Chapter 3

Discrete Random Variables and Probability Distributions

Chaoyue Liu

Department of Mathematics and Statistics

DALHOUSIE UNIVERSITY

Outline

- Introduction to Random Variable
- PMFs and CDFs for Discrete Random Variables
- Expected Value and Variance of Discrete Random Variables
- Binomial Distribution
- Negative Binomial Distribution
- Hypergeometric Distribution
- Poisson Distribution

Introduction to Random Variables

- A random variable assigns a real number to each outcome in the sample space of a random experiment.

- **Definition:** For a given sample space S of an experiment, A random variable X is a function that associates a real number with each outcome in S. That is
  $X : S \to \mathbb{R}$

- examples:
  - Heads -> 1; Tails -> 0
  - the sum of two dice, e.g. X = 4 is the event {(1,3), (2,2), (3,1)}

Notation: we commonly use capital letters (such as X or Y) to denote random variables, and lower case letters (such as x or y) to denote the corresponding values. So, {X = x} is the event that the random variable X takes the specific value x.

Example

- **Experiment:**
  - A process that produces random outcomes
  - e.g. tossing two coins
- **Sample Space:**
  - a set of all possible outcomes produced by an experiment
  - e.g. S={HH, HT, TH, TT}
- **Event:**
  - a subset of the sample space
  - e.g. Let A denote the event that we get at most 1 heads.
- **Random Variable:**
  - Random variable map outcomes to numbers
  - e.g. X = number of heads

Two types of Random Variables

- **Discrete Random Variable: a random variable that has only a specified finite or countably list of possible values, i.e.** $x \in \{x_1, x_2, ...\}$.

  - countable values, e.g.

    - number of customers waiting in line
    - rolling a dice
    - students' grade level

  - Bernoulli Random Variable: the simplest kind of random variable whose only possible values are 0 and 1. e.g. True/False, Head/Tail, Success/Fail …

- **Continuous Random Variables: a random variable which can infinitely many possible values in an interval, i.e.** $x \in \{x : a < x < b; a, b \in \mathbb{R}\}$

  - uncountable value, e.g.

    - waiting time for a bus
    - height, weight

Probability Distributions for Discrete Random variables

- **Definition:**

The probability distribution or probability mass function (pmf) of a discrete random variable X is defined by

$$p(x) = P(X = x) = P(\text{all } s \in S : X(s) = x)$$

which represents the probability that the random variable X equals a specific value x.

Properties of pmf:

- $p(x) = P(X = x) \geq 0$

- $\sum_{\text{all x}} p(x) = 1$

- For any event $A \subset S$

$$P(A) = \sum_{\text{all } s \in A : X(s) = x} p(x)$$

Cumulative Distribution Function for Discrete random variables

- **Definition:** The cumulative distribution function (cdf) $F(x)$ of a discrete random variable X with pdf p(x) is defined by

$$F(x) = P(X \leq x) = \sum_{y: y \leq x} p(y)$$

- **Properties of cdf:**

  - The cumulative distribution function, F(x), is a non-decreasing function ↔ for any $x_1 < x_2, F(x_1) \leq F(x_2)$
  - $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to \infty} F(x) = 1$
  - For any two real numbers $a \leq b$,

    $P(a \leq X \leq b) = F(b) - F(a-)$, where "$a-$" is the largest possible value of X that is strictly less than a.

pmf vs cdf

- **pmf** : $P(X = x)$
- **cdf** : $P(X \leq x)$
- **Let X be a discrete RV with the following pmf:**

X	x_1	x_2	…	x_n
$p(x)$	p_1	p_2	…	p_n

- **The cdf of X is**

$$F(x) = \begin{cases} 0, & x < x_1 \\ p_1, & x_1 \leq x < x_2 \\ p_1 + p_2, & x_2 \leq x < x_3 \\ p_1 + p_2 + p_3, & x_3 \leq x < x_4 \\ \vdots & \vdots \\ p_1 + p_2 + \cdots + p_{n-1}, & x_{n-1} \leq x < x_n \\ 1, & x_n \leq x \end{cases}$$

- $p(x) = F(x) - F(x-)$ **and** $F(x) = P(X \leq x) = \sum_{y \leq x} p(y)$

Example

- **Let x be a random variable with the following mass function**

X	-3	-2	-1	0	1	4	6
P(X=x)	0.13	0.16	0.17	0.2	0.16	0.11	k

- **Calculate**

1) k

2) $P(X \leq -2)$

3) The cdf of X, "F(x)"

4) $P(X > 0)$

5) $P(-2 \leq X \leq 1)$

6) $P(-2 \leq X < 1)$

7) $P(-2 < X \leq 1)$

8) $P(-2 < X < 1)$

Expected Value and Variance of Discrete Random Variables

Chaoyue Liu

Department of Mathematics and Statistics

DALHOUSIE UNIVERSITY

Expected Values of Discrete Random Variables

- **Definition:** Let X be a discrete RV with a set of possible values $D = \{x_1, x_2, \ldots, x_n\}$ and pmf $p(x)$. The expected or mean value of X, denoted as $E(X)$ or μ_x or μ is

$$E(X) = \mu = \sum_{x \in D} x p(x)$$

- **The Expected Value of a random variable gives a measure of the center of X**
- **Example:** The following is the distribution of the number credit cards, X , a person possesses. What is the expected number of credit cards that a randomly selected person will possess?

X	0	1	2	3	4	5
p(x)	0.08	0.28	0.38	0.16	0.06	0.04

Expected Value of functions of random variables

- If X is a discrete rv with a pmf $p(x)$ and $h(x)$ is a function of X, then the expected value of $h(X)$, denoted by $E[h(X)]$ or $\mu_{h(X)}$, is computed by

$$E[h(X)] = \sum_{\text{all } x} h(x)p(x)$$

- The k^{th} moment of a random variable X is defined as $E(X^k)$. Thus, the mean is the first moment of X.

- **Properties of expected value**
  - If c is a constant, then $E(c) = c$.
  - Constants can be factored out of expected values:
    $E[c \cdot g(X)] = c \cdot E[g(X)]$
  - The expected value of a sum is equal to the sum of expected values:
    $\mathrm{E}\left[c_1 g_1(X) + c_2 g_2(X)\right] = c_1 \mathrm{E}\left[g_1(X)\right] + c_2 \mathrm{E}\left[g_2(X)\right]$

The Variance and Standard Deviation

- Definition: Let X be a discrete rv with a pmf $p(x)$ and expected value μ. Then the **variance** of X, denoted by $V(X)$, or σ_x^2 or σ^2, is

$$V(X) = E\left[(X-\mu)^2\right] = \sum_{\text{all } x} (x-\mu)^2 p(x)$$

- The **standard deviation (SD)** of X is $\sigma = \sqrt{\sigma^2}$. Note: $\sigma^2, \sigma \geq 0$.
- The variance and standard deviation give the **measure of spread** for random variables.
- A **shortcut** Formula for σ^2: $V(X) = \sigma^2 = E(X^2) - \mu^2$
- Exercise: show the proof of the shortcut?

The expected value and variance of a linear function of X

- If the function $h(x)$ is a **linear function** of X, i.e. $h(x) = ax + b$, then the expected value and the variance of $h(x)$ can be easily computed as
  - $E[aX + b] = aE[X] + b$
  - $V(aX + b) = a^2V(X)$
  - $SD(aX + b) = |a| \cdot SD(X)$

- Exercise: suppose $E(X) = 5$, $V(X) = 10$ and $h(X) = -4X + 3$, calculate the expected value, variance and standard deviation of $h(x)$?

Example

Suppose there is a non-balance coin which has the probability $P(T) = 2P(H)$. Let's toss this non-balance coin two times independently and let X be the number of heads we observed. Find:

- 1) the pmf and cdf of X
- 2) the expected number of heads and its variance
- 3) to play this game, you should pay $10, but you will gain $20 every time it turns head, what is your expected gain and the corresponding variance?

Discrete Probability Distributions

Chaoyue Liu
Department of Mathematics and Statistics

DALHOUSIE UNIVERSITY

Some important Discrete Probability Distributions

- **Binomial Distribution**
- **Negative Binomial Distribution**
- **Hypergeometric Distribution**
- **Poisson Distribution**

Binomial Distribution

- **A binomial experiment is a random experiment that satisfies the following assumptions:**
  - The experiment consists of n repeated trials.
  - Each trail can only result in two possible outcomes: a success or a failure.
  - The probability of success, denoted by p, is the same on every trial.
  - The trials are independent.
- **Binomial random variable: X = number of successes observed for the n trials in a binomial experiment.**
- **We say that X follows a Binomial distribution, denoted by** $X \sim Bin(n, p)$ **where n, p are parameters.**

- **Example:** Tossing a fair coin 5 times, let X be the number of heads, describe the probability distribution of X?

DALHOUSIE UNIVERSITY

Properties of Binomial Distribution

- **If $X \sim Bin(n, p)$, then we have:**
  - The probability mass function (pmf) of X:

$$b(x; n, p) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & x = 0, 1, 2, \ldots, n \\ 0 & \text{otherwise} \end{cases}$$

  - Expected value (Mean) of X: $E(X) = np$
  - Variance of X: $V(X) = np(1-p)$

- **Example:** Tossing a fair coin n times, let X be the number of heads, what is the probability of X=x?

Example

- **In a restaurant an average of 3 out of every 5 customers ask for water with their meal. A random sample of 10 customers is selected. Find the probability that**
  - exactly 6 ask for water with their meal
  - less than 9 ask for water with their meal
  - What is the expected number of asking for water and what is the corresponding variance?

Table A.1 Cumulative Binomial Probabilities $B(x; n, p) = \sum_{y=0}^{x} b(y; n, p)$

a. $n = 5$

		p														
		0.01	**0.05**	**0.10**	**0.20**	**0.25**	**0.30**	**0.40**	**0.50**	**0.60**	**0.70**	**0.75**	**0.80**	**0.90**	**0.95**	**0.99**
	0	.951	.774	.590	.328	.237	.168	.078	.031	.010	.002	.001	.000	.000	.000	.000
	1	.999	.977	.919	.737	.633	.528	.337	.188	.087	.031	.016	.007	.000	.000	.000
x	2	1.000	.999	.991	.942	.896	.837	.683	.500	.317	.163	.104	.058	.009	.001	.000
	3	1.000	1.000	1.000	.993	.984	.969	.913	.812	.663	.472	.367	.263	.081	.023	.001
	4	1.000	1.000	1.000	1.000	.999	.998	.990	.969	.922	.832	.763	.672	.410	.226	.049

b. $n = 10$

		p														
		0.01	**0.05**	**0.10**	**0.20**	**0.25**	**0.30**	**0.40**	**0.50**	**0.60**	**0.70**	**0.75**	**0.80**	**0.90**	**0.95**	**0.99**
	0	.904	.599	.349	.107	.056	.028	.006	.001	.000	.000	.000	.000	.000	.000	.000
	1	.996	.914	.736	.376	.244	.149	.046	.011	.002	.000	.000	.000	.000	.000	.000
	2	1.000	.988	.930	.678	.526	.383	.167	.055	.012	.002	.000	.000	.000	.000	.000
	3	1.000	.999	.987	.879	.776	.650	.382	.172	.055	.011	.004	.001	.000	.000	.000
	4	1.000	1.000	.998	.967	.922	.850	.633	.377	.166	.047	.020	.006	.000	.000	.000
x	5	1.000	1.000	1.000	.994	.980	.953	.834	.623	.367	.150	.078	.033	.002	.000	.000
	6	1.000	1.000	1.000	.999	.996	.989	.945	.828	.618	.350	.224	.121	.013	.001	.000
	7	1.000	1.000	1.000	1.000	1.000	.998	.988	.945	.833	.617	.474	.322	.070	.012	.000
	8	1.000	1.000	1.000	1.000	1.000	1.000	.998	.989	.954	.851	.756	.624	.264	.086	.004
	9	1.000	1.000	1.000	1.000	1.000	1.000	1.000	.999	.994	.972	.944	.893	.651	.401	.096

Negative Binomial Distribution

- **Negative Binomial experiment:**
  - each trial has only two possible outcomes: success/failure
  - The probability of success is p for all trials
  - repeatedly perform independent trial until having r success.
- **Negative Binomial random variable: X = the number of failures before r^{th} success.**
- **Then we say that X follows a Negative Binomial distribution, $X \sim NB(r, p)$, where r, p are parameters.**

- **Example:** if we keep tossing a fair coin until we get 5 times of heads, let X be the number of tails we got, what is the probability distribution of X?

Properties of negative binomial distribution

- **If $X \sim NB(r, p)$ then we have**
- **the pmf of X is**

$$nb(x; r, p) = \binom{x + r - 1}{r - 1} p^r (1 - p)^x \quad x = 0, 1, 2, \ldots$$

- **Expected value:** $E(X) = r\frac{1-p}{p}$
- **Variance:** $V(X) = r\frac{1-p}{p^2}$
- When $r = 1$ in $NB(r, p)$, then the negative binomial distribution reduces to geometric distribution (X: number of failures before first success occurs).

Relationships with Binomial

- **The binomial distribution and negative binomial distribution are trying to answer somewhat opposite questions:**
- **For the binomial distribution:**
  - the total number of trials is fixed.
  - the number of successes is random.
- **For the negative binomial distribution:**
  - the total number of successes is fixed.
  - the number of trials is random.

Hypergeometric Distribution

- **The assumptions to a hypergeometric distribution:**
  - A population of N objects
  - M objects are characterized as success and $N - M$ objects are characterized as failure
  - Pick randomly n objects without replacement (without replacement, means once we pick an object, we do not put it back to the population.)
- **Hypergeometric random variable: X = the number of successes**
- **Then we say X follows the hypergeometric distribution**

  $X \sim h(n, M, N)$**, where** n, M, N **are parameters**

- **Example:** an urn contains a total of N balls, where M of the balls are red and the remaining N-M balls are blue. Suppose we draw n times without replacement from the urn and X = number of red balls we drew, then X ~ h(n, M, N)

Properties of hypergeometric distribution

- **If $X \sim h(n, M, N)$, then we have**
- **pmf of X is**

$$P(X = x) = h(x; n, M, N) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}, \max(0, n - N + M) \le x \le \min(n, M)$$

Expected value of X is $E(X) = n \times \frac{M}{N}$

Variance of X is $V(X) = (\frac{N-n}{N-1}) \times n \times \frac{M}{N} \times (1 - \frac{M}{N})$

Relationship with Binomial Distribution

- **Binomial Distribution: draw with replacement**
  - Given a population of N objects
  - Randomly draw n times (n repeated trials)
  - M of the objects are characterized as success (the probability of success: $p = \frac{M}{N}$)
  - draw with replacement ($p(success)$ is same for all trials)
  - X = number of success

- **Hypergeometric distribution: draw without replacement**
- **In binomial distribution, each draw is independent (p is same for every trial); in hypergeometric distribution, each draw is not independent (p changes)**

Example

- **Suppose we randomly select 5 cards without replacement from a deck of 52. What is the probability of getting exactly 2 red cards?**

- **If we draw 5 times with replacement** (every time we draw a card, we note down the color and then put it back to the deck)**, what is the probability of observing 2 red cards?**

Poisson Distribution

- **Poisson distribution can be used to model the number of events occurred during a period time.**
- **Definition: A random variable X has a Poisson distribution, with parameter $\lambda > 0$, if its probability mass function is given by**
- $P(X = x) = pois(x; \lambda) = \dfrac{e^{-\lambda}\lambda^{x}}{x!},$ **for** $x = 0, 1, 2, \ldots$
- **denoted by $X \sim Poisson(\lambda)$.**
- Note that e is a mathematical constant ($e \approx 2.718$)
- λ = mean number of occurrences of the event over the interval
- **Expected value and Variance**
  - If $X \sim Poisson(\lambda)$, then $E(X) = V(X) = \lambda$

Example

- **Suppose that the number of typing errors per page has Poisson distribution with average 6 typing errors.**
  1. What is the probability that in a given page, the number of typing errors will be 7?
  2. What is the probability that in a given page, the number of typing errors will be at least 2?
  3. What is the probability that in 2 pages there will be 10 typing errors?
  4. What is the probability that in a half page there will be no typing errors?

Relationship with Binomial

- **The Poisson distribution as a limit Binomial distribution $(n \to \infty, p \to 0)$:**
- **Poisson approximation to the Binomial**
  - $b(x; n, p)$ is difficult to compute when n is large.
  - Binomial can be approximated by a Poisson distribution when n is large (n>50) and p is small (p <5/n).

$$b(x; n, p) \approx pois(x; np)$$

- **Example:** Each computer in a cluster work properly with a probability of 99.9%. The cluster has 100 computers. What is the probability that 2 computers are not working properly.