STA256H5S: Probability and Statistics I

Week 2 & 3: Discrete Random Variables

Thomas Yoon

Department of Mathematical and Computational Sciences
University of Toronto Mississauga

July 2018

Table of Contents

Table of Contents

Definition

A random variable X is **discrete** if it can only take countably infinite (or less) number of distinct values.
An example of countably infinite set is a set of positive integers.
Any set that can form one-to-one correspondence with a set of positive integers is also countably infinite.

Definition

- The probability that the random variable Y takes the value of y is denoted as $P(Y = y) = p_Y(y) = p(y)$.
- This quantity is the sum of all probabilities in sample space S s.t. sample point is y.
- The probability distribution for discrete R.V. Y explains how $P(Y = y)$ is distributed for all possible values of y. It can be represented as a function, table or graph.

Theorem

Recall: Probability Axioms
Following statements are true:

- $0 \leq p(y) \leq 1$
- $\sum_{\text{all } y} p(y) = 1$

Example

Suppose you roll a fair die two times, and you are interested in the sum. Let the sum be the random variable X.

- What are the possible values of X?
- Find the probability that X=3.
- Find the distribution of X.

Example

A computer contains four major components (let's just suppose this is true!). There are two defective components in this particular computer. Michelle tests the components one at a time until both defects are found. Let X denote the number of tests at which she finds the second defect. Find the probability distribution for X.

Example

Megan works for Statistics Canada. She works very hard but she makes mistakes in her calculations 5% of the time. Her supervisor, Elham randomly checks three of her calculations. Let Y be the number of calculations with mistakes Elham found.

- Find the probability distribution of Y.
- Graph the probability distribution of Y.
- Find the probability that Elham will detect at least 2 wrong calculations.
- Find the probability that Elham will detect at most 2 wrong calculations.

Definition

Let Y be a discrete random variable and $p(y)$ denote the probability mass function of Y (Recall: probability mass function is same as probability distribution for discrete random variable). Then the **expected value** of Y, $E(Y)$ is defined as:

$$E(Y) = \sum_{\text{all y}} yp(y)$$

(Note: Intuitively, at least for now, think of $E(Y)$ as the weighted mean of Y.)

Example

Students in STA256 class write a short quiz consisted of two multiple choice questions, each containing 5 options (a to e). Henry didn't study, and decided to guess. Let X be the number of questions he gets correct.

- Find the expected value of X.
- Let $Y = 3X$. Find the expected value of Y

Example

Anthony plays 100 dollar worth of black jack at a Casino almost everyday. He's been unlucky in the past few years. Let Y be the amount of money he walks out with after a visit ($y < 0$ means he lost that much, 0 means he broke even, and > 0 means he made a profit). Set $Z = Y^2$. Find $E(Y), E(Z)$. Consider the following table:

y	p(y)
-100	0.35
-75	0.30
-50	0.10
-25	0.09
0	0.09
25	0.07

UNIVERSITY OF TORONTO MISSISSAUGA

Theorem

Let Y be a discrete random variable with probability mass function $p(y)$, and $g(Y)$ be a real valued function of Y. Then the expected value of $g(Y)$ is:

$$E(g(Y)) = \sum_{\text{all y}} g(y)p(y)$$

Proof

Let Y take values $y_1, ..., y_n$, and surjective function $g(Y)$ take values $g_1, ..., g_m$ where $m \leq n$ ($\because g(Y)$ is a surjective function of Y) .
Then, $E(g(Y)) =$

$= \sum_{j=1}^{n} g(y_j)p(y_j)$

Corollary

Let X be a discrete random variable and $g_1(X), g_2(X)$ be functions of X. Then $E(g_1(X) + g_2(X)) = E(g_1(X) + E(g_2(X))$

Properties of Expected Values

Let X and Y be a random variable and a,b,c,d $\in \mathbb{R}$

1. $E(a) = a$
2. $E(aX) = aE(X)$
3. $E(X + c) = E(X) + c$
4. $E(aX + c) = aE(X) + c$
5. $E(aX + c + bY + d) = aE(X) + c + bE(Y) + d$
6. $X \perp\!\!\!\perp Y \implies E(XY) = E(X)E(Y), E(f(X)g(Y)) = E(f(X)E(g(Y))$

Note: I hope you see that Expected Value preserves linearity.

Proof

1. (Trivial) definition: if a $\in \mathbb{R}$ then $p(a = a) = 1$.
   $E(a) = \sum_{\text{all a}} aP(a = a) = aP(a = a) = a$
2. Let $g(X) = aX$. Then, we know
   $E(aX) = \sum_{\text{all x}} axP(X = x) = a \sum_{\text{all x}} xP(X = x) = aE(X)$

UNIVERSITY OF TORONTO MISSISSAUGA

Definition

Let X be a random variable. Then, **variance** of X, $var(X) = \sigma_X^2 = \sigma^2$ is the weighted average of squared distance from the mean.

$$\sigma_X^2 = E((X - E(X))^2) = E(X^2) - (E(X))^2$$

Standard deviation of X, σ_X is positive square root of variance.
$\sigma_X = |\sqrt{\sigma_X^2}|$

Example

Consider the following table and find $E(X), E(X^2), \sigma_X^2$.

x	p(x)
-5	0.2
-3	0.2
-1	0.2
1	0.2
3	0.2

UNIVERSITY OF TORONTO MISSISSAUGA

Same $E(X)$, same σ^2?

Consider these two cases:

- Student name Pip: she receives 0, 50, and 100%'s on her tests.
- Student name Julia: she receives 25,50, and 75%'s on her tests.

These two students have equal probability of receiving the aforementioned grades with probability of $1/3$ each. Find expected values, standard deviations and variances of these two students.

Example

A potential customer for an 85,000 fire insurance policy possesses a home in an area that, according to experience, may sustain a total loss in a given year with probability of 0.001 and a 50% loss with probability 0.01. Ignoring all other partial losses, what premium should the insurance company charge for a yearly policy in order to break even on all 85,000 policies in this area?

Properties of Variances

Let X be random variable with $E(X) = \mu_X$ and variance of σ_X^2. Let $a, c \in \mathbb{R}$. Then:

1. $var(a) = 0$
2. $var(aX) = a^2\sigma_X^2$
3. $var(X + c) = \sigma_X^2$
4. $var(aX + c) = a^2\sigma_X^2$

Proof

1.
2.
3.
4.

Table of Contents

Bernoulli Distribution

Bernoulli Distribution has following property:

- You conduct a single trial consisted of **two** outcomes (often, success and failure).
- Trivially, probability of success, p + probability of failure, q = 1.
- The random variable that follows Bernoulli distribution can take values of either 0 or 1 (0 if failure is observed and 1 if success is observed)
- Example: A student passing STA256 with probability 65%, Rolling a fair die and observing an even number greater at least 4, etc...

Binomial Distribution

Binomial Distribution has following property:

- Mathematically, binomial distribution is sum of n number of independent Bernoulli Distributions. ie., $Y = \sum_{i=1}^{n} X_i$ where all $X_i \overset{\perp\!\!\!\perp}{\sim} Bernoulli(p)$. This tells us $Y \sim Bin(n, p)$
- The experiment is consisted of 'n' number of identical and independent Bernoulli trials. This means probability of failure and success do not change throughout the whole experiment.
- The random variable that follows binomial distribution represents the number of successes, and it can take values: $0, 1, 2, ..., n$

Binomial Distribution PMF

Random Variable X follows binomial distribution with n number of trials and $0 \leq p \leq 1$ probability of success $\iff$ the probability mass function is $\binom{n}{x} p^x (1-p)^{n-x}$.
In short, $X \sim bin(n, p) \iff P(X = x) = P_X(x) = \binom{n}{x} p^x (1-p)^{n-x}$ for $x = 0, 1, 2, ..., n$ and some $0 \leq p \leq 1$
(Note: Sometimes probability of failure, $1 - p = q$)

Binomial Coefficients

$$
\begin{array}{ccccccccccc}
 & & & & & 1 & & & & & \\
 & & & & 1 & & 1 & & & & \\
 & & & 1 & & 2 & & 1 & & & \\
 & & 1 & & 3 & & 3 & & 1 & & \\
 & 1 & & 4 & & 6 & & 4 & & 1 & \\
1 & & 5 & & 10 & & 10 & & 5 & & 1
\end{array}
$$

Notice that these numbers can be generated by using the choose function, $\binom{n}{k}$.
ie., $\binom{5}{2} = 10$

Binomial Theorem

Ever since you were a baby, you knew $(x+y)^2 = x^2 + 2xy + y^2$ and $(x-y)^3 = x^3 - 3x^2y + 3xy^2 - y^3$.
You can represent it as:
$(x-y)^k = \binom{k}{0}x^k y^0 + (-1)^1\binom{k}{1}x^{k-1}y + (-1)^2\binom{k}{2}x^{k-2}y^2 + \ldots + (-1)^{k-1}\binom{k}{k-1}xy^{k-1} + (-1)^k\binom{k}{k}x^0y^k$
More generally, binomial theorem states:
$(ax+by)^k = \binom{k}{0}(ax)^k(by)^0 + \binom{k}{1}(ax)^{k-1}by + \binom{k}{2}(ax)^{k-2}(by)^2 + \ldots + \binom{k}{k-1}ax(by)^{k-1} + \binom{k}{k}(ax)^0(by)^k = \sum_{n=0}^{k}\binom{k}{n}(ax)^{k-n}(by)^n$

Binomial Distribution: Expected Values and Variances

$$X \ bin(n,p) \implies E(X) = np, \ \sigma_X^2 = np(1-p) = npq$$

Proof: Will be uploaded on Blackboard later in a separate document.
Questions: Come see me during OH.

Example

A certain factory line produces cellphones. At any given time, analysts found that it is prone to make defects 10% of the time. Thomas randomly selects 10 phones. If there is at least one defective phone, Thomas will call the government to close the factory.

- Find the probability that 2 phones were defective.
- Find the probability that Thomas closes the factory.
- What is the expected number and variance in number of defective phones?

Table of Contents

Definition: Geometric Distribution

1. Geometric distribution is a special case of binomial distribution. This means probability of success and failure stays consistent and each trial is independent of others...
2. For 'x' number of trials, every trial except for the last trial (so x-1 total number of trials) is failure, and the last trial (so xth trial) is the first and last success.
3. A random variable X has a geometric distribution $\iff P(X = x) = P_X(x) = (1-p)^{x-1}p = q^{x-1}p$ for $x = 1, 2, ..., 0 \leq p \leq 1$

UNIVERSITY OF TORONTO MISSISSAUGA

Geometric Distribution: Expected Value and Variance

$X \sim Geometric(p) \implies \mu = E(X) = \frac{1}{p}$ and $var(X) = \sigma_X^2 = \frac{1-p}{p^2}$
Proof: Will be uploaded on Blackboard later in a separate document.
Questions: Come see me during OH.

Example

Steve's been trying to pass this machine learning course for very long time. It's been found that the probability he will pass the course is 0.1 each time he takes it. He also tends to forget course materials after the semester so you may assume that whether he passes this time is independent of his previous performances in the course.

1. What is the probability he will finally pass this course on 5th try?
2. What is the expected number of times he will take this course to pass for the first time?
3. What is the variance in number of times he has to take to finally pass this course?

Example

In responding to a survey question on a sensitive topic (such as"Have you ever tried marijuana?"), many people prefer not to respond in the affirmative. Suppose that 80% of the population have not tried marijuana and all of those individuals will truthfully answer no to your question. the remaining 20% of the population have tried marijuana and 70% of those individuals will lie. Derive the probability distribution of Y , the number of people you would need to question in order to obtain a single affirmative response.

Table of Contents

Definition: Negative Binomial Distribution

1. Negative binomial distribution is a special case of binomial distribution. So, each trial is independent. It 'has both' binomial and geometric distribution.
2. For 'x' number of trials, the last 'x'th trial must be a success, and for the remainder 1st to 'x-1'th trial, it is a binomial distribution with (x-1) total number of trials and k-1 number of successes.
3. In this course, we will put randomness in the total number of trials 'x'. As such, following is true:
$X \sim Neg.Bin(k,p) \iff P(X=x) = \binom{x-1}{k-1} p^{k-1}(1-p)^{x-k} p = \binom{x-1}{k-1} p^k (1-p)^{x-k},\ x = k, k+1, ..., 0 \leq p \leq 1$

Negative Binomial Distribution: Expected Value and Variance

Let $X \sim Neg.Bin(k, p)$. Then,

$$\mu = E(X) = \frac{k}{p},\ var(X) = \sigma_X^2 = \frac{k(1-p)}{p^2}$$

Proof: Beyond the scope of this course, involving Probability Generating Function

Example

Hamed in grade 12 got an offer from UofT, but would only accept the offer if he gets 10 confirmations from current students that UofT is a great school. To improve the response rate as he asks, he provides small incentive of 15 dollars to current female students and 10 dollars to male students. He does not prefer to ask more female students than male students. 10% of current students think UofT is a great school.

1. What is the probability that he will need to ask 20 students to get 10th confirmation?
2. What is the expected number of students Hamed will need to ask to get 10th confirmation?
3. Assuming 50% of current students are male, what is the expected expense on his end? variance of the expense?

Table of Contents

Definition: Hypergeometric Distribution

Disclaimer: Don't be fooled. Hypergeometric distribution is not a special case of geometric distribution.

- In hypergeometric distribution, each trial is not independent.
- It is probably easier to see this as a counting problem.
- $X \sim hypergeometric(k, n, N) \iff P(X = x) = \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}$ where $x = 0, 1, ..., n$ subject to restrictions $x \leq k$, and $n - x \leq N - k$
- notations: n number of objects are picked from a box of N objects without replacement. In the original box, there are k objects that meet the conditions, and you wish to find probability you picked x number of objects that meet the conditions.

Hypergeometric Distribution: Expected Value and Variance

X *hypergeometric*$(k, n, N) \implies$
$\mu = E(X) = \frac{nk}{N}$, and $var(X) = \sigma_X^2 = n\frac{k}{N}\frac{N-k}{N}\frac{N-n}{N-1}$

Example

Daniel has total of 10 pets,4 of which are dogs. He has to give away 3 of his pets to his brother.

1. What is the probability that he gives away 2 dogs?
2. What is the probability that he gives away at most 2 dogs?
3. What is the probability that he doesn't give away any dogs?
4. What is the expected number of dogs that he will give away to his brother?
5. What is the variance in number of dogs that he gives away?

Relationship between Binomial and Hypergeometric Distributions

If $1 \leq n << N$, picking an item in each draw would result in very small change in probability of choosing items that meet the conditions. This would also mean that $N - n \approx N$, and $N - 1 \approx N$

Rule of thumb: Hypergeometric distribution can be approximated by using binomial distribution when $\frac{n}{N} \leq 0.05$.

Then, the quantity $\frac{k}{N}$ may be approximated using p.

Notice: $\mu = np = \frac{nk}{N}$, and $\sigma_X^2 = npq = n\frac{k}{N}(1 - \frac{k}{N})$

$\because \frac{N-k}{N} = 1 - \frac{k}{N}$, and $\frac{N-n}{N-1} \approx 1$

So, binomial distribution can be viewed as a large population version of hypergeometric distribution.

Example

$X \sim Bin(n, p)$. Let $Z = 5X$. Find the probability that Z is two standard deviations away from the mean.

Table of Contents

Poisson Distribution: Examples

- Experiments that deal with number of outcomes occurring during a fixed time interval is called Poisson process. The interval can be of any length (in minutes, hours, days, or years)
- The interval does not have to be of time. It can also be of an area, volume, or number of items.
- Examples of Poisson Process: Number of phone calls received per hour, number of customers purchasing any item per hour, number of watermelon yield per 1 acre,...

Poisson Distribution: Properties

1. The number of outcomes occurring in one time interval or specified region is **independent** of the number that occurs in any other disjoint time interval or region of space.
2. The probability that a single outcome will occur during a different length of time interval or region is proportional to the original length of the time interval or the size of the region.
3. The probability that more than one outcome will occur in a very short time interval or in such a small region is negligible.

Poisson Distribution: Definition

$$X \sim Pois(\lambda) \iff P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}, x = 0, 1, 2, ..., \lambda > 0$$

Note: λ parameter represents rate per fixed time interval.

Poisson Distribution: Expected value and Variance

$$X \sim Pois(\lambda) \implies \mu = E(X) = \lambda, \sigma_X^2 = var(X) = \lambda$$

Proof: Will be uploaded on Blackboard later in a separate document.
Questions: Come see me during OH.

Poisson Distribution is a limiting form of binomial distribution

In the case of the Binomial, if n is quite large ($\rightarrow \infty$), p is small ($\rightarrow 0$), and *np* is fixed as constant, the conditions begin to simulate the continuous space or time region implications of the Poisson process. The independence among Bernoulli trials in the Binomial case is consistent with property 2 of the Poisson process. Allowing the parameter p to be close to zero relates to property 3.

Example

Ten customers come to Burger King every hour. The number of customers who came in this hour is independent of the number of customers who came in previous hours or how many will come in the next hour. If more than 10 people come to BK, the manager needs to hire another cashier.

1. What is the probability that the employer does not have to hire another cashier?
2. What is the probability that 2 people will come in the next half hour?
3. Suppose the BK restaurant is open 24/7 and the rate at which customers come stays constant. What is the expected number and variance of number of customers per day?

Table of Contents

Definition

The **kth moment** of a random variable X about its mean, μ is
$\mu_k = E((X - \mu)^k)$

Definition

- **Moment Generating Function** of random variable X is defined to be $M_X(t) = E(e^{Xt})$.
- Moment Generating Function (short: MGF) is said to exist if whenever t is bounded (i.e., $\exists a > 0, |t| < a$), $M_x(t) < \infty$

Finding Moments using MGF's

- If $M_X(t)$ exists, then $\forall k \in \mathbb{N}$,

$$\frac{d^k M_X(t)}{dt^k}|_{t=0} = M_X^{k'}(0) = \mu_k'$$

  In English, the kth derivative of MGF with respect to t then setting $t = 0$, you will obtain μ_k' =kth moment about 0.
- In practice, kth moment about 0 is also called kth moment.

Using Taylor series...

$e^{tX} = 1 + tX + \frac{t^2X^2}{2!} + ...$
$M_X(t) = E(e^{tX}) = 1 + tE(X) + \frac{t^2E(X^2)}{2!} + ...$
This means $\frac{dM_X(t)}{dt}|_{t=0} = E(X)$, $\frac{d^2M_X(t)}{dt^2}|_{t=0} = E(X^2)$,
$\frac{d^3M_X(t)}{dt^3}|_{t=0} = E(X^3)$

Binomial Distribution MGF

Let $X \sim Bin(n, p)$. $M_X(t) = (pe^t + q)^n$

proof

Geometric Distribution MGF

Let $X \sim Geo(p)$. $M_X(t) = \frac{pe^t}{1-qe^t}$.

proof

useful: Sum of geometric series: $S = \frac{1-r^n}{1-r}$
and if $|r| < 1$ then $\lim_{n\to\infty} S = \sum_{n=0}^{\infty} r^n = \frac{1}{1-r}$

Example

Let $M_X(t) = \frac{1}{2}e^t + \frac{2}{5}e^{2t} + ce^{3t}$.

1. Find value of c such that X is a proper discrete random variable, and thus find the distribution of X.
2. Find E(X).
3. Find $E(X^2)$ and thus, obtain σ_X^2

Properties of MGF

- If we can find MGF, then theoretically, we can find any of the moments, $E(X), E(X^2), E(X^3), \ldots$
- MGF's are unique to its own probability distribution. This means that if MGF of random variable X has the same form as MGF of well known probability distributions, we can find the probability distribution of X as well as the values of the parameters (simply by comparison). This also means that if two random variables X and Y have the same MGF, then they must have the same probability distribution.

Example

Let $M_X(t) = (0.3e^t + 0.7)^5$. Find the probability distribution and the values of parameters, n and p.