# STAT1011
# Introduction to Statistics
## Tutorial 4

Chan Chun Ho, Benjamin*†

October 12, 2017

**Abstract**

It aims to introduce basic concepts of discrete probability distributions and incorporate several important discrete probability models for illustration. However, it is by no means a comprehensive list of all discrete distributions. Some exercises are provided for students to practice. The proofs in Section 3 are beyond the scope of this course and they serve the purpose of completeness. Some materials does credit to former TA George and some are extracted from classic textbook "Statistical Inference" (Casella and Berger) used in STAT4003.

Suggestion for future study: If students are interested in statistical theory, they should consider to take STAT2001, 2006 and 4003. There are other popular Statistics courses, including but not limited to STAT3008 Regression and 4005 Time Series. Statistics and Risk Management Science are closely related subjects. RMSC2001 is a good starting point.

# Notations and Definitions

- Summation: $\sum_{i=1}^{n} x_i = x_1 + x_2 + \cdots + x_n$
- Factorial: $n! = n \cdot (n-1) \cdots 2 \cdot 1$; Combination: $\binom{n}{r} = \frac{n!}{r!\,(n-r)!}$
- Set of real numbers: $\mathbb{R} = (-\infty, \infty)$; Set of positive integers: $\mathbb{Z}^+ = \{1, 2, ...\}$
- Set membership: $x \in A$ means "$x$ is an element of the set $A$".
- Sample space: $S$ is the set of all possible outcomes of experiments.
- Expectation: $E(\cdot)$; Variance: $Var(\cdot)$; Standard deviation: $SD(\cdot)$
- Probability distribution: A tilde ($\sim$) means "has the probability distribution of".
- Parameter(s): Population characteristic(s) that can be set to different values to produce different probability distributions

*For enquiry, please email to 1155049861@link.cuhk.edu.hk.

†Personal profile: www.linkedin.com/in/benjamin-chan-chun-ho

# 1 Introduction

## 1.1 Discrete Random Variables

- A **random variable** is a function that maps each element in a sample space $S$ into a real number $x \in \mathbb{R}$.
- A random variable is said to be **discrete** if $S$ is a **countable** set.
- A discrete random variable can have **a finite or an infinite number of values**.

Example: A coin is flipped three times. The sample space is

$$S = \{(H,H,H),(H,H,T),(H,T,H),(H,T,T),(T,H,H),(T,H,T),(T,T,H),(T,T,T)\}.$$

Let $X$ be the total number of heads to land up. Then, $X$ can only be 0, 1, 2 or 3. Thus, $X$ is a discrete random variable.

## 1.2 Discrete Probability Distributions

- For a discrete random variable $X$, the **probability mass function (pmf)** of $X$ is
$$p(x) = P(X = x).$$
- Note that $0 \leq p(x) \leq 1$ and $\sum_{x \in S} p(x) = 1$.
- The **cumulative distribution function (cdf)** of $X$ is
$$F(x) = P(X \leq x) = \sum_{k \leq x,\ k \in S} p(k).$$

## 1.3 Mean and Variance of Discrete Random Variables

- The **mean**, or **expectation**, or **expected value** of $X$ and $g(X)$ are
$$\mu = E(X) = \sum_{x \in S} x\ p(x)$$
$$E[g(X)] = \sum_{x \in S} g(x)\ p(x).$$
- The **variance** of $X$ is
$$\sigma^2 = Var(X) = E(X-\mu)^2 = \sum_{x \in S} (x-\mu)^2 p(x) = \big[\sum_{x \in S} x^2\ p(x)\big] - \mu^2 = E(X^2) - \mu^2$$
- The **standard deviation** of $X$ is simply $\sigma = SD(X) = \sqrt{Var(X)}$.

# 2 Common Discrete Distributions

## 2.1 Bernoulli Distribution

A **Bernoulli random variable** $X$ with parameter $0 < p < 1$ has a pmf as

$$p(0) = 1 - p \quad , \quad p(1) = p$$

A Bernoulli random variable corresponds to a trial where the outcome is either a success ($X = 1$) or a failure ($X = 0$). The probability of success is $p$.

- $E(X) = p$ (Proofs in Section 3)
- $Var(X) = p(1-p)$

## 2.2 Binomial Distribution

A **binomial random variable** $X$ with parameters $(n, p)$, where $n \in \mathbb{Z}^+$ and $0 < p < 1$, has a pmf as

$$p(x) = \binom{n}{x} p^x \ (1-p)^{n-x} \quad , \quad x = 0, 1, ..., n$$

A binomial random variable corresponds to the number of successes in $n$ independent Bernoulli trials, each with probability of success $p$. Note that a Bernoulli distribution is just a binomial distribution with parameters $(1, p)$.

**Remark 1.** In a sequence of $n$ identical, independent Bernoulli trials, each with success probability $p$, define the Bernoulli random variables $X_1, ..., X_n$ as in Section 2.1. The random variable $Y = \sum_{i=1}^{n} X_i$ has the binomial$(n, p)$ distribution.

- $E(X) = np$ (Proofs in Section 3)
- $Var(X) = np(1-p)$

## 2.3 Multinomial Distribution

A **multinomial distribution** with parameters $(n, p_1, ..., p_k)$, where $n \in \mathbb{Z}^+$, $0 < p_i < 1$ for $i = 1, ..., k$ and $\sum_{i=1}^{k} p_i = 1$, has a pmf as

$$p(x_1, ..., x_k) = \frac{n!}{x_1! \cdots x_k!} \, p_1^{x_1} \cdots \ p_k^{x_k} \quad , \quad \sum_{i=1}^{k} x_i = n$$

It corresponds to obtaining $x_1, \cdots, x_k$ of the event each with probability $p_1, \cdots, p_k$ of occurring respectively. A multinomial distribution is a generalization of a binomial distribution, where each trial has $k$ possible outcomes instead of 2.

## 2.4 Poisson Distribution

A **Poisson random variable** $X$ with parameter $\lambda > 0$ has a pmf as

$$p(x) = \frac{\lambda^x}{x!}e^{-\lambda} \quad , \quad x = 0, 1, 2, ...$$

A Poisson distribution with parameter $\lambda$ can be used to approximate a binomial distribution with parameters $(n, p)$ when $n$ is large and $p$ is small such that $np = \lambda$.

**Remark 2.** If $X \sim \text{binomial}(n, p)$ and $Y \sim \text{Poisson}(\lambda)$, for large $n$ and small $p$, with $np = \lambda$, then $P(X = x) \approx P(Y = x)$.[1]

- $E(X) = \lambda$ (Proofs in Section 3)
- $Var(X) = \lambda$

## 2.5 Hypergeometric Distribution

A **hypergeometric random variable** with parameters $(N_1, N_2, n)$ has a pmf as

$$p(x) = \frac{\binom{N_1}{x} \cdot \binom{N_2}{n-x}}{\binom{N}{n}} \quad , \quad x = \max(0, n - N_2), ..., \min(N_1, n) \ , \ N = N_1 + N_2$$

A hypergeometric random variable corresponds to the number of successes in a sample of size $n$ from a finite population of size $N$, where the sampling is without replacement.

If samples of size $n$ are drawn from a population of size $N$, there are $\binom{N}{n}$ combinations. If $x$ of the samples are successes, there are $\binom{N_1}{x}$ combinations, hence $n - x$ of the samples are failures, there are $\binom{N_2}{n-x}$ combinations. Note the range of $X$. Naturally, $0 \leq x \leq n$. Moreover, $\binom{n}{r}$ is defined only if $r \leq n$, so $x \leq N_1$ and $n - x \leq N_2$. Combining ineqalities will yield $x = \max(0, n - N_2), ..., \min(N_1, n)$.

- $E(X) = n(\frac{N_1}{N})$
- $Var(X) = n(\frac{N_1}{N})(\frac{N_2}{N})(\frac{N-n}{N-1})$

Difference between binomial distribution and hypergeometric distribution

Note that both binomial distribution and hypergeometric distribution can be used to model the number of successes. The difference lies in with or without replacement. With replacement, the draws are independent and may be modeled by binomial distribution. Without replacement, the draws are not independent and may be modeled by hypergeometric distribution.

[1](Optional) Interested students may search "moment generating function" and "Poisson approximation".

# 3 (Optional) Proofs

**Theorem 1.** If $X \sim \text{Bernoulli}(p)$, $E(X) = p$ and $Var(X) = p(1-p)$.

*Proof.*

$$E(X) = \sum_{x=0}^{1} x\ p(x) = 0 \cdot p(0) + 1 \cdot p(1) = 0 \cdot (1-p) + 1 \cdot p = p$$

$$Var(X) = \big[\sum_{x=0}^{1} x^2\ p(x)\big] - \mu^2 = 0^2 \cdot (1-p) + 1^2 \cdot p - p^2 = p - p^2 = p(1-p)$$

□

**Theorem 2.** If $X \sim \text{binomial}(n,p)$, $E(X) = np$ and $Var(X) = np(1-p)$.

*Proof.*

$$E(X) = \sum_{x=0}^{n} x \binom{n}{x} p^x (1-p)^{n-x} = \sum_{x=1}^{n} x \binom{n}{x} p^x (1-p)^{n-x} \quad \text{(the } x = 0 \text{ term is 0).}$$

Using the identity

$$x \binom{n}{x} = n \binom{n-1}{x-1},$$

we have

$$\begin{aligned} E(X) &= \sum_{x=1}^{n} n \binom{n-1}{x-1} p^x (1-p)^{n-x} \\ &= \sum_{y=0}^{n-1} n \binom{n-1}{y} p^{y+1} (1-p)^{n-(y+1)} \quad \text{(Substitute } y = x-1\text{)} \\ &= np \sum_{y=0}^{n-1} \binom{n-1}{y} p^y (1-p)^{n-1-y} \\ &= np, \end{aligned}$$

since the last summation must be 1, being the sum over all possible values of a binomial $(n-1,p)$ pmf.

Write

$$x^2\binom{n}{x} = x\frac{n!}{(x-1)!(n-x)!} = xn\binom{n-1}{x-1}.$$

$$\begin{aligned}
E(X^2) &= \sum_{x=0}^{n} x^2\binom{n}{x}p^x(1-p)^{n-x} \\
&= \sum_{x=1}^{n} x^2\binom{n}{x}p^x(1-p)^{n-x} \quad \text{(the } x=0 \text{ term is 0)} \\
&= n\sum_{x=1}^{n} x\binom{n-1}{x-1}p^x(1-p)^{n-x} \\
&= n\sum_{y=0}^{n-1}(y+1)\binom{n-1}{y}p^{y+1}(1-p)^{n-1-y} \quad \text{(Substitute } y=x-1\text{)} \\
&= np\sum_{y=0}^{n-1} y\binom{n-1}{y}p^y(1-p)^{n-1-y} + np\sum_{y=0}^{n-1}\binom{n-1}{y}p^y(1-p)^{n-1-y}.
\end{aligned}$$

The first sum is equal to $(n-1)p$ since it is the mean of a binomial$(n-1,p)$. The second sum equals 1. Hence,

$$E(X^2) = n(n-1)p^2 + np.$$

So,

$$Var(X) = n(n-1)p^2 + np - (np)^2 = -np^2 + np = np(1-p).$$

□

**Theorem 3.** If $X \sim \text{Poisson}(\lambda)$, $E(X) = Var(X) = \lambda$.

*Proof.*

$$E(X) = \sum_{x=0}^{\infty} x \frac{\lambda^x}{x!} e^{-\lambda} = \sum_{x=1}^{\infty} x \frac{\lambda^x}{x!} e^{-\lambda} \quad \text{(the } x = 0 \text{ term is 0).}$$

Using the Taylor series expansion of $e^y$,

$$e^y = \sum_{i=0}^{\infty} \frac{y^i}{i!},$$

$$\begin{aligned} E(X) &= \lambda e^{-\lambda} \sum_{x=1}^{\infty} \frac{\lambda^{x-1}}{(x-1)!} \\ &= \lambda e^{-\lambda} \sum_{y=0}^{\infty} \frac{\lambda^y}{y!} \quad \text{(substitute } y = x - 1\text{)} \\ &= \lambda e^{-\lambda} e^{\lambda} \\ &= \lambda \end{aligned}$$

$$\begin{aligned} E(X^2) &= \sum_{x=0}^{\infty} x^2 \frac{\lambda^x}{x!} e^{-\lambda} \\ &= \sum_{x=1}^{\infty} x^2 \frac{\lambda^x}{x!} e^{-\lambda} \quad \text{(the } x = 0 \text{ term is 0)} \\ &= \lambda \sum_{x=1}^{\infty} x \frac{\lambda^{x-1}}{(x-1)!} e^{-\lambda} \\ &= \lambda \sum_{y=0}^{\infty} (y+1) \frac{\lambda^y}{y!} e^{-\lambda} \quad \text{(substitute } y = x - 1\text{)} \\ &= \lambda \sum_{y=0}^{\infty} y \frac{\lambda^y}{y!} e^{-\lambda} + \lambda \sum_{y=0}^{\infty} \frac{\lambda^y}{y!} e^{-\lambda} \\ &= \lambda^2 + \lambda \end{aligned}$$

The first sum is equal to $\lambda$ since it is the mean of a Poisson($\lambda$). The second sum equals 1. So,

$$Var(X) = \lambda^2 + \lambda - \lambda^2 = \lambda.$$

□

# 4 Exercises

**Exercise 4.1.** The number of calls $X$ to arrive at a switchboard during any 1-minute period is a random variable and has the following pmf. Find the mean and variance of $X$.

| $x$ | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| $p(x)$ | 0.1 | 0.2 | 0.4 | 0.2 | 0.1 |

**Solution.** Let $\mu = E(X)$ be the mean of $X$. Thus,
$E(X) = 0(0.1) + 1(0.2) + 2(0.4) + 3(0.2) + 4(0.1) = 2$
$Var(X) = (0)^2(0.1) + (1)^2(0.2) + (2)^2(0.4) + (3)^2(0.2) + (4)^2(0.1) - 2^2 = 1.2$

---

**Exercise 4.2.** Suppose that the random variable $X$ is equal to the number of hits obtained by a certain baseball player in his next 3 at bats. If $P(X = 1) = 0.3$, $P(X = 2) = 0.2$, and $P(X = 0) = 3 \cdot P(X = 3)$, find $E(X)$.

**Solution.** Since $P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) = 1$, we must have $4 \cdot P(X = 3) + 0.5 = 1$, implying that $P(X = 3) = 0.125$, $P(X = 0) = 0.375$. Hence, $E(X) = 1(0.3) + 2(0.2) + 3(0.125) = 1.075$.

---

**Exercise 4.3.** A fair coin is flipped. If the head lands up, the coin will be flipped twice, and if the tail lands up, the coin will be flipped three times. The random variable, $X$, of interest is the total number of heads to land up. Find the pmf of $X$ and calculate the mean and variance of $X$.

**Solution.** If a head lands up first, the possiblility will be {H,H,H}, {H,H,T}, {H,T,H}, {H,T,T}. If a tail lands up first, the possiblility will be {T,H,H,H}, {T,H,H,T}, {T,H,T,H}, {T,H,T,T}, {T,T,H,H}, {T,T,H,T}, {T,T,T,H}, {T,T,T,T}. The probability of first getting a head is 0.5 and that of first getting a tail is also 0.5. Hence,
$P(X = 0) = 0.5(\frac{1}{8}) = 0.0625$,
$P(X = 1) = 0.5(\frac{1}{4}) + 0.5(\frac{3}{8}) = 0.3125$,
$P(X = 2) = 0.5(\frac{2}{4}) + 0.5(\frac{3}{8}) = 0.4375$,
$P(X = 3) = 0.5(\frac{1}{4}) + 0.5(\frac{1}{8}) = 0.1875$.
Note, $P(X = 0)+P(X = 1)+P(X = 2)+P(X = 3) = 0.0625+0.3125+0.4375+0.1875 = 1$.
Then, $E(X) = 0(0.0625) + 1(0.3125) + 2(0.4375) + 3(0.1875) = 1.75$ and
$Var(X) = (0)^2(0.0625) + (1)^2(0.3125) + (2)^2(0.4375) + (3)^2(0.1875) - 1.75^2 = 0.6875$.

**Exercise 4.4.** If $X$ is a binomial random variable with expected value 6 and variance 2.4, find $P(X = 5)$.

**Solution.** Since $E(X) = np$, $Var(X) = np(1-p)$, we are given that $np = 6$, $np(1-p) = 2.4$. Thus, $1 - p = 0.4$, or $p = 0.6$, $n = 10$. Hence, $P(X = 5) = \binom{10}{5}\,(0.6)^5\,(0.4)^5 = 0.2007$.

---

**Exercise 4.5.** In a batch of 2000 calculators, there are 8 defective ones on average. If a random sample of 150 is selected, find the probability of 5 defective ones.

**Solution.** Let $X$ be the number of defective calculators in the sample, then

$$P(X = 5) = \binom{150}{5}\left(\frac{8}{2000}\right)^5\left(1 - \frac{8}{2000}\right)^{150-5} = 3.39 \times 10^{-4}$$

Or since $n = 150$ is large and $p = \frac{8}{2000}$ is small, we may use the Poisson approximation, with $\lambda = 150 \cdot \frac{8}{2000} = 0.6$, we have

$$P(X = 5) = \frac{0.6^5}{5!}e^{-0.6} = 3.56 \times 10^{-4} \approx 3.39 \times 10^{-4}$$

---

**Exercise 4.6.** A bookstore owner examines 5 books from each lot of 25 to check for missing pages. If he finds at least 2 books with missing pages, the entire lot is returned. If, indeed, there are 5 books with missing pages, find the probability that the lot will be returned.

**Solution.** Let $X$ be the number of books that the owner finds with missing pages, then

$$P(X \geq 2) = 1 - P(X = 0) - P(X = 1) = 1 - \frac{\binom{5}{0}\binom{20}{5}}{\binom{25}{5}} - \frac{\binom{5}{1}\binom{20}{4}}{\binom{25}{5}} = 0.252$$

**Exercise 4.7.** Flaws in a certain type of drapery material appear on the average of one in 150 square feet. If we assume the Poisson distribution, find the probability of at most one flaw in 225 square feet.

**Solution.** Let $X$ be the number of flaws in 225 square feet of drapery material. Given that the mean for 150 square feet is 1, so the mean for 225 square feet is $\frac{225}{150} = 1.5$. Therefore, the distribution of $X$ is Poisson(1.5). Hence, $P(X \leq 1) = e^{-1.5} + e^{-1.5}(1.5) = 0.5578$.

---

**Exercise 4.8.** (13-14 2nd Term Final Q6)
Let $X$ be a Poisson random variable with parameter $\lambda$. Find the value of $i$ such that $P(X = i)$ is maximum.

**Solution.** $X \sim$ Poisson($\lambda$), $x = 0, 1, 2, ...$

$$p(x) = \frac{\lambda^x}{x!} e^{-\lambda}$$

$$p(x-1) = \frac{\lambda^{x-1}}{(x-1)!} e^{-\lambda}$$

$$\frac{p(x)}{p(x-1)} = \frac{\lambda}{x}$$

Therefore, $p(x)$ is maximum when $X$ is the largest integer $\leq \lambda$.

Note that calculation of Poisson probabilities can be done by the recursion relation:

$$p(x) = \frac{\lambda}{x}\ p(x-1)\ ,\ \ x = 1, 2, ...$$

Extension: If $Y \sim$ binomial($n, p$), then

$$p(y) = \frac{n-y+1}{y} \frac{p}{1-p}\ p(y-1)$$

since

$$\begin{aligned}
p(y) &= \binom{n}{y}\ p^y\ (1-p)^{n-y} \\
&= \frac{n!}{y!\ (n-y)!}\ p^y\ (1-p)^{n-y} \\
&= \frac{n-y+1}{y} \frac{p}{1-p} \frac{n!}{(y-1)!\ (n-y+1)!}\ p^{y-1}\ (1-p)^{n-(y-1)} \\
&= \frac{n-y+1}{y} \frac{p}{1-p}\ p(y-1)
\end{aligned}$$