Flexible models for under and over dispersion

Richard P. Waterman

University of Pennsylvania

Hanfeng Gong

University of Pennsylvania

April 23, 1997

Outline

- Two motivating examples.
  - Quality control on count data
  - Modeling square contingency tables – adolescent psych study
  - Plan: extend Rasch model for this data
- Common attribute: mis-specified dispersion
- Approaches to under and over-dispersion
- Parameterized offsets – multinomial embeddings
- Example distributions – what they look like
- Properties of the extensions
- What it does for the quality control
- Latent variable ideas – a Rasch type model

  - Estimation issues
  - Eliminating nuisance parameters
  - Parameter interpretation
- What it does for the square table
- Other ideas
- Multinomial models – inference here
- Code availability

Bug Data: Surveillance vs Screening.

Context: hospital epidemiology. About 6% of people get sick when they are IN hospital.

- Personnel do not have time to screen every positive culture.
- Resource allocation problem.
- Look for indicators of trouble

Daily counts on 17 micro-organisms – bug data.

Best current practice. Monitor data continuously and use an 80 percentile alarm.

Like an empirical one-sided Shewhert chart.

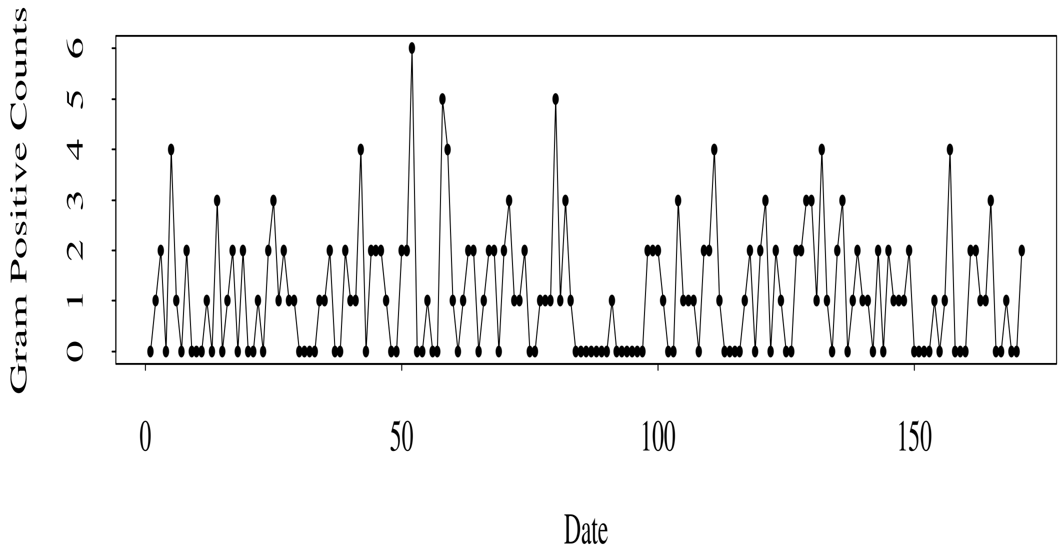
Possible refinements

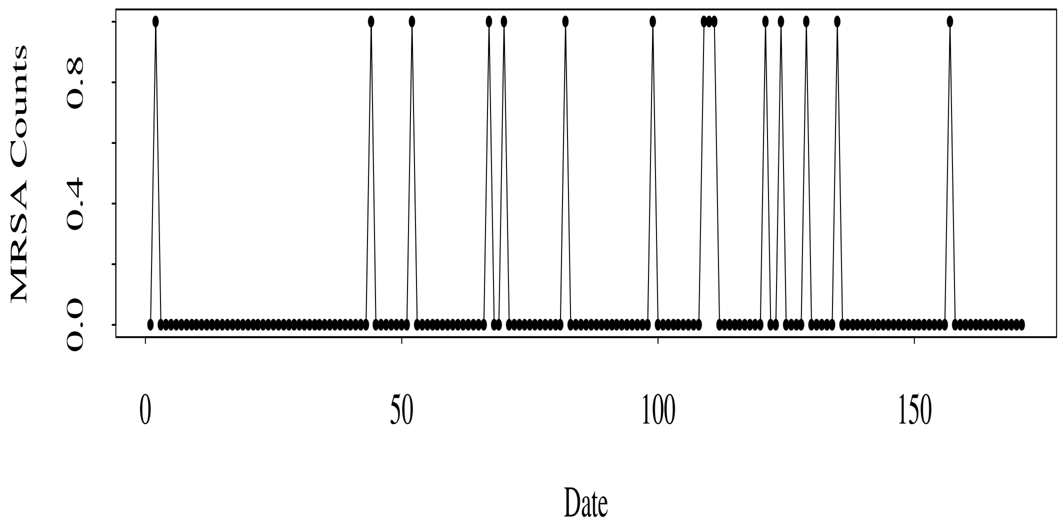
1. Cusum procedure. Power against smaller changes in the underlying process.
2. Take advantage of multivariate nature of the data. Each series by itself may not ring the alarm, but taken together they should.

Examples of the data.

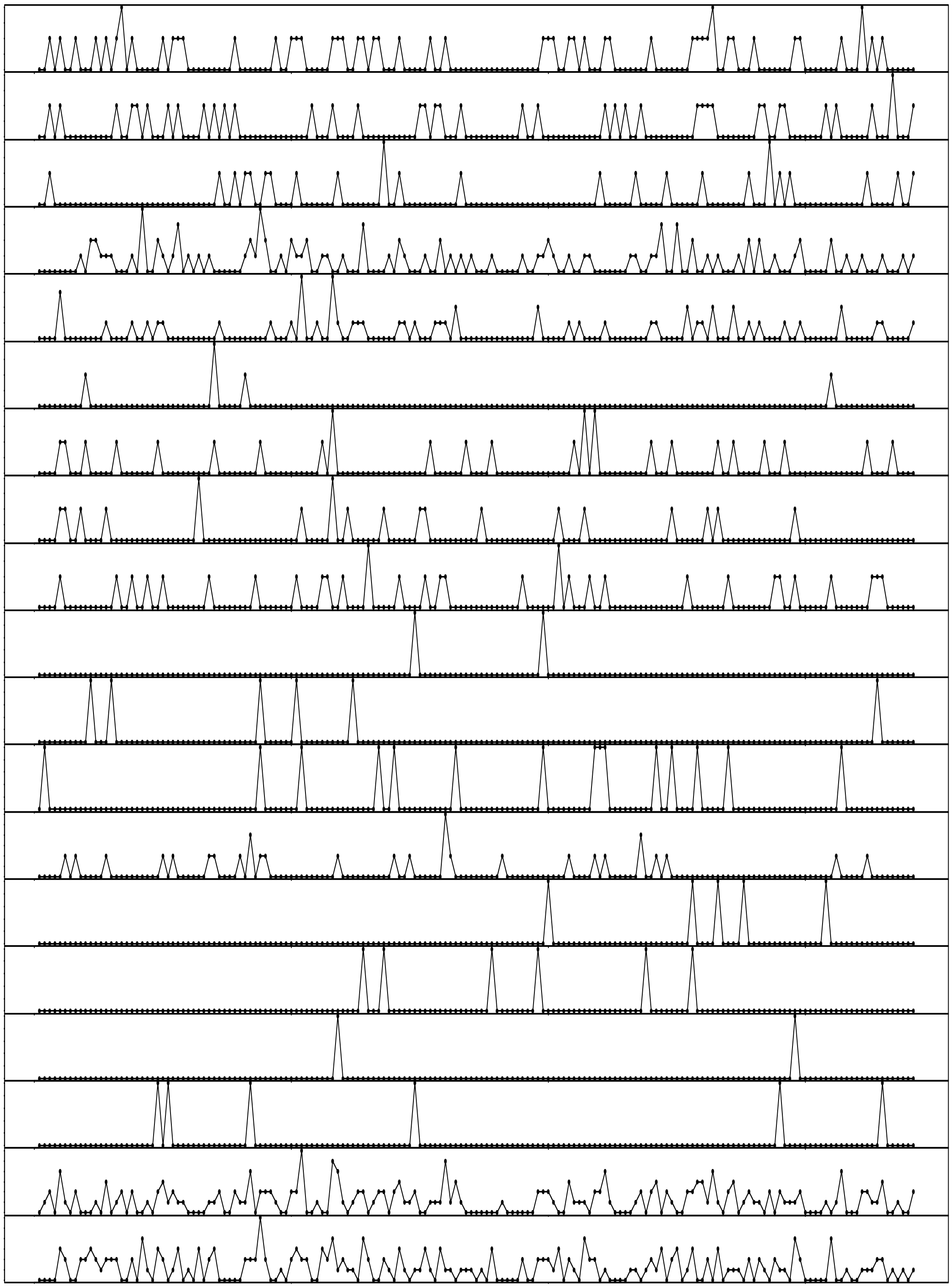
- 17 time series.
- Note: they are small integer counts. Poisson?
- Sometimes may want to group together.
- Gram positive vs Gram negative. Because this is how treatment is determined. Sensible reduction in dimensionality.
- Others should be looked at alone because of inherent interest. MRSA. Resistant to all known treatments!

Bug series.


Gram Positive Counts
0
1
2
3
4
5
6
0
50
100
150
Date


MRSA Counts
0.0
0.4
0.8
0
50
100
150
Date

Bug series (ctd.)



Simple models

Poisson fits?

```
Count     0  1  2 3 4
Freq   107 44 13 6 1
```

Run a poisson regression, with only intercept.

```
Coefficients:
                    Value Std. Error t value
Intercept       0.1000912 0.07259616 1.37874

(Dispersion Parameter for Poisson family taken to be 1 )

Residual Deviance: 250.903 on 170 degrees of freedom
```

Note residual deviance versus residual df.

Some overdispersed others underdispersed.

Observed versus predicted values from the Poisson fit.

```
Count        0    1    2    3    4    5    6
Freq       107   44   13    6    1    0    0
Expected  99.8 53.7 14.5  2.6  0.3  0.0  0.0
```

Summary: doesn't fit so well. Especially in the tails where it matters. Quality control calcs will all be about tail probabilities. Need a good model.

Refs for Poisson quality.

Adolescent psych study.

Context. Questionnaire: 70 items. 4000 adolescents around the world.

Objectives

- Partly exploratory.
- Partly to see if some local theories hold up around the world.
- This part. How do individuals see themselves as compared to their peer groups.

Get at this by asking paired questions.

Example questions. Six such pairs.

The line to the right represents the distance between childhood and adulthood. Place an “X” in the square in which you see yourself today.

Today's question.

How do people rate themselves in terms of maturity as compared to their peers. Comparisons across cultures.

Scale is 1 through 6. Ordinal.

Problems. Different cultures may naturally be at different points on the scale.

Repeated measures categorical data.

A popular model for dichotomous responses is the Rasch. Educational testing.

Would like to extend this.

Bernoulli to Binomial? Since 1 thru 6.

Focus on a single item

Scots:

Value	1	2	3	4	5	6
Freq.	5	10	150	375	214	15
Pred.	6.5	51.8	165.7	265.1	212.1	67.8

US

Value	1	2	3	4	5	6
Freq.	14	37	217	751	353	93
Pred.	10.4	88.0	297.3	502.1	424.0	143.2

Look at distribution compared to Binomial - ignoring covariates. Expect if similar to Rasch "over-dispersion".

Observe: strong central peaks with surprising tails.

Compare with Binomial fit.

Focus on a cross classification.

		PEER					
		1	2	3	4	5	6
	1	13	3	6	5	2	17
S	2	4	55	31	19	5	2
E	3	5	66	495	173	31	12
L	4	7	31	398	1176	165	30
F	5	5	17	91	294	363	27
	6	17	7	21	46	34	105

Note.

- Strong diagonal.
- Movement in one direction – below diagonal.
- Tails heavy.

Plan: capture the qualitative characteristics in a simple parametric model.

Another example: CNN poll.

The communications decency act is constitutional.

0 Agree strongly

1 Agree

2 No opinion

3 Disagree

4 Disagree strongly

Value	0	1	2	3	4
Freq	860	117	39	352	2463
Pred	22.1	232.0	915.6	1605.5	1055.8

Look at the distribution of this one. Ordinal. Strong feelings. Extreme case of what we saw before.

Binomial fit – pretty crazy!

But aim for a parsimonious model.

A simple (very) framework.

Want to capture over **and** under–dispersion.

Poisson and Binomial are both discrete distributions.

Any discrete distribution can be embedded in the multinomial.

Embed in canonically parameterized multinomial.

Canonical parameter: $\eta_i = log\{p(i)/p(0)\}$.

Baseline logits.

Notation.

Observation: y. Probability that $y = i$: $p(i)$.

Poisson

$$p(i) = \exp\{-\exp(\beta_1)\}\exp\{-ln(i!) + i\beta_1\}.$$

Binomial

$$p(i) = \exp[-Kln\{1+\exp(\beta_1)\}]\exp[ln\{B(i,K)\}+i\beta_1]$$

where $B(i,K)$ denotes the Binomial coefficient and β_1 denotes the canonical parameter.

Multinomial representation

$$p(\mathbf{y}) = \sum_{i=1}^{K}\eta_i y_i - ln(1 + \sum_{i=1}^{K}\exp(\eta_i)).$$

Note "redundant parameter"
$\eta_0 = ln(p(0)/p(0)) = 0$ so that η's range from
$\eta_0, \cdots, \eta_K$.

Drop Binomial and Poisson into Multinomial structure

Poisson

$$\eta_i = ln(p(i)/p(0)) = -ln(i!) + i\,\beta_1$$

Binomial

$$\eta_i = ln(p(i)/p(0)) = ln\{B(i, K)\} + i\,\beta_1.$$

Convention for multinomial, say $y \sim Bin(\beta_1, 3)$, then

$$y = 0 \Longrightarrow \mathbf{y} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \quad y = 1 \Longrightarrow \mathbf{y} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$

$$y = 2 \Longrightarrow \mathbf{y} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \quad y = 3 \Longrightarrow \mathbf{y} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

The Poisson and Binomial are both linear models in the canonical parameters.

Special models. They have an offset term. Coefficient is 1.

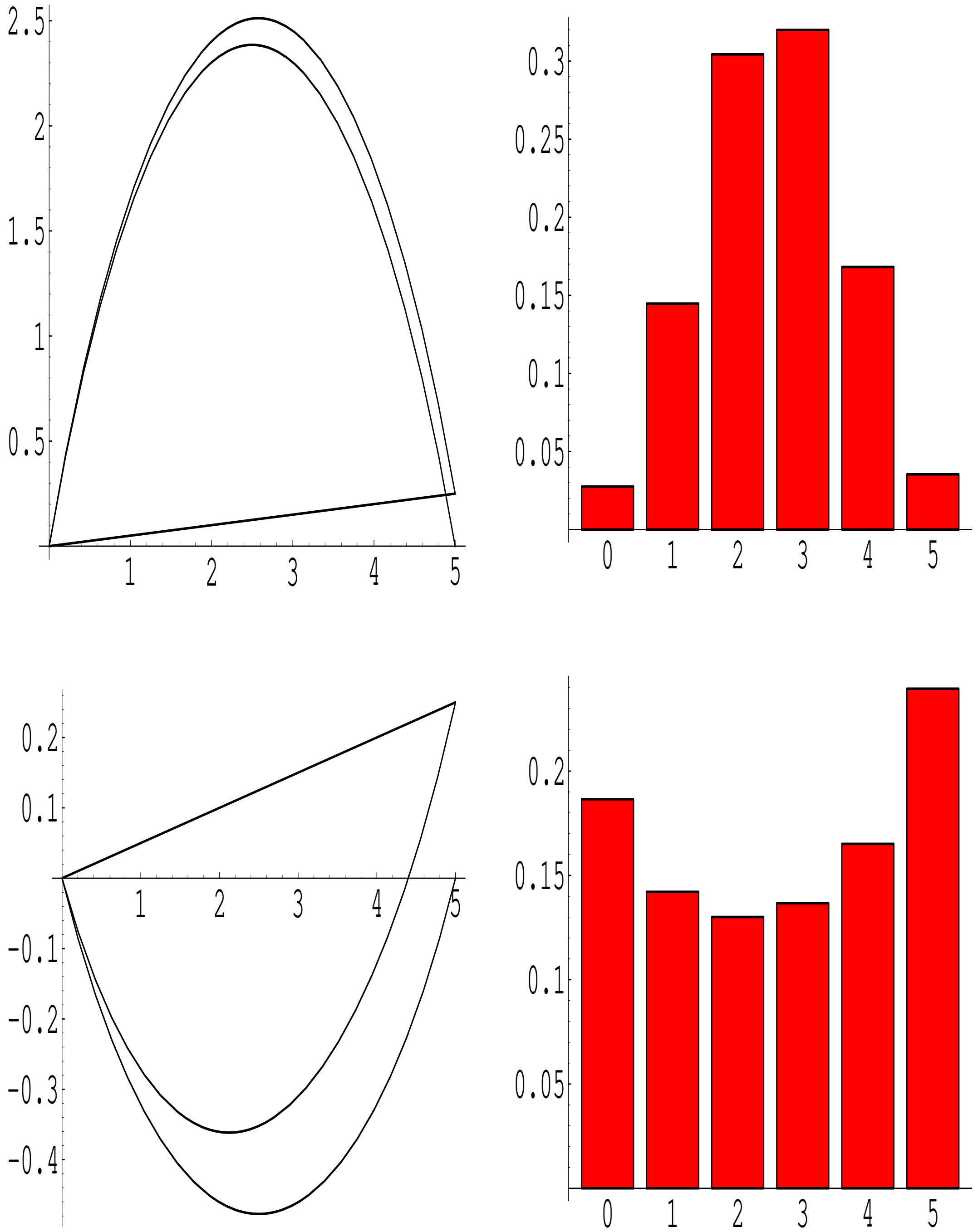
Design matrix looks like (for Binomial): $\eta = X\beta$ where

$$X = \begin{pmatrix} O(1) & 1 \\ O(2) & 2 \\ \vdots & \vdots \\ O(K) & K \end{pmatrix} \quad \text{and} \qquad \beta = (1, \beta_1)^T,$$

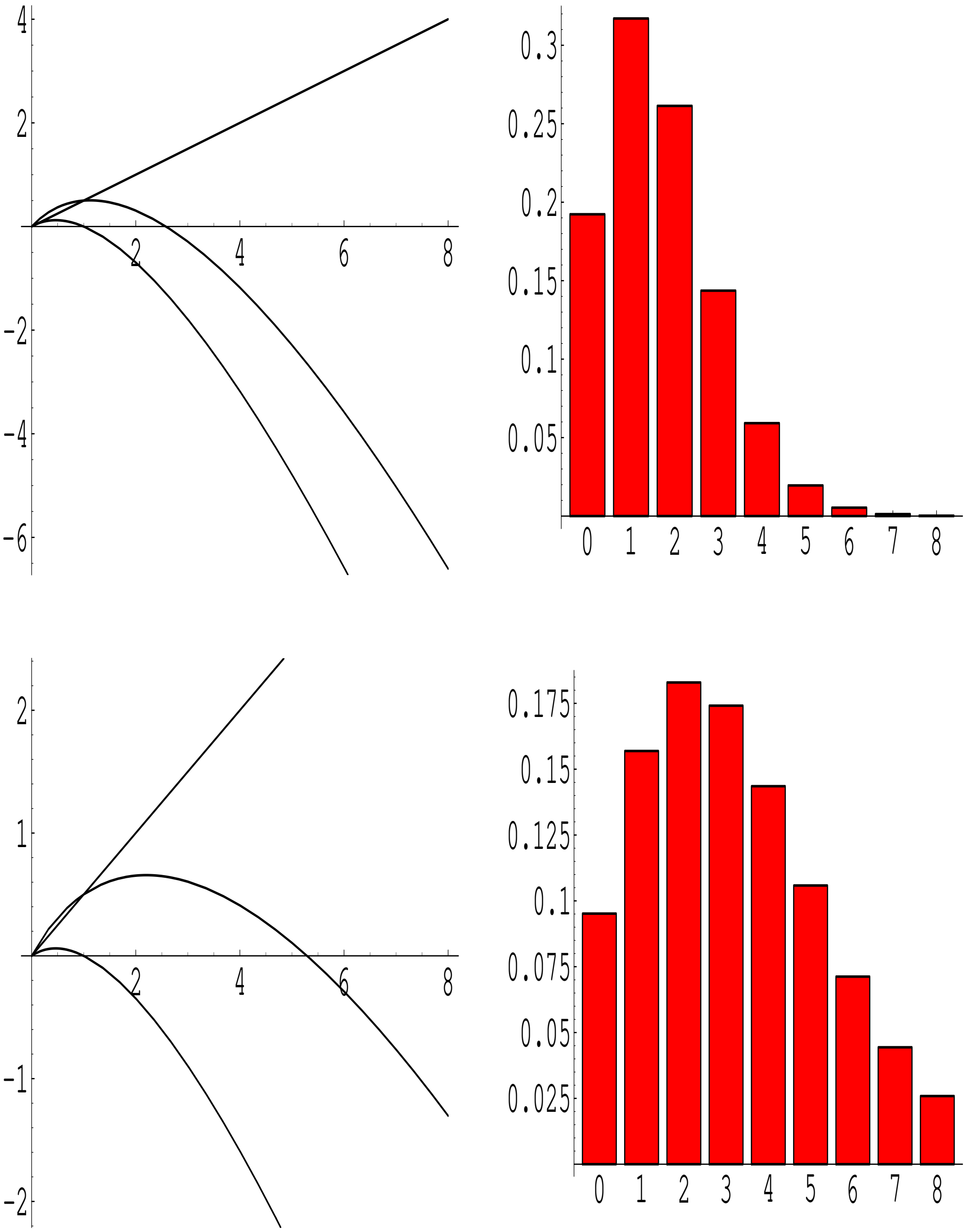
where $O(i)$ is the offset term.

Idea. Consider a parameterized offset. Two parameter exponential family model now. More freedom.

Binomial offset (K=5)



Poisson offset



Comments

Parameter ranges.

Binomial $-\infty < \beta_0 < \infty$, $-\infty < \beta_1 < \infty$.

Poisson $\beta_0 > 0$, $-\infty < \beta_1 < \infty$.

$\beta_0 = 0$, $\beta_1 < 0$.

Some special cases.

$\beta_0 = 1$ then Binomial or Poisson.

Binomial: $\beta_0 = \beta_1 = 0$. Discrete uniform.

$\beta_0 < 0$. “U-shaped” distributions.

Poisson: $\beta_0 = 0$. Geometric distribution
$\beta_1 = \ln(1 - p)$

Asymptotics

Mean variance expansions.

Notation.

$$\begin{pmatrix} \mu \\ \sigma^2 \end{pmatrix} \to \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}.$$

Approximately (for $\sigma^2 \sim \sigma_\mu^2$):

$$\begin{pmatrix} \mu \\ \sigma^2 \end{pmatrix} \to \frac{\sigma_\mu^2}{\sigma^2} \begin{pmatrix} 1 \\ \beta_1 \end{pmatrix}.$$

Dispersion perturbation maps to scale change in multinomial picture.

Example.

Take $\beta_0 = 1$ and $\beta_1 = 0.20067, K = 5, 15, 115.$
Scale change on logit by a factor of 1.01.

K	Mean	Var	New Mean	New Var	Mean ratio	Var ratio
5	2.75	1.2375	2.75062	1.22829	0.9997	1.0075
15	8.25	3.7125	8.25052	3.67830	0.9999	1.0093
115	63.25	28.4625	63.2505	28.1831	0.9999	1.0099

Quick fits

The Poisson-type count data.

$\beta_0 = 0.3192686$ $\beta_1 = -0.8861102$.

Value	0	1	2	3	4	5	6
Freq	107	44	13	6	1	0	0
Expected (Offs)	106.8	44.0	14.5	4.2	1.1	0.3	0.1
Expected (Pois)	99.8	53.7	14.5	2.6	0.3	0.0	0.0

Note: `log(44/107) = -0.8886392`.

The internet data:

$\beta_0 = -1.4782783$ $\beta_1 = 0.2742435$.

Value	0	1	2	3	4
Freq	860	117	39	352	2463
Pred(Offset)	836.4	141.7	102.4	245.3	2505.1
Pred(Binom)	22.1	232.0	915.6	1605.5	1055.8

```
crudefit(y=b2.pos,cats=7,family=poisson,start=c(1,0))
```

Regression set up

$\beta_j = X_j \psi$

Difference in β_1 for two groups with same β_0.

$$\begin{pmatrix} \beta_{0M} \\ \beta_{1M} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} \psi_0 \\ \psi_1 \\ \psi_2 \end{pmatrix}$$

$$\begin{pmatrix} \beta_{0F} \\ \beta_{1F} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & -1 \end{pmatrix} \begin{pmatrix} \psi_0 \\ \psi_1 \\ \psi_2 \end{pmatrix}.$$

Fitting the models: – easy one!

But may want to specify two design matrices. One for each parameter, or could share parameters.

Cusum scheme

or ... the impact of misspecified dispersion on ARL's.

Cusum characteristics.

$$S_t = max(0, S_{t-1} + Y_t - k).$$

- $S_0 = z$.
- Stop if $S_t > h$.
- Choose k by SQLRT.

Objective:

- If process is in control (H_0) then ARL_0 should be large.
- If process is out of control (H_a) then ARL_a should be small.

results2.tex 3/28/97 Simulation for
$\mu_0 = 1, \mu_a = 1.5$. h = 2,5 step 1. rel.disp = .6,1.6, step .1 reps = 100000 per estimate.

$h = 2$. Relative dispersion = 0.6 – 1.6 step 0.1.

Poisson		Poisson Offset	
ARL_0	ARL_a	ARL_0	ARL_a
16.593	6.006	30.073	6.556
16.588	6.006	24.512	6.383
16.601	6.007	20.829	6.221
16.601	6.014	18.424	6.101
16.586	5.996	16.586	5.996
16.590	6.006	15.245	5.920
16.594	6.003	14.238	5.855
16.602	6.007	13.503	5.828
16.600	6.005	12.889	5.793
16.603	6.020	12.411	5.782
16.599	6.004	12.077	5.754

Under-dispersion ARL_0 too small.

Over-dispersion ARL_a too large.

$h = 3$. Relative dispersion = 0.6 – 1.6 step 0.1.

ARL_0	ARL_a	ARL_0	ARL_a
30.993	8.674	72.435	9.690
30.985	8.669	53.598	9.410
31.008	8.677	42.736	9.124
30.982	8.674	36.104	8.857
30.999	8.683	30.999	8.683
30.984	8.665	27.600	8.471
30.983	8.680	24.991	8.333
30.995	8.681	22.969	8.249
30.986	8.670	21.622	8.130
30.993	8.678	20.424	8.036
30.999	8.677	19.361	8.003

$h = 4$. Relative dispersion = 0.6 – 1.6 step 0.1.

ARL_0	ARL_a	ARL_0	ARL_a
54.256	11.695	164.277	13.313
54.256	11.696	110.842	12.836
49.435	11.650	75.803	12.662
54.257	11.694	64.690	12.002
54.225	11.680	54.225	11.680
54.231	11.683	47.441	11.431
54.285	11.699	41.577	11.201
54.256	11.703	37.454	11.044
54.254	11.698	34.393	10.866
54.240	11.690	31.796	10.732
54.259	11.699	29.730	10.633

$h = 5$. Relative dispersion = 0.6 – 1.6 step 0.1.

ARL_0	ARL_a	ARL_0	ARL_a
94.041	15.065	338.193	16.922
94.002	15.052	211.696	16.357
94.023	15.058	153.114	15.972
94.011	15.054	115.510	15.491
94.018	15.071	94.018	15.071
94.052	15.079	76.256	14.722
94.037	15.057	67.083	14.366
94.022	15.068	59.206	14.075
94.041	15.068	53.288	13.853
94.036	15.067	48.502	13.710
93.975	15.051	44.458	13.564

Hopefully

Parameterized offsets will provide a good building block for the CUSUM schemes – allowing a smooth transition between under and over-dispersion.

Latent class models.

Rasch model.

Each person gets their own “intelligence” parameter β_i. Each question has it's own difficulty parameter ψ_j.

Conditionally on β_i subjects provide a set of independent Bernoulli observations with

$$logit(\pi_{ij}) = \beta_i + \psi_j.$$

Rasch extension:

Each person gets their own $\beta_\mathbf{i}$ parameter, β_{i0} and β_{i1}.

Each question pair has its own parameter ψ_j, the second question in each pair has a parameter τ.

Same conditional independence assumptions

$$y_{ijk} \sim BinOffset(\beta_{i0}, \beta_{i1} + \psi_j + \tau I_{k=2}).$$

Additivity on the linear parameter.

Comments

.

Give each person their own β_0 and β_1.

Parameter interpretations – **subject specific**. Want this here. Compare to population averaged interpretation.

- Mixing on the β_1 deals with the diagonal clustering.
- Parameter τ handles the above vs. below diagonal.
- Mixing on the β_0 could explain the **deviants**.

But is it crazy?

Someone with a very negative β_0 is actually out at all the corners – not just the deviant corners.

How do they actually respond?

Surprising (to me).

The deviants, those who made a (1,6)

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]
[1,]	36	7	6	1	5	115
[2,]	2	7	5	3	0	4
[3,]	5	10	6	13	8	5
[4,]	5	4	12	27	8	6
[5,]	8	3	14	11	8	8
[6,]	60	7	20	16	18	89

Those who made a (6,1)

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]
[1,]	76	18	10	8	6	49
[2,]	16	29	22	8	4	7
[3,]	31	35	63	26	13	13
[4,]	28	33	71	69	23	12
[5,]	24	27	54	54	33	23
[6,]	335	31	63	69	54	147

Dealing with the model.

Many nuisance parameters.

Conditional inference – good for the Rasch.

Possible here: YES because linear exp family (that was the point).

A special parameterization of multinomial model.

Polynomial approximations in sufficient statistics to conditional score functions.

Part of: projected scores for multinomial models.

Choice models in marketing research.

Simple conceptual method.

Computationally straightforward.

Nice interpretations.

Flexible family of distributions.