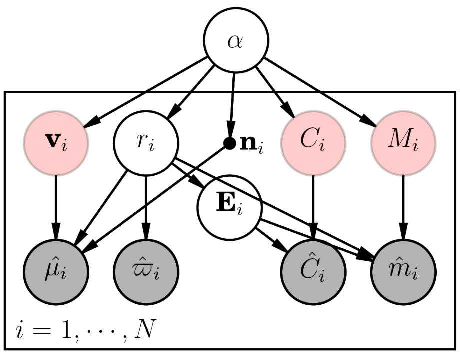
The ultimate (Gaussian) stellar inference machine

This is a brief note describing a model of the positions, velocities, proper motions, and colors of stars. If the likelihood function of those is a multivariate Gaussian (which is the case for Gaia, with strong correlations between parallaxes and proper motions), one can adopte a Gaussian Mixture model for their distributions (joint or split) and analytically marginalize over the true velocity, and colors of each star. As a result, one only need to sample the parameters of the mixture model, as well as the distance and extinction of each star. The effective likelihood function is a simple multivariate Gaussian, derived below. This opens the possibility to implement this model in fast inference/modelling languages like Tensorflow, Stan, or Edward.

A summary of the notation is provided in the table below. Apologies if there are typos in the text or equations!



$\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1, \cdots, \boldsymbol{\alpha}_B)$	All parameters of the mixture model
$\boldsymbol{\alpha}_b = (f_b, \boldsymbol{\xi}_b, \Sigma_b)$	Parameters of the b Gaussian of the mixture
i	Index of the ith star
$\boldsymbol{n}_i = (\alpha_i, \delta_i)$	True/observed angular position
r_i	True distance
$\boldsymbol{v}_i = (v_{x,i}, v_{y,i}, v_{z,i})$	True 3D cartesian velocity
$\hat{\boldsymbol{\mu}}_i = (\mu_{\alpha,i}, \mu_{\delta,i})$	Observed proper motion
$\hat{\varpi}_i$	Observed parallax
$\boldsymbol{E}_i \rightarrow E_{m_i}, E_{C_i}$	True magnitude/color extinction at distance r_i
$C_i, \hat{C}_i$	True and observed color
M_i	True absolute magnitude
$\hat{m}_i$	Observed apparent magnitude

1 Model

Our population/distribution model in 8-dimensional space (3D positions and velocities, plus 2D color–magnitude diagram) is a Gaussian mixture,

$$[\boldsymbol{v}\ \boldsymbol{n}\ r\ C\ M]^T \,|\, \boldsymbol{\alpha} \sim \sum_{b=1}^{B} f_b\, \mathcal{N}^{8\mathrm{D}}\big(\boldsymbol{\xi}_b; \Sigma_b\big). \tag{1}$$

The priors will be specified later. Typically, one would adopt conjugate priors which greatly simply the inference, *i.e.* a Dirichlet prior on the amplitudes $\{f_b\}$, and multivariate Gaussian for each mean $\boldsymbol{\xi}_b$, and Wishard for the covariance Σ_b.

The full 5-dimensional likelihood, accounting for any covariance between the measurements, is

$$[\hat{\mu}_{\alpha,i}\ \hat{\mu}_{\delta,i}\ \hat{\varpi}_i\ \hat{C}_i\ \hat{m}_i]^T \ |\ \boldsymbol{v}_i, \boldsymbol{n}_i, r_i, E_i, C_i, M_i \ \sim\ \mathcal{N}^{5\mathrm{D}}\Big(\boldsymbol{\psi}_i; \Psi_i\Big) \tag{2}$$

with the model vector

$$\boldsymbol{\psi}_i = \begin{bmatrix} \mu_{\alpha,i} = \Pi_\alpha(\boldsymbol{n}_i)\boldsymbol{v}_i/r_i \\ \mu_{\delta,i} = \Pi_\delta(\boldsymbol{n}_i)\boldsymbol{v}_i/r_i \\ 1/r_i \\ C_i + E_{C_i} \\ M_i + 5\log r_i + E_{m_i} \end{bmatrix} \tag{3}$$

where Π_α and Π_δ are projection matrices, projecting the 3D cartesian vector $\boldsymbol{v}$ in spherical coordinates. Ψ_i is the covariance of the measurements, which could be block diagonal.

The full posterior distribution, marginalizing over $(\boldsymbol{v}_i, C_i, M_i)_{i=1,\cdots,N}$ reads

$$
\begin{aligned}
& p\Big(\boldsymbol{\alpha}, \{r_i, \boldsymbol{E}_i\}\Big|\{\hat\mu_{\alpha,i}, \hat\mu_{\delta,i}\ \hat\varpi_i\ \hat C_i\ \hat m_i\}\Big) \\
&= \int \mathrm{d}\{\boldsymbol{v}_i, C_i, M_i\}\ p\Big(\boldsymbol{\alpha}, \{\boldsymbol{v}_i, C_i, M_i, r_i, \boldsymbol{E}_i\}\Big|\{\hat\mu_{\alpha,i}, \hat\mu_{\delta,i}\ \hat\varpi_i\ \hat C_i\ \hat m_i\}\Big) \\
&= \ p(\boldsymbol{\alpha}) \prod_{i=1}^{N} \int \mathrm{d}\boldsymbol{v}_i \mathrm{d}C_i \mathrm{d}M_i\ p\Big(\boldsymbol{v}_i, C_i, M_i, r_i\Big|\boldsymbol{\alpha}\Big) p\Big(\hat{\boldsymbol{\mu}}_i\ \hat\varpi_i\ \hat C_i\ \hat m_i\Big|\boldsymbol{v}_i, C_i, M_i, r_i, \boldsymbol{E}_i\Big) p(\boldsymbol{E}_i) \\
&= \ p(\boldsymbol{\alpha}) \prod_{i=1}^{N}\ p(\boldsymbol{E}_i)\ \sum_{b=1}^{B}\ f_b\ \mathcal{L}_{ib}
\end{aligned} \tag{4}
$$

$$
\mathcal{L}_{ib} \ = \ \int \mathrm{d}\boldsymbol{v}_i \mathrm{d}C_i \mathrm{d}M_i\ \mathcal{N}^{5\mathrm{D}}\Big(\begin{bmatrix} \Pi_\alpha(\boldsymbol{n}_i)\boldsymbol{v}_i/r_i \\ \Pi_\delta(\boldsymbol{n}_i)\boldsymbol{v}_i/r_i \\ 1/r_i \\ C_i + E_{C_i} \\ M_i + 5\log r_i + E_{m_i} \end{bmatrix} - \begin{bmatrix} \hat\mu_{\alpha,i} \\ \hat\mu_{\delta,i} \\ \hat\varpi_i \\ \hat C_i \\ \hat m_i \end{bmatrix}; \Psi_i\Big)\ \mathcal{N}^{8\mathrm{D}}\Big(\begin{bmatrix} \boldsymbol{v}_i \\ \boldsymbol{n}_i \\ r_i \\ C_i \\ M_i \end{bmatrix} - \boldsymbol{\xi}_b; \Sigma_b\Big)
$$

Let's put this integral in a simpler, more symbolic form, using $a = [v_{x,i}\ v_{y,i}\ v_{z,i}\ C_i\ M_i]$. The rest of the identification should be fairly obvious, but it fully detailed below. We exploit a standard Schur complement trick to isolate the contributions of a.

$$
\mathcal{L} \ = \ \int \mathrm{d}a\ \ \mathcal{N}\Big(\begin{bmatrix} Ra - \hat a \\ b \end{bmatrix}; \begin{bmatrix} \Sigma_A & \Sigma_C^T \\ \Sigma_C & \Sigma_B \end{bmatrix}\Big)\ \ \mathcal{N}\Big(\begin{bmatrix} a - \beta \\ c \end{bmatrix}; \begin{bmatrix} \Phi_A & \Phi_C^T \\ \Phi_C & \Phi_B \end{bmatrix}\Big) \tag{5}
$$

$$
\begin{aligned}
\propto & \int \mathrm{d}a\ \exp\Big(-\frac{1}{2}\big(Ra - \hat a - \Sigma_C^T\Sigma_B^{-1}b\big)^T\big(\Sigma_A - \Sigma_C^T\Sigma_B^{-1}\Sigma_C\big)^{-1}\big(Ra - \hat a - \Sigma_C^T\Sigma_B^{-1}b\big)\Big) \\
& \quad \times \exp\Big(-\frac{1}{2}\big(a - \beta - \Phi_C\Phi_B^{-1}c\big)^T\big(\Phi_A - \Phi_C^T\Phi_B^{-1}\Phi_C\big)^{-1}\big(a - \beta - \Phi_C\Phi_B^{-1}c\big)\Big) \\
& \quad \times \exp\Big(-\frac{1}{2}b^T\Sigma_B^{-1}b\Big)\ \exp\Big(-\frac{1}{2}c^T\Phi_B^{-1}c\Big)
\end{aligned} \tag{6}
$$

$$
\begin{aligned}
\propto & \quad \exp\Big(-\frac{1}{2}b^T\Sigma_B^{-1}b\Big)\ \exp\Big(-\frac{1}{2}c^T\Phi_B^{-1}c\Big) \\
& \times \int \mathrm{d}a \exp\Big(-\frac{1}{2}\big(a - \beta - \Phi_C\Phi_B^{-1}c\big)^T\big(\Phi_A - \Phi_C^T\Phi_B^{-1}\Phi_C\big)^{-1}\big(a - \beta - \Phi_C\Phi_B^{-1}c\big)\Big) \\
& \times\ \exp\Big(-\frac{1}{2}\big(a - R^{-1}\hat a - R^{-1}\Sigma_C^T\Sigma_B^{-1}b\big)^T R^T\big(\Sigma_A - \Sigma_C^T\Sigma_B^{-1}\Sigma_C\big)^{-1}R\big(a - R^{-1}\hat a - R^{-1}\Sigma_C^T\Sigma_B^{-1}b\big)\Big)
\end{aligned} \tag{7}
$$

$$
\begin{aligned}
\propto & \quad \exp\Big(-\frac{1}{2}b^T\Sigma_B^{-1}b\Big)\ \exp\Big(-\frac{1}{2}c^T\Phi_B^{-1}c\Big) \\
& \times\ \exp\Big(-\frac{1}{2}\big(R^{-1}\hat a + R^{-1}\Sigma_C^T\Sigma_B^{-1}b - \beta - \Phi_C\Phi_B^{-1}c\big)^T \\
& \qquad \cdot \big(\Phi_A - \Phi_C^T\Phi_B^{-1}\Phi_C + R^{-1}(\Sigma_A - \Sigma_C^T\Sigma_B^{-1}\Sigma_C)R^{T-1}\big)^{-1} \\
& \qquad \cdot \big(R^{-1}\hat a + R^{-1}\Sigma_C^T\Sigma_B^{-1}b - \beta - \Phi_C\Phi_B^{-1}c\big)\Big)
\end{aligned} \tag{8}
$$

We see that the marginalization of a results in a Gaussian term. Back to our original problem, we identify

$$(5\times 1) \quad a \ = \ [v_{x,i}\ v_{y,i}\ v_{z,i}\ C_i\ M_i]^T \tag{9}$$

$$(4\times 1) \quad b \ = \ [\hat\mu_{\alpha,i}\ \ \hat\mu_{\delta,i}\ \ \hat C_i - E_{C_i}\ \ \hat m_i - 5\log r_i - E_{m_i}]^T \tag{10}$$

$$(6\times 1) \quad \beta \ = \ [\xi_{v_x,b}\ \xi_{v_y,b}\ \xi_{v_z,b}\ \xi_{C,b}\ \xi_{M,b}]^T \tag{11}$$

$$(3\times 1) \quad c \ = \ [\alpha - \xi_{\alpha,b}\ \ \ \delta - \xi_{\delta,b}\ \ \ r_i - \xi_{r,b}]^T \tag{12}$$

$$(4\times 5) \quad R \ = \ \begin{bmatrix} \frac{\cos\delta\cos\alpha}{r_i} & \frac{\cos\delta\sin\alpha}{r_i} & \frac{-\sin\alpha}{r_i} & 0 & 0 \\ \frac{-\sin\alpha}{r_i} & \frac{\cos\alpha}{r_i} & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \tag{13}$$

The various matrix blocks are just the original data and model covariances, but partitioned according to those new vectors.

1.1 Velocity model only, no color–magnitude diagram

The posterior distribution reads

$$
\begin{aligned}
p\Big(\boldsymbol{\alpha}, \{r_i\}\Big|\{\hat{\mu}_{\alpha,i}, \hat{\mu}_{\delta,i}\ \hat{\varpi}_i\}\Big) \ &= \int \mathrm{d}\{\boldsymbol{v}_i\}\ p\Big(\boldsymbol{\alpha}, \{\boldsymbol{v}_i, r_i\}\Big|\{\hat{\mu}_{\alpha,i}, \hat{\mu}_{\delta,i}\ \hat{\varpi}_i\}\Big) \\
&= \ p(\boldsymbol{\alpha}) \prod_{i=1}^{N} \int \mathrm{d}\boldsymbol{v}_i\ p(\boldsymbol{v}_i|\boldsymbol{\alpha})\ p(r_i)\ p(\hat{\boldsymbol{\mu}}_i\ \hat{\varpi}_i\ |\boldsymbol{v}_i, r_i) \ = \ p(\boldsymbol{\alpha}) \prod_{i=1}^{N} p(r_i) \sum_{b=1}^{B} f_b \mathcal{L}_{ib}
\end{aligned}
$$

$$
\begin{aligned}
\mathcal{L}_{ib} \quad &= \quad \int \mathrm{d}\boldsymbol{v}_i\ \mathcal{N}^{3\mathrm{D}}\Big(\begin{bmatrix}\Pi_\alpha(\boldsymbol{n}_i)\boldsymbol{v}_i/r_i \\ \Pi_\delta(\boldsymbol{n}_i)\boldsymbol{v}_i/r_i \\ 1/r_i\end{bmatrix} - \begin{bmatrix}\hat{\mu}_{\alpha,i} \\ \hat{\mu}_{\delta,i} \\ \hat{\varpi}_i\end{bmatrix}; \Psi_i\Big)\ \mathcal{N}^{3\mathrm{D}}\Big(\boldsymbol{v}_i - \boldsymbol{\xi}_b; \Sigma_b\Big) & (14) \\
&= \quad \int \mathrm{d}\boldsymbol{v}_i\ \mathcal{N}^{3\mathrm{D}}\Big(\begin{bmatrix}\Pi\boldsymbol{v}_i \\ 1/r_i\end{bmatrix} - \begin{bmatrix}\hat{\boldsymbol{\mu}}_i \\ \hat{\varpi}_i\end{bmatrix}; \begin{bmatrix}\Psi_{\boldsymbol{\mu}\boldsymbol{\mu},i} & \Psi_{\boldsymbol{\mu}\varpi,i} \\ \Psi_{\boldsymbol{\mu}\varpi,i} & \Psi_{\varpi\varpi,i}\end{bmatrix}\Big)\ \mathcal{N}^{3\mathrm{D}}\Big(\boldsymbol{v}_i - \boldsymbol{\xi}_b; \Sigma_b\Big) \\
&\propto \quad \int \mathrm{d}\boldsymbol{v}_i \exp\Big(-\frac{1}{2}\Psi^{-1}_{\varpi\varpi,i}(\hat{\varpi}_i - \frac{1}{r_i})^2\Big)\ \mathcal{N}^{2\mathrm{D}}\Big(\Pi\boldsymbol{v}_i - \hat{\boldsymbol{\mu}}_i; \Psi_{\boldsymbol{\mu}\boldsymbol{\mu},i} - \Psi^T_{\boldsymbol{\mu}\varpi,i}\Psi^{-1}_{\varpi\varpi,i}\Psi_{\boldsymbol{\mu}\varpi,i}\Big)\ \mathcal{N}^{3\mathrm{D}}\Big(\boldsymbol{v}_i - \boldsymbol{\xi}_b; \Sigma_b\Big) & (15) \\
&\propto \quad \exp\Big(-\frac{1}{2}\Psi^{-1}_{\varpi\varpi,i}(\hat{\varpi}_i - \frac{1}{r_i})^2\Big) \int \mathrm{d}\boldsymbol{v}_i \mathcal{N}^{3\mathrm{D}}\Big(\boldsymbol{v}_i - \Pi^{-1}\hat{\boldsymbol{\mu}}_i; \Pi^{-1}(\Psi_{\boldsymbol{\mu}\boldsymbol{\mu},i} - \Psi^T_{\boldsymbol{\mu}\varpi,i}\Psi^{-1}_{\varpi\varpi,i}\Psi_{\boldsymbol{\mu}\varpi,i})\Pi^{-1}\Big)\ \mathcal{N}^{3\mathrm{D}}\Big(\boldsymbol{v}_i - \boldsymbol{\xi}_b; \Sigma_b\Big) \\
&\propto \quad \exp\Big(-\frac{1}{2}\Psi^{-1}_{\varpi\varpi,i}(\hat{\varpi}_i - \frac{1}{r_i})^2\Big)\mathcal{N}^{2\mathrm{D}}\Big(\Pi\boldsymbol{\xi}_b - \hat{\boldsymbol{\mu}}_i; \Psi_{\boldsymbol{\mu}\boldsymbol{\mu},i} - \Psi^T_{\boldsymbol{\mu}\varpi,i}\Psi^{-1}_{\varpi\varpi,i}\Psi_{\boldsymbol{\mu}\varpi,i} + \Sigma_b\Big) & (16) \\
&= \quad \mathcal{N}^{3\mathrm{D}}\Big(\begin{bmatrix}\Pi\boldsymbol{\xi}_b \\ 1/r_i\end{bmatrix} - \begin{bmatrix}\hat{\boldsymbol{\mu}}_i \\ \hat{\varpi}_i\end{bmatrix}; \begin{bmatrix}\Psi_{\boldsymbol{\mu}\boldsymbol{\mu},i} + \Sigma_b & \Psi_{\boldsymbol{\mu}\varpi,i} \\ \Psi_{\boldsymbol{\mu}\varpi,i} & \Psi_{\varpi\varpi,i}\end{bmatrix}\Big) & (17)
\end{aligned}
$$

In other words, we just need to draw the observations as

$$
\begin{bmatrix}\hat{\boldsymbol{\mu}}_i \\ \hat{\varpi}_i\end{bmatrix}\Big|\ r_i, \Psi_i, \boldsymbol{\xi}_b, \Sigma_b \ \sim \ \mathcal{N}^{3\mathrm{D}}\Big(\begin{bmatrix}\Pi\boldsymbol{\xi}_b \\ 1/r_i\end{bmatrix}; \begin{bmatrix}\Psi_{\boldsymbol{\mu}\boldsymbol{\mu},i} + \Pi^T\Sigma_b\Pi & \Psi_{\boldsymbol{\mu}\varpi,i} \\ \Psi_{\boldsymbol{\mu}\varpi,i} & \Psi_{\varpi\varpi,i}\end{bmatrix}\Big) \tag{18}
$$