You are expected to complete this exam during any two-hour window within the exam period, which runs from Saturday 8:30 AM to Sunday 8:30 AM. The exam is open-book and open-note:

- You may use all of this course's resources: the textbook, your class notes, your old homework, and the course web site. You may not share any resources with any other person while you are taking the exam.
- You may cite material (definitions, theorems, examples, etc.) from class, the assigned textbook readings, and the assigned homework problems. You do not have to redevelop or reprove that material. On the other hand, you may not cite results that we have not studied.
- You may not consult any other books, papers, Internet sites, etc. You may use a computer for viewing the course Moodle/web site, running R commands, typing up your answers, and e-mailing with me. If you want to use a computer for other purposes, then check with me first.
- You may not discuss the exam in any way — spoken, written, etc. — with anyone but me, until everyone has handed in the exam.

Feel free to ask clarifying questions in person or over e-mail. You should certainly ask for clarification if you believe that a problem is mis-stated. If you cannot receive clarification, then explain your interpretation in your solution.

Your solutions should be thorough, self-explanatory, neat, concise, and polished. You might want to work first on scratch paper and then recopy your solutions. Alternatively, you might want to type your solutions. Always show enough work and justification so that a typical classmate could understand your solutions. If you cannot solve a problem, then write a brief summary of the approaches you've tried. Partial credit is often awarded. Present your solutions in the order assigned.

Good luck. :)

For the first problem, imagine that you run a health care program. You have n different categories of costs: emergency room visits, prenatal care, cancer screenings, physical therapy, and so on. Let $X_1, \ldots, X_n$ be your costs in the coming year in these n categories. Based on data about previous years, you estimate their means and variances to be $\mu_1, \ldots, \mu_n$ and $\sigma_1^2, \ldots, \sigma_n^2$. Answer the following questions in terms of these quantities. Let $T = X_1 + \cdots + X_n$ be your total cost in the coming year.

A.A. What is the practical value or purpose of knowing $E(T)$? Of knowing $V(T)$?

A.B. Assuming that the X_i are independent, how big might $E(T)$ be?

A.C. Not assuming that the X_i are independent, how big might $E(T)$ be?

A.D. Assuming that the X_i are independent, how big might $V(T)$ be?

A.E. Not assuming that the X_i are independent, how big might $V(T)$ be?

A.F. One of your assistants makes the following argument. "Because we have many categories of costs, T is approximately normal with mean $E(T)$ and variance $V(T)$. This fact can help us simplify a lot of our later calculations about T." What do you think?

For the second problem, let X be a continuous random variable such that its support is contained in $(0, \infty)$. Let $Y = \log X$ (meaning the logarithm base e).

B. Compute the PDF $f_Y(y)$ in terms of $f_X(x)$.

For the third problem, recall that $m_{X+Y}(t) = m_X(t) \cdot m_Y(t)$ if X and Y are independent. Now consider the extreme opposite case, where X and Y are so dependent that in fact $Y = X$.

C. In this extreme opposite case, does $m_{X+Y}(t)$ simplify?

Finally, consider the problem of counting how many earthquakes occur in the USA. Large earthquakes are rare, and they tend to happen in specific locations (e.g., California), while small earthquakes are frequent and more broadly spread. In fact, tiny earthquakes are happening continually all over the USA. So, if we're going to count the number of earthquakes, then we need to first set a threshold M and consider only "large" earthquakes — meaning, of magnitude greater than M. Assume that, for any chosen threshold M, the occurrence of large earthquakes is well modeled as a Poisson process with rate λ (where the unit of time is years). Let X be the number of earthquakes next year.

D.A. As M decreases, how does λ change? (I'm looking for qualitative statements in plain English, rather than precise calculations, which would require detailed data or theory about earthquakes.)

D.B. When M is small, what does the distribution of X look like? Draw or describe its PMF/PDF, and explain its expectation, variance, skewness, and kurtosis. (Hint: There is a hard way to do this problem and an easy way.)