CSC 477 Tutorial
Probability Refresher

Radian Gondokaryono
October 5, 2022

Slides adapted from:
and CSC477 (toronto.edu) Week 7
Outline

• Probability Overview
• Bayes Rule
• Expectation and Variance
• Gaussian Distributions
• Covariance
Motivation

Uncertainty arises through:

• Noisy measurements
• Variability between samples
• Finite size of data sets

Probability provides a consistent framework for the quantification and manipulation of uncertainty.
Sample Space

**Sample space** $\Omega$ is the set of all possible outcomes of an experiment.

**Observations** $\omega \in \Omega$ are points in the space also called sample outcomes, realizations, or elements.

**Events** $E \subset \Omega$ are subsets of the sample space.

In this experiment we flip a coin twice:

- Sample space All outcomes $\Omega = \{HH, HT, TH, TT\}$
- Observation $\omega = HT$ valid sample since $\omega \in \Omega$
- Event Both flips same $E = \{HH, TT\}$ valid event since $E \subset \Omega$
Probability

The probability of an event $E$, $P(E)$, satisfies three axioms:

1: $P(E) \geq 0$ for every $E$

2: $P(\Omega) = 1$

3: If $E_1, E_2, \ldots$ are disjoint then

$$P\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} P(E_i)$$
Joint and Conditional Probabilities

Joint Probability of A and B is denoted $P(A, B)$.
Conditional Probability of A given B is denoted $P(A \mid B)$.

$p(A, B) = p(A \mid B)p(B) = p(B \mid A)p(A)$
Conditional Example

Probability of passing the midterm is 60% and probability of passing both the final and the midterm is 45%. What is the probability of passing the final given the student passed the midterm?

\[ P(F|M) = \frac{P(M,F)}{P(M)} \]
\[ = \frac{0.45}{0.60} \]
\[ = 0.75 \]
Independence

Events A and B are independent if $P(A, B) = P(A)P(B)$.

* Independent: A: first toss is HEAD; B: second toss is HEAD;
  
  $P(A, B) = 0.5 \times 0.5 = P(A)P(B)$

• Not Independent: A: first toss is HEAD; B: first toss is HEAD;
  
  $P(A, B) = 0.5 \neq P(A)P(B)$
Marginalization and Law of Total Probability

Law of Total Probability

\[ P(X) = \sum_{Y} P(X, Y) = \sum_{Y} P(X|Y)P(Y) \]
Bayes’ Rule

\[ P(A|B) = \frac{P(B|A)P(A)}{P(B)} \]

\[ P(\theta|x) = \frac{P(x|\theta)P(\theta)}{P(x)} \]

Posterior = \frac{\text{Likelihood} \times \text{Prior}}{\text{Evidence}}

Posterior \propto \text{Likelihood} \times \text{Prior}
**Mean:** First Moment, $\mu$

\[
E[X] = \sum_{i=1}^{\infty} x_i p(x_i) \quad \text{(univariate discrete r.v.)}
\]

\[
E[X] = \int_{-\infty}^{\infty} xp(x)dx \quad \text{(univariate continuous r.v.)}
\]

**Variance:** Second (central) Moment, $\sigma^2$

\[
\text{Var}(X) = E[(X - E[X])^2]
\]

\[
= E[X^2 - 2XE[X] + E[X]^2]
\]

\[
\]

\[
= E[X^2] - E[X]^2
\]
Univariate Gaussian Distribution

Also known as the Normal Distribution, $N(\mu, \sigma^2)$

$$N(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left( -\frac{(x-\mu)^2}{2\sigma^2} \right)$$
Multivariate Gaussian Distribution

Multidimensional generalization of the Gaussian.

\( \mathbf{x} \) is a D-dimensional vector

\( \mathbf{\mu} \) is a D-dimensional mean vector

\( \Sigma \) is a \( D \times D \) covariance mean matrix with determinant \(|\Sigma|\)

\[
\mathcal{N}(\mathbf{x}|\mathbf{\mu}, \Sigma) = \frac{1}{(2\pi)^{D/2} |\Sigma|^{1/2}} \exp \left( -\frac{1}{2} (\mathbf{x} - \mathbf{\mu})^T \Sigma^{-1} (\mathbf{x} - \mathbf{\mu}) \right)
\]

\[
= -\frac{1}{2} \begin{bmatrix} 3 \\ 2 \end{bmatrix} - \begin{bmatrix} 1 \\ 2 \end{bmatrix}^T \begin{bmatrix} 5 & 2 \\ 2 & 4 \end{bmatrix} \begin{bmatrix} 2 \\ 3 \end{bmatrix} - \begin{bmatrix} 1 \\ 2 \end{bmatrix} \\
= -\frac{1}{2} \begin{bmatrix} 1 \\ 2 \end{bmatrix}^T \begin{bmatrix} 5 & 2 \\ 2 & 4 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix}
\]

Shortcut notation: \( \|\mathbf{x}\|_{\Sigma}^2 = \mathbf{x}^T \Sigma^{-1} \mathbf{x} \)
Multivariate Gaussian Distribution

\[
\mathcal{N}(x|\mu, \Sigma) = \frac{1}{(2\pi)^{D/2} |\Sigma|^{1/2}} \exp \left( -\frac{1}{2} (x - \mu)^T \Sigma^{-1} (x - \mu) \right)
\]

From “Computer Vision: Models, Learning, and Inference” Simon Prince
Covariance

- Measures linear dependence between random variables $X$, $Y$. Does not measure independence.


- Variance of $X$

\[ \text{Var}[X] = \text{Cov}[X] = \text{Cov}[X, X] = E[X^2] - E[X]^2 \]

\[ \text{Cov}[AX + b] = ACov[X]A^T \]

\[ \text{Cov}[X + Y] = \text{Cov}[X] + \text{Cov}[Y] - 2\text{Cov}[X, Y] \]
Covariance Matrix

\[
\text{Var}[X] = \sigma_x^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2
\]

\[
\text{Cov}[X,Y] = \sigma(x,y) = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})
\]

\[
\begin{bmatrix}
\text{Var}[X] & \text{Cov}[X,Y] \\
\text{Cov}[Y,X] & \text{Var}[Y]
\end{bmatrix} = \Sigma = 
\begin{pmatrix}
\sigma(x,x) & \sigma(x,y) \\
\sigma(y,x) & \sigma(y,y)
\end{pmatrix}
\]
Covariance Matrix

- Measures linear dependence between random variables $X, Y$. Does **not** measure independence.

\[
\]

- Entry $(i,j)$ of the covariance matrix measures whether changes in variable $X_i$ co-occur with changes in variable $Y_j$.

- It does not measure whether one causes the other.