

# Section 5: Important Discrete Distributions, More practice with r.v.s

---

## Review of Main Concepts

- **Independence:** Random variable  $X$  and event  $E$  are independent iff

$$\forall x, \quad \mathbb{P}(X = x \cap E) = \mathbb{P}(X = x)\mathbb{P}(E)$$

Random variables  $X$  and  $Y$  are independent iff

$$\forall x \forall y, \quad \mathbb{P}(X = x \cap Y = y) = \mathbb{P}(X = x)\mathbb{P}(Y = y)$$

In this case, we have  $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$  (the converse is not necessarily true).

- **i.i.d. (independent and identically distributed):** Random variables  $X_1, \dots, X_n$  are i.i.d. (or iid) iff they are mutually independent and have the same probability mass function.
- **Independence of functions of a r.v.:** If  $X$  and  $Y$  are independent and  $g(\cdot), h(\cdot)$  are functions mapping real numbers to real numbers, then  $g(X)$  and  $h(Y)$  are independent. (See if you can prove this!)
- **Variance of Independent Variables:** If  $X$  is independent of  $Y$ ,  $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$ . This depends on independence, whereas linearity of expectation always holds. Note that this combined with the above shows that  $\forall a, b, c \in \mathbb{R}$  and if  $X$  is independent of  $Y$ ,  $\text{Var}(aX + bY + c) = a^2\text{Var}(X) + b^2\text{Var}(Y)$ .
- Review: Zoo of Discrete Random Variables

- (a) **Uniform:**  $X \sim \text{Uniform}(a, b)$  ( $\text{Unif}(a, b)$  for short), for integers  $a \leq b$ , iff  $X$  has the following probability mass function:

$$p_X(k) = \frac{1}{b - a + 1}, \quad k = a, a + 1, \dots, b$$

$\mathbb{E}[X] = \frac{a+b}{2}$  and  $\text{Var}(X) = \frac{(b-a)(b-a+1)}{12}$ . This represents each integer from  $[a, b]$  to be equally likely. For example, a single roll of a fair die is  $\text{Uniform}(1, 6)$ .

- (b) **Bernoulli (or indicator):**  $X \sim \text{Bernoulli}(p)$  ( $\text{Ber}(p)$  for short) iff  $X$  has the following probability mass function:

$$p_X(k) = \begin{cases} p, & k = 1 \\ 1 - p, & k = 0 \end{cases}$$

$\mathbb{E}[X] = p$  and  $\text{Var}(X) = p(1 - p)$ . An example of a Bernoulli r.v. is one flip of a coin with  $\mathbb{P}(\text{head}) = p$ .

- (c) **Binomial:**  $X \sim \text{Binomial}(n, p)$  ( $\text{Bin}(n, p)$  for short) iff  $X$  is the sum of  $n$  iid Bernoulli( $p$ ) random variables.  $X$  has probability mass function

$$p_X(k) = \binom{n}{k} p^k (1 - p)^{n-k}, \quad k = 0, 1, \dots, n$$

$\mathbb{E}[X] = np$  and  $\text{Var}(X) = np(1 - p)$ . An example of a Binomial r.v. is the number of heads in  $n$  independent flips of a coin with  $\mathbb{P}(\text{head}) = p$ . Note that  $\text{Bin}(1, p) \equiv \text{Ber}(p)$ . As  $n \rightarrow \infty$  and  $p \rightarrow 0$ , with  $np = \lambda$ , then  $\text{Bin}(n, p) \rightarrow \text{Poi}(\lambda)$ . If  $X_1, \dots, X_n$  are independent Binomial r.v.'s, where  $X_i \sim \text{Bin}(N_i, p)$ , then  $X = X_1 + \dots + X_n \sim \text{Bin}(N_1 + \dots + N_n, p)$ .

- (d) **Geometric:**  $X \sim \text{Geometric}(p)$  ( $\text{Geo}(p)$  for short) iff  $X$  has the following probability mass function:

$$p_X(k) = (1 - p)^{k-1} p, \quad k = 1, 2, \dots$$

$\mathbb{E}[X] = \frac{1}{p}$  and  $\text{Var}(X) = \frac{1-p}{p^2}$ . An example of a Geometric r.v. is the number of independent coin flips up to and including the first head, where  $\mathbb{P}(\text{head}) = p$ .

(e) **Poisson:**  $X \sim \text{Poisson}(\lambda)$  ( $\text{Poi}(\lambda)$  for short) iff  $X$  has the following probability mass function:

$$p_X(k) = e^{-\lambda} \frac{\lambda^k}{k!}, \quad k = 0, 1, \dots$$

$\mathbb{E}[X] = \lambda$  and  $\text{Var}(X) = \lambda$ . An example of a Poisson r.v. is the number of people born during a particular minute, where  $\lambda$  is the average birth rate per minute. If  $X_1, \dots, X_n$  are independent Poisson r.v.'s, where  $X_i \sim \text{Poi}(\lambda_i)$ , then  $X = X_1 + \dots + X_n \sim \text{Poi}(\lambda_1 + \dots + \lambda_n)$ .

(f) **Negative Binomial:**  $X \sim \text{NegativeBinomial}(r, p)$  ( $\text{NegBin}(r, p)$  for short) iff  $X$  is the sum of  $r$  iid Geometric( $p$ ) random variables.  $X$  has probability mass function

$$p_X(k) = \binom{k-1}{r-1} p^r (1-p)^{k-r}, \quad k = r, r+1, \dots$$

$\mathbb{E}[X] = \frac{r}{p}$  and  $\text{Var}(X) = \frac{r(1-p)}{p^2}$ . An example of a Negative Binomial r.v. is the number of independent coin flips up to and including the  $r^{\text{th}}$  head, where  $\mathbb{P}(\text{head}) = p$ . If  $X_1, \dots, X_n$  are independent Negative Binomial r.v.'s, where  $X_i \sim \text{NegBin}(r_i, p)$ , then  $X = X_1 + \dots + X_n \sim \text{NegBin}(r_1 + \dots + r_n, p)$ .

(g) **Hypergeometric:**  $X \sim \text{HyperGeometric}(N, K, n)$  ( $\text{HypGeo}(N, K, n)$  for short) iff  $X$  has the following probability mass function:

$$p_X(k) = \frac{\binom{K}{k} \binom{N-K}{n-k}}{\binom{N}{n}}, \quad k = \max\{0, n+K-N\}, \dots, \min\{K, n\}$$

$\mathbb{E}[X] = n \frac{K}{N}$ . This represents the number of successes drawn, when  $n$  items are drawn from a bag with  $N$  items ( $K$  of which are successes, and  $N - K$  failures) without replacement. If we did this with replacement, then this scenario would be represented as  $\text{Bin}(n, \frac{K}{N})$ .

## 1. Content Review Questions

(a) True or false:  $\text{Var}(A + B) = \text{Var}(A) + \text{Var}(B)$

(b) What is  $\text{Var}(3A + 4)$ ?

- $3\text{Var}(A) + 4$
- $3\text{Var}(A)$
- $9\text{Var}(A)$
- $\text{Var}(A)$

(c) True or false:  $\mathbb{E}[A + B] = \mathbb{E}[A] + \mathbb{E}[B]$

(d) What is  $\mathbb{E}[3A + 4]$ ?

- $3\mathbb{E}[A] + 4$
- $3\mathbb{E}[A]$
- $9\mathbb{E}[A]$
- $\mathbb{E}[A]$

## 2. Pond Fishing

Suppose I am fishing in a pond with  $B$  blue fish,  $R$  red fish, and  $G$  green fish, where  $B + R + G = N$ . Each fish is equally likely to be caught. For each of the following scenarios, identify the most appropriate distribution (with parameter(s)):

(a) How many of the next 10 fish I catch are blue, if I catch and release

$$\text{Bin} \left( 10, \frac{B}{N} \right)$$

$$\text{Ber} \left( \frac{B}{N} \right)$$

$$\text{Bin} \left( 1, \frac{B}{N} \right)$$

(b) How many fish I had to catch until my first green fish, if I catch and release

$$\text{Ber} \left( \frac{G}{N} \right)$$

$$\text{Bin} \left( 1, \frac{G}{N} \right)$$

$$\text{Geo} \left( \frac{G}{N} \right)$$

(c) How many red fish I catch in the next five minutes, if I catch on average  $r$  red fish per minute

$$\text{Poi}(5R)$$

$$\text{Bin} \left( 5, \frac{R}{N} \right)$$

$$\text{Poi}(5r)$$

(d) Whether or not my next fish is blue

$$\text{Poi}(5B)$$

$$\text{Bin} \left( 1, \frac{R}{N} \right)$$

$$\text{Ber} \left( \frac{B}{N} \right)$$

(e) (optional) How many of the next 10 fish I catch are blue, if I do not release the fish back to the pond after each catch

(f) (optional) How many fish I have to catch until I catch three red fish, if I catch and release

### 3. Balls in Bins

*Note: this problem also appeared on the section 4 handout.*

Let  $X$  be the number of bins that remain empty when  $m$  balls are distributed into  $n$  bins randomly and independently. For each ball, each bin has an equal probability of being chosen. (Notice that two bins being empty are not independent events: if one bin is empty, that decreases the probability that the second bin will also be empty. This is particularly obvious when  $n = 2$  and  $m > 0$ .) Find  $\mathbb{E}[X]$ .

### 4. 3-sided Die

*Note: a variation of this problem also appeared on the section 4 handout.* Let the random variable  $X$  be the sum of two independent rolls of a fair 3-sided die. (If you are having trouble imagining what that looks like, you can use a 6-sided die and change the numbers on 3 of its faces.)

(a) What is the probability mass function of  $X$ ?

(b) Find  $\mathbb{E}[X]$ .

(c) What is  $Var(X)$ ?

## 5. Best Coach Ever!!

You are a hardworking boxer. Your coach tells you that the probability of your winning a boxing match is 0.2 independently of every other match.

- (a) How many matches do you expect to fight until you win 10 times and what kind of random variable is this?
- (b) You only get to play 12 matches every year. To win a spot in the Annual Boxing Championship, a boxer needs to win at least 10 matches in a year. What is the probability that you will go to the Championship this year and what kind of random variable is the number of matches you win out of the 12?
- (c) Let  $p$  be your answer to part (b). How many times can you expect to go to the Championship in your 20 year career?

## 6. Variance of a Product

Let  $X, Y, Z$  be independent random variables with means  $\mu_X, \mu_Y, \mu_Z$  and variances  $\sigma_X^2, \sigma_Y^2, \sigma_Z^2$ , respectively. Find  $Var(XY - Z)$ .

## 7. True or False?

Identify the following statements as true or false (true means always true). Justify your answer.

- (a) For any random variable  $X$ , we have  $\mathbb{E}[X^2] \geq \mathbb{E}[X]^2$ .
- (b) Let  $X, Y$  be random variables. Then,  $X$  and  $Y$  are independent if and only if  $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$ .
- (c) Let  $X \sim \text{Binomial}(n, p)$  and  $Y \sim \text{Binomial}(m, p)$  be independent. Then,  $X + Y \sim \text{Binomial}(n + m, p)$ .
- (d) Let  $X_1, \dots, X_{n+1}$  be independent Bernoulli( $p$ ) random variables. Then,  $\mathbb{E}[\sum_{i=1}^n X_i X_{i+1}] = np^2$ .
- (e) Let  $X_1, \dots, X_{n+1}$  be independent Bernoulli( $p$ ) random variables. Then,  $Y = \sum_{i=1}^n X_i X_{i+1} \sim \text{Binomial}(n, p^2)$ .
- (f) If  $X \sim \text{Bernoulli}(p)$ , then  $nX \sim \text{Binomial}(n, p)$ .
- (g) If  $X \sim \text{Binomial}(n, p)$ , then  $\frac{X}{n} \sim \text{Bernoulli}(p)$ .
- (h) For any two independent random variables  $X, Y$ , we have  $Var(X - Y) = Var(X) - Var(Y)$ .

## 8. Fun with Poissons

Let  $X \sim \text{Poisson}(\lambda_1)$  and  $Y \sim \text{Poisson}(\lambda_2)$ , and  $X$  and  $Y$  are independent.

- (a) Show that  $X + Y \sim \text{Poisson}(\lambda_1 + \lambda_2)$  [This was done in class.]
- (b) Show that  $P(X = k | X + Y = n) = P(W = k)$  where  $W \sim \text{Bin}(n, \frac{\lambda_1}{\lambda_1 + \lambda_2})$

## 9. Memorylessness

We say that a random variable  $X$  is memoryless if  $\mathbb{P}(X > k + i \mid X > k) = \mathbb{P}(X > i)$  for all non-negative integers  $k$  and  $i$ . The idea is that  $X$  does not *remember* its history. Let  $X \sim Geo(p)$ . Show that  $X$  is memoryless.