Statistical Rethinking

2023

10. Counts & Hidden Confounds

Generalized Linear Models

Linear Models: Expected value is additive ("linear") combination of parameters

$$Y_i \sim \text{Normal}(\mu_i, \sigma)$$
$$\mu_i = \alpha + \beta_X X_i + \beta_Z Z_i$$

Generalized Linear Models: Expected value is some function of an additive combination of parameters

$$Y_i \sim \text{Bernoulli}(p_i)$$
$$f(p_i) = \alpha + \beta_X X_i + \beta_Z Z_i$$

Generalized Linear Models:
Expected value is some function of an additive combination of parameters

Uniform changes in predictor not uniform changes in prediction

All predictor variables interact, moderate one another

Influences predictions & uncertainty of predictions

$$\text{logit}(p_i) = \alpha + \beta x_i$$

Confounded Admissions

Confounded Admissions

```r
set.seed(12)
N <- 2000 # number of applicants
# even gender distribution
G <- sample( 1:2 , size=N , replace=TRUE )
# sample ability, high (1) to average (0)
u <- rbern(N,0.1)
# gender 1 tends to apply to department 1, 2 to 2
# and G=1 with greater ability tend to apply to 2 as well
D <- rbern( N , ifelse( G==1 , u*1 , 0.75 ) ) + 1
# matrix of acceptance rates [dept,gender]
p_u0 <- matrix( c(0.1,0.1,0.1,0.3) , nrow=2 )
p_u1 <- matrix( c(0.3,0.3,0.5,0.5) , nrow=2 )
p_u <- list( p_u0 , p_u1 )
# simulate acceptance
p <- sapply( 1:N , function(i) p_u[[1+u[i]]][D[i],G[i]] )
A <- rbern( N , p )
```

```r
set.seed(12)
N <- 2000 # number of applicants
# even gender distribution
G <- sample( 1:2 , size=N , replace=TRUE )
# sample ability, high (1) to average (0)
u <- rbern(N,0.1)
# gender 1 tends to apply to department 1, 2 to 2
# and G=1 with greater ability tend to apply to 2 as well
D <- rbern( N , ifelse( G==1 , u*1 , 0.75 ) ) + 1
# matrix of acceptance rates [dept,gender]
p_u0 <- matrix( c(0.1,0.1,0.1,0.3) , nrow=2 )
p_u1 <- matrix( c(0.3,0.3,0.5,0.5) , nrow=2 )
p_u <- list( p_u0 , p_u1 )
# simulate acceptance
p <- sapply( 1:N , function(i) p_u[[1+u[i]]][D[i],G[i]] )
A <- rbern( N , p )
```

```r
set.seed(12)
N <- 2000 # number of applicants
# even gender distribution
G <- sample( 1:2 , size=N , replace=TRUE )
# sample ability, high (1) to average (0)
u <- rbern(N,0.1)
# gender 1 tends to apply to department 1, 2 to 2
# and G=1 with greater ability tend to apply to 2 as well
D <- rbern( N , ifelse( G==1 , u*1 , 0.75 ) ) + 1
# matrix of acceptance rates [dept,gender]
p_u0 <- matrix( c(0.1,0.1,0.1,0.3) , nrow=2 )
p_u1 <- matrix( c(0.3,0.3,0.5,0.5) , nrow=2 )
p_u <- list( p_u0 , p_u1 )
# simulate acceptance
p <- sapply( 1:N , function(i) p_u[[1+u[i]]][D[i],G[i]] )
A <- rbern( N , p )
```

```r
set.seed(12)
N <- 2000 # number of applicants
# even gender distribution
G <- sample( 1:2 , size=N , replace=TRUE )
# sample ability, high (1) to average (0)
u <- rbern(N,0.1)
# gender 1 tends to apply to department 1, 2 to 2
# and G=1 with greater ability tend to apply to 2 as well
D <- rbern( N , ifelse( G==1 , u*1 , 0.75 ) ) + 1
# matrix of acceptance rates [dept,gender]
p_u0 <- matrix( c(0.1,0.1,0.1,0.3) , nrow=2 )
p_u1 <- matrix( c(0.3,0.3,0.5,0.5) , nrow=2 )
p_u <- list( p_u0 , p_u1 )
# simulate acceptance
p <- sapply( 1:N , function(i) p_u[[1+u[i]]][D[i],G[i]] )
A <- rbern( N , p )
```

```
> table(G,D,u)
, , u = 0

   D
G     1   2
  1 908   0
  2 220 645

, , u = 1

   D
G     1   2
  1   0 114
  2  28  85
```

```r
set.seed(12)
N <- 2000 # number of applicants
# even gender distribution
G <- sample( 1:2 , size=N , replace=TRUE )
# sample ability, high (1) to average (0)
u <- rbern(N,0.1)
# gender 1 tends to apply to department 1, 2 to 2
# and G=1 with greater ability tend to apply to 2 as well
D <- rbern( N , ifelse( G==1 , u*1 , 0.75 ) ) + 1
# matrix of acceptance rates [dept,gender]
p_u0 <- matrix( c(0.1,0.1,0.1,0.3) , nrow=2 )
p_u1 <- matrix( c(0.3,0.3,0.5,0.5) , nrow=2 )
p_u <- list( p_u0 , p_u1 )
# simulate acceptance
p <- sapply( 1:N , function(i) p_u[[1+u[i]]][D[i],G[i]] )
A <- rbern( N , p )
```

```
> p_u0
     [,1] [,2]
[1,]  0.1  0.1
[2,]  0.1  0.3
```

```
> p_u1
     [,1] [,2]
[1,]  0.3  0.5
[2,]  0.3  0.5
```

Department 2 discriminates against gender 1

```r
dat_sim <- list( A=A , D=D , G=G )

# total effect gender
m1 <- ulam(
    alist(
        A ~ bernoulli(p),
        logit(p) <- a[G],
        a[G] ~ normal(0,1)
    ), data=dat_sim , chains=4 , cores=4 )

# direct effects - now confounded!
m2 <- ulam(
    alist(
        A ~ bernoulli(p),
        logit(p) <- a[G,D],
        matrix[G,D]:a ~ normal(0,1)
    ), data=dat_sim , chains=4 , cores=4 )
```

total effect shows disadvantage

```
> precis(m1,depth=3)
      mean   sd  5.5% 94.5% n_eff Rhat4
a[1] -1.82 0.09 -1.96 -1.67  1483     1
a[2] -0.89 0.07 -1.00 -0.78  1562     1
```

direct effect confounded

```
> precis(m2,depth=3)
        mean   sd  5.5% 94.5% n_eff Rhat4
a[1,1] -2.05 0.11 -2.23 -1.89  2482     1
a[1,2] -0.66 0.18 -0.96 -0.38  2369     1
a[2,1] -1.77 0.19 -2.06 -1.47  2278     1
a[2,2] -0.68 0.08 -0.81 -0.56  2210     1
```

```
post2 <- extract.samples(m2)
dens(inv_logit(post2$a[,1,1]),lwd=3,col=4,xlim=c(0,0.5),xlab="probability admission")
dens(inv_logit(post2$a[,2,1]),lwd=3,col=4,lty=2,add=TRUE)
dens(inv_logit(post2$a[,1,2]),lwd=3,col=2,add=TRUE)
dens(inv_logit(post2$a[,2,2]),lwd=3,col=2,lty=2,add=TRUE)
```

total effect shows disadvantage

```
> precis(m1,depth=3)
      mean   sd  5.5% 94.5% n_eff Rhat4
a[1] -1.82 0.09 -1.96 -1.67  1483     1
a[2] -0.89 0.07 -1.00 -0.78  1562     1
```

direct effect confounded

You guessed it: Collider bias

Stratifying by D opens non-causal path through u

Can estimate total causal effect of G, but isn't what we want

Cannot estimate direct effect of D or G

Worst place in the world, Collider Bias country

You guessed it: Collider bias

More intuitive explanation:

High ability G1s apply to discriminatory department anyway

G1s in that department are higher ability on average than G2s

High ability compensates for discrimination => masks evidence

MAX-PLANCK-GESELLSCHAFT

Offices & Facilities > Officers of the Max Planck Society > Ce

Gender Gaps at the Academies

David Card, Stefano DellaVigna, Patricia Funk & Nagore Iriberri

WORKING PAPER 30510 **DOI** 10.3386/w30510 **ISSUE DATE** September 2022

Historically, a large majority of the newly elected members of the National Academy of Science (NAS) and the American Academy of Arts and Science (AAAS) were men. Within the past two decades, however, that situation has changed, and in the last 3 years women made up about 40 percent of the new members in both academies. We build lists of active scholars from publications in the top journals in three fields – Psychology, Mathematics and Economics – and develop a series of models to compare changes in the probability of selection of women as members of the NAS and AAAS from the 1960s to today, controlling for publications and citations. In the early years of our sample, women were less likely to be selected as members than men with similar records. By the 1990s, the selection process at both academies was approximately gender-neutral, conditional on publications and citations. In the past 20 years, however, a positive preference for female members has emerged and strengthened in all three fields. Currently, women are 3-15 times more likely to be selected as members of the AAAS and NAS than men with similar publication and citation records.

uniformly high level of Gender Equality work in the Institutes while preserving the professional and cultural

PNAS BRIEF REPORT | SOCIAL SCIENCES COMPUTER SCIENCES

Gendered citation patterns among the scientific elite

Kristina Lerman[a,1], Yulin Yu[b], Fred Morstatter[a], and Jay Pujara[a]

Edited by Susan Fiske, Princeton University, Princeton, NJ; received April 8, 2022; accepted August 22, 2022

Diversity in science is necessary to improve innovation and increase the capacity of the scientific workforce. Despite decades-long efforts to increase gender diversity, however, women remain a small minority in many fields, especially in senior positions. The dearth of elite women scientists, in turn, leaves fewer women to serve as mentors and role models for young women scientists. To shed light on gender disparities in science, we study prominent scholars who were elected to the National Academy of Sciences. We construct author citation networks that capture the structure of recognition among scholars' peers. We identify gender disparities in the patterns of peer citations and show that these differences are strong enough to accurately predict the scholar's gender. In contrast, we do not observe disparities due to prestige, with few significant differences in the structure of citations of scholars affiliated with high-ranked and low-ranked institutions. These results provide further evidence that a scholar's gender plays a role in the mechanisms of success in science.

gender | bibliometrics | science of science | gender disparities

Gender disparities persist in many fields of science. Despite long-running efforts to increase women's representation in the scientific workforce, they continue to face barriers to advancement. Women are less likely than their male peers to be mentored by eminent faculty (1) and to be hired and promoted (2, 3). Women publish in less prestigious journals (4), have fewer collaborators (5), and are underrepresented among journal reviewers and editors (6), and their papers receive fewer citations (7, 8). The multifaceted gender disparities create a "glass ceiling," an invisible barrier that fundamentally limits professional recognition for even the best women scientists (9). As a result, the share of women in higher academic positions decreases steadily (3), with relatively few becoming full professors or receiving prestigious awards. For example, among physics faculty in 4-y colleges and universities, women represent 23% of assistant professors and 18% of associate professors

Causal foundations of bias, disparity and fairness

V.A. Traag [a)] and L. Waltman
Centre for Science and Technology Studies (CWTS), Leiden University, the Netherlands

(Dated: July 27, 2022)

The study of biases, such as gender or racial biases, is an important topic in the social and behavioural sciences. However, the concept of bias is not always clearly defined in the literature. Definitions of bias are often ambiguous, or definitions are not provided at all. To study biases in a precise way, it is important to have a well-defined concept of bias. We propose to define bias as a direct causal effect that is unjustified. We propose to define the closely related concept of disparity as a direct or indirect causal effect that includes a bias. Our proposed definitions can be used to study biases and disparities in a more rigorous and systematic way. We compare our definitions of bias and disparity with various definitions of fairness introduced in the artificial intelligence literature. We also illustrate our definitions in two case studies, focusing on gender bias in science and racial bias in police shootings. Our proposed definitions aim to contribute to a better appreciation of the causal intricacies of studies of biases and disparities. This will hopefully also lead to an improved understanding of the policy implications of such studies.

Keywords: bias; disparity; fairness; causal model; gender bias; racial bias

https://arxiv.org/abs/2207.13665

Citation networks

Citation networks of members of NAS

Women associated with lower lifetime citation rate

Lerman et al 2022 Gendered citation patterns among the scientific elite

Membership

Elections to NAS

Women associated with 3–15 times higher election rate, controlling for citations

Card et al 2022 Gender Gaps at the Academies

citations
quality (hidden)
q
C
G
M
gender
NAS member

Restrict sample to NAS members, examine citations

If men less likely to be elected, then must have higher q,C to compensate

Control for citations, examine elections to NAS

G is "treatment". C is a post-treatment variable! If women less likely to be cited (bias), then women more likely to be elected because they have higher q than indicated by C

No causes in; No causes out

Hyped papers with vague estimands,
unwise adjustment sets

Policy design through collider bias?

We can do better

Strong assumptions required

Qualitative data useful

Sensitivity analysis

What are the implications of what we don't know?

Assume confound exists, model its consequences for different strengths/kinds of influence

How strong must the confound be to change conclusions?

Sensitivity analysis

$$A_i \sim \text{Bernoulli}(p_i)$$

$$\text{logit}(p_i) = \alpha[G_i, D_i] + \beta_{G[i]} u_i$$

Sensitivity analysis

$$A_i \sim \text{Bernoulli}(p_i)$$

$$\text{logit}(p_i) = \alpha[G_i, D_i] + \beta_{G[i]} u_i$$

$$(D_i = 2) \sim \text{Bernoulli}(q_i)$$

$$\text{logit}(q_i) = \delta[G_i] + \gamma_{G[i]} u_i$$

```r
datl <- dat_sim
datl$D2 <- ifelse(datl$D==2,1,0)
datl$N <- length(datl$D)
datl$b <- c(1,1)
datl$g <- c(1,0)

mGDu <- ulam(
    alist(
        # A model
        A ~ bernoulli(p),
        logit(p) <- a[G,D] + b[G]*u[i],
        matrix[G,D]:a ~ normal(0,1),

        # D model
        D2 ~ bernoulli(q),
        logit(q) <- delta[G] + g[G]*u[i],
        delta[G] ~ normal(0,1),

        # declare unobserved u
        vector[N]:u ~ normal(0,1)
    ), data=datl , chains=4 , cores=4 )
```

$$A_i \sim \text{Bernoulli}(p_i)$$

$$\text{logit}(p_i) = \alpha[G_i, D_i] + \beta_{G[i]} u_i$$

$$(D_i = 2) \sim \text{Bernoulli}(q_i)$$

$$\text{logit}(q_i) = \delta[G_i] + \gamma_{G[i]} u_i$$

$$u_j \sim \text{Normal}(0,1)$$

Ignore confound

Assume confound

Sensitivity analysis

What are the implications of what we don't know?

Somewhere between pure simulation and pure analysis

Vary confound strength over range and show how results change –or– vary other effects and estimate confound strength

Confounds persist — don't pretend

More parameters (2006) than observations (2000)!

$$A_i \sim \text{Bernoulli}(p_i)$$

$$\text{logit}(p_i) = \alpha[G_i, D_i] + \beta_{G[i]} u_i$$

$$(D_i = 2) \sim \text{Bernoulli}(q_i)$$

$$\text{logit}(q_i) = \delta[G_i] + \gamma_{G[i]} u_i$$

$$u_j \sim \text{Normal}(0,1)$$

Charlie Chaplin, Modern Times (1936)

PAUSE

Hawaii
Yap
Chuuk
Manus Province
Trobriand Islands
Santa Cruz Islands
Tikopia
Malekula
Fiji
Tonga

Oceanic Technology

How is technological complexity related to population size?

To social structure?

data(Kline)

Influence of **population size** and **contact** on **total tools**

Kline & Boyd 2010 Population size predicts technological complexity in Oceania

Oceanic Technology

data(Kline)

	culture	population	contact	total_tools	mean_TU
1	Malekula	1100	low	13	3.2
2	Tikopia	1500	low	22	4.7
3	Santa Cruz	3600	low	24	4.0
4	Yap	4791	high	43	5.0
5	Lau Fiji	7400	high	33	5.0
6	Trobriand	8000	high	19	4.0
7	Chuuk	9200	high	40	3.8
8	Manus	13000	low	28	6.6
9	Tonga	17500	high	55	5.4
10	Hawaii	275000	low	71	6.6

Kline & Boyd 2010 Population size predicts technological complexity in Oceania

Population
Innovations
Tools
Loss
Contact

contact
location
C
L
P
T
population
tools

Adjustment set for P: L

Also want to stratify by C, to study interaction

Modeling tools

Tool count is not binomial: No maximum

Poisson distribution: Very high maximum and very low probability of each success

Here: Many many possible technologies, very few realized in any one place

Poisson link is log

Poisson distribution takes shape from expected value

Must be positive

Exponential scaling can be surprising!

$$Y_i \sim \text{Poisson}(\lambda_i)$$

$$\log(\lambda_i) = \alpha + \beta x_i$$

$$\lambda_i = \exp(\alpha + \beta x_i)$$

Poisson (poison) priors

Exponential scaling can be surprising

$$\log(\lambda_i) = \alpha$$

$$\alpha \sim \text{Normal}(0,10)$$

Poisson (poison) priors

Exponential scaling can be surprising

$$\log(\lambda_i) = \alpha$$

$$\alpha \sim \text{Normal}(0,10)$$

Poisson (poison) priors

Exponential scaling can be surprising

$$\log(\lambda_i) = \alpha$$

$$\alpha \sim \text{Normal}(0,10)$$

$$\alpha \sim \text{Normal}(3,0.5)$$

Poisson priors

$$\log(\lambda_i) = \alpha + \beta x_i$$

$\alpha \sim \text{Normal}(0,1)$
$\beta \sim \text{Normal}(0,10)$

$\alpha \sim \text{Normal}(3,0.5)$
$\beta \sim \text{Normal}(0,0.2)$

```r
data(Kline)
d <- Kline
d$P <- scale( log(d$population) )
d$contact_id <- ifelse( d$contact=="high" , 2 , 1 )
dat <- list(
    T = d$total_tools ,
    P = d$P ,
    C = d$contact_id )

# intercept only
m11.9 <- ulam(
    alist(
        T ~ dpois( lambda ),
        log(lambda) <- a,
        a ~ dnorm( 3 , 0.5 )
    ), data=dat , chains=4 , log_lik=TRUE )
# interaction model
m11.10 <- ulam(
    alist(
        T ~ dpois( lambda ),
        log(lambda) <- a[C] + b[C]*P,
        a[C] ~ dnorm( 3 , 0.5 ),
        b[C] ~ dnorm( 0 , 0.2 )
    ), data=dat , chains=4 , log_lik=TRUE )

compare( m11.9 , m11.10 , func=PSIS )
```

$$Y_i \sim \text{Poisson}(\lambda_i)$$

$$\log(\lambda_i) = \alpha_{C[i]} + \beta_{C[i]} \log(P_i)$$

$$\alpha_j \sim \text{Normal}(3,0.5)$$

$$\beta_j \sim \text{Normal}(0,0.2)$$

```r
data(Kline)
d <- Kline
d$P <- scale( log(d$population) )
d$contact_id <- ifelse( d$contact=="high" , 2 , 1 )
dat <- list(
    T = d$total_tools ,
    P = d$P ,
    C = d$contact_id )

# intercept only
m11.9 <- ulam(
    alist(
        T ~ dpois( lambda ),
        log(lambda) <- a,
        a ~ dnorm( 3 , 0.5 )
    ), data=dat , chains=4 , log_lik=TRUE )
# interaction model
m11.
```

```
> compare( m11.9 , m11.10 , func=PSIS )
Some Pareto k values are high (>0.5). Set pointwise=TRUE to inspect individual points.
Some Pareto k values are high (>0.5). Set pointwise=TRUE to inspect individual points.
        PSIS    SE dPSIS   dSE pPSIS weight
m11.10  85.9 13.50   0.0    NA   7.3      1
m11.9  141.3 33.69  55.4 33.13   8.0      0
```

```r
    ), data=dat , chains=4 , log_lik=TRUE )

compare( m11.9 , m11.10 , func=PSIS )
```

$$Y_i \sim \text{Poisson}(\lambda_i)$$

$$\log(\lambda_i) = \alpha_{C[i]} + \beta_{C[i]} \log(P_i)$$

$$\alpha_j \sim \text{Normal}(3,0.5)$$

$$\beta_j \sim \text{Normal}(0,0.2)$$

```
> compare( m11.9 , m11.10 , func=PSIS )
Some Pareto k values are high (>0.5). Set pointwise=TRUE to inspect individual points.
Some Pareto k values are high (>0.5). Set pointwise=TRUE to inspect individual points.
        PSIS    SE dPSIS   dSE pPSIS weight
m11.10  85.9 13.50   0.0    NA   7.3      1
m11.9  141.3 33.69  55.4 33.13   8.0      0
```

pPSIS is the “penalty”, the effective number of parameters.

How can a model with more parameters (m11.10) have fewer effective parameters?

m11.9 changes more when individual societies are dropped

```r
k <- PSIS( m11.10 , pointwise=TRUE )$k
plot( dat$P , dat$T , xlab="log population (std)" ,
ylab="total tools" ,
    col=ifelse( dat$C==1 , 4 , 2 ) , lwd=4+4*normalize(k) ,
    ylim=c(0,75) , cex=1+normalize(k) )

# set up the horizontal axis values to compute predictions
at
P_seq <- seq( from=-1.4 , to=3 , len=100 )

# predictions for C=1 (low contact)
lambda <- link( m11.10 , data=data.frame( P=P_seq , C=1 ) )
lmu <- apply( lambda , 2 , mean )
lci <- apply( lambda , 2 , PI )
lines( P_seq , lmu , lty=2 , lwd=1.5 )
shade( lci , P_seq , xpd=TRUE , col=col.alpha(4,0.3) )

# predictions for C=2 (high contact)
lambda <- link( m11.10 , data=data.frame( P=P_seq , C=2 ) )
lmu <- apply( lambda , 2 , mean )
lci <- apply( lambda , 2 , PI )
lines( P_seq , lmu , lty=1 , lwd=1.5 )
shade( lci , P_seq , xpd=TRUE , col=col.alpha(2,0.3))
```

Points scaled by leverage

log scale
Natural scale
Hawaii
Tonga
Yap
Trobriand
total tools
0
20
40
60
-1
0
1
2
log population (std)
Hawaii
Tonga
total tools
0
20
40
60
0
50000
150000
250000
population

Hawaii
Tonga
total tools
0
20
40
60
0
50000
150000
250000
population

Tonga
total tools
0
20
40
60
0
50000
150000
250000

This model is wack

Two immediate ways to improve the model

(1) Use a robust model: gamma-Poisson (neg-binomial)

(2) A principled scientific model

Population
Innovations
Tools
Loss
Contact

$$\Delta T = \alpha P^{\beta} - \gamma T$$

change in tools

$$\Delta T = \alpha P^{\beta} - \gamma T$$

change in tools

innovation rate

$$\Delta T = \alpha P^{\beta} - \gamma T$$

diminishing returns (elasticity)

innovation rate

change in tools

$$\Delta T = \alpha P^{\beta} - \gamma T$$

diminishing returns (elasticity)

innovation rate

change in tools

$$\Delta T = \alpha P^{\beta} - \gamma T$$

rate of loss

diminishing returns depend upon contact

$$\Delta T = \alpha_C P^{\beta_C} - \gamma T$$

innovation depends upon contact

$$\Delta T = \alpha_C P^{\beta_C} - \gamma T$$

Difference equation: Says how tools change, not expected number

What is the equilibrium?

$$\Delta T = \alpha_C P^{\beta_C} - \gamma T$$

```r
f <- function(a=0.02,b=0.5,g=0.2,P=1e4,t_max=50) {
    T <- rep(0,t_max)
    for ( i in 2:t_max )
        T[i] <- T[i-1] + a*P^b - g*T[i-1]
    return(T)
}

plot( NULL , xlim=c(0,50) , ylim=c(0,10) ,
xlab="time" , ylab="Tools" )

T <- f(P=1e3)
lines( 1:50 , T , lwd=3 , col=2 )

T <- f(P=1e4)
lines( 1:50 , T , lwd=3 , col=2 )
```

$$\Delta T = \alpha_C P^{\beta_C} - \gamma T = 0$$

Solve for equilibrium T

$$\Delta T = \alpha_C P^{\beta_C} - \gamma T = 0$$

Solve for equilibrium T

$$\hat{T} = \frac{\alpha_C P^{\beta_C}}{\gamma}$$

$$\hat{T} = \frac{\alpha_C P^{\beta_C}}{\gamma}$$

$$T_i \sim \text{Poisson}(\lambda_i)$$

$$\lambda_i = \hat{T}$$

$$\hat{T} = \frac{\alpha_C P^{\beta_C}}{\gamma}$$

```
# innovation/loss model

dat2 <- list( T=d$total_tools, P=d$population,
C=d$contact_id )

m11.11 <- ulam(
    alist(
        T ~ dpois( lambda ),
        lambda <- exp(a[C])*P^b[C]/g,
        a[C] ~ dnorm(1,1),
        b[C] ~ dexp(1),
        g ~ dexp(1)
    ), data=dat2 , chains=4 , cores=4 )
```

All parameters must be positive

Two ways to do this

(1) use exp()

(2) use appropriate prior

$$\hat{T} = \frac{\alpha_C P^{\beta_C}}{\gamma}$$

```r
# innovation/loss model

dat2 <- list( T=d$total_tools, P=d$population,
C=d$contact_id )

m11.11 <- ulam(
    alist(
        T ~ dpois( lambda )
        lambda <- exp(a[C])
        a[C] ~ dnorm(1,1),
        b[C] ~ dexp(1),
        g ~ dexp(1)
    ), data=dat2 , chains=4
```

```
> precis(m11.11,2)
      mean   sd  5.5% 94.5% n_eff Rhat4
a[1] 0.85 0.68 -0.26  1.90   698     1
a[2] 0.93 0.83 -0.39  2.31   902     1
b[1] 0.26 0.03  0.21  0.32  1149     1
b[2] 0.29 0.10  0.12  0.45   711     1
g    1.11 0.70  0.32  2.43   862     1
```

All parameters must be positive

Two ways to do this

(1) use exp()

propriate prior

$$\hat{T} = \frac{\alpha_C P^{\beta_C}}{\gamma}$$

```r
# innovation/loss model

dat2 <- list( T=d$total_tools, P=d$population,
C=d$contact_id )

m11.11 <- ulam(
    alist(
        T ~ dpois( lambda ),
        lambda <- exp(a[C])*P^b[C]/g,
        a[C] ~ dnorm(1,1),
        b[C] ~ dexp(1),
        g ~ dexp(1)
    ), data=dat2 , chains=4 , cores=4 )
```

Innovation/loss model

Generalized linear model

Still have to deal with location as confound

Count GLMs

Distributions from constraints

Maximum entropy priors: Binomial, Poisson, and extensions

More event types: Multinomial and categorical

Robust regressions:
Beta-binomial, gamma-Poisson

Course Schedule

Week 1	Bayesian inference	Chapters 1, 2, 3
Week 2	Linear models & Causal Inference	Chapter 4
Week 3	Causes, Confounds & Colliders	Chapters 5 & 6
Week 4	Overfitting / MCMC	Chapters 7, 8, 9
Week 5	Generalized Linear Models	Chapters 10, 11
Week 6	Mixtures & ordered categories	Chapters 11 & 12
Week 7	Multilevel models I	Chapter 13
Week 8	Multilevel models II	Chapter 14
Week 9	Measurement & Missingness	Chapter 15
Week 10	Generalized Linear Madness	Chapter 16

https://github.com/rmcelreath/stat_rethinking_2023

BONUS

Simpson's Pandora's Box

Simpson's paradox: Reversal of an association when groups are combined or separated

No way to know which association (combined/separated) is correct without causal assumptions

Infinite evils unleashed

Pandora opens a spreadsheet

Simpson 1951. The Interpretation of Interaction in Contingency Tables

Berkeley Paradox

Unconditional on department:
Women admitted at lower rate

Conditional on department:
Women admitted slightly more

Which is correct? No way to know without assumptions

	dept	admit	reject	applications	gender
1	A	512	313	825	male
2	A	89	19	108	female
3	B	353	207	560	male
4	B	17	8	25	female
5	C	120	205	325	male
6	C	202	391	593	female
7	D	138	279	417	male
8	D	131	244	375	female
9	E	53	138	191	male
10	E	94	299	393	female
11	F	22	351	373	male
12	F	24	317	341	female

Berkeley Paradox

Which is correct? No way to know without assumptions

Mediator (department)

Collider + confound (ability)

	dept	admit	reject	applications	gender
1	A	512	313	825	male
2	A	89	19	108	female
3	B	353	207	560	male
4	B	17	8	25	female
5	C	120	205	325	male
6	C	202	391	593	female
7	D	138	279	417	male
8	D	131	244	375	female
9	E	53	138	191	male
10	E	94	299	393	female
11	F	22	351	373	male
12	F	24	317	341	female

The Fork

$$X \leftarrow Z \rightarrow Y$$

The Pipe

$$X \rightarrow Z \rightarrow Y$$

The Collider

$$X \rightarrow Z \leftarrow Y$$

The Descendant

$$\begin{tikzcd} X \arrow[r] & Z \arrow[r] \arrow[d] & Y \\ & A & \end{tikzcd}$$

$$X \leftarrow Z \rightarrow Y$$

```r
cols <- c(4,2)

N <- 300
Z <- rbern(N)
X <- rnorm(N,2*Z-1)
Y <- rnorm(N,2*Z-1)

plot( X , Y , col=cols[Z+1] , lwd=3 )

abline(lm(Y[Z==1]~X[Z==1]),col=2,lwd=3)

abline(lm(Y[Z==0]~X[Z==0]),col=4,lwd=3)

abline(lm(Y~X),lwd=3)
```

$$X \rightarrow Z \rightarrow Y$$

```r
cols <- c(4,2)

N <- 300
X <- rnorm(N)
Z <- rbern(N,inv_logit(X))
Y <- rnorm(N,(2*Z-1))

plot( X , Y , col=cols[Z+1] , lwd=3 )

abline(lm(Y[Z==1]~X[Z==1]),col=2,lwd=3)

abline(lm(Y[Z==0]~X[Z==0]),col=4,lwd=3)

abline(lm(Y~X),lwd=3)
```

$$X \rightarrow Z \leftarrow Y$$

```r
cols <- c(4,2)

N <- 300
X <- rnorm(N)
Y <- rnorm(N)
Z <- rbern(N,inv_logit(2*X+2*Y-2))

plot( X , Y , col=cols[Z+1] , lwd=3 )

abline(lm(Y[Z==1]~X[Z==1]),col=2,lwd=3)

abline(lm(Y[Z==0]~X[Z==0]),col=4,lwd=3)

abline(lm(Y~X),lwd=3)
```

Non-linear haunting

In event models, effect reversal can arise other ways

Example: Base rate differences

```r
# base rate differences erasing effect of X
N <- 1000
X <- rnorm(N)
Z <- rbern(N)
p <- inv_logit(X + ifelse(Z==1,-1,5))
Y <- rbern(N,p)
```

$$\begin{tikzcd} X \arrow[r] & Y \\ & Z \arrow[u] \end{tikzcd}$$

```r
# base rate differences erasing effect of X
N <- 1000
X <- rnorm(N)
Z <- rbern(N)
p <- inv_logit(X + ifelse(Z==1,-1,5))
Y <- rbern(N,p)

mX <- quap(
    alist(
        Y ~ dbern(p),
        logit(p) <- a + b*X,
        a ~ dnorm(0,1),
        b ~ dnorm(0,1)
    ) , data=list(Y=Y,X=X) )
```

$$\begin{tikzcd} X \arrow[r] & Y \\ & Z \arrow[u] \end{tikzcd}$$

$$\text{logit}(p_i) = \alpha + \beta X_i$$

```r
# base rate differences erasing effect of X
N <- 1000
X <- rnorm(N)
Z <- rbern(N)
p <- inv_logit(X + ifelse(Z==1,-1,5))
Y <- rbern(N,p)

mX <- quap(
    alist(
        Y ~ dbern(p),
        logit(p) <- a + b*X,
        a ~ dnorm(0,1),
        b ~ dnorm(0,1)
    ) , data=list(Y=Y,X=X) )

mXZ <- quap(
    alist(
        Y ~ dbern(p),
        logit(p) <- a + b[Z]*X,
        a ~ dnorm(0,1),
        b[Z] ~ dnorm(0,1)
    ) , data=list(Y=Y,X=X,Z=Z+1) )
```

$$\begin{tikzcd} X \arrow[r] & Y \\ & Z \arrow[u] \end{tikzcd}$$

$$\text{logit}(p_i) = \alpha + \beta X_i$$

$$\text{logit}(p_i) = \alpha + \beta_{Z[i]} X_i$$

$$\text{logit}(p_i) = \alpha + \beta X_i$$

$$\text{logit}(p_i) = \alpha + \beta_{Z[i]} X_i$$

$$\text{logit}(p_i) = \alpha + \beta_{Z[i]} X_i$$

```r
mX <- quap(
    alist(
        Y ~ dbern(p),
        logit(p) <- a + b*X,
        a ~ dnorm(0,1),
        b ~ dnorm(0,1)
    ) , data=list(Y=Y,X=X) )

mXZ <- quap(
    alist(
        Y ~ dbern(p),
        logit(p) <- a + b[Z]*X,
        a ~ dnorm(0,1),
        b[Z] ~ dnorm(0,1)
    ) , data=list(Y=Y,X=X,Z=Z+1) )

mXZ2 <- quap(
    alist(
        Y ~ dbern(p),
        logit(p) <- a[Z] + b[Z]*X,
        a[Z] ~ dnorm(0,1),
        b[Z] ~ dnorm(0,1)
    ) , data=list(Y=Y,X=X,Z=Z+1) )
```

$$\text{logit}(p_i) = \alpha + \beta X_i$$

$$\text{logit}(p_i) = \alpha + \beta_{Z[i]} X_i$$

$$\text{logit}(p_i) = \alpha_{Z[i]} + \beta_{Z[i]} X_i$$

$$\text{logit}(p_i) = \alpha + \beta_{Z[i]} X_i$$

$$\text{logit}(p_i) = \alpha_{Z[i]} + \beta_{Z[i]} X_i$$

$$\text{logit}(p_i) = \alpha + \beta_{Z[i]} X_i$$

$$\text{logit}(p_i) = \alpha_{Z[i]} + \beta_{Z[i]} X_i$$

$$\text{logit}(p_i) = \alpha_{Z[i]} + \beta_{Z[i]} X_i$$

Simpson's Pandora's Box

No paradox, because almost anything can produce it

People do not have intuitions about coefficient reversals

Stop naming statistical paradoxes; start teaching scientific logic

Pandora opens a spreadsheet

Simpson 1951. The Interpretation of Interaction in Contingency Tables