CSE 312: Foundations of Computing II
Instructor: Alex Tsun
Date: 7/29/20
Lecture Topics: 5.6 Moment Generating Functions, 5.7 Limit Theorems

[**Tags:** MGFs]

1. We'll practice using MGFs.
   a. Let $X \sim Geo(p)$. Give a formula for $M_X(t)$, and specify for which values of t the formula converges. Recall the geometric series formula: $\sum_{k=0}^{\infty} r^k = \frac{1}{1-r}$ holds for $|r| < 1$.
   b. Let $Y \sim NegBin(r, p)$. Give a formula for $M_Y(t)$.
   c. Set up formulas to compute $E[Y]$ and $Var(Y)$ using the MGF M_Y (but don't compute anything).

Solution:

a. We have $p_X(k) = (1-p)^{k-1}p$ for $k = 1,2,3, \dots$, so

$$M_X(t) = E[e^{tX}] = \sum_{k=1}^{\infty} e^{tk} p_X(k) = \sum_{k=1}^{\infty} e^{tk}(1-p)^{k-1}p = \sum_{k=1}^{\infty} e^{tk-t}e^t(1-p)^{k-1}p$$

$$= pe^t \sum_{k=1}^{\infty} e^{t(k-1)}(1-p)^{k-1} = pe^t \sum_{k=1}^{\infty} \left((1-p)e^t\right)^{k-1}$$

$$= \frac{pe^t}{1-(1-p)e^t} \quad [geometric\ series]$$

Provided $|(1-p)e^t| < 1 \rightarrow e^t < \frac{1}{|1-p|} \rightarrow \boldsymbol{t < -\log(1-p)}$ to converge.

b. Since $Y = \sum_{i=1}^{r} X_i$ for $X_i \sim Geo(p)$ iid, the MGF of the sum is the product of the MGFs:

$$M_Y(t) = M_{\sum_{i=1}^{r} X_i}(t) = \prod_{i=1}^{r} M_{X_i}(t) = \prod_{i=1}^{r} \frac{pe^t}{1-(1-p)e^t} = \left(\frac{pe^t}{1-(1-p)e^t}\right)^r$$

c. We have

$$E[Y] = M_Y'(0)$$
$$E[Y^2] = M_Y''(0)$$
$$Var(Y) = E[Y^2] - E[Y]^2$$

[**Tags:** CLT]

2. Use the CLT to approximate the following probabilities. Don't forget to apply the continuity correction (only if necessary).
   a. Suppose we roll a fair 10-sided die until we get 100 sevens. What is the probability it takes at least 1050 rolls until this happens?
   b. Let X be the sum of 10,000 real numbers, and Y be the same sum, but with each number rounded to the nearest integer before summing. If the fractions rounded off are independent and each one is uniformly distributed over $(-0.5, +0.5)$, use the Central Limit Theorem to estimate the probability that $|X - Y| > 50$. Noticing that $|X - Y|$ could have been as great as 5,000, look at your answer and think about what it says. (As a small example with sums of 4 real numbers, suppose that $X = 3.2 + 1.92 + (-3.6) + 5.7$. Then Y would be the sum of each of those terms when rounded to the nearest integer: $Y = 3 + 2 + (-4) + 6 = 7$. So, $|X - Y| = 1.3$. The fractions rounded off in this case are $(0.2, -0.08, 0.4, -0.3)$ and the assumption is that these fractions are independent and uniformly distributed in the real interval $(-0.5, +0.5)$.

Solution:

a. We know that $X \sim NegBin\left(r = 100, p = \frac{1}{10}\right)$, so $E[X] = \frac{r}{p} = 1000$ and $Var(X) = \frac{r(1-p)}{p^2} = 9000$. By the CLT, $X \approx N(\mu = 1000, \sigma^2 = 9000)$ since we have the sum of 100 iid Geometric rvs. We have to apply the continuity correction since we are approximating a discrete distribution (NegBin) with a continuous one (Normal).

$$P(X \geq 1050) = P(X \geq 1049.5) = 1 - P(X \leq 1049.5)$$
$$= 1 - P\left(\frac{X - 1000}{\sqrt{9000}} \leq \frac{1049.5 - 1000}{\sqrt{9000}}\right)$$
$$= 1 - P(Z \leq 0.52178) = 1 - \Phi(0.52) = 1 - 0.6985 = 0.3015$$

Imagine summing that Negative Binomial PMF…yuckkk

b. Let $R_1, \ldots . R_n \sim Unif(-0.5, +0.5)$ be the roundoff errors ($n = 10{,}000$). Then,

$$E[R_i] = \frac{a+b}{2} = \frac{-0.5 + 0.5}{2} = 0 \qquad Var(R_i) = \frac{\left(0.5 - (-0.5)\right)^2}{12} = \frac{1}{12}$$

Hence, $R = \sum_{i=1}^{n} R_i$ has $E[R] = 0$ and $Var(R) = \frac{10000}{12}$. By the CLT, $R \approx N\left(\mu = 0, \sigma^2 = \frac{10000}{12}\right)$. Note **we don't use a continuity correction here** since we are approximating a continuous variable.

$$P(|R| > 50) = P(R > 50) + P(R < -50)$$
$$= 2P(R > 50) \approx 2P\left(\frac{R - 0}{\sqrt{\frac{10000}{12}}} \geq \frac{50 - 0}{\sqrt{\frac{10000}{12}}}\right)$$
$$= 2P(Z \geq 1.732) = 2\left(1 - \Phi(1.732)\right) = 0.0832$$

[**Tags:** CLT, Law of Total Expectation]

3. Megha has a health condition that requires unpredictable amounts of medication. Every day, there is a 20% chance that she feels perfectly fine and requires no medicine. Otherwise, she needs to take a dose of medication. The necessary dose is equally likely to be any value in the continuous range 1 to 5 ounces. How much medicine she needs on any given day is independent of all other days. Megha's insurance will fully cover 90 ounces of medicine for each 30-day period. What is the probability that 90 ounces will be enough for the next 30 days? Make your life easier by using Central Limit Theorem.

Solution:

Let $X_1, \ldots, X_{30}$ be the amount of medicine she needs per day. Let $D_1, \ldots, D_{30}$ be the event she needs a dose on day i.

$$E[X_i] = E[X_i | D_i]P(D_i) + E\left[X_i | D_i^C\right]P\left(D_i^C\right) = \left(\frac{1+5}{2}\right) \cdot 0.8 + 0 \cdot 0.2 = 2.4$$

By LOTUS:

$$E[X_i^2] = E[X_i^2 | D_i]P(D_i) + E\left[X_i^2 | D_i^C\right]P\left(D_i^C\right) = \left(\int_1^5 x^2 \frac{1}{4} dx\right) 0.8 + 0 \cdot 0.2 \approx 8.267$$

Hence, $Var(X_i) = E[X_i^2] - E[X_i]^2 \approx 8.267 - 2.4^2 = 2.507$

The total dosage is $X = \sum_{i=1}^{30} X_i$, so $E[X] = 30 \cdot 2.4 = 72$ and $Var(X) = 30 \cdot 2.507 = 75.21$. By the CLT, since X is the sum of iid variables, $X \approx N(\mu = 72, \sigma^2 = 75.21)$, and

$$P(X < 90) = P\left(\frac{X - 72}{\sqrt{75.21}} < \frac{90 - 72}{\sqrt{75.21}}\right) \approx P(Z \leq 2.0755) = \Phi(2.0755) \approx 0.98$$