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) $M_x(t) = E(\sum_{n=0}^{\infty} \frac{x^n t^n}{n!})$ according to Taylor Series
$M_x'(t) = E(0 + x + x^2 t + \frac{x^3 t^2}{2!} + \dots) = E(\sum_{n=0}^{\infty} \frac{x^{n+1} t^n}{n!})$
$\Rightarrow M_x^{(k)}(t) = E(\sum_{n=0}^{\infty} \frac{x^{n+k} t^n}{n!})$

(b) $M_x^{(n)}(0) = E(X^n)$
$\Rightarrow VAR[x] = E[x^2] - E^2[x] = M_x''(0) - [M_x'(0)]^2$

(c) $M_x(t) = M_y(t) \Rightarrow XY$ has same distribution

(d) $M_x(0) = 1$

(e) $M_{ax+b}(t) = M_x(at)e^{bt}$ Proof:
$M_{ax+b}(t) = E(e^{t(ax+b)}) = E(e^{axt}e^{bt}) = E(e^{axt})e^{bt} = M_x(at)e^{bt}$

(f) $M_{X+Y}(bt) = M_X(t)M_Y(t)$ (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
Formula: $\int_0^1 x^{\alpha-1}(1-x)^{\beta-1}dx = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$

3. Basic probability property

(a) Basic Properties
1. $P(E) = \frac{n(E)}{n(S)}, \in [0,1]$
2. $P(\phi) = 0$
3. $A \subseteq B, P(A) \leq P(B)$
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. $E(X) = \sum xP_X(x)$ or $\int_{-\infty}^{\infty} xP_X(x)$
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. Common 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 $X \sim B(n,p)$

$$P_X(x) = \binom{n}{x} p^x (1-p)^{n-x}$$

$$M_X(t) = (1 - p + pe^t)^n$$

$$E(X) = np,\ VAR(X) = np(1-p)$$

(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 $r \geq n$ that is of interest. x is the number of members of the subgroup taken.
Note as $X \sim H(N,n,r)$

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

$$E(X) = n\frac{r}{N},\ VAR(X) = n\frac{r}{N}(1 - \frac{r}{N})(\frac{N-n}{N-1})$$

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)

$$P_X(x) = (1-p)^{x-1}p, x = 1, 2, 3\ldots$$

$$E(X) = 1/p, VAR(X) = \frac{1-p}{p^2}$$

ii. X is the r.v of number of failed trials (x excludes the first success)

$$P_X(x) = (1-p)^x p, x = 0, 1, 2, \ldots$$

$$M_X(t) = \frac{p}{1-e^t(1-p)}$$

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

(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.

$$P_X(x) = \binom{x-1}{r-1}p^r(1-p)^{x-r}, x = r, r+1, r+2\ldots$$

$$M_X(t) = (\frac{p}{1-e^t(1-p)})^r$$

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

Alternatively, X is the r.v of failures before the r^{th} success:

$$P_X(x) = \binom{x+r-1}{r-1}p^r(1-p)^x, x = 0, 1, 2\ldots$$

$$M_X(t) = (\frac{1-p}{1-pe^t})^r$$

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

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. $\int_{-\infty}^{\infty} f_X(x)dx = 1$
4. $f_X(x) = \frac{d}{dx}F_X(x)$

(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:
$VAR(X) = E(X^2) - E^2(X) = \int_{-\infty}^{\infty}[x - E(X)]f_X(x)dx$

7. Common Continuous Distributions

(a) Uniform(rectangle) Distribution

$$f_X(x) = \frac{1}{b-a}, a \leq x \leq b$$

$$F_X(x) = \frac{x-a}{b-a}, a \leq x \leq b$$

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

(b) Exponential Distribution

$$f_X(x) = \frac{1}{\theta}e^{-\frac{x}{\theta}}, x \geq 0$$

$$F_X(x) = 1 - e^{-\frac{x}{\theta}}, x \geq 0$$

$$M_X(t) = \frac{1}{1-\theta t}$$

$$E(X) = \theta, VAR(X) = \theta^2$$

Note as $X \sim exp(\theta)$
In terms of λ: $\lambda = \frac{1}{\theta}$

(c) Gamma Distribution

$$f_X(x) = \frac{1}{\Gamma(n)\theta^n}x^{n-1}e^{-\frac{x}{\theta}}, x \geq 0$$

$$M_X(t) = \frac{1}{(1-\theta t)^n}$$

$$E(X) = n\theta, VAR(X) = n\theta^2$$

Note as $X \sim \Gamma(n, \theta)$
Exponential dist. is a special case of Gamma dist where n = 1. Gamma Distribution can be viewed as a sum of Exponential dists.
$X \sim \Gamma(n, \theta) \Leftrightarrow X = \sum_{i=1}^{n} X_i, X_i \sim exp(\theta)$
In terms of α, β: $(\alpha = n, \beta = \frac{1}{\theta})$

$$f_X(x) = \frac{\beta^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\beta x}, x \geq 0$$

$$M_X(t) = \frac{1}{(1-\frac{t}{\beta})^\alpha}$$

$$E(X) = \frac{\alpha}{\beta}, VAR(X) = \frac{\alpha}{\beta^2}$$

(d) Normal Distribution
Standard normal distribution:
X ~ N(0,1), $\mu = 0, \sigma^2 = 1$
Normal distribution:

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}},\ x \in (-\infty, \infty)$$

$$M_X(t) = e^{\mu t + \frac{1}{2}\sigma^2 t^2}$$

$$E(x) = \mu,\ VAR(X) = \sigma^2$$

(e) Log-normal Distribution:
$Y = e^X, X \sim N(\mu, \sigma^2)$

$$f_Y(y) = \frac{1}{\sigma y\sqrt{2\pi}}e^{-\frac{1}{2}(\frac{lny-\mu}{\sigma})^2}, y \geq 0$$

$$E(Y) = e^{\mu + \frac{\sigma^2}{2}}, VAR(Y) = e^{2\mu+\sigma^2}(e^{\sigma^2} - 1)$$

(f) Pareto Distribution:

$$f_X(x) = \frac{k}{\beta}(\frac{\beta}{x})^{k+1}$$

$$F_X(x) = 1 - (\frac{\beta}{x})^k, k > 2, x \geq \beta > 0$$

$$E(X) = \frac{k\beta^2}{k-1}, VAR(X) = \frac{k\beta^2}{k-2} - (\frac{k\beta}{k-1})^2$$

Other notation: $\alpha = k, \beta = \beta$

(g) Weibull Distribution:

$$f_X(x) = k\lambda x^{k-1} e^{-\lambda k x^k}$$

$$F_X(x) = 1 - e^{-\lambda k x^k}$$

$$E(X) = \frac{\Gamma(1+\frac{1}{k})}{\lambda^{\frac{1}{k}}}, VAR(X) = \frac{1}{\lambda^{\frac{2}{k}}}[\Gamma(1+\frac{2}{k}) - \Gamma^2(1+\frac{1}{k})]$$

Other notation:$k = \alpha, \lambda = \beta$
Note: $\Gamma(\frac{1}{2}) = \sqrt{\pi}$
Proof:
$1 - F_Z(0) = \frac{1}{2}$
$\Rightarrow \frac{1}{\sqrt{2\pi}} \int_0^\infty e^{-\frac{1}{2}z^2} dz = \frac{1}{2}$
Let $u = \frac{1}{2}z^2, du = zdz, z = \sqrt{2u}$
$\Rightarrow \frac{1}{\sqrt{2\pi}} \int_0^\infty e^{-u} \frac{1}{\sqrt{2u}} du = \frac{1}{2}$
$\Rightarrow \int_0^\infty e^{-u} u^{-\frac{1}{2}} = \sqrt{\pi}$
$\Gamma(\frac{1}{2}) = \sqrt{\pi}$

(h) Beta Distribution

$$f_X(x) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1}, x \in (0,1), a > 0, b > 0$$

$$E(X) = \frac{a}{a+b}, VAR(X) = \frac{ab}{(a+b)^2(a+b+1)}$$

(i) Chi-Square χ^2 Distribution

$$f_X(x) = \frac{1}{2^{\frac{k}{2}}\Gamma(\frac{k}{2})} x^{\frac{k}{2}-1} e^{-\frac{x}{2}}, x > 0$$

$$E(X) = k, VAR(X) = 2k$$

$$M_X(t) = \frac{1}{(1-2t)^{\frac{k}{2}}}$$

This is a special case of Gamma Dist. of $\alpha = \frac{k}{2}$, $\beta = 2$
$f_{Z^2}(z) = \frac{1}{\Gamma(\frac{1}{2})2^{\frac{1}{2}}} e^{\frac{y}{2}} y^{\frac{1}{2}-1}$(gamma dist of $\alpha = \frac{1}{2}, \beta = 2$)
$\chi^2 = \sum_{i=1}^k Z_i^2$(sum of k squared standard normal)
THEOREM:

i. $\chi^2(a) + \chi^2(b) = \chi^2(a+b)$
ii. If $X_1, X_2 \ldots, X_k$ is a random sample from $N(\mu, \sigma^2)$
$Z = \sum_{i=1}^k (\frac{x_i - \mu}{\sigma})^2 \sim \chi^2(k)$ (mean)
$Z = \sum_{i=1}^k (\frac{x_i - \bar{x}}{\sigma})^2 \sim \chi^2(k-1)$ (sample mean)
iii. When $n \geq 30$, Normal approximates Chi Square
$\chi^2(k) \sim N(k, 2k)$

8. Central Limit Theorem
Let $\{X_1, X_2, \ldots, X_n\}$ be the independent random variable all of which have the same distribution and mean μ and standard deviation σ. If n is large $n \geq 30$, then

$$S = X_1 + X_2 + \cdots + X_n$$

will be approximately normal with mean $n\mu$ and variance $n\sigma^2$. ($S \sim N(n\mu, n\sigma^2)$)
OR

$$S' = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

will be approximately normal with mean μ, variance $\frac{\sigma^2}{n}$.($S \sim N(\mu, \frac{\sigma^2}{n})$)

9. Finding CDF for $Y = g(X)$
1. $g(x)$ is strictly increasing on the sample space for X
Let $h(y)$ be the inverse function of $g(x)$. $h(x)$ is also strictly increasing.

$$\begin{aligned} F_Y(y) &= P(Y \leq y) \\ &= P(Y \leq y) \\ &= P(g(X) \leq y) \\ &= P[X \leq h(y)] \\ &= F_X(h(y)) \end{aligned}$$

Find the density function by differentiating

$$\begin{aligned} f_Y(y) &= \frac{d}{dy} F_X(h(y)) \\ &= \frac{d}{dx} F_X(h(y)) \frac{dy}{dx} \\ &= \frac{d}{dx} F_X(h(y)) h'(y) \end{aligned}$$

2. $g(x)$ is strictly decreasing on the sample space for X
Let $h(y)$ be the inverse function of $g(x)$. $h(x)$ is also strictly decreasing.

$$\begin{aligned} F_Y(y) &= P(Y \leq y) \\ &= P(Y \leq y) \\ &= P(g(X) \leq y) \\ &= P[X \geq h(y)] \\ &= 1 - F_X(h(y)) \end{aligned}$$

Find the density function by differentiating

$$\begin{aligned} f_Y(y) &= \frac{d}{dy}[1 - F_X(h(y))] \\ &= -\frac{d}{dx} F_X(h(y)) \frac{dy}{dx} \\ &= -\frac{d}{dx} F_X(h(y)) h'(y) \end{aligned}$$

IN ALL:
Density function for $Y = g(X)$
Let g(x) be strictly decreasing or increasing on the domain consisting of the sample space. Then:

$$f_Y(y) = \frac{d}{dx} F_X(h(y)) \cdot |h'(y)|$$

10. Multivariable Distributions

(a) Marginal Distribution

$$p_X(x) = \sum_y p(x,y),\ p_Y(y) = \sum_x p(x,y)$$

$$f_X(x) = \int_{-\infty}^{\infty} p(x,y)dy,\ f_Y(y) = \int_{-\infty}^{\infty} p(x,y)dx$$

(b) Conditional Distribution

$$P(X=x|Y=y) = \frac{p(x,y)}{p_Y(y)},\ P(Y=y|X=x) = \frac{p(x,y)}{p_X(x)}$$

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

(c) Conditional Expected Value

$$E(y|X=x) = \sum_y y \cdot p(y|x),\ E(x|Y=y) = \sum_x x \cdot p(x|y)$$

$$E(y|X=x) = \int_{-\infty}^{\infty} y \cdot f(y|x)dy,\ E(x|Y=y) = \int_{-\infty}^{\infty} x \cdot f(x|y)dx$$

(d) Independence

$$p(x,y) = p_X(x) \cdot p_Y(y)$$

$$f(x|y) = f_X(x),\ f(y|x) = f_Y(y)$$

$$E(X+Y) = E(X) + E(Y)$$

$$COV(X,Y) = 0$$

$$VAR(X \pm Y) = VAR(X) + VAR(Y)$$

(e) Covariance

$$COV(x,y) = E[(X-\mu_X)(Y-\mu_Y)] = E(XY) - E(X)E(Y)$$

Properties:

i. $COV(X,Y) = COV(Y,X)$
ii. $COV(X,X) = VAR(X)$
iii. $COV(X,K) = 0, K$ is constant
iv. $COV(aX,bY) = ab \cdot COV(X,Y)$
v. $COV(X,Y+Z) = COV(X,Y) + COV(X,Z)$
vi. If X,Y are independent, $COV(X,Y) = 0$
vii. $VAR(X \pm Y) = VAR(X) + VAR(Y) \pm 2COV(x,y)$

(f) Correlation Coefficient ρ

$$\rho = \rho_{X,Y} = \frac{COV(X,Y)}{\sigma_X \cdot \sigma_Y}, \in [-1,1]$$

$|\rho_{X,Y}|$ close to 1 : more related
$\rho > 0$ positively related, $\rho < 0$ negatively related

(g) Moment Generating Function of Joint Distribution

$$M_{X,Y}(t_1,t_2) = E(e^{t_1X+t_2Y})$$

$$E(X) = \frac{d}{dt_1} M_{X,Y}(t1,t2)|_{t_1=t_2=0}$$

$$E(Y) = \frac{d}{dt_2} M_{X,Y}(t1,t2)|_{t_1=t_2=0}$$

$$E(X^mY^n) = \frac{\partial^{m+n}}{\partial^m t_1 \partial^n t_2} M_{X,Y}(t_1,t_2)|_{t_1=t_2=0}$$

(h) Multinomial Distribution
Counting Partitions:
The number of Partitions of n objects into k distinct groups of size $n_1, n_2 \dots n_k$ is given by:

$$\binom{n}{n_1, n_2, \dots, n_k} = \frac{n!}{n_1!n_2! \dots n_k!}$$

Suppose that a random experiment has K mutually exclusive outcomes $E_1, E_2, \dots E_k$, with $P(E_i) = p_i, i \in [1,k]$. Suppose you repeat this experiment in n independent trails. Then:
$P(X_1 = E_1, \dots, X_k = E_k) = \binom{n}{n_1,n_2,\dots,n_k} p_1^{n_1} p_2^{n_2} \dots p_k^{n_k}$