ECE368 Cheatsheet

Hanhee Lee

January 10, 2025

Contents

Definition:

Process:

Motivation:

Derivation:

Warning:

Summary:

Algorithm:

Example:

FAQ:

W1 (LG-IPPR 1.1, 1.2; Murphy 2.1 – 2.3)

1 L1: Probability Review

Summary:

FAQ:
- How to study? Practice, practice.
- What textbooks? Use 2024 version of Murphy, Leon Garcia as main reference, Bishop, 4th textbook is intro.
- How is HW graded? Effort, and tutorials are used to explain soln.

1.1 Sample Space

Motivation: If you have 4 sheeps and a flea, the probability that starting from sheep 1, the flea will jump to sheep 4 in 10 steps is 0.2.
- Ambigious as there are 2 different interpretations for the sample space (i.e. space of probability is not clear):
  - Set of sheeps
  - Set of number of steps

1.2 Probability Definitions

Definition:
- **Random Experiment:** An outcome (realization) for each run.
- **Sample Space Ω:** Set of all possible outcomes.
- **Events:** (measurable) subsets of Ω.
- **Probability of Event A:** $P[A] \equiv P[\text{'outcome is in A'}]$.

Example: Roll Fair Die
- $\Omega = \{1, 2, 3, 4, 5, 6\}$.
- $P[\text{'even number'}] = \frac{1}{2}$.

1.3 Axioms of Probability

Definition:
1. $P[A] \geq 0$ for all $A \in \Omega$.
2. $P[\Omega] = 1$.
3. If $A \cap B = \emptyset$, then $P[A \cup B] = P[A] + P[B]$ for all $A, B \in \Omega$.

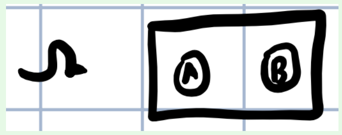

Figure 1: 3rd Axiom

1.4 Conditional Probability

Definition:

$$P[A|B] = \frac{P[A \cap B]}{P[B]} \tag{1}$$

- $|$: Given event (data/obs.).

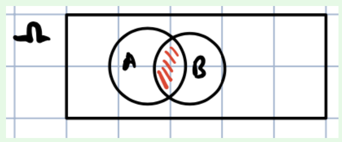

Figure 2: Conditional Probability

Notes:
- Changing sample space to B.
- Conditional probability satisfy the 3 axioms (i.e. are probabilities), can be viewed as probability measure on new sample space B.

1.4.1 Consequences of Conditional Probability

Definition:

$$P[A \cap B] = P[A|B]P[B] = P[B|A]P[A] \tag{2}$$

1.4.2 Independence

Definition: A and B are independent iff

$$P[A \cap B] = P[A]P[B] \iff P[A|B] = P[A] \iff P[B|A] = P[B] \tag{3}$$

1.4.3 Importance of Labelling

Example: **Toss 2 Fair Coins**
1. **Given:** Given that one of the coins is heads, what is the probability that the other coin is tails?
2. **Wrong Solution:** $\frac{1}{2}$ since $\{HH, HT, TH, TT\}$, so $P[T|H] = \frac{1}{2}$, which assumes that the coins are distinguishable (i.e. coin #1 is heads)
3. **Correct Solution:** $\frac{2}{3}$ since $\{HH, HT, TH\}$ as we didn't specify which coin was heads, so $P[T|H] = \frac{2}{3}$, which assumes that the coins are indistinguishable.

2 L2: Probability Review

2.1 Total Probability

Definition: If $H_1, \ldots, H_n$ form a partition of Ω, then

$$P[A] = \sum_{i=1}^{n} P[A|H_i]P[H_i] \tag{4}$$

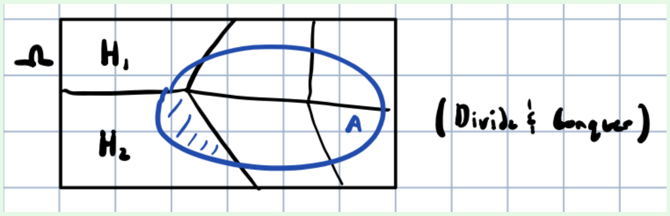

Figure 3: Total Probability

2.2 Bayes' Rule

Definition:

$$P[H_k|A] = \frac{P[H_k \cap A]}{P[A]} = \frac{P[A|H_k]P[H_k]}{\sum_{i=1}^{n} P[A|H_i]P[H_i]} \tag{5}$$

2.2.1 Posteriori Probability, Priori Probability (Prior), Likelihood

Definition:
- **Posteriori:** $P[H_k|A]$.
- **Priori:** $P[H_k]$.
- **Likelihood:** $P[A|H_k]$.

Example: Suppose a lie detector is 95% accurate, i.e. $P[\text{'out=truth'}|\text{'in=truth'}] = 0.95$ and $P[\text{'out=lie'}|\text{'in=lie'}] = 0.95$. It says that Mr. Ernst is lying. What is the probability Mr. Ernst is actually lying.
- **Observation:** $A = \text{'out=lie'}$.
- **Hypothesis:** $H_0 = \text{'in=lie'}$ and $H_1 = \text{'in=truth'}$.
- **Solution:** $P[H_0|A] = \dfrac{P[A|H_0]P[H_0]}{P[A|H_0]P[H_0] + P[A|H_1]P[H_1]} = \dfrac{0.95 \times P[H_0]}{0.95 \times P[H_0] + 0.05 \times (1 - P[H_0])}$.
- $H_0 = 0.01$**:** i.e. 1% of the population are liars, then $P[H_0|A] = \dfrac{0.95 \times 0.01}{0.95 \times 0.01 + 0.05 \times 0.99} = 0.16$.

Warning: Need to know priori probability.

2.2.2 Interpretation of Bayes' Rule

Notes: Taking one component of the total probability and normalizing it by the sum of all components.

2.3 Random Variables

Motivation: Coin Toss Mapping of each outcome to a real number

- $w \in \Omega$ is the outcome of a coin toss, and X is the RV, so $H \to 0$ and $T \to 1$.

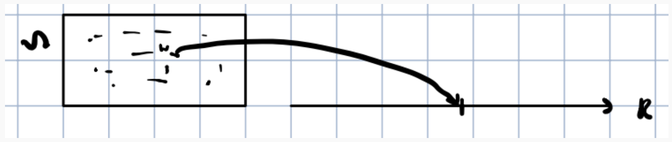

Figure 4: Random Variables

- Mapping is deterministic function. RV is not random or variable.

Definition: Mapping from Ω to $\mathbb{R}$.

2.4 Distribution of RV

2.4.1 Cumulative Distribution Function (CDF) of RV

Definition:

$$F_X(x) \equiv P[X \leq x] \tag{6}$$

2.4.2 Discrete RV Probability Mass Function (PMF)

Definition:

$$P_X(x_j) \equiv P[X = x_j] \quad j = 1, 2, 3, \ldots \tag{7}$$

Example: Binonmial RV w/ (n, p)

$$P_X(k) = \binom{n}{k} p^k (1-p)^{n-k} \tag{8}$$

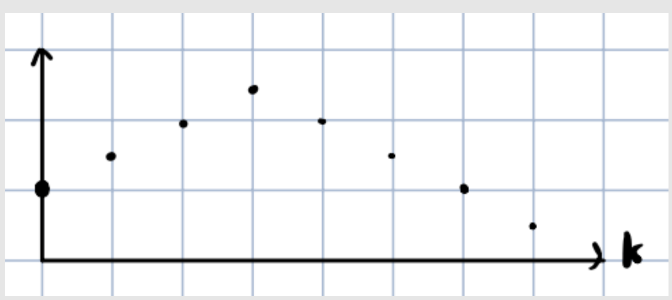

Figure 5: Binomial RV

2.4.3 Continuous RV Probability Density Function (PDF)

Definition:

$$f_X(x) \equiv \frac{d}{dx} F_X(x) \tag{9}$$

$$P[x < X < x + dx] = f_X(x)dx \tag{10}$$

Example: **Gaussian RV w/ (μ, σ^2)**

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{11}$$

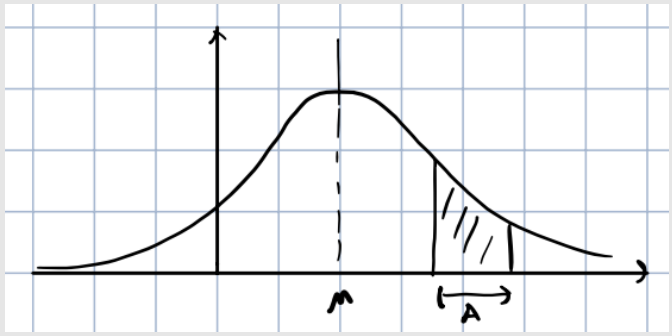

Figure 6: Gaussian RV

- $P[X \in A] = \int_A f_X(x)dx$.

Notes: Discrete RV has pdf w/ δ functions.

2.4.4 Conditional PMF/PDF

Definition:

$$P_X(x|A) \tag{12}$$

$$f_X(x|A) \tag{13}$$

Example: **Continuous**

$$f(x|X>a) = \begin{cases} \frac{f_X(x)}{P[X>a]} & \text{if } x > a \\ 0 & \text{otherwise} \end{cases} \tag{14}$$

Example: **Geometric RV** Geometric RV X w/ success probability p

$$P_X(k) = (1-p)^{k-1}p \tag{15}$$

- **Memoryless Property:** $P_X[k|X>m] = \frac{p(1-p)^{k-1}}{(1-p)^m} = p(1-p)^{k-m-1}$.
  - So it only cares about the additional trials (i.e. same as resetting after m trials).

2.5 Expected Values

Definition:

$$E[X] = \int_{-\infty}^{\infty} x f_X(x)dx \overset{\text{If int. values}}{=} \sum_{k=-\infty}^{\infty} k f_X(k) \tag{16}$$

$$E[h(X)] = \int_{-\infty}^{\infty} h(x) f_X(x)dx \overset{\text{If int. values}}{=} \sum_{k=-\infty}^{\infty} h(k) f_X(k) \tag{17}$$

$$\text{Var}[X] = E[(X-E[X])^2] = E[X^2] - E[X]^2 \tag{18}$$

$$E[X|A] = \int_{-\infty}^{\infty} x f_X(x|A) dx \tag{19}$$

Example: **Lottery Ticket (Geometric RV)**

1. **Given:** Buying one lottery ticket per week
   - Each ticket has $10^{-7} = p$ chance of winning the jackpot.
   - $X =$ '# of weeks to win jackpot'.
2. **Problem:** What is the expected number of weeks to win the jackpot?
3. **Solution:** $E[X] = \sum_{k=1}^{\infty} k(1-p)^{k-1}p = \ldots = \frac{1}{p} = 10^7$ weeks.
4. **Extension (Memoryless Property):** If I have already played for 999999 weeks, what is the expected number of weeks to win the jackpot? $E[X - 999999|X > 999999] = E[X] = 10^7$ weeks.

3 L3: Probability Review

3.1 2 RVs

Notes: RVs are neither random nor a variable.

$$\underline{Z} = (X, Y)$$

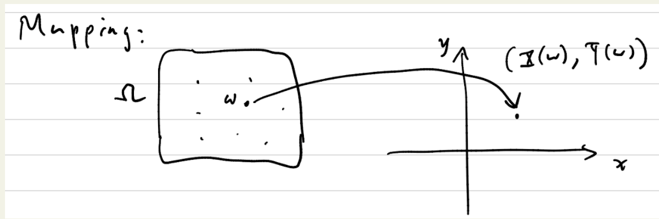

Figure 7: Mapping of RVs

3.2 Joint PMF/PDF

Definition:

$$P_{X,Y}(x, y) = P[X = x, Y = y] \tag{20}$$

$$f_{X,Y}(x, y) = \frac{\partial^2}{\partial x \partial y} F_{X,Y}(x, y) \tag{21}$$

$$P[(X, Y) \in A] = \int\int_{(x,y)\in A} f_{X,Y}(x, y)\, dx\, dy \tag{22}$$

Example: Jointly Gaussian RVs X and Y with $(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$

$$f_{X,Y}(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x-\mu_1}{\sigma_1}\right)\left(\frac{y-\mu_2}{\sigma_2}\right) + \left(\frac{y-\mu_2}{\sigma_2}\right)^2\right]\right\}$$

3.3 Expectations

Definition:

$$E[g(X, Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x, y) f_{X,Y}(x, y)\, dx\, dy$$

Notes:
- $g(X, Y)$ is also an RV, but inside the integral or sum, you use x and y as dummy variables to vary through the values of the RVs.

3.3.1 Correlation

Definition:

$$E[XY] \tag{23}$$

3.3.2 Covariance

Definition:

$$\mathrm{Cov}[X,Y] = E[(X-\mu_X)(Y-\mu_Y)] = E[XY] - \mu_X\mu_Y = E[XY] - E[X]E[Y] \tag{24}$$

Notes:
- Mean shifted to 0.

3.3.3 Correlation Coefficient

Definition:

$$\rho_{X,Y} = E\left[\left(\frac{X-\mu_X}{\sigma_X}\right)\left(\frac{Y-\mu_Y}{\sigma_Y}\right)\right] = \frac{\mathrm{Cov}[X,Y]}{\sigma_X\sigma_Y} \tag{25}$$

- $|\rho_{X,Y}| \leq 1$

Notes:
- Mean shifted to 0 and normalized by the standard deviation.

3.4 Marginal PMF/PDF

Definition:

$$P_X(x) = \sum_{j=1}^{\infty} P_{X,Y}(x, y_j), \quad P_Y(y) = \sum_{j=1}^{\infty} P_{X,Y}(x_j, y) \tag{26}$$

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy, \quad f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx \tag{27}$$

Notes:
- Total probability theorem is being used here.

Example: Jointly Gaussian X and Y:

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy \\ &= \dots \quad \text{(completing the square)} \\ &= \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}}, \quad \text{marginally Gaussian} \end{aligned}$$

- Gaussian RVs has a property that the PDF of a single variable is equal to the marginal Gaussian of two variables.

3.5 Conditional PMF/PDF

Definition:

$$P_{X|Y}(x|y) \triangleq P[X = x | Y = y] = \frac{P_{X,Y}(x,y)}{P_Y(y)} \tag{28}$$

$$f_{X|Y}(x|y) \triangleq \frac{f_{X,Y}(x,y)}{f_Y(y)} \tag{29}$$

3.6 Bayes' Rule

Definition:

$$P_{Y|X}(x|y) = \frac{P_{X,Y}(x,y)}{P_X(x)} = \frac{P_{X|Y}(x|y)P_Y(y)}{\sum_{j=1}^{\infty} P_{X,Y}(x,y_j)P_Y(y_j)} \quad (30)$$

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)} = \frac{f_{X|Y}(x|y)f_Y(y)}{\int_{-\infty}^{\infty} f_{X|Y}(x|y')f_Y(y')\,dy'} \quad (31)$$

3.7 Independent vs. Uncorrelated vs. Orthogonal

Definition:

1. Independent:

$$f_{X|Y}(x|y) = f_X(x)\ \forall y \Leftrightarrow f_{X,Y}(x,y) = f_X(x)f_Y(y) \quad (32)$$

2. Uncorrelated:

$$\text{Cov}[X,Y] = 0 \quad \Leftrightarrow \quad \rho_{X,Y} = 0 \quad (33)$$

3. Orthogonal:

$$E[XY] = 0 \quad (34)$$

Theorem: If independent, then uncorrelated.

Derivation:

$$\begin{aligned}
\text{Independent} \implies E[XY] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy f_{X,Y}(x,y)\,dx\,dy \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy f_X(x)f_Y(y)\,dx\,dy \\
&= \left(\int_{-\infty}^{\infty} x f_X(x)\,dx\right)\left(\int_{-\infty}^{\infty} y f_Y(y)\,dy\right) \\
&\implies E[XY] = E[X]E[Y] \\
&\implies \text{Cov}[X,Y] = 0, \quad \text{uncorrelated} \\
&\not\Leftarrow \text{in general.}
\end{aligned}$$

Example: Jointly Gaussian RVs X and Y: If uncorrelated, i.e. $\rho_{X,Y} = 0$, then X and Y are independent.

$$\begin{aligned}
f_{X,Y}(x,y) &= \frac{1}{2\pi\sigma_1\sigma_2}\exp\left\{-\frac{1}{2}\left[\left(\frac{x-\mu_1}{\sigma_1}\right)^2 + \left(\frac{y-\mu_2}{\sigma_2}\right)^2\right]\right\} \\
&= \frac{1}{\sqrt{2\pi}\sigma_1}e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}} \cdot \frac{1}{\sqrt{2\pi}\sigma_2}e^{-\frac{(y-\mu_2)^2}{2\sigma_2^2}} \\
&= f_X(x)f_Y(y) \quad \text{independent}
\end{aligned}$$

3.8 Conditional Expectation

Definition:

$$E[Y] = E[E[Y|X]] \quad (35)$$

$$E[h(Y)] = E[E[h(Y)|X]] \quad (36)$$

Notes:
- $E[E[Y|X]]$ is w.r.t. X.
- $E[Y|X]$ is w.r.t. Y.

Derivation:

$$\begin{aligned}
E[Y] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y f_{X,Y}(x,y)\,dx\,dy \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y f_{Y|X}(y|x) f_X(x)\,dx\,dy \\
&= \int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty} y f_{Y|X}(y|x)\,dy\right) f_X(x)\,dx \\
&= \int_{-\infty}^{\infty} E[Y|X=x] f_X(x)\,dx \quad \text{(using the total probability theorem)} \\
&= \int_{-\infty}^{\infty} g(x) f_X(x)\,dx \\
&= E[g(X)] \\
&= E[E[Y|X]].
\end{aligned}$$

Example:

1. **Given:** An unknown voltage. $X \sim \text{Uniform}(0,1)$. Measurement from a (bad) voltmeter: $Y \sim \text{Uniform}(0,X)$.

$$f_X(x) = \begin{cases} 1, & 0 < x < 1 \\ 0, & \text{otherwise} \end{cases}$$

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

- **Note:** Area under PDF is 1.

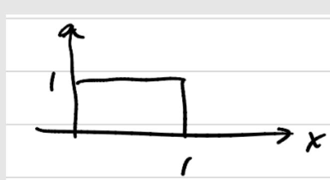

Figure 8: Uniform Distribution of X

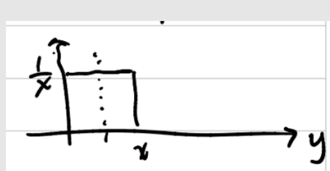

Figure 9: Uniform Distribution of Y

2. **Expected Value (Average Reading of Bad Voltmeter):**

$$\begin{aligned}
E[Y] &= E[E[Y|X]] \\
&= E\left[\frac{X}{2}\right] \quad \text{Since in the middle of 0 and x} \\
&= \frac{1}{2} \cdot E[X] \\
&= \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4} \quad \text{Since } E[X] \text{ (i.e. mean) is 0.5}
\end{aligned}$$

3. **The Long Way:**

$$\begin{aligned}
f_Y(y) &= \int_{-\infty}^{\infty} f_{Y|X}(y|x) f_X(x)\, dx \\
&= \int_{y}^{1} f_{Y|X}(y|x) f_X(x)\, dx \\
&= \int_{y}^{1} \frac{1}{x} \cdot 1\, dx \\
&= -\ln y. \\
E[Y] &= \int_{0}^{1} y \cdot (-\ln y)\, dy = \cdots = \frac{1}{4}
\end{aligned}$$

4. **Question:** Suppose $Y = \frac{1}{8}$. What is "best" given X? This will be the quesiton for the rest of the course.