B555: Homework 2

1. Suppose that X is a discrete n-dimensional random vector

$$X = \begin{pmatrix} X_1 \\ X_1 \\ \vdots \\ X_n \end{pmatrix}$$

and that $a \in \Re^n$ (a is a constant n-dimensional vector). Show $E(a^t X) = a^t E(X)$ where a^t denotes the transpose of a. (The result also holds for continuous random vectors).

2. Let X, Y be independent continuous random variables with density $f(x, y)$, where $f(x, y) = g(x)h(y)$ for some functions g, h. Show that X and Y are independent.

3. For this problem, recall that random variables X, Y are conditionally independent given Z if X and Y are independent under the conditional distribution $f(x, y|z)$ (the distribution of X, Y having observed $Z = z$.

   Suppose that we consider random minutes from randomly chosen drivers to determine three binary random variables:

$$\begin{aligned} H &= \text{the driver is within 1 mile of home} \\ M &= \text{the driver is on a main road} \\ A &= \text{the driver has had an accident in the sampled minute} \end{aligned}$$

   Suppose that $P(H = 1) = .8$ along with $P(M|H)$ and $P(A = 1|, H, M)$ in the following two tables.

$P(M\|H)$	$M = 0$	$M = 1$
$H = 0$	.1	.9
$H = 1$	.6	.4

$P(A = 1\|H, M)$	$M = 0$	$M = 1$
$H = 0$	.0003	.0008
$H = 1$	.0003	.0008

   (a) Compute $P(A = 1|H = 0)$ and $P(A = 1|H = 1)$ Are accidents more likely near home?

   (b) Compute $P(A = 1|M = 0)$ and $P(A = 1|M = 1)$. What conditional independence relation can you conclude?

   (c) Give a cause and effect hypothesis that is consistent with your conditional indepedence statement.

   (d) Suppose that a driver is driving on a stretch of highway (main road) on her way home. As she crosses the 1-mile-to-home road marker does she encounter an elevated probability of accident?

4. Simulation of Binomial

   (a) A random variable has $X \sim \text{Unif}(0, 1)$ when its density function is

$$f(x) = \begin{cases} 1 & 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

   The R command "runif(n)" creates a vector of n independent $\text{Unif}(0, 1)$ outcomes. Use this function to simulate a sequence of 10,000 independent $\text{Binomial}(500, 1/2)$ random variables and plot the resulting histogram. You can plot a histogram of a data vector x with "hist(x)".

   (b) Letting your sequence of Binomial variables be $X_1, \ldots, X_n$, "standardize" these variables by taking $Y_n = (X_n - \mu)/\sigma$ where μ and σ are the mean and standard deviation of the variables. What are the mean and variance of the Y_n variables?

(c) Now simulate 10,000 *standardized* Binomials and plot the empirical cumulative distribution function $\hat{F}(x)$ where

$$\hat{F}(x) = \hat{P}(X \leq x) = \frac{|\{i : X_i \leq x\}|}{n}$$

Use "plot(x,type='l')" to plot with a solid line

(d) On the same plot show the cumulative distribution function of the standard normal. "pnorm(x)" gives the probability that a $N(0,1)$ variable is less than x. You can add to an existing plot with "lines(x)." Your two plots should look very similar. If X is a Binomial with large n, the fact that

$$\frac{X - np}{\sqrt{np(1-p)}} \overset{\text{approx}}{\sim} N(0,1)$$

is known as the normal approximation to the Binomial.

5. It is well-known that independent random variables are also uncorrelated.

Consider the variables $X = (X_1, X_2)$ having probability density function

$$p(x) = \begin{cases} \frac{1}{\pi} & x_1^2 + x_2^2 < 1 \\ 0 & \text{otherwise} \end{cases}$$

Show that X_1 and X_2 are uncorrelated but not independent.

6. Suppose that $X = (X_1, X_2) \sim N(\mu, \Sigma)$ where

$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}$$

(a) Show that in this case X_1, X_2 uncorrelated implies they are also independent. It might help to know that the marginal distribution of X_1 is normal with mean μ_1 and variance $\sigma_1^2 = \sigma_{11}$ with a simliar statement for X_2.

(b) Suppose, $X = (X_1, X_2, \ldots, X_p) \sim N(\mu, \Sigma)$. Show that if $\Sigma_{ij}^{-1} = 0$ then X_i and X_j are conditionally independent given *all* the other variables. (One can also show the stated conditional independence assumption implies $\Sigma_{ij}^{-1} = 0$ for normal random variables).

(c) Suppose $X_1, X_2, \ldots, X_n$ defined by $X_1 = Z_1$ and $X_i = X_{i-1} + Z_i$ for $i = 2, \ldots, n$ where $Z_1, \ldots, Z_n \overset{\text{iid}}{\sim} N(0,1)$. This is known as a random walk.

i. Show that $(X_1, \ldots, X_n) \sim N(0, \Sigma)$ for some covariance matrix Σ.
ii. Without computing argue which elements of Σ^{-1} are 0.

7. Suppose that $X \sim \text{Unif}(0,1)$.

(a) Show that the cdf of $Y = \frac{-\log(X)}{\lambda} = 1 - e^{-\lambda y}$ and conclude that $Y \sim \text{Exponential}(\lambda)$.

(b) Suppose that $X_1, X_2 \ldots \overset{\text{iid}}{\sim} \text{Exponential}(\lambda)$. Interpret X_1 as the time we need to wait for the first event, X_2 as the *additional* time we wait for the 2nd event, X_3 as the additional time for the 3rd event and so on. In R take $\lambda = 3$ and simulate 100,000 times the number of events that occur in the interval $[0,1]$. Compute the empirical pmf, $\hat{p}(n)$ for $n = 0, 1, \ldots$ by

$$\hat{p}(n) = \frac{\#\text{ times experiments gives } n}{100000}$$

On the same plot compare these probabilities with the Poisson(λ) distribution. The process $N(t)$ that counts the number of events that occur before time t is known as a Poisson process.

8. Let M be a discrete variable with probability mass function $p(m)$. Suppose X is a continuous random vector that *depends* on M. Specifically, assume $X|M = m$ is $N(\mu_m, \Sigma_m)$.

(a) Even though M is discrete and X is continuous we can still get the joint distribution of M, X as $p(m, x) = p(m)p(x|m)$. In this problem we'll use the term density for dicrete, continuous and mixed cases. Compute

i. The joint density $p(m, x)$

ii. The marginal density $p(x)$.

iii. The conditional density $p(m|x)$

(b) Suppose that a population is composed of C classes $1, \ldots, C$. let $p(i)$ be the proportion of the population in class i so $p(1) + \ldots + p(C) = 1$. Suppose we take several measurments, X, on the population and that these measurements behave differently under the classes. We model the "class conditional" distribution of X as multivariate normal with mean μ_i and covariance Σ_i.

i. If we observe $X = x$ give the conditional probabilities $P(M = m|X = x)$.

ii. Suppose that you observe $X = x$ and want to choose the value of m, $\hat{M}$, that has the best chance of being right. What should $\hat{M}$ be?