Lecture 29, Mar 23, 2023

Confidence Interval of the Variance

- Recall that $W^2 = \frac{(n-1)S^2}{\sigma^2} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \bar{X})^2$ follows a chi-squared distribution with $v = n - 1$ degrees of freedom, assuming a normal distribution
- The chi-squared distribution is asymmetric, so getting a confidence interval is harder
- Let χ^2_β be the value of χ^2 such that the area under the distribution to the left of it is β
  - $\chi^2_{\alpha/2}$ is the value of χ^2 such that the area under the distribution to the *left* of it is $\alpha/2$
  - $\chi_{1-\alpha/2}$ is the value of χ^2 such that the area under the distribution to the *right* of it is $\alpha/2$
- Denote the CDF of the chi-squared distribution as $F(y; v) = \int_0^y f(x; v)\,\mathrm{d}x$
  - $\chi^2_{\alpha/2} = F^{-1}(\alpha/2; v)$ and $\chi^2_{1-\alpha/2} = F^{-1}(1 - \alpha/2; v)$
- $$\begin{aligned} 1 - \alpha &= P(\chi^2_{\alpha/2} \leq W^2 \leq \chi^2_{1-\alpha/2}) \\ &= P\left(\chi^2_{\alpha/2} \leq \frac{(n-1)S^2}{\sigma^2} \leq \chi^2_{1-\alpha/2}\right) \\ &= P\left(\frac{\chi^2_{\alpha/2}}{(n-1)S^2} \leq \frac{1}{\sigma^2} \leq \frac{\chi^2_{1-\alpha/2}}{(n-1)S^2}\right) \\ &= P\left(\frac{(n-1)S^2}{\chi^2_{1-\alpha/2}} \leq \sigma^2 \leq \frac{(n-1)S^2}{\chi^2_{\alpha/2}}\right) \end{aligned}$$

Equation

Confidence interval of the variance: Given n IID samples, with a sample variance of S^2 and a confidence level of $1 - \alpha$, then the confidence interval for the true variance σ^2 is

$$\left[\frac{(n-1)S^2}{\chi^2_{1-\alpha/2}}, \frac{(n-1)S^2}{\chi^2_{\alpha/2}}\right]$$

where $\chi^2_\beta = F^{-1}(\beta; v)$, and F is the CDF of the chi-squared distribution with $v = n - 1$ degrees of freedom

Maximum Likelihood Estimation

- So far we've relied on intuition to define our estimators (e.g. $\bar{X}$ for μ, S^2 for σ^2, etc)
- Can we find a systematic way to define an estimator for any statistic? (e.g. what if we wanted to estimate v in a chi-squared distribution?)

Definition

The *likelihood function* for an IID sample $x_1, \cdots, x_n$, with each sample distributed according to a PDF $g(x; \theta)$, where θ is a parameter vector, is

$$\begin{aligned} L(x_1, \cdots, x_n; \theta) &= f(x_1, \cdots, x_n; \theta) \\ &= g(x_1; \theta) \cdots g(x_n; \theta) \end{aligned}$$

The *maximum likelihood estimator* is then

$$\hat{\theta} = \max_\theta L(x_1, \cdots, x_n; \theta)$$

- Maximum likelihood estimation estimates the parameter by attempting to maximize the likelihood function, which roughly describes the probability of getting the particular sample
  - In the discrete case, $f(x_1, \cdots, x_n; \theta)$ is exactly the probability of the sample occurring; with a continuous distribution it is more complicated but the intuition still holds
- Example: $n = 1, x_1 = 3, \theta = \mu$, standard normal $f(x_1; \theta) = n(x_1; \theta, 1)$
  - We're trying to move the mean around so that $f(x_1; \theta)$ is maximized
  - The optimal value is $\theta = x_1 = 3$ because the normal distribution peaks at its mean
  - i.e. having a mean of 3 makes it the most likely that we'll get $x_1 = 3$
- Example: Bernoulli distribution, estimating p, given sample $1, 0, 1, 1$
  - We can make this a binomial distribution with 3 successes
  - $L(1, 0, 1, 1; p) = \binom{4}{3} p^3 (1-p)^1 \propto p^3(1-p) = p^3 - p^4$
  - $\frac{\mathrm{d}L}{\mathrm{d}p} \propto 3p^2 - 4p^3 = 0 \implies 3 - 4p = 0 \implies p = \frac{3}{4}$
  - This is exactly what we expect – if we get 3 successes in 4 trials, then we'd estimate the success probability to be $\frac{3}{4}$