

Confidence Intervals: Lecture XXII

Charles B. Moss

October 12, 2026

Just Because - Bayes

- Examination II - 2006 has a question that is a little hard to make out. The parenthesis have vanished. The point is that we assume a prior distribution for p , the parameter in a Bernoulli distribution of

$$g(p) = p(1 - p) \quad (1)$$

First, this prior is an improper prior, it does not integrate to one

$$\int_0^1 p(1 - p) dp = \int_0^1 [p - p^2] dp = \frac{1}{6} \quad (2)$$

- The trick with this formulation is the that the constant of the improper distribution cancels out in the posterior

$$\begin{aligned} f(p|x) &= \frac{6p^x (1-p)^{1-x} [p(1-p)]}{6 \int_0^1 p^x (1-p)^{1-x} [p(1-p)] dp} \\ &= \frac{p^x (1-p)^{1-x} [p(1-p)]}{\int_0^1 p^x (1-p)^{1-x} [p(1-p)] dp} \\ &= \frac{(1-p)^{2-x} p^{1+x}}{\int_0^1 (1-p)^{2-x} p^{1+x} dp} \end{aligned} \tag{3}$$

Graph of Prior

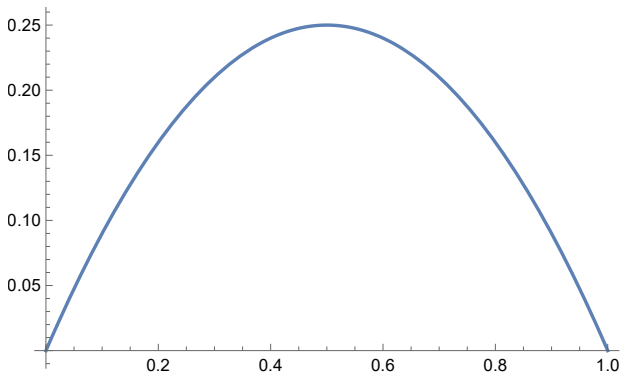


Figure: Graph of Improper Prior

- $E[p] = 1/2$, $V[p] = 1/20$.

Using Mathematica or Copilot

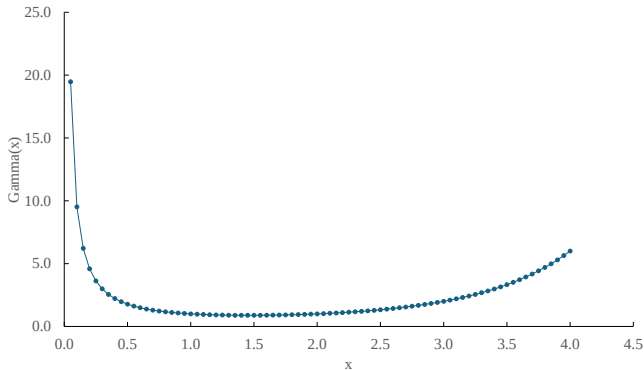
- The integral in the numerator

$$\int_0^1 (1-p)^{2-x} p^{1+x} dp \quad (4)$$

is solved by transforming the parameters into a Beta function

$$\begin{aligned} B(\alpha, \beta) &= \int_0^1 t^{\alpha-1} (1-t)^{\beta-1} dt \\ &= \frac{\Gamma(\alpha) \Gamma(\beta)}{\Gamma(\alpha + \beta)} \\ \Gamma(\alpha) &= \int_0^\infty t^{\alpha-1} e^{-t} dt \end{aligned} \quad (5)$$

The Gamma Function



- First, we rewrite Equation 4 to be in the form of the definition of the Beta function in Equation 5

$$p^{1+x} (1-p)^{2-x} \Rightarrow \begin{cases} p \rightarrow t \\ \alpha - 1 = 1 - x \rightarrow \alpha = 2 + x \\ \beta - 1 = 2 - x \rightarrow \beta = 3 - x \end{cases} \quad (6)$$

Substituting these results into Equation 5

$$\begin{aligned} \int_0^1 p^{1+x} (1-p)^{2-x} &\Rightarrow \frac{\Gamma(2+x) \Gamma(3-x)}{\Gamma([2+x] + [3-x])} \\ &= \frac{\Gamma(2+x) \Gamma(3-x)}{\Gamma[5]} \end{aligned} \quad (7)$$

- Note that $\Gamma[x] = (x - 1)!$ for integer x . Thus, $\Gamma[5] = (5 - 1)! = 24$. Therefore,

$$f(p|x) = \frac{24(1-p)^{2-x}p^{1+x}}{\Gamma(3-x)\Gamma(2+x)} \quad (8)$$

- Suppose that your sample was a tail $x = 0$. The posterior distribution would become

$$f(p|x=0) = 12(1-p)^2p^1 \quad (9)$$

- The expected value of p is then

$$E[p] = 12 \int_0^1 p(1-p)^2p^1dp = \frac{2}{5} \quad (10)$$

1 Interval Estimation

2 Confidence Intervals

Interval Estimation

- As we discussed when we talked about continuous distribution functions, the probability of a specific number under a continuous distribution is zero.
- Thus, if we conceptualize any estimator, either a nonparametric estimate of the mean or a parametric estimate of a function, the probability of the true value equal to the estimated value is obviously zero.
- Thus, usually talk about estimated values in terms of confidence intervals. Specifically, as in the case when we discussed the probability of a continuous variable, we define some range of outcomes. However, this time we usually work the other way around defining a certain confidence level and then stating the values that contain this confidence interval.

Confidence Intervals

- Amemiya notes a difference between confidence and probability. Most troubling is our classic definition of probability as "a probabilistic statement involving parameters." This is troublesome due to our inability without some additional Bayesian structure to state anything concrete about probabilities.
- Example 8.2.1.** Let X_i be distributed as a Bernoulli distribution, $i = 1, 2, \dots, n$. Then,

$$T = \bar{X} \stackrel{A}{\sim} N\left(p, \frac{p(1-p)}{n}\right) \quad (11)$$

Therefore, we have

$$Z = \frac{T - p}{\sqrt{\frac{p(1-p)}{n}}} \stackrel{A}{\sim} N(0, 1) \quad (12)$$

- Why? By the Central Limit Theory.
- Given this distribution, we can ask questions about the probability. Specifically, we know that if Z is distributed $N(0, 1)$, then we can define

$$\gamma_k = P(|Z| < k) \quad (13)$$

Building on the normal probability, the one tailed probabilities for the normal distribution are presented in Table 1.

Table: Confidence Levels

k	$\gamma/2$	γ
1.0000	0.1587	0.3173
1.5000	0.0668	0.1336
1.6449	0.0500	0.1000
1.7500	0.0401	0.0801
1.9600	0.0250	0.0500
2.0000	0.0228	0.0455
2.3263	0.0100	0.0200

- The values of γ_k can be derived from the standard normal table as

$$P \left[\frac{|T - p|}{\sqrt{\frac{p(1-p)}{n}}} < k \right] = \gamma_k \quad (14)$$

- Assuming that the sample of T is t , the confidence interval is defined by

$$C \left[\frac{|t - p|}{\sqrt{\frac{p(1-p)}{n}}} < k \right] = \gamma_k \quad (15)$$

- Building on the first term

$$\begin{aligned} P \left[\frac{|T - p|}{\sqrt{\frac{p(1-p)}{n}}} < k \right] &= P \left[|T - p| < k \sqrt{\frac{p(1-p)}{n}} \right] \\ &= P \left[(T - p)^2 < k^2 \frac{p(1-p)}{n} \right] \\ &= P \left[T^2 - 2Tp + p^2 - \frac{k^2}{n}p + \frac{k^2}{n}p^2 < 0 \right] \\ &= P \left[p^2 \left(1 + \frac{k^2}{n} \right) - p \left(2T + \frac{k^2}{n} \right) + T^2 < 0 \right] < \gamma_k \end{aligned} \tag{16}$$

Using this probability, it is possible to define two numbers $h_1(T)$ and $h_2(T)$ for which this inequality holds.

- Mathematically, applying the quadratic equation

$$P[h_1(T) < p < h_2(T)] \leq \gamma_k \text{ where}$$
$$h_1(T), h_2(T) = \frac{\left(2T + \frac{k^2}{n}\right) \pm \sqrt{\left(2T + \frac{k^2}{n}\right)^2 - 4\left(1 + \frac{k^2}{n}\right)T^2}}{2\left(1 + \frac{k^2}{n}\right)} \quad (17)$$

- For the following table, let S be the sample estimate of the variance $S = \sum_{i=1}^n (X_i - \bar{X})^2 / n$. Thus, we are comparing the confidence interval based on a sample measure with the “true” value of the variance based on the parameter.

Table: Sample Statistics for T for 3 Draws

Draws	Statistic	Draw 1		Draw 2		Draw 3	
		$p = 0.40$	$p = 0.60$	$p = 0.40$	$p = 0.60$	$p = 0.40$	$p = 0.60$
50	T	0.300	0.500	0.480	0.600	0.480	0.660
	S	0.463	0.505	0.505	0.495	0.505	0.479
	$T(1 - T)/\sqrt{50}$	0.030	0.035	0.035	0.034	0.035	0.032
100	T	0.320	0.500	0.430	0.590	0.380	0.590
	S	0.469	0.503	0.498	0.494	0.488	0.494
	$T(1 - T)/\sqrt{100}$	0.022	0.025	0.025	0.024	0.024	0.024
150	T	0.347	0.547	0.393	0.587	0.380	0.593
	S	0.478	0.499	0.490	0.494	0.487	0.493
	$T(1 - T)/\sqrt{150}$	0.018	0.020	0.019	0.020	0.019	0.020
200	T	0.360	0.570	0.400	0.605	0.380	0.575
	S	0.481	0.496	0.491	0.490	0.487	0.496
	$T(1 - T)/\sqrt{200}$	0.016	0.017	0.017	0.017	0.017	0.017

Mathematically derived Bounds

- From Equation 17 we compute two parts of the confidence interval, the linear part

$$\text{linear} = \frac{2t + \frac{k^2}{N}}{2 \left(1 + \frac{k^2}{N}\right)} \quad (18)$$

the quadratic part

$$\text{quadratic} = \frac{\sqrt{\left(2t + \frac{k^2}{N}\right)^2 - 4 \left(1 + \frac{k^2}{N}\right) t^2}}{2 \left(1 + \frac{k^2}{N}\right)} \quad (19)$$

Mathematically derived Bound - continued

- The upper confidence bound is then $\text{linear} + \sqrt{\text{quadratic}}$
- The lower confidence bound is then $\text{linear} - \sqrt{\text{quadratic}}$

Table: Empirical Confidence Intervals for Samples

Draws	Statistic	Draw 1		Draw 2		Draw 3	
		$p = 0.40$	$p = 0.60$	$p = 0.40$	$p = 0.60$	$p = 0.40$	$p = 0.60$
50	Linear	0.3143	0.5000	0.4814	0.5929	0.4814	0.6486
	Square Root	0.1232	0.1336	0.1335	0.1311	0.1335	0.1270
	Lower Bound	0.1910	0.3664	0.3480	0.4618	0.3480	0.5215
	Upper Bound	0.4375	0.6336	0.6149	0.7239	0.6149	0.7756
100	Linear	0.3267	0.5000	0.4326	0.5867	0.3844	0.5867
	Square Root	0.0900	0.0962	0.0953	0.0947	0.0935	0.0947
	Lower Bound	0.2367	0.4038	0.3373	0.4920	0.2910	0.4920
	Upper Bound	0.4166	0.5962	0.5278	0.6813	0.4779	0.6813
150	Linear	0.3505	0.5455	0.3960	0.5845	0.3830	0.5910
	Square Root	0.0753	0.0787	0.0772	0.0778	0.0768	0.0777
	Lower Bound	0.2752	0.4668	0.3188	0.5067	0.3062	0.5133
	Upper Bound	0.4258	0.6242	0.4732	0.6623	0.4598	0.6687
200	Linear	0.3626	0.5687	0.4019	0.6030	0.3823	0.5736
	Square Root	0.0659	0.0680	0.0673	0.0671	0.0667	0.0679
	Lower Bound	0.2967	0.5007	0.3346	0.5359	0.3156	0.5057
	Upper Bound	0.4286	0.6367	0.4692	0.6702	0.4489	0.6415

- **Example 8.2.2** Let $X_i \sim N(\mu, \sigma^2)$, $i = 1, 2, \dots, n$ where μ is unknown and σ^2 is known. We have

$$T = \bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \quad (20)$$

Define

$$P\left[\frac{|T - \mu|}{\sqrt{\frac{\sigma^2}{n}}} < k\right] = \gamma_k. \quad (21)$$

- **Example 8.2.3** Suppose that $X_i \sim N(\mu, \sigma^2)$, $i = 1, 2, \dots, n$ with both μ and σ^2 unknown. Let

$$T = \bar{X} \quad (22)$$

be an estimator of μ and

$$S^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \quad (23)$$

be the estimator of σ^2 . Then the probability distribution

$$t_{n-1} = S^{-1} (T - 1) \sqrt{n-1} \quad (24)$$

or the *Student's t distribution with $n-1$ degrees of freedom* is the probability distribution.

- **Theorem 5.4.1:** Let $X_1, X_2, \dots, X_n$ be a random sample from a $N(\mu, \sigma^2)$ distribution, and let

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \text{ and } S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2. \quad (25)$$

Then,

- $\bar{X}$ and S^2 are independent random variables.
- $\bar{X} \sim N(\mu, \sigma^2/n)$
- $(N-1) S^2/\sigma^2$ has a chi-squared distribution with $n-1$ degrees of freedom.

- The proof of independence is based on the fact that S^2 is a function of the deviations from the mean which, by definition, must be independent of the mean.
- More interesting, is the discussion of the chi-squared statistic.
 - The chi-square distribution is defined as

$$f(x) = \frac{1}{\Gamma\left(\frac{p}{2}\right) 2^{p/2}} x^{\frac{p}{2}-1} e^{-x/2} \quad (26)$$

- In general, the gamma distribution function is defined through the gamma function

$$\Gamma(\alpha) = \int_0^{\infty} t^{\alpha-1} e^{-t} dt \quad (27)$$

- Dividing both sides of the expression by $\Gamma(\alpha)$ yields

$$1 = \frac{1}{\Gamma(\alpha)} \int_0^{\infty} t^{\alpha-1} e^{-t} dt \Rightarrow f(t) = \frac{t^{\alpha-1} e^{-t}}{\Gamma(\alpha)} \quad (28)$$

Substituting $X = \beta t$ gives the traditional two parameter form of the distribution function

$$f(x|\alpha, \beta) = \frac{1}{\Gamma(\alpha) \beta^{\alpha}} x^{\alpha-1} e^{-x/\beta} \quad (29)$$

The expected value of the gamma distribution is $\alpha\beta$ and the variance is $\alpha\beta^2$.

- **Lemma 5.4.1.** (Facts about chi-squared random variables)

We use the notation χ_p^2 to denote a chi-squared random variable with p degrees of freedom.

- If Z is a $N(0, 1)$ random variable, then $Z^2 \sim \chi_1^2$, that is, the square of a standard normal random variable is a chi-squared random variable.
- If $X_1, X_2, \dots, X_n$ are independent, and $X_i^2 \sim \chi_{p_i}^2$, then $X_1^2 + X_2^2 + \dots + X_n^2 \sim \chi_{p_1+p_2+\dots+p_n}^2$, that is, independent chi-squared variables add to a chi-square variable, and the degrees of freedom also add.
- The first part of the Lemma follows from the transformation of random variables for $Y = X^2$ which yields

$$f_Y(y) = \frac{1}{2\sqrt{y}} (f_X(\sqrt{y}) + f_X(-\sqrt{y})) \quad (30)$$

Two Observation Sample

- Let's start with a two observation sample

$$\begin{aligned} S_2^2 &= (X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 \\ &= \left(\frac{2X_1 - X_1 - X_2}{2} \right)^2 + \left(\frac{2X_2 - X_1 - X_2}{2} \right)^2 \\ &= \left(\frac{X_1 - X_2}{2} \right)^2 + \left(\frac{X_2 - X_1}{2} \right)^2 = \frac{(X_1 - X_2)^2}{2} \end{aligned} \quad (31)$$

- To finish the derivation, we multiply S_n^2 by $n/(n-1)$

$$S_2^2 = \frac{2}{2-1} \frac{(X_1 - X_2)^2}{2} = (X_1 - X_2)^2 \quad (32)$$

Decomposing the Variance

- Starting with small sample variance

$$\begin{aligned}(n-1)S_n^2 &= \sum_{i=1}^n \left(x_i - \frac{1}{n} \sum_{j=1}^n x_j \right)^2 \\ &= \sum_{i=1}^n \left(x_i - \frac{1}{n} \sum_{j=1}^{n-1} x_j - \frac{1}{n} x_n \right)^2 \\ &= \left(x_n - \frac{1}{n} \sum_{j=1}^{n-1} x_j - \frac{1}{n} x_n \right)^2 + \sum_{i=1}^{n-1} \left(x_i - \frac{1}{n} \sum_{j=1}^{n-1} x_j - \frac{1}{n} x_n \right)^2\end{aligned}\tag{33}$$

- Working on the first term first

$$\begin{aligned} \left(x_n - \frac{1}{n} \sum_{j=1}^{n-1} x_j - \frac{1}{n} x_n \right)^2 &= \left(\frac{n-1}{n} x_n - \frac{(n-1)}{(n-1)} \frac{1}{n} \sum_{j=1}^{n-1} x_j \right)^2 \\ &= \left(\frac{n-1}{n} \right)^2 (x_n - \bar{x}_{n-1})^2 \end{aligned} \quad (34)$$

- Taking the second term

$$\begin{aligned} \sum_{i=1}^{n-1} \left(x_i - \frac{1}{n} \sum_{j=1}^{n-1} x_j - \frac{1}{n} x_n \right)^2 &= \\ \sum_{i=1}^{n-1} \left(\frac{nx_i - (n-1)\bar{x}_{n-1}}{n} - \frac{1}{n} x_n \right)^2 & \end{aligned} \quad (35)$$

- Of course there is a trick

$$\frac{1}{n} \sum_{j=1}^{n-1} x_j = \frac{1}{n} (n-1) \bar{x}_{n-1} \quad (36)$$

- Hence, the second term becomes

$$\sum_{i=1}^{n-1} \left[(x_i - \bar{x}_{n-1}) - \frac{1}{n} (\bar{x}_{n-1} - x_n) \right]^2 = (n-2) S_{n-1}^2 \quad (37)$$

- Putting the bits together

$$(n-1) S_n^2 = (n-2) S_{n-1}^2 + \left(\frac{n-1}{n} \right) (x_n - \bar{x}_{n-1})^2 \quad (38)$$

- If $n = 2$, we get

$$S_2^2 = \frac{1}{2} (X_2 - X_1)^2 \quad (39)$$

Given $(X_2 - X_1) / \sqrt{2}$ is distributed $N(0, 1)$, $S_2^2 \sim \chi_1^2$ and by extension for $n = k$, $(k-1) S_k^2 \sim \chi_{k-1}^2$.

- Given these results for the chi-square, the distribution of the Student's t then follows.

$$\frac{\frac{\bar{X} - \mu}{S/\sqrt{n}}}{\sqrt{\frac{S^2}{\sigma^2}}} = \frac{(\bar{X} - \mu)}{\sigma/\sqrt{n}} \quad (40)$$

Note that this creates a standard normal random variable in the numerator and

$$\sqrt{\frac{\chi_{n-1}^2}{n-1}} \quad (41)$$

- in the denominator. The complete distribution found by multiplying the standard normal time the chi-squared distribution times the Jacobian of the transformation yields

$$f_T(t) = \frac{\Gamma\left(\frac{p+2}{2}\right)}{\Gamma\left(\frac{p}{2}\right) \frac{1}{(p\pi)^{1/2}}} \frac{1}{(1+t^2/p)^{\frac{p+1}{2}}} \quad (42)$$

Reading

- Moss 8.2
- Salsburg, David “The Lady Tasting Tea” Chapter 2 – The Skew Distributions