180A PRACTICE MIDTERM 2

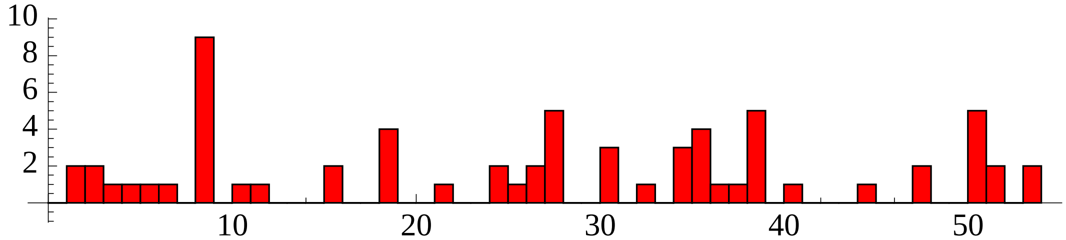
I think these problems are representative, but you may find the actual midterm either more or less difficult than this practice midterm. You may use a calculator.

1. Ecologists want to estimate the number of Northern pike (a non-native invasive fish species) in Lake Davis. They set nets at multiple locations in the lake one day, and catch 1000 Northern pike. Each is tagged and released. The next day they reset the nets and catch 1500 Northern pike, of which 20 have tags from the previous day.
   a. (10 points) If there are N Northern pike in Lake Davis, what is the probability of catching 20 out of 1500 with tags?
   Let $p = 1000/N$. Then $P(20 \text{ tags out of } 1500) = \binom{1500}{20} p^{20}(1-p)^{1500-20}$.
   b. (10 points) How many Northern pike would you estimate there are in Lake Davis?
   Estimate that $20 = 1500p$. Then $N = 75000$.
   c. (5 points) What assumptions are you making in parts (a) and (b)?
   That the fish tagged on the first day are a random sample of the whole population, and that the observed number of recaught fish, 20, is close to the expectation value.

2. Let A_1, A_2, B_1 and B_2 be random variables taking values ± 1. Without knowing anything about the joint distribution of these four random variables:
   a. (10 points) What is the largest and the smallest possible value for $\mathsf{E}[A_1B_1]$?
   The largest and smallest possible values for A_1B_1 are $+1$ and -1, so these also bound the expectation value.
   b. (15 points) What is the largest and the smallest possible value for $\mathsf{E}[A_1B_1 + A_2B_1 + A_1B_2 - A_2B_2]$?
   $A_1B_1+A_2B_1+A_1B_2-A_2B_2 = (A_1+A_2)B_1+(A_1-A_2)B_2$. Exactly one of A_1+A_2 and A_1-A_2 is zero, while the other is ± 2. Thus $-2 \leq \mathsf{E}[A_1B_1+A_2B_1+A_1B_2-A_2B_2] \leq 2$.

3. Let X and Y be random variables taking values in $\{0, 1\}$. Suppose their joint probability distribution is
$$P(X = x, Y = y) = \begin{cases} 1/3 & \text{if } x = y; \\ 1/6 & \text{otherwise.} \end{cases}$$
   a. (10 points) Find the marginal distributions.
   $P(X = x) = \sum_y P(X = x, Y = y) = 1/3 + 1/6 = 1/2 = P(Y = y)$.
   b. (5 points) What is $\mathsf{E}[X]$?
   $\mathsf{E}[X] = \sum_x xP(X = x) = 1/2$.
   c. (10 points) What is the covariance, $\mathsf{Cov}[X, Y]$?
   $\mathsf{Cov}[X,Y] = \mathsf{E}[XY]-\mathsf{E}[X]\mathsf{E}[Y] = \sum_{x,y} xyP(X = x, Y = y)-1/4 = 1/3-1/4 = 1/12$.

4. The graph shows the number of events per day for the first 54 days of this year. There are 67 events.



a. (5 points) Let C be a random variable with probability distribution $P(C = c) =$ (number of days on which there are c events)/54. What is $\mathsf{E}[C]$?

$$P(C = c) = \frac{1}{54}\begin{cases} 25 & c = 0; \\ 13 & c = 1; \\ 8 & c = 2; \\ 2 & c = 3; \\ 2 & c = 4; \\ 3 & c = 5; \\ 1 & c = 9; \\ 0 & \text{otherwise.} \end{cases}$$

So $\mathsf{E}[C] = \sum_c cP(C = c) = (1 \cdot 13 + 2 \cdot 8 + 3 \cdot 2 + 4 \cdot 2 + 5 \cdot 3 + 9 \cdot 1)/54 = 67/54$, which we could also have computed as the number of events divided by the number of days, *i.e.*, the average number of events/day.

b. (10 points) What is $\mathsf{Var}[C]$?

$\mathsf{E}[C] = \sum_c \big(c - \mathsf{E}[C]\big)^2 P(C = c) = \sum_c c^2 P(C = c) - (67/54)^2 = 251/54 - (67/54)^2 \approx 3.109$.

c. (10 points) Do you think these events are a Poisson scatter?

If they were, then when divided into intervals of equal length, *e.g.*, days, the probability distribution for the number of events per interval would be a Poisson distribution. Every Poisson distribution has equal mean and variance, but here $\mathsf{E}[C] = 67/54 \approx 1.241 \ll 3.109 \approx \mathsf{Var}[C]$, so these events are very unlikely to be a Poisson scatter.

d. (5 points, extra credit) What are these events?

These events are, unfortunately, US military casualties in Iraq. See Iraq Coalition Casualty Count, `http://icasualties.org/oif/`.