Jensen's Inequality: E(g(X)) ≥ g(E(X)) (Both exist)
Cauchy-Schwarz: $(E(XY))^2 \le E(X^2)E(Y^2)$ (=:Pr(Y=cX)=1)
Chebyshev: $Pr(|X-u|\ge c\sigma)\le 1/c^2$
Markov: $Pr(|X|\ge c)\le E|X|^r/c^r$ ($E|X|^r$ exists)
General case: $Pr(g(X)\ge c)\le E(g(X))/c$

复合随机变量 Compound r.v.: $S_N=\sum_{[1,N]}X_i$, X_i i.i.d.
$E(S_N)=E(N)E(X)$; $Var(S_N)=E(N)Var(X)+Var(N)E^2(X)$

泊松分布 Poisson distribution:
$f(x|\lambda)=\lambda^x e^{-\lambda}/x!$, $x\in Z$; E(X) = Var(X) = λ
$M_X(t)=\exp(\lambda(e^t-1))$
$X_i\sim Poisson(\lambda_i)\to\sum X_i\sim Poisson(\sum\lambda_i)$
$\sum_{[k,\infty)}Poisson(x|\lambda) = \int_{[0,\lambda]}Gamma(y|k,1)dy$

二项分布 Binomial distribution:
$f(x|n,p)=C_n^x p^x(1-p)^{n-x}$, x = 0,1,...,n
$E(X)=np$; $Var(X)=np(1-p)$; $M_X(t)=(pe^t+1-p)^n$
$X_i\sim Bin(n_i,p)\to\sum X_i\sim Bin(\sum n_i,p)$
$\sum_{[0,k]}Bin(x|n,p)=\int_{[0,1-p]}Beta(x|n-k,k+1)dx$, $0\le k\le n$

负二项分布 Negative Binomial distribution: NB(r,p), $x\in N_r$
$f(x)=C_{x-1}^{r-1}p^r(1-p)^{x-r}$; $E(X)=r/p$; $Var(X)=r(1-p)/p^2$
$M_X(t)=(pe^t/(1-(1-p)e^t))^r$

均匀分布 Uniform distribution:
$E(X)=(a+b)/2$; $Var(X)=(b-a)^2/12$; $M_X(t)=(e^{tb}-e^{ta})/(t(b-a))$

伽玛分布 Gamma distribution:
$\Gamma(\alpha)=\int_{(0,+\infty)}x^{\alpha-1}e^{-x}dx=\alpha\Gamma(\alpha-1)$; $\Gamma(n)=(n-1)!$; $\Gamma(1/2)=\sqrt{\pi}$
$f(x)=\beta^\alpha x^{\alpha-1}e^{-\beta x}/\Gamma(a)$, x>0; $E(X)=\alpha/\beta$; $Var(X)=\alpha/\beta^2$
$M_X(t)=(\beta/(\beta-t))^\alpha$; $cX\sim Gamma(\alpha, \beta/c)$;
$Gamma(1, \beta)=Exp(\beta)$; $Gamma(v/2, 1/2)=\chi^2(v)$
$X\sim Gamma(\alpha,\beta)\to E(\log(X))=\Gamma'(\alpha)/\Gamma(\alpha)-\log(\beta)$

贝塔分布 Beta distribution:
$f(x)=x^{a-1}(1-x)^{b-1}/B(a,b)$; $B(a,b)=\Gamma(a)\Gamma(b)/\Gamma(a+b)$
$E(X)=a/(a+b)$; $E(X^2)=a(a+1)/[(a+b)(a+b+1)]$
$Var(X)=ab/[(a+b)^2(a+b+1)]$
$Y_1\sim Gamma(a,1), Y_2\sim Gamma(b,1)\to Y_1/(Y_1+Y_2)\sim Beta(a,b)$
$X_i\sim U[0,1]\to X_{(k)}\sim Beta(k,n-k+1)$; $Beta(1,1)=U[0,1]$

指数分布 Exponential distribution:
$f(x|\beta)=\beta e^{-\beta x}$, β>0, x≥0; $E(X)=1/\beta$; $Var(X) = 1/\beta^2$
$-\log(U(0,1))/\beta\sim Exp(\beta)$; $X_i\sim Exp(\beta)\to\sum X_i\sim Gamma(n,\beta)$

卡方分布 Chi-square distribution: $\chi^2(v)$ v>0
$f(x|v)=2^{-v/2}x^{v/2-1}e^{-x/2}/\Gamma(v/2)$, x>0; E(X)=v; Var(X)=2v
$Y\sim N(0,1)\to X=Y^2\sim\chi^2(1)$; $X_i\sim\chi^2(v_i)\to\sum X_i\sim\chi^2(\sum v_i)$
$$X_i \overset{i.i.d}{\sim} N(\mu,\sigma^2)\to\frac{(n-1)S^2}{\sigma^2}\sim\chi^2(n-1)$$

正态分布: $N(u,\sigma^2)$
$f(x)=e^{-(x-\mu)(x-\mu)/(2\sigma\sigma)}/(\sigma\sqrt{2\pi})$, $x\in R$; $E(X)=\mu$; $Var(X)=\sigma^2$;
$M_X(t)=\exp(\mu t+\sigma^2t^2/2)$, $E_{N(0,1)}(X^{2n})=(2n-1)!!$, $E_{N(0,1)}(X^{2n+1})=0$
$X_i\sim N(\mu_i,\sigma_i^2)\to\sum a_iX_i\sim N(\sum a_i\mu_i, \sum a_i^2\sigma_i^2)$
$X_1|X_2\sim N(X_2,\sigma_1^2)$, $X_2\sim N(\mu_2,\sigma_2^2)\to X_1\sim N(\mu_2,\sigma_1^2+\sigma_2^2)$

t 分布: t(v) v > 0
$f(x|v)=\Gamma((v+1)/2)(1+x^2/v)^{-(v+1)/2}/(\Gamma(v/2)\sqrt{v\pi})$, $x\in R$
E(X)=0(v>1); Var(X)=v/(v-2) (v>2) (v=1→Cauchy)
$$Z\sim N(0,1), Y\sim\chi^2(v)\to\frac{Z}{\sqrt{Y/v}}\sim t(v)$$
$$X_i\overset{i.i.d}{\sim}N(\mu,\sigma^2)\to\frac{(\overline{X}-\mu)\sqrt{n}}{S}\sim t(n-1)$$

F 分布: $F(n_1,n_2)$ x > 0
$f(x)=(n_1/n_2)^{n_1/2}x^{n_1/2-1}(1+n_1x/n_2)^{-(n1+n2)/2}/B(n_1/2,n_2/2)$
$E(X)=n_2/(n_2-2)$ (n_2 > 2)
$Var(X)=2n_2^2(n_1+n_2-2)/(n_1(n_2-4)(n_2-2)^2)$ (n_2 > 4)
$Y_i\sim\chi^2(n_i)\to Y_1n_2/(Y_2n_1)\sim F(n_1,n_2)$

拉普拉斯分布 Laplace distribution
$f(x)=e^{-|x-\mu|/\sigma}/(2\sigma)$, $x,\mu\in R$ $\mu^{MLE}=med(\mathbf{x})$

单调变换: (y=h(x) 单调且可微)
$g(y)=f(x)\times|dx/dy|=f(h^{-1}(y))\times|dh^{-1}(y)/dy|$
二元变换: $Y_1=h_1(X_1,X_2)$, $Y_2=h_2(X_1,X_2)$;
$J=\partial(x_1,x_2)/\partial(y_1,y_2)$
$g(y_1,y_2)=f(x_1,x_2)\times|J| = f(h_1^{-1}(y_1,y_2),h_2^{-1}(y_1,y_2))\times|J|$

样本方差: $$S^2=\frac{\sum_{i=1}^n(X_i-\overline{X})^2}{n-1}$$

顺序统计量 Order Statistics
$F_{X(n)}(x)=F^n(x)$; $f_{X(n)}(x)=nf(x)F^{n-1}(x)$
$F_{X(1)}(x)=1-(1-F(x))^n$; $f_{X(1)}(x)=nf(x)(1-F(x))^{n-1}$
$F_{X(r)}(x)=\int_{[0,F(x)]}(t^{r-1}(1-t)^{n-r})dt/B(r,n-r+1)$
$=\sum_{[r,n]}C_n^iF^i(x)(1-F(x))^{n-i}$
$f_{X(r)}(x)=n!f(x)F^{r-1}(x)(1-F(x))^{n-r}/((r-1)!(n-r)!)$
$f_{med}(x)=f(x)F^m(x)(1-F(x))^m(2m+1)!/(m!)^2$ (n=2m+1)
$f_{X(1),\dots,X(n)}(\mathbf{x})=n!f_X(x_{(1)})\dots f_X(x_{(n)})$

强收敛: $Pr(X_n=X(n\to\infty))=1$(almost surely) 「a.s.」
均方收敛: $E(X_n-X)^2=0(n\to\infty)$ 「m.s.」
→弱收敛: $Pr(|X_n-X|\ge\varepsilon)=0$ (n→∞) 「P」
→依分布收敛: $F_n(x)=F(x)$ (n→∞) 「L」
弱大数律: $Pr(|X_{nbar}-u|\ge\varepsilon)\le Var(X_{nbar})/\varepsilon^2=\sigma^2/n\varepsilon^2\to 0$
强大数律: $E(X_i)=u<\infty$, X_i i.i.d. → $\sum_{[1,n]}X_i/n\to u$(a.s.)
中心极限定理: X_i i.i.d., $0<\sigma^2<\infty$, Common μ, σ<∞→
$$\frac{\sqrt{n}(\overline{X}-\mu)}{\sigma}\xrightarrow{P}N(0,1)$$

似然函数 Likelihood function
$L(\vec\theta)=f(x_1,\dots,x_n;\vec\theta)=\prod_{[1,n]}f(x_i;\vec\theta)$; $l(\vec\theta)=\log(L(\vec\theta))$

极大似然估计量 Maximum Likelihood Estimator (MLE)
$\vec\theta^*=\arg\max_{\theta\in\Theta}L(\vec\theta)=\arg\max_{\theta\in\Theta}l(\vec\theta)$, 一般有 $\nabla l(\vec\theta^*)=\vec 0$
MLE(θ_{hat}) 是随机变量的函数，mle（θ*）是样本的
定理 3.1 MLE 不变性：若θ^*是 $g(X_1,\dots,X_n)$的 MLE，$\eta=h_{p\times1}(\theta)$是一个一一映射，则$\eta^*=h(\theta^*)$ 是η的 MLE
定理 3.2：若θ^*是 $g(X_1,\dots,X_p)$的 MLE，$\eta_{r\times1}=h(\theta)$，1≤r≤p，则$\eta^*=h(\theta^*)$ 是η的 MLE

矩估计量 Moment Estimator
样本矩：$\sum_{[1,n]}X_i^r/n$ 总体矩：$E(X^r)$，构建二者的相等关系

$f(\mathbf{x};\theta)$=Likelihood×Prior$=f(\mathbf{x}|\theta)\times\pi(\theta)=\prod_{[1,n]}f(x_i|\theta)\times\pi(\theta)$
后验分布 Posterior density：$p(\theta|\mathbf{x})=f(\mathbf{x};\theta)/f(\mathbf{x})\propto f(\mathbf{x};\theta)$
$f(\mathbf{x})=\int_\theta f(\mathbf{x},\theta)d\theta$
贝叶斯估计量 Bayesian Estimator
$E(\theta|\mathbf{x})=\int_\theta\theta\cdot p(\theta|\mathbf{x})d\theta$

无偏估计量：$E(\varphi(x))=\theta$；偏差 $b(\theta)=E(\varphi(x))-\theta$
样本均值和样本方差是总体均值和总体方差的无偏估计量
均方误差 $MSE=E(\varphi(\mathbf{x})-\theta)^2$

Score Function: $S(\theta)=L'(\theta)/L(\theta)=dl(\theta)/d\theta=S(\theta;x)$
Fisher Information: $I_n(\theta) = Var_x(S(\theta;x))$
定理 3.3：CR 下界不等式：$Var(\hat\theta)\ge(\tau'(\theta))^2/I_n(\theta)$
其中，θ_{hat}是τ(θ)的无偏估计，且 f(x;θ)的存在域与θ无关
此时，有 $E(S(\theta;x))=0\to I_n(\theta)=E(S^2(\theta;x))$

定理 3.4: 若 E(S(θ))=0，则 $I_n(\theta)=E(-d^2l(\theta;\mathbf{X})/d\theta^2)=nI(\theta)$，其中 $I(\theta)=E(d\log f(X;\theta)/d\theta)^2=E(-d^2\log f(X;\theta)/d\theta^2)$

有效估计量：θ_{hat}是τ(θ)的无偏估计，且其方差为 CR 下界
有效性 Efficiency: $Eff_{\theta hat}(\theta)=CR/Var(\theta_{hat})=1/I_n(\theta)Var(\theta_{hat})$
UMVUE: 一致最小方差无偏估计量（不存在或唯一）
有效估计量→UMVUE
充分统计量：假如给定统计量 T(**X**)=t，可以使 X 的条件分布不再依赖于参数θ，则称 T(**X**)是θ的充分统计量。
定理 3.5/6 因子分解定理 Factorization：若 X 的分布可写作 f(x;θ)=g(T(X);θ)×h(X)，则 T(X)是θ的充分统计量
完备统计量：假如对于任意θ属于Θ，均有：对于函数 h(T)，E(h(T))=0→Pr(h(T)=0)=1，则 T 是θ的完备统计量。
Lehmann-Scheffé定理：如果 T(X)是θ的完备且充分的统计量，g(T)是τ(θ)的无偏估计量，则 g(T)是τ(θ)唯一的 UMVUE。
MLE 的极限性质：$\theta_n^{MLE}\to\theta(n\to\infty)$「P, L」
MLE 的渐近性质：若 E(S(θ;X))=0，Var(S(θ;X))=nI(θ)，则
$$\frac{S(\theta;X_1,\dots,X_n)}{\sqrt{nI(\theta)}}\xrightarrow{L}N(0,1)(n\to\infty)$$，且
$$(\theta_n^{MLE}-\theta)\sqrt{nI(\theta)}\xrightarrow{L}N(0,1)(n\to\infty)$$，进而对一般函数 g
$$(g(\hat\theta_n)-g(\theta))\sqrt{nI(\theta)}/g'(\theta)\xrightarrow{L}N(0,1)(n\to\infty)$$
这表明θ_n^{MLE}是θ的一个渐近无偏估计量，而且是渐近 UMVUE，同时还渐近地正态分布。(Asymptotic)

枢轴量 Pivotal quantity/Pivot:
X_i i.i.d f(x;θ)，T=T(**X**)是θ的充分统计量，如果 P=P(T,θ)满足 P 的分布与θ无关，则 P 是一个枢轴量。
构造枢轴量：$-2\sum_{[1,n]}\log F(X_i;\theta)\sim\chi^2(2n)$

已知方差构造正态分布均值的置信区间：（中心极限定理）
$$\left[\overline{X}-\frac{z_{\alpha/2}\sigma_0}{\sqrt{n}},\overline{X}+\frac{z_{\alpha/2}\sigma_0}{\sqrt{n}}\right]$$ $z_{\alpha/2}$ 为 N(0,1)的上α/2 分位点

未知方差构造正态分布均值的置信区间：（t 分布的性质）
$$\left[\overline{X}-\frac{t(\alpha/2,n-1)S}{\sqrt{n}},\overline{X}+\frac{t(\alpha/2,n-1)S}{\sqrt{n}}\right]$$

大样本下构造置信区间的三种办法：
I.中心极限定理
II.利用 S(θ,X)的渐近性质 $$1-\alpha\approx Pr(-z_{\alpha/2}\le\frac{S(\theta;X)}{\sqrt{nI(\theta)}}\le z_{\alpha/2})$$
III.利用 MLE 的渐近性质
$$1-\alpha\approx Pr(-z_{\alpha/2}\le(\theta_n^*-\theta)\sqrt{nI(\theta)}\le z_{\alpha/2})$$

构造两个正态分布均值之差的置信区间：
两个方差已知：
$$\sigma_0=\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}$$
枢轴量：$$Z=\frac{(\overline{X}_1-\overline{X}_2)-(\mu_1-\mu_2)}{\sigma_0}\sim N(0,1)$$
方差未知但相等：
$$\frac{(n_i-1)S_i^2}{\sigma^2}\sim\chi^2(n_i-1);\quad S_p^2=\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$$
$$Y=\frac{(n_1+n_2-2)S_p^2}{\sigma^2}\sim\chi^2(n_1+n_2-2)$$
$$\frac{Z}{\sqrt{Y/(n_1+n_2-2)}}=\frac{(\overline{X}_1-\overline{X}_2)-(\mu_1-\mu_2)}{S_p\sqrt{1/n_1+1/n_2}}\sim t(n_1+n_2-2)$$
CI: $$X_1-X_2\mp t(\frac{\alpha}{2},n_1+n_2-2)\cdot S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$$
方差未知：
$$T_{Welch}=\frac{(\overline{X}_1-\overline{X}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}}\sim t(v)$$
$$t(v)=1/(\frac{c^2}{n_1-1}+\frac{(1-c)^2}{n_2-1}), c=\frac{S_1^2/n_1}{S_1^2/n_1+S_2^2/n_2}$$

已知均值构造正态分布方差的置信区间
$$P=\sum_{i=1}^n\left(\frac{X_i-\mu_0}{\sigma}\right)^2\sim\chi^2(n)$$
未知均值构造正态分布方差的置信区间
$$P=\frac{(n-1)S^2}{\sigma^2}\sim\chi^2(n-1)$$
构造两个正态分布方差之比的置信区间
$$\frac{\sigma_2^2S_1^2}{\sigma_1^2S_2^2}\sim F(v_1,v_2); f(1-\frac{\alpha}{2},v_1,v_2)=\frac{1}{f(\frac{\alpha}{2},v_2,v_1)}$$
$v_i = n_i - 1$; $$\left[\frac{S_1^2}{S_2^2}\cdot\frac{1}{f(\alpha/2,v_1,v_2)},\frac{S_1^2}{S_2^2}\cdot f(\alpha/2,v_2,v_1)\right]$$

基础知识：
$Var(\sum c_iX_i)=\sum c_i^2Var(X_i)+2\sum_{i<j}Cov(X_i,X_j)$
$Cov(X_1,X_2)=E(X_1X_2)-E(X_1)E(X_2)$
$Pr(X\le med(X))\ge0.5$; $Pr(X\ge med(X))\ge0.5$
$E(g(X))=E(E(g(X)|Y))=\int E(g(X)|Y=y)f(y)dy$
$Var(g(X))=E(Var(g(X)|Y))+Var(E(g(X)|Y))$
$M_X(t)\ge e^{tE(X)}$ (Jensen)
$\rho=Corr(X,Y)=Cov(X,Y)/\sqrt{Var(X)Var(Y)}$
r 阶矩 r^{th} moment (about the origin): $u'_r=E(X^r)$
r 阶中心矩 r_{th} central moment (about the mean): $u_r=E(X-u)^r$
$\mu_r=\sum_{[0,r]}(-1)^iC_r^i\mu'_{r-i}\mu^i$
期望: μ_1 衡量 Central Location 方差: σ^2 衡量 Dispersion （散度）
斜度 Skewness: μ_3/σ^3 衡量 Asymmetry 左斜-长尾在左-斜度为负
峰度 Kurtosis: μ_4/σ^4 衡量 Flatness

乱七八糟的知识：
Probability Generating Function 概率生成函数 pgf.
$G_X(z)=E(z^X)=\sum_{[x\in Sx]}z^xp_X(x)$ 显然 $G(e^t)=M_X(t)$
Multi-dimensional Normal Distribution:
$$N_d(\vec x|\vec\mu,\mathbf{\Sigma})=\frac{1}{\sqrt{(2\pi)^d|\Sigma|}}e^{-\frac{(\vec x-\vec\mu)^T\Sigma^{-1}(\vec x-\vec\mu)}{2}}$$
其中Σ=(Cov(X_i,X_j))正定
$$M_{(X_1,X_2)}(t_1,t_2)=\exp(\mu_1t_1+\mu_2t_2+(\sigma_1^2t_1^2+\sigma_2^2t_2^2+2\rho\sigma_1\sigma_2t_1t_2)/2)$$
$$a_1X_1+a_2X_2\sim N(a_1\mu_1+a_2\mu_2,a_1^2\sigma_1^2+a_2^2\sigma_2^2+2a_1a_2\rho\sigma_1\sigma_2)$$

Stochastic representation 随机表示
若 X 与 $g(Y_1,\dots,Y_n)$同分布，即 $X\overset{d}{=}g(Y_1,\dots,Y_n)$，则称此式为一个「One-to-many SR」of the r.v. X
Statistic 统计量
X_i i.i.d. F(x). An arbitrary function $T(X_1,\dots,X_n)$ is called a Statistic.
Or: A function of one or more r.v.'s that does not depend on any unknown parameters is called a Statistic.

定理 2.1：
$A_{m\times n}$, $B_{r\times n}$ 是两个标量矩阵，$\mathbf{X}\sim N_n(\mu,\Sigma)$，则
（1）$A\mathbf{X}\sim N_m(A\mu, A\Sigma A^T)$；
（2）$B\mathbf{X}\sim N_r(B\mu, B\Sigma B^T)$；
（3）AX 与 BX 独立当且仅当 $A\Sigma B^T=0_{m\times r}$
定理 2.5：对于任意|t|＜h，$M_{Xi}(t)$均存在时
如果 $\lim_{x\to\infty}M_{X_n}(t)=M_X(t)$，则 $X_n\xrightarrow{L}X(n\to\infty)$。

Newton-Raphson&Fisher Scoring Algorithm
$$x_{t+1}=x_t-\frac{g(x_t)}{g'(x_t)}\longrightarrow\theta_{t+1}=\theta_t-\frac{l'(\theta_t)}{l''(\theta_t)}$$
逆贝叶斯公式: (sampling-wise 为正比关系)(f(x,y)必须存在)
$$f_X(x)f_{Y|X}(y|x)=f_Y(y)f_{X|Y}(x|y)=f(x,y)$$
$$f_X(x)=\frac{f_{X|Y}(x|y_0)}{f_{Y|X}(y_0|x)}/\int_{S_X}\frac{f_{X|Y}(x|y_0)}{f_{Y|X}(y_0|x)}dx$$ (function-wise)
$$f_Y(y)=1/\int_{S_X}\frac{f_{X|Y}(x|y)}{f_{Y|X}(y|x)}dx$$ (point-wise)