8.3 Nonparametric Poisson Regression

Definition. In a **nonparametric Poisson regression model**, the response variable y has Poisson distribution with the probability mass function

$$p(y) = \frac{\lambda^y}{y!}\, e^{-\lambda}, \text{ and}$$

$$\ln \lambda = \beta_0 + \beta_1\, x_1 + \cdots + \beta_m\, x_m + loess(x_{m+1}) + \cdots + loess(x_k)$$

where $x_1, \ldots, x_m$ is a set of **regression variables** (all 0-1 and possibly numeric predictors), and $x_{m+1}, \ldots, x_k$ are **smoothing predictors** (must be numeric).

Example. Going back to the data set "skin_cancer_data.csv", we use a nonparametric Poisson model to regress the number of new cancers on group and gender (regression predictors), and age, number of previous cancers, and year (smoothing predictors).

In SAS:

```
proc import out=cancer_data
datafile="./skin_cancer_data.csv" dbms=csv replace;

/*specifying data for prediction*/
data point4pred;
input group$ gender$ age nprevcancers year;
cards;
Tx F 57 8 3
;

/*fitting nonparametric Poisson model*/
proc gam data=cancer_data;
class group(ref="Tx") gender(ref="F");
model nnewcancers = param(group gender) loess(age)
loess(nprevcancers) loess(year) /link=log dist=poisson;
ods output OutputStatistics=results;
 score data=point4pred out=predicted;
run;
```

Regression Model Analysis Parameter Estimates				
Parameter	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	0.15595	0.32318	0.48	0.6298
group Cx	0.77365	0.10425	7.42	<.0001
group Tx	0	.	.	.
gender M	0.14554	0.09201	1.58	0.1147
gender F	0	.	.	.
Linear(age)	-0.00305	0.00452	-0.68	0.4998
Linear(nprevcancers)	0.06327	0.01289	4.91	<.0001
Linear(year)	-0.08077	0.03014	-2.68	0.0078

Smoothing Model Analysis Analysis of Deviance				
Source	DF	Sum of Squares	Chi-Square	Pr > ChiSq
Loess(age)	2.63991	10.754762	10.7548	0.0093
Loess(nprevcancers)	2.48772	8.660013	8.6600	0.0218
Loess(year)	2.48567	7.344099	7.3441	0.0406

```
proc print data=predicted noobs;
run;
```

group	gender	age	nprevcancers	year	P_nnewcancers	P_age	P_nprevcancers	P_year	LINP_nnewcancers
Tx	F	57	8	3	1.39703	0.088432	0.059870	-0.059779	0.33435

In R:

```
cancer.data<- read.csv(file="./skin_cancer_data.csv", header=TRUE, sep=",")

#specifying reference category
cancer.data$group.rel<- relevel(as.factor(cancer.data$group), ref="Tx")
cancer.data$gender.rel<- relevel(as.factor(cancer.data$gender), ref="F")

#fitting nonparametric Poisson regression
library(gam)
poisson.fit<- gam(nnewcancers ~ group.rel + gender.rel + lo(age) +
lo(nprevcancers) + lo(year), data=cancer.data, family=poisson)
coefficients(summary.glm(poisson.fit))
```

```
                    Estimate  Std. Error    z value     Pr(>|z|)
(Intercept)       0.13051647 0.327459186  0.3985732 6.902077e-01
group.relCx       0.71661329 0.104195477  6.8775854 6.087554e-12
gender.relM       0.16707481 0.091766964  1.8206423 6.866125e-02
lo(age)          -0.00276412 0.004606929 -0.5999920 5.485116e-01
lo(nprevcancers)  0.07033089 0.013268885  5.3004369 1.155259e-07
lo(year)         -0.08045262 0.030026535 -2.6793841 7.375772e-03
```

```
#using the fitted model for prediction
predict(poisson.fit, data.frame(group.rel="Tx", gender.rel="F", age=57,
nprevcancers=8, year=3), type="response")
```

```
1.742775
```

□