A POLYNOMIAL KERNEL

Similar to the derivation of the histogram kernel, we can also derive the polynomial kernel for moment observations. The entries of the polynomial kernel, given by $k(x, x') = (xx' + c)^d$, can be integrated over as,

$$
\begin{aligned}
\kappa\left(\mathbf{R}_x^{(k)}, x'\right) &= \int_0^1 \sum_{i=1}^{d} \binom{d}{i} x^{k+i} x'^i c^{d-i} dx \\
&= \sum_{i=1}^{d} \binom{d}{i} \frac{x'^i c^{d-i}}{k+i+1},
\end{aligned} \tag{19}
$$

$$
\begin{aligned}
\kappa\left(\mathbf{R}_x^{(k)}, \mathbf{R}_{x'}^{(k')}\right) &= \int_0^1 \int_0^1 \sum_{i=1}^{d} \binom{d}{i} x^{k+i} x'^{k'+i} c^{d-i} dx dx' \\
&= \sum_{i=1}^{d} \binom{d}{i} \frac{c^{d-i}}{(k+i+1)(k'+i+1)}.
\end{aligned} \tag{20}
$$

As with the histogram kernel, the infinite sum of the Taylor expansion can also be combined into the Gaussian process,

$$
\begin{aligned}
\kappa\left(\sum_{k=1}^{\infty} \frac{\mathbf{R}_x^{(k)}}{k}, \mathbf{R}_{x'}^{(k')}\right) &= \frac{1}{k} \sum_{k=1}^{\infty} \sum_{i=1}^{d} \binom{d}{i} \frac{c^{d-i}}{(k+i+1)(k'+i+1)} \\
&= \sum_{i=1}^{d} \binom{d}{i} \frac{c^{d-i}\left(\Psi^{(0)}(i+2) + \gamma\right)}{(i+1)(k'+i+1)},
\end{aligned} \tag{21}
$$

$$
\begin{aligned}
\kappa\left(\sum_{k=1}^{\infty} \frac{\mathbf{R}_x^{(k)}}{k}, \sum_{k'=1}^{\infty} \frac{\mathbf{R}_{x'}^{(k')}}{k'}\right) &= \frac{1}{kk'} \sum_{k=1}^{\infty} \sum_{k'=1}^{\infty} \sum_{i=1}^{d} \binom{d}{i} \frac{c^{d-i}}{(k+i+1)(k'+i+1)} \\
&= \sum_{i=1}^{d} \binom{d}{i} \frac{c^{d-i}\left(\Psi^{(0)}(i+2) + \gamma\right)^2}{(i+1)^2}.
\end{aligned} \tag{22}
$$

In the above, $\Psi^{(0)}(\cdot)$ is the digamma function and γ is the Euler-Mascheroni constant. We strongly believe that the polynomial and histogram kernels are not the only kernels which can be analytically derived to include moment observations but act as a reasonable initial choice for practitioners.

B BOUNDS ON LOG DETERMINANTS

For the sake of completeness, we restate the bounds on the log determinants used throughout this paper (Bai & Golub, 1997).

Theorem 1 *Let A be an n-by-n symmetric positive definite matrix, $\mu_1 = Tr(A)$, $\mu_2 = \|A\|_F^2$ and $\lambda_i(A) \in [\alpha; \beta]$ with $\alpha > 0$, then*

$$
\begin{bmatrix} \log\alpha \\ \log t \end{bmatrix}^T \begin{bmatrix} \alpha & t \\ \alpha^2 & t^2 \end{bmatrix}^{-1} \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} \leq Tr(\log(A)) \leq \begin{bmatrix} \log\beta \\ \log\bar{t} \end{bmatrix}^T \begin{bmatrix} \beta & \bar{t} \\ \beta^2 & \bar{t}^2 \end{bmatrix}^{-1} \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}
$$

where,

$$
t = \frac{\alpha\mu_1 - \mu_2}{\alpha n - \mu_1}, \quad \bar{t} = \frac{\beta\mu_1 - \mu_2}{\beta n - \mu_1}
$$

This bound can be easily computed while loading the matrix as both the trace and Frobenius norm can be readily calculated using summary statistics. However, bounds on the maximum and minimum must also be derived. We chose to use Gershgorin intervals to bound the eigenvalues (Gershgorin, 1931).