

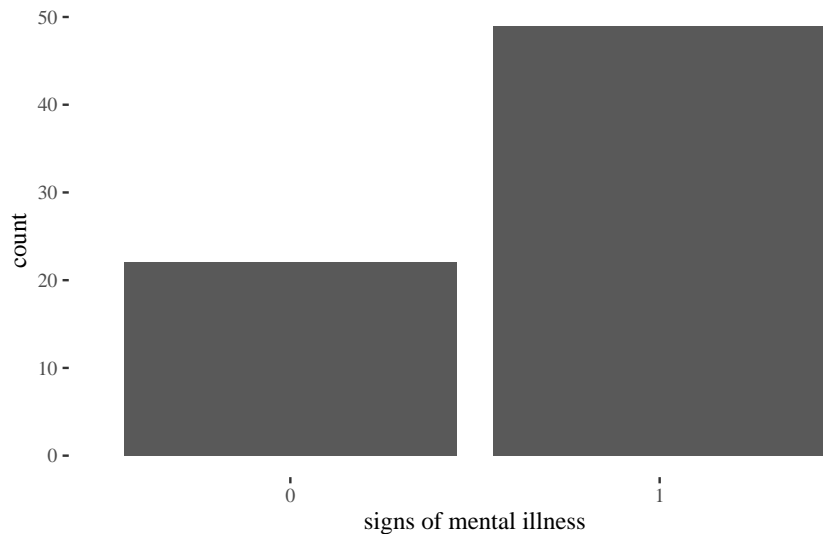
# Sampling and Uncertainty

Rafał Urbaniak, Nikodem Lewandowski  
(LoPSE research group, University of Gdansk)

## Proportion estimates and sampling

### Prior signs of mental illness

(US mass shootings 1982–2015)



## Proportion estimates and sampling

```
p_grid <- seq(0,1, length.out = 1001)
prior <- rep( 1, 1001)

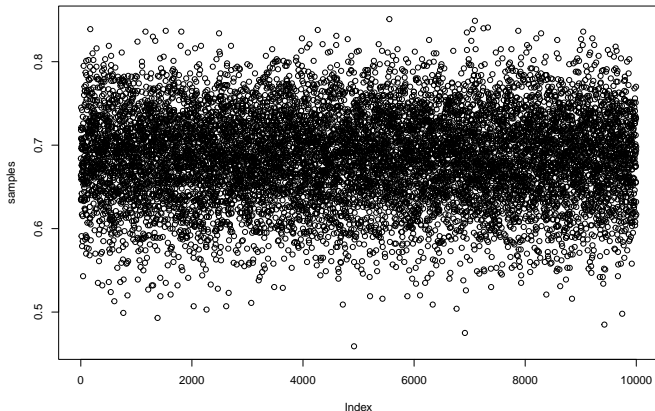
likelihood <- dbinom( sum(sh$mental),
                      size = nrow(sh), prob = p_grid)

posteriorUnnormalized <- prior * likelihood

posterior <- posteriorUnnormalized / sum(posteriorUnnormalized)
```

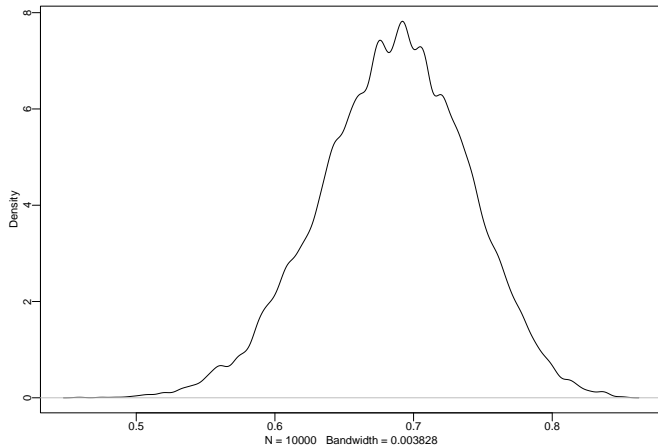
# Proportion estimates and sampling

```
samples <- sample(p_grid, prob = posterior, size = 1e4,  
                 replace = TRUE)  
plot(samples)
```



# Proportion estimates and sampling

```
dens(samples)
```



## Proportion estimates and sampling

```
sum(posterior[p_grid > .6])
```

```
## [1] 0.9357886
```

```
sum(samples > .6) / 1e4
```

```
## [1] 0.9371
```

```
sum(samples > .6 & samples < .7) / 1e4
```

```
## [1] 0.5308
```

## Proportion estimates and sampling

```
quantile( samples, c(.1,.9))
```

```
##    10%    90%  
## 0.615 0.753
```

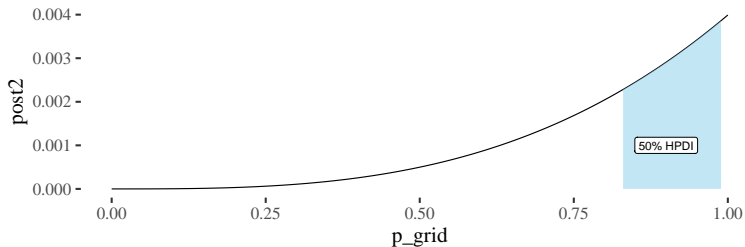
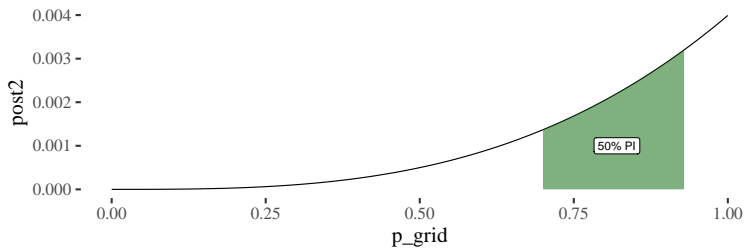
```
PI(samples, .8)
```

```
##    10%    90%  
## 0.615 0.753
```

```
HPDI(samples, .8)
```

```
## |0.8  0.8|  
## 0.614 0.752
```

# PI vs HPDI





# Now with model building

```
sh$WEAPONS OBTAINED LEGALLY
```

```
## [1] "Yes" "Yes" "No" "" "Yes" "Yes" "Yes" "Yes" "Yes" "No" "Yes" "Yes"  
## [13] "Yes" "Yes" "Yes" "Yes" "Yes" "Yes" "Yes" "Yes" "Yes" "No" "Yes" "Yes"  
## [25] "Yes" "Yes" "No" "No" "Yes" "Yes" "No" "Yes" "Yes" "Yes" "No" "Yes"  
## [37] "Yes" "Yes" "Yes" "Yes" "Yes" "Yes" "Yes" "Yes" "No" "No" "No" "Yes"  
## [49] "Yes" "No" "Yes" "Yes" "Yes" "" "Yes" "Yes" "No" "Yes" "Yes" "Yes"  
## [61] "Yes" "Yes" "Yes" "Yes" "Yes" "Yes" "Yes" "Yes" "Yes" "No" "Yes"
```

## Now with model building

```
legal <- sum(sh$WEAPONS OBTAINED LEGALLY == "Yes")
illegal <- sum(sh$WEAPONS OBTAINED LEGALLY == "No")
total <- legal + illegal

datweapons = list (legal = legal, illegal = illegal,
                  total = total)

weaponsModel <- ulam(
  alist(
    legal ~ dbinom( total , theta),
    theta ~ dunif(0,1)
  ) ,
  data= datweapons )
```

# Now with model building

```
precis(weaponsModel)
```

```
##           result
## mean    0.80425077
## sd      0.04415148
## 5.5%    0.73106015
## 94.5%   0.86934906
## n_eff  211.49685081
## Rhat    1.00201182
```

```
weaponsSamples <- as.data.frame(extract.samples(weaponsModel))
```

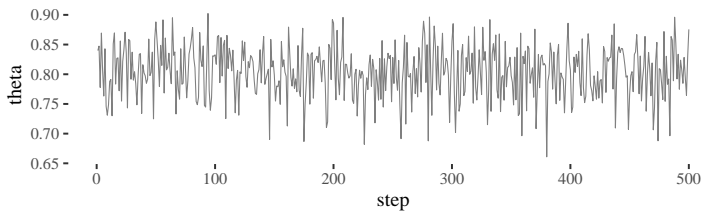
```
weaponsSamples$step <- 1:500
```

```
head(weaponsSamples)
```

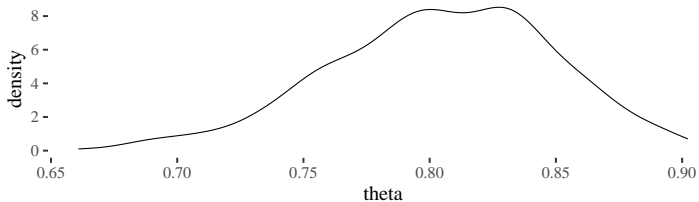
```
##      theta step
## 1 0.839904    1
## 2 0.847359    2
## 3 0.777781    3
## 4 0.869390    4
## 5 0.807456    5
## 6 0.763302    6
```

# Now with model building

Parameter visits



Posterior density



## Beyond binomial: lots of small factors

```
set.seed(312)
runif(1,-1,1)
```

```
## [1] 0.5832919
```

```
person1 <- runif(40,-1,1)
person1[1:15]
```

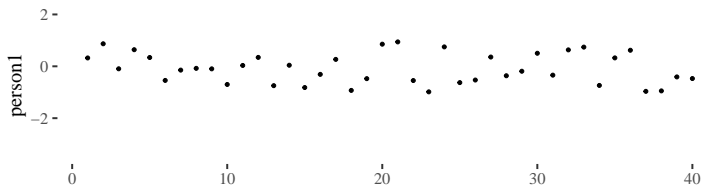
```
## [1] 0.32220310 0.86845868 -0.09645973 0.63976928 0.33923684 -0.54382513
## [7] -0.14640792 -0.07832506 -0.09818215 -0.70031680 0.03328715 0.34389619
## [13] -0.74667117 0.04139647 -0.81864988
```

```
person1pos <- cumsum(person1)
person1pos
```

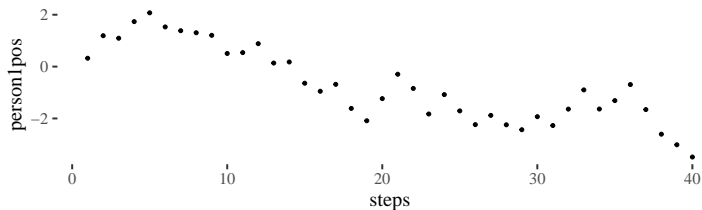
```
## [1] 0.3222031 1.1906618 1.0942021 1.7339713 2.0732082 1.5293830
## [7] 1.3829751 1.3046501 1.2064679 0.5061511 0.5394383 0.8833345
## [13] 0.1366633 0.1780598 -0.6405901 -0.9509429 -0.6835546 -1.6124981
## [19] -2.0885637 -1.2350482 -0.2942103 -0.8431846 -1.8282965 -1.0799415
## [25] -1.7083775 -2.2375286 -1.8804667 -2.2449167 -2.4347412 -1.9316288
## [31] -2.2723666 -1.6391910 -0.8989080 -1.6362535 -1.3121734 -0.6925716
## [37] -1.6578202 -2.6061456 -3.0153709 -3.4872044
```

# Beyond binomial: lots of small factors

40 random steps

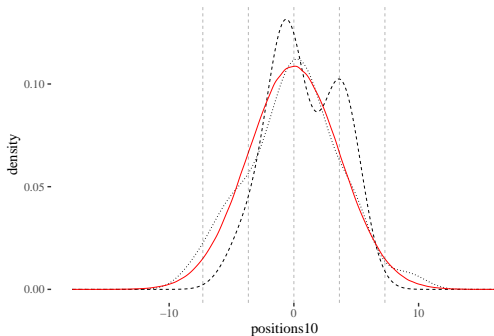


Path through the steps



# Beyond binomial: lots of small factors

Final destinations of 10, 100, and 1e6 drunkards



```
sd(positions1e6)
```

```
## [1] 3.651485
```

```
mean(abs(positions1e6) < abs(sd(positions1e6)) )
```

```
## [1] 0.681449
```

```
mean(abs(positions1e6) < 2 * abs(sd(positions1e6)) )
```

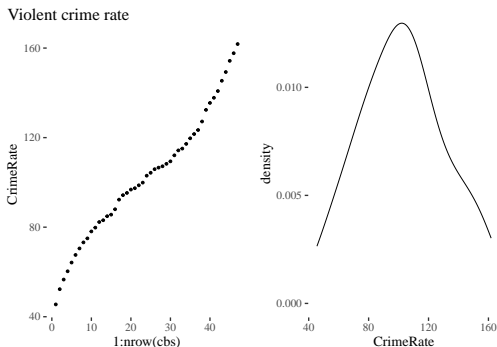
```
## [1] 0.954788
```

# Crime rates and normal distribution

```
cbs <- read.csv(file = "../datasets/CrimeByState.csv")  
#these are registered violent incidents per 100k citizens  
cbs$CrimeRate
```

```
## [1] 45.5 52.3 56.6 60.3 64.2 67.6 70.5 73.2 75.0 78.1 79.8 82.3  
## [13] 83.1 84.9 85.6 88.0 92.3 94.3 95.3 96.8 97.4 98.7 99.9 103.0  
## [25] 104.3 105.9 106.6 107.2 108.3 109.4 112.1 114.3 115.1 117.2 119.7 121.6  
## [37] 123.4 127.2 132.4 135.5 137.8 140.8 145.4 149.3 154.3 157.7 161.8
```

```
cbsPlot <- grid.arrange(ggplot(cbs)+geom_point(aes(x=1:nrow(cbs),y = CrimeRate))+th+  
  ggtitle("Violent crime rate"),  
  ggplot(cbs)+geom_density(aes(x=CrimeRate))+th, ncol=2)
```





## Crime rates and normal distribution

```
dat <- list(  
  rate = cbs$CrimeRate  
)  
  
set.seed(123)  
meanModel <- quap(  
  alist(  
    rate ~ dnorm( mu , sigma ) ,  
    mu ~ dnorm( 100 , 100 ) ,  
    sigma ~ dunif( 0 , 50 )  
  ), data = dat  
)
```

# Crime rates and normal distribution

```
precis(meanModel)
```

```
##           mean      sd      5.5%      94.5%  
## mu      102.79877 4.165765 96.14107 109.45646  
## sigma  28.58386 2.948141 23.87216 33.29556
```

```
pred <- sim(meanModel)
```

```
str(pred)
```

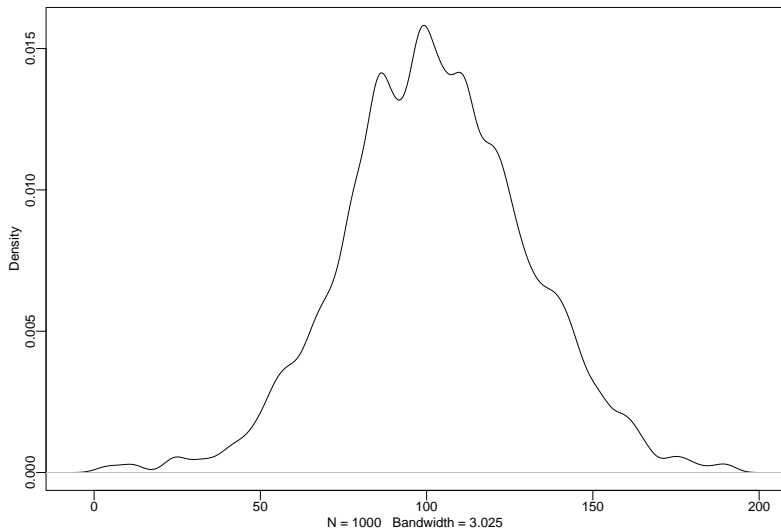
```
## num [1:1000, 1:47] 123 160 110 112 124 ...
```

```
pred[1:5, 1:5]
```

```
##           [,1]      [,2]      [,3]      [,4]      [,5]  
## [1,] 122.5506 67.21098 29.59797 150.41422 158.88619  
## [2,] 159.9094 115.09314 77.61558 193.44488 99.14939  
## [3,] 110.2740 122.49128 45.99694 136.25065 106.91446  
## [4,] 111.9459 112.71564 103.83174 92.10859 76.17974  
## [5,] 123.6168 92.76183 92.35412 106.23952 95.85145
```

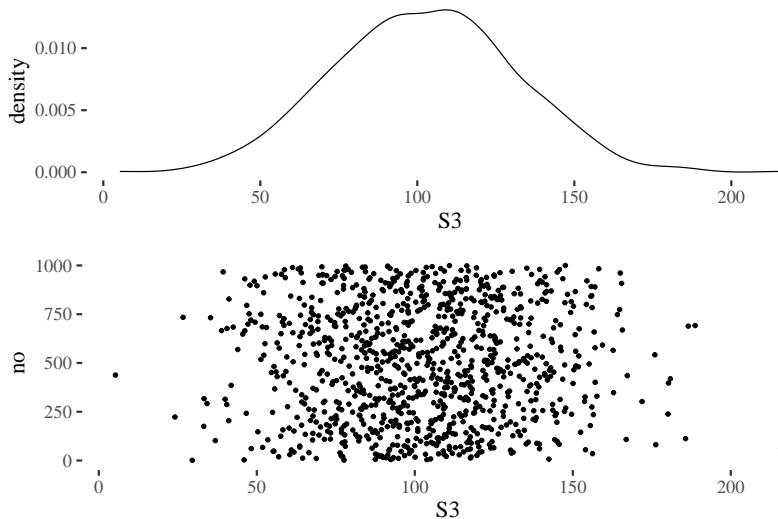
# Crime rates and normal distribution

```
dens (pred[,1])
```



# Crime rates and normal distribution

Simulated parameters for State 3



# Crime rates and normal distribution

```
(meanpreds <- apply(pred, MARGIN = 2, FUN = mean))
```

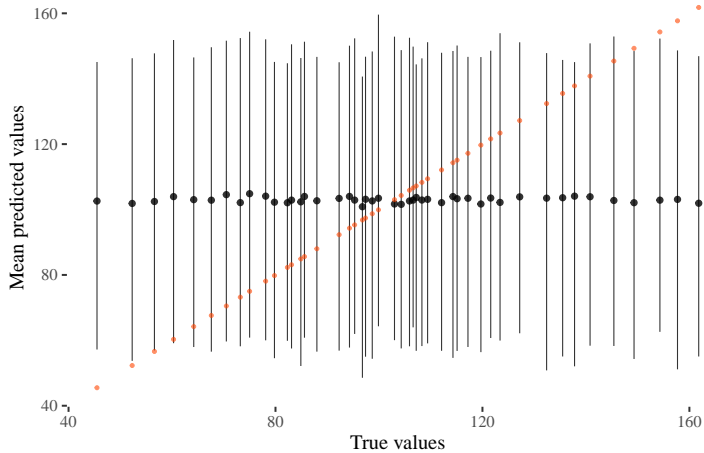
```
## [1] 102.5905 101.8595 102.4534 103.9375 103.0147 102.8535 104.5446 102.1100  
## [9] 104.8375 104.0880 102.2362 102.0704 102.8730 102.3484 103.9993 102.6737  
## [17] 103.4009 104.0101 102.8792 100.8548 103.1379 102.6278 103.4506 101.6893  
## [25] 101.5838 102.6099 102.8824 103.6936 102.8997 103.1187 102.0964 103.9570  
## [33] 103.3063 103.4443 101.6891 103.5112 102.1769 103.8713 103.4748 103.6042  
## [41] 104.0845 103.9004 102.7667 102.0846 102.8636 103.0971 101.9225
```

```
hpdipreds <- as.data.frame(t(apply(pred, MARGIN = 2, FUN = HPDI)))  
head(hpdipreds, n=10)
```

```
##      |0.89      0.89|  
## 1  57.18970 145.1188  
## 2  53.71864 146.2659  
## 3  56.92366 147.7539  
## 4  59.12198 151.8405  
## 5  57.94922 146.5091  
## 6  56.53628 149.6273  
## 7  59.64962 151.6313  
## 8  58.17247 152.4303  
## 9  60.82817 154.3709  
## 10 59.99766 152.0494
```

# Crime rates and normal distribution

## Posterior predictive check



## Levels of uncertainty

```
rate ~ dnorm( mu , sigma ) ,  
mu ~ dnorm( 100 , 100 ) ,  
sigma ~ dunif( 0 , 50 )
```

```
##           mean      sd      5.5%      94.5%  
## mu      102.79877 4.165765 96.14107 109.45646  
## sigma   28.58386 2.948141 23.87216  33.29556
```

# Levels of uncertainty

```
est <- extract.samples( meanModel )  
pred <- sim( meanModel)  
  
head(est)
```

```
##           mu    sigma  
## 1 103.88316 21.73971  
## 2  95.28742 31.64544  
## 3  92.49712 31.12295  
## 4 100.35946 26.08824  
## 5  97.17260 31.52742  
## 6 107.50812 29.07011
```

```
str(pred)
```

```
## num [1:1000, 1:47] 122.3 159.6 126.4 77.5 69.8 ...
```



# Levels of uncertainty

Levels of uncertainty

