

CSE 312: Foundations of Computing II

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

1. Review of Independence and Consequences

- (a) **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).

- (b) **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.
- (c) **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!)
- (d) **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 , $\text{Var}(aX + bY + c) = a^2\text{Var}(X) + b^2\text{Var}(Y)$.

2. 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 λ 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** (not covered in class): $X \sim \text{NegativeBinomial}(r, p)$ ($\text{NegBin}(r, p)$ 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^{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** (not covered in class): $X \sim \text{HyperGeometric}(N, K, n)$ ($\text{HypGeo}(N, K, n)$ 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 $\text{Bin}(n, \frac{K}{N})$.

3. Pond Fishing

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

- how many of the next 10 fish I catch are blue, if I catch and release
- how many fish I had to catch until my first green fish, if I catch and release
- how many red fish I catch in the next five minutes, if I catch on average r red fish per minute
- whether or not my next fish is blue
- (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
- (optional) how many fish I have to catch until I catch three red fish, if I catch and release

4. 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.

- How many matches do you expect to fight until you win 10 times and what kind of random variable is this?
- 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?
- Let p be your answer to part (b). How many times can you expect to go to the Championship in your 20 year career?

5. Variance of a Product

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

6. 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 $Var(X - Y) = Var(X) - Var(Y)$.

7. Fun with Poissons

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

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

8. Memorylessness

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

9. Review of Main Concepts for Continuous Random Variables

- (a) **Cumulative Distribution Function (cdf):** For any random variable (discrete or continuous) X , the cumulative distribution function is defined as $F_X(x) = \mathbb{P}(X \leq x)$. Notice that this function must be monotonically nondecreasing: if $x < y$ then $F_X(x) \leq F_X(y)$, because $\mathbb{P}(X \leq x) \leq \mathbb{P}(X \leq y)$. Also notice that since probabilities are between 0 and 1, that $0 \leq F_X(x) \leq 1$ for all x , with $\lim_{x \rightarrow -\infty} F_X(x) = 0$ and $\lim_{x \rightarrow +\infty} F_X(x) = 1$.
- (b) **Continuous Random Variable:** A continuous random variable X is one for which its cumulative distribution function $F_X(x) : \mathbb{R} \rightarrow \mathbb{R}$ is continuous everywhere. A continuous random variable has an uncountably infinite number of values.
- (c) **Probability Density Function (pdf or density):** Let X be a continuous random variable. Then the probability density function $f_X(x) : \mathbb{R} \rightarrow \mathbb{R}$ of X is defined as $f_X(x) = \frac{d}{dx} F_X(x)$. Turning this around, it means that $F_X(x) = \mathbb{P}(X \leq x) = \int_{-\infty}^x f_X(t) dt$. From this, it follows that $\mathbb{P}(a \leq X \leq b) =$

$F_X(b) - F_X(a) = \int_a^b f_X(x)dx$ and that $\int_{-\infty}^{\infty} f_X(x)dx = 1$. From the fact that $F_X(x)$ is monotonically nondecreasing it follows that $f_X(x) \geq 0$ for every real number x .

If X is a continuous random variable, note that in general $f_X(a) \neq \mathbb{P}(X = a)$, since $\mathbb{P}(X = a) = F_X(a) - F_X(a) = 0$ for all a . However, the probability that X is close to a is proportional to $f_X(a)$: for small δ , $\mathbb{P}(a - \frac{\delta}{2} < X < a + \frac{\delta}{2}) \approx \delta f_X(a)$.

(d) **Discrete to Continuous:**

	Discrete	Continuous
PMF/PDF	$p_X(x) = \mathbb{P}(X = x)$	$f_X(x) \neq \mathbb{P}(X = x) = 0$
CDF	$F_X(x) = \sum_{t \leq x} p_X(t)$	$F_X(x) = \int_{-\infty}^x f_X(t) dt$
Normalization	$\sum_x p_X(x) = 1$	$\int_{-\infty}^{\infty} f_X(x) dx = 1$
Expectation	$\mathbb{E}[X] = \sum_x x \cdot p_X(x)$	$\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} x \cdot f_X(x) dx$
LOTUS	$\mathbb{E}[g(X)] = \sum_x g(x)p_X(x)$	$\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x)f_X(x) dx$

10. Continuous r.v. example

Suppose that X is a random variable with pdf

$$f_X(x) = \begin{cases} 2C(2x - x^2) & 0 \leq x \leq 2 \\ 0 & \text{otherwise} \end{cases}$$

where C is an appropriately chosen constant.

- What must the constant C be for this to be a valid pdf?
- What is $Pr(X > 1)$?
- What is $E(X)$?