

Stat 992: Lecture 35

Smoothing periodic functional data

Moo K. Chung mchung@stat.wisc.edu

April 23, 2004

1. Solution to Problem 34. Diffusion smoothing on a unit circle. Motivation would be when we have periodic functional data $g(\theta) = g(\theta + L)$. This periodic condition can be transformed to

$$g\left(\frac{L}{2\pi} \frac{2\pi}{L} \theta\right) = g\left(\frac{L}{2\pi} \left(\frac{2\pi}{L} \theta + 2\pi\right)\right).$$

So any period L function can be transformed to a periodic function with period 2π . So without loss of generality, we consider periodic data of the form $g(\theta) = g(\theta + 2\pi)$ which is viewed as data defined on a unit circle.

Laplacian on S^1 is trivially $\Delta = \partial_\theta^2$. This can be seen from the line element $dl = d\theta$ so the Laplacian on a circle must be the Laplacian in 1D Euclidean space. We need to solve $\partial_t f = \partial_\theta^2 f$ with initial condition $f(\theta, 0) = g(\theta)$. The solution to an isotropic diffusion equation on manifold is given as a heat kernel convolution by finding the eigenfunctions of Laplacian first. We solve $\partial_\theta^2 H + \lambda H = 0$. Trivially two independent solutions are $H = \cos \sqrt{\lambda} \theta, \sin \sqrt{\lambda} \theta$ with $\lambda > 0$. To grantee the periodicity of the eigenfunctions, $\sqrt{\lambda} = j \in \mathbb{Z}^+$. So we have sequence of eigenfunctions $H_j = \sin j\theta, \tilde{H}_j = \cos j\theta$. $j = 0, 1, 2, \dots$. Note that $\int_0^{2\pi} \sin^2 j\theta d\theta = \int_0^{2\pi} \cos^2 j\theta d\theta = \pi$ from symmetry. So we normalize the eigenfunctions $H_j = \frac{1}{\pi} \sin j\theta, \tilde{H}_j = \frac{1}{\pi} \cos j\theta$. This gives an orthonormal system of basis functions on S^1 .

From Lecture 21, heat kernel expansion,

$$\begin{aligned} K_t(\theta_1, \theta_2) &= \frac{1}{\pi^2} \sum_{j=0}^{\infty} e^{-j^2 t} [\sin(j\theta_1) \sin(j\theta_2) \\ &\quad + \cos(j\theta_1) \cos(j\theta_2)] \\ &= \frac{1}{\pi^2} \sum_{j=0}^{\infty} e^{-j^2 t} \cos[j(\theta_1 - \theta_2)]. \end{aligned}$$

This makes intuitive sense since the kernel must be isotropic. Then the solution to the diffusion

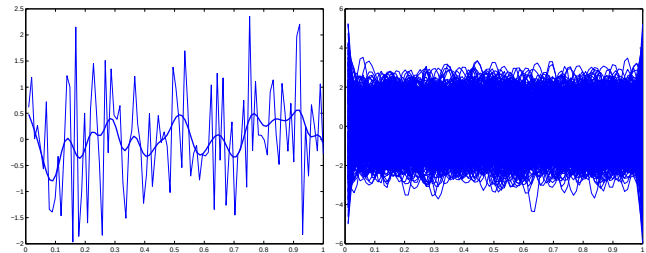


Figure 1: Left: White noise $N(0, 1)$ and $K_\sigma * N(0, 1)$. Right: 1000 zero mean unit variance Gaussian field. We are seeing the boundary effect of kernel smoothing.

smoothing is given by

$$\begin{aligned} f(\theta, t) &= \int_0^{2\pi} K_t(\theta, \phi) g(\phi) d\phi \\ &= \frac{1}{\pi^2} \sum_{j=0}^{\infty} e^{-j^2 t} \int_0^{2\pi} g(\phi) \cos[j(\theta - \phi)] d\phi. \end{aligned}$$

Obviously it is much easier to discretize the diffusion equation than implementing the above analytical solution exactly.

2. Solution to Problem 36 (Tulaya Limpiti). Simulate zero mean isotropic Gaussian field in $[0, 1]$ and find the *corrected* P-value numerically. First we generate 5000 Gaussian white noise $N(0, 1)$ apply iterated Gaussian kernel smoothing.

```
weight= K(1, [-1 0 1])
>>weight =
    0.2119    0.5761    0.2119
GRF=zeros(5000,99);
for k=1:5000
    w=normrnd(0,1,101,1);
    tem=w;
    for j=1:20
        for i=2:100
            tem(i) = dot(w((i-1):(i+1)),weight);
        end;
        w=tem;
    end;
```

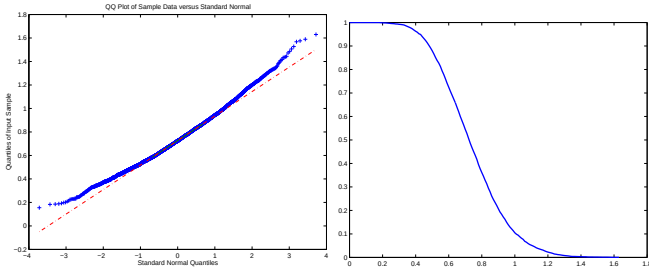


Figure 2: Left: QQ-plot of the suprema Gaussian random field. Right: Monte-Carlo version of P -value of the suprema field.

```
w=w(2:100);
GRF(k,:)=w';
end;
x=[1:99]/99; plot(x,w);
```

Generating random field in this way cause the boundary effect effect (Figure 1.). So we need to chop off fields near boundary. So we will only use $w(10:90)$ and rerun the above code.

```
sup=max(GRF,[],2)
pvalue=inline('length(find(sup>=h))/5000')
x=[0:0.01:1.63];
for i=1:length(x)
    P(i)=pvalue(x(i),sup);
end;
plot(x,P,'linewidth',2)
```

3. *Quantile-quantile plot* (QQ-plot) QQ-plot was first introduced by Wilk and Gnanadesikan (1968). Quantile point q for random variable X is a point that satisfies $P(X \leq q) = F_X(q) = p$, where F_X is the cumulative distribution function (cdf) of X . Assuming we can find the inverse cdf, $q = F_X^{-1}(p)$. There is a way to define the inverse of cdf (see Casella and Berger's Statistical inference). Then we can define the quantile-quantile plot of two distributions X and Y to be a parametric curve $(F_X^{-1}(p), F_Y^{-1}(p)) \in \mathbb{R}^2$ for $p \in [0, 1]$.

Problem 48. Prove that the QQ plot of $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$ is a straight line. What is the slope and the intersection? Bonus problem: Find the QQ plot of two lognormal distributions $\exp(N(\mu_1, \sigma_1^2)), \exp(N(\mu_2, \sigma_2^2))$ analytically.