

## 宿題 1

ガウス混合モデルの EM アルゴリズムを実装し、適当な一次元の確率密度関数を推定する。

今回は、平均 0・分散 1 の正規分布からのサンプリングのうち、7 割に 2 を足し、3 割に  $-2$  を足した分布を用いる。サンプル数 1000 点、および 10000 点について実験を行った。この分布のサンプル数 1000 のときのヒストグラムを図 1 に示す。なお、このヒストグラムは正規化されていることに注意されたい。

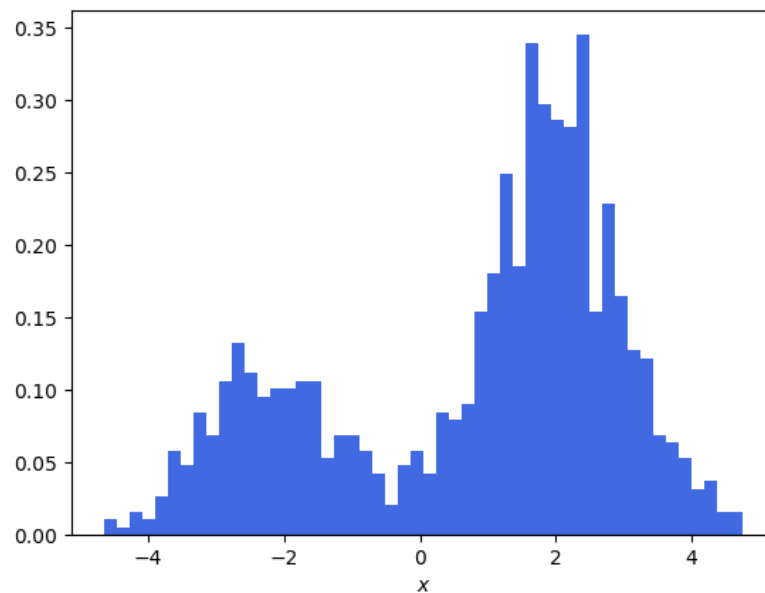


図 1: 実データのヒストグラム (サンプル数 1000)

パラメータの初期化は一様分布により行った。 $w_j$  は総和が 1 となるようにし、 $\mu$  は  $[-1.5, 1.5)$ 、 $\sigma$  は  $[0, 1)$  の一様分布を用いた。また、実データを見ると、2 つの正規分布で近似するのが妥当だと考えられるため、 $m = 2$  とした。

結果は図 2, 3 に示した。また、収束時のパラメータを表 1 に示す。データの作り方から考えて、平均 2・分散 1 の正規分布と平均  $-2$ ・分散 1 の正規分布が 7:3 で重ね合わせられている分布が真の分布であると考えられる。表 1 を見ると、特に 10000 点の場合は、実際そのようなパラメータに近くなっていることがわかる。

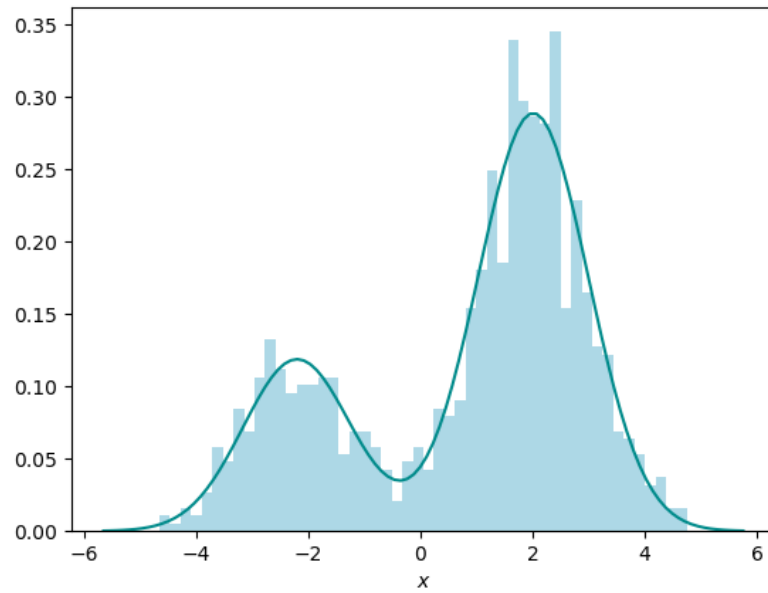


図 2: サンプル数 1000 のときの実データのヒストグラムとガウス混合モデル

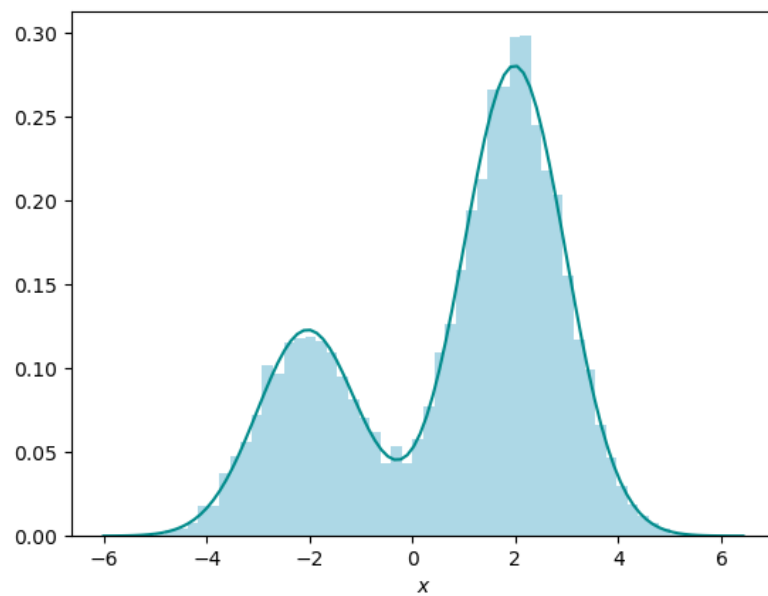


図 3: サンプル数 10000 のときの実データのヒストグラムとガウス混合モデル

表 1: 収束時のパラメータと  $Q$  の値

$n$	$Q$	$Q/n$	$w_1$	$w_2$	$\mu_1$	$\mu_2$	$\sigma_1$	$\sigma_2$
1000	-1996	-1.996	0.7138	0.2862	2.013	-2.201	0.9054	0.9614
10000	-20194	-2.019	0.3020	0.6980	-2.041	1.980	0.9797	0.9916