$X\sim$	Dist.	μ	σ^2	$M_X(t)=$
$\mathrm{Be}(p)$	$\mathbb{P}(X=1)=p=1-\mathbb{P}(X=0)$	p	$p(1-p)$	$1-p+pe^t$
$\mathrm{Bin}(n,p)$	$f_Y(k)=\binom{n}{k}p^k(1-p)^{n-k},\quad k\in\mathbb{Z}_{[0,n]}$	np	$np(1-p)$	$(1-p+pe^t)^n$
$\mathrm{Geom}(p)$	$f_W(k)=(1-p)^{k-1}p,\quad k\in\mathbb{Z}_+$	$\frac{1}{p}$	$\frac{1-p}{p^2}$	$\frac{pe^t}{1-(1-p)e^t}, t<-\ln(1-p)$
$\mathrm{Pois}(\lambda)$	$f_X(k)=\frac{\lambda^k}{k!}e^{-\lambda},\quad k\in\mathbb{Z}_{\geq 0}$	λ	λ	$e^{\lambda(e^t-1)}$
$\mathrm{Exp}(\lambda)$	$\forall x\geq 0, F_X(x)=1-e^{-\lambda x}, f_X(x)=\lambda e^{-\lambda x}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	$\frac{\lambda}{\lambda-t}$
$N(\mu,\sigma^2)$	$f_X(x)=\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$	μ	σ^2	$e^{t\mu+\frac{1}{2}\sigma^2t^2}$

Expectation $\mathbb{E}(X):=\sum_{i=1}x_if_X(x_i)=\sum_{x:f_X(x)>0}xf_X(x)$ only when the sum is abs. conv.

Moments k-th **moment** of X: $m_k=\mathbb{E}X^k$. k-th **central moment**: $\sigma_k=\mathbb{E}[(X-\mathbb{E}X)^k]$. $\mathbb{V}X:=\sigma_2=\mathbb{E}(X-m_1)^2$. Standard deviation of X: $\sigma(X):=\sqrt{\sigma_2}$. $\mathbb{V}X=\mathbb{E}(X^2)-(\mathbb{E}X)^2$.

MGF $M_X(t)=\mathbb{E}[e^{tX}]$. $F_X\equiv F_Y$ iff $M_X(t)=M_Y(t)$ for $t\in(-h,h)$. $M_X^{(k)}(0)=\mathbb{E}[X^k]$.

Mixed R.V.s $F_X(x)=C(x)+D(x)$ where C is continuous and non-decreasing and D is piece-wise constant and non-decreasing. $\mathbb{E}[X]=\int xc(x)dx+\sum xd(x)$. Don't forget the jumps.

Let $p=\int_{-\infty}^{\infty}c(u)dt$. Let $f_U(u)=\frac{c(u)}{p}$ and $f_V(v)=\frac{d(v)}{1-p}$. W be a r.v. such that it equals to U with probability p. Then $F_W(x)=F_X(x)$.

Transformation Let X be a continuous r.v.. Let $Y=g(X)$, where g is bijective and differentiable over the support D_X of X. Then Y is continuous and its pdf is $f_Y(y)=f_X(g^{-1}(y))\cdot\left|\frac{d}{dy}g^{-1}(y)\right|$.

Joint Distribution $F_{X,Y}(x,y)=\mathbb{P}(X\leq x,Y\leq y)$. $f(x,y)=\frac{\partial^2}{\partial x\partial y}F_{X,Y}(x,y)$. Marginal dist. $f_X(x)=\int_{-\infty}^{+\infty}f_{X,Y}(x,y)dy$.

Independence Iff can be factorized. If $X\perp\!\!\!\perp Y$ then $\mathbb{E}[u(X_1)v(X_2)]=\mathbb{E}u(X_1)\,\mathbb{E}v(X_2)$.

Conditional Distribution $f_{Y|X}(y|x)=\frac{f_{X,Y}(x,y)}{f_X(x)}$. $F_{Y|X}(y|x):=\int_{-\infty}^{y}\frac{f_{X,Y}(x,v)}{f_X(x)}dv$. $\mathbb{P}(Y\in A|X=\cdot)\circ X:\Omega\to\mathbb{R}$ as r.v.

Conditional Expectation $\psi(x)=\mathbb{E}(Y|X=x):=\int_{-\infty}^{+\infty}yf_{Y|X}(y|x)dy$. $\mathbb{E}(Y|X)=\psi(X):=\psi\circ X:\Omega\to\mathbb{R}$. $\mathbb{E}[\mathbb{E}(Y|X)]=\mathbb{E}Y$. If $X\perp\!\!\!\perp Y$, then $\mathbb{E}[Y|X=x]=\mathbb{E}Y$.

Conditional Variance $\mathbb{V}(X_2|X_1=x_1)=\mathbb{E}[(X_2-\mathbb{E}[X_2|X_1=x_1])^2|X_1=x_1]=\mathbb{E}[X_2^2|X_1=x_1]-(\mathbb{E}[X_2|X_1=x_1])$. $\mathbb{V}(X_2|X_1)=\mathbb{V}(X_2|X_1=\cdot)\circ X_1:\Omega\to\mathbb{R}$. $\mathbb{V}(X_2)=\mathbb{E}[\mathbb{V}(X_2|X_1)]+\mathbb{V}(\mathbb{E}[X_1|X_1])$.

Joint MGF $M_{X_1,X_2}(t_1,t_2)=\mathbb{E}[e^{t_1X_1+t_2X_2}]$. $X_1\perp\!\!\!\perp X_2$ iff $M_{X_1,X_2}(t_1,t_2)=M_{X_1}(t_1)\cdot M_{X_2}(t_2),\forall(t_1,t_2)$. Marginal $M_{X_1}(t_1)=\mathbb{E}[e^{t_1X_1}]=M_{X_1,X_2}(t_1,0)$. $\mathbb{E}[X_1X_2]=\frac{\partial^2}{\partial t_1\partial t_2}M(t_1,t_2)\Big|_{t_1,t_2=0}$.

Cov. and Corr. Coeff. $\mathrm{Cov}(X,Y)=\mathbb{E}[(X-\mathbb{E}X)\cdot(Y-\mathbb{E}Y)]=\mathbb{E}[XY]-\mathbb{E}X\,\mathbb{E}Y$. $\mathrm{Cov}(X,Y)=\frac{\partial^2}{\partial t_1\partial t_2}M(t_1,t_2)|_{t_1=t_2=0}-\frac{\partial}{\partial t_1}M(t_1,0)|_{t_1=0}\frac{\partial}{\partial t_2}M(0,t_2)|_{t_2=0}$. $\mathrm{Cov}(X,X)=\mathbb{V}X$. $\mathrm{Cov}(aX+b,Y)=a\,\mathrm{Cov}(X,Y)$.

$\rho(X,Y)=\mathrm{Cor}(X,Y)=\mathbb{E}\left[\frac{X-\mathbb{E}X}{\sigma(X)}\frac{Y-\mathbb{E}Y}{\sigma(Y)}\right]=\frac{\mathrm{Cov}(X,Y)}{\sigma(X)\sigma(Y)}$. If $X\perp\!\!\!\perp Y$ then $\rho(X,Y)=0$.

$\mathbb{E}[Y|X]=\mathbb{E}[Y]+\rho(X,Y)\frac{\sigma(Y)}{\sigma(X)}(X-\mathbb{E}[X])$, $\mathbb{E}[\mathbb{V}(Y|X)]=\sigma(Y)^2(1-\rho(X,Y)^2)$.

Transformation $f_{X+Y}(z)=\sum_x f_{X,Y}(x,z-x)\quad\left(=\int_{-\infty}^{+\infty}f_{X,Y}(z-y,y)dy\right)$.

Let $X=(X_1,\dots,X_n)$ be a cont. r.v. with jpdf $f(\cdot)$. Let $\begin{cases}y_1=u_1(x_1,\dots,x_n)\\ y_n=u_n(x_1,\dots,x_n)\end{cases}$, or simply $\boldsymbol{y}=\boldsymbol{u}(\boldsymbol{x})$ be a bijective differentiable transformaion of $\boldsymbol{X}$. Let the inverse of $\boldsymbol{u}$ be $\boldsymbol{w}$, i.e., $\boldsymbol{x}=\boldsymbol{w}(\boldsymbol{y})$. Then, $f_{\boldsymbol{Y}}(\boldsymbol{y})=f_{\boldsymbol{X}}(\boldsymbol{w}(\boldsymbol{y}))\cdot|\det J_{\boldsymbol{w}}(\boldsymbol{y})|$, where $J_{\boldsymbol{w}}(\boldsymbol{y})=\left[\frac{\partial x_i}{\partial y_j}\right]$ is the Jacobian of $\boldsymbol{w}$.

Variance-Covariance Matrix Let $\boldsymbol{W}_1,\boldsymbol{W}_2$ be $m\times n$ matrices of r.v.s. Let $\boldsymbol{A}_1,\boldsymbol{A}_2$ be $k\times m$ matrices of constants. Let $\boldsymbol{B}$ be an $n\times l$ matrix of constants. (1) $\mathbb{E}[\boldsymbol{A}_1\boldsymbol{W}_1+\boldsymbol{A}_2\boldsymbol{W}_2]$, (2) $\mathbb{E}[\boldsymbol{A}_1\boldsymbol{W}_1\boldsymbol{B}]=\boldsymbol{A}_1\,\mathbb{E}[\boldsymbol{W}_1]\boldsymbol{B}$.

$$\mathrm{Cov}(\boldsymbol{X})=\mathbb{E}[(\boldsymbol{X}-\mathbb{E}[\boldsymbol{X}])(\boldsymbol{X}-\mathbb{E}[\boldsymbol{X}])^T]=\begin{pmatrix}\mathbb{V}X_1 & \dots & \mathrm{Cov}(X_1,X_n)\\ \vdots & \ddots & \vdots\\ \mathrm{Cov}(X_n,X_1) & \dots & \mathbb{V}X_n\end{pmatrix}.$$

Linear Combination Let $T=\sum_{i=1}^n a_iX_i$, $W=\sum_{i=1}^m b_iY_i$. If $E[X_i^2]<\infty$ and $E[Y_j^2]<\infty$ for $i=1,\dots,n$ and $j=1,\dots,m$, then $\mathrm{Cov}(T,W)=\sum_{i=1}^n\sum_{j=1}^m a_ib_j\,\mathrm{Cov}(X_i,X_j)$.

Bivariate Normal $f_{X,Y}(x,y)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\cdot\exp\left(-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_1}{\sigma_1}\right)^2-2\rho\left(\frac{x-\mu_1}{\sigma_1}\right)\left(\frac{y-\mu_2}{\sigma_2}\right)+\left(\frac{y-\mu_2}{\sigma_2}\right)^2\right]\right)$

If (X_1,X_2) is a bivariate normal r.v. parametrized by $(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\pi)$, then

- $Y=a_1X_1+a_2X_2$ is normally dist., $\mathbb{E}Y=a_1\mu_1+a_2\mu_2$, $\mathbb{V}Y=a_1^2\sigma_1^2+2a_1a_2\sigma_1\sigma_2\rho+a_2^2\sigma_2^2$.
- The conditional of X_2 given $X_1=x_1$ is also a normal distribution, $\mathbb{E}[X_2|X_1=x_1]=\mu_2+\rho\frac{\sigma_2}{\sigma_1}(x_1-\mu_1)$, $\mathbb{E}[\mathbb{V}(X_2|X_1=x_1)]=\sigma_2^2(1-\rho^2)$.

Multivariate Normal $f_{\boldsymbol{X}}(\boldsymbol{x})=\frac{1}{(2\pi)^{n/2}|\det\boldsymbol{\Sigma}|^{1/2}}\exp\left[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right]$.

If $\boldsymbol{X}\sim N_n(\boldsymbol{\mu},\boldsymbol{\Sigma})$, then $(\boldsymbol{\Sigma}^{1/2})^{-1}(\boldsymbol{X}-\boldsymbol{\mu})\sim N_n(\boldsymbol{0}_n,\boldsymbol{I}_n)$. Conversely, if $\boldsymbol{Z}\sim N_n(\boldsymbol{0}_n,\boldsymbol{I}_n)$, then $\boldsymbol{Y}=\boldsymbol{\Sigma}^{1/2}\boldsymbol{Z}+\boldsymbol{\mu}\sim N_n(\boldsymbol{\mu},\boldsymbol{\Sigma})$.

Suppose $\boldsymbol{W}\sim N_n(\boldsymbol{\mu}_n,\boldsymbol{\Sigma})$. Let $\boldsymbol{V}=\boldsymbol{A}\boldsymbol{W}+\boldsymbol{b}$, where $\boldsymbol{A}\sim\mathbb{R}^{m\times n},\boldsymbol{b}\sim\mathbb{R}^m$. Then, $\boldsymbol{V}\sim N_m(\boldsymbol{A}\boldsymbol{\mu}+\boldsymbol{b},\boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}^T)$.

Auxiliary Results Chebyshev's Inequality Let X be a r.v. with mean μ and variance σ^2 (finite $\mathbb{E}X^2$). Then $\forall\epsilon>0,\mathbb{P}(|X-\mu|\geq a)\leq\frac{\sigma^2}{\epsilon^2}$. $\lim_{n\to\infty}\left(1+\frac{x}{n}\right)^n=e^x$.

Convergence in Probability Let $\{X_n\}$ be a sequence of r.v.s and let X be a r.v. We say $X_n\to X$ in probability, denoted by $X_n\xrightarrow{\mathbb{P}}X$, if $\forall\epsilon>0,\lim_{n\to\infty}\mathbb{P}(|X_n-X|\geq\epsilon)=0$.

$X_n\xrightarrow{\mathbb{P}}X, Y_n\xrightarrow{\mathbb{P}}Y\implies X_n+Y_n\xrightarrow{\mathbb{P}}X+Y$. $\quad X_n\xrightarrow{\mathbb{P}}X, a$ constant $\implies aX_n\xrightarrow{\mathbb{P}}aX$
$(X_n\xrightarrow{\mathbb{P}}X,Y_n\xrightarrow{\mathbb{P}}Y)\implies X_nY_n\xrightarrow{\mathbb{P}}XY$. $(X_n\xrightarrow{\mathbb{P}}X,Y_n\xrightarrow{\mathbb{P}}Y,Y\neq 0)\implies X_n/Y_n\xrightarrow{\mathbb{P}}X/Y$.

The (sequence of) estimator $\hat{\theta}_n$ is said to be a **consistent estimator** of θ if $\hat{\theta}_n\xrightarrow{\mathbb{P}}\theta$.

Convergence in Distribution Let $C(F_X)$ be the set of all points where $F_X(x)$ is continuous. $X_n\to X$ in distribution (/ weakly / in law), denoted by $X_n\xrightarrow{D}X$ (or $X_n\xrightarrow{D}F_X$), if $\forall x\in C(F_X),\lim_{n\to\infty}F_{X_n}(x)=F_X(x)$.

In this case, we say F_X is the **asymptotic distribution** or the **limiting distribution** of $\{X_n\}$.

$(X_n \xrightarrow{\mathbb{P}} X) \implies (X_n \xrightarrow{D} X)$. If $X_n \xrightarrow{D} c$, where c is a constant (i.e., non-random), then $X_n \xrightarrow{\mathbb{P}} c$. Slutsky's $X_n \xrightarrow{D} X, A_n \xrightarrow{D} a, B_n \xrightarrow{D} b \implies A_n + B_n X_n \xrightarrow{D} a + bX$. In general, if $Y_n \xrightarrow{D} Y \not\Longrightarrow X_n + Y_n \xrightarrow{D} X$. $M_{X_n}(t) \to M_X(t), |t| < h \implies X_n \xrightarrow{D} X$.

Let $g: \mathbb{R} \to \mathbb{R}$ be a continuous function. Then $(X_n \xrightarrow{\mathbb{P}} X) \implies g(X_n) \xrightarrow{\mathbb{P}} g(X)$. In paricular, $X_n \xrightarrow{\mathbb{P}} a \implies g(X_n) \xrightarrow{\mathbb{P}} g(a)$. $(X_n \xrightarrow{D} X) \implies g(X_n) \xrightarrow{D} g(X)$.

Weak Law of Large Number Assume that $\mathbb{E}|X_1| < \infty$ and $\mathbb{E}X_1 = \mu$ (some version requires variance $\sigma^2 < \infty$). Then $\overline{X_n} \xrightarrow{D} \mu$where μ represents the constant r.v. μ.

Central Limit Theorem Let $X_1, X_2, \ldots, X_n$ be iid random r.v.s with $\mathbb{E}|X_1|^2 < \infty$ and $\mathbb{E}X_1 = \mu$, $\mathbb{V}X_1 = \sigma^2 > 0$. Let $Sum_n = \sum_{i=1}^n X_i$. Then $\frac{\sqrt{n}(\overline{X_n}-\mu)}{\sigma} = \frac{Sum_n - n\mu}{\sqrt{n}\sigma} \xrightarrow{D} N(0,1)$ where $N(0,1)$ denotes a standard normal r.v.

Δ-method If $\sqrt{n}(X_n - \theta) \xrightarrow{D} N(0, \sigma^2)$, and $g(x)$ is differentiable at $x = \theta$ with $g'(x) \neq 0$, then $\sqrt{n}[g(X_n) - g(\theta)] \xrightarrow{D} N(0, [g'(\theta)]^2\sigma^2)$.

Confidence Interval (L, U) is a $(1-\alpha)100\%$ **confidence interval** for θ if $P_\theta[\theta \in (L, U)] = 1 - \alpha$.

If $Z = \frac{\sqrt{n}(\hat{\theta}-\theta)}{\sigma} \sim N(0,1)$, then the $(1-\alpha)100\%$ CI is $|Z| \leq Z_{\alpha/2}$, i.e., $\hat{\theta} \pm \frac{\sigma}{\sqrt{n}} z_{\alpha/2}$.

Statistics Basic Unbaisedness $\mathbb{E}[T(X_1, \ldots, X_n)] = \theta$ for all $\theta \in \Theta$.

Sample mean, variance: $\overline{X_n} = \frac{1}{n}\sum_{i=1}^n X_i$. $S_n = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X})^2 = \frac{1}{n-1}\sum_{i=1}^n X_i^2 - \frac{n}{n-1}\overline{X}^2$.

Method of Moment Estimation Sample k-th moment $\overline{X^k} := \frac{1}{n}\sum_{i=1}^n X_i^k$.

$\hat{\boldsymbol{\theta}}_{\text{MM}}$ is the solution of the following system $\overline{X} = m_1(\boldsymbol{\theta}), \overline{X^2} = m_2(\boldsymbol{\theta}), \ldots, \overline{X^k} = m_k(\boldsymbol{\theta})$.

Mean Square Est. $\text{MSE}(\hat{\theta}) = \mathbb{E}_\theta[(\hat{\theta} - \theta)^2]$. $\text{Bias}(\hat{\theta}) = \mathbb{E}_\theta[\hat{\theta}] - \theta$. $\text{MSE}(\hat{\theta}) = (\text{Bias}(\hat{\theta}))^2 + \mathbb{V}_\theta(\hat{\theta})$.

Sufficient Statistics $Y = u(X_1, \ldots, X_n)$ is a sufficient statistic for θ if $\mathbb{P}(\boldsymbol{X} = \boldsymbol{x} | Y = u(\boldsymbol{x})) = \frac{\prod_{i=1}^n f(x_i;\theta)}{f_Y(u(x_1,\ldots,x_n))} = H(x_1, \ldots, x_n)$ does not depend on $\boldsymbol{\theta}$ for all $\boldsymbol{x}$ and $\boldsymbol{\theta} \in \Theta$.

Theorem 0.1 (Neyman). *$Y = u(X_1, \ldots, X_n)$ is a sufficient statistic for θ iff there exist two functions g and h s.t. $L(\boldsymbol{\theta}) = g(y; \boldsymbol{\theta}) \cdot h(\boldsymbol{x})$, where h does not depend on $\boldsymbol{\theta}$.*

Order Statistics The i-th smallest, Y_i, is called the i-th order statistic.

$f_{Y_k}(y) = k\binom{n}{k}[F(y)]^{k-1}[1 - F(y)]^{n-k} f(y)$.

Quantiles $\forall p \in (0, 1)$, the p-th quantile of a cont. r.v. X is $\pi_p = F_X^{-1}(p)$.

Let $1 \leq i < j \leq n$. Then (Y_i, Y_j) is a $\mathbb{P}(i \leq \text{Bin}(n, p) \leq j - 1) \times 100\%$ CI for π_p.

Maximum Likelihood Estimator $L(\boldsymbol{\theta}; \boldsymbol{x}) = \prod_{i=1}^n f(x_i; \boldsymbol{\theta})$, $l(\boldsymbol{\theta}) = \log L(\boldsymbol{\theta}) = \sum_{i=1}^n \log f(x_i; \boldsymbol{\theta})$.

$\hat{\boldsymbol{\theta}}_{\text{mle}} = \arg\max_{\boldsymbol{\theta} \in \Theta} L(\boldsymbol{\theta}) = \arg\max_{\boldsymbol{\theta} \in \Theta} l(\boldsymbol{\theta})$. When $l(\boldsymbol{\theta})$ is differentiable, use the first order optimality condition $\frac{\partial}{\partial \boldsymbol{\theta}} l(\boldsymbol{\theta}) = \mathbf{0}$.

Fisher Information $I_X(\theta) = \mathbb{E}\left[\left(\frac{\partial}{\partial\theta}\log f(x;\theta)\right)^2\right]$. Score function $S(x_i;\theta) = \frac{\partial \log f(x_i;\theta)}{\partial\theta}$.

- MLE estimator first order optimality condition $\partial l(\theta)/\partial\theta = 0 \iff \sum_{i=1}^n S(x_i;\theta) = 0$.
- $\mathbb{E}[S(X_i;\theta)] = 0$. Therefore, the MLE estimator for $X_1, \ldots, X_n$ can be viewed as the method of moment estimator for $S(X_1;\theta), \ldots, S(X_n;\theta)$, as $\frac{1}{n}\sum_{i=1}^n S(x_i;\theta) = 0 = \mathbb{E}[S(X;\theta)]$.
- $I_{X_i}(\theta) = \mathbb{E}[S(X_i;\theta)^2] = \mathbb{V}(S(X_i;\theta))$. $I_X(\theta) = \mathbb{E}\left[-\frac{\partial^2}{\partial\theta^2}\log f(X;\theta)\right]$.
- Additivity: If $\boldsymbol{X} = (X_1, \ldots, X_n)$, X_i are iid with $f(x;\theta)$. Then $I_{\boldsymbol{X}}(\theta) = n \cdot I_X(\theta)$.

Rao-Cramer Lower Bound Let $Y = u(X_1, \ldots, X_n)$ be a statistic with mean $\mathbb{E}(Y) = k(\theta)$. Then $\mathbb{V}[Y] \geq \frac{[k'(\theta)^2]}{n \cdot I_X(\theta)}$.

If $\hat{\theta}$ is an unbiased estimator of θ, i.e., $\mathbb{E}\hat{\theta} = \theta$, then $\text{MSE}(\hat{\theta}) = \mathbb{V}(\hat{\theta}) \geq \frac{1}{n \cdot I_X(\theta)}$.

Let $\hat{\theta}$ be an unbiased estimator of θ. Then $\hat{\theta}$ is called an **efficient estimator** of θ if $\mathbb{V}Y$ attains the Rao-Cramer lower bound: $\mathbb{V}(\hat{\theta}) = \frac{1}{n \cdot I_X(\theta)}$.

Limiting Distribution of MLE $\hat{\theta}_{\text{MLE},n}$ of θ satisfies $\sqrt{n}(\hat{\theta}_{\text{MLE},n} - \theta) \xrightarrow{D} N\left(0, \frac{1}{I_X(\theta)}\right)$.

Hypothesis Testing $\Theta = \Theta_0 \sqcup \Theta_1$. $H_0: \theta \in \Theta_0$ **null hypothesis**, $H_1: \theta \in \Theta_1$ **alternative hypothesis**.

Type I error: When we decide to take H_1 but really H_0 is true. **Type II error**: When we decide to take H_0 but really H_1 is true.

Critical (rejection) region: a subset C of D. Then the corresponding **test** of H_0 versus H_1 follows the rule: Reject H_0 if $(X_1, \ldots, X_n) \in C$, accept H_0 if $(X_1, \ldots, X_n) \notin C$.

Size of a test: We say a critical region C is of size α if $\alpha = \max_{\theta\in\Theta_0} \mathbb{P}[(X_1, \ldots, X_n) \in C|\theta]$.

The **power function** of a rejection region C is $r(\theta) = \mathbb{P}[(X_1, \ldots, X_n) \in C|\theta], \forall\theta \in \Theta_1$. $r(\theta)$ is called the **power** of the test at θ, equal to $1 - \mathbb{P}(\text{Type II error}|\theta)$.

p-value: Suppose we observe a value θ_0 of the statistic θ. Then the p-value is defined to be the probability of obtaining values of θ more extreme than θ_0 given H_0. We reject H_0 if the p-value $\leq \alpha$, the size of the test.

Likelihood Ratio Test The **likelihood ratio statistic** is defined as $\Lambda^*(X_1, \ldots, X_n) = \frac{\max_{\theta\in\Theta_0} L(\theta;\boldsymbol{x})}{\max_{\theta\in\Theta_1} L(\theta;\boldsymbol{x})}$, where $L(\cdot;\boldsymbol{x})$ represents the likelihood function (given data $\boldsymbol{x}$).

The **likelihood ratio test** (LRT) rejects H_0 iff $\Lambda^*(X_1, \ldots, X_n) \leq k$, where k a threshold.

We also define $\Lambda(X_1, \ldots, X_n) = \frac{\max_{\theta\in\Theta_0} L(\theta;\boldsymbol{x})}{\max_{\theta\in\Theta} L(\theta;\boldsymbol{x})} = \min\{\Lambda^*(X_1, \ldots, X_n, 1)\}$ and use the criterion $\Lambda(X_1, \ldots, X_n) \leq k$ for simplicity.

Most Powerful Test A subset C of the sample space is called a **best rejection region** with size α if the test with rejection region C (1)has a size α, (2)is more powerful than the test with any other subset A that also has a size α. That is, $\mathbb{P}(\boldsymbol{X} \in C|\theta_1) \geq \mathbb{P}(\boldsymbol{X} \in C|\theta_1)$. The test with the best rejection region with size α is called the **most powerful test** (MPT) with size α.

Neyman-Pearson Theorem: if k satisfies $\alpha = \mathbb{P}(\lambda(x) \leq k|\theta_0)$, then the LRT with the threshold k is the most powerful test with size α.