,

distribution	parameter	support	$f(x\mid\theta)$	$F(x)$	$\mathbb{E}$	Var	$m(t) = \mathbb{E}(e^{tX})$	$\mathcal{I}$	ind. $\sum_{i=1}^{n} X_i \sim$	$\hat{\theta}_{MLE}$
$Bernoulli(p)$	$0 < p < 1$	$k \in \{0, 1\}$	p if $k = 1$ $(1-p)$ if $k = 0$	$1 - p$	p	$p(1-p)$	$(1-p) + pe^t$	$\frac{1}{p(1-p)}$	$Binom(n, p)$	$\overline{X}$
$Binomial(n, p)$	n-# of trial; p- $\mathbb{P}$ of success	$k \in \{0, \cdots, n\}$	$\binom{n}{k} p^k q^{n-k}$	$\sum_{i=1}^{x} \binom{n}{x} p^x q^{n-x}$	np	$np(1-p)$	$(1 - p + pe^t)^n$	$\frac{n}{p(1-p)}$	$Binom(\sum_{i=1}^{n} n_i, p)$	$\overline{X}$
$Geom(p)$	$0 < p \leq 1$; k-# of trials	$k \in \{1, 2, \cdots\}$	$(1-p)^{k-1} p$	$1 - (1-p)^k$	$\frac{1}{p}$	$\frac{1-p}{p^2}$	$\frac{pe^t}{1-(1-p)e^t}$	$\frac{n}{p^2(1-p)}$	$NegBin(n, p)$	$\frac{1}{\overline{X}}$
$NegBin(r, p)$	r-# of success; x- # of fails	$x \in \{0, \cdots\}$	$\binom{x+r-1}{r-1} p^r (1-p)^x$		$\frac{r(1-p)}{p}$	$\frac{r(1-p)}{p^2}$	$\left(\frac{1-p}{1-pe^t}\right)^{n-k}$			
$Poisson(\lambda)$	$\lambda > 0$ - mean	$k \in \mathbb{N} \cup 0$	$\frac{\lambda^k}{k!} e^{-\lambda}$	$e^{-\lambda} \sum_{i=0}^{k} \frac{\lambda^i}{i!}$	λ	λ	$e^{\lambda(e^t - 1)}$	$\frac{n}{\lambda}$	$Poisson(\sum_{i=1}^{n} \lambda_i)$	$\overline{X}$
$Uniform(a, b)$	$-\infty < a < b < \infty$	$x \in [a, b]$	$\frac{1}{b-a}$	$\frac{x-a}{b-a}$	$\frac{1}{2}(a+b)$	$\frac{1}{12}(b-a)^2$	$\frac{e^{tb} - e^{ta}}{t(b-a)}$			$X_{(n)}$
$Exp(\lambda)$	$\lambda > 0$ - rate	$x \in [0, \infty)$	$\lambda e^{-\lambda x}$	$1 - e^{-\lambda x}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	$\frac{\lambda}{\lambda - t}$	$\frac{n}{\lambda^2}$	$Gamma(n, \lambda)$	$\frac{1}{\overline{X}}$
$Gamma(\alpha, \lambda)$	α - shape; λ - rate	$x \in (0, \infty)$	$\frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}$		$\frac{\alpha}{\lambda}$	$\frac{\alpha}{\lambda^2}$	$\left(\frac{\lambda}{\lambda - t}\right)^\alpha$		$Gamma(\sum_{i=1}^{n} \alpha_i, \lambda)$	$\frac{\alpha}{\overline{X}}$
$\mathcal{N}(\mu, \sigma^2)$	μ - mean; σ^2 - variance	$x \in \mathbb{R}$	$\frac{1}{\sqrt{2\sigma^2\pi}} exp\{-\frac{(x-\mu)^2}{2\sigma^2}\}$		μ	σ^2	$exp\{\mu t + \frac{1}{2}\sigma^2 t^2\}$	$\frac{1}{\sigma^2}$, $\frac{n}{2\sigma^4}$	$\mathcal{N}(\sum_{i=1}^{n} \mu_i, \sum_{i=1}^{n} \sigma_i^2)$	
Φ		$x \in \mathbb{R}$	$\frac{1}{\sqrt{2\pi}} e^{-x^2/2}$	$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} dt$	0	1	$exp\{\frac{t^2}{2}\}$			
$Cauchy(\theta)$		$x \in \mathbb{R}$	$\frac{1}{\pi[1 + (x-\theta)^2]}$		n/a	n/a	n/a			
$Multi(n, p_1, \cdots, p_k)$	n- of trial, p_i-event $\mathbb{P}$	$X_i \in \{0, \cdots, n\}$	$\frac{n!}{x_1! \cdots x_k!} p_1^{x_1} \cdots p_k^{x_k}$		np_i	$np_i(1-p_i)$	$\left(\sum_{i=1}^{k} p_i e^{t_i}\right)^n$			

Probability

If $A_i \cap A_j = \emptyset$	$P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$
Cond. probability	$P(A\|B) = \frac{P(A \cap B)}{P(B)}$
If B_i *mutually ind.*	$P(A) = \sum_{i=1}^{n} P(A\|B_i)P(B_i)$
Independence	$P(A \cap B) = P(A)P(B)$
Bayes' Rule	$P(B_j\|A) = \frac{P(A\|B_j)P(B_j)}{\sum_{i=1}^{n} P(A\|B_i)P(B_i)}$

Distribution Function

$F_X(x) = P(X \leq x) = \int_{-\infty}^{x} f_X(x)dx$

$f_X(x) = \frac{d}{dx}F_X(x)$

$F(x_1, \cdots, x_n) = P(X_1 \leq x_1, \cdots X_n \leq x_n)$

If X_i *independent, then* $f_{\underline{X}}(\underline{x}) = \prod_{i=1}^{n} f_{X_i}(x_i)$

Transformation method $f_Y(y) = f_X(g^{-1}(y))|\frac{dg^{-1}(y)}{dy}|$

Order statistc

Let $X_{(1)} = V = min\{X_1, \cdots, X_n\}$
and $X_{(n)} = U = max\{X_1, \cdots, X_n\}$
$F_U(u) = P(U \leq u) = [F_X(u)]^n$
$f_U(u) = nf_X(u)[F_X(u)]^{n-1}$
$F_V(v) = 1 - [1 - F_X(v)]^n$
$f_V(v) = nf_X(v)[1 - F_X(v)]^{n-1}$

Expected Value and Variance

Excepted value $\mathbb{E}(X) = \sum_i x_i f_X(x_i) = \int_{-\infty}^{\infty} xf(x)dx$

Invariance *If* $Y = g(X)$ *then* $\mathbb{E}[Y] = \int_{-\infty}^{\infty} g(x)f(x)dx$

Linearity *If* $Y = a + \sum_{i=1}^{n} b_i X_i$ *then* $\mathbb{E}[Y] = a + \sum_{i=1}^{n} b_i \mathbb{E}[X_i]$

If X_i *independent, then* $\mathbb{E}[\prod_{i=1}^{n} X_i] = \prod_{i=1}^{n} \mathbb{E}[X_i]$

Variance $Var(X) = \mathbb{E}\{(X - \mathbb{E}[X])^2\} = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx$

$Var(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$

If $Y = a + bX$ *then* $Var(Y) = b^2 Var(X)$
$Var(X+Y) = Var(X) + Var(Y) + 2Cov(X,Y)$

If X_i *independent, then* $Var(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} Var(X_i)$

Conditional Expectation of X ***given the event*** $Y = y$

$$\mathbb{E}[X|Y=y] = \sum_{x \in \mathcal{X}} xP(X=x|Y=y) = \sum_{x \in \mathcal{X}} x\frac{P(X=x, Y=y)}{P(Y=y)}$$

$$\mathbb{E}[X|Y=y] = \int_{\mathcal{X}} xf_X(x,y)dx = \int_{\mathcal{X}} x\frac{f_{X,Y}(x,y)}{P(Y=y)}dx$$

Conditional Expectation w.r.t. a random variable Y

$g : y \mapsto \mathbb{E}(X|Y=y) \qquad E(X|Y) = g(Y) : \omega \mapsto \mathbb{E}[X|Y = Y(\omega)]$

Law of total $\mathbb{P}$	$\mathbb{E}[\mathbb{E}[X\|Y]] = \mathbb{E}[X]$
conditional variance	$Var[X\|Y] = \mathbb{E}[(X - \mathbb{E}[X\|Y])^2\|Y)$
Formula for variance	$Var[X\|Y] = E[X^2\|Y] - (\mathbb{E}[X\|Y])^2$
Law of total variance	$Var(X) = \mathbb{E}[Var(X\|Y)] + Var(\mathbb{E}[X\|Y])$

Covariance

$Cov(X,Y) = \mathbb{E}[(X - \mu_X)(Y - \mu_Y)] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$

If $U = a + \sum_{i=1}^{n} b_i X_i$ *and* $V = c + \sum_{j=1}^{m} d_j Y_j$,

then $Cov(U,V) = \sum_{i=1}^{n}\sum_{j=1}^{m} b_i d_j Cov(X_i, Y_j)$

$Cov(aX + bY, cZ) = acCov(X,Z) + bcCov(Y,Z)$

$\rho = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$

Moment generating function

$M(t) = \mathbb{E}[e^{tX}] = \int_{-\infty}^{\infty} e^{tx}f(x)dx$

kth moment $E(X^k) = M^{(k)}(0)$

$\frac{d^k}{dt^k}M(t) = \frac{d^k}{dt^k}\mathbb{E}[e^{tX}] = \frac{d^k}{dt^k}\mathbb{E}[X^k e^{tX}]|_{t=0} = \mathbb{E}[X^k]$

If $Y = a + bX$, *then* $M_Y(t) = e^{at}M_X(bt)$

If X_i *independent,* $M_{\sum_{i=1}^{n} X_i}(t) = \prod_{i=1}^{n} M_{X_i}(t)$

Limit Theorem

Convergence in Probability *A sequence* $\{X_n\}$ *converges in probability to* X *if for all* $\epsilon > 0$ *we have*

$$\lim_{n \to \infty} \mathbb{P}(|X_n - X| \geq \epsilon) = 0 \iff X_n \xrightarrow{p} X$$

Convergence in Distribution *A sequence* $\{X_n\}$ *with cdf* F_n *converges in distribution to* X *with cdf* F *if*

$$\lim_{n \to \infty} F_n(X) = F(x) \iff X_n \xrightarrow{d} X$$

Law of Large Number *If* X_i *be i.i.d. with* $\mathbb{E}[X_i] = \mu$ *and* $Var(X_i) = \sigma^2$. *Let* $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ *Then*

$$P\left(|\overline{X} - \mu| > \epsilon\right) \overset{n \to \infty}{\to} 0 \quad or \quad \overline{X} \xrightarrow{p} \mu$$

Continuity Theorem *If* $M_n(t) \to M(t)$ *then* $F_n(x) \to F(x)$.

Standardization

$$Z = \frac{X - \mathbb{E}[X]}{\sqrt{Var(X)}}$$

Central Limit Theorem *Let* X_i *be i.i.d. RV with* $\mathbb{E}[X_i] = \mu$ *and* $Var[X_i] = \sigma^2$ *and let* $S_n = \sum_{i=1}^{n} X_i$ *then*

$$Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}} \xrightarrow{d} \phi(x) \quad or \quad \sqrt{n}(\overline{X} - \mu) \xrightarrow{d} \mathcal{N}(0, \sigma^2)$$

Distribution from Normal

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \quad S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

If $X_i \sim \mathcal{N}(\mu, \sigma^2)$ *then* $\overline{X} \sim \mathcal{N}(\mu, \frac{\sigma^2}{n})$

Chi-squared *Let* $Z_i \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$ *then* $\chi_n^2 = \sum_{i=1}^{n} Z_i^2$ *with* n *d.f.*

1. $\chi_n^2 \sim \Gamma(\frac{n}{2}, \frac{1}{2})$ *so* $\mathbb{E}[\chi_n^2] = n$ *and* $Var(\chi_n^2) = 2n$
2. *mgf of* $Y \sim \chi_n^2$ *is* $M_Y(t) = (1-2t)^{-n/2}$
3. $\chi_n^2 + \chi_m^2 \sim \chi_{m+n}^2$
4. $\frac{(n-1)S^2}{\sigma^2} \sim \chi_{n-1}^2$ *(simplify RHS and by mgf uniqueness)*

t distribution $t_n = \frac{\mathcal{N}(0,1)}{\sqrt{\chi_n^2/n}}$

1. $\frac{\overline{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}$ *(standardization of sampling mean)*

F distribution $F_{m,n} = \frac{\chi_m^2/m}{\chi_n^2/n}$

1. *If* $X_1, \cdots, X_m, Y_1, \cdots, Y_n \overset{i.i.d.}{\sim} \mathcal{N}(\mu, \sigma^2)$
then $S_X^2/S_Y^2 \sim F_{m-1,n-1}$

Misc

Gamma	$\Gamma(x) = \int_0^\infty u^{x-1}e^{-u}du$ $\Gamma(\alpha) = (\alpha-1)\Gamma(\alpha-1)$ *and* $\Gamma(1/2) = \sqrt{\pi}$ *if* $X \sim \Gamma(\alpha, \lambda)$ *then* $cX \sim \Gamma(\alpha, \lambda/c)$
Binomial coef.	$(a+b)^n = \sum_{k=0}^{\infty} \binom{n}{k} a^k b^{n-k}$
Markov Ineq.	$P(X \geq t) \leq \frac{E(X)}{t}$
Chebyshev Ineq.	$P(\lvert X-\mu\rvert > t) \leq \frac{\sigma^2}{t^2}$

Parameter Estimation

Consistent Estimator $\hat{\theta} \xrightarrow{p} \theta$
$\overline{X} \xrightarrow{p} \mu$ *by WLLN;* $\hat{\sigma}^2$ *and* S^2 *are consistent estimators*

MME *equating moments* $\mu_k = \mathbb{E}[X^k]$ *with* $m_k = \frac{1}{n}\sum_{i=1}^n X_i^k$

1. $m_k \xrightarrow{p} \mu_k$ *for any* k

2. $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n X_i^2 - \overline{X}^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \overline{X})^2$

Property of convergence in probability *Let* $\hat{\theta}_n \xrightarrow{p} \theta$ *and* $\hat{\eta}_n \xrightarrow{p} \eta$

1. $\hat{\theta}_n + \hat{\eta}_n \xrightarrow{p} \theta + \eta$

2. $\hat{\theta}_n \hat{\eta}_n \xrightarrow{p} \theta\eta$

3. $g(\hat{\theta}_n) \xrightarrow{p} g(\theta)$ *for any continuous* g

MLE $\hat{\theta}_{MLE} = \arg\max_\theta \mathcal{L}(\theta)$ *where*

$L(\theta) = f(x_1, \cdots, x_n|\theta) = \prod_{i=1}^n f(X_i = x_i|\theta)$

$l(\theta) = \log L(\theta) = \sum_{i=1}^n \log f(X_i = x_i|\theta)$

Normal: $\hat{\mu}_{MLE} = \overline{X}$ $\hat{\sigma}^2_{MLE} = \frac{1}{n}\sum_i (X_i - \overline{X})^2$

Newton-Raphson Method $\hat{\theta}_{n+1} = \hat{\theta}_n - \frac{l'(\hat{\theta}_n)}{l''(\hat{\theta}_n)}$

Large Number Theory of MLE

Asymptotically Normality $X_i \sim f_\theta$ *then*

$$F_{Z_n}(z) \overset{n\to\infty}{\to} \Phi(z) \iff \sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} \mathcal{N}(0, \sigma^2)$$

where F_{Z_n} *is the cdf of* $Z_n = \frac{\hat{\theta}_n - \theta}{\sigma/\sqrt{n}}$.

Score $u(\theta) = l'(\theta)$.

1. $\mathbb{E}[u(\theta)] = \int \frac{\partial \log f(x|\theta)}{\partial \theta} f(x|\theta)dx = 0$

2. $Var(u(\theta)) = \mathbb{E}[u^2(\theta)] = \mathcal{I}(\theta)$

Fisher Information $\mathcal{I}(\theta) = -\mathbb{E}[l''(\theta)]$
Single observation, $\mathcal{I}^* = -\mathbb{E}[\frac{\partial^2 \log f(x|\theta)}{\partial \theta^2}]$ $\mathcal{I}(\theta) = n\mathcal{I}^*(\theta)$

Slutsky's Theorem *Let* $X_n \xrightarrow{D} X$ *and* $Y_n \xrightarrow{P} c$ *then for continuous* g *we have*

$$g(X_n, Y_n) \xrightarrow{D} g(X, c)$$

$X_n + Y_n \xrightarrow{d} X + c \quad X_n Y_n \xrightarrow{d} cX \quad X_n/Y_n \xrightarrow{d} X/c$

Asymptotic normality of MLE

$$\sqrt{n}(\hat{\theta}_{MLE} - \theta) \xrightarrow{D} \mathcal{N}(0, \frac{1}{\mathcal{I}^*(\theta)}) \iff \hat{\theta}_{MLE} \sim AN(\theta, \mathcal{I}^{-1}(\theta))$$

1. $\frac{u(\theta)}{\sqrt{n}} \xrightarrow{d} \mathcal{N}(0, \mathcal{I}^*(\theta))$ *(asymptotically normal)*

2. $-\frac{1}{n}\frac{\partial^2 l(\theta)}{\partial \theta^2} \xrightarrow{p} \mathcal{I}^*(\theta)$

3. *If* $X_i \overset{i.i.d.}{\sim} Bernoulli(p)$ *then* $\hat{p} \sim AN(p, \frac{p(1-p)}{n})$

Invariance of MLE *let* $\eta = g(\theta)$ *for some transform* g*. Then*

1. $\hat{\eta}_{MLE} = g(\hat{\theta}_{MLE})$

2. *If* g *is differentiable then* $\hat{\eta}_{MLE} \sim AN(\eta, [g'(\theta)]^2 \mathcal{I}^{-1}(\theta))$

Consistency of MLE $\hat{\theta}_{MLE} \xrightarrow{p} \theta$ *(regularity condition)*
Plug-in Principle $\hat{\theta}_{MLE} \sim AN(\theta, \mathcal{I}^{-1}(\hat{\theta}_{MLE}))$ *with* ***estimated standard error*** *of* $\hat{\sigma}_{\hat{\theta}_{MLE}} = \mathcal{I}^{-1/2}(\hat{\theta}_{MLE})$

Confidence Interval & Efficiency

Confidence Interval *A* $100(1-\alpha)\%$ *confidence interval for* θ *is a pair of statistics* L, U *such that* $P(L \leq \theta \leq U) = 1 - \alpha$
Pivot method *Find a pivot* $g(X_1, \cdots, X_n; \theta)$ *and its distribution. such that* $P(a \leq g(X_1, \cdots, X_n) \leq b) = 1 - \alpha$
Normal mean (unknown variance)

$$\frac{\overline{X} - \mu}{S/\sqrt{n}} \sim t_{n-1} \quad \mathbb{P}\left(t_{n-1,\alpha/2} \leq \frac{\overline{X} - \mu}{S/\sqrt{n}} \leq t_{n-1,1-\alpha/2}\right)$$

$$95\%CI \quad \overline{X} \pm \frac{S}{\sqrt{n}} t_{n-1,1-\alpha/2}$$

Normal variance

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1} \quad \mathbb{P}\left(\chi^2_{n-1,\alpha/2} \leq \frac{(n-1)S^2}{\sigma^2} \leq \chi^2_{n-1,1-\alpha/2}\right)$$

$$95\%CI \quad \left[\frac{(n-1)S^2}{\chi^2_{n-1,1-\alpha/2}}, \frac{(n-1)S^2}{\chi^2_{n-1,\alpha/2}}\right]$$

Large number theory method *construct a* $100(1-\alpha)\%$ *CI of the form*

$$\hat{\theta}_{MLE} \sim AN(\theta, \mathcal{I}^{-1}(\hat{\theta}_{MLE})) \quad \hat{\theta}_{MLE} \pm \frac{z_{1-\alpha/2}}{\sqrt{\mathcal{I}(\hat{\theta}_{MLE})}}$$

Goodness-of Estimation

Standard Error loss	$l(\hat{\theta}, \theta) = (\theta - \hat{\theta})^2$
Mean Squared Error	$MSE(\hat{\theta}, \theta) = \mathbb{E}[(\hat{\theta} - \theta)^2]$
Bias	$b(\hat{\theta}, \theta) = \mathbb{E}[\hat{\theta}] - \theta$ *(*$\overline{X}, S^2$ *unbiased)*
Bias-Variance Decomp.	$MSE(\hat{\theta}, \theta) = b^2(\hat{\theta}, \theta) + Var[\hat{\theta}]$

Cramer-Rao Lower Bound *Let* $X_1, \cdots, X_n \sim f_\theta$ *and let* $\hat{\theta}$ *be an unbiased estimator of* θ*. Under some regularity conditions,*

$$Var[\hat{\theta}] \geq \mathcal{I}^{-1}(\theta)$$

For unbiased estimator $\hat{\theta}$

efficient	$Var[\hat{\theta}] = \mathcal{I}^{-1}(\theta)$
asymptotically efficient	$\lim_{n\to\infty} \frac{Var[\hat{\theta}]}{\mathcal{I}^{-1}(\theta)} = 1$
relative efficiency	$eff(\hat{\theta}_1, \hat{\theta}_2) = \frac{Var[\hat{\theta}_2]}{Var[\hat{\theta}_1]}$

$\hat{\lambda}_{MLE} = \overline{X}$ *for Poisson is efficient;* $\hat{\sigma}^2 = S^2$ *is asymptotically efficient. Asymptotically, MLE are unbiased and efficient.*

Sufficiency

Likelihood Principle *All relevant experimental information is contained in the likelihood function for the observed* $(x_1, \cdots, x_n)$
Sufficiency $T(\underline{X}) = T(x_1, \cdots, T_n)$ *is sufficient for an unknown parameter* θ *if the conditional (joint) distribution of* $X_1, \cdots, X_n$ *given* $T(\underline{X}) = t$ *does not depend on* θ *for any given value of* t.

$$P(\underline{X} = \underline{x} | T(\underline{x}) = t) = P(\underline{X} = \underline{x} | T(\underline{x}) = t, \theta)$$

Factorization Theorem $T(\underline{X})$ *is sufficient if and only if for any value of* θ*, There exists functions* g, h *such that*

$$\mathcal{L}(\theta) = g(T(x_1, \cdots, x_n), \theta) h(x_1, \cdots, x_n)$$

Exponential family of distributions *has* $f(x|\theta)$ *of form*

$$f(x|\theta) = \begin{cases} \exp\{c(\theta)T(x) + d(\theta) + S(x)\} & x \in A \\ 0 & \textit{otherwise} \end{cases}$$

where support A *does not depend on* θ. *Then* $\sum_{i=1}^{n} T(x_i)$ *is a sufficient statistic. Normal, Exponential, Gamma, Chi-squared, Bernoulli, and Poisson are examples.*
Rao-Blackwell Theorem *Let the Rao-Blackwell estimator be* $\hat{\theta}^* = \mathbb{E}[\hat{\theta}|T]$ *where* T *is sufficient. Then for all* θ,

$$MSE(\hat{\theta}^*, \theta) \leq MSE(\hat{\theta}, \theta)$$

where equality holds if and only if $\hat{\theta}^* = \hat{\theta}$.

1. $\hat{\theta}^*$ *and the starting estimator* $\hat{\theta}$ *have the same bias (by law of total expectation)*
2. $\hat{\theta}_{RB}$ *is always a function of sufficient statistic* T.

Simple Hypothesis

Type I Error *incorrectly rejecting* $\mathcal{H}_0$ *(false positive)*
Type II Error *incorrectly retaining* $\mathcal{H}_0$ *(false negative)*
Significance Level: *Upper bound on size of test*

$$\alpha = \mathbb{P}(\textit{rejecting} \quad \mathcal{H}_0 | \theta = \theta_0)$$

Power $\pi = 1 - \beta$*: probability of correctly rejecting* $\mathcal{H}_0$

$$\beta = \mathbb{P}(\textit{not rejecting } \mathcal{H}_0 | \theta = \theta_1)$$

Likelihood Ratio Tests *A statistical test based on*

$$\mathcal{C} = \{\underline{x} \in \mathbb{R}^n : \lambda(\underline{x}) = \frac{\mathcal{L}(\theta_1)}{\mathcal{L}(\theta_0)} \geq c\}$$

for some c *satisfying* $\mathbb{P}(\lambda(\underline{x}) \geq c | \theta = \theta_0) = \alpha$. *First find test statistic as a simple function of* $\lambda(\underline{x})$ *and use condition on* α *to determine* c
Most Powerful Test: *We say that the most powerful (MP) test at level* α *if the significant level of the test is* α *and no other test at level* α *has a smaller* β
Neyman-Pearson Lemma *When performing a hypothesis test between two simple hypotheses* $\mathcal{H}_0 : \theta = \theta_0; \mathcal{H}_1 : \theta = \theta_1$*, the likelihood-ratio test based on rejection region*

$$\mathcal{C} = \{\underline{x} \in \mathbb{R}^n : \lambda(\underline{x}) = \frac{\mathcal{L}(\theta_1)}{\mathcal{L}(\lambda_0)} \geq c\} \qquad \mathbb{P}(\lambda(\underline{x}) \geq c | \theta = \theta_0) = \alpha$$

is the most powerful test at significance level α *for a threshold* c
Normal (known σ^2***)*** *Given* $X_1, \cdots, X_n \sim \mathcal{N}(\mu, \sigma^2)$*, with* $\mathcal{H}_0 : \mu = \mu_0; \mathcal{H}_1 : \mu = \mu_1 > \mu_0$ *(or simply* $\mathcal{H}_1 : \mu > \mu_0$*).*

$$\overline{X} \overset{\mathcal{H}_0}{\sim} \mathcal{N}(\mu_0, \frac{\sigma^2}{n}) \quad \mathcal{C} = \{(\underline{x}) : \overline{x} \geq \mu_0 + \frac{\sigma}{\sqrt{n}} z_{1-\alpha}\}$$

$$\pi(\mu^*) = \mathbb{P}(\underline{X} \in \mathcal{C} | \mu = \mu^*) = 1 - \Phi\left(\frac{-\sqrt{n}(\mu^* - \mu_0)}{\sigma} + z_{1-\alpha}\right)$$

Hence larger n *or* σ *or* α*, or further apart null and alternatives are, the larger the power. If we want probability of type II error less than* β*, we have* $n \geq \left\{\frac{\sigma(z_{1-\alpha} + z_{1-\beta})}{\mu_1 - \mu_0}\right\}^2$
Exponential *Let* $X_1, \cdots, X_n \overset{i.i.d.}{\sim} Exp(\lambda)$ *and test* $\mathcal{H}_0 : \lambda = \lambda_0$ *vs.* $\mathcal{H}_1 : \lambda = \lambda_1$. *We have MP test*

$$2\lambda \sum_{i=1}^{n} X_i \sim \chi^2_{2n} \quad \mathcal{C} = \left\{\sum_{i}^{n} X_i \leq \frac{\chi^2_{2n,\alpha}}{2\lambda_0}\right\}$$

$$\pi = \mathbb{P}\left(2\lambda_1 \sum_{i=1}^{n} X_i \leq \frac{\lambda_1 \chi^2_{2n,\alpha}}{\lambda_0} | \lambda_1\right) = F_{\chi^2_{2n}}\left(\frac{\lambda_1 \chi^2_{2n,\alpha}}{\lambda_0}\right)$$

Composite Hypothesis

Power Function $\pi(\theta^*) = \mathbb{P}(\textit{reject } \mathcal{H}_0 \,|\, \theta = \theta^*)$, $\theta^* \in \Theta_1$.
Uniformly Most Poweful (UMP) Test *A test that is MP for every simple alternative* $\theta \in \Theta_1$ *is UMP. A test at level* α *with power function* $\pi(\theta)$ *is a uniformly most powerful (UMP) test, if for any other test at level* α *with power function* $\pi'(\theta)$*, we have* $\pi'(\theta) \leq \pi(\theta)$ *for all* $\theta \in \Theta_1$. *(One tailed test for normal mean is UMP by Neyman-Pearson Lemma, two-tailed test is not)*
p-value *the probability of observing an effect at least as extreme as the one in observed data, assuming the truth of* $\mathcal{H}_0$.

$$p\text{-}value = \mathbb{P}(\textit{Type I Error}) = \mathbb{P}(T(\underline{X}) \geq t(\underline{x}) \,|\, \theta = \theta_0)$$

it is the minimum α *for which* $\mathcal{H}_0$ *will be rejected.*

$$\textit{Reject } \mathcal{H}_0 \textit{ at level } \alpha \iff p - value \leq \alpha$$

Two-tailed test for normal mean (known σ^2***)*** $\mathcal{H}_0 : \mu = \mu_0$ *vs.* $\mathcal{H}_1 : \mu \neq \mu_0$ *two-tailed p-value doubles that of one-tailed, hence harder to reject*

$$\overline{X} \overset{\mathcal{H}_0}{\sim} \mathcal{N}(\mu_0, \frac{\sigma^2}{n}) \quad \mathcal{C} = \left\{|\overline{X} - \mu_0| \geq \frac{\sigma}{\sqrt{n}} z_{1-\alpha/2}\right\}$$

$$p\text{-}value = \mathbb{P}\left(|\overline{X} - \mu_0| \geq |\overline{x} - \mu_0| \middle| \mu = \mu_0\right) = 2(1 - \Phi(\frac{|\overline{x} - \mu_0|}{\sigma/\sqrt{n}}))$$

$$\pi(\mu^*) = 1 - \Phi(-\frac{\sqrt{n}(\mu^* - \mu_0)}{\sigma} + z_{1-\alpha/2}) +$$

$$\Phi(-\frac{\sqrt{n}(\mu^* - \mu_0)}{\sigma} - z_{1-\alpha/2}) \approx 1 - \Phi(-\frac{\sqrt{n}(\mu^* - \mu_0)}{\sigma} + z_{1-\alpha/2})$$

$$95\% CI: \quad \overline{X} - \frac{\sigma}{\sqrt{n}} z_{1-\alpha/2} \leq \mu_0 \leq \overline{X} + \frac{\sigma}{\sqrt{n}} z_{1-\alpha/2}$$

Generalized LRT

Generalized Likelihood Ratio Tests (GLRT) *Consider testing* $\mathcal{H}_0 : \theta \in \Theta_0$ *vs.* $\mathcal{H}_1 : \theta \in \Theta_1$*, such that* $\Theta_0 \cup \Theta_1 = \Theta$ *(entire parameter space), based on* $X_1, \cdots, X_n \sim f_\theta$

1. *The statistic*
$$\Lambda(\underline{X}) = \frac{\sup_{\theta \in \Theta} \mathcal{L}(\theta)}{\sup_{\theta \in \Theta_0} \mathcal{L}(\theta)}$$
is called the ***generalized likelihood ratio (GLR)***
2. *The test based on the rejection region*
$$\mathcal{C} = \{\underline{X} \in \mathbb{R}^n : \Lambda(\underline{X}) \geq c\}$$
$$\sup_{\theta} \{\mathbb{P}(\Lambda(\underline{X}) \geq c \,|\, \theta \in \Theta_0)\} = \alpha$$
is called the ***generalized likelihood ratio tets (GLRT)*** *at level* α

1. ***unrestricted MLE*** $\quad \hat{\theta} = \arg\max_{\theta \in \Theta} \mathcal{L}(\theta)$ *calculate* $\mathcal{L}(\hat{\theta})$
2. ***restricted MLE*** $\quad \hat{\theta}_0 = \arg\max_{\theta \in \Theta_0} \mathcal{L}(\theta_0)$ *calculate* $\mathcal{L}(\hat{\theta}_0)$
3. ***simpler statistic*** $T(\underline{X})$ *such that* $\Lambda(\underline{X})$ *is a monotonically increasing function of* $T(\underline{X})$*, whose distribution when* $\theta = \hat{\theta}_0$ *is known.*
4. ***critical value*** c *such that* $\mathbb{P}(T(\underline{X}) \geq c | \theta = \hat{\theta}_0) = \alpha$

One Sample t test for Normal Mean $\mathcal{H}_0 : \mu = \mu_0$ *vs.* $\mathcal{H}_1 = \mu \neq \mu_0$*, based on* $X_1, \cdots, X_n \overset{i.i.d.}{\sim} \mathcal{N}(\mu, \sigma^2)$*, where* σ^2 *is unknown.*

$$\mathcal{C} = \left\{ |\mathcal{T}| = \left| \frac{\overline{X} - \mu_0}{S/\sqrt{n}} \right| \geq t_{n-1,1-\alpha/2} \right\} \quad \overline{X} \pm \frac{S}{\sqrt{n}} t_{n-1,1-\alpha/2}$$

$$\mathcal{C}_{Right} = \{\mathcal{T} \geq t_{n-1,1-\alpha}\} \quad \text{and} \quad \mathcal{C}_{Left} = \{\mathcal{T} \leq -t_{n-1,1-\alpha}\}$$

$$p-value = \mathbb{P}\left(|\mathcal{T} \geq t(\underline{x})|\mu = \mu_0\right)$$

One sample χ^2 test for Normal Variance *Let* $X_1, \cdots, X_n \sim \mathcal{N}(\mu, \sigma^2)$ *and test* $\mathcal{H}_0 : \sigma^2 = \sigma_0^2$ *vs.* $\mathcal{H}_1 : \sigma^2 \neq \sigma_0^2$

$$\mathcal{C} = \{\mathcal{X}^2 \leq \chi^2_{n-1,\alpha/2}\} \bigcup \{\mathcal{X}^2 \geq \chi^2_{n-1,1-\alpha/2}\}$$

$$\mathcal{X}^2 = \frac{1}{\sigma_0^2} \sum_{i=1}^{n} (X_i - \overline{X})^2$$

T Test for Equality of Normal Means *Suppose* $X_1, \cdots, X_n \overset{i.i.d.}{\sim} \mathcal{N}(\mu_X, \sigma^2)$ *and* $Y_1, \cdots, Y_n \overset{i.i.d.}{\sim} \mathcal{N}(\mu_Y, \sigma^2)$*, assuming same variance for the two population. Now we wish to test* $\mathcal{H}_0 : \mu_X = \mu_Y$ *vs.* $\mathcal{H}_1 : \mu_X \neq \mu_Y$

$$\mathcal{T} = \frac{\overline{X} - \overline{Y}}{S_p \sqrt{\frac{1}{m} + \frac{1}{n}}} \sim t_{m+n-2} \quad S_p^2 = \frac{(m-1)S_X^2 + (n-1)S_Y^2}{m+n-2}$$

$$\mathcal{C} = \{|\mathcal{T}| \geq t_{m+n-2,1-\alpha/2}\} \quad (\overline{X} - \overline{Y}) \pm \sqrt{\frac{1}{m} + \frac{1}{n}} S_p t_{m+n-2,1-\alpha/2}$$

F Test for Equality of Normal Variance *Suppose independent sample* $X_1, \cdots, X_m \overset{i.i.d.}{\sim} \mathcal{N}(\mu_X, \sigma_X^2)$ *and* $Y_1, \cdots, Y_n \overset{i.i.d.}{\sim} \mathcal{N}(\mu_Y, \sigma_Y^2)$ *now test* $\mathcal{H} : \sigma_X^2 = \sigma_Y^2$ *vs.* $\mathcal{H}_1 : \sigma_X^2 \neq \sigma_Y^2$.

$$\mathcal{C} = \{\mathcal{F} \leq F_{m-1,n-1,\alpha/2}\} \bigcup \{\mathcal{F} \geq F_{m-1,n-1,1-\alpha/2}\}$$

where $\mathcal{F} = \frac{S_X^2}{S_Y^2} \sim F_{m-1,n-1}$ $(F_{v_1,v_2,q} = \frac{1}{F_{v_2,v_1,1-q}})$

Paired Sample t Test for Normal Mean *Randomized paired design reduce pair-to-pair variance. Now we test* $\mathcal{H}_0 : \mu_D = 0, \mathcal{H}_1 : \mu_D > 0, \neq, < 0$ *where* $D := X - Y$ *and* $\mu_D := \mu_X - \mu_Y$ *assuming different pairs are independent and that the difference* $X - Y$ *follows a Normal distribution (problem reduces to one sample t test)*

$$\mathcal{T} = \frac{\overline{D}}{S_D/\sqrt{n}} \overset{\mathcal{H}_0}{\sim} t_{n-1}$$

Wilks' Theorem *Let* $X_1, \cdots, X_n \sim f_\theta$ *where* $\theta = (\theta_1, \cdots, \theta_p) \in \Theta$ *is a vector of parameters, and we wish to test the null hypothesis*

$$\mathcal{H}_0 : \theta_1 = \theta_1^0, \theta_2 = \theta_2^0, \cdot, \theta_r = \theta_r^0 (1 \leq r \leq p)$$

against unrestricted alternative for GLRT. Then under some regularity conditions,

$$2 \log \Lambda(\underline{X}) \xrightarrow[\mathcal{H}_0]{\mathcal{D}} \chi_r^2$$

where r is the number of paramters constrained by $\mathcal{H}_0$*, or that the d.f. equal to* $dim\Theta - dim\Theta_0$ *(the dimensinon of free parameters)*

Test for equality of Poisson means *Let* $X_1, \cdots, X_m \overset{i.i.d.}{\sim} Pois(\lambda_X)$ *and* $Y_1, \cdots, Y_n \overset{i.i.d.}{\sim} Pois(\lambda_Y)$ *be two independent samples, suppose we want to test* $\mathcal{H}_0 : \lambda_X = \lambda_Y$ *vs.* $\mathcal{H}_1 : \lambda_X \neq \lambda_Y$ *(Consider* $\mathcal{H}_0 : \theta = \frac{\lambda_X}{\lambda_Y} = 1$ *vs.* $\mathcal{H}_1 : \theta \neq 1$*)*

$$\mathcal{C} = \left\{ 2 \log \left((\overline{X})^{m\overline{X}} (\overline{Y})^{n\overline{Y}} \left(\frac{m\overline{X} + n\overline{Y}}{m+n} \right)^{-m\overline{X} - n\overline{Y}} \right) \geq \chi^2_{1,1-\alpha} \right\}$$

GLRT by parametric bootstrap *We evaluate test directly by taking large number of simulations*

1. *Sample data under* $\mathcal{H}_0$*: two random Poisson samples with same* λ *with size m and n*
2. *Evaluate test statistic* $2 \log \Lambda$ *at simulated data*
3. *repeat for very large number of times*
4. *empirical p-value given by*

$$p - value = \frac{\#\text{ of samples } 2 \log \Lambda \geq t(\underline{x})}{N}$$

Goodness of Fit Test *Suppose* $X_1, \cdots, X_n$ *is a random sample from a discrete distribution with k possible values* $s_1, \cdots, s_k$*, with corresponding probabilities* $p_1, \cdots, p_k$ *i.e.* $p_j = \mathbb{P}(X = s_j)$*. Now we dnote* $O_j = \#\{i : X_i = s_j\}$ *(the observed jth cell count) Now we want to test* $\mathcal{H}_0 : p_1 = p_1^0, \cdots, p_k = p_k^0$ *vs unrestricted alternative. Find the generlized likelihood ratio*

$$\Lambda = \prod_{i=1}^{k} \left(\frac{O_i}{np_i^0} \right)^{O_i} = \prod_{i=1}^{k} \left(\frac{O_i}{\mathbb{E}_i} \right)^{O_i}$$

where $\mathbb{E}_j := np_i^0$ *is the expected jth cell count under* $\mathcal{H}_0$ *By Wilk's Theorem*

$$2 \log \Lambda = 2 \sum_{i=1}^{k} O_i \log \left(\frac{O_i}{\mathbb{E}_i} \right) \xrightarrow[\mathcal{H}_0]{\mathcal{D}} \chi^2_{k-1}$$

Pearson's χ^2 test of goodness of fit *establishes whether an observed freuency distribution differs from a theoretical distribution. Test* $\mathcal{H}_0 : p_1 = p_1^0, \cdots, p_k = p_k^0$ *vs unrestricted alternative.*

$$\mathcal{C} = \{\mathcal{X}^2 \geq \chi^2_{k-1,1-\alpha}\} \quad \mathcal{X}^2 := \sum_{i=1}^{k} \frac{(O_i - \mathbb{E}_i)^2}{\mathbb{E}_i} \xrightarrow[\mathcal{H}_0]{\mathcal{D}} \chi^2_{k-1}$$

d.f. is the number of categories - 1 (dimΘ*) reduced by number of parameters of the fitted distribution (dim*Θ_0*, i.e. number of unknown parameter requiring estimation to compute cell probabilitie. ex. Poisson with unknown* λ $df = k - 2$*).*

Pearson's χ^2 test of independence *assess whether unpaired observations on 2 categorical variables are independent of each other. Test* $\mathcal{H}_0$*: 2 variables are independent of each other vs. unrestricted alternatives. We estimate marginal distribution of two variables using fixed table margin and law of independence* $\mathbb{P}(X, Y) = \mathbb{P}(X)\mathbb{P}(Y)$ *After computing the expected contingency table E. We use*

$$\mathcal{X}^2 = \sum_{i=1}^{\#rows} \sum_{j=1}^{\#cols} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \xrightarrow[\mathcal{H}_0]{\mathcal{D}} \chi^2_{(r-1)(c-1),1-\alpha}$$

$df = (r-1)(c-1)$ *where r is number of categories in one varialbe and c is number of categories in another variable.*

Simple Linear Regression

Method of Least Squares *A method for determining parameters in curve fitting problems. Consider fitting* $Y = \beta_0 + \beta_1 X$ *with i-th residue* $e_i = y_i - \beta_0 - \beta_1 x_i = y_i - \hat{y}_i$*. The least squares estimators of* β_0 *and* β_1 *are the minimizers of the* ***residual sum of squares (RSS)***

$$RSS(\beta_0, \beta_1) := \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2$$

In other words we choose a linear fit $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X$ *such that*

$$(\hat{\beta}_0, \hat{\beta}_1) = \underset{\beta_0, \beta_1}{\arg\min} \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2$$

Least Squares Estimators *of* β_0 *and* β_1 *are given by*

$$\hat{\beta}_1 = \frac{S_{XY}}{S_X^2} \quad \hat{\beta}_0 = \overline{y} - \hat{\beta}_1 \overline{x}$$

$$S_{XY} = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y}) \quad S_X^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \overline{x})^2$$

where S_{XY} is the sample covariance of X and Y and S_X^2 is the sample variance of X. The following properties are handy

$$\sum_{i=1}^n (x_i-\overline{x})^2 = \sum_{i=1}^n x_i^2 - 2n\overline{x}^2 + n\overline{x}^2 = \sum_{i=1}^n x_i^2 - n\overline{x}^2$$

$$\sum_{i=1}^n (x_i-\overline{x})(y_i-\overline{y}) = \sum_{i=1}^n x_iy_i - 2n\overline{xy} + n\overline{xy} = \sum_{i=1}^n x_iy_i - n\overline{xy}$$

The normal equations

$$0 = \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = \sum_{i=1}^n (y_i - \hat{y}_i) = \sum_{i=1}^n e_i$$

$$0 = \sum_{i=1}^n x_i(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = \sum_{i=1}^n x_i(y_i - \hat{y}_i) = \sum_{i=1}^n x_i e_i$$

Standard Statistical Model *stipulates that the observed value of y is a linear function of x plus random noise*

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \quad i = 1, \cdots, n$$

$$\mathbb{E}[\epsilon_i] = 0 \ \forall i \quad Var(\epsilon_i) = \sigma^2 \ \forall i \quad \mathbb{E}[\epsilon_i\epsilon_j] = 0 \text{ for } i \neq j$$

$$\text{hence} \quad Var[y_i] = 0 \qquad Cov(y_i, y_j) = 0 \text{ for } i \neq j$$

Homoscedastic *Variance around regression line is same $\forall X$*

LS estimator as linear estimator *Let $y_1, \cdots, y_n \sim f_\theta$. Any estimator of θ of the form $\hat{\theta} = \sum_{i=1}^n c_i y_i$ is called a linear estimator*

$$\hat{\beta}_1 = \sum_i \left[\frac{(x_i - \overline{x})}{\sum_j (x_j - \overline{x})^2}\right] y_i \quad \hat{\beta}_0 = \sum_i \left[\frac{1}{n} - \frac{\overline{x}(x_i - \overline{x})}{\sum_j (x_j - \overline{x})^2}\right] y_i$$

$$\mathbb{E}[y_i] = \beta_0 + \beta_1 x_i \quad \mathbb{E}[\hat{\beta}_1] = \beta_1 \quad \mathbb{E}[\hat{\beta}_0] = \beta_0$$

$$Var[\hat{\beta}_1] = \frac{\sigma^2}{\sum_i (x_i - \overline{x})^2} \quad Var[\hat{\beta}_0] = \sigma^2 \left[\frac{1}{n} + \frac{\overline{x}^2}{\sum_i (x_i - \overline{x})^2}\right]$$

$$Cov(\hat{\beta}_0, \hat{\beta}_1) = -\frac{\sigma^2 \overline{x}}{\sum_i (x_i - \overline{x})^2} \qquad S_\epsilon^2 = \frac{1}{n-2}\sum_{i=1}^n e_i^2$$

The Gauss-Markov Theorem *Under standard model assumptions, no linear unbiased estimator of β_0 (β_1) has a smaller variance than the least squares estimator $\hat{\beta}_0$ ($\hat{\beta}_1$). Hence LS estimators are Best Linear Unbiased Estimators*

Correlation Coefficient *The correlation coefficient of random variables X and Y is*

$$\rho_{XY} = \frac{Cov(X, Y)}{\sigma_X \sigma_Y}$$

where σ_X and σ_Y are the standard deviations of X and Y, respectively.

Sample Correlation Coefficient *is defined to be*

$$r_{XY} = \frac{S_{XY}}{S_X S_Y} = \frac{\sum_i (x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_i (x_i - \overline{x})^2}\sqrt{\sum_i (y_i - \overline{y})^2}}$$

Explained Variation *Variation in value of Y is the Total Sum of Squares*

$$\sum_{i=1}^n (y_i - \overline{y})^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 + \sum_{i=1}^n (\hat{y}_i - \overline{y})^2$$

$$TSS = RSS + ESS$$

*the **Proportion of explained variance** is defined to be*

$$R^2 = \frac{ESS}{TSS} = \frac{\sum_{i=1}^n (\hat{y}_i - \overline{y})^2}{\sum_{i=1}^n (y_i - \overline{y})^2} \qquad \text{and} \qquad R^2 = r_{XY}^2$$

R^2 is an indication of good linear fit

Statistical Inference under Gaussian Noise *A linear model with following assumptions allows for statistical inference*

1. *$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$ where $i = 1, \cdots, n$*
2. *$\mathbb{E}[\epsilon_i] = 0$ for $i = 1, \cdots, n$*
3. *$Var(\epsilon_i) = \sigma^2$ $i = 1, \cdots, n$ (homoscedastic) and $\mathbb{E}[\epsilon_i\epsilon_j] = 0$ for $i \neq j$ (uncorrelated)*
4. *distribution of ϵ_i is normal for $i = 1, \cdots, n$*
5. *Uncorrelated normal random variable is independent.*

Inference on regression coeffcient *An unbiased estimator of noise variance σ^2 is given by*

$$S^2 = \frac{1}{n-2}\sum_{i=1}^n e_i^2 \quad \text{with} \quad \frac{(n-2)S^2}{\sigma^2} \sim \chi^2_{n-2}$$

Since ϵ_i is normal, we have

$$y_i \sim \mathcal{N}(\beta_0 + \beta_1 x_i, \sigma^2) \quad \hat{\beta}_1 \sim \mathcal{N}(\beta_1, \frac{\sigma^2}{\sum_i (x_i - \overline{x})^2})$$

Hypothesis tests on the slope β_1 to evaluate correlation *Testing $\mathcal{H}_0 : \beta_1 = 0$ vs. $H_1 : \beta_1 \neq 0$, then, by*

$$\mathcal{T} = \frac{\hat{\beta}_1 - \beta_1}{s_{\hat{\beta}_1}} \overset{H_0}{\sim} t_{n-2} \qquad \text{where} \qquad s_{\hat{\beta}_1} = \sqrt{\frac{\frac{1}{n-2}\sum_i e_i^2}{\sum_i (x_i - \overline{x})^2}}$$

and a $100(1-\alpha)\%$ confidence interval for β_1 is given by

$$\hat{\beta}_1 \pm t_{n-2,1-\alpha/2} \frac{S}{\sqrt{\sum_i (x_i - \overline{x})^2}}$$

The prediction at x_0 may be used as an estimator

$$\hat{y}(x_0) = \hat{\beta}_0 + \hat{\beta}_1 x_0$$

$$\mathbb{E}[\hat{y}(x_0)] = \mathbb{E}[\hat{\beta}_0 + \hat{\beta}_1 x_0] = \mathbb{E}[\hat{\beta}_0] + \mathbb{E}[\hat{\beta}_1]x_0 = \beta_0 + \beta_1 x_0$$

$$Var[\hat{y}(x_0)] = Var[\hat{\beta}_0 + \hat{\beta}_1 x_0] = \sigma^2 \left\{\frac{1}{n} + \frac{(x_0 - \overline{x})^2}{\sum_i (x_i - \overline{x})^2}\right\}$$

$$\hat{y}(x_0) \sim \mathcal{N}\left(\mu(x_0), \sigma^2 \left\{\frac{1}{n} + \frac{(x_0 - \overline{x})^2}{\sum_i (x_i - \overline{x})^2}\right\}\right)$$

$$\hat{y}(x_0) \pm t_{n-2,1-\alpha/2} S \sqrt{\frac{1}{n} + \frac{(x_0 - \overline{x})^2}{\sum_i (x_i - \overline{x})^2}}$$

is a $100(1-\alpha)\%$ confidence interval for mean response $\mathbb{E}[y(x_0)]$, where $\hat{y}(x_0) = \hat{\beta}_0 + \hat{\beta}_1 x_0$.

Least Square Estimators under standard normal model are MLEs *Under additional assumption that random noise is Gaussian, the least squares estimators $\hat{\beta}_0^{LS}$ and $\hat{\beta}_1^{LS}$ are maximum likelihood estimators of β_0 and β_1, respectively.*

Diagnostic Plot *Linear regression model relies heavily on assumptions about random errors ϵ_i. the residual e_i should be*

1. *normal*
2. *independent*
3. *homoscedasticitic*
4. *distribution of **standardized residuals** $\frac{e_i}{S} \sim \mathcal{N}(0, 1)$*

Two plots are given

1. ***Residuals vs Fitted Value Plot*** *plot of e_i vs. $\hat{y}_i$.*
   (a) *Symmetry about 0, with homogeneity of the noise variance (homoscedastic), and no trends or pattern implies a good fit for linear models*
   (b) *Streaks of positive/negative residual indicates observation is correlated, violating the independence assumption*
   (c) *The trend resembles an upward or downward curve indicates model misspecification. The assumption of linearity is violated*
   (d) *Increasing variance along the dependent $\hat{y}_i$ axis violates homoscedasticity assumption*

2. ***Quantile-Quantile Plot*** *A plot for comparing two probability distributions by plotting their quantiles against each other. In evaluating good fit for linear model we plot sample quantiles of standardized residues vs. theoretical quantiles of standard normal distribution.*

   (a) *If points approximately lie on the line $y = x$, then the distribution in comparison are similar, i.e.* $\frac{e_i}{S} \sim \mathcal{N}(0, 1)$.

   (b) *If lower quantiles are too small and upper quantiles are too large - a heavy-tailed noise. Perhaps assuming t distribution, thus violating the normality assumption*