Homework Assignment- Extra credit

Due May 5th before midnight via Canvas

SDS 321 Intro to Probability and Statistics

1 Questions

1. (4 pts) Civil engineers believe that W, the amount of weight (in units of 1000 pounds) that a certain span of a bridge can withstand without structural damage resulting, is normally distributed with mean 400 and standard deviation 40. Suppose that the weight (again, in units of 1000 pounds) of a car is a random variable with mean 4 and standard deviation .5. Approximately how many cars would have to be on the bridge span for the probability of structural damage to exceed .05? Let X is random variable of weight that a certain span of a bridge can withstand without structural damage resulting(in units of 1,000 pounds), $X \sim N(400, 40^2)$.) and Y_i is the weigth of a car such that $Y \sim N(4, 1/4)$, then by CLT, $\sum_i Y_i/n \sim N(4, \frac{1}{4n})$, i.e. $\sum_i Y_i \sim N(4n, \frac{n}{4})$ and we want to calculate n such that $P(\sum_i Y_i \geq X) = .05$. Let $U = \sum_i Y_i - X$. Note that $U \sim N(4n - 400, 40^2 + n/4)$.

$$P(\sum_i Y_i \geq X) = P(U \geq 0) = P\left(\frac{U - E[U]}{\sqrt{1600 + n/4}} \geq \frac{400 - 4n}{\sqrt{1600 + n/4}}\right)$$
$$= P(Z \geq \frac{400 - 4n}{\sqrt{1600 + n/4}}) = .05$$

From the normal table

$$\frac{400 - 4n}{\sqrt{1600 + n/4}} = 1.64$$

Now, who is going to solve this horrid quadratic. You can make a pretty nice guess though. n has to be smaller than 10, in which case we can just treat $\sqrt{1600} = 40$, since $n/4 \leq 2.5$ is a ton smaller than 1600. So $400 - 4n \approx 40 * 1.64$ and $n = 100 - 16 \approx 84$.

2. (4 pts) An insurance company has 10,000 automobile policyholders. The expected yearly claim per policyholder is \$250, with a standard deviation of \$750. Approximate the probability that the total yearly claim exceeds \$2.5 million. Let X_i be the yearly income for the i^{th} policyholder. X_i are i.i.d with mean 250 and variance 750^2. By CLT $W = \sum_i X_i$ behaves like a $N(250n, 750^2 n)$ random variable. We want $P(W > 2.5 \times 10^6) = P((W - 250n)/750\sqrt{n} = \frac{2.5\times10^6 - 2.5\times10^6}{750\sqrt{n}}) \approx P(Z > 0) = 0.5$

3. (4 pts) Fifty numbers are rounded off to the nearest integer and then summed. If the individual roundoff errors are uniformly distributed over $[-.5, .5]$, approximate the probability that the resultant sum differs from the exact sum by more than 3. Let each error be $E_i \sim [-1/2, 1/2]$. Let the non rounded number be X_i and the rounded number be Y_i. We have $X_i = Y_i + E_i$. Note that $\bar{E} = \sum_i E_i/n \sim N(0, \frac{1}{12n})$ by CLT. We want

$$P(|\sum_i X_i - \sum_i Y_i| \geq 3) = P(|\bar{E}| \geq 3/n) = P(\frac{|\bar{E}|}{1/\sqrt{12n}} \geq \frac{3/n}{1/\sqrt{12n}}) = P(|Z| \geq 1.5) \approx .06.$$

4. (4 pts) A certain component is critical to the operation of an electrical system and must be replaced immediately upon failure. If the mean lifetime of this type of component is 100 hours and its standard deviation is 30 hours, how many of these components must be in stock so that the probability that the system is in continual operation for the next 2000 hours is at least .95? X_i is the lifetime of each component. X_i are i.i.d. with mean 100 and variance 900. We want n such that $P(\sum_{i=1}^{n} X_i \geq 2000) \geq .95$. By CLT $W = \sum_i X_i$ approaches $N(100n, 900n)$ distribution as n grows. $P(W \geq 2000) = P(\frac{W-100n}{30\sqrt{n}} \geq \frac{2000-100n}{30\sqrt{n}}) = .95$. So $\frac{2000-100n}{30\sqrt{n}} = -1.65$. So $\frac{n-20}{30\sqrt{n}} = .016$ and so $\sqrt{n} - 20/\sqrt{n} \approx .48$. Clealy, $n > 20$. But if $n = 25$ then $\sqrt{n} - 20/\sqrt{n} = 1$. So check out values between 21 and 25. $n = 23$. Or you can solve a quadratic, but its really not necessary.

5. (8 pts) Read about convolution of two i.i.d random variables from B/T 4.2. You can also find more information here: `https://www.dartmouth.edu/~chance/teaching_aids/books_articles/probability_book/Chapter7.pdf` Briefly, if $Z = X+Y$, then for discrete X, Y we have $P(Z = z) = \sum_y P(X = z - y)P(Y = y)$. For continuous X, Y, we replace the sum by integral and PMF's by respective PDFs. $f_Z(z) = \int\limits_{-\infty}^{\infty} f_X(z-y) f_Y(y) dy$.

   (a) If $X \sim Exponential(2)$ and $Y \sim Exponential(2)$ then find the PDF of $Z = X + Y$ when X and Y are independent.

   (b) If $X \sim Uniform([0, 2])$ and $Y \sim Uniform([0, 2])$, then find the PDF of $Z = X + Y$ when X and Y are independent. *Hint: almost identical problems to both are worked out in the above source.*