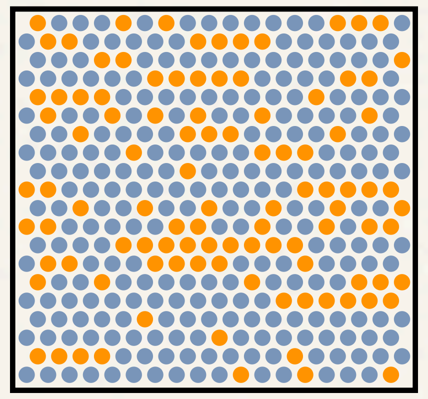
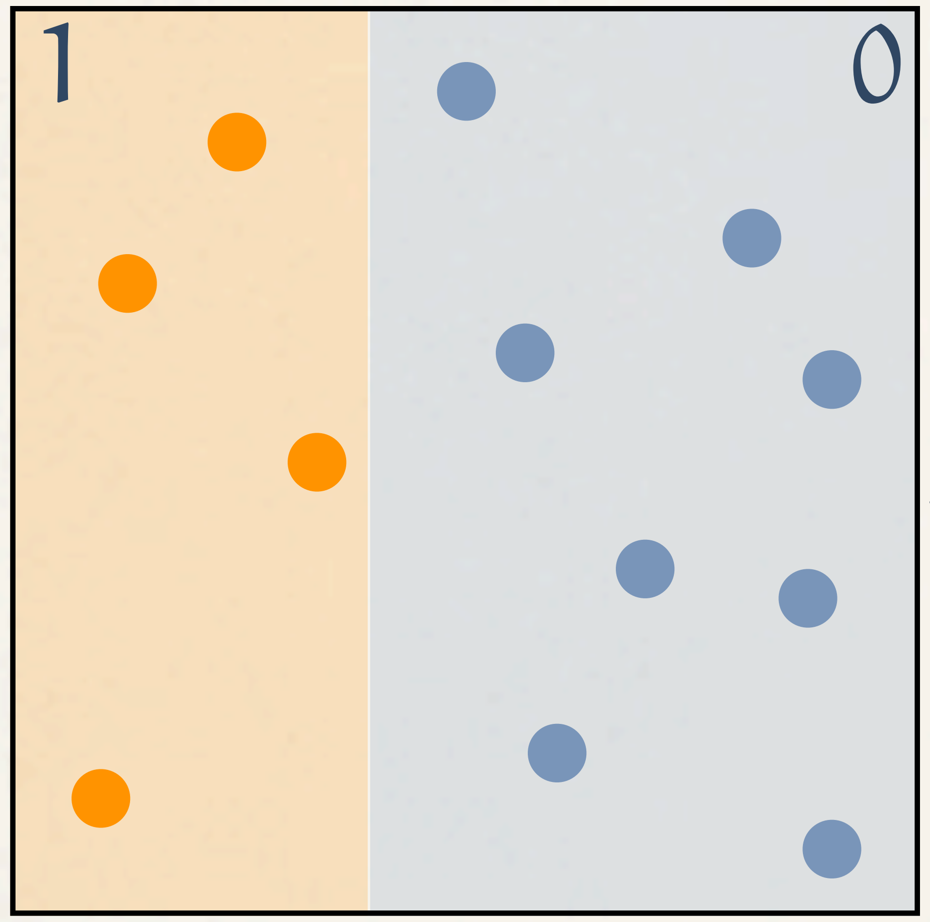
A problem in estimation

A model for a poll



Random sample: repeated independent trials

Fixed but unknown proportion p



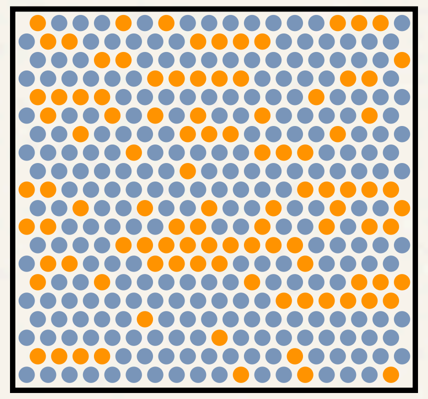
Fixed but unknown proportion $q = 1 - p$

Bernoulli(p) trials: $X_1, X_2, \ldots, X_n = \begin{cases} 1 & \text{with probability } p, \\ 0 & \text{with probability } q. \end{cases}$

Accumulated successes: $S_n = X_1 + X_2 + \cdots + X_n$

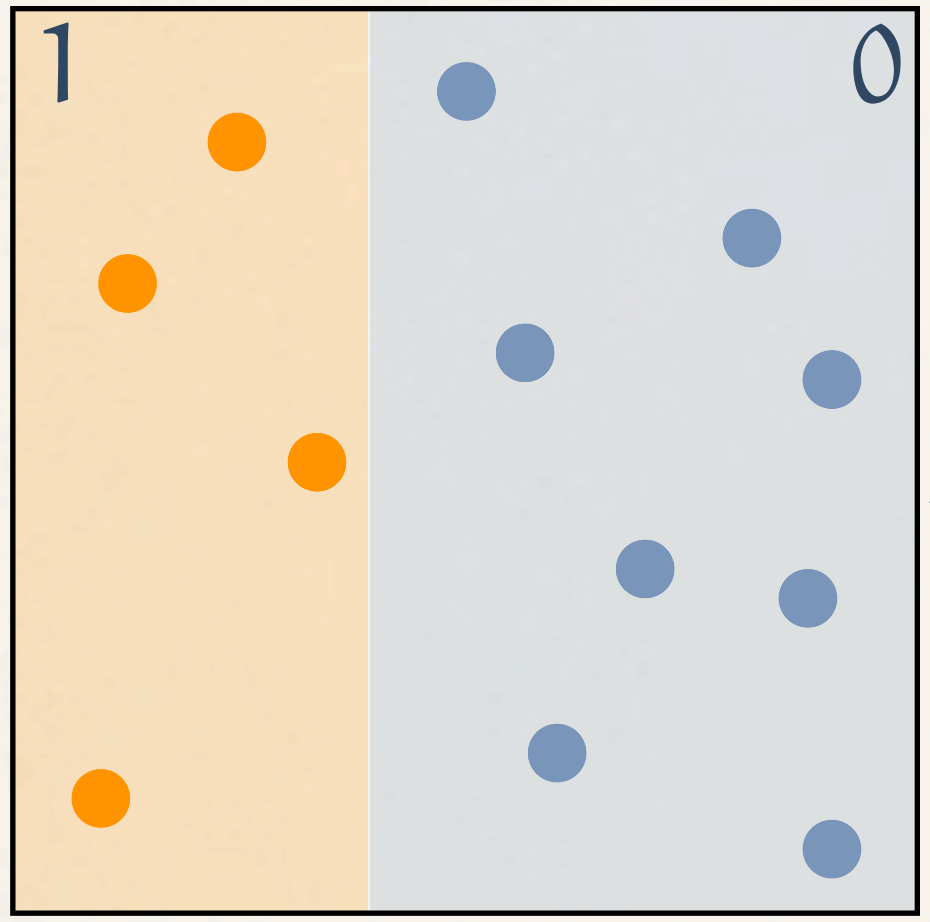
Is $\frac{S_n}{n}$ a good approximation to p?

A model for a poll



Random sample: repeated independent trials

Fixed but unknown proportion p



Fixed but unknown proportion $q = 1 - p$

Bernoulli(p) trials: $X_1, X_2, \dots, X_n = \begin{cases} 1 & \text{with probability } p, \\ 0 & \text{with probability } q. \end{cases}$

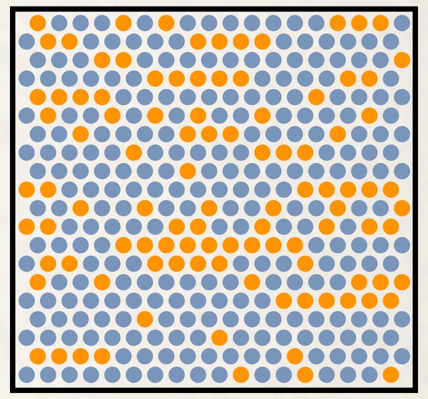
Accumulated successes: $S_n = X_1 + X_2 + \cdots + X_n$

Is $\frac{S_n}{n}$ a good approximation to p?

$$S_n \sim \text{Binomial}(n, p)$$

$$\mathbf{P}\{S_n = k\} = b_n(k; p) = \binom{n}{k} p^k q^{n-k}$$

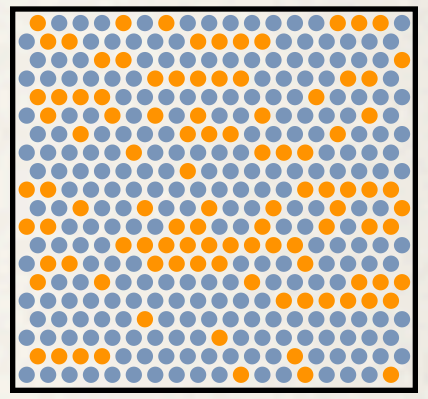
A maximum likelihood principle



$$S_n \sim \text{Binomial}(n, p)$$

$$\mathbf{P}\{S_n = k\} = b_n(k; p) = \binom{n}{k} p^k q^{n-k}$$

A maximum likelihood principle



$$S_n \sim \text{Binomial}(n, p)$$

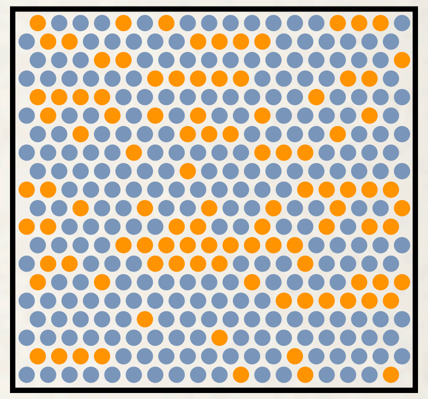
$$\mathbf{P}\{S_n = k\} = b_n(k; p) = \binom{n}{k} p^k q^{n-k}$$

What we know:

The accumulated sum $S_n = S$ is known (but chance-driven).

The bias p is fixed (not chance-driven) but, sadly, unknown.

A maximum likelihood principle



$$S_n \sim \text{Binomial}(n, p)$$

$$\mathbf{P}\{S_n = k\} = b_n(k; p) = \binom{n}{k} p^k q^{n-k}$$

What we know:

The accumulated sum $S_n = S$ is known (but chance-driven).

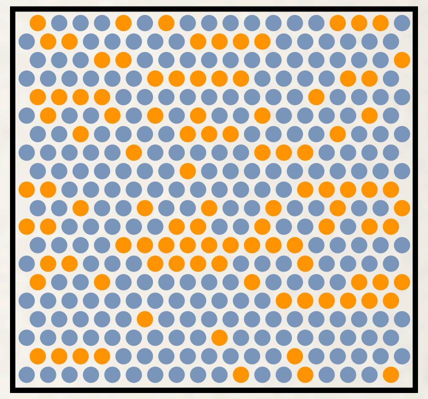
The bias p is fixed (not chance-driven) but, sadly, unknown.

R. A. Fisher's maximum likelihood principle:

Estimate the bias by that value which best explains the observation.

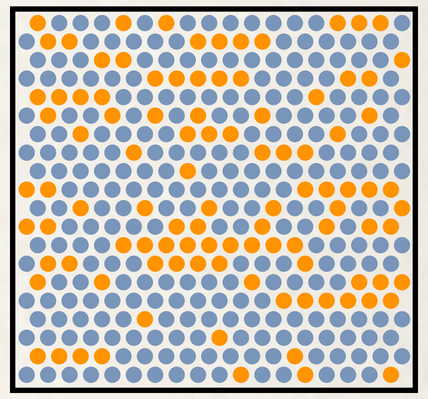
$$S_n \sim \mathrm{Binomial}(n, p)$$

$$\mathbf{P}\{S_n = k\} = b_n(k; p) = \binom{n}{k} p^k q^{n-k}$$



$$S_n \sim \mathrm{Binomial}(n, p)$$

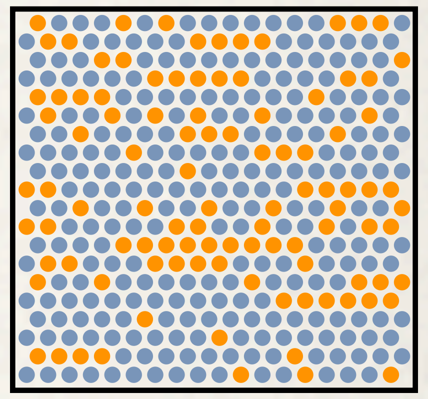
$$\mathbf{P}\{S_n = k\} = b_n(k; p) = \binom{n}{k} p^k q^{n-k}$$



Treat p as a variable x to be determined (estimated).
Given S, what value of x maximises $b_n(S; x)$?

$$S_n \sim \text{Binomial}(n, p)$$

$$\mathbf{P}\{S_n = k\} = b_n(k; p) = \binom{n}{k} p^k q^{n-k}$$

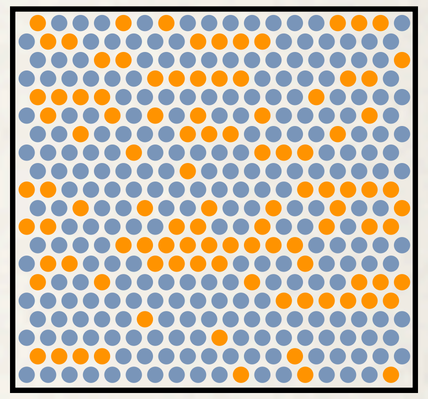


Treat p as a variable x to be determined (estimated).
Given S, what value of x maximises $b_n(S; x)$?

$$b_n(S; x) = \binom{n}{S} x^S (1 - x)^{n-S}$$

$$S_n \sim \text{Binomial}(n, p)$$

$$\mathbf{P}\{S_n = k\} = b_n(k; p) = \binom{n}{k} p^k q^{n-k}$$

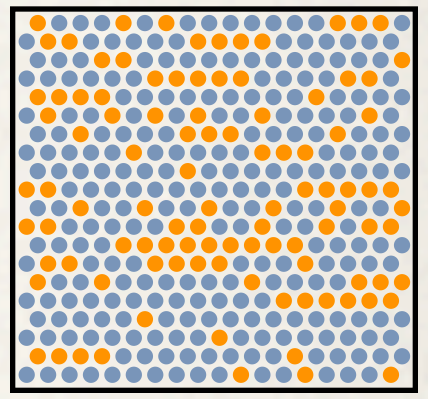


Treat p as a variable x to be determined (estimated).
Given S, what value of x maximises $b_n(S; x)$?

$$\begin{aligned} b_n(S; x) &= \binom{n}{S} x^S (1-x)^{n-S} \\ &= \binom{n}{S} x^{n \cdot \frac{S}{n}} (1-x)^{n(1-\frac{S}{n})} \end{aligned}$$

$$S_n \sim \mathrm{Binomial}(n, p)$$

$$\mathbf{P}\{S_n = k\} = b_n(k; p) = \binom{n}{k} p^k q^{n-k}$$

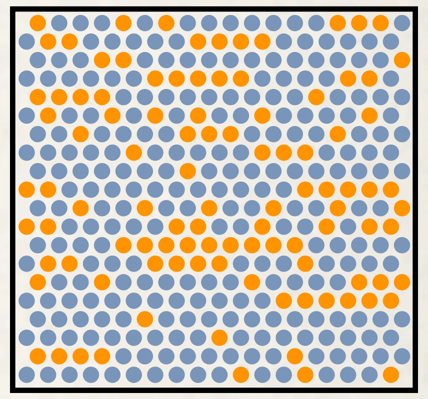


Treat p as a variable x to be determined (estimated).
Given S, what value of x maximises $b_n(S; x)$?

$$\begin{aligned} b_n(S; x) &= \binom{n}{S} x^S (1 - x)^{n-S} \\ &= \binom{n}{S} x^{n \cdot \frac{S}{n}} (1 - x)^{n(1-\frac{S}{n})} \\ &= \binom{n}{S} \left[x^{\frac{S}{n}} (1 - x)^{1-\frac{S}{n}} \right]^n \end{aligned}$$

$$S_n \sim \mathrm{Binomial}(n, p)$$

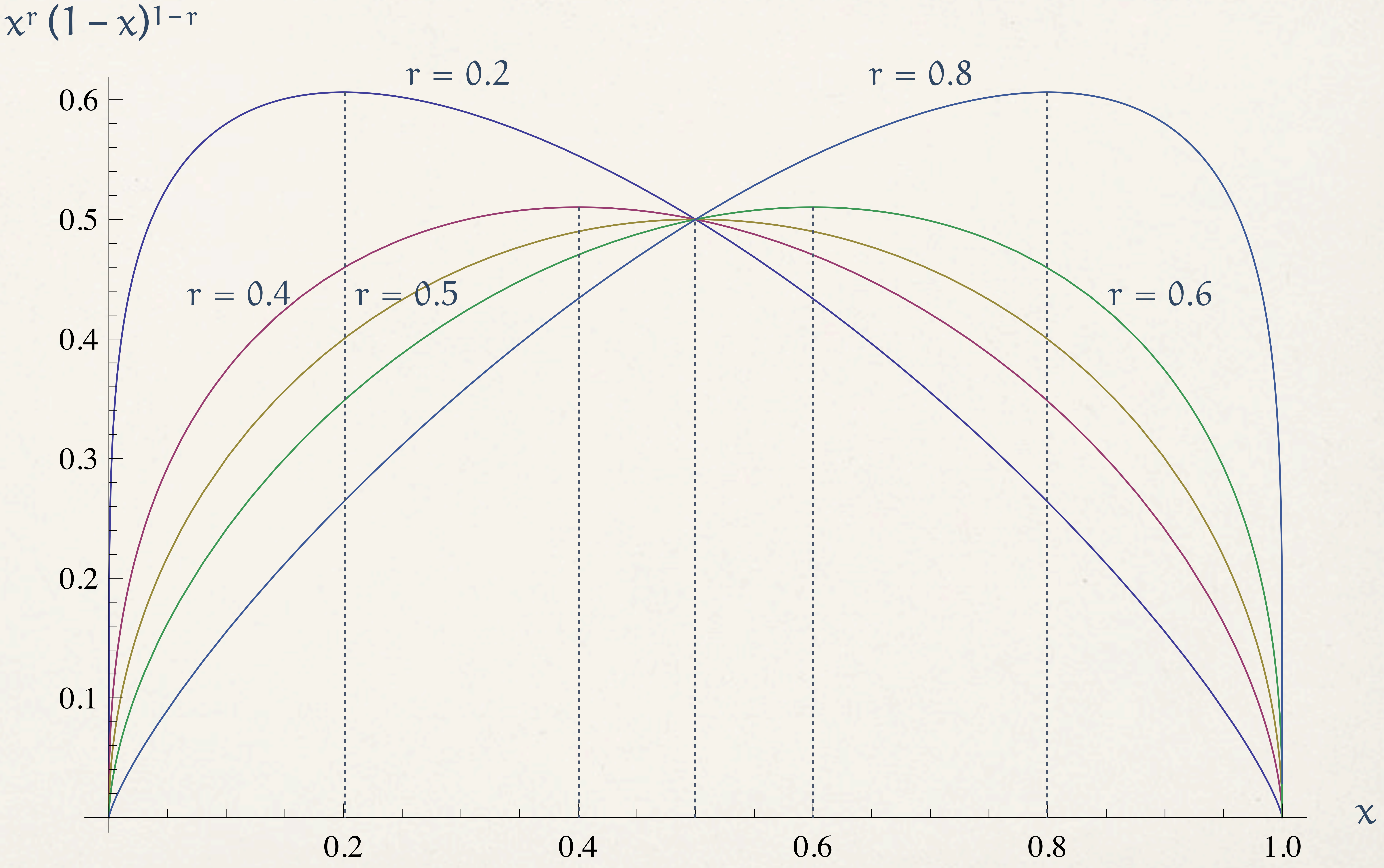
$$\mathbf{P}\{S_n = k\} = b_n(k; p) = \binom{n}{k} p^k q^{n-k}$$



Treat p as a variable x to be determined (estimated).
Given S, what value of x maximises $b_n(S; x)$?

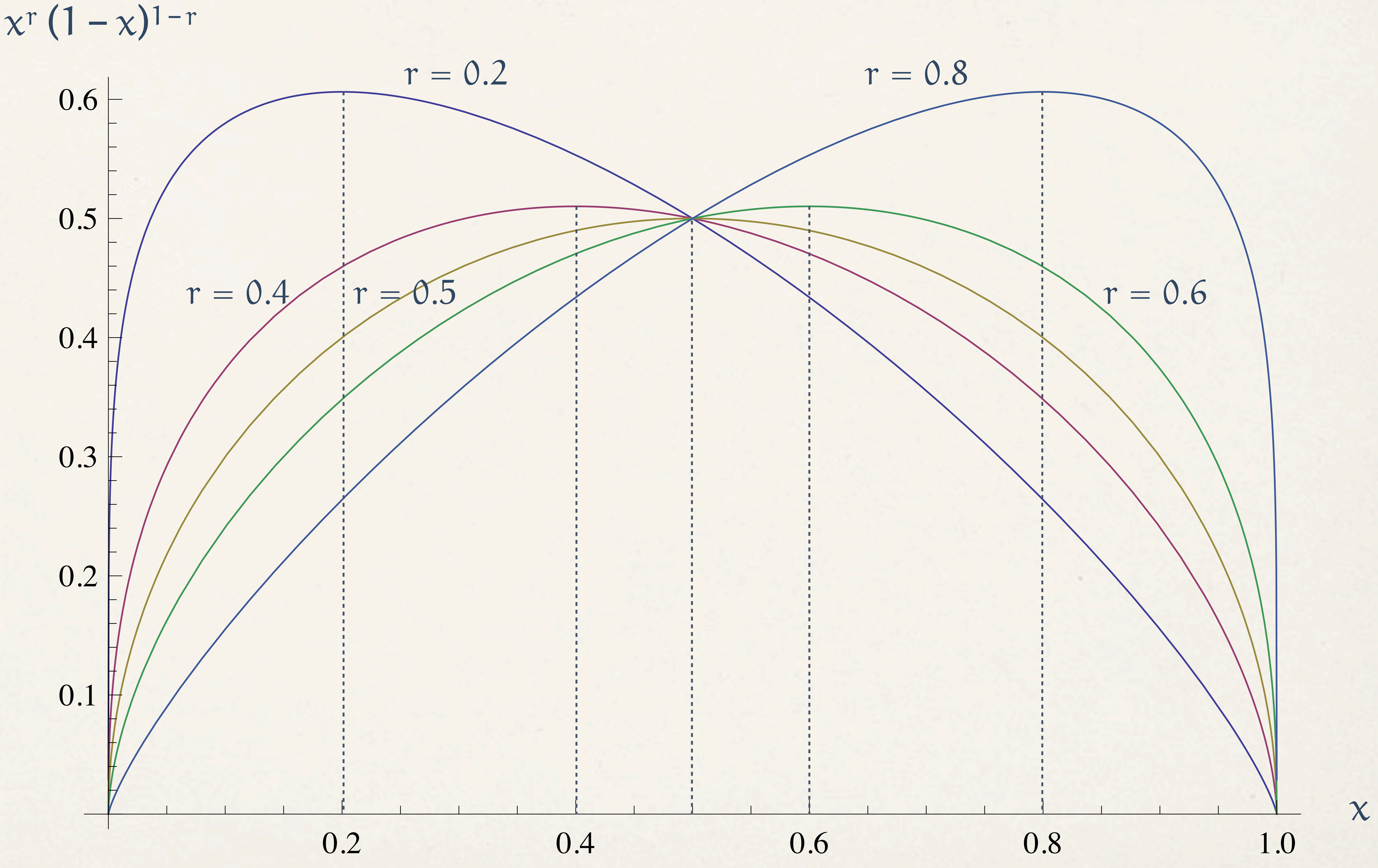
$$\begin{aligned} b_n(S; x) &= \binom{n}{S} x^S (1-x)^{n-S} \\ &= \binom{n}{S} x^{n \cdot \frac{S}{n}} (1-x)^{n(1-\frac{S}{n})} \\ &= \binom{n}{S} \left[x^{\frac{S}{n}} (1-x)^{1-\frac{S}{n}} \right]^n \end{aligned}$$

The principle: Given $r = S/n$, select x to maximise $f(x) = x^r (1-x)^{1-r}$.



$x^r (1-x)^{1-r}$
r = 0.2
r = 0.8
r = 0.4
r = 0.5
r = 0.6
0.6
0.5
0.4
0.3
0.2
0.1
0.2
0.4
0.6
0.8
1.0
x

Given $r = S/n$, the function $f(x) = x^r (1-x)^{1-r}$ is maximised at $x = r$.



Slogan

Given S_n, the maximum likelihood estimate for the bias p is S_n/n.

Slogan

Given S_n, the maximum likelihood estimate for the bias p is S_n/n.

New questions

How large should n be? And what guarantees, if any, can we give?