Chapter 2

Introduction to Discrete Random Variables

Wei-Yang Lin

Department of Computer Science

& Information Engineering

mailto:wylin@cs.ccu.edu.tw

- **2.1 Probabilities involving random variables**
- 2.2 Discrete random variables
- 2.3 Multiple random variables
- 2.4 Expectation

- 2.1 Probabilities involving random variables
- **2.2 Discrete random variables**
- 2.3 Multiple random variables
- 2.4 Expectation

- We say X is a **discrete random variable** if there exist distinct real numbers x_i such that

$$\sum_i \mathsf{P}(X = x_i) = 1$$

Probability mass function

- When X is a discrete random variable taking distinct values x_i, we define its **probability mass function** (pmf) by

$$p_X(x_i) := \mathsf{P}(X = x_i). \tag{1}$$

- Since $p_X(x_i)$ is a probability, it is a number satisfying

$$0 \leq p_X(x_i) \leq 1, \tag{2}$$

and

$$\sum_i p_X(x_i) = 1. \tag{3}$$

Uniform random variable

- We say X is uniformly distributed on $1, \ldots, n$ if

$$\mathsf{P}(X = k) = \frac{1}{n}, \quad k = 1, \ldots, n.$$

- In other words, its pmf takes only two values:

$$p_X(k) = \begin{cases} 1/n, & k = 1, \ldots, n, \\ 0, & \text{otherwise.} \end{cases}$$

The Poisson random variable

- The **Poisson** random variable is used to model many different physical phenomena.
- A random variable X is said to have a Poisson pmf with parameter $\lambda > 0$,denoted by $X \sim \text{Poisson}(\lambda)$, if

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

- To see that these probabilities sum to one, recall that the power series for e^z is

$$e^z = \sum_{k=0}^{\infty} \frac{z^k}{k!}.$$

Example 2.7

- The number of hits to a popular website during a 1-minute interval is given by a Poisson(λ) random variable.
- Find the probability that there is at least one hit between 3:00 am and 3:01 am if $\lambda = 2$.
- Then find the probability that there are at least 2 hits during this time interval.

Solution

Let X denote the number of hits. Then

$$\mathsf{P}(X \geq 1) = 1 - \mathsf{P}(X = 0) = 1 - e^{-\lambda} = 1 - e^{-2} \approx 0.865.$$

Similarly,

$$\begin{aligned} \mathsf{P}(X \geq 2) &= 1 - \mathsf{P}(X = 0) - \mathsf{P}(X = 1) \\ &= 1 - e^{-\lambda} - \lambda e^{-\lambda} \\ &= 1 - e^{-2} - 2e^{-2} \approx 0.594 \end{aligned}$$

Homework

- Problems 10, 11.

- 2.1 Probabilities involving random variables
- 2.2 Discrete random variables
- **2.3 Multiple random variables**
- 2.4 Expectation

Independence

- We say that X and Y are **independent random variables** if and only if

$$\mathsf{P}(X \in B, Y \in C) = \mathsf{P}(X \in B)\mathsf{P}(Y \in C)$$

for all sets B and C.

Example 2.8

- On a certain aircraft, the main control circuit on an autopilot fails with probability p.
- A redundant backup circuit fails independently with probability q.
- The aircraft can fly if at least one of the circuits is functioning.
- Find the probability that the aircraft cannot fly.

Solution

- We introduce two random variables, X and Y.
- We set $X = 1$ if the main circuit fails, and $X = 0$ otherwise.
- We set $Y = 1$ if the backup circuit fails, and $Y = 0$ otherwise.
- Then $\mathsf{P}(X = 1) = p$ and $\mathsf{P}(Y = 1) = q$.
- We assume X and Y are independent random variables.
- Using the independence of X and Y,

$$\mathsf{P}(X = 1, Y = 1) = \mathsf{P}(X = 1)\mathsf{P}(Y = 1) = pq.$$

Bernoulli random variable

- The random variables X and Y of the preceding example are said to be **Bernoulli**.
- To indicate the relevant parameters, we write $X \sim \text{Bernoulli}(p)$ and $Y \sim \text{Bernoulli}(q)$.
- Bernoulli random variables are good for modelling the result of an event having two possible outcomes.

Independence

- Given any finite number of random variables, say $X_1, \ldots, X_n$, we say they are **independent** if

$$\mathsf{P}\Big(\bigcap_{j=1}^{n} \{X_j \in B_j\}\Big) = \prod_{j=1}^{n} \mathsf{P}(X_j \in B_j),$$

  for all choices of the sets $B_1, \ldots, B_n$.

Example 2.9

- Let X, Y, and Z be the number of hits at a website on three consecutive days.
- Assume they are independent Poisson(λ) random variables.
- Find the probability that on each day the number of hits is at most n.

Solution

- The probability that on each day the number of hits is at most n is

$$\mathsf{P}(X \leq n, Y \leq n, Z \leq n).$$

- By independence, this is equal to

$$\mathsf{P}(X \leq n)\mathsf{P}(Y \leq n)\mathsf{P}(Z \leq n).$$

- Since the random variables are Poisson(λ), each factor is equal to

$$\mathsf{P}(X \le n) = \sum_{k=0}^{n} \mathsf{P}(X = k) = \sum_{k=0}^{n} \frac{\lambda^k}{k!} e^{-\lambda},$$

and so

$$\mathsf{P}(X \le n, Y \le n, Z \le n) = \Big(\sum_{k=0}^{n} \frac{\lambda^k}{k!} e^{-\lambda} \Big)^3.$$

Max and min problems

- Calculations similar to those in the preceding example can be used to find probabilities involving the maximum or minimum of several independent random variables.

Example 2.11

- For $i = 1, \ldots, n$, let X_i model the yield on the ith production run of an IC manufacturer.
- Assume yields on different runs are independent.
- Find the probability that the highest yield obtained is less than or equal to z, and find the probability that the lowest yield obtained is less than or equal to z.

Solution

- Observer that $\max(X_1, \ldots, X_n) \le z$ if and only if all of the X_k are less than or equal to z; i.e.,

$$\{\max(X_1, \ldots, X_n) \le z\} = \bigcap_{k=1}^{n} \{X_k \le z\}.$$

- It then follows that

$$\mathsf{P}(\max(X_1, \ldots, X_n) \le z) = \mathsf{P}\Big(\bigcap_{k=1}^{n} \{X_k \le z\}\Big) = \prod_{k=1}^{n} \mathsf{P}(X_k \le z),$$

where the second equation follows by independence.

- For the min problem, observer that $\min(X_1, \ldots, X_n) \le z$ if and only if at least one of the X_i is less than or equal to z; i.e.,

$$\{\min(X_1, \ldots, X_n) \le z\} = \bigcup_{k=1}^{n} \{X_k \le z\}.$$

Hence,

$$\begin{aligned}
\mathsf{P}(\min(X_i, \ldots, X_n) \le z) &= \mathsf{P}\Big(\bigcup_{k=1}^{n} \{X_k \le z\}\Big) \\
&= 1 - \mathsf{P}\Big(\bigcap_{k=1}^{n} \{X_k > z\}\Big) \\
&= 1 - \prod_{k=1}^{n} \mathsf{P}(X_k > z)
\end{aligned}$$

Joint probability mass functions

- The **joint probability mass function** of X and Y is defined by

$$p_{XY}(x_i, y_j) := \mathsf{P}(X = x_i, Y = y_j). \tag{4}$$

- An example is sketched in Figure 2.8.

- When X and Y take finitely many values, say $x_1, \ldots, x_m$ and $y_1, \ldots, y_n$, respectively, we can arrange the probabilities $p_{XY}(x_i, y_j)$ in the $m \times n$ matrix

$$\begin{bmatrix} p_{XY}(x_1, y_1) & p_{XY}(x_1, y_2) & \cdots & p_{XY}(x_1, y_n) \\ p_{XY}(x_2, y_1) & p_{XY}(x_2, y_2) & \cdots & p_{XY}(x_2, y_n) \\ \vdots & \vdots & \ddots & \vdots \\ p_{XY}(x_m, y_1) & p_{XY}(x_m, y_2) & \cdots & p_{XY}(x_m, y_n) \end{bmatrix}.$$

- Notice that the sum of the entries in the top row is

$$\sum_{j=1}^{n} p_{XY}(x_1, y_j) = p_X(x_1).$$

- In general, the sum of the entries in the ith row is $p_X(x_i)$, and sum of the entries in the jth column is $p_Y(y_j)$.
- It turns out that we can extract the **marginal probability mass function** $p_X(x_i)$ and $p_Y(y_j)$ from the joint pmf $p_{XY}(x_i, y_j)$ using the formulas

$$p_X(x_i) = \sum_j p_{XY}(x_i, y_j) \tag{5}$$

and

$$p_Y(y_j) = \sum_i p_{XY}(x_i, y_j). \tag{6}$$

Homework

- Problems 14, 15, 17, 18, 19, 20, 23, 25, 26.

- 2.1 Probabilities involving random variables
- 2.2 Discrete random variables
- 2.3 Multiple random variables
- **2.4 Expectation**

- If X is a discrete random variable taking distinct values x_i with probabilities $\mathsf{P}(X = x_i)$, we define the **expectation** or **mean** of X by

$$\mathsf{E}[X] := \sum_i x_i \mathsf{P}(X = x_i), \tag{7}$$

or, using the pmf notation,

$$\mathsf{E}[X] := \sum_i x_i p_X(x_i). \tag{8}$$

Example 2.21

Find the mean of a Bernoulli(p) random variable X.

Solution

Since X takes only the values $x_0 = 0$ and $x_1 = 1$,we can write

$$\mathsf{E}[X] = \sum_{i=0}^{1} i\mathsf{P}(X = i) = 0 \cdot (1 - p) + 1 \cdot p = p.$$

Expectation of a function of a random variable

- Given a random variable X, we will often have to define a new random variable by $Z := g(X)$.
- If we want to compute $\mathsf{E}[Z]$, it might seem that we first have to find pmf of Z.
- However, we can compute $\mathsf{E}[Z] = \mathsf{E}[g(X)]$ without actually finding the pmf of Z.
- The formula is
$$\mathsf{E}[g(X)] = \sum_i g(x_i) p_X(x_i). \tag{9}$$

- As an example of its use, we can write, for a constant a,

$$\mathsf{E}[aX] = \sum_i a x_i p_X(x_i) = a \sum_i x_i p_X(x_i) = a\mathsf{E}[X].$$

- In other words, the constant factors can be pulled out of the expectation.

- Also, it is a simple exercise to show that $\mathsf{E}[X + Y] = \mathsf{E}[X] + \mathsf{E}[Y]$.

- Thus, expectation is a **linear operator**; i.e., for constants a and b,

$$\mathsf{E}[aX + bY] = \mathsf{E}[aX] + \mathsf{E}[bY] = a\mathsf{E}[X] + b\mathsf{E}[Y]. \tag{10}$$

Moments

- The nth **moment**, $n \geq 1$, of a random variable X is defined to be $\mathsf{E}[X^n]$.
- The first moment of X is its mean, $\mathsf{E}[X]$.
- Letting $m = \mathsf{E}[X]$, we define the **variance** of X by

$$\mathsf{var}(X) := \mathsf{E}[(X - m)^2]. \tag{11}$$

Variance formula

- It is often convenient to use the variance formula

$$\mathsf{var}(X) = \mathsf{E}[X^2] - (\mathsf{E}[X])^2. \tag{12}$$

- To derive the variance formula, write

$$\begin{aligned}
\mathsf{var}(X) &:= \mathsf{E}[(X-m)^2] \\
&= \mathsf{E}[X^2 - 2mX + m^2] \\
&= \mathsf{E}[X^2] - 2m\mathsf{E}[X] + m^2, \quad \text{by linearity,} \\
&= \mathsf{E}[X^2] - m^2 \\
&= \mathsf{E}[X^2] - (\mathsf{E}[X])^2.
\end{aligned}$$

Example 2.28

Find the second moment and the variance of X if $X \sim \text{Bernoulli}(p)$.

Solution

- Since X takes only the values 0 and 1, it has the unusual property that $X^2 = X$.
- Hence, $\mathsf{E}[X^2] = \mathsf{E}[X] = p$.
- It follows that

$$\mathsf{var}(X) = \mathsf{E}[X^2] - (\mathsf{E}[X])^2 = p - p^2 = p(1-p).$$

Correlation and covariance

- The **correlation** between two random variables X and Y is defined to be $\mathsf{E}[XY]$.
- The correlation is important because it determines when two random variables are linearly related; namely, when one is a linear function of the other.

Example 2.36

- Let X have zero mean and unit variance, and put $Y := 3X$.
- Find the correlation between X and Y.

Solution

- First note that since X has zero mean, $\mathsf{E}[X^2] = \mathsf{var}(X) = 1$.
- Then write $\mathsf{E}[XY] = \mathsf{E}[X \cdot 3X] = 3\mathsf{E}[X^2] = 3$.

Cauchy-Schwarz inequality

- An important property of correlation is the **Cauchy-Schwarz inequality**, which says that

$$|\mathsf{E}[XY]| \leq \sqrt{\mathsf{E}[X^2]\mathsf{E}[Y^2]}, \tag{13}$$

where the equality holds if and only if X and Y linearly related.

Cauchy-Schwarz inequality

- To derive (13), let λ be a constant and write

$$\begin{aligned} 0 &\leq \mathsf{E}[(X-\lambda Y)^2] \\ &= \mathsf{E}[X^2 - 2\lambda XY + \lambda^2 Y^2] \\ &= \mathsf{E}[X^2] - 2\lambda \mathsf{E}[XY] + \lambda^2 \mathsf{E}[Y^2]. \end{aligned}$$

- Take

$$\lambda = \frac{\mathsf{E}[XY]}{\mathsf{E}[Y^2]},$$

then

$$0 \leq \mathsf{E}[X^2] - 2\frac{\mathsf{E}[XY]^2}{\mathsf{E}[Y^2]} + \frac{\mathsf{E}[XY]^2}{\mathsf{E}[Y^2]} = \mathsf{E}[X^2] - \frac{\mathsf{E}[XY]^2}{\mathsf{E}[Y^2]}.$$

Cauchy-Schwarz inequality

- The previous result can be rearranged to get

$$\mathsf{E}[(XY]^2 \leq \mathsf{E}[X^2]\mathsf{E}[Y^2]. \tag{14}$$

- Taking square roots yields (13).
- We can also show that if (13) holds with equality, then X and Y are linearly related.
- If (13) holds with equality, then so does (14).
- It follows that

$$\mathsf{E}[(X - \lambda Y)^2] = 0 \quad \Rightarrow \quad X = \lambda Y.$$

Correlation coefficient

- The **correlation coefficient** of random variables X and Y is defined to be the correlation of their normalized versions,

$$\rho_{XY} := \mathsf{E}\Big[\Big(\frac{X - m_X}{\sigma_X}\Big)\Big(\frac{Y - m_Y}{\sigma_Y}\Big)\Big]. \tag{15}$$

- Furthermore, $|\rho_{XY}| \leq 1$, with equality if and only if X and Y are related by a linear function plus a constant.

- A pair of random variables is said to be **uncorrelated** if their correlation coefficient is zero.

Covariance

- The **covariance** between X and Y is defined by

$$\mathsf{cov}(X,Y) := \mathsf{E}[(X - m_X)(Y - m_Y)]. \tag{16}$$

- With this definition, we can write

$$\rho_{XY} = \frac{\mathsf{cov}(X,Y)}{\sqrt{\mathsf{var}(X)\mathsf{var}(Y)}}.$$

- Hence, X and Y are uncorrelated if and only if their covariance is zero.

Exercise

- Let $X_1, X_2, \ldots, X_n$ be a sequence of uncorrelated random variables.
- More precisely, for $i \neq j$, X_i and X_j are uncorrelated.
- Please show that

$$\mathsf{var}\Big(\sum_{i=1}^{n} X_i\Big) = \sum_{i=1}^{n} \mathsf{var}(X_i).$$

Homework

- Problems 32, 34, 35, 36, 37, 45.