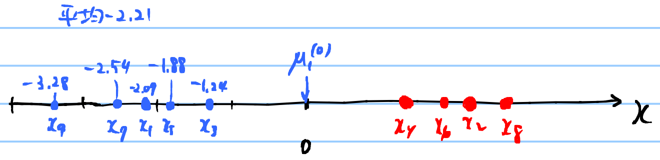
9.2

(xが1次元の場合 (9.5) を試してみる。)



$\eta_n = \frac{1}{n}$, $\mu_1^{(0)} = 0$ とする

$$\mu_1^{(1)} = \mu_1^{(0)} + 1 \cdot \overbrace{(x_1 - \mu_1^{(0)})}^{z(\mu_1^{(0)})} = 0 + 1 \cdot (-2.09 - 0) = -2.09$$

$$\mu_1^{(2)} = \mu_1^{(1)} + \frac{1}{3} \cdot \overbrace{(x_3 - \mu_1^{(1)})}^{z(\mu_1^{(1)})} = -2.09 + \frac{1}{3}(-1.24 + 2.09) = -1.81 \quad \leftarrow x_1 \text{と} x_3 \text{の平均} = -1.67$$

$$\mu_1^{(3)} = \mu_1^{(2)} + \frac{1}{5} \cdot \overbrace{(x_5 - \mu_1^{(2)})}^{z(\mu_1^{(2)})} = -1.81 + \frac{1}{5}(-1.88 + 1.81) = -1.82 \quad \leftarrow x_1, x_3, x_5 \text{の平均} = -1.74$$

$$\mu_1^{(4)} = \mu_1^{(3)} + \frac{1}{7} \cdot \overbrace{(x_7 - \mu_1^{(3)})}^{z(\mu_1^{(3)})} = -1.82 + \frac{1}{7}(-2.54 + 1.82) = -1.92 \quad \leftarrow x_1, x_3, x_5, x_7 \text{の平均} = -1.94$$

$$\mu_1^{(5)} = \mu_1^{(4)} + \frac{1}{9} \overbrace{(x_9 - \mu_1^{(4)})}^{z(\mu_1^{(4)})} = -1.92 + \frac{1}{9}(-3.28 + 1.92) = -2.07 \quad \leftarrow x_1, x_3, x_5, x_7, x_9 \text{の平均} = -2.21$$

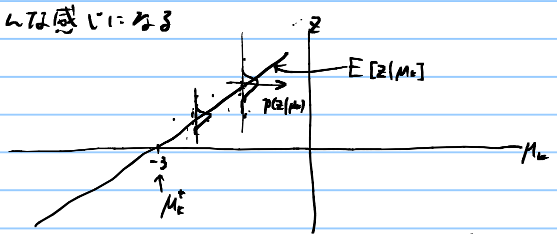
大体平均の近くの値になる

((9.5) と Robbins-Monro の関係)

確率変数x_nの分布を $p(x_n) = N(x_n | -3, 1)$ とすると
確率変数 $z = x_n - \mu_k$ の分布は $p(z|\mu_k) = N(z | -3 + \mu_k, 1)$ となる。
このとき

$$E[z|\mu_k] = \int z\, p(z|\mu_k)\, dz = \int z\, N(z | -3 + \mu_k, 1)\, dz = -3 + \mu_k$$

グラフは、こんな感じになる



これと図2.10を見比べて Robbins-Monro の手続きを適用する方法を考えると

$f(\mu_k) = \tilde{x}_n - \mu_k$, $\tilde{x}_n = \{x_n\}$の平均, $\{x_n\}$は $k=1$ に属するデータ点の集合(要素の順番を固定した集合)

$$\mu_k^{(N)} = \mu_k^{(N-1)} - a_{N-1} z(\mu_k^{(N-1)}) = \mu_k^{(N-1)} - a_{N-1}\{x_n^{(N)} - \mu_k^{(N-1)}\}, \quad x_n^{(N)} \text{は} \{x_n\} \text{の} N \text{番目の要素}$$

という反復式で $f(\mu_k^*) = 0$ の解が求まる。

（解答）

(9.1) $J = \sum_{n=1}^{N} \sum_{k=1}^{K} r_{nk} \|x_n - \mu_k\|^2$

$\|x_n - \mu_k\|^2 = (x_n - \mu_k)^T (x_n - \mu_k) = x_n^T x_n - 2 x_n^T \mu_k + \mu_k^T \mu_k$

$\frac{\partial}{\partial \mu_k} \|x_n - \mu_k\|^2 = -2x_n + 2\mu_k$

(C.19)

$\frac{\partial}{\partial x} x^T A x = (A + A^T) x$ の $A = I$ の場合

μ_k で偏微分して

$$\frac{\partial J}{\partial \mu_k} = \sum_{n=1}^{N} r_{nk} \cdot 2(\mu_k - x_n)$$

$$= 2 \sum_{n=1} r_{nk} (\mu_k - x_n)$$

$$= 2 \sum_{n \text{ of } r_{nk}=1} (\mu_k - x_n)$$ ← k に属する x_n のみ集計することによって r_{nk} を消す

なので J の極小点は

$$\sum_{n \text{ of } r_{nk}=1} (x_n - \mu_k) = 0$$

$$\therefore \sum x_n - N\mu_k = 0, \quad N = \sum_{n \text{ of } r_{nk}=1} 1$$

$$\therefore \frac{\sum x_n}{N} - \mu_k = 0$$

$$\therefore \tilde{x}_n - \mu_k = 0$$, $\tilde{x}_n$ は $r_{nk} = 1$ である x_n の平均

で与えられる。

これを解くために Robbins-Monro 法を適用するには、回帰関数を

$$f(\mu_k) = \tilde{x}_n - \mu_k$$

としたい。回帰関数がこうなるためには

x_n, μ_k が独立な確率変数であると仮定し、確率変数 z

$$z = x_n - \mu_k$$

を考えればよい。

なぜなら、このとき

$$E[z|\mu_k] = \int z\, p(z|\mu_k)\, dz$$

$$= \int z \frac{p(z, \mu_k)}{p(\mu_k)} dz$$

$$= \int z \frac{p(x_n = z + \mu_k)\, p(\mu_k)}{p(\mu_k)} dz$$

$$= \int z\, p(x = z + \mu_k)\, dz$$

$$= \int (x_n - \mu_k)\, p(x_n = x_n - \mu_k + \mu_k)\, dx_n$$ ← $x_n = z + \mu_k$ と変数変換した

$$= \int (x_n - \mu_k)\, p(x_n)\, dx_n$$

$$= \tilde{x}_n - \mu_k$$

$$= f(\mu_k)$$

となり、目的の回帰関数を得ることができるからである。

（$z = x - y$, x と y が独立のときの $p(z, y)$ について）

例えば、x, y が2値の場合の同時分布の表を考えよ

x	y	$p(x,y)$	$p(x)$	$p(y)$	z
0	0	0.18			0
0	1	0.42	$p(x=0)=0.6$	$p(y=0)=0.3$	-1
1	0	0.12	$p(x=1)=0.4$	$p(y=1)=0.7$	1
1	1	0.28			0

このとき z の取り得る値は、$0, -1, 1$ で $p(z, y)$ は

z	y	$p(z,y)$
0	0	$p(x=0, y=0) = 0.18$
0	1	$p(x=1, y=1) = 0.28$
-1	0	$p(x=-1, y=0) = 0$
-1	1	$p(x=0, y=1) = 0.42$
1	0	$p(x=1, y=0) = 0.12$
1	1	$p(x=2, y=1) = 0$

となる。

これを一般化して

x, y が独立なので

$$p(z, y) = p(x = z + y, y) = p(x = z + y)\, p(y)$$

となる。

この z を (2.129) に入れて

$$\mu_k^{(N)} = \mu_k^{(N-1)} - a_{N-1}\left(x_n^{(N)} - \mu_k^{(N-1)}\right)$$, x_n は $r_{nk} = 1$ であるもの

という反復式を得る。

この反復式は、$a_{N-1} = \eta_n$ とすれば、(9.5) と同じになる。

上の反復式を利用したK-meansアルゴリズムは以下のようになる

① μ_kの初期値を適当に決める

② 新しいx_nデータがくると、その時点の一番近いμ_kにx_nが属するとする。→ 逐次Eステップ

③ μ_k, x_nを上の反復式に入れてμ_kを更新する → 逐次Mステップ

②,③をデータが来る度にくり返す。

反復式はx_n,μ_kが確率変数であると仮定して、Robbins-Monroの手続きで得られたものなので
このアルゴリズムは確率的アルゴリズムである。