Example 9 Suppose that we toss a bent coin 100 times and get 72 heads. Find a 95% confidence interval for the probability of a head.

Let Y = number of heads out of the n = 100 tosses
and p = probability of a head on a single toss.

Then $Y \sim \text{Bin}(n,p)$, with realised value $y = 72$.

Now $Y \overset{\cdot}{\sim} N(np, np(1-p))$.

(This follows by the central limit theorem, since n = 100 is large.)

So $\dfrac{Y-np}{\sqrt{np(1-p)}} \overset{\cdot}{\sim} N(0,1)$.

So $1-\alpha \doteq P\left(-z_{\alpha/2} < \dfrac{Y-np}{\sqrt{np(1-p)}} < z_{\alpha/2}\right)$ (where it's hard to get p in the middle) (1)

$$\doteq P\left(-z_{\alpha/2} < \frac{Y-np}{\sqrt{n\hat{p}(1-\hat{p})}} < z_{\alpha/2}\right), \text{where } \hat{p} = \frac{Y}{n} \text{ (an unbiased estimate of } p\text{)} \quad (2)$$

$$= P\left(\frac{Y}{n} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} < p < \frac{Y}{n} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right).$$

So a $100(1-\alpha)$% CI for p is $\left(\hat{p} \pm z_{\alpha/2}\sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}\right)$. (3)

In our case $\hat{p} = y/n = 72/100 = 0.72$, and so the required 95% CI is

$$\left(0.72 \pm 1.96\sqrt{\frac{0.72(1-0.72)}{100}}\right) = (0.72 \pm 0.09) = (0.63, 0.81).$$

Note 1: Since this interval is entirely above 0.5, we suspect that the coin is not fair and that heads are more likely to come up than tails.

Note 2: CI (3) above is called ***the standard CI for a binomial proportion***. It is only one amongst many that have been proposed in the statistical literature. The following is another one which is more complicated but arguably better.

Another approach

One way to 'improve' CI (3) above is to go back to equation (1) and try harder to get p in the middle, without the 'fudge' of approximating p by $\hat{p} = Y/n$ at (2).

Observe that

$$\begin{aligned}
1-\alpha &\doteq P\left(-z_{\alpha/2} < \frac{Y-np}{\sqrt{np(1-p)}} < z_{\alpha/2}\right) \\
&= P\left(\left(\frac{Y-np}{\sqrt{np(1-p)}}\right)^2 < z^2\right) \quad \text{where } z = z_{\alpha/2} \\
&= P\left(Y^2 - 2npY + n^2p^2 < z^2np - z^2np^2\right) \\
&= P\left(p^2(1+z^2/n) - p(2Y/n + z^2/n) + Y^2/n^2 < 0\right) \\
&= P\left(p^2(1+z^2/n) - p(2\hat{p} + z^2/n) + \hat{p}^2 < 0\right) \quad \text{where } \hat{p} = Y/n \\
&= P\left(a < p < b\right),
\end{aligned}$$

where a and b are the roots of the quadratic $p^2(1+z^2/n) - p(2\hat{p} + z^2/n) + \hat{p}^2$.

These roots are given by

$$\begin{aligned}
&\frac{(2\hat{p}+z^2/n) \pm \sqrt{(2\hat{p}+z^2/n)^2 - 4\hat{p}^2(1+z^2/n)}}{2(1+z^2/n)} \\
&= \frac{\hat{p}+z^2/(2n) \pm (1/2)\sqrt{\cancel{4\hat{p}^2} + 4\hat{p}z^2/n + z^4/n^2 - \cancel{4\hat{p}^2} - 4\hat{p}^2z^2/n}}{1+z^2/n} \\
&= \frac{\hat{p}+z^2/(2n) \pm \sqrt{\hat{p}z^2/n + z^4/(4n^2) - \hat{p}^2z^2/n}}{1+z^2/n}.
\end{aligned}$$

Rearranging,we find that another $100(1-\alpha)$% CI for p is given by

$$\left(\frac{\hat{p} + \dfrac{z_{\alpha/2}^2}{2n} \pm z_{\alpha/2}\sqrt{\dfrac{\hat{p}(1-\hat{p})}{n} + \dfrac{z_{\alpha/2}^2}{4n^2}}}{1+\dfrac{z_{\alpha/2}^2}{n}}\right). \tag{4}$$

This interval is called ***the Wilson CI for a binomial proportion***. Note that if n is large this CI (4) is approximately the same as the standard CI (3), $\left(\hat{p} \pm z_{\alpha/2}\sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}\right)$.

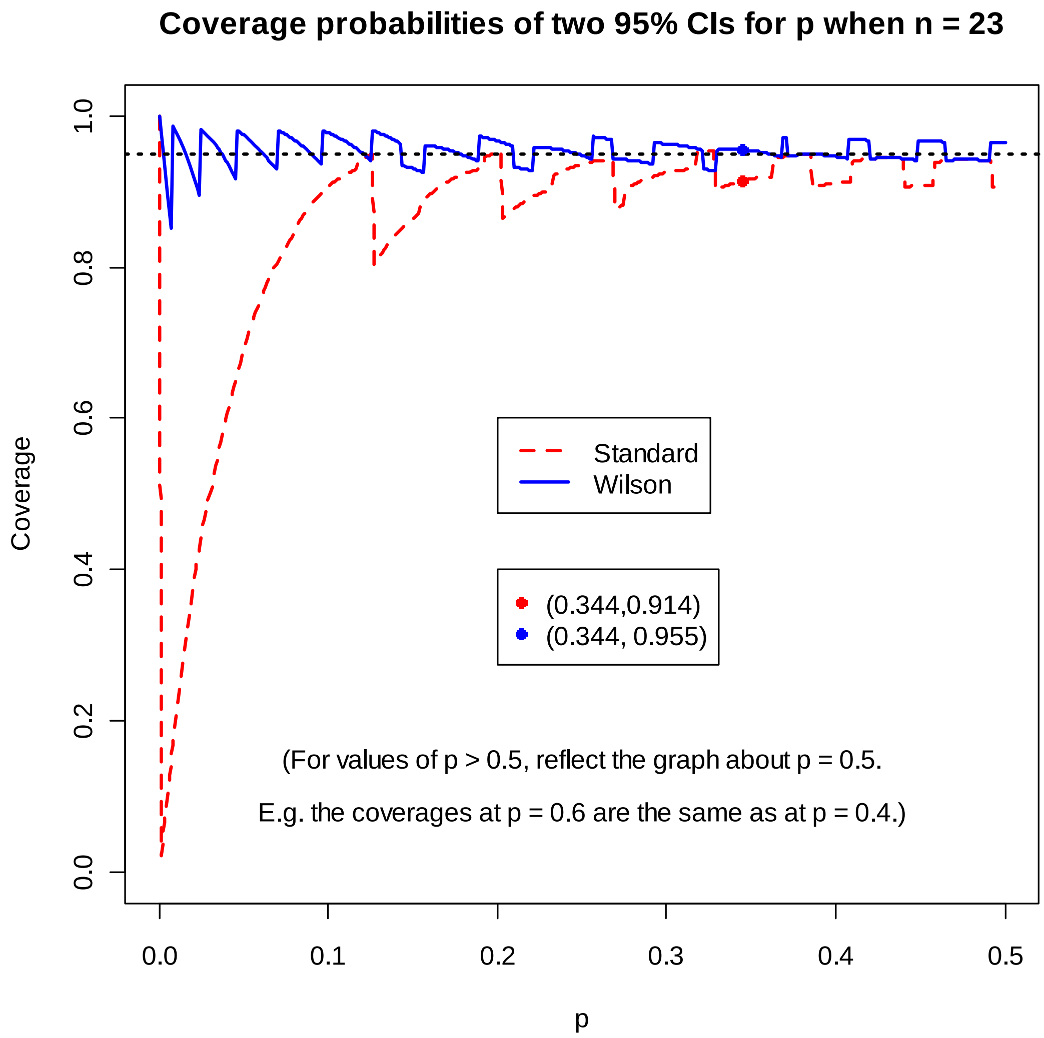
For example, if n = 100 trials result in y = 72 successes (as in Example 9), the 95% Wilson CI for p is (0.62, 0.80), which is very similar to the standard CI, (0.63, 0.81).

Discussion (non-assessable)

The Wilson CI is considered 'better' than the standard CI because its coverage probabilities are overall closer to the desired $1-\alpha$. For example, if $n = 23$ and $p = 0.344$, the coverage probability of the standard 95% CI is 91.4%, and the coverage probability of the Wilson 95% CI is 95.5%.

These calculations can be repeated for all values of p on a grid (p = 0.001,0.002,..., 1), with $n = 23$ held the same in each case. The results are shown in the following figure.

We see that for most values of p the Wilson CI has a coverage which is closer to 0.95 than that of the standard CI; however, the opposite holds for some values of p. Another pattern to note is that the Wilson CI performs far better than the standard CI when p is near 0 or 1. For any particular value of p, the difference between the two coverage probabilities converges to zero as n tends to infinity.



R Code (non-assessable)

```
n=23; p=0.344; yv=0:n; fyv=dbinom(yv,n,p); phatv=yv/n;  z=qnorm(0.975);
L1v=phatv-z*sqrt(phatv*(1-phatv)/n)
U1v=phatv+z*sqrt(phatv*(1-phatv)/n)
L2v = (phatv+z^2/(2*n)-z*sqrt(phatv*(1-phatv)/n+z^2/(4*n^2)))/(1+z^2/n)
U2v = (phatv+z^2/(2*n)+z*sqrt(phatv*(1-phatv)/n+z^2/(4*n^2)))/(1+z^2/n)
I1v=rep(0,n+1); I2v=rep(0,n+1)
for(i in 1:(n+1)){        if((L1v[i]<=p)&&(p<=U1v[i])) I1v[i] = 1
                          if((L2v[i]<=p)&&(p<=U2v[i])) I2v[i] = 1    }
CP1 = sum(fyv*I1v); CP2 = sum(fyv*I2v); c(CP1,CP2)
        # 0.9136211 0.9547925

pvec=seq(0,0.5,0.001); m=length(pvec); CP1vec=rep(NA,m); CP2vec=rep(NA,m)
for(j in 1:m){
        p=pvec[j]; yv=0:n; fyv=dbinom(yv,n,p); phatv=yv/n;  z=qnorm(0.975);
        L1v=phatv-z*sqrt(phatv*(1-phatv)/n)
        U1v=phatv+z*sqrt(phatv*(1-phatv)/n)
        L2v = (phatv+z^2/(2*n)-z*sqrt(phatv*(1-phatv)/n+z^2/(4*n^2)))/(1+z^2/n)
        U2v = (phatv+z^2/(2*n)+z*sqrt(phatv*(1-phatv)/n+z^2/(4*n^2)))/(1+z^2/n)
        I1v=rep(0,n+1); I2v=rep(0,n+1)
        for(i in 1:(n+1)){        if((L1v[i]<=p)&&(p<=U1v[i])) I1v[i] = 1
                                  if((L2v[i]<=p)&&(p<=U2v[i])) I2v[i] = 1    }
        CP1vec[j] = sum(fyv*I1v); CP2vec[j] = sum(fyv*I2v); c(CP1,CP2)    }

plot(c(0,0.5),c(0,1),type="n",xlab="p",ylab="Coverage",
   main="Coverage probabilities of two 95% CIs for p when n = 23")
lines(pvec,CP1vec,lty=2,lwd=2,col="red")
lines(pvec,CP2vec,lty=1,lwd=2,col="blue")
abline(h=0.95,lty=3,lwd=2);
points(c(0.344,0.344),c(0.9136211, 0.9547925),
        pch=c(16,16),col=c("red","blue"),cex=c(1,1))
legend(0.2,0.6,c("Standard","Wilson"),
        lty=c(2,1),lwd=c(2,2),col=c("red","blue"))
legend(0.2,0.4,   c("(0.344,0.914)","(0.344, 0.955)"),  pch=16, col=c("red","blue"))
text(0.25,0.15,"(For values of p > 0.5, reflect the graph about p = 0.5.")
text(0.25,0.08,"E.g. the coverages at p = 0.6 are the same as at p = 0.4.)")
```

Example 10 You have a bent \$1 coin and a bent \$2 coin.

You toss the \$1 coin 200 times and get 108 heads.
You toss the \$2 coin 300 times and get 141 heads.

Find a 90% CI for the difference between the probability of a head on the \$1 coin and the probability of a head on the \$2 coin.

Let $X \sim Bin(n, p)$ and $Y \sim Bin(m,q)$, where both n and m are large and $X \perp Y$.
Next let $\hat{p} = X/n$ and $\hat{q} = Y/m$.
Then it can be shown that an approximate $100(1-\alpha)$% CI for $p - q$ is

$$\left(\hat{p} - \hat{q} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n} + \frac{\hat{q}(1-\hat{q})}{m}}\right).$$

In our case: p = probability of heads on a toss of the \$1 coin
q = probability of heads on a toss of the \$2 coin
n = 200, x = 108, $\hat{p}$ = 108/200 = 0.54
m = 300, y = 141, $\hat{q}$ = 141/300 = 0.47
α = 0.1, $z_{\alpha/2} = z_{0.05}$ = 1.645.

So a 90% CI for $p - q$ is

$$\left(0.54 - 0.47 \pm 1.645\sqrt{\frac{0.54(1-0.54)}{200} + \frac{0.47(1-0.47)}{300}}\right) = (0.07 \pm 0.075)$$

$$= (-0.005, 0.145).$$

Since this CI contains 0 we suspect the chance of heads is the same for both coins.

Example 11 Suppose that $Y_1, \dots, Y_n \sim iid\ N(\mu, \sigma^2)$.
Find a $100(1-\alpha)$% CI for σ^2.

Recall that $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$.

Therefore $1-\alpha = P\left(\chi^2_{1-\alpha/2}(n-1) < \frac{(n-1)S^2}{\sigma^2} < \chi^2_{\alpha/2}(n-1)\right)$

$$= P\left(\frac{(n-1)S^2}{\chi^2_{\alpha/2}(n-1)} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{1-\alpha/2}(n-1)}\right).$$

($\chi^2_p(m)$ is the upper p quantile of the chi square dsn with m degrees of freedom.)

So a $100(1-\alpha)$% CI for σ^2 is $\left(\frac{(n-1)S^2}{\chi^2_{\alpha/2}(n-1)}, \frac{(n-1)S^2}{\chi^2_{1-\alpha/2}(n-1)}\right)$.

Confidence intervals for the difference between two population means

Suppose that: $X_1,\ldots,X_n \sim iid\ (\mu_X,\sigma_X^2)$ (1st sample)
$Y_1,\ldots,Y_m \sim iid\ (\mu_Y,\sigma_Y^2)$ (2nd sample)

$(X_1,\ldots,X_n)\perp(Y_1,\ldots,Y_m)$ (the two samples are independent)

$\bar{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$ (1st sample mean)
$\bar{Y}=\frac{1}{m}\sum_{i=1}^{m}Y_i$ (2nd sample mean)

$S_X^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2$ (1st sample variance)
$S_Y^2=\frac{1}{m-1}\sum_{i=1}^{m}(Y_i-\bar{Y})^2$ (2nd sample variance).

We wish to construct a $100(1-\alpha)\%$ CI for $\mu_X-\mu_Y$. A suitable CI will depend on:

- what assumptions we can make regarding the distributions of the X_i and Y_i values
- what knowledge we have regarding the two population variances, σ_X^2 and σ_Y^2
- whether the two sample sizes *n* and *m* are 'large' (so that we can apply the CLT).

If *n* and *m* are large, and σ_X^2 and σ_Y^2 are known, a suitable approximate CI is

$$\left(\bar{X}-\bar{Y}\pm z_{\alpha/2}\sqrt{\frac{\sigma_X^2}{n}+\frac{\sigma_Y^2}{m}}\right).$$

If *n* and *m* are both large, and σ_X^2 and σ_Y^2 are unknown, a suitable approximate CI is

$$\left(\bar{X}-\bar{Y}\pm z_{\alpha/2}\sqrt{\frac{S_X^2}{n}+\frac{S_Y^2}{m}}\right).$$

If the X_i and Y_i values are normally distributed, and σ_X^2 and σ_Y^2 are known, a suitable exact CI is

$$\left(\bar{X}-\bar{Y}\pm z_{\alpha/2}\sqrt{\frac{\sigma_X^2}{n}+\frac{\sigma_Y^2}{m}}\right).$$

If the X_i and Y_i values are normally distributed, σ_X^2 and σ_Y^2 are unknown, and *n* and *m* are both large, a suitable approximate CI is

$$\left(\bar{X}-\bar{Y} \pm z_{\alpha/2}\sqrt{\frac{S_X^2}{n}+\frac{S_Y^2}{m}}\right).$$

If the X_i and Y_i values are normally distributed, all with common population variance $\sigma^2 = \sigma_X^2 = \sigma_Y^2$ which however is unknown, a suitable exact CI is

$$\left(\bar{X}-\bar{Y} \pm t_{\alpha/2}(n+m-2)S_p\sqrt{\frac{1}{n}+\frac{1}{m}}\right),$$

where $S_p^2 = \dfrac{(n-1)S_X^2+(m-1)S_Y^2}{n+m-2}$ is the *pooled sample variance*.

(*Exercise:* Show that S_p^2 is unbiased for σ^2.)

If the X_i and Y_i values are normally distributed, nothing at all is known about σ_X^2 and σ_Y^2, and *n* and *m* are not both large, then the construction of a suitable CI is beyond the scope of this course. Interested students can find out more by researching *the Behrens-Fisher problem*, e.g. on Wikipedia (this topic is non-assessable).

Estimation via Monte Carlo methods

Example 12 (*Buffon's needle problem)*

A kitchen floor has a pattern of parallel lines that are 10 cm apart. You have a needle in your hand that is also 10 cm long. If you randomly throw the needle onto the floor, what is the probability *p* that it will cross a line?

It is difficult to find *p* exactly. However, it can be approximated simply, as follows.

We throw the needle onto the floor *n* = 1000 times, and find that the needle crosses a line 651 times (say).Then an unbiased estimate of *p* is $\hat{p} = 651/1000 = 0.651$,
and a 95% CI for *p* is $(0.651 \pm 1.96\sqrt{0.651(1-0.651)/1000})$ = (0.621, 0.681).

Note that we could get a narrower CI simply by increasing *n*. We could also use a computer to simulate the throwing of the needle.

The above solution via simulation is an example of *Monte Carlo methods*.

This term refers generally to any technique for solving a problem which involves the use of randomly generated numbers.

Exercise: Find the exact value of p in Example 12 analytically.

Solution: Let: X = perpendicular distance from centre of needle to nearest line in units of 5 cm

Y = acute angle between lines and needle in radians

A = "Needle crosses a line".

Then: $X \sim U(0,1)$, $f(x) = 1,\ 0 < x < 1$

$Y \sim U(0, \pi/2)$, $f(y) = 2/\pi,\ 0 < y < \pi/2$

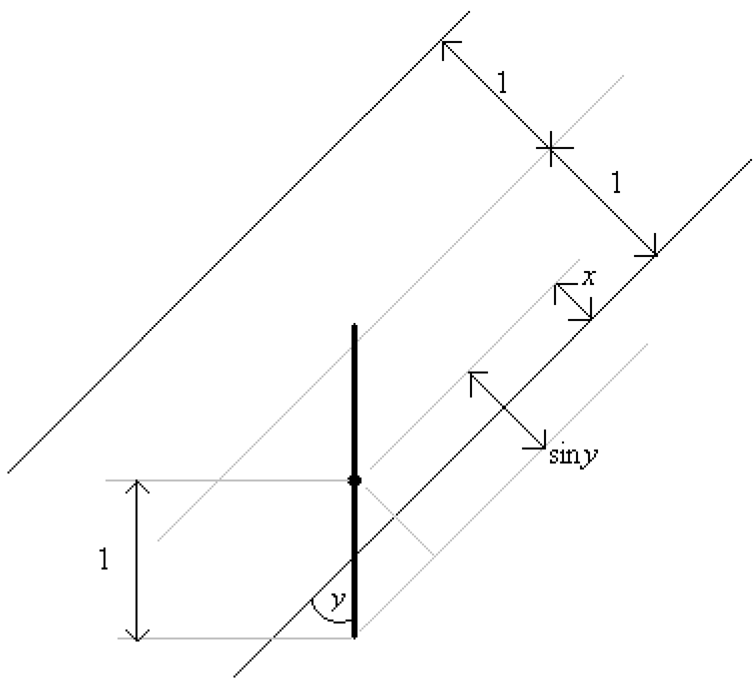
$X \perp Y$

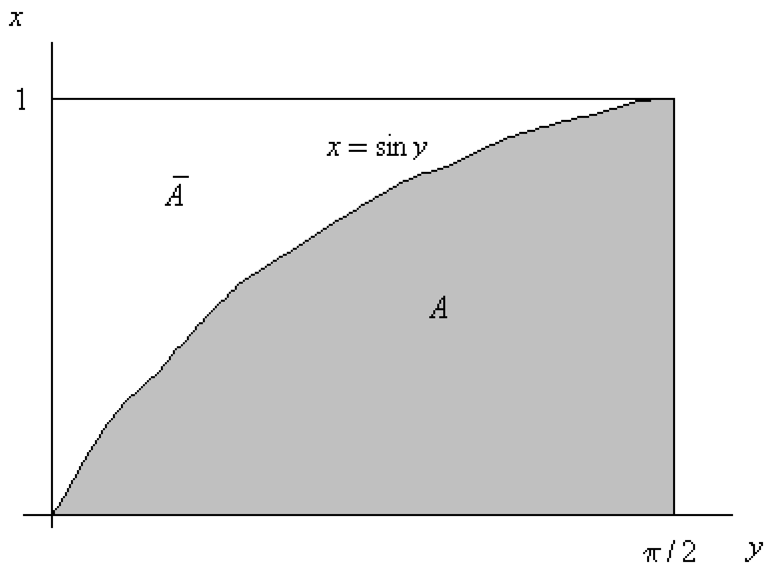
$f(x,y) = f(x)f(y) = 1 \times 2/\pi,\quad 0 < x < 1,\ 0 < y < \pi/2$

$A = \{(x,y): x < \sin y\}$.

So: $$p = P(A) = \iint_A f(x,y)\,dx\,dy = \frac{2}{\pi}\int_{y=0}^{\pi/2}\left(\int_{x=0}^{\sin y} dx\right)dy = \frac{2}{\pi}\int_{y=0}^{\pi/2} \sin y\,dy$$

$$= \frac{2}{\pi}\left[-\cos y\Big|_0^{\pi/2}\right] = \frac{2}{\pi}\left(-\cos\left(\frac{\pi}{2}\right) - (-\cos 0)\right) = \frac{2}{\pi}\left(-0-(-1)\right) = \frac{2}{\pi} = 0.637.$$





Alternatively, $P(A\,|\,y) = P(X < \sin y) = \sin y$, since $(X\,|\,y) \sim X \sim U(0,1)$.

So $p =$ $P(A) = EP(A\,|\,y) = E\sin Y = \int_0^{\pi/2} (\sin y)\frac{2}{\pi}\,dy = \frac{2}{\pi}$, as before.

Generalisation: If the length of the needle is *r* times the distance between lines, it can be shown that the probability that the needle will cross a line is

$$p = \begin{cases} 2r/\pi, & r \le 1 \\ 1 - \dfrac{2}{\pi}\left(\sqrt{r^2 - 1} - r + \sin^{-1}\left(\dfrac{1}{r}\right)\right), & r > 1 \end{cases}$$

(As an exercise, students should check that this formula is true for the case $r < 1$.
In an exam, students would not be expected to prove the result for the case $r > 1$.)