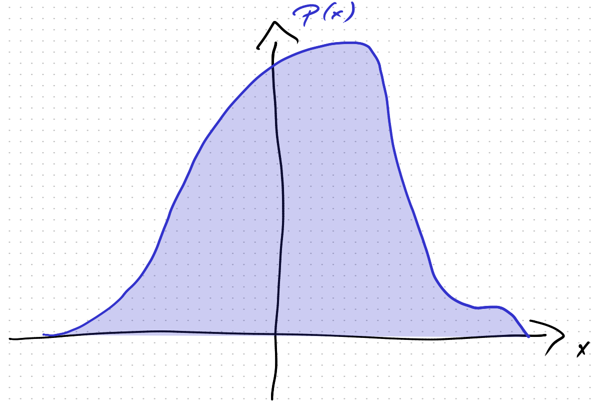
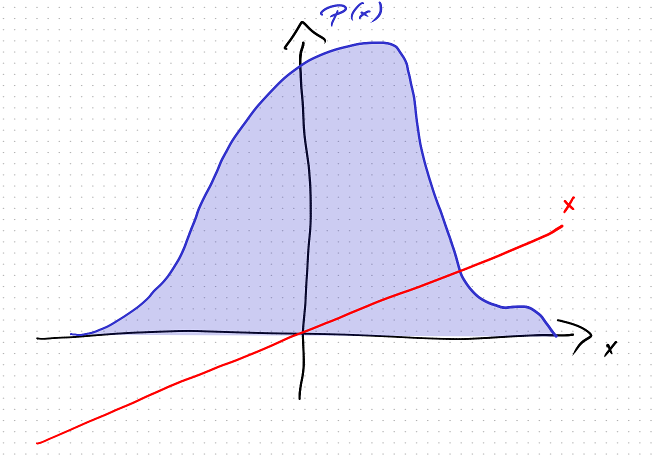
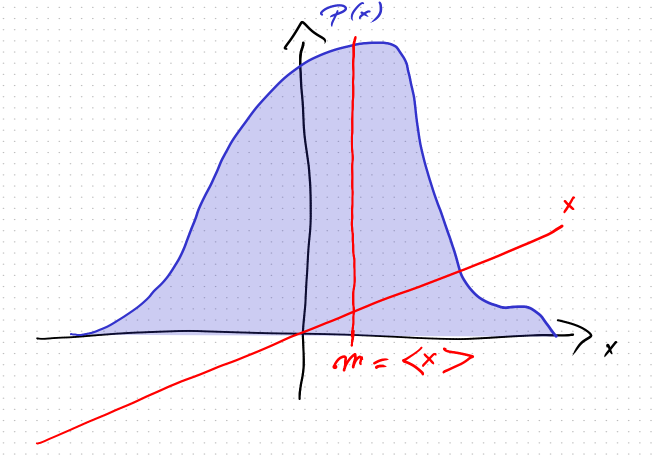
4.8 Maximum Entropy with known 1st and 2nd Moments



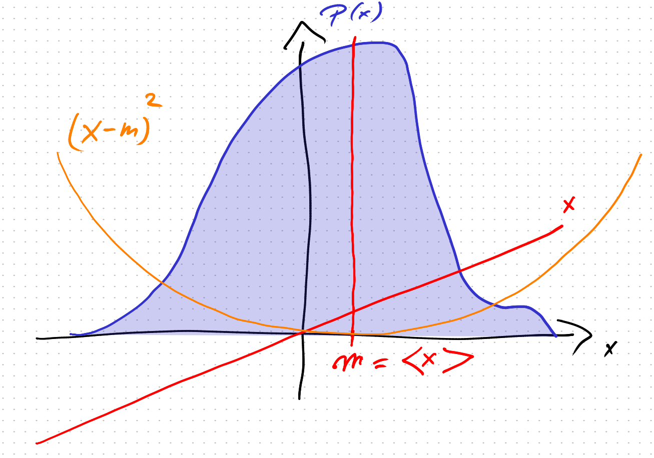
4.8 Maximum Entropy with known 1st and 2nd Moments



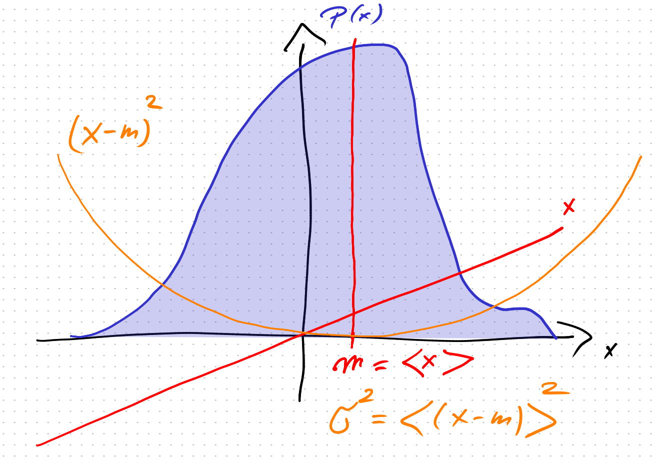
4.8 Maximum Entropy with known 1st and 2nd Moments



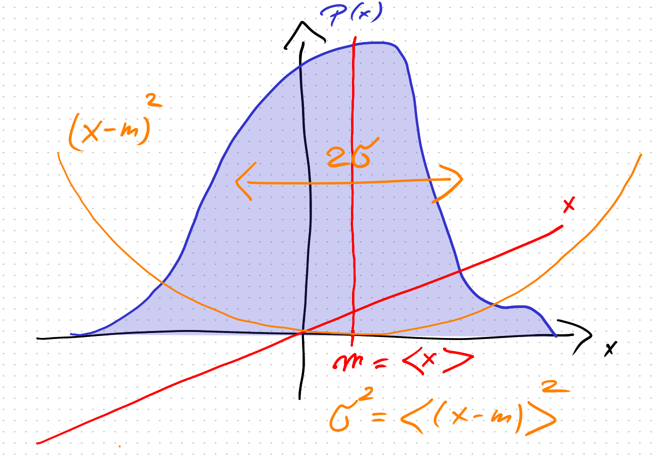
4.8 Maximum Entropy with known 1st and 2nd Moments



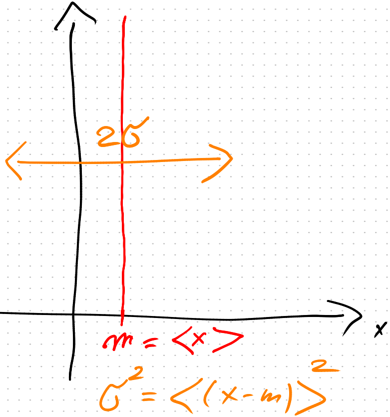
4.8 Maximum Entropy with known 1st and 2nd Moments



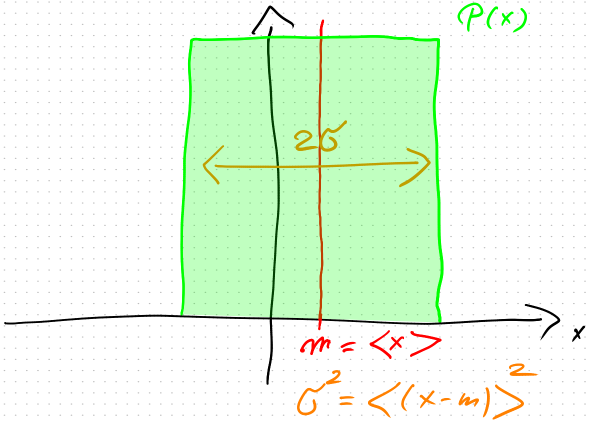
4.8 Maximum Entropy with known 1st and 2nd Moments



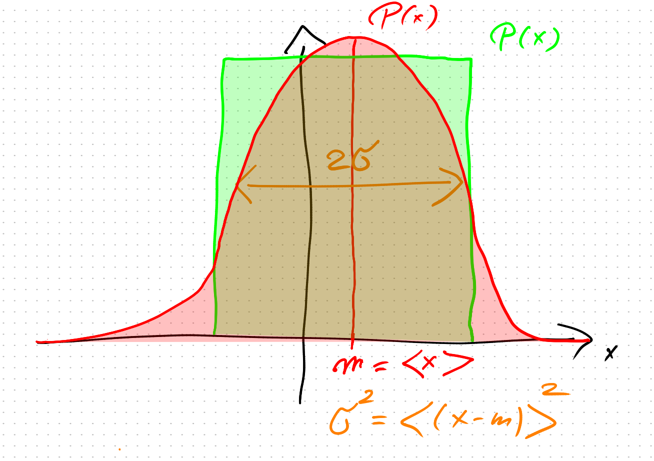
4.8 Maximum Entropy with known 1st and 2nd Moments



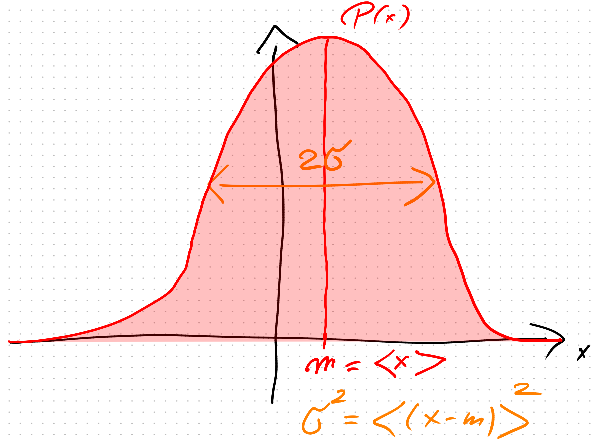
4.8 Maximum Entropy with known 1st and 2nd Moments



4.8 Maximum Entropy with known 1st and 2nd Moments



4.8 Maximum Entropy with known 1st and 2nd Moments



4.8 Maximum Entropy with known 1^{st} and 2^{nd} Moments

Prior information I: $x \in \mathbb{R}$
Prior knowledge: $q(x) := \mathcal{P}(x|I) = \text{const.}$
Updating information J: $\langle x \rangle_{(x|J,I)} = m$, $\langle (x-m)^2 \rangle_{(x|J,I)} = \sigma^2$
Posterior knowledge: $p(x) := \mathcal{P}(x|J,\,I) = \frac{e^{\alpha x + \beta (x-m)^2}}{\mathcal{Z}(\alpha,\,\beta)}$
1. calculate $\mathcal{Z}(\alpha,\,\beta)$:

$$
\begin{aligned}
\mathcal{Z}(\alpha,\,\beta) &= \int_{-\infty}^{\infty} dx\, e^{\alpha x + \beta (\underbrace{x-m}_{=x'})^2} \\
&= \int_{-\infty}^{\infty} dx'\, e^{\alpha x' + \alpha m + \beta x'^2}
\end{aligned}
$$

4.8 Maximum Entropy with known 1st and 2nd Moments

$$\mathcal{Z}(\alpha,\,\beta) = \int_{-\infty}^{\infty} dx'\, e^{\alpha x' + \alpha m + \beta x'^2}$$

Completing the square:
$$= e^{\alpha m} \int_{-\infty}^{\infty} dx'\, e^{\beta\left(x'^2 + \frac{2\alpha x'}{2\beta} + \frac{\alpha^2}{(2\beta)^2}\right) - \frac{\alpha^2}{4\beta}}$$

$$= e^{\alpha m - \frac{\alpha^2}{4\beta}} \int_{-\infty}^{\infty} dx'\, e^{\beta\left(x' + \frac{\alpha}{2\beta}\right)^2}$$

Claiming $\beta < 0$:
$$= e^{\alpha m + \frac{\alpha^2}{4\,|\beta|}} \int_{-\infty}^{\infty} dx'\, e^{-|\beta|\,\left(x' - \frac{\alpha}{2|\beta|}\right)^2}$$

$$= e^{\alpha m + \frac{\alpha^2}{4\,|\beta|}} \sqrt{\frac{\pi}{-\beta}}$$

4.8 Maximum Entropy with known 1st and 2nd Moments

2. determine α and β:

$$
\begin{aligned}
\ln \mathcal{Z}(\alpha,\, \beta) &= \alpha m - \frac{\alpha^2}{4\beta} + \frac{1}{2}\ln\left(\frac{\pi}{-\beta}\right) \\
\frac{\partial \ln \mathcal{Z}(\alpha,\, \beta)}{\partial \alpha} &= m - \frac{\alpha}{2\beta} \stackrel{!}{=} m \\
\Rightarrow \alpha &= 0 \\
\frac{\partial \ln \mathcal{Z}(\alpha = 0,\, \beta)}{\partial \beta} &= -\frac{1}{2\beta} \stackrel{!}{=} \sigma^2 \\
\Rightarrow \beta &= -\frac{1}{2\sigma^2}
\end{aligned}
$$

Insert in $\mathcal{Z}(\alpha,\, \beta)$:

$$
\mathcal{Z} = \sqrt{2\pi\sigma^2}
$$

4.8 Maximum Entropy with known 1st and 2nd Moments

3. calculate $p(x) = \mathcal{P}(x|J, I)$**:**

$$
\begin{aligned}
P(x|J, I) &= \left.\frac{e^{\alpha x+\beta(x-m)^2}}{\mathcal{Z}(\alpha, \beta)}\right|_{\alpha=0,\beta=-1/(2\sigma^2)} \\
&= \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-m)^2}{2\sigma^2}} \\
&= \mathcal{G}(x-m, \sigma^2)
\end{aligned}
$$

$\Rightarrow$ Maximum Entropy PDF $P(x|J, I)$ for known 1st and 2nd moments (and flat prior) is Gaussian distribution

5 Gaussian Distribution

- maximum Entropy solution if only 1^{st} and 2^{nd} moments known
- emerges according to central limit theorem
- mathematically convenient

5.1 One dimensional Gaussian distribution:

$$\mathcal{G}(x-m,\,\sigma_x^2) \quad = \quad \frac{1}{\sqrt{2\pi\sigma_x^2}} \exp\left(-\frac{(x-m)^2}{2\,\sigma_x^2}\right)$$

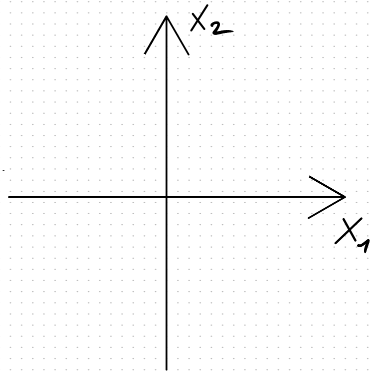
5.2 Multivariate Gaussian Distribution

$x = (x_1, \dots x_n)^{\mathrm{t}}$: zero centered independent Gaussian distributed variables
$\sigma_1^2, \dots \sigma_n^2$: corresponding variances
$X = \mathrm{diag}(\sigma_1^2, \dots \sigma_n^2)$: diagonal covariance matrix

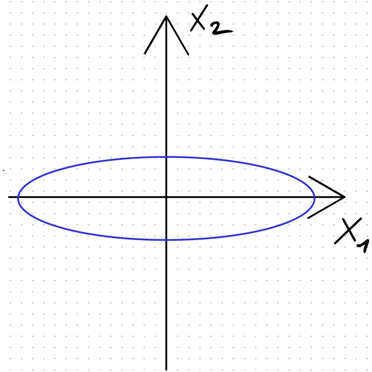
$$\begin{aligned}
\text{Joint probability: } \mathcal{P}(x) &= \prod_{i=1}^{n} \mathcal{P}(x_i) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{(x_i)^2}{2\,\sigma_i^2}\right) \\
&= \frac{1}{\prod_{i=1}^{n} \sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{1}{2}\sum_{i=1}^{n} \frac{x_i^2}{\sigma_i^2}\right) = \frac{1}{\sqrt{|2\pi X|}} \exp\left(-\frac{1}{2} x^{\dagger} X^{-1} x\right)
\end{aligned}$$

$$\text{Multivariate Gaussian: } \mathcal{G}(x,\, X) = \frac{1}{\sqrt{|2\pi X|}} \exp\left(-\frac{1}{2} x^{\dagger} X^{-1} x\right)$$

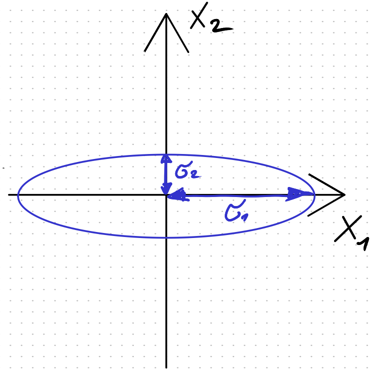
Orthonormal transformation



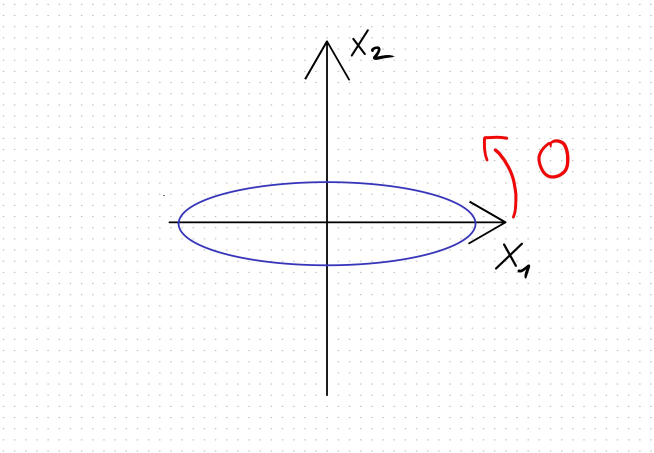
Orthonormal transformation



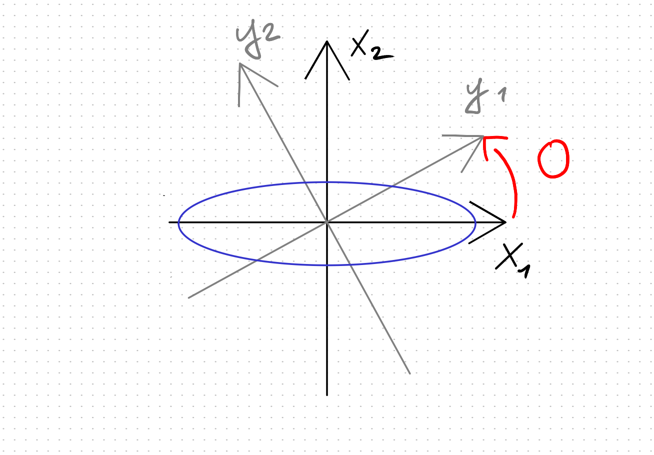
Orthonormal transformation



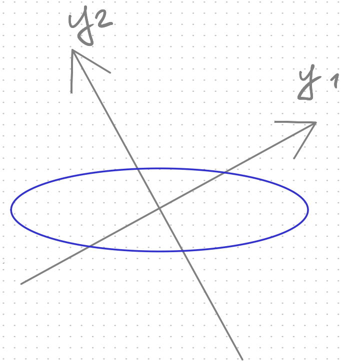
Orthonormal transformation



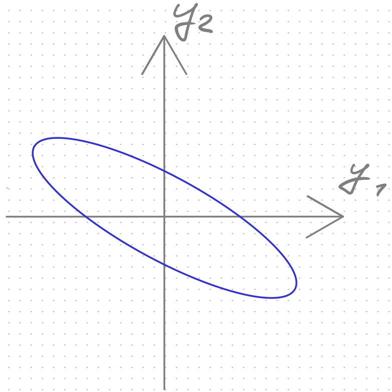
Orthonormal transformation



Orthonormal transformation



Orthonormal transformation



From independent to dependent coordinates

Orthonormal basis transformation in n-dim. space:

$$\begin{aligned} y &= O\,x \\ O^{-1} &= O^{\dagger} \end{aligned}$$

$$\begin{aligned} &\Rightarrow \quad |O| = |O^{\dagger}| = |O^{-1}| = 1/|O| \\ &\Rightarrow \quad |O|^2 = 1 \\ &\Rightarrow \quad \|O\| = \|O^{\dagger}\| = 1 \end{aligned}$$

Conservation of probability mass:

$$\mathcal{P}(y|I)\,dy \quad = \quad \mathcal{P}(x|I)\,dx\big|_{x=O^{\dagger}y}$$

From independent to dependent coordinates

$$
\begin{aligned}
\Rightarrow \mathcal{P}(y|I) &= \mathcal{G}(x, X) \left\| \frac{\partial x}{\partial y} \right\| \Bigg|_{x=O^\dagger y} = \mathcal{G}(O^\dagger y, X) \underbrace{\|O^\dagger\|}_{=1} \\
&= \frac{1}{\sqrt{|2\pi X|}} \exp\left(-\frac{1}{2} \underbrace{(O^\dagger y)^\dagger}_{x^\dagger = y^\dagger O} X^{-1} \underbrace{O^\dagger y}_{x} \right) \\
&= \frac{1}{\sqrt{|2\pi X|}} \exp\left(-\frac{1}{2} y^\dagger \underbrace{O X^{-1} O^\dagger}_{Y^{-1}} y \right) \\
&= \frac{1}{\sqrt{|2\pi X|}} \exp\left(-\frac{1}{2} y^\dagger Y^{-1} y \right)
\end{aligned}
$$

From independent to dependent coordinates

$$
\begin{aligned}
|Y| &= |Y^{-1}|^{-1} \\
&= |OX^{-1}O^{\dagger}|^{-1} \\
&= (\underbrace{|O|}_{=\pm 1}\,|X^{-1}|\,\underbrace{|O^{\dagger}|}_{=\pm 1})^{-1} \\
&= |X|
\end{aligned}
$$

Generic multivariate Gaussian:

$$\Rightarrow \mathcal{P}(y) = \mathcal{G}(y,\, Y) = \frac{1}{\sqrt{|2\pi Y|}} \exp\left(-\frac{1}{2} y^{\dagger} Y^{-1} y\right)$$

Moments of the multivariate Gaussian

Normalization:

$$
\begin{aligned}
\langle 1 \rangle_{\mathcal{G}(y,\,Y)} &= \int dy\, 1\, \mathcal{G}(y,\, Y) = \int dx\, 1\, \mathcal{G}(x,\, X) = 1 \\
1 &= \frac{1}{\sqrt{|2\pi Y|}} \underbrace{\int dy \,\exp\left(-\frac{1}{2} y^\dagger Y^{-1} y\right)}_{=\sqrt{|2\pi Y|}}
\end{aligned}
$$

1^{st} Moment:

$$
\begin{aligned}
\langle y \rangle_{\mathcal{G}(y,\,Y)} &= \int dy\, y\, \mathcal{G}(y,\, Y) \\
&= \int dy' \,(-y')\, \mathcal{G}(-y',\, Y) \| -\mathbb{1} \| \\
&= -\langle y' \rangle_{\mathcal{G}(y',\,Y)} = 0
\end{aligned}
$$

Moments of the multivariate Gaussian

2nd Moment:

$$\left\langle y\, y^\dagger \right\rangle_{\mathcal{G}(y,Y)} = \int dy\, y\, y^\dagger \mathcal{G}(y,\, Y)$$

$$= \int dx\, \mathcal{G}(x,\, X) O x x^\dagger O^\dagger$$

$$= O \int dx\, x\, x^\dagger \mathcal{G}(x,\, X)\, O^\dagger \stackrel{?}{=} O X O^\dagger = Y$$

$$\int dx\, x_i\, x_j\, \mathcal{G}(x,\, X) = \left[\prod_{k=1}^{n} \int dx_k \mathcal{G}(x_k, \sigma_k^2) \right] x_i\, x_j$$

$$= \begin{cases} \left[\int dx_i \mathcal{G}(x_i, \sigma_i^2)\, x_i \right] \left[\int dx_j \mathcal{G}(x_j, \sigma_j^2)\, x_j \right] & \text{if } i \neq j \\ \int dx_i \mathcal{G}(x_i, \sigma_i^2)\, x_i^2 & \text{if } i = j \end{cases}$$

$$= \begin{cases} 0 & \text{if } i \neq j \\ \sigma_i^2 & \text{if } i = j \end{cases} = \delta_{ij} \sigma_i^2 = X_{ij} \;\square$$

Moments of the multivariate Gaussian

$$\begin{aligned}
\langle y \rangle_{\mathcal{G}(y,\,Y)} &= 0 \\
\langle f(y) \rangle_{\mathcal{G}(y,\,Y)} &= 0, \ \text{if} \, f(-y) = -f(y) \\
\langle y y^{\dagger} \rangle_{\mathcal{G}(y,\,Y)} &= Y
\end{aligned}$$

Wick theorem

Wick theorem:

$\mathbb{P}$: set of all possible ways to partition $\{i_1, \ldots, i_{2n}\}$ into pairs

$$\langle y_{i_1} \ldots y_{i_{2n}} \rangle_{\mathcal{G}(y,\,Y)} = \langle \prod_{j=1}^{2n} y_{i_j} \rangle_{\mathcal{G}(y,\,Y)} = \sum_{p \in \mathbb{P}} \prod_{(i',j') \in p} Y_{i_{i'} i_{j'}}$$

Examples:

- $\langle y_{i_1} y_{i_2} \rangle_{\mathcal{G}(y,\,Y)} = Y_{i_1 i_2}$
- $\langle y_{i_1} y_{i_2} y_{i_3} y_{i_4} \rangle_{\mathcal{G}(y,\,Y)} = Y_{i_1 i_2} Y_{i_3 i_4} + Y_{i_1 i_3} Y_{i_2 i_4} + Y_{i_1 i_4} Y_{i_2 i_3}$

$$\Rightarrow \quad \langle y_i^{2n} \rangle_{\mathcal{G}(y,\,Y)} = \frac{(2n)!}{2^n n!} (Y_{ii})^n$$

$$\Rightarrow \quad \langle y_i^{2n+1} \rangle_{\mathcal{G}(y,\,Y)} = 0$$

Maximum Entropy with known n-dimensional 1st and 2nd Moments

Prior information I: $s \in V$ (e.g. $\mathbb{R}$, $\mathbb{R}^n$, $C(\mathbb{R}^n)$)
Prior knowledge: $q(s) := \mathcal{P}(s|I) = \text{const.} = 1$
Updating information J: $\langle s \rangle_{(s|J,I)} = m$, $\langle (s-m)(s-m)^\dagger \rangle_{(s|J,I)} = S$
Posterior: $p(s) = \frac{1}{\mathcal{Z}} \exp[\sum_i \mu_i (s-m)_i + \sum_{ij} \Lambda_{ij} \underbrace{((s-m)_i(s-m)_j - S_{ji})}_{=B_{ji}(s)}]$

1. calculate $\mathcal{Z}(\mu, \Lambda)$:

$$
\begin{aligned}
\mathcal{Z}(\mu, \Lambda) &= \int ds \exp\left[\mu^\dagger \underbrace{(s-m)}_{s'} + \text{Tr}[\Lambda B(s)]\right] \\
&= \int ds' \exp\left[\mu^\dagger s' + \text{Tr}[\Lambda(s' s'^\dagger - S)]\right] \\
&= \int ds' \exp\left[\mu^\dagger s' + s'^\dagger \Lambda s' - \text{Tr}[\Lambda S]\right] = e^{-\text{Tr}[\Lambda S]} \int ds' \, e^{\mu^\dagger s' + s'^\dagger \Lambda s'}
\end{aligned}
$$

Maximum Entropy with known n-dimensional 1st and 2nd Moments

2. determine μ and Λ:

$$\ln \mathcal{Z}(\mu,\, \Lambda) = -\mathrm{Tr}[\Lambda S] + \ln\left(\int ds' \, \exp(\mu^\dagger s' + s'^\dagger \Lambda s')\right)$$

$$\frac{\partial \ln \mathcal{Z}(\mu,\, \Lambda)}{\partial \mu} = \left(\frac{\partial \ln \mathcal{Z}}{\partial \mu_i}\right)_i = \frac{\int ds' \, s' \exp(\mu^\dagger s' + s'^\dagger \Lambda s')}{\int ds' \, \exp(\mu^\dagger s' + s'^\dagger \Lambda s')} \stackrel{!}{=} 0$$

$$\Rightarrow \mu = 0$$

$$\frac{\partial \ln \mathcal{Z}(\mu,\, \Lambda)}{\partial \Lambda} = \left(\frac{\partial \ln \mathcal{Z}}{\partial \Lambda_{ij}}\right)_{ij} = \underbrace{-(S_{ji})_{ij}}_{=-S} + \left(\frac{\int ds' \, s'_i s'_j \exp(s'^\dagger \Lambda s')}{\int ds' \, \exp(s'^\dagger \Lambda s')}\right)_{ij} \stackrel{!}{=} 0$$

$$\Rightarrow S = \frac{\int ds' \, s' s'^\dagger \exp\left(-\frac{1}{2} s'^\dagger (-\frac{1}{2}\Lambda^{-1})^{-1} s'\right)}{\int ds' \exp\left(-\frac{1}{2} s'^\dagger (-\frac{1}{2}\Lambda^{-1})^{-1} s'\right)} = \frac{\int ds' \, s' s'^\dagger \mathcal{G}\left(s', -\frac{1}{2}\Lambda^{-1}\right)}{\int ds' \mathcal{G}\left(s', -\frac{1}{2}\Lambda^{-1}\right)}$$

$$= -\frac{1}{2}\Lambda^{-1} \Rightarrow \Lambda = -\frac{1}{2} S^{-1}$$

Maximum Entropy with known n-dimensional 1st and 2nd Moments

Insert in $\mathcal{Z}(\mu, \Lambda)$:

$$\begin{aligned}\mathcal{Z}(\mu, \Lambda) &= \int ds' \exp\left[-\frac{1}{2} s'^{\dagger} S^{-1} s' + \frac{1}{2} \mathrm{Tr}[\underbrace{S^{-1} S}_{=\mathbb{1}}] \right] \\ &= |2\pi S|^{1/2} e^{\frac{1}{2}\mathrm{Tr}[\mathbb{1}]}\end{aligned}$$

3. calculate $p(s) = \mathcal{P}(s|J, I)$**:** remember: $s' = s - m$

$$\begin{aligned}P(s|J, I) &= \frac{1}{\sqrt{|2\pi S|}} \exp\left(-\frac{1}{2} (s-m)^{\dagger} S^{-1} (s-m) \right) \\ &= \mathcal{G}(s-m, S)\end{aligned}$$

⇒use Gaussian distribution $\mathcal{G}(s-m, S)$ given the n-dim. mean m and variance S