# Statistical Literature and Problems

STAT 501

29 June, 2020

## (Albert & Chib, 1993)

James H. Albert & Siddhartha Chib (1993) Bayesian Analysis of Binary and Polychotomous Response Data, Journal of the American Statistical Association, 88:422, 669-679, DOI: 10.1080/01621459.1993.10476321

### Data augmentation and Gibbs sampling

$$f(k) = \binom{n}{k} p^k (1-p)^{n-k} \tag{1}$$

$$\begin{aligned} g(X_n) &= g(\theta) + g'(\tilde{\theta})(X_n - \theta) \\ \sqrt{n}[g(X_n) - g(\theta)] &= g'\left(\tilde{\theta}\right)\sqrt{n}[X_n - \theta] \end{aligned}$$

$$\begin{aligned} p(\theta_{1:k}|x_{1:k}) &\propto p(x_{1:k}|\theta_{1:k})p(\theta_{1:k}) && (2) \\ &\propto \frac{\Gamma\left(\sum_{i=1}^k \alpha_i\right)}{\prod_{i=1}^k \Gamma(\alpha_i)} \prod_{i=1}^k \theta_i^{\alpha_i + n_k - 1} \\ &\propto \text{Dirichlet}(\vec{\alpha} + \vec{n}) && (3) \end{aligned}$$

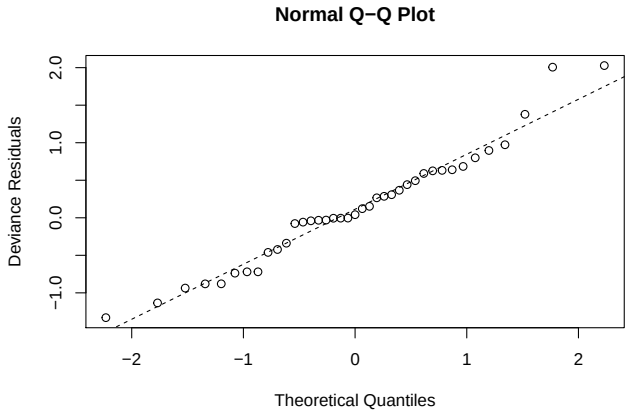
### 5.1 Finney Data

```
## $Vol
##  [1] 3.70 3.50 1.25 0.75 0.80 0.70 0.60 1.10 0.90 0.90 0.80 0.55 0.60 1.40 0.75 2.30 3.20 0.85 1.70 
##
## $Rate
##  [1] 0.825 1.090 2.500 1.500 3.200 3.500 0.750 1.700 0.750 0.450 0.570 2.750 3.000 2.330 3.750 1.640
##
## $Resp
##  [1] 1 1 1 1 1 1 0 0 0 0 0 0 0 1 1 1 1 1 0 1 0 0 0 0 1 0 1 0 1 0 1 0 0 1 1 1 0 0 1
##
## $plotFdat
## function(Rs=Resp,lV=lVol,lR=lRate,zc,zr,rob=FALSE,cont=FALSE){
## # Rs <- Resp; lV <- log(Vol); lR <- log(Rate)
## plot(lV,lR,type="n",xlab="log(Vol)",ylab="log(Rate)")
## points(lV[Rs==0],lR[Rs==0],pch=5, cex=1.2)
## points(lV[Rs==1],lR[Rs==1],pch=16,cex=1.2)
```

```
## if (cont) {
##  lV.new <- seq(from=min(lV), to=max(lV ),len=10)
##  lR.new <- seq(from=min(lR), to=max(lR), len=10)
##  lGrid  <- expand.grid(lVol=lV.new,lRate=lR.new)
##  TT.glm <- exp(coef(zc)[1]+coef(zc)[2]*lGrid[1]+coef(zc)[3]*lGrid[2]) ; TT.glm = TT.glm$lVol
##  PP.glm <- matrix( TT.glm/(1+TT.glm), nrow=10, ncol=10, byrow=F)
##  if (rob) {
##    TT.rob <- exp(coef(zr)[1]+coef(zr)[2]*lGrid[1]+coef(zr)[3]*lGrid[2]); TT.rob = TT.rob$lVol
##    PP.rob <- matrix(TT.rob/(1+TT.rob), nrow=10, byrow=F)
##  }
##  contour(lV.new, lR.new, PP.glm, levels=c(0.25,0.50),add=TRUE,col=2)
##  if (rob) contour(lV.new,lR.new,PP.rob,levels=c(0.25,0.50),add=TRUE,col=4)
## }
## invisible() }

##
## Call:
## glm(formula = Resp ~ lVol + lRate, family = binomial)
##
## Deviance Residuals:
##       Min         1Q     Median         3Q        Max
## -1.448059  -0.604886   0.098552   0.610092   2.290514
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -2.9239     1.2878 -2.2704 0.023182 *
## lVol          5.2205     1.8581  2.8096 0.004960 **
## lRate         4.6312     1.7891  2.5886 0.009636 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 54.0398  on 38  degrees of freedom
## Residual deviance: 29.2640  on 36  degrees of freedom
## AIC: 35.264
##
## Number of Fisher Scoring iterations: 6

##             (Intercept)       lVol      lRate
## (Intercept)   1.6584256 -1.9356315 -2.1393640
## lVol         -1.9356315  3.4523951  2.6753814
## lRate        -2.1393640  2.6753814  3.2007547

## [1] 3

##      NULL
## 10.005871

##             1             2             3             4             5             6             7
## 0.04455271688 0.01740458143 0.07102437364 0.06730379485 0.16848730078 0.19515806153 0.00098307137 0.1

##
```

```
## Call:
## glm(formula = Resp ~ lVol + lRate, family = binomial, method = "cubinf",
##     ufact = 3.2)
##
## Deviance Residuals:
##       Min        1Q    Median        3Q       Max
## -1.330434 -0.380446  0.041256  0.607839  2.027545
##
## Coefficients:
##               Value Std. Error t value
## (Intercept) -6.3360     2.7881 -2.2725
## lVol         9.8529     4.2632  2.3111
## lRate        8.7606     3.7015  2.3668
##
## (Dispersion Parameter for binomial family taken to be 1 )
##
##     Null Deviance: 54.03984 on 38 degrees of freedom
##
## Residual Deviance: 22.88263 on 36 degrees of freedom
##
## Number of Iterations: 21
##
## Correlation of Coefficients:
##       (Intercept) lVol
## lVol  -0.94748
## lRate -0.97456     0.93620


##
## Calls:
## Name
## z.glm    glm(formula = Resp ~ lVol + lRate, family = binomial)
## z.cub    glm(formula = Resp ~ lVol + lRate, family = binomial, method = "cubinf",
##     ufact = 3.2)
##
##
## Residual Statistics
##            Min        1Q      Median      3Q     Max
## z.glm -0.64952 -0.167682 0.004844458 0.17004 0.92743
## z.cub -0.58730 -0.021774 0.000096875 0.10063 0.99638
##
##
## Number of Parameter in each Model
##       Nobs Resid df Model Parameters Est. Parameters
## z.glm   39       36                3               3
## z.cub   39       36                3               3
##
##
## Coefficients:
##                   Estimate Std. Error t value
## (Intercept):z.glm -2.92385  1.26661   -2.30840
## (Intercept):z.cub -6.33596  2.78810   -2.27250
## lVol:z.glm         5.22049  1.82750    2.85664
## lVol:z.cub         9.85293  4.26324    2.31114
## lRate:z.glm        4.63123  1.75964    2.63193
```

```
## lRate:z.cub        8.76056  3.70151    2.36675
##
##
## Residual Scale Estimates:
## z.glm : 10.006
## z.cub : 1
##
## Correlation of Coefficients:
##
## Model =  z.glm
##       (Intercept) lVol
## lVol  -0.80894
## lRate -0.92856     0.80482
##
## Model =  z.cub
##       (Intercept) lVol
## lVol  -0.94748
## lRate -0.97456     0.93620
```




```
##                   1                   2                   3                   4                   5
## 0.00589234388614778 0.94452236088465491 0.07732888915470906 0.99997031333451014 0.00360061134648365 0
```

### 5.2 Election Data

### 5.3 A Trivariate Probit Example

## (Polson etal, 2013)

Nicholas G. Polson, James G. Scott & Jesse Windle (2013) Bayesian Inference for Logistic Models Using Pólya–Gamma Latent Variables, Journal of the American Statistical Association, 108:504, 1339-1349, DOI: 10.1080/01621459.2013.829001