NYU MathUA-233 Theory of Probability (Summer 2017 Session II)

(Week 3) Mon July 17- Thurs July 20, 2017

1 Week 3 Topics: Chapter 4: Discrete Random Variables

1. **What is a random variable?** It's a function from the sample space to the real line. Then where does the randomness come from? It comes from the random experiment whose outcomes are points in the sample space.

2. Types of random variables: discrete, continuous, joint randoms. Dirac delta function (connects discrete and continuous- see Week 4 notes).

3. **What is a probability distribution of a random variable?** it's a specification of the probabilities of a random variable.

   Probability Mass Functions and Cumulative Distribution Function

   - PMF (only for discrete random variables): $p(x) = P(X = x)$; $p(x) \geq 0$ and $\sum_x p(x) = 1$.
   - CDF: $F(x) = P(X < x)$. This function is: (a) Increasing (not necessarily strict); (b) Right continuous, (c) $\to 0$ as $x \to -\infty$ and $\to 1$ as $x \to \infty$. This is an *iff* statement.

4. **Independence of random variables**: X and Y are independent if $P(X \leq, Y \leq y) = P(X \leq x)P(Y \leq y)$ for all x and y. In the discrete case this is equivalent to $P(X = x, Y = y) = P(X = x)P(Y = y)$ (knowing the value of X tells us nothing about knowing the value of Y).

5. **Expectation of a random variable**: (Averaging- means- expected values) helpful to make predictions *before* performing the experiment. Discrete:

$$E(X) = \sum_x xP(X = x) \text{ where } \{x : P(X = x) > 0\}$$

   - Bridge between probabilities and expected values $E(X) = P(A)$ where X is the indicator random variable for the event A (Bernoulli).
   - **Linearity of expectation** Very important $E(X + Y) = E(X) + E(Y)$ even if X and Y are dependent. $E(cX) = cE(X)$ where c is a constant. Linearity remains true even when the random variables are not independent.

6. **Variance of a random variable** $Var(X) = E((X - E(X))^2) = E(X^2) - E(X)^2$.

7. **Famous and useful discrete random variables** For each, you need to know its meaning, PMF, expectation and variance. You also need to know how they relate to each other. These are summarized at the inside of the front cover of the book.

   (a) **Bernoulli**(p)

(b) **Binomial**(n, p): The number of successes in n independent Bernoulli trials.

(c) **Geometric**(p) Independent Bernoulli trials. Number of trials to obtain a first success.

(d) **Negative Binomial**(r, p)

(e) **Hypergeometric**(m, n, N)

(f) **Negative Hypergeometric** (m, n, r)

(g) **Poisson**(λ) (Very widely used). (Relationship to Binomial(n, p), Poisson paradigm, birthday problem example, St. Petersberg paradox, applications.)

8. **Useful combinatorial identities** (Helpful when computing expectations and variances of discrete random variables.) Important to understand the logic behind these identities.

- $i \binom{m}{i} = m \binom{m-1}{i-1}$ or $n \binom{N}{n} = N \binom{N-1}{n-1}$
- Vandermonde's identity: $\binom{N}{n} = \sum_{j=0}^{n} \binom{m}{j} \binom{N-m}{n-j}$
- $\binom{n}{i} = \binom{n}{n-i}$
- $\binom{n}{r} = \binom{n-1}{r-1} + \binom{n-1}{r}$, $1 \leq r \leq n$.

2 Reading assignment

Chapter 4 sections: 4.1 through 4.10.

3 Homework Assignment: Due Monday July 24, 2017

1. Book problems page 163: 4.10, 4.13, 4.17, 4.30, 4.57, 4.63. Page 170: 4.19, 4.21, 4.26.

2. There are 100 shoelaces in a box. At each stage, you pick two random ends and tie them together. Either this results in a longer shoelace (if the two ends came from different pieces), or it results in a loop (if the two ends came from the same piece). What are the expected number of steps until everything is in loops, and the expected number of loops after everything is in loops? Hint: for each step, create an indicator R.V. for whether a loop was created then, and note that the number of free ends goes down by 2 after each step.

3. (Putnam problem) You have a random permutation of the integers $1, 2, \ldots, n$, where $n \geq 2$. Find the expected number of local maxima (for example in $3, 2, 1, 4, 7, 5, 6$, a local maximum is any number that is bigger than its neighbors, in this case: 3, 7 and 6 are local maxima.) (*Answer*: (n+1)/3.)

4 Test 3 on Thursday July 20

Name:

1. (1 point) *Prosecutor's fallacy*: A crime-scene DNA sample is compared against a database of 20,000 men. A match is found, that man is accused and at his trial, it is testified that the probability that two DNA profiles match by chance is only 1 in 10,000. This does not mean the probability that the suspect is innocent is 1 in 10,000. Show that even if none of the men in the database left the crime-scene DNA, a match by chance to an innocent is almost 86%.

2. (1 point) A crime is committed by one of two suspects, A and B. Initially, there is equal evidence against both of them. In further investigation at the crime scene, it is found that the guilty party had a blood type found in 15% of the population. Suspect A does match this blood type, whereas the blood type of Suspect B is unknown.

   (a) Given this new information, what is the probability that A is the guilty party?

   (b) Given this new information, what is the probability that B?s blood type matches that found at the crime scene?

3. (3 points) The hypergeometric random variable X represents the number of white balls in a random sample of n balls chosen without replacement from an urn of N balls of which m are white.

   (a) Find the probability mass function of the hypergeometric random variable.

   (b) Prove that the expectation $E(X) = \frac{mn}{N}$. (Hint: you will need to use the following combinatorial identities: $i\begin{pmatrix} m \\ i \end{pmatrix} = m\begin{pmatrix} m-1 \\ i-1 \end{pmatrix}$ and $n\begin{pmatrix} N \\ n \end{pmatrix} = N\begin{pmatrix} N-1 \\ n-1 \end{pmatrix}$ then Vandermonde's identity: $\begin{pmatrix} N-1 \\ n-1 \end{pmatrix} = \sum_{j=0}^{n-1}\begin{pmatrix} m-1 \\ j \end{pmatrix}\begin{pmatrix} N-m \\ n-1-j \end{pmatrix}$.)

   (c) Explain the result of part (b) in terms of the probability p of drawing a white ball.

   (d) Show that if the sample is drawn with replacement, then X will be a Binomial(n, p) random variable.

   (e) Explain the logic behind Vandermonde's identity.

4. (1 point) Find the expected value of the sum obtained when n fair dice are rolled. (Hint: write as sum of simple random variables X_i = value on die i.)

5. (2 points) A ball is drawn from an urn containing 3 white and 3 black balls. After the ball is drawn, it is replaced and another ball is drawn. This process goes on indefinitely. What is the probability that of the first 4 balls drawn, exactly 2 are white?

6. (2 points) When coin 1 is flipped, it lands on heads with probability 0.4. When coin 2 is flipped, it lands on heads with probability 0.7. One of these coins is randomly chosen and flipped 10 times.

   (a) What is the probability that the coin lands on heads on exactly 7 of the 10 flips?

   (b) Given that the first of these 10 flips lands on heads, what is the conditional probability that exactly 7 out of the 10 flips lands on heads?