Basics

- **Power laws:** $x^a x^b = x^{a+b}$, $\frac{x^a}{x^b} = x^{a-b}$, $(x^a)^b = x^{ab}$
- **Log rules:** $\log(xy) = \log(x) + \log(y)$, $\log(x/y) = \log(x) - \log(y)$, $\log(x^a) = a\log(x)$, $\ln 1 = 0$, $\ln e = 1$, $\ln\sqrt{x} = \frac{1}{2}\ln x$
- **DeMorgan:** $(A \cup B)^c = A^c \cap B^c$ and $(A \cap B)^c = A^c \cup B^c$
- **Derrivatives:** $(fg)' = f'g + fg'$, $(f/g)' = (f'g - fg')/g^2$, $(f(g))' = f'(g)g'$, $(ae^x)' = ae^x$, $(\ln x)' = 1/x$
- **Integrals:** $\int f(x)g'(x)dx = f(x)g(x) - \int f'(x)g(x)dx$, $\int e^{ax}dx = \frac{1}{a}e^{ax} + C$, $\int \ln(x)dx = x\ln(x) - x + C$
- **Integration by parts:** $\int u\,dv = uv - \int v\,du$. $u :=$ fn that simp. when different.
- **QF:** $x = \frac{-b\pm\sqrt{b^2-4ac}}{2a}$ for $ax^2 + bx + c = 0$.
- **MatMul:** If $A : m \times n, B : n \times p$ and $C = AB, c_{ij} = \sum_{k=1}^n a_{ik}b_{kj}$.
- **Determinant** of 2×2 mat is $ad - bc$, 3×3 is $a(ei - fh) - b(di - fg) + c(dh - eg)$.
- **Eigenvalue** of A is λ where $Av = \lambda v$ for some non-zero v. λ found by solving $\det(A - \lambda I) = 0$.
- **Spectral:** If A sym., then $A = Q\Lambda Q'$ with orthogonal Q of v's and $\text{diag}(\lambda_i)$.

Probability

- **Probability measure** P on a σ-algebra $\mathcal{C}$ over sample space Ω is fn $P : \mathcal{C} \to [0,1]$ that satisfies (1) $\forall A \in \mathcal{C}$, $P(A) \geq 0$, (2) $P(\Omega) = 1$, (3) $\forall i \in \mathbb{N}, A_i \in \mathcal{C}$, $A_i \cap A_j = \emptyset$ for $i \neq j$ implies $P\left(\bigcup_{i=1}^\infty A_i\right) = \sum_{i=1}^\infty P(A_i)$. This implies (1) $P(\emptyset) = 0$, (2) $P(A^c) = 1 - P(A)$, (3) $P(A\cup B) = P(A) + P(B) - P(A\cap B)$.
- **Bayes' rule:** For events A, B where $P(B) > 0$, $P(A|B) = \frac{P(B|A)P(A)}{P(B)}$. $P(A|B)$ satisfies the axioms of a probability measure.
- **Law of total probability:** For events $B, A_1, A_2, \ldots$ that form a partition of the sample space, $P(B) = \sum_{i=1}^\infty P(B|A_i)P(A_i)$.
- Events A, B are **independent** if $P(A \cap B) = P(A)P(B) \implies$ if $P(B) > 0$ independence is given when $P(A|B) = P(A)$

Random Variables/Vectors

- A **random variable** is a measurable fn $X : \Omega \to \mathbb{R}$.
- The **cdf** of X is $F_X(x) = P(X \leq x)$. The **pmf** of a discrete r.v. X is $p_X(x) = P(X = x)$. The **pdf** of a continuous r.v. X is $f_X(x)$ where $F_X(x) = \int_{-\infty}^x f_X(t)dt$. A r.v. is contінous if its cdf can be expressed as an integral of a pdf.
- For r.v.s X, Y, the **joint cdf** is $F_{X,Y}(x,y) = P(X \leq x, Y \leq y) = P(\{\omega : X(\omega) \leq x\} \cap \{\omega : Y(\omega) \leq y\})$. The **marginal cdf** is $F_X(x) = \lim_{y\to\infty} F_{X,Y}(x,y)$. The **marginal pdf** is $f_X(x) = \int_{-\infty}^\infty f_{X,Y}(x,y)dy$.
- Two r.v.s are **independent** if $F_{X,Y}(x,y) = F_X(x)F_Y(y) \iff f_{X,Y}(x,y) = f_X(x)f_Y(y)\ \forall x, y$. Two r.v.s can not be independent if their support depends on each other.
- **Conditional pdf:** $f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$ if $f_Y(y) > 0$. The conditional cdf is $F_{X|Y}(x|y) = \int_{-\infty}^x f_{X|Y}(t|y)dt$. Independence implies $f_{X|Y}(x|y) = f_X(x)$.
- **Transformation**: If X is r.v. and $Y = g(X)$ then $f_Y(y) = f_X(g^{-1}(y))\,|\det(J)|$ with $J = \left(\frac{\partial g_i^{-1}(y)}{\partial y_j}\right)_{i,j=1}^n$.
- Probability integral transform: If X is continuous r.v. with cdf F_X, then $F_X(X) \sim \text{Uniform}(0,1)$.

Moments

- If X has pdf $f_X(x)$, then the **expectation** $\mathbb{E}[X] = \int_{-\infty}^\infty x f_X(x)dx$ provided $\int_{-\infty}^\infty |x| f_X(x)dx < \infty$.
- The expectation of a function $g(X)$ is $\mathbb{E}[g(X)] = \int_{-\infty}^\infty g(x)f_X(x)dx$.
- The **conditional expectation** is $\mathbb{E}[Y|X = x] = \int_{-\infty}^\infty y f_{Y|X}(y|x)dy$.
- The **law of iterated expectation** is $\mathbb{E}[Y] = \mathbb{E}[\mathbb{E}[Y|X]]$.
- The k-th **moment** of X is $\mu_k = \mathbb{E}[X^k]$. The k-th **central moment** is $\mu_k' = \mathbb{E}[(X - \mathbb{E}[X])^k]$.
- **Variance** is $\sigma_X^2 = \text{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$.
- **Covariance** is $\text{Cov}[X,Y] = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$. $X \perp\!\!\!\perp Y \implies \text{Cov}[X,Y] = 0$. $\text{Cov}[aU + bV, cW + dZ] = ac\,\text{Cov}[U,W] + ad\,\text{Cov}[U,Z] + bc\,\text{Cov}[V,W] + bd\,\text{Cov}[V,Z]$.
- The **correlation** is $\rho_{X,Y} = \frac{\text{Cov}[X,Y]}{\sigma_X\sigma_Y}$.
- The **law of total variance** is $\text{Var}(Y) = \mathbb{E}[\text{Var}(Y|X)] + \text{Var}(\mathbb{E}[Y|X])$.
- The **covariance matrix** of r.v. X is $\Sigma = \mathbb{E}[(X - \mathbb{E}[X])(X - \mathbb{E}[X])']$, with $\Sigma_{ij} = \text{Cov}[X_i, X_j]$ and $\Sigma_{ii} = \sigma_{X_i}^2$.
- Linear transformations of r.v.: If $Y = AX + b$ then $\mathbb{E}[Y] = A\mathbb{E}[X] + b$ and $\text{Var}(Y) = A\,\text{Var}(X)A'$.
- The **mgf** of X is $M_X(t) = \mathbb{E}[e^{tX}] = \int_{-\infty}^\infty e^{tx}f_X(x)dx$. The k-th moment is $\mu_k = M_X^{(k)}(0)$. The mgf of independent r.v.s X, Y satisfies $M_{X+Y}(t) = M_X(t)M_Y(t)$.
- **chf** of X is $\phi_X(t) = \mathbb{E}[e^{itX}] = M_X(it) = \mathbb{E}[i\sin(tX) + \cos(tX)] = \mathbb{E}[\cos(tX)] + i\mathbb{E}[\sin(tX)]$.

Common Distributions

- **Bernoulli**(p): $P(X = 1) = p$, $P(X = 0) = 1 - p$. $\mathbb{E}[X] = p$, $\text{Var}(X) = p(1-p)$, $M_X(t) = 1 - p + pe^t$, $\phi_X(t) = 1 - p + pe^{it}$.
- **Binom**(n,p): $P(X = x) = \binom{n}{x}p^x(1-p)^{n-x}$. $\mathbb{E}[X] = np$, $\text{Var}(X) = np(1-p)$, $M_X(t) = (1 - p + pe^t)^n$, $\phi_X(t) = (1 - p + pe^{it})^n$. If $Y = \sum_{i=1}^m X_i, X_i \overset{i.i.d.}{\sim} \text{Binom}(n,p) \implies Y \sim \text{Binom}(\sum^m n_i, p)$.
- **Poisson**(λ): $P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!}$. $\mathbb{E}[X] = \lambda$, $\text{Var}(X) = \lambda$, $M_X(t) = e^{\lambda(e^t-1)}$, $\phi_X(t) = e^{\lambda(e^{it}-1)}$. If $Y = \sum_{i=1}^m X_i, X_i \overset{i.i.d.}{\sim} \text{Poisson}(\lambda) \implies Y \sim \text{Poisson}(\sum^m \lambda_i)$.
- **Uniform**(a,b): $f_X(x) = \frac{1}{b-a}$. $\mathbb{E}[X] = \frac{a+b}{2}$, $\text{Var}(X) = \frac{(b-a)^2}{12}$.
- **Standard Normal**$(0,1)$: $f_X(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$. $\mathbb{E}[X] = 0 = \mathbb{E}[X^{2m+1}]$, $\text{Var}(X) = \mathbb{E}[X^2] = 1$, $\mathbb{E}[X^4] = 3$, $M_X(t) = e^{t^2/2}$, $\phi_X(t) = e^{-t^2/2}$.
- **Normal**(μ,σ^2): $f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}}e^{-(x-\mu)^2/(2\sigma^2)}$. $M_X(t) = e^{\mu t + \sigma^2 t^2/2}$, $\phi_X(t) = e^{i\mu t - \sigma^2 t^2/2}$. If $Y = \sum_{i=1}^m c_i X_i, X_i \overset{i.i.d.}{\sim} \text{Normal}(\mu_i, \sigma_i^2) \implies Y \sim \text{Normal}(\sum^m c_i\mu_i, \sum^m c_i^2\sigma_i^2)$.
- **Var** of $aX_1 + bX_2$ w/ $X_j \sim \mathcal{N}(\mu_j, \sigma_j^2)$ is $a^2\sigma_1^2 + b^2\sigma_2^2 + 2ab\,\text{Cov}[X_1, X_2]$, where $\text{Cov}[X_1, X_2] = \rho\sigma_1\sigma_2$.
- If $Z \sim \mathcal{N}(0,1)$, then $X = \mu + \sigma Z \sim \mathcal{N}(\mu, \sigma^2)$. Conversely, if $X \sim \mathcal{N}(\mu,\sigma^2)$, then $Z = \frac{X-\mu}{\sigma} \sim \mathcal{N}(0,1)$.
- $\chi^2(k)$: $X = \sum_{i=1}^k Z_i^2, Z_i \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$. $\mathbb{E}[X] = k$, $\text{Var}(X) = 2k$, $M_X(t) = (1-2t)^{-k/2}$, $\phi_X(t) = (1-2it)^{-k/2}$.
- $F(d_1, d_2)$: $X = \frac{Y_1 d_2}{Y_2 d_1}$, $Y_i \overset{i.i.d.}{\sim} \chi^2(d_i)$.
- $t(d)$: $X = \frac{Z}{\sqrt{Y/d}}$, $Z \sim \mathcal{N}(0,1)$, $Y \sim \chi^2(d)$ independent.
- **Student's Theorem:** If $X_1, \ldots, X_n \overset{i.i.d.}{\sim} \mathcal{N}(\mu, \sigma^2)$, then $\frac{1}{n}\sum X_i = \bar{X} \sim \mathcal{N}(\mu, \sigma^2/n)$, $S^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2 \sim \frac{\sigma^2}{n-1}\chi^2(n-1)$, and $\frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t(n-1)$, even for small n.
- **Bivariate standard normal** (X,Y) has $f_{X,Y}(x,y) = \frac{1}{2\pi\sqrt{1-\rho^2}}\exp\left(-\frac{1}{2(1-\rho^2)}u\right)$, where $u = x^2 + y^2 - 2\rho xy$.
- **Bivariate normal** (X,Y) has $f_{X,Y}(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}}\exp\left(-\frac{1}{2(1-\rho^2)}u\right)$, where $u = \left[\frac{(x-\mu_X)^2}{\sigma_X^2} + \frac{(y-\mu_Y)^2}{\sigma_Y^2} - \frac{2\rho(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y}\right]$. $\rho = 0 \Leftarrow X \perp\!\!\!\perp Y \iff f_{X,Y}(x,y) = f_X(x)f_Y(y)$.

Inequalities

- **Markov:** For non-negative r.v. X and $a > 0$, $P(X \geq a) \leq \frac{\mathbb{E}[X]}{a}$.
- **Chebyshev:** For r.v. X with mean μ and variance σ^2, and $k > 0$, $P(|X - \mu| \geq k) \leq \frac{\sigma^2}{k^2}$.
- **Jensen:** For convex fn g and r.v. X, $\mathbb{E}[g(X)] \geq g(\mathbb{E}[X])$. If g is concave, the inequality is reversed.
- **Cauchy-Schwarz:** For r.v.s X, Y, $|\mathbb{E}[XY]| \leq \sqrt{\mathbb{E}[X^2]\mathbb{E}[Y^2]}$. If $\mathbb{E}[X^2] = 0$

or $\mathbb{E}[Y^2] = 0$, then equality holds. Otherwise, equality holds iff there exist constants a, b, not both zero, such that $P(aX + bY = 0) = 1$.

Convergence

- **Almost sure convergence:** $X_n \xrightarrow{a.s.} X$ if $P(\{\omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\}) = 1$.
- **Convergence in probability:** $X_n \xrightarrow{p} X$ if $\forall \epsilon > 0, \lim_{n\to\infty} P(|X_n - X| \geq \epsilon) = 0$.
- For constant b and r.v. X_n, Y_n: If $X_n \xrightarrow{p} X$ and $Y_n \xrightarrow{p} Y$, then $bX_n + Y_n \xrightarrow{p} bX + Y$, $X_nY_n \xrightarrow{p} XY$.
- **Convergence in r-th mean:** $X_n \xrightarrow{r-m} X$ if $\lim_{n\to\infty} \mathbb{E}[|X_n - X|^r] = 0$ for $r > 0$.
- **Convergence in distribution:** $X_n \xrightarrow{d} X$ if $\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$ at all points x where F_X is continuous. $\forall$ fn g : bounded continuous, $X_n \xrightarrow{d} X \iff \lim_{n\to\infty} \mathbb{E}[g(X_n)] = \mathbb{E}[g(X)]$.
- $X_n \xrightarrow{a.s.} X \implies X_n \xrightarrow{p} X$; $X_n \xrightarrow{r-m} X \implies X_n \xrightarrow{t-m} X$ for $r \geq t > 1$; $X_n \xrightarrow{r-m} X \implies X_n \xrightarrow{p} X$ for $r \geq 1$, $X_n \xrightarrow{p} X \implies X_n \xrightarrow{d} X$.
- **CMT:** If $X_n \to X$ and continuous g, then $g(X_n) \to g(X)$, for a.s., p, d, and to r.m. if fn is Lipschitz. This applies too if X is constant.
- **Slutsky's theorem:** If $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{p} c$, then $X_n + Y_n \xrightarrow{d} X + c$, $X_nY_n \xrightarrow{d} cX$, and if $c \neq 0$, $X_n/Y_n \xrightarrow{d} X/c$.
- $\mathbf{o_p}$: We say $X_n = o_p(a_n)$ if $X_n/a_n \xrightarrow{p} 0 \implies P(|X_n/a_n| \geq \epsilon) \to 0 \;\forall\; \epsilon > 0$.
- $\mathbf{O_p}$: We say $X_n = O_p(a_n)$ if $\forall \epsilon > 0, \exists M > 0, N > 0$ s.t. $P(|X_n/a_n| > M) < \epsilon \;\forall\; n$.

LLN, CLT, Delta Method

- **WLLN:** If X_i with $\mathbb{E}[X_i] = \mu$, $\text{Var}(X_i) = \sigma^2$, $\text{Cov}[X_i, X_j] = 0$ for $i \neq j$. Let $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$, $\bar{\mu} = \frac{1}{n}\sum_{i=1}^n \mu_i$, and $\bar{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n \sigma_i^2$. If $\frac{\bar{\sigma}^2}{n} \to 0$ as $n \to \infty$, then $\bar{X}_n - \bar{\mu} \xrightarrow{p} 0$.
- **LL CLT:** If X_i iid w/ $\mathbb{E}[X_i] = \mu$, $\text{Var}(X_i) = \sigma^2$. Let $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$. Then $\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{d} \mathcal{N}(0, \sigma^2)$ and $\sqrt{n}/\sigma(\bar{X}_n - \mu) \xrightarrow{d} \mathcal{N}(0, 1)$.
- **LF CLT:** For $X \sim iid(\mu, \sigma^2)$, $\frac{1}{\sqrt{n}}\sum_{i=1}^n (X_i - \mu) \xrightarrow{d} \mathcal{N}(0, \sigma^2)$.
- **Probabilities from normal approx:** For any $X \sim \mathcal{N}(\mu, \sigma^2)$ and $a \in \mathbb{R}$: $P(X > a) = P\left(\frac{X-\mu}{\sigma} > \frac{a-\mu}{\sigma}\right) = 1 - \Phi\left(\frac{a-\mu}{\sigma}\right) \implies P(X \leq a) = \Phi\left(\frac{a-\mu}{\sigma}\right)$
- **Delta Method:** Let X_n be a sequence of r.v.s and g a fn differentiable at point a with $g'(a) \neq 0$. If $\sqrt{n}(X_n - a) \xrightarrow{d} \mathcal{N}(0, \sigma^2)$, then $\sqrt{n}(g(X_n) - g(a)) \xrightarrow{d} \mathcal{N}(g(a) \cdot 0, (g'(a))^2\sigma^2)$.
- **MV Delta Method:** Let X_n be a sequence of $k \times 1$ random vectors such that $\sqrt{n}(X_n - a) \xrightarrow{d} \mathcal{N}(0, \Sigma)$ and let $g : \mathbb{R}^k \to \mathbb{R}^m$ be a function that is differentiable at $a \in \mathbb{R}^k$ with A the $m \times k$ Jacobian matrix of first derivatives of g at a. Then, $\sqrt{n}(g(X_n) - g(a)) \xrightarrow{d} \mathcal{N}(0, A\Sigma A')$

Estimation

- **Loss:** A loss fn $L(\theta, a)$ measures the cost of taking action a when the true parameter is θ. The risk fn is $R(\theta, \delta) = \mathbb{E}_\theta[L(\theta, \delta(X))]$.
- **Estimator:** Let X r.v. length n with cdf $F_X(x, \theta)$ depending on parameter $\theta \in \Theta \subseteq \mathbb{R}^k$. An estimator of θ is a fn $\hat{\theta} = g(X)$.
- **Bias:** $\text{Bias}(\hat{\theta}) = \mathbb{E}[\hat{\theta}] - \theta$. If $\text{Bias}(\hat{\theta}) = 0$, $\hat{\theta}$ is unbiased.
- **MSE:** $\text{MSE}(\hat{\theta}) = \mathbb{E}[(\hat{\theta} - \theta)^2] = \text{Var}(\hat{\theta}) + \text{Bias}(\hat{\theta})^2$.
- **Consistency:** $\hat{\theta}_n$ is consistent for θ if $\hat{\theta}_n \xrightarrow{p} \theta$.
- **Asymptotic normality:** $\hat{\theta}_n$ is asymptotically normal if $\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} \mathcal{N}(0, \sigma^2)$ for some $\sigma^2 > 0$.
- MLE is invariant to **reparametrizations**: if $\gamma = h(\theta)$ for some one-to-one function h, then the MLE of γ is $\hat{\gamma} = h(\hat{\theta})$.
- **Score:** i.i.d. sample $X_1, \ldots, X_n$ with pdf $f(x, \theta)$ with $L(\theta) = \sum_{i=1}^n \log f(X_i, \theta)$. The score is $S(\theta) = \frac{\partial L(\theta)}{\partial \theta}$, i.e. $s_i = \frac{\partial \log f(X_i, \theta)}{\partial \theta} = \frac{1}{f(x,\theta)}\frac{\partial f(x,\theta)}{\partial \theta}$. At the true parameter, $\mathbb{E}[S(\theta)] = 0$.
- **Fisher information:** $I(\theta) = -\mathbb{E}_\theta\left[\frac{\partial^2 \log f(X,\theta)}{\partial \theta^2}\right] = \text{Var}(S(\theta))$.
- **Cramer-Rao lower bound:** For unbiased estimator $\hat{\theta}$, $\text{Var}(\hat{\theta}) \geq \frac{1}{I(\theta)}$, with $I(\theta) = n\bar{I}(\theta)$ w/ $\bar{I}(\theta) = -\frac{1}{n}\mathbb{E}[\frac{\partial s_i(\theta)}{\partial \theta}]$.
- Under some regularity conditions, the MLE $\hat{\theta}_{MLE}$ satisfies $\sqrt{n}(\hat{\theta}_{MLE} - \theta) \xrightarrow{d} \mathcal{N}\left(0, \bar{I}(\theta)^{-1}\right)$, as well as $\hat{\theta}_{MLE} \xrightarrow{p} \theta_0$.

Testing

- **Classification errors:** Type I error: Reject H_0 when true. Type II error: Fail to reject H_0 when false.
- **Test** is a fn $\phi : \mathbb{R}^n \to \{0, 1\}$ where $\phi(x) = 1$ means reject H_0 and $\phi(x) = 0$ means fail to reject H_0.
- **Level:** ϕ has level α if $\mathbb{E}_0[\phi(X)] \leq \alpha$. The **size** of ϕ is $\sup_{\theta\in\Theta_0} \mathbb{E}_\theta[\phi(X)] = \sup_{\theta\in\Theta_0} \int \phi(x)f(x,\theta)dx$.
- **Power:** The power function $\beta(\phi) = \mathbb{E}_1[\phi(X)] = \int \phi(x)f(x,\theta)dx$.
- **p-value:** The p-value is the smallest level at which the test rejects the null hypothesis: $p = \sup_{\theta\in\Theta_0} P_\theta(X \in W) = \sup_{\theta\in\Theta_0} \int_W f(x,\theta)dx$ where W is the rejection region.
- A test is said to be **biased** if its power is less than its size for some $\theta \in \Theta_1$.
- **Neyman-Pearson Lemma:** For simple hypotheses $H_0 : \theta = \theta_0$ vs. $H_1 : \theta = \theta_1$, the most powerful level α test rejects H_0 when $\Lambda(x) = \frac{f(x,\theta_1)}{f(x,\theta_0)} > k$ for some k such that $\mathbb{E}_0[\phi(X)] = P_0(\Lambda(x) > k) = \alpha$. This test does not depend on θ_1, can be used for composite H_1, making it the UMP test.
- **Confidence set:** Map $S(X)$ such that $P(\theta_0 \in S(X)) \geq C$ for all $\theta_0 \in \Theta$. The level is C. The random set contains the true parameter with probability at least C. Can be constructed by inverting a family of $\alpha = 1 - C$ tests: $S(X) = \{\theta : \phi(x, \theta) = 0\}$.
- **Log-likelihood ratio test** $\xi_{LR} = 2\left[\log \mathcal{L}(\hat{\theta}) - \log \mathcal{L}(\theta_0)\right]$. Under regularity conditions, if H_0 is true, $\xi_{LR} \xrightarrow{d} \chi_k^2$ where k is the dimension of θ.
- **Wald test** $\xi_W = \sqrt{n}(\hat{\theta} - \theta_0)'\left(-\frac{1}{n}\frac{\partial^2 \log \mathcal{L}(\hat{\theta})}{\partial\theta\partial\theta'}\right)^{-1}\sqrt{n}(\hat{\theta} - \theta_0)$. Under regularity conditions, if H_0 is true, $\xi_W \xrightarrow{d} \chi_k^2$ where k is the dimension of θ.
- **Lagrange multiplier test:** $\xi_{LM} = \frac{1}{\sqrt{n}}S(\theta_0)'\left(-\frac{1}{n}\frac{\partial^2 \log \mathcal{L}(\theta_0)}{\partial\theta\partial\theta'}\right)^{-1}\frac{1}{\sqrt{n}}S(\theta_0)$. Under regularity conditions, if H_0 is true, $\xi_{LM} \xrightarrow{d} \chi_k^2$ where k is the dimension of θ.
- Asymptotically, $\xi_{LR}, \xi_W, \xi_{LM}$ are equivalent.

Bayes

- **Prior** distribution $\pi(\theta)$ represents beliefs about θ before seeing data. Prior is conjugate if posterior is in same family as prior.
- **Posterior** distribution $\pi(\theta|x) = \frac{f(x|\theta)\pi(\theta)}{\int f(x|\theta)\pi(\theta)d\theta}$.
- **Bayes action:** $a^* = \arg\min_a \int L(\theta, a)\pi(\theta|x)d\theta$. Estimation is a special case with action space Θ and the loss typically squared error or absolute error. Testing is a special case with action space $\{0, 1\}$ and loss matrix $L(\theta, a)$.
- **Credible sets** are Bayesian analog of confidence sets. A 100% credible set $S(x)$ satisfies $P(\theta \in S(x)|x) = C$.
- **Complete class theorem:** Under mild regularity conditions, every admissible decision rule is a Bayes rule or a limit of Bayes rules.