The likelihood principle
(quotes from Berger and Wolpert, 1988)

- "Among all prescriptions for statistical behavior, the Likelihood Principle (LP) stands out as the simplest and yet most farreaching."

- "It essentially states that **all evidence**, which is obtained from an experiment, **about an unknown quantity θ, is contained in the likelihood function of θ for the given data**."

- "The implications of this are profound, since most **non-Bayesian** approaches to statistics and indeed most standard statistical measures of evidence (such as coverage probability, error probabilities, significance level, frequentist risk, etc.) are then **contraindicated**."

- “Birnbaum (1962) showed it to be a consequence of the more commonly trusted **Sufficiency Principle** (that a sufficient statistic summarizes the evidence from an experiment) and **Conditionalilty Principle** (that experiments not actually performed should be irrelevant to conclusions).”

- “Acceptance of such a thesis radically alters the way one views statistics. Indeed, to many Bayesians, belief in the LP is the big difference between Bayesians and frequentists, not the desire to involve prior information. Thus Savage said (in the discussion of Birnbaum (1962a)

  > I, myself, came to take...Bayesian statistics...seriously only through recognition of the likelihood principle.

  Many Bayesians became Bayesians only because the LP left them little choice.”

Example 2 (Berger and Wolpert)

"The next example also seems intuitively clear, yet is the key to all that follows".

- Flip a fair coin.
  If heads, perform experiment $X \sim N(\theta, 1)$
  If tails, perform experiment $X \sim N(\theta, 4)$

- If $C(x) = [-2, 2]$, then

$$Pr(\theta \in C) = 0.5(0.68 + 0.95) = 0.815$$

- Is .815 relevant? Why average over experiments which could have been, but were not performed.

- Should *condition* on chosen experiment.
  If heads coverage is 95%, if tails 68%.
  Or, of course, use a different interval depending on the experiment run.

Example 5 (Berger and Wolpert)

- $Y_1, Y_2, \ldots \sim Bernoulli(\theta)$
- Experiment 1 (E1): $n = 12$ is fixed
  - If $X = \sum_{i=1}^{12} Y_i = 9$, then
  $$l^1(\theta) = \theta^9(1-\theta)^3$$
- Experiment 2 (E2): Sample until $\sum Y_i = 9$
  - If $n = 12$, then $l^2(\theta) \propto l^1(\theta)$
- If $H_0 : \theta = 0.5$ versus $H_a : \theta > 0.5$, then
  - Under $(E1, X = 9, \theta)$ p-value=0.073
  - Under $(E2, n = 12, \theta)$ p-value=0.034
- *violation of LP!!*

Experiment, E: triple $(X, \theta, \{f_\theta\})$.

$Ev(E, x)$: evidence about θ arising from experiment E and observation x.

Recall: T is sufficient if the conditional distribution of X given T does not depend on θ.

Sufficiency Principle: Consider the experiment $(X, \theta, \{f_\theta\})$ and suppose that $T(X)$ is a sufficient statistic for θ. If x and y are sample points satisfying $T(x) = T(y)$, then

$$Ev(E, x) = Ev(E, y)$$

Conditionality Principle: Let

$$E_i = (X_i, \theta, \{f_\theta^i\})$$

for $i = 1, 2$ be two experiments, where only the unknown parameter θ need be common between the two experiments. Consider the mixed experiment in which the random variable J is observed, where $P(J = 1) = P(J = 2) = 1/2$ (independent of θ, X_1, or X_2), and the experiment E_J is performed. Formally, the experiment performed is

$$E^* = (X^*, \theta, \{f_\theta^*\})$$

where $X^* = (J, X_J)$ and

$$f_\theta^*(x^*) = f_\theta^*((j, x_j)) = \frac{1}{2} f_\theta^j(x_j)$$

Then,

$$Ev(E^*, (j, x_j)) = Ev(E_j, x_j)$$

The Conditionality Principle simply says that if one of two experiments is randomly chosen and the chosen experiment is done, yielding data x, the information about θ *depends only on the experiment performed*.

$\Rightarrow$ The Likelihood Principle can be derived from the Sufficiency Principle and the Conditionality Principle

Likelihood Principle: Consider the experiments $E_i = (X_i, \theta, \{f_\theta^i\})$ for $i = 1, 2$, where the unknown parameter θ is the same in both experiments. Suppose x_1^* and x_2^* are sample points from E_1 and E_2, respectively, such that

$$L_{x_1^*}(\theta) = c\, L_{x_2^*}(\theta)$$

(i.e. $f_\theta^1(x_1^*) = c f_\theta^2(x_2^*)$ for all θ) Then,

$$Ev(E_1, x_1^*) = Ev(E_2, x_2^*)$$

Likelihood Principle Corolary: If

$$E = (X, \theta, \{f_\theta\})$$

is an experiment, then $Ev(E, x)$ should depend on E and x only through $L_x(\theta)$.

Birnbaum's Theorem: The Likelihood Principle follows from the Sufficiency Principle and the Conditionality Principle. The converse is also true.

Proof:

Let E_1 and E_2 be two experiments and E^* be the mixed experiment as defined in the conditionality principle.

Now let $T(X^*) = T(J, X_J)$ be $(1, x_1^*)$ if $J = 2, X_2 = x_2^*$ and (J, X_J) otherwise.

That is, if you get x_2^* from experiment 2, report that you got x_1^* from experiment 1, otherwise, just report what happened.

If you get T which is not $(1, x_1^*)$, then you know X^* is T.

If you get $T = (1, x_1^*)$, then X^* is either $(1, x_1^*)$ or $(2, x_2^*)$ and,

$$P_\theta(X^* = (1, x_1^*) \,|\, T = (1, x_1^*)) =$$

$$\frac{.5 f_\theta^1(x_1^*)}{.5 f_\theta^1(x_1^*) + .5 f_\theta^2(x_2^*)} = \frac{c}{c+1}$$

Hence the conditional distribution of X^* does not depend on θ.

T is sufficient.

Hence,

$$Ev(E_1, x_1^*) = Ev(E^*, (1, x_1^*)) =$$

$$Ev(E^*, (2, x_2^*)) = Ev(E_2, x_2^*)$$

where the second equality is by sufficiency and the other two by conditioning.