Inferences Based on a Single Sample

Some background material [additional non-examinable notes] on an introduction to statistical inference is available at
"`http://www.engr.mun.ca/~ggeorge/3423/handout/H10aInfer.doc`".

Confidence Intervals

- Almost any parameter that we might wish to estimate has its set of possible values.
- The point estimate, a single number, can be replaced by an entire interval of plausible values. This interval is the **confidence interval**.
- A confidence interval is an interval of plausible values for the parameter being estimated. The degree of plausibility is specified by a confidence level, such as 95% or 99%.

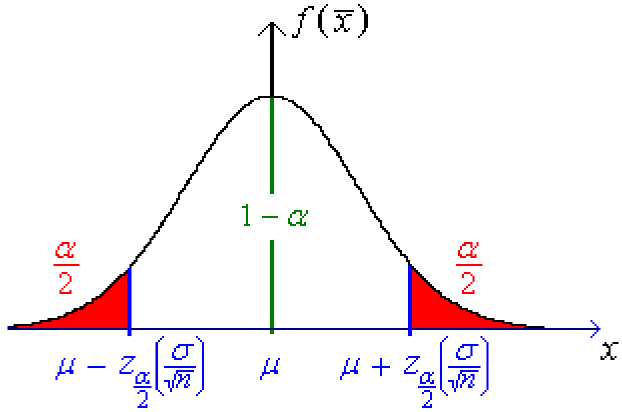
Calculation of the classical confidence interval

Suppose that the parameter of interest is a population mean μ and that

- the population distribution is normal, and
- the value of the population standard deviation σ is known.

If $X_i \sim N(\mu, \sigma^2)$, then $\overline{X} \sim \mathrm{N}\left(\mu, \dfrac{\sigma^2}{n}\right)$

and $Z = \dfrac{\overline{X} - \mu}{\left(\frac{\sigma}{\sqrt{n}}\right)} \sim \mathrm{N}(0, 1)$



Because the area under the normal curve between

$\mu - z_{\alpha/2}\left(\dfrac{\sigma}{\sqrt{n}}\right)$ and $\mu + z_{\alpha/2}\left(\dfrac{\sigma}{\sqrt{n}}\right)$ is $1 - \alpha$,

$$\mathrm{P}\left[\mu - z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right) < \overline{X} < \mu + z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)\right] = 1 - \alpha$$

Subtract $\mu, \overline{X}$:

$$P\left[-\overline{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < -\mu < -\overline{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$$

Multiply by (−1):

$$P\left[+\overline{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}} > +\mu > +\overline{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$$

Rearrange the inequalities to obtain:

The **confidence interval estimator** for μ (at a level of confidence of $(1 - \alpha)$) is

$$\overline{X} - z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right) < \mu < \overline{X} + z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)$$

The $(1 - \alpha)$ confidence interval estimator for μ is a random interval

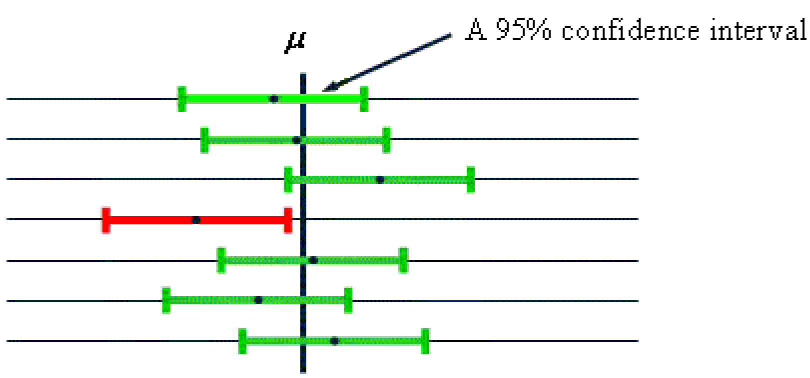
$$\left[\overline{X} - z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right), \overline{X} + z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)\right]$$

The probability is $(1 - \alpha)$ that the above random interval includes the true value of μ. . $(1 - \alpha)$ of all random samples will produce an inequality, (the $(1 - \alpha)$ **confidence interval estimate** for μ)

$$\bar{x} - z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right) < \mu < \bar{x} + z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)$$

that is true. Note that the confidence interval estimate contains no random quantities at all! The statement is either absolutely certain to be true or absolutely certain to be false, (depending on the values of μ, σ, $\bar{x}$, n and α).

Interpretation of confidence interval [= confidence interval **estimate**]



Only 5% of all 95% confidence interval estimates for μ fail to include μ.

A concise expression for the C.I. (confidence interval estimate for μ) is

$$\bar{x} \;\pm\; z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

A Bayesian view of interval estimation:

If the only quantity among { μ, σ, $\bar{x}$, n and α } that we don't know is μ, then represent the unknown μ by the random quantity A. Then

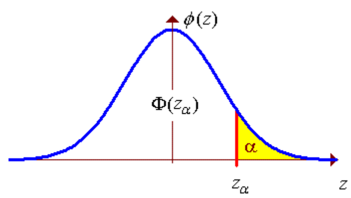
$$\mathrm{P}\left[-z_{\alpha/2} < \frac{\bar{x} - A}{\left(\frac{\sigma}{\sqrt{n}}\right)} < +z_{\alpha/2} \right] = 1 - \alpha$$

$$\rightarrow \qquad \mathrm{P}\left[\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < A < \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \right] = 1 - \alpha$$

which is a valid probability statement about the random quantity A (→ decision theory).

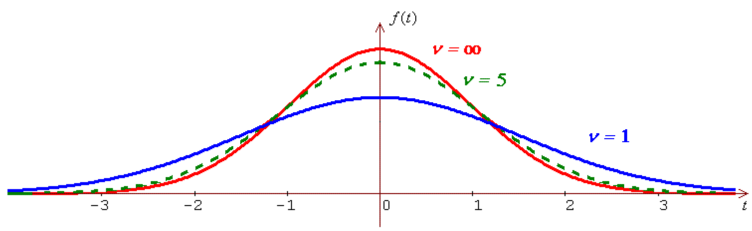
A note about the standard normal distribution and the *t* distribution

Let $Z \sim$ N(0, 1) (standard normal distribution), so that $\mathrm{P}[Z \le z] = \Phi(z)$ (cumulative distribution function for the standard normal distribution).
Then the $(1 - \alpha)\times 100^{\text{th}}$ percentile of the standard normal distribution is z_α , which satisfies $\mathrm{P}[Z > z_\alpha] = \alpha$.
It also follows that $1 - \Phi(z_\alpha) = \Phi(-z_\alpha) = \alpha$.



$\Phi(z_\alpha) = 1 - \alpha$

The t distribution with v degrees of freedom is also a bell shaped curve, with a mean, median and mode at $t = 0$, but with a greater variance than the standard normal distribution. As the number of degrees of freedom increases, the t distribution approaches the z (standard normal) distribution. The graphs of t_1 and t_5 are shown here, together with $\phi(z)$, which is indistinguishable to the eye from t_ν for ν above 30 or so.



Therefore $\lim_{\nu \to \infty} t_{\alpha,\nu} = t_{\alpha,\infty} = z_\alpha$.

To find the $(1 - \alpha)\times 100^{\text{th}}$ percentile z_α , use the final row in the table of critical values of the t distribution (on page 15-02): $z_\alpha = t_{\alpha,\infty}$.

The final row of the table on page 15-02 (the t tables) is

	α				
ν	**0.1**	**0.05**	**0.025**	**0.01**	**0.005**
∞	**1.28155**	**1.64485**	**1.95996**	**2.32635**	**2.57583**

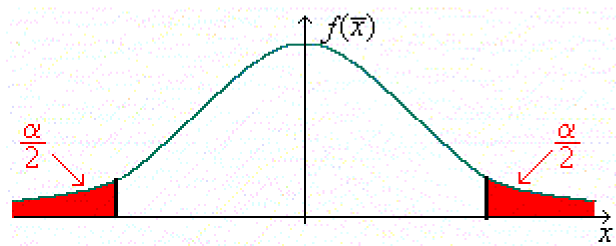
Therefore $P[Z > 1.645] \approx .05$ or equivalently $z_{.05} \approx 1.645$;
$P[Z > 1.960] \approx .025$ or equivalently $z_{.025} \approx 1.960$; etc.

Example 10.01

The rate of energy loss X (watt) in a motor is known to be a normally distributed random quantity with standard deviation $\sigma = 3.0$ W. A random sample of 100 such motors produces a sample mean rate of energy loss of 58.3 W. Find a 99% confidence interval estimate for the true mean rate of energy loss μ.

Endpoints of classical CI:

$$\bar{x} \pm z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$



$$\frac{\alpha}{2} = \frac{1}{2}(1 - .99) = .005$$

$$z_{.005} = t_{.005,\infty} = \mathbf{2.57583}$$

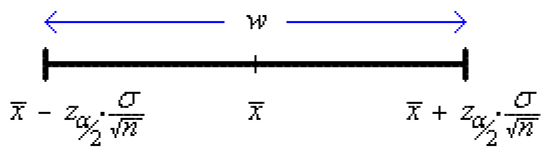
Therefore the endpoints are $58.3 \pm 2.57583 \times \left(\frac{3.0}{\sqrt{100}}\right) = 58.3 \pm 0.772\ldots$

⇒ **99% CI for μ is** <u>**$57.53 \le \mu \le 59.07$**</u> **(W) (to 2 d.p.)**

[Reason for "≤" instead of "<": The stated endpoints are just inside the exact interval. The exact lower limit is 57.527..., which places 57.53 inside. The answer could also be quoted correctly as "$57.52 < \mu < 59.08$" or "$57.53 \le \mu < 59.08$".]

Example 10.01 (continued)

How large must n be for the width of the 99% confidence interval estimate for μ to be less than 1.0?



$$w = \left(\bar{x} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right) - \left(\bar{x} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right) = 2z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

$$\therefore \quad w = 2 \times 2.57\ldots \times \frac{3.0}{\sqrt{n}}$$

We require $w < 1 \quad \Rightarrow \quad 15.45498 \le \sqrt{n}$

$\Rightarrow \quad n \ge 238.8...$

Therefore $n_{\min} = \underline{\underline{239}}$.

Choice of sample size

The width of the confidence interval $\left(\bar{x} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}},\ \bar{x} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right)$ is

$$w = 2 \cdot z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \qquad \Rightarrow \qquad \boxed{n = \left(2z_{\alpha/2} \cdot \frac{\sigma}{w}\right)^2}$$

The sample size is inversely related to the square of the desired width.

Endpoints of a (1 – α) CI for μ:

(a) σ^2 known:

$$\bar{x} \pm z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

(b) σ^2 unknown, n large:

$$\bar{x} \pm z_{\alpha/2} \cdot \frac{s}{\sqrt{n}}; \qquad s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2} \quad \text{(sample s.d.)}$$

(c) σ^2 unknown, n small:

$$\bar{x} \pm t_{\alpha/2,\, n-1} \cdot \frac{s}{\sqrt{n}}$$

When n is small, X must be [nearly] normal.

Example 10.02

The lifetime X of a particular brand of filaments is known to be normally distributed. A random sample of six filaments is tested to destruction and they are found to last for an average of 1,007 hours with a sample standard deviation of 6.2 hours.

(a) Find a 95% confidence interval estimate for the population mean lifetime μ.
(b) Is the evidence consistent with $\mu \neq 1000$?
(c) Is the evidence consistent with $\mu > 1000$?

$$\overline{X} \sim \mathrm{N}\left(\mu, \tfrac{\sigma^2}{n}\right) \quad \Rightarrow \quad Z = \frac{\overline{X} - \mu}{\left(\frac{\sigma}{\sqrt{n}}\right)} \sim \mathrm{N}(0,1)$$

but σ^2 **is not known. Use** $T = \dfrac{\overline{X} - \mu}{\left(\frac{S}{\sqrt{n}}\right)} \sim t_{n-1} \leftarrow \{\text{degrees of freedom, } \nu\}$

(a) A 95% CI for μ **is**

$$\overline{x} \pm t_{\alpha/2,\, n-1} \cdot \frac{s}{\sqrt{n}} = 1007 \pm t_{.025,\,5}\left(\frac{6.2}{\sqrt{6}}\right)$$

From the table on page 15-02, $t_{.025,\,5} = 2.57...$

$$\therefore \quad \overline{x} \pm t_{\alpha/2,\, n-1} \cdot \frac{s}{\sqrt{n}} = 1007 \pm 2.57\ldots \times 2.531\ldots = 1007 \pm 6.506\ldots$$

The 95% CI for μ **is**

$\underline{1000.49 < \mu < 1013.51}$ **(hours) (2 d.p.)**

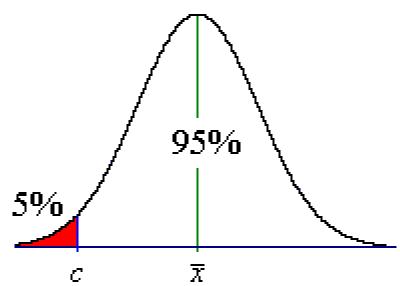
(b) $\mu = 1000$ **is *not* in the CI.**
Therefore yes, the evidence is consistent with $\mu \neq 1000$

(c) We need a one-sided CI (to test $\mu > 1000$**):**

$t_{.05,\,5} = 2.01505$

$$c = \overline{x} - t_{\alpha,\nu} \cdot \frac{s}{\sqrt{n}} = 1007 - 2.015\ldots \times \frac{6.2}{\sqrt{6}}$$

$= 1007 - 5.10...$

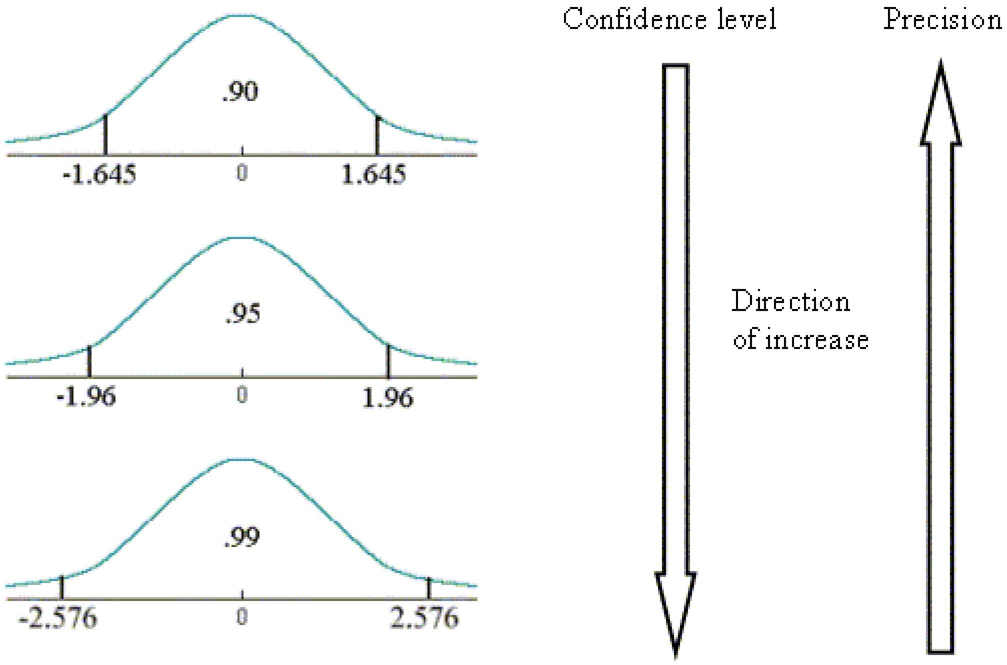


$\Rightarrow$ **95% CI is** $\mu > 1001.90$

[Expressed loosely, "we are 95% sure that $\mu > 1001.90$**".]**

Yes, the CI is consistent with $\mu > 1000$.

Properties of a confidence interval



If we think of the length of the confidence interval as specifying its precision, then the confidence level (or reliability) of the interval is inversely related to its precision.

Estimation of Population Proportion

When a random sample of size n is drawn from a population in which a proportion p of the items are "successes", each item in the sample is a Bernoulli random quantity, with

$$P[\text{"success"}] = p \quad \text{and} \quad P[\text{"failure"}] = q = 1 - p$$

Let X = number of successes in the random sample, then the probability distribution of X is

$$X \sim \text{bin}(n, p)$$

with $\quad E[X] = np \quad$ and $\quad V[X] = npq = \sigma^2$

But, for n, np, nq all large, $\text{bin}(n, p) \to N(\mu, \sigma^2)$

Therefore

$$X \sim N(np, npq) \quad \text{(approximately)}$$

The **sample proportion** $\hat{P}$ is a point estimator for the population proportion p.

$$\mathrm{E}\left[\hat{P}\right] = \mathrm{E}\left[\frac{X}{n}\right] = \frac{1}{n}\mathrm{E}[X] = \frac{np}{n} = p$$

[Compare with Bayesian $\frac{x+1}{n+2}$

$\Rightarrow$ $\hat{P}$ is an unbiased estimator of p.

− if x & n large enough, close to x / n .]

$$\mathrm{V}\left[\hat{P}\right] = \mathrm{V}\left[\frac{X}{n}\right] = \frac{1}{n^2}\mathrm{V}[X] = \frac{npq}{n^2} = \frac{pq}{n}$$

Therefore, for sufficiently large n, np and nq , (namely, $np > 10$ and $nq > 10$),

$$\hat{P} \sim N\left(p, \frac{pq}{n}\right)$$

Compare this with $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$, for which the corresponding confidence interval has endpoints $\bar{x} \pm z_{\alpha/2} \cdot \left(\frac{s}{\sqrt{n}}\right)$.

Therefore, the 100(1- α)% confidence interval estimator for p is

$$\boxed{\hat{P} \pm z_{\alpha/2}\sqrt{\frac{\hat{P}\hat{Q}}{n}}}$$

and the 100(1- α)% confidence interval **estimate** for p is

$$\boxed{\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}}$$

Note also the more precise confidence interval quoted in the course text,
(Devore, sixth edition, section 7.2, page 266):

$$\boxed{\frac{\hat{p} + \frac{z_{\alpha/2}^2}{2n} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n} + \frac{z_{\alpha/2}^2}{4n^2}}}{1 + \frac{z_{\alpha/2}^2}{n}}}$$

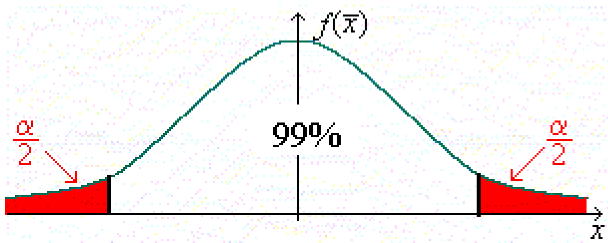
Example 10.03

From a random sample of one thousand silicon wafers, 750 pass a quality control test. Find a 99% confidence interval estimate for p (the true proportion of wafers in the population that are good).

$n = 1000$ and $x = 750$

$$\Rightarrow \quad \hat{p} = \frac{x}{n} = \frac{750}{1000} = \frac{3}{4}$$

$$\Rightarrow \quad \hat{q} = 1 - \hat{p} = \frac{1}{4}$$



$\alpha/2 = .005$

$z_{.005} = t_{.005,\infty} \approx 2.576$

Endpoints of the C.I.:

$$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}\,\hat{q}}{n}} = .75 \pm 2.576\sqrt{\frac{.75 \times .25}{1000}}$$

$= .75 \pm .035\,27...$

Therefore the 99% confidence interval estimate for p is

$$\underline{71.5\% \le p \le 78.5\%}$$

correct to three significant figures.

Using the more precise version of the confidence interval yields

$$\frac{\hat{p} + \dfrac{z_{\alpha/2}^2}{2n} \pm z_{\alpha/2}\sqrt{\dfrac{\hat{p}\hat{q}}{n} + \dfrac{z_{\alpha/2}^2}{4n^2}}}{1 + \dfrac{z_{\alpha/2}^2}{n}} = \frac{.75 + \dfrac{2.576^2}{2000} \pm 2.576\sqrt{\dfrac{.75 \times .25}{1000} + \dfrac{2.576^2}{4 \times 10^6}}}{1 + \dfrac{2.576^2}{1000}}$$

$= .7483519\ldots \pm .035195\ldots \quad \Rightarrow \quad \underline{\underline{71.3\% < p < 78.4\%}}$

[With a sample size of 1000 and the observed numbers of successes and failures both exceeding 100 by a large margin, it is no surprise that the two versions of the classical confidence interval for *p* agree to approximately 1%.]

(1−α)×100% Bayesian interval for μ

Suppose that previous evidence leads us to believe that $\mu = \mu_o$. The strength of this belief is represented by the variance σ_o^2 (lower variance corresponds to stronger belief). We wish to update that estimate after a random sample of size n has been examined. Assume that $n >> 30$ (so that the Central Limit Theorem will apply).

Prior distribution:

$$\bar{X} \sim \text{N}\left(\mu_o, \sigma_o^2\right)$$

New evidence:

Sample size = n
Sample mean = $\bar{x}$
Sample standard deviation = s

Calculate

$$\boxed{\mu^* = \frac{w_d \bar{x} + w_o \mu_o}{w_d + w_o}}, \quad \boxed{\left(\sigma^*\right)^2 = \frac{1}{w_d + w_o}}$$

where w_d, w_o are the weights of the data and original information respectively, given by

$$\boxed{w_d = \frac{1}{\left(\frac{s^2}{n}\right)}, \quad w_o = \frac{1}{\sigma_o^2}}$$

[weights ~ precision]

Posterior distribution:

$$\bar{X} \sim \text{N}\left(\mu^*, \left(\sigma^*\right)^2\right)$$

→ (1−α)×100% Bayesian interval for μ:

$$\mu = \mu^* \pm z_{\alpha/2}\, \sigma^*$$

Compare with the classical (1−α)×100% confidence interval for μ:

$$\mu = \bar{x} \pm z_{\alpha/2} \frac{s}{\sqrt{n}} \quad (n >> 30) \quad \text{or} \quad \mu = \bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

In many applications, the Bayesian interval is often narrower than the classical confidence interval, because the Bayesian interval incorporates more information (previous evidence or belief about the true value of μ).

[Note: it is easy to show that as $\sigma_o^2 \to \infty$ (or if $\bar{x} = \mu_o$ then), $\mu^* = \bar{x}$
and that as $\sigma_o^2 \to \infty$, $\sigma^{*2} \to s^2 / n$, which are the classical expressions.]

Examples of Bayesian Confidence Intervals

These examples are modifications of the previous examples of classical confidence intervals for μ.

Example 10.04

The rate of energy loss X (watt) in a motor is known to be a normally distributed random quantity and prior experience suggests that the mean is $\mu = 60$ W with standard deviation $\sigma = 3.0$ W. A random sample of 100 such motors produces a sample mean rate of energy loss of 58.3 W with sample standard deviation 2.8 W. Find a 99% confidence interval estimate for the true mean rate of energy loss μ.

Prior: $X \sim \mathrm{N}\left(60, 3^2\right)$ Weight: $w_\circ = \dfrac{1}{\sigma_\circ^2} = \dfrac{1}{9}$

Data: $\bar{x} = 58.3, \quad s = 2.8, \quad n = 100$ Weight: $w_d = \dfrac{1}{\left(\frac{s^2}{n}\right)} = \dfrac{100}{2.8^2}$

$$\Rightarrow \quad \left(\sigma^*\right)^2 = \frac{1}{w_d + w_\circ} = \frac{1}{\frac{100}{2.8^2} + \frac{1}{9}} = \frac{9 \times 7.84}{900 + 7.84} = 0.077\ldots$$

and $$\mu^* = \frac{w_d \bar{x} + w_\circ \mu_\circ}{w_d + w_\circ} = \frac{\frac{100}{7.84} \times 58.3 + \frac{1}{9} \times 60}{\frac{100}{7.84} + \frac{1}{9}} = \frac{900 \times 58.3 + 7.84 \times 60}{900 + 7.84} = 58.314\ldots$$

The 99% CI for μ is

$$\mu^* \pm z_{.005}\sigma^* = 58.3\ldots \pm 2.57583\sqrt{0.077\ldots} = 58.3\ldots \pm 0.718\ldots$$

$=$ **[57.60, 59.03]** **(2 d.p.)**

Compare this with the classical CI: [57.53, 59.07] .

[The Bayesian CI is more precise, due to good use of prior information.]

[Note that the true value of σ^2 is not known. Therefore the true number of degrees of freedom is between $n-1 = 99$ and infinity.
However, $t_{.005,99} = 2.626$ and $z_{.005} = 2.575...$, which shifts the boundaries of the CI by less than 0.015.
The error caused in replacing t by the approximation z is therefore negligible.]

Example 10.05

The lifetime X of a particular brand of filaments is known to be normally distributed. Prior experience suggests that $\mu = 1000$ and $\sigma = 6.0$. A random sample of six filaments is tested to destruction and they are found to last for an average of 1,007 hours with a sample standard deviation of 6.2 hours.

(a) Find a 95% confidence interval estimate for the population mean lifetime μ.

(b) Is the evidence consistent with $\mu \neq 1000$?

(a) $\mu_\circ = 1000, \quad \sigma_\circ^2 = 6^2, \quad \bar{x} = 1007, \quad s = 6.2, \quad n = 6$

$$\Rightarrow \quad w_\circ = \frac{1}{36}, \quad w_d = \frac{6}{(6.2)^2} \quad \Rightarrow \quad (\sigma^*)^2 = \frac{1}{\frac{6}{38.44} + \frac{1}{36}} = 5.438\ldots$$

$$\mu^* = \frac{\frac{6}{38.44} \times 1007 + \frac{1}{36} \times 1000}{\frac{6}{38.44} + \frac{1}{36}} = 1005.942\ldots$$

The 95%CI is

$$\mu^* \pm t_{.025,5}\sigma^* = 1005.9\ldots \pm 2.57\ldots \times \sqrt{5.4\ldots} = \mathbf{[999.95,\ 1011.94)} \quad (2 \text{ d.p.})$$

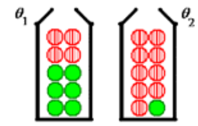
[Note that the true number of degrees of freedom is between $(n-1)$ and ∞, because of the presence of the prior information.
The CI quoted here is therefore conservative (wider than the true interval).]

(b) **$\mu = 1000$ is inside the CI; therefore NO.**

[This is the opposite conclusion from that drawn from the classical confidence interval for μ: (1000.49, 1013.51). A fairly strong prior belief in $\mu = 1000$, together with a small sample size, has dragged the Bayesian CI down closer to $\mu = 1000$; close enough to bring $\mu = 1000$ inside the CI.]

See also `www.engr.mun.ca/~gggeorge/3423/demos/BayesCI.xls.`

Example 10.A1



Suppose that we have 10,000 urns of the following types:
Urn type θ_1 : 8,000 urns, each with 4 red (R) and 6 non-red ($\tilde{R}$) balls
and urn type θ_2 : 2,000 urns, each with 9R and 1$\tilde{R}$
An urn of one of these two types is placed in front of us. We have paid a fee to withdraw two balls from the mystery urn in order to improve our chances of guessing correctly which type of urn it is.

The consequences of our guess are included in a choice between two contracts:

α_1 : gain of +\$28 if the urn is θ_1, gain of –\$32 if the urn is θ_2.

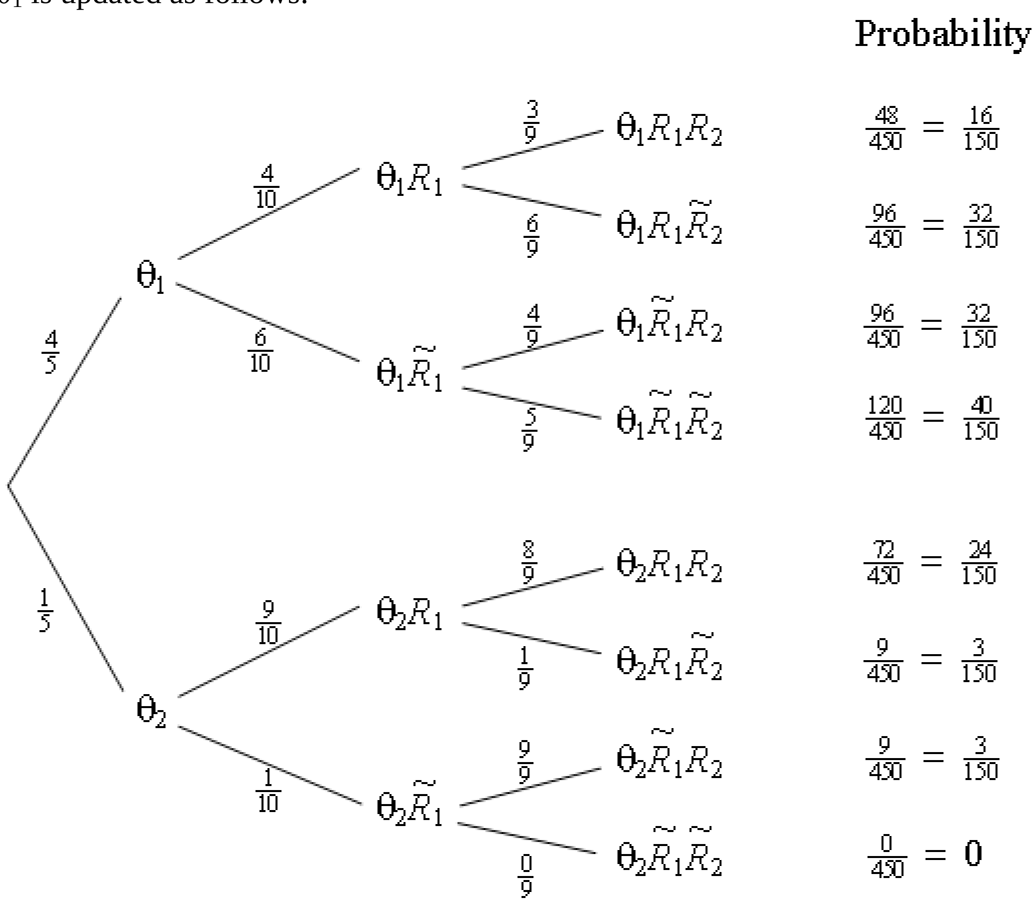
and

α_2 : gain of –\$17 if the urn is θ_1, gain of +\$88 if the urn is θ_2.

Prior:
Before any balls are withdrawn, our best guess at the probability that the urn is of type θ_1 is

$$P[\theta_1] = \frac{8000}{10000} = \frac{4}{5} = .8000$$

Posterior:
After the two balls are withdrawn, our best guess at the probability that the urn is of type θ_1 is updated as follows:



Note that events R_1 and R_2 are not independent, yet they are **exchangeable**;

$$R_1\tilde{R}_2 \approx \tilde{R}_1 R_2$$

The data allow us to update our assessment of the probabilities of θ_1 and θ_2.

$$\mathrm{P}\left[\theta_1 \middle| R_1 R_2\right] = \frac{\mathrm{P}\left[\theta_1 R_1 R_2\right]}{\mathrm{P}\left[R_1 R_2\right]} = \frac{\mathrm{P}\left[\theta_1 R_1 R_2\right]}{\mathrm{P}\left[\theta_1 R_1 R_2\right] + \mathrm{P}\left[\theta_2 R_1 R_2\right]}$$

$$= \frac{\frac{4}{5}\times\frac{4}{10}\times\frac{3}{9}}{\frac{4}{5}\times\frac{4}{10}\times\frac{3}{9}+\frac{1}{5}\times\frac{9}{10}\times\frac{8}{9}} = \frac{\frac{16}{150}}{\frac{16}{150}+\frac{24}{150}} = \frac{16}{40} = \frac{2}{5}$$

$$\mathrm{P}\left[\theta_1 \middle| R_1 \tilde{R}_2\right] = \frac{\mathrm{P}\left[\theta_1 R_1 \tilde{R}_2\right]}{\mathrm{P}\left[R_1 \tilde{R}_2\right]} = \frac{\mathrm{P}\left[\theta_1 R_1 \tilde{R}_2\right]}{\mathrm{P}\left[\theta_1 R_1 \tilde{R}_2\right] + \mathrm{P}\left[\theta_2 R_1 \tilde{R}_2\right]} = \frac{\frac{32}{150}}{\frac{32}{150}+\frac{3}{150}} = \frac{32}{35}$$

$$\mathrm{P}\left[\theta_1 \middle| \tilde{R}_1 R_2\right] = \frac{\mathrm{P}\left[\theta_1 \tilde{R}_1 R_2\right]}{\mathrm{P}\left[\tilde{R}_1 R_2\right]} = \frac{\mathrm{P}\left[\theta_1 \tilde{R}_1 R_2\right]}{\mathrm{P}\left[\theta_1 \tilde{R}_1 R_2\right] + \mathrm{P}\left[\theta_2 \tilde{R}_1 R_2\right]} = \frac{\frac{32}{150}}{\frac{32}{150}+\frac{3}{150}} = \frac{32}{35}$$

$$\mathrm{P}\left[\theta_1 \middle| \tilde{R}_1 \tilde{R}_2\right] = \frac{\mathrm{P}\left[\theta_1 \tilde{R}_1 \tilde{R}_2\right]}{\mathrm{P}\left[\tilde{R}_1 \tilde{R}_2\right]} = \frac{\mathrm{P}\left[\theta_1 \tilde{R}_1 \tilde{R}_2\right]}{\mathrm{P}\left[\theta_1 \tilde{R}_1 \tilde{R}_2\right] + \mathrm{P}\left[\theta_2 \tilde{R}_1 \tilde{R}_2\right]} = \frac{\frac{40}{150}}{\frac{40}{150}+\frac{0}{150}} = \frac{40}{40} = 1$$

Data if	then $\mathrm{P}[\theta_1] \rightarrow$	Decision choose	E[gain]	Actual gain
RR	2/5 = .4000	α_2	\$46	–\$17 or +\$88
$R\tilde{R} \vee \tilde{R}R$	32/35 ≈ .9143	α_1	\$22.86	+\$28 or –\$32
$\tilde{R}\tilde{R}$	1	α_1	\$28	+\$28 (certain)

The expected gains are calculated from the posterior probabilities:

Data = RR ⇒ E[Gain] = **.4×(–\$17) + .6×(+\$88) = –\$6.80 + \$52.80 = +\$46.00**

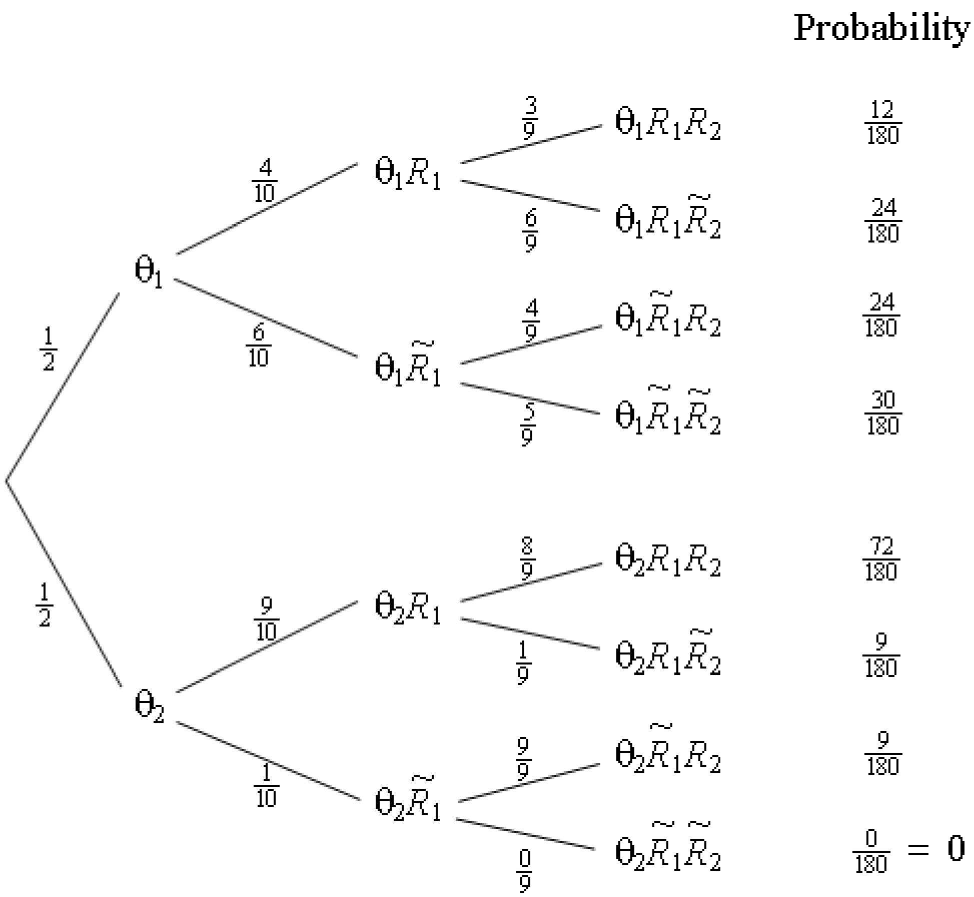
Data = $R\tilde{R}$ or $\tilde{R}R$ ⇒ E[Gain] = **32/35×(+\$28) + 3/35×–(\$32) ≈ +\$22.86**

These revised probabilities incorporate both information from the data (the two balls drawn from the urn) and the prior information (our previous belief that $\mathrm{P}[\theta_1] = 4/5$).

Now let us examine what happens if we have no preconceptions as to the relative numbers of urns of each type. We still know that each urn of type θ_1 contains 4 red and 6 non-red balls and that each urn of type θ_2 contains 9 red and 1 non-red balls. However, we do not know how many of each type of urn there may be. We may express this lack of prior information (an indifference between θ_1 and θ_2) by considering each type to be equally likely before the two balls are drawn. Therefore the prior probability is now

$$\mathrm{P}[\theta_1] = 1/2 .$$

The tree diagram changes to



But $P[\theta_1] = P[\theta_2]$ has the following consequences:
$P[\theta_1 R_1 R_2] = P[R_1 R_2 | \theta_1]\, P[\theta_1] = P[R_1 R_2 | \theta_1] / 2$,
$P[\theta_2 R_1 R_2] = P[R_1 R_2 | \theta_2]\, P[\theta_2] = P[R_1 R_2 | \theta_2] / 2$, so that

$$P\left[\theta_1 \middle| R_1 R_2\right] = \frac{P\left[\theta_1 R_1 R_2\right]}{P\left[\theta_1 R_1 R_2\right] + P\left[\theta_2 R_1 R_2\right]} = \frac{P\left[R_1 R_2 \middle| \theta_1\right]}{P\left[R_1 R_2 \middle| \theta_1\right] + P\left[R_1 R_2 \middle| \theta_2\right]}$$

(and similarly for the other three updated probabilities for θ_1).
Therefore the updated probabilities are:

$$P\left[\theta_1 \middle| R_1 R_2\right] = \frac{\cancel{\frac{1}{2}} \times \frac{4}{10} \times \frac{3}{9}}{\cancel{\frac{1}{2}} \times \frac{4}{10} \times \frac{3}{9} + \cancel{\frac{1}{2}} \times \frac{9}{10} \times \frac{8}{9}} = \frac{\frac{4}{10} \times \frac{3}{9}}{\frac{4}{10} \times \frac{3}{9} + \frac{9}{10} \times \frac{8}{9}} = \frac{12}{84} = \frac{1}{7}$$

$$P\left[\theta_1 \middle| \tilde{R}_1 R_2\right] = P\left[\theta_1 \middle| R_1 \tilde{R}_2\right] = \frac{\frac{4}{10} \times \frac{6}{9}}{\frac{4}{10} \times \frac{6}{9} + \frac{9}{10} \times \frac{1}{9}} = \frac{24}{33}$$

$$P\left[\theta_1 \middle| \tilde{R}_1 \tilde{R}_2\right] = \frac{\frac{6}{10} \times \frac{5}{9}}{\frac{6}{10} \times \frac{5}{9} + \frac{1}{10} \times \frac{0}{9}} = \frac{30}{30} = 1$$

Bayesian inference generally incorporates prior information whereas classical inference uses information only from the data. In the special case of no prior information, the two methods often produce the same prediction.

[Not examinable:]

<u>Inference on population proportion *p*</u> :

Let p = proportion of "successes" in the population.

Draw a random sample, size n (with replacement and/or from a huge population).

Let X = the number of successes in the sample
Let P = (a random quantity representing the unknown p)
then $\text{P}[X=x \mid P=p] = \text{b}(x;\, n,\, p)$ (binomial)

Let our prior belief about p be the probability mass function $\text{P}[P=p] = f_o(p)$

Then the posterior (updated) distribution, using both our prior belief and the result from the random sample, is

$$\text{P}\left[P = p \mid X = x\right] = \frac{\text{P}\left[(P = p) \wedge (X = x)\right]}{\text{P}\left[X = x\right]}$$

$$= \frac{\text{P}\left[X = x \mid P = p\right] \cdot \text{P}\left[P = p\right]}{\sum_p \text{P}\left[X = x \mid P = p\right] \cdot \text{P}\left[P = p\right]} \qquad (\text{Bayes})$$

$= f_U(p)$

Therefore

$$f_U(p) = \frac{b(x;\, n,\, p) \cdot f_\circ(p)}{\sum_p b(x;\, n,\, p) \cdot f_\circ(p)}$$

As P becomes continuous, the probability point mass $f_o(p)$ spreads out over an infinitesimal interval dp, with mass $f_o(p)\,dp$, where $f_o(p)$ is now the probability *density* function.

$$f_U(p)\,dp = \text{P}\left[(p \le P < p + dp) \mid (X = x)\right] = \frac{b(x;\, n,\, p) \cdot f_\circ(p)\,dp}{\int_0^1 b(x;\, n,\, p) \cdot f_\circ(p)\,dp}$$

<u>Example 10.A2</u>

Suppose that we have a complete lack of prior knowledge of the true value of p, so that we consider any one value of p to be just as likely as any other in the interval $0 \le p \le 1$.
Then $f_o(p) \sim U(0, 1)$ (continuous uniform distribution), that is

$$f_\circ(p) = \begin{cases} 1 & (0 \le p \le 1) \\ 0 & (\text{otherwise}) \end{cases}$$

$\Rightarrow \quad b(x;\, n,\, p)\; f_o(p) = {}^nC_x\, p^x\, (1-p)^{\,n-x} \qquad (0 \le p \le 1)$

$$\Rightarrow \int_0^1 b(x;\, n,\, p) \cdot f_\circ(p)\, dp = {}^nC_x \int_0^1 p^x (1-p)^{n-x}\, dp = \frac{n!}{x!(n-x)!} \cdot \frac{x!(n-x)!}{(n+1)!} = \frac{1}{n+1}$$

Upon observing x successes in n trials, our assessment of the value of p is updated to

$$\boxed{f_U(p) = (n+1)\, {}^nC_x\, p^x (1-p)^{n-x} \qquad (0 \le p \le 1)}$$

The population mean $\mu_P = E[P]$ for this distribution (which is also our point estimate for the unknown p) is

$$\int_{-\infty}^{+\infty} p \cdot f_U(p)\, dp = \frac{(n+1)\, n!}{x!(n-x)!} \cdot \int_0^1 p^{x+1} (1-p)^{n-x}\, dp = \frac{(n+1)!}{x!(n-x)!} \cdot \frac{(x+1)!(n-x)!}{(n+2)!} = \underline{\underline{\boldsymbol{\frac{x+1}{n+2}}}}$$

The most likely value of p (the mode) is found by solving

$$\frac{d}{dp}\left(f_U(p) \right) = 0 \quad \text{for } p \qquad \Rightarrow \quad \text{mode} = \underline{\underline{\boldsymbol{\frac{x}{n}}}}$$

(which is the classical point estimate for the unknown p).

The Bayesian and classical point estimates for p therefore disagree,
(except when $x = n / 2$).

If a random sample of size $n = 10$ is drawn, then the updated p.d.f. for p is:

Observed value of x:	$f_U(p)$	$E[P]$		mode
0	$11\,(1-p)^{10}$	1/12		0
1	$110\,p\,(1-p)^9$	2/12		1/10
2	$495\,p^2\,(1-p)^8$	3/12		2/10
3	$1320\,p^3\,(1-p)^7$	4/12		3/10
4	$2310\,p^4\,(1-p)^6$	5/12		4/10
5	$2772\,p^5\,(1-p)^5$	6/12	=	5/10
6	$2310\,p^6\,(1-p)^4$	7/12		6/10
7	$1320\,p^7\,(1-p)^3$	8/12		7/10
8	$495\,p^8\,(1-p)^2$	9/12		8/10
9	$110\,p^9\,(1-p)$	10/12		9/10
10	$11\,p^{10}$	11/12		1

Similar tables can be constructed for any other sample size n.

Using integration by parts, the following recurrence relation is established:

$$I(m,n,a,b) = \int_a^b p^m (1-p)^n \, dp \;=\; \left[\frac{-p^m (1-p)^{n+1}}{n+1}\right]_a^b + \frac{m}{n+1} \cdot I(m-1, n+1, a, b)$$

When m, n are both positive integers, this leads to

$$I(m,n,a,b) = (1-a)^{n+1} \sum_{k=0}^{m} \frac{m!\,n!}{(m-k)!\,(n+k+1)!} a^{m-k} (1-a)^k$$
$$- (1-b)^{n+1} \sum_{k=0}^{m} \frac{m!\,n!}{(m-k)!\,(n+k+1)!} b^{m-k} (1-b)^k$$

Special cases:

$$I(m,n,0,b) = \frac{m!\,n!}{(m+n+1)!} - (1-b)^{n+1} \sum_{k=0}^{m} \frac{m!\,n!}{(m-k)!\,(n+k+1)!} b^{m-k} (1-b)^k$$

$$I(m,n,0,1) = \frac{m!\,n!}{(m+n+1)!} = \int_0^1 p^m (1-p)^n \, dp = B(m+1, n+1) \quad \text{[beta function]}$$

Also

$$I(0,n,a,b) = \frac{(1-a)^{n+1} - (1-b)^{n+1}}{n+1}$$

$$I(m,0,a,b) = \frac{b^{m+1} - a^{m+1}}{m+1}$$

With $f_\circ(p) = \mathrm{U}(0,1) \quad \rightarrow \quad f_{\mathrm{U}}(p) = (n+1)\,{}^nC_x \, p^x (1-p)^{n-x} \quad (0 \le p \le 1)$,

the updated cumulative distribution function is

$$F_U(p) = \int_{-\infty}^{p} f_U(t)\, dt = (n+1)\,{}^nC_x \, I(x, n-x, 0, p)$$

$$\Rightarrow \quad F_U(p) = 1 - (1-p)^{n+1-x} \sum_{k=0}^{x} {}^{n+1}C_k \; p^k (1-p)^{x-k}$$

From the c.d.f., the boundaries of any desired confidence intervals on *p* can be calculated.

For example, with $n = 25$, $x = 20$ and using the QuickBasic program "`www.engr.mun.ca/~ggeorge/3423/demos/BayeCIp2.bas`", the following results were reported:

```
Mode of  p = .800000    Mean of  p = .777778     Median of  p = .784706
Lower quartile = .727898 ,      Upper quartile = .835014

The Bayesian 95% confidence interval for  p  is
      (.606506, .910263)
The Bayesian 99% confidence interval for  p  is
      (.545044, .936493)

The Classical 95% confidence interval for  p  is
      (.608691, .911394)
The Classical 99% confidence interval for  p  is
      (.543389, .9307
```