Chapter C: Random Variables

In the definitions and theorems in this chapter, we assume a probability space $(S, \mathcal{A}, P)$ is given.

C.I Discrete Random Variables

Loosely speaking, a discrete random variable is a map $X : S \to \mathbb{R}$.

Example 1. Suppose a fair coin is tossed 3 times in a row. Then our sample space S are all the possible triplets of letters t or h (for tail and head respectively). The event space is $\mathcal{A} = \mathcal{P}(S)$, and every outcome have an equal probability, so that $P(A) = 1/8$ for every atomic event A.

Let X denotes "the number of heads appearing", then X is a discrete random variable.

The possible values of X, called the image of X, is $\text{im}X = \{0, 1, 2, 3\}$. Each value can be assigned a probability:

$$P(X = 0) := P(\{s \in S \,:\, X(s) = 0\}) = P(\{ttt\}) = \frac{1}{8}$$
$$P(X = 1) := P(\{s \in S \,:\, X(s) = 1\}) = P(\{htt, tht, tth\}) = \frac{3}{8}$$
$$P(X = 2) := P(\{s \in S \,:\, X(s) = 2\}) = P(\{hht, hth, thh\}) = \frac{3}{8}$$
$$P(X = 3) := P(\{s \in S \,:\, X(s) = 3\}) = P(\{ttt\}) = \frac{1}{8}.$$

Definition 1. A discrete random variable X on a probability space $(S, \mathcal{A}, P)$ is defined to be a mapping $X : S \to \mathbb{R}$ such that

a) the set $\text{Im}\, X$ is a countable subset of $\mathbb{R}$;

b) $X^{-1}(\{x\}) := \{s \in S \,:\, X(s) = x\}$ is an event for every $x \in \mathbb{R}$.

Note:

- To simplify the notation, we will use the notation $\{X = x\}$ to denote the set $X^{-1}(\{x\})$.
- We also generalize this notation to include pre-images of intervals. For $x \in \mathbb{R}$,

$$\{X \leq x\} := \{s \in S \,:\, X(s) \leq x\}$$

  and analogously for $\{X \geq x\}$.

Example 2. Let $(S, \mathcal{A}, P)$ be a probability space in which

$$S = \{1, 2, 3, 4, 5, 6\}, \quad \mathcal{A} = \{\varnothing, \{2, 4, 6\}, \{1, 3, 5\}, S\}.$$

Let U and V be defined by

$$U(s) = s, \quad V(s) = \begin{cases} 1 & \text{if } s \text{ is even,} \\ 0 & \text{if } s \text{ is odd} \end{cases}$$

for $s \in S$.

a) Is U a discrete random variable?

b) Is V a discrete random variable?

Solution.

a) The set $\operatorname{Im} U = \{1, 2, 3, 4, 5, 6, \}$, which is discrete. The problem is the set $U^{-1}(\{u\})$ for some real values of u. If $u = 1$, then $U^{-1}(\{1\}) = \{1\}$ which is not in the event space. Therefore, $U^{-1}(\{1\})$ is not an event and U is not a discrete random variable.

b) The set $\operatorname{Im} V = \{0, 1\}$, which is discrete. If $v \neq 0, 1$, then $V^{-1}(\{v\}) = \varnothing$ because there is no s such that $V(s) = v$. If $v = 0$, then $V^{-1}(\{0\}) = \{1, 3, 5\}$ which is an event from the event space. If $v = 1$, then $V^{-1}(\{1\}) = \{2, 4, 6\}$ which is an event from the event space. Therefore, V is a discrete random variable.

THEOREM 1. If X and Y are two discrete random variable, then the mapping $Z : S \to \mathbb{R}$ defined by $Z(s) = X(s) + Y(s)$ is a discrete random variable.

Proof. First, since X and Y are discrete random variable, the $\operatorname{Im} Z$ will be discrete necessarily. Let $z \in \mathbb{R}$. If $\{Z = z\} = \varnothing$ and since $\varnothing$ is an event, $\{Z = z\}$ is an event. So assume that $\{Z = z\} \neq \varnothing$. We can show that

$$\{Z = z\} = \{s \in S \,:\, X(s) + Y(s) = z\} = \bigcup_{y \in \operatorname{Im} Y} \Big(\{X = z - y\} \cap \{Y = y\} \Big).$$

Since $\operatorname{Im} Y$ is discrete, the union on the left hand side is an infinite union of countable sets. We know that $\{X = z - y\}$ and $\{Y = y\}$ are events, for any choices of y and z. Therefore $\{X = z - y\} \cap \{Y = y\}$ is an event. Since the union of countably many events is still an event, we conclude that $\{Z = z\}$ is an event. $\square$

C.II Probability Mass Functions

We will use the following notations:

- $P(X = x)$ for $P(X^{-1}\{x\})$.
- Also, we will write $P(X \leq x)$ for $P(\{X \leq x\})$.

DEFINITION 2. Let X be a discrete random variable. The probability mass function (abbreviated pmf) p_X of X is the function defined by

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

EXAMPLE 3. In Example 1, the pmf is

$$p_X(0) = \frac{1}{8}, \quad p_X(1) = \frac{3}{8} = p_X(2), \textit{ and } \quad p_X(3) = \frac{1}{8}.$$

Note:

- For a discrete random variable X, the $\operatorname{Im} X$ is countable, so we can write $\operatorname{Im} X = \{x_1, x_2, \ldots\}$ in a list.
- Also, for any $x \notin \operatorname{Im} X$, we have $p_X(x) = 0$.

- In this case, we see that

$$\sum_{i=1}^{\infty} p_X(x_i) = \sum_{x \in \operatorname{Im} X} p_X(x) = P(S) = 1.$$

Theorem 2. Let $S = \{s_i : i \in I\}$ be a countable set of distinct real numbers where I is an index set, and let $\{\pi_i : i \in I\}$ be a collection of real numbers satisfying

$$\pi_i \geq 0, \quad \forall i \in I, \text{ and } \quad \sum_{i \in I} \pi_i = 1.$$

Then there exists a probability space $(S, \mathcal{A}, Q)$ and a discrete random variable $X : S \to \mathbb{R}$ such that the pmf of X is given by

$$p_X(s) = \begin{cases} \pi_i, & s = s_i \\ 0, & s \neq s_i. \end{cases}$$

Proof. Assume $S = \{s_i : i \in I\}$ be a countable set and assume the existence of the numbers π_i. Define $\mathcal{A} = \mathcal{P}(S)$ and define the probability measure Q to be $Q(\{A\}) = \sum_{i \text{i} A} \pi_i$ for $A \subset S$. Then $(S, \mathcal{A}, Q)$ is a probability space from the assumptions on π_i and the fact that S is discrete. Now, let $X : S \to \mathbb{R}$ be $X(s) = s$. Then, we have $\operatorname{Im} X = S$, $p_X(x) = 0$ if $x \notin S$ and if $x \in \operatorname{Im} S$, say $x = s_i$ for some i, then

$$p_X(x) = P(X = s_i) = P(\{s_i\}) = \pi_i. \qquad \square$$

.

Note: With this theorem, it is enough to say "Let X be a random variable taking the value s_i with probability π_i, for $i \in I$" and forget about the probability space.

Example 4. Let $S = \{0, 1, 2, \ldots\}$ and let $X : S \to \mathbb{R}$ be a discrete random variable with $\operatorname{Im} X = \{0, 1, 2, 3, 4, \ldots\}$. Define the function $p : S \to [0, 1]$ by $p(x) = c2^x/x!$, for $x = 0, 1, 2, \ldots$,

a) For what value of c is the function p a pmf?

b) Find $P(X = 0)$.

c) Find $P(X > 2)$.

Solution.

a) We must have $\sum_{x \in \operatorname{Im} X} p(x) = 1$. But

$$\sum_{x=0}^{\infty} \frac{2^x}{x!} = e^2$$

and therefore

$$1 = \sum_{x=0}^{\infty} p(x) = ce^2 \quad \Rightarrow \quad c = e^{-2}.$$

b) We have $P(X = 0) = p(0) = e^{-2}2^0/0! = e^{-2} \approx 0.1353$.

c) We have $P(X > 2) = 1 - P(X \leq 2)$. But,

$$P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2) = e^{-2}\left(\frac{2^0}{0!} + \frac{2^1}{1!} + \frac{2^2}{2!}\right) \approx 0.6767$$

and therefore $P(X > 2) = 1 - 0.6767 = 0.3233$. $\triangle$

C.III Functions of Discrete Random Variables

Let X be a discrete random variable and let $g : \mathbb{R} \to \mathbb{R}$ be a function. Then we can show that the function $Y : S \to \mathbb{R}$ defined by

$$Y(s) = g(X(s))$$

is a discrete random variable[1]. We usually write $Y = g(X)$.

Example 5. Let X be a discrete random variable.

a) Let $g(x) = ax + b$. Then $Y = g(X) = aX + b$. In this case, we have

$$P(Y = y) = P(aX + b = y) = P(X = (y - b)/a) = p_X((y - b)/a).$$

b) Let $g(x) = x^2$. Then $Y = g(X) = X^2$. In this case, for $y > 0$,

$$\begin{aligned} P(Y = y) = P(X^2 = y) = P(\{s \, : \, X(s) = \sqrt{y}\} \cup \{s \, : \, X(s) = -\sqrt{y}\}) \\ = P(X = \sqrt{y}) + P(X = -\sqrt{y}) \\ = p_X(\sqrt{y}) + p_X(-\sqrt{y}). \end{aligned}$$

Theorem 3. Let X be a discrete random variable and $g : \mathbb{R} \to \mathbb{R}$ be a function. Then the pmf of Y is

$$p_Y(y) = \sum_{x \in g^{-1}(y)} P(X = x)$$

for $y \in \mathbb{R}$.

Proof. By definition, for a given $y \in \operatorname{Im} Y$, we have

$$p_Y(y) = P(Y = y) = P(g(X) = y).$$

But,

$$\{s \in S \, : \, g(X(s)) = y\} = \{x \in \operatorname{Im} X \, : \, g(x) = y\} = g^{-1}(\{y\}) \cap \operatorname{Im} X.$$

Since $\overline{\operatorname{Im} X}$ does not contribute to the value of the probability, we can therefore write

$$p_Y(y) = P(\{x \in \operatorname{Im} X \, : \, g(x) = y\}) = P(g^{-1}(\{y\}) \cap \operatorname{Im} X) = \sum_{x \in g^{-1}(y)} P(X = x). \qquad \square$$

C.IV Expectation and Variance

Expected Value

Example 6. Consider a fair 6-faced die and the following game. After tossing the die, if the face lands on an even number, then you win 2 US dollars. But if the face lands on even, then you loose 1 US dollar. Would you like to play this game?

[1] Trick: $\operatorname{Im} Y$ is discrete because $\operatorname{Im} X$ is. Also, if $x \in \mathbb{R}$, then $Y^{-1}(\{x\}) = X^{-1}(g^{-1}(\{x\})) = X^{-1}(\operatorname{Im} X \cap g^{-1}(\{x\})) \cup X^{-1}(\overline{\operatorname{Im} X} \cap g^{-1}(\{x\})) = X^{-1}(\operatorname{Im} X \cap g^{-1}(\{x\})) \cup \varnothing = X^{-1}(\operatorname{Im} X \cap g^{-1}(\{x\})) = \cup_{a \in A} X^{-1}(\{a\})$, where $A = \operatorname{Im} X \cap X^{-1}(\{x\})$ is a countable set.

Solution. Let $S = \{⚀, ⚁, ⚂, ⚃, ⚄, ⚅\}$ and $P(A) = \frac{1}{6}$ for every atomic event A. Define the discrete random variable X in the following way:

$$X(s) = \begin{cases} 2 & \text{if } s \text{ is even,} \\ -1 & \text{if } s \text{ is odd.} \end{cases}$$

Then, we see that

$$E(X) = 2P(X = 2) + (-1)P(X = -1) = 2\left(\frac{1}{2}\right) - \left(\frac{1}{2}\right) = \frac{1}{2}\$. \qquad \triangle$$

DEFINITION 3. If X is a discrete random variable, then the <u>expectation</u> or <u>mean</u> of X is denoted by $E(X)$ and defined by

$$E(X) = \sum_{x \in \operatorname{Im} X} xP(X = x)$$

whenever this sum converges absolutely, meaning $\sum_{x \in \operatorname{Im} X} |xP(X = x)| < \infty$.

<u>Note:</u> Using the pmf of X, the expected value can be rewritten as

$$E(X) = \sum_{x \in \operatorname{Im} X} xp_X(x).$$

THEOREM 4. If X is a discrete random variable and $g : \mathbb{R} \to \mathbb{R}$, then

$$E(g(X)) = \sum_{x \in \operatorname{Im} X} g(x)P(X = x),$$

whenever the sum converges absolutely.

Proof. From the definition of the expected value, we have

$$\begin{aligned} E(g(X)) &= \sum_{y \in \operatorname{Im} g(X)} yP(g(X) = y) \\ &= \sum_{y \in \operatorname{Im} g(X)} y \sum_{x \in \operatorname{Im} X : g(x) = y} P(X = x) \\ &= \sum_{y \in \operatorname{Im} g(X)} \sum_{x \in \operatorname{Im} X : g(x) = y} g(x)P(X = x) \\ &= \sum_{x \in \operatorname{Im} X} g(x)P(X = x). \end{aligned} \qquad \square$$

An application of the previous theorem to the expectation gives the following result.

COROLLARY 5. Let X be a discrete random variable and let $a, b \in \mathbb{R}$.

a) If $P(X \geq 0) = 1$ and $E(X) = 0$, then $P(X = 0) = 1$.

b) $E(aX + b) = aE(X) + b$.

Proof. a) Assume that $P(X \geq 0) = 1$ and $E(X) = 0$. Notice that

$$P(X < 0) = 1 - P(X \geq 0) = 0.$$

Therefore, the values of X should all be positive or zero. By definition of $E(X)$, we have that $\sum_{x \in \operatorname{Im} X} xP(X = x) = 0$. Since every $x \in \operatorname{Im} X$ is positive, we must have $xP(X = x) = 0$ for $x > 0$ and therefore $P(X = x) = 0$ in this case. Now,

$$1 = P(X \geq 0) = P(X = 0) + P(X > 0) = P(X = 0) + 0 \quad \Rightarrow \quad P(X = 0) = 1.$$

b) This is a consequence of Theorem 4. $\square$

Variance

Another important statistics of a discrete random variable to know about is the variance.

DEFINITION 4. The variance of a discrete random variable X, denoted by $\mathrm{var}(X)$ is defined by

$$\mathrm{var}(X) = E\Big([X - E(X)]^2\Big).$$

Note: The standard deviation of a random variable is $\sqrt{\mathrm{var}(X)}$, usually denoted by σ.

Here is an easier expression to compute the variance of a discrete random variable.

THEOREM 6. Let X be a discrete random variable. Then,

$$\mathrm{var}(X) = E(X^2) - \mu^2,$$

where $\mu := E(X)$.

Proof. Let $\mu := E(X)$. From Theorem 4 with $g(x) = (x - \mu)^2$, we have

$$\begin{aligned}
\mathrm{var}(X) &= \sum_{x \in \mathrm{Im}\, X} g(x) P(X = x) = \sum_{x \in \mathrm{Im}\, X} (x - \mu)^2 P(X = x) \\
&= \sum_{x \in \mathrm{Im}\, X} \Big(x^2 - 2x\mu + \mu^2\Big) P(X = x) \\
&= \sum_{x \in \mathrm{Im}\, X} x^2 P(X = x) - 2\mu \sum_{x \in \mathrm{Im}\, X} x P(X = x) + \mu^2 P(X = x) \\
&= \sum_{x \in \mathrm{Im}\, X} x^2 P(X = x) - 2\mu \sum_{x \in \mathrm{Im}\, X} x P(X = x) + \mu^2 \sum_{x \in \mathrm{Im}\, X} P(X = x) \\
&= E(X^2) - 2\mu^2 + \mu^2 = E(X^2) - \mu^2
\end{aligned}$$ $\square$

EXAMPLE 7. The manager of an industrial plant is planning to buy a new machine of either type a or type b. If t denotes the number of hours of daily operations, the number of daily repairs Y_1 required to maintain a machine of type a is a random variable with mean and variance both equal to $t/10$. The number of daily repairs Y_2 for a machine of type b is a random variable with mean and variance both equal to $3t/25$. The daily cost of operating a is $C_a(t) = 10t + 30Y_1^2$; for b it is $C_b(t) = 8t + 30Y_2^2$. Assume that the repairs take negligible time and that each night the machines are tuned so that they operate essentially like new machines at the start of the next day. Which machine minimizes the expected daily cost if a workday consisting of

a) 10 hours.

b) 20 hours.

C.V Conditional Expectation and the Partition Theorem

When a condition is added, then the additional information will influence the probability $P(X = x)$ and therefore directly the expectation of X. We therefore introduce the conditional expectation.

Definition 5. Let X be a discrete random variable and B be an event with $P(B) > 0$. The conditional expectation of X given B is denoted by $E(X|B)$ and is defined by

$$E(X|B) = \sum_{x \in \operatorname{Im} X} xP(X = x|B),$$

whenever this sum converges absolutely.

We therefore have an analogous result to the Partition Theorem, but for the conditional expectation.

Theorem 7. Let X be a discrete random variable and B_1, B_2, ... be mutually exclusive events such that $\cup_{i=1}^{\infty} B_i = S$ and $P(B_i) > 0$ for each i. Then

$$E(X) = \sum_{i=1}^{\infty} E(X|B_i)P(B_i),$$

whenever the sum converges absolutely.

Proof. We can partition S as $S = \cup_{i=1}^{\infty} B_i$, where $B_i \cap B_j = \varnothing$, when $i \neq j$. Therefore,

$$\begin{aligned} E(X) = \sum_{x \in \operatorname{Im} X} xP(X = x) &= \sum_{x \in \operatorname{Im} X} x\Big(\sum_{j=1}^{\infty} P(X = x|B_j)P(B_j)\Big) \\ &= \sum_{j=1}^{\infty} \sum_{x \in X} xP(X = x|B_j)P(B_j) \\ &= \sum_{j=1}^{\infty} E(X|B_j)P(B_j). \end{aligned}$$

□

C.VI Examples of Discrete Random Variables

Let X be a discrete random variable.

Bernouilli distribution

A discrete random variable X has the Bernouilli distribution with parameter $p \in [0,1]$ if $\operatorname{Im} X = \{0,1\}$ and

$$P(X=1) = p \quad \text{and} \quad P(X=0) = 1-p.$$

Used Scenarios: The Bernouilli distribution is usually used to model experiment in which the outcome is "success" or "failure".

Binomial Distribution

Let n be an integer and $q \in [0,1]$. X has the binomial distribution with parameters n and q if $\operatorname{Im} X = \{0,1,2,\ldots,n\}$ and

$$P(X=k) = \frac{n!}{k!(n-k)!} q^k (1-q)^{n-k}, \quad k = 0,1,2,\ldots,n.$$

Used Scenarios: Experiments where the goal is to obtain a certain number of successes in n trials.

Example 8. There are $n=6$ machines to test if they are working properly or not. According to a recent survey, a machine is working properly in 75% of the time. What is the probability that 4 machines are working properly.

Solution. We have $q = 0.75$ and $n = 6$. Let X be the discrete random variable given the number of machines that are working properly. Then $X \sim Bi(6, 0.75)$. Therefore,

$$P(X=4) = \binom{6}{4}(0.75)^4(0.25)^2 = \frac{6!}{4!2!}(0.75)^4(0.25)^2 \approx 0.2966. \qquad \triangle$$

.

Poisson Distribution

Let $\lambda > 0$. X has the Poisson distribution if $\operatorname{Im} X = \{0,1,2,\ldots\}$ and

$$p_X(k) = \frac{1}{k!}\lambda^k e^{-\lambda}, \quad k = 0,1,2,\ldots.$$

Used Scenarios: Experiments where the goal is to obtain a certain number of successes in n trials, with n large.

Note: The parameter λ usually refers to the expected number of successes in an experiment (justified later when we introduce expectation of discrete random variables).

Example 9. Consider an experiment that consists of counting the number of α-particles given off in a 1-second interval by 1 gram of radioactive material. If we know from past experience that, on the average, 3.2 such α-particles are given off, what is a good approximation to the probability that no more than 2 α-particles will appear?

Solution. We think of a the surface of the material as a composition of a high number n of particular, that has $3.2/n$ chance of given off. We therefore can approximate the desire probability by a Poisson distribution with parameter $\lambda = nq = 3.2$. Then,

$$P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2) = \frac{e^{-3.2}3.2^0}{0!} + \frac{e^{-3.2}3.2^1}{1!} + \frac{e^{-3.2}3.2^2}{2!} \approx 0.3799.$$

$\triangle$

THEOREM 8. Let X be a discrete random variable which follows a binomial distribution with parameters n and q and let $\lambda = nq$. Then

$$p_X(k) \approx \frac{1}{k!}\lambda^k e^{-\lambda}, \quad k = 0, 1, 2, \ldots,$$

when n is large enough and q is small enough.

Negative Binomial Distribution

Let $q \in (0, 1)$ and $n \geq 0$ be an integer. Then X has the negative binomial distribution with parameters q and n if $\operatorname{Im} X = \{n, n+1, n+2, \ldots\}$ and

$$p_X(k) = \frac{(k-1)!}{(n-1)!(k-n)!}q^n(1-q)^{k-n}, \quad k = n, n+1, n+2, \ldots$$

Used-case Scenarios: Experiments where the goal is to find the probability of having the n-th success after k trials.

EXAMPLE 10. A geological study indicates that an exploratory oil well drilled in a particular region should strike oil with probability 0.2. Find the probability that the third oil strike comes on the fifth well drilled.

Solution. Let X be the number of strikes needed to obtain a third oil strike. In this case, we have $q = 0.2$ and $n = 3$. We are searching for $P(X = 5)$. Then

$$P(X = 5) = \frac{4!}{2!2!}(0.2)^3(0.8)^2 = 0.03072.$$

$\triangle$

.

Geometric Distribution

Let $q \in (0, 1)$. Then X has the geometric distribution with parameter q if $\operatorname{Im} X = \{1, 2, \ldots\}$ and

$$p_X(k) = (1-q)^{k-1}q, \quad k = 1, 2, 3, \ldots.$$

Used-case Scenarios: Experiments where the goal is to find the probability of the first success to occur within k tries.

EXAMPLE 11. An urn contains 10 red balls and 20 blue balls. Ball are randomly selected, one at a time, until a red one is obtained. If we assume that each selected ball is replaced before the next one is drawn, what is the probability that

a) exactly 3 draws are needed?

b) at least 6 draws are needed.

Solution. Let X be the discrete random variable counting the number of time needed to get a red ball. The random variable X follows a geometric distribution with parameter q, giving the probability of selecting a red ball.

Since the ball is replaced in the urn, the probability of selecting a red ball is always the same, that is 1/3/. Therefore, $q = 1/3$.

a) Let $k = 3$, so that $P(X = k) = (1 - 1/3)^2(1/3) = 4/27$.

b) What is $P(X \geq 6)$? Using the complement, this is $1 - P(X < 6)$. Therefore,

$$P(X \geq 6) = 1 - P(X = 1) - P(X = 2) - P(X = 3) - P(X = 4) - P(X = 5) \approx 0.8683.$$

△

Summary

The table below is a summary of the expected value and variance of each of the examples presented in this section.

Distribution	Expected Value	Variance
$B(q)$	q	$q(1-q)$
$B(n,q)$	nq	$nq(1-q)$
$\mathcal{P}(\lambda)$	λ	λ
$G(q)$	$1/q$	$(1-q)/q^2$
$NB(n,q)$	n/q	$n(1-q)/q^2$

Table 1: Table of Mean and Variance of different distributions