ECE2720: Data Science for Engineers

Instructor: Jayadev Acharya, Cornell

Notes by Joyce Shen, FA24

1 Numerical Optimization

2 Gaussian Distribution

2.1 Probability Intro

Want to to model uncertainy in outcomes + situations where outcomes are not known prior.

Discrete probability:- $\mathcal{X}$:= set of outcomes, where p is the probability of the outcome happening

- $p(x) \geq 0$ for all $x \in \mathcal{X}$
- $\sum_{x \in \mathcal{X}} p(x) = 1$ – the **Law of Total Probability**

Random variable X has outcomes $\mathcal{X} \sim p$ when $\mathcal{X}$ is the (apriori unknown) outcome drawn from p. An **event** is any subset of $\mathcal{X}$ (all possible outcomes), usually denoted by A or E.

The **expectation** $\mathbb{E}$ of a random variable X is its expected value. Note that

$$\mathbb{E}[x] = \sum_{x \in \mathcal{X}} \mathbb{P}(X = x) \times x = \sum_{x \in \mathcal{X}} p(x) \times x$$

Additionally,

- For a function $g(x)$, $\mathbb{E}[g(x)] = \sum_{x \in \mathcal{X}} p(x) \times g(x)$.
- **Linearity of Expectation** states that for any random variables X_1 and X_2, $\mathbb{E}[X_1 + X_2] = \mathbb{E}[X_1] + \mathbb{E}[X_2]$

Two random variables are **independent** if

$$\mathbb{P}(X_1 = x_1, X_2 = x_2) = \mathbb{P}(X_1 = x_1) \times \mathbb{P}(X_2 = x_2)$$

2.2 Discrete + Continuous Distributions

Discrete Distributions

$\mathcal{X}$ - a discrete set
p:-

- $p(x) \geq 0$ for all $x \in \mathbb{R}$
- $\sum_{x \in \mathcal{X}} p(x) d(x) = 1$

$$\mathbb{E}[X] = \sum_{x \in \mathcal{X}} x \times p(x)$$

For an event $A \subseteq \mathcal{X}$,

$$\mathbb{P}(X \in A) = \mathbb{P}(A) = \sum_{x \in A} p(x)$$

Continuous Distributions

$\mathcal{X} \sim$ a continuous set of real numbers
p:-

- $p(x) \geq 0$ for all $x \in \mathbb{R}$
- $\int_{-\infty}^{\infty} p(x) \times x = 1$

$$\mathbb{E}[x] = \int_{\mathbb{R}} p(x) \times x dx$$

For an event $A \subseteq \mathcal{X}$,

$$\mathbb{P}(X \in A) = \mathbb{P}(A) = \int_{x \in A} p(x) dx$$

2.3 Gaussian Distributions

This is the most important distribution, also known as the **bell curve** or **normal distribution**.

Definition 2.1. A **Gaussian distribution**with mean μ and variance $\sigma^2 > 0$ has the following distribution:

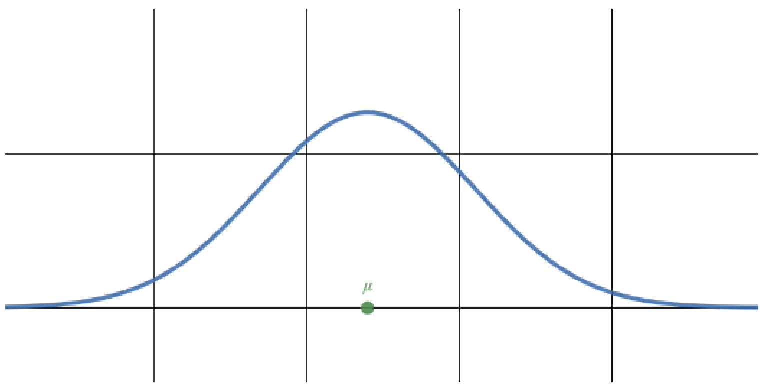
$$p(x) = \phi(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{(x-\mu)^2}{2\sigma^2})$$

From MATH2411 – The **normal distribution**, also called the **Gaussian distribution**, is the most important continuous probability distribution because of its ubiquity in the real world.

Definition 2.2 (Normal Distribution)**.** A **normal distribution** has notation: $X \sim N(\mu, \sigma^2)$ where $\mu \in (-\infty, \infty)$ is the *mean* and $\sigma^2 \in (0, \infty)$ is the **variance**. A normal distribution has pdf:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} \text{ for all real values x}$$

A normal distribution has expectation $E(X) = \mu$ and variance $Var(X) = \sigma^2$. Note that all normal distributions have a bell-shaped density curve regardless of the values of μ and σ.



Observe that $\phi_{\mu,\sigma^2}(x)$ is symmetric about μ and that $X \sim \phi_{\mu,\sigma^2} \rightarrow \text{N}(\mu, \sigma^2)$

Definition 2.3. The **central limit theorem (CLT)** states that a random quantity that is the sum of many small independent random quantities will have follow a Gaussian distribution.

By definition of **variance**: $\mathbb{E}[(x-\mu)^2]$, a normal distribution has

$$\mathbb{E}[x] = \mu, \text{Var}(x) = \sigma^2$$

Note that with expectation and variance, given an $X \sim p(x)$, $Y = aX + b$,

$$\mathbb{E}[Y] = a\mathbb{E}[X] + b, \text{Var}(Y) = a^2\text{Var}(X)$$

2.4 Linear Models

A large part of data science is reasoning about input/output relationships in a system. We're going to look at input which follows a Gaussian distribution.

Theorem 2.1. *With $X \sim N(\mu, \sigma^2)$ and $Y = aX + b$, then*

$$Y \sim N(a\mu + b, a^2\sigma^2)$$

Note that we can standardize Gaussian distributions: From MATH2411-

Definition 2.4 (Standard Normal Distribution)**.** The **standard normal distribution** has mean 0 and variance 1, i.e., $N(0, 1)$. The random variable following the standard normal distribution is often denoted by Z in probability and statistics. Its probability density function (pdf) is

$$\varphi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$

and its distribution function (cdf) is

$$\Phi(z) = \int_{-\infty}^{z} f(t)dt = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt = \mathbb{P}(X \leq z)$$

Note that $\lim_{x\to\infty} \Phi_{\mu,\sigma^2}(x) = 1$ and recall that Φ_{μ,σ^2} is an increasing function (cdf!).

To **standardize** a distribution, we use

$$\frac{x - \mu}{\sigma} \sim N(0, 1)$$

2.5 Random Vectors

Tool to analyze how different random quantities are related (e.g., current and voltage, image darkness and malignancy, etc).

Definition 2.5. A **random vector** is a placeholder for an unrevealed vector-valued random quantity: either thought of as a *single* random entity or a *vector* of random variables.

For any p over $\mathbb{R} \times \mathbb{R}$,

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(x_1, x_2) dx_1 dx_2 = 1$$

2.5.1 Bivariate Gaussian

Definition 2.6. The bivariate $\mathcal{N}$ distribution with paraneters $\mu_1 \in \mathbb{R}, \mu_2 \in \mathbb{R}, \sigma_1^2 > 0, \sigma_2^2 > 0, -1 < \rho < 1$ is

$$\phi_{\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho}(\overrightarrow{x}) = \frac{\exp[-\frac{1}{2(1-\rho^2)}((\frac{x_1-\mu_1}{\sigma_1})^2 - 2\rho(\frac{x_1-\mu_1}{\sigma_1})(\frac{x_2-\mu_2}{\sigma_2}) + (\frac{x_2-\mu_2}{\sigma_2})^2)]}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}$$

ρ is the parameter that characterizes how X_1 and X_2 relate to each other. If

- $\rho = 0$, then X_1 and X_2 are independent (Gaussian)
- $\rho > 0$, then X_1 and X_2 are *positively* correlated
- $\rho < 0$, then X_1 and X_2 are *negatively* correlated

If $A \subseteq \mathbb{R} \times \mathbb{R}$, then

$$\mathbb{P}((x_1, x_2) \in A) = \iint_{(x_2,x_2)\in A} \phi_{\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho}(x_1, x_2) dx_1 dx_2$$

2.5.2 Marginalization

X_1 over $\mathcal{X}_1$, X_2 over $\mathcal{X}_2$, where $p(x_1, x_2)$ = joint distirbution of X_1, X_2.