Econ 270 Homework 3

Expectation

The following pmf describes the distribution of weights of fresh black truffles in grams. Calculate the expected cost of a randomly selected black truffle if truffles cost \$100 per ounce (there are 28 grams in an ounce)

$p(14) = .1, p(21) = .15, p(28) = .5, p(35) = .15, p(42) = .1$

We can either calculate the raw expectation and translate it, or directly translate the pmf and calculate the expectation

For the first method we have $E[X] = 14 * p(14) + 21 * p(21) + 28 * p(28) + 35 * p(35) + 42 * p(42) = 28$. So the mean weight is 28 grams. Price is given by $P = 100\frac{X}{28}$ where X is weight in grams, so we have $E[P] = \frac{100}{28}E[X] = 100$

For the second method, we can directly translate the weights to prices so that $p(50) = .1, p(75) = .15), p(100) = .5, p(125) = .15, p(150) = .42$, when yields the same value of 100

Suppose that $E[X] = 10$ and $E[Y] = -5$. Calculate $E[2X + 3Y]$

$E[2X + 3Y] = 2E[X] + 3E[Y] = 2(10) + 3(-5) = 5$

Book Questions:

3.31, 3.33, 4.15

Variance

Let X be a random variable with $E[X] = 1, var(X) = 1$. Let X_1, X_2, X_3 be independent random variable with the same distribution as X. Calculate $var(X_1 + X_2 + X_3)$ and $var(3X)$

$var(X_1 + X_2 + X_3) = var(X_1) + var(X_2) + var(X_3) = 3var(X) = 3$. $var(3X) = 9var(X) = 9$. Note that these are very different distributions (and that we don't need $E[X]$)

The standard deviation of height in adult males is 3 inches. Calculate the variance of height in centimeters.

It's easier to use the variance formula and note that height in centimeters (Y) = 2.54 times height in inches (X): $var(Y) = var(2.54X) = 2.54^2 var(X) = 2.54^2 * 3^2 \approx 58.06$. Taking the square root yields $\sigma_Y = 7.62$. Note that we could have also directly taken $2.54 * 3$

You are given that X and Y are independent with $var(X) = 1, var(Y) = 1$. Calculate $var(X - Y)$

$var(X-Y) = var(X)+var(-Y) = var(X)+(-1)^2 var(Y) = var(X)+var(Y) = 2$. Note that $var(X+Y) = var(X - Y)$ for independent events

You are given that $E[X] = 5, E[Y] = 10, var(X) = 2, var(Y) = 3, cov(X, Y) = -1$. Calculate $var(2X - 3Y)$

$var(2X - 3Y) = 4var(X) + 9var(Y) + 2(2)(-3)cov(X, Y) = 8 + 27 + 12 = 47$

You are given the following joint probability mass function for x and y, $p(x, y)$: $p(0, 0) = .1, p(1, 0) = .3, p(0, 1) = .5, p(1, 1) = .1$. Calculate the following:

- $E[X], E[Y]$
- $E[X^2], E[Y^2]$
- $var(x), var(y)$
- $E[XY], cov(X, Y), \rho_{xy}$

$E[X] = 0 * .1 + 1 * .3 + 0 * .5 + 1 * .1 = .4, E[Y] = 0 * .1 + 0 * .3 + 1 * .5 + 1 * .1 = .6$

$E[X^2] = 0^2 * .1 + 1^2 * .3 + 0^2 * .5 + 1^2 * .1 = .4, E[Y] = 0^2 * .1 + 0^2 * .3 + 1^2 * .5 + 1^2 * .1 = .6$

$var(X) = E[X^2] - E[X]^2 = .4 - .4^2 = .24. var(Y) = E[Y^2] - E[Y]^2 = .6 - .6^2 = .24$

$E[XY] = 0*0*.1 + 1*0*.3 + 0*1*.5 + 1*1*.1 = .1$

$cov(X,Y) = E[XY] - E[X]E[Y] = .1 - .4*.6 = -.14$

$\rho_{xy} = \frac{cov(X,Y)}{\sigma_X \sigma_Y} = \frac{-.14}{\sqrt{.24}\sqrt{.24}} = \frac{-.14}{.24} = -.5833$

Note that these variables are fairly strongly negatively correlated. We can infer that from the joint pmf: the probability that X and Y are both low (0,0) or both high (0,0) is only .2, compared to the .8 that one is high and the other is low (1,0) and (0,1)

[hard] The returns of a specific stock has variance of 9, while the returns on a bond has variance 4. The correlation coefficient between the two is -0.5. You own one unit of the stock. Calculate the number of bonds you must purchase to minimize your total portfolio variance

This is a fairly classic modern portfolio theory question. The variance of our portfolio is $var(X + kY)$ where X represents our stock, and kY represents our bond porfolio when we purchase k bonds. We then have $var(X + kY) = var(X) + k^2 var(Y) + 2k cov(X,Y) = 9 + 4k^2 - 3*2k$, where the last part is calculated with $-.5 = \frac{cov(X,Y)}{\sigma_X \sigma_Y}$. We then need to maximize the expression with respect to k, i.e. $argmax_k\ 9 - 6k + 4k^2$. We can solve this using calculus, but without calculus can note that this is the same as maximizing $2k(2k-3)$. This has zeroes at $k = 0, k = 1.5$, and by symmetry the maximum must be in the center of these two, which is $\frac{3}{4}$. So we should buy $\frac{3}{4}$ of a bond (in practice this would mean buying 100 stocks and 75 bonds if fractional shares aren't available)

Book Questions:

3.34, 3.35, 3.43, 3.44, 4.16, 4.19a

3.34

$E[X] = .34*25 + .12*35 = 12.7, E[X^2] = .34*25^2 + .12*35^2 = 359.5, \sigma_X = \sqrt{E[X^2] - E[X]^2} = 14.079$

$E[\sum_1^{120} X_i] = 120E[X] = 1524$, $\sigma_{\sum X_i} = \sqrt{120}\sigma_X = 154.22$

This requires independence. This is unlikely to hold in practice - for instance, domestic and international flights may tend to carry very different passengers (with respect to luggage preference), so that people with similar baggage preferences will tend to cluster onto individual flights. This would result in a positive correlation rather than independence, so that our standard deviation is underestimated. The mean calculation is still valid.

3.44

$E[X+Y_1+Y_2+Y_3] = 48+6 = 54$. $\sigma_{X+Y_1+Y_2+Y_3} = \sqrt{var(X) + 3var(Y)} = 1.0897$. Note that we're assuming that each scoop is independent here. If the random variation occurs due to variation in the size of the ice cream scoop, rather than the method of scooping, then this would be σ_{X+3Y} instead. These are different due to their differing independence assumptions

$E[X - Y] = 48 - 2 = 46$. $var(X - Y) = var(X) + (-1)^2 var(Y) = var(X) + var(Y) \implies \sigma_{x-y} = 1.03$

There's a lot of ways to answer part C, but one thing to note here is that when we subtract ice cream, we're adding entropy (randomness) to the process, so we expect the variance to increase.

4.16

$E[X] = 0*(1-p) + 1*p = p, E[X^2] = 0^2*(1-p) + 1^2*p = p. var(X) = p - p^2 = p(1-p). \sigma_X = \sqrt{p(1-p)}$

The following are more abstract, but very useful going forward: 3.45,3.46,3.47

3.46

$var(\frac{X_1+X_2+X_3}{3}) = \frac{1}{9}var(X_1 + X_2 + X_3) = \frac{1}{9}3var(X) = \frac{\sigma^2}{3}$

CDF

A distribution has CDF $F(x) = 1 - e^{-x}$. Calculate the median of X

Solve $F(X) = .5$: $1 - e^{-x} = .5 \implies .5 = e^{-x} \implies log(.5) = -x \implies x = log(2) = .693$

A distribution has CDF $F(x) = 1 - (\frac{2}{x})^2$ for $x \geq 2$. Calculate the interquartile range

Solve $F(x_1) = .25, F(x_2) = .75$ to get Q_1, Q_3: $.25 = 1 - (\frac{2}{x_1})^2 \implies .75 = (\frac{2}{x_1})^2 \implies \sqrt{.75}x_1 = 2 \implies x_1 = 2.309$. Similarly, $x_2 = 4$ so that $IQR = x_2 - x_1 = 1.691$

[moderately hard] X is a distribution with CDF $F(x) = x, 0 \leq x \leq 1$. Calculate the probability that the maximum from a sample of two from this distribution is at least 0.75, i.e. $P(max(X_1, X_2) \geq .75)$

In order for the max of two values to be less than .75, it needs to be the case that both X_1 and X_2 are less than .75. We therefore have $P(max(X_1, X_2) \leq .75) = P(X_1 \cap X_2 \leq .75) = P(X_1 \leq .75)P(X_2 \leq .75) = F(.75)F(.75) = .75^2$. We then have $P(max(X_1, X_2) \geq .75) = 1 - P(max(X_1, X_2) \leq .75) = 1 - .75^2 = .4375$. Note that this uses independence.

We can use this general principle to calculate arbitrary order statistics for random variables that are independent by invoking order statistics. This yields the beta distribution when the underlying distribution is uniform.

Book Questions:

3.37, 3.38

3.38:

(a) The 9.7% in the 100k+ bucket tells us this is rightward skewed (also that the data is censored)
(b) This is just the cumulative sum: $2.2 + 4.7 + 15.8 + 18.3 + 21.2 = 62.2\%$
(c) $.622 * .41 = 25.5\%$. This requires independence of gender and income, which is very unlikely
(d) This obviously invalidates (c). To be valid it needs to be the case that the unconditional 62.2% is also the percent of females under this threshold.

Normal Distribution

Book Questions:

4.1 (no need to draw a graph), 4.3, 4.5, 4.6, 4.7, 4.10, 4.37

4.6

$-1.645 * 583 + 4313 = 3354$ seconds. Note that this is the 5th percentile since lower numbers are faster

$1.28 * 807 + 5261 = 6294$ seconds

4.10

First, we calculate the 81.5th percentile for a standard normal: .8965. Then we know that $185 + .8965\sigma = 220 \implies \sigma = 39.04$

Binomial Distribution

Book Questions:

4.17, 4.21, 4.22

4.22

$P(X \geq 1) = 1 - P(X = 0) = 1 - .93^{10} = .516$

$P(X = 2) = .07^2 * .93^8 * \binom{10}{2} = .07^2 * .93^8 * 45 = .123$

$P(X \leq 1) = P(X = 0) + P(X = 1) = .516 + .07 * .93^9 * \binom{10}{1} = .516 + .07 * .93^9 * 10 = .88$

If there are 5 tents, there is about a 50/50 chance that at least 1 tent will have more than 1 student with arachnophobia. This is not very optimal for relatively large cohorts and would be better off to assign students based on a survey of arachnopobia if this is a big concern. Note that the probability that more than 5 students out of 50 have arachnophobia is much less likely (13.5% compared to 50%), and these numbers diverge as cohort size increases (since p=.07<.1)