This exam covers the material discussed in lecture from Chapters 1,2 and 3. The best way to study is to (listed in order of importance) understand the 4 problems here, the homework problems, the examples presented in lecture, and the extra problems provided from the book.

Topics NOT covered:

- challenge problems from beginning of class
- Benford's law
- Medians
- Moser paper on birthday problem
- Proofs

Topics covered:

1. Counting
   - Permutations and combinations
   - Birthday problem
   - Card questions
2. Independence
   - definition
   - expected value
   - Poisson approximation
   - Unions
     - Inclusion-exclusion
     - Bonferroni
     - Union bound
     - Joe-dimagio
   - Poisson thinning
3. Random variables
   - Discrete r.v.'s
   - Continuous r.v.'s
     - Density functions
     - Distribution functions
     - Functions of random variables
   - Joint distributions

Additional book questions:

Section 1.5 24, 30, 34, 36, 51, 66

Section 2.5 2, 9, 12, 17, 34, 45

Section 3.5 8, 12, 32,

Additional questions:

1. What is the probability a group of 30 people has exactly one pair of birthday twins and no other birthday overlaps?

> **Solution:** We can compute this directly. $C_{30,2}C_{365,1}364 \cdot 363 \cdots (364-28+1)/365^{30}$.

2. 10 baby, 20 adult female, and 30 adult male ducks are flying past Poi(10) photojournalists. Each uniformly selects a single duck to photograph. (a) What is the expected number of ducks photographed? (b) What is the probability that exactly 3 different baby ducks, 4 different female ducks, and 5 different male ducks are photographed? (c) The last question was too not what I intended to ask. What I meant to ask is what is the probability that 3 photos are takes of baby ducks, 4 of females, and 5 of males. (So, duplicate photos are allowed.)

> **Solution:** (a) Let D_i be the number of times duck i is photographed. By thinning we have $D_i =$ Poi(10/60)$_=$ Poi(1/6). Let $X_i = 1$ if $D_i > 0$ and $X_i = 0$ if $D_i = 0$. So $X_i = \text{Ber}(1-e^{-1/6})$. If N is the number of distinct ducks photographed then $N = \sum_1^{60} X_i = \text{Bin}(60, 1-e^{-1/6})$. It is a binomial because the X_i are iid by Poisson thinning. We then have $EN = np = 60(1-e^{-1/6}) \approx 9.2$.
>
> (b) Let B be the number of baby ducks photographed, F the number of female, and M the number of male. Let $p = e^{-1/6}$. We can break N from above up into three different parts so that
>
> $$B = \text{Bin}(10, p), F = \text{Bin}(20, p), M = \text{Bin}(30, p).$$
>
> These are independent because the X_i are iid. So, we get
>
> $$P(B = 3, F = 4, M = 5) = C_{10,3}p^3(1-p)^7 C_{20,4}p^4(1-p)^6 C_{30,5}p^5(1-p)^5.$$
>
> (c) Let $B = \text{Poi}(100/60) = \text{Poi}(5/3), F = \text{Poi}(10/3)$ and $M = \text{Poi}(5)$ be the number of babies, females and males photographed, respectively. By Poisson thinning we have
>
> $$P(B = 3, F = 4, M = 5) = [e^{-5/3}(5/3)^3/3!][e^{-10/3}(10/3)^4/4!][e^{-5}(5)^5/5!].$$

3. A skin cell reproduces by splitting into two new cells. Each of the two cells independent of one another move one unit to the left, or one unit to the right of the parent cell. The 2 new cells also split and the two new cells randomly displace left or right in the same fashion. Cells keep splitting like this indefinitely. Suppose left displacement has probability 1/5 and right has probability 4/5.

   (a) The number of cells in the nth generation is 2^n. Let $A_{i,n}$ be the event that the ith cell of these is at $-n$. What is $P(A_{i,n})$?

   > **Solution:** $(1/5)^n$.

(b) Are the events $A_{i,n}$ independent? Give an intuitive explanation in words with no calculation.

Solution: No. If you know that a cell is at $-n$, it makes it much more likely that, say, the immediate sibling is also at $-n$.

(c) Let E_n be the event that there are any cells from the nth generation at $-n$. Use a union bound to estimate $P(E_n)$. Please simplify your answer so there is no summation. (*Hint: there are* 2^n *cells in the nth generation.*)

Solution: $\sum_{i=1}^{2^n}(1/5)^{-n} = (2/5)^n$.

(d) Use another union bound to estimate the probability that there exists a time $n \geq 1$ with a cell at $-n$. You may leave your answer as a sum.

Solution: $\sum_{n=1}^{\infty}(2/5)^n = 2/3$

4. You are admitting undergraduates to graduate school. There are 300 applicants and you give out 30 acceptances. After doing so, you notice that a randomized list of all 300 applicants has 5 names in a row from the 30 you selected. Use the Bonferoni inequalities to estimate the probability this occurs. Instead of numbers you may use $p = 30 \cdot 29 \cdots 26/300^5 \approx 7 \times P_{300,5}$, $q = 270 \cdot 30 \cdot 29 \cdots 26/P_{300,6} \approx 6 \times 10^{-6}$.

Solution: Let B be the event of having 5 names in a row from your list of 30 and B_i the event that i is not from the list, but $i+1, \dots, i+5$ are. Set B_0 to be the event of $1, 2, 3, 4, 5$ being from your list.

Observe that $B = B_0 \cup \bigcup_1^{295} B_i$. We have $P(B_0) = p$ and $P(B_i) = q$. So $P(B) \leq p + 295q$. For the lower bound we use the bounds $P(B_0 \cap B_i) \leq P(B_0)P(B_i) = pq$ and $P(B_i \cap B_j) \leq P(B_i)P(B_j) = q^2$ to write

$$\sum_1^{295} P(B_0 \cap B_i) + \sum_{1 \leq i < j \leq 295} P(B_i \cap B_j) \leq 295pq + C_{295,2}q^2 \approx 0.$$

By the Bonferoni-inequality we have

$$P(B) \approx p + 295q.$$

Redemption questions:

Section 1.5 24, 30, 36, 51

Section 2.5 9, 12, 34, 45

Section 3.5 8, 12, 32

1. A bacteria colony is multiplying. It starts with a single bacterium. Each bacterium takes Ber(p) minutes with $p = 1/3600$ to split into three (yes, usually it takes 0 minutes). The children then take an independent Ber(p) minutes to split in three and so on.

   A biologist is interested in studying all of the bacteria in G_n, the nth generation of this colony (i.e. the bacteria formed after n splits). So there are 3^n of them, each with an ancestry of n splits back to the original.

   (a) Let $T_1, \dots, T_{3^n}$ be the time it took for each bacterium in G_n to be born starting from the first bacterium. Explain why each T_i is a Bin(n, p) random variable.

Solution: The age is a sum of n independent Bernoulli(p) random variables.

(b) Are the T_i independent? Briefly explain.

Solution: No. For example, siblings have the same lineage back to the original, so these are highly dependent.

(c) Let $E_{i,n,k}$ for $i = 1, 2, \ldots, 3^n$ be the event that the ith bacteria from G_n took k minutes to be born. What is $P(E_{i,n,k})$?

Solution: $P(E_{i,n,k}) = C_{n,k}p^k(1-p)^{n-k}$

(d) Let $E_{n,k}$ be the event that there exists a bacterium from G_n that is k minutes old. Use a union bound to upper bound $P(E_{n,k})$.

Solution: $P(E_{n,k}) \leq \sum_1^{3^n} P(E_{i,n}) = 3^n C_{n,k}p^k(1-p)^{n-k}$.

(e) Use the bound $C_{n,k} \leq 4^n$ and $1 - p \leq 1$ to give an upper bound on $P(E_{n,k})$ that does not depend on k.

Solution: $P(E_{n,k}) \leq 3^n 4^n p^k 1^{n-k} = 12^n p^k$.

(f) Let E_n be the event that there exists a bacterium from G_n that is at least $n/2$ minutes old. Use the previous part and the bound $k \geq n/2$ and $n/2 \leq n$ to show that $P(E_n) \leq n(12\sqrt{p})^n$.

Solution: $P(E_n) \leq \sum_{k=n/2}^{n} 12^n p^k \leq (n/2)12^n p^{n/2} \leq n(12\sqrt{p})^n$

(g) Let E be the event that there exists an n such that a bacterium from G_n is at least $n/2$ minutes old. Write E in terms of the E_n and apply a union bound to upper bound $P(E)$. Simplify by using the formula $\sum_{n=1}^{\infty} na^n = \frac{a}{(1-a)^2}$.

Solution: $P(E) = P(\cup_{n=1}^{\infty} E_n) \leq \sum_{n=1}^{\infty} n(12\sqrt{p})^n = 4/9$.