

8 Sampling Theoretic Inference

Sampling-theoretic, or *frequentist* approaches to statistical inference are based on the sampling distributions of random variables constructed as statistics, such as y or $t = y/n$. One way this is used is to produce *confidence intervals* for the unknown parameter θ , which provide rule-of-thumb estimates of *margins of error* about the estimate t .

A 95% equal-tails **confidence interval** is a random interval $[L_Y, R_Y]$ whose left and right end-points L_Y and R_Y are random variables, functions of Y , that satisfy the bounds $P[\theta < L_Y \mid \theta] \leq .025$ and $P[\theta > R_Y \mid \theta] \leq .025$ to ensure that $P[L_Y \leq \theta \leq R_Y \mid \theta] \geq 0.95$. Choosing numbers $0 \leq L_0 \leq L_1 \leq \dots \leq L_n \leq 1$ such that $P[\theta < L_Y \mid \theta] \leq .025$ is the same as finding numbers L_j such that $P[Y \geq j \mid \theta] \leq .025$ for each $\theta < L_j$; in S-Plus the requirement is $1 - \text{pbinom}(j-1, n, \text{theta}) \leq .025$ for $\text{theta} < L[j]$; since $1 - \text{pbinom}(j-1, n, \text{theta}) = \text{pbeta}(\text{theta}, j, n-j+1)$ for all $0 < \text{theta} < 1$ and integers $0 < j < n$, the best choice is $L_j = \text{qbeta}(.025, j, n-j+1)$. Similarly $R_j = 1 - L_{n-j} = \text{qbeta}(.975, j+1, n-j)$, so the *exact 95% confidence interval* is given by

$$0.95 \leq P[\text{qbeta}(.025, y, n-y+1) \leq \theta \leq \text{qbeta}(.975, y+1, n-y)].$$

For example, if we were to observe *one* Bernoulli trial (i.e., Binomial with $n = 1$), an exact 95% equal-tails confidence interval would be $[0.0, 0.975]$ if $Y = 0$ and $[0.025, 1.0]$ if $Y = 1$; whatever the value of θ , these choices of $[L_0, R_0]$, $[L_1, R_1]$ satisfy $P[\theta < L_Y] \leq 0.025$ and $P[\theta > R_Y] \leq 0.025$. With $n = 2$ we would have $L_0 = 0$, $L_1 = 1 - \sqrt{.975}$, $L_2 = \sqrt{.025}$, and $R_j = 1 - L_{2-j}$ for $j = 0, 1, 2$.

By the normal approximation to the binomial distribution we have, approximately, $Y \sim \text{No}(n\theta, n\theta(1-\theta))$, so $t = y/n \sim \text{No}(\theta, \theta(1-\theta)/n)$; if n is large enough and θ not too close to 0 or 1 then we can justify the further approximation that $\tau^2 = \theta(1-\theta)/n \approx t(1-t)/n$, leading to approximate confidence intervals of the form $L_y \approx t - 1.96\sqrt{t(1-t)/n}$, $R_y \approx t + 1.96\sqrt{t(1-t)/n}$ with $t = y/n$, i.e.

$$0.95 \approx P[t - 1.96\tau \leq \theta \leq t + 1.96\tau]$$

where $t = y/n$ and $\tau^2 = t(1-t)/n$. This normal approximation is used far more often than the exact calculations above, in part because it requires no computation beyond a normal distribution table and a bit of arithmetic. Of course either of these methods may be used to calculate 90%, 99%, or in general $100(1-\alpha)\%$ confidence intervals for any $0 < \alpha < 1$: simply replace .025 and .975 by $\alpha/2$ and $1-\alpha/2$, and replace 1.96 by the number $z_{\alpha/2} = \text{qnorm}(1-\alpha/2)$ with probability $\alpha/2$ to its right under the standard normal distribution.

Note that these probabilities are conditional on θ and not on y , so they are quite awkward to interpret or appreciate when y is observed and θ is unknown. Two possible interpretations are:

- Before seeing the data y or t , there is (at least approximately) a 95% chance that the interval $[L_y, R_y]$ will cover the true value of θ .

- Imagine repeating the binomial experiment many times, with possibly different values of θ value; if you compute the interval for each case, then about 95% of them will cover their true θ values.

These are of little comfort in most applied statistical problems, where the data y have already been observed (so the first point above does not apply) and where we have no intention of repeating the experiment a large number of times (so the second doesn't, either). Indeed, we would argue that:

- Once y or t are observed and fixed, any probabilities assigned to possible outcomes initially are now irrelevant. A computed interval $t \pm a$ either covers the true value or not.
- The problem at hand has one θ value to be estimated, and one observed t value; the consideration of hypothetical future experiments and data sets is irrelevant to the current inference problem.

For these reasons, we approach inference via likelihood and Bayesian methods that are explicitly *conditional* in nature: inference is based on reasoning *conditional* on the observed data at hand.

In the present problem, for example, we would prefer the Bayesian reference analysis presented earlier: with a uniform prior density $p(\theta) = 1$ for θ , conditionally on the observation of $Y = y$ for a binomial random variable $Y \sim \text{Bi}(n, \theta)$ we have a Beta posterior distribution $[\theta \mid Y = y] \sim \text{Be}(y + 1, n - y + 1)$, hence have exact and approximate 95% intervals of the form

$$\begin{aligned} 0.95 &\leq \text{P}[\text{qbeta}(.025, y + 1, n - y + 1) \leq \theta \leq \text{qbeta}(.975, y + 1, n - y + 1)] \\ &\approx \text{P}[\bar{\theta}_n - 1.96 \sigma \leq \theta \leq \bar{\theta}_n + 1.96 \sigma] \end{aligned}$$

where $\bar{\theta}_n = \text{E}[\theta \mid y] = (y + 1)/(n + 2)$ is our usual point estimate of θ , the posterior mean, and where $\sigma^2 = \bar{\theta}_n(1 - \bar{\theta}_n)/(n + 3)$ is the posterior variance. That said, it is very often the case, at least in simple and standard models such as the binomial here, that the *numerical* summaries of inferences, such as likelihood or Bayesian intervals, often coincide, at least approximately, with sampling theoretic approaches— the left and right endpoints of both the exact and approximate intervals above will be quite close when y and $n - y$ are large, since then $\bar{\theta}_n \approx t = y/n$ and $\sigma^2 \approx \tau^2 = t(1 - t)/n$.

In the more complicated models that arise in statistical analysis of complex, real-world scientific problems, this is not true, however, and indeed sampling theoretic approaches rapidly become intractable, while modern simulation-based Bayesian methods are able to support scientific inference for a wide and growing range of problems.