Econ. 240A, Spring 2000 D. McFadden

PROBLEM SET 2 (Properties of Special Distributions)

(Due Monday, Feb. 6)

1. Suppose that the duration of a spell of unemployment (in days) can be described by a geometric distribution, $\text{Prob}(k) = p^k(1-p)$, where $0 < p < 1$ is a parameter and k is a non-negative integer. What is the expected duration of unemployment? What is the probability of a spell of unemployment lasting longer than K days? What is the conditional expectation of the duration of unemployment, given the event that $K > m$, where m is a positive integer? [Hint: Use formulas for geometric series, see 2.1.10.]

2. Use the moment generating function to find $\mathbf{E}X^3$ when X has density $e^{-x/\lambda}/\lambda$, $x > 0$.

3. A log normal random variable Y is one that has log(Y) normal. If log(Y) has mean μ and variance σ^2, find the mean and variance of Y. [Hint: It is useful to find the moment generating function of $Z = \log(Y)$.]

4. If X and Y are independent normal, then X+Y is again normal, so that one can say that *the normal family is closed under addition*. (Addition of random variables is also called convolution, from the formula for the density of the sum.) Now suppose X and Y are independent and have extreme value distributions, $\text{Prob}(X \le x) = \exp(-e^{a-x})$ and $\text{Prob}(Y \le y) = \exp(-e^{b-y})$, where a and b are location parameters. Show that max(X,Y) once again has an extreme value distribution (with location parameter $c = \log(e^a+e^b)$), so that *the extreme value family is closed under maximization*.

5. If X is standard normal, derive the density and characteristic function of $Y = X^2$, and confirm that this is the same as the tabled density of a chi-square random variable with one degree of freedom. If X is normal with variance one and a mean μ that is not zero, derive the density of Y, which is non-central chi-square distributed with one degree of freedom and noncentrality parameter μ^2.

OVER!

6. The data set nyse.txt in the class data area of the class home page contains daily observations on stock market returns from Jan. 2, 1968 through Dec. 31, 1998, a total of 7806 observations corresponding to days the market was open. There are four variables, in columns delimited by spaces. The first variable (DAT) is the date in *yymmdd* format, the second variable (RNYSE) is the daily return to the NYSE market index, defined as the log of the ratio of the closing value of the index today to the closing index on the previous day the market was open, with distributions (dividends) factored in. The third variable (SP500) is the S&P500 market index, an index of a majority of the high market value stocks in the New York stock exchange. The fourth variable (RTB90) is the rate of interest in the secondary market for 90-day Treasury Bills, converted to a daily rate commensurate with RNYSE..

a. Let $\mathbf{E}_n$ denote a sample average (empirical expectation). Find the sample mean $\mu = \mathbf{E}_n X$, variance $\sigma^2 = \mathbf{E}_n(X - \mu)$, skewness $\mathbf{E}_n(X - \mu)^3/\sigma^3$, and kurtosis $\mathbf{E}_n(X - \mu)^4/\sigma^4 - 3$, for the variables RNYSE and RTB90. Normally distributed random variables have zero skewness and kurtosis in the population. Making an "eyeball" comparison, do the sample moments appear to be consistent with the proposition that RNYSE and RTB90 are normally distributed?

b. For RNYSE, form the *standardized* variable $Z = (\text{RNYSE} - \mu)/\sigma$, by subtracting this variable's sample mean and then dividing by the square root of its variance (or standard deviation). Sort the values of Z from low to high, and then construct a new variable Y that equals i/7806 for $1 \leq i \leq 7806$. The values of Z are called the *order statistics* of the sample, and Y is the empirical CDF, a CDF that puts 1/7806 probability at each observed value of RNYSE. Plot Y against $\Phi(Z)$, where Φ is the standard normal CDF. If RNYSE is normal, then these curves will differ only because of sampling noise in Y. Does it appear by eyeball comparison that they are likely to be the same? A particular issue is the theoretical question of whether the distribution of returns has fat tails, so that the variance and higher moments are hard to estimate precisely or may fail to exist. In a normal sample, one would expect that on average 99 percent of standardized observations are less than 2.575 in magnitude. Do the standardized values Z appear to be consistent with this frequency?

c. A claim in the analysis of stock market returns is that the introduction of financial derivatives and index funds through the 1980's made it easier for arbitragers to close windows of profit opportunity. The argument is made that the resulting actions of arbitragers have made the market more volatile. Compare the subsamples of NYSE excess returns (EXCESS = RNYSE - NRTB90) for the periods 1968-1978 and 1988-1998. By eyeball comparison, were there differences in mean excess return in the two decades? In the variance (or standard deviation) of excess return? Now do a 2×2 table of sample means classified by the two decades above and by whether or not the previous day's excess return was above its decade average. Does it appear that the gap between mean excess returns on days following previous rises and falls has increased or shrunk in the decade of the 90's?