Probability

2017-02-20

- Axioms of probability
  - For any event A of the **sample space** S, the probability of A is greater than or equal to 0, i.e. $P(A) \geq 0$.
  - The probability of the sample space S is 1, i.e. $P(S) = 1$.
  - For any countable mutually exclusive events $A_1, A_2, \ldots, A_n$, the probability of their union is the sum of individual probability, i.e. $P(A_1 \cup A_2 \cup \ldots \cup A_n) = P(A_1) + P(A_2) + \ldots + P(A_n)$.
- Axioms will not be proved. All the theorems, definitions, and properties will be based on the axioms.

2017-03-06

- Events A and B are **exclusive** if and only if $P(A \cup B) = P(A) + P(B)$.
- Events $A, B, C, \ldots$ are **mutually exclusive** if and only if the following two conditions hold:
  - $A, B, C, \ldots$ are pairwise exclusive.
  - $P(A \cup B \cup C \cup \ldots) = P(A) + P(B) + P(C) + \ldots$
- Events A and B are **independent** if and only if $P(A \cap B) = P(A) \cdot P(B)$.
- Events A, B, and C are **mutually independent** if and only if the following two conditions hold:
  - $A, B, C, \ldots$ are pairwise independent.
  - $P(A \cap B \cap C \cap \ldots) = P(A) \cdot P(B) \cdot P(C) \cdot \ldots$
- **Conditional probability** also satisfies the axioms of probability.
- Events that are independent are sometimes called **statistically independent**, **stochastically independent**, or **independent in a probabilistic sense**.
- If A and B are independent events, then the following pairs of events are also independent: (a) A and B'; (b) A' and B; (c) A' and B'.
- [x] Homework: 1.3-13, 1.4-15, 1.5-4
- Solutions: Solutions1.pdf

2017-03-13

- **Bayes' theorem**:
  - For an event of interest, we have its **prior** probability, given another event that has happened, we look for its **posterior** probability.
  - $P(B_i|A) = \frac{P(A|B_i)P(B_i)}{P(A|B_1)P(B_1)+...+P(A|B_n)P(B_n)}$, where $\{B_1, ... B_n\}$ is a partition of the sample space. $P(B_i)$ is the **prior**, and $P(B_i|A)$ is the **posterior**.
- A **random variable** is a function that maps the possible outcomes of an experiment to real numbers.
- If S is a *finite* or *countable infinite* set, then X is said to be a **discrete random variable**.
- A set is said to be **countable infinite** if
  - it contains infinite number of elements.
  - there exists a *one-to-one* mapping between each element of the set and the positive integers.
- Examples of coutable infinite sets: (a) integer numbers; (b) fractional numbers. In the case of real numbers, it is an **uncountable infinite** set.
- **Probability mass function (PMF)** v.s. **Probability density function (PDF)**
  - A **probability mass function (PMF)** is a function that gives the probability that a *discrete random variable* is exactly equal to some value.
  - A **probability density function (PDF)** can be interpreted as providing a relative likelihood that the value of the *continuous random variable* would equal that sample.
- The **cumulative distribution function (CDF)**, or **probability distribution function** is the probability that X will take a value less than or equal to x: $F(x) = P(X \leq x)$
- PMF and CDF of a discrete random variable:
  - PMF: $f(x) = P(X = x)$
  - CDF: $F(x) = P(X \leq x)$
- *Discrete* probability distributions: **uniform**, **binomial**, **negative binomial**, **geometric**, **hypergeometric**, and **Poisson**.
- **Uniform distribution**: $f(x) = \frac{1}{b-a+1} = \frac{1}{n}, x = [a, b] \in \mathbb{Z}$ if we let $n = b - a + 1$
  - Mean: $\frac{a+b}{2}$
  - Variance: $\frac{(b-a+1)^2-1}{12} = \frac{n^2-1}{12}$
  - MGF: $\frac{e^{at}-e^{(b+1)t}}{(b-a+1)(1-e^t)} = \frac{e^{at}-e^{(b+1)t}}{n(1-e^t)}$
  - Function in `Python`: `scipy.stats.randint`
- **Hypergeometric distribution**: $f(x) = \binom{n_1}{x} \cdot \binom{n_2}{n-x} / \binom{n_1+n_2}{n}, x = [0, \min\{n_1, n\}] \in \mathbb{Z}$

- Literal: x successes in a sample of size n drawn without replacement.
- Mean: $n(\frac{n_1}{n_1+n_2}) = np$ if we let $p = \frac{n_1}{n_1+n_2}$
- Variance: $n(\frac{n_1}{n_1+n_2})(\frac{n_2}{n_1+n_2})(\frac{n_1+n_2-n}{n_1+n_2-1}) = np(1-p)(\frac{n_1+n_2-n}{n_1+n_2-1})$ if we let $p = \frac{n_1}{n_1+n_2}$
- MGF: (ignored)
- Function in `Python`: `scipy.stats.hypergeom`

2017-03-20

- **Expected value**: $E[x] = \sum x \cdot f(x)$
- $E[u(x)] = \sum u(x) f(x)$
- Theorems about the expected value
  - If c is a constant, $E[c] = c$
  - If c is a constant and u is a function, then $E[c \cdot u(x)] = c \cdot E[u(x)]$
  - If c_1 and c_2 are constants and u_1 and u_2 are functions, then $E[c_1 \cdot u_1(x) + c_2 \cdot u_2(x)] = c_1 \cdot E[u_1(x)] + c_2 \cdot E[u_2(x)]$
- The **variance** of a random variable is defined to be $E[(X-\mu)^2]$ and is typically denoted by σ^2 or $\mathrm{var}(X)$. σ is normally called the **standard deviation**.
- $E[X]$ is the value of b that minimizes $E[(X-b)^2]$.
- Let $Y = aX + b$
  - $\mu_Y = a \cdot \mu_X + b$
  - $\sigma_Y^2 = a^2 \cdot \sigma_X^2$
- In many distributions, the *mean* and *variance* together uniquely determine the *parameters* of the random variables.
- **Moment** of a distribution:
  - $E[X^k]$: the k-th moment of the distribution about *origin*.
  - $E[(X-b)^k]$: the k-th moment of the distribution about b.
  - $E[(X)_k] = E[X(X-1)(X-2)\dots(X-k+1)]$: the k-th **factorial moment**.
- Some properties about moment:
  - *Mean* is the first moment about 0.
  - *Variance* is (1) the second moment about mean, or (2) the difference of the second moment and the square of the first moment about 0.
  - The *second factorial moment* is the difference of the second and first moments about 0.
- **Moment-generating function (MGF)**:
  - $M(t) = E[e^{tX}] = \sum e^{tx} f(x)$
  - $M^{(r)}(t) = E[X^r e^{tX}] = \sum x^r e^{tx} f(x)$

- $M^{(r)}(0) = E[X^r] = \sum x^r f(x)$
- The MGF of a random variable uniquely determines the distribution of that random variable.
- If the MGF exists, there is one and only one distribution of probability associated with that MGF.
- If two random variables (or two distributions of probability) have the same MGF, they must have the same distribution of probability.

- **Bernoulli experiment**:
  - The outcome can be classified in one of two mutually exclusive and exhaustive ways, say, success and failure.
  - Performed several independent times.
- **Binomial distribution**: $f(x) = \binom{n}{x} p^x (1-p)^{n-x}$
  - Literal: x successes in a sample of size n drawn with replacement.
  - Mean: np
  - Variance: $np(1-p)$
  - MGF: $(1-p+pe^t)^n$
  - Function in Python: `scipy.stats.binom`
- **Geometric distribution**: $f(x) = p(1-p)^{x-1}$
  - Literal: 1 success in a sample of size x drawn with replacement.
  - Mean: $\frac{1}{p}$
  - Variance: $\frac{1-p}{p^2}$
  - MGF: $\frac{pe^t}{1-(1-p)e^t}$
  - Function in Python: `scipy.stats.geom`
- [x] Homework: 2.1-16, 2.2-10, 2.3-11, 2.4-20
- Solutions: [Solutions2.pdf](), [hw2_b00401062.pdf]()

2017-04-10

- **Negative binomial distribution**: $f(x) = \binom{x-1}{r-1} p^r (1-p)^{x-r}$
  - Literal: r successes in a sample of size x drawn with replacement.
  - Mean: $\frac{r}{p}$
  - Variance: $\frac{r(1-p)}{p^2}$
  - MGF: $(\frac{pe^t}{1-(1-p)e^t})^r$
  - Function in Python: `scipy.stats.nbinom`

- An **approximate Poisson process** has to satisfy the following conditions:
    - The numbers of occurrences in nonoverlapping subintervals are independent.
    - The probability of exactly one occurrence in a sufficiently short subinterval of length h is approximately λh.
    - The probability of two or more occurrences in a sufficiently short subinterval is essentially 0.
- λ can be understood as the average (expected) number of occurences in an interval of length 1.
- The PMF of Poisson distribution $f(x)$ can be solved by $\lim_{n\to\infty}\binom{n}{x}(\frac{\lambda}{n})^x(1-\frac{\lambda}{n})^{n-x}=\frac{e^{-\lambda}\lambda^x}{x!}$, which can be understood as a binomial distribution whose $p=\frac{\lambda}{n}$.
- **Poisson distribution**: $f(x)=\frac{e^{-\lambda}\lambda^x}{x!}$
    - Mean: λ
    - Variance: λ
    - MGF: $e^{\lambda(e^t-1)}$
    - Function in `Python`: `scipy.stats.poisson`
- PMF and CDF of a continuous random variable:
    - PMF: $f(x)=\mathrm{d}F(x)/\mathrm{d}x$
    - CDF: $F(x)=P(X\leq x)=\int_{-\infty}^{x}f(t)\mathrm{d}t$