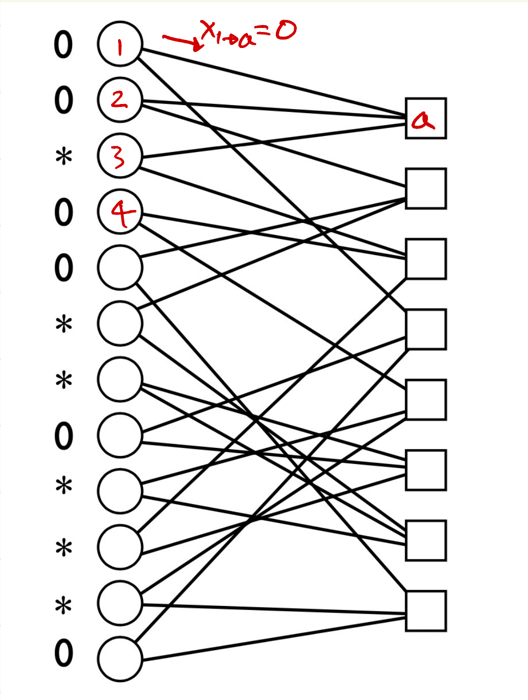
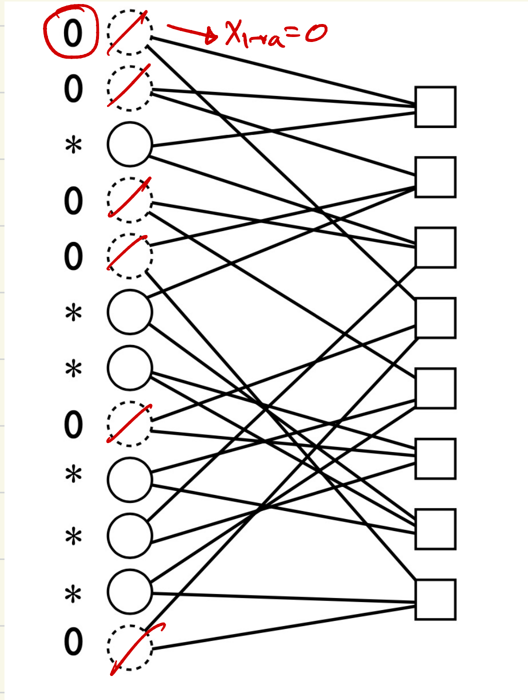
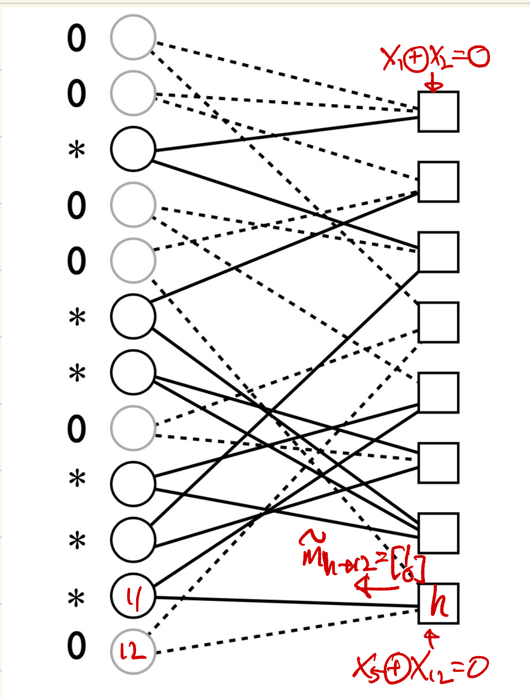
Without loss of generality, suppose all 0's were sent.

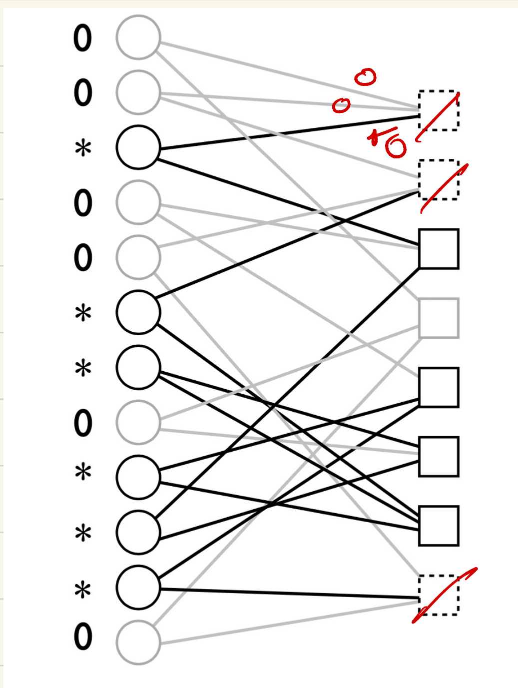
Peel nodes/edges whose messages are $\begin{bmatrix}0\\1\end{bmatrix}$ or $\begin{bmatrix}1\\0\end{bmatrix}$.



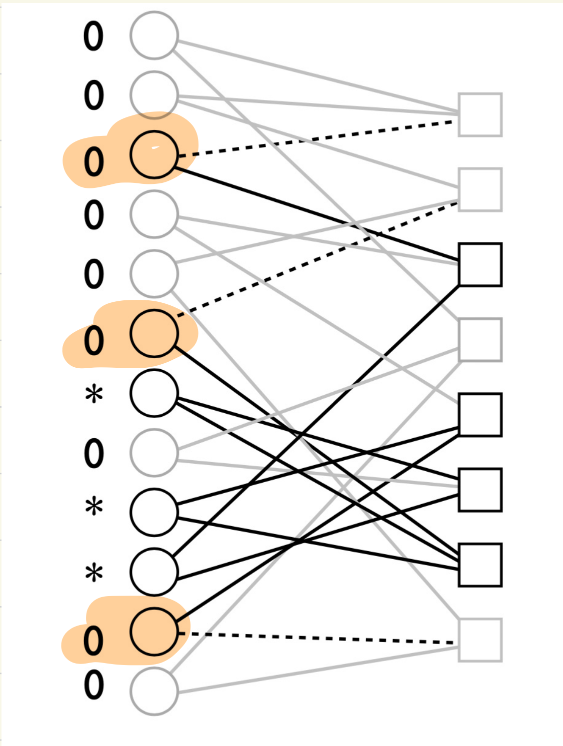


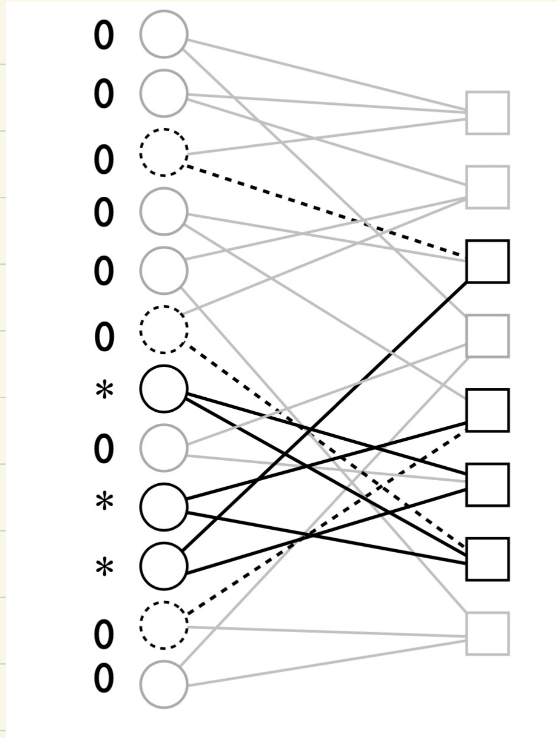
$x_1 = 0$

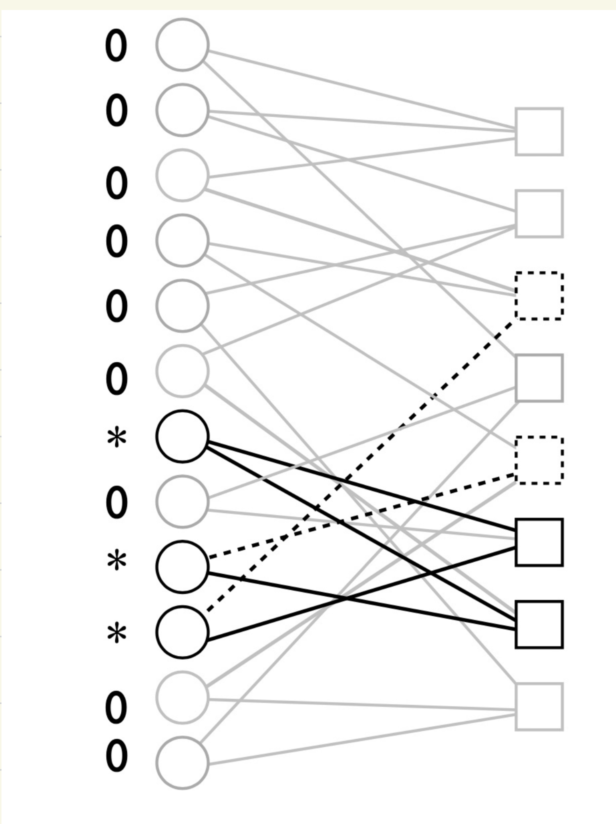


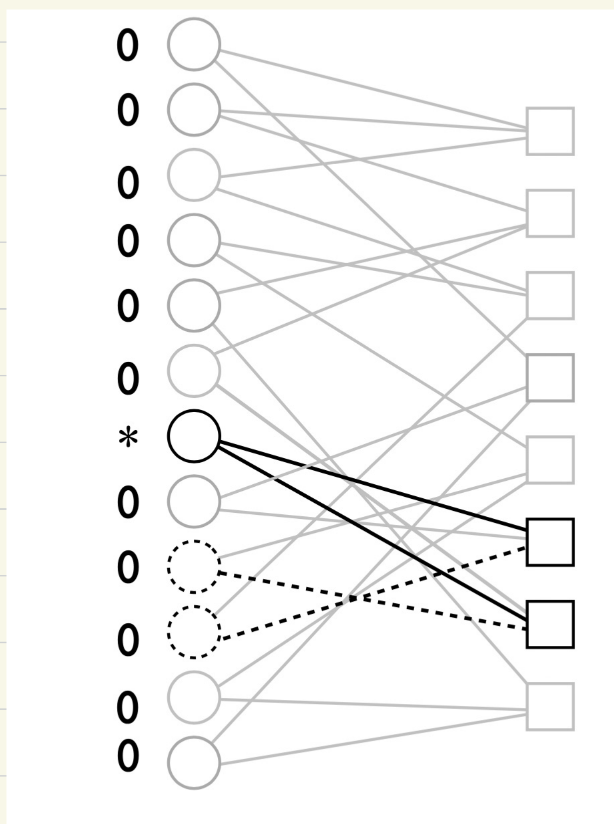


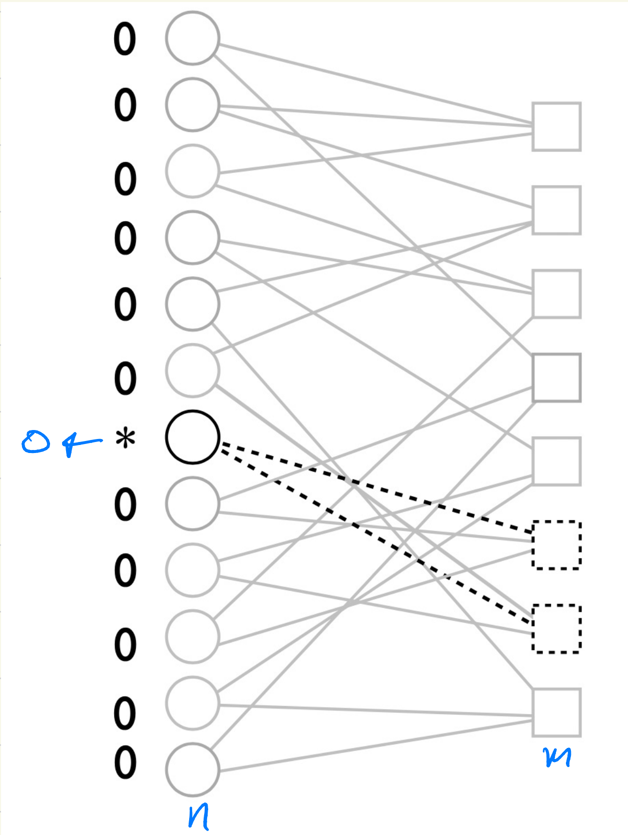
* factor nodes with 1 remaining edge can be decoded.


0
0
0
0
0
0
*
0
*
*
0
0


0
0
0
0
0
0
*
0
*
*
0
0

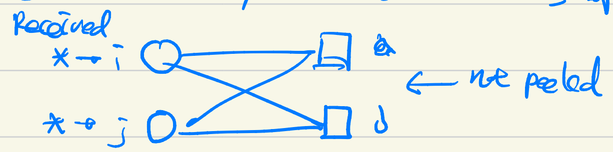

0
0
0
0
0
0
*
0
*
*
0
0


0
0
0
0
0
0
*
0
0
0
0
0



* If all nodes are peeled, then Decoding success.

* If a subset of nodes are left whose uncertainty is not resolves ef



then those bits cannot be decoded.

* Design Goal: for a given n & m, find a graph that minimizes the error probability.

↑ Density Evolution is critical in achieving this breakthrough

---

Strategy under Random Factor Graph with $m, n \to \infty$.

① suppose we update BP as follows.

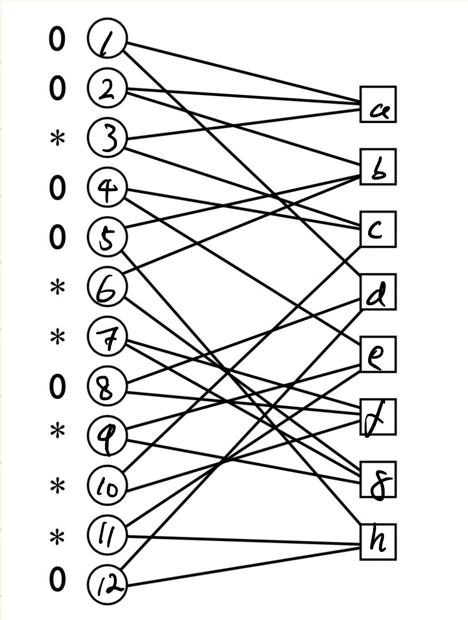
repeat:
- Update $\{m_{i\to a}^{(t)}\}$
- Draw fresh random graph
- Update $\{\hat{m}_{a\to i}^{(t)}\}$
- Draw fresh random graph
- $t = t+1$

and analyze this process

② Justify it is accurate if FG is random & $m, n \uparrow \infty$. for fixed $t \le \log n$

* Computation tree for $m^{(t)}_{i\to a}(x_i)$

message from $i \xrightarrow{\text{to}} a$, after t-iterations of BP.

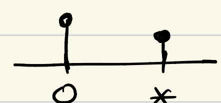


for t=1

$$\begin{pmatrix} m^{(1)}_{1\to a}(x_1) = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \\ \vdots \\ m^{(1)}_{3\to c}(x_3) = \begin{bmatrix} 1/2 \\ 1/2 \end{bmatrix} \\ \vdots \end{pmatrix} \Rightarrow \begin{pmatrix} x^{(1)}_{1\to a} = 0 \\ \vdots \\ x^{(1)}_{3\to c} = * \\ \vdots \end{pmatrix}$$

Set. Set $\{x^{(1)}_{i\to a}\}$

Histogram of this $\{x^{(1)}_{i\to a}\}$



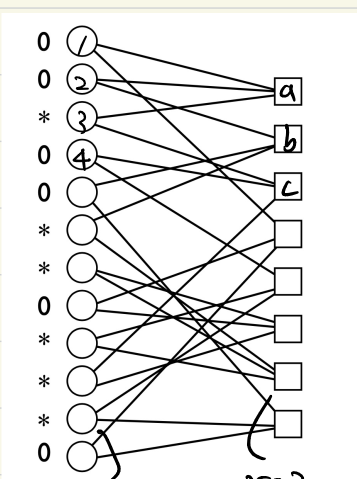
Density : $\hat{\mathbb{P}}^{(1)}_{G_{m,n}, BEC(\varepsilon)}(x=*)$ $\begin{matrix} 0 \\ * \end{matrix}\begin{bmatrix} \\ \bigcirc \end{bmatrix} = \xi^{(1)}$ keep.

for a randomly chosen edge

Q. $\mathbb{E}\left[\hat{\mathbb{P}}^{(1)}_{G_{m,n},BEC(\varepsilon)}\right] =$

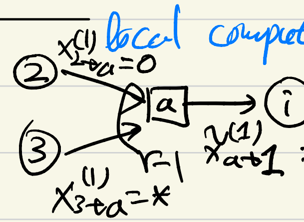
Q. $\lim_{m,n\to\infty} \hat{\mathbb{P}}^{(1)}_{G_{m,n},BEC(\varepsilon)} =$

for $t=1$, $\tilde{m}^{(1)}_{a\to i}(x_i)$



$\tilde{x}_{a\to 1} \in \{0, 1, *\}$

local computation tree. at depth t=1



$$\tilde{x}^{(1)}_{a\to 1} = \begin{cases} * & \text{if at least 1 star} \\ \bigoplus x_{\partial a \setminus i \to a} \end{cases} = *$$

Histogram $\{\tilde{x}^{(1)}_{a\to i}\} = \tilde{\mathbb{P}}^{(1)}_{G_{m,n}, BEC(\varepsilon), l, r}(x=*) = \tilde{\xi}^{(1)}$

$$\mathbb{P}(\tilde{X}^{(1)}_{a\to 1} = *) = \mathbb{P}(X^{(1)}_{2\to a} = * \text{ or } X^{(1)}_{3\to a} = *)$$

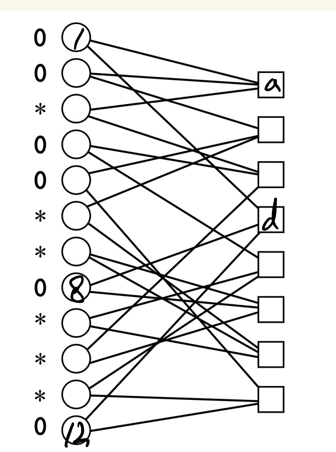
($X^{(1)}_{2\to a} \leftarrow z^{(1)}$)

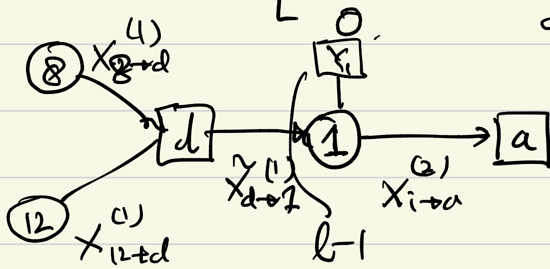
$$\tilde{z}^{(1)} = 1 - (1 - z^{(1)})^{r-1}, \quad r-1 = 2$$

Completely justified if
$\{m^{(1)}_{i\to a}\}$ was computed
and then we draw random edges again,
and compute $\{\tilde{m}^{(1)}_{a\to i}\}$ and repeat...

$$z^{(2)} = \mathbb{P}(X^{(2)}_{i\to a} = *) = (\tilde{z}^{(1)})^{l-1} \cdot \varepsilon$$

$$X^{(2)}_{i\to a} = \begin{cases} * & \text{incoming is } * \text{ \& } Y_i = * \\ 0 & \text{otherwise} \end{cases}$$





Evolution of messages		Evolution of density (Histogram)
$\{m^{(1)}_{i\to a}\}$	⟷	$z^{(1)} \in [0,1]$
$\{\tilde{m}^{(1)}_{a\to i}\}$		$\tilde{z}^{(1)}$
$\{m^{(2)}_{i\to a}\}$		$z^{(2)}$
$\{\tilde{m}^{(2)}_{a\to i}\}$		$\tilde{z}^{(2)}$
⋮		

$z^{(1)} = \varepsilon$

$\tilde{z}^{(t)} = 1 - (1 - z^{(t)})^{r-1}$

$z^{(t)} = \varepsilon \cdot (\tilde{z}^{(t-1)})^{l-1}$

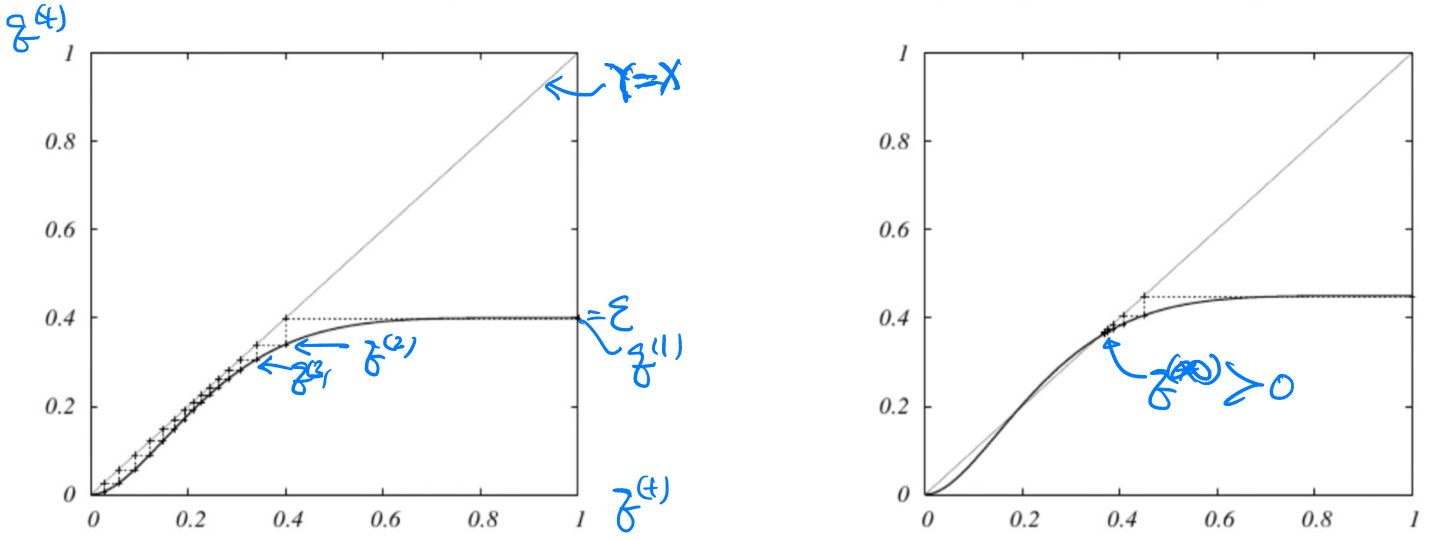
Density Evolution Update

$$z^{(1)} = \varepsilon, \quad z^{(t)} = \varepsilon \cdot \left(1 - (1 - z^{(t-1)})^{r-1}\right)^{l-1}$$

$$z^{(t+1)} = \varepsilon \cdot \left(1 - (1 - z^{(t)})^{r-1}\right)^{\ell-1}$$

density evolution for (3,6) code with $\epsilon = 0.4$(left) and 0.45(right)

($\ell = 3$, $r = 6$)



rate of this code = 0.5, threshold $\epsilon^* \simeq 0.4$xxx,

Eventually $z^{(t)} \to 0$.
$\mathbb{P}(X_{i\to a} = *) \to 0$
Perfect decoding.
if $\varepsilon = 0.4$

$z^{(\infty)} > 0$, 0.38xxx
$\lim_{t\to\infty} \lim_{n\to\infty} \mathbb{P}(X_{i\to a} = *) = 0.38$
if $\varepsilon = 0.45$

bit-error-rate.

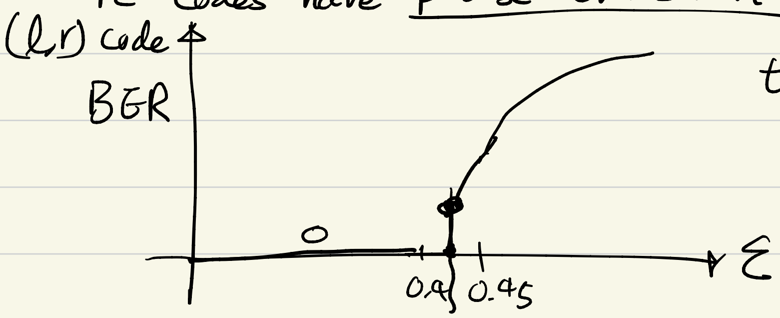
$\mathbb{P}(X_i = *) = \varepsilon \cdot \left(1 - (1 - z)^{r-1}\right)^{\ell}$ (product ℓ)

$= \prod_{a \in \partial i} \hat{x}_{a\to i} =$

---

* Important Consequence ①.

LDPC Codes have phase transition



the asymptotic BER curve has a discontinuity

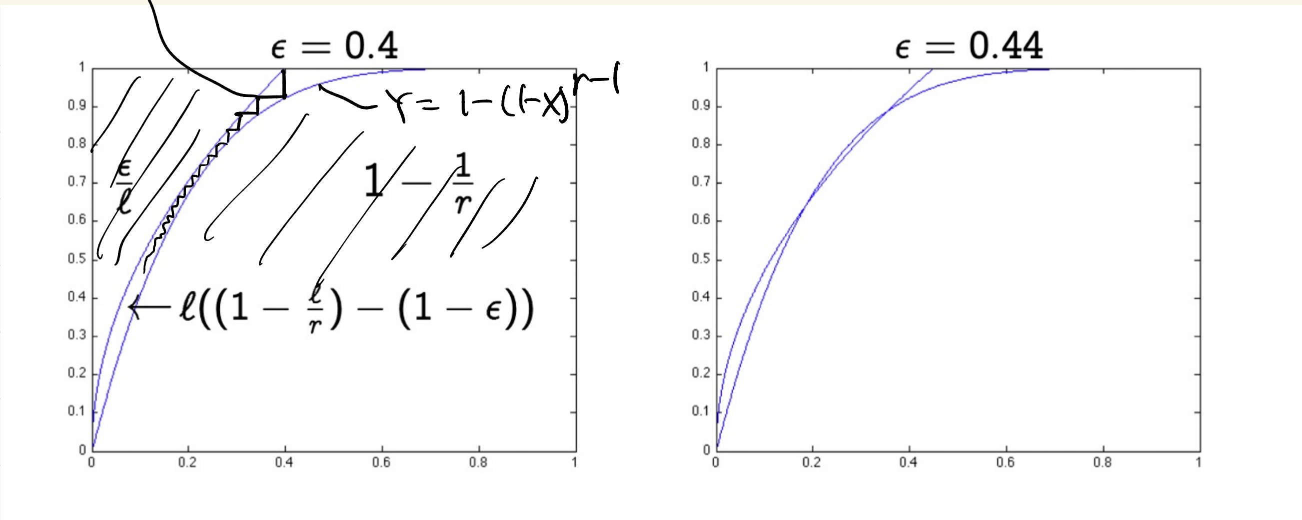
ε^* ← can compute with density evolution

* Important Consequence ②

we had $\xi = \varepsilon(1-(1-\xi)^{r-1})^{\ell-1}$ ← the solution is the final error rate.

let's change it to $\left(\frac{\xi}{\varepsilon}\right)^{\frac{1}{\ell-1}} = 1-(1-\xi)^{r-1}$

$Y=\left(\frac{x}{\varepsilon}\right)^{\frac{1}{\ell-1}}$ plot $\quad \| \; Y \quad\quad \| \; Y$



If we can design degree distribution $\ell = [\ell_1 \; \ell_2 \; \ell_3 \; \ldots \; \ell_{max}]$
$\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\; 0 \;\; 0.3 \;\; 0.4$
$r = [\quad\quad\quad\quad]$

such that the Gap closes ⟶ then we have a capacity achieving code.

Claim: for fixed ℓ, r, t, the random graph is locally-treelike with probability one, as $m, n \to \infty$.

$\lim_{n\to\infty} \mathbb{P}(\text{local tree of depth } t \text{ is a tree}) \to 0$

⇒ t-iterations of BP is performed on a random tree, so the 2 processes are statistically equivalent.

Gaussian Graphical Models.

So far we learned discrete distributions in $\mathcal{X}^n$

- any factor $f_a(x_1, \dots, x_l)$ can be written in a table of size $|\mathcal{X}|^l$.
- any algorithm we learned only uses $+$, $\times$, and accesses to these tables.

Now. we learn continuous distributions in $\mathbb{R}^n$

- we need a parametric family to be able to efficiently store the local factors and compute messages.

Def. $X = [X_1, \dots, X_n]$ is Gaussian ($N(\mu, \Sigma)$) if
$\mu \in \mathbb{R}^n$, $\Sigma \in \mathbb{R}^{n \times n}$

$$P(x) = \frac{1}{(2\pi)^{n/2} |\Sigma|^{1/2}} \cdot \exp\left(-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)\right)$$

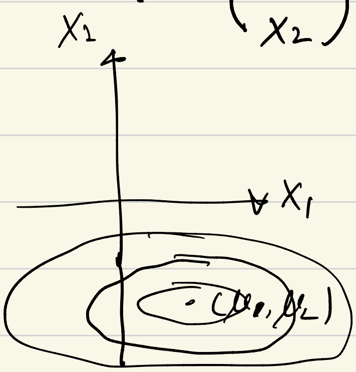
$P(x)$ ↑ Probability Density Function

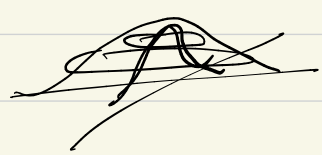
$|\Sigma|$ ↑ determinant $= \prod_i \sigma_i(\Sigma)$

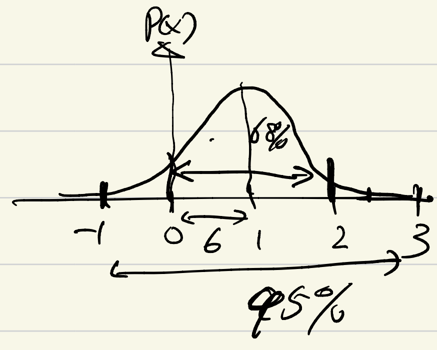
where $\mu = \mathbb{E}_p[X]$, and $\Sigma = \mathrm{Cov}(X) = \mathbb{E}[(X-\mu)(X-\mu)^T]$

- Σ is symmetric and positive-definite
  - $x^T \Sigma x > 0$, $\forall x \neq 0$.
  - $\sigma_{min}(\Sigma) > 0$, → invertible.

ex> $\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim N\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}\right)$, $X_1 \sim N(1, 1 = \sigma^2)$







Def. Covariance form of a multivariate Gaussian is

$$X \sim N(\mu, \Sigma) \iff P(x) = \frac{1}{(2\pi)^{n/2} \cdot |\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)\right)$$

(Σ: P.D.)

the exponent is $-\frac{1}{2} x^T \Sigma^{-1} x + \frac{1}{2}\left(x^T \Sigma^{-1} \mu + \mu^T \Sigma^{-1} x\right) + C.$

$$= -\frac{1}{2} x^T \Sigma^{-1} x + (\Sigma^{-1}\mu)^T \cdot x + C$$

Def. Information form of a multivariate Gaussian is

$$X \sim N^{-1}(h, J) \longleftrightarrow P(x) = \frac{1}{Z_{h,J}} \exp\left(-\frac{1}{2} x^T J x + h^T x\right)$$

(J: P.D.)

claim. $N(\mu, \Sigma) = N^{-1}(h, J)$ iff. $\begin{cases} \Sigma^{-1} = J \\ \Sigma^{-1}\mu = h. \end{cases}$

Remark. ① marginalization is easy with covariance form.

② conditioning is easy with Information form.

① $\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \sim N\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}\right)$ ($X_1 \in \mathbb{R}^{n_1}$, $X_2 \in \mathbb{R}^{n_2}$)

What is marginal distribution of $P(x_1) = ?$

$X_1 \sim N(\mu_1, \Sigma_{11})$

proof> $P(x_1) = \int P(x_1, x_2)\, dx_2 \ldots$

but by definition

$\mathbb{E}[X_1] = \mu_1$

$\mathbb{E}[(X_1-\mu_1)(X_1-\mu_1)^T] = \Sigma_{11}$

Just need to prove that X_1 is Gaussian.

$\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \sim N^{-1}\left(\begin{bmatrix} h_1 \\ h_2 \end{bmatrix}, \begin{bmatrix} J_{11} & J_{12} \\ J_{21} & J_{22} \end{bmatrix}\right)$

what is marginal $P(X_1)$ in terms of $h_1, h_2, J_{11}, J_{12}, J_{21}, J_{22}$?

$$X_1 \sim N^{-1}\left(h_1 - J_{12} J_{22}^{-1} h_2,\ J_{11} - J_{12} J_{22}^{-1} J_{21}\right)$$

complicated → Requires J_{22}^{-1} which can be costly what is conditional dist. $P(x_1 | x_2) = ?$

② Covariance form

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \sim N\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \right)$$

$$P(x_1 | x_2) \sim N\left(\mu_1 + \Sigma_{12} \Sigma_{22}^{-1} (x_2 - \mu_2), \; \Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} \right).$$

* complicated → Requires Σ_{22}^{-1} which can be costly.

Information form

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \sim N^{-1}\left(\begin{bmatrix} h_1 \\ h_2 \end{bmatrix}, \begin{bmatrix} J_{11} & J_{12} \\ J_{21} & J_{22} \end{bmatrix} \right)$$

$$P(x_1 | x_2) \sim N^{-1}\left(h_1 - J_{12} x_2, \; J_{11} \right)$$

proof>

$$P(x_1 | x_2) \propto \exp\left(-\frac{1}{2} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}^T \begin{bmatrix} J_{11} & J_{12} \\ J_{21} & J_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} h_1^T & h_2^T \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \right)$$

$$\propto \exp\left(-\frac{1}{2} x_1^T J_{11} x_1 + \left(-\frac{1}{2} J_{12} x_2 - \frac{1}{2} J_{21}^T x_2 + h_1 \right)^T x_1 \right)$$

Claim. For Covariance form, independence is easy to read.
For Information form, conditional independence is easy to read.