ISTA 410/510 Homework II

For contribution to the final grade, due dates, current late policy, and instructions for handing the assignment in, see the assignment web page.

Create a PDF document with your answers and/or the results of any programs that you write. You should also hand in your programs, but I won't necessarily look at them. Your PDF should be named <first_name>-<last_name>-<assignment>.pdf (e.g., kobus-barnard-hw2.pdf).

You should explain, in your PDF, where the results come from (e.g., "These plots are the result of running the program hw1_part2.m with parameters 1,2,4, and 8 respectively."). Please use this course as an opportunity to learn how to write better figure captions. They should tell the reader how to interpret the figures, and the answer to obvious questions the reader might have which are not readily available from the figure.

For simplicity, problems are generally all worth the same, except ones marked by "+" that are expected to substantively more time consuming, and are worth double. Additional "+" means what you expect.

Questions marked by * are required for grad students only. They count as challenge problems for undergraduates.

Questions marked by ** are challenge problems for both grads and undergraduates.

Any non-challenge problem can be replaced by challenge problems with collective value is at least that of the problem being replace (e.g., an undergraduate might replace a non starred "+" problem with two "*" problems without "+"). Please make it clear that this is what you are doing (e.g., for a required problem you could answer "see optional problem #3"). The point here is to enable students to avoid problems that they feel are not instructive.

For a complete assignment, undergraduates need to hand in problems that have total value of at least 9. (There are 8 problems without "*", one of which has a "+"). Grad students need four more.

Extra problems (please indicated in your answer when you are doing an extra problem) are eligible for modest extra credit. The maximum score for an assignment will be capped at 120%. The maximum score for all assignments taken together is capped at 65/60.

Hints or answers to many of these problems can be looked up. If you are stuck and make use of a resource, simply make a note of it. For example, you might say that you had a glance at the solution to the same or similar problem solution in a particular source, and then attempted to recreate for yourself. This is better than being completely stuck, or copying the answer blindly.

Mathematical content. This is a mathematical subject and there is a wide variance in backgrounds of students who take this course. For example, there may be problems in the assignments which seem more difficult than they really are simply because you are not used to the kind of problem. In general, I am very willing to give hints, consider other work in exchange, and grade holistically, focused on effort and progress from whatever level you are at. However, this works best if you start the assignment early, and work through it steadily over time, rather than do the last minute thing.

1. Consider throwing a 6 sided die two times in succession. Let A be the number of the first throw and B be the number of the second throw. Let S be the random variable that is the sum of the numbers.What is:
   a) P(A=3, B=3 | S=6)
   b) P(S=6 | A=3, B=3)
   c) P(A=1, B=1 | S=6)
   d) P(A=1|S=2)
   e) P(A=3|S=2)
   f) P(A=2|S=6)
   g) P (S=12)
   h) P(S=6)

2. (From Bishop. In Bishop this problem is a double star one, but I don't think it is so bad, and it has no stars for us.)

Table 8.2 The joint distribution over three binary variables.

a	b	c	$p(a,b,c) \times 1000$
0	0	0	192
0	0	1	144
0	1	0	48
0	1	1	216
1	0	0	192
1	0	1	64
1	1	0	48
1	1	1	96

8.3 ($\star\star$) Consider three binary variables $a, b, c \in \{0, 1\}$ having the joint distribution given in Table 8.2. Show by direct evaluation that this distribution has the property that a and b are marginally dependent, so that $p(a, b) \neq p(a)p(b)$, but that they become independent when conditioned on c, so that $p(a, b|c) = p(a|c)p(b|c)$ for both $c = 0$ and $c = 1$.

3. (*) Consider these propositions for conditional independence:

   (A) P(X|Y,Z)=P(X|Z) or P(Y,Z)=0.

   (B) P(Y|X,Z)=P(Y|Z) or P(X,Z)=0

   (C) P(X,Y|Z)=P(X|Z)P(Y|Z)

   Show that (A) implies (B); (A) implies (C), and (C) implies (A). (By symmetry, the third case means that (C) also implies (B), so no need to show that).

(If you like, simply pretend that probabilities are never zero, rather than fretting about the special cases of P(X,Z) or P(Y,Z) being zero. I am not looking for a tight proof. The point of this problem is to get used to manipulating expressions using basic probability definitions.)

4. Using Matlab, plot three univariate Gaussians on one graph. Specifically, plot Gaussians with means and variances (2, 0.2), (1, 0.5), and (0, 2).

5. Using Matlab, make a 3D plot of p(x,y) for a bivariate Gaussian with mean (0,0) and covariance matrix.

$$\begin{vmatrix} 0.5 & 0.3 \\ 0.3 & 2.0 \end{vmatrix}$$

4. For the bivariate Gaussian above, using numerical integration, approximate p(x) for a sensible stepping of x values and plot the result. Does it have the shape you expect, and what is that shape?

5. (**) Perhaps proof by programming is not for you. Show that for a bivariate Gaussian p(x,y), p(x) is also Gaussian. What do you expect the area of the curve to be? It may be easier to work with the precision matrix which is the inverse of the covariance matrix and which is also symmetric.

6. For the bivariate Gaussian above, using the result of the previous question, plot p(y|X=0.5). Does it have the shape you expect, and what is that shape? What do you expect the area of the curve to be?

7. (**) Perhaps proof by programming is not for you. Show that for a bivariate Gaussian p(x,y), p(x|y) is also Gaussian. It may be easier to work with the precision matrix which is the inverse of the covariance matrix and which is also symmetric.

8. (**) Can you transform (x,y) so that the the above Gaussian has diagonal covariance in a different space? If so, do so.

9. Recall that, for a particular form of the likelihood, a conjugate prior is one where the posterior distribution is the same kind of function, just with different parameters. Fill in a few steps to show that the Beta distribution is a conjugate prior of the Binomial distribution. Make sure you specify the new parameters of the distribution for the posterior.

10. (*) Remind yourself about the Poisson distribution. Show that a conjugate prior for it is the Gamma distribution.

11. (**) Most familiar distributions are in the exponential family, which means that they have the form:

$$p(\mathbf{x} \mid \boldsymbol{\eta}) = h(\mathbf{x}) g(\boldsymbol{\eta}) \exp\{\boldsymbol{\eta}^T \mathbf{u}(\mathbf{x})\}$$

where x may be scalar or vector, discrete or continuous. The parameters η are called the natural parameters of the distribution. $g(\eta)$ is the the normalizing constant.

a. Show that the multinomial, Gaussian, gamma, and Poisson distributions are part of this family.

b. Suggest a generalized conjugate prior for this family, and show that it has the conjugacy property.

12. (+) Implement the method covered in class for sampling a generic distribution. Use this to sample a univariate Gaussian distribution with mean 0 and variance 1. Plot histograms normalized to unit area of 10, 100, 1000, and 10000 samples. Plot the source distribution (the univariate Gaussian) on top of the histograms.

13. (*) Using your code from the previous problem as a starting point, do the same for the Beta distributions with alpha=2 and beta=5.

14. (*) Consider the simplest D dimension Gaussians that have mean vector **0** and variance vector **1**. Derive an expression, $p(r)$, for the probability mass at a radius r. In other words, $p(r)\Delta r$ is the probability that a sampled point is between r and $r+\Delta r$, where, for a D dimensional point, **X**, $r^2 = \mathbf{X}^{\mathbf{T}}\mathbf{X}$. Plot p(r) for D=1, D=2, D=10, D=20, and D=30. Comment on the result.

15. (**) The vision lab software has a straightforward C implementation of the Box-Muller algorithm for sampling simple Gaussian distributions. According to Wikipedia, the Ziggurat algorithm is faster. Investigate this conjecture in some way.