UNIVERSITY OF TORONTO | Statistical Sciences

# DoSS Summer Bootcamp Probability Module 6

Ichiro Hashimoto

University of Toronto

July 17, 2026

# Recap

Learnt in last module:

- Moments
  - ▷ Expectation, Raw moments, central moments
  - ▷ Moment-generating functions
- Change-of-variables using MGF
  - ▷ Gamma distribution
  - ▷ Chi square distribution
- Conditional expectation
  - ▷ Law of total expectation
  - ▷ Law of total variance

# Outline

- Covariance
  - Covariance as an inner product
  - Correlation
  - Cauchy-Schwarz inequality
  - Uncorrelatedness and Independence
- Concentration
  - Markov's inequality
  - Chebyshev's inequality
  - Chernoff bounds

# Covariance

**Recall the property of expectation:**

$$\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y).$$

# Covariance

**Recall the property of expectation:**

$$\mathbb{E}(X+Y)=\mathbb{E}(X)+\mathbb{E}(Y).$$

**What about the variance?**

$$\begin{aligned} Var(X+Y) &= \mathbb{E}(X+Y-\mathbb{E}(X)-\mathbb{E}(Y))^2 \\ &= \mathbb{E}(X-\mathbb{E}(X))^2+\mathbb{E}(Y-\mathbb{E}(Y))^2+2\mathbb{E}((X-\mathbb{E}(X))(Y-\mathbb{E}(Y))) \\ &= Var(X)+Var(Y)+2\,\underbrace{\mathbb{E}((X-\mathbb{E}(X))(Y-\mathbb{E}(Y)))}_{?} \end{aligned}$$

# Covariance

**Intuition:**

A measure of how much $X, Y$ change together.

# Covariance

**Intuition:**

A measure of how much $X, Y$ change together.

## Covariance

For two jointly distributed real-valued random variables $X, Y$ with finite second moments, the covariance is defined as

$$Cov(X, Y) = \mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))).$$

**Simplification:**

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

$$\mathbb{E}\left((X-\mathbb{E}X)(Y-\mathbb{E}Y)\right) = \mathbb{E}\left(XY - X\cdot\mathbb{E}Y - (\mathbb{E}X)\cdot Y + \mathbb{E}X\cdot\mathbb{E}Y\right)$$

$$= \mathbb{E}(XY) - \mathbb{E}(X\cdot\mathbb{E}Y) - \mathbb{E}[(\mathbb{E}X)\cdot Y] + \mathbb{E}(\mathbb{E}X\cdot\mathbb{E}Y)$$

constants

$$= \mathbb{E}(XY) - \mathbb{E}X\cdot\mathbb{E}Y \cancel{- \mathbb{E}X\cdot\mathbb{E}Y} \cancel{+ \mathbb{E}X\cdot\mathbb{E}Y}$$

# Covariance

**Properties:**

- $Cov(X, X) = Var(X) \geq 0$;
- $Cov(X, a) = 0$, $\quad a$ is a constant;
- $Cov(X, Y) = Cov(Y, X)$;
- $Cov(X + a, Y + b) = Cov(X, Y)$;
- $Cov(aX, bY) = abCov(X, Y)$.

$Cov(X, a) = \mathbb{E}\left((X - \mathbb{E}X)\underbrace{(a - \mathbb{E}a)}_{=0}\right) = \mathbb{E}(0) = 0$

$Cov(X+a, Y+b)$
$= \mathbb{E}\left((X+\cancel{a} - \mathbb{E}(X+\cancel{a}))(Y+\cancel{b} - \mathbb{E}(Y+\cancel{b}))\right)$
$= \mathbb{E}\left((X - \mathbb{E}X)(Y - \mathbb{E}Y)\right) = Cov(X, Y)$

$Cov(aX, bY) = \mathbb{E}\left((aX - \mathbb{E}(aX))(bY - \mathbb{E}(bY))\right) = \mathbb{E}\left(a(X - \mathbb{E}X) \cdot b(Y - \mathbb{E}Y)\right)$

$= ab\,\mathbb{E}\left((X - \mathbb{E}X)(Y - \mathbb{E}Y)\right) = ab\,Cov(X, Y)$

UNIVERSITY OF TORONTO

# Covariance

**Properties:**

- $Cov(X, X) = Var(X) \geq 0$;
- $Cov(X, a) = 0, \quad a$ is a constant;
- $Cov(X, Y) = Cov(Y, X)$;
- $Cov(X + a, Y + b) = Cov(X, Y)$;
- $Cov(aX, bY) = abCov(X, Y)$.

**Corollary about variance:**

$$Var(aX + b) = a^2 Var(X).$$

# Covariance

## Relate covariance to inner product:

**Inner product (not rigorous)**

Inner product is a operator from a vector space $V$ to a field $F$ (use $\mathbb{R}$ here as an example): $< \cdot, \cdot >: V \times V \to \mathbb{R}$ that satisfies:

- Symmetry: $< x, y > = < y, x >$;

$\mathbb{R}$

- Linearity in the first argument: $< ax + by, z > = a < x, z > + b < y, z >$;
- Positive-definiteness: $< x, x > \geq 0$, and $< x, x > = 0 \Leftrightarrow x = 0$

$\langle x, x \rangle \neq 0 \Leftrightarrow x \neq 0$

# Covariance

## Relate covariance to inner product:

**Inner product (not rigorous)**

Inner product is a operator from a vector space $V$ to a field $F$ (use $\mathbb{R}$ here as an example): $< \cdot, \cdot >$: $V \times V \to \mathbb{R}$ that satisfies:

- Symmetry: $< x, y > = < y, x >$;
- Linearity in the first argument: $< ax + by, z > = a < x, z > + b < y, z >$;
- Positive-definiteness: $< x, x > \geq 0$, and $< x, x > = 0 \Leftrightarrow x = 0$

**Remark:**

Covariance defines an inner product over the quotient vector space obtained by taking the subspace of random variables with finite second moment and identifying any two that differ by a constant.

# Covariance

### Properties inherited from the inner product space

Recall in Euclidean vector space:

- $< x, y >= x^\top y = \sum_{i=1}^{n} x_i y_i$;
- $||x||_2 = \sqrt{< x, x >}$;
- $< x, y >= ||x||_2 \cdot ||y||_2 \cos(\theta)$.

$\theta$ is angle between $x, y$

$$\cos\theta = \frac{< x, y >}{||x|| \, ||y||}$$

Respectively:

- $< X, Y >= Cov(X, Y)$;
- $||X|| = \sqrt{Var(X)}$;

# Covariance

**A substitute for** $\cos(\theta)$**:**

**Correlation**

For two jointly distributed real-valued random variables $X, Y$ with finite second moments, the correlation is defined as

$$Corr(X, Y) = \rho_{XY} = \frac{Cov(X, Y)}{\sqrt{Var(X) \cdot Var(Y)}}. = \frac{\langle X, Y \rangle}{\|X\| \|Y\|}$$

# Covariance

**A substitute for $\cos(\theta)$:**

**Correlation**

For two jointly distributed real-valued random variables $X, Y$ with finite second moments, the correlation is defined as

$$Corr(X, Y) = \rho_{XY} = \frac{Cov(X, Y)}{\sqrt{Var(X) \cdot Var(Y)}}.$$

**Uncorrelatedness:**

$$X, Y \text{ uncorrelated} \quad \Leftrightarrow \quad Corr(X, Y) = 0.$$

# Covariance

implies $|Corr(X,Y)| \leq 1$

**Cauchy-Schwarz inequality**

$$|Cov(X, Y)| \leq \sqrt{Var(X)Var(Y)}.$$

**Proof:** $\hat{X} = X - \mathbb{E}X$, $\hat{Y} = Y - \mathbb{E}Y$

$$0 \leq \mathbb{E}(\hat{X} + t\hat{Y})^2 = \mathbb{E}\hat{X}^2 + 2(\mathbb{E}(\hat{X}\cdot\hat{Y}))\cdot t + (\mathbb{E}\hat{Y}^2)\cdot t^2$$

$t \in \mathbb{R}$

$$= \underbrace{Var(X) + 2\,Cov(X,Y)\cdot t + Var(Y)\cdot t^2}_{\text{quadratic}}$$

Since this inequality holds for any $t$,

we must have $\quad D/4 = Cov(X,Y)^2 - Var(X)\cdot Var(Y) \leq 0$

$\therefore |Cov(X,Y)| \leq \sqrt{Var(X)\cdot Var(Y)}$

# Covariance

$ax^2+bx+c=0 \qquad D = b^2-4ac.$

**Uncorrelatedness and Independence:**

Observe the relationship:

$$Corr(X,Y)=0 \quad \Leftrightarrow \quad Cov(X,Y)=0 \quad \Leftrightarrow \quad \mathbb{E}(XY)=\mathbb{E}(X)\mathbb{E}(\overset{Y}{\cancel{X}})$$

$$\Uparrow \qquad \cancel{\Downarrow}$$

X and Y are independent.

UNIVERSITY OF TORONTO

# Covariance

**Uncorrelatedness and Independence:**

Observe the relationship:

$$Corr(X, Y) = 0 \quad \Leftrightarrow \quad Cov(X, Y) = 0 \quad \Leftrightarrow \quad \mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(X)$$

**Conclusions:**

- Independence $\Rightarrow$ Uncorrelatedness
- Uncorrelatedness $\not\Longrightarrow$ Independence

**Remark:**
Independence is a very strong assumption/property on the distribution.

# Covariance

### Special case: multivariate normal

**Multivariate normal**

A $k$-dimensional random vector $\mathbf{X} = (X_1, X_2, \cdots, X_k)^\top$ follows a multivariate normal distribution $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, if

$$f_{\mathbf{X}}(x_1, \ldots, x_k) = \frac{\exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})\right)}{\sqrt{(2\pi)^k |\boldsymbol{\Sigma}|}},$$

$:$ determinant of $\Sigma$

where $\boldsymbol{\mu} = \mathbb{E}[\mathbf{X}] = (\mathbb{E}[X_1], \mathbb{E}[X_2], \ldots, \mathbb{E}[X_k])^\top$, and $[\boldsymbol{\Sigma}]_{i,j} = \Sigma_{i,j} = Cov(X_i, X_j)$.

**Observation:**

The distribution is decided by the covariance structure.

# Covariance

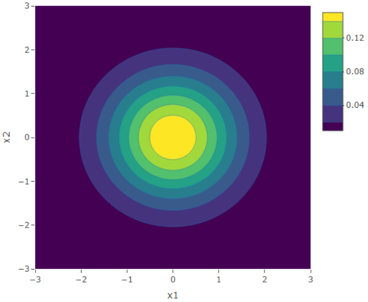
$$X_i, i = 1, \cdots k \text{ independent} \Leftrightarrow f_{\mathbf{X}}(x_1, \ldots, x_k) = \prod_{i=1}^{k} f_{X_i}(x_i)$$

$$\Leftrightarrow \mathbf{\Sigma} \cancel{= I_k} \Leftrightarrow Cov(X_i, X_j) = 0, i \neq j.$$

is diagonal

**Example:**

- $Corr(X, Y) = 0$




UNIVERSITY OF TORONTO

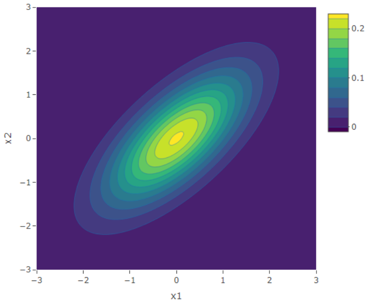
# Covariance

$$X_i, i = 1, \cdots k \text{ independent} \Leftrightarrow f_{\mathbf{X}}(x_1, \ldots, x_k) = \prod_{i=1}^{k} f_{X_i}(x_i)$$
$$\Leftrightarrow \mathbf{\Sigma} = I_k \Leftrightarrow Cov(X_i, X_j) = 0, i \neq j.$$

**Example:**

- $Corr(X, Y) = \underline{0.7}$

positive



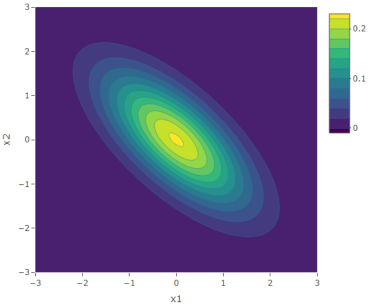
# Covariance

$$X_i, i = 1, \cdots k \text{ independent} \Leftrightarrow f_{\mathbf{X}}(x_1, \ldots, x_k) = \prod_{i=1}^{k} f_{X_i}(x_i)$$

$$\Leftrightarrow \boldsymbol{\Sigma} = I_k \Leftrightarrow Cov(X_i, X_j) = 0, i \neq j.$$

**Example:**

- $Corr(X, Y) = -0.7$




UNIVERSITY OF TORONTO

# Concentration

## Measures of a distribution:

- $\mathbb{E}(X^k)$, $\mathbb{E}(X)$, $Var(X)$;
- $Cov(X, Y)$ and $Corr(X, Y)$.

UNIVERSITY OF TORONTO

# Concentration

**Measures of a distribution:**

- $\mathbb{E}(X^k)$, $\mathbb{E}(X)$, $Var(X)$;
- $Cov(X, Y)$ and $Corr(X, Y)$.

**Tail probability: $\mathbf{P}(|X| > t)$**

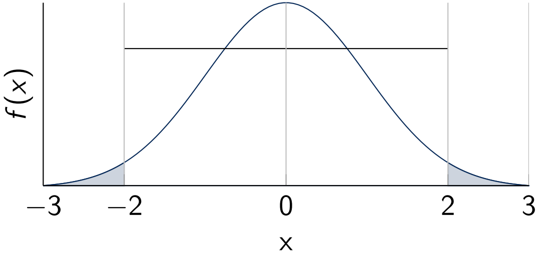



Figure: Probability density function of $\mathcal{N}(0, 1)$

# Concentration

## Concentration inequalities:

- Markov inequality
- Chebyshev inequality
- Chernoff bounds

# Concentration

## Concentration inequalities:

- Markov inequality
- Chebyshev inequality
- Chernoff bounds

### Markov inequality

Let $X$ be a random variable that is non-negative (almost surely). Then, for every constant $a > 0$,

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}(X)}{a}.$$

**Proof:** We use the following fact:

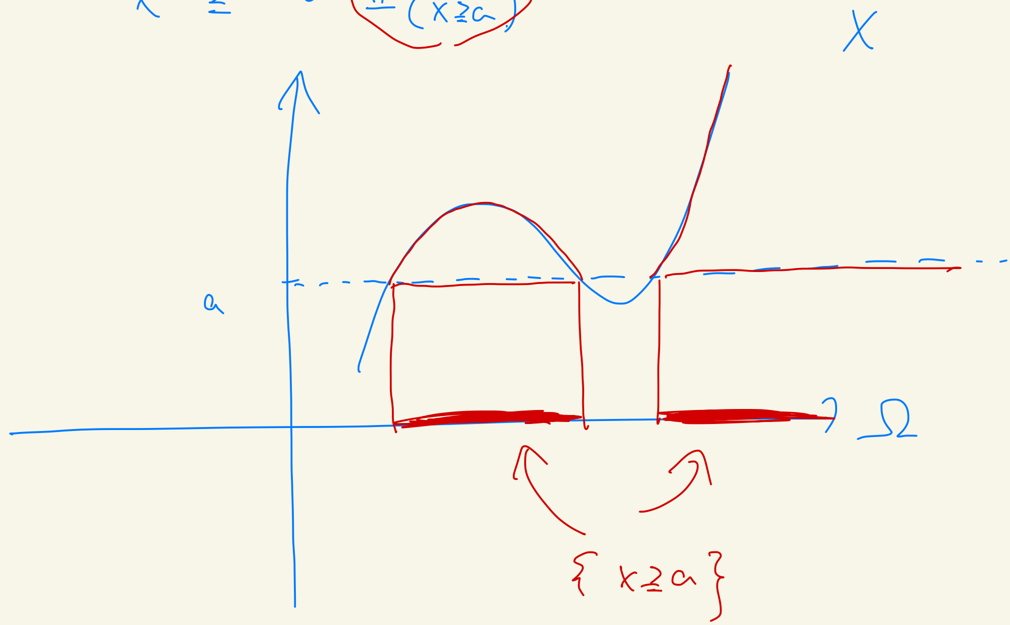
if $X \leq Y$, then $\mathbb{E} X \leq \mathbb{E} Y$ (monotonicity of expectation)

UNIVERSITY OF TORONTO

Observe that

$$X \geq a \, \mathbb{1}_{(X \geq a)}$$

$$\mathbb{1}_A(\omega) = \begin{cases} 1 \cdots \text{if } \omega \in A \\ 0 \cdots \text{otherwise} \end{cases}$$




By the monotonicity of expectation

$$\mathbb{E}X \geq \mathbb{E}(a \mathbb{1}_{X \geq a})$$

$$= a \, \mathbb{E}(\mathbb{1}_{X \geq a})$$

$$= a \left( 1 \cdot \mathbb{P}(X \geq a) + 0 \cdot \mathbb{P}(X < a) \right)$$

$$= a \, \mathbb{P}(X \geq a)$$

$$\therefore \quad \mathbb{P}(X \geq a) \leq \frac{\mathbb{E}X}{a}$$

# Concentration

**Markov inequality (continued)**

Let $X$ be a random variable, then for every constant $a > 0$,

$$\mathbb{P}(|X| \geq a) \leq \frac{\mathbb{E}(|X|)}{a}.$$

This applies to any random variable

## A more general conclusion:

**Markov inequality (continued)**

Let $X$ be a random variable, if $\Phi(x)$ is monotonically increasing on $[0, \infty)$, then for every constant $a > 0$,

apply Markov

$$\mathbb{P}(|X| \geq a) = \mathbb{P}(\Phi(|X|) \geq \Phi(a)) \leq \frac{\mathbb{E}(\Phi(|X|))}{\Phi(a)}.$$

# Concentration

## Chebyshev inequality

Let $X$ be a random variable with finite expectation $\mathbb{E}(X)$ and variance $Var(X)$, then for every constant $a > 0$,

choose $\phi(x) = x^2$

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq a) \leq \frac{Var(X)}{a^2}, = \frac{1}{b^2}$$

Then

$a^2 = b^2 Var(x)$

$a = b\sqrt{Var(X)}$

or equivalently,

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq a\sqrt{Var(X)}) \leq \frac{1}{a^2}.$$

$\mathbb{P}(|X - \mathbb{E}X| \geq a)$

$= \mathbb{P}(|X - \mathbb{E}X| \geq b\sqrt{Var(X)})$

$\leq \frac{1}{b^2}$

**Example:**

Take $a = 2$,

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq 2\sqrt{Var(X)}) \leq \frac{1}{4}.$$

UNIVERSITY OF TORONTO

# Concentration

## Chernoff bound (general)

Let $X$ be a random variable, then for $t \geq 0$,

choose $\phi(x) = e^x$.

$$\mathbb{P}(X \geq a) \overset{\leq}{=} \mathbb{P}(e^{t\cdot X} \geq e^{t\cdot a}) \leq \frac{\mathbb{E}\left[e^{t\cdot X}\right]}{e^{t\cdot a}},$$

← use Markov

and

$$\mathbb{P}(X \geq a) \leq \boxed{\inf_{t\geq 0} \frac{\mathbb{E}\left[e^{t\cdot X}\right]}{e^{t\cdot a}}.}$$

Find $t$ which minimizes

$$\frac{\mathbb{E}[e^{tx}]}{e^{ta}}$$

**Remark:**

This is especially useful when considering $X = \sum_{i=1}^{n} X_i$ with $X_i$'s independent,

$$\mathbb{P}(X \geq a) \leq \inf_{t\geq 0} \frac{\mathbb{E}\left[\prod_i e^{t\cdot X_i}\right]}{e^{t\cdot a}} = \inf_{t\geq 0} e^{-t\cdot a} \prod_i \mathbb{E}\left[e^{t\cdot X_i}\right].$$

Suppose $X_i \overset{i.i.d}{\sim} Z$, $i = 1, \dots, n$

$$\mathbb{P}\Big(\underbrace{X}_{\sum_{i=1}^n X_i} \geq a\Big) \leq \inf_{t>0} \frac{(\mathbb{E}\, e^{tZ})^n}{e^{t \cdot a}}$$

Further assume $Z \sim \text{Bern}(\tfrac{1}{2})$

$$\mathbb{E}\, e^{tZ} = \frac{e^{t} + e^{-t}}{2} \leq e^{t}$$

This is not useful

Then $\mathbb{P}\left(\sum_{i=1}^n X_i \geq a\right) \leq \inf_{t>0} \dfrac{e^{tn}}{e^{ta}}$.

$$= \inf_{t>0} e^{t(n-a)}$$

$$\mathbb{P}\left(\sum_{i=1}^n X_i \geq a\right) = \inf_{t>0} \frac{\left(\frac{e^{t} + e^{-t}}{2}\right)^n}{e^{ta}}$$

$\Downarrow$.

Hoeffding's bound

# Problem Set

**Problem 1:** Let

$$f_{X,Y}(x,y) = \begin{cases} 2 & 0 \leq y \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases},$$

compute $Cov(X, Y)$.

**Problem 2:** For $X \sim \mathcal{N}(0, 1)$, compute the Chernoff bound.