

8 Discrete Random Variables

Intuitively, to tell whether a random variable is discrete, we simply consider the possible values of the random variable. If the random variable is limited to only a finite or countably infinite number of possibilities, then it is discrete.

Example 8.1. Voice Lines: A voice communication system for a business contains 48 external lines. At a particular time, the system is observed, and some of the lines are being used. Let the random variable X denote the number of lines in use. Then, X can assume any of the integer values 0 through 48. [15, Ex 3-1]

Definition 8.2. A random variable X is said to be a **discrete random variable** if there exists a countable number of distinct real numbers x_k such that

$$\sum_k P[X = x_k] = 1. \quad (11)$$

└ finite
└ countably infinite

In other words, X is a discrete random variable if and only if X has a countable support.

Example 8.3. For the random variable N in Example 7.5 (Three Coin Tosses),

The possible values: 0, 1, 2, 3
four values finite $\Rightarrow N$ is discrete.

For the random variable S in Example 7.6 (Sum of Two Dice),

11 values finite $\Rightarrow S$ is discrete.
 The possible values: 2, 3, 4, ..., 12

8.4. Although the support S_X of a random variable X is defined as any set S such that $P[X \in S] = 1$. For discrete random variable, S_X is usually set to be $\{x : p_X(x) > 0\}$, the set of all “possible values” of X .

Definition 8.5. Important Special Case: An **integer-valued random variable** is a discrete random variable whose x_k in (11) above are all integers.

8.6. Recall, from 7.14, that the *probability distribution* of a random variable X is a description of the probabilities associated with X .

For a discrete random variable, the distribution is often characterized by just a list of the possible values (x_1, x_2, x_3, \dots) along with the probability of each:

$$(P[X = x_1], P[X = x_2], P[X = x_3], \dots, \text{ respectively}).$$

In some cases, it is convenient to express the probability in terms of a formula. This is especially useful when dealing with a random variable that has an unbounded number of outcomes. It would be tedious to list all the possible values and the corresponding probabilities.

8.1 PMF: Probability Mass Function

Definition 8.7. When X is a discrete random variable satisfying (11), we define its **probability mass function** (pmf) by²⁹

$$p_X(x) = P[X = x].$$

↑ subscript indicates the name of the RV.

- Sometimes, when we only deal with one random variable or when it is clear which random variable the pmf is associated with, we write $p(x)$ or p_x instead of $p_X(x)$.
- The argument (x) of a pmf ranges over all real numbers. Hence, the pmf is defined for x that is not among the x_k in (11). In such case, the pmf is simply 0. This is usually expressed as “ $p_X(x) = 0$, otherwise” when we specify a pmf for a particular r.v.

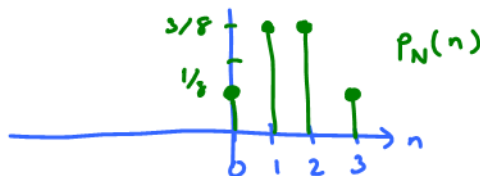
²⁹Many references (including [15] and MATLAB) use $f_X(x)$ for pmf instead of $p_X(x)$. We will NOT use $f_X(x)$ for pmf. Later, we will define $f_X(x)$ as a probability density function which will be used primarily for another type of random variable (continuous r.v.)

Example 8.8. Continue from Example 7.5. N is the number of heads in a sequence of three coin tosses.

$$N \in \{0, 1, 2, 3\}$$

n	$P[N=n]$
0	$1/8$
1	$3/8$
2	$3/8$
3	$1/8$

$$P_N(n) = P[N=n] = \begin{cases} 1/8, & n = 0, 3, \\ 3/8, & n = 1, 2, \\ 0, & \text{otherwise.} \end{cases}$$



8.9. Graphical Description of the Probability Distribution: Traditionally, we use **stem plot** to visualize p_X . To do this, we graph a pmf by marking on the horizontal axis each value with nonzero probability and drawing a vertical bar with length proportional to the probability.

8.10. Any pmf $p(\cdot)$ satisfies two properties:

(a) $p(\cdot) \geq 0$

(b) there exists numbers x_1, x_2, x_3, \dots such that $\sum_k p(x_k) = 1$ and $p(x) = 0$ for other x .

When you are asked to verify that a function is a pmf, check these two properties.

8.11. Finding probability from pmf: for any subset B of \mathbb{R} , we can find

$$P[X \in B] = \sum_{x_k \in B} P[X = x_k] = \sum_{x_k \in B} p_X(x_k).$$

In particular, for integer-valued random variables,

$$P[X \in B] = \sum_{k \in B} P[X = k] = \sum_{k \in B} p_X(k).$$

8.12. Steps to find probability of the form P [some condition(s) on X] when the pmf $p_X(x)$ is known.

- Find the support of X .
- Consider only the x inside the support. Find all values of x that satisfies the condition(s).
- Evaluate the pmf at x found in the previous step.
- Add the pmf values from the previous step.

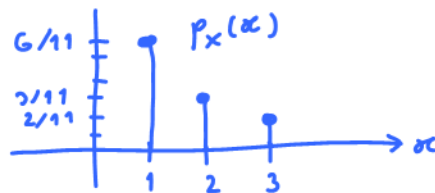
Example 8.13. Suppose a random variable X has pmf

$$p_X(x) = \begin{cases} c/x, & x = 1, 2, 3, \\ 0, & \text{otherwise.} \end{cases}$$

- (a) The value of the constant c is

$$\frac{c}{1} + \frac{c}{2} + \frac{c}{3} = 1 \Rightarrow c = \frac{6}{11}$$

- (b) Sketch of pmf



(c) $P[X = 1] = \frac{6}{11}$

(f) $P[X \leq 1.5] = \frac{6}{11}$

(g) $P[X \leq 5] = 1$

(d) $P[X \geq 2] = \frac{3}{11} + \frac{2}{11} = \frac{5}{11}$

(h) $P[X \leq 0] = 0$

(e) $P[X > 3] = 0$

8.14. Any function $p(\cdot)$ on \mathbb{R} which satisfies

- (a) $p(\cdot) \geq 0$, and
- (b) there exists numbers x_1, x_2, x_3, \dots such that $\sum_k p(x_k) = 1$ and $p(x) = 0$ for other x

is a pmf of some discrete random variable.

8.2 CDF: Cumulative Distribution Function

Definition 8.15. The **(cumulative) distribution function (cdf)** of a random variable X is the function $F_X(x)$ defined by

$$F_X(x) = P[X \leq x].$$

- The argument (x) of a cdf ranges over all real numbers.
- From its definition, we know that $0 \leq F_X \leq 1$.
- Think of it as a function that collects the “probability mass” from $-\infty$ up to the point x .

8.16. From pmf to cdf: In general, for any discrete random variable with possible values x_1, x_2, \dots , the cdf of X is given by

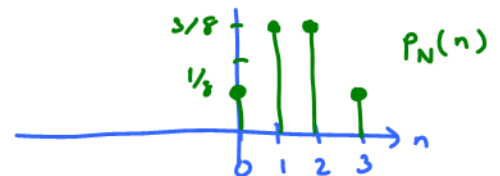
$$F_X(x) = P[X \leq x] = \sum_{x_k \leq x} p_X(x_k).$$

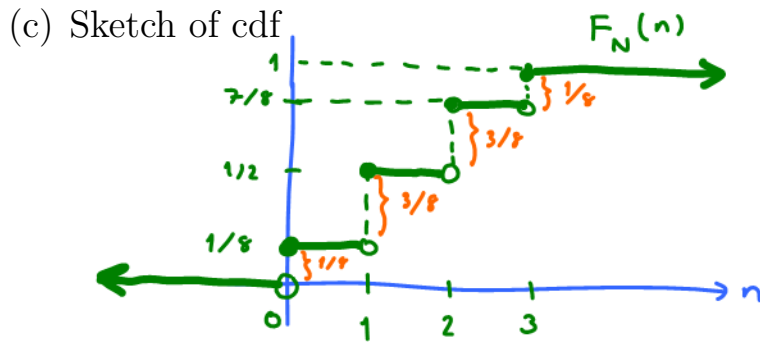
Example 8.17. Continue from Examples 7.5, 7.11, and 8.8 where N is defined as the number of heads in a sequence of three coin tosses. We have

$$p_N(0) = p_N(3) = \frac{1}{8} \text{ and } p_N(1) = p_N(2) = \frac{3}{8}.$$

(a) $F_N(0) = P[N \leq 0] = \frac{1}{8}$

(b) $F_N(1.5) = P[N \leq 1.5] = \frac{1}{8} + \frac{3}{8} = \frac{4}{8} = \frac{1}{2}$





8.18. Facts:

- For any discrete r.v. X , F_X is a right-continuous, **staircase function** of x with jumps at a countable set of points x_k .
- When you are given the cdf of a discrete random variable, you can derive its pmf from the locations and sizes of the jumps. If a jump happens at $x = c$, then $p_X(c)$ is the same as the amount of jump at c . At the location x where there is no jump, $p_X(x) = 0$.

Example 8.19. Consider a discrete random variable X whose cdf $F_X(x)$ is shown in Figure 9.

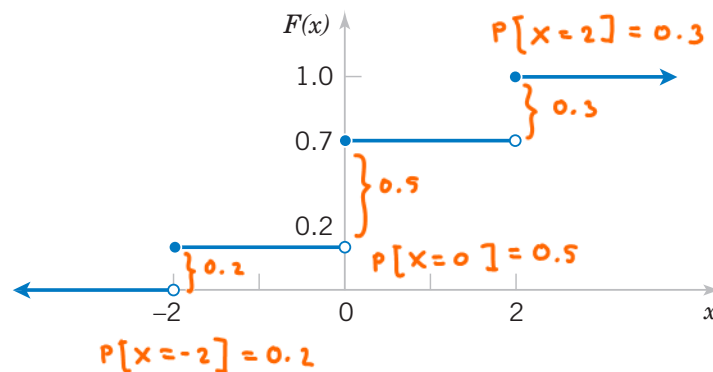


Figure 9: CDF for Example 8.19

Determine the pmf $p_X(x)$.

$$p_X(x) = \begin{cases} 0.2, & x = -2, \\ 0.5, & x = 0, \\ 0.3, & x = 2, \\ 0, & \text{otherwise.} \end{cases}$$

8.20. Characterizing³⁰ properties of cdf:

CDF1 F_X is **non-decreasing** (monotone increasing)

$$P[X \leq a] \leq P[X \leq b]$$

$$F_X''(a) \text{ vs. } F_X''(b)$$

$$b > a$$



CDF2 F_X is right continuous (continuous from the right)

Recall: $f(x)$ is continuous iff $\forall x_0, \lim_{x \rightarrow x_0} f(x) = f(x_0)$

$$\lim_{x \rightarrow x_0} f(x) = f(x_0)$$

$$\lim_{x \rightarrow x_0^+} f(x) = \lim_{x \rightarrow x_0^-} f(x)$$

$f(x)$ is right continuous iff $\forall x_0, \lim_{x \rightarrow x_0^+} f(x) = f(x_0)$

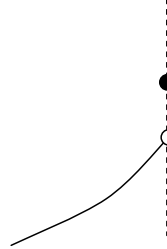


Figure 10: Right-continuous function at jump point

CDF3 $\lim_{x \rightarrow -\infty} F_X(x) = 0$ and $\lim_{x \rightarrow \infty} F_X(x) = 1$.

8.21. F_X can be written as

$$F_X(x) = \sum_{x_k} p_X(x_k) u(x - x_k),$$

where $u(x) = 1_{[0, \infty)}(x)$ is the unit step function.

³⁰These properties hold for any type of random variables. Moreover, for any function F that satisfies these three properties, there exists a random variable X whose CDF is F .

ECS315 2013/1 Part III.2 Dr.Prapun

Useful and Popular

8.3 Families of Discrete Random Variables

Many physical systems can be modeled by the same or similar random experiments and random variables. In this subsection, we present the analysis of several discrete random variables that frequently arise in applications.³¹

Remember the pmf of these RVs.

Uniform


Bernoulli

Binomial

Geometric

Poisson

Definition 8.22. X is *uniformly distributed* on a finite set S if

$$p_X(x) = P[X = x] = \begin{cases} \frac{1}{|S|}, & x \in S, \\ 0, & \text{otherwise,} \end{cases}$$


- We write $X \sim \mathcal{U}(S)$ or $X \sim \text{Uniform}(S)$.
- Read “ X is uniform on S ” or “ X is a uniform random variable on set S ”.
- The pmf is usually referred to as the uniform discrete distribution.
- Simulation: When the support S contains only consecutive integers³², it can be generated by the command `randi` in MATLAB (R2008b).

³¹As mention in 7.12, we often omit a discussion of the underlying sample space of the random experiment and directly describe the distribution of a particular random variable.

³²or, with minor manipulation, only uniformly spaced numbers

Example 8.23. X is uniformly distributed on $1, 2, \dots, n$ if

In MATLAB, X can be generated by `randi(10)`.

Example 8.24. Uniform pmf is used when the random variable can take finite number of “equally likely” or “totally random” values.

- Classical game of chance / classical probability
- Fair gaming devices (well-balanced coins and dice, well-shuffled decks of cards)

Example 8.25. Roll a fair dice. Let X be the outcome.

Definition 8.26. X is a **Bernoulli** random variable if

$$p_X(x) = \begin{cases} 1 - p, & x = 0, \\ p, & x = 1, \\ 0, & \text{otherwise,} \end{cases} \quad p \in (0, 1)$$

- Write $X \sim \mathcal{B}(1, p)$ or $X \sim \text{Bernoulli}(p)$
- X takes only two values: 0 or 1

Definition 8.27. X is a **binary** random variable if

$$p_X(x) = \begin{cases} 1 - p, & x = a, \\ p, & x = b, \\ 0, & \text{otherwise,} \end{cases} \quad p \in (0, 1), \quad b > a.$$

- X takes only two values: a or b

Definition 8.28. X is a *binomial* random variable with size $n \in \mathbb{N}$ and parameter $p \in (0, 1)$ if

$$p_X(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x}, & x \in \{0, 1, 2, \dots, n\}, \\ 0, & \text{otherwise} \end{cases} \quad (12)$$

- Write $X \sim \mathcal{B}(n, p)$ or $X \sim \text{binomial}(p)$.
 - Observe that $\mathcal{B}(1, p)$ is Bernoulli with parameter p .
- To calculate $p_X(x)$, can use `binopdf(x,n,p)` in MATLAB.
- Interpretation: X is the number of successes in n independent Bernoulli trials.

Example 8.29. An optical inspection system is to distinguish among different part types. The probability of a correct classification of any part is 0.98. Suppose that three parts are inspected and that the classifications are independent.

- (a) Let the random variable X denote the number of parts that are correctly classified. Determine the probability mass function of X . [15, Q3-20]
- (b) Let the random variable Y denote the number of parts that are incorrectly classified. Determine the probability mass function of Y .

Solution:

- (a) X is a binomial random variable with $n = 3$ and $p = 0.98$. Hence,

$$p_X(x) = \begin{cases} \binom{3}{x} 0.98^x (0.02)^{3-x}, & x \in \{0, 1, 2, 3\}, \\ 0, & \text{otherwise} \end{cases} \quad (13)$$

In particular, $p_X(0) = 8 \times 10^{-6}$, $p_X(1) = 0.001176$, $p_X(2) = 0.057624$, and $p_X(3) = 0.941192$. Note that in MATLAB, these probabilities can be calculated by evaluating `binopdf(0:3,3,0.98)`.

(b) Y is a binomial random variable with $n = 3$ and $p = 0.02$. Hence,

$$p_Y(y) = \begin{cases} \binom{3}{y} 0.02^y (0.98)^{3-y}, & y \in \{0, 1, 2, 3\}, \\ 0, & \text{otherwise} \end{cases} \quad (14)$$

In particular, $p_Y(0) = 0.941192$, $p_Y(1) = 0.057624$, $p_Y(2) = 0.001176$, and $p_Y(3) = 8 \times 10^{-6}$. Note that in MATLAB, these probabilities can be calculated by evaluating `binopdf(0:3,3,0.02)`.

Alternatively, note that there are three parts. If X of them are classified correctly, then the number of incorrectly classified parts is $n - X$, which is what we defined as Y . Therefore, $Y = 3 - X$. Hence, $p_Y(y) = P[Y = y] = P[3 - X = y] = P[X = 3 - y] = p_X(3 - y)$.

Example 8.30. Daily Airlines flies from Amsterdam to London every day. The price of a ticket for this extremely popular flight route is \$75. The aircraft has a passenger capacity of 150. The airline management has made it a policy to sell 160 tickets for this flight in order to protect themselves against no-show passengers. Experience has shown that the probability of a passenger being a no-show is equal to 0.1. The booked passengers act independently of each other. Given this overbooking strategy, what is the probability that some passengers will have to be bumped from the flight?

Solution: This problem can be treated as 160 independent trials of a Bernoulli experiment with a success rate of $p = 9/10$, where a passenger who shows up for the flight is counted as a success. Use the random variable X to denote number of passengers that show up for a given flight. The random variable X is binomial distributed with the parameters $n = 160$ and $p = 9/10$. The probability in question is given by

$$P[X > 150] = 1 - P[X \leq 150] = 1 - F_X(150).$$

In MATLAB, we can enter `1-binocdf(150,160,9/10)` to get 0.0359. Thus, the probability that some passengers will be bumped from any given flight is roughly 3.6%. [22, Ex 4.1]

Definition 8.31. A geometric random variable X is defined by the fact that for some constant $\beta \in (0, 1)$,

$$p_X(k+1) = \beta \times p_X(k)$$

for all $k \in S$ where S can be either \mathbb{N} or $\mathbb{N} \cup \{0\}$.

(a) When its support is $\mathbb{N} = \{1, 2, \dots\}$,

$$p_X(x) = \begin{cases} (1 - \beta) \beta^{x-1}, & x = 1, 2, \dots \\ 0, & \text{otherwise.} \end{cases}$$

- Write $X \sim \mathcal{G}_1(\beta)$ or $\text{geometric}_1(\beta)$.
- In MATLAB, to evaluate $p_X(x)$, use `geopdf(x-1, 1-beta)`.
- Interpretation: X is the number of trials required in Bernoulli trials to achieve the first success.

In particular, in a series of Bernoulli trials (independent trials with constant probability p of a success), let the random variable X denote the number of trials until the first success. Then X is a geometric random variable with parameter $\beta = 1 - p$ and

$$\begin{aligned} p_X(x) &= \begin{cases} (1 - \beta) \beta^{x-1}, & x = 1, 2, \dots \\ 0, & \text{otherwise} \end{cases} \\ &= \begin{cases} p(1 - p)^{x-1}, & x = 1, 2, \dots \\ 0, & \text{otherwise.} \end{cases} \end{aligned}$$

(b) When its support is $\mathbb{N} \cup \{0\}$,

$$\begin{aligned} p_X(x) &= \begin{cases} (1 - \beta) \beta^x, & x = 0, 1, 2, \dots \\ 0, & \text{otherwise} \end{cases} \\ &= \begin{cases} p(1 - p)^x, & x = 0, 1, 2, \dots \\ 0, & \text{otherwise.} \end{cases} \end{aligned}$$

- Write $X \sim \mathcal{G}_0(\beta)$ or $\text{geometric}_0(\beta)$.
- In MATLAB, to evaluate $p_X(x)$, use `geopdf(x, 1-beta)`.
- Interpretation: X is the number of failures in Bernoulli trials before the first success occurs.

8.32. In 1837, the famous French mathematician Poisson introduced a probability distribution that would later come to be known as the Poisson distribution, and this would develop into one of the most important distributions in probability theory. As is often remarked, Poisson did not recognize the huge practical importance of the distribution that would later be named after him. In his book, he dedicates just one page to this distribution. It was Bortkiewicz in 1898, who first discerned and explained the importance of the Poisson distribution in his book *Das Gesetz der Kleinen Zahlen* (*The Law of Small Numbers*). [22]

Definition 8.33. X is a **Poisson** random variable with **parameter** $\alpha > 0$ if

$$p_X(x) = \begin{cases} e^{-\alpha} \frac{\alpha^x}{x!}, & x = 0, 1, 2, \dots \\ 0, & \text{otherwise} \end{cases}$$

- In MATLAB, use `poisspdf(x,alpha)`.
- Write $X \sim \mathcal{P}(\alpha)$ or $\text{Poisson}(\alpha)$.
- We will see later in Example 9.7 that α is the “average” or expected value of X .
- Instead of X , Poisson random variable is usually denoted by Λ . The parameter α is often replaced by $\lambda\tau$ where λ is referred to as the **intensity/rate parameter** of the distribution

Example 8.34. The first use of the Poisson model is said to have been by a Prussian (German) physician, Bortkiewicz, who found that the annual number of late-19th-century Prussian (German) soldiers kicked to death by horses fitted a Poisson distribution [6, p 150],[3, Ex 2.23]³³.

³³I. J. Good and others have argued that the Poisson distribution should be called the Bortkiewicz distribution, but then it would be very difficult to say or write.

Example 8.35. The number of hits to a popular website during a 1-minute interval is given by $N \sim \mathcal{P}(\alpha)$ where $\alpha = 2$.

(a) Find the probability that there is at least one hit between 3:00AM and 3:01AM.

(b) Find the probability that there are at least 2 hits during the time interval above.

8.36. One of the reasons why Poisson distribution is important is because many natural phenomena can be modeled by *Poisson processes*.

Definition 8.37. A *Poisson process* (PP) is a random arrangement of “marks” (denoted by “ \times ” below) on the time line.

The “marks” may indicate the arrival times or occurrences of event/phenomenon of interest.

Example 8.38. Examples of processes that can be modeled by *Poisson process* include

(a) the sequence of times at which lightning strikes occur or mail carriers get bitten within some region

(b) the emission of particles from a radioactive source

(c) the arrival of

- telephone calls at a switchboard or at an automatic phone-switching system
- urgent calls to an emergency center
- (filed) claims at an insurance company
- incoming spikes (action potential) to a neuron in human brain

(d) the occurrence of

- serious earthquakes
- traffic accidents
- power outages

in a certain area.

(e) page view requests to a website

8.39. It is convenient to consider the Poisson process in terms of customers arriving at a facility.

We focus on a type of Poisson process that is called *homogeneous Poisson process*.

Definition 8.40. For *homogeneous Poisson process*, there is only one parameter that describes the whole process. This number is called the *rate* and usually denoted by λ .

Example 8.41. If you think about modeling customer arrival as a Poisson process with rate $\lambda = 5$ customers/hour, then it means that during any fixed time interval of duration 1 hour (say, from noon to 1PM), you expect to have about 5 customers arriving in that interval. If you consider a time interval of duration two hours (say, from 1PM to 3PM), you expect to have about $2 \times 5 = 10$ customers arriving in that time interval.

8.42. One important fact which we will revisit later is that, for a homogeneous Poisson process, the number of arrivals during a time interval of duration T is a Poisson random variable with parameter $\alpha = \lambda T$.

Example 8.43. Examples of Poisson *random variables*:

- #photons emitted by a light source of intensity λ [photons/second] in time τ
- #atoms of radioactive material undergoing decay in time τ
- #clicks in a Geiger counter in τ seconds when the average number of click in 1 second is λ .
- #dopant atoms deposited to make a small device such as an FET
- #customers arriving in a queue or workstations requesting service from a file server in time τ
- Counts of demands for telephone connections in time τ
- Counts of defects in a semiconductor chip.

Example 8.44. Thongchai produces a new hit song every 7 months on average. Assume that songs are produced according to a Poisson process. Find the probability that Thongchai produces more than two hit songs in 1 year.

8.45. Poisson approximation of Binomial distribution: When p is small and n is large, $\mathcal{B}(n, p)$ can be approximated by $\mathcal{P}(np)$

- (a) In a large number of independent repetitions of a Bernoulli trial having a small probability of success, the total number of successes is approximately Poisson distributed with parameter $\alpha = np$, where n = the number of trials and p = the probability of success. [22, p 109]

(b) More specifically, suppose $X_n \sim \mathcal{B}(n, p_n)$. If $p_n \rightarrow 0$ and $np_n \rightarrow \alpha$ as $n \rightarrow \infty$, then

$$P[X_n = k] = \binom{n}{k} p_n^k (1 - p_n)^{n-k} \rightarrow e^{-\alpha} \frac{\alpha^k}{k!}.$$

Example 8.46. Consider $X_n \sim \mathcal{B}(n, 1/n)$.

Example 8.47. Recall that Bortkiewicz applied the Poisson model to the number of Prussian cavalry deaths attributed to fatal horse kicks. Here, indeed, one encounters a very large number of trials (the Prussian cavalymen), each with a very small probability of “success” (fatal horse kick).

8.48. Summary:

$X \sim$	Support S_X	$p_X(x)$
Uniform \mathcal{U}_n	$\{1, 2, \dots, n\}$	$\frac{1}{n}$
$\mathcal{U}_{\{0,1,\dots,n-1\}}$	$\{0, 1, \dots, n-1\}$	$\frac{1}{n}$
Bernoulli $\mathcal{B}(1, p)$	$\{0, 1\}$	$\begin{cases} 1-p, & x=0 \\ p, & x=1 \end{cases}$
Binomial $\mathcal{B}(n, p)$	$\{0, 1, \dots, n\}$	$\binom{n}{x} p^x (1-p)^{n-x}$
Geometric $\mathcal{G}_0(\beta)$	$\mathbb{N} \cup \{0\}$	$(1-\beta)\beta^x$
Geometric $\mathcal{G}_1(\beta)$	\mathbb{N}	$(1-\beta)\beta^{x-1}$
Poisson $\mathcal{P}(\alpha)$	$\mathbb{N} \cup \{0\}$	$e^{-\alpha} \frac{\alpha^x}{x!}$

Table 3: Examples of probability mass functions. Here, $p, \beta \in (0, 1)$. $\alpha > 0$. $n \in \mathbb{N}$

8.4 Some Remarks

8.49. Sometimes, it is useful to define and think of pmf as a vector \underline{p} of probabilities.

When you use **MATLAB**, it is also useful to keep track of the values of x corresponding to the probabilities in p . This can be done via defining a vector \underline{x} .

Example 8.50. For $\mathcal{B}(3, \frac{1}{3})$, we may define

$$\underline{x} = [0, 1, 2, 3]$$

and

$$\begin{aligned}\underline{p} &= \left[\binom{3}{0} \left(\frac{1}{3}\right)^0 \left(\frac{2}{3}\right)^3, \binom{3}{1} \left(\frac{1}{3}\right)^1 \left(\frac{2}{3}\right)^2, \binom{3}{2} \left(\frac{1}{3}\right)^2 \left(\frac{2}{3}\right)^1, \binom{3}{3} \left(\frac{1}{3}\right)^3 \left(\frac{2}{3}\right)^0 \right] \\ &= \left[\frac{8}{27}, \frac{4}{9}, \frac{2}{9}, \frac{1}{27} \right]\end{aligned}$$

8.51. At this point, we have a couple of ways to define probabilities that are associated with a random variable X

- (a) We can define $P[X \in B]$ for all possible set B .
- (b) For discrete random variable, we only need to define its pmf $p_X(x)$ which is defined as $P[X = x] = P[X \in \{x\}]$.
- (c) We can also define the cdf $F_X(x)$.

Definition 8.52. If $p_X(c) = 1$, that is $P[X = c] = 1$, for some constant c , then X is called a ***degenerated*** random variable.