02/05/2026: Continuous Random Variables (Part 2)

CSCI 546: Diffusion Models

Textbook reference: Sec 4.6-4.10

Announcement (Sign-in Sheet)

Please sign the sign-in sheet.

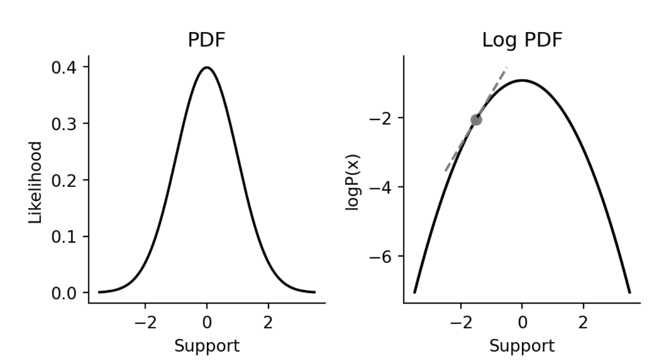
Review Problem Set #6

Concepts for Problem Set #7

Stein's Identity

Let X be random variable with unknown probability density function p. Score-based diffusion models (Module 2B) need to learn the *score function*

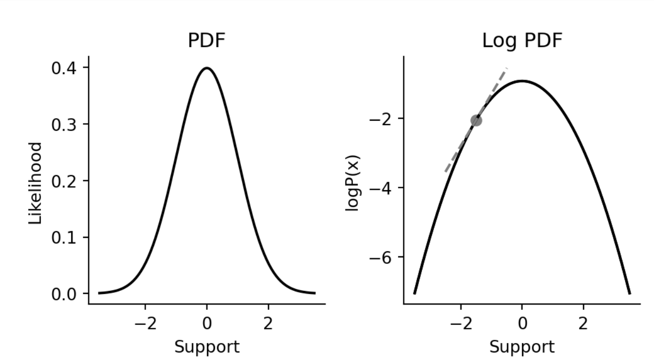
$$\nabla_x \log p(x)$$



Stein's Identity

Let X be random variable with unknown probability density function p. Score-based diffusion models (Module 2B) need to learn the *score function*

$$\nabla_x \log p(x)$$



Problem. How to learn the score function without knowing the density p?

Stein's Identity

Let X be random variable with unknown probability density function p. Score-based diffusion models (Module 2B) need to learn the *score function*

$$\nabla_x \log p(x)$$

Stein's Identity

With some abuse of notation, we write Stein's Identity as

$$\mathbb{E}[g(X)\nabla \log p(X)] = -\mathbb{E}[\nabla g(X)].$$

for sufficiently "nice" density p and test function g.

Stein's Identity

Let X be random variable with unknown probability density function p. Score-based diffusion models (Module 2B) need to learn the *score function*

$$\nabla_x \log p(x)$$

Stein's Identity

With some abuse of notation, we write Stein's Identity as

$$\mathbb{E}[g(X)\nabla \log p(X)] = -\mathbb{E}[\nabla g(X)].$$

for sufficiently "nice" density p and test function g.

Implications.

- The RHS does not involve p explicitly.
- It only involves expectations over samples (which can be approximated empirically).

Multivariate Normal distribution

The random vector $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ has the **multivariate normal distribution** (or **multivariate Gaussian distribution**), written $\mathcal{N}(\boldsymbol{\mu}, \Sigma)$, if it has the joint probability density function

$$f(\boldsymbol{x}) = \frac{1}{(2\pi)^{d/2}\,|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^\top \Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right), \qquad \boldsymbol{x} \in \mathbb{R}^n.$$

Multivariate Normal distribution

The random vector $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ has the **multivariate normal distribution** (or **multivariate Gaussian distribution**), written $\mathcal{N}(\boldsymbol{\mu}, \Sigma)$, if it has the joint probability density function

$$f(\boldsymbol{x}) = \frac{1}{(2\pi)^{d/2}\,|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^\top \Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right), \qquad \boldsymbol{x} \in \mathbb{R}^n.$$

- $\boldsymbol{\mu}$ is the mean $\boldsymbol{E}[\boldsymbol{X}]$.

Multivariate Normal distribution

The random vector $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ has the **multivariate normal distribution** (or **multivariate Gaussian distribution**), written $\mathcal{N}(\boldsymbol{\mu}, \Sigma)$, if it has the joint probability density function

$$f(\boldsymbol{x}) = \frac{1}{(2\pi)^{d/2}\,|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^\top \Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right), \qquad \boldsymbol{x} \in \mathbb{R}^n.$$

- $\boldsymbol{\mu}$ is the mean $\boldsymbol{E}[\boldsymbol{X}]$.
- $\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_{nn} \end{pmatrix}$ is the covariance matrix.

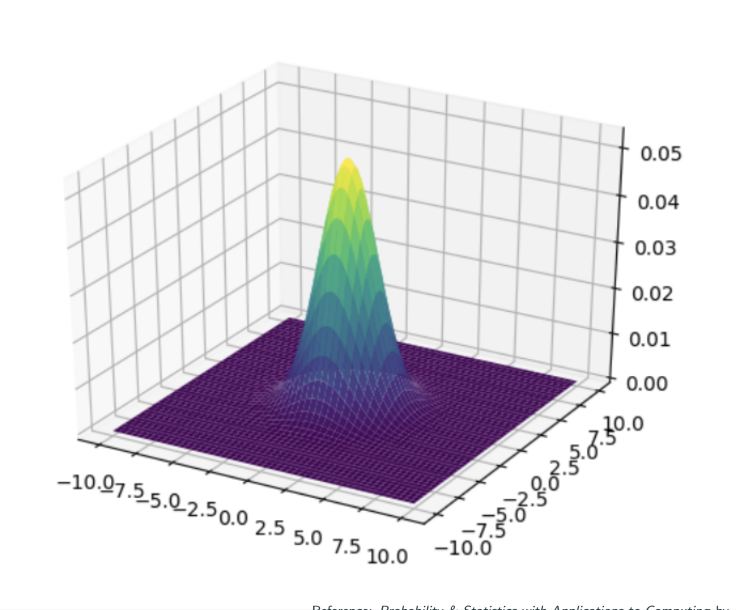
Multivariate Normal distribution

The random vector $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ has the **multivariate normal distribution** (or **multivariate Gaussian distribution**), written $\mathcal{N}(\boldsymbol{\mu}, \Sigma)$, if it has the joint probability density function

$$f(\boldsymbol{x}) = \frac{1}{(2\pi)^{d/2} |\Sigma|^{1/2}} \exp\left(-\frac{1}{2} (\boldsymbol{x} - \boldsymbol{\mu})^\top \Sigma^{-1} (\boldsymbol{x} - \boldsymbol{\mu}) \right), \qquad \boldsymbol{x} \in \mathbb{R}^n.$$

- $\boldsymbol{\mu}$ is the mean $\boldsymbol{E}[\boldsymbol{X}]$.
- $\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_{nn} \end{pmatrix}$ is the covariance matrix.
    - $\sigma_{11} = \mathrm{Cov}(X_1, X_1) = \mathrm{Var}(X_1)$
    - $\sigma_{12} = \mathrm{Cov}(X_1, X_2)$
    - $\vdots$

Example: Bivariate Normal distribution



Reference: *Probability & Statistics with Applications to Computing* by Alex Tsun

Example: Bivariate Normal distribution

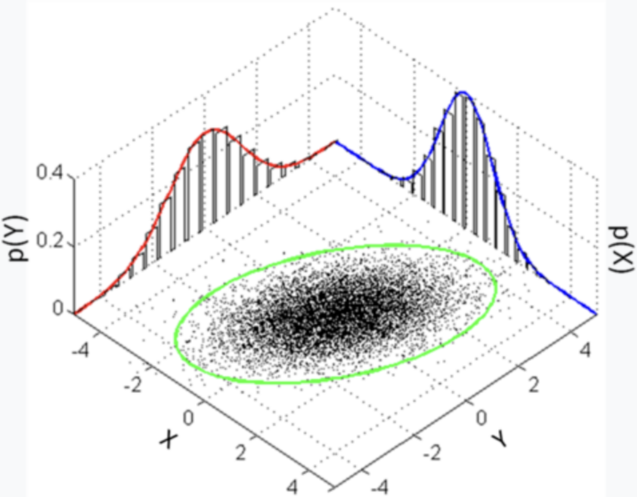
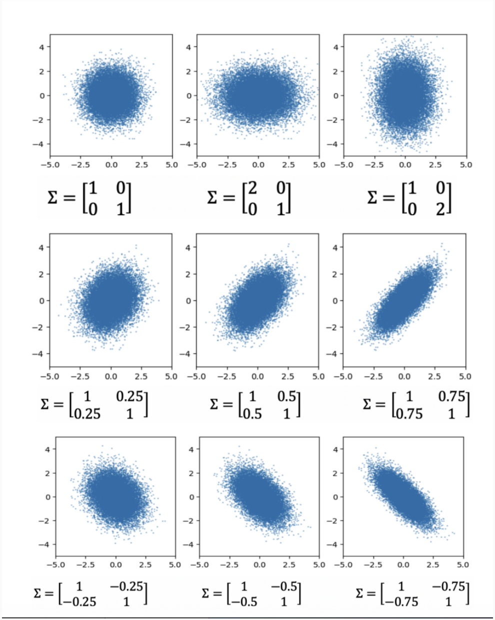


Figure 1: Many sample points from a bivariate normal distribution with $\boldsymbol{\mu} = \left[\begin{smallmatrix}0\\0\end{smallmatrix}\right]$ and $\Sigma = \begin{bmatrix} 1 & 3/5 \\ 3/5 & 2 \end{bmatrix}$, shown along with the 3-sigma ellipse, the two marginal distributions, and the two 1-d histograms.

Reference: Wikipedia



Poll. What is happening in each row?

Properties of covariance matrices

Covariance matrices Σ have two useful properties

1. **Symmetry.** A $d \times d$ matrix Σ is symmetric if $\Sigma^\top = \Sigma$. That is,

$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_{nn} \end{pmatrix} = \begin{pmatrix} \sigma_{11} & \sigma_{21} & \cdots & \sigma_{n1} \\ \sigma_{12} & \sigma_{22} & \cdots & \sigma_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1n} & \sigma_{2n} & \cdots & \sigma_{nn} \end{pmatrix} = \Sigma^\top.$$

Properties of covariance matrices

Covariance matrices Σ have two useful properties

1. **Symmetry.** A $d \times d$ matrix Σ is symmetric if $\Sigma^\top = \Sigma$. That is,

$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_{nn} \end{pmatrix} = \begin{pmatrix} \sigma_{11} & \sigma_{21} & \cdots & \sigma_{n1} \\ \sigma_{12} & \sigma_{22} & \cdots & \sigma_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1n} & \sigma_{2n} & \cdots & \sigma_{nn} \end{pmatrix} = \Sigma^\top.$$

2. **Positive semi-definiteness.** A $n \times n$ matrix Σ is positive semi-definite if $\boldsymbol{x}^\top \Sigma \boldsymbol{x} \geq 0$ for all $\boldsymbol{x} \in \mathbb{R}^n$.

Matrix Square Roots

Theorem. A symmetric matrix Σ has a symmetric square root $\boldsymbol{Q}$ (where $\boldsymbol{Q}^2 = \Sigma$) if it is positive semi-definite.

Matrix Square Roots

Theorem. A symmetric matrix Σ has a symmetric square root $\boldsymbol{Q}$ (where $\boldsymbol{Q}^2 = \Sigma$) if it is positive semi-definite.

Method.

1. Diagonalize the symmetric matrix $\Sigma = \boldsymbol{U}\boldsymbol{D}\boldsymbol{U}^\top$, where $\boldsymbol{U}$ is an orthonormal and $\boldsymbol{D}$ is a diagonal matrix.

Matrix Square Roots

Theorem. A symmetric matrix Σ has a symmetric square root $\boldsymbol{Q}$ (where $\boldsymbol{Q}^2 = \Sigma$) if it is positive semi-definite.

Method.

1. Diagonalize the symmetric matrix $\Sigma = \boldsymbol{U}\boldsymbol{D}\boldsymbol{U}^\top$, where $\boldsymbol{U}$ is an orthonormal and $\boldsymbol{D}$ is a diagonal matrix.
   - $\boldsymbol{U}$ is orthonormal means $\boldsymbol{U}^\top\boldsymbol{U} = \boldsymbol{I} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}$.

Matrix Square Roots

Theorem. A symmetric matrix Σ has a symmetric square root $\boldsymbol{Q}$ (where $\boldsymbol{Q}^2 = \Sigma$) if it is positive semi-definite.

Method.

1. Diagonalize the symmetric matrix $\Sigma = \boldsymbol{U}\boldsymbol{D}\boldsymbol{U}^\top$, where $\boldsymbol{U}$ is an orthonormal and $\boldsymbol{D}$ is a diagonal matrix.
   - $\boldsymbol{U}$ is orthonormal means $\boldsymbol{U}^\top\boldsymbol{U} = \boldsymbol{I} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}$.
   - $\boldsymbol{D}$ is diagonal means $\boldsymbol{D} = \begin{pmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & d_n \end{pmatrix}$.

Matrix Square Roots

Theorem. A symmetric matrix Σ has a symmetric square root $\boldsymbol{Q}$ (where $\boldsymbol{Q}^2 = \Sigma$) if it is positive semi-definite.

Method.

1. Diagonalize the symmetric matrix $\Sigma = \boldsymbol{U}\boldsymbol{D}\boldsymbol{U}^\top$, where $\boldsymbol{U}$ is an orthonormal and $\boldsymbol{D}$ is a diagonal matrix.
   - $\boldsymbol{U}$ is orthonormal means $\boldsymbol{U}^\top \boldsymbol{U} = \boldsymbol{I} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}$.
   - $\boldsymbol{D}$ is diagonal means $\boldsymbol{D} = \begin{pmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & d_n \end{pmatrix}$.
2. Compute the diagonal matrix of square roots: $\boldsymbol{D}^{1/2}$. (Just take the square root of each diagonal element.)

Matrix Square Roots

Theorem. A symmetric matrix Σ has a symmetric square root $\boldsymbol{Q}$ (where $\boldsymbol{Q}^2 = \Sigma$) if it is positive semi-definite.

Method.

1. Diagonalize the symmetric matrix $\Sigma = \boldsymbol{U}\boldsymbol{D}\boldsymbol{U}^\top$, where $\boldsymbol{U}$ is an orthonormal and $\boldsymbol{D}$ is a diagonal matrix.
   - $\boldsymbol{U}$ is orthonormal means $\boldsymbol{U}^\top \boldsymbol{U} = \boldsymbol{I} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}$.
   - $\boldsymbol{D}$ is diagonal means $\boldsymbol{D} = \begin{pmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & d_n \end{pmatrix}$.
2. Compute the diagonal matrix of square roots: $\boldsymbol{D}^{1/2}$. (Just take the square root of each diagonal element.)
3. Compute $\boldsymbol{Q} = \Sigma^{1/2} = \boldsymbol{U}\boldsymbol{D}^{1/2}\boldsymbol{U}^\top$.

Matrix Square Roots

Theorem. A symmetric matrix Σ has a symmetric square root $\boldsymbol{Q}$ (where $\boldsymbol{Q}^2 = \Sigma$) if it is positive semi-definite.

Method.

1. Diagonalize the symmetric matrix $\Sigma = \boldsymbol{U}\boldsymbol{D}\boldsymbol{U}^\top$, where $\boldsymbol{U}$ is an orthonormal and $\boldsymbol{D}$ is a diagonal matrix.
   - $\boldsymbol{U}$ is orthonormal means $\boldsymbol{U}^\top\boldsymbol{U} = \boldsymbol{I} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}$.
   - $\boldsymbol{D}$ is diagonal means $\boldsymbol{D} = \begin{pmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & d_n \end{pmatrix}$.
2. Compute the diagonal matrix of square roots: $\boldsymbol{D}^{1/2}$. (Just take the square root of each diagonal element.)
3. Compute $\boldsymbol{Q} = \Sigma^{1/2} = \boldsymbol{U}\boldsymbol{D}^{1/2}\boldsymbol{U}^\top$.

Random Groups

Aubrey Williams: 4
Austin Barton : 3
Blake Sigmundstad: 6
Diego Moylan: 1
Dillon Shaffer: 7
Ismoiljon Muzaffarov: 3
Jacob Tanner: 4

Josh Stoneback: 6
Joshua Bowen: 8
Joshua Culwell: 2
Laura Banaszewski: 3
Lina Hammel: 2
Logan Racz: 1

Matt Hall: 2
Micah Miller: 5
Mike Kadoshnikov: 1
Owen Cool: 7
Racquel Bowen: 4
Samuel Mocabee: 5
Tatiana Kirillova: 8

Group exercises - Problem Set 7

1. (Sec 4.6) **Conditional distributions.** Suppose X, Y has the joint density

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

   (a) Find the *marginal density* $f_Y(y)$. (b) Compute the *conditional density* $f_{X|Y}(x \mid y)$.

2. (4.7.25a) **Stein's Identity**

   a. Justify Stein's identity, which can be expressed with some abuse of notation as

$$\mathbb{E}[g(X)\nabla \log f(X)] = -\mathbb{E}[\nabla g(X)],$$

   where here we assume sufficiently "nice" density f and test function g. (Hint: Use Integration By Parts.)

   b. Show that in the special case where $X \sim N(\mu, \sigma^2)$, Stein's identity reduces to

$$\mathbb{E}[g(X)(X-\mu)] = \sigma^2 E[g'(X)]$$

3. (4.9.2) **Standard Multivariate Normals: Whitening/Decorrelation.** If $\boldsymbol{X}$ is a random vector with the $N(\boldsymbol{\mu}, \boldsymbol{V})$ distribution where $\boldsymbol{V}$ is non-singular, show that $\boldsymbol{Y} = (\boldsymbol{X} - \boldsymbol{\mu})\boldsymbol{V}^{-\frac{1}{2}}$ is has the $N(\boldsymbol{0}, \boldsymbol{I})$ distribution, where $\boldsymbol{I}$ is the identity matrix. The random vector $\boldsymbol{Y}$ is said to have the *standard* multivariate normal distribution.