A Note on Marginalization

If we have a probability density $P(x, y)$, how do we derive $P(x)$? First, what do we mean by both of these objects? By $P(x, y)$ we mean that the probability that x is between x' and $x' + dx$ **and** that y is between y' and $y' + dy$ is $P(x', y')dxdy$. Likewise, the probability that x is in between x' and $x' + dx$ (regardless of the value of y) is $P(x')dx$.

Second, some nomenclature. We call $P(x, y)$ a joint probability distribution and we call the dimension-reduced partner to it, $P(x)$, a marginal distribution.

We can derive the marginal distribution $P(x)$ from the joint distribution $P(x, y)$ by integration: $P(x) = \int dyP(x, y)$. This makes sense because we're adding up the probability from all the different ways x can have a particular value; i.e., at fixed x we integrate over y.

This simple procedure receives a fancy name: "marginalization."

Now, if you've run a Markov chain then marginalization is so slick you might not even know you're doing it. Imagine you've run the chain for parameters x and y and you want to plot an (un-normalized) $P(x)$. All you do is make a histogram over x of the chain elements; i.e., for each bin of x, add up the number of chain elements that have x in that bin. That's equivalent to doing the integration over y because you are adding up everything in that bin regardless of what the value of y is.