Study Guide for Week 5 and 6

Botao Jin

University of California, Santa Barbara — April 29, 2024

Random Variables

1. Discrete random variables: $3.5, 2.38$
2. Continuous random variables: $3.7, 3.20$

Distributions

1. Binomial distribution: $2.21, 2.62$
2. Geometric distribution: $2.20, 2.22$
3. Negative binomial distribution: Example 7.7
4. Hyper-geometric distribution: $2.24, 2.28$
5. Poisson distribution: $4.10, 4.33, 4.34$
   Remark: The mean of Poisson random variable is exactly same as the parameter λ
6. Uniform distribution: $1.9, 1.11, 3.4, 3.41$
7. Exponential distribution: $4.49, 4.50$
8. Normal distribution: $3.17, 3.18$

Binomial Approximation

See exercise: 4.35

Discrete Random Variables

1. Exercise 3.5: The support for the random variable X is $\{1, 4/3, 3/2, 9/5\}$ with p.m.f:

X	1	4/3	3/2	9/5
P	1/3	1/6	1/4	1/4

2. Exercise 2.38:

   (a) Use the law of total probability:

$$\begin{aligned} P(R) &= P(R|SOME)P(SOME) + P(R|DOGS)P(DOGS) \\ &\quad + P(R|ARE)P(ARE) + P(R|BROWN)P(BROWN) \\ &= 0 + 0 + \frac{1}{3} \times \frac{1}{4} + \frac{1}{5} \times \frac{1}{4} = \frac{2}{15} \end{aligned}$$

   (b) Support of X is $S_X = \{3, 4, 5\}$, the probability mass function is

X	3	4	5
P	1/4	1/2	1/4

   (c) Note that

   - $\{X > 3\} = \{\text{SOME, DOGS, BROWN}\}$
   - $\{X = 4\} = \{\text{SOME, DOGS}\}$
   - $\{X = 5\} = \{\text{BROWN}\}$

   By definition of conditional probability, we obtain

$$P(X = 3|X > 3) = 0$$
$$P(X = 4|X > 3) = \frac{2}{3}$$
$$P(X = 5|X > 3) = \frac{1}{3}$$

   (d) Note that $\{X = k\}$ for $k = 3, 4, 5$ form a partition for the whole sample space Ω, then followed from hint, we have

$$\begin{aligned} P(R|X > 3) &= \sum_{k=3}^{5} P(R \cap \{X = k\}|X > 3) \\ &= P(R \cap \{X = 4\}|X > 3) + P(R \cap \{X = 5\}|X > 3) \\ &= P(R|X = 4)P(X = 4|X > 3) + P(R|X = 5)P(X = 5|X > 3) \end{aligned}$$

   where

$$\begin{aligned} P(R|X = 4) &= P(R \cap SOME|X = 4) + P(R \cap DOGS|X = 4) \\ &= P(R|SOME)P(SOME|X = 4) + P(R|DOGS)P(DOGS|X = 4) \\ &= 0 \end{aligned}$$

   and

$$P(R|X = 5) = P(R|BROWN) = \frac{1}{5}$$

   Using the result from part (c), we have

$$P(R|X > 3) = \frac{1}{5} \times \frac{1}{3} = \frac{1}{15}$$

   (e) By Bayes formula:

$$\begin{aligned} P(BROWN|R) &= \frac{P(R|BROWN)P(BROWN)}{P(R)} \\ &= \frac{(1/5)(1/4)}{2/15} \\ &= \frac{3}{8} \end{aligned}$$

Continuous Random Variables

1. Exercise 3.7: Since X is a continuous random variable, we have $P(a \le X \le b) = F(b) - F(a)$
   (a) $F(b) - F(a) = P(a \le X \le b) = 1$ implies that $F(b) = 1$ and $F(a) = 0$, which further implies $b \ge \sqrt{3}$ and $a \le \sqrt{2}$, thus the smallest interval is $[\sqrt{2}, \sqrt{3}]$
   (b)
   $$\begin{aligned} P(X = 1.6) &= P(X \le 1.6) - P(X < 1.6) \\ &= F(1.6) - \lim_{x \to 1.6^-} F(x) = 0 \end{aligned}$$
   The last equality holds since the cumulative density function F is continuous for any x.
   Remark. For any continuous random variable X, the probability that X takes on one special value is equal to zero.
   (c) Since X is a continuous Random Variable, we have $P(X = 1) = 0$ so that $P(1 \le X \le 3/2) = P(1 < X \le 3/2)$, hence
   $$P(1 \le X \le 3/2) = F(3/2) - F(1) = (3/2)^2 - 2 - 0 = .25$$
   (d) Since F is continuous, and it is differential apart from $\sqrt{2}$ and $\sqrt{3}$. So, we can differentiate F to obtain the density:
   $$f(x) = \begin{cases} 2x & x \in (\sqrt{2}, \sqrt{3}) \\ 0 & \text{otherwise} \end{cases}$$
2. Exercise 3.20: To show that Y follows the distribution Unif$[0, c]$, we should figure out the C.D.F or P.D.F:
   $$\begin{aligned} F_Y(y) &= \mathbb{P}(Y \le y) \\ &= \mathbb{P}(c - X \le y) \\ &= \mathbb{P}(X \ge c - y) \\ &= \begin{cases} 1 & c - y \le 0 \\ \frac{c-(c-y)}{c} & c - y \in (0, c) \\ 0 & c - y \ge c \end{cases} \\ &= \begin{cases} 1 & y \ge c \\ \frac{y}{c} & y \in (0, c) \\ 0 & y \le 0 \end{cases} \end{aligned}$$
   This is a C.D.F of Unif$[0, c]$.

Binomial Distribution

1. Exercise 2.21: Suppose X be the number of problems she correctly answer, so $X \sim Bin(4, .8)$.
   (a) $P(\text{She got an A}) = P(X \ge 3) = \binom{4}{3}(.8)^3(.2) + \binom{4}{4}(.8)^4 = .8192$
   (b) X_1 be the indicator of whether she gets the first problem correct, and Y be the number of the remaining three problems she correctly answer, so $X \sim Bern(.8)$, $Y \sim Bin(3, .8)$, X and Y are independent, and $X = X_1 + Y$
   $$\begin{aligned} P(X \ge 3 | X_1 = 1) &= \frac{P(X \ge 3, X_1 = 1)}{P(X_1 = 1)} \\ &= \frac{P(Y \ge 2, X_1 = 1)}{P(X_1 = 1)} \\ &= \frac{P(Y \ge 2)P(X_1 = 1)}{P(X_1 = 1)} \\ &= P(Y \ge 2) = 3(.8)^2(.2) + (.8)^3 = .896 \end{aligned}$$

2. Exercise 2.62: Note that

$$p = P(\text{Three marble are blue}) = \frac{\binom{9}{3}}{\binom{9+4}{3}} = \frac{42}{143}$$

Let X be the number of times we got three blue marbles, then $X \sim Bin(20, p)$ with p.m.f

$$P(X = k) = \binom{20}{k} p^k (1-p)^{20-k}$$

for $p = \frac{42}{143}$ and $k = 0, 1, \dots, 20$.

Geometric Distribution

1. Exercise 2.20: A fair dice is rolled repeatedly.

   (a) Let X be the number of threes in the first five rolls, then

   $$P(\text{get a three for at most two times}) = P(X \leq 2) = \sum_{k=0}^{2} \binom{5}{k} (1/6)^k (5/6)^{5-k}$$

   (b) Let N be the number of times we needed to see the first three, then $N \sim Geo(1/6)$

   $$\begin{aligned} P(\text{No three before the fifth roll}) &= P(\text{No three in the first four rolls}) \\ &= P(N \geq 5) = P(N > 4) = (5/6)^4 \end{aligned}$$

   (c) Consider the event that {The first three appears before the twentieth but not in the fifth roll}, which is equivalent to {The first three appears between the fifth roll and the nineteenth roll}, hence

   $$P(5 \leq N < 20) = P(N \geq 5) - P(N \geq 20) = P(N > 4) - P(N > 19) = (5/6)^4 - (5/6)^{19}$$

2. Exercise 2.22: Let X be the number of rounds we needed to see the first success, then $X \sim Geo(p)$

   (a) By symmetric, $P(\text{Anne wins}) = P(\text{Anne loses}) = P(\text{tie}) = 1/3$, so $p = 1/3$
   Remark. You can also check it by the law of total probability

   (b) $P(\text{The first win in the fourth round}) = P(X = 4) = (2/3)^3(1/3) = 8/81$

   (c) $P(\text{The first win comes after the fourth round}) = P(\text{She didn't win in the first four round}) = (2/3)^4 = 16/81$

Hyper-geometric Distribution

1. Exercise 2.24: Three people are randomly chosen from a group of two men and four women. Let X be the number of women, then

   (a) $X \sim hypergeo(6, 4, 3)$

   (b) The p.m.f of X is

   $$P(X = k) = \frac{\binom{4}{k}\binom{2}{3-k}}{\binom{6}{3}}$$

   for $k = 1, 2, 3$

2. Exercise 2.28:

   (a) Hyper-geometric distribution with parameters $(52, 4, 13)$

(b) Binomial distribution with parameters $n = 50$ and

$$p = 1 - \frac{\binom{4}{0}\binom{48}{13}}{\binom{52}{13}}$$

You can directly leave it as the fraction.

(c) Binomial distribution with parameters $n = 50$ and

$$p = \frac{\binom{4}{1}\binom{13}{13}}{\binom{52}{13}} = \frac{4}{\binom{52}{13}}$$

(d) Hyper-geometric distribution with parameters $(52, 13, 13)$

Poisson distribution

1. Exercise 4.10: Suppose X is the score, then $X \sim Poi(\lambda)$. Given the information, we have $.5 = P(X \geq 1)$, and in other words, $P(X = 0) = e^{-\lambda} = .5$. Thus $\lambda = \log 2$ and $P(X = 3) = e^{-\lambda}\lambda^3/3! = .028$

2. Exercise 4.33: Let X be the number of calls received with $X \sim Poi(\lambda)$. Then $0.5\% = P(X = 0) = e^{-\lambda}$, thus average number of calls $\lambda = -\log(.005) = 5.298$

3. Exercise 4.34: Suppose X be the number of accidents with $X \sim Poi(\lambda)$. Average $\lambda = 3$ implies that $P(X \leq 2) = e^{-3}(1 + 3/1! + 3^2/2!) = 8.5e^{-3}$

Uniform Distribution

1. Exercise 1.9: To get the shorter piece less than 1/5 of the original, the distance between the location we choose and left/right end of the stick should be lower than 1/5 of the origin, thus the probability is 2/5.

2. Exercise 1.11: $p = \frac{2^\pi}{20^2} = \frac{\pi}{100}$

3. Exercise 3.4: $X \sim Unif[4, 10]$, then the p.d.f of X is

$$f(x) = \begin{cases} 1/6 & 4 < x < 10 \\ 0 & \text{Otherwise} \end{cases}$$

(a) $P(X < 6) = \int_4^6 1/6dx = 1/3$

(b) $P(|X - 7| > 1) = P(X < 6) + P(X > 8) = 2/3$

(c) For $4 < t < 6$, we have

$$P(X < t|X < 6) = \frac{P(X < t)}{P(X < 6)} = \frac{t - 4}{6} \cdot 3 = \frac{t - 4}{2}$$

4. Exercise 3.41: Note that given $Y = k$, $X \sim Unif(0, k]$ for $k = 1, 2, 3, 4, 5, 6$. Given $s \in (3, 4)$, we have $P(X \leq s|Y = k) = 1$ for $k = 1, 2, 3$ since $X \leq k < s$ in this case. For $k = 4, 5, 6$, $s < k$ implies $P(X \leq s|Y = k) = s/k$. Thus by law of total probability, we have

$$P(X \leq s) = \sum_{k=1}^{6} P(X \leq s|Y = k)P(Y = k) = \sum_{k=1}^{3} 1 \cdot \frac{1}{6} + \sum_{k=4}^{6} \frac{s}{k} \cdot \frac{1}{6} = \frac{1}{2} + \frac{s}{6}(\frac{1}{4} + \frac{1}{5} + \frac{1}{6}) = \frac{1}{2} + \frac{37s}{360}$$

Thus $F(s) = \frac{1}{2} + \frac{37s}{360}$ and $f(s) = 37/360$

Exponential Distribution

1. Exercise 4.49: Let T be the life time of the stove, C be the cost of warranty you charge from customer, and X be the profit of the extended warranty. Thus we have

$$X = \begin{cases} C & T \geq r \\ C - 800 & T < r \end{cases} = C \cdot 1_{T \geq r} + (C - 800) 1_{T < r}$$

where 1_A represent the indicator function of the event A and $1_A \sim Bern(P(A))$. Want $E[X] = 0$, we have

$$E[X] = CP(T \geq r) + (C - 800)P(T \leq r) = Ce^{-r/10} + (C - 800)(1 - e^{-r/10}) = 0$$

The tuple (C, r) satisfies the equation above is the reasonable choices.

2. Exercise 4.50: By memoryless property of $T \sim exp(1/3)$, we have

$$P(\text{Wait for at least addition three hours}|\text{already waited for 7 hours}) = P(T \geq 3) = e^{-1}$$

More generally,

$$P(\text{Wait for at lesat x additional hours}|\text{already waited for 7 hours}) = P(T \geq x) = e^{-x/3}$$

Normal Distribution

1. Exercise 3.17: $X \sim N(-2, 7)$ and let Φ be the distribution function of $Z \sim N(0, 1)$

   (a)
$$P(X > 3.5) = P\left(\frac{X+2}{\sqrt{7}} > \frac{3.5+2}{\sqrt{7}}\right) = 1 - \Phi\left(\frac{5.5}{\sqrt{7}}\right) = .0188$$

   (b)
$$P(-2.1 < X < -1.9) = P\left(\frac{-2.1+2}{\sqrt{7}} < \frac{X+2}{\sqrt{7}} < \frac{-1.9+2}{\sqrt{7}}\right) = \Phi\left(\frac{-.1}{\sqrt{7}}\right) - \Phi\left(\frac{.1}{\sqrt{7}}\right) = .032$$

   Note that $\Phi(-z) = 1 - \Phi(z)$ and $\Phi(z) - \Phi(-z) = \Phi(z) - (1 - \Phi(z)) = 2\Phi(z) - 1$.

   (c)
$$P(X < 2) = P\left(\frac{X+2}{\sqrt{7}} 2 \frac{2+2}{\sqrt{7}}\right) = \Phi\left(\frac{4}{\sqrt{7}}\right) = .9345$$

   (d)
$$P(X < -10) = P\left(\frac{X+2}{\sqrt{7}} < \frac{-10+2}{\sqrt{7}}\right) = \Phi\left(\frac{-8}{\sqrt{7}}\right) = 1 - \Phi\left(\frac{8}{\sqrt{7}}\right) = .0013$$

   (e)
$$P(X > 4) = P\left(\frac{X+2}{\sqrt{7}} > \frac{4+2}{\sqrt{7}}\right) = 1 - \Phi\left(\frac{6}{\sqrt{7}}\right) = .0116$$

2. Exercise 3.18: $X \sim N(3, 4)$ and let Φ be the distribution function of $Z \sim N(0, 1)$

   (a)
$$P(2 < X < 6) = P\left(\frac{2-3}{2} < \frac{X-3}{2} < \frac{6-3}{2}\right) = \Phi(1.5) - \Phi(-.5) = \Phi(1.5) - 1 + \Phi(.5) = .6247$$

   (b)
$$P(X > c) = P\left(\frac{X-3}{2} > \frac{c-3}{2}\right) = 1 - \Phi\left(\frac{c-3}{2}\right) = .33$$

   By normal table, we solved for $c = 3.88$

   (c) $E[X^2] = Var(X) + (E[X])^2 = 4 + 9 = 13$

Binomial Approximation

1. Exercise 4.35: $X \sim bin(n, p)$, for $n = 365$ and

$$p = P(\text{All ten are heads or tails}) = P(\text{All heads}) + P(\text{All tails}) = \frac{1}{2^{10}} + \frac{1}{2^{10}} = \frac{1}{512}$$

   (a) Given the information on the random variable X, we have

$$P(X > 1) = \sum_{k=2}^{365} \binom{365}{k} (1/512)^k \left(1 - \frac{1}{512}\right)^{365-k}$$

   or

$$P(X > 1) = 1 - P(X = 0) - P(X = 1) = 1 - \left(1 - \frac{1}{512}\right)^{365} - \frac{365}{512}\left(1 - \frac{1}{512}\right)^{364}$$

   (b) Since p is relatively small, consider Poisson approximation is better. $\lambda = np = \frac{365}{512}$ implies

$$\begin{aligned} P(X > 1) &= 1 - P(X = 0) - P(X = 1) \\ &= 1 - e^{-\lambda} - \lambda e^{\lambda} \\ &= 1 - e^{-365/512} - (365/512)e^{365/512} = .1603 \end{aligned}$$

Note: This study guide is used for Botao Jin's sections only. Comments, bug reports: b_jin@ucsb.edu