Estimation in the continuous case, 3.5 and sampling distributions

Grethe Hystad

October 2, 2012

The likelihood function

Recall the likelihood function is defined by,

$$L(\theta) = \prod_{i=1}^{n} f(x_i, \theta) \quad \text{for} \quad \theta \in \Omega$$

The maximum likelihood estimator, MLE, of θ is

$$\hat{\theta} = \hat{\theta}(X_1, ..., X_n) = \arg\max_{\theta}(L(\theta))$$

which maximizes $L(\theta)$ in Ω.

The exponential function

Example

Consider the exponential distribution with p.d.f. $f(x;\theta) = \frac{1}{\theta}e^{-\frac{x}{\theta}}$ for $0 < x < \infty$ with $\Omega = \{\theta : 0 < \theta < \infty\}$. Determine the maximum likelihood estimator of θ

The exponential function

Solution

- *Let $X_1, ..., X_n$ be independent and identically distributed random variables from an exponential distribution with parameter θ.*
- *Then* $L(\theta) = \frac{1}{\theta}e^{-\frac{x_1}{\theta}} \cdots \frac{1}{\theta}e^{-\frac{x_n}{\theta}} = \frac{1}{\theta^n}e^{-\frac{\sum_{i=1}^{n} x_i}{\theta}}$ *for* $0 < \theta < \infty$.
- *Setting* $\frac{d}{d\theta}\ln(L(\theta)) = 0$, *we obtain that the estimate of* θ *is* $\theta = \overline{x}$.
- *Hence the maximum likelihood estimator of* θ *is* $\hat{\theta} = \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$. *(More details is given in class)*

MLE, Normal distribution with known variance and unknown mean

- Let $X_1, ..., X_n$ be independent and identically distributed random variables from $N(\theta, \sigma^2)$, where $\theta \in \Omega = \{\theta : -\infty < \theta < \infty\}$ is **unknown** and σ^2 is **known**.
- The maximum likelihood estimator of θ is $\hat{\theta} = \overline{X}$.
- More details are given in class.

The likelihood function for two and more parameters

- Let $X_1, ..., X_n$ be independent and identically distributed random variables with common parameters θ_1 and θ_2 and p.d.f. denoted by $f(x; \theta_1, \theta_2)$.
- Suppose that θ_1 and θ_2 are unknown parameters.
- Define the likelihood function as

$$L(\theta_1, \theta_2) = \prod_{i=1}^{n} f(x_i; \theta_1, \theta_2).$$

- Then $\hat{\theta}_1 = \mu_1(X_1, ..., X_n)$ and $\hat{\theta}_2 = \mu_2(X_1, ..., X_n)$ are the maximum likelihood estimators of θ_1 and θ_2 respectively if $\theta_1 = \mu(x_1, ..., x_n)$ and $\theta_2 = \mu(x_1, ..., x_n)$ maximizes $L(\theta_1, \theta_2)$.

MLE, Normal distribution with unknown mean and unknown variance

- Let $X_1, ..., X_n$ be independent and identically distributed random variables from $N(\theta_1, \theta_2)$, where the parameter space is $\Omega = \{(\theta_1, \theta_2) : -\infty < \theta_1 = \mu < \infty,\ \ 0 < \theta_2 = \sigma^2 < \infty\}$.

MLE, Normal distribution with unknown mean and unknown variance

- Let $X_1, ..., X_n$ be independent and identically distributed random variables from $N(\theta_1, \theta_2)$, where the parameter space is $\Omega = \{(\theta_1, \theta_2) : -\infty < \theta_1 = \mu < \infty, \ 0 < \theta_2 = \sigma^2 < \infty\}$.
- Both θ_1 and θ_2 are unknown.

MLE, Normal distribution with unknown mean and unknown variance

- Let $X_1, ..., X_n$ be independent and identically distributed random variables from $N(\theta_1, \theta_2)$, where the parameter space is $\Omega = \{(\theta_1, \theta_2) : -\infty < \theta_1 = \mu < \infty, \ \ 0 < \theta_2 = \sigma^2 < \infty\}$.
- Both θ_1 and θ_2 are unknown.
- The maximum likelihood estimator of $\theta_1 = \mu$ is $\hat{\theta_1} = \overline{X}$.

MLE, Normal distribution with unknown mean and unknown variance

- Let $X_1, ..., X_n$ be independent and identically distributed random variables from $N(\theta_1, \theta_2)$, where the parameter space is $\Omega = \{(\theta_1, \theta_2) : -\infty < \theta_1 = \mu < \infty,\ \ 0 < \theta_2 = \sigma^2 < \infty\}$.
- Both θ_1 and θ_2 are unknown.
- The maximum likelihood estimator of $\theta_1 = \mu$ is $\hat{\theta_1} = \overline{X}$.
- The maximum likelihood estimator of $\theta_2 = \sigma^2$ is $\hat{\theta_2} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$.

MLE, Normal distribution with unknown mean and unknown variance

- Let $X_1, ..., X_n$ be independent and identically distributed random variables from $N(\theta_1, \theta_2)$, where the parameter space is $\Omega = \{(\theta_1, \theta_2) : -\infty < \theta_1 = \mu < \infty, \ \ 0 < \theta_2 = \sigma^2 < \infty\}$.
- Both θ_1 and θ_2 are unknown.
- The maximum likelihood estimator of $\theta_1 = \mu$ is $\hat{\theta_1} = \overline{X}$.
- The maximum likelihood estimator of $\theta_2 = \sigma^2$ is $\hat{\theta_2} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$.
- $\hat{\theta_1} = \overline{X}$ is an **unbiased estimator** of $\theta_1 = \mu$ since $E(\hat{\theta_1}) = \theta_1$.

MLE, Normal distribution with unknown mean and unknown variance

- Let $X_1, ..., X_n$ be independent and identically distributed random variables from $N(\theta_1, \theta_2)$, where the parameter space is $\Omega = \{(\theta_1, \theta_2) : -\infty < \theta_1 = \mu < \infty,\ \ 0 < \theta_2 = \sigma^2 < \infty\}$.
- Both θ_1 and θ_2 are unknown.
- The maximum likelihood estimator of $\theta_1 = \mu$ is $\hat{\theta_1} = \overline{X}$.
- The maximum likelihood estimator of $\theta_2 = \sigma^2$ is $\hat{\theta_2} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$.
- $\hat{\theta_1} = \overline{X}$ is an **unbiased estimator** of $\theta_1 = \mu$ since $E(\hat{\theta_1}) = \theta_1$.
- $\hat{\theta_2} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$ is a **biased estimator** of $\theta_2 = \sigma^2$ since $E(\hat{\theta_2}) = \frac{n}{n-1}\theta_2 \neq \theta_2$.

MLE, Normal distribution with unknown mean and unknown variance

- Let $X_1, ..., X_n$ be independent and identically distributed random variables from $N(\theta_1, \theta_2)$, where the parameter space is $\Omega = \{(\theta_1, \theta_2) : -\infty < \theta_1 = \mu < \infty, \ \ 0 < \theta_2 = \sigma^2 < \infty\}$.
- Both θ_1 and θ_2 are unknown.
- The maximum likelihood estimator of $\theta_1 = \mu$ is $\hat{\theta_1} = \overline{X}$.
- The maximum likelihood estimator of $\theta_2 = \sigma^2$ is $\hat{\theta_2} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$.
- $\hat{\theta_1} = \overline{X}$ is an **unbiased estimator** of $\theta_1 = \mu$ since $E(\hat{\theta_1}) = \theta_1$.
- $\hat{\theta_2} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$ is a **biased estimator** of $\theta_2 = \sigma^2$ since $E(\hat{\theta_2}) = \frac{n}{n-1}\theta_2 \neq \theta_2$.
- Recall that $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$ is the **unbiased estimator** of σ^2 but is not the MLE.

MLE, Normal distribution with known mean and unknown variance

- Let $X_1, ..., X_n$ be independent and identically distributed random variables from $N(\mu, \theta)$, where the parameter space is $\Omega = \{\theta : 0 < \theta = \sigma^2 < \infty\}$.

MLE, Normal distribution with known mean and unknown variance

- Let $X_1, ..., X_n$ be independent and identically distributed random variables from $N(\mu, \theta)$, where the parameter space is $\Omega = \{\theta : 0 < \theta = \sigma^2 < \infty\}$.
- Then $\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2$ is a maximum likelihood estimator of $\theta = \sigma^2$.

MLE, Normal distribution with known mean and unknown variance

- Let $X_1, ..., X_n$ be independent and identically distributed random variables from $N(\mu, \theta)$, where the parameter space is $\Omega = \{\theta : 0 < \theta = \sigma^2 < \infty\}$.
- Then $\hat{\theta} = \frac{1}{n} \sum_{i=1}^{n} (X_i - \mu)^2$ is a maximum likelihood estimator of $\theta = \sigma^2$.
- $\hat{\theta}$ is an **unbiased estimator** of $\theta = \sigma^2$ since $E(\hat{\theta}) = \theta$.

MLE, Normal distribution with known mean and unknown variance

- Let $X_1, ..., X_n$ be independent and identically distributed random variables from $N(\mu, \theta)$, where the parameter space is $\Omega = \{\theta : 0 < \theta = \sigma^2 < \infty\}$.
- Then $\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2$ is a maximum likelihood estimator of $\theta = \sigma^2$.
- $\hat{\theta}$ is an **unbiased estimator** of $\theta = \sigma^2$ since $E(\hat{\theta}) = \theta$.
- To prove this is left as a homework problem.

Sampling distribution

- Suppose we have a large population and draw all possible samples of size n from the population.
- Suppose for each sample, we compute a statistics (for example the sample mean).
- The sampling distribution is the probability distribution of this statistics considered as a random variable.

Sampling distribution

The sampling distribution depends on:

- The underlying population distribution.
- The statistics being computed.
- The sample size.
- The sampling procedure.

We measure the variability of the sampling distribution by its variance or its standard deviation.

The normal distribution, sampling

Example

- Consider a normal population with mean μ and variance σ^2.
- Suppose we repeatedly take samples of size n from this population and calculate the statistics, the sample mean $\overline{x}$ for each sample.
- Recall $E(\overline{X}) = \mu$ and $\text{Var}(\overline{X}) = \frac{\sigma^2}{n}$
- The sample mean $\overline{X}$ is normal, $N(\mu, \frac{\sigma^2}{n})$, since the underlying population is normal.

Rolling a die

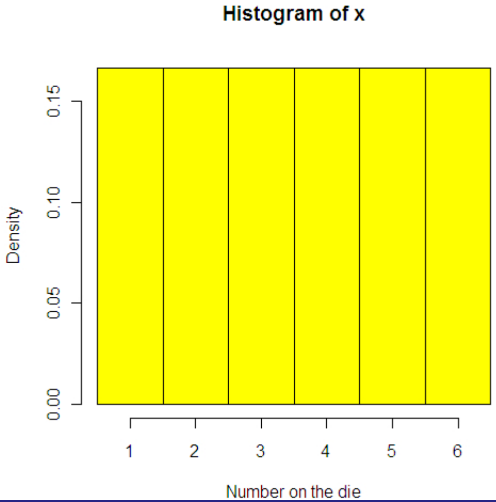
Example

Suppose a fair die is rolled an infinitely number of times. Let the random variable X be the number on the die on any throw.
The probability distribution of X is given by:

X	1	2	3	4	5	6
$P(X = x)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

- We have that the population mean is $\mu = \frac{1}{6}(1 + 2 + 3 + 4 + 5 + 6) = 3.5$
- The population standard deviation is $\sigma^2 = 2.92$.

Rolling a die



Rolling two dice

Example

We will now roll two dice ($n = 2$). Let the random variable X_1 be the number on the first die and let the random variable X_2 be the number on the second die. We will look at the probability distribution of the mean $\overline{X}$ of rolling two dice.

Sample	$\overline{x}$	Sample	$\overline{x}$	Sample	$\overline{x}$	Sample	$\overline{x}$
$(1,1)$	1.0	$(2,1)$	1.5	..	..	$(6,1)$	3.5
$(1,2)$	1.5	$(2,2)$	2.0	..	..	$(6,2)$	4.0
$(1,3)$	2.0	$(2,3)$	2.5	$(5,3)$	4.0	$(6,3)$	4.5
$(1,4)$	2.5	$(2,4)$	3.0	$(5,4)$	3.5	$(6,4)$	5.0
$(1,5)$	3.0	$(2,5)$	3.5	$(5,5)$	5.0	$(6,5)$	5.5
$(1,6)$	3.5	$(2,6)$	4.0	$(5,6)$	5.5	$(6,6)$	6.0

Rolling two dice

Example

We have the following sampling distribution of $\overline{X}$.

$\overline{x}$	$f(\overline{x})$
1.0	1/36
1.5	2/36
2.0	3/36
2.5	4/36
3.0	5/36
3.5	6/36
4.0	5/36
4.5	4/36
5.0	3/36
5.5	2/36
6.0	1/36

Rolling two dice

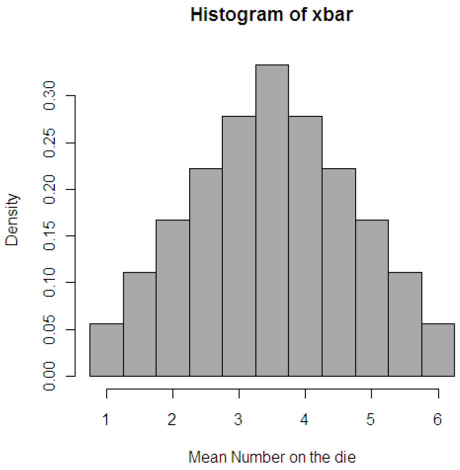
Example

We have

$$E(\overline{X}) = \sum_{\overline{x}} \overline{x} f(\overline{x})$$
$$= (1.0)(1/36) + (1.5)(2/36) + \cdots + (6.0)(1/36) = 3.5 = \mu.$$

$$\text{Var}(\overline{X}) = \sum_{\overline{x}} (\overline{x} - 3.5)^2 f(\overline{x})$$
$$= (1.0 - 3.5)^2 (1/36) + (1.5 - 3.5)^2 (2/36) +$$
$$+ \cdots + (6.0 - 3.5)^2 (1/36) = 1.46 = \frac{\sigma^2}{2}.$$

Rolling two dice



Rolling n dice

- We see from the histogram that $\overline{X}$ looks approximately normal.
- In general, roll a die n times. Then $E(\overline{X}) = \mu$ and $\text{Var}(\overline{X}) = \frac{\sigma^2}{n}$.
- We will see in chapter 3.6 that $\overline{X}$ is approximately normal, $N(\mu, \frac{\sigma^2}{n})$, for large enough n (Central limit theorem).

Maximum likelihood estimators

- Maximum likelihood estimators in regular cases are approximately normal.

Maximum likelihood estimators

- Maximum likelihood estimators in regular cases are approximately normal.
- The mean $\overline{X}$ of a random sample of size n from any distribution with mean μ and finite variance σ^2 is approximately $N(\mu, \frac{\sigma^2}{n})$ for large enough sample size n (Central limit theorem, more about this later).

Maximum likelihood estimators

- Maximum likelihood estimators in regular cases are approximately normal.
- The mean $\overline{X}$ of a random sample of size n from any distribution with mean μ and finite variance σ^2 is approximately $N(\mu, \frac{\sigma^2}{n})$ for large enough sample size n (Central limit theorem, more about this later).
- For the exponential distribution, with parameter θ, $\overline{X}$ is approximately $N(\theta, \frac{\theta^2}{n})$ since $E(\overline{X}) = \theta$ and $\text{Var}(\overline{X}) = \frac{\theta^2}{n}$.

Maximum likelihood estimators

- Maximum likelihood estimators in regular cases are approximately normal.
- The mean $\overline{X}$ of a random sample of size n from any distribution with mean μ and finite variance σ^2 is approximately $N(\mu, \frac{\sigma^2}{n})$ for large enough sample size n (Central limit theorem, more about this later).
- For the exponential distribution, with parameter θ, $\overline{X}$ is approximately $N(\theta, \frac{\theta^2}{n})$ since $E(\overline{X}) = \theta$ and $\text{Var}(\overline{X}) = \frac{\theta^2}{n}$.
- For the binomial distribution, $\overline{X}$ is approximately $N(p, \frac{p(1-p)}{n})$ since $E(\overline{X}) = p$ and $\text{Var}(\overline{X}) = \frac{p(1-p)}{n}$, where p is the success probability.

Confidence interval

- Let's say that an estimator $U = u(X_1, ..., X_n)$ of θ has a normal or approximately normal distribution with unknown mean θ and known variance σ^2. Then

$$\frac{U - E(U)}{\sqrt{\text{Var}(U)}} = \frac{U - \theta}{\sigma}$$

is approximately $N(0, 1)$.

Confidence interval

- Let's say that an estimator $U = u(X_1, ..., X_n)$ of θ has a normal or approximately normal distribution with unknown mean θ and known variance σ^2. Then

$$\frac{U - E(U)}{\sqrt{\text{Var}(U)}} = \frac{U - \theta}{\sigma}$$

  is approximately $N(0, 1)$.

- Then using the standard normal distribution, $P(-2 \leq \frac{U-\theta}{\sigma} \leq 2) \approx 0.95$ which implies that $P(U - 2\sigma \leq \theta \leq U + 2\sigma) \approx 0.95$.

Confidence interval

- Let $x_1, ..., x_n$ be the observed values, so the estimate is $u = u(x_1, ..., x_n)$.

Confidence interval

- Let $x_1, ..., x_n$ be the observed values, so the estimate is $u = u(x_1, ..., x_n)$.
- Thus, we are 95% confident that θ is in the interval, $(u(x_1, ..., x_n) - 2\sigma, u(x_1, ..., x_n) + 2\sigma)$.

Confidence interval

- Let $x_1, ..., x_n$ be the observed values, so the estimate is $u = u(x_1, ..., x_n)$.
- Thus, we are 95% confident that θ is in the interval, $(u(x_1, ..., x_n) - 2\sigma, u(x_1, ..., x_n) + 2\sigma)$.
- We say that the interval, written as, $\mu \pm 2\sigma$, is an approximately 95% confidence interval for θ.

Confidence interval

Let $\overline{X}$ be the mean of n independent and identically distributed random variables from a distribution with **unknown** mean μ and **known** variance σ^2. Then an approximately 95% confidence interval for μ is

$$\overline{x} \pm 2\frac{\sigma}{\sqrt{n}}$$

provided n is large enough.

Confidence interval

Let $\overline{X}$ be the mean of n independent and identically distributed random variables from a distribution with **unknown** mean μ and **unknown** variance σ^2. Recall that $\frac{s}{\sqrt{n}}$ is an estimate of $\frac{\sigma}{\sqrt{n}}$, the standard deviation of $\overline{X}$. Thus, an approximately 95% confidence interval for μ is then

$$\overline{x} \pm 2\frac{s}{\sqrt{n}}$$

provided n is large enough.

Confidence interval

If $X \in b(n,p)$, recall that $\sqrt{\frac{(\frac{x}{n})(1-\frac{x}{n})}{n}}$ is an estimate of $\sqrt{\frac{p(1-p)}{n}}$, the standard deviation of $\frac{X}{n}$. Thus, an approximately 95% confidence interval for p is then

$$\frac{x}{n} \pm 2\sqrt{\frac{(\frac{x}{n})(1-\frac{x}{n})}{n}}$$

provided n is large enough.

Confidence interval

Example

Let
8.6, 6.1, 11.9, 11.0, 11.3, 9.3, 15.8, 10.3, 6.1, 7.0, 12.0, 5.9, 13.0, 11.1, 9.3
11.6, 12.7, 8.3, 12.6, 11.5, 7.5, 11.2, 9.7, 10.5, 9.2, 9.4, 8.2, 8.8, 11.7, 8.3
be a random sample of 30 observations from a distribution with **unknown** mean μ and **known** variance $\sigma^2 = 4$. Then an approximately 95% confidence interval for μ is

$$\bar{x} \pm 2\frac{\sigma}{\sqrt{n}} = 9.9 \pm 2\frac{2}{\sqrt{30}} = 9.9 \pm 0.73.$$

Thus we are approximately 95% confident that μ is within the interval,

$$(9.2,\ 10.6)$$

The sample mean

Example

The math SAT 1 scores among U.S. college students is normally distributed with a mean of 500 and standard deviation of 100.
A. What is the probability that the SAT score for a randomly selected student is greater than 600?
B. Suppose we randomly select 50 students from U.S. colleges independently from each other. What is the probability that the mean SAT score of those 50 students is greater than 600?

Solution

- *A. Let X be the SAT score for the student. Define $Z = \frac{X-500}{100}$.*
- *Then*

$$P(X > 600) = P\left(\frac{X-500}{100} > \frac{600-500}{100}\right)$$
$$= P\left(Z > \frac{600-500}{100}\right) = P(Z > 1) = 0.16.$$

Solution

- *B. We have that* $\overline{X}$ *is* $N(500, \frac{100^2}{50})$*. Define*

$$Z = \frac{\overline{X} - 500}{\frac{100}{\sqrt{50}}}.$$

- *Then*

$$P(\overline{X} > 600) = P\left(\frac{\overline{X} - 500}{\frac{100}{\sqrt{50}}} > \frac{600 - 500}{\frac{100}{\sqrt{50}}}\right) = P\left(Z > \frac{600 - 500}{\frac{100}{\sqrt{50}}}\right)$$
$$= P(Z > 7.1) = 1 - P(Z \leq 7.1) \approx 0.$$