



UNIVERSITY OF
TORONTO

Statistical Sciences

DoSS Summer Bootcamp Probability Module 6

Ichiro Hashimoto

University of Toronto

July 17, 2025

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?

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

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

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

Simplification:

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

Covariance

Properties:

- $\text{Cov}(X, X) = \text{Var}(X) \geq 0$;
- $\text{Cov}(X, a) = 0$, a is a constant;
- $\text{Cov}(X, Y) = \text{Cov}(Y, X)$;
- $\text{Cov}(X + a, Y + b) = \text{Cov}(X, Y)$;
- $\text{Cov}(aX, bY) = ab\text{Cov}(X, Y)$.

Covariance

Properties:

- $\text{Cov}(X, X) = \text{Var}(X) \geq 0$;
- $\text{Cov}(X, a) = 0$, a is a constant;
- $\text{Cov}(X, Y) = \text{Cov}(Y, X)$;
- $\text{Cov}(X + a, Y + b) = \text{Cov}(X, Y)$;
- $\text{Cov}(aX, bY) = ab\text{Cov}(X, Y)$.

Corollary about variance:

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

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): $\langle \cdot, \cdot \rangle: V \times V \rightarrow \mathbb{R}$ that satisfies:

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

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): $\langle \cdot, \cdot \rangle: V \times V \rightarrow \mathbb{R}$ that satisfies:

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

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:

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

Respectively:

- $\langle X, Y \rangle = \text{Cov}(X, Y)$;
- $\|X\| = \sqrt{\text{Var}(X)}$;

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

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

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

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

Uncorrelatedness:

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

Covariance

Cauchy-Schwarz inequality

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

Proof:

Covariance

Uncorrelatedness and Independence:

Observe the relationship:

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

Covariance

Uncorrelatedness and Independence:

Observe the relationship:

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

Conclusions:

- Independence \Rightarrow Uncorrelatedness
- Uncorrelatedness \nRightarrow Independence

Remark:

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

Covariance

Special case: multivariate normal

Multivariate normal

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

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

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

Observation:

The distribution is decided by the covariance structure.

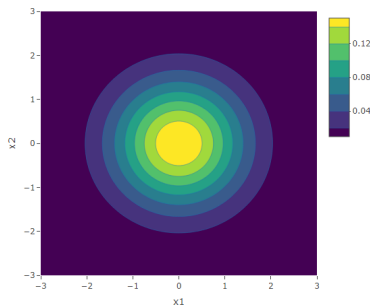
Covariance

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

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

Example:

- $\text{Corr}(X, Y) = 0$



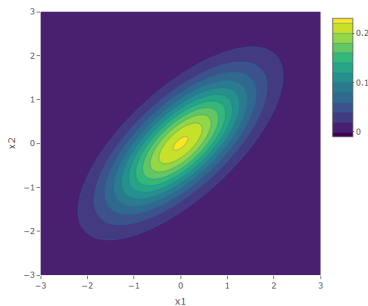
Covariance

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

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

Example:

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



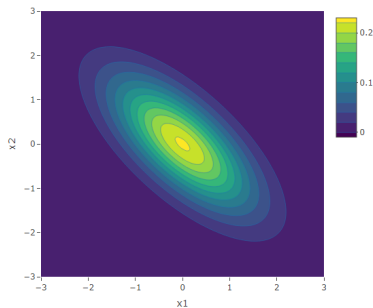
Covariance

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

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

Example:

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



Concentration

Measures of a distribution:

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

Concentration

Measures of a distribution:

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

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

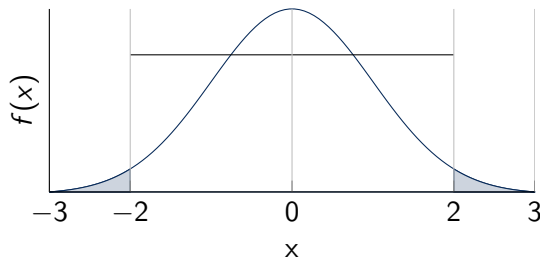


Figure: Probability density function of $\mathcal{N}(0, 1)$

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:

Concentration

Markov inequality (continued)

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

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

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 $a > 0$,

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

Concentration

Chebyshev inequality

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

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

or equivalently,

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

Example:

Take $a = 2$,

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

Concentration

Chernoff bound (general)

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

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

and

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

Remark:

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

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

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 $\text{Cov}(X, Y)$.

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