Probability Theory Notes

Preface

This is my personal study notes of Probability Theory course, instructed by Hou Lin, at Tsinghua University, during the autumn semester of 2023. The notes is formulated throughout the semester, so it's also a personal study profile. The formal version is polished in the winter vacation of 2024. The last update is on 2024.7.12.

Disclaimer

The notes is just a personal study profile. Mistakes are unavoidable, and points are incomplete. It's suggested to use it as supplementary material when going over a course, or a quick guide to relevant domains. It's more recommended to use textbooks or professional online tutorials for systematical study purpose.

Contents

1 Fundamentals

1.1 Axioms of probability (Kolmogorov axioms)

1. Nonnegativity: The probability of an event is always a non-negative real number.

$$P(E) \in R, P(E) \geq 0$$

2. Unitarity: The probability of the entire sample space Ω is 1.

$$P(\Omega) = 1$$

3. Additivity: Any countable sequence of disjoint sets $E_1, E_2, \cdots$ satisfies

$$P(\bigcup_{i=1}^{\infty} E_i) = \sum_{i=1}^{\infty} P(E_i)$$

1.2 Conditional probability

1. Definition: the probability of an event occurring, given that another event occurs.
2. Properties:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

$$P(\cap_{i=1}^{n} A_i) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \cdots P(A_n| \cap_{i=1}^{n} A_i)$$

3. Law of total probability: If $B_1, B_2, \cdots, B_n$ is a partition of a sample space Ω, then for any event A, we have

$$P(A) = P(A|B_1) + P(A|B_2) + \cdots + P(A|B_n)$$

4. Bayes' Rule:

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}$$

$$P(B_i|A) = \frac{P(A|B_i)P(B_i)}{\sum_i P(A|B_i)P(B_i)}$$

$$f_{X|Y}(x|y) = \frac{f_X(x)f_{Y|X}(y|x)}{f_Y(y)}$$

$$f_{X|Y}(x|y) = \frac{f_X(x)f_{Y|X}(y|x)}{\int_{-\infty}^{\infty} f_X(t)f_{Y|X}(y|t)dt}$$

1.3 Independence

1. Definition: Two events A and B are independent if and only if

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

2. Properties:
   (a) If A and B are independent, then A and B^c are independent, A^c and B are independent, A^c and B^c are independent.
   (b) Disjoint events are not independent.
   (c) Necessary and sufficient condition:$F(x, y) = F(x)F(y)$; $f(x, y) = f(x)f(y)$; $\phi(x, y) = \phi(x)\phi(y)$
3. Pairwise independence: A finite set of events $\{A_i\}_{i=1}^{n}$ is pairwise independent if every pair of events are independent.

$$P(A_i \cap A_j) = P(A_i)P(A_j)$$

4. Mutual independence: A finite set of events $\{A_i\}_{i=1}^n$ is mutually independent if every event is independent of any intersection of the other events.

$$P(\cap_{i\in S}A_i) = \prod_{i\in S} P(A_i)$$

For more than two events, a mutually independent set of events is pairwise independent, but the converse is not necessarily true.

5. Conditional independence: If $P(A\cap B|C) = P(A|C)P(B|C)$, then A and B are conditionally independent given the condition C. Conditional independence can't bring independence, and vice versa.

6. Independence of random variables: If two random variables X and Y satisfies $P_{X,Y}(x,y) = p_X(x)P_Y(y)$, then X and Y are independent random variables. In this case, for any function $g(\cdot)$, $h(\cdot)$, $E[g(X)h(Y)] = E[g(X)]E[h(Y)]$.

2 Measurements

2.1 Expectation

1. Definition: In a discrete case, when $\sum_{i=1}^{\infty}|x_i|p_i$ converges,

$$E(X) = \sum_{i=1}^{\infty} x_i p_i$$

In a continuous case,

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

Especially,

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

2. Properties:

   (a) $E(cX+d) = cE(X)+d$

   (b) $E(X_1+X_2+\cdots+X_n) = E(X_1)+E(X_2)+\cdots+E(X_n)$

   (c) Assuming $X_1, X_2, \cdots, X_n$ are independent, then

   $$E(X_1X_2\cdots X_n) = E(X_1)E(X_2)\cdots E(X_n)$$

   (d) Law of total expectation:

   $$E(X) = \sum_y P_Y(y)E(X|Y=y)$$

   $$E(Y) = E[E(Y|X)]$$

2.2 Variance

1. Definition: $Var(X) = E[(X-EX)^2] = E(X^2) - E(X)^2$. The standard deviation $\sigma = \sqrt{Var(X)}$.

2. Properties:

   (a) $Var(X+c) = Var(X)$

   (b) $Var(cX+d) = c^2Var(X)$

   (c) Assuming $X_1, X_2, \cdots, X_n$ are independent, then

   $$Var(X_1+X_2+\cdots+X_n) = Var(X_1)+Var(X_2)+\cdots+Var(X_n)$$

   (d) For any random variables $X_1, X_2, \cdots, X_n$,

   $$Var(X_1+X_2+\cdots+X_n) = \sum_{i=1}^{n} Var(X_i) + \sum_{1\le i,j\le n, i\ne j} Cov(X_i, X_j)$$

(e) For independent and identical distributed random variables $X_1, X_2, \cdots, X_n$,

$$Var(\overline{X}) = Var(\frac{X_1 + X_2 + \cdots + X_n}{n}) = \frac{var(X)}{n}$$

(f) For any $c \in R$,

$$Var(X) = E(X - E(X))^2 \leq E(X - c)^2$$

(g) Law of total variance:

$$Var(X) = E[Var(X|Y)] + Var[E(X|Y)]$$

2.3 Covariance

1. Definition: $Cov(X, Y) = E[(X - EX)(Y - EY)] = E(XY) - E(X)E(Y)$.
2. Properties:
   (a) $Cov(X, X) = Var(X)$
   (b) $Cov(aX + b, cY + d) = acCov(X, Y)$
   (c) $Cov(X + Y, Z) = Cov(X, Z) + Cov(Y, Z)$
   (d) $Cov(\mathbf{X}) = \mathbf{\Sigma} \implies Cov(\mathbf{AX}) = \mathbf{A\Sigma A^T}$

2.4 Median

1. Definition:
   In a discrete case, if the data size is odd, the median of a dataset is the middle value, and if the data size is even, the median is usually defined as the arithmetic mean of the two middle values;
   In a continuous case, a median is defined as any real number m that satisfies the inequalities:

$$\int_{(-\infty, m]} dF(x) \geq \frac{1}{2} \quad \int_{[m, +\infty]} dF(x) \geq \frac{1}{2}$$

2. Properties:
   (a) If the distribution has finite variance, then the distance between the median $\tilde{X}$ and the mean $\overline{X}$ is bounded by one standard deviation:

$$|\overline{X} - \tilde{X}| \leq \sigma$$

   (b) If $\tilde{X}$ is a sample median, then it minimizes the arithmetic mean of the absolute deviations:

$$E(|X - \tilde{X}|) \leq E(|X - c|), \forall c \in R$$

2.5 Mean absolute deviation

1. Definition: Mean absolute deviation(MAD) is the average of the absolute deviations from the mean, which is $E(|X - \mu|)$.
2. Properties:
   (a) Mean absolute deviation is less than or equal to the standard deviation. For a normal distribution, the ratio of MAD and standard deviation is $\sqrt{\frac{2}{\pi}} \approx 0.7979$
   (b) The median is the point about which the average absolute deviation is minimized.

2.6 Skewness

1. Definition: For a probability distribution, the skewness is $\frac{E(X-E(X))^3}{\sigma^3}$.
2. Properties:
   (a) If the skewness is positive, the PDF curve leans leftward and the right tail is longer. If the skewness is negative, the PDF curve leans rightward and the left tail is longer.
   (b) A distribution with negative skewness can have its mean greater than or less than the median, and likewise for positive skewness.
   (c) A symmetric distribution has zero skewness. But zero skewness doesn't mean a symmetric distribution.

2.7 Kurtosis

1. Definition: For a probability distribution, the kurtosis is $\frac{E(X-E(X))^4}{\sigma^4}$
2. Properties:
   (a) A higher value of kurtosis indicates a higher, sharper peak of PDF, while a lower value indicates a shorter, fatter peak.
   (b) The kurtosis of a standard normal distribution is 3.
   (c) The kurtosis of a uniform distribution is 1.8.
   (d) The kurtosis of an exponential distribution is 9.
   (e) The range of kurtosis is $[1, +\infty)$.

2.8 Other measurements

1. Range: The difference between the largest and smallest values; the result of subtracting the sample maximum and minimum.
2. Interquartile range(IQR): The difference between the 75th and 25th percentiles of the data. It is used to describe statistical dispersion.
3. Mode: The value that appears most often in a dataset.

3 Distributions

3.1 Discrete random variable

3.1.1 Bernoulli distribution

1. Definition: A discrete probability distribution of a random variable which takes the value 1 with probability p and the value 0 with probability $1-p$. This is usually denoted as $X \sim B(1,p)$ or $X \sim B(p)$.
2. Expectation: p
3. Variance: $p(1-p)$

3.1.2 Binomial distribution

1. Definition: In a binomial distribution with parameters n and p, $P(X = k) = C_n^k p^k (1-p)^{n-k} (k \in \{0, 1, \cdots, n\})$. This is usually denoted as $X \sim B(n,p)$.
2. Expectation: p
3. Variance: $p(1-p)$
4. Properties:
   (a) If $X_1, X_2, \cdots, X_n$ are independent Bernoulli random variables with parameter p, then $X_1 + X_2 + \cdots + X_n \sim B(n,p)$.
   (b) If X, Y are independent and $X \sim B(n,p)$, $Y \sim B(m,p)$, then
   $X + Y \sim B(m+n, p)$.
   (c) The integer k_0 that maximizes P(X=k) satisfies:

$$k_0 = \begin{cases} (n+1)p,\ (n+1)p-1, \text{ if (n+1)p is integer} \\ \lfloor (n+1)p \rfloor, \text{ if (n+1)p is not integer} \end{cases}$$

3.1.3 Multinomial distribution

1. Definition: In a multinomial distribution with parameters N and $p_1, p_2, \cdots, p_n$, the n-dimensional random vector X follows this rule:

$$P(X_i = k) = \sum_{k_1+k_2+\cdots+k_n=N, k_i=k} \frac{N!}{k_1!k_2!\cdots k_n} p_1^{k_1} p_2^{k_2} \cdots p_n^{k_n}$$

2. Expectation: $E(X_i) = np_i$
3. Variance: $Var(X_i) = np_i(1-p_i)$, $Cov(X_i, X_j) = -np_ip_j (i \neq j)$

3.1.4 Poisson distribution

1. Definition: In a Poisson distribution with parameter λ, $P(X=i)=\frac{e^{-\lambda}\lambda^i}{i!}(i=0,1,\cdots)$. This is usually denoted as $X \sim P(\lambda)$.
2. Expectation: λ
3. Variance: λ
4. Properties:
   (a) Asymptotic property: when n is big and p is small and $np=\lambda$ is a proper integer, then
   $$B(n,p) \sim P(\lambda)$$

3.1.5 Geometric distribution

1. Definition: In a geometric distribution with parameter p, $P(X=k)=p(1-p)^{k-1}(i=1,2,\cdots)$. This is usually denoted as $X \sim G(p)$.
2. Expectation: $\frac{1}{p}$
3. Variance: $\frac{1-p}{p^2}$
4. Properties:
   (a) Memoryless property: For any positive integer m and n,
   $$P(X>m+n|X>n)=P(X>m)$$

3.1.6 Hypergeometric distribution

1. Definition: In a hypergeometric distribution with parameter n, M, N, $P(X=k)=\frac{C_M^k C_{N-M}^{n-k}}{C_N^n}$. This is usually denoted as $X \sim H(n,M,N)$.
2. Expectation: $\frac{nM}{N}$
3. Variance: $\frac{Mn(N-M)(N-n)}{N^2(N-1)}$
4. Properties:
   (a) Asymptotic property: Let $p=\frac{M}{N}$, when $N \to \infty$,
   $$H(n,M,N) \sim B(n,p)$$
   .

3.2 Continuous random variable

3.2.1 Uniform distribution

1. Definition: In a uniform distribution with parameters a and b, the probability density function is $f(x)=\frac{1}{b-a}(a \le x \le b)$. This is usually denoted as $X \sim U(a,b)$.
2. Expectation: $\frac{a+b}{2}$
3. Variance: $\frac{(b-a)^2}{12}$

3.2.2 Normal distribution

1. Definition: In a normal distribution with parameters μ and σ, the probability density function is

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} (-\infty < x < \infty)$$

2. Expectation: μ

3. Variance: σ^2

4. Properties:

   (a) If $X \sim N(\mu, \sigma^2)$, then $Y = aX + b \sim N(a\mu + b, a^2\sigma^2)$

   (b) If $X_1 \sim N(\mu_1, \sigma_1^2)$, $X_2 \sim N(\mu_2, \sigma_2^2)$, then

$$X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$$

$$\frac{X_1}{X_2} \sim Cauchy(0, 1)$$

   (c) if X_1 and X_2 are independent random variables and their sum $X_1 + X_2$ has a normal distribution, then both X_1 and X_2 must be normal deviates.

   (d) The inflection points are $x = \mu - \sigma$ and $x = \mu + \sigma$.

   (e) $E(X^2) = \mu^2 + \sigma^2$, $E[(X - \mu)^2] = \sigma^2$,
   $E(X^3) = \mu^3 + 3\mu\sigma^2$, $E[(X - \mu)^3)] = 0$,
   $E(|X|) = \sqrt{\frac{2}{\pi}}$

   (f) For two normal random variable X and Y with mean 0 and variance σ^2, $Y = X^2 + Y^2$ follows an exponential distribution with parameter $\frac{1}{2\sigma^2}$, i.e. $Y \sim \varepsilon(\frac{1}{2\sigma^2})$. $f_Y(y) = \frac{1}{2\sigma^2} e^{-\frac{y}{2\sigma^2}}, y \geq 0$.
   Let $R = \sqrt{Y} = \sqrt{X^2 + Y^2}$, then $f_R(r) = \frac{r}{\sigma^2} e^{-\frac{r^2}{2\sigma^2}}$. This is usually called R follows a Rayleigh distribution.

   (g) If $X, Y \sim N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$, then X, Y are independent $\Leftrightarrow \rho = 0$.

5. Multivariate normal distribution: for a k-dimensional multivariate normal random vector $\mathbf{x}$ with mean μ and covariance matrix Σ, the probability density function is:

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{k}{2}} |\Sigma|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(\mathbf{x} - \mu)\Sigma^{-1}(\mathbf{x} - \mu)\right)$$

   especially, in the bivariate case, let σ_X, σ_Y be the variance and ρ be the covariance, then

$$f(x, y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}\left[(\frac{x-\mu_1}{\sigma_X})^2 - 2\rho(\frac{x-\mu_1}{\sigma_X})(\frac{y-\mu_2}{\sigma_Y}) + (\frac{y-\mu_2}{\sigma_Y})^2\right]\right)$$

$$Y|X \sim N(\mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x - \mu_1), (1 - \rho^2)\sigma_2^2)$$

$$X + Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2 + 2\rho\sigma_1\sigma_2)$$

3.2.3 Exponential distribution

1. Definition: In an exponential distribution with parameter λ, the probability density function is

$$f(x) = \lambda e^{-\lambda x} (0 \leq x < \infty)$$

   This is usually denoted as $X \sim \varepsilon(\lambda)$.

2. Expectation: $\frac{1}{\lambda}$

3. Variance: $\frac{1}{\lambda^2}$

4. Properties:

   (a) Memoryless property: For any positive real number m and n,

$$P(X > m + n|X > n) = P(X > m)$$

3.2.4 Chi-squared distribution

1. Definition: If $Z_1, Z_2, \cdots, Z_k$ are independent, standard normal random variables, then the sum of their squares, $Q = \sum_{i=1}^{k} Z_i^2$ is distributed according to the chi-square distribution with k degrees of freedom. This is usually denoted as $Q \sim \chi^2(k)$
2. Expectation: k
3. Variance: $2k$
4. Properties:
   (a) Probability density function:
   $$f(x) = \frac{1}{2^{\frac{k}{2}}\Gamma(\frac{k}{2})} x^{\frac{k}{2}-1} \mathrm{e}^{-\frac{x}{2}} (x > 0)$$
   (b) Additivity: If $X_1 \sim \chi^2(k_1), X_2 \sim \chi^2(k_2)$, then $X_1 + X_2 \sim \chi^2(k_1 + k_2)$

3.2.5 t-distribution

1. Definition: If $X \sim N(0,1)$,$Y \sim \chi_n^2$ and X and Y are independent, then $T = \frac{X}{\sqrt{\frac{Y}{n}}}$ is a t-deviate with n degrees of freedom. This is usually denoted as $T \sim t_n$
2. Properties:
   (a) Probability density function:
   $$f(x) = \frac{\Gamma(\frac{k+1}{2})}{\sqrt{k\pi}\Gamma(\frac{k}{2})} \left(1 + \frac{x^2}{k}\right)^{-\frac{k+1}{2}}$$

3.2.6 f-distribution

1. Definition: If $X \sim \chi_m^2$,$Y \sim \chi_n^2$ and X and Y are independent, then $F = \frac{X}{m}/\frac{Y}{n}$ is a f-deviate with m and n degrees of freedom. This is usually denoted as $F \sim f_{m,n}$.
2. Properties:
   (a) Probability density function:
   $$f(x) = \frac{\Gamma(\frac{m+n}{2}) m^{\frac{m}{2}} n^{\frac{n}{2}} x^{\frac{m}{2}-1}}{\Gamma(\frac{m}{2})\Gamma(\frac{n}{2})(mx+n)^{\frac{m+n}{2}}}, \ x \geq 0$$

4 Moment Generating Function

For a random variable X, the moment generating function is $M_X(t) = E[e^{tX}]$.

4.1 Properties

1. $M_X(t) = 1 + tE(X) + \frac{1}{2!}t^2E(X^2) + \frac{1}{3!}t^3E(X^3) + \cdots + \frac{1}{n!}t^nE(X^n)$
2. $E(X^n) = M_X^{(n)}(0)$
3. $M_{\alpha X+\beta} = e^{\beta t} M_X(\alpha t)$
4. For independent random variables $X_1, X_2, \cdots, X_n$, let $S = \sum_{i=1}^{n} X_i$, then $M_S(t) = M_{X_1}(t)M_{X_2}(t)\cdots M_{X_n}(t)$
5. When X only takes nonnegative integer values, $P(X = 0) = \lim_{s\to-\infty} M(s)$
6. If $M_X(t)$ can be written as $M_X(t) = \sum p_k e^{kt}$, $P(X = k) = p_k$

5 Useful conclusions

Distribution	PDF	Expectation	Variance	MGF	C.F
Uniform	$\frac{1}{b-a}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$	$\frac{e^{tb}-e^{ta}}{t(b-a)}$	$\frac{e^{itb}-e^{ita}}{it(b-a)}$
Bernoulli		p	$p(1-p)$	$1-p+pe^t$	$1-p+pe^{it}$
Binomial		np	$np(1-p)$	$(1-p+pe^t)^n$	$(1-p+pe^{it})^n$
Poisson		λ	λ	$e^{\lambda(e^t-1)}$	$e^{\lambda(e^{it}-1)}$
Geometric		$\frac{1}{p}$	$\frac{1-p}{p^2}$	$\frac{pe^t}{1-(1-p)e^t}$	$\frac{pe^{it}}{1-(1-p)e^{it}}$
Gaussian		μ	σ^2	$e^{\mu t+\frac{1}{2}\sigma^2 t^2}$	$e^{i\mu t-\frac{1}{2}\sigma^2 t^2}$
Exponential		$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	$\frac{1}{1-\frac{t}{\lambda}}$	$\frac{1}{1-\frac{it}{\lambda}}$

5.1 Inequalities

1. Markov's inequality: for a nonnegative random variable X and any positive number a,
$$P(X \geq a) \leq \frac{E(X)}{a}$$
2. Chebyshev's inequality: for a random variable X with finite expectation and any positive number a,
$$P(|X - E(X)| \geq a) \leq \frac{Var(X)}{a^2}$$
3. Jensen's inequality: let $f : R \to R$ be a convex function and X a random variable with finite expectation, then
$$f(E(X)) \leq E(f(X))$$
Let $f(x) = x^2$, we get:
$$(E(X))^2 \leq E(X^2)$$
4. Cauchy-Schwarz inequality:
$$|E(XY)|^2 \leq E(X^2)E(Y^2)$$

5.2 Calculation

1. Maximum of multiple random variables: let $X = max\{X_1, X_2, \cdots, X_n\}$ and $F_X(x)$ be the cumulative distribution function, then the CDF of X is
$$F(x) = F_{X_1}(x)F_{X_2}(x)\cdots F_{X_n}(x)$$
By differential or difference, we can get $f_X(x)$.
2. Minimum of multiple random variables: let $X = min\{X_1, X_2, \cdots, X_n\}$ and $F_X(x)$ be the cumulative distribution function, then the CDF of X is
$$F(x) = 1 - [1 - F_{X_1}(x)][1 - F_{X_2}(x)]\cdots[1 - F_{X_n}(x)]$$
3. Let X_1, $X - 2$, ..., X_n be i.i.d random variables with CDF $F(x)$ and PDF $f(x)$, then the joint pdf of $X_{(1)}$ and $X_{(n)}$ is
$$f(x, y) = n(n-1)[F(y) - F(x)]^{n-2}f(y)f(x)(y > x)$$
4. PDF of the function of random variable: let X be a random variable and $f_X(x)$ be the PDF, then for $Y = g(X)(g(x)$ is monotonic), the PDF is
$$f_Y(Y) = f_X(h(y))|h'(y)|$$

where $h(y)$ is the inverse function of $g(x)$.

Generally, if $g(x)$ is not monotonic, for every monitonic interval, let the inverse function of $g(x)$ be $h_i(y)$, then the PDF of Y is

$$f_Y(Y) = \sum_i f_X(h_i(y))|h_i'(y)|$$

5. When $Y = aX + b$, then the pdf of X and Y satisfies:

$$f_Y(y) = \frac{1}{|a|} f_X(\frac{y-b}{a})$$

6. Let the joint pdf of X, Y be $f(x, y)$. If $x = x(u, v)$, $y = y(u, v)$, then the joint pdf of U, V is

$$g(u, v) = f(x(u, v), y(u, v))|\frac{\partial(x, y)}{\partial(u, v)}|$$

7. For random variables X and Y, the pdf of $X + Y$ is
$f(k) = \int_{-\infty}^{\infty} f(x, k-x)dx$

8. For random variables X and Y, the pdf of $X - Y$ is
$f(k) = \int_{-\infty}^{\infty} f(x, x-k)dx$

9. For random variables X and Y, the pdf of Y/X is $f(k) = \int_{-\infty}^{\infty} |x| f(x, kx)dx$

10. For random variables X and Y, the pdf of XY is $f(k) = \int_{-\infty}^{\infty} \frac{1}{|x|} f(x, \frac{k}{x})dx$

5.3 Combinatorics

1. Binomial expansion:

$$(x + y)^n = \sum_{i=0}^{n} C_n^i x^i y^{n-i}$$

$$\sum_{k=0}^{n} C_n^k p^k (1-p)^{n-k} = 1$$

2. Multinomial expansion:

$$(x_1 + x_2 + \cdots + x_m)^n = \sum_{k_1+k_2+\cdots+k_m=n; k_1, k_2, \cdots, k_m \geq 0} \frac{n!}{k_1!k_2!\cdots k_m!} x_1^{k_1} x_2^{k_2} \cdots x_m^{k_m}$$

3. $C_n^0 + C_n^2 + C_n^4 + \cdots + C_n^m = 2^{n-1}$, for even n, $m = n$, for odd n, $m = n - 1$.

4. $C_n^1 + C_n^3 + C_n^5 + \cdots + C_n^m = 2^{n-1}$, for even n, $m = n - 1$, for odd n, $m = n$.

5. $C_m^n = C_{m-1}^{n-1} + C_{m-1}^n$

6. $C_n^n + C_{n+1}^n + C_{n+2}^n + \cdots + C_m^n = C_{m+1}^{n+1}$

7. $(C_n^0)^2 + (C_n^1)^2 + \cdots + (C_n^n)^2 = C_{2n}^n$