Metropolis-Hastings algorithm, why it works:

(See Gelman et al book 'Bayesian data analysis'). Two conditions needed: (1) simulated sequence is a Markov chain with a unique stationary distribution, (2) the stationary distribution needs to be equal to $\pi(x \mid \text{data})$, our target density. The first step holds if the Markov chain is irreducible, aperiodic and not transient. The latter two conditions hold for a random walk on any proper distribution, and irreducibility holds if the random walk has positive probability of eventually reaching any state from any other state. This is true for all sensible proposal densities we could use. It remains to be shown that the stationary distribution is really $\pi(x \mid \text{data})$. We consider first the Metropolis algorithm.

Metropolis algorithm, stationary density:

Special case of Metropolis-Hastings algorithm. In Metropolis algorithm, the proposal density $Q(x^* \mid X_{i-1})$ is symmetric. A new value is proposed from Q, and it is accepted with probability

$$r = \min\Big(\frac{\pi(x^* \mid \text{data})}{\pi(x_{i-1} \mid \text{data})}, 1\Big)$$

If x^* is not accepted, the old value x_{i-1} is taken. So x_i will be either x^* or x_{i-1}. The proposed value is always accepted if the target density is higher at that point. If it is lower, the proposal may still be accepted, but with a probability which is the ratio of the density values in r above. To see that the stationary density is the required target density, think of starting the algorithm at step $i-1$ with a draw x_{i-1} from the target density $\pi(x \mid \text{data})$. Then consider two points x_a and x_b drawn from $\pi(x \mid \text{data})$ and labeled so that let's say $\pi(x_b \mid \text{data}) \geq \pi(x_a \mid \text{data})$. Then:

$$\begin{aligned}
&P(\text{to be at } x_a \text{ and to move to } x_b \text{ over steps i-1,i}) \\
&= P(x_{i-1} = x_a, x_i = x_b) \\
&= \underbrace{\pi(x_a \mid \text{data})}_{P(\text{to be at } x_a)} \underbrace{Q(x_b|x_a)}_{P(\text{proposal})} \underbrace{\min\Big(\frac{\pi(x_b \mid \text{data})}{\pi(x_a \mid \text{data})}, 1\Big)}_{P(\text{acceptance})=1}
\end{aligned}$$

The acceptance probability here was one, because of the assumption that the target density is higher at point x_b.

Similarly, we have:

$$\begin{aligned}
&P(\text{to be at } x_b \text{ and to move to } x_a \text{ over steps i-1,i}) \\
&= P(x_i = x_a, x_{i-1} = x_b) \\
&= \underbrace{\pi(x_b \mid \text{data})}_{P(\text{to be at } x_b)} \underbrace{Q(x_a|x_b)}_{P(\text{proposal})} \underbrace{\frac{\pi(x_a \mid \text{data})}{\pi(x_b \mid \text{data})}}_{P(\text{acceptance})<1} \\
&= \pi(x_a \mid \text{data}) Q(x_a \mid x_b)
\end{aligned}$$

Now, recall that in Metropolis algorithm, the proposal density is symmetric $Q(x_a \mid x_b) = Q(x_b \mid x_a)$. Hence we get that

$$P(x_{i-1} = x_a, x_i = x_b) = P(x_i = x_a, x_{i-1} = x_b)$$

So that the joint distribution of two consecutive samples is symmetric. Therefore, x_i and x_{i-1} have the same marginal distributions and $\pi(x \mid \text{data})$ is the stationary distribution of the Markov chain of x.

Metropolis-Hastings algorithm, stationary density:

Generalization of Metropolis algorithm. The proposal density Q is not required to be symmetric. To correct this asymmetry in the jumping rule, the ratio r is defined as a ratio of ratios:

$$r = \min\left(\frac{\pi(x^* \mid \text{data})/Q(x^* \mid x_{i-1})}{\pi(x_{i-1} \mid \text{data})/Q(x_{i-1} \mid x^*)}, 1\right)$$

Now, to prove that the stationary density is the required target density, consider again starting the chain at step $i-1$ by drawing from the target $\pi(x \mid \text{data})$. Consider two such points x_a and x_b labeled so that let's say $\pi(x_b \mid \text{data})Q(x_a \mid x_b) \geq \pi(x_a \mid \text{data})Q(x_b \mid x_a)$. The same proof as with Metropolis algorithm can be repeated. Start by writing:

$$\begin{aligned}
&P(\text{to be at } x_a \text{ and to move to } x_b \text{ over steps i-1,i}) \\
&= P(x_{i-1} = x_a, x_i = x_b) \\
&= \underbrace{\pi(x_a \mid \text{data})}_{P(\text{to be at } x_a)} \underbrace{Q(x_b|x_a)}_{P(\text{proposal})} \underbrace{\min\left(\underbrace{\frac{\pi(x_b \mid \text{data})/Q(x_b \mid x_a)}{\pi(x_a \mid \text{data})/Q(x_a \mid x_b)}}_{>1}, 1\right)}_{P(\text{acceptance})=1} \\
&= \pi(x_a \mid \text{data})Q(x_b \mid x_a)
\end{aligned}$$

because, due to our labeling, we have $P(\text{acceptance}) = 1$ according to the MH rule.

And in the other direction:

$$\begin{aligned}
&P(\text{to be at } x_b \text{ and to move to } x_a \text{ over steps i-1,i}) \\
&= P(x_i = x_a, x_{i-1} = x_b) \\
&= \underbrace{\pi(x_b \mid \text{data})}_{P(\text{to be at } x_b)} \underbrace{Q(x_a|x_b)}_{P(\text{proposal})} \underbrace{\frac{\pi(x_a \mid \text{data})/Q(x_a \mid x_b)}{\pi(x_b \mid \text{data})/Q(x_b \mid x_a)}}_{P(\text{acceptance})<1} \\
&= \pi(x_a \mid \text{data})Q(x_b \mid x_a)
\end{aligned}$$

because, due to our labeling, here $P(\text{acceptance}) < 1$ and it takes the expression defined in the Metropolis-Hastings acceptance rule. Again, the joint density of x_{i-1} and x_i is symmetric and the same conclusion follows: $\pi(x \mid \text{data})$ is the stationary distribution of the Markov chain.

Gibbs as a special case of Metropolis-Hastings:

Consider $(x^1, \ldots, x^n)$ as n-dimensional parameter vector to be sampled. Choose the proposal density of the jth parameter x^j to be the same as the conditional posterior density:
$\pi(x^j \mid x^1, \ldots, x^{j-1}, x^{j+1}, \ldots, x^n, \text{data})$.
Then, the acceptance probabilities become one, and we have Gibbs.