18. Interval Estimation LM 5.3

18.1 Uncertainty in point estimates

(i) When we have a point estimate, for example, $\overline{x_n}$, for the mean θ of a distribution (Poisson, Normal, ...), we would like to know how close to the true θ-value it is likely to be.

(ii) But the estimate is just a number: to consider this problem we have to consider the random variable – the *estimator*, such as $\overline{X_n}$.

(iii) We have previously considered the distributions of the estimator, computing such things as mean square error. This is one way of considering how close to the true θ-value our estimate might be.

18.2 Confidence intervals

(i) An alternative way is to compute a *confidence interval*.

(ii) Definition: Given an n-sample $X_1, ..., X_n$ from some (discrete or continuous) pdf/pmf indexed by parameter θ, let $L(X_1, ..., X_n)$ and $U(X_1, ..., X_n)$ be two functions of the data random variables, such that $L \leq U$ for all possible samples $x_1, ..., x_n$. If

$$P_\theta(L(X_1, ..., X_n) \leq \theta \leq U(X_1, ..., X_n)) \;=\; (1 - \alpha) \qquad \text{for all } \theta$$

then $(L(x_1, ..., x_n),\ U(x_1, ..., x_n))$ is a $(1 - \alpha)$-level confidence interval for θ.

(iii) Note: the interval $(L(x_1, ..., x_n),\ U(x_1, ..., x_n))$ is just numbers; there are no probabilities associated with these numbers, **nor with** θ. The probabilities concern the random intervals $(L(X_1, ..., X_n),\ U(X_1, ..., X_n))$ which we might get on repeating the sample and interval construction procedure many times.

(iv) Generally we think of α, the probability the random interval does **not** cover the true θ as small (e.g. 0.05), and construct (e.g.) 95% confidence intervals. It is conventional to have $\alpha/2$ of the probability falling at each end, although this is not actually the shortest interval if the distribution is not symmetric.

(v) Even then, there are many, many possible functions L and U of the data random variables. We use **sufficiency** to consider only ones that are functions of the sufficient statistic.

18.3 Confidence interval for the mean of a Normal distribution

We will do this one, because it is the classic example. We will pretend we know that the average of independent Normal random variables is Normal (this should be proved in Stat 342).

Suppose $X_1, ..., X_n$ are i.i.d from $N(\theta, \sigma^2)$ where σ^2 is known. We **do** know that $\overline{X_n}$ is sufficient for θ, that $\mathrm{E}(\overline{X_n}) = \theta$ and $\mathrm{var}(\overline{X_n}) = \sigma^2/n$. So assuming $\overline{X_n}$ is Normal, $(\overline{X_n} - \theta)\sqrt{n}/\sigma$ is $N(0, 1)$,

$$P(\overline{X_n} - z_{\alpha/2}\sigma/\sqrt{n} \leq \theta \leq \overline{X_n} + z_{\alpha/2}\sigma/\sqrt{n}) \;=\; P(-z_{\alpha/2} \leq (\overline{X_n} - \theta)\sqrt{n}/\sigma \leq z_{\alpha/2}) \;=\; (1 - \alpha)$$

where $z_{\alpha/2}$ is such that $P(Z > z_{\alpha/2}) = \alpha/2$ if $Z \sim N(0, 1)$.

18.4 Confidence interval for θ given a sample from $U(0, \theta)$.

Suppose $X_1, ..., X_n$ are i.i.d. $U(0, \theta)$. We know $W = \max(X_i)$ is sufficient for θ and $P(W \leq k\theta) = k^n$, $(0 < k < 1)$. So

$$P(W < \theta < W/k) \;=\; P(W > k\theta) \;=\; 1 - k^n$$

So if $k = \alpha^{1/n}$, $(\max(x_i), \max(x_i)/k)$ is $(1 - \alpha)$-level confidence interval for θ.

Note: This is an example where it is more natural NOT to use a symmetric interval.

Example: $\alpha = 0.05$, $n = 25$; interval is $(\max(x_i), 1.127 \max(x_i))$.

19. Examples of interval estimation

19.1 Confidence interval for a binomial proportion

(i) $X_1, ..., X_n$ are i.i.d $Bin(1, \theta)$. We know $T = \sum_1^n X_i \ \sim Bin(n, \theta)$ is sufficient for θ.

(ii) Discrete distributions are a nuisance because first, we have to sum probabilities, and, second we have to distinguish $\leq$ and $<$. Instead we assume n is "large enough" and use a Normal approximation.

(iii) So T/n is approx $N(\theta, \theta(1-\theta)/n)$, so

$$P(-z_{\alpha/2} \ < \ (T/n \ - \ \theta)/\sqrt{\theta(1-\theta)/n} \ < \ z_{\alpha/2}) \ = \ (1-\alpha)$$

(iv) In principle this could be solved (numerically at least) to give an interval for θ. However, since this Normal distribution is already an approximation, it is simple to approximate the variance by $(t/n)(1-t/n)/n$ where t is the value of T. Then

$$(t/n - z_{\alpha/2}\sqrt{(t/n)(1-t/n)/n}, \ (t/n) + z_{\alpha/2}\sqrt{(t/n)(1-t/n)/n})$$

is a $(1-\alpha)$-level confidence interval for θ.

19.2 Confidence interval for the rate parameter of an exponential

$Y_1, ..., Y_n$ i.i.d exponential $\mathcal{E}(\lambda)$. We know $T = \sum_1^n Y_i$ is sufficient for λ, and $T \ \sim \ G(n, \lambda)$ so $\lambda T \ \sim \ G(n, 1)$. This is now a fixed distribution, and we can find the $\alpha/2$ and $(1-\alpha/2)$ points of the distribution, say $g^-_{\alpha/2}$ and $g^+_{\alpha/2}$ (These will depend on n.) Then

$$P(g^-_{\alpha/2}/T \ < \ \lambda \ < \ g^+_{\alpha/2}/T) \ = \ P(g^-_{\alpha/2} \ < \ \lambda T \ < \ g^+_{\alpha/2}) \ = \ (1-\alpha)$$

So $(g^-_{\alpha/2}/t, \ g^+_{\alpha/2}/t)$ is $(1-\alpha)$-level confidence interval for λ.

19.3 Confidence interval for the variance in $N(0, \sigma^2)$.

$X_1, ..., X_n$ i.i.d $N(0, \sigma^2)$. We know $T = \sum_1^n X_i^2$ is sufficient for σ^2 and $T \ \sim \ \sigma^2\chi^2_n$ or $T/\sigma^2 \ \sim \ \chi^2_n \equiv G(n/2, 1/2)$. This is now a fixed distribution, and we can find the $\alpha/2$ and $(1-\alpha/2)$ points of the distribution, say $c^-_{\alpha/2}$ and $c^+_{\alpha/2}$ (These will depend on n.) Then

$$P(T/c^+_{\alpha/2} \ < \ \sigma^2 \ < \ T/c^-_{\alpha/2}) \ = \ P(c^-_{\alpha/2} \ < \ T/\sigma^2 \ < \ c^+_{\alpha/2}) \ = \ (1-\alpha)$$

So $(t/c^+_{\alpha/2}, \ t/c^-_{\alpha/2})$ is $(1-\alpha)$-level confidence interval for λ.

19.4 Sample size

(i) The length of the confidence interval, in each given example, will depend on n. If n is larger the confidence interval will be shorter.

(ii) We can (try to) find the n to make the confidence interval no more than a given length. For example, in the case of $N(\theta, \sigma^2)$ with σ^2 known, the length is $2z_{\alpha/2}\sigma/\sqrt{n}$.

(iii) For the case of a binomial proportion, σ is replaced by the Bernoulli standard deviation $\sqrt{\theta(1-\theta)}$, which depends on the unknown θ. One solution, for determining sample size, is to take the θ for which this is largest: i.e. $\theta = 1/2$, $\sqrt{\theta(1-\theta)} = 1/2$ (see LM P.373). Then if we compute the required sample size for $\theta = 1/2$ we can be sure it will be good enough whatever the true θ.