FALL 2013
Stat 8003: Statistical Methods I
Lecture 14: Review Session

Jichun Xie

In this lecture, we will discuss some comprehensive application problems which can be solved by what have been discussed before. As you know, a good applied statistician should be able to use existing methods or techniques to formulate complicated problems and solve them reasonably.

1 Example 1

Positron Emission Tomography (PET) is performed by introducing a radioactive tracer into an animal or human subject. Radioactive emissions are then used to assess levels of metabolic activity and blood flow in organs of interest. Positrons emitted by the tracer annihilate with nearby electrons, giving pairs of photons that fly off in opposite directions. Some of these are counted by bands of gamma ray detectors placed around the subjects body, but the others (photons) miss the detectors. The detected counts are used to form an image of the level of metabolic activity in the organs based on the estimate spatial concentration of the isotope.

For a statistical model, the region of interest is divided into n pixels or voxels and it is assumed that the number of emmissions U_{ij} from the jth pixel detected at the ith detector is a Poisson random variable with mean $p_{ij}\lambda_j$; here λ_j is the intensity of emissions from that pixel and p_{ij} is the probability that a single emission is detected at the ith detector: The $\lambda_1, \ldots, \lambda_n$ are unknown and our parameters of interest; the p_{ij}, $i = 1, \ldots, d$, $j = 1, \ldots, n$ depend on the geometry of the detection system, the isotope and other factors, but can be taken to be known. The U_{ij} are unknown but can be plausibly be assumed independent. The counts Y_i ($i = 1, \ldots d$) at the ith detector are observed and have independent Poisson distributions with mean $\sum_{j=1}^{n} p_{ij}\lambda_j$.

What are the maximum likelihood estimates for emission intensities $\lambda_1, \ldots, \lambda_n$ based on the observed data $Y_1, \ldots, Y_d$ and the known detection probabilities p_{ij}, $i = 1, \ldots, n$, $j = 1, \ldots, d$?

Consider viewing the U_{ij}, $i = 1, \ldots, n$, $j = 1, \ldots, d$ as missing data. The complete data is then

$$(\mathbf{u}, \mathbf{y}) = (\{u_{ij}, i = 1, \ldots, n, j = 1, \ldots, d\}, \{y_1, \ldots, y_n\}),$$

where $y_i = \sum_{j=1}^{n} u_{ij}$. The complete data log likelihood is

$$l_{\mathbf{u},\mathbf{y}} = \sum_{i=1}^{d} \sum_{j=1}^{n} \{u_{ij} \log(p_{ij}\lambda_j) - p_{ij}\lambda_j\}.$$

This is an exponential family. The expected complete data log liklihood, where the missing data follows its conditional distribution given the observed data and the current parameter estimates of $\hat{\lambda}_1^{(k-1)}, \ldots, \hat{\lambda}_n^{(k-1)}$, is

$$\begin{aligned} Q & \left\{(\lambda_1, \ldots, \lambda_n), (\hat{\lambda}_1^{(j-1)}, \ldots, \hat{\lambda}_n^{(k-1)})\right\} \\ & = \sum_{i=1}^{d} \sum_{j=1}^{n} \left\{ \mathtt{E}(u_{ij} \mid y_1, \ldots, y_n, \hat{\lambda}_1^{(k-1)}, \ldots, \hat{\lambda}_n^{(k-1)}) \log(p_{ij}\lambda_j) - p_{ij}\lambda_j \right\} \end{aligned}$$

To complete the E step, we need to calculate

$$\mathtt{E}(U_{ij} \mid y_1, \ldots, y_d, \hat{\lambda}_1^{(k-1)}, \ldots, \hat{\lambda}_n^{(k-1)}).$$

As $Y_i = \sum_{j=1}^{n} U_{ij}$, the conditional density of U_{ij} given $Y_i = y_i$ is binomial with y_i trials and probability of success

$$\frac{p_{ij}\hat{\lambda}_j^{(k-1)}}{\sum_{h=1}^{n} p_{ih}\hat{\lambda}_h^{(k-1)}};$$

thus,

$$\mathtt{E}(U_{ij} \mid y_1, \ldots, y_d, \hat{\lambda}_1^{(k-1)}, \ldots, \hat{\lambda}_n^{(k-1)}) = \frac{p_{ij}\hat{\lambda}_j^{(k-1)}}{\sum_{h=1}^{n} p_{ih}\hat{\lambda}_h^{(k-1)}} \cdot y_i.$$

The M step yields

$$\begin{aligned} \hat{\lambda}_j^{(k)} &= \frac{\sum_{i=1}^{d} \mathtt{E}(U_{ij} \mid y_1, \ldots, y_d, \hat{\lambda}_1^{(k-1)}, \ldots, \hat{\lambda}_n^{(k-1)})}{\sum_{i=1}^{d} p_{ij}} \\ &= \frac{\sum_{i=1}^{d} (p_{ij}\hat{\lambda}_j^{(k-1)} y_i)/\{\sum_{h=1}^{n} p_{ih}\hat{\lambda}_h^{(k-1)}\}}{\sum_{i=1}^{d} p_{ij}} \\ &= \hat{\lambda}_j^{(k-1)} \frac{1}{\sum_{i=1}^{d} p_{ij}} \sum_{i=1}^{d} \frac{p_{ij} y_i}{\sum_{h=1}^{n} \hat{\lambda}_h^{(k-1)} p_{ih}}, \quad j = 1, \ldots, n. \end{aligned}$$

It has been shown (Vardi *et al.*, 1985, *Journal of the American Statistical Association*) that $\hat{\lambda}_1^{(k)}, \ldots, \hat{\lambda}_n^{(k)}$ converges to the maximizer of the likelihood function of $\lambda_1, \ldots, \lambda_n$ given the observed data $y_1, \ldots, y_n$ and the known detection probabilities $p_{ij}, i = 1, \ldots, n, j = 1, \ldots, d$ but there will not be a unique MLE if the number of pixels n is greater than the number of detectors d.

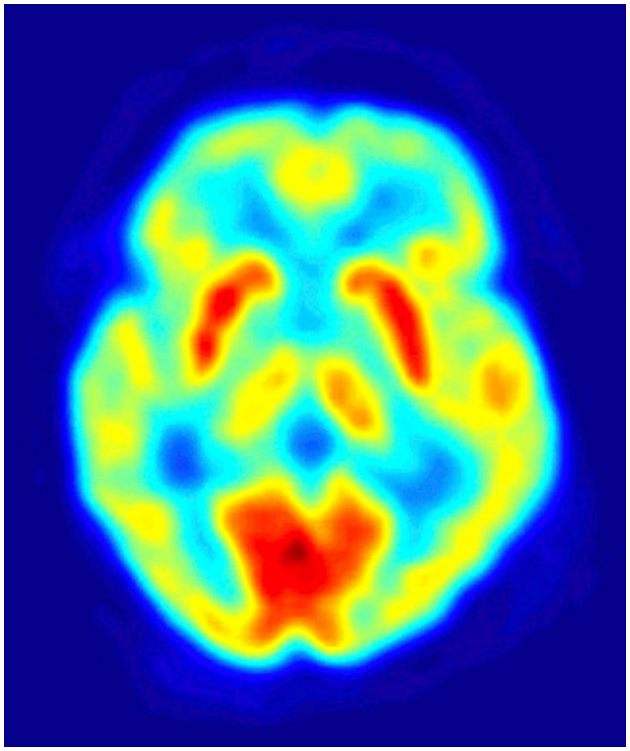

Figure 1: PET Scan of the Human Brain

2 Example 2

When we discuss linear models, we usually assume that the covariates $X_1, \ldots, X_p$ are given, or in other words, fixed. We also mentioned that even if X_i is viewed as a random variable, some of the main inference (like the estimates of $\boldsymbol{\beta}$) won't change. But we also mentioned in the previous lecture that random variable itself will bring extra noise. How to perform the linear model analysis when the covariates are random variables? Here comes an exmaple.

Selling Price of Antique Grandfather Clocks. The data give the selling price at auction of 32 antique grandfather clocks. Also recorded is the age of the clock and the number of people who made a bid.

Variable	Description
Age	Age of the clock (years)
Bidders	Number of individuals participating in the bidding
Price	Selling price (pounds sterling)

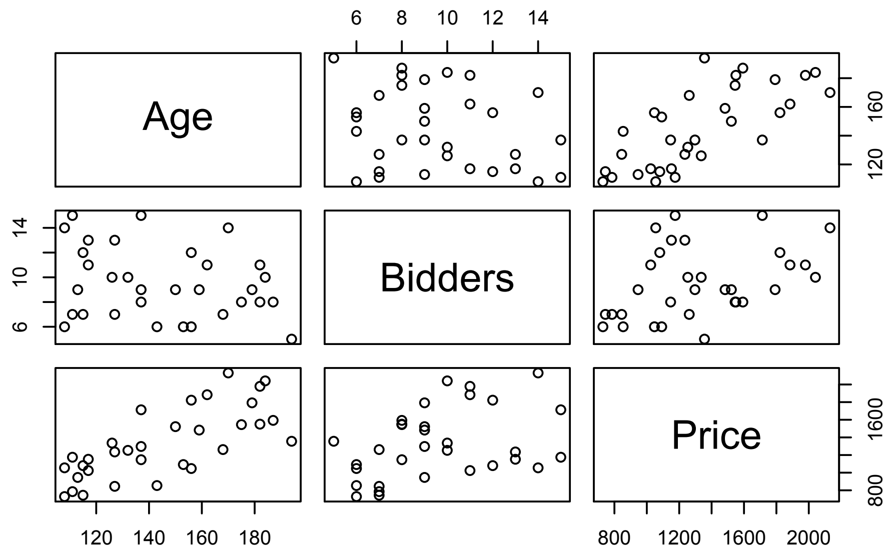

Figure 2: Scatter Plot of Antique Clocks Auction

Now suppose the goal is to investigate the relationship between auction price and the age of clock and the number of bidders. It is reasonable to treat both the age of clock and the number of bidders as known. We can assume the following linear model:

$$Y_k = \beta_0 + \beta_1 X_{k,1} + \beta_2 X_{k,2} + \epsilon_k, \quad k = 1, \ldots, n \tag{1}$$

Here Y_k is auction price for the kth clock, $X_{k,1}$ is age of the kth clock and $X_{k,2}$ is the number of bidders for the kth clock. Here ϵ_k is the random error term measuring other unknown random factors that will affect the price of the kth clock. Assume $\epsilon_k \sim \mathrm{N}(0, \sigma^2)$.

```
> fit1 = lm(Price~Age+Bidders, data = auction)
> summary(fit1)

Call:
lm(formula = Price ~ Age + Bidders, data = auction)

Residuals:
   Min      1Q  Median      3Q     Max
-207.2 -117.8    16.5   102.7   213.5
```

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -1336.7221   173.3561  -7.711 1.67e-08 ***
Age            12.7362     0.9024  14.114 1.60e-14 ***
Bidders        85.8151     8.7058   9.857 9.14e-11 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 133.1 on 29 degrees of freedom
Multiple R-squared:  0.8927,  Adjusted R-squared:  0.8853
F-statistic: 120.7 on 2 and 29 DF,  p-value: 8.769e-15
```

Both factors (age and number of bidders) are significant. Is there any serious collinarity problem here?

Now suppose there is another antique clock ready for auction. Some antique investor would like to predict the price of this clock. Based on the advertisement, the clock is 153 years old. How to predict the price of the clock?

Note that before the auction, the number of bidders is unknown. We use model (1) to investigate the relationship between price and age and number of bidders. But due to the randomness of bidders, the model and the previous analysis cannot help the investor much. Instead, it is reasonable to consider the following model here:

$$Y_k = \alpha_0 + \alpha_1 X_{k,1} + \alpha_2 Z_k + \epsilon_k, \quad k = 1, \dots, n \tag{2}$$

Here Y_k is still the price and $X_{k,1}$ is the age of the kth clock. $X_{k,1}$ is fixed (known before the auction). Z_k is the *centered and standardized* number of bidders of the kth clock, which is random (unknown before the auction). We assume $Z_k \sim \mathrm{N}(0, 1)$. And we also assume $\epsilon_k \sim \mathrm{N}(0, \sigma^2)$ and $\epsilon_k \perp\!\!\!\perp Z_k$.

Now try to answer the following questions first.

(a). What are the expectation and variance of Y_k?

(b). The variance has two component. What is your interpretation of each?

(c). What is the joint distribution of Y_k and Z_k?

(d). What are the maximum likelihood estimators of α_0, α_1, α_2 and σ^2?

(e). Construct a Wald test of the null hypothesis $\mathrm{H}_0 : \alpha_1 = 0$.

(f). Predict the auction price of the clock with age equal to 153, and construct a 95% confidence interval.

Here come the solutions.

(a). It is easy to see that under model (2),

$$\mathtt{E}(Y_k) = \alpha_0 + \alpha_1 X_1, \quad \mathtt{Var}(Y_k) = \sigma^2 + \alpha_2^2.$$

Actually $Y_k \sim \mathrm{N}(\alpha_0 + \alpha_1 X_1, \sigma^2 + \alpha_2^2)$. Note the distribution of Y_k is different from that in model (1).

(b). The variance has two components. The σ^2 comes from the random error ϵ_k; α_2^2 comes from the randomness of the bidders.

(c). The joint distribution of Y_k and Z_k is bi-variate normal. We then need to investigate their mean, variance and $\mathtt{Cov}(Y_k, Z_k)$.

$$\mathtt{Cov}(Y_k, Z_k) = \mathtt{Cov}(\alpha_0 + \alpha_1 X_{k,1} + \alpha_2 Z_k + \epsilon_2, Z_k) = \alpha_2.$$

Combined with the model setting and problem (a), we have

$$\begin{pmatrix} Y_k \\ Z_k \end{pmatrix} \sim \mathrm{N}\left\{ \begin{pmatrix} \alpha_0 + \alpha_1 X_{k,1} \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2 + \alpha_2^2 & \alpha_2 \\ \alpha_2 & 1 \end{pmatrix} \right\}$$

You can use $f_{Y,Z}(y,z) = f_{Y|Z}(y \mid z) f_Z(z)$ to get the joint distribution too.

$$f_{Y,Z}(y,z) = \frac{1}{2\pi\sigma} \exp\left\{ -\frac{1}{2} z^2 - \frac{1}{2\sigma^2}(y - \alpha_0 - \alpha_1 x_1 - \alpha_2 z)^2 \right\}.$$

Remark: Bivariate normal distribution of $(Y,Z)^{\mathrm{T}}$. Suppose

$$\mathtt{E}\begin{pmatrix} Y \\ Z \end{pmatrix} = \boldsymbol{\mu} = \begin{pmatrix} \mu_Y \\ \mu_Z \end{pmatrix}, \quad \text{and } \mathtt{Var}\left\{ \begin{pmatrix} Y \\ Z \end{pmatrix} \right\} = \boldsymbol{\Sigma} = \begin{pmatrix} \sigma_Y^2 & \rho\sigma_Y\sigma_Z \\ \rho\sigma_Y\sigma_Z & \sigma_Z^2 \end{pmatrix}.$$

Here ρ is the correlation between Y and Z. Then the density function of $(Y,Z)^{\mathrm{T}}$ is

$$f_{Y,Z}(y,z) = \frac{1}{2\pi\sigma_Y\sigma_Z\sqrt{1-\rho^2}} \exp\left\{ -\frac{1}{2(1-\rho^2)} \left[\frac{(y-\mu_Y)^2}{\sigma_Y^2} + \frac{(z-\mu_Z)^2}{\sigma_Z^2} - \frac{2\rho(y-\mu_Y)(z-\mu_Z)}{\sigma_Y\sigma_Z} \right] \right\},$$

In the clock auction example,

$$\rho = \mathtt{Cor}(Y_k, Z_k) = \frac{\alpha_2}{(\sigma^2 + \alpha_2^2)^{1/2}}.$$

(d). To obtain MLE, we start from the joint distribution obtained in (c).

$$
\begin{aligned}
L(\alpha_0, \alpha_1, \alpha_2, \sigma^2) &= (2\pi\sigma)^{-n} \exp\left\{ -\frac{1}{2}\sum_{k=1}^{n} z_k^2 - \frac{1}{2\sigma^2}\sum_{k=1}^{n}(y_k - \alpha_0 - \alpha_1 x_{k,1} - \alpha_2 z_k)^2 \right\} \\
\log L(\alpha_0, \alpha_1, \alpha_2, \sigma^2) &= -\frac{n}{2}\log(\sigma^2) - \frac{1}{2}\sum_{k=1}^{n} z_k^2 - \frac{1}{2\sigma^2}\sum_{k=1}^{n}(y_k - \alpha_0 - \alpha_1 x_{k,1} - \alpha_2 z_k)^2 \\
&= -\frac{n}{2}\log(\sigma^2) - \frac{1}{2}\mathbf{Z}^{\mathrm{T}}\mathbf{Z} - \frac{1}{2\sigma^2}(\mathbf{Y} - \mathbf{X}\boldsymbol{\alpha})^{\mathrm{T}}(\mathbf{Y} - \mathbf{X}\boldsymbol{\alpha}),
\end{aligned}
$$

where $\mathbf{x}$ is the design matrix for $(1, X_1, Z)$. Take the derivate and set it to zero, we can get

$$
\hat{\alpha} = (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{Y} \tag{3}
$$

$$
\hat{\sigma}^2 = \frac{1}{n}(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\alpha}})^{\mathrm{T}}(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\alpha}}). \tag{4}
$$

And then we check the second derivative to confirm that (3) and (4) are MLEs. It is easy to see that the MLE for model (2) is the same of those for model (1).

(e). Let $\mathbf{e}_2 = (0, 1, 0)^{\mathrm{T}}$. Then $\hat{\alpha}_1 = \mathbf{e}_2^{\mathrm{T}}\hat{\boldsymbol{\alpha}}$.

We first derive $\mathtt{Var}(\hat{\boldsymbol{\alpha}})$. Note that the design matrix $\mathbf{X}$ involves random variable $\mathbf{Z}$.

$$
\begin{aligned}
\mathtt{Var}(\hat{\boldsymbol{\alpha}} \mid \mathbf{Z}) &= \sigma^2(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1} \\
\mathtt{E}(\hat{\boldsymbol{\alpha}} \mid \mathbf{Z}) &= \boldsymbol{\alpha}.
\end{aligned}
$$

Thus,

$$
\mathtt{Var}(\hat{\boldsymbol{\alpha}}) = \mathtt{E}\{\mathtt{Var}(\hat{\boldsymbol{\alpha}} \mid \mathbf{Z})\} + \mathtt{Var}\{\mathtt{E}(\hat{\boldsymbol{\alpha}} \mid \mathbf{Z})\} = \sigma^2 \mathtt{E}_Z\left\{(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\right\}.
$$

Consequently,

$$
\mathtt{Var}(\hat{\alpha}_1) = \mathbf{e}_2^{\mathrm{T}}\mathtt{Var}(\hat{\boldsymbol{\alpha}})\mathbf{e}_2 = \sigma^2 \mathtt{E}_Z\left\{\mathbf{e}_2^{\mathrm{T}}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{e}_2\right\},
$$

It is not easy to calculate $\mathtt{E}_Z\left\{\mathbf{e}_2^{\mathrm{T}}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{e}_2\right\}$. Suppose we can obtain an estimate for $\widehat{\mathtt{Var}}(\hat{\alpha}_1)$. Then to test $\mathrm{H}_0: \ \alpha_1 = 0$, the Wald test is constucted by

$$
W = \frac{\hat{\alpha}_1}{\sqrt{\widehat{\mathtt{Var}}(\hat{\alpha}_1)}}.
$$

Under H_0, we assume that W asymptotically follows $\mathrm{N}(0, 1)$ distribution.

In practice, $\mathtt{Var}(\hat{\alpha}_1)$ can be estimated by some numerical methods. But when n is large, it is not easy to get the numerical results. In that case, we can consider using conditional Wald test, that is,

$$
W \mid \mathbf{Z} = \frac{\hat{\alpha}_1 - \mathtt{E}(\hat{\alpha}_1 \mid \mathbf{Z})}{\sqrt{\widehat{\mathtt{Var}}(\hat{\alpha}_1 \mid \mathbf{Z})}}.
$$

This conditional Wald test is the same as the Wald test we discussed before for fixed design linear models.

(f). When a new clock comes in, we observe its age, $X_{0,1} = 153$. It is reasonable to use

$$\hat{Y}_0 = \hat{\alpha}_0 + \hat{\alpha}_1 X_{0,1}$$

to estimate its auction price Y_0, since $Z_0 \sim \mathrm{N}(0, 1)$ and $\epsilon_0 \sim \mathrm{N}(0, \sigma^2)$.

How to construct a 95% prediction CI for Y_0? Note that now

$$Y_0 = \alpha_0 + \alpha_1 X_{0,1} + \alpha_2 Z_0 + \epsilon_0.$$

We need to calculate $\mathtt{E}\{(\hat{Y}_0 - Y_0)^2\}$.

$$\begin{aligned}
\mathtt{E}\{(\hat{Y}_0 - Y_0)^2\} &= \mathtt{E}\left[\mathtt{E}\left\{(Y_0 - \hat{Y}_0)^2 \mid \mathbf{Z}, Z_0\right\}\right] \\
&= \mathtt{E}\left[\mathtt{E}\left(\left\{X_0^{\mathrm{T}}(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}) + (\hat{\epsilon}_0 - \epsilon_0)\right\}^2 \mid \mathbf{Z}, Z_0\right)\right] \\
&= \mathtt{E}\{X_0^{\mathrm{T}} \mathtt{Var}(\hat{\boldsymbol{\alpha}} \mid \mathbf{Z}) X_0 + \sigma^2\} \\
&= \mathtt{E}\{(\sigma^2 X_0^{\mathrm{T}} (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1} X_0 + \sigma^2) \mid \mathbf{Z}, Z_0\},
\end{aligned}$$

Again, if we can get the expression of $\mathtt{E}\{(\hat{Y}_0 - Y_0)^2\}$, we are able to build up a $(1 - \alpha)\%$ prediction CI:

$$\left(\hat{Y}_0 - z_{(1-\alpha)/2} \cdot [\hat{\mathtt{E}}(\hat{Y}_0 - Y_0)^2]^{1/2}, \quad \hat{Y}_0 + z_{(1-\alpha)/2} \cdot [\hat{\mathtt{E}}(\hat{Y}_0 - Y_0)^2]^{1/2}\right)$$

Note that for the prediction error, we cannot use the conditional probabability trick since Z_0 is not observed.