Section 3: Random variables – discrete

STA 35C – Statistical Data Science III

Instructor: Akira Horiguchi

Fall Quarter 2025 (Sep 24 – Dec 12)
MWF, 12:10 PM – 1:00 PM, Olson 158
University of California, Davis

Overview

Based on Chapter 3 of textbook: `https://www.probabilitycourse.com/`

- Contains problems with solutions, and problems without solutions.

1. Basic concepts
2. Independent random variables
3. Special distributions
4. Cumulative distribution function
5. Expected value
6. Functions of random variables
7. Variance

Basic concepts

Random variables

We usually focus on some *numerical aspects* of a random experiment.

- For example, in a soccer game we may be interested in the number of goals, shots, shots on goal, corners kicks, fouls, etc.
- On any given day at UCD, we may be interested in the number of Cheeto sightings.
- These are examples of *random variables*.

Definition 1: Random variable

A *random variable* $X\colon \Omega \to \mathbb{R}$ is a function from the sample space Ω to the real numbers.

- E.g., toss a coin three times. Sample space is

$$\Omega = \{TTT, TTH, THT, THH, HTT, HTH, HHT, HHH\}.$$

  We can define a random variable X whose value is the number of observed heads.
- Usually denote random variables by capital letters such as X, Y, and Z.
- The *range* of a random variable X is the set of possible values for X. For example:
  - I toss a coin 100 times. Let X be the number of heads I observe.
  - I toss a coin until the first heads appears. Let Y be the total number of coin tosses.
  - The random variable T is defined as the time (in hours) from now until the next earthquake occurs in a certain city.

Discrete random variables

A random variable X is *discrete* if its range is countable.

- Recurring examples:
  1. number of heads after two coin flips,
  2. number of coin flips needed before a heads turns up.
- Here probabilities can be assigned to each realizable value.
  1. For $\{0, 1, 2\}$ (finite), we can assign probabilities $1/4$, $1/2$, and $1/4$.
  2. For $\mathbb{N}$ (countably infinite), we can assign probabilities $(1/2)^k$ to each $k \in \mathbb{N}$.
- For a discrete r.v. X with range $\{x_1, x_2, x_3, \cdots\}$, the function $f_X(\cdot)$ defined as

$$f_X(x_k) = P(X = x_k), \quad \text{for } k = 1, 2, 3, \cdots,$$

  is called the *probability mass function (PMF)* of X.
  1. $f_X(0) = 1/4$, $f_X(1) = 1/2$, and $f_X(2) = 1/4$.
  2. $f_X(k) = (1/2)^k$ for each $k \in \mathbb{N}$.

  Here $f_X(a)$ is *"the probability that X equals a."*

PMF of the number of heads after two flips of a fair coin.

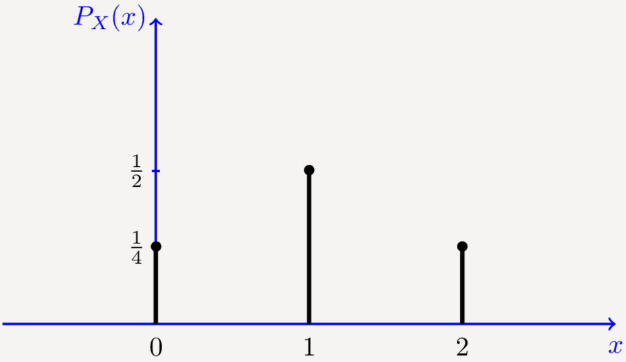


Fig.3.1 - PMF for random Variable X in Example 3.3.

The PMF of a discrete random variable is also called the r.v.'s *probability distribution*.

Discrete random variables: PMF

A PMF is a probabilty measure, so it satisfies Definition 1 from Section 2.

- In particular, it satisfies countable additivity.
- This lets us deduce the probability $P(X \in A)$ that a discrete r.v. X lies in an event A:

$$P(X \in A) = P\left(\bigcup_{a \in A} [X = a]\right) = \sum_{a \in A} f_X(a), \tag{1}$$

Independent random variables

Independent random variables

When dealing with more than one random variable, often need to consider the *dependence/correlation* between them.

- Concept of *independent random variables* is similar to that of independent events.
- Two random variables are independent if knowing the value of one does not change the probabilities for the other.

Two independent random variables

Definition 2: Two independent random variables

Two discrete random variables X and Y are *independent* if

$$P(X = a, Y = b) = P(X = a)\, P(Y = b) \tag{2}$$

for all a, b.

If two random variables are independent, then we can write

$$P(X \in A, Y \in B) = P(X \in A)\, P(Y \in B)$$

for all sets A, B. We can also write

$$P(Y = b | X = a) = P(Y = b)$$

for all a, b.

Toss a fair coin four times.

- Let X be the number of heads observed in the first and second coin flips.
- Let Y be the number of heads observed in the third and fourth coin flips.

Find $\underbrace{P\Big((X < 2) \text{ and } (Y > 1)\Big)}$.

$$P(X<2, Y>1) = P(X<2)\,P(Y>1)$$

$$= P(X \in \{0,1\})\,P(Y \in \{2\})$$

$$= \Big[1 - \underbrace{P(X \in \{2\})}_{\frac{1}{2}\cdot\frac{1}{2}}\Big]\,\underbrace{P(Y \in \{2\})}_{\frac{1}{2}\cdot\frac{1}{2}}$$

$$= \frac{3}{4}\cdot\frac{1}{4} = \boxed{\frac{3}{16}}$$

≥ 2 independent random variables

Definition 3: ≥ 2 independent random variables

Discrete random variables $X_1, X_2, X_3, \cdots, X_n$ are *independent* if

$$\underbrace{P(X_1 = x_1, X_2 = x_2, \cdots, X_n = x_n)} = \underbrace{P(X_1 = x_1)}\,\underbrace{P(X_2 = x_2)} \cdots \underbrace{P(X_n = x_n)} \tag{3}$$

for all $x_1, x_2, \cdots, x_n$.

Special distributions

Uniform distribution

A random variable X with values in a finite set M is *uniformly* distributed if each element in M has the same probability:

$$P(X = k) = \frac{1}{|M|} \quad \text{for all } k \in M$$

- Such distributions occur when all possible outcomes are equally likely.
- We write $X \sim U(M)$ or $X \sim Unif(M)$.
- Nine random draws in R: `sample(c(1,2,3,4,5,6), size=9, replace=T)`

TRUE

nine rolls of a fair six-sided die

Bernoulli distribution

A random variable X is *Bernoulli* distributed with parameter $p \in (0, 1)$, if $P(X = 1) = p$ and $P(X = 0) = 1 - p$.

- For when our random experiment has only two possible outcomes ("success" and "failure").
- Example: flip a coin with probability p of heads ("success"). Is it heads?
- We write $X \sim Ber_p$ or $X \sim Bern(p)$.
- Nine random draws in R: `rbinom(n=9, size=1, prob=1/3)`

p

Binomial distribution

A random variable X is *Binomial* distributed with parameters $n \in \mathbb{N}$ and $p \in (0, 1)$ if

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k} \quad \text{for all } k = 0, \ldots, n.$$

- We think of n as the number of experiments and p the success probability. In the above equation, k is the number of successes.
- For measuring the probability of the number of successes of n independent Bernoulli experiments with parameter p.
- Example: flip a coin n times, each flip with probability p of heads ("success"). How many heads?
- We write $X \sim Bin_{n,p}$ or $X \sim Bin(n, p)$.
- A random draw in R: `rbinom(n=3, size=1, prob=0.25) |> sum()`

  n P

Cumulative distribution function

Motivation

The PMF is one way to describe the distribution of a discrete random variable.

- Pro: intuitive.
- Con: it cannot be defined for continuous random variables.

The *cumulative distribution function (CDF)* can characterize the distribution of *any kind* of random variable (discrete, continuous, mixed).

Cumulative distribution function

The *CDF* of a random variable X is the function $F_X\colon \mathbb{R} \to [0, 1]$ defined by

$$F_X(a) := P(X \le a), \quad a \in \mathbb{R}. \tag{4}$$

This is "the probability that X is less than or equal to a."

- Definition holds regardless of whether X is discrete, continuous, or mixed.
- In the discrete case – recall Eq. (1) – holds for any $a \in \mathbb{R}$,

$A = (-\infty, a)$

$F_X(a) = P(X \in A) = \sum_{s \le a} f_X(s)$

$$F_X(a) = \sum_{s \le a} f_X(s).$$

$F_X(a) = P(X \le a)$

$= P\left(\bigcup_{s \le a} [X = s]\right)$

countable additivity,

$= \sum_{s \le a} f_X(s)$

- For any $a, b \in \mathbb{R}$ with $b > a$ holds,

$$P(a < X \le b) = F_X(b) - F_X(a).$$

Cumulative distribution function

From the definition of F_X in Eq. (4) come the following properties:

1. F_X is *right-continuous* and monotonically increasing,
   "non-decreasing"
2. $\lim_{a\to-\infty} F_X(a) = 0$,
3. $\lim_{a\to+\infty} F_X(a) = 1$.

4

$$\overbrace{P(X \le x_1)}^{\text{CDF at } x_1} + \overbrace{P(X = x_2)}^{\text{PMF at } x_2} = \overbrace{P(X \le x_2)}^{\text{CDF at } x_2}$$

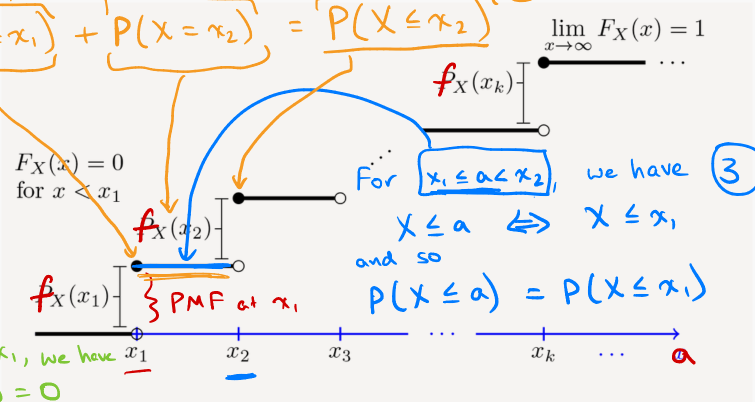


Fig.3.4 - CDF of a discrete random variable.

3

For $x_1 \le a < x_2$, we have

$X \le a \iff X \le x_1$

and so

$P(X \le a) = P(X \le x_1)$

1

For $a < x_1$, we have

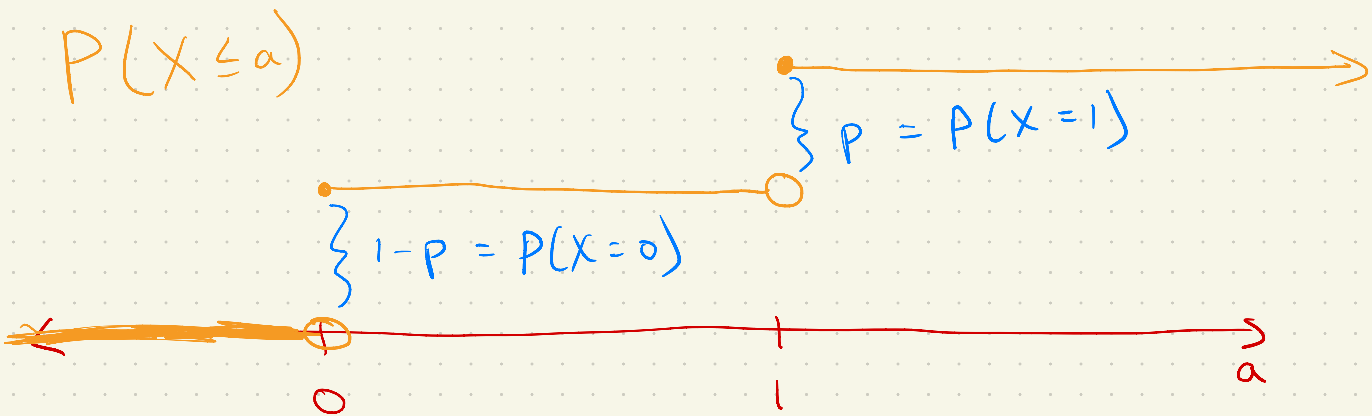
$P(X \le a) = 0$

2

$$\underbrace{P(X \le x_1)}_{\text{CDF at } x_1} = \underbrace{P(X = x_1)}_{\text{PMF at } x_1} + \underbrace{P(X < x_1)}_{0}$$

CDF of Bernuoilli random variable with parameter p

- For $a < 0$, we have $P(X \le a) = 0$.
- For $a = 0$, we have $P(X \le a) = \underbrace{P(X=0)}_{1-p} + \underbrace{P(X<0)}_{0} = 1-p$
- For $0 \le a < 1$, we have $P(X \le a) = P(X \le 0) + \underbrace{P(0 < X \le a)}_{0}$
- For $a = 1$, we have $P(X \le a) = \underbrace{P(X=1)}_{p} + \underbrace{P(X<1)}_{P(X \le 0) = 1-p} = 1.$
- For $a > 1$, we have $P(X \le a) = P(X \le 1) = 1.$



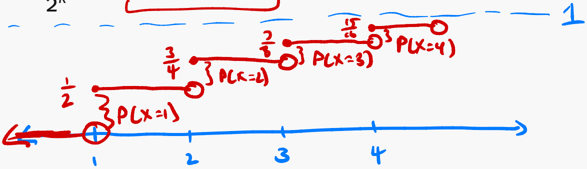
Example

Suppose the PMF of a discrete random variable X is given by

$$P(X=k) = f_X(k) = \frac{1}{2^k} \quad \text{for } k = 1, 2, 3, \ldots$$

$$\boxed{F_X(k) = 1 - \frac{1}{2^k}}$$
CDF

1. Find and plot the CDF.
2. Find $P(2 < X \le 5)$.
3. Find $P(X > 4)$.



$a < 1$: $P(X \le a) = 0$

$a = 1$: $P(X \le 1) = P(X=1) + P(X<1) = \frac{1}{2} + 0 = \frac{1}{2}$

$1 < a < 2$: $P(X \le a) = P(X \le 1) = \frac{1}{2}$

$a = 2$: $P(X \le 2) = P(X=2) + \underbrace{P(X<2)}_{P(X \le 1)} = \frac{1}{4} + \frac{1}{2} = \frac{3}{4}$

$2 < a < 3$: $P(X \le a) = P(X \le 2) + \underbrace{P(2 < X < a)}_{0} = \frac{3}{4} + 0 = \frac{3}{4}$

$a = 3$: $P(X \le 3) = \underbrace{P(X=3)}_{\frac{1}{2^3}} + \underbrace{P(X<3)}_{P(X \le 2)} = \frac{1}{8} + \frac{3}{4} = \frac{7}{8}$

Example

Suppose the PMF of a discrete random variable X is given by

$$f_X(k) = \frac{1}{2^k} \quad \text{for } k = 1, 2, 3, \ldots$$

1. Find and plot the CDF.
2. Find $P(2 < X \leq 5)$.
3. Find $P(X > 4)$.

Closed form expression for CDF of X for $k \in \mathbb{N}$:

$$F_X(k) = P(X \leq k) = 1 - P(X > k)$$

$$= 1 - \sum_{a=k+1}^{\infty} P(X = a)$$

$$= 1 - \sum_{a=k+1}^{\infty} \frac{1}{2^a}$$

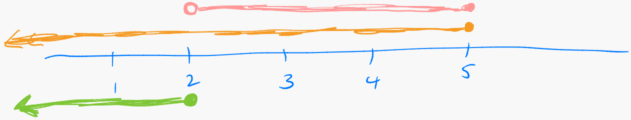
$$= \boxed{1 - \frac{1}{2^k}}$$

Example

Suppose the PMF of a discrete random variable X is given by

$$f_X(k) = \frac{1}{2^k} \quad \text{for } k = 1, 2, 3, \ldots$$

1. Find and plot the CDF.
2. Find $P(2 < X \le 5)$.
3. Find $P(X > 4)$.



$$P(2 < X \le 5) = P(X \le 5) - P(X \le 2)$$
$$= F_X(5) - F_X(2)$$
$$= \left[1 - \frac{1}{2^5}\right] - \left[1 - \frac{1}{2^2}\right] = \frac{1}{2^2} - \frac{1}{2^5}$$

$$P(X > 4) = 1 - P(X \le 4)$$
$$= 1 - F_X(4) = 1 - \left\{1 - \frac{1}{2^4}\right\} = \frac{1}{2^4}$$

Expected value

Given some numbers, we often want a descriptive summary of these values.

- Their *average* is a single number that represents/describes the whole collection.
- How might we describe a “representative value” for a random variable?
- With a random variable, some values occur more often than other values.
- We might want to weight the values more if they occur more often.

Example: suppose we have a fair die. How might we summarize the outcomes of this die using a single number? What about for an unfair die?

The *expected value* of a random variable is the weighted average of all of its values, where the weights are the probabilities that these values occur.

Definition 2: Expected value $E(\cdot)$

Let X be a discrete random variable. Then the *expected value* of X is defined as

"E[x]" or "EX"

$$E(X) = \sum_{\text{all } k} P(X = k) \cdot k \tag{5}$$

Example: Let $X \sim Bernoulli(p)$. Find $E(X)$.

$$E[X] = \sum_{k=0}^{1} P(X=k) \cdot k$$

$$= \underbrace{P(X=0) \cdot 0}_{0} + \underbrace{P(X=1) \cdot 1}$$

$$= P(X=1)$$

$$= p$$ (the Bernoulli parameter)

If X is a random variable, then any function of X is also a random variable.

- For example, if $Y = aX + b$, we can talk about $EY = E[aX + b]$.

Theorem 3.2: Expectation is linear

We have $E[aX+b] = E[aX] + b = aE[X] + b$

- $E[aX + b] = aEX + b$, for all $a, b \in \mathbb{R}$;
- $E[X_1 + X_2 + \cdots + X_n] = EX_1 + EX_2 + \cdots + EX_n$, for any set of random variables $X_1, X_2, \ldots, X_n$.

Example: Let $X \sim Binomial(n, p)$. Find $E(X)$.

If $Y \sim Bern(p)$, then $EY = p$

We can think of X as being $Y_1 + Y_2 + \cdots + Y_n$
where Y_i's are independent Bernouilli trials w/param p

$$E(X) = E(Y_1 + Y_2 + \cdots + Y_n)$$
$$= \underbrace{EY_1}_{p} + \underbrace{EY_2}_{p} + \cdots + \underbrace{EY_n}_{p}$$
$$= \boxed{np}$$

Functions of random variables

$Y = X^2 \qquad R_X = \mathbb{Z} \qquad R_Y = \{a^2 \mid a \in \mathbb{Z}\} = \{0\} \cup \mathbb{N}$

If X is a r.v., then ~~and~~ any *function $Y = g(X)$ of X* is itself a random variable.

- Range of Y is
$$R_Y := \{g(a) | a \in R_X\}$$
where R_X is the range of X.

$\to \{a \mid g(a) = b\}$

- PMF of Y is
$$f_Y(b) \overset{\text{definition of PMF}}{=} P(Y = b) \overset{Y=g(X)}{=} P\Big(g(X) = b\Big) = \sum_{a:\, g(a)=b} \overbrace{f_X(a)}^{P(X=a)}.$$

- Expected value of Y is
$$EY = \sum_{b \in R_Y} b\, f_Y(b).$$

In practice, usually easier to use the *law of the unconscious statistician (LOTUS)*:
$$EY = E\big[g(X)\big] = \sum_{a \in R_X} g(a)\, f_X(a).$$

Example

Find $E[\sin(X)]$, where X is a discrete random variable with range

$$= \left[\frac{1}{5}+\frac{1}{5}\right]\cdot 0 + \left[\frac{1}{5}+\frac{1}{5}\right]\cdot\frac{\sqrt{2}}{2} + \frac{1}{5}\cdot 1$$

$$R_X = \left\{0, \frac{\pi}{4}, \frac{\pi}{2}, \frac{3\pi}{4}, \pi\right\}$$

a	$\sin(a)$
0	0
$\pi/4$	$\sqrt{2}/2$
$\pi/2$	1
$3\pi/4$	$\sqrt{2}/2$
π	0

and PDF values

$$\boxed{f_X(0) = f_X\left(\frac{\pi}{4}\right) = f_X\left(\frac{\pi}{2}\right) = f_X\left(\frac{3\pi}{4}\right) = f_X(\pi) = \frac{1}{5}.}$$

$$E[\sin(X)] = \sum_{a\in R_X} \sin(a)\cdot f_X(a) = \frac{1}{5}\sum_{a\in R_X}\sin(a)$$

$$= \Big[\underbrace{f_X(0)}_{\frac{1}{5}} + \underbrace{f_X(\pi)}_{\frac{1}{5}}\Big]\cdot 0 + \Big[\underbrace{f_X\left(\tfrac{\pi}{4}\right)}_{\frac{1}{5}} + \underbrace{f_X\left(\tfrac{3\pi}{4}\right)}_{\frac{1}{5}}\Big]\cdot\frac{\sqrt{2}}{2} + \underbrace{f_X\left(\tfrac{\pi}{2}\right)}_{\frac{1}{5}}\cdot 1$$

$$= \frac{1}{5}\left(0 + \sqrt{2} + 1\right).$$

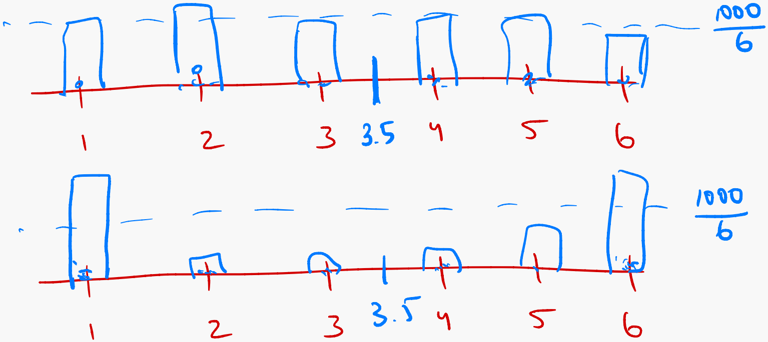
Variance

Intuition

Often summarize a probability distribution by its center and spread.

- Center: expected value
- Spread: *variance*

Think of variance as "how much a random variable varies about its mean."



Definition 3: Variance $Var(\cdot)$

Let X be a random variable with $E(X^2) < \infty$. Then the *variance* of X is defined as

$$Var(X) := E\left[\{X - E(X)\}^2\right]. \qquad E\left[\{X - EX\}^2\right] \tag{6}$$

- A *large value* of $Var(X)$ means that $\{X - E(X)\}^2$ is often large, so X often takes values far from its mean.
  - Implies that the distribution is very spread out.

$\{X - EX\}^2$

- A *small value* of $Var(X)$ means that $\{X - E(X)\}^2$ is often small, so X often takes values close to its mean.
  - Implies that the distribution is concentrated around its average.

Standard deviation

$Var(X)$ has a different unit than X. E.g., if X is a stock price.

- Can instead measure spread using the square root of variance:

Definition 4: Standard deviation $Var(\cdot)$

Let X be a r.v. with $E\left(X^2\right) < \infty$. Then the *standard deviation* of X is defined as

$$SD(X) := \sqrt{Var(X)} \tag{7}$$

- Despite having the same unit of X, the variance is easier to mathematically find the minimum of (i.e., take the derivative of).
- Usually we will describe a distribution's spread using the variance.

Properties and calculation tools

From Definition 3, we can deduce the following properties:

- $Var(X) \geq 0$.
- If $Var(X) = 0$, then X is constant.
- The variance of X can also be calculated as

$$Var(X) = E(X^2) - (E[X])^2. \tag{8}$$

Properties of $Var(\cdot)$

Let $c \in \mathbb{R}$ be a constant, and let X be a random variable with $E(X^2) < \infty$. Then

i) $Var(c) = 0$;

ii) $Var(X + c) = Var(X)$;

iii) $Var(cX) = c^2 Var(X)$;

Example: consider $c = 5$, $Var(X) = 1$.

$Var(X+5) = Var(X) = 1$

$Var(5X) = 25\,Var(X) = 25$

Theorem: variance of sum of independent random variables

If $X_1, X_2, \ldots, X_n$ are *independent* random variables, then

$$Var\Big(X_1 + X_2 + \cdots + X_n\Big) = Var(X_1) + Var(X_2) + \cdots + Var(X_n). \quad (9)$$

Example: if $X \sim Binomial(n, p)$, find $Var(X)$.

$$Var(X) = Var(Y_1 + \cdots + Y_n)$$
$$= Var(Y_1) + \cdots + Var(Y_n)$$
$$= n\,Var(Y_1)$$

$$\boxed{\text{If } X_1 = X_2, \text{ then } Var(X_1 + X_2) = Var(2X_1) = 4Var(X_1) \neq Var(X_1) + Var(X_2)}$$

LOTUS

$$E(Y^2) \overset{\downarrow}{=} 0^2 \cdot P(Y=0) + 1^2 \cdot P(Y=1)$$
$$= P(Y=1)$$
$$= p$$

$$Var(Y) = \underbrace{E(Y^2)}_{p} - \underbrace{(EY)^2}_{p^2}$$

$$\boxed{= p - p^2 = p(1-p)}$$