

# Stat 992: Lecture 18

## Multiple comparison

Moo K. Chung `mchung@stat.wisc.edu`

March 1, 2004

**Problem 6.** Is the variance of signal smaller after smoothing?

**Theorem.** Assume the covariance function of field  $Y$  to be isotopic, i.e.  $R_Y(x, y) = f(\|x - y\|)$  and  $f$  decreasing. Then

$$\text{Var}[K_\sigma * Y(x)] \leq \text{Var}Y(x) \text{ for each } x \in \mathbb{R}^n.$$

*Proof.* (Shijie Tang) Note that  $R_Y(x', y') = f(\|x' - y'\|) \leq f(0) = R_Y(x', x')$ . Let us denote  $\text{Var}Y(x) = f(0) = \zeta^2$ . Isotropic covariance function basically implies stationary uniform variance field.

The covariance function  $R$  of  $K_\sigma * Y(x)$  is given by

$$\begin{aligned} R(x, y) &= \int K_\sigma(x - x')K_\sigma(y - y')R_Y(x', y') dx' dy' \\ &\leq \int K_\sigma(x - x')K_\sigma(y - y')\zeta^2 dx' dy' \\ &\leq \zeta^2. \end{aligned}$$

By letting  $x = y$ , we prove the statement.  $\square$

While investigating the above theorem, I realized that we did not actually prove the iterated kernel identity for truncated and normalized kernel! Note that

$$\tilde{K}_\sigma * \tilde{K}_\sigma \neq \tilde{K}_{\sqrt{2}\sigma}.$$

But our numerical implementation of kernel smoothing is based on iterated kernel smoothing so how we guarantee the convergence to the integral version of kernel smoothing?

**Theorem.**

$$K_\sigma * K_\sigma(x) \leq \tilde{K}_\sigma * \tilde{K}_\sigma(x) \leq \frac{1}{\alpha^2(\Omega)} K_\sigma * K_\sigma(x), x \in \Omega$$

where  $\alpha(\Omega) = \int_\Omega K_\sigma(x) dx$ . Obviously this can be generalized into  $k$ -iterated truncated normalized kernel smoothing so that

$$K_\sigma^{(k)}(x) \leq \tilde{K}_\sigma^{(k)}(x) \leq \frac{1}{\alpha^k(\Omega)} K_\sigma^{(k)}(x), x \in \Omega$$

For the proper choice of  $\sigma$  and  $\Omega$ , we can make  $\alpha = 0.9999$ . The usual number of iterations is less than 100-200 in practical application. That gives  $\alpha^{100} = 0.9900$

and  $\alpha^{200} = 0.9802$ . For  $\alpha = 0.999$ ,  $\alpha^{100} = 0.9048$  and  $\alpha^{200} = 0.8186$  Hence, decreasing the size of bandwidth and increasing the number of iterations would perform better.

*Proof.*

$$K_\sigma(x) = \alpha \tilde{K}_\sigma(x) + K_\sigma(x)I_{\Omega^c}(x).$$

Apply convolution to the above equation.

$$\begin{aligned} K_\sigma * K_\sigma(x) &\geq \alpha K_\sigma * \tilde{K}_\sigma(x) \\ &\geq \alpha^2 \tilde{K}_\sigma(x) * \tilde{K}_\sigma(x). \end{aligned}$$

Trivially  $K_\sigma(x) \leq \tilde{K}_\sigma(x), x \in \Omega$  since  $0 < \alpha < 1$ .  $\square$  This theorem shows that the iterated truncated kernel smoothing can approximate the integral version of kernel smoothing within one percent error if one desires it in practical applications.

1. *Multiple comparison.* We have an image  $Y$  of size  $95 \times 68$  where each pixel intensity is assumed to follow Gaussian. Suppose we have model

$$Y(x) = \mu(x) + \epsilon(x)$$

where  $\epsilon$  is a zero mean unit variance Gaussian field. In imaging, one of the most important problem is that of detection, which can be stated as the problem of identifying the region of statistically significant region. So it can be formulated as an inference problem

$$H_0 : \mu(x) = 0 \text{ for all } x \text{ vs. } H_1 : \mu(x) > 0 \text{ for some } x.$$

Note that the null hypothesis is based on multiple hypotheses. Let  $H_0^j : \mu(x_j) = 0, x_j \in \Omega$  and assume  $\Omega = \{x_1, \dots, x_m\}$ . Then

$$H_0 = \bigcap_{j=1}^m H_0^j.$$

We also have corresponding alternate hypothesis  $H_1^j : \mu(x_j) > 0$ . For simplicity, use  $Z$ -statistic for test and we will reject each  $H_0^j$  if  $Z > h$  for some

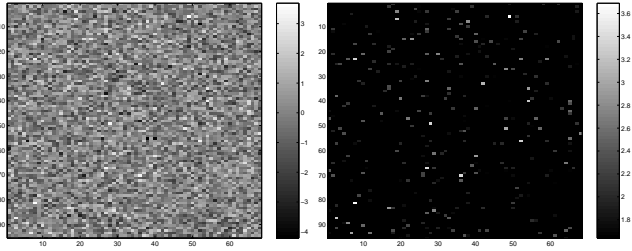


Figure 1: Left: original  $N(0,1)$  white noise image. Right: 5% of pixels are false positives at  $\alpha = 0.05$  level tests.

threshold  $h$ . For level  $\alpha = 0.05$  test,  $h = 1.64$  (Figure 1). The type I error for the multiple hypotheses testing would be

$$\begin{aligned} \alpha &= P(\text{reject } H_0 | H_0 \text{ true}) \\ &= P(\text{reject some } H_0^j | H_0 \text{ true}) \\ &= P\left(\bigcup_{j=1}^m \{Y(x_j) > h\} \mid \mathbb{E}Y = 0\right). \end{aligned}$$

Unfortunately,  $Y(x_j)$ 's are correlated which makes the computation of type I error almost intractable for random fields other than Gaussian. Read Keith Worsley's papers at <http://www.math.mcgill.ca/keith>. He made random field approach for multiple comparison very popular in imaging.

2. *i.i.d. simulation.* For simplicity of argument, assume there is a single image as an observation. For  $Z \sim N(0,1)$ ,

$$P\text{-value} = P(Z > 1.64) = 0.05.$$

So we expect 5% of observations to be false positive.

```
Y=normrnd(0,1,95,68); imagesc(Y);
colorbar; colormap('bone')
[Yl,Yh]=image_threshold(Y, 1.64);
imagesc(Yh);colorbar; n=95*68
size(find(reshape(Y,n,1)>=1.64),1)/n
>ans =
    0.05
```

3. *Random field simulation.*

```
Y_smoos=gaussblur(Y,4); imagesc(Y_smoos);
colorbar; colormap('bone');
[Yl,Yh]=image_threshold(Y_smoos, 1.64);
imagesc(Yh);colorbar; colormap('bone')
```

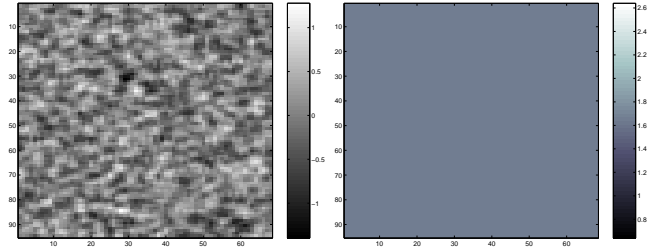


Figure 2: Left: Gaussian kernel smoothing with 4mm FWHM filter. FWHM=full width at half maximum. Right: threshold at 1.64 shows no statistically significant regions. So kernel smoothing can be considered as an ad-hoc approach for removing false positives.

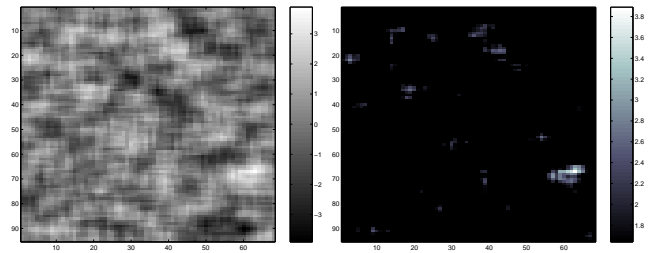


Figure 3: Left: zero mean unit variance Gaussian field constructed from 12mm fWHM filter. Right: threshold at 1.64 showing statistically significant clustered regions

```
Y_smoos=Y_smoos/std(reshape(Y_smoos,n,1));
>size(find(reshape(Y_smoos,n,1)>=1.64),1)/n
ans =
    0.0424
```

**Problem 28.** Let  $[Y(x_1), \dots, Y(x_m)]' \sim N(0, \Sigma)$  and  $\mathbb{E}[Y(x_i)Y(x_j)] = \sigma_{ij}$ . Compute the probability

$$\alpha(h) = P(\text{for some } j, Y(x_j) > h).$$

If you can't solve it analytically, try Monte-Carlo simulation for many  $h$ 's and based on the simulation plot  $\alpha(h)$ .