UNIVERSITY OF TORONTO | Statistical Sciences

DoSS Summer Bootcamp Probability Module 6

Miaoshiqi (Shiki) Liu

University of Toronto

July 21, 2022

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?

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

UNIVERSITY OF TORONTO

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:

$$Var(Y) = E((X - E(X))(X - E(X)))$$

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

$$Cov(X, X) = Var(X)$$

$$Cov(Y, Y) = Var(Y)$$

Covariance

Properties:

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

$$Var(a) = 0$$

$$Cov(X,a) = E\big((\underbrace{X - E(X)})(a-a)\big) = 0.$$

$$Cov(X+a, Y+b) = E\big(((X+a) - E(X+a))(Y+b-E(Y+b))\big)$$

$$= E\big((X-E(X))(Y-E(Y))\big)$$

$$Cov(aX, bY) = E\big((aX - aEX)(bY - bEY)\big)$$

$$= E\big(ab(X-EX)(Y-EY)\big)$$

$$= ab\, Cov(X,Y).$$

Covariance

Properties:

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

Corollary about variance:

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

$$Cov(ax+b, ax+b) = Cov(aX, ax)$$
$$= a^2 Var(X)$$

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:

$\mathbb{C}$

$<x,y> = \overline{<y,x>}$

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

UNIVERSITY OF TORONTO

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: $< x, y > = < y, x >$; $Cov(X,Y) = Cov(Y,X)$ $Cov(aX+bY, Z)$
- Linearity in the first argument: $< ax + by, z > = a < x, z > + b < y, z >$; $= \cdots$
- Positive-definiteness: $< x, x > \geq 0$, and $< x, x > = 0 \Leftrightarrow x = 0$

$Cov(X,X) = Var(X) \geq 0$ $Var(X) = 0$

$X = EX$ a.s.

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.

$X, \quad X+a. \quad \forall$ constant a.

$X = 0$. a.s.

$Cov(X,X) = 0 \Rightarrow X = a$. a.s

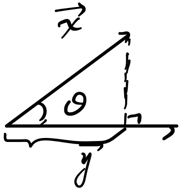
Covariance

Properties inherited from the inner product space

Recall in Euclidean vector space:

$\mathbb{R}^n \quad x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \quad y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$

- $< x, y >= x^\top y = \sum_{i=1}^n x_i y_i;$
- $||x||_2 = \sqrt{< x, x >};$ $= \sqrt{\sum_{i=1}^n x_i^2}$
- $< x, y >= ||x||_2 \cdot ||y||_2 \cos(\theta).$



$\|\vec{x}\|_2 \|\vec{y}\|_2 \cos(\theta)$

$\underbrace{\|\vec{x}\| \cos\theta} \cdot \|\vec{y}\|$

Respectively:

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

$? = \dfrac{Cov(X,Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}}$

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

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

UNIVERSITY OF TORONTO

Covariance

A substitute for $\cos(\theta)$**:** $|\cos\theta| \leq 1$

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:

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

$$\Longleftrightarrow \quad Cov(X, Y) = 0.$$

Covariance

$|\rho| \le 1$

Cauchy-Schwarz inequality

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

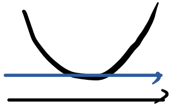
Proof:

$$Var(X + aY) = Cov(X+aY,\, X+aY)$$

$$= Var(X) + \underbrace{Cov(X, aY)} + \underbrace{Cov(aY, X)} + a^2 Var(Y)$$

$$= \underbrace{Var(X) + 2a\,Cov(X,Y) + a^2 Var(Y)}_{f(a)} \ge 0$$

1° $Var(Y) = 0$. $Y = \text{constant. a.s.}$ $Cov(X,Y) = 0$.



2° $Var(Y) > 0$. $f(a) \ge 0$

$$f(a) = \underbrace{Var(Y)}_{A} \cdot a^2 + \underbrace{2\,Cov(X,Y)}_{B} \cdot a + \underbrace{Var(X)}_{C}.$$

$$\Delta = B^2 - 4AC \le 0$$

$$4\,Cov^2(X,Y) \le 4\,Var(X)\,Var(Y)$$

Covariance

Uncorrelatedness and Independence:

Observe the relationship:

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

$E(X) \cdot E(Y)$.

Covariance

Uncorrelatedness and Independence:

X, Y ind

$\Rightarrow E(XY) = E(X)E(Y)$

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

(μ, σ^2) mean variance.

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

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.

$$\begin{pmatrix} Cov(X_1, X_1) & \cdots & Cov(X_1, X_k) \\ \vdots & & \\ Cov(X_k, X_1) & \cdots & Cov(X_k, X_k) \end{pmatrix}$$

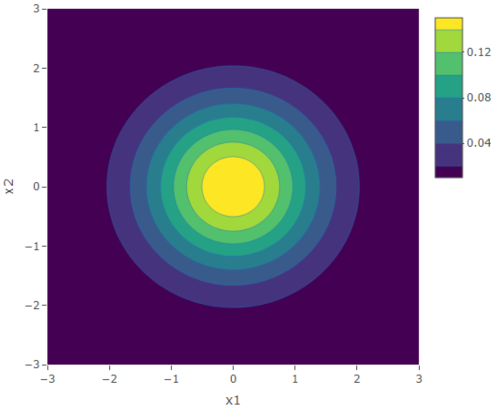
Covariance

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

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

Example:

- $Corr(X, Y) = 0$



joint

UNIVERSITY OF TORONTO

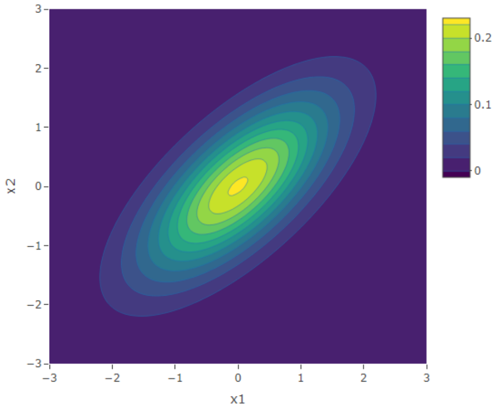
Covariance

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

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

Example:

- $Corr(X, Y) = 0.7$



UNIVERSITY OF TORONTO

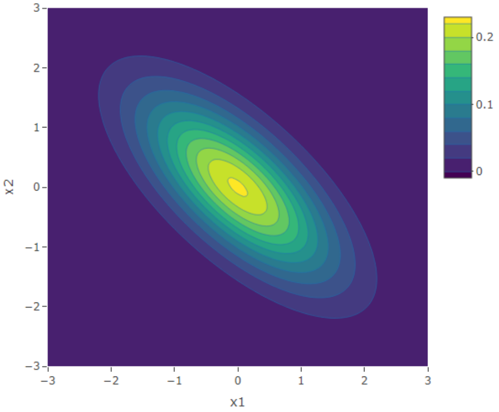
Covariance

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

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

Example:

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



UNIVERSITY OF TORONTO

Concentration

Measures of a distribution:

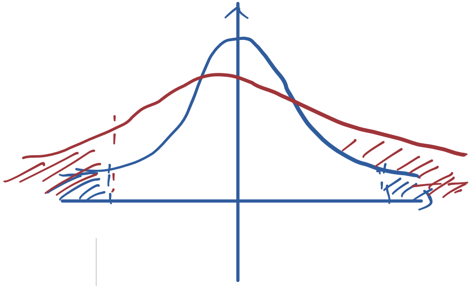
- $\mathbb{E}(X^k)$, $\mathbb{E}(X)$, $Var(X)$;
- $Cov(X, Y)$ and $Corr(X, Y)$.

Concentration

Measures of a distribution:

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

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



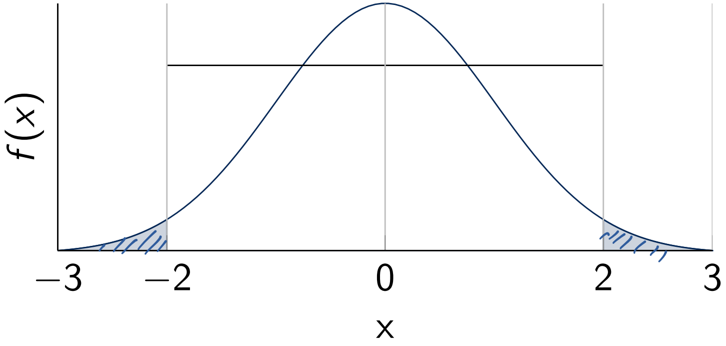


Figure: Probability density function of $\mathcal{N}(0, 1)$

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 $a > 0$,

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

continuous.

Proof:

$$E(X) = \int_{-\infty}^{+\infty} x f_X(x)\,dx = \int_{\{x \geq a\}} \underbrace{x}_{\geq a} f_X(x)\,dx + \int_{\{x<a\}} \underbrace{x f_X(x)}_{\geq 0}\,dx$$

$$\geq \int_{\{x \geq a\}} x f_X(x)\,dx$$

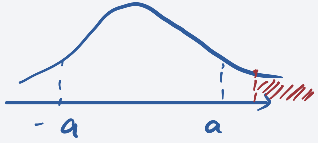
$$\geq \int_{\{x \geq a\}} a f_X(x)\,dx \rightarrow a P(X \geq a)$$

Concentration

Markov inequality (continued)

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

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



A more general conclusion:

≥ 0

Markov inequality (continued)

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

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

Concentration

Chebyshev inequality

Let X be a random variable with finite expectation $\mathbb{E}(X)$ and variance $Var(X)$, then for every constant $a > 0$, $P((X - E(X))^2 \geq a^2) \leq \frac{E((X - E(X))^2)}{a^2}$

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

or equivalently,

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

Example:

Take $a = 2$,

$P(|X| \geq 2\sqrt{Var(X)}) \approx 10\%$

$X \sim N(0,1)$

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

UNIVERSITY OF TORONTO

Concentration

Chernoff bound (general)

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

$f(x) = e^{tx}$

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

$M_X(t)$

and

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

Remark:

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

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

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.