Math 170E

Lecture Notes Section 5.7 *†

Approximating discrete random variables

Instructor: Swee Hong Chan

NOTE: Materials that appear in the textbook but do not appear in the lecture notes might still be tested. Please send me an email if you find typos.

*Version date: Friday 12th March, 2021, 14:53.

†This notes is based on March Boedihardjo's and Jamie Haddock's notes from the past quarters, and I would like to thank them for their generosity. "*Nanos gigantum humeris insidentes* (I am but a dwarf standing on the shoulders of giants)".

1 Binomial RVs: Recap

Let $X_1, X_2, \ldots$ be independent Bernoulli random variables with parameter p.

Each X_i has mean $\mu = p$ and variance $\sigma^2 = p(1-p))$.

Let

$$Y \quad := \quad X_1 + X_2 + \ldots + X_n.$$

Recall that Y is the binomial random variables with parameter n and p.

2 Binomial RVs: Example

In a roulette game, the probability of winning with a bet on red is $p = \frac{18}{38}$.

Let Y equal the number of winning bets out of 1000 independent bets that are placed.

Calculate $P(Y = 500)$ and $P(Y \leq 500)$.

3 Binomial RVs: Answer

Y is the binomial RV with $n = 1000$ and $p = \frac{18}{38}$. So the first probability is equal to

$$P(Y = 500) = \binom{1000}{500} \left(\frac{18}{38}\right)^{500} \left(\frac{20}{38}\right)^{500}.$$

The second probability is equal to

$$P(Y \le 500) \quad = \quad \sum_{x=0}^{500} \binom{1000}{x} \left(\frac{18}{38}\right)^{x} \left(\frac{20}{38}\right)^{1000-x}.$$

By using **computer**, the two probabilities are 0.0063 and 0.95524.

However, direct computation is not possible when you have **no access to computers**, e.g., during final exam or actuarial exam.

4 Central limit theorem: recap, binomial

Recall the central limit theorem from Section 5.6

Theorem 1. *Let $X_1, X_2, \ldots$ be independent Bernoulli RVs with parameter p. Let $Y = X_1 + X_2 + \ldots + X_n$. Let*

$$W = \frac{Y - np}{\sqrt{np(1-p)}}.$$

Then, for n ***large enough****, we have W is approximately the standard normal RVs*

Remark 2. For binomial RVs, we usually consider n to be **large enough** when $np \geq 5$ and $n(1-p) \geq 5$.

5 Binomial approximations for pmf

Input: Binomial random variables Y with parameter n and p.

Question: The probability $P(Y = k)$, where k is an integer.

6 Half-unit correction: Motivation

For any integer k,

$$Y = k \qquad \Rightarrow \qquad W = \frac{k - np}{\sqrt{np(1-p)}}.$$

Central limit theorem says that W is approximately standard normal RV, so

$$P(Y = k) \quad = \quad P\left(W = \frac{k - np}{\sqrt{np(1-p)}}\right).$$

But the right-hand side is 0 since W is a continuous RV!

This contradiction is created because we try to approximate **discrete RVs** by **continuous RVs**.

Hence we need to apply **half-unit correction** for continuity.

7 Half-unit correction: Method

Since k is integer, we have

$$P(Y = k) \quad = \quad P\left(k - \frac{1}{2} < Y < k + \frac{1}{2}\right).$$

Now note that

$$\begin{aligned} Y = k - \frac{1}{2} \qquad &\Rightarrow \qquad W = \frac{k - \frac{1}{2} - np}{\sqrt{np(1-p)}}, \\ Y = k + \frac{1}{2} \qquad &\Rightarrow \qquad W = \frac{k + \frac{1}{2} - np}{\sqrt{np(1-p)}}. \end{aligned}$$

By central limit theorem, we then have

$$\begin{aligned} P(Y = k) \quad &= \quad P\left(\frac{k - \frac{1}{2} - np}{\sqrt{np(1-p)}} < W < \frac{k + \frac{1}{2} - np}{\sqrt{np(1-p)}}\right) \\ \approx \quad &\Phi\left(\frac{k + \frac{1}{2} - np}{\sqrt{np(1-p)}}\right) \; - \; \Phi\left(\frac{k - \frac{1}{2} - np}{\sqrt{np(1-p)}}\right). \end{aligned}$$

8 Binomial approximations for pmf: Method

Input: Binomial random variables Y with parameter n and p.

Question: The probability $P(Y = k)$, where k is an integer.

Answer: Approximate values of $P(Y = k)$ by

$$P(Y = k) \quad \approx \quad \Phi\left(\frac{k + \frac{1}{2} - np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{k - \frac{1}{2} - np}{\sqrt{np(1-p)}}\right).$$

9 Binomial approximations: Example 1

Let Y be the binomial random variable with $n = 1000$ and $p = \frac{18}{38}$.

Approximate $P(Y = 500)$ by the central limit theorem.

Answer: We have $P(Y = 500)$ is approximately

$$\Phi\left(\frac{500.5 - (1000)\left(\frac{18}{38}\right)}{\sqrt{(1000)\left(\frac{18}{38}\right)\left(\frac{20}{38}\right)}}\right) \quad - \quad \Phi\left(\frac{499.5 - (1000)\left(\frac{18}{38}\right)}{\sqrt{(1000)\left(\frac{18}{38}\right)\left(\frac{20}{38}\right)}}\right)$$

$$= \quad \Phi(1.7) - \Phi(1.64) \quad = \quad 0.0059.$$

This is very close to our answer by using computer (0.0063)!

10 Binomial approximations for cdf

Input: Binomial random variables Y with parameter n and p.

Question: The probability $P(Y \leq k)$ where k is an integer.

We have

$$
\begin{aligned}
P(Y \le k) &= \sum_{i=-\infty}^{k} P(Y=i) \\
\approx & \sum_{i=-\infty}^{k} \left(\Phi\left(\frac{i+\frac{1}{2}-np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{i-\frac{1}{2}-np}{\sqrt{np(1-p)}}\right) \right)
\end{aligned}
$$

Writing out the sum,

$$
\begin{aligned}
&\Phi\left(\frac{k+\frac{1}{2}-np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{k-\frac{1}{2}-np}{\sqrt{np(1-p)}}\right) \\
&\quad + \Phi\left(\frac{k-\frac{1}{2}-np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{k-\frac{3}{2}-np}{\sqrt{np(1-p)}}\right) \\
&\quad + \Phi\left(\frac{k-\frac{3}{2}-np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{k-\frac{5}{2}-np}{\sqrt{np(1-p)}}\right) + \ldots
\end{aligned}
$$

Every term except for the first cancels each other, so

$$
P(Y \le k) = \Phi\left(\frac{k+\frac{1}{2}-np}{\sqrt{np(1-p)}}\right)
$$

11 Binomial approximations for cdf: Method

Input: Binomial random variables Y with parameter n and p.

Question: The probability $P(Y \le k)$ where k is an integer.

Answer: Approximate values of $P(Y \le k)$ by

$$P(Y \le k) \quad \approx \quad \Phi\left(\frac{k + \frac{1}{2} - np}{\sqrt{np(1-p)}}\right).$$

12 Binomial approximations: Example 2

Let Y be the binomial random variable with $n = 1000$ and $p = \frac{18}{38}$.

Approximate $P(Y \leq 500)$ by the central limit theorem.

Answer: We have $P(Y \leq 500)$ is approximately

$$\Phi\left(\frac{500.5 - (1000)\left(\frac{18}{38}\right)}{\sqrt{(1000)\left(\frac{18}{38}\right)\left(\frac{20}{38}\right)}}\right) = \Phi(1.7) = 0.9554.$$

This is equal to the answer we computed before using the computer, up to 4 digits of error!

13 Binomial approximations: Example 3

Let Y be the binomial random variable with $n = 1000$ and $p = \frac{18}{38}$.

Approximate $P(Y < 500)$ and $P(480 < Y \leq 500)$ by the central limit theorem.

14 Binomial approximations: Answer 3

We rewrite all these probabilities in terms of pmf and cdf.

We have

$$
\begin{aligned}
P(Y < 500) &= P(Y \leq 500) - P(Y = 500) \\
&\approx \Phi(1.7) - (\Phi(1.7) - \Phi(1.64)) \\
&= \Phi(1.64) = 0.9495.
\end{aligned}
$$

We also have

$$
\begin{aligned}
P(480 < Y \leq 500) &= P(Y \leq 500) - P(Y \leq 480) \\
&\approx \Phi\left(\frac{500.5 - (1000)\left(\frac{18}{38}\right)}{\sqrt{(1000)\left(\frac{18}{38}\right)\left(\frac{20}{38}\right)}}\right) - \Phi\left(\frac{480.5 - (1000)\left(\frac{18}{38}\right)}{\sqrt{(1000)\left(\frac{18}{38}\right)\left(\frac{20}{38}\right)}}\right) \\
&= \Phi(1.7) - \Phi(0.43) = 0.289.
\end{aligned}
$$

15 Poisson RVs: Recap

Let $X_1, X_2, \ldots$ be independent Poisson random variables with $\lambda = 1$.

Each X_i has mean 1 and variance 1.

Let

$$Y = X_1 + X_2 + \ldots + X_n.$$

Recall that Y is the Poisson random variables with parameter $\lambda = n$.

16 Central limit theorem: Poisson

Theorem 3. *Let $X_1, X_2, \ldots$ be independent Poisson RVs with $\lambda = 1$. Let $Y = X_1 + X_2 + \ldots + X_n$. Let*

$$W = \frac{Y - n}{\sqrt{n}}.$$

Then, for n ***large enough****, we have W is approximately the standard normal RVs*

Remark 4. Since Poisson is discrete random variables, we still need to apply half-unit correction for continuity.

Remark 5. In general, we need to apply half-unit correction for continuity whenever the random variable is discrete with integers as support.

17 Poisson approximations

Input: Poisson random variable Y with parameter λ.

Question: The probability $P(Y = k)$ and $P(Y \leq k)$ where k is an integer.

Answer: Approximate the probabilities by

$$\begin{aligned} P(Y = k) &\approx \Phi\left(\frac{k + \frac{1}{2} - \lambda}{\sqrt{\lambda}}\right) - \Phi\left(\frac{k - \frac{1}{2} - \lambda}{\sqrt{\lambda}}\right) \\ P(Y \leq k) &\approx \Phi\left(\frac{k + \frac{1}{2} - \lambda}{\sqrt{\lambda}}\right). \end{aligned}$$

18 Poisson approximations: Example

Let Y be the Poisson random variable with $\lambda = 20$.

Approximate $P(17 \leq Y \leq 21)$ by central limit theorem.

Answer: We have:

$$\begin{aligned} P(17 \leq Y \leq 21) &= P(Y \leq 21) - P(Y \leq 16) \\ &= \Phi\left(\frac{(21.5) - 20}{\sqrt{20}}\right) - \Phi\left(\frac{(16.5) - 20}{\sqrt{20}}\right) \\ &= \Phi(0.335) - \Phi(-0.783) = 0.6312 - 0.2177 \\ &= 0.4135. \end{aligned}$$