* Announcement
① HW6 due today
② OH 9AM ~ 6PM, 10PM ~ 11PM
③ Exam-prep section 4-6PM today
④ Midterm 9PM Thu - 9PM Fri

# STAT 88: Lecture 18

**Contents**



**During the Exam**

- Start with the easy questions. Don't over-think straightforward questions.
- Read each question carefully.
- Provide reasoning or a calculation in all questions.
- Don't simplify any arithmetic or algebra in your answers unless a question explicitly asks you to.
- If an answer is taking you numerous lines of calculation, or complicated algebra/calculus, you've probably missed something.
- If you don't know how to do a problem, try not to leave it blank.

**Problem Solving Suggestions** If a question asks for a distribution, start by listing all the possible values of the random variable. Take the time to do this carefully, identifying the minimum possible value (if there is one) and the maximum possible value (if there is one). Not only will it help focus your calculation of the probabilities, it might also get you partial credit for having understood the variable even if you didn't get the probabilities correctly.

Ex: A deck consists of 12 cards, 5 of which are blue and 7 green. I draw from the deck at random without replacement till both colors have appeared among the draws. Let $D$ be the number of draws. Find the distribution of $D$.

- Possible values of $D = 2, 3, 4, \dots, 8$

$P(D=2) \quad P(D=3) \quad \cdots \quad P(D=8)$

- Find $P(D=k)$.

$k=3$ · $P(D=3) = P(bbg) + P(ggb)$

$$= \frac{\binom{5}{2}\binom{7}{0}}{\binom{12}{2}} \cdot \frac{7}{10} + \frac{\binom{5}{0}\binom{7}{2}}{\binom{12}{2}} \cdot \frac{5}{10}$$

$k=8$ · $P(D=8) = P(gggggggb) = \dfrac{\binom{5}{0}\binom{7}{7}}{\binom{12}{7}} \cdot \dfrac{5}{5}$

If a question asks for an expectation, don't immediately try to find the distribution of the random variable and then apply the definition of expectation. Try properties of expectation first. The most common are additivity (see if you can write the random variable as a sum of simpler ones) and conditioning (see if you would know the expectation if you were given the result of an early stage of the experiment).

Ex: A fair die is rolled 14 times. Let $X$ be the number of faces that appear exactly 2 times. Find $E(X)$.

① $X$ = # faces that appear twice

② $I_j = \begin{cases} 1 & j\text{th face appears twice} \\ 0 & \text{else} \end{cases}$

③ $p = P(I_j = 1) = \binom{14}{2}\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)^{12}$

④ $X = I_1 + \cdots + I_6$

⑤ $E(X) = 6 \cdot p$

The conditional distribution of $Y$ given $X = x$ is just an ordinary distribution. You have to first recognize that "given $X = x$" means you can treat the random variable $X$ as the constant $x$. You therefore have to provide the possible values of $Y$ (under the condition that $X = x$) and the corresponding conditional probabilites given $X = x$. How you calculate those probabilities depends on the setting. Sometimes you can just see what they are because the condition $X = x$ simplifies your outcome space, and sometimes you have to use the division rule $P(Y = y|X = x) = P(X = x, Y = y)/P(X = x)$. This is not particular to finding conditional distributions. It's a feature of finding conditional probabilities in general.

$X_1+X_2 \sim \text{Poi}(a+b)$

Ex: Let $X_1 \sim \text{Pois}(a)$ and $X_2 \sim \text{Pois}(b)$ and they are independent. Let $m$ be a fixed positive integer. Find the distribution of $X_1|X_1 + X_2 = m$. Recognize this as one of the famous ones and provide its parameters.

$$\underset{k \le m}{P(X_1=k \mid X_1+X_2=m)} = \frac{P(X_1=k, X_1+X_2=m)}{P(X_1+X_2=m)} = \frac{P(X_1=k, X_2=m-k)}{P(X_1+X_2=m)}$$

$$\left\{ P(X_1=a \mid X_2=b) = \frac{P(X_1=a, X_2=b)}{P(X_2=b)} \right\}$$

$$\overset{\text{independence}}{=} \frac{P(X_1=k)\, P(X_2=m-k)}{P(X_1+X_2=m)}$$

$$\overset{\text{poisson formula}}{=} \frac{\frac{e^{-a}a^k}{k!}\,\frac{e^{-b}b^{(m-k)}}{(m-k)!}}{\frac{e^{-(a+b)}\cdot(a+b)^m}{m!}} = \binom{m}{k}\left(\frac{a}{a+b}\right)^k\left(\frac{b}{a+b}\right)^{m-k}$$

$$\sim \text{Binom}\left(m, \frac{a}{a+b}\right)$$

Ex: Let $X_1 \sim \text{Pois}(a)$ and $X_2 \sim \text{Pois}(b)$ and they are independent. Find $E(X_1X_2)$.

$$E(X_1X_2) = \sum_k E(X_1 \overset{=k}{X_2} \mid X_2=k)\, P(X_2=k)$$

$$= \sum_k E(kX_1 \mid X_2=k)\, P(X_2=k)$$

$$= \sum_k E(kX_1) \cdot P(X_2=k)$$

$$= \sum_k E(X_1) \cdot \underbrace{k \cdot P(X_2=k)}_{= E(X_2)}$$

$$= E(X_1)E(X_2)$$

$$= a \cdot b$$

## More Examples

Example: There are three events $A, B, C$ such that

- $P(A)$ = Chance that you catch bus to school = 30%.
- $P(B)$ = Chance that it rains = 50%.
- $P(C)$ = Chance that you make it to class on time = 65%.

Find the best lower and upper bounds for $P(A \cup B)$, $P(A \cup B \cup C)$.

$\rightarrow \underbrace{\max(P(A), P(B))}_{= 50\%} \leq P(A \cup B) \leq \underbrace{P(A) + P(B)}_{= 80\%}$

$\rightarrow \underbrace{\max(P(A), P(B), P(C))}_{= 65\%} \leq P(A \cup B \cup C) \leq \underbrace{P(A) + P(B) + P(C)}_{= 145\%}$

More than 100%, so upper bound is 100%

Review page 4-5 of lecture note 2 for $P(A \cap B)$ and $P(A \cap B \cap C)$

Example: You roll a die and see $N$ dots, and then flip a $p = 1/3$ coin $N$ times. What is the expected number of heads you get?

① $X$ = # heads you get

$X | N=n \sim \text{Binomial}(n, p)$

$E(X) = \sum_n E(X|N=n) P(N=n)$

$= \sum_n np \cdot P(N=n) = p \cdot E(N) = \frac{1}{3} \cdot \frac{1+6}{2} = \frac{7}{6}$

Don't use this approach.

② $I_j = \begin{cases} 1 & j\text{th trial is head} \\ 0 & \text{else.} \end{cases}$ $j = 1, \dots, N$

$X = I_1 + I_2 + \cdots + I_N$

$E(X | N=n) = E(I_1 + I_2 + \cdots + \underset{\overset{\shortparallel}{n}}{I_N} \mid \boxed{N=n})$

$= E(\underbrace{I_1 + I_2 + \cdots + I_n}_{\text{Binomial}(n,p)} \mid N=n)$

$= n \cdot p.$

Example: Let $X$ be a random variable and let $A$ be a subset of the real line. Let $g(x)$ be the indicator function for $A$:

$$g(x) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{otherwise} \end{cases}$$

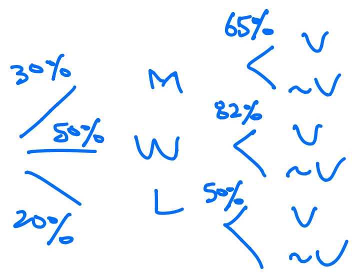
Let $Y = g(X)$. Find an expression for the cumulative distribution of $Y$.

$$Y = \begin{cases} 1 & \text{if } X \in A \\ 0 & \text{else} \end{cases}$$

$P(Y=1) = P(X \in A)$

So $Y \sim \text{Ber}(P(X \in A))$ $\rightarrow$ get CDF of $Y$

Example: Suppose that 30 percent of computer owners use a Macintosh, 50 percent use Windows, and 20 percent use Linux. Suppose that 65 percent of the Mac users have succumbed to a computer virus, 82 percent of the Windows users get the virus, and 50 percent of the Linux users get the virus. We select a person at random and learn that her system was infected with the virus. What is the probability that she is a Windows user?




$$P(W|V) = \frac{P(V|W)\,P(W)}{P(V|W)\,P(W) + P(V|L)\,P(L) + P(V|M)\,P(M)}$$

$$= \frac{0.82(0.5)}{0.82(0.5) + 0.50(0.2) + 0.65(0.3)}$$

Example: Jamarcus is writing some code for his CS 170 project. Jamarcus is a really good student so each line he codes has a 99% chance of having no bugs (errors), independent from all other lines. The project has 1000 lines of code. Assume that each line of code can only have one bug.

(a) Find approximately the chance that exactly 40 lines of code have a bug.

(b) Find approximately the chance that there are greater than 40 lines with bugs in them.

(a) $X$ = # lines w/ bugs out of 1000 lines

$X \sim \text{Binom}(1000, \underbrace{\tfrac{1}{100}}) \sim \text{Pois}(10)$

$$P(X=40) = \frac{e^{-10} 10^{40}}{40!}$$

(b) $P(X>40) = 1 - P(X \le 40)$

$= 1 - F(40)$

CDF of poisson dist'n.

Example: Class A includes 300 students with 4 seniors. Class B includes 250 students with 100 seniors. I randomly select 10 students from class A and 5 students from class B, indepednetly of each other. What is the chance that the number of seniors from class A is equal to the number of seniors from class B?

$X =$ # seniors from class A $\sim HG(300, 4, 10)$

$Y =$ " B $\sim HG(250, 100, 5)$

$$P(X=Y) = \sum_{k=0}^{4} P(X=k, Y=k)$$

$$= \sum_{k=0}^{4} P(X=k)\,P(Y=k)$$

$$= \sum_{k=0}^{4} \frac{\binom{4}{k}\binom{6}{10-k}}{\binom{300}{10}} \cdot \frac{\binom{100}{k}\binom{95}{5-k}}{\binom{250}{5}}$$

Example: A fair coin is tossed 3 times. Let

- $X$ be the number of heads in the first two tosses;
- $Y$ be the number of heads in the last two tosses.

Are $X$ and $Y$ independent?

$P(X=a, Y=b) = P(X=a)P(Y=b)$ for all $a,b$?

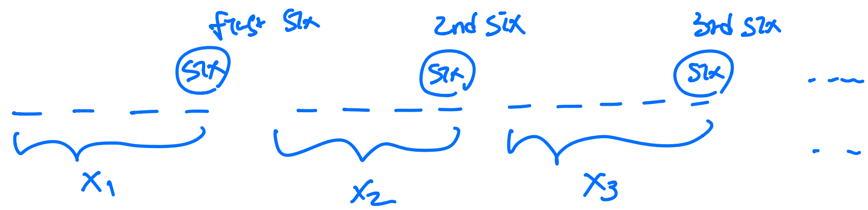
$P(X=2, Y=2) = P(\{HHH\}) = \frac{1}{8}$

$\neq$

$P(X=2) \cdot P(Y=2) = \frac{1}{4} \cdot \frac{1}{4} = \frac{1}{16}$

HHH
HHT
HTH
HTT
THH
THT
TTH
TTT

Example: (Lecture Note 16) A die is rolled repeatedly. Find the expected number of rolls till a total of 5 sixes appear.




$X = X_1 + X_2 + X_3 + X_4 + X_5.$ $\quad X_i \sim \text{Geom}(p)$

$E(X) = 5 \cdot \frac{1}{p}$

$= 5 \cdot 6 = 30$

"Random walk"

Example: A particle starts at the origin of the real line and moves along the line in jumps of one unit. For each jump the probability is $p$ that the particle will jump one unit to the left and the probability is $1-p$ that the particle will jump one unit to the right. Let $X$ be the position of the particle after $n$ units. Find $E(X)$.

Exercise.

$$I_j = \begin{cases} 1 & \text{w.p. } p \quad \text{left} \\ -1 & \text{w.p. } 1-p \quad \text{right} \end{cases} \quad \rightsquigarrow \quad E(I_j) = 1 \cdot p + (-1) \cdot (1-p) = 2p-1$$

$$X = I_1 + I_2 + \cdots + I_n$$

$$\Rightarrow E(X) = E(I_1 + \cdots + I_n) = n \cdot (2p-1)$$