MAXIMUM ENTROPY ON A SIMPLEX: AN EXPOSITORY NOTE

MARK FISHER

Preliminary and incomplete

ABSTRACT. The Gibbs distribution $f(x) = e^{-\lambda^\top x}/Z(\lambda)$ for $x = \{x_1, \ldots, x_n\}$ defined over the region where $x_i \geq 0$ and $\sum_{i=1}^n x_i \leq 1$ characterizes the maximum entropy distribution on a simplex subject to $E[x] = \mu$. An explicit representation for $Z(\lambda)$ is derived.

1. PRELIMINARIES

Let $x = \{x_i\}_{i=1}^n$ where $x_i \geq 0$ and $\sum_{i=1}^n x_i \leq b$ for some $b \geq 0$ and define $x_{n+1} := b - \sum_{i=1}^n x_i$.[1] Then $\widetilde{x} := \{x_i\}_{i=1}^{n+1}$ lies on an n-dimensional (generalized) simplex denoted Δ_b^n. Let $\Delta^n \equiv \Delta_1^n$ denote the n-dimensional (ordinary) simplex. We can express any function $\widetilde{g}(\widetilde{x}) = \widetilde{g}(x_1, \ldots, x_n, x_{n+1})$ subject to $x_{n+1} = b - \sum_{i=1}^n x_i$ as $g(x) = g(x_1, \ldots, x_n) := g(x_1, \ldots, x_n, b - \sum_{i=1}^n x_i)$. Moreover, $\int_{\Delta_b^n} \widetilde{g}(\widetilde{x})\, d\widetilde{x} = \int_{\Delta_b^n} g(x)\, dx$, which can be computed as follows. Let $w = \{w_1, \ldots, w_n\}$ be a permutation of $\{1, \ldots, n\}$, so that w is a list of indices in some fixed order. Then[2]

$$\int_{\Delta_b^n} g(x)\, dx = \int_0^b dx_{w_1} \int_0^{b - x_{w_1}} dx_{w_2} \cdots \int_0^{b - \sum_{i=1}^{n-1} x_{w_i}} dx_{w_n}\, g(x). \tag{1.1}$$

The order of integration in (1.1) is from right to left; i.e., from x_{w_n} first to x_{w_1} last.

Let $f(x)$ denote the joint probability density for x so that $\int_{\Delta^n} f(x)\, dx = 1$. Let μ denote the mean of x,

$$\mu = \langle x \rangle := E[x] := \int_{\Delta^n} x\, f(x)\, dx, \tag{1.2}$$

and let Σ denote the covariance matrix of x,

$$\Sigma = E[(x - \mu)\,(x - \mu)^\top] = E[x\, x^\top] - \mu\, \mu^\top = \int_{\Delta^n} \left(x\, x^\top\right) f(x)\, dx - \mu\, \mu^\top, \tag{1.3}$$

where $x^\top$ denotes the transpose of x, so that $\Sigma_{ij} = E[x_i\, x_j] - \mu_i\, \mu_j = \langle x_i\, x_j \rangle - \langle x_i \rangle\, \langle x_j \rangle$.[3]

Date: March 19, 2006.

The views expressed herein are the author's and do not necessarily reflect those of the Federal Reserve Bank of Atlanta or the Federal Reserve System.

[1] If $n = 0$, then $x_{n+1} = b$.

[2] We are using the notation $\int dx_1 \int dx_2\, g(x_1, x_2) \equiv \iint g(x_1, x_2)\, dx_2\, dx_1$ on the right-hand side of (1.1).

[3] The mean of x_{n+1} is given by $\mu_{n+1} = 1 - \sum_{i=1}^n \mu_i$. The covariance between x_{n+1} and x_i equals $-\sum_{j=1}^n \Sigma_{ij}$ and the variance of x_{n+1} equals $\sum_{i=1}^n \sum_{j=1}^n \Sigma_{ij}$.

The classic distribution for x on Δ^n is the Dirichlet distribution, for which

$$f(x) = \frac{\Gamma\left(\sum_{i=1}^{n+1} \alpha_i\right)}{\prod_{i=1}^{n+1} \Gamma(\alpha_i)} x_1^{\alpha_1 - 1} \cdots x_n^{\alpha_n - 1} (1 - \textstyle\sum_{i=1}^n x_i)^{\alpha_{n+1}-1}, \tag{1.4}$$

where $\alpha_i > 0$. In this note we consider an alternative distribution.

2. Maximum entropy distributions

Here we outline the derivation of the maximum entropy distribution for x over a generic region $\mathcal{R}$.[4] In Section 3 we will specialize to $\mathcal{R} = \Delta^n$.

The object is to find the continuous function f that maximizes the entropy

$$H = -\int_{\mathcal{R}} \log(f(x))\, f(x)\, dx \tag{2.1}$$

subject to $\int_{\mathcal{R}} q(x)\, f(x)\, dx = \theta$ and $\int_{\mathcal{R}} f(x)\, dx = 1$ where $q(x)$ is a vector function of x and θ is given. (We will be especially interested in $q(x) = x$.) To this end, form the Lagrangian

$$\mathcal{L} = -\int_{\mathcal{R}} \log(f(x))\, f(x)\, dx - \lambda^\top \left(\int_{\mathcal{R}} q(x)\, f(x)\, dx - \theta \right) - \varphi \left(\int_{\mathcal{R}} f(x)\, dx - 1 \right), \tag{2.2}$$

where $\lambda = \{\lambda_i\}_{i=1}^k$ is a vector of Lagrange multipliers, φ is a scalar Lagrange multiplier, and θ is a k-dimensional vector. To apply the calculus of variations, express (2.2) as

$$\mathcal{L} = \int_{\mathcal{R}} g(x, f(x))\, dx + \left(\lambda^\top \theta + \varphi\right), \tag{2.3}$$

where

$$g(x, y) = -\log(y)\, y - \left(\lambda^\top q(x)\right) y - \varphi\, y. \tag{2.4}$$

In this case, the first-order (Euler–Lagrange) condition is $\partial g(x, y)/\partial y = 0$, or[5]

$$-\log(f(x)) - 1 - \lambda^\top q(x) - \varphi = 0. \tag{2.5}$$

Exponentiating both sides of (2.5) and rearranging produces[6]

$$f(x) = e^{-(1+\varphi) - \lambda^\top q(x)}. \tag{2.6}$$

Define

$$Z(\lambda) := \int_{\mathcal{R}} e^{-\lambda^\top q(x)}\, dx. \tag{2.7}$$

Since $\int_{\mathcal{R}} f(x)\, dx = 1$, we have $e^{1+\varphi} = Z(\lambda)$, and we obtain the Gibbs distribution

$$f(x) = \frac{e^{-\lambda^\top q(x)}}{Z(\lambda)}, \tag{2.8}$$

[4]See Jaynes (2003) for a discussion of maximum entropy.

[5]One would obtain this condition if one differentiated $\mathcal{L}$ with respect to the 'probabilities' $f(x)$, treating $\int_{\mathcal{R}} dx$ as a summation operator.

[6]The second-order (Legendre) condition for a maximum is $\partial^2 g(x, y)/\partial y^2 < 0$, which is satisfied since $-f(x)^{-1} < 0$.

where $Z(\lambda)$ is known as the *partition function*.[7]

Define

$$m(\lambda) := -\nabla_\lambda \log\big(Z(\lambda)\big) \qquad \text{and} \qquad S(\lambda) := \nabla^2_\lambda \log\big(Z(\lambda)\big). \tag{2.10}$$

We now show that $m(\lambda) = \theta$ and $S(\lambda) = E[q(x)\,q(x)^\top] - \theta\,\theta^\top$:

$$\begin{aligned} m(\lambda) = \frac{-\nabla_\lambda Z(\lambda)}{Z(\lambda)} &= \frac{-\nabla_\lambda \int_{\mathcal{R}} e^{-\lambda^\top q(x)}\,dx}{Z(\lambda)} = \frac{\int_{\mathcal{R}} \Big(-\nabla_\lambda e^{-\lambda^\top q(x)}\Big)\,dx}{Z(\lambda)} \\ &= \frac{\int_{\mathcal{R}} q(x)\,e^{-\lambda^\top q(x)}\,dx}{Z(\lambda)} = \int_{\mathcal{R}} q(x)\,f(x)\,dx = \theta \end{aligned} \tag{2.11}$$

and

$$\begin{aligned} S(\lambda) = -\nabla_\lambda m(\lambda) &= -\nabla_\lambda \int_{\mathcal{R}} \frac{q(x)\,e^{-\lambda^\top q(x)}}{Z(\lambda)}\,dx = \int_{\mathcal{R}} \left(-\nabla_\lambda \frac{q(x)\,e^{-\lambda^\top q(x)}}{Z(\lambda)}\right)dx \\ = \int_{\mathcal{R}} \Big(q(x)\,q(x)^\top - q(x)\,m(\lambda)^\top\Big)\,\frac{e^{-\lambda^\top q(x)}}{Z(\lambda)}\,dx &= \int_{\mathcal{R}} \Big(q(x)\,q(x)^\top\Big)\,f(x)\,dx - \theta\,\theta^\top. \end{aligned} \tag{2.12}$$

For $q(x) = x$, we have $m(\lambda) = \mu$ and $S(\lambda) = \Sigma$.

Two illustrations. Consider the following two illustrations for which $n = 1$. First, let $g(x) = x_1$ and let $\mathcal{R} = [0, \infty)$. In this case $Z(\lambda) = \lambda_1^{-1}$. We can solve $m(\lambda) = \theta$ for $\theta = \lambda_1^{-1}$. Consequently, $\lambda_1\,e^{-\lambda_1\,x_1} = e^{-\theta^{-1}\,x_1}/\theta$ is the exponential distribution.

Second, let $g(x) = (x_1, x_1^2)^\top$ and let $\mathcal{R} = (-\infty, \infty)$. In this case

$$Z(\lambda) = \frac{e^{\lambda_1^2/(4\,\lambda_2)}\sqrt{\pi}}{\sqrt{\lambda_2}}. \tag{2.13}$$

Letting $\theta_1 = \mu$ and $\theta_2 = \mu^2 + \sigma^2$, we can solve $m(\lambda) = \theta$ for

$$\lambda_1 = -\frac{\mu}{\sigma^2} \qquad \text{and} \qquad \lambda_2 = \frac{1}{2\,\sigma^2} \tag{2.14}$$

and consequently we obtain the Gaussian distribution:

$$\frac{e^{-\lambda_1\,x_1 - \lambda_2\,x_1^2}}{Z(\lambda)} = \frac{e^{-\frac{(x_1-\mu)^2}{2\,\sigma^2}}}{\sqrt{2\,\pi}\,\sigma}. \tag{2.15}$$

We note that λ_2 equals one-half the *precision* $1/\sigma^2$ and λ_1 equals the negative of the mean times the precision.

[7] The density for x is related to the density for $\widetilde{x}$ as follows. Define $\widetilde{\lambda} := \{\widetilde{\lambda}_i\}_{i=1}^{n+1}$. Then, for $x_{n+1} = 1 - \sum_{i=1}^n x_i$, $\widetilde{f}(\widetilde{x}) = e^{-\widetilde{\lambda}^\top \widetilde{x}} / \int_{\Delta^n} e^{-\widetilde{\lambda}^\top \widetilde{x}}\,d\widetilde{x} = e^{-\lambda^\top x}/Z(\lambda) = f(x)$, where $\lambda_i = \widetilde{\lambda}_i - \widetilde{\lambda}_{n+1}$. More generally, let $\widetilde{\lambda}^{(j)} := \widetilde{\lambda} \setminus \{\lambda_j\}$ and $\widetilde{x}^{(j)} := \widetilde{x} \setminus \{x_j\}$. Then

$$\widetilde{f}(\widetilde{x}) = f^{(j)}(\widetilde{x}^{(j)}) = \frac{e^{-(\widetilde{\lambda}^{(j)} - \lambda_j)^\top \widetilde{x}^{(j)}}}{e^{\lambda_j}\,Z(\widetilde{\lambda}^{(j)} - \lambda_j)}. \tag{2.9}$$

Likelihood. Given N independent observations $\{X_i\}_{i=1}^N$, the likelihood for λ is

$$\prod_{i=1}^N f(X_i) = \prod_{i=1}^N \frac{e^{-\lambda^\top g(X_i)}}{Z(\lambda)} = \left(\frac{e^{-\lambda^\top \overline{g}}}{Z(\lambda)}\right)^N, \tag{2.16}$$

where $\overline{g} = \frac{1}{N}\sum_{i=1}^N g(X_i)$. The log-likelihood is

$$\ell(\lambda) = -N\left(\lambda^\top \overline{g} + \log\big(Z(\lambda)\big)\right). \tag{2.17}$$

Thus

$$\nabla_\lambda \ell(\lambda) = -N\left(\overline{g} + \nabla_\lambda \log\big(Z(\lambda)\big)\right) = -N\left(\overline{g} - m(\lambda)\right) \tag{2.18}$$

$$\nabla_\lambda^2 \ell(\lambda) = -N\,\nabla_\lambda^2 \log\big(Z(\lambda)\big) = -N\,S(\lambda). \tag{2.19}$$

The maximum likelihood value for λ can be computed by solving $\nabla_\lambda \ell(\widehat{\lambda}) = 0$ for $\widehat{\lambda} = m^{-1}(\overline{g})$.[8] In addition, the Gaussian approximation to the likelihood is proportional to

$$\exp\left(-\frac{N}{2}\left(\lambda - \widehat{\lambda}\right)^\top S(\widehat{\lambda})\left(\lambda - \widehat{\lambda}\right)\right), \tag{2.20}$$

where $N^{-1}\,S(\widehat{\lambda})^{-1}$ is the covariance matrix for λ. The maximum likelihood value for $\theta = \langle g(x)\rangle$ is $\widehat{\theta} = m(\widehat{\lambda}) = \overline{g}$.

If $g(x) = x$, the maximum likelihood value for $\theta = \mu$ is $\widehat{\mu} = m(\widehat{\lambda}) = \overline{X} = \frac{1}{N}\sum_{i=1}^N X_i$ and the Gaussian-approximation covariance matrix for μ is $N^{-1}\,S(m^{-1}(\overline{X}))$.[9]

Marginal and conditional distributions. Here we suppose $q(x) = x$.

Partition the set of indices $I = \{1,\ldots,n\}$ into α and β, where $\alpha \cup \beta = I$ and $\alpha \cap \beta = \varnothing$. Let $x_\alpha = \{x_i : i \in \alpha\}$, $x_\beta = \{x_i : i \in \beta\}$, etc. The marginal distribution of x_α is

$$f(x_\alpha) = \int\limits_{\mathcal{R}_\beta(x_\alpha)} f(x)\,dx_\beta = \frac{e^{-\lambda_\alpha^\top x_\alpha}}{Z(\lambda)} \int\limits_{\mathcal{R}_\beta(x_\alpha)} e^{-\lambda_\beta^\top x_\beta}\,dx_\beta = \frac{e^{-\lambda_\alpha^\top x_\alpha}\,Z\big(\lambda_\beta, x_\alpha\big)}{Z(\lambda)}, \tag{2.21}$$

where $\mathcal{R}_\beta(x_\alpha)$ denotes the domain of x_β as a function of x_α and

$$Z\big(\lambda_\beta, x_\alpha\big) := \int\limits_{\mathcal{R}_\beta(x_\alpha)} e^{-\lambda_\beta^\top x_\beta}\,dx_\beta. \tag{2.22}$$

Therefore, the distribution of x_β conditional on x_α is

$$f(x_\beta \mid x_\alpha) = \frac{f(x)}{f(x_\alpha)} = \frac{e^{-{\lambda_\beta}^\top x_\beta}}{Z\big(\lambda_\beta, x_\alpha\big)}, \tag{2.23}$$

[8] Given $z = \{z_1,\ldots,z_n\}$ where $z_i > 0$ (for $i = 1,\ldots,n$) and $\sum_{i=1}^n z_i < 1$, $m^{-1}(z)$ exists and is unique. (Need to show this.)

[9] Note that $\widehat{\lambda}$ maximizes the entropy of the distribution given $\mu = \overline{X}$.

which evidently is the maximum entropy distribution for x_β over $\mathcal{R}_\beta(x_\alpha)$ subject to the conditional mean

$$\mu_{\beta|x_\alpha} = \int\limits_{\mathcal{R}_\beta(x_\alpha)} x_\beta\, f(x_\beta \mid x_\alpha)\, dx_\beta = -\nabla_{\lambda_\beta} \log\big(Z\big(\lambda_\beta, x_\alpha\big)\big). \tag{2.24}$$

3. Maximum entropy on a simplex

First we deal with a possible source of confusion. The set $\widetilde{x}$ can be interpreted as a discrete probability measure, the entropy of which is $-\sum_{i=1}^{n+1} x_i \log(x_i)$. This is distinct from the entropy of $\widetilde{x}$, namely $-\int_{\Delta^n} \widetilde{f}(\widetilde{x}) \log\big(\widetilde{f}(\widetilde{x})\big)\, d\widetilde{x} = -\int_{\Delta^n} f(x) \log\big(f(x)\big)\, dx$, that we are interested in here.[10]

Computing the normalization constant. Here we specialize to $\mathcal{R} = \Delta^n$ and $q(x) = x$. Define

$$\zeta_b(\lambda) := \int_{\Delta_b^n} e^{-\lambda^\top x}\, dx = \Big(\prod_{i=1}^n \lambda_i\Big)^{-1} - \sum_{i=1}^n \Big(\lambda_i\, e^{\lambda_i\, b} \prod_{\substack{j=1\\ j\neq i}}^n (\lambda_j - \lambda_i)\Big)^{-1}, \tag{3.1}$$

and let $\zeta(\lambda) := \zeta_1(\lambda)$. Given $\mathcal{R} = \Delta^n$, we have $Z(\lambda) = \zeta(\lambda)$ and thus

$$f(x) = \frac{e^{-\lambda^\top x}}{\zeta(\lambda)}. \tag{3.2}$$

Given this solution, the first-order series expansions for $Z(\lambda)$ and $m_i(\lambda)$ around $\lambda = 0$ are

$$Z(\lambda) = \frac{1}{n!} - \frac{\sum_{i=1}^n \lambda_i}{(n+1)!} + \mathcal{O}(\lambda^2) \tag{3.3}$$

$$m_i(\lambda) = \frac{1}{n+1} + \frac{\sum_{j=1}^n \lambda_j}{(n+1)^2 + (n+1)^3} - \frac{\lambda_i}{(n+1)\,(n+2)} + \mathcal{O}(\lambda^2). \tag{3.4}$$

Define $m_{n+1}(\lambda) := 1 - \sum_{i=1}^n m_i(\lambda)$. Then $\mu_{n+1} = m_{n+1}(\lambda)$. The first-order series expansion for $m_{n+1}(\lambda)$ around $\lambda = 0$ is

$$m_{n+1}(\lambda) = \frac{1}{n+1} + \frac{\sum_{i=1}^n \lambda_i}{(n+1)^2 + (n+1)^3} + \mathcal{O}(\lambda^2). \tag{3.5}$$

In fact, $\lambda_i = 0 \iff m_i(\lambda) = m_{n+1}(\lambda)$.

[10]Nevertheless, the relative entropy of the discreet distribution can be used as a prior for $\widetilde{x}$ (just as the maximum entropy distribution can). Consider

$$\widetilde{h}(\widetilde{x}) = -\sum_{i=1}^{n+1} x_i \log(x_i/m_i),$$

where $\widetilde{m} = \{m_1, \ldots, m_{n+1}\}$ is some base *measure* such that $m_i > 0$ and $\sum_{i=1}^{n+1} m_i = 1$. We can write this as

$$h(x) = \widetilde{h}(\widetilde{x})\Big|_{\substack{x_{n+1} = 1 - \sum_{i=1}^n x_i \\ m_{n+1} = 1 - \sum_{i=1}^n m_i}}$$

Now let $f(x) = e^{\alpha\, h(x)} / \int_{\Delta^n} e^{\alpha\, h(x)}\, dx$ be the distribution for x, where $\alpha \geq 0$ is a scalar parameter that controls how tightly the distribution is concentrated around its mode at $x = m$.

Marginal and conditional distributions. Let $s_\alpha = \sum_{i\in\alpha} x_i$ and let n_β denote the number of elements in β. Given $\mathcal{R} = \Delta^n$, it follows that $\mathcal{R}_\beta(x_\alpha) = \Delta^{n_\beta}_{1-s_\alpha}$ and therefore

$$Z(\lambda_\beta, x_\alpha) = \zeta_{1-s_\alpha}(\lambda_\beta). \tag{3.6}$$

Consequently (2.21) and (2.23) become

$$f(x_\alpha) = \frac{e^{-\lambda_\alpha^\top x_\alpha}\, \zeta_{1-s_\alpha}(\lambda_\beta)}{\zeta(\lambda)} \tag{3.7}$$

and

$$f(x_\beta \mid x_\alpha) = \frac{e^{-\lambda_\beta^\top x_\beta}}{\zeta_{1-s_\alpha}(\lambda_\beta)}. \tag{3.8}$$

In particular, note

$$f(x_\beta \mid x_\alpha = 0) = \frac{e^{-\lambda_\beta^\top x_\beta}}{\zeta(\lambda_\beta)}, \tag{3.9}$$

for which the conditional mean is

$$\mu_{x_\beta|(x_\alpha=0)} = m(\lambda_\beta). \tag{3.10}$$

Drawing from the distribution. We can draw from the joint distribution via the Gibbs sampler, drawing cyclically from the univariate conditional distributions. Let $\beta = \{i\}$. Then

$$f(x_i \mid x_{-i}) = \frac{e^{-\lambda_i\, x_i}}{\zeta_{1-s_{-i}}(\{\lambda_i\})} = \frac{\lambda_i\, e^{-\lambda_i\, x_i}}{1 - e^{-\lambda_i\,(1-s_{-i})}}. \tag{3.11}$$

where $f(x_i \mid x_{-i}) \equiv f(x_\beta \mid x_\alpha)$ and $s_{-i} \equiv s_\alpha$. Define the conditional cdf

$$F(x_i \mid x_{-i}) := \int_0^{x_i} f(t|x_{-i})\, dt = \frac{1 - e^{-\lambda_i\, x_i}}{1 - e^{-\lambda_i\,(1-s_{-i})}} \tag{3.12}$$

for $x_i \le 1 - s_{-i}$. Solving $F(x_i|x_{-i}) = u$ for x_i produces

$$\begin{aligned} x_i &= -\lambda_i^{-1} \log\big(1 + (e^{-\lambda_i\,(1-s_{-i})} - 1)\, u\big) \\ &= (1 - s_{-i})\, u - \frac{1}{2}\,(1 - s_{-i})^2\, u\,(1-u)\,\lambda_i + \mathcal{O}(\lambda_i^2). \end{aligned} \tag{3.13}$$

We can obtain independent draws from $f(x_i|x_{-i})$ via independent draws of $u \sim U(0,1)$. By initializing the Gibbs sampler at μ, the target distribution appears to be reached in about n draws.

Alternative representations for the distribution of $\widetilde{x}$. In addition to $f(x)$ there n ways to represent the distribution of $\widetilde{x} = \{x_1, \ldots, x_{n+1}\}$: $f(x^{(j)})$ for $j = 1, \ldots, n$, where $x^{(j)} = \{x_1^{(j)}, \ldots, x_n^{(j)}\}$ denotes the vector where x_{n+1} replaces x_j in x:

$$x_i^{(j)} = \begin{cases} x_i & i \ne j \\ x_{n+1} & i = j. \end{cases} \tag{3.14}$$

Note that $x_j = 1 - \sum_{i=1}^n x_i^{(j)}$. Changing variables from x to $x^{(j)}$ produces $f(x^{(j)}) = e^{-\lambda^{(j)\top} x^{(j)}}/\zeta(\lambda^{(j)})$, where $\lambda^{(j)} = \{\lambda_1^{(j)}, \dots, \lambda_n^{(j)}\}$ and

$$\lambda_i^{(j)} = \begin{cases} \lambda_i - \lambda_j & i \neq j \\ \quad - \lambda_j & i = j. \end{cases} \tag{3.15}$$

Note that $\mu_j = 1 - \sum_{i=1}^n \mu_i^{(j)}$, where $\mu^{(j)} = m(\lambda^{(j)})$. Moreover, for $i \in \{1, \dots, n\} \setminus \{j\}$, $m(\lambda) = m(\lambda^{(j)})$.

As an example, let $n = 1$. Given $f(x_1) = e^{-\lambda_1 x_1}/\zeta(\lambda_1)$, then $f(x_2) = e^{\lambda_1 x_2}/\zeta(-\lambda_1)$ and $m(-\lambda_1) = 1 - m(\lambda_1)$.

We can use these alternative representations to compute the marginal distribution of $x_{n+1} = x_j^{(j)}$:

$$f(x_j^{(j)}) = \frac{e^{-\lambda_j x_j^{(j)}}\, \zeta_{1-x_j^{(j)}}(\lambda_{-j}^{(j)})}{\zeta(\lambda^{(j)})}. \tag{3.16}$$

To condition on a subset of $\widetilde{x}$ that includes x_{n+1}, first change variables to $x^{(j)}$ for some j such that $x_\alpha^{(j)}$ is the appropriate subset and then apply (3.8):

$$f(x_\beta^{(j)} \mid x_\alpha^{(j)}) = \frac{e^{-\lambda_\beta^{(j)\top} x_\beta^{(j)}}}{\zeta_{1-s_\alpha^{(j)}}(\lambda_\beta^{(j)})}. \tag{3.17}$$

In particular, for $x_\alpha^{(j)} = 0$

$$f(x_\beta^{(j)} \mid x_\alpha^{(j)} = 0) = \frac{e^{-\lambda_\beta^{(j)\top} x_\beta^{(j)}}}{\zeta(\lambda_\beta^{(j)})}. \tag{3.18}$$

4. Other regions

We can apply the foregoing to other regions. Consider the region of stationarity for an autoregressive process: $A(L)\, z_t = \varepsilon_t$, where $\varepsilon_t \sim$ iid $N(0, \sigma^2)$ and $A(L) = 1 - x_1 L - x_2 L^2 - \dots - x_n L^n$ is a polynomial in the lag operator. The region of stationarity is characterized by those $x \in \mathbb{R}^n$ such that all of the roots of $A(L) = 0$ lying outside the unit circle. For $n = 1$ we have $-1 < x_1 < 1$ and for $n = 2$ we have $x_1 + x_2 < 1$ and $-1 < x_2 < 1 + x_1$. In this latter case,

$$Z(\lambda) = \frac{e^{2\lambda_1 + \lambda_2}\,(\lambda_1 - \lambda_2) + e^{\lambda_2 - 2\lambda_1}\,(\lambda_1 + \lambda_2) - 2\, e^{-\lambda_2} \lambda_1}{\lambda_1 \left(\lambda_1^2 - \lambda_2^2\right)}. \tag{4.1}$$

With $\lambda_1 = 0$ and $\lambda_2 = -1.344$ we obtain $\mu_1 = \mu_2 = 0$.

References

Jaynes, E. T. (2003). *Probability Theory: The Logic of Science.* Cambridge University Press.

Federal Reserve Bank of Atlanta, Research Department, 1000 Peachtree Street N.E., Atlanta, GA 30309–4470
E-mail address: mark.fisher@atl.frb.org
URL: http://www.markfisher.net