PH 712 Probability and Statistical Inference

Part I: Random Variable

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

2024/09/08 17:08:32

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
  - $\Omega = \{$"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 sample spaces,
  - $(0, 10]$: the survival time ≤ 10
  - $\{$"yes"$\}$: the treatment is effective
- Occurrence of event: the outcome is part of the event
- Probability (denoted by Pr): a function quantifying the likelihood of the occurrence of events
  - Input: an event
  - Output: a real number
  - Requirements:
    * $\Pr(A) \geq 0$ 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):
    * $\Pr(A) = 1 - \Pr(A^c)$, where the superscript denotes the complement set
    * $\Pr(\emptyset) = 0$
    * $\Pr(A) \leq \Pr(B)$ if $A \subset B$
    * $0 \leq \Pr(A) \leq 1$ for each A
    * $\lim_{n\to\infty} \Pr(A_n) = \Pr(\lim_{n\to\infty} A_n) = \Pr(\bigcup_{n=1}^{\infty} A_n)$ if $\{A_n\}_{n=1}^{\infty}$ is nondecreasing (i.e., $A_1 \subset A_2 \subset \cdots$)
    * $\lim_{n\to\infty} \Pr(A_n) = \Pr(\lim_{n\to\infty} A_n) = \Pr(\bigcap_{n=1}^{\infty} A_n)$ if $\{A_n\}_{n=1}^{\infty}$ is nonincreasing (i.e., $A_1 \supset A_2 \supset \cdots$)
    * $\Pr(A \bigcup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B)$ 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)

- Conditional probability of B given A (with $\Pr(A) > 0$): $\Pr(B \mid A) = \Pr(A \cap B)/\Pr(A)$
  - 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., $\{A_{n_1}\}_{n=1}^{N}$ 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 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 $\Pr(A) \neq 0$)
- 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 $2 \leq K \leq N$, $\Pr(\bigcap_{k=1}^{K} A_{n_k}) = \prod_{k=1}^{K} \Pr(A_{n_k})$

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

- An RV: a function encoding the entries of Ω
    - Input: an entry of Ω
    - Output: a real number
    - Usage: any event may be expressed in term of
- 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 distribution of an RV $\Leftrightarrow$ knowing the cdf

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.
    - Show that F_X is a cdf. (Hint: Check all the three properties of cdf, especially the right-continuity of F at positive integers.)

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 distribution of a discrete RV $\Leftrightarrow$ knowing the pmf
    - 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: the number of successes in a fixed number of independent Bernoulli trials.

          - https://en.wikipedia.org/wiki/Binomial_distribution
          - E.g., flipping a coin 10 times and counting the number of heads.
      * Negative binomial: the number of trials until a specified number of successes is achieved.
          - https://en.wikipedia.org/wiki/Negative_binomial_distribution
          - E.g., the number of coin flips until you get 3 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.
      * Hypergeometric: the number of successes in a sample drawn without replacement from a finite population.
          - https://en.wikipedia.org/wiki/Hypergeometric_distribution
          - E.g., drawing a certain number of red balls from a bag containing both red and blue balls without replacement.
      * 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): $f_X(t) = \mathrm{d}F_X(t)/\mathrm{d}t = \lim_{\delta\to 0^+} \Pr(t < X \le t+\delta)/\delta (\ge 0)$.
      * $\int_{-\infty}^{\infty} f_X(x)\mathrm{d}x = 1$
  - Knowing the distribution of a continuous RV $\Leftrightarrow$ knowing the pdf
  - 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: the time between events in a Poisson process, often used to describe waiting times.
          - https://en.wikipedia.org/wiki/Exponential_distribution
      * Chi-squared: sum of squared standard normal RVs; arising in hypothesis testing, particularly in tests of independence and goodness of fit.
          - https://en.wikipedia.org/wiki/Chi-squared_distribution
      * Cauchy: known for its heavy tails and undefined mean and variance; used in robust statistics.
          - https://en.wikipedia.org/wiki/Cauchy_distribution
      * Weibull: a generalization of the exponential distribution, used in reliability engineering and failure time analysis.
          - https://en.wikipedia.org/wiki/Weibull_distribution
      * Log-normal: $\exp(\mathcal{N}(0, 1))$; commonly used to model stock prices and other financial data.
          - https://en.wikipedia.org/wiki/Log-normal_distribution
      * (Student's) t: used in hypothesis testing, particularly for small sample sizes.
          - https://en.wikipedia.org/wiki/Student%27s_t-distribution

Example Lec1.2

- 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 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 $Binom(n,p)$ is $\{0, \ldots, n\}$
  - $\int_{\mathrm{supp}(X)} f_X(x)\mathrm{d}x = 1$
- 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 $\mathbb{R}$
  - $\sum_{x \in \mathrm{supp}(X)} p_X(x) = 1$

Example Lec1.3

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

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

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

Expectations (HMC Sec. 1.8–1.9)

- Given RV X and function g, the expectation of $g(X)$ is $\mathrm{E}\{g(X)\}$
  - $= \int_{-\infty}^{\infty} g(x) f_X(x)\mathrm{d}x$ for continuous X
  - $= \sum_{x \in \mathrm{supp}(X)} g(x) p_X(x)$ for discrete X
  - Weighted average of values of $g(X)$
  - $\mathrm{E}\{a_1 g_1(X) + a_2 g_2(X)\} = a_1 \mathrm{E}\{g_1(X)\} + a_2 \mathrm{E}\{g_2(X)\}$
- Mean of X (a.k.a. the 1st raw moment/moment about 0 of X): $\mathrm{E}(X)$
- Variance of X (a.k.a. the 2nd central moment of X): $\mathrm{Var}(X) = \mathrm{E}\{X - \mathrm{E}(X)\}^2$
  - $\mathrm{Var}(X) = \mathrm{E}(X^2) - \{\mathrm{E}(X)\}^2$
  - $\mathrm{Var}(aX+b) = a^2 \mathrm{Var}(X)$
- Standard deviation of X: square root of the variance of X

Example Lec1.4

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

Distribution of an RV (con'd)

- Moment generating function (MGF, HMC Sec. 1.9/CB Sec. 2.3)
  - $M_X(t) = \mathrm{E}\{\exp(tX)\}$
    * Continuous X: $M_X(t) = \int_{-\infty}^{\infty} \exp(tx) f_X(x) \mathrm{d}x$
    * Discrete X: $M_X(t) = \sum_{u \in \mathrm{supp}(X)} \exp(tx) p_X(x)$
  - The MGF of X is $M_X(t)$, $t \in A$, $\Leftrightarrow$ $M_X(t)$ is finite for t in a neighborhood of 0, say A; otherwise the MGF does NOT exist or is NOT well defined.
  - $M_{aX+b}(t) = \exp(bt) M_X(at)$
  - Knowing the distribution of an RV $\Leftrightarrow$ knowing the MGF (if any)
- Characteristic function (CF, optional)
  - $\varphi_X(t) = \mathrm{E}\{\exp(itX)\}$
    * Continuous X: $\varphi_X(t) = \int_{-\infty}^{\infty} \exp(itu) f_X(u) \mathrm{d}u$
    * Discrete X: $\varphi_X(t) = \sum_{u \in \mathrm{supp}(X)} \exp(itu) p_X(u)$
  - Always well-defined
  - $\varphi_{aX+b}(t) = \exp(bt) \varphi_X(at)$
  - Knowing the distribution of an RV $\Leftrightarrow$ knowing the CF

Example Lec1.5

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

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 Lec1.6

- 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 $A = \{x : x \geq 1\}$.