Section 2: Probability Review

1 Notation and Basics

Events An event A is a subset of the sample space Ω (the set of all possible outcomes). We assign probabilities $P(A) \in [0, 1]$ to events.

Random Variables A random variable $X : \Omega \to \mathbb{R}$ is a function that assigns a number to each outcome.

- Discrete X: takes values in a countable set $\mathcal{X}$
- Continuous X: takes values in $\mathbb{R}^d$

Example (dice): Sample space $\Omega = \{1, 2, 3, 4, 5, 6\}$.
Random variable $X(\omega) = \begin{cases} 1 & \text{if } \omega \text{ is even} \\ 0 & \text{otherwise} \end{cases}$ maps outcomes to $\{0, 1\}$.

PMF / PDF / CDF

PMF (Probability Mass Function)	$p_X(x) = \Pr(X = x)$	[discrete]
PDF (Probability Density Function)	$p_X(x)$ s.t. $\Pr(X \in A) = \int_A p_X(x)\,dx$	[continuous]
CDF (Cumulative Distribution Function)	$F_X(x) = \Pr(X \le x)$	[both]

Normalization: $\sum_x p(x) = 1$ (discrete) or $\int_{-\infty}^{\infty} p(x)\,dx = 1$ (continuous)

Notation: When clear from context, we write $p(x)$ instead of $p_X(x)$.

Joint, Marginal, and Conditional Probabilities

Joint: $p(x, y)$

Marginal: $p(x) = \sum_y p(x, y)$ or $p(x) = \int p(x, y)\,dy$

Conditional: $p(x \mid y) = \dfrac{p(x, y)}{p(y)}$ (when $p(y) > 0$)

Independence vs Conditional Independence

Independence: $X \perp Y$

$$p(x, y) = p(x)\,p(y) \iff p(x \mid y) = p(x)$$

Conditional independence: $X \perp Y \mid Z$

$$p(x, y \mid z) = p(x \mid z)\,p(y \mid z) \iff p(x \mid y, z) = p(x \mid z)$$

RL: "Future independent of past given present state (and action)" is conditional independence → Markov property.

2 Distributions

Bernoulli and Categorical

Bernoulli: $X \in \{0, 1\}$, parameter p

$$\Pr(X = 1) = p, \qquad \Pr(X = 0) = 1 - p$$

Categorical: $A \in \{1, \ldots, K\}$, probabilities $\pi_1, \ldots, \pi_K$ with $\sum_k \pi_k = 1$

RL: Discrete-action policies $\pi_\theta(a \mid s)$ are typically categorical distributions.

Gaussian (Normal)

Multivariate Gaussian: $a \sim \mathcal{N}(\mu, \Sigma)$

- μ = mean vector
- Σ = covariance matrix

Diagonal covariance (common in RL):

$$\Sigma = \mathrm{diag}(\sigma_1^2, \ldots, \sigma_d^2)$$

Independent noise per action dimension. Policy outputs $\mu(s)$ and either:

- state-independent $\log \sigma$ (one learnable vector), or
- state-dependent $\log \sigma(s)$

RL: Continuous-action policies are typically Gaussian: $\pi_\theta(a \mid s) = \mathcal{N}(\mu_\theta(s), \Sigma)$.

3 Expectation and Variance

Definition of Expectation

Discrete: $\mathbb{E}[X] = \sum_x x\, p(x)$

Continuous: $\mathbb{E}[X] = \int_{-\infty}^{\infty} x\, p(x)\, dx$

Function of X: $\mathbb{E}[f(X)] = \sum_x f(x)\, p(x)$ or $\int_{-\infty}^{\infty} f(x)\, p(x)\, dx$

Linearity of Expectation

For *any* random variables X, Y and constants a, b:

$$\mathbb{E}[aX + bY] = a\,\mathbb{E}[X] + b\,\mathbb{E}[Y]$$

Variance

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

RL: High variance of Monte Carlo estimators is a core issue in RL → motivates baselines and advantage functions.

4 Conditioning and Bayes

Conditional Probability

$$\Pr(A \mid B) = \frac{\Pr(A \cap B)}{\Pr(B)} \qquad (\text{assuming } \Pr(B) > 0)$$

Chain Rule (rearranging the conditional definition):

$$p(x, y) = p(x)\,p(y \mid x) = p(y)\,p(x \mid y)$$

General form:

$$p(x_1, \ldots, x_n) = \prod_{i=1}^{n} p(x_i \mid x_{1:i-1})$$

RL: Trajectory probability is a direct application: $p(s_1, a_1, s_2, a_2, \ldots) = p(s_1) \prod_t \pi(a_t|s_t)\, p(s_{t+1}|s_t, a_t)$

Law of Total Probability (marginalization + chain rule):

$$p(y) = \sum_x p(x, y) = \sum_x p(x)\,p(y \mid x) = \mathbb{E}_{x\sim p(x)}[p(y \mid x)]$$

RL: Used in Bellman equations: $V(s) = \mathbb{E}_{a\sim\pi}[Q(s, a)] = \sum_a \pi(a|s)Q(s, a)$

Conditional Expectation

$\mathbb{E}[X \mid Y]$ = "expected value of X after observing Y"

Tower property (law of total expectation):

$$\mathbb{E}[X] = \mathbb{E}\big[\mathbb{E}[X \mid Y]\big]$$

Bayes Rule

$$p(x \mid y) = \frac{p(y \mid x)\, p(x)}{p(y)}$$

5 Markov Property and MDPs

Markov Property

In an MDP, the transition dynamics satisfy the Markov property: the next state depends only on the current state and action, not the full history.

$$p(s_{t+1} \mid s_{1:t}, a_{1:t}) = p(s_{t+1} \mid s_t, a_t)$$

Equivalently (conditional independence):

$$s_{t+1} \perp (s_{1:t-1}, a_{1:t-1}) \,\big|\, (s_t, a_t)$$

Markov Chain vs MDP

	Markov Chain	**MDP**
Actions	None	$a_t \sim \pi(a_t \mid s_t)$
Transition	$p(s_{t+1} \mid s_t)$	$p(s_{t+1} \mid s_t, a_t)$