# STAT2006 Basic Concepts in Statistics and Probability II
# Review of Selected Discrete Distributions

Chan Chun Ho, Benjamin*†

January 26, 2018

**Abstract**

It aims to review basic concepts of discrete probability distributions and incorporate several important discrete probability models for illustration. It serves as optional reading materials. It would not be covered in tutorials. Note that it is by no means a comprehensive list of all discrete distributions. No attempt is made at completeness or full rigor. Some materials does credit to former TA George and some are extracted from classic textbook "Statistical Inference" (Casella and Berger) used in STAT4003.

# Notations and Definitions

- Summation: $\sum_{i=1}^{n} x_i = x_1 + x_2 + \cdots + x_n$; Open interval: $(a, b) = \{x \in \mathbb{R} \mid a < x < b\}$
- 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 natural numbers: $\mathbb{N} = \{1, 2, \dots\}$
- Set membership: $x \in A$ means "$x$ is an element of the set $A$".
- Expectation: $E(\cdot)$; Variance: $Var(\cdot)$; Standard deviation: $SD(\cdot)$; Covariance: $Cov(\cdot, \cdot)$
- Moment-generating function (mgf): $M_X(t)$; Random variable: $X$; Observed value: $x$
- Sample space: $S$ is the set of all possible outcomes of experiments.
- Probability distribution: A tilde ($\sim$) means "has the probability distribution of".
- Parameter(s): $\theta$ denotes population characteristic(s) that can be set to different values to produce different probability distributions.
- Parameter space: $\Omega$ denotes the set of all possible values for all the different parameters in a distribution

*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 $X$ that maps each element $s$ in a sample space $S$ into a real number $x \in \mathbb{R}$, i.e. $X(s) = x$.
- A random variable is said to be **discrete** if the set $\{x : X(s) = x, s \in S\}$ is **countable**.
- A discrete random variable can have **a finite or an infinite number of values**.

## 1.2 Discrete Probability Distributions

- For a discrete random variable $X$, the **probability mass function (pmf)** of $X$ is

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

- Denote $\{x : X(s) = x, s \in S\}$ by $\{x\}$ and $\{k : X(s) = k, k \leq x, s \in S\}$ by $\{k : k \leq x\}$.
- Note that

$$0 \leq p(x) \leq 1 \quad \text{and} \quad \sum_{\{x\}} p(x) = 1.$$

- The **cumulative distribution function (cdf)** of $X$ is

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

## 1.3 Mean and Variance of Discrete Random Variables

- The **mean**, or **expectation**, or **expected value** of $X$, $g(X)$ and $e^{tX}$ are

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

$$E[g(X)] = \sum_{\{x\}} g(x)\ p(x) \quad \text{and} \quad M_X(t) = E(e^{tX}) = \sum_{\{x\}} e^{tx}\ p(x).$$

- The **variance** of $X$ is

$$\sigma^2 = Var(X) = E(X-\mu)^2 = \sum_{\{x\}} (x-\mu)^2 p(x) = \big[\sum_{\{x\}} x^2\ p(x)\big] - \mu^2 = E(X^2) - \mu^2.$$

- 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 $p \in (0,1)$ has a pmf as

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

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 Theorem 3.1)
- $Var(X) = p(1-p)$
- $M_X(t) = pe^t + (1-p)$

## 2.2 Binomial Distribution

A **binomial random variable** $X$ with parameters $(n, p)$, where $n \in \mathbb{N}$ and $p \in (0,1)$, 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.[1]

- $E(X) = np$ (Proofs in Theorem 3.4)
- $Var(X) = np(1-p)$
- $M_X(t) = [pe^t + (1-p)]^n$

[1] "Statistical Inference" 2nd ed. (Casella and Berger) Ch3 p.90

## 2.3 Multinomial Distribution

A **multinomial distribution** with parameters $(n, p_1, ..., p_k)$, where $n \in \mathbb{N}$, $p_i \in (0, 1)$ 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.

- $E(X_i) = np_i$
- $Var(X_i) = np_i(1 - p_i)$
- $Cov(X_i, X_j) = -np_i p_j \; (i \neq j)$
- $M_X(t_1, \ldots, t_k) = (\sum_{i=1}^{k} p_i e^{t_i})^n$

## 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 $X \sim$ binomial$(n, p)$ and $Y \sim$ Poisson$(\lambda)$, for large $n$ and small $p$, with $np = \lambda$, then $P(X = x) \approx P(Y = x)$.[2]

- $E(X) = \lambda$ (Proofs in Theorem 3.6)
- $Var(X) = \lambda$
- $M_X(t) = e^{\lambda(e^t - 1)}$

[2] Interested students may read "Statistical Inference" 2nd ed. (Casella and Berger) Ch2 p.66-67.

## 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 $x = \max(0, n - N_2), ..., \min(N_1, n)$.[3]

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

<u>Difference between binomial distribution and hypergeometric distribution</u>

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.

[3] "Statistical Inference" 2nd ed. (Casella and Berger) Ch3 p.86-88

## 3 Theorems and Proofs

**Theorem 3.1.** If $X \sim \text{Bernoulli}(p)$, $E(X) = p$, $Var(X) = p(1-p)$ and $M_X(t) = pe^t + (1-p)$.

*Proof.*

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

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

$$M_X(t) = \sum_{x=0}^{1} e^{tx}\ p(x) = e^{t \cdot 0} \cdot p(0) + e^{t \cdot 1} \cdot p(1) = 1 \cdot (1-p) + e^t \cdot p = pe^t + (1-p)$$

□

---

**Theorem 3.2.** If $X \sim \text{Bernoulli}(p)$, $\sum_{x=0}^{1} p(x) = 1$.

*Proof.*

$$\sum_{x=0}^{1} p(x) = p(0) + p(1) = (1-p) + p = 1.$$

□

---

**Theorem 3.3. (Binomial Theorem).** For any real numbers $u$ and $v$ and integer $n \geq 0$,[4]

$$(u+v)^n = \sum_{x=0}^{n} \binom{n}{x} u^x v^{n-x}.$$

*Proof.* Write

$$(u+v)^n = (u+v)(u+v) \cdot \dots \cdot (u+v),$$

and consider how the right-hand side would be calculated. From each factor $(u+v)$ we choose either an $u$ or $v$, and multiply together the $n$ choices. For each $x = 0, 1, \dots, n$, the number of such terms in which $u$ appears exactly $x$ times is $\binom{n}{x}$. Therefore, this term is of the form $\binom{n}{x} u^x v^{n-x}$ and the result follows. □

---

[4] "Statistical Inference" 2nd ed. (Casella and Berger) Ch3 p.90

**Theorem 3.4.** If $X \sim \text{binomial}(n,p)$, $E(X) = np$, $Var(X) = np(1-p)$ and $M_X(t) = [pe^t + (1-p)]^n$.

*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. Next, 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,

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

Finally,

$$M_X(t) = \sum_{x=0}^{n} e^{tx} \binom{n}{x} p^x (1-p)^{n-x} = \sum_{x=0}^{n} \binom{n}{x} (pe^t)^x (1-p)^{n-x}.$$

By Theorem 3.3,

$$\sum_{x=0}^{n} \binom{n}{x} u^x v^{n-x} = (u+v)^n.$$

Hence, letting $u = pe^t$ and $v = 1-p$, we have[5]

$$M_X(t) = [pe^t + (1-p)]^n.$$

□

---

**Theorem 3.5.** If $X \sim$ binomial$(n, p)$, $\sum_{x=0}^{n} p(x) = 1$.

*Proof.* By Theorem 3.3, letting $u = p$ and $v = 1-p$, we get

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

□

---

**Theorem 3.6.** If $X \sim$ Poisson$(\lambda)$, $E(X) = Var(X) = \lambda$.[6]

*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 $e^y$,

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

[5] "Statistical Inference" 2nd ed. (Casella and Berger) Ch2 p.64
[6] "Statistical Inference" 2nd ed. (Casella and Berger) Ch3 p.92-93

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

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

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

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

$\square$

**Question.** 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

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