binomial

适用条件: ① n & p 固定 ② outcome 仅成功/否 ③ 事件互相 independent.

$X \sim B(n, p)$: $P(X=r) = \binom{n}{r} p^r (1-p)^{n-r}$

$E(aX_1 + bX_2) = aE(X_1) + bE(X_2)$

$Var(aX_1 + bX_2) = a^2 Var(X_1) + b^2 Var(X_2)$

mean of $X = E(X) = \mu = np$

variance of $X = Var(X) = \sigma^2 = np(1-p)$

poisson

适用条件 : ① events independent ② 单独出现 ③ occur at a constant average rate

$X \sim Po(\lambda)$: $P(X=x) = \frac{e^{-\lambda}\lambda^x}{x!}$

$X+Y \sim Po(\lambda+\mu)$ X & Y independent.

mean of $X = E(X) = \lambda$

variance of $X = Var(X) = \sigma^2 = \lambda$

$$\text{type of data}\begin{cases}\text{qualitative}\\[2ex] \text{quantitative}\begin{cases}\text{discrete}\\ \text{continuous}\end{cases}\end{cases}$$

$$\bar{x} = \frac{\sum x}{n}$$

variance $= \dfrac{S_{xx}}{n} = \dfrac{\sum(x-\bar{x})^2}{n} = \dfrac{\sum x^2}{n} - (\bar{x})^2$ $\qquad S_{xx} = \sum(x-\bar{x})^2 = \sum x^2 - \dfrac{(\sum x)^2}{n}$

s.d. $\quad \sigma = \sqrt{Var}$

1. Probability

- probability = $\frac{\text{\# event can occur}}{\text{\# total outcomes}}$

- relative frequency : proportion of times that an event occur in huge repetitions of experiment

  概率随着实验次数增加无限接近一个数 eg. P(骰子 1 朝上) → $\frac{1}{6}$

- subjective frequency : (相对主观) eg. P(天气预报为晴)

  the probability of an event is a measure of how sure the person think the event will happen.

2. Probability Model

- sample space (S) : a set of distinct outcomes for an experiment
  所有可能性. ☆不重复. 不遗漏

  - discrete sample space : finite or "countably finite" $S = \{1, 2, 3, \cdots\}$
  - continuous sample space $S = \{x : x > 0\}$

- event (A) : a subset of sample space $A \subseteq S$
  - simple event event contains only one point $A = \{a_1\}$
  - compound event A is made up by more than 1 simple events $A = \{a_1, a_2\}$

☆ sample point 样本点 : sample space 里的个体

- probability distribution : Let $S = \{a_1, a_2, \cdots\}$ be a discrete sample space. Assign probabilities $P(a_i)$ for $i = 1, 2, \cdots$ to a_i's

  s.t. 1) $0 \leq P(a_i) \leq 1$

  2) $\sum_{\forall i} P(a_i) = 1$.

  The set of probabilities $\{P(a_i) : i = 1, 2, \cdots\}$ is called probability distribution of S.

- probability of an event $P(A)$: the sum of probabilities for all simple events that make up A. $P(A) = \sum_{a \in A} P(a)$.

  eg. $A = \{2, 4, 6\}$. $P(A) = P(2) + P(4) + P(6)$

simple : $P(a_i) = \frac{1}{|S|}$

compound : $P(A) = \sum_{a \in A} P(a_i) = \frac{|A|}{|S|}$ ← A事件发生的次数

- Odds

$\begin{cases} \text{odds in favour 赔率} & \dfrac{P(A)}{1-P(A)} \\ \text{odds against 胜率} & \dfrac{1-P(A)}{P(A)} \end{cases}$

Q. roll a die

a) P(odds in favour of rolling a "6") $= \dfrac{\frac{1}{6}}{1-\frac{1}{6}} = \dfrac{1}{5}$

b) P(odds against of rolling a "6") $= \dfrac{1-\frac{1}{6}}{\frac{1}{6}} = 5$

3. Probability and counting techniques

- addition rule.

  If A & B are disjoint ($A \cap B = \emptyset$), then $|A \cup B| = |A| + |B|$

- multiplication rule.

  $A = (a_1, a_2, \cdots, a_k)$ with n_i choices for a_i

  $|A| = n_1 \cdot n_2 \cdots n_k = \prod_{i=1}^{k} n_i$

- **Number of subsets of size k:** We use the combinatorial symbol $\binom{n}{k}$ ("n *choose* k") to denote the number of subsets of size k that can be selected from a set of n objects. By an argument similar to that above, if m denotes the number of subsets of size k that can be selected from n things, then $m \times k! = n^{(k)}$ and so we have

$$m = \binom{n}{k} = \frac{n^{(k)}}{k!}$$ ← binomial coefficient

permutation: $n^{(k)} = n(n-1)\cdots(n-k-1) = \frac{n!}{(n-k)!}$ n个空填k个数 (数字不同)

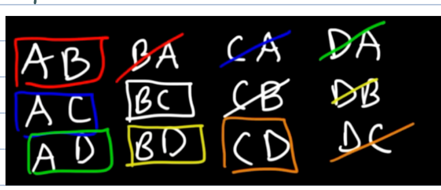
• order matters. $ABC \neq CAB$

combination: $\underline{\binom{n}{k}} = \frac{n^{(k)}}{k!} = \frac{n!}{(n-k)!\,k!}$ n个空填k个数 (数字相同)

n choose k

• order doesn't matter $ABC = CAB$

eg. 从ABCD中选2个字母



$P = 12$ $_nP_R = \frac{n!}{(n-R)!}$ 在4个字母中任选2个 $_4P_2 = \frac{4!}{(4-2)!}$

$C = 6$ $_nC_R = \frac{_nP_R}{R!} = \frac{n!}{(n-R)!\,R!}$ $_4C_2 = \frac{4!}{(4-2)!\,2!}$

将这R个排列组合的可能性

Properties of $\binom{n}{k}$: You should be able to prove the following for n and k non-negative integers:

1. $n^{(k)} = \frac{n!}{(n-k)!} = n(n-1)^{(k-1)}$ for $k \geq 1$
2. $\binom{n}{k} = \frac{n!}{k!(n-k)!} = \frac{n^{(k)}}{k!}$
3. $\binom{n}{k} = \binom{n}{n-k}$ for all $k = 0, 1, \ldots, n$
4. If we define $0! = 1$, then the formulas hold with $\binom{n}{0} = \binom{n}{n} = 1$.
5. $\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}$
6. **Binomial Theorem:** $(1+x)^n = \binom{n}{0} + \binom{n}{1}x + \binom{n}{2}x^2 + \ldots + \binom{n}{n}x^n$

0 — 1
1 — 1 1
2 — 1 2 1 $\binom{3}{1}$
$\binom{3}{0}$ 1 3 3 1
1 4 6 4 1
1 5 10 10 5 1
1 6 15 20 15 6 1
1 7 21 35 35 21 7 1

$\binom{x}{n} - \binom{x-1}{n} = \binom{x-1}{n-1}$

☆ Q Suppose you have 20 distinct books, 7 of which are written by Mark Twain.

a) How many ways can you arrange 12 books on a shelf if the order they are on the shelf matters?

b) How many ways can you arrange 12 books on a shelf if exactly 3 of them must be Mark Twain books?

c) A monkey picks books at random from the 20 books and puts them on the shelf until it contains 12 books. What is the probability that at least 3 of the books on the shelf are written by Mark Twain?

a) $20^{(12)}$

b) $\underbrace{\binom{7}{3}\binom{13}{9}}_{\text{选}} \cdot \underbrace{12!}_{\text{arrange}}$

c) $P(2本) = \binom{7}{2}\cdot\binom{13}{10}$

$P(1本) = \binom{7}{1}\cdot\binom{13}{11}$

$P(0本) = \binom{7}{0}\cdot\binom{13}{12}$

$1 - P(2本) - P(1本) - P(0本)$

Q. H E L L O K I T T Y 任意组合形式排列

a) how many ways can you this be done?

$\frac{10!}{2!\,2!}$ ← 将 L & T 区别成 L_1 L_2 T_1 T_2

← # ways to arrange L.

ways to arrange T

b) P (all letters appear in alphabetic order).

$\frac{1}{\frac{10}{2!\,2!}}$

c) P (begin and ends with T)

$\frac{\frac{8!}{2!}}{\frac{10!}{2!\,2!}}$ ⇒ 8 position but 2L we need to consider

Number of arrangements when symbols are repeated:

If we have n_i symbols of type i, $i = 1, 2, \ldots, k$ with $n_1 + n_2 + \cdots + n_k = n$, then the number of arrangements using all of the symbols is

$$\binom{n}{n_1}\binom{n-n_1}{n_2}\binom{n-n_1-n_2}{n_3}\cdots\binom{n_k}{n_k}$$

$$= \frac{n!}{n_1!n_2!\cdots n_k!}$$

← if they were distinguishable

undistinguishable array

multinomial coefficient $\binom{n}{n_1, n_2, \cdots, n_k} = \frac{n!}{n_1!n_2!\cdots n_k!}$

a) **Addition rule**: something OR $(+)$ something else

b) **Multiplication rule**: something AND $(\times)$ something else

c) **Factorial**: $n! =$ Number of arrangements of n distinct objects when order matters

d) **Permutation**: $n^{(k)} = \frac{n!}{k!} =$ Number of ways to pick k objects from n distinct objects when order matters (without replacement)

$13 \neq 31$

e) **Combination**: $\binom{n}{k} = \frac{n!}{k!(n-k)!} =$ Number of ways to choose k objects from n objects when order doesn't matter Without replacement

f) **Multinomial coefficient**: $\binom{n}{n_1,\ldots,n_k} = \frac{n!}{n_1!\ldots n_k!} =$ Number of ways to arrange n_1 objects of type 1, ..., n_k objects of type k, where $n_1 + n_2 + \cdots + n_k = n$.

4. Probability rules and conditional probability

- union $A \cup B = \{ s \in S : s \text{ belong to either } A \text{ or } B \}$

intersection. $A \cap B = \{ s \in S : s \text{ belongs to both } A \,\&\, B \}$

complement $\bar{A} = \{ s \in S : s \text{ doesn't belong to } A \}$

- fundamental laws of set algebra

(a) commutativity

$A \cup B = B \cup A \qquad A \cap B = B \cap A$

(b) associativity

$(A \cup B) \cup C = A \cup (B \cup C) \qquad (A \cap B) \cap C = A \cap (B \cap C)$

(c) distributivity

$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$
$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$

- DeMorgan's law $\overline{A \cup B} = \bar{A} \cap \bar{B}$

$\overline{A \cap B} = \bar{A} \cup \bar{B}$.

Exercise: for events *A*,*B*, and *C* from S, find representations of the following events:

- *A* occurs but not *B* $\longrightarrow$ $A \cap \bar{B}$
- *A* occurs but not *B* nor *C* $\longrightarrow$ $A \cap (\bar{B} \cap \bar{C})$
- *A* and *C* occur but not *B* $\longrightarrow$ $A \cap C \cap \bar{B}$
- *A* or *B* occur but not *C* $\longrightarrow$ $(A \cup B) \cap \bar{C}$

- probability rules:

  R_1 : $P(S) = 1$

  R_2 : $0 \le P(A) \le 1$.

  R_3 : $A \subseteq B \Rightarrow P(A) \le P(B)$

- union rules

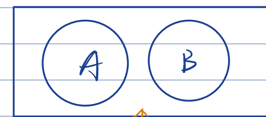
  $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ 画 Venn 图中

  $P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(AB) - P(BC) - P(AC) + P(ABC)$

- mutually exclusive (disjoint) A & B has no common points.

  $A \cap B = \phi$.

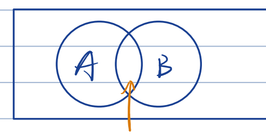
  $P(A_1 \cup A_2 \cup \cdots \cup A_n) = \sum_{i=1}^{n} P(A_i)$



无重叠

- indepent events A 与 B 之间没有关系

  $P(A \cap B) = P(A)P(B)$



$A \cap B$

- mutually independent $P(A_{i_1} \cap \cdots \cap A_{i_k}) = P(A_{i_1}) \cdots P(A_{i_k})$

  → proposition: A & B mutually exclusive ∧ independent. $\Rightarrow P(A) = 0$ or $P(B) = 0$

  proof. ∵ A & B mutually exclusive. & independent

  $\therefore P(A \cap B) = P(A) \cdot P(B) = P(\phi) = 0$

  $\therefore P(A) \cdot P(B) = 0$

  $\therefore P(A) = 0$ or $P(B) = 0$

  → proposition: A & B are independent $\Rightarrow \bar{A}$ & B are independent

- conditional probability

$P(A|B) = \dfrac{P(A\cap B)}{P(B)}$ (A given B) 在B发生的情况下 A发生的概率.

↳ 用于定义 independence

A & B are independent $\Leftrightarrow$ $P(A|B) = P(A) \vee P(B|A) = P(B)$

$\cancel{= \dfrac{P(A)P(B)}{P(B)}}$

→ $0 \le P(A|B) \le 1$

→ $P(\bar{A}|B) = 1 - P(A|B)$

※ → 若 A_1 & A_2 disjoint, 则 $P(A_1 \cap A_2 | B) = P(A_1|B) + P(A_2|B)$

→ $P(S|B) = 1 = P(B|B)$

- product rule: $P(A\cap B) = P(A|B)P(B) = P(B|A)P(A)$

$P(A\cap B\cap C) = P(A)\cdot P(B|A)\cdot P(C|AB)$

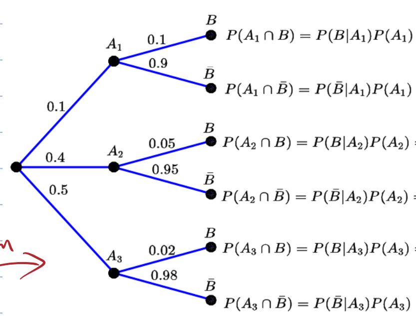
proof: $RHS = P(A)\dfrac{P(AB)}{P(A)} P(C|AB)$

$= P(AB)\cdot P(C|AB)$

$= P(AB)\cdot \dfrac{P(CAB)}{P(AB)}$

$= LHS$

e.g. tree diagram



- law of total probability

Let $A_1, A_2 \cdots A_k$ be a partition of the sample space S into disjoint events (mutually exclusive). $A_1 \cup \cdots \cup A_k = S$. $A_i \cap A_j = \emptyset$. $\forall i \ne j$

Let B be arbitrary event in S, then $P(B) = P(BA_1) + P(BA_2) + \cdots + P(BA_k)$

$= \sum_{i=1}^{k} P(B|A_i)P(A_i)$

- Bayes's theorem

$P(A|B) = \dfrac{P(B|A)P(A)}{P(B)} = \dfrac{P(B|A)\cdot P(A)}{P(B|\bar{A})P(\bar{A}) + P(B|A)\cdot P(A)}$

5. Discrete Random variables

- random variable

$X: S \to \mathbb{R}$

a function maps from sample space S to set of real numbers $\mathbb{R}$.

ep. 抛硬币

$S = \{H, T\}$

$X(s) = \begin{cases} 1 & \text{if } s = H \\ 0 & \text{if } s = T \end{cases}$

range (of random variable): the value that a random variable takes

$\begin{cases} \text{discrete} & \text{range is discrete subset of } \mathbb{R} \\ \text{continuous} & \text{range is an interval} \quad \text{ep. } X(s) = [0, 1] \end{cases}$

- Probability (mass) function p.f.

$f(x) = P(X = x) \quad (x \in \text{range}(X))$

probability distribution $\{(x, f(x)) : x \in \text{range}(X)\}$

Properties of p.f.: ① $f(x) \geq 0 \quad \forall x \in A$

② $\sum_{\forall x \in A} f(x) = 1$

ep. 抛3次硬币. $X = \#$ heads occur.

x	0	1	2	3
$P(X=x) = f(x)$	$(\frac{1}{2})^3$	$3 \cdot (\frac{1}{2})^3$	$3 \cdot (\frac{1}{2})^3$	$(\frac{1}{2})^3$

- Cumulative distribution function (cdf)

$$F(x) = P(X \le x) \qquad x \in \mathbb{R}$$
$$= P(\{w \in S : X(w) \le x\})$$

If X is discrete with probability function f, then $F(x) = P(X \le x) = \sum_{y \le x} f(y)$

Properties of cdf: ① $0 \le F(x) \le 1$

② $F(x) \le F(y) \quad \forall x < y$

③ $\lim_{x \to -\infty} F(x) = 0 \qquad \lim_{x \to \infty} F(x) = 1$

* 证明 cdf. ① $\lim_{x \to \infty} F_X(x) = 1$ ② $\lim_{x \to -\infty} F_X(x) = 0$ ③ non-decreasing

eg. 掷骰子. X = 掷出之数.

x	1	2	3	4	5	6	
$f(x) = P(X = x)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	

X	$x<1$	$1 \le x < 2$	$2 \le x < 3$	$3 \le x < 4$	$4 \le x < 5$	$5 \le x < 6$	$x \ge 6$
$F(x) = P(X \le x)$	0	$\frac{1}{6}$	$\frac{2}{6}$	$\frac{3}{6}$	$\frac{4}{6}$	$\frac{5}{6}$	1

$$F(2) = P(X \le 2) = P(X=1) + P(X=2) = \frac{2}{6}$$

$$f_X(x) = P(X \le x) - P(X \le x-1) = F_X(x) - F_X(x-1)$$

Example

Suppose students A,B and C each independently answer a question on a test. The probability of getting the correct answer is 0.9 for A, 0.7 for B and 0.4 for C. Let X denote the number of people who get the answer correct.

a) Compute the probability function of X

b) Compute the cdf of X.

c) Compute the probability that at least one person gets the answer correct.

$P(A)=0.9 \qquad P(\bar{A})=0.1$

$P(B)=0.7 \qquad P(\bar{B})=0.3$

$P(C)=0.4 \qquad P(\bar{C})=0.6$

a)

x	0	1	2	3
$f(x)$	0.018	0.216	0.514	0.252

$0.1\times0.3\times0.6$

$0.9\times0.3\times0.6+0.7\times0.1\times0.6+0.4\times0.1\times0.3$

$0.9\times0.7\times0.6+0.9\times0.3\times0.4+0.1\times0.7\times0.4$

$0.1\times0.3\times0.6$

b)

x	0	1	2	3
$F_X(x)=P(X\le x)$	0.018	0.234	0.748	1

c) $1-P(\text{no one})=1-0.018=0.982$.

Example

Suppose that N balls labelled $1,2,\ldots,N$ are placed in a box, and n balls $(n\le N)$ are randomly selected without replacement. Define the random variable X = largest number selected. Find the probability function of X.

方法1: $x=\{n,n+1,\ldots,N\}$.

$f_X(x)=P(X=x)$

$=P(\text{one}=x \wedge \text{other}<x)$

$=\underbrace{n}_{n\text{种选法}}\cdot\frac{1}{N}\cdot\frac{x-1}{N-1}\cdots\frac{x-n+1}{x-n+1}$

方法2: $|S|=\binom{N}{n}$ ← N中选n个 without replacement

Let A_x = maximum is x.

$|A_x|=\underbrace{1}_{\#\text{ ways choose } x}\cdot\binom{x-1}{n-1}$ ← 选起n-1个球

$f_X(x)=P(X=x)=\dfrac{\binom{x-1}{n-1}}{\binom{N}{n}}$

Question 1: For which $x\in\mathbb{N}$ is $f_X(x)=P(X=x)>0$? In other words, what is the range $X(S)$?

$x\le r \wedge x\le n$. $\qquad x\le\min\{r,n\}$

$n>N-r \Rightarrow n-(N-r)>0$. $\qquad x\ge\max\{0,\underbrace{n-(N-r)}_{\#\text{fail in all}}\}$.

$\therefore \max\{0,n-(N-r)\}\le x\le\min\{r,n\}$.

Question 2: Compute $f_X(x)=P(X=x)$.

N中选n的方式: $\binom{N}{n}$

x中选 r(success) 的方式: $\binom{r}{x}$

(N−r) fail 中选 (n−x) fail 的方式: $\binom{N-r}{n-x}$

$$\boxed{f_X(x)=\frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}}}$$

- discrete uniform variable $X \sim U[a,b]$

一共有事件 $a, \cdots, b$. 每件事发生概率一样, 为 $\frac{1}{b-a+1}$

$X: \{a, a+1, a+2, \cdots, b\}$. $a, b \in \mathbb{Z}$. all values are equally likely.

X has discrete uniform distribution on $\{a, a+1, \cdots b\}$. $X \sim U[a,b]$.

$$P(X=x) = \begin{cases} \frac{1}{b-a+1} & x \in \{a, \cdots, b\} \\ 0 & \text{otherwise} \end{cases}$$

$X \sim Y$.

$\forall x \in \mathbb{R} \quad F_X(x) = F_Y(x)$

- Hypergeometric distribution $X \sim Hyp(N, r, n)$

一共 N 个 item, 包含 r 个 Success, 抽取 n 个.

条件:
- 在 N 中选 n 个 without replacement.
- # success $= r$ # fail $= N - r$ ← binary responses

→ 当 N large, n small 的话, 抽中 1 个东西两次的概率极小.

则 $\frac{r}{N} = p$. $\Rightarrow$ Binomial $X \sim Bin \sim (n, p)$

$P(X \le k) \approx P(Y \le k)$

- Bernoulli distribution $X \sim Bernoulli(p)$

做一次实验. 成功概率为 p. $(X \in \{0, 1\})$

条件: either success or fail

pf.: $f(x) = p^x (1-p)^{1-x}$ $x \in \{0, 1\}$

- Binomial distribution. $X \sim Bin(n, p)$

↖ 多次 Bernoulli

已知一共 n 次实验. 未知 # S

做 n 次实验. 每次成功概率为 p.

条件:
- either success or fail
- with replacement
- for every try. $P(S) = p$ $P(F) = 1 - p$. $0 < p < 1$

pf : $f(x) = \binom{n}{x} p^x (1-p)^{n-x}$ $\quad x \in \{0, 1, \cdots, n\}$.

Q. Use binomial theorem to show the pf sums up to 1.

$$\sum_{x=0}^{n} \binom{n}{x} a^x b^{n-x} = (a+b)^n = 1.$$

$\uparrow$ $a=p \quad b=1-p$

- Negative Binomial distribution $\quad X \sim NegBin(k, p)$

实验中有k次 success. 每次 success 的概率为p.

已知k次 Success. 未知一共实验次数

条件: • either success or fail

• for every try. $P(S)=p \quad P(F)=1-p. \quad 0<p<1$

• 重复实验，直到有k次 success

p.f : $P(X=x) = P(x$ fail before the kth success$)$

$$= \binom{x+k-1}{x} p^x (1-p)^x$$

$$= \binom{x+k-1}{k-1}$$

$$\underbrace{\_\ \_\ \cdots\ \_\ \_}_{x+k}\ \underline{S}$$

- Geometric distribution $\quad X \sim Geo(p)$

实验中有1次 success. 每次 success 的概率为p.

条件: • 相当于 $X \sim NegBin(1, p)$

pf : $f(x) = \binom{x+1-1}{x} p^1 (1-p)^x = p(1-p)^x$

cdf : $F(x) = P(X \le x) = P(X \le \lfloor x \rfloor) = \sum_{k=0}^{\lfloor x \rfloor} (1-p)^k p = 1-(1-p)^{\lfloor x \rfloor+1} \quad (x \ge 0)$

Q. Show the pf sums up to 1.

$$\sum_{x=0}^{\infty} f(x) = \sum_{x=0}^{\infty} (1-p)^x p = p \sum_{x=0}^{\infty} (1-p)^x = \frac{p}{1-(1-p)} = 1$$

- Memoryless property of Geo(p)

Let $X \sim Geo(p)$ and s, t be non-negative integers, then $P(X \geq s+t \mid X \geq s) = P(X \geq t)$

proof: $P(X \geq s+t \mid X \geq s)$

$$= \frac{P(X \geq s+t \cap X \geq s)}{P(X \geq s)} = \frac{P(X \geq s+t)}{P(X \geq s)} = \frac{1-P(X \leq s+t-1)}{1-P(X \leq s-1)} = \frac{1-F(s+t-1)}{1-F(s-1)}$$

$$= \frac{(1-p)^{s+t}}{(1-p)^{s}} = (1-p)^{t} = P(X \geq t)$$

- Poisson distribution $X \sim Poi(\mu)$ $\mu > 0$

在单位时间内发生 μ 次数

p 很小 in binomial

条件:
- either success or fail
- independence n 次实验相互独立 (non-overlap intervals)
- individuality 事件都不在同一时刻发生
- homogeneity / uniformity 一个区间内事件发生概率与区间大小成正比.
- n 大 p 小

{ 代替一些 Binomial Distribution $n \to \infty$ $p \to 0$ $np \to \lambda$
{ # occurrence until t in Possion process

pf. $P(X=x) = f(x) = e^{-\mu} \cdot \frac{\mu^x}{x!}$ $(x = 0, 1, 2, \cdots)$

Q. Verify f is a valid probability function

→ Obviously, $f(x) \geq 0$ $\forall x$.

→ $\sum_{x=0}^{\infty} f(x) = \sum_{x=0}^{\infty} e^{-\mu} \frac{\mu^x}{x!} = e^{-\mu} \sum_{x=0}^{\infty} \frac{\mu^x}{x!} = e^{-\mu} e^{\mu} = 1$

exponential sum: $e^{\mu} = \sum_{x=0}^{\infty} \frac{\mu^x}{x!}$

* Poission approximation of Binomial

$$Bin(n, p) \xrightarrow{np = \mu} Poi(\mu)$$

Let $\mu = np$. $n \to \infty$, $p \to 0$.

$$\binom{n}{x} p^x (1-p)^{n-x} \longrightarrow e^{-\mu} \frac{\mu^x}{x!}$$

Example

Consider drawing a 5 card hand at random from a standard 52 card deck.

What is the probability that the hand contains at least 3 Kings?

$X \sim Hyp(52, 4, 5)$

$$f_X(3) + f_X(4) = \frac{\binom{4}{3}\binom{52-4}{5-3}}{\binom{52}{5}} + \frac{\binom{4}{4}\binom{52-4}{5-4}}{\binom{52}{5}}$$

Example

Suppose that in a weekly lottery you have probability 0.02 of winning a prize with a single ticket. If you buy 1 ticket per week for 52 weeks, what is the probability that

a) you win no prizes?

b) you win 2 or more prizes?

$$\begin{cases} W & 0.02 \\ L & 0.98 \end{cases}$$

$X \sim Bin(52, 0.02)$

a) $P(X=0) = \binom{52}{0} \times 0.98^{52} \times 0.02^{0} \approx 0.3497$

b) $P(X \geq 2) = 1 - P(X=1) - P(X=0)$

$= 1 - \binom{52}{0} \cdot 0.98^{52} \cdot 0.02^{0} - \binom{52}{1} \cdot 0.98^{51} \cdot 0.02^{1}$

≈ 0.491

Example

A bit error occurs for a given data transmission method independently in one out of every 1000 bits transferred. Suppose a 64 bit message is sent using the transmission system.

a) What is the probability that there are exactly 2 bit errors?

b) Approximate the probability in a) using a Poisson approximation.

a) $X \sim Bin(64, \frac{1}{1000})$

$P(X=2) = \binom{64}{2}\left(\frac{1}{1000}\right)^2\left(1 - \frac{1}{1000}\right)^{64-2} \approx 0.00189$

b) $X \sim Po(\frac{64}{1000})$ $\mu = np$

almost the same

$P(X=2) = e^{-\frac{64}{1000}} \dfrac{\left(\frac{64}{1000}\right)^2}{2!} \approx 0.00192$

Website hits for a given website occur according to a Poisson process with a rate of 100 hits per minute. We say a second is a "break" if there are no hits in that second.

a) What is the probability p of a break in any given second?

b) Compute the probability of observing exactly 10 breaks in 60 consecutive seconds.

c) Compute the probability that one must wait for 30 seconds to get 2 breaks.

(a) $X \sim Po\left(\frac{100}{60}\right) \rightarrow X \sim Po\left(\frac{5}{3}\right)$

$p = P(X=0) = e^{-\frac{5}{3}} \frac{\left(\frac{5}{3}\right)^0}{0!} \approx 0.189$

(b) $Y \sim Bi\left(60, \frac{5}{3}\right)$

$P(Y=10) = \binom{60}{10} \cdot \left(\frac{5}{3}\right)^{10} \cdot \left(1-\frac{5}{3}\right)^{50} = 0.124$

(c) $Z \sim NegBin(k=2, p)$

$P(Z=30) = \binom{30+2-1}{2-1} p^2 (1-p)^{30} \approx 0.002.$

At a super busy coffee chain, customers arrive at a rate of 5 per minute.

a) Find the probability that that there are more than 2 customers in one minute.

b) Suppose you record the number of customers in 5 consecutive one-minute intervals. What is the probability that in at least 3 of them there were more than 2 customers?

c) Suppose you are waiting until finally, there is one minute with more than 2 customers. Denote by X the the number of minutes you need to wait. Find the probability function of X.

d) Find the probability that a minute with more than 2 customers actually had 6 customers.

e) Suppose in 3 minutes, there were n customers. Find the probability that x of these came in the first two minutes.

a) $X \sim Po(5)$

$P(X>2) = 1-P(X\le 2) = 1-P(X=2) - P(X=1) - P(X=0)$

b) $Y \sim Bi(5, \ )$

$P(Y \ge 3) = 1-P(Y=2) - P(Y=1) - P(Y=0)$

7. Expected Value and Variance

- Sample

mean: $\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$ $x_1, \cdots, x_n$: n outcomes for a random variable X.

median: a value s.t half of results are below it. and half are above

mode: the most frequently-occuring value in a sample.

- Expected Value (mean / expantation / first moment of X)

$$E(X) = \sum_{x \in X(S)} x f(x)$$ 加权平均值

Property:

1. Law of unconscious statistician

$$E[g(X)] = \sum_{x \in X(S)} g(x) f(x)$$

2. Linearity of expectation

$$E(aX+b) = aE(X)+b \qquad E[a\,g(X)+b] = aE[g(X)]+b$$

proof: $E(aX+b) = \sum_{x \in X(S)} (ax+b) f(x)$

$$= a \underbrace{\sum_{x \in X} x f(x)}_{\downarrow} + b \underbrace{\sum_{x \in X} f(x)}_{=1}$$

$$= a\, E(X) + b.$$

$$g(E[X]) \neq E[g(X)]$$

$X \sim Bi(n,p)$	$E(X) = np$
$X \sim Po(\lambda)$	$E(X) = \lambda$
$X \sim Hyp(N,r,n)$	$E(X) = n\frac{r}{N}$
$X \sim NB(k,p)$	$E(X) = \frac{k(1-p)}{p}$
$X \sim U(a,b)$	$E(X) = \frac{a+b}{2}$

proof:

① $W \sim Bi(n,p)$ $\quad E(W)=np$

proof. $E(W)=\sum_{x=0}^{n} x\binom{n}{x}p^{x}(1-p)^{n-x}$

$$=\sum_{x=1}^{n} x\cdot\frac{n!}{(n-x)!\,x!}\,p^{x}(1-p)^{n-x}$$

$$=\sum_{x=1}^{n} \frac{n!}{(n-x)!\,(x-1)!}\,p^{x}(1-p)^{n-x}$$

$$=n\cdot p\sum_{x=1}^{n}\frac{(n-1)!}{(n-1-(x-1))!\,(x-1)!}\,p^{x-1}(1-p)^{n-1-(x-1)}$$

$$=np\underbrace{\sum_{y=0}^{n-1}\binom{n-1}{y}p^{y}(1-p)^{n-1-y}}_{=\text{ sum of p.f of } Bi(n-1,p) \;=1}$$

$$=np$$

② $Z\sim Po(\mu)$ $\quad E(Z)=\mu$

proof: $E(Z)=\sum_{z=0}^{\infty} z\, e^{-\mu}\cdot\frac{\mu^{z}}{z!}$

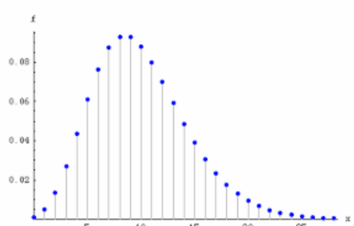
$$=\sum_{z=1}^{\infty} z\cdot e^{-\mu}\cdot\frac{\mu^{z}}{z(z-1)!}$$

$$=\mu\cdot\underbrace{\sum_{z=1}^{\infty} e^{-\mu}\cdot\frac{\mu^{z-1}}{(z-1)!}}_{=1}$$

$$=\mu$$

Interpretations of the expected value: $E(X)$

a) $E(X)$ is the "balancing point" of the probability function $f_X(x)$



b) $E(X)$ is what the average of many, many independent realizations of the random variable X would approach (Law of large numbers).

- Variance 方差

$$Var(X) = E[(X - E[X])^2]$$
$$= E(X^2) - [E(X)]^2 \qquad (Var(X) \geq 0)$$

kth moment of random variable X $\qquad E(X^k)$

Property:

① $Var(X) \geq 0$

② Var 越大，偏离越大

③ $Var(X) = 0 \iff P(X = E(X)) = 1$

proof ③:

$(\Rightarrow)$ Let $Var(X) = 0$ $\quad E(X - E(X))^2 = 0$

$\sum_{x \in X} (x - E(X))^2 f(x) = 0$ only possible if $x = E(X)$ $\forall x \in X(S)$ with $f(x) > 0$

$\therefore P(X = E(X)) = 1$

$(\Leftarrow)$ Let $P(X = E(X)) = 1$. Then

$Var(X) = E((X - E(X))^2) = \sum_{x \in X} (x - E(X))^2 f(x) = 0$

★ There exist distributions without $E(X)$:

$f_X(x) = \frac{2}{3} \cdot \frac{1}{x^2}$. $\qquad x = 1, 2, \cdots$

$E(X) = +\infty \qquad Var(X)$ not defined

$X \sim Bi(n, p)$	$Var(X) = np(1-p)$
$X \sim Po(\lambda)$	$Var(X) = \lambda$
$X \sim Hyp(N, r, n)$	$Var(X) = n\frac{r}{N}(1 - \frac{r}{N})(\frac{N-n}{N-1})$
$X \sim NB(k, p)$	$Var(X) = \frac{k(1-p)}{p^2}$
$X \sim U(a, b)$	$Var(X) = \frac{(b-a)^2}{12}$

- Standard deviation 标准差 σ s.d

$$\sigma = \sqrt{Var(X)}$$

Q Suppose the discrete random variable X has probability function

$$f(-1) = 0.15, \quad f(0) = 2c, \quad f(1) = 0.5, \quad f(2) = 0.05, \quad f(3) = c.$$

where c is a constant making f a valid probability function. Compute $E(X)$ and $E(\exp(X))$.

$\rightarrow \because \sum_{i=-1}^{3} f(i) = 1 \quad \rightarrow c = 0.1$

$\rightarrow E(X) = (-1)\cdot 0.15 + 0 \cdot 0.2 + 1 \cdot 0.5 + 2 \cdot 0.05 + 3 \cdot 0.1 = 0.75$

$\rightarrow E(\exp(X)) = \exp(-1)\cdot 0.15 + \exp(0)\cdot 0.2 + \cdots + \exp(3)\cdot 0.1 \approx 3.992$

ep. Suppose X denote the outcome of one fair six sided die roll.

$E(X) = \frac{1}{6} \times (1+2+\cdots+6) = 3.5$

Q. Let X denote the outcome of a fair six sided die roll. Compute $Var(X)$.

$E(X) = \sum_{x=1}^{6} x \cdot \frac{1}{6} = 3.5$

$E(X^2) = 1^2 \cdot \frac{1}{6} + 2^2 \cdot \frac{1}{6} + \cdots + 6^2 \cdot \frac{1}{6} = \frac{91}{6}$

$Var(X) = E(X^2) - (E(X))^2 = \frac{91}{6} - 3.5^2 = \frac{35}{12}$

Definition

The k^{th} **moment** *of a random variable X is defined by*

$$E(X^k)$$

8. Continuous Random Variable

- def. cont random variable

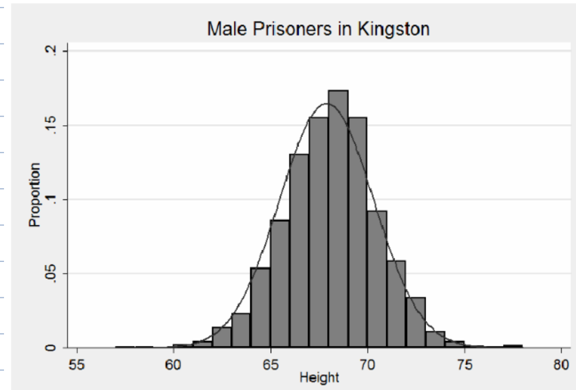
range $X(S)$ is an interval $(a,b) \in \mathbb{R}$

现实中不存在. 只能不断近似 得到

ep. 测时间 / 长度 ...

$f_X(x) = P(X=x)$

$P(a \le X \le b) = \int_a^b f(x)\,dx$



- Probability density function pdf.

X is cont. random variable with pdf $f(x)$. $g: \mathbb{R} \to \mathbb{R}$.

$$E(g(X)) = \int_{-\infty}^{\infty} g(x) f(x)\,dx$$

$$E(X) = \int_{-\infty}^{\infty} x f(x)\,dx$$

$$Var(X) = E([X-E(X)]^2) = \int_{-\infty}^{\infty} (x-E(X))^2 f(x)\,dx$$

properties:

1) $f(x) \ge 0$
2) $\int_{-\infty}^{\infty} f(x)\,dx = 1$
3) $P(a \le X \le b) = \int_a^b f(x)\,dx$

- cumulative distribution function cdf. 累积概率

$F_X(x) = P(X \le x) = \int_{-\infty}^{x} f(y)\,dy$ $\qquad \frac{d}{dx} F_X(x) = f(x)$

properties:

1. $F(x)$ is defined for all $x \in \mathbb{R}$.
2. $F(x)$ non-decreasing
3. $\lim_{x\to-\infty} F(x) = 0 \qquad \lim_{x\to\infty} F(x) = 1$
4. $P(a \le X \le b) = F(b) - F(a) = \int_a^b f(y)\,dy$

- Probability density function

$f(x) = \frac{d}{dx} F(x)$

proof:

$P(x \le X \le x+\Delta x) = F(x+\Delta x) - F(x)$

$\lim_{\Delta x\to 0} \frac{P(x \le X \le x+\Delta x)}{\Delta x} = \lim_{\Delta x \to 0} \frac{F(x+\Delta x)-F(x)}{\Delta x} = F'(x)$ 求导.

- Support of pdf.

$\mathrm{supp}(f) = \{x \in \mathbb{R} : f(x) \ne 0\}$

求 f 在 domain D 中的积分: $D \cap \mathrm{supp}(f)$

- $E(X)$ & $Var(X)$ for cts RV

pdf: f. $\quad g: \mathbb{R} \to \mathbb{R}$.

$E[g(x)] = \int_{-\infty}^{\infty} g(x) f(x)\,dx$

$E(X) = \int_{-\infty}^{\infty} x f(x)\,dx$

$Var(X) = E(X^2) - E(X)$ also holds

$Var(X) = E[(X-E(X))^2] = \int_{-\infty}^{\infty} (x-E(X))^2 f(x)\,dx$

Suppose that X is a continuous random variable with probability density function

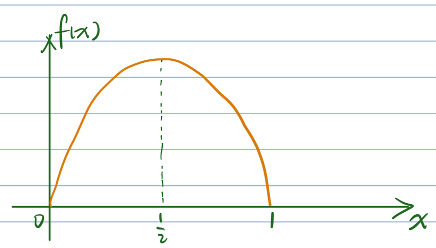
$$f(x) = \begin{cases} cx(1-x) & \text{if } 0 \le x \le 1, \\ 0 & \text{otherwise} \end{cases}$$

a) Compute c so that this is a valid pdf $\int_{-\infty}^{\infty} f = 1$
b) Graph $f(x)$
c) Compute $P(X \ge 1/2)$
d) Compute $P(1/4 \le X \le 3/4)$
e) Compute $P(X = 1/2)$

a) $\int_0^1 f(x)\,dx = \int_0^1 cx(1-x)\,dx = \frac{c}{6} = 1$

$c = 6$

b)



c) $P(X \ge \frac{1}{2}) = \int_{\frac{1}{2}}^{1} f(x)\,dx = \int_{\frac{1}{2}}^{1} 6x(1-x)\,dx = \frac{1}{2}$

d) $P(\frac{1}{4} \le X \le \frac{3}{4}) = \int_{\frac{1}{4}}^{\frac{3}{4}} f(x)\,dx = \frac{4}{16}$

e) $P(X = \frac{1}{2}) = 0$ due to $f(x)$ is cts.

Example

Suppose X has pdf

$$f(x) = \begin{cases} 6x(1-x) & \text{if } 0 \le x \le 1, \\ 0 & \text{otherwise} \end{cases}$$

Compute $E[X]$ and $Var(X)$

$E[X] = \int_0^1 x \cdot 6x(1-x)\,dx = \frac{1}{2}$

$E[X^2] = \int_0^1 x^2\, 6x(1-x)\,dx = \frac{3}{10}$

$Var[X] = E[X^2] - (E[X])^2 = \frac{3}{10} - \frac{1}{4} = \frac{1}{20}$

Random Variable

discrete

$f(x) = P(X=x)$

$P(X \in A) = \sum_{x \in S \cap A} f(x)$

$E(X) = \sum_{x \in S} x f(x)$

cts.

$P(X=x) = 0$ $\quad f(x) = F'(x)$

$P(X \in A) = \int_A f(x)\,dx$

$E(X) = \int_{-\infty}^{\infty} x f(x)\,dx$

* pdf of $f(x) \neq P(X=x)$

$\delta > 0$ 足够小.

$P(X \in (x-\frac{\delta}{2}, x+\frac{\delta}{2})) = F(x+\frac{\delta}{2}) - F(x-\frac{\delta}{2})$

$= f(x)\,\delta$

- cts uniform distribution $\quad X \sim U(a,b)$

$$f(x) = \begin{cases} \frac{1}{b-a} & x \in (a,b) \\ 0 & \text{otherwise.} \end{cases}$$

$E(X) = \frac{a+b}{2}$ $\quad Var(X) = \frac{(b-a)^2}{12}$

$$E(X) = \int_{-\infty}^{\infty} x f(x)\,dx = \int_a^b x \frac{1}{b-a}\,dx = \frac{1}{2}\cdot\frac{1}{b-a}[x^2]_a^b = \frac{a+b}{2}$$

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f(x)\,dx = \int_a^b x^2 \cdot \frac{1}{b-a}\,dx = \frac{1}{3(b-a)}(b^3-a^3)$$

$$Var(X) = E(X^2) - E(X)^2 = \frac{(b-a)^2}{12}$$

☆ If the function g has an inverse over X, then we have a easy way of obtaining the distribution of $Y = g(X)$

Suppose that the random variable X had the following pdf:

f(x) = 2x , 0 ≤ x ≤ 1; 0, otherwise.

Let Y = $\sqrt{X}$

a. Determine the pdf of Y (be sure to include the range).

① 用 $F_X(y)$ 表示 $F_Y(y)$. （从 $P(Y \le y)$ 入手）

$F_Y(y) = P(Y \le y)$

$F_Y(y) = P(\sqrt{X} \le y) = P(X \le y^2) = F_X(y^2)$

② 用 cdf of X to find cdf of y.

$$pdf_X: \ f(x) = \begin{cases} 2x & 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

$$cdf_X: \ F_X(x) = P(X \le x) = \begin{cases} 0 & x \le 0 \\ x^2 & 0 < x \le 1 \\ 1 & x > 1 \end{cases}$$

$$cdf_Y: \ F_Y(y) = F_X(y^2) = (y^2)^2 = y^4$$

③ 对 cdf. ($F_Y(y)$) 求导 得 pdf.

$$pdf_Y: \ f_Y(y) = \frac{dF_Y(y)}{dy} = 4y^3.$$

④ 找到 Y in range

$x = 0 \quad y = 0$

$x = 1 \quad y = 1 \qquad 0 \le y \le 1$

$$f_Y(y) = \begin{cases} 4y^3 & 0 \le y \le 1 \\ 0 & \text{otherwise} \end{cases}$$

Let X be a continuous random variable with pdf

$$f(x) = \lambda e^{-\lambda x} \qquad x > 0,$$

and 0 otherwise, where $\lambda > 0$ is a parameter and $c > 0$ is a constant to be determined.

a) Determine c so that f is a valid pdf.

b) Determine the cdf of X.

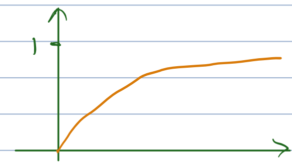
c) What distribution does the random variable $Y = e^{-\lambda X}$ have?

density function

a) $\int_{-\infty}^{\infty} f(x)\, dx = 1$.

$\int_{-\infty}^{\infty} ce^{-\lambda x} dx = c\left[-\frac{1}{\lambda} e^{-\lambda x}\right]_0^{\infty} = \frac{c}{\lambda} = 1 \qquad \Rightarrow \quad c = \lambda$

b) $F(x) = \int_{-\infty}^{x} f(t)\, dt = \int_0^x \lambda e^{-\lambda t}\, dt = 1 - e^{-\lambda x} \quad (x \geq 0)$



c) distribution of $Y = e^{-\lambda x}$

range of y : $(0, 1)$

$P(Y \leq y) = P(e^{-\lambda x} \leq y)$

$= P(-\lambda x \leq \log y)$

$= P(X \geq -\frac{1}{\lambda} \log y)$

$= 1 - P(X \leq -\frac{1}{\lambda} \log y)$

$= 1 - F(-\frac{1}{\lambda} \log y)$

$= y$.

pdf of Y : $f_Y(y) = \begin{cases} 1 & y \in (0, 1) \\ 0 & \text{otherwise} \end{cases}$

$Y \sim U(0, 1)$

- exponential distribution $X \sim Exp(\lambda)$

适用条件: Possion process with respect to time.

If a possion process has rate λ. X = waiting time between 2 ocurences.

$X \sim Exp(\lambda)$ λ : rate of occurence in Po.

occurences in $(0,t)$ follows $Po(\lambda t)$

X = time taken until 1st event occur.

pdf: $f(x) = \begin{cases} \lambda e^{-\lambda x} & x > 0 \\ 0 & x \le 0 \end{cases}$

cdf: $F(x) = \begin{cases} 1 - e^{-\lambda x} & x \ge 0 \\ 0 & \text{otherwise} \end{cases}$

proof: $\int_{-\infty}^{x} f(t)\, dt = \int_0^x \lambda e^{-\lambda t} dt = 1 - e^{-\lambda x} = P(X \le x)$

$F(x) = P(X < x)$

$= P(\text{time to 1st occurence} \le X)$

$= 1 - P(\text{time to 1st occurence} > X)$

$= 1 - P(\text{no occurence between } (0, x))$

$= 1 - \dfrac{e^{-\lambda x}(\lambda x)^0}{0!}$ $(x > 0)$ 服从Po分布

$= 1 - e^{-\lambda x}$

$f(x) = \frac{d}{dx} F(x)$

$= \lambda e^{-\lambda x}$

$X \sim Exp(\theta)$ $\theta = \frac{1}{\lambda}$: waiting time until 1st occur

$f(x) = \begin{cases} \frac{1}{\theta} e^{-\frac{x}{\theta}} & x > 0 \\ 0 & x \le 0 \end{cases}$

$E(X) = \int_0^\infty x \cdot \frac{1}{\theta} e^{-\frac{x}{\theta}} dx$

$E(X^2) = \int_0^\infty x^2 \cdot \frac{1}{\theta} e^{-\frac{x}{\theta}} dx$.

- Gamma function

$$\Gamma(\alpha) = \int_0^\infty y^{\alpha-1} e^{-y}\, dy \quad (\alpha > 0)$$

property : 1. $\Gamma(\alpha) = (\alpha-1)\Gamma(\alpha-1) \quad (\alpha > 1)$

2. $\Gamma(\alpha) = (\alpha-1)! \quad (\alpha \in N)$

3. $\Gamma(\frac{1}{2}) = \sqrt{\pi}$

proof : 1. $\int_0^\infty y^{\alpha-1} e^{-y}\, dy = -\lim_{y\to\infty} y^{\alpha-1} e^{-y} + (\alpha-1)\int_0^\infty y^{\alpha-2} e^{-y}\, dy$ 分部积分

$\because \alpha > 1 \quad \lim_{y\to\infty} y^{\alpha-1} - e^{-y} = 0$

$\therefore \int_0^\infty y^{\alpha-1} e^{-y}\, dy = (\alpha-1)\int_0^\infty y^{\alpha-2} e^{-y}\, dy = (\alpha-1)\Gamma(a-1)$

2. $\Gamma(2) = 1\,\Gamma(1) = 1$ induction

$\Gamma(3) = 2\Gamma(2) = 2!$

$\Gamma(4) = 3\Gamma(3) = 3!$

$\vdots$

$\Gamma(n+1) = n!$

3. $E(X) = \int_{-\infty}^\infty x f(x)\, dx$

$= \int_0^\infty x \cdot \frac{1}{\theta} e^{-\frac{x}{\theta}}\, dx \quad$ let $y = \frac{x}{\theta} \quad dx = \theta\, dy$

$= \int_0^\infty y e^{-y} \theta\, dy$

$= \theta.$

$\boxed{\text{If } X \sim Exp(\theta), \text{ then } E(X) = \theta \quad Var(X) = \theta^2.}$

proof : $E(X) = \int_0^\infty x \cdot \frac{1}{\theta} e^{-\frac{x}{\theta}}\, \underline{dx}$

$= \theta \int_0^\infty y^{2-1} e^{-y}\, \underline{dy} \quad (\text{Let } y = \frac{x}{\theta}. \quad dx = \theta\, dy)$

$= \theta \cdot \Gamma(2) = \theta \cdot (2-1)! = \theta$

$E(X^2) = \int_0^\infty x^2 \cdot \frac{1}{\theta} e^{-\frac{x}{\theta}}\, dx = \int_0^\infty \theta^2 y^{3-1} e^{-y}\, dy. = \theta^2 \Gamma(3) = 2\theta^2$

$Var(X) = E(X^2) - (E(X))^2 = 2\theta^2 - \theta^2 = \theta^2$

Example

Nupur decided to enjoy a relaxing Summer away from student housing, so he rented a place in Simcoe, Ontario. However, the busses there are far and few between. Suppose busses arrive according to a Poisson process with an average of 3 busses per hour.

a) Find the probability of waiting at least 15 minutes.

b) Find the probability of waiting at least another 15 minutes given that you have already waiting for 6 minutes.

a) rate $= \lambda = 3$ bus/h $= \frac{1}{20}$ bus/min

X = waiting time until 1st bus

$X \sim Exp(\lambda = \frac{1}{20})$

$P(X>15) = 1 - P(X \le 15) = 1 - (1 - e^{-\frac{1}{20}\cdot 15}) = e^{-\frac{3}{4}}$

b) $P(X>21 \mid X>6) = \frac{P(X>21)}{P(X>6)} = \frac{e^{-\frac{21}{20}}}{e^{-\frac{6}{20}}} = e^{-\frac{3}{4}}$

Example

Exponential distribution is also very useful in reliability engineering. The lifetime of a seat belt motor on a 1994 Saturn GL is known to follow an exponential distribution with mean 14 years. $\theta = 14$

a) What is the standard deviation of the lifetime of a seat belt motor on a 1994 Saturn GL?

b) Compute the probability that the lifetime of the seat belt motor will last more than 6 years.

c) If a seat belt motor has lasted 14 years, what is the probability that it will last another 6 years?

$X \sim Exp(\theta = 14)$

a) $\sigma = \sqrt{\sigma^2} = \sqrt{\theta^2} = 14$

b) $P(X>6) = e^{-\frac{6}{14}}$

c) $P(X>20 \mid X>14) = \frac{e^{-\frac{20}{14}}}{e^{-\frac{14}{14}}} = e^{-\frac{6}{14}}$

- Memoryless property of Exp(θ)

$P(X>s+t \mid X>s) = P(X>t)$

Prove: If $X \sim Exp(\lambda)$. $s,t>0$, then $P(X>s+t \mid X>s) = P(X>t)$

proof: $\because X \sim Exp(\lambda)$

$\therefore F(x) = P(X \le x) = \int_0^x \lambda e^{-\lambda t}\, dt = 1-e^{-\lambda t} \quad x \ge 0$

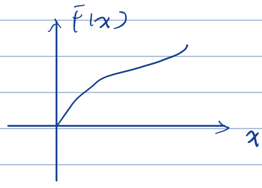
$$P(X>t+s \mid X>s) = \frac{P(X>t+s \wedge X>s)}{P(X>s)}$$
$$= \frac{P(X>t+s)}{P(X>s)}$$
$$= \frac{1-F(t+s)}{1-F(s)}$$
$$= e^{-\lambda t}$$
$$= 1-F(t)$$
$$= P(X>t)$$

- Percentile

def. The $100 \times q$th percentile of the distribution of X with cdf F_X is C_q.

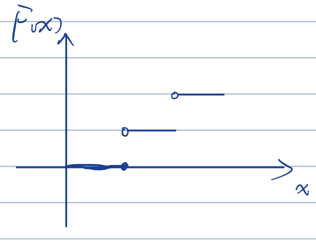
s.t $F_X(C_q) = q$

* median   5% quantile
* mode   "most likely"
* mean   average



case 1: F strictly increasing & cts.

$F(S_q) = q \qquad S_q = F^{-1}(q)$



case 2: F jumps / flat.

$F^{-1}(y) = \inf_{x \in \mathbb{R}} \{F(x) \ge y\}$

Example

Suppose $X \sim Exp(\theta = 5)$, calculate the ① 25th percentile and the ② median of the distribution of X.

pdf. $f(x) = \frac{1}{5} e^{-\frac{x}{5}} \quad (x > 0)$.

cdf. $F(x) = \int_{-\infty}^{x} f(t)\,dt = 1 - e^{-\frac{x}{5}}$

$F(c_q) = q \qquad q \in (0,1)$

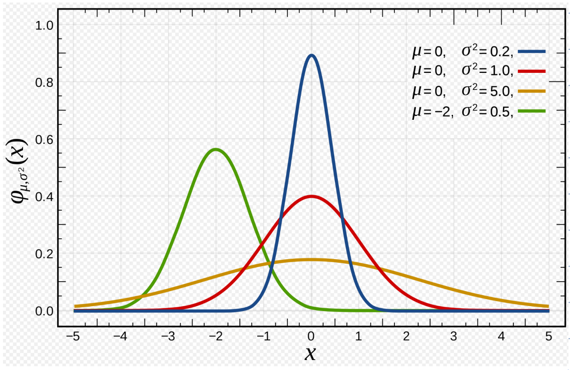
$1 - e^{-\frac{c_q}{5}} = q$

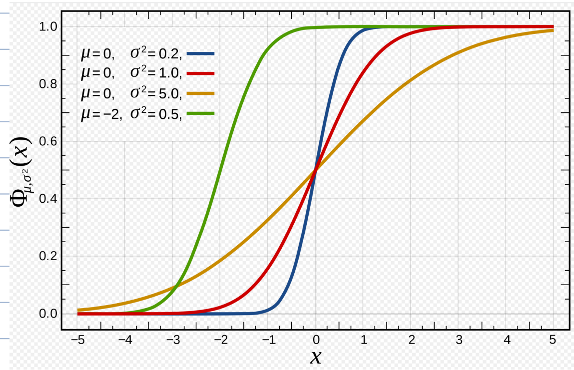
$c_q = -5 \log(1-q)$

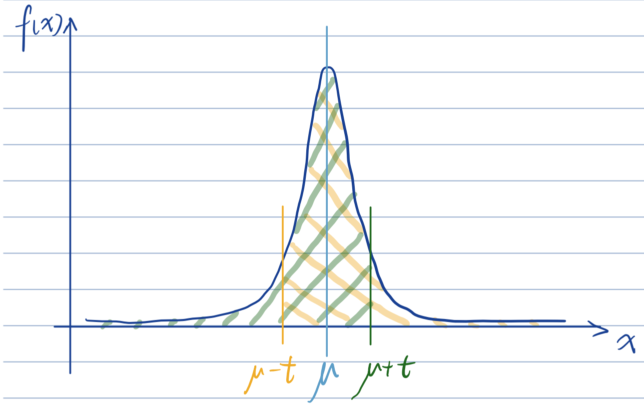
① $q = 0.25 \qquad c_{0.25} = -5 \log(0.75) \approx 1.438$

② $q = 0.5 \qquad c_{0.5} = -5 \log(0.5) \approx 3.466$

– Normal distribution







- standardising normal random variable

$X \sim N(\mu, \sigma^2)$, $Z = \frac{X-\mu}{\sigma} \sim N(0,1)$ → z-score of x.

unit free standardized

$E(X) = \mu \qquad Var(X) = \sigma^2$

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \qquad x \in \mathbb{R}$$

$$F(x) = P(X \le x) = P\left(Z \le \frac{x-\mu}{\sigma}\right)$$

$$P(a \le X \le b) = \int_a^b \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \phi\left(\frac{X-\mu}{\sigma}\right)$$

通过normalizing得到

- standard normal random variable : $X \sim N(0,1)$ 标准正态分布

$f(x)$: $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ → Since $\int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx = \sqrt{2\pi}$

$F(x)$: $\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy$ → percentile

– Properties

1. symmetric

$$\begin{cases} \text{symmetric about } \mu : P(X>\mu+t) = P(X<\mu-t) \\ \text{symmetric about } \mu=0 \text{ (mean)}: P(Z\le -x) = \phi(-x) = 1-\phi(x) \end{cases}$$

unimodal: (① one mode ② peaks at μ)

2. μ

$$\begin{cases} \mu\uparrow. \ \text{curve 向右移} \\ \mu\downarrow. \ \text{curve 向左移} \end{cases}$$

3. σ^2

σ^2越高代表 curve越分散

Example

Compute the

a) 75th percentile of the standard normal distribution

b) 58th percentile of the $N(5,9)$ distribution

c) Let $Z \sim N(0,1)$. Find c such that

$$P(-c \le Z \le c) = 0.95$$

a) $\phi^{-1}(0.75) = \int_{-\infty}^{0.75} \frac{1}{\sqrt{2\pi}} e^{-0.75^2} dy$

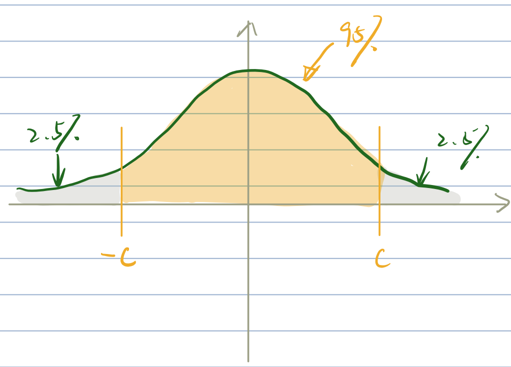
b) $5+\sqrt{9}\cdot \phi(0.58) = 5+\sqrt{9}\int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dy$

$= 5\times 3\times 0.2019$

c) $P(-c\le Z\le c) = \phi(c)-\phi(-c)$

$= \phi(c) - (1-\phi(c))$

$= 2\phi(c) - 1 \quad = 0.95$



$\phi(c) = 0.975$

$c = \phi^{-1}(0.975) = 1.96$

9. Multivariate distributions 多元分布

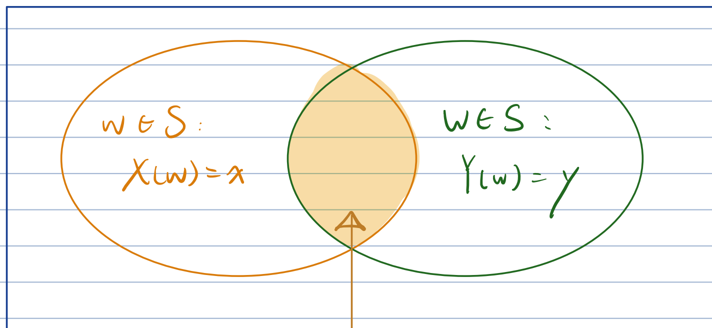
- def. joint probability function of X & Y

Suppose X & Y are discrete random variables defined on the same sample space.

The joint probability function of X & Y is

$$f(x,y) = P(\{X=x\} \cap \{Y=y\}) \qquad x \in X(S) \quad y \in Y(S)$$

简写成: $f(x,y) = P(X=x,\ Y=y)$



$P(X=x \wedge Y=y)$

What does a 2-variable probability distribution look like?

Let $X \in \{1,2,3\}$ and $Y \in \{1,2\}$, and suppose that every outcome of (X,Y) is equally likely.

$$f(x,y) = P(X=x \wedge Y=y) \begin{cases} = \frac{1}{6} & x \in \{1,2,3\} \wedge y \in \{1,2\} \\ 0 & ow \end{cases}$$

a table of joint probability distribution of (X,Y):

$f(x,y)$	$x=1$	$x=2$	$x=3$
$y=1$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$
$y=2$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

– Properties of joint probability function

1) $f(x,y) \geq 0$ 这是废话

2) $\sum_{x,y} f(x,y) = 1$ 这也是，因为 joint probability func 属于 probability func

– marginal probability function 边际概率

Suppose X & Y are discrete random variables with joint probability func $f(x,y)$

The marginal probability function of X is $f_X(x) = P(X=x) = \sum_{y \in Y(S)} f(x,y)$

The marginal probability function of Y is $f_Y(y) = P(Y=y) = \sum_{y \in Y(S)} f(x,y)$

这俩玩意儿加和没啥意义
乘积也没意义

* marginal distribution 与 probability distribution 是一个东西

Example

Suppose X and Y have joint probability function

$$f(x,y) = \frac{1}{6}\left(\frac{1}{2}\right)^x \left(\frac{2}{3}\right)^y, \quad x, y = 0, 1, 2 \ldots$$

Compute the marginal probability functions $f_X(x)$ and $f_Y(y)$.

marginal probability of X : $f_X(x) = \sum_{y=0}^{\infty} \frac{1}{6}\left(\frac{1}{2}\right)^x \left(\frac{2}{3}\right)^y$

$= \frac{1}{6}\left(\frac{1}{2}\right)^x \sum_{y=0}^{\infty} \left(\frac{2}{3}\right)^y$

$= \frac{1}{6}\left(\frac{1}{2}\right)^x \frac{1}{1-\frac{2}{3}}$ $\quad \left(\sum_{k=0}^{\infty} q^k = \frac{1}{1-q}\right)$ $0<q<1$

$= \frac{1}{2} \cdot \left(\frac{1}{2}\right)^x$

marginal probability of Y : $f_Y(y) = \frac{1}{3}\left(\frac{2}{3}\right)^y$

- independent random variable

If $f(x,y) = f_X(x) f_Y(y)$

then X & Y are independent random variable.

若 $X_1, X_2, \cdots, X_n$ independent, 则 $f(x_1 x_2 \cdots x_n) = f(x_1) f(x_2) \cdots f(x_n)$

Q. Suppose $X \sim Poi(2)$, $Y \sim Poi(3)$, and that X and Y are independent. What is the joint probability function of X and Y?

$$f(x,y) = f_X(x) \cdot f_Y(y) = e^{-2} \cdot \frac{2^x}{x!} \cdot e^{-3} \cdot \frac{3^y}{y!}$$

- Conditional distributions.

对 event A & B $P(B) \neq 0$. $P(A|B) = \frac{P(A \cap B)}{P(B)}$

The conditional probability function of X given $Y=y$ $f_Y(y) > 0$ is:

$$f_X(x|y) = P(X=x \mid Y=y) = \frac{P(X=x, Y=y)}{P(Y=y)} = \frac{f(x,y)}{f_Y(y)}$$

Example

Suppose a fair coin is tossed 3 times. Define the random variables $X =$ "number of Heads", and

$$Y = \begin{cases} 1 & \text{Head occurs on the first toss,} \\ 0 & \text{Tail occurs on the first toss.} \end{cases}$$

Find $f_X(x|y)$ for $y = 0, 1$.

		X				
		0	1	2	3	$\Sigma P(Y=y)$
y.	0	$\frac{1}{8}$	$\frac{2}{8}$	$\frac{1}{8}$	0	$\frac{1}{2}$
	1	0	$\frac{1}{8}$	$\frac{2}{8}$	$\frac{1}{8}$	$\frac{1}{2}$
$\Sigma P(X=x)$		$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$	1

	$x=0$	$x=1$	$x=2$	$x=3$
$f_X(x\|y=0)$	$\frac{\frac{1}{8}}{\frac{1}{2}} = \frac{1}{4}$	$\frac{\frac{3}{8}}{\frac{1}{2}} = \frac{1}{2}$	$\frac{1}{4}$	0
$f_X(x\|y=1)$	0	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

我们在意的几个事件合成的整体.

做题时可以把几个小事件当作一个大事件.

$$P(U=u) = \sum_{\substack{(x_1,\dots,x_n) \\ \text{s.t. } g(x_1,\dots,x_n)=u}} f(x_1,\dots,x_n)$$

- Sum of independent

Possion.	$X+Y \sim Po(\lambda_1+\lambda_2)$
Binomial	$X+Y \sim Bi(n+m, p)$
Bernoulli	$X_1+X_2+\cdots+X_n \sim Bi(n,p)$
Geometric	$X_1+X_2+\cdots+X_k \sim NB(k,p)$

- Theorem

$X \sim Po(\lambda_1)$ $\quad Y \sim Po(\lambda_2)$ $\quad X \mid X+Y=n \sim Bi\left(n, \frac{\lambda_1}{\lambda_1+\lambda_2}\right)$

proof: Let $X \sim Po(\lambda_1)$ $\quad Y \sim Po(\lambda_2)$ $\quad X \& Y$ indep.

Then $X+Y \sim Po(\lambda_1+\lambda_2)$

$$P(X=x \mid X+Y=n) = \frac{P(X=x, X+Y=n)}{P(X+Y=n)}$$

$$= \frac{P(X=x, Y=n-x)}{P(X+Y=n)}$$

$$= \frac{e^{-\lambda_1}\cdot\frac{\lambda_1^x}{x!}\cdot e^{-\lambda_2}\frac{\lambda_2^{n-x}}{(n-x)!}}{e^{-(\lambda_1+\lambda_2)}\frac{(\lambda_1+\lambda_2)^n}{n!}} \quad (\text{Since } X \& Y \text{ indep})$$

$$= \frac{n!}{(n-x)!\,x!}\left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^x\left(\frac{\lambda_2}{\lambda_1+\lambda_2}\right)^{n-x}$$

$$= \text{p.f of } Bi\left(n, \frac{\lambda_1}{\lambda_1+\lambda_2}\right) \text{ at } x.$$

☆ roll a die n times.

掷到数字:	1	2	3	4	5	6
次数	X_1	X_2	X_3	X_4	X_5	X_6

$X_j = Bi(n, \frac{1}{6})$ ☆ not independent. 因为 $X_1+\cdots+X_6=n$

- Multinomial distribution

def: multinomial distribution

an experiment satisfy

① individual trials have k possible outcomes.

p_i: probability of each individual outcome $p_1 + \cdots + p_k = 1$

② trials are independently repeated n times.

X_i: denoting the number of times outcome i occurred $X_1 + \cdots + X_k = n$

$X_1, \cdots, X_k$ have multinomial distribution with n & $p_1 \cdots p_k$

def. multinomial coefficients. $(X_1, \cdots, X_k) \sim \text{Mult}(n, p_1, \cdots, p_k)$

If $X_1, \cdots, X_k$ have a joint multinomial distribution with parameters n and $p_1 \cdots p_k$. then their joint probability function is $f(x_1, \cdots, x_k) = \underbrace{\frac{n!}{x_1! \cdots x_k!}}_{\text{multinomial coefficients}} p_1^{x_1} \cdots p_k^{x_k}$

$x_1 \cdots x_k$ satisfy $x_1 + \cdots + x_k = n$ $x_i \geq 0$

$x_1 + \cdots + x_k = n$

Q. Consider drawing 5 cards from a standard 52 card deck of playing cards (4 suits, 13 kinds) **with replacement**. What is the probability that 2 of the drawn cards are hearts, 2 are spades, and 1 is a diamond?

$(H, S, D, C) \sim \text{Mult}(5, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})$

$P(H=2, S=2, D=1, C=0) = \frac{5!}{2!2!1!0!} (\frac{1}{4})^5$

- Marginal distribution of multinomial

Let $(X_1, X_2, \cdots, X_k) \sim \text{Mult}(n, p_1, \cdots, p_k)$,

Then $X_j \sim Bi(n, p_j)$ $j \in 1, \cdots, k$

- Conditional distribution of multinomial

Let $(X_1, X_2, \cdots, X_k) \sim \text{Mult}(n, p_1, \cdots, p_k)$,

Then $X_i | X_i + X_j = t \sim Bi(t, \frac{p_i}{p_i + p_j})$ $(i \neq j)$

Q. We can model n rounds of fair, independent rock-paper-scissors game using multinomial distribution:

$$(R, P, C) \sim Mult(n, 1/3, 1/3, 1/3).$$

Suppose that I play 5 games of R-P-S. Given that the sum of Rocks and Papers is 4, what would be the distribution of the number of Rocks I played?

$R \mid R+P=4 \quad \sim Bi\left(4, \frac{\frac{1}{3}}{\frac{1}{3}+\frac{1}{3}}\right)$

- Sum of individual rvs in multinomial

Let $(X_1, X_2, \cdots, X_k) \sim Mult(n, p_1, \cdots, p_k)$,

Then $X_i + X_j \sim Bi(n, p_i + p_j) \quad (i \neq j)$

在 multinomial distribution 中，事件是独立的，但 marginal random variables 不是.

- Joint expectations

- def.

Suppose X & Y are discrete random variables with joint probability function $f(x, y)$

Then for a function $g: \mathbb{R}^2 \to \mathbb{R}$. $\quad E[g(X,Y)] = \sum_{(x,y)} g(x,y) f(x,y)$

Example

Suppose X and Y have joint probability function given by the following table:

	f(x,y)	x = 0	x = 1	x = 2
y	0	.2	.3	.1
	2	.25	.13	.02

Compute E$[XY]$ and E$[Y]$. Compute $Cov(X, Y)$.

$E(Y) = 0 \cdot 0.6 + 2 \cdot 0.4 = 0.8$

$E(X \cdot Y) = 0 \cdot 0 \cdot 0.2 + \cdots + 1 \cdot 2 \cdot 0.13 + \overset{x}{\underline{2}} \cdot \overset{y}{\underline{2}} \cdot \overset{f(x,y)}{\underline{0.02}} = 0.34$

$cov(X, Y) = E(XY) - E(X)E(Y) = 0.4 - 0.8 \cdot 0.67 = -0.446$

property:

1. $E[a g_1(X,Y) + b g_2(X,Y)] = a\cdot E[g_1(X,Y)] + b\cdot E[g_2(X,Y)]$.

2. $E(X+Y) = E(X) + E(Y)$ 　　　 $\ast$ $E(X\cdot Y) \neq E(X)E(Y)$

proof: 1.

Let $g(x,y) = a g_1(X,Y) + b\cdot g_2(X,Y)$

then $E(a\cdot g_1(X,Y) + b_2 g_2(X,Y))$

$= \sum_{(x,y)} (a g_1(x,y) + b g_2(x,y)) f(x,y)$ 　 (by $E(g(X,Y)) = \sum_{(x,y)} g(x,y) f(x,y)$)

$= a \sum_{(x,y)} g_1(x,y) f(x,y) + b \sum_{(x,y)} g_2(x,y) f(x,y)$

Example

Let $(X_1, X_2, X_3) \sim Mult(n, p_1, p_2, p_3)$. Then

$$E[X_1X_2] = n(n-1)p_1p_2.$$

proof: $E(X_1) = np_1$
$E(X_2) = np_2$

- def. covariance.

If X & Y are jointly distributed, then $cov(X,Y)$ denotes the covariance between X & Y.

$cov(X,Y) = E[X-E(X)][Y-E(Y)]$.

$= E(XY) - E(X)E(Y)$

ep. X temperature in trt. Y temperature in loo.

$\begin{cases} 若\ X > E(X),\ 可能\ Y > E(Y). \\ 若\ X < E(X),\ 可能\ Y < E(Y). \end{cases}$

$\Rightarrow (X - E(X))(Y - E(Y)) > 0$

- Theorem

If X & Y are independent, then $cov(X,Y)=0$

* 若 $cov(X,Y)=0$, X & Y not necessarily indep.

counter example: $Y=X^2-1$

proof: Let X, Y be indep. Then $f(x,y)=f_X(x)f_Y(y)$

$$E(X\cdot Y)=\sum_{(x,y)} x\cdot y\cdot f(x,y)$$
$$=\sum_{(x,y)} x\cdot y\cdot f_X(x)\cdot f_Y(y)$$
$$=\left(\sum_x x f_X(x)\right)\left(\sum_y y f_Y(y)\right)$$
$$=E(X)E(Y)$$

$$cov(X,Y)=E(XY)-E(X)E(Y)=E(X)E(Y)-E(X)E(Y)=0$$

- correlation $corr(X,Y)$

$$corr(X,Y)=\rho=\frac{cov(X,Y)}{\sigma(X)\sigma(Y)}$$

uncorrelated: $cov(X,Y)=0$. / $corr(X,Y)=0$

* X & Y indep $\Rightarrow$ X & Y uncorrelated.
* $cov(X,X)=Var(X)$
* $corr(X,Y)$ 没有单位

Property:
1. $\rho=corr(X,Y)$ 与 $cov(X,Y)$ 正负性相同
2. $-1\le\rho\le 1$
3. $|\rho|=1 \Leftrightarrow X=aY+b$

→ 通过 Cauchy-Schwarz inequality 证明

4. X, Y indep $\Rightarrow corr(X,Y)=0$
5. $corr(X,X)=\frac{cov(X,X)}{\sigma^2(X)}=1$

若 $\sigma(X)=0$ 或 $E(X)$. $E(Y)$. $Var(X)$. $Var(X)$ 中有 not well defined or.
则 $corr(X,Y)$ 不成立

- def. linear combination

Suppose $X_1, \cdots, X_n$ are joint distributed random variables with joint probability function $f(x_1, \cdots, x_n)$.

A linear combination of the RVs $X_1 \cdots X_n$ is any random variable of the form
$\sum_{i=1}^{n} a_i X_i$ $\quad (a_1, \cdots, a_n \in \mathbb{R})$

$X = (X_1 \cdots X_n)^T \quad a = (a_1 \cdots a_n)^T$. linear combination: $X^T a$.

* total $\quad T = \sum_{i=1}^{n} X_i, \quad a_i = 1$
* sample mean $\quad \bar{X} = \sum_{i=1}^{n} \frac{1}{n} X_i, \quad a_i = \frac{1}{n}$
* expected value $\quad E(\sum_{i=1}^{n} a_i X_i) = \sum_{i=1}^{n} a_i E(X_i)$

Example

Let $P_1, P_2, \ldots, P_7$ represent the number of cans of pop that Harold drinks each day (from day 1 to day 7). If each random variable P_i has mean $\mu = 6$, what is the expected number of cans consumed per day during those 7 days?

$$E\left(\frac{P_1 + \cdots + P_7}{7}\right) = \frac{1}{7} \sum_{i=1}^{7} E(P_i) = \frac{1}{7} \cdot \sum_{i=1}^{7} 6 = \frac{1}{7} \times 7 \times 6 = 6.$$

Example

Suppose $X \sim N(1, 1)$ and $Y \sim U(0, 1)$. Compute $E(2X - 4Y)$.

$$E(2X - 4Y) = 2E(X) - 4E(Y) = 2 \times 1 - 4 \times \frac{1}{2} = 0.$$

- Variance of linear combination

Let X, Y be random variables, $a, b \in \mathbb{R}$.

$$Var(aX + bY) = a^2 Var(X) + b^2 Var(Y) + 2ab\, Cov(X, Y).$$

$$Var(\sum_{i=1}^{n} a_i X_i) = \sum_{i=1}^{n} a_i^2 Var(X_i) + 2 \sum_{i<j\le n} a_i a_j\, Cov(X_i, X_j)$$

$$Var(aX + bY) = Cov(aX + bY, aX + bY)$$
$$= a \cdot a\, Cov(X, X) + a \cdot b\, Cov(X, Y) + b \cdot a\, Cov(X, Y) + b \cdot b\, Cov(Y, Y)$$
$$= a^2 Var(X) + 2ab\, Cov(X, Y) + b^2 Var(Y)$$

What about the distribution of linear combination?

So far, $E(X)$ & $Var(X)$ of linear combination.

We have seen some special cases for the sum $X + Y$:

- Sum of independent Poisson is Poisson
- Sum of independent Binomial with same p is again Binomial
- Sum of independent Hypergeometric with same p is again Hypergeometric
- ...

- Theorem: a linear function of normal is normal.

Let $X \sim N(\mu, \sigma^2)$ $\quad Y = aX + b.$ $\quad Y \sim N(a\mu + b, a^2\sigma^2)$

Let $X_i \sim N(\mu_i, \sigma_i^2)$ $\quad \sum_{i=1}^{n} a_i X_i + b_i \sim N\left(\sum_{i=1}^{n} a_i\mu_i + b_i, \sum_{i=1}^{n} a_i^2\sigma_i^2\right)$

- Theorem: Sample mean of normal is normal.

Let $X_i \sim N(\mu, \sigma^2)$ $\quad i = 1 \cdots n$ independently.

Then $\sum_{i=1}^{n} X_i \sim N(n\mu, n\sigma^2)$

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

Example

Suppose that $X_i \sim N(\mu, \sigma^2)$, and that $X_1, \ldots, X_n$ are independent. Compute $Var\left(\sum_{i=1}^{n} X_i\right)$ and $Var(\bar{X})$.

$$Var\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} Var(X_i) = \sum_{i=1}^{n} r^2 = n \cdot r^2.$$

$$Var(\bar{X}) = Var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2} \cdot n \cdot r^2 = \frac{r^2}{n}$$

Example

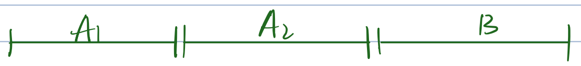
Three cylindrical parts are joined end to end to make up a shaft in a machine: 2 type-A parts and 1 type-B part. The lengths of the parts vary a little, and have the following distributions:

$$A \sim N(6, 0.4), \quad B \sim N(35.2, 0.6).$$

The overall length of the assembled shaft must lie between 46.8 and 47.5 or else the shaft has to be scrapped. Assume the lengths of different parts are independent.

a) What percent of assembled shafts have to be scrapped?

b) What is the scrapping percentage if we reduce the variance of A and B by 50% each?



a) Let L = total length $= A_1 + A_2 + B$.

$A_1 \sim N(6, 0.4)$ $\quad A_2 \sim N(6, 0.4)$ $\quad B \sim N(35, 0.6)$

$L \sim N(6+6+35,\ 0.4+0.4+0.6) \sim N(47.2, 1.4)$

$$\begin{aligned} P(\text{scrapped}) &= P(L < 46.8) + P(L > 47.5) \\ &= P\left(Z < \frac{46.8-47.2}{\sqrt{1.4}}\right) + P\left(Z > \frac{47.5-47.2}{\sqrt{1.4}}\right) \\ &= \phi(-0.37) + (1 - \phi(0.27)) \\ &\approx 0.749. \end{aligned}$$

b) $L' \sim N(47.2, 0.7)$ $\qquad P(L' < 46.8) + P(L' > 47.5)$

- def. indicator. (Bernoulli) variable

Let $A \subset S$ be an event. We say $\mathbb{1}_A$ is the indicator random variable of the event A.

$\mathbb{1}_A$ is defined by: $\mathbb{1}_A(\omega) = \begin{cases} 1 & \omega \in A \\ 0 & \omega \notin A. \end{cases}$

同时 $\mathbb{1}_A$ 也定义 Bernoulli random variable

property:

1. $E(\mathbb{1}_A) = P(A)$
2. $Var(\mathbb{1}_A) = P(A)(1-P(A))$
3. $cov(\mathbb{1}_A, \mathbb{1}_B) = P(A \cap B) - P(B)P(B)$

proof: a) $E(\mathbb{1}_A) = 1 \cdot P(A) + 0 \cdot P(A^c) = P(A)$

b) $E(\mathbb{1}_A^2) = 1^2 P(A) + 0 P(A^c) = P(A)$

$Var(\mathbb{1}_A) = P(A) - P(A)^2 = P(A)(1-P(A))$

c) $E(\mathbb{1}_A \mathbb{1}_B) = 1\, P(A \cap B)$.

$cov(\mathbb{1}_A \mathbb{1}_B) = E(\mathbb{1}_A \mathbb{1}_B) - E(\mathbb{1}_A)E(\mathbb{1}_B) = P(A \cap B) - P(A)P(B)$

Example

Let $p = p(A)$. At which value of p is $Var(\mathbb{1}_A)$ maximized?

$p \cdot (1-p)$ max at $p = \frac{1}{2}$

所以这是 variance?

$\Rightarrow$ $X \sim Bi(n,p)$ $\quad E(X) = np$. $\quad Var(X) = np(1-p)$

proof: Let A_i = "success in trial i" $\quad p = P(A_i)$

Then $X = \sum_{i=1}^{n} \mathbb{1}_{A_i}$ $\quad \mathbb{1}_{A_1}, \cdots, \mathbb{1}_{A_n}$ are indep.

$E(X) = E(\sum_{i=1}^{n} \mathbb{1}_{A_i}) = \sum_{i=1}^{n} E(\underbrace{\mathbb{1}_{A_i}}_{p}) = np$.

Similarly, $Var(X) = Var(\sum_{i=1}^{n} \mathbb{1}_{A_i}) = \sum \underbrace{Var(\mathbb{1}_{A_i})}_{p(1-p)} = np(1-p)$

10. Central Limit theorem & Moment generating functions

- Central Limit theorem.

Suppose $X_1, \cdots, X_n$ are independent random variables, each with a common cumulative distribution function F.

Suppose further that $E(X_i) = \mu$. $Var(X_i) = \sigma^2 < \infty$.

Then for all $x \in \mathbb{R}$. $n \to \infty$ $P\left(\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \le x\right) \to \phi(x)$.

若 n 足够大. 则 $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$ $\sum_{i=1}^{n} X_i \sim N(n\mu, n\sigma^2)$

As long as we have:

① a set of independent and identically distributed random variables

② a finite common mean μ and finite common variance.

The distribution of their sample mean can be approximated by normal distribution.

random variable 不 normal, sample mean 可以 normal.

Example

Jason rolls a six sided die 1000 times, and records the results. If the die is a fair die, estimate the probability that the sum of the die rolls is less than 3400.

· Let $X_1, \cdots, X_{1000}$ denote the outcomes of the 1st, 2nd, $\cdots$, 1000th die roll and note that they are independent discrete uniforms on $\{1, 2, \cdots, 6\}$ with $\mu = 3.5$, $\sigma^2 = \frac{35}{12}$.

· By CLT. $\sum_{i=1}^{n} X_i \sim N(1000 \cdot \mu, 1000 \cdot \sigma^2)$

$$P\left(\sum_{i=1}^{n} X_i \le 3400\right) = P\left(\frac{\sum X_i - 1000 \cdot \mu}{\sqrt{1000\sigma^2}} \le \frac{3400 - 1000\mu}{\sqrt{1000\sigma^2}}\right)$$

$$\approx P(Z \le -1.85)$$

$$\approx 0.032.$$

- Theorem.

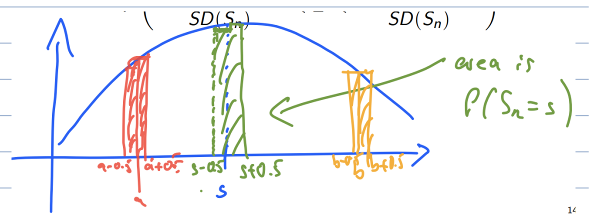
$X_n \sim Bi(n,p)$ n large, $\dfrac{X_n - np}{\sqrt{np(1-p)}} \sim N(0,1)$

$X_\lambda \sim Po(\lambda)$ λ large, $\dfrac{X_\lambda - \lambda}{\sqrt{\lambda}} \sim N(0,1)$

How to correct for continuity?

$P(a \le S_n \le b) \approx P(a-0.5 < S_n < b+0.5)$

$P(S_n = s) = P(s-0.5 < S_n < s+0.5)$



$$\approx P\left(\frac{(s-0.5)-\mu_{S_n}}{\sigma(S_n)} < Z < \frac{(s+0.5)-\mu_{S_n}}{\sigma(S_n)}\right)$$

CLT in 检验方法：（中心极限定理）

1. 观测数据足够多，CLT能提供合理的近似值。
2. 若观测值的分布是 ⌒，则对于较小的 n (5-15)，CLT也适用。
3. 若分布高度偏离(skew)或离散(discrete)，则常需要更大的 n 值 (n>50)
4. 当 approximating normal distribution 时，我们不必使用 continuity correlation

- Moment generating functions (MGF)

functions that can define distribution of RV:

1) pf. $f(x)$

2) cdf. $F(x)$

3) MGF. $M_X(t) = E[e^{tX}] = \sum_{x \in X(S)} e^{tx} f(x)$ $t \in \mathbb{R}$.

properties of mgf.

1) $M_X(0) = 1$

2) $M_X(t) = \sum_{j=0}^{\infty} \dfrac{t^j E[X^j]}{j!}$

3). $M_X(t)$ is defined in a neighborhood of $t=0$.

$\dfrac{d}{dt^k} M_X(0) = E[X^k]$

- Continuity Theorem

If X & Y have MGF. $M_X(t)$ & $M_Y(t)$ defined in neighbourhoods of the origin.
$M_X(t) = M_Y(t)$ for all t where they are defined, then $X \stackrel{D}{=} Y$.

↳ MGF uniquely characterizes a distribution

P3 $M_X(t) = E(e^{tX}) = e^{t\cdot 1} P(X=1) + e^{t\cdot 0} P(X=0)$
$= pe^t + (1-p)$

$M_X'(t) = pe^t \longrightarrow E(X) = M_X'(0) = p$

$M_X''(t) = pe^t \longrightarrow E(X^2) = M_X''(0) = p$

$Var(X) = p - p^2 = p(1-p)$

P5 T in general: independent $\Rightarrow$ uncorrelated
F
F correlated $\Rightarrow$ dependent.

P8 $F(x) = 0.5$ $\int_0^x e^{-x} dx = 0.5$

P12 $E(X_i) = \frac{\frac{1}{2}}{\frac{1}{2}} = 1$ $Var(X) = 2.$

by CLT, $\sum^{50}$