$P(C_j|C) = \frac{P(C\cap C_j)}{P(C)} = \frac{P(C|C_j)P(C_j)}{\sum_{i=1}^{k} P(C|C_i)P(C_i)}$
where $\Omega = \bigcup_{i=1}^{k} C_i,\ C_i \cap C_j = \phi$

mgf of X is $M(t) \equiv E[\exp(tX)]$
mgf is unique and 1:1 with distribution
$M(t) = M(t_1, \ldots, t_n) \equiv E[\exp(t'X)]$

(Def.) X_1 and X_2 are **independent** if
(1) $f(x_1, x_2) = f_1(x_1) f_2(x_2)$
(2) $\exists g, h$ s.t $f(x_1, x_2) = g(x_1) h(x_2)$
(3) $M(t_1, t_2) = M(t_1, 0) \cdot M(0, t_2)$

(Thm.) If X_1 and X_2 are independent
(1) $\Pr(a < X_1 < b, c < X_2 < d)$
$= \Pr(a < X_1 < b) \cdot \Pr(c < X_2 < d)$
(2) $E[u(X_1)v(X_2)] = E[u(X_1)]E[v(X_2)]$
(3) $Cov(X_1, X_2) = 0$

Let $Y = u(X)$ and u is 1:1 (monotone)
$f_Y(y) = f_X[u^{-1}(y)]$ (discrete)
$f_Y(y) = f_X[u^{-1}(y)] \left|\frac{d}{dy}u^{-1}(y)\right|$ (cont.)
where $u^{-1}(y) = \binom{V_1(y)}{V_2(y)}$
$$\left|\frac{d}{dy}u^{-1}(y)\right| = \left|det\begin{pmatrix} \frac{\partial V_1}{\partial y_1} & \frac{\partial V_1}{\partial y_2} \\ \frac{\partial V_2}{\partial y_1} & \frac{\partial V_2}{\partial y_2} \end{pmatrix}\right|$$

$\{Y_n\}$ **converges in probability** to Y
if $\lim_{n\to\infty} \Pr[|Y_n - Y| \geq \varepsilon] = 0, \forall \varepsilon > 0$

$\{Y_n\}$ **converges almost sure** to Y
if $P[\omega|\lim_{n\to\infty} Y_n(\omega) = Y(\omega)] = 1$

$\{Y_n\}$ **converges in distribution** to Y
if $\lim_{n\to\infty} F_n(y) = F(y)$
for every continuous point of F

Law of Large Numbers $(Y_i \sim iid)$
$E[Y_i] = \mu$ and $Var(Y_i) = \sigma^2 < \infty$
We then have $\text{plim}_{n\to\infty} \bar{Y}_n = \mu$

Central Limit Theorem $(X_i \sim iid)$
$E[X_i] = \mu$ and $Var(X_i) = \sigma^2$ or Σ
$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_i - \mu) = \sqrt{n}\left(\bar{X} - \mu\right) \xrightarrow{d} N(0, \sigma^2) \quad \text{or} \quad N(0, \Sigma)$$
$\sqrt{n}$ is rate of convergence.

· X_i : independent & $Y = \sum_{i=1}^{n} a_i X_i$
$\Rightarrow M_Y(t) = \prod_{i=1}^{n} M_i(a_i t)$
· $Y_n \xrightarrow{d} c \Leftrightarrow Y_n \xrightarrow{p} c$
· $\lim_{n\to\infty} M_n(t) = M_Y(t) \Rightarrow Y_n \xrightarrow{d} Y$

Markov's inequality:
$\Pr[u(X) \geq c] \leq \frac{E[u(X)]}{c} \quad \forall c > 0$

Chebyshev's inequality:
$\Pr[|X - \mu| \geq c\sigma] \leq \frac{1}{c^2}$
$\Pr[|X - \mu| \geq \varepsilon] \leq \frac{\sigma^2}{\varepsilon^2}$

Jensen's ineqaulity:
$g(E[X]) \leq E[g(X)] \quad \forall g'' > 0$

$X \sim b(n, p)$
$f(x) = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}$
$M(t) = (1 - p + pe^t)^n$
$E[X] = np, Var(X) = np(1-p)$

$X \sim Poisson(\lambda)$
$f(x) = \frac{\lambda^x e^{-\lambda}}{x!}$
$M(t) = \exp[\lambda(e^t - 1)]$
$E[X] = Var(X) = \lambda$

$X \sim \Gamma(\alpha, \beta)$
$f(x) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-\frac{x}{\beta}}, (x > 0)$
$M(t) = (1 - \beta t)^{-\alpha}$
$E[X] = \alpha\beta, Var(X) = \alpha\beta^2$
where $\Gamma(\alpha) = \int_0^\infty y^{\alpha-1} e^{-y} dy$
$\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$,
$\Gamma(1) = 1$, $\Gamma(\frac{1}{2}) = \sqrt{\pi}$
$\alpha = 1, \beta = \frac{1}{\lambda}$ implies
$f(x) = \lambda e^{-\lambda x}$ (exponential)
$\alpha = \frac{n}{2}, \beta = 2$ implies χ^2_n

$X \sim beta(\alpha, \beta)$
$f(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}$
$E = \frac{\alpha}{\alpha+\beta}, V = \frac{\alpha\beta}{(\alpha+\beta+1)(\alpha+\beta)^2}$

$X \sim N(\mu, \sigma^2)$
$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]$
$M(t) = \exp\left[\mu t + \frac{\sigma^2}{2} t^2\right]$

$X \sim N(\mu, \Sigma)$
$$f(x) = \frac{1}{(2\pi)^{\frac{n}{2}} \sqrt{|det\Sigma|}} \exp\left[-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)\right]$$
$M(t) = \exp\left[\mu' t + \frac{t'\Sigma t}{2}\right]$

$X \sim U(a, b)$
$f = \frac{1}{b-a}, E = \frac{a+b}{2}, V = \frac{(b-a)^2}{12}$

$X \sim t_n$
$f(x) = \frac{\Gamma[(n+1)/2]}{\sqrt{\pi n}\Gamma[n/2]} \cdot \left(1 + \frac{y^2}{n}\right)^{-\frac{n+1}{2}}$

conditional pdf of X_2 given X_1
$f(x_2|x_1) = \frac{f(x_1, x_2)}{f_1(x_1)}$
$E[u(X_2)|x_1] = \int u(x_2) f(x_2|x_1) dx_2$

Law of Iterated Expectation :
$E[X_2] = E[E[X_2|X_1]]$

Correlation coefficient:
$\rho = \frac{E[(X-\mu_X)(Y-\mu_Y)]}{\sigma_X \sigma_Y}$

$X \sim N(\mu, \sigma^2) \Rightarrow Y = \frac{X-\mu}{\sigma} \sim N(0, 1)$

$X \sim N(0, 1) \Rightarrow X^2 \sim \chi^2(1)$.

$X_i \overset{iid}{\sim} N(0, 1) \Rightarrow \sum_{i=1}^{n} X_i^2 \sim \chi^2(n)$.

$X_i \overset{iid}{\sim} N(0, 1) \Rightarrow X_{n\times 1} \sim N(0, I_n)$

$X \sim N(\mu, \Sigma) \Rightarrow$ for $m \times n$ A,
$Y = AX \sim N(A\mu, A\Sigma A')$

$X = \binom{X_1}{X_2} \sim N(\mu, \Sigma)$, then
X_1 and X_2 are independent iff
$E\left[(X_1 - \mu_1)(X_2 - \mu_2)'\right] = 0$

$X = \binom{X_1}{X_2} \sim N(\mu, \Sigma) \Rightarrow X_1|X_2 \sim N$
with mean= $\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(X_2 - \mu_2)$
variance= $\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$

$X \sim N(0, I_n)$, then $AB' = 0$ iff
AX and BX are independent

$X \sim N(0, I_n) \Rightarrow X'X \sim \chi^2(n)$

$Z \sim N(0, 1), nY \sim \chi^2_n \Rightarrow \frac{Z}{\sqrt{Y}} \sim t_n$

· $var(X) = E[XX'] - \mu\mu'$
· $var(AX) = Avar(X)A'$
· $U \sim U(0, 1), F$: cdf
$\Rightarrow Y = F^{-1}(U)$ has cdf F
· $\sum_{x=0}^{\infty} \frac{k^x}{x!} = e^k$
· $\lim_{n\to\infty} \left(1 + \frac{b}{n}\right)^{cn} = e^{bc}$
· $\frac{d}{dx} \int_a^{b(x)} F(t) dt = F(b(x)) \cdot b'(x)$

$h(\cdot)$: continuous, $\text{plim}_{n\to\infty} Y_n = c$
$\Rightarrow \text{plim}_{n\to\infty} h(Y_n) = h(c)$

Slutsky: $Y_n \xrightarrow{p} c, X_n \xrightarrow{d} X$
$\Rightarrow Y_n X_n \xrightarrow{d} cX$

Delta: $\sqrt{n}(\bar{X} - \mu) \xrightarrow{d} N(0, \sigma^2)$
$\Rightarrow \sqrt{n}[u(\bar{X}) - u(\mu)]$
$\xrightarrow{d} N(0, [u'(\mu)]^2\sigma^2)$
$(or) \quad \xrightarrow{d} N(0, G\Sigma G')$
for multivariate: $G = \frac{du(\mu)}{dX'}$

$\hat{\theta}(X)$ is unbiased if $E[\hat{\theta}(X)] = \theta$
$\hat{\theta}(X)$ is consistent if $\hat{\theta}(X) \xrightarrow{p} \theta$
$\hat{\theta}(X)$ is MLE
if $\hat{\theta}(X) = \arg\max_{\theta\in\Theta} L(X, \theta)$

Invariance Theorem:
$\hat{\theta}$: MLE of θ, and h:monotone
$\Rightarrow h(\hat{\theta})$: MLE of $h(\theta)$

Method of Moments:
$E[X_i^j] = M_j(\theta_1, \cdots, \theta_K)$ $_{j=1,\cdots,K}$
M is 1-1 $\Rightarrow (\hat{\theta}_1, \cdots, \hat{\theta}_K)$
$= M^{-1}\left(\frac{1}{n}\sum X_i, \cdots, \frac{1}{n}\sum X_i^K\right)$

Measures of **estimator quality**:
$\hat{\theta}$ is UMVE of θ if $E[\hat{\theta}] = \theta$ and
$var(\hat{\theta}) \le var(\tilde{\theta})$ when $E[\tilde{\theta}] = \theta$

Loss fnction: $L[\theta, \hat{\theta}(x)]$
Risk: $R(\theta, \hat{\theta}) = E[L(\theta, \hat{\theta}[X])]$
Risk Minimum Estimator :)

MSE $= E[\{\hat{\theta}(X) - \theta\}^2]$
$= var(\hat{\theta}) + (bias)^2$

BLUE: best linear unbiased e...
$\hat{\beta} = (X'X)^{-1}X'y$ is BLUE

$X \sim f(x, \theta), Y = u(X) \sim g(y, \theta)$
$E[X|Y = y]$ doesn't rely on θ
$\Rightarrow Y$ is **sufficient** for θ

Factorization theorem:
$Y = u(X)$ is sufficient for θ
$\Leftrightarrow f(x, \theta) = h_1[u(x), \theta]h_2(x)$

Rao-Blackwell theorem:
$Y = u(X)$ is sufficient for θ
$E[\hat{\theta}] = \theta$, $\phi(y) = E[\hat{\theta}|Y = y]$
$\Rightarrow E[\phi(Y)] = \theta, var(\cdot) \le var(\hat{\theta})$

Score:
$S(x, \theta) = \frac{\partial log f(X,\theta)}{\partial\theta}$
$E[S(X, \theta)] = 0$

Fisher information $I(\theta)$:
$I(\theta) = E\left[\left(\frac{\partial log f(X,\theta)}{\partial\theta}\right)^2\right]$
$= -E\left[\frac{\partial^2 log f(X,\theta)}{\partial\theta^2}\right]$
$= E\left[S(X, \theta), S(X, \theta)'\right]$
$= -E\left[\frac{\partial^2 log f(X,\theta)}{\partial\theta\partial\theta'}\right]$

$X = (X_1, \cdots, X_n)'$ and
$X_i \sim$ iid $\Rightarrow I(\theta) = nJ(\theta)$

Cramer-Rao inequality:
$E[\hat{\theta}] = \theta$
$\Rightarrow var(\hat{\theta}) \ge I(\theta)^{-1}$
$I(\theta)^{-1}$: CR lower bound

Unbiased $\hat{\theta}$ is **efficient**
if $var(\hat{\theta}) = I(\theta)^{-1}$

$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} N(0, \sigma^2)$
$\Rightarrow \hat{\theta}_n$ is asympt. efficient
if $\sigma^2 = I(\theta)^{-1}$
* Info Mat of original pdf

limiting dist. of MLE
MLE is asympt. efficient
$\sqrt{n}(\hat{\theta}_{MLE} - \theta)$
$\xrightarrow{d} N(0, I(\theta)^{-1})$
* Info Mat of original pdf

Bayesian Estimation:
$K(\theta|x) = \frac{L(x|\theta)h(\theta)}{\int_{-\infty}^{\infty} L(x|\theta)h(\theta)d\theta}$
(posterior)

C: critical region
reject H_0 when $x \in C$
Power$(\theta) = Pr(X \in C|\theta)$
Pr(Type I error)$=Pr(X \in C|\theta \in \Theta_0)$
Pr(Type II error)$=Pr(X \in C^c|\theta \in \Theta_1)$

Size α of a test is
$\alpha = \max_{\theta\in\Theta_0} Pr(X \in C)$

C is optimal for tesing H_0 against H_1
if for any critical region A of size α,
$Pr(X \in C) \ge Pr(X \in A)$ under H_1

Neyman-Pearson theorem:
The optimal test has
$C = \left\{x \,\middle|\, \frac{f(x,\theta_1)}{f(x,\theta_0)} \ge k\right\}$

Uniformly Most Powerful test: a test with
C optimal against any $\mu = \mu_1 \in H_1$

LR (Likelyhood Ratio) test:
$X \sim f(x, \theta), H_0 : \theta \in \Theta_0, H_1 : \theta \in \Theta_1$
$LR = \frac{\sup_{\theta\in\Theta} f(x,\theta)}{\sup_{\theta\in\Theta_0} f(x,\theta)}$
LR test is $C = \{x|LR \ge k\}$
where k is chosen so that the size $= \alpha$

Asymptotic test: LR,Wald,LM$\sim \chi_k^2$
$2\left(\sum_{i=1}^n \log f(X_i, \hat{\theta}) - \sum_{i=1}^n \log f(X_i, \theta_0)\right)$
$n(\hat{\theta} - \theta_0)'I(\hat{\theta})(\hat{\theta} - \theta_0)$
$\frac{1}{n}[\sum_{i=1}^n S(X_i, \theta_0)]'I(\theta_0)^{-1}[\sum_{i=1}^n S(X_i, \theta_0)]$

Linear Regression Model:
$$y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}, X = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$
$y = X\beta + \varepsilon$, $\varepsilon \sim N(0, \sigma^2 I)$
$\hat{\beta} = (X'X)^{-1}X'y = \beta + (X'X)^{-1}X'\varepsilon$
$\sim N(\beta, \sigma^2(X'X)^{-1})$
$\hat{u} = y - X\hat{\beta} = y - X(X'X)^{-1}X'y$
$= My = M(X\beta + \varepsilon)$
$= (I - X(X'X)^{-1}X')X\beta + M\varepsilon = M\varepsilon$
$\frac{(n-k)s^2}{\sigma^2} = \frac{\hat{u}'\hat{u}}{\sigma^2} = (\frac{\varepsilon}{\sigma})'M(\frac{\varepsilon}{\sigma}) \sim \chi_{n-k}^2$

$\hat{\beta}$ is independent of s^2
$c'\hat{\beta} \sim N(c'\beta, \sigma^2 c'(X'X)^{-1}c)$
$t = \frac{c'(\hat{\beta}-\beta)}{s\sqrt{c'(X'X)^{-1}c}} \sim t_{n-k}$

$\frac{da'x}{dx} = \frac{dx'a}{dx} = a \qquad \frac{dAx}{dx} = \frac{dx'A'}{dx} = A'$

$\frac{d(Ax)'(Bx)}{dx} = (A'B + B'A)x$