Random variables

Random variables

- A random variable is a numeric quantity whose value depends on the outcome of a random event.

- Mathematically, a random variable is a mapping that maps an event in the sample space to a numeric value.

Random variables

- We use a capital letter, like X, to denote a random variable (RV).
- The value of a RV is denoted with a lowercase letter, e.g., $P(X = x)$
- $P(X = 2)$ is the probability that X has the value 2.
- $P(X < 10)$ is the probability that X is less than 10.

Random variables

- Discrete random variables often take only integer values
    - Example: number of students present, gender of an unborn baby
- Continuous random variables take real (decimal) values
    - Example: tomorrow's PM 2.5 level, your final grade

Probability Mass Functions

- The probability mass function (PMF) of a discrete RV describes the probability of each value.

- For example, If X is the result of one roll of a fair die:

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

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

$$\cdots$$

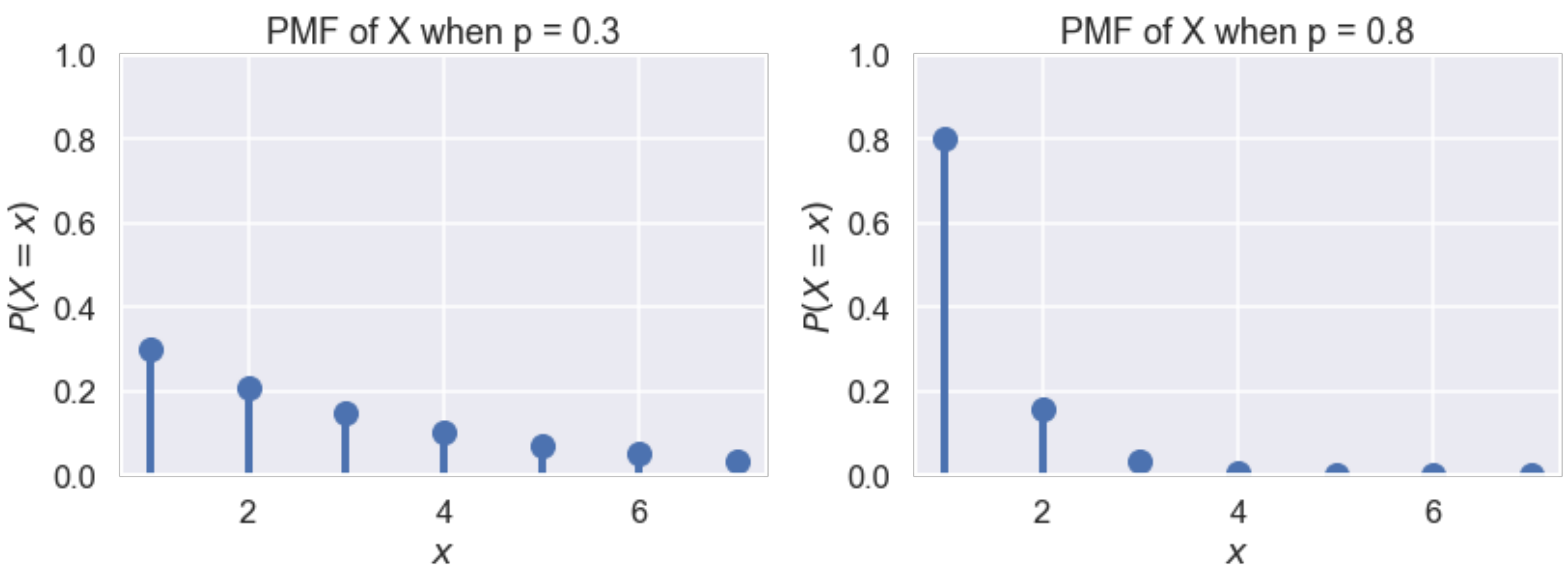
$$P(X=6)=\frac{1}{6}$$

X	P(X)
1	1/6
2	1/6
⋮	⋮
6	1/6

Probability Mass Functions

- A PMF is often a parameterized function:

$$P(X = k) = (1 - p)^k p$$



Bernoulli distribution

$X \sim \text{Bern}(p)$

$P(X = 1) = p, \quad P(X = 0) = 1 - p$

- Distribution used to describe a trial with binary outcomes
  - e.g. let X be 1 if a coin lands head
  - fundamental tool for logistic regression

PMF for two RVs

- Joint PMF: the list of $P(X = x, Y = y)$
- Marginal PMF: $P(X = x) = \sum_{y \in \mathbb{Y}} P(X = x, Y = y)$
- Conditional PMF: $P(Y = y \mid X = x) = \dfrac{P(X = x, Y = y)}{P(X = x)}$

Independence

- Two RVs X and Y are independent iff

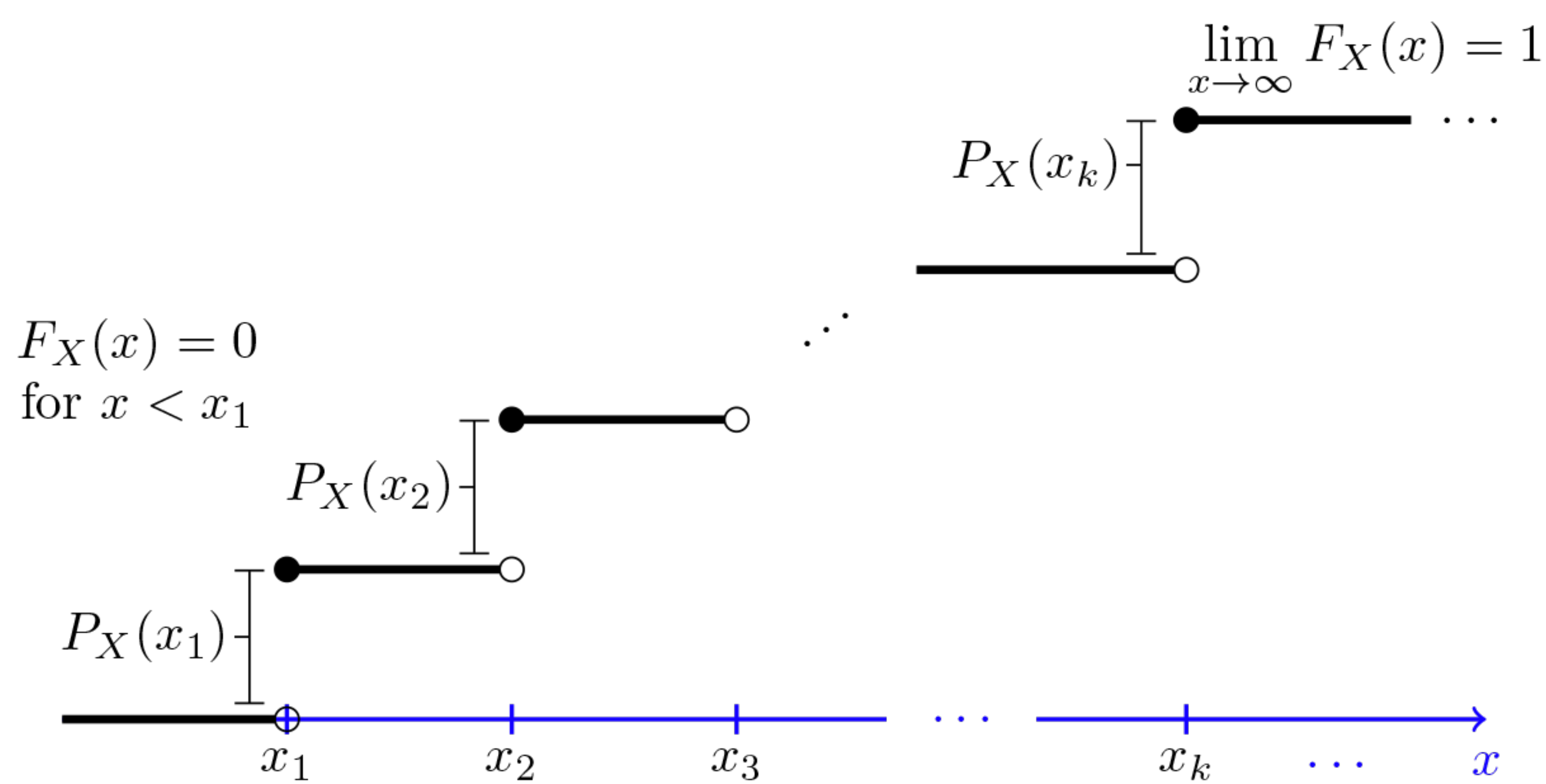
$$P(X = x, Y = y) = P(X = x) \cdot P(Y = y)$$

for all $x \in \mathbb{X}$ and $y \in \mathbb{Y}$

Cumulative Distribution Function

- The cumulative distribution function (CDF) for a discrete RV is

$$F(x) = P(X \leq x) = \sum_{t \leq x} f(t)$$



Probability density function

A probability density function (PDF) of a continuous RV is a function satisfying the following properties:

1. $f_X(x) \geq 0, -\infty < x < \infty$
2. $\int_{-\infty}^{\infty} f_X(x)dx = 1$
3. $P(a < X \leq b) = \int_a^b f_X(x)dx$

Probabilities from continuous distributions

- For a continuous random variable X, the probability $P(X = x)$ is defined as 0 (recall the definition of a distribution)
- $P(a < X \leq b) = F_X(b) - F_X(a) = \int_a^b f(x)dx$

Normal

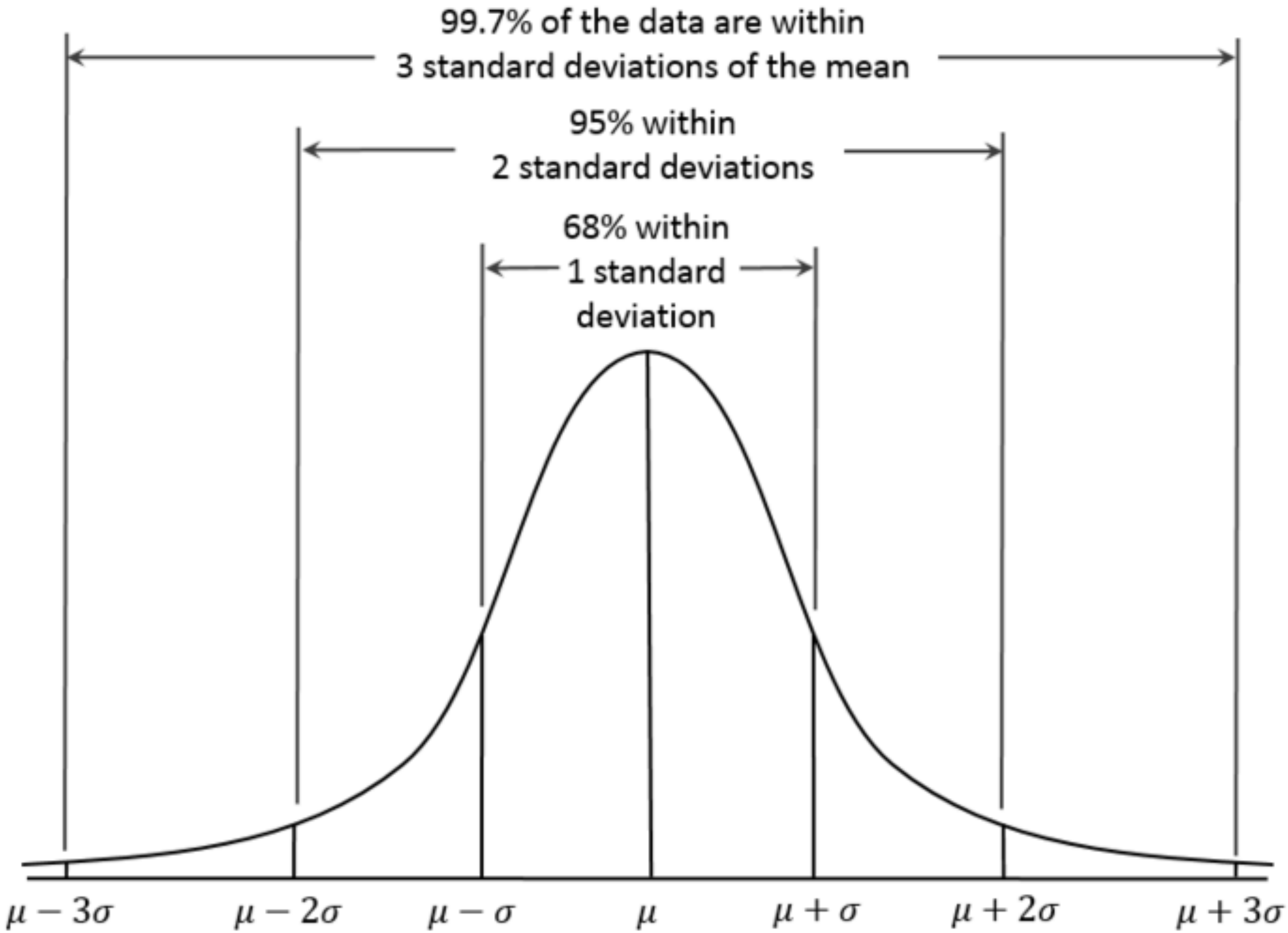
- If X is a continuous RV with pdf

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x-\mu)^2},$$

then $X \sim N\left(\mu, \sigma^2\right)$

- Useful for regression

Normal



99.7% of the data are within
3 standard deviations of the mean
95% within
2 standard deviations
68% within
1 standard
deviation
$\mu - 3\sigma$
$\mu - 2\sigma$
$\mu - \sigma$
μ
$\mu + \sigma$
$\mu + 2\sigma$
$\mu + 3\sigma$

CDF for a continuous RV

- The CDF for a continuous random variable is

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt$$

Transformation of a RV

- A Transformation of a RV is another RV, e.g. X^2 is another RV if X is a RV

Expectation

- The expectation $E(X)$ is the **long-run average** value of X; that is, the average value when the trial is repeated for infinitely many times.

- For a discrete RV,

$$E(X) = \sum_{x \in \mathbb{X}} x \cdot P(X = x)$$

- For a continuous RV,

$$E(X) = \int x \cdot f_X(x) dx$$

Example: Bernoulli RV

$X \sim \text{Bern}(p)$

$$\begin{aligned} E(X) &= \sum_{x \in \mathbb{X}} x \cdot P(X = x) \\ &= (1)(p) + (0)(1-p) \\ &= p \end{aligned}$$

- Note that $E(X)$ does not have to be a value that X can take.

  - In this case X is either 0 or 1, but $E(X) = p$

Linearity of expectation

- For any RVs X and Y and constants a and b:

$$E(aX + bY) = aE(X) + bE(Y)$$

- This holds even when X and Y are not independent!

Variance

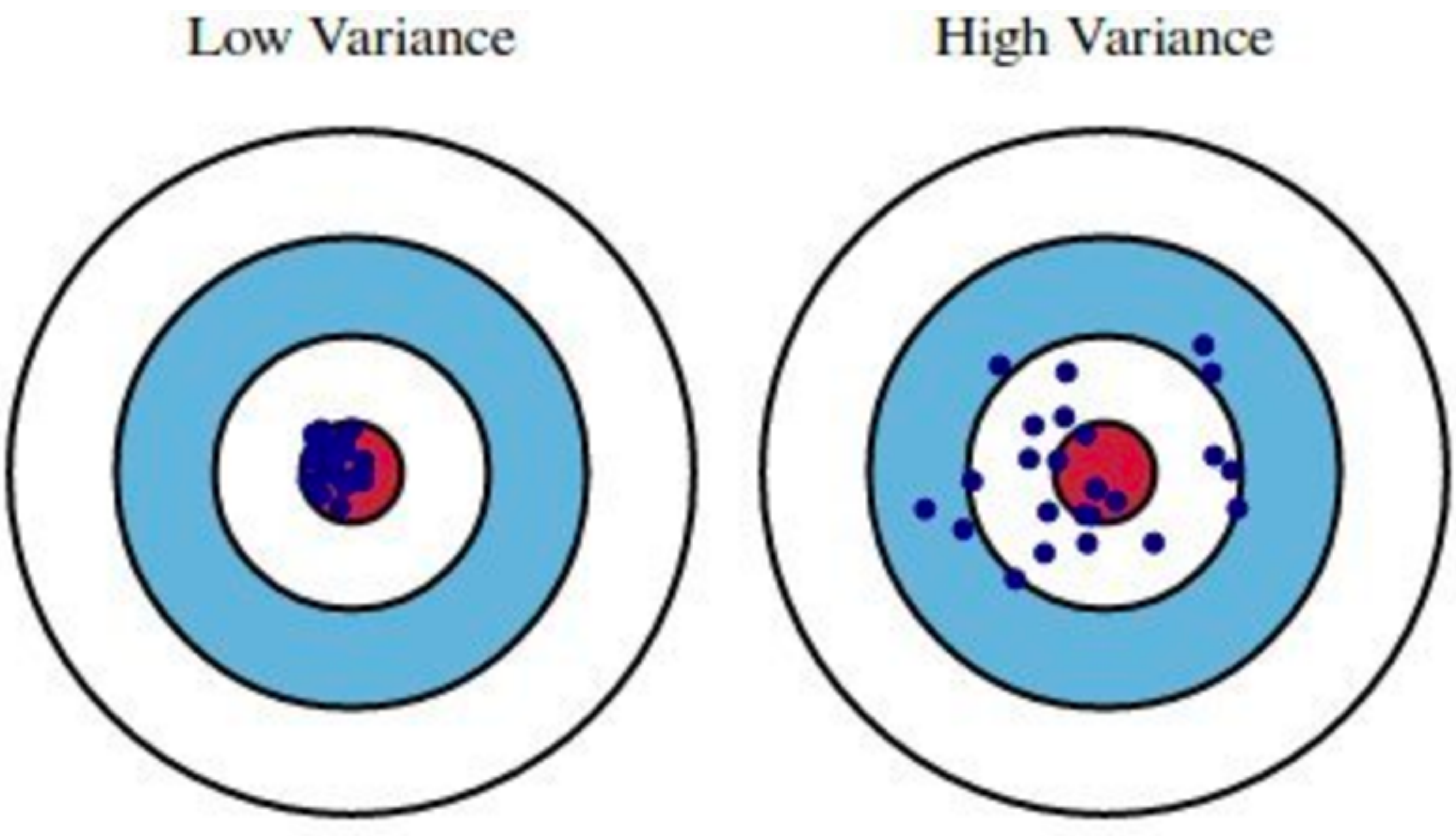
- The variance of a RV is the average squared deviation of the variable around its mean (its expected value):

$$\text{Var}(X) = E\left([X - E(X)]^2\right)$$
$$= E\left((X - \mu)^2\right)$$

- With some algebra, you can show that

$$\text{Var}(X) = E\left(X^2\right) - E(X)^2$$

Mean vs variance



Example: Bernoulli RV

$$E\left(X^2\right) = \left(1^2\right)p + \left(0^2\right)(1-p) = p$$

$$\begin{aligned}\text{Var}(X) &= E\left(X^2\right) - E(X)^2 \\ &= p - p^2 = p(1-p)\end{aligned}$$

Variance properties

- If X is a RV and a is a constant, this is always true:

$$\text{Var}(aX) = a^2\text{Var}(X)$$

- This is true only if X and Y are independent:

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$$

-

Covariance

- Covariance measures the joint variablility of two RVs X and Y,

$$\text{Cov}(X, Y) = E\left([X - E(X)][Y - E(Y)]\right)$$

- With some algebra, you can show that

$$\text{Cov}(X, Y) = E(XY) - E(X)E(Y)$$

Covariance properties

If X, Y, W, and V be random variables and a, b, c, d be constants,

- $\text{Cov}(X, a) = 0$
- $\text{Cov}(X, X) = \text{Var}(X)$
- $\text{Cov}(X, Y) = \text{Cov}(Y, X)$
- $\text{Cov}(X + a, Y + b) = \text{Cov}(X, Y)$

Covariance properties

- $\text{Cov}(aX, bY) = ab\text{Cov}(X, Y)$

- $\text{Cov}(aX + bY, cW + dV) = ac\text{Cov}(X, W) + ad\text{Cov}(X, V)$
  $+ bc\text{Cov}(Y, W) + bd\text{Cov}(Y, V)$

- The above equation implies that

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y)$$

- $\text{Cov}(X, Y) = 0$ iff X and Y are independent

Readings

- Chapter 12.1–12.2 of our textbook

Random vector

- The collection of multiple random variables
- Mean value $\longrightarrow$ mean vector
- Variance $\longrightarrow$ covariance matrix
- k–th moment $\longrightarrow$ tensor of order k
- Fundamental building block in machine learning
- 詳見應用多變量分析

Random function

- A random function occurs when the result of a random trial is a function, e.g.
  - tomorrow's power generation curve by wind energy
  - an image by an automatic vehicle
  - random field
- 詳見隨機過程、函數型資料分析

Others

- Random intervals (symbolic data analysis)
- Random graphs
- …
- Refer to object-oriented data analysis