Math 218 Mathematical Statistics

Prof. D. Joyce, Clark University

11 Feb 2009

Due Today. From page 658, exercises 1, 3.

First test. Feb. 20.

Last meeting. Discussed maximum likelihood estimators for discrete distributions.

Today. Maximum likelihood estimators for continuous distributions, confidence intervals.

Likelihood functions for continuous distributions. For continuous distributions, we can't use probability, because the probability of any particular outcome is 0. But we can use the density function. Thus, for a continuous distribution, the likelihood of a parameter θ for a given random sample $X_1 = x_1$, $X_2 = x_2$, ..., $X_n = x_n$, also denoted $L(\theta|x_1, x_2, \ldots, x_n)$ is the product of values of the density function f as

$$f(x_1 \,|\, \theta) f(x_2 \,|\, \theta) \cdots f(x_n \,|\, \theta).$$

(So, the same formula, but the symbol f now denotes a probability density function instead of a probability mass function.)

The likelihood function for the normal distribution and its maximum likelihood estimators. Since the probability density function for a normal($\mu, \sigma^)$ distribution is

$$f(x \,|\, \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right),$$

the likelihood function is

$$\begin{aligned} &L(\mu, \sigma^2|x_1, x_2, \ldots, x_n) \\ = &\prod_{i=1}^n \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \\ = &\left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i-\mu)^2\right) \end{aligned}$$

Next, to find the maximum likelihood estimators $\hat\mu$ and $\hat\sigma^2$ for the parameters μ and σ^2, we just have to find those values of μ and σ^2 that maximize the function $L(\mu, \sigma^2|x_1, x_2, \ldots, x_n)$. We need to compute its derivative to find the critical points so we can find where the maximum occurs. But, since the function $L(\mu, \sigma^2)$ at the same places where its log does, that is, where $\ln L(\mu, \sigma^2)$ does, we'll use its log instead, because its easier to find the derivative of its log. Its log is

$$\ln L(\mu, \sigma^2) = -n \ln\sqrt{2\pi} - \frac{n}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^n (x_i-\mu)^2.$$

Here we have two parameters, μ and σ^2, so we need to set both derivatives of $\ln L(\mu, \sigma^2)$ to 0. First, the derivative with respect to μ

$$\frac{\partial}{\partial\mu}\ln L(\mu, \sigma^2) = \frac{1}{\sigma^2}\sum_{i=1}^n (x_i-\mu) = 0.$$

Since $\sum_{i=1}^n (x_i - \mu) = 0$, therefore the critical value for μ is $\frac{1}{n}\sum_{i=1}^n x_i$, which is the sample mean $\overline{x}$. This is the only critical value, so it maximizes $L(\mu, \sigma^2)$. Therefore, the maximum likelihood estimator $\hat\mu$ for the mean μ is the sample mean $\overline{x}$.

Second, the derivative with respect to σ^2

$$\frac{\partial}{\partial\sigma^2}\ln L(\mu, \sigma^2) = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^n (x_i-\mu)^2 = 0.$$

We can simplify that equation a bit to get

$$\frac{1}{\sigma^2}\sum_{i=1}^n (x_i-\mu)^2 = n$$

Therefore

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2$$

and, as we're solving these equations simultaneously, we've already determined the solution has $\mu = \overline{x}$, so we can rewrite that as

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

Again, as this is the only critical value for σ^2, it maximizes $L(\mu, \sigma^2)$. Therefore, the maximum likelihood estimator $\hat{\sigma}^2$ for the population variance σ^2 is the statistic

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

As we discussed before, sometimes this statistic is called the sample variance, but our text uses $n-1$ in the denominator for the sample variance.

Introduction to confidence intervals. Although estimating a parameter θ by a particular number $\hat{\theta}$ may be the simplest kind of statistical inference, that often is not very satisfactory. Some indication of the spread of the likely values of θ explains a lot more. One way that's done is with confidence intervals. A typical confidence interval is a 95% confidence interval $[L, U]$ for θ and that's given by two statistics, L and U such that

$$P(L \leq \theta \leq U) = 0.95.$$

Other confidence levels besides 95% are defined similarly.

This concept is best explained with an example. Let's take a normal distribution with a known value for σ^2, but an unknown value for μ, and our job is to come up with a confidence interval for μ. The sample mean $\overline{X}$ is a point estimator for μ, and we know that $\overline{X}$ is a normal distribution with mean μ and variance σ^2/n. From the table for the standard normal distribution, the probability that a standard normal random variable Z lies between -1.96 and 1.96 is 95%. Therefore,

$$P\left(\mu - 1.96\,\frac{\sigma}{\sqrt{n}} \leq \overline{X} \leq \mu + 1.96\,\frac{\sigma}{\sqrt{n}}\right) = 0.95.$$

We can rewrite the first inequality $\mu - 1.96\,\frac{\sigma}{\sqrt{n}} \leq \overline{X}$ as $\mu \leq \overline{X} + 1.96\,\frac{\sigma}{\sqrt{n}}$, and we can rewrite the second inequality $\overline{X} \leq \mu + 1.96\,\frac{\sigma}{\sqrt{n}}$ as $X - 1.96\,\frac{\sigma}{\sqrt{n}} \leq \mu$. Therefore, the statement of probability can be rewritten as

$$P\left(X - 1.96\,\frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{X} + 1.96\,\frac{\sigma}{\sqrt{n}}\right) = 0.95.$$

We now have two statistics, $L = X - 1.96\,\frac{\sigma}{\sqrt{n}}$ and $U = X + 1.96\,\frac{\sigma}{\sqrt{n}}$, so that $P(L \leq \mu \leq U) = 0.95$.