# STA250 Probability and Statistics

# Chapter 6 Notes

**Some Discrete Probability Distributions**


***Asst. Prof. Abdullah YALÇINKAYA***

*Ankara University, Faculty of Science, Department of Statistics*

*ayalcinkaya@ankara.edu.tr*


*2023*

# STA250 Probability and Statistics

## Reference Book

This lecture notes are prepared according to the contents of

“**PROBABILITY & STATISTICS FOR ENGINEERS & SCIENTISTS** by Walpole, Myers, Myers and Ye”

# Discrete Probability Distributions

- **The probability distribution for a discrete variable $X$ can be represented by a formal, a table, or a graph that provides $p(x) = P\ (X = x)$.**
- **Some common discrete distribution models:**
- **<u>Uniform</u>: All outcomes are equally likely.**
- **<u>Binomial</u>: Number of successes in n independent trials, with each trial having probability of success $p$ and probability of failure $q\ (= 1 - p)$.**
- **<u>Multinomial</u>: # of outcomes in $n$ trials, with each of $k$ possible outcomes having probabilities $p_1, p_2, \ldots, pk$.**

# Discrete Probability Distributions

- **Common discrete distribution models, continued:**
- **<u>Hypergeometric</u>: A sample of size n is selected from $N$ items <u>without replacement</u>, and k items are classified as successes ($N - k$ are failures).**
- **<u>Negative Binomial</u>: In n independent trials, with probability of success $p$ and probability of failure $q$ ($q = 1 - p$) on each trial, the probability that the $k$th success occurs on the $x$th trial.**
- **<u>Geometric</u>: Special case of the negative binomial. The probability that the 1st success occurs on the $x$th trial.**
- **<u>Poisson</u>: If $\lambda$ is the rate of occurrence of an event (number of outcomes per unit time), the probability that $x$ outcomes occur in a time interval of length $t$.**

# Discrete Uniform Distribution

- When $X$ assumes the values $x_1, x_2, \ldots, x_k$ and each outcome is equally likely. Then

$$f(x\,;k) = \frac{1}{k}, x = x_1, x_2, \ldots, x_k,$$

- and

$$\mu = \frac{\sum_{i=1}^{k} x_i}{k}$$

$$\sigma^2 = \frac{\sum_{i=1}^{k} (x_i - \mu)^2}{k}$$

- Since all observations are equally likely, this is similar to the mean and variance of a sample of size $k$, but note that we use $k$ rather than $k - 1$ to calculate variance.

# Binomial Distribution

- **Binomial: Number of successes in n independent trials, with each trial having probability of success $p$ and probability of failure $q\ (= 1 - p)$.**
  - Each trial is called a Bernoulli trial.
  - Experiment consists of n repeated trials.
  - Two possible outcomes, called success or failure.
  - $P(success) = p$, constant from trial to trial.
  - The repeated trials are independent.
- **The number of successes in $n$ Bernoulli trials is a called binomial random variable. The probability distribution of this discrete random variable called the binomial distribution and its values will be denoted by $b(x; n, p)$.**

# Binomial Distribution

- **Binomial: If $x$ is the number of successes in $n$ trials, each with two outcomes where $p$ is the probability of success and $q = 1 - p$ is the probability of failure, the probability distribution of $X$ is**

$$b(x\,;n\,,p) = \binom{n}{x} p^x q^{n-x}, x = 0,1,2,\ldots,n.$$

**the number of ways a given outcome $x$ can occur times the probability of that outcome occurring, and**

$$\mu = np$$

$$\sigma^2 = npq$$

# Binomial Distribution Example 1.

Example 5.1: The probability that a certain kind of component will survive a shock test is 3/4. Find the probability that exactly 2 of the next 4 components tested survive.

*Solution*: Assuming that the tests are independent and $p = 3/4$ for each of the 4 tests, we obtain

$$b\left(2; 4, \frac{3}{4}\right) = \binom{4}{2}\left(\frac{3}{4}\right)^2\left(\frac{1}{4}\right)^2 = \left(\frac{4!}{2!\,2!}\right)\left(\frac{3^2}{4^4}\right) = \frac{27}{128}.$$

# Binomial Distribution Example 2.

**Example 5.2:** The probability that a patient recovers from a rare blood disease is 0.4. If 15 people are known to have contracted this disease, what is the probability that (a) at least 10 survive, (b) from 3 to 8 survive, and (c) exactly 5 survive?

***Solution*:** Let $X$ be the number of people who survive.

$$\text{(a)}\qquad P(X \geq 10) = 1 - P(X < 10) = 1 - \sum_{x=0}^{9} b(x; 15, 0.4) = 1 - 0.9662$$
$$= 0.0338$$

$$\text{(b)}\quad P(3 \leq X \leq 8) = \sum_{x=3}^{8} b(x; 15, 0.4) = \sum_{x=0}^{8} b(x; 15, 0.4) - \sum_{x=0}^{2} b(x; 15, 0.4)$$
$$= 0.9050 - 0.0271 = 0.8779$$

$$\text{(c)}\qquad P(X = 5) = b(5; 15, 0.4) = \sum_{x=0}^{5} b(x; 15, 0.4) - \sum_{x=0}^{4} b(x; 15, 0.4)$$
$$= 0.4032 - 0.2173 = 0.1859$$

# Binomial Distribution Example 3.

**Example 5.2:** The probability that a patient recovers from a rare blood disease is 0.4. If 15 people are known to have contracted this disease, what is the probability that (a) at least 10 survive, (b) from 3 to 8 survive, and (c) exactly 5 survive?

***Solution*:** Let $X$ be the number of people who survive.

(a) $$P(X \geq 10) = 1 - P(X < 10) = 1 - \sum_{x=0}^{9} b(x; 15, 0.4) = 1 - 0.9662$$
$$= 0.0338$$

(b) $$P(3 \leq X \leq 8) = \sum_{x=3}^{8} b(x; 15, 0.4) = \sum_{x=0}^{8} b(x; 15, 0.4) - \sum_{x=0}^{2} b(x; 15, 0.4)$$
$$= 0.9050 - 0.0271 = 0.8779$$

(c) $$P(X = 5) = b(5; 15, 0.4) = \sum_{x=0}^{5} b(x; 15, 0.4) - \sum_{x=0}^{4} b(x; 15, 0.4)$$
$$= 0.4032 - 0.2173 = 0.1859$$

**Find the mean and variance of the binomial random variable of Example 5.2.**

**Solution:** Since Example 5.2 was a binomial experiment with $n$ = 15 and $p$ = 0.4,

$$\mu = (15)(0.4) = 6 \text{ and } \sigma^2 = (15)(0.4)(0.6) = 3.6$$

# Binomial Distribution Example 4.

**(Chebyshev's Theorem)** The probability that any random variable $X$ will assume a value within $k$ standard deviations of the mean is at least $1 - 1/k^2$. That is,

$$P(\mu - k\sigma < X < \mu + k\sigma) \geq 1 - \frac{1}{k^2}.$$

**Example 5.5:** Find the mean and variance of the binomial random variable of Example 5.2, and then use Chebyshev's theorem (on page 137) to interpret the interval $\mu \pm 2\sigma$.

***Solution*:** Since Example 5.2 was a binomial experiment with $n = 15$ and $p = 0.4$, by Theorem 5.1, we have

$$\mu = (15)(0.4) = 6 \text{ and } \sigma^2 = (15)(0.4)(0.6) = 3.6.$$

Taking the square root of 3.6, we find that $\sigma = 1.897$. Hence, the required interval is $6 \pm (2)(1.897)$, or from 2.206 to 9.794. Chebyshev's theorem states that the number of recoveries among 15 patients who contracted the disease has a probability of at least 3/4 of falling between 2.206 and 9.794 or, because the data are discrete, between 2 and 10 inclusive.

$$P(\mu - 2\sigma < X < \mu + 2\sigma) \geq 1 - \frac{1}{4}$$

$$P(2 < X < 10) \geq \frac{3}{4}$$

# Multinomial Distribution

**Multinomial Distribution**

If a given trial can result in the $k$ outcomes $E_1, E_2, \ldots, E_k$ with probabilities $p_1, p_2, \ldots, p_k$, then the probability distribution of the random variables $X_1$, $X_2$, $\ldots$, $X_k$, representing the number of occurrences for $E_1, E_2, \ldots, E_k$ in $n$ independent trials, is

$$f(x_1, x_2, \ldots, x_k; p_1, p_2, \ldots, p_k, n) = \binom{n}{x_1, x_2, \ldots, x_k} p_1^{x_1} p_2^{x_2} \cdots p_k^{x_k},$$

with

$$\sum_{i=1}^{k} x_i = n \text{ and } \sum_{i=1}^{k} p_i = 1.$$

# Multinomial Distribution Example

**Example 5.7:** The complexity of arrivals and departures of planes at an airport is such that computer simulation is often used to model the "ideal" conditions. For a certain airport with three runways, it is known that in the ideal setting the following are the probabilities that the individual runways are accessed by a randomly arriving commercial jet:

$$\begin{aligned} &\text{Runway 1:} \quad p_1 = 2/9,\\ &\text{Runway 2:} \quad p_2 = 1/6,\\ &\text{Runway 3:} \quad p_3 = 11/18. \end{aligned}$$

What is the probability that 6 randomly arriving airplanes are distributed in the following fashion?

$$\begin{aligned} &\text{Runway 1:} \quad \text{2 airplanes},\\ &\text{Runway 2:} \quad \text{1 airplane},\\ &\text{Runway 3:} \quad \text{3 airplanes} \end{aligned}$$

***Solution*:** Using the multinomial distribution, we have

$$\begin{aligned} f\left(2, 1, 3; \frac{2}{9}, \frac{1}{6}, \frac{11}{18}, 6\right) &= \binom{6}{2, 1, 3} \left(\frac{2}{9}\right)^2 \left(\frac{1}{6}\right)^1 \left(\frac{11}{18}\right)^3 \\ &= \frac{6!}{2!\,1!\,3!} \cdot \frac{2^2}{9^2} \cdot \frac{1}{6} \cdot \frac{11^3}{18^3} = 0.1127. \end{aligned}$$

# Hypergeometric Distribution

- **Hypergeometric: The distribution of the number of successes, $x$, in a sample of size $n$ is selected from $N$ items without replacement, where $k$ items are classified as successes (and $N-k$ as failures), is**

$$h(x\,;N\,,n,k)=\frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}},$$

$$\max\{0,n-(N-k)\}\leq x\leq \min\{k,n\}$$

**then**

$$\mu=\frac{nk}{N}$$

**and**

$$\sigma^2=\frac{N-n}{N-1}\cdot n\cdot\frac{k}{N}(1-\frac{k}{N})$$

# Hypergeometric Distribution Example

**Example 5.9:** Lots of 40 components each are deemed unacceptable if they contain 3 or more defectives. The procedure for sampling a lot is to select 5 components at random and to reject the lot if a defective is found. What is the probability that exactly 1 defective is found in the sample if there are 3 defectives in the entire lot?

***Solution:*** Using the hypergeometric distribution with $n = 5$, $N = 40$, $k = 3$, and $x = 1$, we find the probability of obtaining 1 defective to be

$$h(1; 40, 5, 3) = \frac{\binom{3}{1}\binom{37}{4}}{\binom{40}{5}} = 0.3011.$$

Once again, this plan is not desirable since it detects a bad lot (3 defectives) only about 30% of the time.

**Example 5.11:** Find the mean and variance of the random variable of Example 5.9

***Solution:*** Since Example 5.9 was a hypergeometric experiment with $N = 40$, $n = 5$, and $k = 3$, by Theorem 5.2, we have

$$\mu = \frac{(5)(3)}{40} = \frac{3}{8} = 0.375,$$

and

$$\sigma^2 = \left(\frac{40-5}{39}\right)(5)\left(\frac{3}{40}\right)\left(1 - \frac{3}{40}\right) = 0.3113.$$

# Binomial Approximation to Hypergeometric

- **If n is small compared with $N$, then the hypergeometric distribution can be approximated using the binomial distribution.**
- **The rule of thumb is that this is valid if $(n/N) \leq 0.05$. In this case, we can use the binomial distribution with parameters $n$ and $p = k/N$.**
- **Then**

$$\mu = np = \frac{nk}{N}$$

$$\sigma^2 = npq = n \cdot \frac{k}{N}(1 - \frac{k}{N})$$

# Binomial Approximation to Hypergeometric

**Example 5.12:** A manufacturer of automobile tires reports that among a shipment of 5000 sent to a local distributor, 1000 are slightly blemished. If one purchases 10 of these tires at random from the distributor, what is the probability that exactly 3 are blemished?

***Solution*:** Since $N = 5000$ is large relative to the sample size $n = 10$, we shall approximate the desired probability by using the binomial distribution. The probability of obtaining a blemished tire is 0.2. Therefore, the probability of obtaining exactly 3 blemished tires is

$$h(3; 5000, 10, 1000) \approx b(3; 10, 0.2) = 0.8791 - 0.6778 = 0.2013.$$

On the other hand, the exact probability is $h(3; 5000, 10, 1000) = 0.2015$.

# Negative Binomial Distribution

- **Negative Binomial: In n independent trials, with probability of success $p$ and probability of failure $q$ $(q = 1 - p)$ on each trial, the probability that the $kth$ success occurs on the $xth$ trial.**

$$b^*(x\,;k\,,p) = \binom{x-1}{k-1} p^k q^{x-k}, x = k, k+1, k+2, \ldots$$

- **Again we have the number of ways an outcome $x$ can occur times the probability of that outcome occurring.**

- **The above formula comes from the fact that in order to get the $kth$ success on the $xth$ trial, we must have $k - 1$ successes in the first $x - 1$ trials, and then the final trial must also be a success.**

# Negative Binomial Distribution Example

**Example 5.14:** In an NBA (National Basketball Association) championship series, the team that wins four games out of seven is the winner. Suppose that teams $A$ and $B$ face each other in the championship games and that team $A$ has probability 0.55 of winning a game over team $B$.

(a) What is the probability that team $A$ will win the series in 6 games?

(b) What is the probability that team $A$ will win the series?

(c) If teams $A$ and $B$ were facing each other in a regional playoff series, which is decided by winning three out of five games, what is the probability that team $A$ would win the series?

*Solution:* (a) $b^*(6; 4, 0.55) = \binom{5}{3} 0.55^4 (1 - 0.55)^{6-4} = 0.1853$

(b) $P$(team $A$ wins the championship series) is

$$\begin{aligned} &b^*(4; 4, 0.55) + b^*(5; 4, 0.55) + b^*(6; 4, 0.55) + b^*(7; 4, 0.55) \\ &= 0.0915 + 0.1647 + 0.1853 + 0.1668 = 0.6083. \end{aligned}$$

(c) $P$(team $A$ wins the playoff) is

$$\begin{aligned} &b^*(3; 3, 0.55) + b^*(4; 3, 0.55) + b^*(5; 3, 0.55) \\ &= 0.1664 + 0.2246 + 0.2021 = 0.5931. \end{aligned}$$

# Geometric Distribution

- **<u>Geometric</u>: Special case of the negative binomial with $k = 1$. The probability that the *<u>1st</u>* success occurs on the $xth$ trial is**

$$g(x\,;p) = pq^{x-1}, \qquad x = 1,2,3,\ldots$$

**then**

$$\mu = \frac{1}{p}$$

**and**

$$\sigma^2 = \frac{1-p}{p^2}$$

# Geometric Distribution Example

**Example 5.15:** For a certain manufacturing process, it is known that, on the average, 1 in every 100 items is defective. What is the probability that the fifth item inspected is the first defective item found?

***Solution*:** Using the geometric distribution with $x = 5$ and $p = 0.01$, we have

$$g(5; 0.01) = (0.01)(0.99)^4 = 0.0096.$$

**Example 5.16:** At a "busy time," a telephone exchange is very near capacity, so callers have difficulty placing their calls. It may be of interest to know the number of attempts necessary in order to make a connection. Suppose that we let $p = 0.05$ be the probability of a connection during a busy time. We are interested in knowing the probability that 5 attempts are necessary for a successful call.

***Solution*:** Using the geometric distribution with $x = 5$ and $p = 0.05$ yields

$$P(X = x) = g(5; 0.05) = (0.05)(0.95)^4 = 0.041.$$

# Poisson Distribution

- **Poisson distribution: If $\lambda$ is the average # of outcomes per unit time (arrival rate), the Poisson distribution gives the probability that $x$ outcomes occur in a given time interval of length $t$.**
- **A Poisson process has the following properties:**
  - Memoryless: the number of occurrences in one time interval is independent of the number in any other disjoint time interval.
  - The probability that a single outcome will occur during a very short time interval is proportional to the size of the time interval and independent of other intervals.
  - The probability that more than one outcome will occur in a very short time interval is negligible.
- **Note that the rate could be per unit length, area, or volume, rather than time.**

# Poisson Distribution

- **<u>Poisson distribution</u>: If $\lambda$ is the rate of occurrence of an event (average # of outcomes per unit time), the probability that $x$ outcomes occur in a time interval of length $t$ is**

$$p(x\,;\lambda\,t) = \frac{e^{-\lambda t}(\lambda t)^x}{x!}, x = 0,1,2,\ldots$$

**then**

$$\mu = \sigma^2 = \lambda t$$

# Poisson Distribution Example 1

Suppose that a random system of police patrol is devised so that a patrol officer may visit a given beat location $Y = 0, 1, 2, 3, \ldots$ times per half-hour period, with each location being visited an average of once per time period. Assume that $Y$ possesses, approximately, a Poisson probability distribution. Calculate the probability that the patrol officer will miss a given location during a half-hour period. What is the probability that it will be visited once? Twice? At least once?

**Solution** For this example the time period is a half-hour, and the mean number of visits per half-hour interval is $\lambda = 1$. Then

$$p(y) = \frac{(1)^y e^{-1}}{y!} = \frac{e^{-1}}{y!}, \qquad y = 0, 1, 2, \ldots.$$

The event that a given location is missed in a half-hour period corresponds to $(Y = 0)$, and

$$P(Y = 0) = p(0) = \frac{e^{-1}}{0!} = e^{-1} = .368.$$

Similarly,

$$p(1) = \frac{e^{-1}}{1!} = e^{-1} = .368,$$

and

$$p(2) = \frac{e^{-1}}{2!} = \frac{e^{-1}}{2} = .184.$$

The probability that the location is visited *at least* once is the event $(Y \geq 1)$. Then

$$P(Y \geq 1) = \sum_{y=1}^{\infty} p(y) = 1 - p(0) = 1 - e^{-1} = .632. \quad \blacksquare$$

Source: ***Mathematical Statistics with Applications,*** **Seventh Edition**
*Dennis D. Wackerly, William Mendenhall III, Richard L. Scheaffer*

# Poisson Distribution Example 2

**Example 5.17:** During a laboratory experiment, the average number of radioactive particles passing through a counter in 1 millisecond is 4. What is the probability that 6 particles enter the counter in a given millisecond?

***Solution*:** Using the Poisson distribution with $x = 6$ and $\lambda t = 4$ and referring to Table A.2, we have

$$p(6;4) = \frac{e^{-4}4^6}{6!} = \sum_{x=0}^{6} p(x;4) - \sum_{x=0}^{5} p(x;4) = 0.8893 - 0.7851 = 0.1042.$$

**Example 5.18:** Ten is the average number of oil tankers arriving each day at a certain port. The facilities at the port can handle at most 15 tankers per day. What is the probability that on a given day tankers have to be turned away?

***Solution*:** Let $X$ be the number of tankers arriving each day. Then, using Table A.2, we have

$$P(X > 15) = 1 - P(X \leq 15) = 1 - \sum_{x=0}^{15} p(x;10) = 1 - 0.9513 = 0.0487.$$

Table A.2 contains Poisson probability sums,

$$P(r;\lambda t) = \sum_{x=0}^{r} p(x;\lambda t),$$

# Poisson Approximation to Binomial

- **For a set of Bernoulli trials with $n$ very large and $p$ small, the Poisson distribution with mean $np$ can be used to approximate the binomial distribution.**
  - Needed since binomial tables only go up to $n = 20$.
- **The rule of thumb is that this approximation is valid if $n \geq 20$ and $p \leq 0.05$. (*If* $n \geq 100$, the approximation is excellent if $np \leq 10$). In this case, we can use the Poisson distribution with**

$$\mu = \sigma^2 = np$$

  - A different approximation for the binomial can be used for large $n$ if $p$ is not small.

Let $X$ be a binomial random variable with probability distribution $b(x; n, p)$. When $n \to \infty$, $p \to 0$, and $np \overset{n\to\infty}{\longrightarrow} \mu$ remains constant,

$$b(x; n, p) \overset{n\to\infty}{\longrightarrow} p(x; \mu).$$

# Poisson Approximation to Binomial Example

**Example 5.19:** In a certain industrial facility, accidents occur infrequently. It is known that the probability of an accident on any given day is 0.005 and accidents are independent of each other.

(a) What is the probability that in any given period of 400 days there will be an accident on one day?

(b) What is the probability that there are at most three days with an accident?

***Solution*:** Let $X$ be a binomial random variable with $n = 400$ and $p = 0.005$. Thus, $np = 2$. Using the Poisson approximation,

(a) $P(X = 1) = e^{-2}2^1 = 0.271$ and

(b) $P(X \leq 3) = \sum_{x=0}^{3} e^{-2}2^x/x! = 0.857$.

# Poisson Approximation to Binomial Example

**Example 5.20:** In a manufacturing process where glass products are made, defects or bubbles occur, occasionally rendering the piece undesirable for marketing. It is known that, on average, 1 in every 1000 of these items produced has one or more bubbles. What is the probability that a random sample of 8000 will yield fewer than 7 items possessing bubbles?

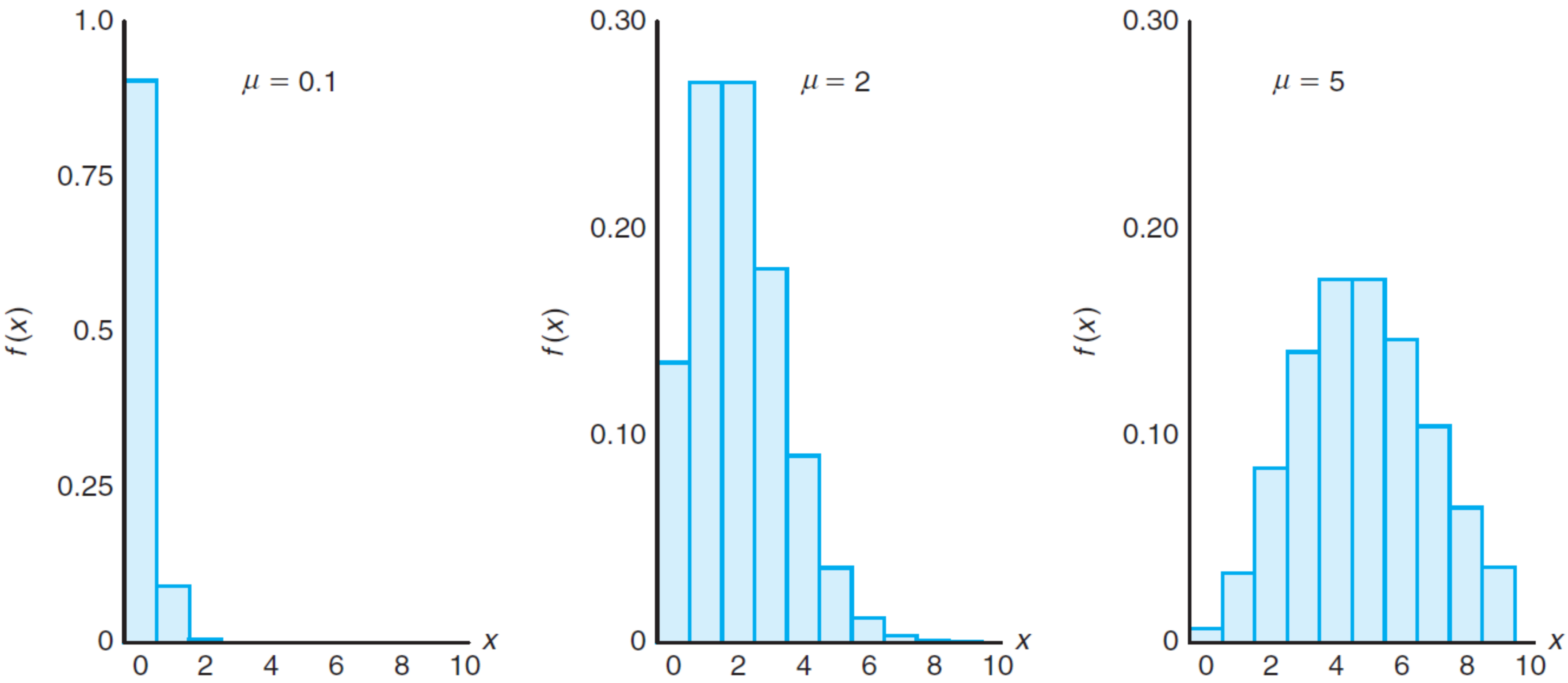
***Solution*:** This is essentially a binomial experiment with $n = 8000$ and $p = 0.001$. Since $p$ is very close to 0 and $n$ is quite large, we shall approximate with the Poisson distribution using

$$\mu = (8000)(0.001) = 8.$$

Hence, if $X$ represents the number of bubbles, we have

$$P(X < 7) = \sum_{x=0}^{6} b(x; 8000, 0.001) \approx p(x; 8) = 0.3134.$$

# Nature of Poisson Probability Function




The nearness to symmetry when μ becomes as large as 5.

# Next Lesson

- **Some Continuous Probability Distributions**

**See you☺**