PH 712 Probability and Statistical Inference

Part I: Random Variable

Zhiyang Zhou (zhou67@uwm.edu, zhiyanggeezhou.github.io)

2025/09/24 13:05:00

Probability (HMC Sec. 1.1–1.3)

- Sample space (denoted by Ω): the set of all the possible outcomes, e.g.,
  - $\Omega = \mathbb{R}^+$ if investigating survival times of cancer patients
  - $\Omega = \{\text{yes, no}\}$ if investigating whether a treatment is effective
- Event (denoted by capital Roman letters, e.g., A): a subset of the sample space, e.g., corresponding to the previous two examples of sample spaces,
  - $A = (0, 10]$: the survival time ≤ 10
  - $B = \{\text{yes}\}$: the treatment is effective
- An event A occurs $\Leftrightarrow$ the outcome belongs to A, e.g.,
  - The survival time is 11: A does happens
  - The treatment outcome is "yes": B happens
- Probability (denoted by Pr): a function quantifying the occurrence likelihood of an event
  - E.g.,
    * $\Pr(A)$: the probability (occurrence likelihood) of event A
    * $\Pr(A^c)$: the probability that event A does NOT occur ($A^c = \Omega \setminus A$ denoting the complement set of A)
    * $\Pr(A \cup B)$: the probability of either A or B
    * $\Pr(A \cap B)$: the probability of both A and B
  - Input: an event
  - Output: a real number (the occurrence probability of the input event)
  - Requirements (definition in math):
    * $\Pr(A) \geq 0$ for any event A
    * $\Pr(\Omega) = 1$ (i.e., the sample space as a special event always occurs)
    * (The probability of the union of mutually exclusive countably events is the sum of the probability of each event) If $\{A_n\}_{n=1}^\infty$ is a sequence of events with $A_{n_1} \bigcap A_{n_2} = \emptyset$ for all $n_1 \neq n_2$, then $\Pr(\bigcup_{n=1}^\infty A_n) = \sum_{n=1}^\infty \Pr(A_n)$
  - More properties (deduced from the above requirements):
    * $\Pr(A) = 1 - \Pr(A^c)$
    * $\Pr(\emptyset) = 0$
    * $\Pr(A) \leq \Pr(B)$ if $A \subset B$
    * $0 \leq \Pr(A) \leq 1$ for each A
    * $\Pr(A \bigcup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B)$ for any events A and B regardless if they are disjoint or not
    * $\Pr(\bigcup_{n=1}^\infty A_n) \leq \sum_{n=1}^\infty \Pr(A_n)$ for arbitrary sequence $\{A_n\}_{n=1}^\infty$

Conditional probability and independence (HMC Sec. 1.4)

- Motivating example

- A: the event that a given person recovers from a disease
- B: the event that a given person has received a certain treatment
- $\Pr(A)$: the probability that a given person recovers from the disease
- $\Pr(A \mid B)$: the probability that a given person recovers from the disease, given that the person has received the treatment
- If $\Pr(A \mid B) = \Pr(A)$, then the treatment is NOT effective for the disease

- Conditional probability of B given A (with $\Pr(A) > 0$): $\Pr(B \mid A) = \Pr(A \cap B)/\Pr(A)$
  - Interpretation: the occurrence probability of B, given that A has already occurred.
  - Properties:
    - $\Pr(B \mid A) \geq 0$
    - $\Pr(A \mid A) = 1$
    - $\Pr(\bigcup_{n=1}^{\infty} B_n \mid A) = \sum_{n=1}^{\infty} \Pr(B_n \mid A)$ if $\{B_n\}_{n=1}^{\infty}$ are mutually exclusive
    - (Law of total probability) $\Pr(B) = \sum_{n=1}^{N} \Pr(A_n) \Pr(B \mid A_n)$ if $\{A_n\}_{n=1}^{N}$ form a partition of Ω (i.e., $\{A_n\}_{n=1}^{N}$ are mutually exclusive and $\Omega = \bigcup_{n=1}^{N} A_n$)
    - (Bayes' theorem) $\Pr(A_i \mid B) = \Pr(A_i)\Pr(B \mid A_i)/\sum_{n=1}^{N} \Pr(A_n)\Pr(B \mid A_n)$ if $\{A_n\}_{n=1}^{N}$ form a decomposition/partition of Ω
- Independence between two events B and A (i.e., $B \perp A$): $\Pr(B \cap A) = \Pr(A)\Pr(B)$
  - $\Leftrightarrow B \perp A^c$
  - $\Leftrightarrow \Pr(B \mid A) = \Pr(B)$ (if $\Pr(A) \neq 0$)
  - $\Leftrightarrow \Pr(A \mid B) = \Pr(A)$ (if $\Pr(B) \neq 0$)
- Mutual independence among N events $A_1, \ldots, A_N$: for arbitrary subset of $\{A_1, \ldots, A_N\}$, say $\{A_{n_1}, \ldots, A_{n_K}\}$ with $2 \leq K \leq N$, $\Pr(\bigcap_{k=1}^{K} A_{n_k}) = \prod_{k=1}^{K} \Pr(A_{n_k})$

HMC Ex. 1.4.31

- A French writer, Chevalier de Méré, had asked a famous mathematician, Pascal, to explain why the following two probabilities were different (the difference had been noted from playing the game many times): (1) at least one six in four independent casts of a six-sided die; (2) at least a pair of sixes in 24 independent casts of a pair of dice. From proportions it seemed to Mr. de Méré that the two probabilities should be the same. Compute the probabilities of (1) and (2).
  - Hint: Pr(no six in one cast of a die) $= 5/6$, Pr(no six in one cast of a pair of dice) $= (5/6)^2$, and Pr(only one six in one cast of a pair of dice) $= 2 \times (1/6) \times (5/6)$.

RV and events

- RV: an encoder (function) mapping entries of sample space to real numbers,
  - Input: an element of sample space
  - Output: a real number
- Example of RVs: Severity of a patient's cold symptoms
  - Sample space $\Omega = \{$no reaction, mild, moderate, severe$\}$
  - RV X: X(no reaction) $= 0$, X(mild) $= 1$, X(moderate) $= 2$, X(severe) $= 3$
- Using values of an RV to define events
  - For the above example, $\{X \leq .7\} = \{\text{no reaction}\}$, $\{X \leq 2.3\} = \{\text{no reaction}, \text{mild}, \text{moderate}\}$
  - What is $\{1.1 \leq X < 2\}$? How about $\{1.1 \leq X < 2.1\}$?

Distribution of an RV (HMC Chp. 1.5–1.7)

- The cumulative distribution function (cdf) of RV X, say F_X, is defined as

$$F_X(t) = \Pr(X \leq t), \quad t \in \mathbb{R}.$$

  - F_X satisfies following three properties:
    - (Right continuous) $\lim_{x \to t^+} F_X(x) = F_X(t)$ (p.s., $\lim_{x \to t^-} F_X(x) = \Pr(X < t)$);
    - (Non-decreasing) $F_X(t_1) \leq F_X(t_2)$ for $t_1 \leq t_2$;

    * (Ranging from 0 to 1) $F_X(-\infty) = 0$ and $F_X(\infty) = 1$.
  - Reversely, a function satisfying the three above properties must be a cdf for certain RV.
    * Indicating an one-to-one correspondence between the set of all the RVs and the set of all the cdfs
  - Knowing the cdf of an RV $\Leftrightarrow$ knowing its distribution

Example Lec1.1

- Given $p \in (0, 1)$, suppose
$$F_X(x) = \begin{cases} 1 - (1-p)^{\lfloor x \rfloor}, & x \geq 1, \\ 0, & \text{otherwise}, \end{cases}$$
where $\lfloor x \rfloor$ represents the integer part of real x.
  - Plot the curve of F_X.

```
p = .3
F_X = function(x) {
  return((1 - (1- p)^floor(x))*ifelse(x >=1, 1, 0))
}
curve(F_X, from = -10, to = 10, n = 1000, col = "blue", lwd = 2,
      xlab = "x", ylab = expression(F[X](x)), main = "Cumulative Distribution Function")
```

- Given $\lambda > 0$, suppose
$$F_X(x) = \begin{cases} 1 - \exp(-x/\lambda), & x > 0, \\ 0, & \text{otherwise}, \end{cases}$$
  - Plot the curve of F_X.

```
lambda = 2
F_X = function(x) {
  return((1 - exp(- x/lambda))*ifelse(x >0, 1, 0))
}
curve(F_X, from = -10, to = 10, n = 1000, col = "blue", lwd = 2,
      xlab = "x", ylab = expression(F[X](x)), main = "Cumulative Distribution Function")
```

Distribution of an RV (con'd)

- Discrete RV
  - RV X merely takes countably different values
  - Probability mass function (pmf): $p_X(t) = \Pr(X = t)$
    * $F_X(t) = \sum_{x \leq t} p_X(x)$
    * $p_X(t) = F_X(t) - \Pr(X < t) = F_X(t) - \lim_{x \to t^-} F_X(x)$
  - Knowing the pmf of a discrete RV $\Leftrightarrow$ knowing its distribution
  - Examples:
    * Bernoulli: a discrete RV with two possible outcomes, typically coded as 0 (failure) and 1 (success).
      · https://en.wikipedia.org/wiki/Bernoulli_distribution
    * Binomial (denoted by $B(n, p)$): the number of successes in n independent Bernoulli trials.
      · https://en.wikipedia.org/wiki/Binomial_distribution
      · E.g., flipping a coin 10 times and counting the number of heads.
    * Geometric: the number of trials until the first success in a series of independent Bernoulli trials.
      · https://en.wikipedia.org/wiki/Geometric_distribution
      · E.g., the number of coin flips needed until the first head appears.
    * Poisson: the number of events that occur in a fixed interval of time or space, where events happen independently.
      · https://en.wikipedia.org/wiki/Poisson_distribution

      - E.g., the number of emails you receive in an hour.
    - Uniform (the discrete version): each outcome in a finite set has an equal probability.
      - https://en.wikipedia.org/wiki/Discrete_uniform_distribution
      - E.g., rolling a fair dice, where each of the six faces has an equal chance of landing.
- Continuous RV
  - RV X is continuous $\Leftrightarrow$ its cdf F_X is absolutely continuous, i.e., there exists f_X such that

$$F_X(t) = \int_{-\infty}^{t} f_X(x)\mathrm{d}x, \quad \forall t \in \mathbb{R}.$$

    - Probability density function (pdf): $f_X(t) = \mathrm{d}F_X(t)/\mathrm{d}t$ (nonnegative for all t).
      - $\int_{-\infty}^{\infty} f_X(x)\mathrm{d}x = \lim_{t\to\infty} \int_{-\infty}^{t} f_X(x)\mathrm{d}x = \lim_{t\to\infty} F_X(t) = 1$
    - $\Pr(X = x_0) = 0$ for all $x_0 \in \mathbb{R}$
      - Because $\Pr(X = x_0) = \Pr(X \le x_0) - \Pr(X < x_0) = F_X(x_0) - \lim_{x\to x_0^-} F_X(x) = 0$
  - Knowing the pdf of a continuous RV $\Leftrightarrow$ knowing its distribution
  - Examples:
    - Uniform (the continuous version): all outcomes in a continuous range are equally likely.
      - https://en.wikipedia.org/wiki/Uniform_distribution_(continuous)
    - Normal/Gaussian (denoted by $\mathcal{N}(\mu, \sigma^2)$): the most important and widely used distributions, where data is symmetrically distributed around the mean.
      - https://en.wikipedia.org/wiki/Normal_distribution
    - Exponential: often used to describe waiting times.
      - https://en.wikipedia.org/wiki/Exponential_distribution

Example Lec1.2

- Given $\lambda > 0$, suppose

$$F_X(x) = \begin{cases} 1 - \exp(-x/\lambda), & x > 0, \\ 0, & \text{otherwise}, \end{cases}$$

  - What is the type of X, discrete or continuous?
- Given $p \in (0, 1)$, suppose

$$F_X(x) = \begin{cases} 1 - (1-p)^{\lfloor x \rfloor}, & x \ge 1, \\ 0, & \text{otherwise}, \end{cases}$$

  where $\lfloor x \rfloor$ represents the integer part of x.
  - What is the type of X, discrete or continuous?

Support of RV (CB pp. 50 & HMC pp. 46)

- For discrete RV X with pmf p_X
  - $\mathrm{supp}(X) = \{x \in \mathbb{R} : p_X(x) > 0\}$
  - E.g., support of $B(n, p)$ is $\{0, \dots, n\}$
  - $\sum_{x\in \mathrm{supp}(X)} p_X(x) = 1$
- For continuous RV X with pdf f_X
  - $\mathrm{supp}(X) = \{x \in \mathbb{R} : f_X(x) > 0\}$
  - E.g., support of $\mathcal{N}(0, 1)$ is $\mathbb{R}$
  - $\int_{\mathrm{supp}(X)} f_X(x)\mathrm{d}x = 1$

Example Lec1.3

- Revisit F_X defined in Example Lec1.1, i.e.,

$$F_X(x) = \begin{cases} 1 - (1-p)^{\lfloor x \rfloor}, & x \ge 1, \\ 0, & \text{otherwise}, \end{cases}$$

  where $\lfloor x \rfloor$ represents the integer part of real x.
  - What is the support of X?

Indicator function

Given a set A, the indicator function of A is

$$\mathbf{1}_A(x) = \begin{cases} 1, & x \in A, \\ 0, & \text{otherwise.} \end{cases}$$

Example Lec1.4

- Revisit F_X defined in Example Lec1.1, i.e.,

$$F_X(x) = \left\{ \begin{array}{ll} 1-(1-p)^{\lfloor x \rfloor}, & x \geq 1, \\ 0, & \text{otherwise,} \end{array} \right.$$

  where $\lfloor x \rfloor$ represents the integer part of x.
  - Please reformulate F_X with the indicator function of $A = \{x : x \geq 1\}$.

Indicating the support when writing pmf and pdf

- Bernoulli: https://en.wikipedia.org/wiki/Bernoulli_distribution
- Binomial (denoted by $B(n,p)$): https://en.wikipedia.org/wiki/Binomial_distribution
  - $p_X(k) = \binom{n}{k} p^k (1-p)^{n-k} \cdot \mathbf{1}_{\{0,1,\ldots,n\}}(k)$
    * OR $\binom{n}{k} p^k (1-p)^{n-k}$, $k \in \{0,1,\ldots,n\}$
- Geometric: https://en.wikipedia.org/wiki/Geometric_distribution
  - $p_X(k) = (1-p)^{k-1} p \cdot \mathbf{1}_{\mathbb{Z}^+}(k)$
    * OR $(1-p)^{k-1}p$, $k \in \mathbb{Z}^+$
- Poisson: https://en.wikipedia.org/wiki/Poisson_distribution
  - $p_X(k) = \lambda^k \exp(-\lambda)/k! \cdot \mathbf{1}_{\{0,1,2,\ldots\}}(k)$
    * OR $\lambda^k \exp(-\lambda)/k!$, $k \in \{0,1,2,\ldots\}$
- Uniform (the discrete version; denoted by $U([a,b])$ with integers $a < b$): https://en.wikipedia.org/wiki/Discrete_uniform_distribution
  - $p_X(k) = 1/(b-a+1) \cdot \mathbf{1}_{\{a,a+1,\ldots,b-1,b\}}(k)$
    * OR $1/(b-a+1)$, $k \in \{a, a+1, \ldots, b-1, b\}$
- Uniform (the continuous version): https://en.wikipedia.org/wiki/Uniform_distribution_(continuous)
- Normal/Gaussian (denoted by $\mathcal{N}(\mu, \sigma^2)$): https://en.wikipedia.org/wiki/Normal_distribution
- Exponential: https://en.wikipedia.org/wiki/Exponential_distribution
  - $f_X(x) = \lambda \exp(-\lambda x) \cdot \mathbf{1}_{[0,\infty)}(x)$
    * OR $\lambda \exp(-\lambda x)$, $x \geq 0$

Expectations (HMC Sec. 1.8–1.9)

- Given RV X and function g, the expectation of $g(X)$ is

$$\mathrm{E}\{g(X)\} = \begin{cases} \sum_{x \in \mathrm{supp}(X)} g(x) p_X(x) & \text{for discrete } X \\ \int_{x \in \mathrm{supp}(X)} g(x) f_X(x) \mathrm{d}x & \text{for continuous } X \end{cases}$$

  - Weighted average of values of $g(X)$
  - $\mathrm{E}\{a_1 g_1(X) + a_2 g_2(X)\} = a_1 \mathrm{E}\{g_1(X)\} + a_2 \mathrm{E}\{g_2(X)\}$ for constants a_1 and a_2
- Examples

- Taking $g(X) = X$

$$\mathrm{E}(X) = \begin{cases} \sum_{x \in \mathrm{supp}(X)} x p_X(x) & \text{for discrete } X \\ \int_{x \in \mathrm{supp}(X)} x f_X(x) \mathrm{d}x & \text{for continuous } X \end{cases}$$

  * The mean of X (a.k.a. the 1st raw moment/moment about 0 of X)
  * $\mathrm{E}(aX + b) = a\mathrm{E}(X) + b$ for constants a and b
- Taking $g(X) = X^k$ with positive integer k:

$$\mathrm{E}(X^k) = \begin{cases} \sum_{x \in \mathrm{supp}(X)} x^k p_X(x) & \text{for discrete } X \\ \int_{x \in \mathrm{supp}(X)} x^k f_X(x) \mathrm{d}x & \text{for continuous } X \end{cases}$$

  * The kth raw moment/moment about 0 of X
- Taking $g(X) = \{X - \mathrm{E}(X)\}^2$:

$$\mathrm{Var}(X) = \mathrm{E}[\{X - \mathrm{E}(X)\}^2] = \begin{cases} \sum_{x \in \mathrm{supp}(X)} \{x - \mathrm{E}(X)\}^2 p_X(x) & \text{for discrete } X \\ \int_{x \in \mathrm{supp}(X)} \{x - \mathrm{E}(X)\}^2 f_X(x) \mathrm{d}x & \text{for continuous } X \end{cases}$$

  * Variance of X (a.k.a. the 2nd central moment of X)
  * Measuring how spread out the data are if they are independently generated following F_X
  * $\mathrm{Var}(X) = \mathrm{E}(X^2) - \{\mathrm{E}(X)\}^2$
  * $\mathrm{Var}(aX + b) = a^2 \mathrm{Var}(X)$
  * $\mathrm{sd}(X) = \sqrt{\mathrm{Var}(X)}$: the standard deviation of X
- Taking $g(X) = \mathbf{1}_A(X)$:

$$\mathrm{E}\{\mathbf{1}_A(X)\} = \Pr(X \in A)$$

Example Lec1.5

- Find the mean and variance of $X \sim \mathcal{N}(0, 1)$, i.e., $f_X(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$

$$\mathrm{E}(X) = \int_{\mathbb{R}} x f_X(x) \mathrm{d}x \overset{x\exp(-x^2/2) \text{ is odd}}{=} \int_{\mathbb{R}} \frac{x}{\sqrt{2\pi}} \exp(-x^2/2) \mathrm{d}x = 0$$

$$\mathrm{Var}(X) \overset{\text{even } x^2 \exp(-x^2/2)}{=} 2\int_0^\infty \frac{x^2 \exp(-x^2/2)}{\sqrt{2\pi}} \mathrm{d}x \overset{u = x^2/2}{=} 2\int_0^\infty \frac{2u \exp(-u)}{\sqrt{2\pi}} \mathrm{d}\sqrt{2u} = \frac{2\Gamma(3/2)}{\sqrt{\pi}} = 1$$

- Find the mean and variance of $X \sim \mathcal{N}(\mu, \sigma^2)$ with $\mu \in \mathbb{R}$ and $\sigma \in \mathbb{R}^+$, i.e., $f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$ (p.s. $X \sim \mathcal{N}(\mu, \sigma^2) \Leftrightarrow Z = (X - \mu)/\sigma \sim \mathcal{N}(0, 1)$)
- Find the mean and variance of Cauchy distribution, i.e., $f_X(x) = \{\pi(1 + x^2)\}^{-1}$, $x \in \mathbb{R}$

$$\int_1^\infty \frac{x^2}{\pi(1 + x^2)} \mathrm{d}x \geq \int_1^\infty \frac{x}{\pi(1 + x^2)} \mathrm{d}x = \infty$$