VII. Inflation

Readings

Alan Guth's "Inflationary Universe" paper (Phys Rev D, Vol. 23, p. 347, 1981) is a classic, well worth reading.

The basics are well covered by Ryden, Chapter 11.

For more physics background, I recommend Kolb & Turner, Chapter 8, or Peacock's treatment, which is somewhat more concise and up-to-date.

Extending the standard model

Strengths of the "standard cosmological model" (a.k.a. the big bang theory):

Given GR, the cosmological principle, and known atomic/nuclear/particle physics, it explains

- the dark night sky
- Hubble expansion
- Approximate agreement between $1/H_0$ and the ages of the oldest stars.
- the existence of a CMB with a blackbody spectrum
- the primordial helium abundance (at the 10% level)
- the primordial D , ^{3}He , and ^{7}Li abundances (at the order-of-magnitude level)

Shortcomings of the standard model:

It doesn't explain

- the baryon asymmetry
- the homogeneity of the universe
- the fact that $\Omega \approx 1$ and, consequently, that the entropy ($\sim$ number of photons) within the curvature radius is enormous
- the origin of irregularities leading to the formation of galaxies and galaxy structures

To explain these, one looks to extensions of the standard model, just as particle physicists look to extensions of the standard particle physics model to explain particle masses and interactions.

Indeed, the two extensions may be linked, which is one reason particle physicists are interested in cosmology and cosmologists are interested in particle physics.

In each case, the motivation for extension of the standard model is its incomplete explanatory power rather than a demonstrable failure to fit observations.

Inflation is an extension of the standard model intended to explain the last 3 of the above items. Let's first consider the homogeneity problem and the Ω problem, a.k.a. the "horizon problem" and the "flatness problem"

The Horizon Problem

A photon is emitted on a radial path from a selected origin at $t = 0$ in a FRW universe. It follows a null geodesic:

$$ds^2 = -c^2 dt^2 + a^2(t)\left[dr^2 + S_k^2(r)d\gamma^2\right] = 0 \quad \Longrightarrow \quad dr = \frac{c\,dt}{a(t)}.$$

The photon covers physical distance $a(t)dr$ at the speed of light.

Suppose that $a(t) = At^\alpha$ with $0 < \alpha < 1$. At time t the photon has gone a physical distance

$$R_H(t) \equiv a(t)r = a(t)\int_0^t \frac{c\,dt'}{a(t')} = At^\alpha \int_0^t \frac{c\,dt'}{At'^\alpha} = \frac{ct}{1-\alpha}.$$

$R_H(t)$ is the radius of the *particle horizon*, the maximum distance over which a causal signal can propagate in the age of the universe.

If $0 < \alpha < 1$ and α is not very close to 1, R_H is always $\sim ct$.

Radiation dominated $\Longrightarrow \alpha = 1/2$, $R_H = 2ct$.

Matter dominated $\Longrightarrow \alpha = 2/3$, $R_H = 3ct$.

$R_H(t)$ grows faster than $a(t)$, and the ratio of the *comoving* horizon volume today to some earlier time is

$$\frac{R_H^3(t_0)/a^3(t_0)}{R_H^3(t)/a^3(t)} = \left(\frac{a_0}{a}\right)^{3/\alpha-3} = (1+z)^{3/\alpha-3}.$$

For $\alpha = 2/3$ and $z = 1000$, this ratio is $1000^{3/2} \approx 3\times 10^4$, so at last scattering of the CMB the presently observable universe contained $\sim 30,000$ causally disconnected patches.

Why is the CMB uniform?

The horizon problem in a nutshell: In the standard cosmological model, no causal process can establish homogeneity within the presently observable universe. Homogeneity must be accepted as a mysterious initial condition.

The flatness problem

The Dicke coincidence argument

Friedmann equation:

$$\left(\frac{\dot a}{a}\right)^2 - \frac{8\pi G}{3}\left[\rho_{m,0}\left(\frac{a_0}{a}\right)^3 + \rho_{r,0}\left(\frac{a_0}{a}\right)^4 + \rho_{\Lambda,0}\right] + \frac{kc^2}{a^2} = 0.$$

The value of a has changed by a factor $\sim 0.1\text{MeV}/10^{-4}\text{eV} \sim 10^9$ since big bang nucleosynthesis, and $\sim 10^{19}\text{GeV}/10^{-4}\text{eV} \sim 10^{32}$ since the Planck epoch.

It would therefore be a surprising coincidence to come along at a time when the curvature term kc^2/a^2 is of the same order as the gravitational term.

Specifically, if the curvature term is comparable to the gravitational term today, then it had to be extremely small compared to the gravitational term in the very early universe.

For example, to have $0.1 \lesssim \Omega \lesssim 10$ today requires $\Omega \approx 1 \pm 10^{-4}$ at $z = 10^4$ (matter-radiation equality), $\Omega \approx 1 \pm 10^{-14}$, at $z = 10^9$ (nucleosynthesis), and still finer tuning at higher redshift.

Dicke's conclusion: Probably $\Omega = 1$ ($k = 0$) or Ω very close to 1 ($a \gg cH^{-1}$) so that we are not living at a "special" time.

Whether or not Dicke's conclusion is correct, the curvature term was tiny compared to the gravitational term in the early universe.

One would like a physical explanation for this, since naively one might expect "random" initial conditions to have the two terms roughly equal.

The entropy argument

A different (and perhaps more compelling) view of the flatness problem:

The entropy per unit volume is $s \sim n_\gamma \sim \left(\frac{kT}{\hbar c}\right)^3$.

Since $T \propto 1/a$, the entropy within the curvature radius (assuming $k \neq 0$) is conserved, $sa^3 =$ constant $= S$.

Since Ω is not far from 1, the curvature radius is at least $\sim ct = 3000h^{-1}$ Mpc$\sim 10^{28}$ cm, and $n_\gamma \sim 400$ cm^{-3}, so $S \gtrsim 10^{86}$.

A closed universe with a much smaller S would have collapsed long ago, while an open universe with much smaller S would have become freely expanding long ago.

We would like a physical explanation for this enormous dimensionless number.

High entropy within the curvature radius necessarily entails a high ratio of the gravitational term in the Friedmann equation to the curvature term, so these two versions of the flatness problem are closely related.

The Inflationary Solution

The inflation scenario, proposed first by Guth in 1981 (though some of the key ideas were developed earlier and independently by Starobinsky and others), attempts to solve these problems by positing a phase in the very early universe when the dominant form of stress-energy is a component with negative pressure that produces accelerated expansion.

Let's assume that this component has constant energy density ρ_ϕ. (From our earlier discussion, notes p. 21, it must therefore have pressure $p_\phi = -\rho_\phi$.)

Ignoring matter and radiation but including ρ_ϕ, the Friedmann equation is

$$\left(\frac{\dot{a}}{a}\right)^2 - \frac{8\pi G}{3}\rho_\phi + \frac{kc^2}{a^2} = 0.$$

In contrast to the matter- or radiation-dominated case, the gravitational term *grows* relative to the curvature term.

Thus, Ω is driven *towards* one instead of away from one.

Once the curvature term is negligible, the solution to the Friedmann equation is exponential expansion:

$$a = a_0 e^{t/t_H}, \qquad t_H = \frac{a}{\dot{a}} = H^{-1} = \left(\frac{8\pi G\rho_\phi}{3}\right)^{-1/2}.$$

The scale factor can therefore grow by an enormous factor in a moderate number of Hubble times t_H.

Solving the horizon problem

Our argument for the existence of particle horizons assumed $a(t) \propto t^\alpha$ with $\alpha \leq 1$.

For a model with faster growth of $a(t)$ (exponential or a steeper power law), the argument reverses, and instead of particle horizons there are event horizons.

If two observers are separated by a particle horizon, then it is impossible for one to have causally affected the other in the past, i.e., no light signal can have been sent between them.

If two observers are separated by an event horizon, then it is impossible for them to causally affect each other in the future, i.e., neither can send a light signal that will be received by the other.

In an exponentially expanding universe, the radius of the event horizon is $\sim cH^{-1}$.

If the inflation phase lasts long enough, then it can solve the horizon problem.

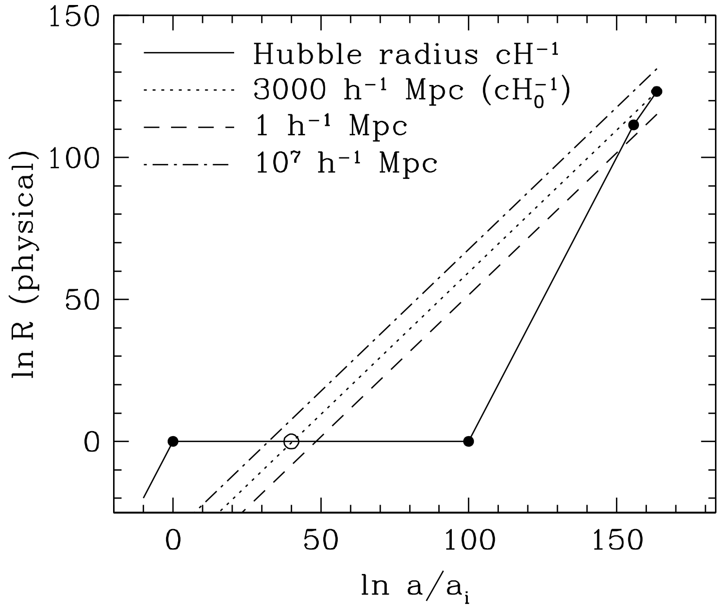
During inflation observers "lose contact" with other observers as they pass beyond event horizons, but they regain contact after inflation ends, when they reenter the particle horizon.

The presently observable universe, with Hubble radius $\sim ct$, can grow from a single patch that was causally connected before inflation occurred. Putting in numbers shows that $N \gtrsim 60$ e-folds of inflation are required to achieve this.

In effect, inflation answers the horizon paradox by saying that we calculated the particle horizon incorrectly, assuming that the expansion of the universe was always decelerating.

The last scales to leave the event horizon during the inflationary epoch are the first ones to reenter the particle horizon during the subsequent radiation/matter dominated era.

This is sometimes referred to as the "LIFO" rule (last-in, first-out).



(Above) The evolution of the Hubble radius (solid line) during inflation (flat), radiation domination, and matter domination (note inflection). Dashed, dotted, and dot-dashed lines show the physical length of three constant comoving scales. The scale corresponding to the current Hubble radius cH_0^{-1} first "left the horizon" about 60 e-folds before the end of inflation (open circle).

Solving the flatness problem

$N \gtrsim 60$ e-folds of exponential expansion makes the universe *very* flat at the end of inflation, since the curvature term $-kc^2/a^2$ drops while the gravitational term $8\pi G\rho_\phi/3$ remains constant.

This can all happen in a time $\sim 60H^{-1} \sim 10^{-34}$ seconds if inflation occurs at $T \approx 10^{15}$ GeV, the temperature when a grand unification symmetry might break.

When inflation ends, energy conservation implies that the enormous energy in ρ_ϕ must be converted into radiation. An enormous amount of entropy is *created* during this "reheating" epoch, explaining the large entropy within the curvature radius.

While we have considered ρ_ϕ =const., and thus $p_\phi = -\rho_\phi$ and exponential expansion, solving the horizon and flatness problems in this way requires only that the expansion be accelerating (growing faster than $a \propto t$).

The acceleration Friedmann equation is $\ddot{a} \propto (\rho + 3p)$, so accelerated expansion only requires $w \equiv p_\phi/\rho_\phi < -1/3$.

For $-1 < w < -1/3$, ρ_ϕ falls with time but slower than a^{-2}, so the gravitational term still grows relative to the curvature term.

Of course, the weaker the acceleration, the longer inflation must last to solve the horizon and flatness problems.

Inflation and Scalar Fields

Scalar field dynamics

A scalar field ϕ with a potential $V(\phi)$ [$V(\phi) = \lambda\phi^4$ is a commonly used example] has energy density and pressure

$$\rho = \frac{1}{2}\dot{\phi}^2 + V(\phi) + \frac{1}{2}(\vec{\nabla}\phi)^2$$
$$p = \frac{1}{2}\dot{\phi}^2 - V(\phi) - \frac{1}{6}(\vec{\nabla}\phi)^2.$$

Here we are using $\hbar = c = 1$ units, for which ϕ has units of GeV, and $V(\phi)$, ρ, and p all have units of GeV^4.

If the $\dot{\phi}^2$ and $\vec{\nabla}\phi$ terms are much smaller than $V(\phi)$, then we have $p \approx -\rho$ and thus a component that can produce exponential expansion.

Successful inflation thus requires a phase in which $V(\phi)$ dominates the energy and pressure budget for a sufficiently long time.

Once inflation starts, gradients are rapidly suppressed by the exponential expansion, so these terms are unlikely to be a problem.

Smallness of the $\dot{\phi}^2$ terms will arise for suitable forms of the potential $V(\phi)$ and starting values ϕ_i of the field.

The requirement of slow evolution of ϕ is known as the *slow roll* condition.

The dynamical equation for the classical evolution (i.e., ignoring quantum fluctuations) of ϕ is

$$\ddot{\phi} + 3H\dot{\phi} - \nabla^2\phi + \frac{dV}{d\phi} = 0.$$

Ignoring the gradient term, this resembles the equation for a ball rolling down a potential hill $V(\phi)$ with a friction term $3H\dot{\phi}$ caused by the expansion of the universe.

The inflationary phase

If $\dot{\phi}^2$ and $(\vec{\nabla}\phi)^2$ are small compared to $V(\phi)$, as required for inflation to occur, then the first and third terms above can be neglected, and we get the slow roll evolution equation:

$$3H\dot{\phi} = -\frac{dV}{d\phi}.$$

Inflation occurs during the slow roll phase, while ϕ and the energy density $V(\phi)$ are approximately constant.

In units with $\hbar = c = 1$ and $G = m_{\rm Pl}^{-2}$ where $m_{\rm Pl} = (\hbar c/G)^{1/2}$ is the Planck mass, the Friedmann equation during inflation is

$$H^2 = \frac{8\pi}{3m_{\rm Pl}^2} V(\phi).$$

Reheating

Inflation ends when ϕ begins to oscillate about the minimum of $V(\phi)$, so that kinetic energy terms dominate the density and pressure.

If there is some coupling between the field ϕ (known as the *inflaton*) and other fields and particles, then these oscillations will be damped and the energy will be dumped into these other fields and particles.

This is the reheating epoch.

Energy conservation suggests that the energy density in photons and relativistic particles at the end of reheating should be close to the value of $V(\phi)$ during inflation, but the $3H\dot{\phi}$ friction term in the equation of motion allows energy to be dissipated (effectively into expansion).

In simple reheating models, the reheat temperature depends on the strength of the coupling between ϕ and matter/radiation fields; more energy can be dissipated in $H\dot{\phi}$ friction if the coupling is weak.

However, Traschen & Brandenberger have argued that a proper analysis always yields energy density $\sim V(\phi)$, and as far as I know this is the accepted view.

After reheating, the universe evolves as a normal radiation-dominated FRW universe. Any prior inhomogeneities have been erased by the enormous expansion, and the curvature radius has been inflated to an enormous scale, containing an enormous entropy.

Note that the value of $V(\phi_{\rm min})$ must be $\lesssim \rho_{\Lambda,0}$, since otherwise the universe would enter a Λ-dominated phase when the temperature fell to $T \sim [V(\phi_{\rm min})]^{1/4}$.

It is generally assumed that $V(\phi_{\rm min}) = 0$, but there is no clear physical argument that this must be so.

Inflation Models

Requirements for successful inflation

Starting: Inflation will start if a large enough volume of space (size $\gtrsim cH^{-1}$) has $V(\phi)$ dominating over $\dot{\phi}^2$ and $(\vec{\nabla}\phi)^2$.

Lasting: The slow rolling phase must last at least ~ 60 $e-$folds to solve the horizon problem.

Ending: A successful inflation model also requires a "graceful exit" in which the universe (or at least some part of it) stops inflating and reheats.

Smoothness: As discussed below, quantum fluctuations during inflation produce density fluctuations, which must be at (or at least not exceed) the $\sim 10^{-5}$ level implied by CMB anisotropy.

Old, new, and chaotic inflation

Inflation models differ in what the potential $V(\phi)$ is and how these requirements are met.

In Guth's 1981 paper, the form of $V(\phi)$ was motivated by GUT models, and inflation lasted a long time because ϕ was trapped in a metastable state by a potential barrier.

Inflation ends by ϕ tunneling through the barrier and rolling to $V(\phi) = 0$.

Unfortunately, this mechanism leads to nucleation of bubbles, and these bubbles must collide to reheat the universe.

But the nucleation centers are being carried apart by inflation, and bubbles never collide.

Thus, Guth's model, usually known as *old inflation*, fails the graceful exit test. Guth recognized this, but wrote the paper anyway, commenting (correctly) that a successful model incorporating similar principles might be found.

In *new inflation*, $V(\phi)$ has a flat plateau but no barrier; ϕ starts near zero and slowly rolls down the plateau until it reaches the potential minimum, where it oscillates and reheats.

However, it is difficult to arrange ϕ starting near zero for a field that is weakly coupled as required by the smoothness constraint.

The most popular models now are variants of *chaotic inflation*, in which the minimum of $V(\phi)$ is at $\phi = 0$, e.g., $V(\phi) = \frac{1}{2}m^2\phi^2$ or $V(\phi) = \lambda\phi^4$.

If ϕ starts far from the minimum, then a long slow roll period can ensue.

In old and new inflation, the inflationary phase is preceded by a "normal" radiation-dominated phase, but in chaotic inflation it need not be.

The initial conditions are imagined to be "chaotic," with large spatial variations in ϕ.

Inflation takes place in those regions where ϕ happens to be large.

Stochastic, eternal inflation

The classical equation of motion is modified by quantum fluctuations, which can carry ϕ up or down the potential hill.

At large ϕ, the fluctuations dominate over the classical evolution terms (which always drive ϕ down the potential hill).

Causally separated volumes will experience independent quantum fluctuations.

The regions where ϕ is largest will inflate most rapidly and acquire most of the volume.

This leads to Andrei Linde's picture of *eternal inflation*: some regions of the universe (occupying most of the volume) are always inflating because quantum fluctuations have pushed ϕ to high values.

FRW expansion arises only in those places where ϕ actually reaches the potential minimum.

The cosmological principle thus breaks down dramatically, though only on scales much larger than the present Hubble radius cH^{-1}.

If particle physics symmetries can break in different ways, or the multi-dimensional space incorporated in string theory can compactify in different ways, then these causally disconnected regions could have effectively different physical laws, perhaps even different numbers of dimensions.

This idea is often referred to as the *multiverse.*

If a multiverse exists, then there is room for some aspects of physics to be explained by anthropic arguments (the requirement that intelligent life be able to arise), which is a good or bad thing depending on your point of view.

Primordial density fluctuations from inflation

Origin of fluctuations

The exponential expansion during inflation "irons out" any pre-existing irregularities.

Inflation was invented to explain homogeneity, i.e., to solve the horizon problem.

But within a year, several groups (Guth and Pi; Hawking; Starobinsky; Bardeen, Steinhardt, & Turner) showed that it can also explain the origin of *inhomogeneities.*

The field ϕ experiences quantum fluctuations, as the uncertainty principle tells us it must.

The exponential expansion stretches these fluctuations to large scales, and they "freeze in" as real density fluctuations.

A proper quantum mechanical treatment is subtle, but one can think of ϕ fluctuations as slightly shifting the time at which reheating occurs in different parts of the universe, by $\delta t = \delta\phi/\dot{\phi}$.

Quantum fluctuations "ripple" the surface on which the universe is homogeneous; on the constant t surfaces of the unperturbed universe, the density is inhomogeneous.

Magnitude of fluctuations

The typical quantum fluctuations in ϕ are $\delta\phi \sim H$.

One can guess this by recognizing that, in an exponentially expanding universe, H is the only characteristic quantity with the same units as $\delta\phi$ (energy). Since $\hbar = 1$, there is no large multiplicative constant that comes in with quantum effects in these units.

More specifically (following an argument given in Peebles, p. 402), the uncertainty principle gives $\delta E \delta t \sim \hbar \sim 1$.

For flat $V(\phi)$, the energy density associated with a fluctuation field $\psi \equiv \delta\phi$ is $(\vec{\nabla}\psi)^2 \sim$

$(\psi/H^{-1})^2 \sim H^2\psi^2$ and the energy is $\sim H^{-3} \cdot H^2\psi^2 = H^{-1}\psi^2$, since the lengthscale of a typical fluctuation is $\sim H^{-1}$.

The timescale is also $\sim H^{-1}$, so $H^{-1}\psi^2 H^{-1} \sim 1$ and thus $\delta\phi = \psi \sim H$.

Alternatively, quantum fluctuations in ϕ can be likened to Hawking radiation, in this case associated with the event horizon of the expanding universe instead of the event horizon of a black hole.

Hawking radiation has a characteristic wavelength of order the event horizon radius (there's no other lengthscale in town), which in this case is $\sim cH^{-1} = H^{-1}$.

This corresponds to energy fluctuations $\delta\phi \sim hc/\lambda \sim H$.

The typical fractional density fluctuation at the time a scale leaves the horizon during inflation is

$$\frac{\delta\rho}{\rho} = \left[1 + \left(\frac{\delta a}{a}\right)\right]^4 - 1 \approx 4\frac{\delta a}{a} \sim \frac{\dot{a}}{a}\delta t = H\delta t,$$

where δt is the perturbation to the time at which inflation in this region ends.

Using

$$\delta t = \frac{\delta\phi}{\dot{\phi}} \sim \frac{H}{\dot{\phi}}$$

for $\delta\phi \sim H$ and $\dot{\phi} \sim V'/H$ from the slow roll equation yields

$$\frac{\delta\rho}{\rho} \sim \frac{H^3}{V'},$$

where $V' = dV/d\phi$.

Predicted fluctuation properties

Outside the horizon, perturbations do not grow in amplitude (though this statement is complicated by the freedom to choose perturbed coordinates), so perturbations also *reenter* the horizon during the subsequent FRW phase with amplitude δ_H.

After reentering the horizon, perturbations can grow if the universe is matter dominated.

To the extent that expansion is truly exponential, with constant H, δ_H is independent of scale.

Fluctuations of this form are called *scale-invariant*, and are sometimes referred to as having a Zel'dovich or Peebles-Harrison-Zel'dovich spectrum.

In addition to having equal density contrast at the time they enter the horizon, such fluctuations are scale-invariant in the sense that each logarithmic range of length scales makes equal contribution to rms gravitational potential fluctuations at fixed time.

Fluctuations on scales that we can probe observationally left the inflationary horizon $\sim 50 - 60$ $e-$folds before the end of inflation; scales that left the horizon earlier are sub-galactic.

Since these last 60 $e-$folds may be near the end of the inflationary epoch, we expect small departures from exponential expansion and thus small departures from perfect scale invariance.

The predicted departures depend on the form of $V(\phi)$.

The small amplitude of CMB anisotropies implies $\delta_H \sim 10^{-5}$. This observation places stringent limits on $V(\phi)$, generally requiring the potential to be extremely flat.

Inflationary fluctuations are also "adiabatic," an unfortunate term that means equal fluctuations in all forms of energy (photons, neutrinos, dark matter, baryons, etc.) and thus a perturbation to spacetime curvature.

We will discuss the distinction between adiabatic (a.k.a. curvature) and isocurvature fluctuations when we get to CMB anisotropy.

Finally, quantum fluctuations of a weakly coupled field are Gaussian, meaning that the distribution $P(\delta)$ is Gaussian and any joint distribution $P(\delta_1, \delta_2, ...\delta_n)$ of density contrasts at points $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n$ is a multi-variate Gaussian.

Thus, standard versions of inflation predict Gaussian fluctuations for the origin of structure.

Assessment of Inflation

Inflation offers a plausible solution to the horizon and flatness problems.

It generically predicts that space should be flat to within practical limits of measurement.

This prediction looks good at the moment, though improved measurements could challenge it.

Inflation also generically predicts a nearly scale-invariant spectrum of adiabatic primordial fluctuations.

Combined with the assumptions of cold dark matter and dark energy, this prediction appears very successful in matching a variety of observations.

So far so good, **BUT**

Inflation appears to require unusual initial conditions and an unusually flat potential $V(\phi)$, so it is often criticized as being *fine-tuned*.

Also, each of its predictions is of a generic form, and many theoretical arguments assumed a flat universe with scale-invariant Gaussian fluctuations on grounds of simplicity before inflation came along.

One could thus imagine some alternative model coming along to explain the same things (the proponents of the cyclic model claim that it does so, see various papers by Steinhardt & Turok).

There are hopes for obtaining stronger confirmation, especially if a gravity wave contribution to the CMB anisotropy can be detected and shown to have the relation to density fluctuations predicted by generic inflationary models.

Inflation is a rather flexible scenario, so it is harder to see what observations would rule it out.

However, if a competing theory with comparable explanatory power comes to the fore, it might make predictions that could distinguish between the two.

For now, inflation is the extension of the standard cosmological model that looks most likely to survive.