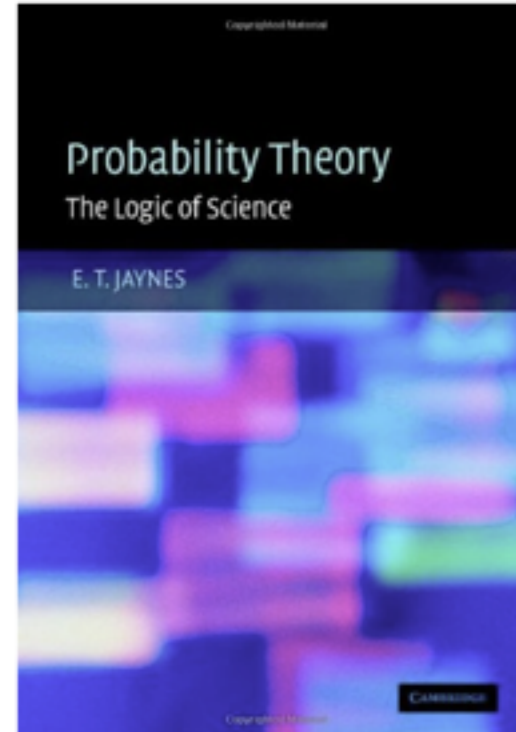
Maximum entropy

- Due to Edwin T. Jaynes (1922–1998)
- The maxent principle:
  - *Distribution with largest entropy is distribution most consistent with stated assumptions*
  - *Can happen the largest number of ways*
- For parameters, provides way to understand priors
- For observations, way to understand likelihood
- Also reproduces Bayesian updating as special case (*minimum cross-entropy*)



E. T. Jaynes (1922–1998)

Maximum entropy

- Ye olde information entropy:

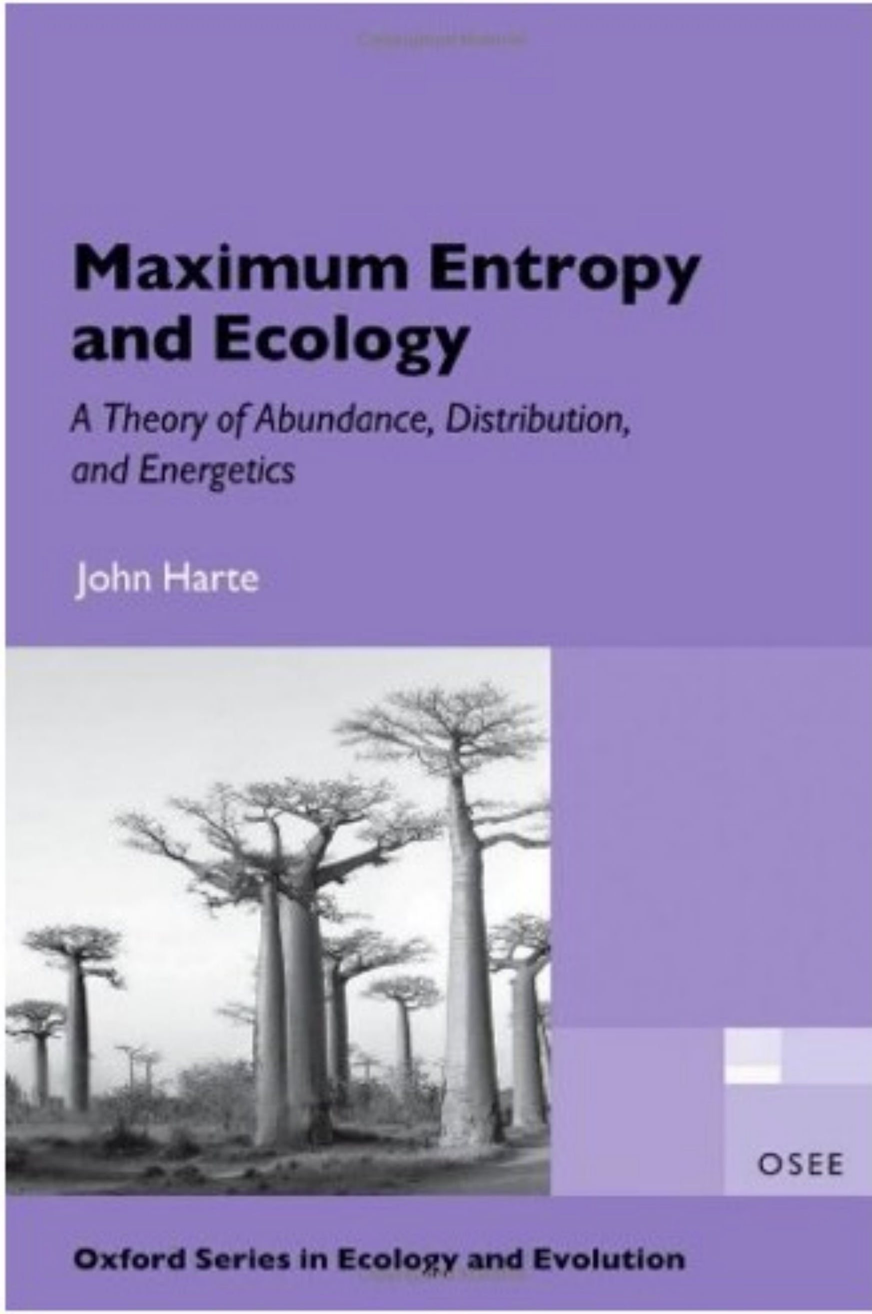
$$H(p) = -\sum_i p_i \log p_i$$

- Q: What kind of distribution maximizes this quantity?
- A: Flattest distribution still consistent with **constraints**. This is the distribution that can happen the most unique ways.
- Whatever does happen, bound to be one of those ways.

Maximum entropy

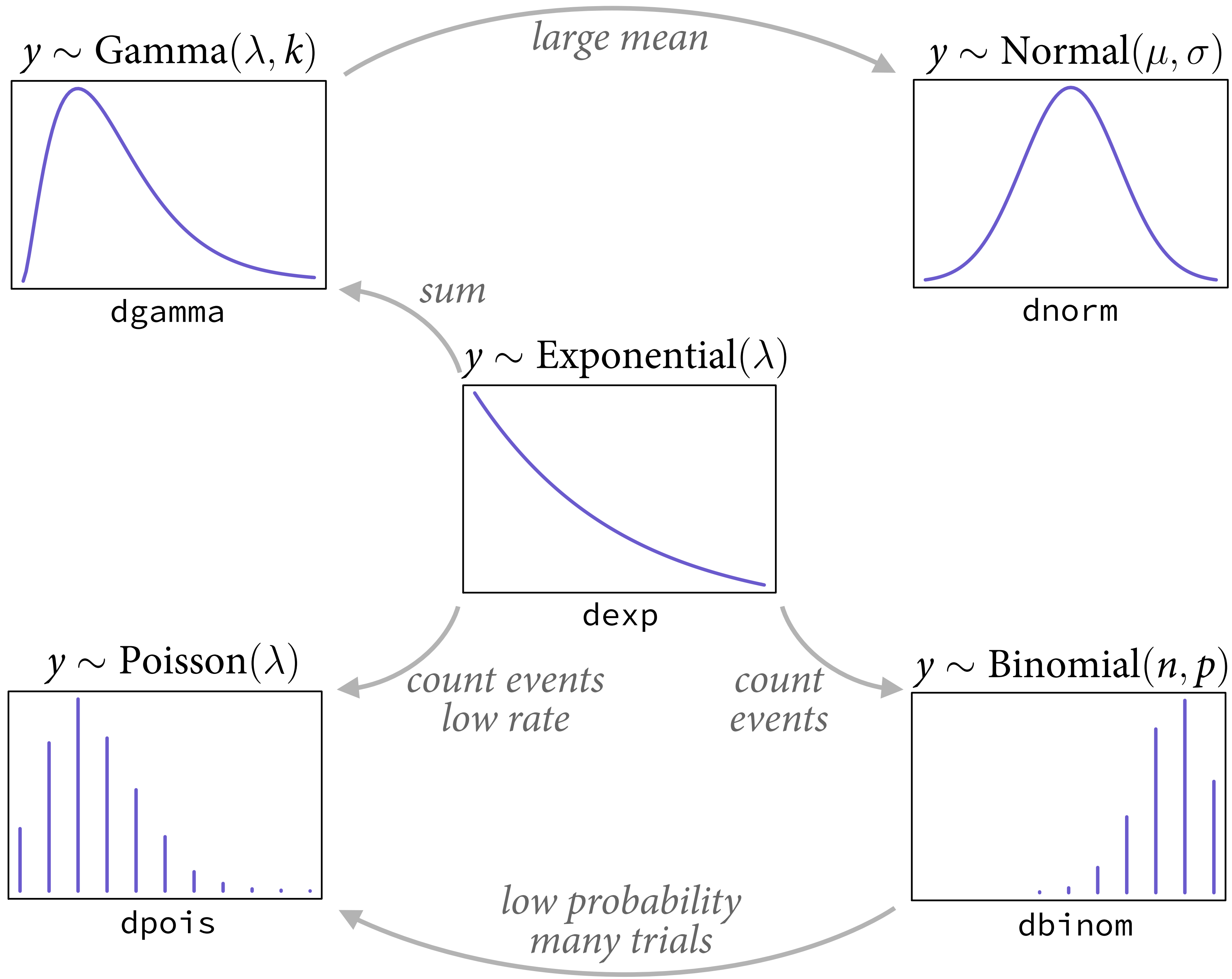
Constraints	Maxent distribution
Real value in interval	Uniform
Real value, finite variance	Gaussian
Binary events, fixed probability	Binomial
Non-negative real, has mean	Exponential

Please read the paper: The common patterns of nature!



Generalized Linear Models

- Goal: Connect linear model to outcome variable
- Still geocentric!
- Strategy:
  1. Pick an outcome distribution
  2. Model its parameters using links to linear models
  3. Compute posterior
- Can model multivariate relationships and non-linear responses
- Building blocks of multilevel models



$y \sim \text{Gamma}(\lambda, k)$
large mean
$y \sim \text{Normal}(\mu, \sigma)$
dgamma
dnorm
sum
$y \sim \text{Exponential}(\lambda)$
dexp
$y \sim \text{Poisson}(\lambda)$
count events
low rate
count
events
$y \sim \text{Binomial}(n, p)$
dpois
dbinom
low probability
many trials

Figure 9.5

Generalized Linear Models

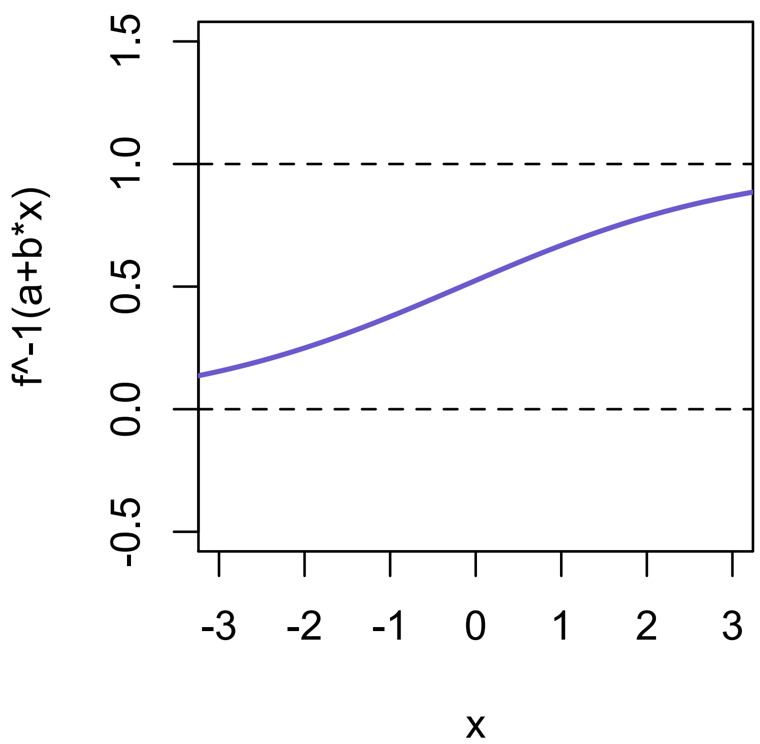
same units

$$y_i \sim \text{Normal}(\mu_i, \sigma),$$
$$\mu_i = \alpha + \beta x_i.$$

count

$$y_i \sim \text{Binomial}(n_i, p_i)$$
$$f(p_i) = \alpha + \beta x_i$$

link function



Need a link

$$y \sim \text{Binomial}(n, p)$$

- y and p on different scales
- y: count
- p: probability
- Want to model p as function of predictor variables
- Must bound it to [0,1] interval

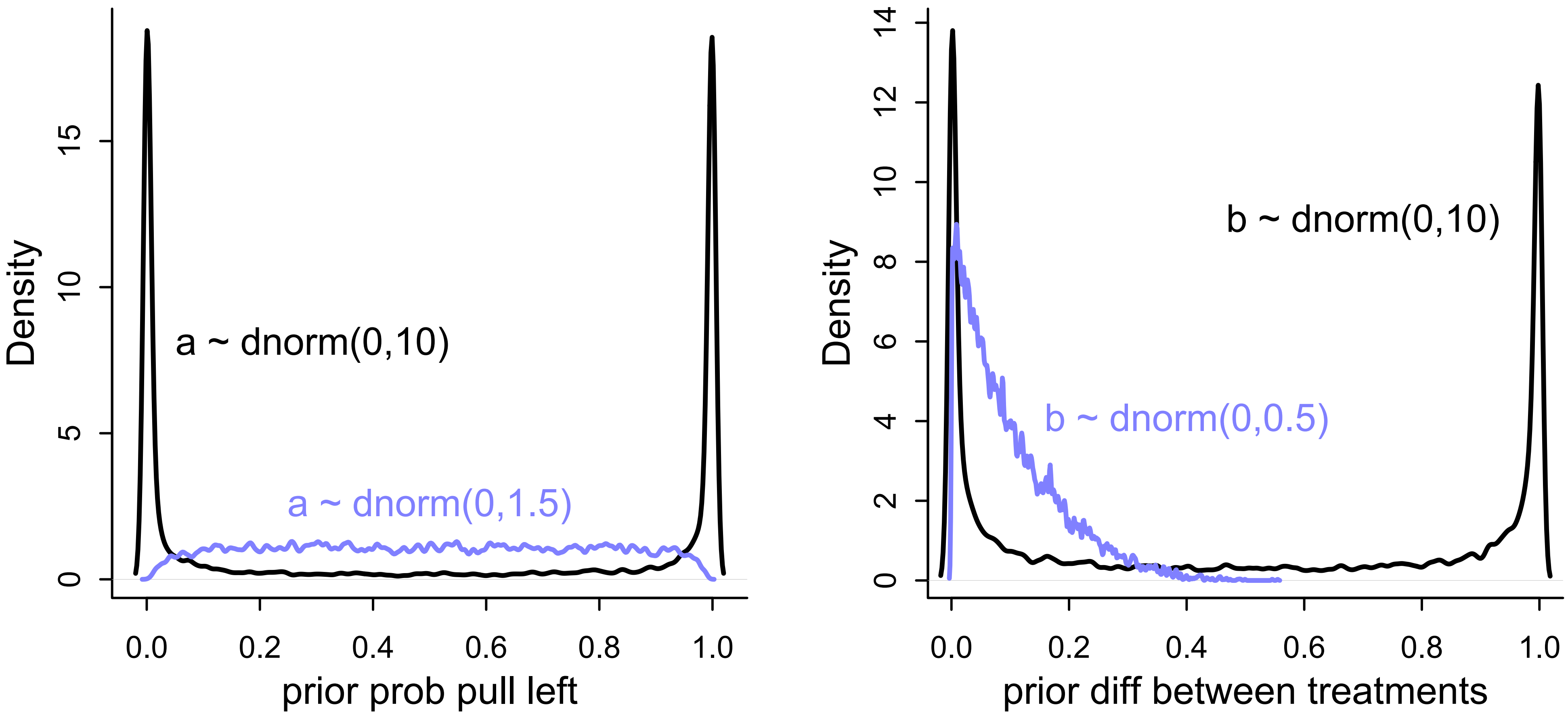

a ~ dnorm(0,10)
a ~ dnorm(0,1.5)
Density
prior prob pull left
0
5
10
15
0.0
0.2
0.4
0.6
0.8
1.0
b ~ dnorm(0,10)
b ~ dnorm(0,0.5)
Density
prior diff between treatments
0
2
4
6
8
10
12
14
0.0
0.2
0.4
0.6
0.8
1.0

Figure 11.3

Relative and absolute effects

- Parameters on *relative* effect scale
- Predictions on *absolute* effect scale
- Proportional odds: Relative effect measure

R code 11.22

```
post <- extract.samples(m11.4)
mean( exp(post$b[,4]-post$b[,2]) )
```

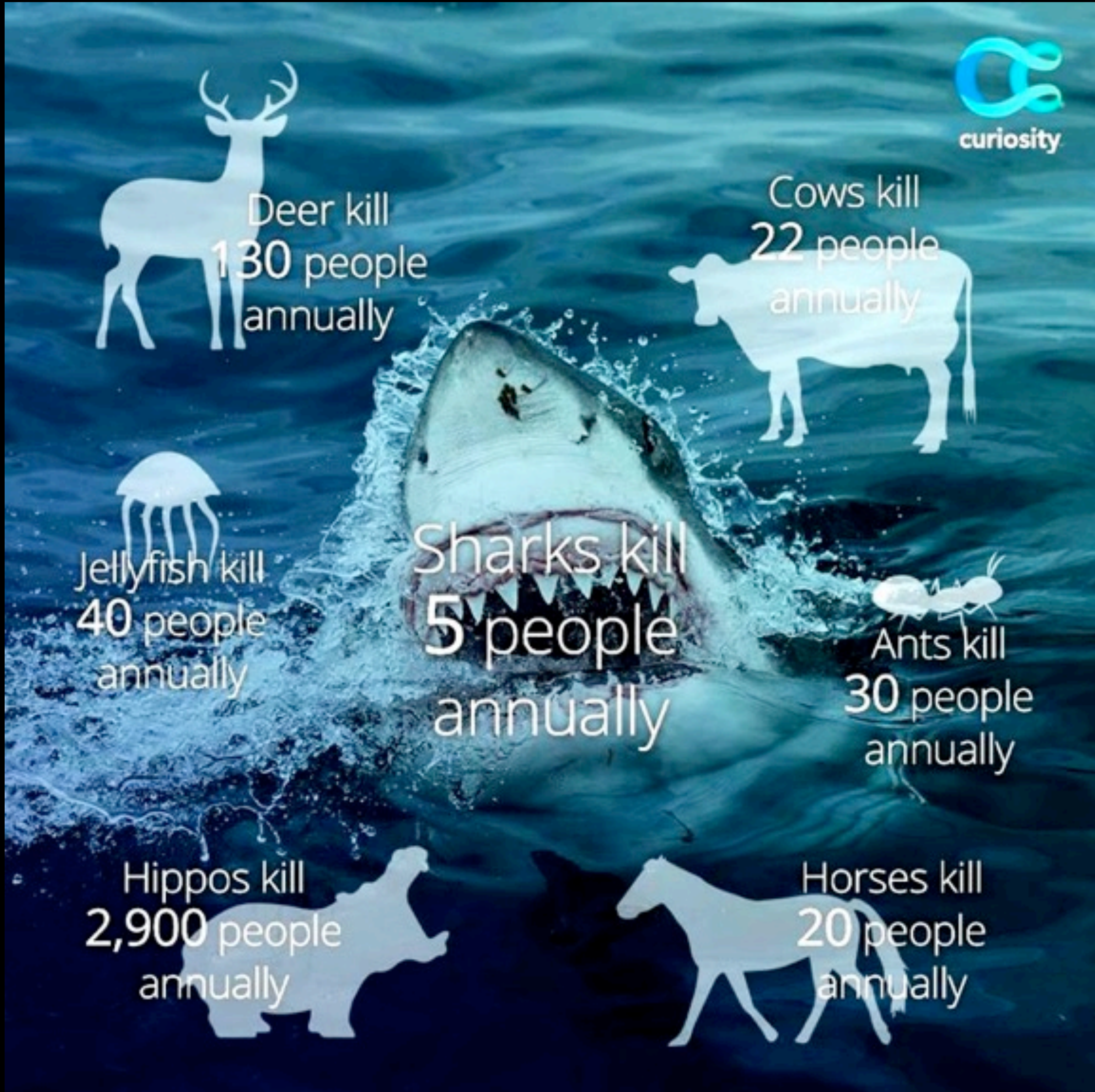
```
[1] 0.9206479
```

Relative and absolute effects

- Parameters on *relative* effect scale
- Predictions on *absolute* effect scale
- Using relative effects may exaggerate importance of predictor
    - Good for scaring people, getting published
    - Not so good for public health, scientific progress
    - But needed for causal inference

relative shark

absolute penguin


curiosity
Deer kill 130 people annually
Cows kill 22 people annually
Jellyfish kill 40 people annually
Sharks kill 5 people annually
Ants kill 30 people annually
Hippos kill 2,900 people annually
Horses kill 20 people annually

Risk communication

- Many people mistake relative risk for absolute risk
- Example:
  - 1/1000 women develop blood clots
  - 3/1000 women on birth control develop blood clots
  - => 200% increase in blood clots!
  - Change in probability is only 0.002
  - Pregnancy much more dangerous than blood clots

Aggregated Binomial

R code 11.28

```
library(rethinking)
data(UCBadmit)
d <- UCBadmit
```

- Numbers accepted/rejected to 6 PhD programs at UC Berkeley (largest depts in 1973)
- Evidence of gender discrimination? Dean was afraid of lawsuit.
- Call in the statisticians!

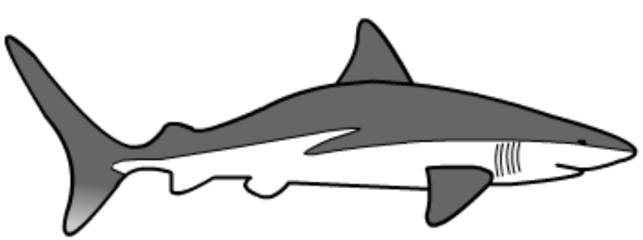
```
m11.7 <- ulam(
    alist(
        admit ~ dbinom( applications , p ) ,
        logit(p) <- a[gid] ,
        a[gid] ~ dnorm( 0 , 1.5 )
    ) , data=dat_list , chains=4 )
precis( m11.7 , depth=2 )
```

Posterior contrast

- Compute the contrast between genders
- On both logit (shark) and prob (penguin) scales

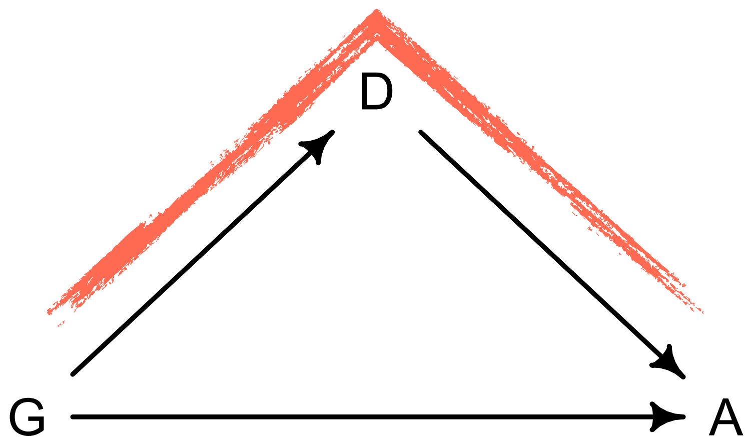
```
post <- extract.samples(m11.7)
diff_a <- post$a[,1] - post$a[,2]
diff_p <- inv_logit(post$a[,1]) - inv_logit(post$a[,2])
precis( list( diff_a=diff_a , diff_p=diff_p ) )
```

```
'data.frame': 10000 obs. of 2 variables:
       mean   sd 5.5% 94.5%     histogram
diff_a 0.61 0.06 0.51  0.71    ▁▁▁▂▇▇▃▁▁▁▁
diff_p 0.14 0.01 0.12  0.16  ▁▁▁▂▃▇▇▅▂▁▁▁▁
```



Backdoor admissions

- Backdoor path through department
- Use unique intercepts to adjust for that path



$$A_i \sim \text{Binomial}(N_i, p_i)$$
$$\text{logit}(p_i) = \alpha_{\text{GID}[i]} + \delta_{\text{DEPT}[i]}$$
$$\alpha_j \sim \text{Normal}(0, 1.5)$$
$$\delta_k \sim \text{Normal}(0, 1.5)$$

Stratification by department

$$
\begin{aligned}
A_i &\sim \text{Binomial}(N_i, p_i) \\
\text{logit}(p_i) &= \alpha_{\text{GID}[i]} \\
\alpha_j &\sim \text{Normal}(0, 1.5)
\end{aligned}
$$

Stat Q: What are the average probabilities of admission for females and males across all departments?

Causal Q: What is the TOTAL influence of gender?

$$
\begin{aligned}
A_i &\sim \text{Binomial}(N_i, p_i) \\
\text{logit}(p_i) &= \alpha_{\text{GID}[i]} + \delta_{\text{DEPT}[i]} \\
\alpha_j &\sim \text{Normal}(0, 1.5) \\
\delta_k &\sim \text{Normal}(0, 1.5)
\end{aligned}
$$

Stat Q: What is the average difference in probability of admission for females and males within departments?

Causal Q: What is the DIRECT influence of gender?

Posterior contrast

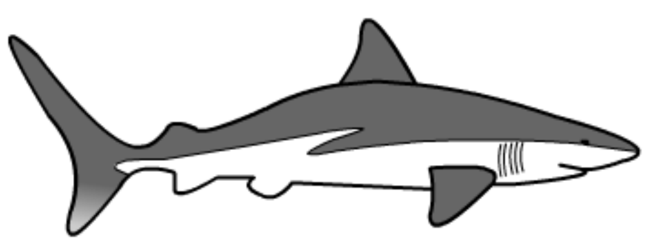
- Compute the contrast between genders
- On both logit (shark) and prob (penguin) scales

```
post <- extract.samples(m11.7)
diff_a <- post$a[,1] - post$a[,2]
diff_p <- inv_logit(post$a[,1]) - inv_logit(post$a[,2])
precis( list( diff_a=diff_a , diff_p=diff_p ) )
```

On the outcome (absolute effect)

log-odds scale (relative effect)

```
'data.frame': 10000 obs. of 2 variables:
       mean   sd 5.5% 94.5%    histogram
diff_a 0.61 0.06 0.51  0.71
diff_p 0.14 0.01 0.12  0.16
```



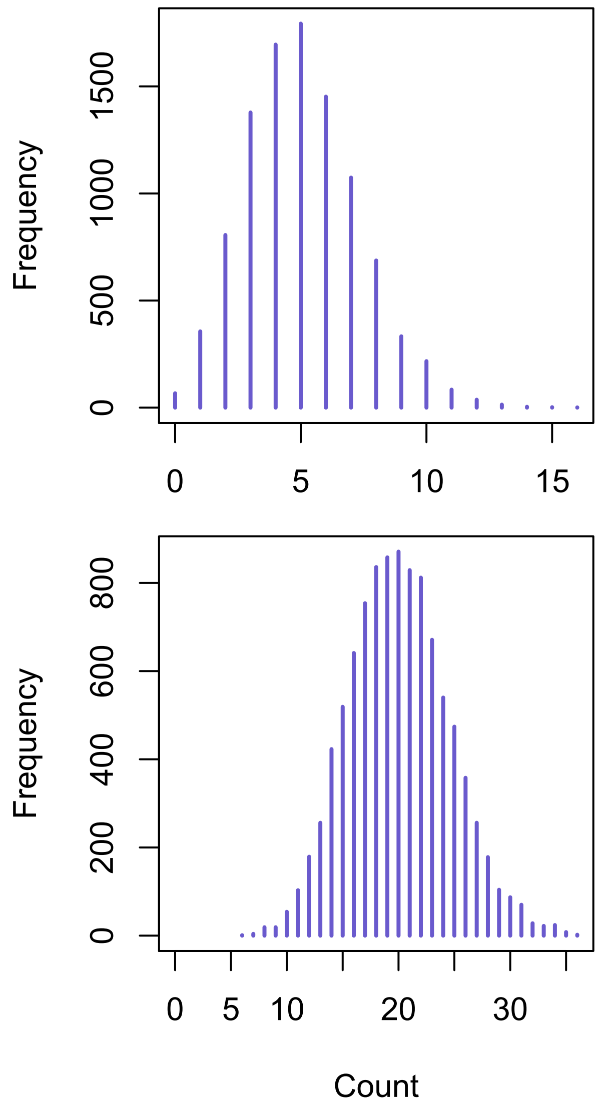
$$\mathbb{E}(y) = \bar{y} = \mu_x$$

Poisson GLMs

$$y \sim \text{Poisson}(\lambda)$$

$$E(y) = \lambda$$
$$\text{var}(y) = \lambda$$

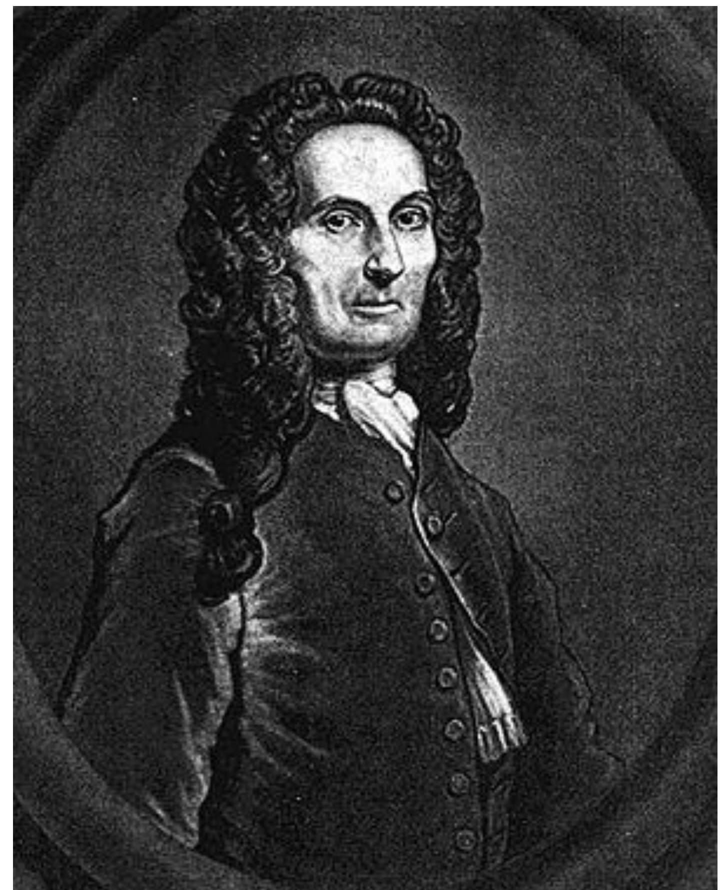
- Counts without upper limit, constant expected value
- Single parameter: events per unit time/distance
- Variance equal to mean



Poisson GLMs

- Examples: Soccer goals, fission events, photons striking a detector, DNA mutations, soldiers killed by horses

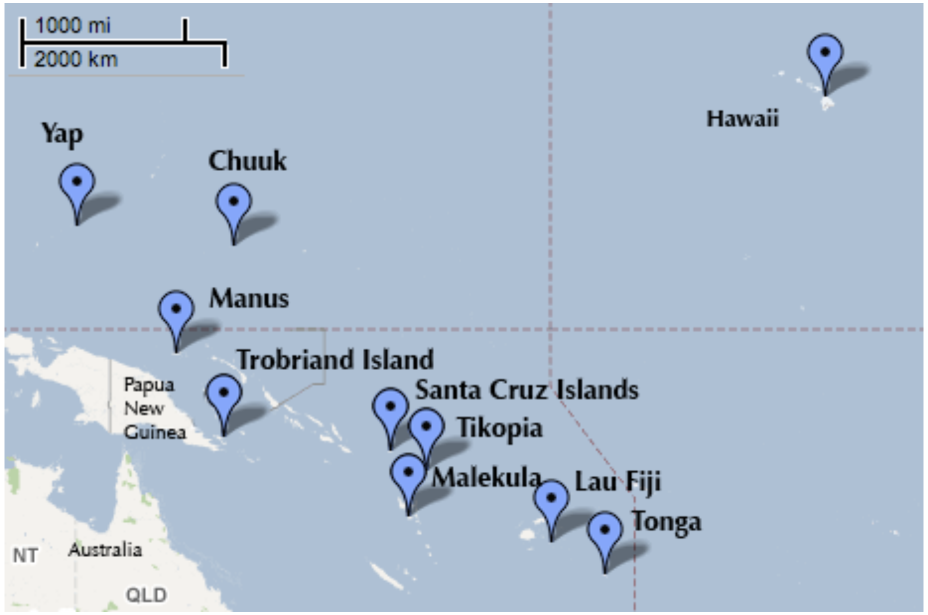
Siméon Denis Poisson (1781–1840)



Abraham de Moivre (1667–1754)

Oceanic tool complexity

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



(1) Complexity of toolkit proportional to magnitude of population?

(2) Contact with other islands moderates impact?

Dr. Michelle Kline (Simon Fraser U)

Anatomy of Poisson GLM

total_tools
(*outcome*)

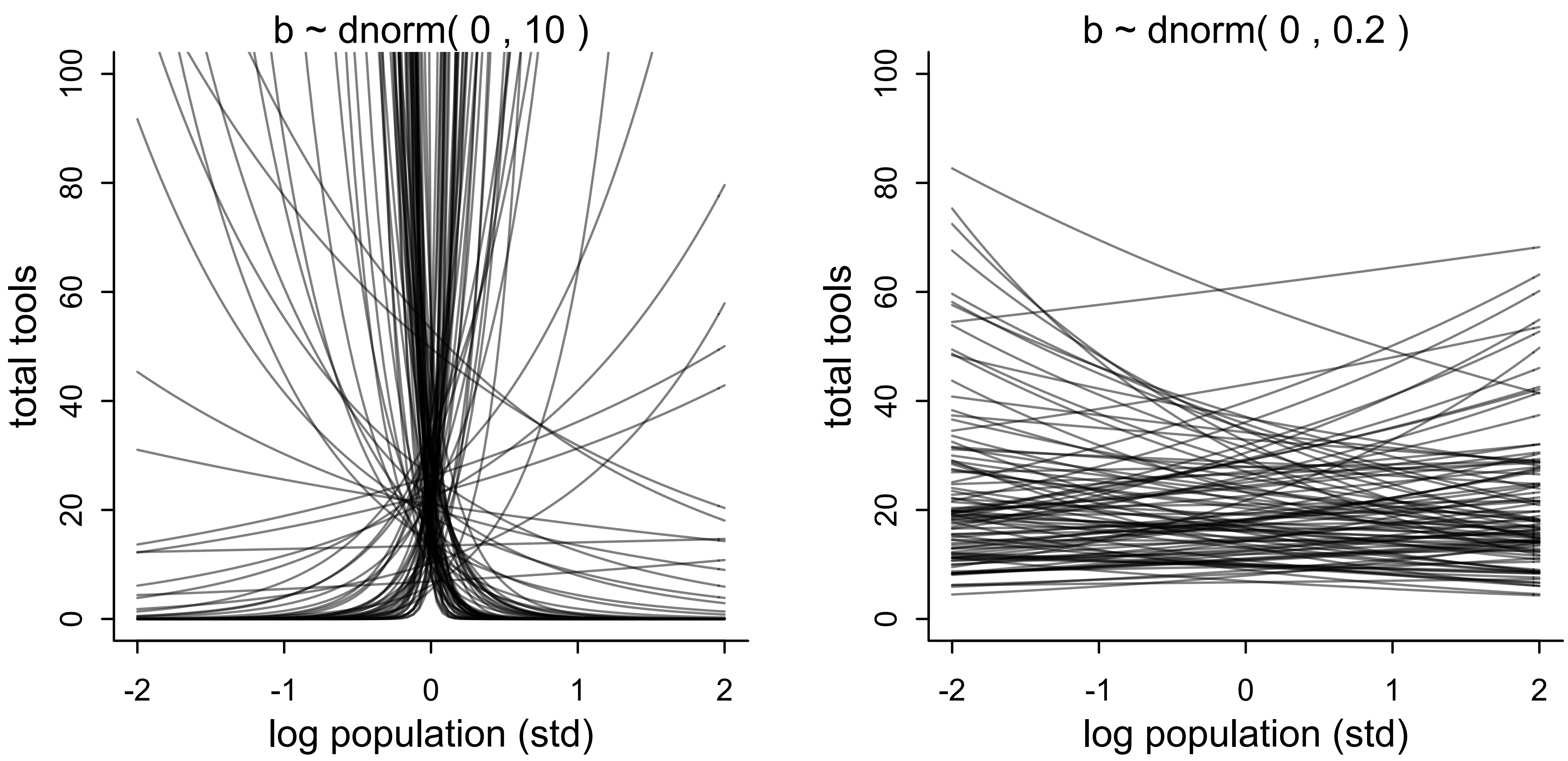
expected tools
for case *i*

log link

$$
\begin{aligned}
T_i &\sim \text{Poisson}(\lambda_i) \\
\log \lambda_i &= \alpha_{\text{CID}[i]} + \beta_{\text{CID}[i]} \log P_i \\
\alpha_j &\sim \text{to be determined} \\
\beta_j &\sim \text{to be determined}
\end{aligned}
$$

Priors & the log link

- Slopes equally unintuitive



Scientific model

- Change in tools per unit time:

Diminishing returns
("elasticity")

Innovation rate

Loss rate

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

Population

Tools

Scientific model

- Solve for steady state expected number of tools
- Where $\Delta T = 0$

$$\hat{T} = \frac{\alpha P^{\beta}}{\gamma}$$

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

$$\lambda_i = \alpha P_i^{\beta} / \gamma$$

No *ad hoc* link function!

Scientific model

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

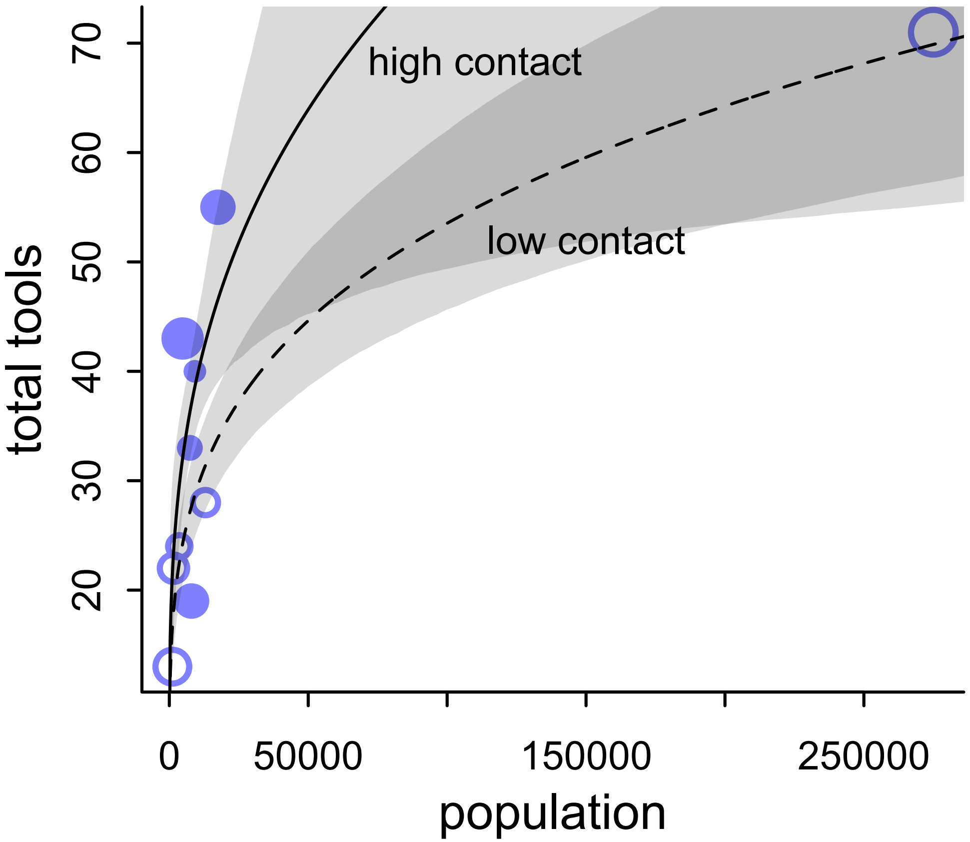
$$\lambda_i = \alpha P_i^{\beta} / \gamma$$

R code 11.52

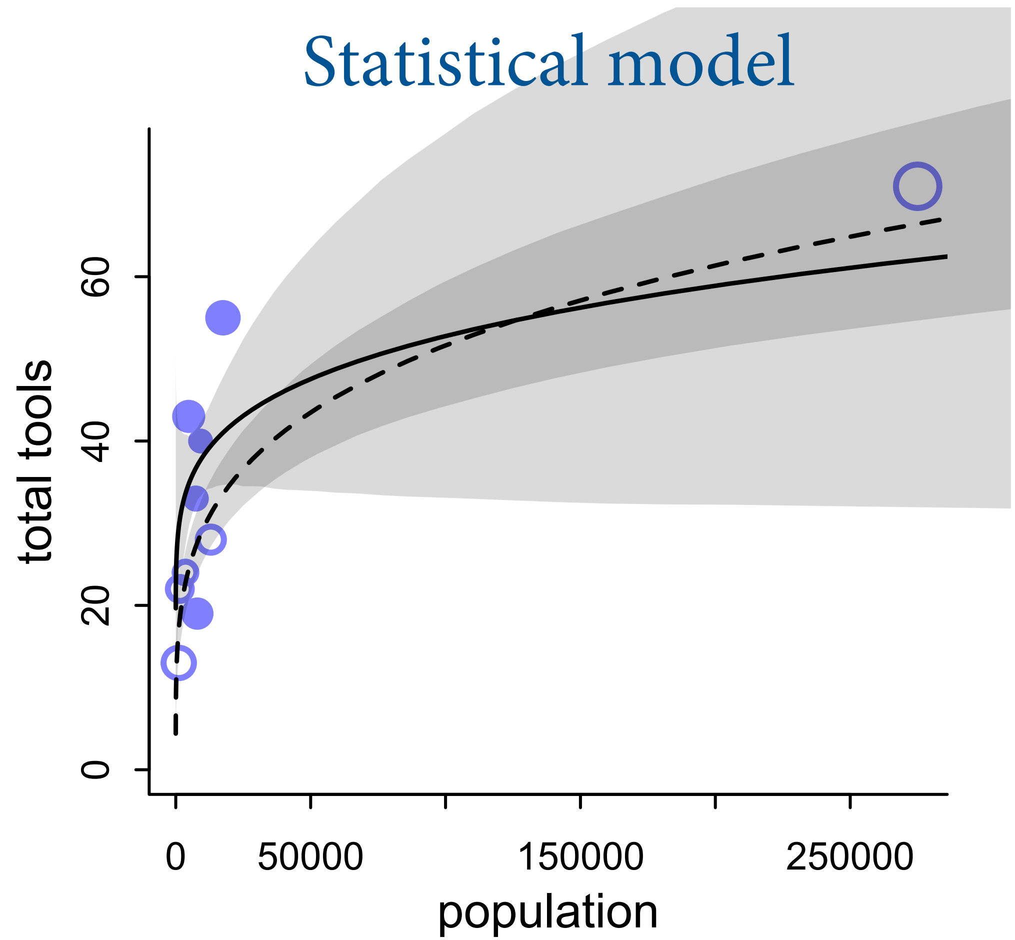
```
dat2 <- list( T=d$total_tools, P=d$population, cid=d$contact_id )
m11.11 <- ulam(
    alist(
        T ~ dpois( lambda ),
        lambda <- exp(a[cid])*P^b[cid]/g,
        a[cid] ~ dnorm(1,1),
        b[cid] ~ dexp(1),
        g ~ dexp(1)
    ), data=dat2 , chains=4 , log_lik=TRUE )
```

Science pays

Scientific model



Statistical model



Model violations now mean something.

Parameters now mean something.

Additional count distributions

- Multinomial/categorical: generalized binomial, more than 2 un-ordered outcomes
- Geometric: number of trials until specific event
- Mixtures, coping with heterogeneity:
  - Beta-binomial: varying probabilities
  - gamma-Poisson: aka negative-Binomial, varying rates
  - others (e.g. Dirichlet-multinomial)