ESTIMATION OF MULTIVARIATE PDF USING THE MAXIMUM ENTROPY PRINCIPLE

Zois Boukouvalas[1] , *Charles Cavalcante*[2]

[1]University of Maryland, College Park
[2]Federal University of Ceara, Fortaleza-CE

1. MAXIMUM ENTROPY DISTRIBUTIONS

The maximum entropy density, subject to known moment constraints, problem can be written as the following optimization problem [**?**]:

$$\begin{aligned}\max_{p(\mathbf{x})} H(p(\mathbf{x})) &= -\int_{\Omega(\mathbf{x})} p(\mathbf{x}) \log p(\mathbf{x}) d\mathbf{x} \\ s.t. &\int_{\Omega(\mathbf{x})} p(\mathbf{x}) d\mathbf{x} = 1, \\ &\int_{\Omega(\mathbf{x})} r_i(\mathbf{x}) p(\mathbf{x}) d\mathbf{x} = \alpha_i, \text{for } i = 1, \ldots, M, \end{aligned} \tag{1}$$

where $\mathbf{x} \in \mathbb{R}^K$, $p(\mathbf{x}) \geq 0$, with equality outside the support set $\Omega(\mathbf{x})$, $r_i(\mathbf{x}) \in C(\mathbb{R}^K, \mathbb{R}), \forall i$, $C(\mathbb{R}^K, \mathbb{R})$ represents the space of continue functions which have domain $\mathbb{R}^K$ and codomain $\mathbb{R}$. The first constraint need to be $\int_{\Omega(\mathbf{x})} p(\mathbf{x}) d\mathbf{x} = 1$, equivalently $r_0 = 1$ and $\alpha_0 = 1$, in order for $p(\mathbf{x})$ to be a valid pdf. Thus, p is a density on support set $\Omega(\mathbf{x})$ meeting certain moment constraints $\alpha_1, \ldots, \alpha_M$. The optimization problem in (1) can be rewritten as a Lagrangian function:

$$L(p) = -\int p \log p + \lambda_0 \left(\int p - 1 \right) + \sum_{i=1}^{M} \lambda_i \left(\int r_i p - \alpha_i \right), \tag{2}$$

where λ_i, $i = 0, \ldots, M$, are the Lagrangian multipliers. By setting $\partial L(p)/\partial p(\mathbf{x}) = 0$, we obtain the form of the maximizing density

$$p(\mathbf{x}) = e^{-1+\lambda_0+\sum_{i=1}^{M} \lambda_i r_i(\mathbf{x})}, \tag{3}$$

where Lagrangian multipliers are chosen so that p satisfies the constraints. We can solve for them by the following Newton iteration:

$$\begin{aligned}\boldsymbol{\lambda}_{t+1} &= \boldsymbol{\lambda}_t - \mathbf{J}^{-1} E_{p_t}\{\mathbf{r} - \boldsymbol{\alpha}\} \\ &= \boldsymbol{\lambda}_t - \left(E_{p_t}\left\{\mathbf{r}\mathbf{r}^T\right\}\right)^{-1} E_{p_t}\{\mathbf{r} - \boldsymbol{\alpha}\}, \end{aligned} \tag{4}$$

where $\mathbf{r} = [r_0, \ldots, r_M]^T \in C^{M+1}(\mathbb{R}^K, \mathbb{R})$, $\boldsymbol{\lambda} = [\lambda_0, \ldots, \lambda_M]^T \in \mathbb{R}^{M+1}$, $\boldsymbol{\alpha} = [\alpha_0, \ldots, \alpha_M]^T \in \mathbb{R}^{M+1}$, and $\mathbf{J} \in \mathbb{R}^{K \times K}$ is the Jacobian matrix. The ijth entry of $\mathbf{J}$ is

$$\begin{aligned}\mathbf{J}_{ij} &= \frac{\partial \int_{\Omega(\mathbf{x})} r_i(\mathbf{x}) p(\mathbf{x}) d\mathbf{x} - \alpha_i}{\partial \lambda_j} \\ &= \int_{\Omega(\mathbf{x})} r_i(\mathbf{x}) \frac{\partial p(\mathbf{x})}{\partial \lambda_j} d\mathbf{x} \\ &= \int_{\Omega(\mathbf{x})} r_i(\mathbf{x}) r_j(\mathbf{x}) p(\mathbf{x}) d\mathbf{x} \\ &= E_{p_t}\left\{r_i r_j\right\} \end{aligned} \tag{5}$$

The ith entry of $E_{p_t}\{\mathbf{r} - \boldsymbol{\alpha}\}$ is

$$E_{p_t}\{\mathbf{r} - \boldsymbol{\alpha}\}_i = E_{p_t}\{r_i - \alpha_i\} = \int_{\Omega(\mathbf{x})} r_i(\mathbf{x}) p(\mathbf{x}) d\mathbf{x} - \alpha_i \tag{6}$$

The entropy corresponding to this density function will be

$$\begin{aligned} H(p) &= -\int p \log p \\ &= -\int \left(-1 + \lambda_0 + \sum_{i=1}^{M} \lambda_i r_i\right) p \\ &= 1 - \lambda_0 - \sum_{i=1}^{M} \lambda_i \alpha_i. \end{aligned} \tag{7}$$

2. SELECTION OF CONSTRAINTS

2.1. Constrained functions

In order to evaluate the joint entropy bound for random vector $\mathbf{s} \in \mathbb{R}^K$, assuming all the marginals have the similar distribution, we can use maximum entropy distributions method. By selecting a set of constraints, we can obtain the maximum entropy distribution which satisfy these constraints. Before evaluating the maximum entropy distribution, it will make the problem easier by whitening the data.

$$\begin{aligned} H(\mathbf{s}) &= H(\mathbf{x}) - \log|\det(\mathbf{\Sigma}^{-1/2})| \\ &= H(\mathbf{x}) + \frac{1}{2}\log(\det(\mathbf{\Sigma})), \end{aligned} \tag{8}$$

where $\mathbf{\Sigma} = E\{\mathbf{s}\mathbf{s}^T\}$ is the covariance matrix of $\mathbf{s}$, and $\mathbf{x} = \mathbf{\Sigma}^{-1/2}\mathbf{s}$.

We select the following constraints

1. $\int_{\Omega(\mathbf{x})} p(\mathbf{x}) d\mathbf{x} = 1$
2. $\int_{\Omega(\mathbf{x})} \mathbf{1}^T \mathbf{x} p(\mathbf{x}) d\mathbf{x} = 0$
3. $\int_{\Omega(\mathbf{x})} \mathbf{x}^T \mathbf{x} p(\mathbf{x}) d\mathbf{x} = K$
4. $\int_{\Omega(\mathbf{x})} \mathbf{1}^T \mathbf{g}(\mathbf{x}) p(\mathbf{x}) d\mathbf{x} = \alpha_g$
5. $\int_{\Omega(\mathbf{x})} q(\mathbf{x}) p(\mathbf{x}) d\mathbf{x} = \alpha_q$

where α_g and α_q can be evaluated by using sample average, $\mathbf{g} = [g_1, \dots, g_K]^T \in C^K(\mathbb{R}, \mathbb{R})$, and $q \in C(\mathbb{R}^K, \mathbb{R})$. For any subset of random vector $\mathbf{x}$, $\tilde{\mathbf{x}}$ and $\check{\mathbf{x}}$, which have same dimension, will have same distribution since

$$p(\tilde{\mathbf{x}}) = \int_{\Omega(\tilde{\mathbf{x}}^c)} p(\mathbf{x}) d\tilde{\mathbf{x}}^c = \int_{\Omega(\check{\mathbf{x}}^c)} p(\mathbf{x}) d\check{\mathbf{x}}^c = p(\check{\mathbf{x}})$$

where $\tilde{\mathbf{x}}^c$ and $\check{\mathbf{x}}^c$ are the complement set for $\tilde{\mathbf{x}}$ and $\check{\mathbf{x}}$, respectively. Few comments on this set of constraints:

- The first four constraints enforce the validity of the pdf, data are zero mean, and that the covariance is the identity matrix.
- We end up with a multivariate Gaussian if we only use first four constraints.
- The fifth constraint is added in order to capture the non-Gaussianity.
- We end up with a multivariate distribution with independent marginals if we only use first five constraints since we can always write the joint pdf as a product of all marginal pdfs.

$$p(\mathbf{x}) = e^{-1+\lambda_0+\lambda_1\sum_{k=1}^K x_k + \lambda_2\sum_{k=1}^K x_k^2 + \lambda_3\sum_{k=1}^K g(x_k) + +\lambda_4 q(\mathbf{x})}$$

In order to evaluate the multidimensional integral easily, constrained function $q(\mathbf{x})$ can be define as following,

$$q(\mathbf{x}) = \begin{cases} 1, & \text{if } \max_i |x_i| < \gamma,\ \Omega_\gamma \\ 0, & \text{otherwise, } \Omega_\gamma^c \end{cases} \tag{9}$$

Let $p = e^{\lambda_5 q}\tilde{p}$. Then, for Newton iteration, the integral could be calculated as following

$$\begin{aligned} E\{r_i r_j\} &= \int_\Omega r_i r_j p \\ &= \int_{\Omega_\gamma} r_i r_j p + \int_{\Omega_\gamma^c} r_i r_j p \\ &= \int_{\Omega_\gamma} r_i r_j e^{\lambda_5 1}\tilde{p} + \int_{\Omega_\gamma^c} r_i r_j e^{\lambda_5 0}\tilde{p} \\ &= e^{\lambda_5}\int_{\Omega_\gamma} r_i r_j \tilde{p} + \left(\int_\Omega r_i r_j \tilde{p} - \int_{\Omega_\gamma} r_i r_j \tilde{p}\right) \\ &= \int_\Omega r_i r_j \tilde{p} - (1 - e^{\lambda_5})\int_{\Omega_\gamma} r_i r_j \tilde{p} \end{aligned} \tag{10}$$

where both integrals can be separate into product of a lot of one dimensional integrals.

3. NOTES

1. We can generate data using a different implementation. Mixture of multivariate GGDs.
2. We need a way to quantify the effectiveness of our approximation
3. So far the position of the local constraint is placed on a manually selected position. We need to make it automated.
4. If we decide to use Gaussian kernels for local constraints we can use multivariate integration using Bayesian techniques.