Discrete Random Variables

CSGE603012 – Statistics and Probability

Fakultas Ilmu Komputer

Universitas Indonesia

2017

Credits

These course slides were prepared by Alfan F. Wicaksono. **Suggestions**, **comments**, and **criticism** regarding these slides are welcome. Please kindly send your inquiries to alfan@cs.ui.ac.id.

Technical questions regarding the topic should be directed to current lecturer team members:

- Ika Alfina, S.Kom., M.Kom.
- Prof. T. Basaruddin, Ph.D.
- Alfan F. Wicaksono

The content was based on previous semester's (odd semester 2013/2014) course slides created by **all previous lecturers**.

References

- Introduction to Probability and Statistics for Engineers & Scientists, 4th ed.,
  - Sheldon M. Ross, Elsevier, 2009.
- A First Course in Probability, 8th Edition.
  - Sheldon M. Ross
- Applied Statistics for the Behavioral Sciences, 5th Edition,
  - Hinkle., Wiersma., Jurs., Houghton Mifflin Company, New York, 2003.
- Probability and Statistics for Engineers & Scientists, 4th Edition
  - Anthony J. Hayter, Thomson Higher Education

Outline

- Bernoulli Random Variables
- Binomial Random Variables
- Geometric Random Variables
- Hypergeometric Random Variables
- Poisson Random Variables

Bernoulli and Binomial Random Variables

Bernoulli Random Variables

Let **X** be the random variable with only two possible values.

X = 1 when the outcome is a "**success**"

X = 0 when the outcome is a "**failure**"

To model:

- The outcome of a coin toss {**Head**, **Tail**}
- Whether a valve is **open** or **shut**
- Whether an item is **defective** or **not**

www.silvermagpie.co.uk

Bernoulli Random Variables

The probability mass function (PMF) of **X** is given by

$$P(X=1)=p$$
$$P(X=0)=1-p$$

or

$$P(X=x)=p^x(1-p)^{1-x} \qquad x=0,1$$

where **p**, **0 ≤ p ≤ 1**, is the probability that the trial is a "success".

If **X** is a bernoulli R.V. (variable that has bernoulli distribution) with parameter **p**, we can write $X \sim Ber(p)$

Not really useful, but serves as a basis for other important random var !

Bernoulli Random Variables

Expectation/Mean

$$\mu = E[X] = 1.P(X=1) + 0.P(X=0) = p$$

Variance

$$\begin{aligned}\sigma^2 &= Var(X) \\ &= E[X^2] - (E[X])^2 \\ &= 1^2.P(X=1) + 0^2.P(X=0) - (p)^2 \\ &= p(1-p)\end{aligned}$$

CDF

$$F_X(x) = \begin{cases} 0 & x < 0 \\ 1-p & 0 \le x < 1 \\ 1 & x \ge 1 \end{cases}$$

Binomial Random Variables

Consider an experiment consisting of

- ***n*** Bernoulli Trials (X_1, X_2, ..., X_n)
- that are independent, and
- that each has a **constant** probability ***p*** of success.

If **X** represents the number of successes that occur in the **n trials**, i.e., $X = X_1 + X_2 + ... + X_n$, then **X** is said to be a *binomial* random variable (variable that has *binomial distribution*) with parameters **(n, p)**.

$$X \sim B(n, p)$$

Binomial Random Variables

The probability mass function of **X**, where **X ~ B(n, p)**

$$P(X=x)=\binom{n}{x}p^{x}(1-p)^{n-x} \qquad x=0,1,2,\ldots,n$$

A Binomial R.V. can be viewed as the **sum** of **n independent** Bernoulli R. V.

$$X=X_1+X_2+X_3+\ldots+X_n \qquad X_i \sim Ber(p)$$

Expectation

$$\begin{aligned}E[X]&=E[X_1+X_2+\ldots+X_n]\\&=E[X_1]+\ldots+E[X_n]=np\end{aligned}$$

Variance

$$\begin{aligned}Var(X)&=Var(X_1+X_2+\ldots+X_n)\\&=Var(X_1)+\ldots+Var(X_n)=np(1-p)\end{aligned}$$

Binomial Random Variables

Example:

1000 people are polled in a survey and asked if they support Jokowi.

- The responses are **YES** or **NO**
- The probability of a person saying **YES** is **p**

If **X** represents the number of **YES** occurs, then **X** is a binomial random variable with parameter **p**.

$$X \sim B(1000, p)$$

Why ? Because three characteristics of Binomial R.V. hold !

Binomial Random Variables

Example of Non Binomial R.V. !

- You stand in front of FASILKOM. Let,

  X = number of students passing by in the next 5 minutes

  (no fixed number of trials !)

- Gather a random sample of five men and five women. Let,

  X = number of persons out of 10 who are more than 170 cm tall

  (successes on trials do not have equal probability, men & women have different probability)

- Draw 4 cards (**without replacement**) from a deck of 52 cards. Let,

  X = number of aces among the four

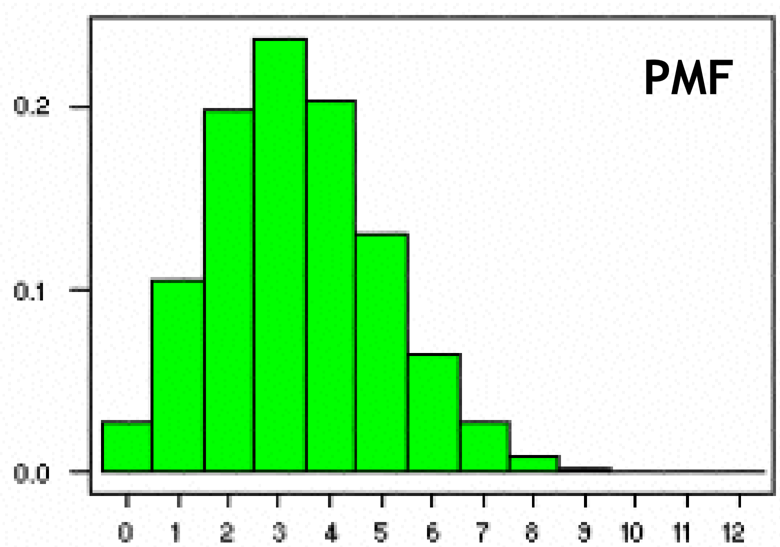
  (Trials are not independent due to replacement)

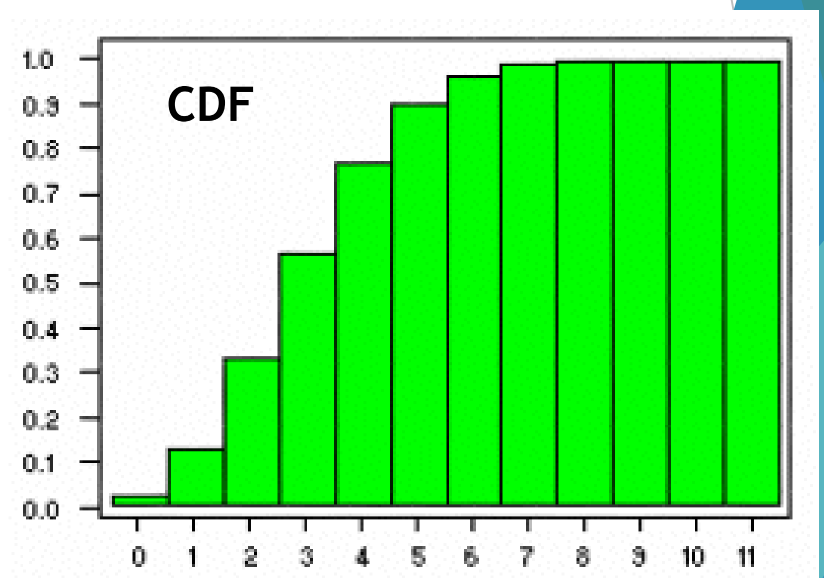
Binomial Random Variables

CDF of Binomial R.V.

$$F_X(i) = P(X \le i) = \begin{cases} 0 & i < 0 \\ \sum_{k=0}^{\lfloor i \rfloor} \binom{n}{k} p^k (1-p)^{n-k}, & 0 \le i < n \\ 1 & i \ge n \end{cases}$$

PMF & CDF of the **B(20, 1/6):**





http://www.stat.yale.edu/Courses/1997-98/101/binom.htm

Binomial Random Variables

Example: **X ~ B(8, 0.5)**

P (X = 3) ?   P (X ≤ 1) = ?

$$P(X=3)=\binom{8}{3}(0.5)^3(1-0.5)^5=0.219$$

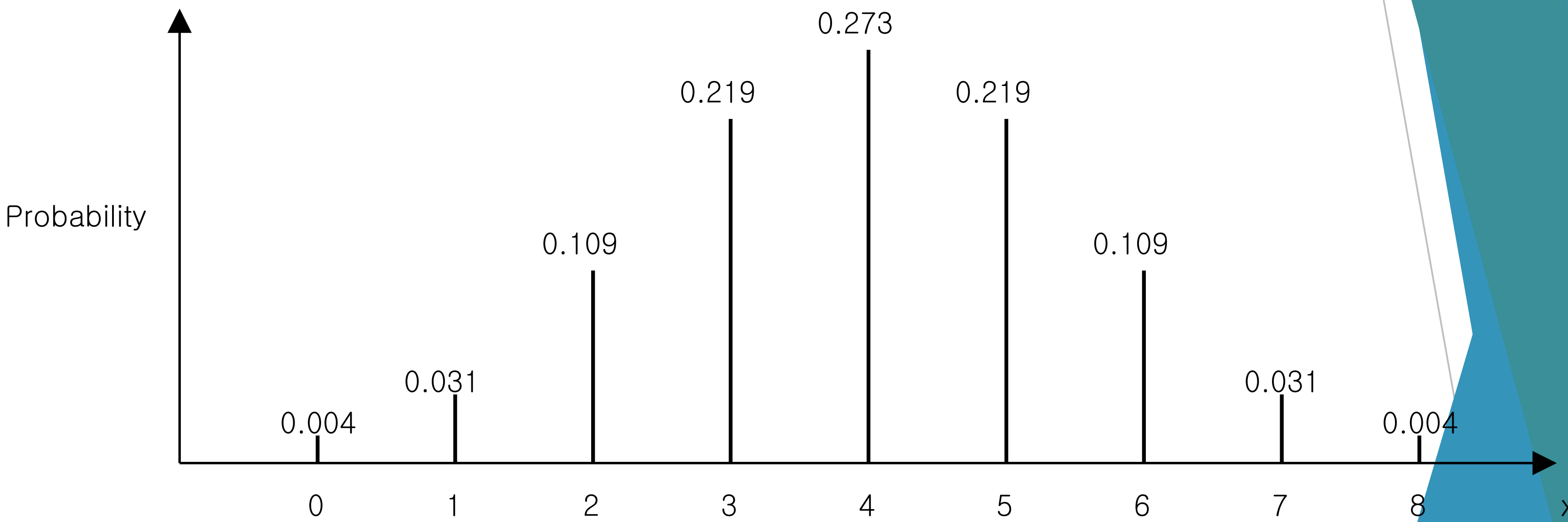
$$P(X\le 1)=P(X=0)+P(X=1)$$

$$=\binom{8}{0}(0.5)^0(1-0.5)^8+\binom{8}{1}(0.5)^1(1-0.5)^7$$

$$=0.035$$

Binomial Random Variables

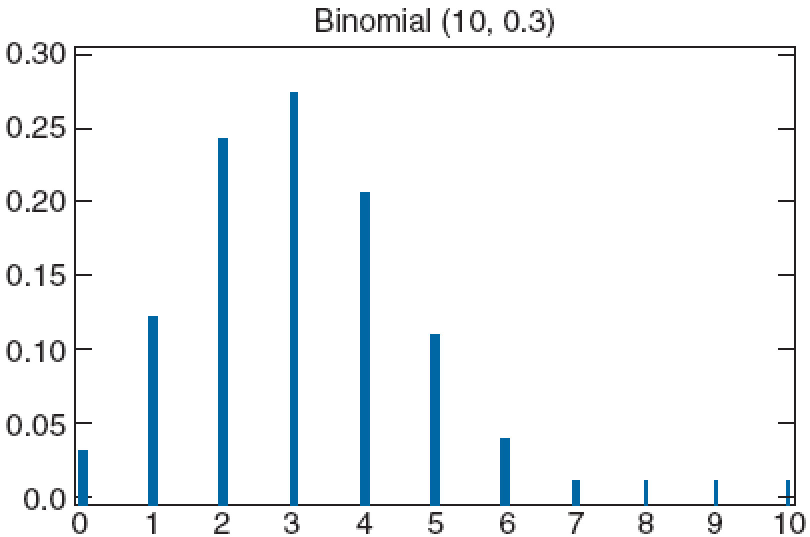
Example: **X ~ B(8, 0.5)**

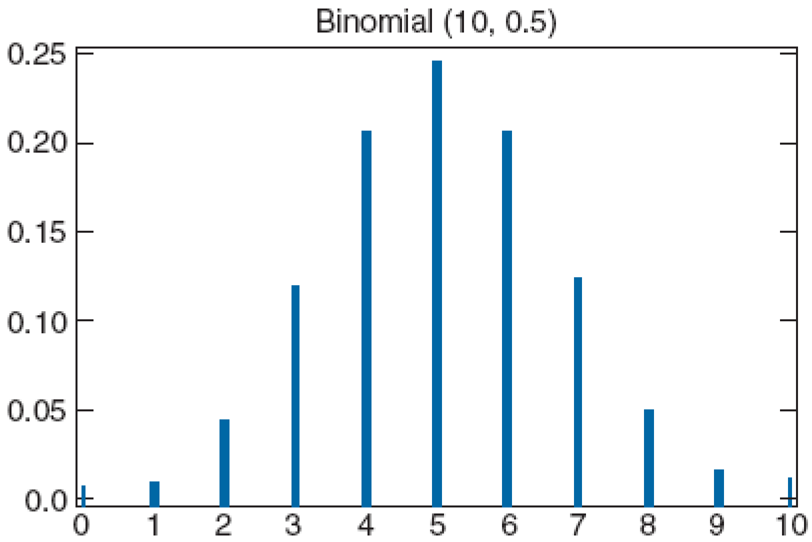


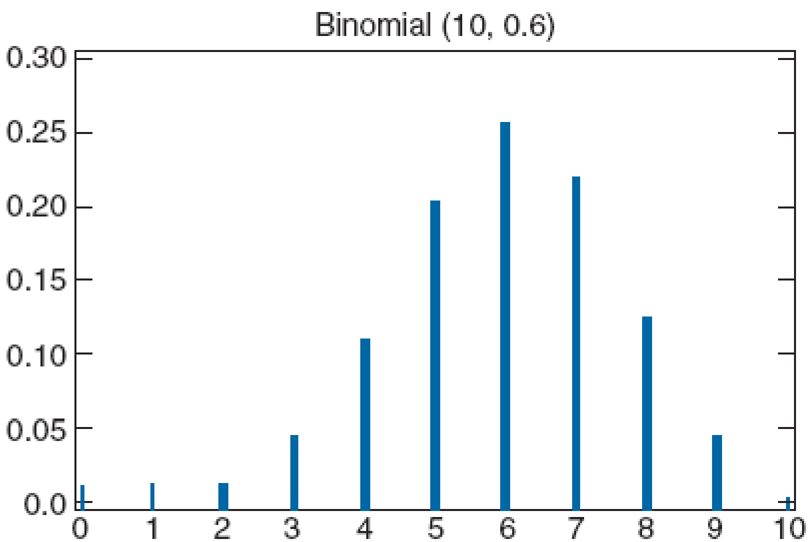
x	0	1	2	3	4	5	6	7	8
$P(X \leq x)$	0.004	0.035	0.144	0.363	0.636	0.855	0.965	0.996	1.0000

Probability & statistics course slides, by Prof. Kil Rhee-Man, NIPRL Lab, KAIST, 2011

Binomial Random Variables







Binomial Random Variables

Let random variable **X** be the number of **Heads** obtained from an experiment of tossing **three** coins. Suppose, probability that **Head** occurs when we toss a coin is **p**.

(a) What are possible values for **X** ?

(b) What kind of distribution does **X** follow ?

(c) Determine **PMF** of X !

(d) Determine **CDF** of X !

Binomial Random Variables

(a) 0, 1, 2, 3 => why ?

(b) $$X \sim B(3, p)$$

(c) **PMF**

$$P(X = x) = \binom{3}{x} p^x (1-p)^{3-x} \qquad x = 0, 1, 2, 3$$

(d) **CDF**

$$F_X(a) = P(X \le a) = \begin{cases} 0 & a < 0 \\ \sum_{k=0}^{\lfloor a \rfloor} \binom{3}{k} p^k (1-p)^{3-k}, & 0 \le a < 3 \\ 1 & a \ge 3 \end{cases}$$

Practice

Diketahui bahwa 40% dari tikus yang disuntik dengan sejenis serum ternyata terlindung dari serangan penyakit. Bila 5 tikus disuntik, berapakah peluang bahwa:

- Tidak ada yang terserang penyakit tersebut
- Kurang dari 2 yang terserang penyakit
- Antara 2 sampai 4 tikus yang terserang penyakit
- Paling sedikit 3 yang terserang penyakit

Binomial Random Variables

A company produces disks with the defect rate of 0.01. One pack of disk contains 10 discs. The company policy is to give money back if one pack contains more than 1 defect disc.

If someone buy three packs, what is the probability that exactly one pack is to be returned ?

Let **X** = the number of defect disc in one pack

It's clear that **X ~ B(10, 0.01)**

$$P(X > 1) = 1 - P(X = 0) - P(X = 1)$$

$$= 1 - \binom{10}{0}(0.01)^0(0.99)^{10} - \binom{10}{1}(0.01)^1(0.99)^9$$

$$\approx 0.005$$

Binomial Random Variables

After we compute the probability that a package will have to be replaced is **P(X > 1)** = 0.005.

Now,

Y = the number of packages that the person will have to return when he buys three packs.

Y ~ B(3, 0.005)

So,

$$P(Y=1)=\binom{3}{1}(0.005)^1(0.995)^2=0.015$$

Binomial Random Variables

Testing your understanding !

If **X** and **Y** are independent binomial random variables having respective parameters (n_x, p) and (n_y, p).

What is the probability distribution of new random variable **X + Y** ?

And what are the parameters of this distribution ?

Binomial Random Variables

X ~ B(n, p)

Computing **P(X = x)** recursively !

It can be shown that

$$P(X = k+1) = \frac{p}{1-p}\frac{n-k}{k+1}P(X = k)$$

For example, **X ~ B(6, 0.4)**

$$P(X = 0) = 0.046$$

$$P(X = 1) = \frac{4}{6}\frac{6}{1}P(X = 0) = 0.1866$$

$$P(X = 2) = \frac{4}{6}\frac{5}{2}P(X = 1) = 0.311$$

Geometric Random Variables

Geometric Random Variable

Let **X** be the number of trials **up to and including** the **first success** in a sequence of independent Bernoulli trials, each having a probability ***p*** of being success.

We say **X** is a **Geometeric Random Variable** (variable that has geometric distribution) with parameter **p**.

$$X \sim Geo(p)$$

The PMF of **X** is

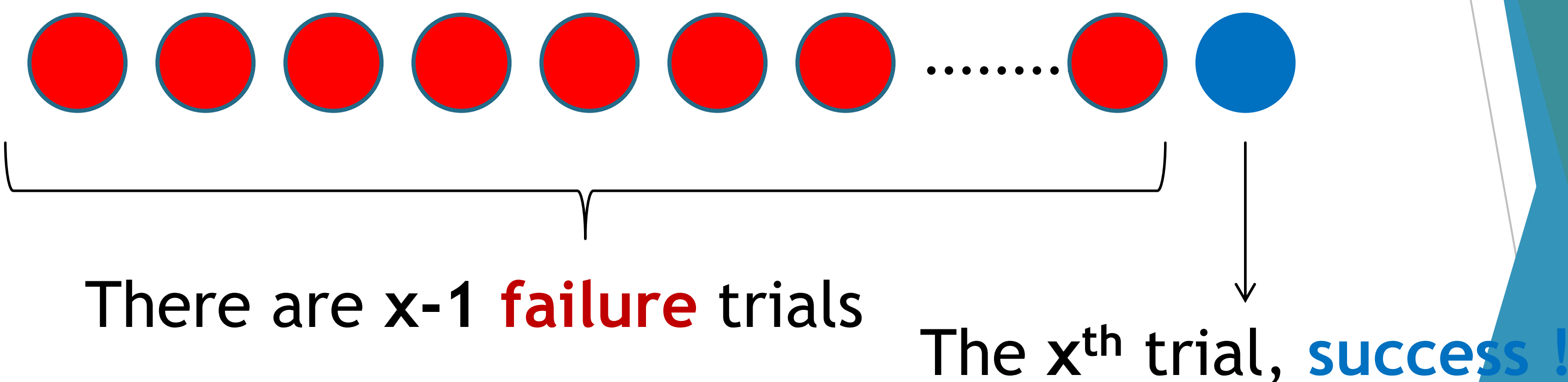
$$P(X = x) = (1 - p)^{x-1} p$$

$$x = 1,2,3,4,\ldots$$

Geometric Random Variable

The number of trials until the first success, with the probability of success in each trial ***p***.

In this case, we also count the first success !



The probability until the **first success** is clearly:

$$(1-p)(1-p)\ldots(1-p)\,p$$

That's why

$$P(X=x)=(1-p)^{x-1}\,p$$

Geometric Random Variable

The Cumulative Distribution Function (CDF) is

$$F_X(x) = P(X \le x) = \sum_{r=1}^{\lfloor x \rfloor} p(1-p)^{r-1} = 1-(1-p)^{\lfloor x \rfloor}, \qquad x \ge 1$$

Or, the complete form => $$F_X(x) = \begin{cases} 0 & x < 1 \\ 1-(1-p)^{\lfloor x \rfloor} & x \ge 1 \end{cases}$$

The mean (expectation) & variance:

$$\mu = E[X] = \frac{1}{p} \qquad \sigma^2 = Var(X) = \frac{1-p}{p^2}$$

Geometric Random Variable

$$E[X] = \sum_{x=1}^{\infty} x(1-p)^{x-1} p = p \sum_{x=1}^{\infty} x(1-p)^{x-1} = p.\frac{1}{p^2} = \frac{1}{p}$$

$$Var(X) = E[X^2] - (E[X])^2$$

$$E[X^2] = \sum_{x=1}^{\infty} x^2(1-p)^{x-1} p = p \sum_{x=1}^{\infty} x^2(1-p)^{x-1}$$

$$\sum_{x=1}^{\infty} x^2(1-p)^{x-1} = \frac{1}{p}\left(\frac{1}{p} + \frac{2(1-p)}{p^2}\right) = \frac{2-p}{p^3}$$

$$Var(X) = p\left(\frac{2-p}{p^3}\right) - \frac{1}{p^2} = \frac{1-p}{p^2}$$

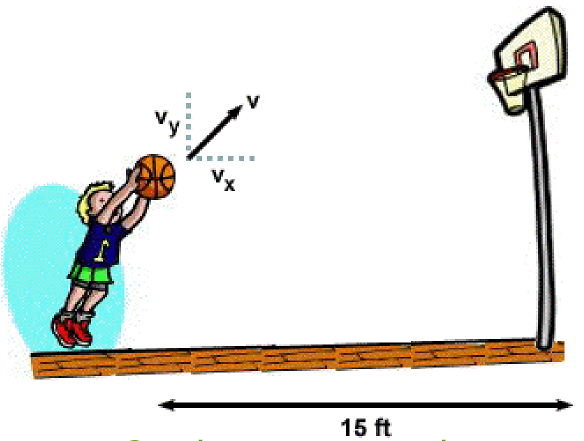
Geometric Random Variable

Peluang bahwa Budi berhasil memasukkan sebuah bola ke dalam ring bola basket adalah 0.6. misal, lemparan yang satu dengan yang lainnya bersifat *independent*.

1. Berapa peluang bahwa Budi membutuhkan 3 lemparan hingga akhirnya memasukkan bola ?

2. Berapa peluang bahwa Budi membutuhkan paling sedikit 3 lemparan untuk memasukkan bola ?

3. Perkirakan berapa banyak lemparan yang harus dilakukan agar lemparannya berhasil masuk ring ?



Gambar: www.sparknotes.com

Geometric Random Variable

Jawab:

Misal, **X** : R.V. yang menyatakan banyaknya lemparan yang dilakukan hingga akhirnya bola berhasil masuk ring.

$$X \sim Geo(0.6)$$

1. $P(X = 3) = (1-p)^2 p = (0.4)^2(0.6) = 0.096$

2. $$\begin{aligned} P(X \geq 3) &= 1 - P(X=1) - P(X=2) \\ &= 1 - (0.4)^0(0.6) - (0.4)^1(0.6) \\ &= 1 - 0.4 - 0.24 = 0.36 \end{aligned}$$

3. $E[X] = \dfrac{1}{p} = \dfrac{1}{0.6} = 1.67$

Geometric Random Variable

Example:

If a person is unsuccessful in starting the old car's engines, then he must wait 10 minutes before trying again. In each attempt, the success probability is 0.75.

Q1 What is the probability that the old car's engines start on **the third** attempt ?

X : the number of trials until the engines start

$$P(X = 3) = (0.25)^2.(0.75) = 0.047$$

Geometric Random Variable

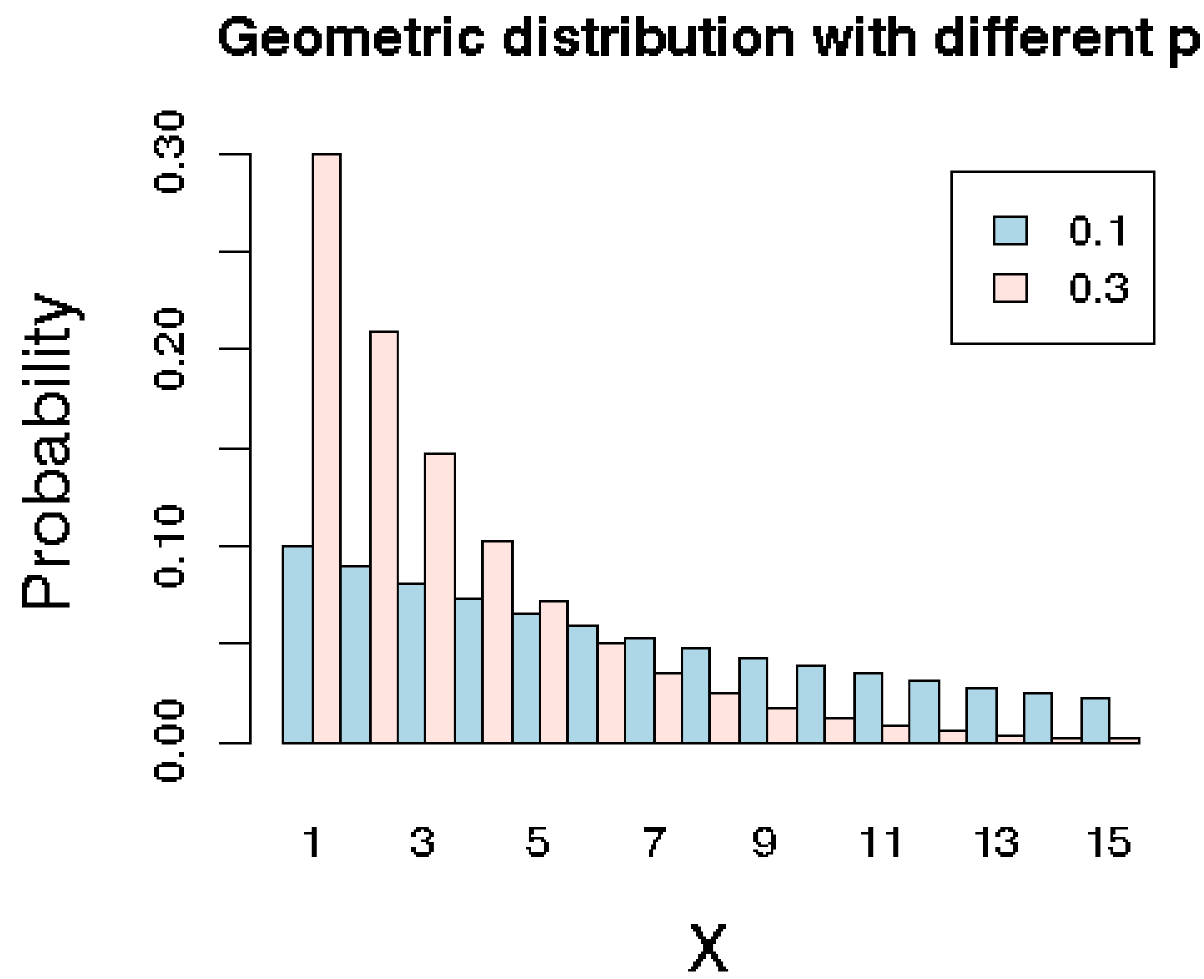
Q2 What is the probability that the engines start **within 20 minutes** of the first attempt ?

Q3 The expected number of attempts required to start the engines ?

$$P(X \leq 3) = 1 - (1 - 0.75)^3 = 0.984$$

$$E[X] = \frac{1}{p} = \frac{1}{0.75} = 1.33$$

Geometric Random Variable



http://raven.iab.alaska.edu/~ntakebay/teaching/programming/probability/node8.html

Geometric Random Variable

Another example ....

- It is known that in a large accounting population 3% of accounts are in error. Accounts are inspected until first account in error is encountered.

  Let X = number of inspections to obtain first account in error.

- Roll a die until first 5.

  Let X = number of roll until first 5.

Geometric Random Variable has *Memoryless Property*

A nonnegative random variable **X** is *memoryless* if

$$P(X > r+m \mid X > m) = P(X > r) \qquad r, m \geq 0$$

This property states that the value of **r** in the probability **P(X = r)** is always calculated from **current condition** (we don't care about how many previous failures).

- condition has occurred: failure in all previous **m** trials, so that **"X > m"** as a prior condition, and
- we will calculate the probability of success will happen in the **(m+r)**th experiment with this prior condition.

Geometric Random Variable has *Memoryless Property*

Proof:

$$
\begin{aligned}
P(X > r+m \mid X > m) &= \frac{P(X > r+m, X > m)}{P(X > m)} \\
&= \frac{P(X > r+m)}{P(X > m)} \\
&= \frac{1 - P(X \le r+m)}{1 - P(X \le m)} \\
&= \frac{1-(1-(1-p)^{r+m})}{1-(1-(1-p)^{m})}
\end{aligned}
$$

$$
\begin{aligned}
&= (1-p)^r \\
&= 1-(1-(1-p)^r) \\
&= P(X > r)
\end{aligned}
$$

Other consequence:

$$P(X > r+m) = P(X > r)P(X > m)$$

Test your understanding !

In a sequence of bernoulli experiments to get number ‘6’ in rolling a die:

- What is the probability to get 6 successes in 10 experiments?
- What is the probability to get the first success at the 10th experiment?
- What is the average number of successes in 10 experiments?
- What is the average number of experiments to get the first success?

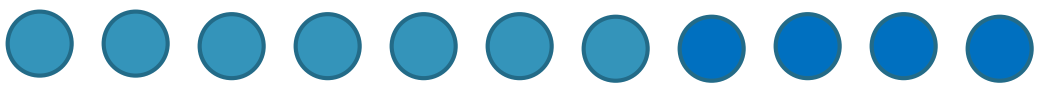
Hypergeometric Random Variables

Hypergeometric Random Variables

Definition

Consider a collection of **N** items of which **r** are of a certain/special kind.

r special items



There are **N** items, all

If one of the **N** items is chosen at random, the probability that it is a special kind is:

$$p = \frac{r}{N}$$

So, if **n** items are chosen at random **with replacement**, it is clear that **X**, the number of special items chosen, is

$$X \sim B(n, \frac{r}{N})$$

Hypergeometric Random Variables

Definition

However, if **n** items are chosen at random **without replacement**, then **X**, the number of special items chosen, is the **hypergeometric R. V.**

$$X \sim H(n, r, N)$$

The PMF:

$$P(X = x) = \frac{\binom{r}{x} \times \binom{N-r}{n-x}}{\binom{N}{n}}$$

for

$$\max\{0, n+r-N\} \leq x \leq \min\{n, r\}$$

Hypergeometric Random Variables

The Expectation & Variance

$$E[X] = n\frac{r}{N}$$

Prove it !

$$Var(X) = \frac{nr(N-n)(N-r)}{N^2(N-1)}$$

It represents the distribution of the number of items of a certain kind in a random sample of size **n** drawn ***without replacement*** from a population of size **N** that contains **r** items of this kind.

Hypergeometric Random Variables

From inside the box containing **10 ping pong balls**, **4 balls** are taken randomly. Among those 10 balls, there are **3 red** and **7 white** balls. Find the probability of the 4 balls have been taken, there is **1 red ball** at the most !

This question is related to Hypergeometric R.V. with **x=0** or **x=1**. Given the information that **N=10**, **n=4**, and **r=3**.

$$P(X \le 1) = P(X = 0) + P(X = 1) = \frac{\binom{3}{0} \times \binom{7}{4}}{\binom{10}{4}} + \frac{\binom{3}{1} \times \binom{7}{3}}{\binom{10}{4}} = \frac{2}{3}$$

Hypergeometric Random Variables

Binomial Approximation

If the population size **N** is **much bigger** than the number of items taken **n**, then Binomial R.V. will be a reasonably good approximation for hypergeometric R.V.

Let ***p*** **= r / N**,

$$E[X] = n\frac{r}{N} = np$$

$$Var(X) = \frac{nr(N-n)(N-r)}{N^2(N-1)} = np(1-p)\frac{N-n}{N-1}$$

When, **N goes to infinity**, then

$$Var(X) = np(1-p)$$

Hypergeometric Random Variables

Binomial Approximation

From the previous example, if **N = 100**, **r = 30**, and **n = 4**.

We see that **N >> n !** We can approximate using binomial R.V.

$$P(X \le 1) = \binom{n}{0} p^0 (1-p)^n + \binom{n}{1} p^1 (1-p)^{n-1}$$

$$= \binom{4}{0} (0.3)^0 (1-0.3)^4 + \binom{4}{1} (0.3)^1 (1-0.3)^3 = 0.6517$$

If you use **hypergeometric R.V.**, you will get **0.6516 !** Not much different !

Hypergeometric Random Variables

Just another example ...

A small lake contains 50 fish. One day, a fisherman catches 10 of these fish and tags them so that they can be recognized if they are caught again. The tagged fish are released back into the lake. The next day, the fisherman goes out and catches 8 fish, which are kept in the fishing boat until they are all released at the end of the day.

What is the distribution of **X**, the number of tagged fish caught on the second day ?

Variable **X** is clearly hypergeometric R.V. with **N = 50**, **r = 10**, and **n = 8**.

Hypergeometric Random Variables

Probability that 3 tagged fish are caught on the second day:

$$P(X=3)=\frac{\binom{10}{3}\times\binom{40}{5}}{\binom{50}{8}}=0.147$$

The expected number of tagged fish recaptured:

$$E[X]=\frac{nr}{N}=\frac{8\times 10}{50}=1.6$$

Suatu kotak berisi 40 suku cadang.

Cara sampling kotak ialah dengan memilih 5 suku cadang secara acak dari dalamnya dan menolak kotak tersebut bila diantaranya ada yang cacat.

Berapa peluang kotak tersebut ditolak bila kotak tersebut berisi 3 yang cacat ?

Hypergeometric R.V. & Binomial R. V.

Lex **X** and **Y** be independent binomial random variables:

$$X \sim B(r, p)$$

$$Y \sim B(N - r, p)$$

Let us consider the **conditional probability mass function** of **X** given that **X + Y = n**, denoted by

$$P(X = x \mid X + Y = n)$$

Hypergeometric R.V. & Binomial R. V.

$$P(X = x \mid X + Y = n) = \frac{P(X = x, X + Y = n)}{P(X + Y = n)} = \frac{P(X = x, Y = n - x)}{P(X + Y = n)}$$

$$P(X = x \mid X + Y = n) = \frac{P(X = x)P(Y = n - x)}{P(X + Y = n)}$$

$$= \frac{\binom{r}{x} p^x (1 - p)^{r-x} \binom{N - r}{n - x} p^{n-x} (1 - p)^{N-r-n+x}}{\binom{N}{n} p^n (1 - p) N - n}$$

$$= \frac{\binom{r}{x}\binom{N - r}{n - x}}{\binom{N}{n}}$$

The conditional distribution of **X** given the value of **X + Y** is **hypergeometric !**

Di kampung "senyum", ada 40 wanita dan 60 pria.

Banyaknya "senyuman" yang dilemparkan per hari oleh seorang pria adalah Poisson R.V. dengan rata-rata 10.

Sedangkan, banyaknya "senyuman" yang dilemparkan per hari oleh seorang wanita juga merupakan Poisson R.V. dengan rata-rata 5.

Misalkan, seseorang diambil secara acak dari kampung tersebut.

- Jika diketahui orang tersebut adalah pria, berapa peluang ia **tidak** tersenyum hari ini.
- Berapa peluang ia tersenyum 2 kali hari ini.

Practice

Sebuah grup terdiri dari 6 pria dan 6 wanita. Misalkan dipilih 4 orang secara acak dari 12 orang tersebut untuk membentuk sebuah kepanitiaan.

Berapakah peluang:

- Ada sebanyak minimal 2 pria pada kepanitiaan tersebut
- Ada pasangan pria-wanita pada kepanitiaan tersebut

Poisson Random Variables

It is often useful to define a random variable that **counts the number of "events"** that occur **within certain specified boundaries/interval.**

- The distribution of
  - The number of defects in an item
  - The number of emails coming in an hour
  - The number of telephone calls received by an operator with a certain time limit
  - The number of wrong telephone numbers that are dialed in a day
  - The number of misprints on a page (or a group of pages) of a book
  - The number of customers entering a post office on a given day
  - The number of transistors that fail on their first day of use
  - The number of people in a community living to 100 years of age

A random variable **X** distributed as a **Poisson** random variable with parameter **α**, which is written

$$X \sim P(\alpha) \qquad \alpha = \lambda t$$

has a probability mass function:

$$P(X = x) = e^{-\alpha} \frac{\alpha^x}{x!} \qquad x = 0, 1, 2, \ldots$$

X : Random variable that describes the **number of events/arrivals** within interval ***t***.

λ : **Average number** of **events** occuring **per unit** of interval (**rate**)

t : Length of **interval**

α : **Average number** of **events** occuring in **any** interval of length ***t***.

It is useful to model the number of **times that a certain event occurs per unit of time, distance, or volume.**

CDF:

$$F_X(x) = P(X \le x) = \begin{cases} 0 & i < 0 \\ \displaystyle\sum_{i=0}^{\lfloor x \rfloor} e^{-\alpha} \frac{\alpha^i}{i!} & i \ge 0 \end{cases}$$

Mean & Variance

$$E[X] = Var[X] = \alpha = \lambda t$$

Proof (optional):

Let's first determine its moment generating function

Poisson VS Binomial

Binomial

1. Discrete/whole numbers
2. A certain number of **opportunities** for the occurrence are given.
   - You flip a coin certain times and ask how many times Head appeared. "Opportunity = flipping coin"
3. Each trial/"flip" is independent of the others

Poisson

1. Discrete/whole numbers, but **Infinite**
2. **No special opportunities** for the events
   - An accident may happen at any times. No special opportunity when an accident happens or not.
3. An occurrence happening at one time interval is independent of another occurrence happening at another time interval.

Suppose that the **average number** of accidents occurring **weekly** on a particular stretch of a highway equals **3**. Calculate the probability that there is at least one accident **this week** !

Let,

X = the number of accidents during **this week**

t = 1 week

We can assume that **X** is a poisson random variable. The average number of accidents **per week** is **3** (**λ = 3**). so, we can write **X ~ P(λt) ~ P(3)**.

$$\alpha = \lambda t = 3(1) = 3$$

$$P(X \geq 1) = 1 - P(X = 0)$$

$$= 1 - e^{-3}\frac{3^0}{0!}$$

$$\approx 0.9502$$

A quality inspector at a glass manufacturing company inspects sheets of glass to check for any slight imperfections.

Suppose that the number of these flaws **X** in **a sheet of glass** has a Poisson distribution with **λ = 0.5** flaws per sheet.

Q1. The probability that there is no flaw in a sheet ?

Q2. The probability that there are two or more flaws in a sheet ?

λ = 0.5 implies that the expected number of flaws per sheet is only **0.5**.

***t* = 1 sheet**

$$\alpha = \lambda t = 0.5$$

$$X \sim P(0.5)$$

$$P(X=0) = \frac{e^{-0.5} \times (0.5)^0}{0!} = e^{-0.5} = 0.607$$

$$P(X \geq 2) = 1 - P(X=0) - P(X=1)$$

$$= 1 - \frac{e^{-0.5} \times (0.5)^0}{0!} - \frac{e^{-0.5} \times (0.5)^1}{1!} = 0.090$$

The arrival of a customer in a store has a Poisson distribution with $\lambda = 5$ visitors per hour.

1. The probability that visitors coming **less than 5** in **one hour** is ?
2. The number of visitors > 5 in an hour ?
3. The number of visitors =10 in a **day** ?
4. The expected number of visitor in an hour ?
5. The expected number of visitor in a day ?

Answer:

Only #1 and #3

(1) Let **X** : R.V. that describes the number of visitors coming in **one hour**

$$\lambda = 5, t = 1 \text{ hour}, \alpha = \lambda t = 5,\ X \sim P(5)$$

$$P[X < 5] = \frac{5^0 e^{-5}}{0!} + \frac{5^1 e^{-5}}{1!} + \frac{5^2 e^{-5}}{2!} + \frac{5^3 e^{-5}}{3!} + \frac{5^4 e^{-5}}{4!}$$

$$= e^{-5}(1 + 5 + 25/2 + 125/6 + 625/24)$$

$$= 0.440493$$

(3) Let **Y** : R.V. that describes the number of visitors coming in **a day**

$$\lambda = 5, t = 24 \text{ hour}, \alpha = \lambda t = 120,\ Y \sim P(120)$$

$$P[Y = 10] = \frac{120^{10} e^{-120}}{10!}$$

Misalkan, banyaknya tikus ladang di suatu area mengikuti distribusi Poisson. Rata-rata banyaknya tikus ladang per meter-persegi di suatu kampung ditaksir 12 ekor. Carilah:

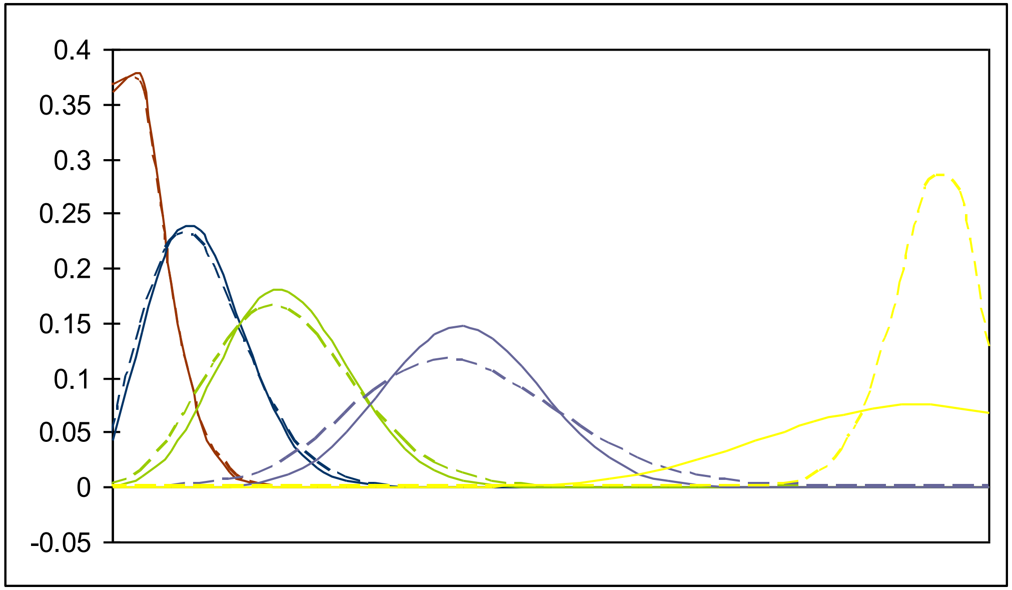
1) Peluang bahwa antara 2 hingga 4 tikus ditemukan pada daerah seluas 1 meter-persegi ?

2) Misal, daerah 3 meter-persegi dibagi menjadi 3 buah **area** A, B, dan C yang **tidak** beririsan (masing-masing 1 meter-persegi).

Berapa peluang bahwa tidak ada tikus yang ditemukan pada 2 dari 3 area yang diperiksa !

Approximation for a binomial R.V.

- Binomial pmf (full line) and Poisson pmf (dashed line) for $n = 30$; each colour for $np = \alpha = 1, 3, 6, 12,$ and 28.



- Both function graphs are nearly the same for $\alpha = \lambda t \leq 3$, while for other, the bigger the value of α the lower the Poisson pmf graphs compared to Binomial graphs.

Approximation for a binomial R.V.

Proof:

Let, **X ~ B(n, p)** and α **= np**.

$$P(X=i)=\frac{n!}{(n-i)!i!}p^i(1-p)^{n-i}$$

$$=\frac{n!}{(n-i)!i!}\left(\frac{\alpha}{n}\right)^i\left(1-\frac{\alpha}{n}\right)^{n-i}=\frac{n(n-1)\ldots(n-i+1)}{n^i}\frac{\alpha^i}{i!}\frac{(1-\alpha/n)^n}{(1-\alpha/n)^i}$$

Now, for **n large** and **p small**,

$$\left(1-\frac{\alpha}{n}\right)^n\approx e^{-\alpha},\quad \frac{n(n-1)\ldots(n-i+1)}{n^i}\approx 1,\quad \left(1-\frac{\alpha}{n}\right)^i\approx 1$$

So,

$$P(X=i)\approx e^{-\lambda t}\frac{\lambda t^i}{i!}$$

Approximation for a binomial R.V.

Poisson random variable can be used as an approximation for a binomial random variable with parameters **(n, p)** when **n** is **large** and **p** is **small**.

When we approximate using Poisson R.V., we use

$$\alpha = \lambda t = np$$

in some literatures, it is recommended for **n ≥ 30** and **np ≤ 3**.

Suppose the probability that an item produced by a certain machine will be defective is **0.1**.

Find the probability that a sample of **10 items will contain at most one defective item** !

Assume that the quality of successive items is independent.

X : the number of defective item on the sample

It's clear that **X ~ Binomial(10, 0.1)**, so

$$P(X \le 1) = \binom{10}{0}(0.1)^0(0.9)^{10} + \binom{10}{1}(0.1)^1(0.9)^9 = 0.7361$$

Whereas, we can also approximate using Poisson approximation:

$$P(X \le 1) \approx e^{-1}\frac{1^0}{0!} + e^{-1}\frac{1^1}{1!} \approx 0.7358$$

Dalam suatu proses produksi yang menghasilkan barang dari gelas, kadang-kadang terjadi cacat.

Diketahui bahwa 1 dari 1000 barang yang dihasilkan mempunyai kecacatan.

Berapakah peluang bahwa dalam sampel acak sebesar 3000 barang akan berisi 10 barang yang rusak ?

Diketahui dalam setengah hari sebuah toko kamera mampu menjual rata-rata 1,5 unit kamera.

Tentukan:

- Peluang toko tersebut mampu menjual lebih dari 1 unit kamera dalam suatu hari ?

  0,800852

- Peluang dalam kurun 5 hari, sebanyak 3 hari toko tersebut mampu menjual lebih dari 1 unit kamera per harinya ?

  0,203709

Di kampung "senyum", ada 40 wanita dan 60 pria.

Banyaknya "senyuman" yang dilemparkan per hari oleh seorang pria adalah Poisson R.V. dengan rata-rata 10.

Sedangkan, banyaknya "senyuman" yang dilemparkan per hari oleh seorang wanita juga merupakan Poisson R.V. dengan rata-rata 5.

Misalkan, seseorang diambil secara acak dari kampung tersebut.

- Jika diketahui orang tersebut adalah pria, berapa peluang ia **tidak** tersenyum hari ini.
- Berapa peluang ia tersenyum 2 kali hari ini.

Computing the Poisson Distribution Function

if it is calculated directly with a big enough **α**, it yields unstable results (i.e., due to real number rounding).

Instead, compute it recursively

$$\frac{P(X=i+1)}{P(X=i)} = \frac{e^{-\alpha}\alpha^{i+1}/(i+1)!}{e^{-\alpha}\alpha^{i}/i!} = \frac{\alpha}{i+1}$$

$$P(X=i+1) = \frac{\alpha}{i+1} P(X=i)$$