Gaussian is not good at describing distributions with multiple modes ($\geq$ bumps).

Example:

- $x = 1$
- $x = -1$
- $x = 0$

But could describe this as some combination of two Gaussians, $N(x|1, \sigma^2)$ and $N(x|-1, \sigma^2)$.

Alternative:

"Splicing" $P(x) \propto \sum_{n} N(x|\mu_n, \sigma^2)$

"Max" $P(x) = \max_{n} N(x|\mu_n, \sigma^2)$

"Upper convex" $P(x) \leq \max_{n} N(x|\mu_n, \sigma^2)$, $N(x|\mu, \sigma^2)$

Linear combination (aka mixture of Gaussians!)

$$P(x) = \frac{1}{2} N(x|1, \sigma^2) + \frac{1}{2} N(x|-1, \sigma^2)$$

Note that normalization constant for first two alternatives may be a problem to compute. But our linear combination is already correctly normalized!

$$\sum_{\infty} = \frac{1}{2} + \frac{1}{2} = 1$$
In general, \( K \) components with means \( \mu_k \), 

\[ \text{covariance matrices } \Sigma_k, \text{ and } \text{"mixing weights" } w_k \] 

such that \( \sum_{k=1}^{K} w_k = 1 \)

\[ p(x) = \sum_{k=1}^{K} w_k \mathcal{N}(x | \mu_k, \Sigma_k) \]

again, this is easily verified to be a properly normalized density function if \( \sum_{k=1}^{K} w_k = 1 \)

to generate samples from an MCG

For \( i = 1 \) to \( N \)

- Generate \( u \) = uniform random number \( U(0, 1) \)

- if \( u < w_1 \)

  - generate \( x_i \sim \mathcal{N}(x | \mu_1, \Sigma_1) \)

else if \( u < w_1 + w_2 \)

  - generate \( x_i \sim \mathcal{N}(x | \mu_2, \Sigma_2) \)

else if \( u < w_1 + w_2 + \ldots + w_{k-1} \)

  - generate \( x_i \sim \mathcal{N}(x | \mu_{k-1}, \Sigma_{k-1}) \)

else

  - generate \( x_i \sim \mathcal{N}(x | \mu_k, \Sigma_k) \)

end if

end for
This corresponds to the point in the mixture of Gaussian powerpoint slides where I say I’ll derive $E_x[\log P(X,Z|\theta)]$ on the board.

\[
P(X,Z|\theta) = \prod_k \prod_{n_k} \left[ \prod_{i=1}^k P(X_i|\mu_i, \Sigma_i) \right]
\]

\[
\log P(X,Z|\theta) = \sum_{n_k} \log P(X_i|\mu_i, \Sigma_i) + \sum_{n_k} \log P(Z|\mu_i, \Sigma_i)
\]

\[
E_x[\log P(X,Z|\theta)] = \text{Aside: } E_x[f(c_x)] = f(0)P(Z=0) + f(1)P(Z=1)
\]

\[
= \sum_{n_k} \log P(X_i|\mu_i, \Sigma_i)
\]

And this $\delta_{n_k}$ notice that discrete variable $Z_{n_k}$ has been replaced with a continuous $0 \leq \delta_{n_k} \leq 1$ value.

Now what is $\delta_{n_k} = P(Z_{n_k} = 1)$?

\[
P(X,Z_{n_k}) = \prod_{i=1}^k \prod_{n_k} \left[ \prod_{i=1}^k P(X_i|\mu_i, \Sigma_i) \right] = \prod_{i=1}^k P(X_i|\mu_i, \Sigma_i)
\]

\[
P(X,Z_{n_k})' = \prod_{i=1}^k P(X_i|\mu_i, \Sigma_i)
\]

\[
P(Z_{n_k} = 1|X) = \frac{P(X,Z_{n_k})}{P(X,Z_{n_k})} = \frac{P(X,Z_{n_k})}{P(X)}
\]

Note: There is a coupling between $Z_{n_1}, Z_{n_2}, \ldots, Z_{n_k}$ (Only one of them can be 1, the rest are 0).

So we have to be careful with this summation. We sum over all $k$ of them.

\[
P(Z_{n_k} = 1|X) = \prod_{i=1}^k \prod_{n_k} \left[ \prod_{i=1}^k P(X_i|\mu_i, \Sigma_i) \right] = \prod_{i=1}^k P(X_i|\mu_i, \Sigma_i)
\]

And although we don’t need it:

\[
P(Z_{n_k} = 0|X) = 1 - P(Z_{n_k} = 1|X)
\]