Questions and Comments from Students:

1. Want more exercises from book, with solutions.
2. Want more Power-Point summaries, like the one for Ch 4-5.
3. IQR – uses fourths or quartiles?
4. Don't know when to use conditional probability.
5. What is covariance?
6. What is the connection between strong and weak relationship and the value of the correlation coefficient?
7. What population assumptions are needed to compute a probability about a sample mean?
8. How is standard error different from standard deviation?
9. How can I learn to apply the techniques of this course if the text does not provide the whole context for the examples it describes?
10. What is it about the Poisson Process that would work for a short duration during the day but not for the whole day (as in traffic for example, or calls to a help line)?
11. p 74 Ex 43 (prob of 6,7,8,9,10 in a 5 card hand) . Why 4^5 ?
12. Expl, chi-squared, and gamma – what is the connection?
13. More examples of exponential, gamma, and use of pdf, cdf.
14. Material of Ch 5 covered too fast.
15. Connections between gamma, exponential, neg. binomial and geometric.
16. Uniform vs binomial – confusing.
17. In the chapter 5 sampling theory notes you mentioned on page 2 that " samples of 99 mean and 101 women are impossible" I think what you meant was samples of 99 men and women?
18. Want more examples like the investment examples that are useful for life.
19. Waiting time for TA is too long.
20. What is the expected value of a uniform distribution not really "expected"?
21. Clarify the definition and characteristics of the uniform distribution.
22. Material is dry and difficult, want more exercises solved in class.
23. What is the general definition of variance?
24. What are quartiles and how do you calculate IQR?
25. What does covariance calculate? Same for correlation?
26. Provide exercise examples for Ch 5 and 6.
27. How do we know what values to use for alpha and beta in the gamma model?
28. Is alpha in gamma analogous to k in neg Binomial? What about beta in gamma?
29. Other than double integrals, what sort of questions relating to joint probabilities of continuous variables, might we see on the midterm?

Answers:

1. There are lots of exercises with answers in the book. If you cannot get an answer, or get the wrong answer, you need to speak to me or the TA, not simply read the solution. Since the exercises usually use the techniques of the chapter, if you cannot solve you are missing some understanding, and reading the solution often will not help. Of course, as a method of learning the material the first time, reading solutions to exercises can help – that is why I have provided many of these in the notes. The notes are important, and solving exercises is only a part of the course. You need to understand the ideas, not just learn how to solve exercises.

2. I will do this for future chapters 7-9.

3. The box on p 41 about fourth spread is just non-standard terminology for quartiles and the interquartile range (IQR). IQR = Q3-Q1. It is probably simpler just to use the quartile terminology – ignore the "fourths".

4. In mathematics, you have if-then proofs. In statistics, you have if-then probability calculations. P(statement about outcome | assumption about outcome) is the format of conditional probability. If you have P(statement) it is not conditional!

5. Covariance is just a term for the numerator of the correlation coefficient. Don't worry about what it means, but you should think about when it is positive and when it is negative. If you take more stats courses, you will run into it again and it will play a larger role.

6. The correlation coefficient measures the strength of the linear association between two variables. If it is near ± 1, it is a strong association (meaning knowledge of one variable would allow you to predict the other one). If it is near 0, it is a weak relationship.

7. To compute P(sample mean < 10) for example, for a sample that is fairly large (say n>30) all you need is the sample itself – you don't need any other info about the population (such as mean, sd or distribution shape). From the sample, you compute the mean and sd and since the CLT says the distribution of the sample mean is approx normal, and since all you need to compute probabilities for a normal distribution is its mean and sd, you can then compute the P(sample mean < 10) or whatever. Actually, you only estimate P(sample mean < 10) since you are using estimated pop mean and sd, but it usually is good enough.

8. Standard error is the sd of the distribution of a statistic (ie, of something computed from sample data). For example, if the sd of a sample is 5, and the sample size is 25, the standard error of the sample mean is 5/sqrt(25) = 1. A typical deviation from the mean of the sample mean is 1 in this case, whereas a typical deviation from the mean of the sample data itself was 5. If you are still confused, throw out the jargon "standard error" and only talk about standard deviations. The point is that sd of X is different than sd of

$\overline{X}$. The latter is smaller (and so commonly used that some people thought it would help to give it a different name).

9. Come to lectures and read the lecture notes. And pay special attention to the detailed examples discussed in class (like the investment examples).

10. The Poisson Process is the most common application of the Poisson distribution. The assumptions of it are described on p 137. One assumption is that the rate at which events occur must be constant (in the sense that, in any small interval Δt, the probability of an event is proportional to Δt, and the constant of proportionality must not change over time.) So the Poisson model only works when the rate of events can be assumed to be constant over time. If we are talking about something like traffic, in which the rate changes a lot during the day, the Poisson would not work because the λ actually depends on t. But for a short time period, the assumption of constant rate may be good enough.

11. P(6,7,8,9,10) = ? This is one of those questions where each 5-card poker hand has the same probability ($=1/C_{5,52}$) so the prob calculation requires a count of the number of ways the event can occur. How many of the $C_{5,52}$ hands have 6,7,8,9,10 in them (any order)? There are 4 6s, 4, 7s, etc and so 4^5 is the total number of hands like this. (m ways to do one thing, n ways to do another thing, then there are mn ways to do both things).

12. The exponential and chi-squared are special cases of the gamma. This can be seen analytically by comparing the density functions. In addition, there is a model connection between the gamma and the exponential – seen via the Poisson process. The exponential is the time between successive events, and the gamma is the time until the kth event. The reason the Poisson is called the Poisson is another connection – the number of events that have occurred is a specific time has a Poisson distribution.

13. The exponential has a pdf and cdf that can be written explicitly and computed easily. The gamma has a pdf but the cdf requires a table to look up. Of course, you can't look up the cdf for the gamma unless you know the parameters (so you can enter the table). The details are given in the notes.

14. Ch 5 is always confusing to students. We need to repeat this several times in the course. Going slower is not as helpful as you might think. Nevertheless, we are spending two lectures this week when we will go over the material (Today and Wed).

15. These connections are outlined in the notes: I'll copy the notes here:

The title of these notes was: " The Gamma Distribution (and connections with Exponential, Geometric, Negative Binomial, Chi-Squared, and Normal Distributions)." Now we outline these connections.

Recall that the Geometric distribution was a natural model for the number of Bernoulli trials that were failures before the first success occurred. Now to find the number of Bernoulli trials that were failures until the rth success occurred, we really have to wait for r Geometrics. That is, if Y is Negative Binomial with parameters r and p, and X is Geometric with parameter p, then

$$Y = \sum_{i=1}^{r} X_i \qquad (*)$$

The sum of r Geometrics (with the same parameter p) is Negative Binomial.

Now recall that the **Geometric is the discrete analogue of the Exponential distribution**. It turns out that (*) is also true if the X_i are Exponential with parameter λ and Y is Gamma with parameters r and λ^{-1}. (The parametrization of the exponential on p 177 of your book defines λ in such a way that the mean is λ^{-1}). Look at the mean and SD of the Gamma. These can also be written
mean = r λ^{-1} and SD = $r^{1/2}\lambda^{-1}$
in the parameterization that relates best to the exponential.

So we have a sort of box of links:

Geometric	**Exponential**
Neg Binomial	**Gamma**

16. Binomial vs Uniform The binomial models the number of heads in n tosses of a coin. The uniform on (0,1) models the number between 0 and 1 selected in such a way that every possible number has the same chance of occurring. (That is the continuous uniform, which is the usual one referred to by "uniform distribution". The discrete uniform on 0,1,2,..9 is the distribution giving equal probability to each of 0,1,2,...9.)
I can't see the source of confusion!

17. Yes, than "mean" should be "men".

18. I agree 100% but I am constrained by the requirements of the course contents that I must cover. However, now that we have covered most of the basics, we can spend more time on those examples I outlined the first day – and thanks for your suggestions about gas consumption ...

19. Please feel free to come to my office hours or any time after class. Even before class outside of my office hours is OK occasionally.

20. The expected value of the uniform RV X that is defined on (0,1), in other words E(X), is ½. However, the probability P(X=1/2)=0, so it can hardly be said that the value X=1/2 is "expected" in the ordinary sense of the word. Even for the discrete uniform distribution on 0,1,2,...9, E(X)=4.5, and yet 4.5 is an impossible value for X, and again, is certainly not "expected" in the ordinary sense. But this is not just a problem with the uniform distributions – it is rare that the E(X) value would be a common value. The expected size of a household, E(H), in Canada might be 3.3, but one would not expect to find 3.3 individuals in a household!

21. There are two classes of uniform distribution:
i) the continuous class $X \sim U(a,b)$
and
ii) the discrete class $Y \sim U(a,b)$

X here takes values anywhere in the interval (a,b) and nowhere else. It takes every one of the values in (a,b) with equal density (relative probability).

Y here takes values on any integer in {a,a+1,a+2, ..., b}. Every one of these values has probability (1/(b-a+1)). Think of the case a=0,b=9 to understand this.

The mean and variance of X is $(a+b)/2$, $((b-a)^2)/12$.
The mean and variance of Y is $(a+b)/2$, $((b-a+1)^2-1)/12$.
The pdf of the continuous uniform is $f_X(x) = 1$ on (0,1) and =0 otherwise.
The cdf of X is $F_X(x) = x$ on (0,1)

The probability law of Y is $P(Y=y) = 1/(b-a+1)$ for y = a, a+1, a+2, ...,b
The cdf of Y is $F_Y(y)$ = (the number of values in a, a+1, ..., b less than y)/(b-a+1) for y in (a,b).

If you put a=1, b=n, the info about Y will seem simpler.

22. The dry aspect is partly unavoidable – this course is the only required course in stats in some programs and we need to cover the basics. But we will be doing more applications like the investment scenario, and I hope this helps with motivation. The "difficult" is not surprising but avoidable. Students need to understand that the course is not about solving numerical problems – but rather about understanding the basic strategies of mathematical statistics. Solving more exercises is useful but not nearly enough since it does not force to understand what you are doing with the calulations. Most students look for a similar problem and copy the numerical procedure. This gets the right answer but does not encourage understanding. That is why I keep harping on
i) attending lectures and reading through the lecture notes
ii) resolving confusion by talking to someone who understand the material

The material is not difficult if you consistently review and clarify material after every lecture. The material is cumulative and a big push at the end will not work.

23. Variance of $X = E(X-E(X))^2$. This should be more than a formula to you. X-E(X) is the deviation from the mean. So variance is the average value of the squared deviation.

24. See 3. above.

25. See 5. and 6. above.

26. Exercises for Ch 5 and 6 – there are lots of them with answers in the book. Also, for Ch 5 I posted several exercises Feb 19, 21, and 23. Morevoer, the text solves 26 more exercises in detail from sections 5.1-5.4. In Ch 6 I did not emphasize the calculations and doing these exercises would not be a useful strategy for this chapter. Read the notes – ask questions if necessary.

There is a perception that if you just learn the calculation procedures for the common questions, you will be able to do well on the midterm and the final exam. But the quiz should have been a wake-up call! I require understanding of the concepts, and the tests will try to focus on that and in so doing, encourage the kind the learning that will be useful to you (as opposed to memorizing a few calculation procedures that will be useless in practice.)

Read the notes! Ask questions!

27-28. In the notes for Ch 4 section 4, I show how to compute alpha and beta from the mean and sd. In most applications, reasonable values for the mean and sd can be specified, and so you can specify alpha and beta this way. Alpha in the gamma is the number of events that occur to make the gamma time duration have this gamma distribution – yes it is analogous to k in the neg binomial, where k is the number of successes to occur before the number of trials will have that neg binomial distribution. The beta in the gamma model is a scaling factor for time. The analogue in the neg binomial would be the rate at which we observe trial outcomes, but since we do not usually vary this in the neg binomial model, there is no parameter for it. We observe every trial in the neg binomial, not every second or every third trial ...

29. You could be asked what the double integral represents (a volume calculation where volume is probability), you could be asked to compute a joint density in terms of a marginal and conditional density, you could be asked to compute P(A and B) where events A and A|B have probabilities that are known or could be looked up in a table, you could be asked for E(Y|X=a) which is a single integral, or the variance of Y|X, to explain what the graph looks like for the joint density of X and Y, to explain how, in graphical terms, the joint density relates to the marginal density or conditional density, ...

Basically, you need to have a way of verbalizing the meaning of joint, conditional and marginal probabilities as well as representing them and their relationships mathematically.