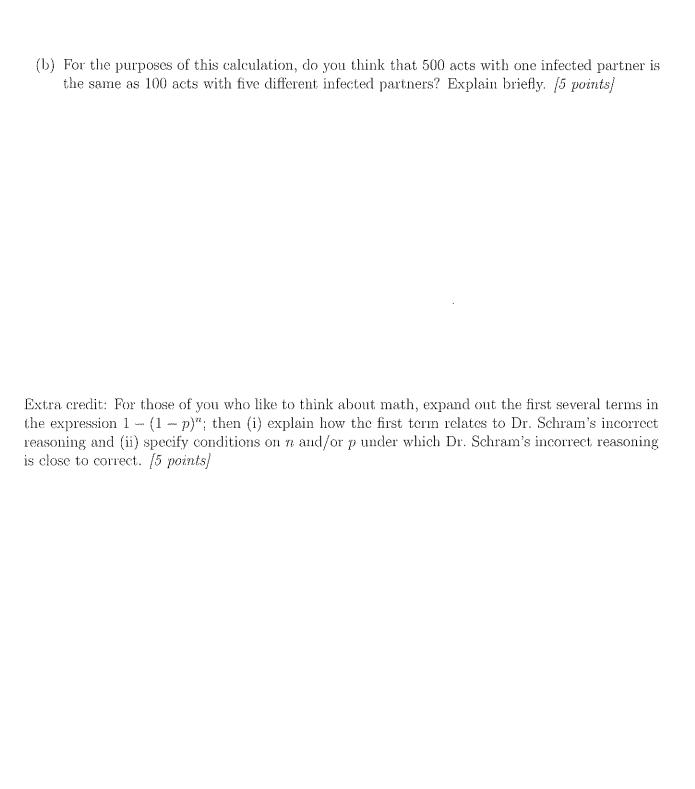2. (medicine) AIDS is one of the largest public health problems facing the United States today. As such it receives a large amount of coverage in the media, unfortunately not all of it accurate. On 24 Aug 1987, the *Los Angeles Times* published a column on its editorial page by Dr. Neil Schram, chairman of the Los Angeles City/County AIDS Task Force, that contained the following passage:

> Several studies of sexual partners of people infected with the virus show that a single act of unprotected vaginal intercourse has a surprisingly low risk of infecting the uninfected partner — perhaps one in 100 to one in 1,000. For an average, consider the risk to be one in 500. If there are 100 acts of intercourse with an infected partner, the chance of infection increases to one in five. Statistically, 500 acts of intercourse with one infected partner or 100 acts with five different infected partners lead to a 100% probability of infection (statistically, not necessarily in reality).

Evidently Dr. Schram would say that 600 acts would lead to a 120% probability of infection (statistically speaking, of course :-).

(a) Work out the right answers to his probability estimates (that is, what really is the chance of infection in 100 and then 500 acts, assuming his $\frac{1}{500}$ single-episode risk figure?), being explicit about all assumptions you make in your calculations and commenting on their plausibility. *[15 points]*

(b) For the purposes of this calculation, do you think that 500 acts with one infected partner is the same as 100 acts with five different infected partners? Explain briefly. *[5 points]*

Extra credit: For those of you who like to think about math, expand out the first several terms in the expression $1 - (1 - p)^n$; then (i) explain how the first term relates to Dr. Schram's incorrect reasoning and (ii) specify conditions on n and/or p under which Dr. Schram's incorrect reasoning is close to correct. *[5 points]*

(b) In the game of craps, somebody rolls a pair of dice, and that person plus a bunch of people standing around watching bet on the sum of the faces pointing up when the dice come to rest. Each die has six faces, which (when the dice are thrown fairly) come up equally often. Two outcomes of particular interest are rolling a three (that is one way to lose, or "crap out") and rolling a seven, which at some points in the game is one way to win. What is the chance of getting a three on any single roll of the dice? How about a seven?

6. It is now generally accepted that cigarette smoking causes heart disease, lung cancer, and many other diseases. However, in the 1950s, this idea was controversial. Epidemiological studies in the late 1950s established a strong association between smoking and ill-health, but association is of course not the same thing as causation. The statistician and geneticist R. A. Fisher advanced the "constitutional hypothesis": there is some genetic factor that disposes you both to smoke and to die.

To test Fisher's idea, epidemiologists used twin studies. They identified sets of smoking-discordant monozygotic twin pairs ("monozygotic" twins come from one egg and have identical genetic makeup; "smoking-discordant" means that one twin in the pair smokes and the other doesn't). This sets up a race: which twin dies first — the smoker or the non-smoker? Data from a twin study in Finland were as follows:

Cause of Death	Smokers	Non-smokers
Coronary heart disease	9	0
Lung cancer	2	0

In other words, there were 9 twin pairs where at least one twin died of coronary heart disease, and in all 9 the smoker died first (and similarly for lung cancer, which is a rare disease, even among smokers).

Under Fisher's constitutional hypothesis H, each twin in any given pair is equally likely to die first, so that the number of pairs in which the smoker dies first is like the number of heads in IID tossing of a fair (50/50) coin.

(a) Assuming H, what is the chance of both smokers dying first of lung cancer? Show your work.

(b) Repeat for the 9 deaths from coronary heart disease.

(c) On the basis of these data, can the difference in the rate of first dying for smoking and non-smoking twins be explained by

   (i) chance?
   (ii) genetics?
   (iii) the adverse health effects of smoking?

Explain briefly.

R-42