Discrete Random Variables II (3.4 - 3.6)

Binomial Random Variables

1. **Binomial Random Variable:** A binomial random variable is a discrete random variable that results from the conduct of a binomial experiment.

   (a) **Binomial Experiment:** An experiment, for which the following four conditions are satisfied, is called a **binomial experiment**:

      i. The experiment consists of a sequence of n independent trials, where n is fixed prior to the conduct of the experiment.

      ii. Each trial is **identical** and must result in one of the same **two possible outcomes**, denoted "success" (S) or "failure" (F).

      iii. The trials are independent, which can result from either:
      - Random Sampling **with replacement** from a dichotomous population - one with only two types of items (presence or absence of attribute).
      - Sampling **without replacement** from a very large population with only two types of items. Sampling not independent, but approximately so if $n/N \leq .05$, where n is the number of trials(or sample size) and N is the population size.

      iv. The probability of success, $P(S)$, is the same for each trial, $P(S) = p$.

   (b) **Binomial random variable,** X, is a count of the number of successes in n trials:

   $$X \equiv \{ \text{ number of S's in } n \text{ trials}\} = \{0, 1, 2, ..., n\}$$

   X has a **binomial distribution** with parameters n and p. The pmf is denoted as:

   $$X \sim b(x; n, p)$$

   (c) The Binomial pmf has the following functional form:

   $$b(x; n, p) = P(X = x) = \left(\begin{array}{c} n \\ x \end{array} \right) p^x (1-p)^{(n-x)}$$

   - there are "n choose x" or $\left(\begin{array}{c} n \\ x \end{array} \right)$ ways to arrange x successes and $n - x$ failures
   - each such arrangement has probability $p^x(1-p)^{(n-x)}$
   - the binomial coefficient

   $$\left(\begin{array}{c} n \\ x \end{array} \right) = \frac{n!}{x!(n-x)!}$$

   recall that $\left(\begin{array}{c} n \\ 0 \end{array} \right) = 1$, $\left(\begin{array}{c} n \\ 1 \end{array} \right) = n$, $\left(\begin{array}{c} n \\ n \end{array} \right) = 1$,
   $n! = n(n-1)\ldots(2)1$ and $0! = 1$

(d) Example: For a three coin toss experiment, let random variable $X = \{$number of heads in three tosses$\}$. X is then a Binomial random variable with $n = 3$ and $p = 0.5$. It can be shown that there are 8 possible outcomes. First consider the general case of a binomial experiment where P(H)$= p$ and P(T)$= (1-p)$.

Outcome	Probability	
HHH	P(H)P(H)P(H)	p^3
HHT	P(H)P(H)P(T)	$p^2(1-p)$
HTH	P(H)P(T)P(H)	$p(1-p)p$
HTT	P(H)P(T)P(T)	$p(1-p)^2$
THH	P(T)P(H)P(H)	$p^2(1-p)$
THT	P(T)P(H)P(T)	$p(1-p)^2$
TTH	P(H)P(T)P(T)	$p(1-p)^2$
TTT	P(T)P(T)P(T)	$(1-p)^3$

- P(X=0) = P(TTT) $= (1-p)^3$
- P(X=1) = P(HTT)+P(THT)+P(TTH), or

$$\begin{aligned} \text{P(X=1)} &= p(1-p)^2 + p(1-p)^2 + p(1-p)^2 \\ &= 3p(1-p)^2 \end{aligned}$$

- P(X=2) = P(HHT)+P(THH)+P(HTH), or

$$\begin{aligned} \text{P(X=2)} &= p^2(1-p) + p^2(1-p) + p^2(1-p) \\ &= 3p^2(1-p) \end{aligned}$$

- and P(X=3) = P(HHH) $= p^3$

- Compare the above with probabilities calculated using the binomial probability density function

$$p(x) = \left(\begin{array}{c} n \\ x \end{array} \right) p^x(1-p)^{(n-x)}$$

- here $n = 3$ and P(H)$= p$, so that

$$P(X=0) = \left(\begin{array}{c} 3 \\ 0 \end{array} \right) 1 \cdot p^0(1-p)^{(3-0)} = (1-p)^3$$

- and for X=1

$$P(X=1) = \left(\begin{array}{c} 3 \\ 1 \end{array} \right) 3 \cdot p^1(1-p)^2 = 3p(1-p)^2$$

- and for X=2

$$P(X=2) = \left(\begin{array}{c} 3 \\ 2 \end{array} \right) 3 \cdot p^2(1-p)^1 = 3p^2(1-p)^1$$

- and for X=3

$$P(X=3) = \left(\begin{array}{c} 3 \\ 3 \end{array} \right) 1 \cdot p^3(1-p)^0 = 3p^3$$

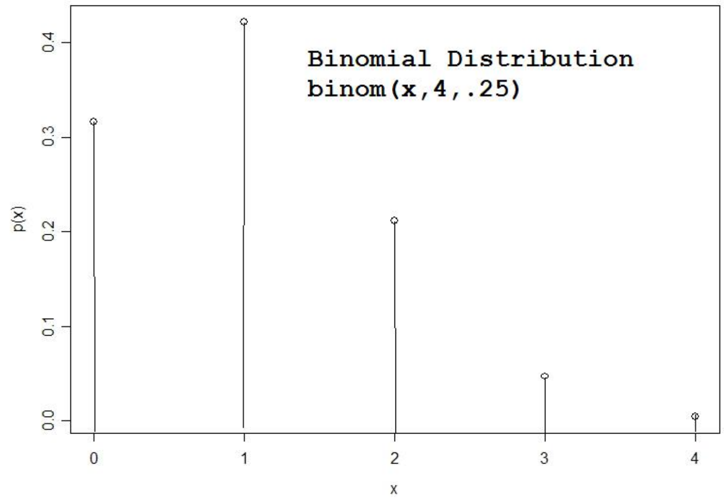
- Note that these values are identical to those computed above in the table of 8 possible outcomes.
- When each of the 8 outcomes are equally probable, the probability distribution for random variable X, the number of heads in 3 tosses, is as shown in the table below.

x	0	1	2	3
p(x)	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

(e) Example: Cystic Fibrosis is an autosomal recessive genetic disease occurring when a child receives two copies of a faulty gene. In Caucasian populations, nearly 1 in 25 are carriers (one faulty gene) and 1 in 2500 are(were) born with the disease. Suppose both parents are carriers and they have 4 children. Each parent passes, independently, either a faulty gene (+) or a healthy gene (−) to each child born to them. As carriers, the Mother has (+, −) and the Father has (+, −).

What is the probability that at least one of their 4 children gets CF?

- $P(child\ gets\ CF) = P(+_M\ \cap\ +_F) = \left(\frac{1}{2}\right) \cdot \left(\frac{1}{2}\right) = 1/4$
- X, the number of children with CF is then a binomial random varaible with $n = 4$ and $p = .25$ distributed as $b(x;\ 4,\ .25)$.
- Now determine pmf values for X, where $X = \{0, 1, 2, 3, 4\}$
  - $P(X = 0) = \begin{pmatrix} 4 \\ 0 \end{pmatrix} .25^0(.75^4) = .75^4 = 0.3164$ and $P(X > 0) = 1 - .75^4 = 0.6836$
  - $P(X = 1) = \begin{pmatrix} 4 \\ 1 \end{pmatrix} .25^1(.75^3) = 4(.25)(.75^3) = .4219$
  - $P(X = 2) = \begin{pmatrix} 4 \\ 2 \end{pmatrix} .25^2(.75^2) = .2109$
  - $P(X = 3) = \begin{pmatrix} 4 \\ 3 \end{pmatrix} .25^3(.75^1) = .0469$
  - $P(X = 4) = \begin{pmatrix} 4 \\ 4 \end{pmatrix} .25^4(.75^1) = .25^4 = .0039$



(f) **Cumulative Distribution Function(cdf):** When X is a binomial random variable with parameters n and p. The cdf is denoted as:

$$F(x) = P(X \leq x) = B(x; n, p) = \sum_{y=0}^{x} b(y; n, p)$$
$$\text{where } x = 0, 1, 2, ..., n$$

(g) Example: Studies show that, in Truro, only 40% of drivers signal before turning. If 20 turns are monitored at a particular Truro intersection

i. What is the probability that no more than 4 drivers signal?

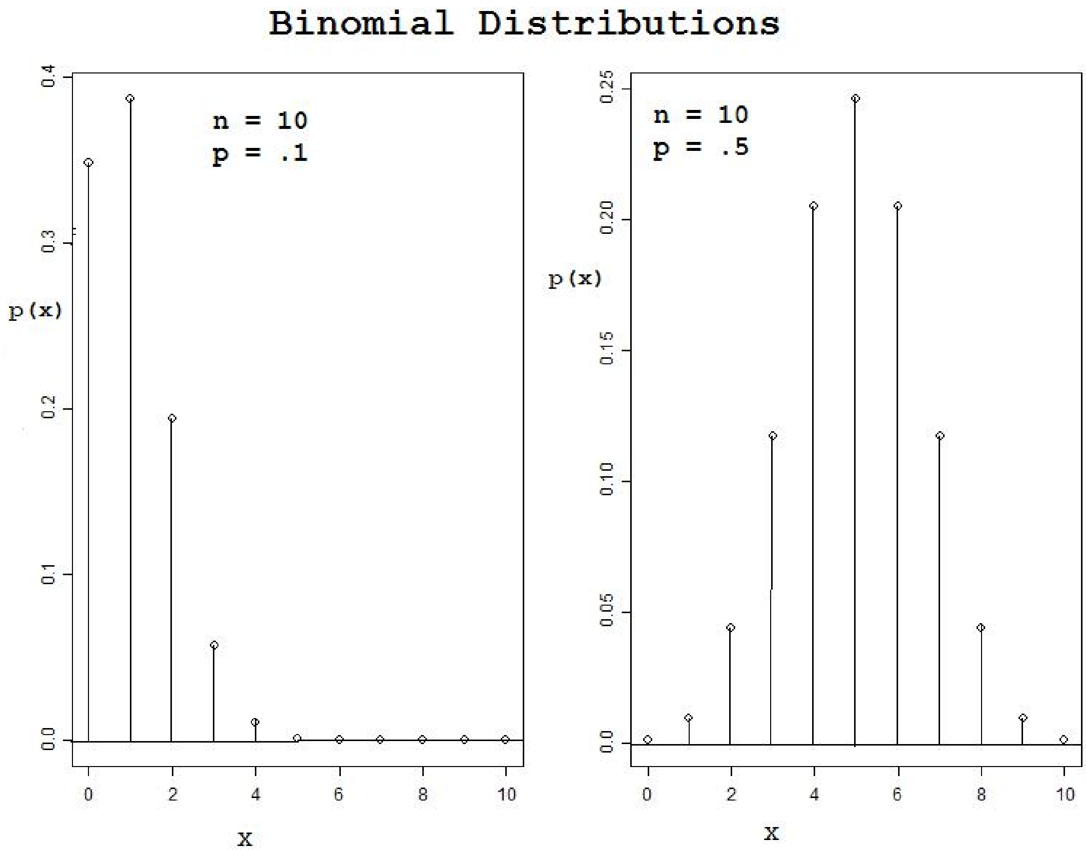
- Random variable X, the number of drivers signalling out of 20, is b(x;20,.40)
- we want to determine the value of $P(X \leq 4) = B(4; 20, .40)$
- Could be computed as: $B(4; 20, .40) = \sum_{x=0}^{4} b(x; 20, .40) = p(0) + p(1) + p(2) + p(3) + p(4)$
- or, looked up directly from Table A.1: Cumulative Binomial Probabilities (Text p.737), using $n = 20$ and $p = .40$: $P(X \leq 4) = 0.051$

ii. What is the probability that at least 6 drivers signal? (using Table A.1)

- want $P(X \geq 6) = 1 - P(X \leq 5) = 1 - 0.126 = .874$

(h) **Shape of the Binomial Distribution(pmf):** Shape varies with the value of parameter, p.

- for $p = .1$ the distribution is skewed to the right
- for $p = .5$ the distribution is symmetric, nearly normal



- Under certain conditions we can approximate the binomial distribution using the normal distribution with correct mean and variance.

2. **Expected Value for Binomial Random Variable, X:** The expected value (or mean) of any discrete random variable is a **probability weighted** sum of its values.

$$\begin{aligned} E(X) = \mu_X &= x_1 p(x_1) + x_2 p(x_2) + \ldots + x_k p(x_k) \\ &= \sum x_i p(x_i) \end{aligned}$$

   (a) When the binomial pmf is substituted for $p(x_i)$ above, the following value results for the mean of X:

$$E(X) = \mu_X = \sum_{x=0}^{n} x \left(\begin{array}{c} n \\ x \end{array} \right) p^x (1-p)^{(n-x)} = np$$

   (b) The expected value or mean is generally not one of the values random variable X takes on.

   (c) Example: Once again, consider a three coin toss experiment with $X = \{$ number of heads in three tosses$\}$. Here, X is Binomial with $n = 3$ and $p = 0.5$. There are eight possible outcomes. The probability distribution(pmf) for X determined earlier, is shown in the table below.

x	0	1	2	3
p(x)	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

(d) When the expected value of X, $E(X)$ or μ_X, is computed from the pmf:

$$E(X) = \mu_X = \sum x_i p(x_i) = 0\frac{1}{8} + 1\frac{3}{8} + 2\frac{3}{8} + 3\frac{1}{8} = 1.5$$

(e) When the expected value of binomial random variable X is computed as $E(X) = np$:

$$E(X) = \mu_X = np = 3(0.5) = 1.5$$

Note that the two expected values computed above are identical.

3. **Variance of Binomial Random Variable, X:** The *variance* of a discrete random variable X is the probability weighted average of squared deviations of its values from their mean.

$$\begin{aligned} Var(X) = \sigma_X^2 &= (x_1 - \mu)^2 p(x_1) + \ldots + (x_k - \mu)^2 p(x_k) \\ &= \sum (x_i - \mu)^2 p(x_i) \end{aligned}$$

(a) When the binomial pmf and expected value is substituted for $p(x_i)$ and μ above, the following relationship results for the variance of X:

$$Var(X) = \sigma_X^2 = \sum_{x=0}^{n} (x - np)^2 \binom{n}{x} p^x (1-p)^{(n-x)} = np(1-p)$$

(b) The Standard Deviation of binomial random variable X, denoted $SD(X)$, is computed as the positive square root of the variance of X, $Var(X)$.

$$SD(X) = \sqrt{Var(X)} = \sqrt{np(1-p)}$$

(c) For our previous example, the variance of random variable X is first calculated from the definition of variance for a discrete random variable.

$$\begin{aligned} Var(X) = \sigma_X^2 &= (0 - 1.5)^2 \frac{1}{8} + (1 - 1.5)^2 \frac{3}{8} \\ &\quad + (2 - 1.5)^2 \frac{3}{8} + (3 - 1.5)^2 \frac{1}{8} = 0.75 \end{aligned}$$

and

$$SD(X) = \sigma_X = .866$$

(d) When the variance of binomial random variable X is computed as $Var(X) = np(1-p)$:

$$Var(X) = \sigma_X^2 = np(1-p) = 3(0.5)(0.5) = 0.75$$

Note that the two variances computed above are identical.

(e) Example: Cards are drawn from a standard 52 card deck. If drawing a club is considered a success, determine the following probabilities.

i. What is the probability of exactly one success in four draws (with replacement)?

- Let $X \equiv \{$ number of clubs drawn$\}$
- X is a binomial rv with $n = 4$ and $p = 13/52 = 1/4$
- $P(X = 1) = b(1; 4, 0.25) = \left(\begin{array}{c} 4 \\ 1 \end{array} \right) \left(\frac{1}{4}\right)^1 \cdot \left(\frac{3}{4}\right)^3 = 0.422$

ii. What is the probability of no successes in five draws(with replacement)?

- X is a binomial rv with $n = 5$ and $p = 13/52 = 1/4$
- $P(X = 0) = b(0; 5, 0.25) = \left(\begin{array}{c} 5 \\ 0 \end{array} \right) \left(\frac{1}{4}\right)^0 \cdot \left(\frac{3}{4}\right)^5 = 0.237$

iii. If X is the number of S's(clubs) drawn, what is the expected value of X for five draws(with replacement)?

- X is binomial with $n = 5$ and $p = 13/52 = 1/4$
- $E(X) = np = 5(1/4) = 1.25$

iv. If X is the number of S's(clubs) drawn, what is the variance and standard deviation of X for five draws(with replacement)?

- X is binomial with $n = 5$ and $p = 13/52 = 1/4$
- $Var(X) = np(1 - p) = 5(1/4)(3/4) = 0.9375$
- $SD(X) = \sqrt{Var(X)} = \sqrt{0.9375} = 0.968$

Hypergeometric Random Variables

4. **Hypergeometric Random Variable:** A hypergeometric random variable is a discrete random variable that results from the sampling of certain finite populations without replacement.

   (a) There are three specific sampling assumptions necessary for formation of random variables that follow the hypergeometric distribution:

   - The population to be sampled consists of N individuals, a finite population.
   - The population to be sampled is dichotomous in that each individual in the population can be characterized as either an S or an F, and there are M S's in the population.
   - A random sample of size n is selected without replacement from the population.

   (b) Random variable X is then the total number of S's in the sample and the probability mass function (pmf) for X is given by the **hypergeometric distribution**.

$$h(x; n, M, N) = P(X = x) = \frac{\left(\begin{array}{c} M \\ x \end{array} \right) \left(\begin{array}{c} N-M \\ n-x \end{array} \right)}{\left(\begin{array}{c} N \\ n \end{array} \right)}$$

   (c) **Hypergeometric random variable,** X, is a count of the number of successes in a sample of size, n:

$$max(0, n - N + M) \ \leq x \leq \ min(n, N)$$

   pmf parameters are n, M, and N. The pmf is denoted as:

$$X \sim h(x; n, M, N)$$

   (d) Combinations in numerator and denominator of hypergeometric pmf.

      i. $\left(\begin{array}{c} M \\ x \end{array} \right) \equiv$ {# ways to select x S's from a total of M S's }

      ii. $\left(\begin{array}{c} N-M \\ n-x \end{array} \right) \equiv$ {# ways to select $(n-x)$ F's from a total of $(N-M)$ F's }

      iii. $\left(\begin{array}{c} N \\ n \end{array} \right) \equiv$ {# ways to select sample of size n from finite population of size N}

(e) Example: Five individuals from an animal population thought to be near extinction in Northern British Columbia have been caught, tagged, and released to mix into the population. After they have had an opportunity to mix, a random sample of 10 of these animals is selected. Let $X \equiv \{$number of tagged animals in second sample$\}$.

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

If there are a total of 25 animals of this type in the region, what is the probability that $X = 2$?

- Population: $N = 25$ total animals, $M = 5$ tagged animals

- Sample: $n = 10$ sample size, $x = 2$

$$P(X = 2) = h(2; 10, 5, 25) = \frac{\binom{5}{2}\binom{20}{8}}{\binom{25}{10}} = 0.385$$

(f) **Cumulative Distribution Function(cdf):** When X is a hypergeometric random variable with parameters n, M, and N. The cdf is denoted as:

$$F(x) = P(X \leq x) = \sum_{y=0}^{x} h(y; n, M, N)$$

(g) Example: Continuing with the animal tagging example in Northern BC, where $n = 10$, $M = 5$, and $N = 25$. Again, let $X \equiv \{$number of tagged animals in second sample$\}$.
What is the probability that $X \leq 2$?

- $P(X \leq 2) = F(2) = \sum_{x=0}^{2} h(x; 10, 5, 25) = p(0) + p(1) + p(2)$

$$P(X = 0) = h(0; 10, 5, 25) = \frac{\binom{5}{0}\binom{20}{10}}{\binom{25}{10}} = 0.057$$

$$P(X = 1) = h(1; 10, 5, 25) = \frac{\binom{5}{1}\binom{20}{9}}{\binom{25}{10}} = 0.257$$

$$P(X = 2) = h(2; 10, 5, 25) = \frac{\binom{5}{2}\binom{20}{8}}{\binom{25}{10}} = 0.385$$

- $P(X \leq 2) = \sum_{x=0}^{2} h(x; 10, 5, 25) = 0.057 + 0.257 + 0.385 = 0.699$

(h) **Expected Value for Hypergeometric Random Variable, X:** As noted earlier, the expected value (or mean) of any discrete random variable is a **probability weighted** sum of its values. When the hypergeometric pmf is substituted into the general relationship for expected value, the following result for the mean of X is obtained:

$$E(X) = \mu_X = n \cdot \frac{M}{N}$$

(i) The expected value or mean is generally not one of the values random variable X takes on.

(j) **Variance of Hypergeometric Random Variable, X:** As before, the *variance* of any discrete random variable X is the probability weighted average of squared deviations of its values from their mean. When the hypergeometric pmf and expected value are substituted into the definition, the following relationship results for the variance of X:

$$Var(X) = \sigma_X^2 = \left(\frac{N-n}{N-1}\right) \cdot n \cdot \frac{M}{N} \cdot \left(1 - \frac{M}{N}\right)$$

(k) The Standard Deviation of hypergeometric random variable X, denoted $SD(X)$, is computed as the positive square root of the variance of X, $Var(X)$.

$$SD(X) = \sqrt{Var(X)}$$

(l) Example: Continuing with the animal tagging example in Northern BC, where $n = 10$, $M = 5$, and $N = 25$. Again, let $X \equiv$ {number of tagged animals in second sample}.
What is the Expected Value of hypergeometric random variable X?

- $E(X) = n \cdot \left(\frac{M}{N}\right) = 10 \cdot \left(\frac{5}{25}\right) = 2.0$

What is the Variance of hypergeometric random variable X?

- $Var(X) = \left(\frac{N-n}{N-1}\right) \cdot n \cdot \frac{M}{N} \cdot \left(1 - \frac{M}{N}\right) = \frac{15}{24}(10)(.2)(.8) = 0.625(1.6) = 1.$

(m) **Finite Population Correction:** The ratio $\left(\frac{M}{N}\right)$ is the proportion of S's in the population. If we replace $\left(\frac{M}{N}\right)$ with p in the above relationships for $E(X)$ and $Var(X)$, we get

$$\begin{aligned} E(X) &= n \cdot \left(\frac{M}{N}\right) = np \\ Var(X) &= \left(\frac{N-n}{N-1}\right) \cdot n \cdot \frac{M}{N} \cdot \left(1 - \frac{M}{N}\right) \\ &= \left(\frac{N-n}{N-1}\right) \cdot np(1-p) \end{aligned}$$

- For expected values, the hypergeometric rv has $E(X) = np$, which is identical to that of a binomial rv. The means are equal.

- But for variances, the hypergeometric and binomial differ by a factor of $\left(\frac{N-n}{N-1}\right)$. This factor is sometimes called the **finite population correction factor**.

- The factor is always less than 1 in value, so the hypergeometric rv has a smaller variance than the binomial. This factor can also be written in the following form:

$$\frac{(1 - n/N)}{(1 - 1/N)} < 1$$

- When n is very small relative to N, the above factor approximately equals 1 and the two rv variances are approximately equal. Under these conditions

$$\text{(Hypergeometric rv)} \approx \text{(Binomial rv)}$$

- Comparing the two types of random variables:

Binomial	Hypergeometric
$E(X) = np$	$E(X) = n \cdot \left(\frac{M}{N}\right) = np$
$Var(X) = np(1-p)$	$Var(X) = \left(\frac{N-n}{N-1}\right) \cdot np(1-p)$
Sampling dichotomous rv	Sampling dichotomous rv
(with replacement)	(without replacement)
Independent trials	
Constant p	

(n) Example: A digital camera comes in either a 3-megapixel version or a 4-megapixel version. A camera store has just recieved a shipment of 15 of these cameras, of which 6 have 3-megapixel resolution. Suppose that 5 of these cameras are randomly chosen to be stored behind the counter, while the other 10 are placed in a storeroom. Let $X \equiv \{$number of 3-megapixel cameras selected for storage behind the counter$\}$. So we'll consider choosing a 3-megapixel camera a success, S.

i. If X is a hypergeometric rv, what are its distribution parameter values?
   - Sample size: $n = 5$; Number of S's in population: $M = 6$; Population size: $N = 15$.

ii. What is $P(X = 2)$?

$$P(X = 2) = h(2; 5, 6, 15) = \frac{\binom{6}{2}\binom{9}{3}}{\binom{15}{5}} = 0.280$$

iii. What is $P(X \leq 2)$?

$$P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2)$$

$$= \sum_{x=0}^{2} h(x; 5, 6, 15) = \frac{\binom{6}{0}\binom{9}{5}}{\binom{15}{5}} + \frac{\binom{6}{1}\binom{9}{5}}{\binom{15}{5}} + 0.280 = \frac{126 + 756 + 840}{3003} = 0.573$$

iv. What is $P(X \geq 2)$?

$$P(X \geq 2) = 1 - P(X \leq 1) = 1 - (P(X = 0) + P(X = 1))$$
$$= 1 - (\frac{126 + 756}{3003}) = 0.706$$

v. Calculate the mean and standard deviation of X.

- $E(X) = n \cdot \frac{M}{N} = 5(6/15) = 2.$
- $Var(X) = \left(\frac{N-n}{N-1}\right) \cdot n \cdot \frac{M}{N} \cdot \left(1 - \frac{M}{N}\right) = 2(1 - \frac{6}{15}) \cdot \frac{10}{14} = 0.857$
- $SD(X) = \sqrt{0.857} = 0.926$

Poisson Random Variables

5. **Poisson Random Variables:** Poisson random variables are discrete random variables that result from the occurrence of events that occur over time or space.

   (a) Poisson random variable X is a count of the number of events occuring either during time period of length t, or within area, A.

   (b) Poisson random variable X can only take on positive integer values, that is $X = 0, 1, 2, ...$

   (c) The Poisson distribution is a 1 parameter distribution, depending only on the parameter λ where $\lambda = \alpha \cdot t$ (or $\mu = \lambda A$).

   (d) Parameter λ has the same units as random variable X, number of counts.

   (e) α is the rate of the process and has units of either $time^{-1}$ or $area^{-1}$.

6. **Poisson distribution(pmf):** The probability of x events occuring within time period t for a Poisson random variable with parameter λ is:

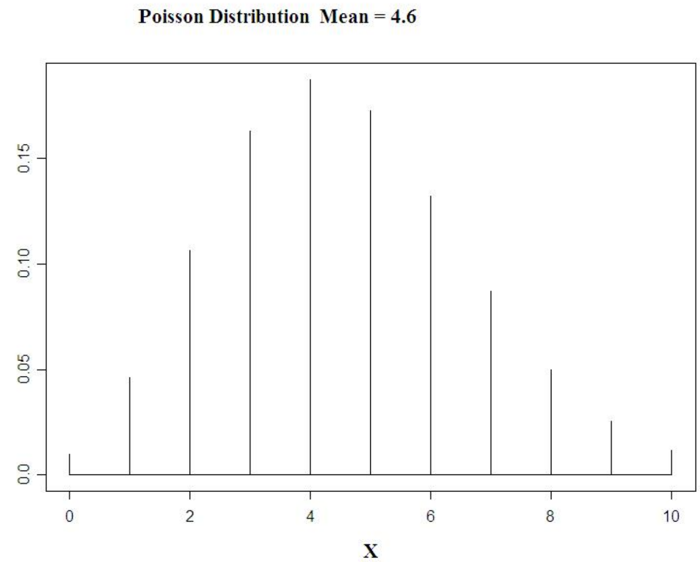
$$P(X = x) = p(x; \lambda) = \frac{e^{-\lambda}\lambda^x}{x!}$$

   - Where $X = 0, 1, 2, ...$ and $\lambda = \alpha \cdot t$
   - Note that λ is the expected value or mean of X, $E(X) = \lambda$.
   - here e is approximately 2.71828, $x! = x(x-1)\ldots(2)1$ and $0! = 1$
   - Poisson random variable X is distributed as:

$$X \sim p(x; \lambda)$$

Example(time): Suppose the nationwide number of deaths from typhoid fever over a 1-year period is Poisson distributed with $\lambda = 4.6$. What is the probability distribution of the number of deaths over a 6 months period?

- To evaluate X, the number of deaths in 6 months, first consider that $\lambda = 4.6$ for $t = 1$ yr so $\alpha = 4.6\ yr^{-1}$.
- for $t = 6$ months: $\lambda = 4.6(0.5) = 2.3$
- $P(X = 0) = \frac{e^{-2.3}(2.3)^0}{0!} = 0.100$
- $P(X = 1) = \frac{e^{-2.3}(2.3)^1}{1!} = 0.231$
- $P(X = 2) = \frac{e^{-2.3}(2.3)^2}{2!} = 0.265$
- $P(X = 3) = \frac{e^{-2.3}(2.3)^3}{3!} = 0.203$
- $P(X = 4) = \frac{e^{-2.3}(2.3)^4}{4!} = 0.117$
- $P(X = 5) = \frac{e^{-2.3}(2.3)^5}{5!} = 0.054$
- $P(X \geq 6) = 1 - (0.100 + 0.231 + 0.265 + 0.203 + 0.117 + 0.054) = 0.030$
- This Poisson distribution, with $\lambda = 4.6$ is plotted in the figure below. Note that it is a discrete distribution, a probability mass function



Example(area): If the distribution of the number of bacterial colonies on an agar plate of area A follows a Poisson distribution with $A = 100cm^2$ and rate parameter $\alpha = .02$ colonies/cm, calculate the probability distribution of the number of bacterial colonies.

(a) Let X be the number of colonies, with $\lambda = \alpha A = 100(.02) = 2.$.

- $P(X = 0) = \frac{e^{-2.}(2.)^0}{0!} = 0.135$
- $P(X = 1) = \frac{e^{-2.}(2.)^1}{1!} = 0.271$
- $P(X = 2) = \frac{e^{-2.}(2.)^2}{2!} = 0.271$
- $P(X = 3) = \frac{e^{-2.}(2.)^3}{3!} = 0.180$
- $P(X = 4) = \frac{e^{-2.}(2.)^4}{4!} = 0.090$
- $P(X \geq 5) = 1 - (0.135 + 0.271 + 0.271 + 0.180 + 0.090) = 0.053$

7. **Cumulative Distribution Function(cdf):** When X is a poisson random variable with parameter λ. The cdf is denoted as:

$$F(x) = P(X \leq x) = \sum_{y=0}^{x} p(y; \lambda)$$

(a) Example: Let $X \equiv$ {number of flaws on surface of a randomly selected boiler or a certain type}, have a Poisson distribution with parameter, $\lambda = 5$. Compute the following probabilities. Use Table A.2 in the text(pp. 738-9).

a.) What is $P(X \leq 8)$?

- $P(X \leq 8) = F(8; 5) = \sum_{x=0}^{8} p(x; 15) = 0.932$

b.) What is $P(X = 8)$?

- $P(X = 8) = P(X \leq 8) - P(X \leq 7) = F(8; 5) - F(7; 5) = 0.932 - 0.867 = 0.065$

c.) What is $P(X \geq 9)$?

- $P(X \geq 9) = 1 - P(X \leq 8) = 1 - 0.932 = 0.068$

d.) What is $P(5 \leq X \leq 8)$?

- $P(5 \le X \le 8) = P(X \le 8) - P(X \le 4) = 0.932 - 0.440 = 0.492$

e.) What is $P(5 \ < \ X \ < \ 8)$?

- $P(5 \ < \ X \ < \ 8) = P(X \le 7) - P(X \le 5) = 0.867 - 0.616 = 0.251$

(b) **Poisson Mean and Variance:** For a Poisson distribution with parameter λ, both the mean (expected value) and the variance are equal to λ.

$$E(X) = Var(X) = \lambda$$

- Discrete data(counts) where the mean and variance are approximately equal, are often Poisson distributed. These are sometimes called unbounded counts.

(c) **Poisson approximation to Binomial:** A Binomial distribution with large n and small p can be accurately approximated by the Poisson distribution with parameter $\lambda = np$.

i. Use this approximation when $n \ge 100$ and $p \le .01$

ii. Binomial probabilities may become more difficult to compute than Poisson when n gets very large.

Example: Suppose we conduct a study concerning genetic susceptibility to breast cancer and find that 4 out of 1000 women aged 40-49 whose mothers have had breast cancer also develop breast cancer over the next year of life. The probability of developing breast cancer over a 1 year period has been determined from large population studies to be 1 in 1000 for women from this group. How extreme are our findings?

- Let X be the number of women developing cancer over the next year out of $n = 1000$ trials with $p = \frac{1}{1000} = .001$. So X is a binomial random variable with $n = 1000$ and $p = .001$.
- What is the probability of a result more extreme than ours?
  $P(X \ge 4) = 1 - P(X \le 3) = 1 - (P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3))$
- $P(X = 0) = \begin{pmatrix} 1000 \\ 0 \end{pmatrix} .001^0(.999^{1000}) = .3677$
- $P(X = 1) = \begin{pmatrix} 1000 \\ 1 \end{pmatrix} .001^1(.999^{999}) = .3681$
- $P(X = 2) = \begin{pmatrix} 1000 \\ 2 \end{pmatrix} .001^2(.999^{998}) = .1840$
- $P(X = 3) = \begin{pmatrix} 1000 \\ 3 \end{pmatrix} .001^3(.999^{997}) = .0613$
- $P(X \ge 4) = 1 - P(X \le 3) = 1 - (.3677 + .3681 + .1840 + .0613) = .0189$

- These Binomial probabilities aren't so easy to calculate.

iii. Use the Poisson approximation to compute the above binomial probabilities.

- Since $n \geq 100$ and $p \leq .01$, the Poisson approximation with $\lambda = np = 1000(.001) = 1$ should work well.
- We can use Table A.2 in the text with $\lambda = 1$ to get:
- $P(X = 0) = P(X \leq 0) = .368$
- $P(X = 1) = P(X \leq 1) - P(X \leq 0) = .736 - .368 = .368$
- $P(X = 2) = P(X \leq 2) - P(X \leq 1) = .920 - .736 = .184$
- $P(X = 3) = P(X \leq 3) - P(X \leq 2) = .981 - .920 = .061$
- $P(X \geq 4) = 1 - P(X \leq 3) = 1 - .981 = .019$

x	Exact Binomial	Poisson Approx.
0	.368	.368
1	.368	.368
2	.184	.184
3	.061	.061
≥ 4	.019	.019