Polytechnic Institute of NYU Prof. Özgür Özkan CS6003—Fall 2013

Quiz. December 12th, 2013

- You have two hours.
- Use a word processor to write your responses. Clearly label your solutions. Once you are done with the exam, submit the file as a private post on Piazza in one of the following formats: .doc, .docx, .pdf, .tex, .txt (if using Piazza)
- You can also use pen and paper and scan. You will get 15 extra minutes for the scanning and uploading.
- Write your name and ID number on the first page.
- Answer **all** questions.
- Explain your answers. If you do not show work, you will not get credit.
- 120 points total.
- Closed book, NO notes. You are allowed to use a word processor to write your answers, and a pdf reader to view the exam questions. You may use a SIMPLE calculator app that allows arithmetic operations.
- No cheat sheet.

You MUST show work to get credit.

Reference

Recall that if $X_1, X_2, ..., X_n$ are independent Poisson trials with $Pr[X_i = 1] = p_i$ where $X = \sum_i X_i$ and $\mu = E[X]$, then for any $\delta \in (0, 1]$:

$$Pr[X > (1 + \delta)\mu] < e^{-\mu\delta^2/4}$$

(5pts) 1. Give two events A and B, $p(A \cup B) = p(A) + p(B)$ iff

$\surd$ **A and B are disjoint.**
○ A and B are dependent

(5pts) 2. Give two events A and B, $p(A \cap B) = p(A)p(B)$ iff

○ A and B are conditional.
$\surd$ **A and B are independent**

(10pts) 3. I throw three 7 sided die. The sides are labeled with integers from 1 to 7. The even labeled sides are twice as likely to come up as the odd labeled sides. What is the expected sum of the throw? (e.g. if the first dice comes up 3, the second dice comes up 1, and the third dice comes up 7, the sum would be 11).

Solution: $((1+3+5+7+2+2+4+4+6+6)/10) \cdot 3 = 12$.

(10pts) 4. You purchase a box of 6 cups. Each has a 0.1 probability of being defective independent of the others. What is the probability that exactly 2 are defective?

Solution: $\binom{6}{2} \cdot \left(\frac{1}{10}\right)^2 \cdot \left(\frac{9}{10}\right)^4$

(10pts) 5. Suppose that out of every 100 emails you receive 1 is spam. Your spam filter can filter spam emails with 90 percent probability. It also lets through legitimate emails to your mailbox with 95 percent probability. What is the probability that a filtered email is actually spam? What is the probability that an email that ends up in your mailbox is not spam?

Solution: Baye's theorem.

$$P(S|F) = \frac{P(F|S)P(S)}{P(F|S)P(S) + P(F|\bar{S})P(\bar{S})} = \frac{0.9 \cdot 0.01}{0.9 \cdot 0.01 + 0.05 \cdot 0.99} = 0.15384615384$$

$$P(\bar{S}|\bar{F}) = \frac{P(\bar{F}|\bar{S})P(\bar{S})}{P(\bar{F}|\bar{S})P(\bar{S}) + P(\bar{F}|S)P(S)} = \frac{0.95 \cdot 0.99}{0.95 \cdot 0.99 + 0.1 \cdot 0.01} = 0.9989378651$$

(10pts) 6. Given n mutually independent Bernoulli trials each with success probability of $1/\sqrt{n}$, what is the expected number of successes?

> **Solution:** Let X be a r.v. that is the sum of the outcomes of the trials. Then, $E[X] = np = n/\sqrt{n} = \sqrt{n}$.

(10pts) 7. Let X be a random variable. If $E[X] = \sqrt{n}$, what is $E[\beta]$ where $\beta = 3X + X/4 + 5$.

> **Solution:** By linearity of expectation, $E[\beta] = 3E[X] + \frac{1}{4}E[X] + 5 = 3.25 \cdot \sqrt{n} + 5$.

(10pts) 8. A file has to be downloaded from a malfunctioning server. At each attempt, the file gets corrupted with probability 0.9. How many attempts are needed in expectation to download the file without the file becoming corrupted?

> **Solution:** This is a geometric random variable with expectation $1/p$ where p in this case is 0.9. Let X be the number of attempts. $E[X] = 1/0.9 = 1.1111...$

(10pts) 9. An unbiased coin is to be flipped 100 times. Let X be a random variable that counts the number of tails. Note that $E[X] = 50$ and $V(X) = 25$. Give an upper bound on the probability that the coin comes up tails at most 25 times.

> **Solution:** Chebyshev: $P(|X - E[X]| > t) \leq V(X)/t^2 = 25/25^2 = 1/25$

(10pts) 10. Suppose that the number of requests a server receives per day is a random variable with an expected value of 1000. Give an upper bound on the probability that the server will receive more than 1100 requests. Hint: lack of knowledge and assumptions limit your options as to what inequalities to use.

> **Solution:** Markov: $Pr[X \geq 1100] \leq 1000/1100 = 10/11$. Why can we use Markov? Because X is positive.

(10pts) 11. Suppose that the number of requests a server receives per day is a random variable with an expected value of 1000. Furthermore, each request is received independently of others. Give an upper bound on the probability that the server will receive more than 1100 requests. Hint: knowing about the independence of the requests will allow you to give a bound much tighter than the similar problem above.

Solution: Chernoff: $Pr[X > (1+0.1)\mu] < \exp(-10^3 \cdot 0.01/4) = \exp(-2.5) = 0.08208499862$. Independence allows us to use Chernoff. Depending on your interpretation of the problem, the requests are either Poisson trials or they are Bernoulli trials – a special case of Poisson trials.

(20pts) 12. A new web service decides to use a program called BinCodes to generate coupon codes for its customers. BinCodes generates 10 digit binary strings as coupon codes where each digit is 0 or 1 uniformly at random. If two users get the same coupon code, one of them cannot use it and must be refunded one dollar for the inconvenience. If BinCodes is used to generate a coupon code for all of the 5000 customers, how many dollars in total are expected to be refunded to users due to such collisions?

Solution: You need to show work but you also need to be rigorous here. Here is a sketch (minimum I should see). There are $2^{10} = 1024$ equally likely codes (bins). How many pairs of customers are there? $\binom{5000}{2}$. What is the probability of a pair colliding in a single bin? $(1/1024)^2$. What is the probability of a pair colliding at any bin? $1024 \cdot (1/1024)^2$.

Now, let $X_{i,j}$ be a random variable where $X_{i,j} = 1$ if customers i and j collide and $X_{i,j} = 0$ otherwise. Let $X = \sum_{i \neq j} X_{i,j}$ be the total number of collisions (hence, total number of dollars to be refunded). Thus, not that $P(X_{i,j} = 1) = 1/1024$. This yields:

$$\begin{aligned}
E[X] &= E\left[\sum_{i\neq j} X_{i,j}\right] && \text{by definition} \\
&= \sum_{i\neq j} E[X_{i,j}] && \text{linearity of expectation} \\
&= \sum_{i\neq j} P(X_{i,j} = 1) \\
&= \sum_{i\neq j} 1/1024 && \text{established above} \\
&= \binom{5000}{2} 1/1024 \\
&= \frac{3124375}{256} = 12204.589...
\end{aligned}$$