r (as in the first few parts of this problem). If $X \leq t$, Sam gets a score of $S = 1/X$. Otherwise, his score is 0. What is the cumulative distribution function of S?

$$0 \leq X \leq t \rightarrow \frac{1}{t} \leq S \leq \infty$$

$$X > t \rightarrow S = 0$$



$$F_S(z) = Pr(S \leq z) = \begin{cases} 0 & z < 0 \\ 1 - \frac{t^2}{r^2} & 0 \leq z < \frac{1}{t} \\ 1 - \frac{t^2}{r^2} + Pr\left(\frac{1}{t} \leq S \leq z\right) = 1 - \frac{t^2}{r^2} + Pr\left(\frac{1}{z} \leq X \leq t\right) & \\ = 1 - \frac{t^2}{r^2} + \frac{t^2}{r^2} - \frac{1}{z^2 r^2} = & \frac{1}{t} \leq z \end{cases}$$

4. (30 points) Circle the best single answer, one of (i) through (v), for each question. No explanations needed. You may want to refer to the table of values for the CDF of a normal distribution on the first page of the exam.

   (a) A bus route has interarrival times that are exponentially distributed with parameter $\lambda = 0.05\text{min}^{-1}$. The probability of waiting an hour or more for a bus is:

   (i) $1 - e^{-1200}$ (ii) e^{-1200} (iii) e^{-3} (iv) $1 - e^{-3}$ (v) With my luck, nearly 1

   $\lambda = \frac{1}{0.05}$ $\Pr(X \geq 60) = e^{-\lambda 60} = e^{-0.05 \cdot 60} = e^{-3}$

   (b) For a random variable Z which has a standard normal distribution, $P(Z < 2) =$

   (i) 0.0228 (ii) .0179 (iii) .9821 (iv) 0.9772 (v) none of these

   $= 1 - \Pr(Z < -2) = 1 - 0.0228 = 0.9772$

   (c) Suppose the random variable X has a normal distribution with mean 5 and variance 64. The probability that X takes on a value of at least 21 is approximately equal to

   (i) 0.05 (ii) .0985 (iii) .4015 (iv) .9015 (v) none of these

   $\Pr\left(\frac{X-5}{8} \geq \frac{21-5}{8}\right) = \Pr(Z \geq 2) = 0.0228$

   (d) Chebychev's Inequality implies that the proportion of observations that are at most 3 standard deviations of the mean is at least

   (i) 1/3 (ii) 2/3 (iii) 1/9 (iv) 8/9 (v) none of these

   $\Pr(|X - \mu| \geq 3\sigma) \leq \frac{1}{9}$

(e) The probability that a normal random variable X falls within 2 standard deviations of its mean is approximately

(i) .0228 (ii) 0.05 (iii) .9772 (iv) .0456 (v) .9544

(f) A continuous random variable X has density $f_X(x)$. In terms of $f_X(\cdot)$, the density $f_Y(y)$, where $Y = 10X - 3$ is

(i) $10f_X(y) - 3$ (ii) $f_X(10y - 3)$ (iii) $\frac{1}{10}f_X(\frac{y+3}{10})$ (iv) $10f_X(\frac{y+3}{10})$ (v) none of these

$$F_Y(y) = Pr(10X - 3 \le y) = Pr\left(X \le \frac{3+y}{10}\right) = F_X\left(\frac{3+y}{10}\right)$$

$$f_Y(y) = \frac{d}{dy} F_Y(y) = \frac{1}{10} f_X\left(\frac{3+y}{10}\right)$$

5. (25 points)

Let $k_1, \ldots, k_n$ be n samples from a geometric distribution with unknown parameter θ. (So k_1 is the number of tosses until the first heads in the first trial, k_2, the number of tosses until the first heads in the second trial, and so on.)

(a) What is the likelihood function $L(k_1, \ldots, k_n|\theta)$?

$$L(\theta) = (1-\theta)^{k_1-1}\theta\,(1-\theta)^{k_2-1}\theta \cdots (1-\theta)^{k_n-1}\theta = (1-\theta)^{(\sum k_i)-n}\,\theta^n$$

(b) What is the log-likelihood function?

$$LL(\theta) = \left(\sum k_i - n\right)\log(1-\theta) + n\log(\theta)$$

(c) What is the maximum likelihood estimator for θ? (You do not need to verify second order conditions, but you do need to show all other aspects of your derivation.)

$$LL'(\theta) = -\frac{\sum k_i - n}{1-\theta} + \frac{n}{\theta}$$

set $LL'(\theta) = 0$

$$\theta\left(\sum k_i - n\right) = (1-\theta)n$$

$$\theta = \frac{n}{\sum k_i}$$