# [Homework] CLT

1. Calculate the following integrals;

$$\int_{-\infty}^{\infty} xe^{-x^2-2x}dx \quad \text{and} \quad \int_{-\infty}^{\infty} x^2 e^{-x^2-2x}dx.$$

**Sol.**

$$\begin{aligned}
\int_{-\infty}^{\infty} xe^{-x^2-2x}dx &= e \cdot \int_{-\infty}^{\infty} xe^{-(x+1)^2}dx \\
&= e \cdot \sqrt{2\pi\frac{1}{2}} \int_{-\infty}^{\infty} \sqrt{2\pi\frac{1}{2}} xe^{-\frac{(x+1)^2}{2\cdot\frac{1}{2}}}dx \\
&= -e\sqrt{\pi}
\end{aligned}$$

$$\int_{-\infty}^{\infty} x^2 e^{-x^2-2x}dx = \frac{3}{2}e\sqrt{\pi}$$

2. From the fact that the MGF of $Geo(p)$ is

$$\varphi_{Geo(p)}(t) = \frac{pe^t}{1-qe^t} \quad \text{for } t < -\ln q,$$

drive the MGF of $Exp(\lambda)$.

**Sol1.**

$$\begin{aligned}
\varphi_{Exp(\lambda)}(t) &= E(e^{tX}) \\
&= \int_{-\infty}^{\infty} e^{tX} f_X(x)dx \\
&= \int_{-\infty}^{0} e^{tX} \lambda e^{-\lambda x}dx \\
&= \frac{\lambda}{\lambda-t}
\end{aligned}$$

**Sol2.** Since $np = \lambda$

$$\begin{aligned}
\frac{pe^t}{1-qe^t} &= \lim_{n\to\infty} \frac{\frac{\lambda}{n}e^{\frac{t}{n}}}{1-(1-\frac{\lambda}{n})e^{\frac{t}{n}}} \\
&= \lim_{n\to\infty} \frac{\lambda e^{\frac{t}{n}}}{n(1-e^{\frac{t}{n}})+\lambda e^{\frac{t}{n}}} \\
&= \frac{\lambda}{\lambda-t}
\end{aligned}$$

1

3. From the fact that the MGF of $Exp(\lambda)$ is

$$\varphi_{Exp(\lambda)}(t) = \frac{\lambda}{\lambda - t} \qquad \text{for } t < \lambda,$$

drive the MGF of $\Gamma(n, \lambda)$ and from this derivation of the MGF of $\Gamma(n, \lambda)$ guess the MGF of $\Gamma(\alpha, \lambda)$.

**Sol.** $X_1, \cdots, X_n$ : $n$ mutually independent random variable $\sim Exp(\lambda)$

$$
\begin{aligned}
\varphi_{Gamma(n,\lambda)}(t) = \frac{\lambda}{\lambda - t} &= \prod_{i=1}^{n} \varphi_{Exp(\lambda)}(t) \\
&= \prod_{i=1}^{n} \frac{\lambda}{\lambda - t} \\
&= \left(\frac{\lambda}{\lambda - t}\right)^n
\end{aligned}
$$

4. Using the following MGF recover the mean and variance.

| Distribution | MGF | Mean | Variance |
|:---:|:---:|:---:|:---:|
| $B(p)$ | $1 + p(e^t - 1)$ | $p$ | $p(1 - p)$ |
| $B(n, p)$ | $(1 + p(e^t - 1))^n$ | $np$ | $np(1 - p)$ |
| $Po(\lambda)$ | $e^{\lambda(e^t - 1)}$ | $\lambda$ | $\lambda$ |
| $N(\mu, \sigma^2)$ | $e^{\mu t + \frac{1}{2}\sigma^2 t^2}$ | $\mu$ | $\sigma^2$ |

5. An insurance company has 10,000 automobile policy holders. The expected yearly claim per policy holder is \$240, with a standard deviation of \$ 800. Approximate the probability that the total yearly claim exceeds \$2.7 million.

**Sol.** Let $X_i(i = 1, \cdots 10000)$ be the yearly claim of the $i$th policyholder. Note that $X_1, \cdots X_{10000}$ are independent random variables with mean 240 and variance $800^2$.

From the central limit theorem, it follows

$$Z = \frac{\sum_{i=1}^{10000} X_i - 10000 \cdot 240}{800 \cdot 100}$$

has approximately a standard normal distribution. Hence

$$P\left(\sum_{i=1}^{10000} X_i > 2700000\right) \approx P\left(Z \geq \frac{2700000 - 2400000}{800 \cdot 100}\right) = P(Z \geq 3.75) = 1 - \Phi(3.75)$$

6. A person has 100 light bulbs whose lifetimes are independent exponentials with mean 5 hours.

   (a) If the bulbs are used one at a time, with a failed bulb being replaced immediately by a new one, approximate the probability that there is still a working bulb after 525 hours.

      **Sol.** Let $X_1, \cdots, X_{100}$ be the lifetimes of the bulbs and $X$ their sum. Then $E[X] = 500$ and $Var(X) = 2500$. So

      $$P(X \geq 525) = P(\frac{X - 500}{50} \geq \frac{1}{2}) \approx 1 - \Phi(\frac{1}{2})$$

   (b) Suppose that it takes a random time, uniformly distributed over $(0, 1)$, to replace a failed bulb. Approximate the probability that all bulbs have failed by time 550.

      **Sol.** We include a minor error in our computations if we include a replacement time for the very last bulb in the total time of operation of the 100 bulbs. Let $X_i$ be the operation time of the ith bulb, this has Exponential($\frac{1}{5}$) distribution. Let also $Y_i$ be the replacement time for bulb $i$. In this case we need to take the sum of 100 independent $Z_i = X_i + Y_i$ variables into account. For any of these variables, we have

      $$E(Z) = E(\sum_{i=1}^{100} X_i) + E(\sum_{i=1}^{99} Y_i) = 5 \cdot 100 + 0.5 \cdot 99 = \frac{1099}{2}$$

      $$Var(Z) = Var(\sum_{i=1}^{100} X_i) + Var(\sum_{i=1}^{99} Y_i) = 25 \cdot 100 + \frac{1^2}{12} \cdot 99 = \frac{30099}{12}$$

      $$P(Z < 550) = P(\frac{Z - \frac{1099}{2}}{\sqrt{\frac{30099}{12}}} < \frac{550 - \frac{1099}{2}}{\sqrt{\frac{30099}{12}}}) \approx \Phi(\quad)$$

7. The ideal size of a first-year class at a particular college is 150 students. The college, knowing from past experience that, on the average, only 30 percent of those accepted for admission will actually attend, uses a policy of approving the applications of 450 students. Compute the probability that more than 150 first-year students attend this college.

   **Sol.** Let $X$ denote the number of students that attend, then $X$ is a binomial random variable with parameter $n = 450$ and $p = 0.3$. The normal approximation yields that

$$P(X \geq 150) = P(\frac{X - 450 \cdot 3}{\sqrt{450 \cdot 0.3 \cdot 0.7}} \geq \frac{150 - 450 \cdot 0.3}{\sqrt{450 \cdot 0.3 \cdot 0.7}})$$
$$\approx 1 - \Phi(\frac{15}{\sqrt{94.5}})$$

8. $A$ has 20 jobs that she must do in sequence, with the times required to do each of these jobs being independent random variables with mean 50 minutes and standard deviation 10 minutes. $B$ has 20 jobs that he must do in sequence, with the times required to do each of these jobs being independent random variables with mean 52 minutes and standard deviation 15 minutes.

   (a) Find the probability that $A$ finishes in less than 900 minutes.
      **Sol.** Let $X_i$ be the time it takes to finish job $i$ and $X = \sum_{i=1}^{20} X_i$.

      $$EX = 20 \cdot 50 = 1000, VarX = 20 \cdot 100 = 2000$$

      $$P(X < 900) = P(\frac{X - 1000}{\sqrt{2000}} \leq \frac{900 - 1000}{\sqrt{2000}})$$
      $$= P(Z \leq -\frac{100}{\sqrt{2000}})$$
      $$= 1 - \Phi(\frac{100}{\sqrt{2000}})$$

   (b) Find the probability that $B$ finishes in less than 900 minutes.
      **Sol.** Let $Y_i$ be the time it takes to finish job $i$ and $Y = \sum_{i=1}^{20} Y_i$.

      $$EY = 20 \cdot 52 = 1040, VarY = 20 \cdot 225 = 4500$$

      $$P(Y < 900) = P(\frac{Y - 1040}{\sqrt{4500}} \leq \frac{900 - 1040}{\sqrt{4500}})$$
      $$= P(Z \leq -\frac{140}{\sqrt{4500}})$$
      $$= 1 - \Phi(\frac{140}{\sqrt{4500}})$$

   (c) Find the probability that $A$ finishes before $B$.
      **Sol.**
      $$E(X - Y) = -40, Var(X - Y) = 6500$$

$$
\begin{aligned}
P(X < Y) &= P(X - Y < 0) \\
&= P(\frac{(X - Y) - (-40)}{\sqrt{6500}} < \frac{0 - (-40)}{\sqrt{6500}}) \\
&= P(Z < \frac{4}{\sqrt{65}}) = \Phi(\frac{4}{\sqrt{65}})
\end{aligned}
$$

9. (Binomial model of the stock price movement) Let's consider a stock. Comparing to the previous trading day close $S$ of the stock, at the end of the trading day we model that this stock will move up $uS$ with probability $p$ or down $dS$ with probability $1 - p$ independently, where $u = e^{0.01}$, $d = e^{-0.01}$, $p = 0.5$. Today's close of this stock is \$100. What is the probability that the stock price is more than \$110 one month later, or after 21 trading days approximately?

   **Sol.**
$$
X_i = \begin{cases} 0.01 & p = 0.5 \\ -0.01 & q = 0.5 \end{cases}
$$

   let $X = \sum_{i=1}^{21} X_i$ Then, $EX = 0, VarX = 0.0021$

$$
\begin{aligned}
P(100 \cdot e^{\sum_{i=1}^{21} X_i} \geq 110) &= P(\sum_{i=1}^{21} X_i \geq \ln\frac{110}{100}) \\
&= P(\frac{\sum_{i=1}^{21} X_i - 0}{\sqrt{0.0021}} \geq \frac{\ln\frac{110}{100} - 0}{\sqrt{0.0021}}) \\
&= P(Z \geq \frac{\ln\frac{110}{100} - 0}{\sqrt{0.0021}}) = 1 - \Phi(\frac{\ln\frac{110}{100}}{\sqrt{0.0021}})
\end{aligned}
$$

10. One thousand independent rolls of a fair die will be made. Compute an approximation to the probability that the number 6 will appear between 150 and 200 times inclusively. If the number 6 appears exactly 200 times, find the probability that the number 5 will appear less than 150 times.

   **Sol.** First, let's compute an approximation to the probability that number 6 will appear between 150 and 200 times. The die is rolled 1,000 times. These are independent trials which are Bernoulli trials since a die shows either a 6 or not. Thus, $n = 1000$. The probability of success is constant, $p = \frac{1}{6}$. We can use the binomial model.

   Let $X$ be the number of times the die shows a 6. Again, consider how difficult it

would be to use the binomial. We would have to compute the following.

$$
\begin{aligned}
P(150 \leq X \leq 200) &= P(X \leq 200) - P(X \leq 150)\\
&= \sum_{i=0}^{200} \binom{200}{i}(\tfrac{1}{6})^i(\tfrac{5}{6})^{200-i} - \sum_{i=0}^{0} \binom{150}{i}(\tfrac{1}{6})^i(\tfrac{5}{6})^{150-i}
\end{aligned}
$$

We can use the normal approximation to the binomial model because $np = 1000 \cdot \frac{1}{6} > 10$ and $n(1-p) = 1000 \cdot \frac{5}{6} > 10$. But, using the normal approximation, let $Y$ be normally distributed using the parameters from the binomial model. That is, in the normal model, $\mu = np$ and $\sigma = \sqrt{np(1-p)}$. So, let $Y \sim N(\frac{1000}{6}, \frac{5000}{36})$.

$$
\begin{aligned}
P(150 \leq X \leq 200) &= P(X \leq 200) - P(X \leq 150)\\
&= P\left(Z \leq \frac{200 - \frac{1000}{6}}{\sqrt{\frac{5000}{36}}}\right) - P\left(Z \leq \frac{150 - \frac{1000}{6}}{\sqrt{\frac{5000}{36}}}\right)\\
&= \Phi\left(\frac{200 - \frac{1000}{6}}{\sqrt{\frac{5000}{36}}}\right) - \Phi\left(\frac{150 - \frac{1000}{6}}{\sqrt{\frac{5000}{36}}}\right)
\end{aligned}
$$

Let $Y$ be the number of 6s that appear and let $X$ be the number of 5s that appear. Then we need to compute $P(Y < 150 | X = 200)$. That is, we consider that 200 of the rolls show a 5, therefore, we have 800 trials remaining that can display anything except a 5, so $n = 800$ now. Since we have already considered those 200 trials where a 5 appears, the probability of success (a 6 appears) is $p = \frac{1}{5}$.

Again, we use the normal approximation due to the large value for $n$ and since the success/failure conditions hold $(np > 10$ and $n(1-p) > 10)$. We know that $\mu = np = 800 \cdot \frac{1}{5} = 160$ and $\sigma = \sqrt{800 \cdot \frac{1}{5}\frac{4}{5}} = \sqrt{128}$.

Let $Y^* \sim N(160, 128)$. Then,

$$
P(Y^* < 150) = P\left(Z < \frac{150 - 160}{\sqrt{128}}\right) = \Phi\left(\frac{150 - 160}{\sqrt{128}}\right)
$$

# [Extra]

1. Fill the below blank;

| Distribution | MGF | Mean | Variance |
|---|---|---|---|
| $Geo(p)$ | | | |
| $Exp(\lambda)$ | | | |
| $\Gamma(\alpha, \lambda)$ | | | |

2. 100 numbers are rounded off to the nearest integer and then summed. If the individual round-off errors are uniformly distributed over $(-0.5, 0.5)$, approximate the probability that the resultant sum differs from the exact sum by more than 1.

3. We take turns tossing a fair coin, over and over. When the coin lands heads, I win \$1; when the coin lands tails, you win \$1. Approximately compute the probability that after 2000 tosses you are up by \$36.

4. 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 400. Suppose that the weight (again, in units of 1000 pounds) of a car is a random variable with mean 3 and standard deviation 0.3. Approximately how many cars would have to be on the bridge span for the probability of structural damage to exceed 2.5%?

5. To determine the effectiveness of a certain diet in reducing the amount of cholesterol in the bloodstream, 100 people are put on the diet. After they have been on the diet for a sufficient length of time, their cholesterol count will be taken. The nutritionist running this experiment has decided to endorse the diet if at least 65 percent of the people have a lower cholesterol count after going on the diet. What is the probability that the nutritionist endorses the new diet if, in fact, it has no effect on the cholesterol level?

6. Twelve percent of the population is left handed. Approximate the probability that there are at least 20 left-handers in a school of 200 students.

7. Explain why $\Gamma(t, \lambda)$ has an approximately normal distribution when $t$ is large.

8. An examination is frequently regarded as being good (in the sense of determining a valid grade spread for those taking it) if the test scores of those taking the examination can be approximated by a normal density function. (In other words, a graph of the frequency of grade scores should have approximately the bell-shaped form of the normal density.) The instructor often uses the test scores to estimate the normal parameters $\mu$ and $\sigma^2$ and then assigns the letter grade $A$ to those whose test score is greater than $\mu + \sigma$, $B$ to those whose score is between $\mu$ and $\mu + \sigma$, $C$ to those whose score is between $\mu - \sigma$ and $\mu$, $D$ to those whose score is between $\mu - 2\sigma$ and $\mu - \sigma$, and $F$ to those getting a score

below $\mu - 2\sigma$. (This strategy is sometimes referred to as grading "on the curve.") Approximately calculate how many percent of the class will get $A$?

9. Suppose that a binary message-either 0 or 1-must be transmitted by wire from location $A$ to location $B$. However, the data sent over the wire are subject to a channel noise disturbance, so, to reduce the possibility of error, the value 2 is sent over the wire when the message is 1 and the value $-2$ is sent when the message is 0. If $x$, $x = \pm 2$, is the value sent at location $A$, then $R$, the value received at location $B$, is given by $R = x + Z$, where $Z$ is the channel noise disturbance. When the message is received at location $B$, the receiver decodes it according to the following rule:

$$\text{If } R \geq 0.5, \quad \text{then 1 is concluded.}$$
$$\text{If } R < 0.5, \quad \text{then 0 is concluded.}$$

Because the channel noise is often normally distributed, we will determine the error probabilities when $Z$ is a standard normal random variable. Calculate the error probability when the original binary message is 1.