Section 4.3. Confidence Intervals for Parameters of Discrete Distributions

Note. We now consider a sample $X_1, X_2, \ldots, X_n$ on a discrete random variable X with probability mass function $p(x; \theta)$, where $\theta \in \Omega$ and Ω is an interval of real numbers. We want a confidence interval for parameter θ.

Note 4.2.A/Definition. Let $T = T(X_1, X_2, \ldots, X_n)$ be an estimator of θ where the cumulative distribution function of T is $F_T(t; \theta)$. *Assume* that $F(t; \theta)$ is a nonincreasing and continuous function of θ for every t in the support of T. For a given realization $x_1, x_2, \ldots, x_n$ of the sample, let t be the realized value of the statistic T (so $t = T(x_1, x_2, \ldots, x_n)$). Let $\alpha_1 > 0$ and $\alpha_2 > 0$ be given such that $\alpha = \alpha_1 + \alpha_2 < 0.50$. Let $\underline{\theta}$ and $\overline{\theta}$ be the solutions of the equations

$$F_T(t-; \underline{\theta}) = 1 - \alpha_2 \text{ and } F_T(t; \overline{\theta}) = \alpha_1,$$

where $T-$ is the statistic whose support lags by one value of T's support. That is, if $y_i < t_{i+1}$ are consecutive support values of T, then $T = t_{i+1}$ if and only if $T- = t_i$. The interval $(\underline{\theta}, \overline{\theta})$ is a *confidence interval* for θ with confidence coefficient of at least $1 - \alpha$.

Note. The proof that $(\underline{\theta}, \overline{\theta})$ has the level of claimed confidence will be discussed at the end of this section.

Example 4.3.1. We now illustrate the process of solving the two equations of Note 4.2.A to determine $\underline{\theta}$ and $\overline{\theta}$ in the continuous case (of course, the "$t-$" stuff does not apply in the continuous case). Let $X_1, X_2, \ldots, X_n$ be a random sample from $N(\theta, \sigma^2)$ distribution where σ^2 is known. Let $\overline{X}$ be the sample mean and $\overline{x}$ is realized value. As argued in Example 4.2.2 (which gives an approximate $(1-\alpha)100\%$ confidence interval, but have here we know σ^2 so that there is no need for approximation and the Central Limit Theorem), a $(1-\alpha)1--\%$ confidence interval for mean θ is $(\overline{x} - z_{\alpha/2}(\sigma/\sqrt{n}), \overline{x} + z_{\alpha/2}(\sigma/\sqrt{n}))$. With θ as the true mean and $\Phi(z)$ as the cumulative distribution function of a standard normal distribution, we have the cumulative distribution function of $\overline{X}$ is $F_{\overline{X};\theta}(t) = \Phi((t-\theta)/(\sigma/\sqrt{n}))$. Then equation $F_{\overline{X};\theta}(t-;\underline{\theta}) = F_{\overline{X};\theta}(t;\theta) = 1-\alpha_2$ gives $\Phi((\overline{x}-\underline{\theta})/((\sigma/\sqrt{n})) = 1-\alpha/2$ since we have the realization $t = \overline{x}$, and $\alpha_2 = \alpha/2$. This becomes $(\overline{x}-\underline{\theta})/(\sigma/\sqrt{n}) = \Phi^{-1}(1-\alpha/2) = z_{\alpha/2}$. Solving for $\underline{\theta}$ we have $\underline{\theta} = \overline{x} - z_{\alpha/2}(\sigma/\sqrt{n})$. Equation $F_T(t;\theta) = F_{\overline{X};\theta}(t;\overline{\theta}) = \alpha_1$ gives $\Phi((\overline{x}-\overline{\theta})/(\sigma/\sqrt{n})) = \alpha/2$ since we have the realization $t = \overline{x}$, and $\alpha_1 = \alpha/2$. This becomes $(\overline{x} - \overline{X})/(\sigma/\sqrt{n}) = \Phi(\alpha/2) = -z_{\alpha/2}$. Solving for $\overline{\theta}$ we have $\overline{\theta} = \overline{x} + z_{\alpha/2}(\sigma/\sqrt{n})$. Therefore the $(1-\alpha)100\%$ confidence interval for mean θ is

$$(\underline{\theta}, \overline{\theta}) = (\overline{x} - z_{\alpha/2}(\sigma/\sqrt{n}), \overline{x} + z_{\alpha/2}(\sigma/\sqrt{n})),$$

as expected.

Note. Solving the two equations of Note 4.2.A in the discrete case will often require a numerical technique. Since $F_T(T;\overline{\theta})$ is, in practice, often strictly decreasing and continuous in θ, we apply the simple bisection algorithm. This is covered in Numerical Analysis (MATH 4257/5257); see my online notes for Numerical Analysis on Section 2.1. The Bisection Method. We briefly describe it here. Consider the equation $g(x) = d$ where g is continuous and strictly decreasing. Assume we know that $g(a) > d > g(b)$. Then by the Intermediate Value Theorem, $g(x) = d$ has some solution in the interval (a, b). Let $c = (a + b)/2$ be the midpoint of this interval. We make the following conditional replacements:

$$\text{if } g(c) > d \text{ then replace } a \text{ with } c$$
$$\text{if } g(c) < d \text{ then replace } b \text{ with } c.$$

If the new value $|a - b| < \varepsilon$, where ε is some specified tolerance, then we set $x = (a + b)/2$ so that x is within ε (actually, $\varepsilon/2$) of the exact solutions x_e where $g(x_e) = d$. Otherwise, we iterate the process with the new values of a and b. We illustrate the use of the bisection algorithm below.

Example 4.3.2. Let X have a Bernoulli distribution with θ as the probability of success. let $\Omega = (0, 1)$ Suppose $X_1, X_2, \ldots, X_n$ is a random sample on X. Notice that

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{\#\text{ successes}}{n} = \text{ sample proportion.}$$

So we use $\overline{X}$ as the estimator of θ. The cumulative density function of $n\overline{X}$ is the binomial $b(n, \theta)$. So

$$\begin{aligned}
F_{\overline{X}}(\overline{x};\theta) &= P(n\overline{X} \le n\overline{m}) \\
&= \sum_{j=0}^{n\overline{x}} \binom{n}{j} \theta^j (1-\theta)^{n-1} \text{ (notice that } n\overline{x} \text{ is an integer)} \\
&= 1 - \sum_{j=n\overline{x}+1}^{n} \binom{n}{j} \theta^j (1-\theta)^{n-j}
\end{aligned}$$

$$= \; 1 - \int_0^\theta \frac{n!}{(n\overline{x})!(n-(n-(n\overline{x}+1)!} z^{n\overline{x}}(1-z)^{n-(n\overline{x}+1)}\,dz \text{ by Exercise 4.3.6.}$$

So by the Fundamental Theorem of Calculus (Part 1) (see my online Calculus 1 [MATH 1910] on Section 5.4. The Fundamental Theorem of Calculus; notice Theorem 5.4(a)), we have

$$\frac{d}{d\theta}[F_{\overline{X}}(\overline{x};\theta)] = \frac{d}{d\theta}\left[1 - \int_0^\theta \frac{n!}{(n\overline{x})!(n-(n-(n\overline{x}+1)!} z^{n\overline{x}}(1-z)^{n-(n\overline{x}+1)}\,dz\right]$$
$$= \frac{-n!}{(n\overline{x})!(n-(n\overline{x}+1)!}\theta^{n\overline{x}}(1-\theta)^{n-(n\overline{x}+1)} < 0.$$

Since this derivative is negative then $F_{\overline{X}}(\overline{x};\theta)$ is a strictly decreasing function of θ (for each $\overline{x}$). Now let $\alpha_1, \alpha_2 > 0$ where $\alpha - 1 + \alpha_2 < 1/2$ be given and let $\underline{\theta}$ and $\overline{\theta}$ be solutions to the equations

$$F_{\overline{X}}(\overline{x}-;\theta) = 1 - \alpha_2 \text{ and } F_{\overline{X}}(\overline{x};\overline{\theta}) = \alpha_1.$$

Then, as described above, $(\underline{\theta}, \overline{\theta})$ is a confidence interval for θ with confidence coefficient at least $1-\alpha$ where $\alpha = \alpha_1 + \alpha_2$. We next illustrate a numerical solution with specific numbers.

Example 4.3.2 (continued). Let X have a Bernoulli distribution with θ as the probability of a success. Suppose a sample of size $n = 30$ is taken with realizes sample mean $\overline{x} = 0.60$ (so that $n\overline{x} = 18$). Let $\alpha_1 = \alpha_2 = 0.05$. With $n\overline{x} = 18$ we have $n\overline{x}- = 17$, the equations of Example 4.3.2 become

$$F_{\overline{X}}(\overline{x}-;\theta) = \sum_{j=0}^{17} \binom{30}{j} \underline{\theta}^j (1-\underline{\theta})^{n-j} = 0.95$$

and

$$F_{\overline{X}}(\overline{x};\theta) = \sum_{j=0}^{18} \binom{30}{j} \overline{\theta}^j (1-\overline{\theta})^{n-j} = 0.05.$$

We borrow some information from Hogg, McKean, and Craig and use the Binomial Probability Calculator of StatTrek.com. With $\theta = 0.4$ (as the text suggests) we have $F_{\overline{X}}(\overline{x}-;\theta) = 0.9788 > 0.95$ (to four decimal places) and with $\theta = 0.45$ (also suggested by the text) we have $F_{\overline{X}}(\overline{x}-;\theta) = 0.9286 < 0.95$. So we must have $\underline{\theta}$ that yields $F_{\overline{X}}(\overline{x}-;\underline{\theta}) = 095$ is between 0.40 and 0.45. We now apply the bisection algorithm to find this $\underline{\theta}$.

STEP	a	b	$c = (a+b)/2$	$g(a)$	$g(b)$	$g(c)$
1	0.4	0.45	0.425	0.9788	0.9286	0.9595
2	0.425	0.45	0.4375	0.9595	0.9286	0.9458
3	0.425	0.4375	0.4313	0.9595	0.9458	0.9530
4	0.4313	0.4375	0.4344	0.9530	0.9458	0.9495
5	0.4313	0.4344	0.4329	0.9530	0.9495	0.9512
6	0.4329	0.4344	0.4337	0.9512	0.9495	0.9503
7	0.4337	0.4344	0.4341	0.9603	0.9495	0.9498
8	0.4337	0.4341	0.4339	0.9503	0.9498	0.9499
9	0.4339	0.4341	0.4340	0.9501	0.9498	0.9499
10	0.4339	0.4340	—	0.9501	0.9499	—

So we take $\underline{\theta}$ as 0.4339 or 0.4340 (we could keep up with more decimal places to get a decision between these two); notice that Hogg, Mckean, and Craig give 0.4339417 (which is very much in agreement with our computations). To three decimal places we have 0.434 (as does Hogg, McKean, and Craig). Similarly (again borrowing from Hogg, Mckean, and Craig) with $\overline{\theta} = 0.7$ we have $F_{\overline{X}}(\overline{x};\theta) = 0.1593 > 0.05$ and with $\overline{\theta} = 0.8$ we have $F_{\overline{X}}(\overline{x};\theta) = 0.0095 < 0.05$. So we must have $\overline{\theta}$ that yields $F_{\overline{X}}(\overline{x};\overline{\theta}) = 0.05$ is between 0.7 and 0.8. The bisection method similarly (and tediously, if done by hand) leads to $\overline{\theta} = 0.750$ (to three decimal places). So the confidence interval for proportion θ is $(\underline{\theta}, \overline{\theta}) = (0.434, 0.750)$, where the level of confidence is at least $(1-\alpha)100^{\%} = 90\%$.

Example 4.3.3. Let $X_1, X_2, \ldots, X_n$ be a random sample on Poisson distributed random variable X with mean θ. We take $\overline{X} = \frac{1}{n}\sum_{i=1}^n X_i$ as the estimator of θ. As in Example 4.3.2, $n\overline{X}$ has a Poisson distribution with mean $n\theta$. The cumulative distribution function of $\overline{X}$ is then

$$F_{\overline{X}}(\overline{x};\theta) = \sum_{j=0}^{n\overline{x}} e^{-n\theta}\frac{(n\theta)^j}{j!} = \frac{1}{\Gamma(n\overline{x})}\int_{n\theta}^{\infty} x^{n\overline{x}}e^{-x}\,dx \text{ by Exercise 4.3.7.}$$

By the Fundamental Theorem of Calculus (Part 1),

$$\frac{d}{d\theta}[F_{\overline{X}}(\overline{x};\theta)] = \frac{d}{d\theta}\left[\frac{1}{\Gamma(m\overline{x})}\int_{n\theta}^{\infty} x^{n\overline{x}}e^{-x}\,dx\right] = \frac{1}{\Gamma(n\overline{x})}\frac{d}{d\theta}\left[-\int_{\infty}^{n\theta} x^{n\overline{x}}e^{-x}\,dx\right]$$

$$= \frac{-1}{\Gamma(n\overline{x})}(n\theta)^{n\overline{x}}e^{-(n\theta)}[n] = \frac{-n}{\Gamma(n\overline{x})}(n\theta)^{n\overline{x}}e^{-n\theta} < 0.$$

Therefore $F_{\overline{X}}(\overline{x};\theta)$ is a strictly decreasing function of θ for every given $\overline{x}$. With $\overline{x}$ realized from the sample, for $\alpha_1, \alpha_2 > 0$ such that $\alpha_1 + \alpha_2 < 1/2$, the confidence interval for θ is $(\underline{\theta}, \overline{\theta})$ where

$$F_{\overline{X}}(\overline{x}-;\theta) = \sum_{j=0}^{n\overline{x}-1} e^{-n\underline{\theta}}\frac{(n\underline{\theta})^j}{j!} = 1-\alpha_2$$

and

$$F_{\overline{X}}(\overline{x};\overline{\theta}) = \sum_{j=1}^{n\overline{x}} e^{-n\overline{\theta}}\frac{(n\overline{\theta})^j}{j!} = \alpha_1.$$

The confidence is at least $1-\alpha = 1-(\alpha_1+\alpha_2)$. We again give a numerical solution.

Example 4.3.3 (continued). Suppose now that a sample is taken from a Poisson distributed random variable X with (unknown) mean θ. Let the sample size be $n = 25$ and let the realized value of $\overline{X}$ be $\overline{x} = 5$. Then $n\overline{x} = 125$. Let $\alpha_1 = \alpha_2 = 0.5$. Then the equation of Example 4.3.3 are

$$\sum_{j=0}^{124} e^{-n\underline{\theta}} \frac{(n\underline{\theta})^j}{j!} = 0.95 \text{ and } \sum_{j=0}^{125} e^{-n\overline{\theta}} \frac{(n\overline{\theta})^j}{j!} = 0.05.$$

We borrow some information from Hogg, Mckean, and Craig and using the Poisson Distribution Calculator of StatTrek.com. With $\theta = 4.0$, $n\theta = (25)(4) = 100$, and $n\overline{x} = (25)(5) = 125$ we have

$$F_{\overline{X}}(\overline{x}-;\theta) = \sum_{j=0}^{124} e^{-100} \frac{(100)^j}{j!} = 0.9912 > 0.95$$

(there is an error in Hogg, McKean, and Craig; they give a value of 0.9932, which is based on summing to 125 not 124), and with $\theta = 4.4$, $n\theta = (25)(4.4) = 110$, and $n\overline{x} = 125$ we have

$$F_{\overline{X}}(\overline{x}-;\theta) = \sum_{j=0}^{124} e^{-110} \frac{(110)^j}{j!} = 0.9145 < 0.95.$$

So we must have $\underline{\theta}$ that yields $F_{\overline{X}}(\overline{x}-;\theta) = 0.95$ is between 4.0 and 4.4. We now apply the bisection algorithm to find this $\underline{\theta}$ (see below). Based on these computations, we take $\underline{\theta}$ as 4.2878. Similarly (again borrowing from Hogg, McKean, and Craig) with $\theta = 5.5$ we have $F_{\overline{X}}(\overline{x};\theta) = 0.1528 > 0.5$ and with $\theta = 6.0$ we have $F_{\overline{X}}(\overline{x};\theta) = 0.0204 < 0$. So we must have $\overline{\theta}$ that yields $F_{\overline{X}}(\overline{x};\theta) = 0/05$ is between 5.5 and .0. The bisection method similarly leads to $\overline{\theta} = 5.8006$. So the confidence interval for proportion θ is $(\underline{\theta}, \overline{\theta}) = (4.287, 5.800)$, where the level of confidence is at least $(1-\alpha)100^{\%} = 90\%$.

STEP	a	b	$c = (a+b)/2$	$g(a)$	$g(b)$	$g(c)$
1	4.0	4.4	4.2	0.9912	0.9145	0.9688
2	4.2	4.4	4.3	0.9688	0.9145	0.9468
3	4.2	4.3	4.25	0.9688	0.9468	0/9590
4	4.25	4.3	4.275	0.9590	0.9468	0.9532
5	4.275	4.3	4.2875	0.9532	0.9468	0.9501
6	4.2875	4.3	4.2938	0.9501	0.9468	0.9485
7	4.2875	4.2938	4.2907	09501	0.9485	0.9493
8	4.2875	4.2907	4.2891	0.9501	0.9493	0.9497
9	4.2875	4.2891	4.2883	0.9501	0.9497	0.9499
10	4.2875	4.2883	4.2879	0.9501	0.9499	0.9500
11	4.2875	4.2879	4.2877	0.9501	0.9500	0.9500
12	4.2877	4.2879	4.2878	0.9500	0.9500	09500

Note. We now formalize the results of this section in a theorem and offer a "brief sketch" of the proof.

Theorem 4.2.A. Consider a sample $X_1, X_2, \ldots, X_n$ on a discrete random variable X with probability mass function $p(x;\theta)$, where $\theta \in \Omega$ and Ω is an interval of real numbers. Let $T = T(X_1, X_2, \ldots, X_n)$ be an estimator of θ where the cumulative distribution function of T is $F_T(t;\theta)$. Suppose that $F(t;\theta)$ is a nonincreasing and continuous function of θ for every t in the support of T. For a given realization $x_1, x_2, \ldots, x_n$ of the sample, let t be the realized value of the statistic T (so $t =$

$T(x_1, x_2, \ldots, x_n))$. Let $\alpha_1 > 0$ and $\alpha_2 > 0$ be given such that $\alpha = \alpha_1 + \alpha_2 < 0.50$. Let $\underline{\theta}$ and $\overline{\theta}$ be the solutions of the equations

$$F_T(t-; \underline{\theta}) = 1 - \alpha_2 \text{ and } F_T(t; \overline{\theta}) = \alpha_1,$$

where $T-$ is the statistic whose support lags by one value of T's support. The interval $(\underline{\theta}, \overline{\theta})$ is a *confidence interval* for θ with confidence coefficient of at least $1 - \alpha$.

Revised: 3/15/2022