采样

- $z^{(t)}$，从 $g_k(z|z^{(t)})$ 中抽取一个样本 z^*，

以概率 $A_k(z^*, z^{(t)})$ 接受它.

$$A_k(z^*, z^{(t)}) = \min\left(1, \frac{p(z^*)g_k(z^{(t)}|z^*)}{p(z^{(t)})g_k(z^*|z^{(t)})}\right).$$

算法定义的马尔可夫链是一个
不变的概率分布

- $p(z)q_k(z'|z)A_k(z',z) = \min(p(z)q_k(z'|z), p(z')q_k(z|z'))$
$= \min(p(z')q_k(z|z'), p(z)q_k(z'|z))$
$= p(z')q_k(z|z')A_k(z,z')$.

detail balance

$z \to z'$
$z' \to z$
同

提议分布

- 提议分布不再是参数的一个对称函数,

Q - 提议分布的具体的选择对算法的表现产生重要影响?

A - 提议分布的具体的选择会对算法的表现产生重要的影响。对于连续状态空间来说，一个常见的选择是一个以当前状态为中心的高斯分布，这会在确定分布的方差参数时需要进行一个重要的折中。如果方差过小，那么接受的转移的比例会很高，但是遍历状态空间的形式是一个缓慢的随机游走过程，导致较长的时间开销。然而，如果方差过大，那么拒绝率会很高，因为在我们考虑的这种复杂问题中，许多的步骤会到达$p(\mathbf{z})$很低的状态。考虑一个多元概率分布$p(\mathbf{z})$，它在$\mathbf{z}$的元素之间具有很强的相关性，如图11.10所示。提议分布的标度ρ应该尽可能大，同时要避免达到较高的拒绝率。这表明，ρ应该与最小的长度标度$\sigma_{\min}$是同一个量级的。然后，系统通过随机游走的方式探索伸长的方向，因此到达一个与原始状态或多或少独立的状态所需的步骤数量是$(\sigma_{\max}/\sigma_{\min})^2$量级的。事实上，在二维的情形下，随着$\rho$的增加，拒绝率的增加会被接收的转移步骤数的增加所抵消。更一般地，对于多元高斯分布，得到独立样本所需的步骤的数量的增长量级是$(\sigma_{\max}/\sigma_2)^2$的，其中$\sigma_2$是第二小的标准差（Neal, 1993）。抛开这些细节不谈，如果概率分布在不同的方向上的差异非常大，那么Metropolis-Hastings算法的收敛速度会非常慢。

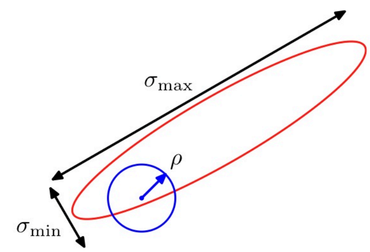

图 11.10: 使用Metropolis-Hastings算法，用一个各项同性的高斯提议分布（蓝色圆圈）从一个具有相关性的多元高斯分布（红色椭圆）中采样，这个多元高斯分布在不同的方向上的标准差的数值相当不同。为了让拒绝率较低，提议分布的标度ρ应该与最小的标准差$\sigma_{\min}$处于同一个量级，这会产生随机游走的行为，达到独立的状态所需的步骤数的量级为$(\sigma_{\max}/\sigma_{\min})^2$，其中$\sigma_{\max}$是最大的标准差。