SERIK SAGITOV, Chalmers Tekniska Högskola, January 28, 2008

Chapter 8. Estimation of parameters and fitting of probability distributions

Given a parametric model with unknown parameter(s) θ estimate θ from a random sample $(X_1, \ldots, X_n)$

Two basic methods of finding good estimates

1. method of moments, simple, first approximation for
2. max likelihood method, good for large samples

1. Parametric models

Binomial $\mathrm{Bin}(n, p)$: no. successes in n Bernoulli trials

$f(k) = \binom{n}{k} p^k q^{n-k}$, $0 \le k \le n$, $\mu = np$, $\sigma^2 = npq$

Hypergeometric $\mathrm{Hg}(N, n, p)$: sampling without repl

$f(k) = \frac{\binom{Np}{k}\binom{Nq}{n-k}}{\binom{N}{n}}$, $\mu = np$, $\sigma^2 = npq(1 - \frac{n-1}{N-1})$

Geometric $\mathrm{Geom}(p)$: no. trials untill first success

$f(k) = pq^{k-1}$, $k \ge 1$, $\mu = \frac{1}{p}$, $\sigma^2 = \frac{q}{p^2}$

Poisson $\mathrm{Pois}(\lambda)$: no. rare events $\approx \mathrm{Bin}(n, \lambda/n)$

$f(k) = \frac{\lambda^k}{k!} e^{-\lambda}$, $k \ge 0$, $\mu = \sigma^2 = \lambda$

Exponential $\mathrm{Exp}(\lambda)$: Poisson waiting times

$f(x) = \lambda e^{-\lambda x}$, $x > 0$, $\mu = \sigma = \frac{1}{\lambda}$

Normal $\mathrm{N}(\mu, \sigma^2)$: many small independent contributions

$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}$, $-\infty < x < \infty$

Gamma(α, λ): shape parameter α, scale parameter λ

$f(x) = \frac{1}{\Gamma(\alpha)} \lambda^\alpha x^{\alpha-1} e^{-\lambda x}$, $x \ge 0$, $\mu = \frac{\alpha}{\lambda}$, $\sigma^2 = \frac{\alpha}{\lambda^2}$

2. Method of moments

IID sample $(X_1, \ldots, X_n)$ from PD(θ_1, θ_2)

pop. moments $\mathrm{E}(X) = f(\theta_1, \theta_2)$, $\mathrm{E}(X^2) = g(\theta_1, \theta_2)$

MME $(\tilde\theta_1, \tilde\theta_2)$

solve equations $\bar X = f(\tilde\theta_1, \tilde\theta_2)$ and $\overline{X^2} = g(\tilde\theta_1, \tilde\theta_2)$

Ex 1: red mites

(6 apple trees) $\times$ (25 leaves) were selected

$(X_1, \ldots, X_{150})$ = numbers of red mites on 150 leaves

no. mites	0	1	2	3	4	5	6	7	Total
no. leaves	70	38	17	10	9	3	2	1	150

Poisson model $X \sim \mathrm{Pois}(\lambda)$: constant infestation rate λ

$\mathrm{E}(X) = \lambda$, MME $\tilde\lambda = \bar X = \frac{172}{150} = 1.147$

To measure the Poisson model fit to the data compute

$$\text{Chi-square test statistic: } X^2 = \Sigma \frac{(O_j - E_j)^2}{E_j}$$

$$E_j = 150 \cdot \frac{(1.147)^{j-1}}{(j-1)!} \cdot e^{-1.147}, \; E_5 = 150 - E_1 - \ldots - E_4$$

cell j	observed O_j	expected E_j	$\frac{(O_j-E_j)^2}{E_j}$
1	70	47.7	10.4
2	38	54.6	5.0
3	17	31.3	6.5
4	10	12.0	0.3
5	15	4.4	30.6
Total	150	150	$X^2 = 52.8$

Ex 2: bird hops

X_i = no. hops that a bird does between flights

No. hops	1	2	3	4	5	6	7	8	9	10	11	12	Tot
Frequency	48	31	20	9	6	5	4	2	1	1	2	1	130

Summary statistics

$\bar{X} = \frac{\text{total number of hops}}{\text{number of birds}} = \frac{363}{130} = 2.79$

$\overline{X^2} = 1^2 \cdot \frac{48}{130} + 2^2 \cdot \frac{31}{130} + \ldots + 11^2 \cdot \frac{2}{130} + 12^2 \cdot \frac{1}{130} = 13.20$

$s^2 = \frac{130}{129}(\overline{X^2} - \bar{X}^2) = 5.47$

$s_{\bar{X}} = \sqrt{\frac{5.47}{130}} = 0.205$

An approximate 95% CI for μ

$\bar{X} \pm z_{0.025} \cdot s_{\bar{X}} = 2.79 \pm 1.96 \cdot 0.205 = 2.79 \pm 0.40$

Geometric model $X \sim \text{Geom}(p)$

$\mu = 1/p, \qquad \tilde{p} = 1/\bar{X} = 0.358$

approx. 95% CI for p: $(\frac{1}{2.79+0.40}, \frac{1}{2.79-0.40}) = (0.31, 0.42)$

Model fit

j	1	2	3	4	5	6	7+
O_j	48	31	20	9	6	5	11
E_j	46.5	29.9	19.2	12.3	7.9	5.1	9.1

$E_j = 130 \cdot (0.642)^{j-1}(0.358)$

$E_7 = 130 - E_1 - \ldots - E_6$

chi-square test statistic $X^2 = 1.86$

3. Maximum Likelihood method

Before sampling

$X_1, \ldots, X_n$ have joint pmf/pdf $f(x_1, \ldots x_n|\theta)$

- draw three pdf curves for $\theta_1 < \theta_2 < \theta_3$

After sampling

$x_1, \ldots, x_n$ are the observed sample values (fixed)

likelihood $L(\theta) = f(x_1, \ldots x_n|\theta)$ is a function of θ

- likelihood curve connects pdf values for $\theta_1 < \theta_2 < \theta_3$

MLE $\hat{\theta}$ of θ is the value of θ that maximizes $L(\theta)$

Large sample properties of MLE

If sample is iid, then

$L(\theta) = f(x_1|\theta) \ldots f(x_n|\theta)$ which implies for large n

Normal approximation $\hat{\theta} \in \mathrm{N}(\theta, \frac{1}{nI(\theta)})$

Fisher information in a single observation

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

MLE $\hat{\theta}$ is asymptotically unbiased, consistent, and asymptotically efficient (minimal variance)

Cramer-Rao inequality:

$\mathrm{Var}(\theta^*) \geq \frac{1}{nI(\theta)}$ if θ^* is an unbiased estimate of θ

Approximate $100(1-\alpha)\%$ CI for θ: $\hat{\theta} \pm \frac{z_{\alpha/2}}{\sqrt{nI(\hat{\theta})}}$

Ex 3: bike helmets

Data: $n = 10$ new bike helmets are tested

$X = 3$ helmets are flawed

Binomial model $X \sim \text{Bin}(n, p)$

$p =$ population proportion of flawed helmets

MME: sample proportion $\tilde{p} = \frac{X}{n} = 0.3$, since $\mu = np$

Bin(n, p): sample proportion is MME and MLE of p

For what value of p is the observed $X = 3$ most likely?

likelihood $L(p) = \text{P}(X = 3) = 120p^3(1-p)^7$

Maximize log-likelihood

$\log L(p) = c + 3\log(p) + 7\log(1-p)$

$\frac{d}{dp}(3\log(p) + 7\log(1-p)) = 0$

$\frac{3}{p} = \frac{7}{1-p}$ so that $\hat{p} = 3/10$

Ex 4: lifetimes

Lifetimes of five batteries measured in hours

$x_1 = 0.5, x_2 = 14.6, x_3 = 5.0, x_4 = 7.2, x_5 = 1.2$

Exponential model $X \sim \text{Exp}(\lambda)$: $\lambda =$ death rate per hour

$\mu = 1/\lambda$, $\tilde{\lambda} = 1/\bar{X} = \frac{5}{28.5} = 0.175$

Likelihood function

$L(\lambda) = \lambda e^{-\lambda x_1} \lambda e^{-\lambda x_2} \lambda e^{-\lambda x_3} \lambda e^{-\lambda x_4} \lambda e^{-\lambda x_5}$

$= \lambda^n e^{-\lambda(x_1 + \ldots + x_n)} = \lambda^5 e^{-\lambda \cdot 28.5}$

It grows from 0 to $2.2 \cdot 10^{-7}$ and then falls down

likelihood maximum is reached at $\hat{\lambda} = 0.175$

MLE $\hat{\lambda} = 1/\bar{X}$ is biased but asymptotically unbiased

$E(\hat{\lambda}) \approx \lambda$ for large samples since $\bar{X} \approx \mu$

Fisher information

$\frac{\partial^2}{\partial \lambda^2} \log f(X|\lambda) = -1/\lambda^2$, $I(\lambda) = \frac{1}{\lambda^2}$

$\mathrm{Var}(\hat{\lambda}) \approx \frac{\lambda^2}{n}$

Approximate 95% CI for λ

$0.175 \pm 1.96 \frac{0.175}{\sqrt{5}} = 0.175 \pm 0.153$

Ex 5: male heights

Male height sample of size $n = 24$

170,175,176,176,177,178,178,179,179,180,180,180,

180,180,181,181,182,183,184,186,187,192,192,199

Summary statistics

$\bar{X} = 181.46$, $\overline{X^2} = 32964.2$, $\overline{X^2} - \bar{X}^2 = 37.08$

Gamma model $X \sim \mathrm{Gamma}(\alpha, \lambda)$

method of moments: $E(X) = \frac{\alpha}{\lambda}$, $E(X^2) = \frac{\alpha(\alpha+1)}{\lambda^2}$ imply

$\tilde{\alpha} = \bar{X}^2/(\overline{X^2} - \bar{X}^2) = 887.96$, $\tilde{\lambda} = \tilde{\alpha}/\bar{X} = 4.89$

Maximum likelihood method

$\frac{\partial}{\partial \alpha} \log L(\alpha, \lambda) = n \log(\lambda) + \Sigma \log X_i - n \frac{\Gamma'(\alpha)}{\Gamma(\alpha)}$

$\frac{\partial}{\partial \lambda} \log L(\alpha, \lambda) = \frac{n\alpha}{\lambda} - \Sigma X_i$

Solve numerically two equations

$\log(\hat{\alpha}/\bar{X}) = -\frac{1}{n} \Sigma \log X_i + \Gamma'(\hat{\alpha})/\Gamma(\hat{\alpha})$

$\hat{\lambda} = \hat{\alpha}/\bar{X}$

with initial values $\tilde{\alpha} = 887.96$, $\tilde{\lambda} = 4.89$

Mathematica: $\hat{\alpha} = 908.76, \hat{\lambda} = 5.01$

FindRoot[Log[a] == 0.00055+Gamma'[a]/Gamma[a], {a, 887.96}]

Parametric bootstrap

Simulate

1000 samples of size 24 from Gamma(908.76; 5.01)

find 1000 estimates $\hat{\alpha}_j$ and plot a histogram

Use the simulated sampling distribution of $\hat{\alpha}$ and $\hat{\lambda}$

to find $\bar{\alpha} = 1039.0$ and $s_{\hat{\alpha}} = \sqrt{\frac{1}{999}\Sigma(\hat{\alpha}_j - \bar{\alpha})^2} = 331.29$

large standard error because of small $n = 24$

Bootstrap algorithm to find approximate 95% CI:

$\hat{\alpha} \to \hat{\alpha}_1, \ldots, \hat{\alpha}_B \to$ sampling distribution of $\hat{\hat{\alpha}}$

$\to$ 95% brackets c_1, c_2

$$
\begin{aligned}
0.95 &\approx \mathrm{P}(\ c_1 < \hat{\hat{\alpha}} < c_2) \\
&= \mathrm{P}(c_1 - \hat{\alpha} < \hat{\hat{\alpha}} - \hat{\alpha} < c_2 - \hat{\alpha}) \\
&\approx \mathrm{P}(c_1 - \hat{\alpha} < \hat{\alpha} - \alpha < c_2 - \hat{\alpha}) \\
&= \mathrm{P}(2\hat{\alpha} - c_2 < \alpha < 2\hat{\alpha} - c_1)
\end{aligned}
$$

Matlab commands

gamrnd(908.76*ones(1000,24), 5.01*ones(1000,24))

prctile(x,2.5), prctile(x,97.5)

4. Exact CI

Assumption on the PD

IID sample $(X_1, \ldots, X_n)$ is taken from $\mathrm{N}(\mu, \sigma^2)$

with unspecified parameters μ and σ

Exact distributions $\frac{\bar{X}-\mu}{s_{\bar{X}}} \sim t_{n-1}$ and $\frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{n-1}$

t_{n-1}-distribution curve looks similar to N(0,1)-curve
symmetric around zero, larger variance $= \frac{n-1}{n-3}$
If $Z, Z_1, \ldots, Z_k$ are N(0,1)
and independent, then $\frac{Z}{\sqrt{(Z_1^2+\ldots+Z_k^2)/n}} \sim t_k$
Different shapes of χ^2_k-distribution
$\mu = k, \sigma^2 = 2k$, pdf $f_1(0) = \infty$, $f_2(0) = 0.5$, $f_3(0) = 0$
if $Z_i \sim$ N(0,1) are IID, then $Z_1^2 + \ldots + Z_k^2 \sim \chi^2_k$

Exact $100(1-\alpha)\%$ CI for μ: $\bar{X} \pm t_{n-1}(\alpha/2) \cdot s_{\bar{X}}$

Exact CI for μ is wider than the approximate CI

$\bar{X} \pm 1.96 \cdot s_{\bar{X}}$	approximate CI for large n
$\bar{X} \pm 2.26 \cdot s_{\bar{X}}$	exact CI for $n = 10$
$\bar{X} \pm 2.13 \cdot s_{\bar{X}}$	exact CI for $n = 16$
$\bar{X} \pm 2.06 \cdot s_{\bar{X}}$	exact CI for $n = 25$
$\bar{X} \pm 2.00 \cdot s_{\bar{X}}$	exact CI for $n = 60$

Exact $100(1-\alpha)\%$ CI for σ^2: $\left(\frac{(n-1)s^2}{\chi^2_{n-1}(\alpha/2)}; \frac{(n-1)s^2}{\chi^2_{n-1}(1-\alpha/2)}\right)$

Non-symmetric CI for σ^2

$(0.47s^2, 3.33s^2)$ for $n = 10$	$(0.55s^2, 2.40s^2)$ for $n = 16$
$(0.61s^2, 1.94s^2)$ for $n = 25$	$(0.72s^2, 1.49s^2)$ for $n = 60$
$(0.94s^2, 1.07s^2)$ $n = 2000$	$(0.98s^2, 1.02s^2)$ $n = 20000$

5. Sufficiency

Definition

$T = T(X_1, \ldots, X_n)$ is a sufficient statistic for θ
if given $T = t$ conditional distribution of
$(X_1, \ldots, X_n)$ does not depend on θ

A sufficient statistic T contains all the information in the sample about θ

Factorization criterium

$f(x_1, \ldots, x_n|\theta) = g(t, \theta)h(x_1, \ldots, x_n)$
$\mathrm{P}(\mathbf{X} = \mathbf{x}|T = t) = \frac{h(\mathbf{x})}{\Sigma_{\mathbf{x}:T(\mathbf{x})=t}\, h(\mathbf{x})}$ independent of θ

If T is sufficient for θ, the MLE is a function of T

Bernoulli distribution

$\mathrm{P}(X_i = x) = \theta^x(1-\theta)^{1-x}$
$f(x_1, \ldots, x_n|\theta) = \Pi_{i=1}^n\, \theta^{x_i}(1-\theta)^{1-x_i} = \theta^{n\bar{x}}(1-\theta)^{n-n\bar{x}}$
sufficient statistic $T = n\bar{X}$ number of successes
$g(t, \theta) = \theta^{n\bar{x}}(1-\theta)^{n-n\bar{x}}$

Normal distribution $\mathrm{N}(\mu, \sigma^2)$

$\Pi_{i=1}^n\, \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}} = \frac{1}{\sigma^n(2\pi)^{n/2}} e^{-\frac{t_2 - 2\mu t_1 + n\mu^2}{2\sigma^2}}$
sufficient statistic $(t_1, t_2) = (\Sigma_{i=1}^n\, x_i, \Sigma_{i=1}^n\, x_i^2)$

Rao-Blackwell theorem

two estimates of θ: $\hat{\theta}$ and $\tilde{\theta} = \mathrm{E}(\hat{\theta}|T)$
if $\mathrm{E}(\hat{\theta}^2) < \infty$, then $\mathrm{MSE}(\tilde{\theta}) \leq \mathrm{MSE}(\hat{\theta})$