Maximum Likelihood
Large Sample Theory

MIT 18.443

Dr. Kempthorne

Spring 2015

Outline

1. Large Sample Theory of Maximum Likelihood Estimates
   - Asymptotic Distribution of MLEs
   - Confidence Intervals Based on MLEs

Asymptotic Results: Overview

Asymptotic Framework

- Data Model : $\mathbf{X}_n = (X_1, X_2, \dots, X_n)$ i.i.d. sample with pdf/pmf $f(x_1, \dots, x_n \mid \theta) = \prod_{i=1}^n f(x_i \mid \theta)$
- Data Realization: $\mathbf{X}_n = \mathbf{x}_n = (x_1, \dots, x_n)$
- **Likelihood of** θ (given $\mathbf{x}_n$):
$$lik(\theta) = f(x_1, \dots, x_n \mid \theta)$$
- $\hat{\theta}_n$: **MLE** of θ given $\mathbf{x}_n = (x_1, \dots, x_n)$
- $\{\hat{\theta}_n, n \to \infty\}$: sequence of MLEs indexed by sample size n

Results:

- **Consistency:** $\hat{\theta}_n \xrightarrow{\mathcal{L}} \theta$
- **Asymptotic Variance:** $\sigma_{\hat{\theta}_n} = \sqrt{Var(\hat{\theta}_n)} \xrightarrow{\mathcal{L}} \sqrt{\kappa(\theta)/n}$ where $\kappa(\theta)$ is an explicit function of the pdf/pmf $f(\cdot \mid \theta)$.
- **Limiting Distribution:** $\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{\mathcal{L}} N(0, \kappa(\theta))$.

Theorems for Asymptotic Results

Setting:

- $x_1, \ldots, x_n$ a realization of an i.i.d. sample from distribution with density/pmf $f(x \mid \theta)$.
- $\ell(\theta) = \sum_{i=1}^{n} \ln f(x_i \mid \theta)$
- θ_0: true value of θ
- $\hat{\theta}_n$: the MLE

Theorem 8.5.2.A Under appropriate smoothness conditions on f, the MLE $\hat{\theta}_n$ is consistent, i.e., for any true value θ_0, for every $\epsilon > 0$,

$$P(|\hat{\theta}_n - \theta_0| > \epsilon) \longrightarrow 0.$$

Proof:

- Weak Law of Large Numbers (WLLN)

  $$\frac{1}{n}\ell(\theta) \xrightarrow{\mathcal{L}} E[\log f(x \mid \theta) \mid \theta_0] = \int \log[f(x \mid \theta)] f(x \mid \theta_0) dx$$

  (**Note!!** statement holds for every θ given any value of θ_0.)

Theorem 8.5.2A (continued)

Proof (continued):

- The MLE $\hat{\theta}_n$ maximizes $\frac{1}{n}\ell(\theta)$
- Since $\frac{1}{n}\ell(\theta) \longrightarrow E[\log f(x \mid \theta) \mid \theta_0]$,

  $\hat{\theta}_n$ is close to θ^* maximizing $E[\log f(x \mid \theta) \mid \theta_0]$
- Under smoothness conditions on $f(x \mid \theta)$,

  θ^* maximizes $E[\log f(x \mid \theta) \mid \theta_0]$

  if θ^* solves

  $$\frac{d}{d\theta}\left(E[\log f(x \mid \theta) \mid \theta_0]\right) = 0$$
- Claim: $\theta^* = \theta_0$:

$$\begin{array}{rcl}
\frac{d}{d\theta}\left(E[\log f(x \mid \theta) \mid \theta_0]\right) & = & \frac{d}{d\theta}\int \log[f(x \mid \theta)] f(x \mid \theta_0)dx \\
 & = & \int \left(\frac{d}{d\theta}\log[f(x \mid \theta)]\right) f(x \mid \theta_0)dx \\
 & = & \int \left(\frac{\frac{d}{d\theta}[f(x|\theta)]}{f(x|\theta)}\right) f(x \mid \theta_0)dx \\
(\text{at } \theta = \theta_0) & = & [\int \frac{d}{d\theta}[f(x \mid \theta)]dx]|_{\theta=\theta_0} \\
 & = & [\frac{d}{d\theta}\int f(x \mid \theta)dx]|_{\theta=\theta_0} = \frac{d}{d\theta}(1) \equiv 0.
\end{array}$$

Theorems for Asymptotic Results

Theorem 8.5.B Under smoothness conditions on $f(x \mid \theta)$,

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{\mathcal{L}} N(0, 1/\mathrm{I}(\theta_0))$$

where $\mathrm{I}(\theta) = E\left(-\left[\frac{\partial^2}{\partial\theta^2}\log f(X \mid \theta)\right] \mid \theta\right)$

Proof: Using the Taylor approximation to $\ell'(\theta)$, centered at θ_0 consider the following development:

- $$0 = \ell'(\hat{\theta}) \approx \ell'(\theta_0) + (\hat{\theta} - \theta)\ell''(\theta_0)$$
  $$\Longrightarrow (\hat{\theta} - \theta) \approx \frac{\ell'(\theta_0)}{-\ell''(\theta_0)}$$
  $$\Longrightarrow \sqrt{n}(\hat{\theta} - \theta) \approx \frac{\sqrt{n}[\frac{1}{n}\ell'(\theta_0)]}{\frac{1}{n}[-\ell''(\theta_0)]}$$
- By the CLT $\sqrt{n}[\frac{1}{n}\ell'(\theta_0)] \xrightarrow{\mathcal{L}} N(0, \mathrm{I}(\theta_0))$ (*Lemma A*)
- By the WLLN $\frac{1}{n}[-\ell''(\theta_0)] \xrightarrow{\mathcal{L}} \mathrm{I}(\theta_0)$

Thus: $\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{\mathcal{L}} (1/\mathrm{I}(\theta_0))^2 N(0, \mathrm{I}(\theta_0)) = N(0, 1/\mathrm{I}(\theta_0))$

Theorems for Asymptotic Results

Lemma A (extended). For the distribution with pdf/pmf $f(x \mid \theta)$ define

- The **Score Function**

$$U(X;\theta) = \frac{\partial}{\partial\theta} log f(X \mid \theta)$$

- The **(Fisher) Information** of the distribution

$$\mathrm{I}(\theta) = E\left(-\left[\frac{\partial^2}{\partial\theta^2}\log f(X \mid \theta)\right] \mid \theta\right)$$

Then under sufficient smoothness conditions on $f(x \mid \theta)$

$$\begin{array}{lrcl} \textbf{(a).} & E[U(X;\theta) \mid \theta] & = & 0 \\ \textbf{(b).} & Var[U(X;\theta) \mid \theta] & = & \mathrm{I}(\theta) \\ & & = & E\left([U(X;\theta)]^2 \mid \theta]\right) \end{array}$$

Proof:

- Differentiate $\int f(x \mid \theta)dx = 1$ with respect to θ two times.
- Interchange the order of differentiation and integration.
- (a) follows from the first derivative.
- (b) follows from the second derivative.

Theorems for Asymptotic Results

Qualifications/Extensions

- Results require true θ_0 to lie in interior of the parameter space.
- Results require that $\{x : f(x \mid \theta) > 0\}$ not vary with θ.
- Results extend to multi-dimensional θ

  Vector-valued **Score Function**

  Matrix-valued **(Fisher) Information**

Outline

1. Large Sample Theory of Maximum Likelihood Estimates
   - Asymptotic Distribution of MLEs
   - Confidence Intervals Based on MLEs

Confidence Intervals

Confidence Interval for a Normal Mean

- $X_1, X_2, \ldots, X_n$ i.i.d. $N(\mu, \sigma_0^2)$, unknown mean μ (known σ_0^2)
- Parameter Estimate: $\overline{X}$ (sample mean)

$$\begin{aligned} \overline{X} &= \frac{1}{n}\sum_{i=1}^{n} X_i \\ E[\overline{X}] &= \mu \\ Var[\overline{X}] &= \sigma_{\overline{X}}^2 = \sigma_0^2/n \end{aligned}$$

- A 95% confidence interval for μ is a random interval, calculated from the data, that contains μ with probability 0.95, no matter what the value of the true μ.

Confidence Interval for a Normal Mean

- Sampling distribution of $\overline{X}$

$$\begin{aligned} \overline{X} &\sim N(\mu, \sigma^2_{\overline{X}}) = N(\mu, \sigma_0^2/n) \\ \Longrightarrow \quad Z = \frac{\overline{X} - \mu}{\sigma_{\overline{X}}} &\sim N(0,1) \text{ (standard normal)} \end{aligned}$$

- For any $0 < \alpha < 1$, (e.g. $\alpha = 0.05$)
  - Define $z(\alpha/2) : P(Z > z(\alpha/2)) = \alpha/2$
  - By symmetry of the standard normal distribution

    $$P\left(-z(\alpha/2) < Z < +z(\alpha/2)\right) = 1 - \alpha$$

    i.e.,

    $$P\left(-z(\alpha/2) < \frac{\overline{X} - \mu}{\sigma_{\overline{X}}} < +z(\alpha/2)\right) = 1 - \alpha$$

- Re-express interval of Z as interval of μ:

$$P\left(\overline{X} - z(\alpha/2)\sigma_{\overline{X}} < \mu < \overline{X} + z(\alpha/2)\sigma_{\overline{X}}\right) = 1 - \alpha$$

- The interval given by $[\overline{X} \pm z(\alpha/2)\ \sigma_{\overline{X}}]$ is the
  $100(1-\alpha)\%$ **confidence interval** for μ

Confidence Interval for a Normal Mean

Important Properties/Qualifications

- The confidence interval is random.
- The parameter μ is not random.
- $100(1-\alpha)\%$: the confidence-level of the confidence interval is the probability that the random interval contains the fixed parameter μ (the **"coverage probability"** of the confidence interval)
- Given a data realization of $(X_1, \ldots, X_n) = (x_1, \ldots, x_n)$ μ is either inside the confidence interval or not.
- The confidence level scales the reliability of a sequence of confidence intervals constructed in this way.
- The confidence interval quantifies the uncertainty of the parameter estimate.

Confidence Intervals for Normal Distribution Parameters

Normal Distribution with unknown mean (μ) and variance σ^2.

- $X_1, \ldots, X_n$ i.i.d. $N(\mu, \sigma^2)$, with MLEs:

$$\begin{aligned} \hat{\mu} &= \overline{X} \\ \hat{\sigma}^2 &= \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2 \end{aligned}$$

- Confidence interval for μ based on the T-Statistic

$$T = \frac{\overline{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}$$

  where t_{n-1} is Student's t distribution with $(n-1)$ degrees of freedom and $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$.

- Define $t_{n-1}(\alpha/2) : P(t_{n-1} > t_{n-1}(\alpha/2)) = \alpha/2$

  By symmetry of Student's t distribution

$$P\left[-t_{n-1}(\alpha/2) < T < +t_{n-1}(\alpha/2)\right] = 1 - \alpha$$

  i.e., $$P\left[-t_{n-1}(\alpha/2) < \frac{\overline{X} - \mu}{S\sqrt{n}} < +t_{n-1}(\alpha/2)\right] = 1 - \alpha$$

Confidence Interval For Normal Mean

- Re-express interval of T as interval of μ:

$$P\left[-t_{n-1} < \frac{\overline{X} - \mu}{S\sqrt{n}} < +t_{n-1}\right] = 1 - \alpha$$

$\Longrightarrow$

$$P\left[\overline{X} - t_{n-1}(\alpha/2)S/\sqrt{n} < \mu < \overline{X} + t_{n-1}(\alpha/2)S/\sqrt{n}\right] = 1 - \alpha$$

- The interval given by $[\overline{X} \pm t_{n-1}(\alpha/2)\ S/\sqrt{n}]$ is the

  $100(1-\alpha)\%$ **confidence interval** for μ

- Properties of confidence intervals for μ
  - Center is $\overline{X} = \hat{\mu}_{MLE}$
  - Width proportional to $\hat{\sigma}_{\hat{\mu}_{MLE}} = S/\sqrt{n}$ (random!)

Confidence Interval For Normal Variance

Normal Distribution with unknown mean (μ) and variance σ^2.

- $X_1, \ldots, X_n$ i.i.d. $N(\mu, \sigma^2)$, with MLEs:
$$\begin{aligned} \hat{\mu} &= \overline{X} \\ \hat{\sigma}^2 &= \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2 \end{aligned}$$
- Confidence interval for σ^2 based on the sampling distribution of the MLE $\hat{\sigma}^2$.
$$\Omega = \frac{n\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{n-1}.$$
where χ^2_{n-1} is Chi-squared distribution with $(n-1)$ d.f.
- Define $\chi^2_{n-1}(\alpha^*) : P(\chi^2_{n-1} > \chi^2_{n-1}(\alpha^*)) = \alpha^*$
Using $\alpha^* = \alpha/2$ and $\alpha^* = (1-\alpha/2)$,
$$P\left(+\chi^2_{n-1}(1-\alpha/2) < \Omega < +\chi^2_{n-1}(\alpha/2)\right) = 1-\alpha$$
i.e., $$P\left(+\chi^2_{n-1}(1-\alpha/2) < \frac{n\hat{\sigma}^2}{\sigma^2} < +\chi^2_{n-1}(\alpha/2)\right) = 1-\alpha$$

Confidence Interval for Normal Variance

- Re-express interval of Ω as interval of σ^2:

$$P[+\chi^2_{n-1}(1-\alpha/2) < \Omega < +\chi^2_{n-1}(\alpha/2)] = 1-\alpha$$
$$P[+\chi^2_{n-1}(1-\alpha/2) < \frac{n\hat{\sigma}^2}{\sigma^2} < +\chi^2_{n-1}(\alpha/2)] = 1-\alpha$$
$$P[\frac{n\hat{\sigma}^2}{\chi^2_{n-1}(\alpha/2)} < \sigma^2 < \frac{n\hat{\sigma}^2}{\chi^2_{n-1}(1-\alpha/2)}] = 1-\alpha$$

- The $100(1-\alpha)\%$ **confidence interval** for σ^2 is given by

$$[\frac{n\hat{\sigma}^2}{\chi^2_{n-1}(\alpha/2)} < \sigma^2 < \frac{n\hat{\sigma}^2}{\chi^2_{n-1}(1-\alpha/2)}]$$

- Properties of confidence interval for σ^2
  - Asymmetrical about the MLE $\hat{\sigma}^2$.
  - Width proportional to $\hat{\sigma}^2$ (random!)
  - $100(1-\alpha)\%$ **confidence interval** for σ immediate:

$$[\hat{\sigma}(\sqrt{\frac{n}{\chi^2_{n-1}(\alpha/2)}}) < \sigma < \hat{\sigma}(\sqrt{\frac{n}{\chi^2_{n-1}(1-\alpha/2)}})]$$

Confidence Intervals For Normal Distribution Parameters

Important Features of Normal Distribution Case

- Required use of exact sampling distributions of the MLEs
  $\hat{\mu}$ and $\hat{\sigma}^2$.
- Construction of each confidence interval based on **pivotal quantity**: a function of the data and the parameters whose distribution does not involve any unknown parameters.
- Examples of **pivotals**

$$T = \frac{\overline{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}$$

$$\Omega = \frac{n\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{n-1}$$

Confidence Intervals Based On Large Sample Theory

Asymptotic Framework (Re-cap)

- Data Model : $\mathbf{X}_n = (X_1, X_2, \ldots, X_n)$ i.i.d. sample with pdf/pmf $f(x_1, \ldots, x_n \mid \theta) = \prod_{i=1}^n f(x_i \mid \theta)$
- **Likelihood of** θ given $\mathbf{X}_n = \mathbf{x}_n = (x_1, \ldots, x_n)$:
$$lik(\theta) = f(x_1, \ldots, x_n \mid \theta)$$
- $\hat{\theta}_n = \hat{\theta}_n(\mathbf{x}_n)$: **MLE** of θ given $\mathbf{x}_n = (x_1, \ldots, x_n)$
- $\{\hat{\theta}_n, n \to \infty\}$: sequence of MLEs indexed by sample size n

Results (subject to sufficient smoothness conditions on f)

- **Consistency:** $\hat{\theta}_n \xrightarrow{\mathcal{L}} \theta$
- **Asymptotic Variance:** $\sigma_{\hat{\theta}_n} = \sqrt{Var(\hat{\theta}_n)} \xrightarrow{\mathcal{L}} \sqrt{\dfrac{1}{n\mathrm{I}(\theta)}}$

  where $\mathrm{I}(\theta) = E[(\frac{d}{d\theta}[\log f(x \mid \theta)])^2] = E[-\left(\frac{d^2}{d\theta^2}[\log f(x \mid \theta)]\right)]$
- **Limiting Distribution:** $\sqrt{n\mathrm{I}(\theta)}(\hat{\theta}_n - \theta) \xrightarrow{\mathcal{L}} N(0, 1)$.

Confidence Intervals Based on Large Sample Theory

Large-Sample Confidence Interval

- Exploit the limiting **pivotal quantity**

$$\mathcal{Z}_n = \sqrt{n\mathrm{I}(\theta)}(\hat{\theta}_n - \theta) \xrightarrow{\mathcal{L}} N(0,1).$$

- I.e.

$$\begin{array}{rccccccc} & P(-z(\alpha/2) & < & \mathcal{Z}_n & < & +z(\alpha/2)) & \approx & 1-\alpha \\ \iff & P(-z(\alpha/2) & < & \sqrt{n\mathrm{I}(\theta)}(\hat{\theta}_n - \theta) & < & +z(\alpha/2)) & \approx & 1-\alpha \\ \iff & P(-z(\alpha/2) & < & \sqrt{n\mathrm{I}(\hat{\theta}_n)}(\hat{\theta}_n - \theta) & < & +z(\alpha/2)) & \approx & 1-\alpha \end{array}$$

  Note (!): $\mathrm{I}(\hat{\theta}_n)$ substituted for $\mathrm{I}(\theta)$

- Re-express interval of $\mathcal{Z}_n$ as interval of θ:

$$P(\hat{\theta}_n - z(\alpha/2)\frac{1}{\sqrt{n\mathrm{I}(\hat{\theta})}} < \theta < \hat{\theta}_n + z(\alpha/2)\frac{1}{\sqrt{n\mathrm{I}(\hat{\theta})}}) \approx 1-\alpha$$

- The interval given by $[\hat{\theta}_n \pm z(\alpha/2)\frac{1}{\sqrt{n\mathrm{I}(\hat{\theta})}}]$ is the

  $100(1-\alpha)\%$ **confidence interval** (large sample) for θ

Large-Sample Confidence Intervals

Example 8.5.B. Poisson Distribution

- $X_1, \ldots, X_n$ i.i.d. $Poisson(\lambda)$
- $f(x \mid \lambda) = \frac{\lambda^x}{x!} e^{-\lambda}$
- $\ell(\lambda) = \sum_{i=1}^{n} [x_i \ln(\lambda) - \lambda - \ln(x!)]$

MLE $\hat{\lambda}$

- $\hat{\lambda}$ solves: $\dfrac{d\ell(\lambda)}{d\lambda} = \sum_{i=1}^{n} [\dfrac{x_i}{\lambda} - 1] = 0$; $\hat{\lambda} = \overline{X}$.
- $\mathrm{I}(\lambda) = E[-\dfrac{d^2}{d\lambda^2} \log f(x \mid \lambda)] = E[\dfrac{X_i}{\lambda^2}] = \dfrac{1}{\lambda}$
- $\mathcal{Z}_n = \sqrt{n\mathrm{I}(\hat{\lambda})}(\hat{\lambda} - \lambda) = \dfrac{\hat{\lambda} - \lambda}{\sqrt{\hat{\lambda}/n}} \xrightarrow{\mathcal{L}} N(0, 1)$

Large-Sample Confidence Interval for λ

- Approximate $100(1-\alpha)\%$ confidence interval for λ

$$[\hat{\lambda} \pm \hat{\sigma}_{\hat{\lambda}}] = [\overline{X} \pm z(\alpha/2)\sqrt{\frac{\overline{X}}{n}}]$$

Confidence Interval for Poisson Parameter

Deaths By Horse Kick in Prussian Army (Bortkiewicz, 1898)

- Annual Counts of fatalities in 10 corps of Prussian cavalry over a period of 20 years.
- $n = 200$ corps-years worth of data.

Annual Fatalities	Observed
0	109
1	65
2	22
3	3
4	1

- Model $X_1, \ldots, X_{200}$ as i.i.d. $Poisson(\lambda)$.
- $\hat{\lambda}_{MLE} = \overline{X} = \frac{122}{200} = 0.61$
- $\hat{\sigma}_{\hat{\lambda}_{MLE}} = \sqrt{\hat{\lambda}_{MLE}/n} = .0552$
- For an 95% confidence interval, $z(\alpha/2) = 1.96$ giving

$$0.61 \pm (1.96)(.0552) = [.5018, .7182]$$

Confidence Interval for Multinomial Parameter

Definition: Multinomial Distribution

- $W_1, \ldots, W_n$ are iid $Multinomial(1, probs = (p_1, \ldots, p_m))$ r.v.s
- The sample space of each W_i is $\mathcal{W} = \{1, 2, \ldots, m\}$, a set of m distinct outcomes.
- $P(W_i = k) = p_k$, $k = 1, 2, \ldots, m$.

Define Count Statistics from Multinomial Sample:

- $X_k = \sum_{i=1}^n 1(W_i = k)$, (sum of indicators of outcome k), $k = 1, \ldots, m$
- $\mathbf{X} = (X_1, \ldots, X_m)$ follows a multinomial distribution
$$f(x_1, \ldots, x_n \mid p_1, \ldots, p_m) = \frac{n!}{\prod_{j=1}^m x_i!} \prod_{j=1}^m p_j^{x_j}$$
where $(p_1, \ldots, p_m)$ is the vector of cell probabilities with $\sum_{i=1}^m p_i = 1$ and $n = \sum_{j=1}^m x_i$ is the total count.

Note: for $m = 2$, the W_i are $Bernoulli(p_1)$
X_1 is $Binomial(n, p_1)$ and $X_2 \equiv n - X_1$

MLEs of Multinomial Parameter

Maximum Likelihood Estimation for Multinomial

- Likelihood function of counts
$$\begin{aligned} lik(p_1,\ldots,p_m) &= log[f(x_1,\ldots,x_m \mid p_1,\ldots,p_m)] \\ &= log(n!) - \sum_{j=1}^m log(x_j!) + \sum_{j=1}^m x_j log(p_j) \end{aligned}$$
- Note: Likelihood function of Multinomial Sample $w_1, \ldots, , w_n$
$$\begin{aligned} lik^*(p_1,\ldots,p_m) &= log[f(w_1,\ldots,w_n \mid p_1,\ldots,p_m)] \\ &= \sum_{i=1}^n [\sum_{j=1}^m \log(p_j) \times 1(W_i = j)] \\ &= \sum_{j=1}^m x_j log(p_j) \end{aligned}$$
- Maximum Likelihood Estimate (MLE) of $(p_1, \ldots, p_m)$
  maximizes $lik(p_1, \ldots, p_m)$ (with $x_1, \ldots, x_m$ fixed!)
  - Maximum achieved when differential is zero
  - Constraint: $\sum_{j=1}^m p_j = 1$
  - Apply method of Lagrange multipliers

    Solution: $\hat{p}_j = x_j/n$, $j = 1, \ldots, m$.

    Note: if any $x_j = 0$, then $\hat{p}_j = 0$ solved as limit

MLEs of Multinomial Parameter

Example 8.5.1.A Hardy-Weinberg Equilibrium

- Equilibrium frequency of genotypes: AA, Aa, and aa
- $P(a) = \theta$ and $P(A) = 1 - \theta$
- Equilibrium probabilities of genotypes: $(1-\theta)^2$, $2(\theta)(1-\theta)$, and θ^2.
- Multinomial Data: (X_1, X_2, X_3) corresponding to counts of AA, Aa, and aa in a sample of size n.

Sample Data

Genotype	*AA*	*Aa*	*aa*	*Total*
Count	X_1	X_2	X_3	n
Frequency	342	500	187	1029

Hardy-Weinberg Equilibrium

Maximum-Likelihood Estimation of θ

- $(X_1, X_2, X_3) \sim Multinomial(n, p = ((1-\theta)^2, 2\theta(1-\theta), \theta^2))$
- Log Likelihood for θ

$$\begin{aligned}
\ell(\theta) &= log(f(x_1, x_2, x_3 \mid p_1(\theta), p_2(\theta), p_3(\theta))) \\
&= log(\tfrac{n!}{x_1!x_2!x_3!} p_1(\theta)^{x_1} p_2(\theta)^{x_2} p_3(\theta)^{x_3}) \\
&= x_1 log((1-\theta)^2) + x_2 log(2\theta(1-\theta)) \\
&\qquad + x_3 log(\theta^2) + (\text{non-}\theta\ terms) \\
&= (2x_1 + x_2) log(1-\theta) + (2x_3 + x_2) log(\theta) + (\text{non-}\theta\ terms)
\end{aligned}$$

- First Differential of log likelihood:

$$\ell'(\theta) = -\frac{(2x_1 + x_2)}{1-\theta} + \frac{(2x_3 + x_2)}{\theta}$$

$$\Longrightarrow \quad \hat{\theta} = \frac{2x_3 + x_2}{2x_1 + 2x_2 + 2x_3} = \frac{2x_3 + x_2}{2n} = 0.4247$$

- Asymptotic variance of MLE $\hat{\theta}$:

$$Var(\hat{\theta}) \longrightarrow \frac{1}{E[-\ell''(\theta)]}$$

- Second Differential of log likelihood:

$$\begin{aligned}\ell''(\theta) &= \frac{d}{d\theta}[-\frac{(2x_1+x_2)}{1-\theta} + \frac{(2x_3+x_2)}{\theta}] \\ &= -\frac{(2x_1+x_2)}{(1-\theta)^2} - \frac{(2x_3+x_2)}{\theta^2}\end{aligned}$$

- Each of the X_i are $Binomial(n, p_i(\theta))$ so

$$\begin{aligned}E[X_1] &= np_1(\theta) = n(1-\theta)^2 \\ E[X_2] &= np_2(\theta) = n2\theta(1-\theta) \\ E[X_3] &= np_3(\theta) = n\theta^2\end{aligned}$$

- $E[-\ell''(\theta)] = \dfrac{2n}{\theta(1-\theta)}$

- $\hat{\sigma}_{\hat{\theta}} = \sqrt{\dfrac{\hat{\theta}(1-\hat{\theta})}{2n}} = \sqrt{\dfrac{.4247(1-.4247)}{2 \times 1029}} = 0.0109$

Hardy-Weinberg Model

Approximate 95% **Confidence Interval for** θ

Interval : $\hat{\theta} \pm z(\alpha/2) \times \hat{\sigma}_{\hat{\theta}}$, with

- $\hat{\theta} = 0.4247$
- $\hat{\sigma}_{\hat{\theta}} = 0.0109$
- $z(\alpha/2) = 1.96$ (with $\alpha = 1 - 0.95$)

Interval : $0.4247 \pm 1.96 \times (0.0109) = [0.4033, .4461]$

Note (!!) : Bootstrap simulation in R of $\hat{\theta}$ the $RMSE(\hat{\theta}) = 0.0109$ is virtually equal to $\hat{\sigma}_{\hat{\theta}}$