430.457

Introduction to Intelligent Systems

Prof. Songhwai Oh

ECE, SNU

Chapter 12. Quantifying Uncertainty

PROBABILITY

“Very” Simple Example

Trial: Flip two identical fair coins sequentially

- Possible outcomes:
  - {HH, HT, TH, TT}
- What is the probability that the two coins face the same direction?
  - This event is {HH,TT}
  - P({HH,TT}) = ?
- What if the set of outcomes is not discrete? Can we generalize the concept?

Probability Space

A **probability space** is a triple (Ω, $\mathcal{F}$, P)

- Ω is a set of outcomes
- $\mathcal{F}$ is a set of events
- P: $\mathcal{F} \rightarrow [0,1]$ is a function that assigns probabilities to events

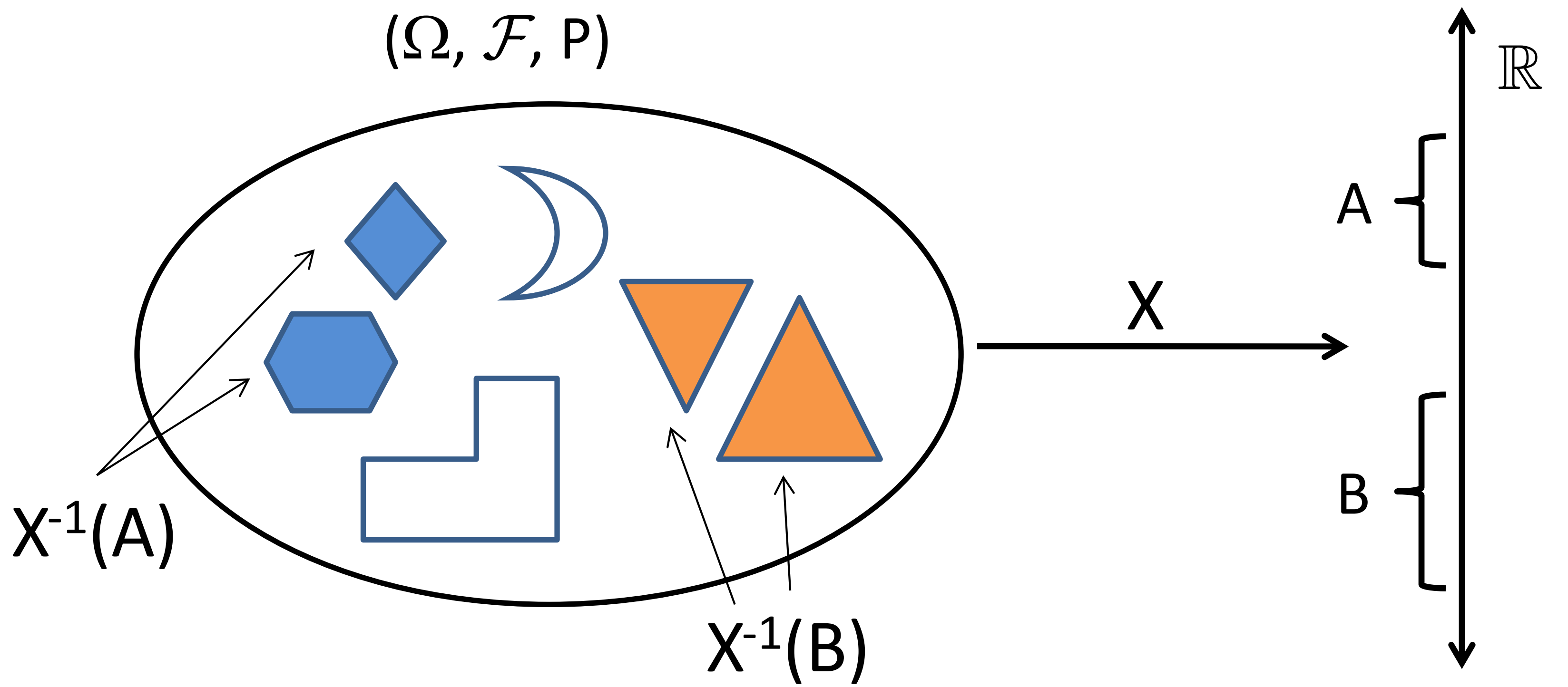
Note: $\mathcal{F}$ is a σ-field, i.e., collection of subsets of Ω such that

- If $A \in \mathcal{F}$ then $A^c \in \mathcal{F}$
- If $A_i \in \mathcal{F}$ is a countable sequence of sets then $\cup_i A_i \in \mathcal{F}$

Random Variables (r.v.)

- $X: \Omega \to \mathbb{R}$ is a **random variable** if for every measurable set $B \subset \mathbb{R}$,

$$X^{-1}(B) = \{\omega : X(\omega) \in B\} \in \mathcal{F} .$$



Probability Axioms

1. Nonnegativity

   $P(A) \geq 0 \qquad\qquad \forall A \in \mathcal{F}$

2. Normalization

   $P(\Omega) = 1$

3. Additivity

   $A_1, A_2, \ldots$ disjoint events

   $P(\cup_i A_i) = \sum_i P(A_i)$

Conditional Probability

Conditional probability

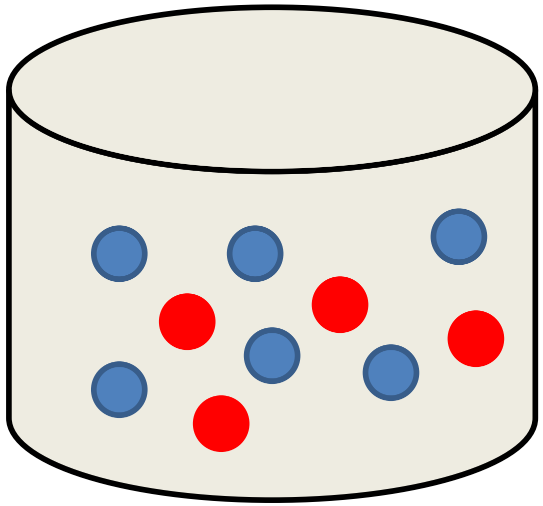
$$P(B|A) = \frac{P(B \cap A)}{P(A)}$$

Multiplication rule

$$P(A \cap B) = P(A)\, P(B|A)$$

Bayes rule

$$P(B|A) = \frac{P(B \cap A)}{P(A)} = \frac{P(A|B)P(B)}{P(A)}$$



Urn

<u>Example</u>

A: blue on first draw

B: blue on second draw

P(A,B) = ?

P(A)=

P(B|A)=

P(A,B)=

Independence

Events A and B are **independent** if $P(B|A) = P(B)$,
i.e., $P(A \cap B) = P(A)\, P(B)$.

<u>Example</u>

Roll 2 dice

A = 1st die is 4
B_1 = 2nd die is 2
B_2 = sum of 2 dice is 9
B_3 = sum of 2 dice is 7

Q1: Are A and B_1 independent?
Q2: Are A and B_2 independent?
Q3: Are A and B_3 independent?

Some Discrete Random Variables

- **Bernoulli distribution**: $X \sim Bernoulli(p)$

$$\begin{aligned} P(X=1) &= p \\ P(X=0) &= 1-p \end{aligned}$$

- **Binomial distribution**: $X \sim Binomial(n,p)$

$$P(X=k) = \binom{n}{k} p^k (1-p)^{n-k} \qquad \text{for } k = 0, \ldots, n$$

- **Geometric distribution**: $X \sim Geometric(p)$

$$P(X=n) = (1-p)^{n-1} p \qquad \text{for } n = 1, 2, \ldots$$

- **Poisson distribution**: $X \sim Poisson(\lambda)$, where $\lambda > 0$

$$P(X=k) = e^{-\lambda} \frac{\lambda^k}{k!} \qquad \text{for } k = 0, 1, 2, \ldots$$

Density, Distribution

Density function (f)

$$\forall x \quad f(x) \geq 0 \qquad \int_{-\infty}^{\infty} f(x)dx = 1 \qquad P(a \leq X \leq b) = \int_a^b f(x)dx$$

Distribution function (F)

$$\begin{aligned} F(x) &= P(X \leq x) \\ &= P(-\infty \leq X \leq x) \\ &= \int_{-\infty}^{x} f(x)dx \end{aligned}$$

$$P(a \leq X \leq b) = F(b) - F(a)$$

Joint Distribution, Marginal Distribution

- Two random variables X and Y.
- **Joint** distribution:

$$P(X, Y)$$

- **Marginal** distribution:

$$P(X = x) = \sum_y P(X = x, Y = y)$$

- If X and Y are independent,

$$\begin{aligned} P(X + Y = z) &= \sum_x P(X = x, Y = z - x) \\ &= \sum_x P(X = x) P(Y = z - x) \end{aligned}$$

Expected Value

X discrete r.v.

$$\mathbb{E}h(X) = \sum_x h(x)P(X = x)$$

$h(x) = x$	$\mathbb{E}X$	expected value of X
$h(x) = x^k$	$\mathbb{E}X^k$	k-th moment of X
$h(X) = (X - \mathbb{E}(X))^2$	$\mathbf{var}(X)$	variance of X

Standard deviation: $\sigma(X) = \sqrt{\mathbf{var}(X)}$.

Poisson Distribution

For $k = 0, 1, 2, \ldots,$

$$P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}$$

$$\begin{aligned}
\mathbb{E}(X) &= \textstyle\sum_{k=0}^{\infty} kP(X = k) \\
&= \sum_{k=0}^{\infty} ke^{-\lambda} \frac{\lambda^k}{k!} = \lambda \sum_{k=1}^{\infty} e^{-\lambda} \frac{\lambda^{k-1}}{(k-1)!} = \lambda \sum_{n=0}^{\infty} e^{-\lambda} \frac{\lambda^n}{n!} \\
&= \lambda \left(\sum_{n=0}^{\infty} e^{-\lambda} \frac{\lambda^n}{n!} \right) \\
&= \lambda
\end{aligned}$$

Practice: $\mathbf{var}(X) = \lambda$

Some Facts about Expectation

- $X_1, X_2, \ldots, X_n$, random variables
- Linearity of expectation

$$\mathbb{E}(X_1 + \cdots + X_n) = \mathbb{E}X_1 + \cdots + \mathbb{E}X_n$$

- If $X_1, X_2, \ldots, X_n$ are independent,

$$\begin{aligned} \mathbb{E}(X_1 \cdots X_n) &= \mathbb{E}X_1 \cdots \mathbb{E}X_n \\ \mathbf{var}(X_1 + \cdots + X_n) &= \mathbf{var}(X_1) + \cdots + \mathbf{var}(X_n) \end{aligned}$$

- For $c \in \mathbb{R}$,

$$\begin{aligned} \mathbb{E}(X + c) &= \mathbb{E}(X) + c \\ \mathbf{var}(X + c) &= \mathbf{var}(X) \\ \mathbb{E}(cX) &= c\mathbb{E}(X) \\ \mathbf{var}(cX) &= c^2\mathbf{var}(X) \end{aligned}$$

Moment-Generating Function

$$
\begin{array}{ll}
\varphi_X(t) = \mathbb{E}(e^{tX}) & \\
\varphi'_X(t) = \mathbb{E}(Xe^{tX}) & \varphi'_X(0) = \mathbb{E}(X) \\
\varphi''_X(t) = \mathbb{E}(X^2e^{tX}) & \varphi''_X(0) = \mathbb{E}(X^2) \\
\vdots & \vdots \\
\varphi^{(k)}_X(t) = \mathbb{E}(X^ke^{tX}) & \varphi^{(k)}_X(0) = \mathbb{E}(X^k)
\end{array}
$$

Theorem: X and Y are independent. Then $\varphi_{X+Y}(t) = \varphi_X(t)\varphi_Y(t)$.

$$
\begin{aligned}
\varphi_{X+Y}(t) &= \mathbb{E}\left(e^{t(X+Y)}\right) \\
&= \mathbb{E}\left(e^{tX}e^{tY}\right) \\
&= \mathbb{E}\left(e^{tX}\right)\mathbb{E}\left(e^{tY}\right) \\
&= \varphi_X(t)\varphi_Y(t)
\end{aligned}
$$

Important property of MGF: it determines the distribution of the r.v.

Laws of Large Numbers

Strong law of large numbers:

- $X_1, X_2, \ldots$, i.i.d. (independently and identically distributed) with $E(X_i) = \mu$ and integrable.
- $S_n = X_1 + \cdots + X_n$
- Then w.p. (with probability) one

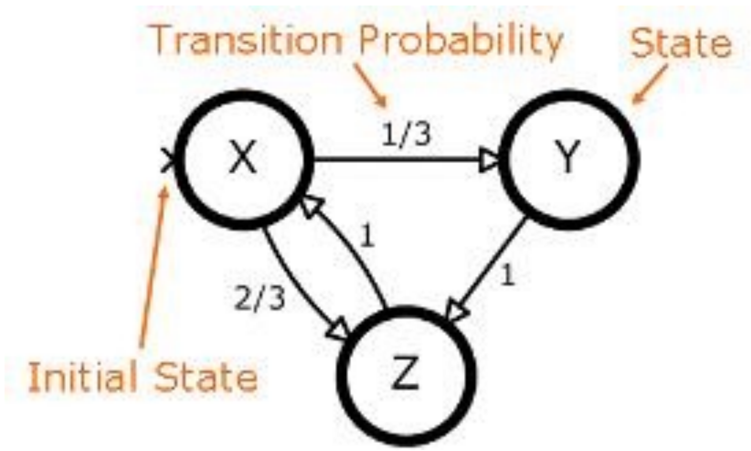
$$\frac{S_n}{n} \to \mu \text{ as } n \to \infty.$$

Central limit theorem:

- Sum of a sufficiently large number of independent random variables, each with finite mean and variance, will be approximately **normally** distributed.

Markov Chains

- An important tool in probability



- Markov property
- Classification of states
- Limit behavior, convergence, stationary distribution
- Applications: Hidden Markov model, Markov random field, Markov chain Monte Carlo, …