ST2132

AY23/24 SEM 1

github/jovyntls

01. PROBABILITY

Expectation

discrete: (mass)

$$E(X) := \sum_{i=1}^{n} x_i p_i$$

continuous: (density)

$$E(X) := \int_{-\infty}^{\infty} x f(x)\, dx$$

expectation of a function $h(X)$

$$E\{h(X)\} = \begin{cases} \sum_{i=1}^{n} h(x_i) p_i & X \text{ is discrete} \\ \int_{-\infty}^{\infty} h(x) f(x)\, dx & X \text{ is continuous} \end{cases}$$

Variance

variance, $\text{var}(X) := E\{(X-\mu)^2\}$

$$= E(X^2) - E(X)^2$$

standard deviation, $SD(X) := \sqrt{\text{var}(X)}$

useful cases

- $E\{X(X-\mu)\} = E(X^2) - \mu^2$
- $\text{var}(X - c) = \text{var}(X)$
- variance of sum = sum of variances
  $\text{var}(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} \text{var}(x_i)$

Law of Large Numbers

LLN: for a function h, as realisations $r \to \infty$,

$$\frac{1}{r}\sum_{i=1}^{r} h(x_i) \to E\{h(X)\}$$

$$\bar{x} \to E(X), \quad v \to \text{var}(X)$$

Monte Carlo approximation

simulate $x_1, \ldots, x_r$ from X. by LLN, as $r \to \infty$, the approximation becomes exact

$$E\{h(X)\} \approx \frac{1}{r}\sum_{i=1}^{r} h(x_i)$$

Joint Distribution

(**discrete**) mass function:

$$P(X = x_i, Y = y_j) = p_{ij}$$

(**continuous**) density function:

$$f : \mathbb{R}^2 \to [0, \infty), \quad \int_{-\infty}^{\infty} f(x,y)\, dx\, dy = 1$$

(**expectation**) for $h : \mathbb{R}^2 \to \mathbb{R}$,

$$E\{h(X,Y)\} = \begin{cases} \sum_{i=1}^{I}\sum_{j=1}^{J} h(x_i, y_j) p_{ij} & X \text{ is discrete} \\ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(x,y) f(x,y)\, dx\, dy & Y \text{ is continuous} \end{cases}$$

Covariance

let $\mu_X = E(X)$, $\mu_Y = E(Y)$.

covariance

$$\begin{aligned} \text{cov}(X,Y) &= E\{(X-\mu_X)(Y-\mu_Y)\} \\ &= E(XY) - \mu_X\mu_Y \\ &= \text{cov}(Y,X) \end{aligned}$$

$$\text{cov}(W, aX + bY + c) = a\,\text{cov}(W,X) + b\,\text{cov}(W,Y)$$

variance

$$\text{var}(X) = \text{cov}(X,X)$$

$$\text{var}(\textstyle\sum_{i=1}^{N} a_i X_i) = \sum_{i=1}^{N} a_i^2\, \text{var}(X_i) + 2\sum_{1 \le i < j \le N} a_i a_j\, \text{cov}(X_i, X_j)$$

joint $=$ marginal $\times$ conditional distributions

$$\begin{aligned} f(x,y) &= f_X(x) f_Y(y|x) \\ &= f_Y(y) f_X(x|y), \quad x, y \in \mathbb{R} \end{aligned}$$

- $f(x,y)$ is the *joint density*
- $f_X(x), f_Y(y)$ are the *marginal densities*
- $f_X(\cdot|y)$ is the **conditional** density of X given $Y = y$
- for discrete case, *density* $\equiv$ *probability*, $x \equiv x_i$, $y \equiv y_j$

Independence

- X, Y are independent $\iff \forall x, y \in \mathbb{R}$,
  1. $f(x,y) = f_X(x) f_Y(y)$
  2. $f_Y(y|x) = f_Y(y)$
  3. $f_X(x|y) = f_Y(x)$
- X, Y are independent $\Rightarrow$
  - $E(XY) = E(X)E(Y)$
  - $\text{cov}(X,Y) = 0$

  (the converse does not hold)

Conditional expectation

discrete case

let $f_Y(\cdot|x_i)$ be the conditional pmf of Y given $X = x_i$.

$$E[Y|x_i] := \sum_{j=1}^{J} y_j f_Y(y_j|x_i)$$

$$\text{var}[Y|x_i] := \sum_{j=1}^{J} (y_j - E[Y|x_i])^2 f_Y(y_j|x_i)$$

$E[Y|x_i]$ is like $E(Y)$, with conditional distribution replacing marginal distribution $f_Y(\cdot)$. likewise, $\text{var}[Y|x_i]$ like $\text{var}(Y)$.

continuous case

$$E[Y|x] := \int_{-\infty}^{\infty} y f_Y(y|x)\, dy$$

$$\begin{aligned} \text{var}[Y|x] &:= \int_{-\infty}^{\infty} (y - E[Y|x])^2 f_Y(y|x)\, dy \\ &= E(Y^2|x) - \{E(Y|x)\}^2 \end{aligned}$$

Distributions

if X is iid with expectation μ, SD σ and $S_n = \sum_{i=1}^{n} X_i$,

distribution of X	$E(X)$	$\text{var}(X)$
$Bernoulli(p)$	p	$p(1-p)$
$Binomial(n,p)$	np	$np(1-p)$
$Geometric(n,p)$	$1/p$	$(1-p)/p^2$
$Multinomial(n, \mathbf{p})$	$\begin{bmatrix} np_1 \\ np_2 \\ \vdots \\ np_k \end{bmatrix}$	$\text{var}(X_i) = np_i(1-p_i)$ $\text{var}(X) =$ *covariance matrix* M with $m_{ij} = \begin{cases} \text{var}(X_i) & \text{if } i = j \\ \text{cov}(X_i, X_j) & \text{if } i \neq j \end{cases}$

- binomial: n coin flips (bernoulli) with probability p
  - $X \sim Bin(n,p) \quad \Rightarrow X_i \overset{i.i.d.}{\sim} Bernoulli(p)$
  - $P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$
  - $\text{cov}(X, n - X) = -\text{var}(X)$
- multinomial: tally of k possible outcomes (n events)
  - $\text{cov}(X_i, X_j) < 0$
  - $X_i \sim Bin(n, p_i)$
  - $X_i + X_j \sim Bin(n, p_i + p_j)$

02. PROBABILITY (2)

Mean Square Error (MSE)

$$\begin{aligned} MSE &= E\{(Y-c)^2\} \\ &= \text{var}(Y) + \{E(Y) - c\}^2 \end{aligned}$$

$\min MSE = \text{var}(Y)$ when $c = E(Y)$

if Y and X are correlated:

$$MSE = \text{var}[Y|x] + \{E[Y|x] - c\}^2$$

mean MSE

$$\frac{1}{n}\sum_{i=1}^{n} \text{var}[Y|x_i] \approx E\{\text{var}[Y|X]\}$$

random conditional expectations

- $E[Y|X]$ is a r.v. which takes value $E[Y|x]$ with probability/density $f_X(x)$
- $\text{var}[Y|X]$ is a r.v. which takes value $\text{var}[Y|x]$ with probability/density $f_X(x)$

$$E(E[X_2|X_1]) = E(X_2)$$

$$\text{var}(E[X_2|X_1]) + E(\text{var}[X_2|X_1]) = \text{var}(X_2)$$

CDF (cumulative distribution function)

for r.v. X, let $F(x) = P(X \le x)$

- domain: $\mathbb{R}$; codomain: $[0,1]$

$$F(x) = \int_{-\infty}^{\infty} f(x)\, dx$$

Standard Normal Distribution

$Z \sim N(0,1)$ has density function

$$\phi(z) = \frac{1}{\sqrt{2\pi}} \exp\{-\frac{z^2}{2}\}, \quad -\infty < z < \infty$$

$$E(Z) = 0, \quad \text{var}(Z) = 1$$

CDF, $\Phi(x) = P(Z \le x) = \int_{-\infty}^{x} \phi(z)\, dz$

- $E(Z^2) = 1$

general normal distribution

standardisation: $\frac{X-\mu}{\sigma} \sim N(0,1)$

- for $W = a + bX$,
  - density, $f_W(w) = \frac{d}{dw} F_W(w)$
  - CDF, $F_W(w) = P(X \le \frac{w-a}{b}) = \Phi(\frac{w-a}{b})$

Central Limit Theorem

CLT

as $n \to \infty$, the distribution of the standardised $S_n = \frac{S_n - n\mu}{\sqrt{n}\sigma}$ converges to $N(0,1)$

for large n, approximately $S_n \sim N(n\mu, n\sigma^2)$

Distributions

chi-square (χ^2)

let $Z \sim N(0,1)$. $\Rightarrow$ then $Z^2 \sim \chi_1^2$ (1 degree of freedom)

- degrees of freedom = number of RVs in the sum

$$E(Z^2) = 1, \quad E(Z^4) = 3$$

$$\text{var}(Z^2) = E(Z^4) - \{E(Z^2)\}^2 = 2$$

let $V_1, \ldots, V_n$ be iid χ_1^2 RVs and $V = \sum_{i=1}^{n} V_i$. then

$$V \sim \chi_n^2$$

$$E(V) = n \quad \text{var}(V) = 2n$$

gamma

let shape parameter $\alpha > 0$, rate parameter $\lambda > 0$.

The $Gamma(\alpha, \lambda)$ density is

$$\frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}, \quad x > 0$$

$\Gamma(\alpha)$ is a number that makes density integrate to 1

$$E(X) = \frac{\alpha}{\lambda}, \quad \text{var}(X) = \frac{\alpha}{\lambda^2}$$

$$\Gamma(\alpha+1) = \alpha\Gamma(\alpha)$$

- if $X_1 \sim Gamma(\alpha_1, \lambda)$ and $X_2 \sim Gamma(\alpha_2, \lambda)$ are independent, then $X_1 + X_2 \sim Gamma(\alpha_1 + \alpha_2, \lambda)$

t distribution

let $Z \sim N(0,1)$ and $V \sim \chi_n^2$ be independent.

$$\frac{Z}{\sqrt{V/n}} \sim t_n$$

has a t distribution with n degrees of freedom.

- t distribution is symmetric around 0
- $t_n \to Z$ as $n \to \infty$ (because $\frac{V}{n} \to 1$)

F distribution

let $V \sim \chi_m^2$ and $W \sim \chi_n^2$ be independent.

$$\frac{V/m}{W/n} \sim F_{m,n}$$

has an F distribution with (m,n) degrees of freedom.

- even if $m = n$, still two RVs V, W as they are independent

IID Random Variables

let $X_1, \ldots, X_n$ be iid RVs with mean $\bar{X}$.

sample variance, $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$

$E(S^2) = \sigma^2$ but $E(S) < \sigma$

more distributions:

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$$
$\bar{X}$ and S^2 are independent

$$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$$
$$\frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t_{n-1}$$

Multivariate Normal Distribution

let $\boldsymbol{\mu}$ be a $k \times 1$ vector and $\boldsymbol{\Sigma}$ be a *positive-definite* symmetric $k \times k$ matrix.

the random vector $\boldsymbol{X} = (X_1, \dots, X_k)'$ has a multivariate normal distribution $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$
$$E(\boldsymbol{X}) = \boldsymbol{\mu}, \quad \text{var}(\boldsymbol{X}) = \boldsymbol{\Sigma}$$

- two multinomial normal random vectors $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$, sizes h and k, are independent if $\text{cov}(\boldsymbol{X}_1, \boldsymbol{X}_2) = \boldsymbol{0}_{h \times k}$

03. POINT ESTIMATION

for a variable v in population N,
$$\mu = \frac{1}{N}\sum_{i=1}^{N} v_i \qquad \sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(v_i - \mu)^2$$

- $\mu,\ \sigma^2$ are **parameters** (unknown constants)

draws with replacement

$$\text{random sample mean, } \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$
$$E(\bar{X}) = \mu,\ \text{var}(\bar{X}) = \frac{\sigma^2}{n}$$
$$E(X_i) = \mu, \qquad \text{var}(X_i) = \sigma^2$$

- same distribution: x_i, X_i, population distribution
- the error in $\bar{x}$ is $\mu - \bar{x}$; it cannot be estimated

representativeness

- $X_1, \dots, X_n$ is **representative** of the population
  - as n gets larger, $\bar{X}$ gets closer to μ
- $x_1, \dots, x_n$ are *likely* representative of the population

Point estimation of mean

a population (size N) has unknown mean μ, variance σ^2.

standard error

SE is a constant by definition:
$$SE = SD(\bar{X}) = \frac{\sigma}{\sqrt{n}}$$
point estimation of mean: SE ($\bar{x}$) is estimated as $\frac{s}{\sqrt{n}}$

Simple random sampling (SRS)

n random draws *without replacement* from a population

for $i \neq j$, $\text{cov}(X_i, X_j) = -\frac{\sigma^2}{N-1}$

- if n/N is relatively large, account for $\text{cov}(X_i, X_j)$
$$E(\bar{X}) = \mu, \quad \text{var}(\bar{X}) = \frac{N-n}{N-1}\frac{\sigma^2}{n}$$
- if $n << N$, then SRS is like sampling *with replacement* (treat the data as IID RVs $X_1, \dots, X_n$)
$$E(\bar{X}) = \mu, \quad \text{var}(\bar{X}) = \frac{\sigma^2}{n}$$

estimating proportion p

- the estimate of σ is $\hat{\sigma}$, not s
- unbiased estimator $\hat{p}$
  - $E(\hat{p}) = p, \quad \text{var}(\hat{p}) = \frac{p(1-p)}{n}, \quad SE = SD(\hat{p})$

04. ESTIMATION (SE, bias, MSE)

for random draws $X_1, \dots, X_n$ *with replacement*

MSE and bias

suppose measurements were from a population with mean $w + b$ where b is a constant: $\quad x_i = w + b + \epsilon_i$

- $E(\bar{X}) = w + b, \quad SD(\bar{X}) = \frac{\sigma}{\sqrt{n}}$
  - $SE = \frac{\sigma}{\sqrt{n}}$ measures how far $\bar{x}$ is from $w + b$, not w
- if $b \neq 0$, then $\bar{x}$ is a biased estimate for w
- $MSE = E\{(\bar{X} - w)^2\} = \frac{\sigma^2}{n} + b^2$

general case

let θ be a parameter and $\hat{\theta}$ be an estimator (RV).
$$SE = SD(\hat{\theta}), \qquad \text{bias} = E(\hat{\theta}) - \theta,$$
$$MSE = E\{(\hat{\theta} - \theta)^2\} = SE^2 + bias^2$$
as $n \to \infty,\ MSE \to b^2$

05. INTERVAL ESTIMATION

let $x_1, \dots, x_n$ be realisations of IID RVs $X_1, \dots, X_n$ with unknown $\mu = E(X_i)$ and $\sigma^2 = \text{var}(X_i)$.

point estimation: $\mu \approx \bar{x} \pm \frac{s}{\sqrt{n}}$
interval estimation: interval contains μ with some confidence level

interval estimation works well if
- X_i has a normal distribution, for any $n > 1$
- X_i has any other distribution but n is large

normal "upper-tail quantile" z_p

let $Z \sim N(0,1)$. let z_p be the $(1-p)$-quantile of Z.
$$p = \Pr(Z > z_p)$$

(case 1) normal distribution with known σ^2

$X_1, \dots, X_n \overset{i.i.d.}{\sim} N(0,1)$ with known σ^2.
for $0 < \alpha < 1$, $\Pr(-z_{\frac{\alpha}{2}} \leq Z \leq z_{\frac{\alpha}{2}}) = 1 - \alpha$

confidence interval for μ: the random interval
$$\left(\bar{X} - z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}, \bar{X} + z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right)$$
contains μ with probability (confidence level) $1 - \alpha$

(case 2) normal distribution with unknown σ^2

replace σ with S and use t distribution:

for $0 < p < 1$, let $t_{p,n}$ be such that
$$\Pr(t_n > t_{p,n}) = p$$
as $n \to \infty, \quad t_{n,p} \to z_p$

the random interval
$$\left(\bar{X} - t_{\frac{\alpha}{2}, n-1}\frac{S}{\sqrt{n}}, \bar{X} + t_{\frac{\alpha}{2}, n-1}\frac{S}{\sqrt{n}}\right)$$
contains μ with probability $1 - \alpha$.

(case 3) general distribution with unknown σ^2

- CLT: for large n, approximately $\frac{S_n - n\mu}{\sqrt{n}\sigma} \sim N(0,1)$
- since $\frac{S_n - n\mu}{\sqrt{n}\sigma} = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$,

for large n, the random interval
$$\left(\bar{X} - z_{\frac{\alpha}{2}}\frac{S}{\sqrt{n}}, \bar{X} + z_{\frac{\alpha}{2}}\frac{S}{\sqrt{n}}\right)$$
contains μ with probability $\approx 1 - \alpha$

- for SRS, multiply SE by correction factor $\sqrt{\frac{N-n}{N-1}}$
- contains μ with probability $< 1 - \alpha$
- probability $\to 1 - \alpha$ as $n \to \infty$
- **exception**: for Bernoulli, $\sigma = \sqrt{p(1-p)}$ is not estimated by s, but by replacing p with the sample proportion

06. METHOD OF MOMENTS

modified notation of mass/density functions:
- **bernoulli**: $f(x|p) = p^x(1-p)^{1-x}, \quad x = 0, 1$
  - parameter space is $(0, 1)$
- **poisson**: $f(x|\lambda) = \frac{\lambda^x e^{-\lambda}}{x!}, \quad x = 0, 1, \dots$
  - parameter space is $\mathbb{R}_+$

parameter estimation

assuming data $x_1, \dots, x_n$ are realisations of IID RVs $X_1, \dots, X_n$ with mass/density function $f(x|\theta)$, where θ is unknown in parameter space Θ.
- 2 methods to estimate θ :
  - method of moments (MOM)
  - method of maximum likelihood (MLE)
- the estimate of θ is a realisation of an estimator $\hat{\theta}$
- parameter space Θ: set of values that can be used to estimate the real parameter value θ
  - e.g. for $N(\mu, \sigma^2)$, parameter space $\Theta = \mathbb{R} \times \mathbb{R}_+$

Moments of an RV

the k-th moment of an RV X is
$$\mu_k = E(X^k), \quad k = 1, 2, \dots$$

estimating moments

let $X_1, \dots, X_n$ be IID with the same distribution as X.

the k-th sample moment is
$$\hat{\mu}_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$$
$$E(\hat{\mu}_k) = E(\tfrac{1}{n}\sum_{i=1}^{n} x_i^k) = \mu_k \quad \Rightarrow \text{unbiased!}$$

MOM: general

let $X \sim Distribution(\theta)$. to obtain $\bar{x}$ and SE:

1. $\mu = \mu_1, \quad \sigma^2 = \mu_2 - \mu_1^2$
2. express parameters in terms of moments
3. estimate MOM estimator using sample mean $\bar{x}$: $\hat{\theta} = \hat{\mu}_1 = \bar{X}$
4. obtain $SE = SD(\hat{\theta}) = \sqrt{\text{var}(\hat{\theta})} = \sqrt{\frac{1}{n}\text{var}(X)}$
$$\theta \approx \bar{x} \pm \sqrt{\frac{\text{var}(X)}{n}}$$

07. MLE

Likelihood function

let $x_1, \dots, x_n$ be realisations of iid rvs $X_1, \dots, X_n$ with density $f(x|\theta),\ \theta \in \Theta \subset \mathbb{R}^k$.

likelihood function $L : \Theta \to \mathbb{R}_+$ is
$$L(\theta) = \prod_{i=1}^{n} f(x_i|\theta)$$
$$= f(x_1|\theta) \times \dots \times f(x_n|\theta)$$

loglikelihood function $\ell : \Theta \to \mathbb{R}$ is
$$\ell(\theta) = \log L(\theta) = \sum_{i=1}^{n} \log f(x_n|\theta)$$
(can omit additive constants (ℓ)/constant factors (L))

Maximum Likelihood Estimation (MLE)

- **maximiser** of $L \to$ the maximum likelihood estimate of θ (a realisation of the MLEstimator $\hat{\theta}$)
  - maximiser of loglikelihood $\ell = \log L$ over Θ

find the value of θ that maximises (log)likelihood:
1. calculate likelihood L, loglikelihood ℓ
2. differentiate loglikelihood ℓ: $\ell'(\theta) = 0$
3. confirm max point: $\ell''(\theta) < 0$

ML vs MOM

- MOM estimates can always be written in terms of the data (sample moments)
  - ML uses *
- ML has better (smaller) SE and bias than MOM
- MOM/ML estimates are asymptotically unbiased
  - as $n \to \infty$, $E(\hat{\theta}_n) \to \theta$

Kullback-Liebler divergence (KL)

let $\mathbf{q} = (q_1, \dots, q_k)$ and $\mathbf{p} = (p_1, \dots, p_k)$ be strictly positive probability vectors.

the **KL divergence** between $\mathbf{q}$ and $\mathbf{p}$ is
$$d_{KL}(\mathbf{q}, \mathbf{p}) = \sum_{i=1}^{k} q_i \log(\frac{q_i}{p_i})$$
- $d_{KL}(\mathbf{q}, \mathbf{p}) \geq 0 \qquad$ (equality $\iff \mathbf{q} = \mathbf{p}$)
- $d_{KL}(\mathbf{q}, \mathbf{p}) \neq d_{KL}(\mathbf{p}, \mathbf{q})$

- used to maximise ℓ to find MLE for multinomial
- let $\mathbf{q}$ be the MOM estimate for $\mathbf{p}$. for any $\mathbf{p}$,
$$\ell(\mathbf{q}) - \ell(\mathbf{p}) = \sum_{i=1}^{k} x_i \log q_i - \sum_{i=1}^{k} x_i \log p_i$$
$$= n\, d_{KL}(\mathbf{q}, \mathbf{p}) \geq 0$$
  - $\ell(\mathbf{q}) - \ell(\mathbf{p}) = 0 \iff \mathbf{p} = \mathbf{q} = \frac{\mathbf{x}}{n}$

Hardy-Weinberg equilibrium (HWE)

let θ be the proportion of a.

the population is in **HWE** if
$$f(aa) = \theta^2, \quad f(aA) = 2\theta(1-\theta), \quad f(AA) = (1-\theta)^2$$

- (e.g. genotypes) Under HWE, the number of a alleles in an individual has a $Binom(2, \theta)$ distribution
  - for n randomly chosen people, number of a alleles $(AA, Aa, aa) \sim Multinomial(n, \theta)$

Multinomial ML estimation

for $(X_1, X_2, X_3) \sim Multinomial(n, \mathbf{p})$
where $p_1 = (1-\theta)^2$, $p_2 = 2\theta(1-\theta)$, $p_3 = \theta^2$

- $L(\theta) = p_1^{x_1} p_2^{x_2} p_3^{x_3} \quad = 2^{x_2}(1-\theta)^{2x_1+x_2}\theta^{x_2+2x_3}$
- $\ell(\theta) = x_2 \log 2 + (2x_1+x_2)\log(1-\theta) + (x_2+2x_3)\log\theta$
- ML estimator: $\hat\theta = \frac{X_2+2X_3}{2n}$
- SE estimation: $\sqrt{\frac{\theta(1-\theta)}{2n}}$
  - $X_2 + 2X_3$ is the number of a alleles: $Binom(2n, \theta)$
    $\Rightarrow \text{var}(\hat\theta) = \frac{\theta(1-\theta)}{2n}$

08. LARGE-SAMPLE DISTRIBUTION OF MLEs

asymptotic normality of ML estimator

let $\hat\theta_n$ be the ML estimator of $\theta \in \Theta \subset \mathbb{R}$, based on iid RVs $X_1, \dots, X_n$ with density $f(x|\theta)$.

> for large n, approximately
> $$\hat\theta_n \sim N(\theta, \frac{\mathcal{I}(\theta)^{-1}}{n})$$

Fisher Information

let X have density $f(x|\theta)$, $\theta \in \Theta \subset \mathbb{R}^p$.

> the **Fisher information** is the $p \times p$ matrix
> $$\mathcal{I}(\theta) = -E\left[\frac{d^2 \log f(X|\theta)}{d\theta^2}\right]$$

- $\mathcal{I}(\theta)$ is symmetric, with (ij)-entry $-E\left[\frac{\delta^2 \log f(X|\theta)}{\delta\theta_i\delta\theta_j}\right]$
- $\mathcal{I}(\theta)$ measures the information about θ in one sample X.

Asymptotic normality: Bernoulli

$X \sim Bernoulli(p): \; f(x|p) = p^x(1-p)^{1-x}, \; x = 0, 1$

Fisher information

- $\log f(X|p) = X\log p + (1-X)\log(1-p)$
- differentiate $\frac{d}{dp}$: $\frac{X}{p} - \frac{1-X}{1-p}$
- differentiate $\frac{d^2}{dp^2}$: $-\frac{X}{p^2} - \frac{1-X}{(1-p)^2}$
- $\mathcal{I}(p) = -E(\frac{d^2 \log f(X|p)}{dp^2}) = \frac{1}{p(1-p)}$
  - minimised at $p = 0.5$

Asymptotic normality

for $X_1, \dots, X_n$ iid $Bernoulli(p)$ RVs,
Fisher information in each X_i: $\mathcal{I}(p) = \frac{1}{p(1-p)}$

- ML estimator $\hat p = \bar X$
- for large n, $\hat p \approx N\left(p, \frac{p(1-p)}{n}\right)$
  - $E(\hat p) = p, \quad \text{var}(\hat p) = \frac{p(1-p)}{n}$

Asymptotic normality: Geometric

$X \sim Geometric(p): \; f(x|p) = p(1-p)^{1-x}$

Fisher information

- $\log f(X|p) = \log p + (X-1)\log(1-p)$
- differentiate $\frac{d}{dp}$: $\frac{1}{p} - \frac{X-1}{1-p}$
- differentiate $\frac{d^2}{dp^2}$: $-\frac{1}{p^2} - \frac{X-1}{(1-p)^2}$
- $\mathcal{I}(p) = -E(\frac{d^2 \log f(X|p)}{dp^2}) = \frac{1}{p(1-p)} + \frac{1}{p^2} = \frac{1}{p^2(1-p)}$

Asymptotic normality

for $X_1, \dots, X_n$ iid $Geometric(p)$ RVs,
Fisher information in each X_i, $\mathcal{I}(p) = \frac{1}{p^2(1-p)}$

- ML estimator $\hat p = \frac{1}{\bar X}$
- for large n, $\hat p \approx N\left(p, \frac{p^2(1-p)}{n}\right)$
  - $E(\hat p) > p$ since $E(\hat p) = E(\frac{1}{\bar X}) > \frac{1}{E(\bar X)} = p$
  - likely $\text{var}(\hat p) \neq \frac{p^2(1-p)}{n}$

Asymptotic normality: Normal

Fisher information

$X \sim N(\mu, \sigma^2), \theta = (\mu, \sigma)$.

$f(x|p) = \frac{1}{\sqrt{2\pi}\sigma}\exp\{-\frac{(x-\mu)^2}{2\sigma^2}\}, \quad x \in \mathbb{R}$

- $\log f(X|p) = \frac{1}{2}\log 2\pi - \log\sigma - \frac{(x-\mu)^2}{2\sigma^2 n}$
  $= c - \log\sigma - \frac{(X-\mu)^2}{2\sigma^2 n}$
- differentiate $\frac{d}{dp}$: $\quad \frac{\delta}{\delta\mu} = \frac{X-\mu}{\sigma^2}, \quad \frac{\delta}{\delta\sigma} = -\frac{1}{\sigma} + \frac{(X-\mu)^2}{\sigma^3}$
- differentiate $\frac{d^2}{dp^2}$: $\begin{bmatrix} \frac{\delta^2}{\delta\mu^2} & \frac{\delta^2}{\delta\mu\delta\sigma} \\ \frac{\delta^2}{\delta\sigma\delta\mu} & \frac{\delta^2}{\delta\sigma^2} \end{bmatrix}$
- $\mathcal{I}(p) = -E(\frac{d^2 \log f(X|\theta)}{d\theta^2}) = \begin{bmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{2}{\sigma^2} \end{bmatrix}$

Asymptotic normality

for $X_1, \dots, X_n$ iid $N(\mu, \sigma^2)$ RVs, $\theta = (\mu, \sigma)$,

Fisher information in each $X_i : \mathcal{I}(\theta) = \begin{bmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{2}{\sigma^2} \end{bmatrix}$

- ML estimator $\hat\theta = \begin{bmatrix} \bar X \\ \hat\sigma \end{bmatrix}$
- for large n, $\hat\theta \approx N\left(\begin{bmatrix} \mu \\ \sigma \end{bmatrix}, \begin{bmatrix} \frac{\sigma^2}{n} & 0 \\ 0 & \frac{\sigma^2}{2n} \end{bmatrix}\right)$

are expectation and variance exact?

- a random variable cannot be exactly normal! (cannot be negative)
  - $\bar X \sim N(\mu, \frac{\sigma^2}{n})$
  - $\hat\sigma \sim N(\sigma, \frac{\sigma^2}{2n})$ approximately; $E(\hat\sigma) \neq \sigma$

normal data

for $x_1, \dots, x_n$ IID $N(\mu, \sigma^2)$ RVs with large n,
ML estimates of μ and σ are $\bar x = \dots$ and $\hat\sigma = \dots$

- for approximate variance $\begin{bmatrix} \frac{\sigma^2}{n} & 0 \\ 0 & \frac{\sigma^2}{2n} \end{bmatrix}$,

  SEs of $\bar x$ and $\hat\sigma$ are estimated as $\frac{\hat\sigma}{\sqrt n}$ and $\frac{\hat\sigma}{\sqrt{2n}}$
- approximate $(1-\alpha)$-CI:

  $\mu : \left(\bar x - z_{\frac{\alpha}{2}}\frac{\hat\sigma}{\sqrt n}, \bar x + z_{\frac{\alpha}{2}}\frac{\hat\sigma}{\sqrt n}\right)$

  $\sigma : \left(\hat\sigma - z_{\frac{\alpha}{2}}\frac{\hat\sigma}{\sqrt{2n}}, \hat\sigma + z_{\frac{\alpha}{2}}\frac{\hat\sigma}{\sqrt{2n}}\right)$

Gamma distribution

$X \sim Gamma(\alpha, \lambda)$,

$f(x|\alpha, \lambda) = \frac{\lambda^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\lambda x}, \; x > 0$

$\log f(X) = \alpha\log\lambda - \log\Gamma(\alpha) + (\alpha-1)\log X - \lambda X$

let $\psi(\alpha) = \frac{d}{d\alpha}\log\Gamma(\alpha)$:

($\psi(\alpha)$ = digamma function, $\psi'(\alpha)$ = trigamma function)

- $\frac{\delta \log f(X)}{\delta\alpha} = \log\lambda - \psi(\alpha) + \log X$
- $\frac{\delta \log f(X)}{\delta\lambda} = \frac{\alpha}{\lambda} - X$
- $\frac{\delta^2 \log f(X)}{\delta\alpha^2} = -\psi'(\alpha)$
- $\frac{\delta^2 \log f(X)}{\delta\lambda^2} = -\frac{\alpha}{\lambda^2}$
- $\frac{\delta^2 \log f(X)}{\delta\alpha\delta\lambda} = \frac{\delta^2 \log f(X)}{\delta\lambda\delta\alpha} = \frac{1}{\lambda}$

$$\mathcal{I}(\alpha, \lambda) = \begin{bmatrix} \psi'(\alpha) & -\frac{1}{\lambda} \\ -\frac{1}{\lambda} & \frac{\alpha}{\lambda^2} \end{bmatrix}$$

Approximate CI with ML estimate

$\hat\theta_n$ is the ML estimator of $\theta \in \Theta \subset \mathbb{R}$ based on iid RVs $X_1, \dots, X_n$. $0 < \alpha < 1$

- for large n, approximately $\hat\theta_n \sim N(\theta, \frac{\mathcal{I}(\theta)^{-1}}{n})$.
  for $0 < \alpha < 1$,
  $$1 - \alpha \approx \Pr\left(-z_{\frac{\alpha}{2}} \leq \frac{\hat\theta_n - \theta}{\sqrt{\mathcal{I}(\theta)^{-1}/n}} \leq z_{\frac{\alpha}{2}}\right)$$
- the random interval
  $$\left(\hat\theta_n - z_{\frac{\alpha}{2}}\sqrt{\frac{\mathcal{I}(\theta)^{-1}}{n}}, \hat\theta_n + z_{\frac{\alpha}{2}}\sqrt{\frac{\mathcal{I}(\theta)^{-1}}{n}}\right)$$
  covers θ with probability $\approx 1 - \alpha$
- **MLE**: ML estimate of θ, **SE**: $\sqrt{\frac{\mathcal{I}(\theta)^{-1}}{n}}$ with θ replaced by MLE
  - approxiate $(1-\alpha) - CI$ for θ is
    $(MLE - z_{\frac{\alpha}{2}}SE, MLE + z_{\frac{\alpha}{2}}SE)$

Scope of asymptotic normality of ML estimators

- for iid normal RVs, let $\hat\sigma$ be the ML estimator of σ. then $\hat\sigma^2$ is the ML estimator of σ^2
  - both $\hat\sigma$ and $\hat\sigma^2$ are asymptotically normal
  - $\frac{1}{\hat\sigma}$ is also asymptotically normal
- let $\hat\theta^n$ be the ML estimator of θ. For strictly increasing or strictly decreasing $h : \Theta \to \mathbb{R}$, $h(\hat\theta^n)$ is the ML estimator of $h(\theta)$.
  - for large n, $h(\hat\theta^n)$ is approximately normal

population mean vs parameter

for n random draws with replacement from a population with mean μ and variance σ^2,

Estimator	E	var	Distribution
random sample mean, $\hat\mu$	μ	$\frac{\sigma^2}{n}$	$\approx$ normal
ML estimator, $\hat\theta_n$	$\approx\theta$	$\approx \frac{\mathcal{I}(\theta)^{-1}}{n}$	$\approx$ normal

$\hat\theta_n$ is not normal (but may approach normal for large n)

summary

> let X have density $f(x|\theta), \theta \in \Theta \subset \mathbb{R}^k$.
> The **Fisher information** at θ in X is the $k \times k$ matrix
> $-E\left[\frac{d^2 \log f(X|\theta)}{d\theta^2}\right]$.
>
> let $\hat\theta_n$ be the ML estimator of θ based on iid RVs $X_1, \dots, X_n$ with density $f(x|\theta)$.
> For large n, the distribution of $\hat\theta_n$ is approximately
> $N\left(\theta, \frac{\mathcal{I}(\theta)^{-1}}{n}\right)$

$\Rightarrow$ SE can be estimated without monte carlo
$\Rightarrow$ accurate CIs are available
skipped: Fisher information in IID samples; binomial fisher information, MLE; HWE trinomial fisher information
$E(\frac{d\log f(X|\lambda)}{d\lambda}) = 0$

09. HYPOTHESIS TESTING

let $x_1, \dots, x_n$ be realisations of IID $N(\mu, \sigma^2)$ RVs $X_1, \dots, X_n$ where μ is a parameter and σ is known.

null hypothesis, $H_0 : \mu = \mu_0$
alternative hypothesis, $H_1 : \mu = \mu_1$

It is believed that $\mu = \mu_0$, but it might be μ_1. 2 methods to test if H_0 should be rejected in favour of H_1 using $\bar x$:

- if $\bar x$ falls inside the **rejection region**, we reject H_0
  - based on a choice of α (type I error)
- **P value** $\to$ the probability that $\bar X$ is more extreme than $\bar x$, assuming H_0 is true. (if small, doubt H_0)
  - based on an observed test statistic

if σ is unknown or $x_1, \dots, x_n \not\sim N(\mu, \sigma^2)$, we can use CLT

Rejection region

$x_1, \dots, x_n$ are from IID $N(\mu, \sigma^2)$ RVs, with σ known

One-tailed test

$$H_0 : \mu = \mu_0, \quad H_1 : \mu = \mu_1 > \mu_0$$

under H_0,
$$\bar X \sim N(\mu_0, \frac{\sigma^2}{n}), \quad \frac{\bar X - \mu_0}{\sigma/\sqrt n} \sim N(0, 1)$$

$$\alpha = P_{H_0}(\bar X > \mu_0 + c) = \Pr(Z > \frac{c}{\sigma/\sqrt n}) \quad \Rightarrow c = z_\alpha$$

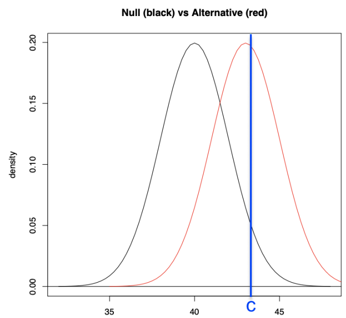
- reject H_0 if $\bar x - \mu_0 > c$ (for some $c > 0$)
  - $\bar x$ is the **test statistic**
  - interval $(\mu_0 + z_\alpha\frac{\sigma}{\sqrt n}, \infty)$ is the **rejection region**
    - for a test of **size** α, $c = z_\alpha\frac{\sigma}{\sqrt n}$

Hypothesis	$\bar x < \mu_0 + c$	$\bar x > \mu_0 + c$
$H_0 : \mu = \mu_0$	✓ not reject H_0	×(I) reject H_0
$H_1 : \mu = \mu_1$	×(II) not reject H_0	✓ reject H_0

- type I error: rejecting H_0 when it is true
- type II error: not rejecting H_0 when it is false

Size and power

- **size** of a test $\to$ probability of a Type I error
  - $\alpha := P_{H_0}(\bar X > \mu_0 + c)$
  - aka **level**
- **power** of a test $\to 1-$ probability of a Type II error
  - $\beta := P_{H_1}(\bar X > \mu_0 + c) \quad \Rightarrow$ power $= 1 - \beta$
  - as $n \to \infty$, power $\to 1$
    - increasing power of rejecting H_0
- α and β are both about the same event ($\bar X$ is in the rejection region), but calculated under different hypotheses (H_0, H_1)
- $\uparrow c : \; \downarrow\alpha, \downarrow\beta$ ($\downarrow$ type I error, $\uparrow$ type II error)
- commonly $\alpha = 0.05$
  - keep α small since H_0 is the default hypothesis



Two-tailed test

$x_1, \dots, x_n$ are from iid $N(\mu, \sigma^2)$ RVs, σ known

$$H_0 : \mu = \mu_0, \quad H_0 : \mu = \mu_0, H_1 : \mu = \mu_1 \neq \mu_0$$

- reject H_0 if $|\bar{x} - \mu_0| > c$, for some $c > 0$
  - **rejection region:** $(-\infty, \mu_0 - c)$ and $(\mu_0 + c, \infty)$
- $\alpha = P_{H_0}(|\bar{X} - \mu_0| > c) = \Pr\left(|Z| > \frac{c}{\sigma/\sqrt{n}}\right)$
  $= 2\Pr\left(Z > \frac{c}{\sigma/\sqrt{n}}\right)$
- $c = z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$
  - **rejection region:** $(-\infty, \mu_0 - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}) \wedge (\mu_0 + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \infty)$

Composite hypothesis

- **simple** hypothesis $\rightarrow$ specify a single value ($H_0 : \mu = \mu_0, H_1 : \mu = \mu_1$)
- **composite** hypothesis $\rightarrow$ range of values
  - one-tailed test: $H_0 : \mu = \mu_0, \; H_1 : \mu > \mu_0$
    - rejection region: $(\mu_0 + z_\alpha \frac{\sigma}{\sqrt{n}}, \; \infty)$
      $\Rightarrow$ no change since it doesn't involve μ_1
  - two-tailed test: $H_0 : \mu = \mu_0, \; H_1 : \mu \neq \mu_0$
    - rejection region:
      $(-\infty, \mu_0 - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}) \wedge (\mu_0 + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \infty)$
      $\Rightarrow$ no change since it doesn't involve μ_1
    - if $\bar{x}$ falls *outside* the rejection region, i.e.
      $\mu_0 - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \leq \bar{x} \leq \mu_0 + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$
      - then H_0 is NOT rejected at level α
      - μ_0 lies in the $(1-\alpha)$-CI for μ
  - as $n \to \infty$, power $\to 1$

Hypothesis testing and CI

the $(1-\alpha)$-CI for μ, $\left(\bar{x} - z_{\frac{\alpha}{2}} \frac{\hat{\sigma}}{\sqrt{n}}, \bar{x} + z_{\frac{\alpha}{2}} \frac{\hat{\sigma}}{\sqrt{n}}\right)$ consists of the values μ_0 for which the test $H_0 : \mu = \mu_0, \; H_1 : \mu \neq \mu_0$ is not rejected at level α.

P-value

- **P-value** $\rightarrow$ the probability under H_0 that the random test statistic is more extreme than the observed test statistic
  - small p-value = more "extreme" (more doubt)
- reject H_0 at level $\alpha \iff P < \alpha$
- generally, P-value for two-tailed test is double that of one-tailed test

formulae for P-value

$H_1 : \mu > \mu_0$

$$P = P_{H_0}(\bar{X} > \bar{x}) = \Pr\left(Z > \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}\right)$$

$H_1 : \mu < \mu_0$

$$P = P_{H_0}(\bar{X} < \bar{x}) = \Pr\left(Z < \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}\right)$$

$H_1 : \mu \neq \mu_0$

$$P = P_{H_0}(|\bar{X} - \mu_0| > |\bar{x} - \mu_0|) = \Pr\left(|Z| > \frac{|\bar{x} - \mu_0|}{\sigma/\sqrt{n}}\right)$$

10. GOODNESS-OF-FIT

- **likelihood ratio** (LR) test $\rightarrow$ based on the ratio of likelihoods
  - P-value can be approximated using χ^2 distribution for a large sample size

multinomial

let $X \sim Trinomial(n, \mathbf{p})$. by HWE, $\mathbf{p}$ is a function of θ as follows: $p_1 = (1-\theta)^2, \; p_2 = 2\theta(1-\theta), \; p_3 = \theta^2$

let L_1 and L_0 be the maximum likelihood value for the general model ($Trinomial(n, \mathbf{p})$) and the HWE.

- $L_1 \geq L_0$ (L_0 is the maximum over a subset of L_1)
  - general trinomial
    - likelihood, $L(\mathbf{p}) = p_1^{x_1} p_2^{x_2} p_3^{x_3}$
    - ML estimate of $\mathbf{p}$ is $\frac{\mathbf{x}}{n}$
    - $\log L_1 = x_1 \log(\frac{x_1}{n}) + x_2 \log(\frac{x_2}{n}) + x_3 \log(\frac{x_3}{n})$
  - HWE:
    - likelihood, $L(\theta) = p_1(\theta)^{x_1} p_2(\theta)^{x_2} p_3(\theta)^{x_3}$
    - ML estimate of θ is $\frac{x_2 + 2x_3}{2n}$
- larger $L_1/L_0 \Rightarrow$ poorer fit for HWE

LR test

- null hypothesis: HWE holds
  $H_0 : p_1 = (1-\theta)^2, \; p_2 = 2\theta(1-\theta), \; p_3 = \theta^2$
- LR test statistic: $2\log\left(\frac{L_1}{L_0}\right) = 2(\log L_1 - \log L_0)$
- degree of freedom = difference in the number of parameters between the models
  - general model has 2 params, HWE has 1 param
- P-value $= \Pr\left(\chi_1^2 > 2\log(\frac{L_1}{L_0})\right)$

Nested models

the set of all $Trinomial(n, \mathbf{p})$ distributions can be represented by

$$\Omega_1 = \left\{(p_1, p_2, p_3) : p_i > 0, \; \textstyle\sum_{i=1}^3 p_i = 1\right\}$$

which has dimension 2 $\quad (\dim \Omega_1 = 2)$

- by HWE, $\mathbf{p}$ is in the subset
  $\Omega_0 = \{((1-\theta)^2, 2\theta(1-\theta), \theta^2) : 0 < \theta < 1\}$ ($\dim \Omega_0 = 1$)
- Ω_0 is **nested** in Ω_1
- measure goodness-of-fit of HWE by testing $H_0 : \mathbf{p} \in \Omega_0$

General Multinomial LR test

let $(X_1, \dots, X_k) \sim Multinomial(n, \mathbf{p})$. then $\mathbf{p} \in \Omega_1$, the set of all positive probability vectors of length k.

to test if $\mathbf{p}$ is in a subspace
$\Omega_0 = \left\{(p_1(\theta), \dots, p_k(\theta)) : \theta \in \Theta \subset \mathbb{R}^h\right\}$
with $\dim \Omega_0 < \dim \Omega_1 = k - 1$

let L_j be the maximum likelihood value under Ω_j. To test $H_0 : \mathbf{p} \in \Omega_0$, we use the **LR statistic**,

$$G = 2\log(\frac{L_1}{L_0})$$

- for Ω_1: $\log L_1 = \sum_{i=1}^k X_i \log(\frac{X_i}{n})$
- for Ω_0: $\log L_0 = \sum_{i=1}^k X_i \log p_i(\hat{\theta})$

$$G = 2\sum_{i=1}^k X_i \log\left(\frac{X_i}{np_i(\hat{\theta})}\right)$$

given data $(x_1, \dots, x_n)$, let g be a realisation of G. P-value $P_{H_0}(G > g)$ is approximately $\Pr(\chi^2_{k-1-\dim \Omega_0} > g)$ for large n.

- to compute g, replace
  - X_i with *observed count* x_i
  - $np_i(\hat{\theta})$ with *expected count*, calculated using ML estimate of θ

Test of independence

for a population with attributes q and r, let p_{ij} be the population proportion of people with $q = q_i$ and $r = r_j$. for any i, j, $p_{ij} = q_i \times r_j$.

- let $(X_{ij}, 1 \leq i \leq I, 1 \leq j \leq J) \sim Multinomial(n, \mathbf{p})$. $\mathbf{p} \in \Omega_1$, where $\dim \Omega_1 = IJ - 1 = k - 1$.
- H_0 : the two categories q, r are independent
  - if q, r are independent, then $\exists$ positive numbers $\sum_{i=1}^I q_i = \sum_{j=1}^J r_j = 1$ such that $p_{ij} = q_i \times r_j$, $1 \leq i \leq I, 1 \leq j \leq J$
- $\dim \Omega_0 = (I-1) + (J-1) = I + J - 2$
- $\dim \Omega_1 - \dim \Omega_0 = (I-1)(J-1)$
- under independence (H_0), for large n, approximately $G \sim \chi^2_{(I-1)(J-1)}$

G statistic

for any i, let $X_{i+} = \sum_{j=1}^J X_{ij}$.
for any j, let $X_{+j} = \sum_{i=1}^I X_{ij}$.

- $\Omega_1 : \log L_1 = \sum_{ij} X_{ij} \log\left(\frac{X_{ij}}{n}\right)$
- Ω_0 :
  $\log L_0 = \sum_i X_{i+} \log\left(\frac{X_{i+}}{n}\right) + \sum_{+j} X_{+j} \log\left(\frac{X_{+j}}{n}\right)$
- $G = 2(\log L_1 - \log L_0) = 2\sum_{ij} X_{ij} \log\left(\frac{X_{ij}}{X_{i+}X_{+j}/n}\right)$
- the data x_{ij} are the *observed counts*
- the data $x_{i+}x_{+j}/n$ are the *expected counts*
- P-value $= \Pr\left(\chi^2_{(I-1)(J-1)} > g\right)$

General LR test

we have n iid RVs with density defined by $\theta \in \Omega_1$ of dimension k_1; nested in Ω_1 is a smaller model Ω_0 of dimension k_0.

$$H_0 : \theta \in \Omega_0 \qquad H_1 : \theta \in \Omega_1 \backslash \Omega_0$$

to test $H_0 : \theta \in \Omega_0$, we use LR statistic

$$G = 2\log\left(\frac{L_1}{L_0}\right)$$

where L_j is the maximum likelihood value over Ω_j.

for large n, the P-value can be approximately computed, because:

if $\theta \in \Omega_0$, as $n \to \infty$, the distribution of G converges to $\chi^2_{k_1 - k_0}$

Normal LR test

$x_1, \dots, x_n$ are form iid $N(\mu, \sigma^2)$ RVs. to test $H_0 : \mu = 0$:

σ	Ω_1	k_1	Ω_0	k_0
known	$\mathbb{R}$	1	$\{0\}$	0
unknown	$\mathbb{R} \times \mathbb{R}_+$	2	$\{0\} \times \mathbb{R}_+$	1

under H_0, for large n, approximately $G \sim \chi_1^2$

- **case 1**: σ known
  - $\Omega_1 : \log L_1 = -\frac{n\hat{\sigma}^2}{2\sigma^2}$
  - $\Omega_0 : \log L_0 = -\frac{n\hat{\mu}^2}{2\sigma^2}$
  - $G = 2(\log L_1 - \log L_0) = \frac{n\bar{X}^2}{\sigma^2}$
    - if H_0 holds ($\mu = 0$), then $\bar{X} \sim N(0, \frac{\sigma^2}{n})$. for any n, $G \sim \chi_1^2$ exactly.
- **case 2**: σ unknown
  - $\Omega_1 : \log L_1 = -\frac{n}{2}\log\hat{\sigma}^2 - \frac{n}{2}$
  - $\Omega_0 : \log L_0 = -\frac{n}{2}\log\hat{\mu}_2 - \frac{n}{2}$
  - $G = 2(\log L_1 - \log L_0) = n\log(\frac{\hat{\mu}_2}{\hat{\sigma}^2})$
  - if H_0 holds ($\mu = 0$), for large n, $G \sim \chi_1^2$ approximately

Summary

- LR test applies when the investigator wants to know the goodness-of-fit of a model relative to a larger model, of dimensions $k_0 < k_1$.
- test statistic, $G = 2\log\left(\frac{L_1}{L_0}\right)$
  - L_0, L_1 are the maximum likelihood value under the small and large models
- if n is large, the P-value $\Pr(G > g)$ (computed provided H_0 is true) can be approximated by a $\chi^2_{k_1 - k_0}$ distribution