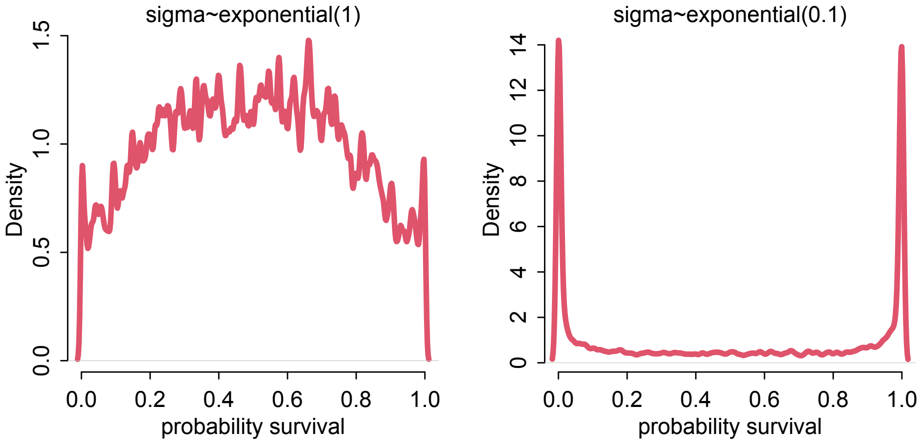
STATISTICAL RETHINKING 2022
WEEK 6 SOLUTIONS

1. Simulating varying effect priors is in principle like simulating any other priors. The only difference is that the parameters have an implied order now, because some parameters depend upon others. So in this problem we must simulate σ and $\bar{\alpha}$ first, and then we can simulate the individual tank α_T variables. Here is how I did it:

```
n <- 1e4
sigma <- rexp(n,1)
abar <- rnorm(n,0,1)
aT <- rnorm(n,abar,sigma)
dens(inv_logit(aT),xlim=c(0,1),adj=0.1,lwd=4,col=2)
```

This simulates 1e4 tanks from the prior. Then it plots the implied survival distribution on the probability scale. Before showing that plot, I'll also simulate from a prior in which $\sigma \sim \text{Exponential}(0.1)$. Showing both:



Increasing the variantion across tanks, by making the σ distribution wider, pushes prior survival up against the floor and ceiling of the outcome space. This is the same phenomenon you saw before for ordinary logit models. The key lesson again is that flat priors on one scale are not necessarily flat on another.

2. I will use the suggested matrix of parameters approach.

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

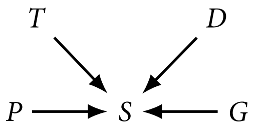
```
dat <- list(
    S = d$surv,
    D = d$density,
    T = 1:nrow(d),
    P = ifelse( d$pred=="no" , 1L , 2L ),
    G = ifelse( d$size=="small" , 1L , 2L ) )

m2 <- ulam(
    alist(
        S ~ binomial( D , p ),
        logit(p) <- a[T] + b[P,G],
        a[T] ~ normal( 0 , sigma ),
        matrix[P,G]:b ~ normal( 0 , 1 ),
        sigma ~ exponential( 1 )
    ), data=dat , chains=4 , cores=4 , log_lik=TRUE )
precis(m2,3,pars=c("b","sigma"))
```

```
         mean   sd  5.5% 94.5% n_eff Rhat4
b[1,1]   2.36 0.29  1.89  2.84  1108  1.01
b[1,2]   2.50 0.31  2.00  2.98  1282  1.00
b[2,1]   0.43 0.25  0.06  0.81   673  1.00
b[2,2]  -0.43 0.25 -0.81 -0.03   720  1.00
sigma    0.74 0.15  0.51  0.98   455  1.01
```

The parameters are in order from top to bottom: no-pred/small, no-pred/large, pred/small, pred/large. The curious thing is not that survival is lower with predation, but rather that it is lowest for large tadpoles, `b[2,2]`. This is a strong interaction that would be missed if we had made the effects purely additive with one another (on the log-odds scale). The Vonesh & Bolker paper that these data come from goes into this interaction in great depth.

The problem asked for a justification of the model in terms of the DAG. Here's the DAG:



This is an experiment, so we know the treatments *P*, *G*, and *D* are not confounded. At least not in any obvious way. And then unobserved tank effects *T* also moderate the influence of the treatments. The model I used tries to estimate how *P* and *G* moderate one another. It ignores *D*, which we are allowed to do, because it is not a confound, just a competing cause. But I include tanks, which is also just a competing cause. Including competing causes helps with precision, if nothing else.

What the DAG does not justify is the full interaction matrix that I used. DAGs contain no information about how variables interact or not to produce outcomes.

They just show inputs and outputs. To justify any particular statistical model, you need more than the DAG.

3. Density is an important factor in these experiments. So let's include it finally. I will do something simple, just include it as an additive effect that interacts with predators. But I will use the logarithm of density, so that it has implied diminishing returns on hte log-odds scale.

```
dat$Do <- standardize(log(d$density))
m3 <- ulam(
    alist(
        S ~ binomial( D , p ),
        logit(p) <- a[T] + b[P,G] + bD[P]*Do,
        a[T] ~ normal( 0 , sigma ),
        matrix[P,G]:b ~ normal( 0 , 1 ),
        bD[P] ~ normal(0,0.5),
        sigma ~ exponential( 1 )
    ), data=dat , chains=4 , cores=4 , log_lik=TRUE )
precis(m3,3,pars=c("b","bD","sigma"))
```

```
        mean   sd  5.5% 94.5% n_eff Rhat4
b[1,1]  2.34 0.29  1.89  2.80  1632     1
b[1,2]  2.48 0.29  2.03  2.97  1648     1
b[2,1]  0.55 0.24  0.16  0.94  1097     1
b[2,2] -0.38 0.23 -0.73 -0.02  1236     1
bD[1]   0.14 0.21 -0.21  0.47  1340     1
bD[2]  -0.47 0.17 -0.75 -0.20  1375     1
sigma   0.64 0.13  0.44  0.87   417     1
```

Again an interaction. Higher densities are worse for survival, but only in the presence of predators. The other estimates are not changed much.

The σ estimate here is a little smaller than in the previous problem. This is just because density is an real cause of survival, so it explains some of the variation that was previously soaked up by tanks with different densities.

4. This last problem is not easy. But the steps are similar to the causal simulation examples I have shown in lecture. First we set up the status quo (pre-intervention) groups.

```
# setup status quo groups
nreps <- 20
D <- rep( c(10,35,10,35) , times=nreps )
Do <- rep( ( D - mean(d$density) ) / sd(d$density) , times=nreps )
G <- rep( c(1,1,2,2) , times=nreps )
P <- rep( rep(2,4) , times=nreps )
```

The above specifies 80 groups with the right proportions of density and size. You can verify by doing the cross-tabs. Next we simulate unobserved group (tank) effects.

We need 100 groups, but we want replicates from the posterior to marginalize over. So we'll do 2000 samples from the posterior for each of the 100 groups.

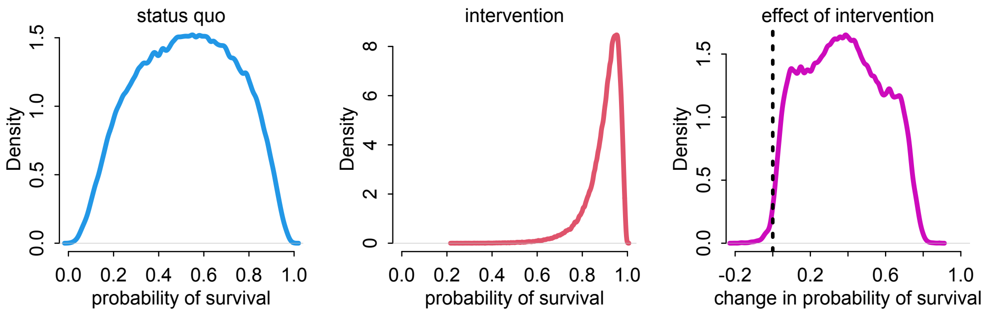
```
# simulate tanks
post <- extract.samples(m3)
aT <- replicate( length(D) , rnorm(2000,0,post$sigma) )
```

Now we can simulate survival. For both status quo (predators) and absence.

```
# simulate survival status quo
p1 <- sapply( 1:length(D) ,
    function(i)
        inv_logit(
            aT[,i] +
            post$b[,P[i],G[i]] +
            post$bD[,P[i]]*Do[i] ) )

# simulate survival intervention
P <- rep( rep(1,4) , times=nreps )
p2 <- sapply( 1:length(D) ,
    function(i)
        inv_logit(
            aT[,i] +
            post$b[,P[i],G[i]] +
            post$bD[,P[i]]*Do[i] ) )
```

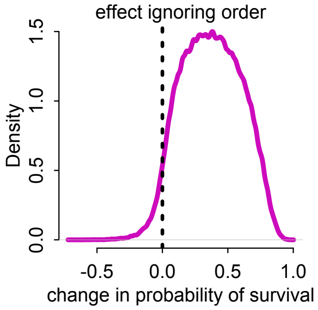
This code uses the same groups in each scenario, changing only the presence of predators. So the intervention isolates *P*. Now we can plot the distributions of survival under each scenario and the difference.



The intervention helps a lot, but because the survival varies so much under status quo, the causal effect ranges from almost nothing to at most 80%.

You might notice that just trying to visually subtract the blue from the red density above doesn't really do the job. Why does that weird bulge form in the middle of the purple density? It's because we are comparing specific groups in both scenarios. If we shuffle the group indices, you'll get a smoother curve:

```
# shuffle and take difference
s1 <- sample( 1:80 , size=80 )
s2 <- sample( 1:80 , size=80 )
dens( p2[,s2] - p1[,s1] , lwd=4 , col=6 )
```



This is not the expectation of the specified intervention. It currepsinds instead to something in which the group effects also change (are shuffled). Notice that it actually predicts a larger maximum effect. It is wrong.