1. Moment Generating Function $M_X(t) = E(e^{tx})$
   continuous: $M_X(t) = \int_{-\infty}^{\infty} e^{tx} P_x(x) dx$
   discrete: $M_X(t) = \sum e^{tx} P_x(x)$
   $\frac{dM_x(t)}{dt} = E(\frac{de^{tx}}{dt}) = E(xe^{tx}) \Rightarrow M_x^{(n)}(t) = E(x^n e^{tx})$
   Properties:
   (a) $M_x(t) = (1 - p + pe^t)^n$ for binomial(n,p)
   (b) $M_x(t) = E(\sum_{n=0}^{\infty} \frac{x^n t^n}{n!})$ according to Taylor Series
   $M_x'(t) = E(0 + x + x^2 t + \frac{x^3 t^2}{2!} + \dots) = E(\sum_{n=0}^{\infty} \frac{x^{n+1} t^n}{n!})$
   $\Rightarrow M_x^{(k)}(t) = E(\sum_{n=0}^{\infty} \frac{x^{n+k} t^n}{n!})$
   (c) $M_x^{(n)}(0) = E(X^n)$
   $\Rightarrow VAR[x] = E[x^2] - E^2[x] = M_x''(0) - [M_x'(0)]^2$
   (d) $M_x(t) = M_y(t) \Rightarrow XY$ has same distribution
   (e) $M_x(0) = 1$
   (f) $M_{ax+b}(t) = M_x(at)e^{bt}$ Proof:
   $M_{ax+b}(t) = E(e^{t(ax+b)}) = E(e^{axt} e^{bt}) = E(e^{axt})e^{bt} = M_x(at)e^{bt}$
   (g) $M_{X+Y}(bt) = M_X(t) M_Y(t)$ (X, Y independent)

2. Gama Function $\Gamma(n) = \int_0^{\infty} u^{n-1} e^{-u} du = (n-1)!$
   Proof:
   1. $\int_0^{\infty} e^{-x} dx = -e^{-x}|_0^{\infty} = 1$
   2. $\Gamma(n) = -\int_0^{\infty} x^{n-1} de^{-x}$
   $= -(x^{n-1} e^{-x}|_0^{\infty} - \int_0^{\infty} e^{-x}(n-1)x^{n-1} dx)$
   $= 0 + \int_0^{\infty} e^{-x}(n-1)x^{n-2} dx$
   $= (n-1)\int_0^{\infty} e^{-x} x^{n-2} dx$
   $= (n-1)\Gamma(n-1)$
   Use Induction to prove the formula

3. Basic probability property
   (a) Basic Properties
   1. $P(E) = \frac{n(E)}{n(S)}, \in [0, 1]$
   2. $P(\phi) = 0$
   3. $A \subseteq B, P(A) \le P(B)$
   4. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
   (b) Conditional Probability and Bayles Theorem
   1. $P(A|B) = \frac{P(A \cup B)}{P(B)}$
   2. $P(A|B) = \frac{P(B|A)P(A)}{P(B)} = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A')P(A')}$
   (c) PDF and CDF
   PDF: probability distribution function $\Rightarrow P_X(x)$
   CDF: cummulative distribution function $\Rightarrow F_X(x) = \phi_X(x) = N_X(x) = P(X < x)$
   (d) Expectation
   1. $E(X) = \sum x P_X(x)$ or $\int_{-\infty}^{\infty} x P_X(x)$
   2. $E(c) = c$
   3. $E(aX) = aE(X)$
   4. $E(X + Y) = E(X) + E(Y)$
   5. $E(X) = M_X'(0)$
   (e) Variance and Standard Deviation
   1. $VAR(X) = E(X^2) - E^2(X) = \sum[(x - E(X)]P_X(x) = E[(x-\mu)^2] = \sigma^2$
   2. $VAR(c) = 0$
   3. $VAR(aX) = a^2 VAR(x)$
   4. $VAR(X \pm Y) = VAR(X) + VAR(Y) \pm 2COV(x, y)$, $COV(x, y) = E(XY) - E(X)E(Y)$
   (f) z-score
   $Z = \frac{x-\mu}{\sigma}$, measures the distance of x from expected value in standard units.

4. Discrete Distributions
   (a) Binomial Distribution
   DEF: n time Bernoulli trials combined. probability of success and fail is (p, 1-p). Probability of success remains the same through the trails. X is the r.v of success times.
   Note as $B(n, p)$
   $P_X(x) = \binom{n}{x} p^x (1-p)^{n-x}$
   $M_X(t) = (1 - p - pe^t)^n$
   $E(X) = np, VAR(X) = np(1-p)$
   (b) Hyper Geometric Distribution
   DEF: A sample of size n taken from a finite population of size N. The population has a subgroup of size $r \ge n$ that is of interest. x is the number of members of the subgroup taken.
   Note as $H(N, n, r)$
   $P_X(x) = \frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}}$
   $E(X) = n\frac{r}{N}$
   $VAR(X) = n\frac{r}{N}(1 - \frac{r}{N})(\frac{N-n}{N-1})$
   When $x \to \infty$, H(N,n,r) $\to$ B(n,$\frac{r}{N}$). H samples without replacement while B samples with replacement.
   (c) Poisson Process
   DEF: model the number of random occurance of some phenomenon in specific unit of space or time.
   $P_X(x) = \frac{e^{-\lambda}\lambda^x}{x!}$
   λ: average arrival in given time or space
   $E(X) = \lambda, VAR(X) = \lambda$
   Poisson simulates Binomial when n is large. usually when $\lambda = np < 10$, we use poisson as an approximation of Binomial.
   (d) Geometric Distribution DEF: number of trials to get the first success in a sequence of Bernoulli trials where p is the success probability.
   i. X is the r.v of number of total trials (x includes the first success)
   $P_X(x) = (1-p)^{x-1} p, x = 1, 2, 3\dots$
   $E(X) = 1/p, VAR(X) = \frac{1-p}{p^2}$
   ii. X is the r.v of number of failed trials (x excludes the first success)
   $P_X(x) = (1-p)^x p, x = 0, 1, 2, \dots$
   $M_X(t) = \frac{p}{1 - e^t(1-p)}$
   $E(X) = \frac{1-p}{p}, VAR(X) = \frac{1-p}{p^2}$
   (e) Negative Binomial Distribution
   DEF: X is the r.v of number of trials need to observe the r^{th} success in a sequence of Bernoulli trails where p is the success probability.
   $P_X(x) = \binom{x-1}{r-1} p^r (1-p)^{x-r}, x = r, r+1, r+2\dots$
   $M_X(t) = (\frac{p}{1 - e^t(1-p)})^r$
   $E(X) = \frac{r}{p}, VAR(X) = \frac{r(1-p)}{p^2}$
   Alternatively, X is the r.v of failures before the r^{th} success:
   $P_X(x) = \binom{x+r-1}{r-1} p^r (1-p)^x, x = 0, 1, 2\dots$

$M_X(t) = (\frac{1-p}{1-pe^t})^r$
$E(X) = \frac{r(1-p)}{p}, VAR(X) = \frac{r(1-p)}{p^2}$

5. Chebychev's Theorem
   $P(\mu - k\sigma \leq x \leq \mu + k\sigma) \geq 1 - \frac{1}{k^2}$

6. Continuous Random Variable

   (a) Basic Properties
   1. $pdf : f_X(x) \geq 0$
   2. $cdf : F_X(x) = \int_{-\infty}^{\infty} f_X(x)dx$
   3. $\int_{-\infty}^{\infty} f_X(x)dx = 1$

   (b) Mean, Medium and Variance
   mean: $\mu = E(x) = \int_{-\infty}^{\infty} x f_X(x)dx$
   medium m: solve function $\int_{-\infty}^{m} f_X(x)dx = \frac{1}{2}$
   Variance:
   $VAR(X) = E(X^2) - E^2(X) = \int_{-\infty}^{\infty} [x - E(X)] f_X(x)dx$